ESSAYS ON SOCIAL INTERACTIONS, COMPETITION AND MARKETS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Qi Wu

May 2021

ESSAYS ON SOCIAL INTERACTIONS, COMPETITION AND MARKETS

Qi Wu, Ph.D.

Cornell University 2021

This dissertation consists of three essays in the areas of Industrial Organization and Applied Microeconomics, examining the role of social interactions on market outcomes and the welfare consequences.

The first essay studies the role of social influence on consumer demand and firm competition through pricing strategies. Social influence is an important driver of consumption behavior, but its effect on firm competition and pricing is understudied. 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.

In the second essay, my co-authors and I examine the role of social referrals and information exchange in urban labor markets. We use the universe of de-identified and geocoded cellphone records for over a million individuals from a major Chinese telecommunication provider. We find that information flows, as measured by call volume, correlates strongly with worker flows, a pattern that persists at different levels of geographic aggregation. Conditional on information flow, socioeconomic diversity of the social contacts, especially that associated with the working population, helps to predict the worker flows. We supplement the phone records with administrative data on firm attributes and auxiliary data on job postings and residential housing prices. Referred jobs are associated with higher monetary gains, a higher likelihood to transition from part-time to full-time, reduced commuting time, and a higher probability of entering desirable jobs.

The third essay studies the effects of parental retirement on adult children's labor supply through intergenerational time and monetary transfer. My coauthor and I exploit the mandatory retirement age in China as the cut-off point and apply a regression discontinuity (RD) approach to four waves of the China Family Panel Studies (CFPS) Dataset. Our findings suggest that parental retirement reduces adult children's annual hours of labor supply by 3 to 4 percent. This reduction is especially pronounced for female children. We find that the reduction can be explained by parents' increasing demand for time and care from children due to the significant drop in parents' self-rated health upon retirement. Although both male and female children increased their monetary and time transfers to parents, we find that parents tend to make more transfers to sons compared to daughters. Daughters are also more likely to make transfers to parents after they retire, both in terms of money and in terms of time.

BIOGRAPHICAL SKETCH

Qi Wu is currently in her sixth year of study in the Department of Economics at Cornell University. Her areas of specialization include empirical industrial organization, social network economics, and applied economics. She is interested in understanding the role of social interactions on market outcomes and the welfare consequences.

Before coming to Cornell, she received a bachelor's degree in Economics from Renmin University of China and a master's degree in Applied Economics from the University of Michigan, Ann Arbor.

This document is dedicated to all Cornell graduate students.

ACKNOWLEDGEMENTS

First and foremost, I am extremely grateful to my thesis advisors, Panle Jia Barwick, Eleonora Patacchini, Benjamin Leyden, and Giulia Brancaccio for their invaluable advice, continuous support, and patience during my PhD study. I owe my development throughout graduate school, both professionally and personally, to them, and I aspire to continue my career as an economist after them.

I would like to thank Nahim Bin Zahur, Jean-Francois Houde, Evan Riehl, Seth Sanders, Doug Miller, Michael Lovenheim, Zhuan Pei, Michele Belot, Matteo Benetton, Michael Zheng Song, Qinshu Xue, Penny Sanders, Zihan Hu, Jorgen Harris for their comments and suggestions on my dissertation. I am grateful to Miriam Larson-Koester, my graduate mentor in the Graduate Student Association for Economics Mentorship Program, who provided me valuable support and guidance at the early stage of my graduate PhD life. I would like to thank Anne Burton, Xiaolu Wang, Fikri Pitsuwan, Amanda Eng, Katherine Wen, Sam Dodini, Sunwoo Lee, Abhishek Ananth, George Orlov, Stephenson Strobel for their comments on my writing of the paper draft. I also thank my friends Yu She, Yimeng Tang, Xin Gao, Jee-Hun Choi, Esteban Méndez at the economics department for their help and company. Financial support from the Small Grant in Labor Economics, C.V. Starr Fellowship, and Sage Fellowship are gratefully acknowledged.

Finally, I would like to express my gratitude to my parents, Xiqing Wu and Lili Zhang, and my boyfriend Hongtu Zhao. Without their tremendous understanding and encouragement in the past few years, it would be impossible for me to complete my study and thrive.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

SOCIAL INFLUENCE IN PRODUCT CHOICE AND MARKET COMPETITION: EVIDENCE FROM A MOBILE COMMUNICATION NETWORK

1.1 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

[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.

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 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

[3]The change rate is consistent with a national marketing survey conducted by Penguin Intelligence in September 2017 as described in Section 1.2

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 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

[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.

multiplier effect and 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 1.2, I give background on the

industry and describe the data and sample. In Section 1.3, 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 1.4 outlines the demand model, the two-period pricing model and Section 1.5 describes the estimation method and results. In Section 1.6, I compute demand, prices, firm profits and consumer surplus in the counterfactual scenario. Section 3.6 provides a few robustness checks. Section 3.7 concludes.

1.2 Industry Background and Data

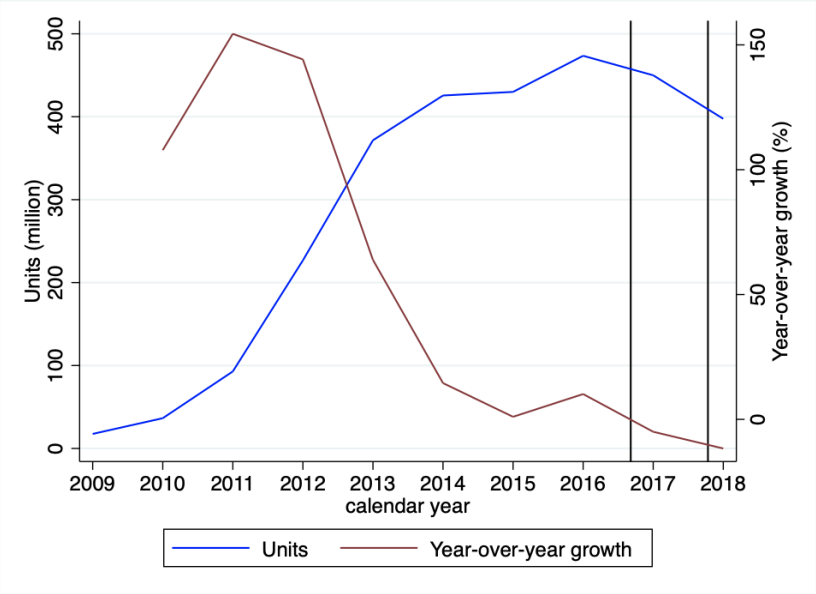
1.2.1 Overview

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 1.1.[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

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

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.

Figure 1.1: 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.

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

1.2.2 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

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

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 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

[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

[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.

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.

1.2.3 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 A.9.

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 and the new device are held for at least one month, respectively. I identify

[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.

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 A.2 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 A.10 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 1.1a 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

[13] I use one-way contact as the baseline definition of a friend. An alternative definition of reciprocal communications delivers robust results in Section 3.6.

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 1.1b 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 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 A.4 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

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

Table 1.1: 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.

by brand and consumer phone changing behavior by operating systems to see the sample representativeness. Table 1.2 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

Table 1.2: 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.

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 A.2.3, ending up with 62 models. Table 1.3a 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 1.3b 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.

Table 1.3: Summary Statistics: Product Attributes

(a) Phones: Product Attributes

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

(b) New Buyers: Old Phone vs New Phone

	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 clock 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 1.3a 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 A.2.3. Composite model "other" in each market is also included. The table 1.3b compares the attributes of the old phones and new phones among all new buyers.

1.3 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:

$$y_{ijt} = \beta s_{m(i),j,t-3} + \boldsymbol{Z_i X_j \gamma}_1 + \boldsymbol{Z_i \gamma}_2 + \boldsymbol{Z_{m(i),j} \gamma}_3 + \lambda_{R(i)f(j)} + \eta_{jt} + \varepsilon_{ijt} \tag{1.1}$$

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 in Section 1.2.3, $m(i)$ 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

[15] The data is organized at individual by alternative level as in Table A.1.

[16] Robustness checks are available when using $t-1$, $t-2$ etc. as the end period in Section 3.6.

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. $\boldsymbol{Z}_{\boldsymbol{m(i),j}}$ 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.

1.3.1 Addressing Sorting on Correlated Tastes

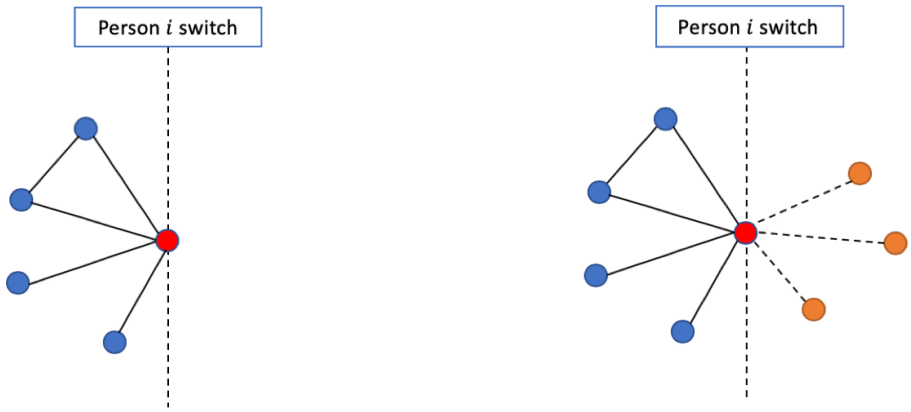
A long-standing identification challenge with observational data is to differentiate social influence from correlated tastes, which render individuals and their contacts' to form a friendship as well as to conduct similar behavior. For example, one chooses a phone because his friends are using that particular 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 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 1.2. 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.

Figure 1.2: Falsification Test Illustration



(a) Current Friends - Peer Group

(b) Future Friends

Notes: Figure 1.2 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.

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 1.1 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 formal presentation, in model 1.2, I test the difference between β_1 and β_2.

$$y_{ijt} = \beta_1 s_{m(i)j,t-3} + \beta_2 s_{m'(i)j,t-3} + \boldsymbol{Z_i X_j \gamma_1} + \boldsymbol{Z_i \gamma_2} + \boldsymbol{Z_{m(i),j} \gamma_3} + \lambda_{R(i)f(j)} + \eta_{jt} + \varepsilon_{ijt} \quad (1.2)$$

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 A.5a and A.5b 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). 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.

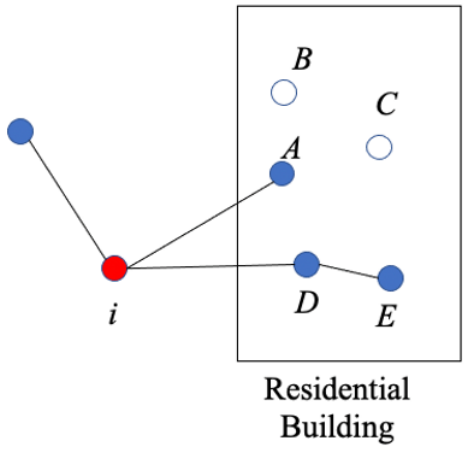
The second control variable is the share of same-old-brand non-contacts, the share of non-contact consumers who replace from the same old brand as the new

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 1.3 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 1.1 can be defined as $s_{NB(m(i)),j}$, the share of i's contacts' residential neighbors using j:

Figure 1.3: Instrumental Variables Illustration



Notes: Figure 1.3 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.[17] Research shows that phone's antenna performance is 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 1.1 also controls for the

[17] 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 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.

time-invariant neighborhood-specific common preferences.

1.3.2 Results

Now I present the results for the baseline model with gradual controls. Table 1.4 reports the results for the linear probability models for smartphone model choice (see equation 1.1). In column 1, I only include the residential neighborhood fixed effects to control for spatial and income related factors. In column 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 columns 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 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 reports 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%).

Table 1.4: Effects of Social Influence on 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. Neighborhood 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 1.3 Model 1.1. Residential neighborhood-by-brand fixed effects are included in all columns. Product-by-month fixed effects are included in Columns 3-6. Column 6 is the preferred specification. Standard errors are clustered by neighborhood-model pair and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.5 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.[18] Although the two regressors have similar means and standard deviations as reported in Appendix Table A.2, columns 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 columns 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.[19] 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.

In Table 1.6, I report the 2SLS results in comparison to the OLS results. Columns 1 and 2 show the comparison with only residential fixed effects, while columns 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

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

[19]The result of the falsification test remain similar when controlling for current and future friend characteristics.

Table 1.5: 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**	0.01	0.01
	(0.004)	(0.01)	(0.003)	(0.01)
Observations	4,218,170	2,082,518	4,218,170	2,082,518
R-squared	0.065	0.072	0.098	0.105
Resid. Neighborhood x brand FE	Yes	Yes	Yes	Yes
Controls	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 1.3 Model 1.1. Residential neighborhood-by-brand fixed effects and product-by-month fixed effects are included in all columns. Standard errors are clustered by neighborhood-model pair and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

and 4G compatibility) as the instrumental variables. The F-tests for the significance of the instruments reported at the bottom of Table 1.6 suggest that the instrumental variables are strong and statistically significant. Column 1 reports the estimates using specification of Table 1.4 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 1.4 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 about 0.10 across several different specifications.

Table 1.6: Effects of Social Influence on Product Choice: IV Results

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*** (0.01)	0.13*** (0.02)	0.11*** (0.01)		0.11*** (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 1.4 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 1.3 Model 1.1. Residential neighborhood-by-brand fixed effects are included in all columns. Product-by-month fixed effects are controlled in Columns 3-5. Standard errors are clustered by neighborhood-model pair and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

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.

1.3.3 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 1.7 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 1.24. 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 1.7 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.

Table 1.7: Social Influence By Peers' Income Levels

Dep. var. Prob i chooses phone j at time t	(1)	(2)
Share high-income friend	0.05*** (0.01)	
Share middle-income friend	0.05*** (0.01)	
Share low-income friend	0.03*** (0.01)	
Share friend of higher income		0.06*** (0.01)
Share friend of similar or lower income		0.02** (0.01)
Observations	4,002,782	4,002,782
R-squared	0.096	0.098
Residence Neighborhood x brand FE	Yes	Yes
Controls	Yes	Yes
Product x month	Yes	Yes

Notes: The table compares the social influence of friends in different income levels. One unit of observation is an individual-model pair. In Column 1, independent variable "Share high-income friend" is the share of friends in high income group whose house price per square meter exceed the 75th percentile of the house price distribution. Analogously, "Share middle-income friend" and "Share low-income friend" are the share of friends whose house price between the 25th and 75th percentile and below the 25th percentile respectively. The cutoff values i.e. the 25th and 75th percentile are 818 USD (5300 RMB) and 2935.72 USD (19000 RMB). In Column 2, "Higher" refers to friends whose house prices per square meters are at least one standard deviation 309 USD (2000 RMB) higher than new buyer i's house price, otherwise belongs to "Similar or Lower". Own house price are included in Column 2. Standard errors are clustered by neighborhood-model pair and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

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 1.8 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 1.8, 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

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

Table 1.8: Social Influence By Visual and Hidden Phone Attributes

Dep. var.	Visual Attributes			Hidden Attributes	
Prob i chooses phone j at time t	(1)	(2)	(3)	(4)	(5)
Share Friend	0.06***	0.09***	0.07***	0.09***	0.12***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Share Friend x Bigger Screen	0.06***				
	(0.01)				
Share Friend x More color option		0.04*			
		(0.02)			
Share Friend x Three cameras			0.05***		
			(0.02)		
Share Friend x High CPU Speed				0.02	
				(0.03)	
Share Friend x Better Screen Resolution					-0.02**
					(0.01)
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 phone age 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, the base level of the interaction term, 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 are controlled. Brand-by-vintage-by-month fixed effects are controlled. Standard errors are clustered by neighborhood-brand pair and reported in parentheses and clustered. *** p<0.01, ** p<0.05, * p<0.1.

at the intersection of friends and coworkers. Newly joined coworkers are possible sources 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 1.9 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

Table 1.9: Social Influence By Relationship

Dep. var. Prob i chooses phone j at time t	(1)	(2)
Share Friend and Existing Coworker	0.04***	
	(0.01)	
Share Friend and New Coworker	0.01	
	(0.02)	
Share Friend of Longer Relationship		0.08***
		(0.03)
Share Friend of Shorter Relationship		0.05
		(0.03)
Observations	273,358	4,218,170
R-squared	0.099	0.096
Resid. Neighborhood x brand FE	Yes	Yes
Controls	Yes	Yes
Product x 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 and Existing Coworker" is the share of friends who are existing coworkers using phone j three months prior to time t. "Share Friend and New Coworker" is the share of friends who are new coworkers moving in and 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 clustered by neighborhood-model pair and reported in parentheses and. *** p<0.01, ** p<0.05, * p<0.1.

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 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

[21]Others includes Unix, BlackberryOS etc.

statistically significant and the main coefficient to decrease. Table 1.10 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 1.4 column 6. Table 1.10 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 10 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.

Table 1.10: Social Influence and Same Operating System Effects

Dep. var. Prob i chooses phone j at time t	(1)	(2)	(3)
Share Friend	0.10***	0.09***	0.09***
	(0.01)	(0.01)	(0.01)
Share Same OS as j	0.004**		0.001
	(0.002)		(0.001)
Share Same Brand as j		0.02***	0.02***
		(0.01)	(0.004)
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 clustered by neighborhood-model pair and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

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.

1.4 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.

1.4.1 Demand

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 as suggested in Section 1.3.3.

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 product attributes, share of peers beforehand and individual demo-

[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.

graphics:

$$u_{ijt} = \bar{u}(p_{jt}, X_{jt}, \xi_{jt}, s_{m(i),j,t-3}, D_i) + \varepsilon_{ijt} \tag{1.3}$$

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

$$\bar{U}_{ijt} = \alpha_i p_{jt} + \sum_{k=1}^{K} X_{jk}\beta_{ik} + \theta s_{m(i),j,t-3} + \zeta_{f(j)} + \eta_t + \xi_{jt} + \varepsilon_{ijt} \tag{1.4}$$

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

$$\bar{U}_{ijt} = \alpha_i p_{jt} + \sum_{k=1}^{K} X_{jk}\beta_{ik} + \bar{\theta} s_{m(i),j,t-3} + \theta_{inc} s_{m(i),j,t-3} \mathbf{1}\{y_i > p75\} + \theta_{use} s_{m(i),j,t-3} Ncalls_i + \zeta_{f(j)} + \eta_t + \xi_{jt} + \varepsilon_{ijt} \quad (1.5)$$

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

$$\alpha_i = \bar{\alpha} + \alpha_1 \mathbf{1}\{y_i > p75\} + \sigma_p \nu_{ip} \quad (1.6)$$

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:

$$\beta_{ik} = \bar{\beta} + \boldsymbol{D}_i \boldsymbol{\beta}_{Dk} + \sigma_k \nu_{ik} \tag{1.7}$$

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 1.3, 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:

$$u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt} \tag{1.8}$$

where

$$\delta_{jt} = \bar{\alpha} p_{jt} + \sum_{k=1}^{K} X_{jk} \bar{\beta}_k + \zeta_{f(j)} + \eta_t + \xi_{jt} \tag{1.9}$$

[23] Interactions with call minutes are rarely significant.

$$\mu_{ijt} = (\alpha_1 \mathbf{1}\{y_i > p75\} + \sigma_p \nu_{ip}) p_{jt} + \sum_{k=1}^{K} X_{jk}(\boldsymbol{D}_i \boldsymbol{\beta}_{Dk} + \sigma_k \nu_{ik})$$
$$+ \bar{\theta} s_{m(i),j,t-3} + \theta_{inc} s_{m(i),j,t-3} \mathbf{1}\{y_i > p75\} + \theta_{use} s_{m(i),j,t-3} Ncalls_i \quad (1.10)$$

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:

$$P_{ijt}(Y_i = j | \boldsymbol{X}, \boldsymbol{p}, s_{m(i),j,t-3}, w_i, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \frac{exp(\delta_{jt} + \mu_{ijt})}{\sum_{j'=1}^{J} exp(\delta_{j't} + \mu_{ij't})} \quad (1.11)$$

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:

$$s_{jt}(\boldsymbol{X}, \boldsymbol{p}, \boldsymbol{s_m}, \theta_1, \theta_2) = \int_{i \in t, A_{jt}} P_{ijt}(\boldsymbol{X}, \boldsymbol{p}, \nu_i, s_{m(i)}, \theta_1, \theta_2) dF(\nu_i, s_{m(i)}) \tag{1.12}$$

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 1.2, 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.

[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.

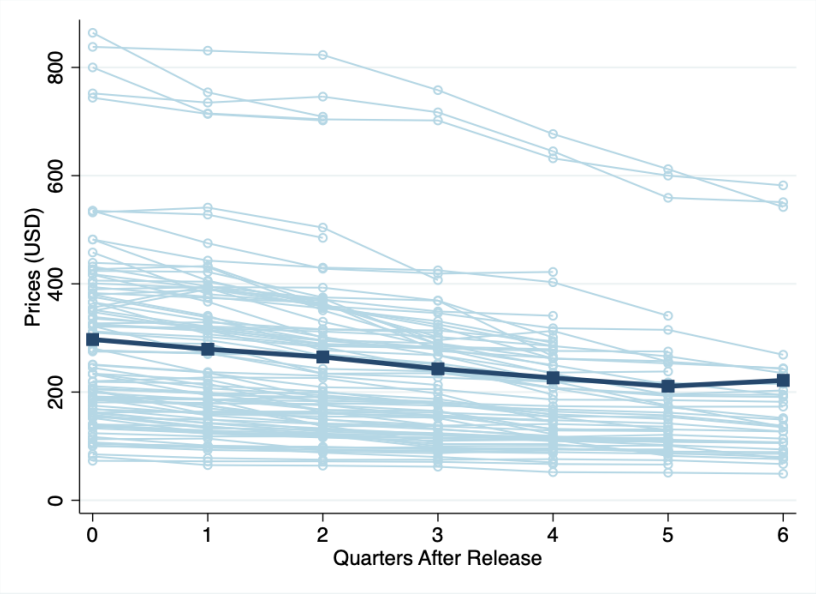
[25] Mobile phone internet user penetration in China 2015-2025, Published by Statista Digital Market Outlook, July 17, 2020, https://www.statista.com/statistics/309015/china-mobile-phone-internet-user-penetration/

1.4.2 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 1.4), the release prices for top-five brands remain stable and slightly increase over time, as suggested in Figure 1.5. Figure 1.4 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 1.5, 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

Figure 1.4: Median Prices Since Release

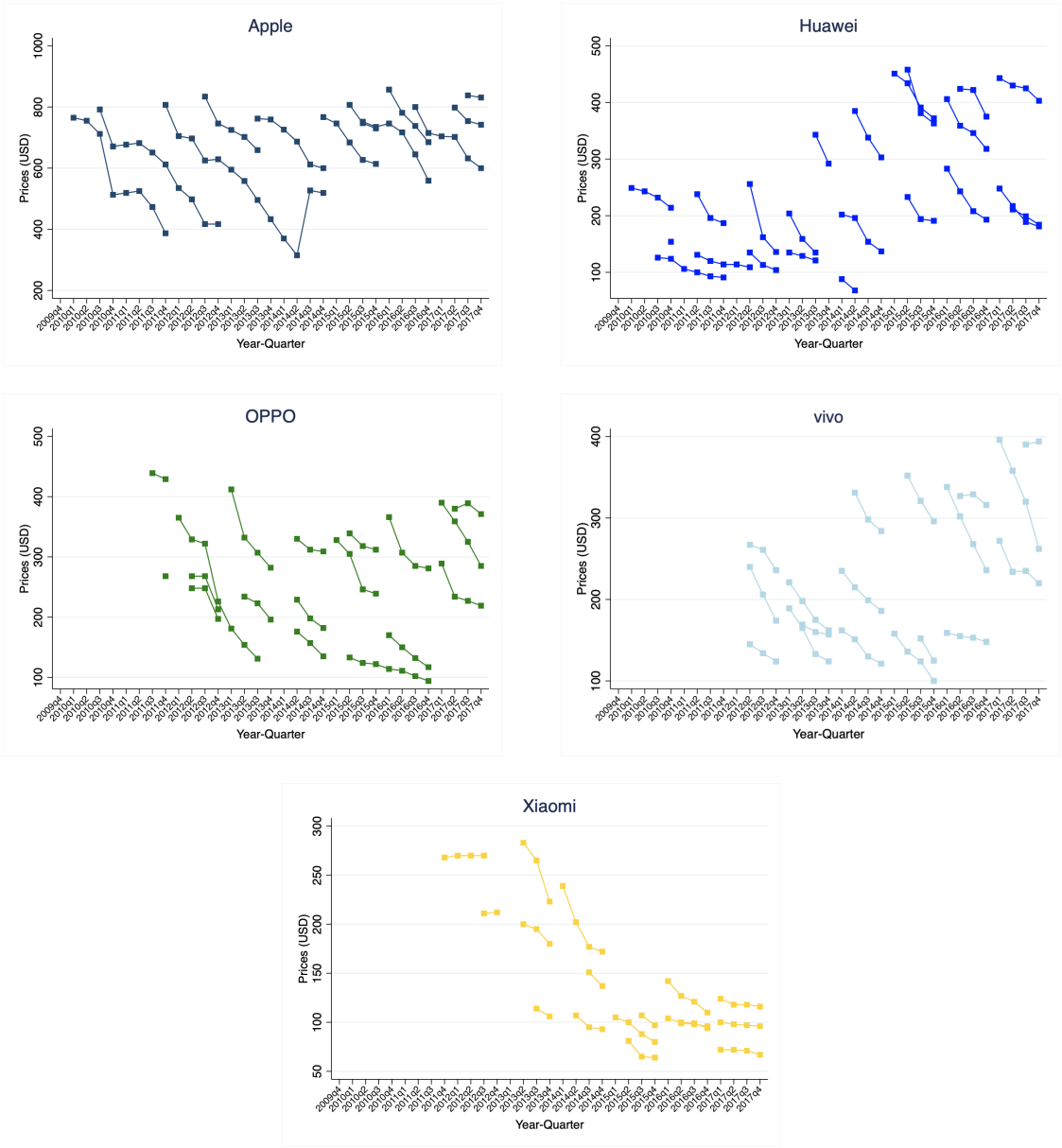


Notes: The figure plots the prices by quarters from release for models released after 2015. The dark blue line represents the median prices among all these models. Each light blue indicates a model. Data Source: IDC Quarterly Mobile Phone Tracker.

profits. In the 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.

[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).

Figure 1.5: 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.

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:

$$W_f = \sum_{j \in J_f} (p_{j1} - mc_{j1}) Q_{j1} + \delta (p_{j2} - mc_{j2}) Q_{j2} \tag{1.13}$$

where

$$Q_{j1} = \sum_{t \in \{T=1\}} s_{jt} M_t, \qquad Q_{j2} = \sum_{t \in \{T=2\}} s_{jt} M_t \tag{1.14}$$

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 1.12. 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

$$p^*_{j2} = mc_{j2} - \underbrace{\frac{\partial Q_{j2}(\boldsymbol{p}, \boldsymbol{X}, \boldsymbol{s_{m1}})}{\partial p_{j2}} \times Q_{j2}}_{\text{Price Markup due to Social Differentiation Effect}} = p^*_{j2} + [\Delta^{-1}_{f2} \times \boldsymbol{Q}_2]_j \tag{1.15}$$

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 1.15 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 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

$$p_{j1}^{*} = mc_{j1} - [\Delta_{f1}^{-1} \times \tilde{\boldsymbol{Q}}_1]_j \tag{1.16}$$

$$\tilde{\boldsymbol{Q}}_1 = \boldsymbol{Q}_1 - \delta \underbrace{\frac{\partial \boldsymbol{Q}_2(\boldsymbol{p}, \boldsymbol{X}, \boldsymbol{s_{m0}})}{\partial \boldsymbol{p}_1}}_{\substack{\text{Inter-temporal} \\ \text{Social Multiplier Effect}}} \times \underbrace{\left\{\Delta_{f2}^{-1} \times \boldsymbol{Q}_2 + Diag(\Delta_{f2}^{-1} \times \boldsymbol{Q}_2)\right\}}_{\substack{\text{Price Markup due to Social Differentiation Effect} \\ \text{in Period 2}}} \tag{1.17}$$

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 1.16, 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

$$\frac{\partial Q_{j2}}{\partial p_{r1}} = \int_i \frac{ds_{ij2}}{dp_{r1}} dF(i) = \int_i \theta s_{ij2}(1 - s_{ij2}) \left(\sum_{l \in m(i)} \frac{\partial s_{lj1}}{p_{r1}} \right) dF(i) \tag{1.18}$$

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

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 1.18, 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 1.17 thus Equation 1.16 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 A.4.

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

$$ln(mc_{jT}) = W_{jT}\phi + \omega_{jT} \tag{1.19}$$

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 1.15, 1.16 and 1.19 yields

$$ln \begin{bmatrix} p_{j1} + [\Delta_{f1}^{-1} \times \tilde{\mathbf{Q}}_1]_j \\ p_{j2} + [\Delta_{f2}^{-1} \times \mathbf{Q}_2]_j \end{bmatrix} = \begin{bmatrix} W_{j1} \\ W_{j2} \end{bmatrix} \phi + \begin{bmatrix} \omega_{j1} \\ \omega_{j2} \end{bmatrix} \tag{1.20}$$

which I bring to data for estimation.

1.5 Estimation

1.5.1 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 1.3, 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 the differential income effects at a cruder level than the neighborhood. The mean utility part in Equation 1.9 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.

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

Estimation In the demand model, there are two sets of parameters to be estimated. $\boldsymbol{\theta}_1$ collects parameters in δ_{jt} (Equation 1.9), also called "linear parameters"; $\boldsymbol{\theta}_2$ collects parameters in μ_{ijt} (Equation 1.10), also called "nonlinear parameters". $\boldsymbol{\theta}_1 = \{\bar{\alpha}, \bar{\beta}_1, \bar{\beta}_2, \bar{\beta}_3, \bar{\beta}_4, \bar{\beta}_5, \boldsymbol{\gamma}\}$, $\boldsymbol{\theta}_2 = \{\bar{\theta}, \theta_{inc}, \theta_{use}, \alpha_1, \beta_{age,2}, \beta_{age,3}, \beta_{age,4}, \sigma_p, \sigma_2, \sigma_3, \sigma_4, \sigma_5\}$, where 2 to 5 represents phone screen size, camera resolution, CPU clock speed, weight and battery capacity. $\boldsymbol{\gamma}$ 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 $(\boldsymbol{\theta}_2, \boldsymbol{\delta})$ 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

$$\delta_{jt}^{h+1}(\boldsymbol{\theta}_2) = \delta_{jt}^{h}(\boldsymbol{\theta}_2) + lns_{jt}^N - lns_{jt}(\boldsymbol{\theta}_2, \delta_{jt}) \tag{1.21}$$

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

[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.

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 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 A.3.

On the supply side, as reported in Table A.6, 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 1.15 and 1.16. 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 1.20 when assuming a normal distributed cost shock.

1.5.2 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 1.11 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 A.8 reports alternative model specifications and the main estimates are quite stable. The log-likelihood is highest in the specification in Table 1.11.

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 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

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

Table 1.11: 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.

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 1.12 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 1.13 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 1.12: Model Fit: Share Among New Buyers

	N. models	Data	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 1.13: 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 1.14 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 1.14: 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 1.15 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.

[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.

Table 1.15: Marginal Costs

Y = ln(mc)	Full Dynamics (T =2)	
ln(W)	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).

1.6 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.

1.6.1 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

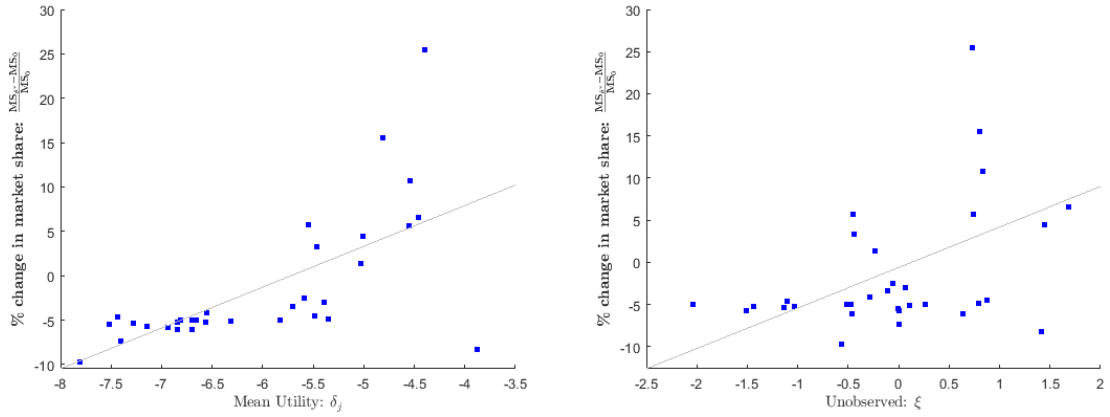
(Amaldoss and Jain, 2005b; Smallwood and Conlisk, 1979). As high-tech products like smartphones are a combination of several key features which essentially determine the product quality, it is important to empirically evaluate the impact of social influence for the demand of quality. To address this 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 1.4, 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 1.6 reports the percent change in demand by quality when social influence is present compared to the counterfactual case when social influence is absent. Figure 1.6a 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 1.6b 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 hetero-

Figure 1.6: Social Influence on Demand By Quality



(a) Quality: Mean utility

(b) Quality: Unobserved 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.

geneous 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 1.16 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 1.7 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.

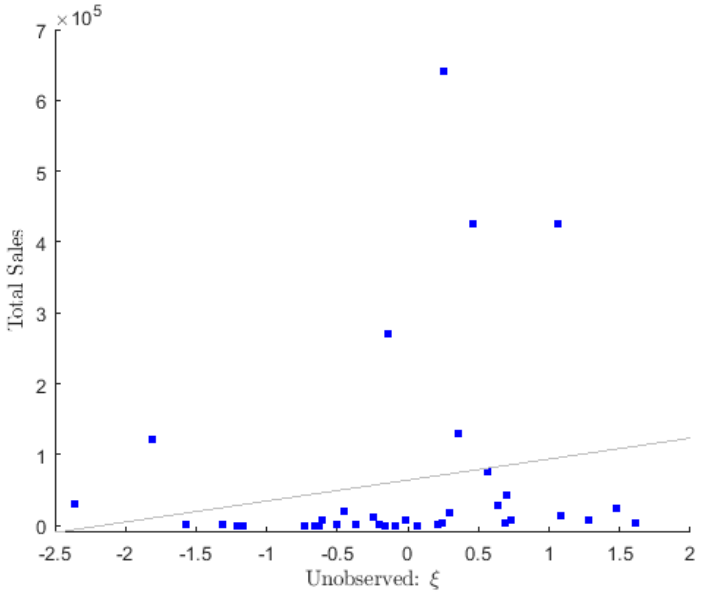
Consumers benefit as social influence increases the average perceived quality level.

Table 1.16: 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 1.7: 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.

1.6.2 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 1.15 and Equation 1.16. 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 A.3.2.

Table 1.17 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

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

row 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.

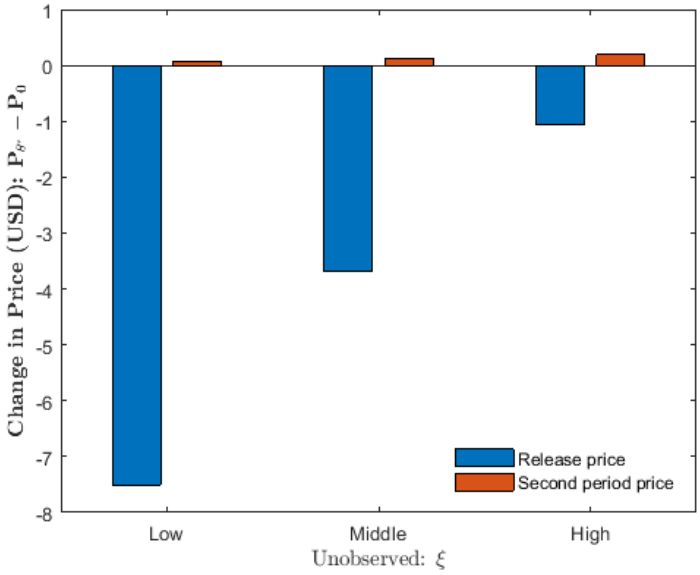
Table 1.17: Counterfactual Prices, Profits and CS Without Social Influence

	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.

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 1.18 reports the heterogeneous price adjustment by the unobserved product quality, visualized in

Figure 1.8: 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.

Figure 1.8. 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).

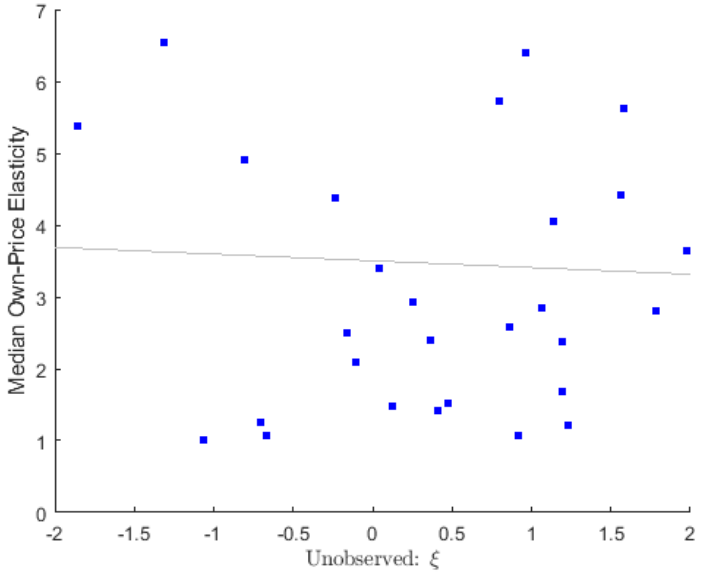
Table 1.18: Heterogeneous Price Changes Due to Social Influence

		ξ		
	Average	Low	Middle	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.

The heterogeneous price adjustments by unobserved quality can be explained by their difference in price elasticity. Figure 1.9 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).

Figure 1.9: 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.

Next I explore heterogeneous effects in profits among products to understand firm's incentive to alter the pricing strategy with social influence. Table 1.19 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.

Table 1.19: Heterogeneous Average Profit Changes Due to Social Influence

Average Profits		ξ		
(Million)	ALL	Low	Middle	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.

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 1.20 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.

Table 1.20: 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.

1.7 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.

1.7.1 Alternative Friend Definition: Reciprocal Contacts

As discussed in Section 1.2, 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 A.11 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 A.12 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 1.21 reports the baseline result using reciprocal contacts. The OLS results are similar to estimates in Table 1.4. 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.

Table 1.21: 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)	
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 1.1 using reciprocal contact definition. Columns 1 and 2 report the OLS estimates specified as Table 1.4 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 clustered by neighborhood-model pair and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

1.7.2 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 1.22 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 1.23 reports the result of using three strategies discussed in Section 1.3 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.

Table 1.22: 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)
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
Resid. Neighborhood x brand 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 uses a Friend dummy as the key regressor and the same specification in Table 1.4. '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. Standard errors are clustered by neighborhood-model pair and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.23: IV Results Robustness: 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)	
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 1.22. Columns 1 and 2 report the OLS estimates specified as columns 5 and 6 Table 1.22. 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 clustered by neighborhood-model pair and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

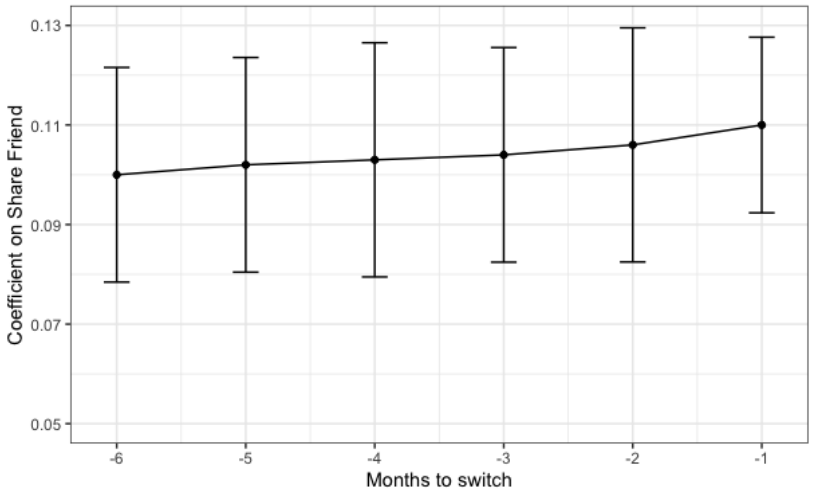
1.7.3 Other Robustness

Alternative Time Lags Figure 1.10 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 1.24 shows the heterogeneous effects of peers using average monthly plan fee as income proxy. It shows similar result as in Table 1.7.

Alternative Demand Specifications Table A.8 reports alternative model specifications. The main estimates on Share Friend, and phone attributes are quite stable.

Figure 1.10: Robustness: Social Influence Using Alternative Time Lags



Notes: The figure plots the coefficient on "Share Friend" using alternative time lags $t-6$, $t-5$, $t-4$, $t-3$, $t-2$ and $t-1$. That is, the share of friends using phone j 6, 5, 4, 3 ,2 and 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 1.4.

Table 1.24: Social Influence By Peers' Income Levels: Robustness

Dep. var. Prob i chooses phone j at time t	(1)	(2)
Share high-fee friend	0.07***	
	(0.01)	
Share middle-fee friend	0.06***	
	(0.01)	
Share low-fee friend	0.05***	
	(0.01)	
Higher Fee		0.08***
		(0.01)
Similar or Lower Fee		0.05***
		(0.01)
Observations	4,128,580	4,128,580
R-squared	0.096	0.098
Resid. Neighborhood x brand FE	Yes	Yes
Controls	Yes	Yes
Product x 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 high/middle/low-fee friend" is the share of friends in above the 75th percentile, between 25th an 75th percentile, and below the 25th percentile of the distribution of monthly plan fee. 2.78 USD (18 RMB) and 21 USD (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 (6.2 USD or 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 clustered by neighborhood-model pair and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

1.8 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.

CHAPTER 2

INFORMATION, MOBILE COMMUNICATION AND SOCIAL REFERRALS

2.1 Introduction

Information affects every aspect of economic decisions, from firm production to household consumption, from government regulation to international treaty negotiations. Classical analysis assumes that agents choose actions to maximize payoff under *perfect* information (Arrow and Debreu, 1954). In reality, information is rarely perfect. Agents' information sets differ substantially, as highlighted by the influential literature on information asymmetry (Akerlof, 1970; Rothschild and Stiglitz, 1976; Spence, 1973). In addition, information exchange and acquisition are costly and crucially depend on social interactions among individuals.

Quantifying the effect of information exchange among social entities and individuals on economic outcomes is challenging because it is difficult to measure the extent of information exchange, and even more so the quality of information that is passed on from one agent to another. The widespread use of location-aware and Global Positioning System (GPS) technologies in mobile phone devices provides a novel avenue that helps researchers to quantify the extent of information flow among individuals, while also tracking their movements in physical space. Datasets derived from these geocoded phone communication records present three unique

Joint with Panle Jia Barwick (Cornell University), Eleonora Patacchini (Cornell University), and Yanyan Liu (International Food Policy Research Institute)

advantages over traditional ones. First, the frequency and intensity of calling records provide a direct measure of information exchange. Second, the panel data nature of these datasets make it feasible to follow individuals over time and space and control for individual unobserved attributes. Third, such data portray a more accurate profile of individuals' social networks than do surveys commonly used in the literature. Existing research has documented that mobile phone usage predicts human mobility (Gonzalez et al., 2008), migration (Blumenstock et al., 2019), poverty and wealth (Blumenstock et al., 2015), credit repayment (Bjorkegren and Grissen, 2018), restaurant choices (Athey et al., 2018), and residential location choices (Buchel et al., 2019).

In this paper, we analyze the impact of information exchange on labor market dynamics. Our empirical research has the following goals. First, we investigate the extent to which information flow is accompanied by worker flows. Second, we examine how information flow among social contacts affects job transitions and the efficiency of worker-vacancy matches.

To this end, we exploit the universe of de-identified and geocoded cellphone records from a major Chinese telecommunication service provider over the course of twelve months in a city. These detailed records enable us to construct measures of information flow between geographic areas and among individuals, as well as variables on employment status, history of work locations, home locations, and demographic attributes. We supplement our phone records with administrative data on firm attributes (industry and payroll) and auxiliary data sets on residential housing prices and job postings for additional socioeconomic measures.

We proceed in several steps. Our analysis begins with documenting that infor-

mation flow as measured by the frequency of phone calls correlates strongly with worker flows. Such a correlation persists at different levels of spatial aggregation. Conditional on the number of phone calls exchanged, the diversity of individuals' social contacts (sources of information) also matters. Within different diversity measures, diversity in socioeconomic status is more valuable than diversity in spatial locations. As far as job mobility is concerned, diversity in the information sources possessed by the working population is far more critical than that by the residential population. Surprisingly, in terms of the relationship between information diversity and economic development, our data exhibit remarkable similarity to the UK data analyzed by Eagle et al. (2010), highlighting the potentially wide applicability of this finding in different settings.

Having illustrated the importance of information flow with respect to worker flow, we examine the role of job-related information shared by social contacts, or friends, on job switches.[1] When an individual moves to a pre-existing friend's workplace, we define such a friend as 'a referral'. We first document that the intensity of information flow between workers and their referrals exhibits an inverted-U shape that peaks at the time of the job switch. In contrast, the information flow between workers and non-referral friends remains stable throughout the sample period, with no noticeable differences during the months that precede job switches. The distinctions in mobile phone calling patterns are not driven by changes in the number of social contacts, which is steady throughout our sample period.

Next, we define the referral effect as the effect of having social contacts in workplaces on individuals' work location choices. We quantify the referral effect using the difference in a job seeker's propensity to switch to 1) a friend's workplace, and

[1]We use *social contacts* and *friends* interchangeably in this paper.

2) a work location in the same neighborhood but without a friend.

One might be concerned that our definition of the referral effect suffers from several confounding factors. First, firms sometimes relocate, consolidate, or open new plants in different areas. If employer relocate employees in different time periods, we might observe workers moving to the neighborhood of pre-existing social contacts. This is unlikely to be important in our study since multi-plant firms are a rare phenomenon in the Chinese manufacturing industry. However, to the extent that it matters, we tackle it by adding the *interaction* of the origin and destination neighborhood fixed effects. In other words, we compare individuals who share the same origin-destination neighborhood pair but have different social networks and examine their choices of workplace locations with and without friends. These origin-destination-neighborhood interactions also control for geospatial attributes that are correlated with job flows (commercial centers, industrial clusters, etc.)

The second confounding factor, a long-standing challenge in the literature that examines observational data, is the difficulty in distinguishing a referral effect from homophily and sorting. If individuals share similar skills and preferences with their friends, then an individual might move to a location where a friend works, not because of the referral information but because the vacant position requests certain skill sets. In addition, not all locations have desirable openings. Leveraging the richness and structure of our data, we conduct a battery of tests. First, we limit our analysis to individuals for whom there is at least one additional location within the same neighborhood that has vacancy listings in the same occupation as the occupation that the job switcher takes. This mitigates the concern that individuals sort into friends' locations as a result of the availability of job opportunities, rather than useful information provided by referrals.

Second, we distinguish between friends who are currently working in the location and friends who used to work there but moved away prior to the job switch. Given that sorting into friends' location by unobserved preferences or skills should happen regardless of a friend's *current* location, we would expect to find similar estimates for both types of friends if our estimated referral effects primarily reflect sorting.

Third, we compare friends who work with friends who live at the location where a job switcher moves to. Larger estimates for friends working in the location would be consistent with referrals: affiliation with the workplace enables friends working there an information advantage of job openings over others. Our results from these different tests illustrate that friends currently work in the location are indeed much more important than friends who moved away prior to the job switch and friends who live but not work there, indicating that our estimated referral effects are unlikely to be driven by sorting.

According to our definition, at least one in every four jobs are based on referrals. Having a referral in a location increases by close to four times the likelihood that an individual moves there – a pattern that is robust across a host of specifications and consistent with previous studies carried out in various countries (Ioannides and Loury, 2004). Referrals are particularly important for young workers, people switching jobs from suburbs to the inner city, and those who change sectors. These results are in line with the observation that information asymmetries are more severe in these settings.

Job information passed on via referrals is valuable for workers. Specifically, referral jobs are associated with higher wages and non-wage benefits, shorter commutes, and a greater likelihood to transition from part-time to full-time and from regular

jobs to premium ones. Information transmitted through the referral networks is also valuable for firms. We find suggestive evidence that firms whose employees have a larger social network are more likely to have successful recruits, achieve higher retention rates, and experience faster growth. Finally, referrals improve labor market efficiency by providing better matches between workers and vacancies, and mitigate labor market inequality, as women and migrants are more likely to find jobs through referrals.

Our work contributes to the emerging literature that demonstrates how the widespread use of electronic technologies, and, consequently the wealth of information on individual digital footprints, opens new frontiers for urban economics (Bailey et al., 2018; Donaldson and Storeygard, 2016; Glaeser et al., 2015). A pioneering study by Henderson et al. (2012) exploits satellite data to conduct an analysis on urban economic activities at a finer level of spatial disaggregation than traditional studies. Using predicted travel time from Google Maps, Akbar et al. (2018) construct city-level vehicular mobility indices for 154 Indian cities and propose new methodologies to improve our understanding of urban development. Other studies examine housing decisions (Bailey et al., 2018b), households' responses to income shocks (Baker, 2018), and entrepreneurship and investment (Jeffers, 2018). Our work contributes to this literature by combining mobile phone records with traditional socioeconomic data to shed light on urban labor market mobility at fine geographical and temporal scales.

Another relevant strand of literature examines the role of social networks in job searches (Schmutte, 2016; Topa, 2011). To identify referred workers, this literature uses surveys or assumes interactions and exchange of job information between social ties, such as fellow workers, family ties, ethnic groups, residential neighbors, and

Facebook friends.[2] The paper closest to ours is Bayer et al. (2008), who also study the importance of referral effects in an urban market. Using Census data on residential and employment locations, they document that individuals who reside in the same city block are more likely to work together than those who live in nearby blocks, and they interpret these findings as evidence of social interactions. Another related paper Gee et al. (2017) uses facebook friends as a measure of social network and documents that strong ties are more important than weak ties in job finding at the margin, but collectively weak ties are more important because they are numerous. We contribute to this literature by providing a more refined measure of social networks and information exchange among individuals, and we introduce complementary data on vacancies and firm attributes to cover a diverse set of economic outcomes.

Finally, our work is related to the empirical literature on information economics. Recent studies have shown that increasing information transparency (for example, through better labels and postings) helps consumers' perceptions of product attributes (e.g., Smith and Johnson 1988), improves consumer choices (e.g., Hastings and Weinstein 2008), and drives up average product quality (e.g., Jin and Leslie 2003; Bai 2018). Our analysis contributes to this strand of literature by quantifying the importance of information exchange through referrals in facilitating urban labor market mobility. Our study is also related to the literature on diversity, including Page (2007) and Eagle et al. (2010). We propose novel measures for the diversity of socioeconomic outcomes and illustrate the important role they play in shaping worker flows.

[2]In the existing literature researchers have proposed several proxies for social networks, such as former fellow workers (Cingano and Rosolia 2012; Giltz 2017; Saygin et al. 2018), family ties (Kramarz and Skans 2014), individuals who belong to the same immigrant community or ethnic group (Edin et al. 2003 ; Munshi and Rosenzweig 2013; Beaman 2012; Dustmann et al. 2016; Aslund et al. 2014), residential neighbors (Bayer et al. 2008; Hellerstein et al. 2011; Hellerstein et al. 2014; Schmutte 2015), and Facebook friends (Gee et al. 2017)

The paper proceeds as follows. Section 2.2 discusses data and the institutional background. Section 2.3 presents motivating evidence that information flow strongly correlates with the flow of workers. Section 2.4 presents the regression framework and reports results from the empirical analysis. Section 2.5 concludes.

2.2 Data and Institutional Background

We have compiled a large number of data sets for our analysis. Besides data on geocoded phone records, we have assembled administrative data on firm attributes and auxiliary data on neighborhood attributes, residential housing prices, and vacancies (job postings).

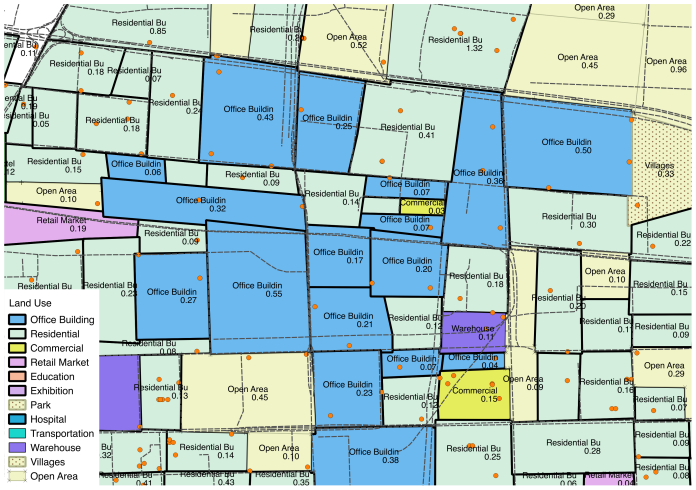
Geographical Units At the highest level, the city we study is divided into twenty-three administrative districts and counties.[3] These districts and counties are further broken into 1,406 neighborhoods that are delineated by major roads. A neighborhood is similar to but is smaller in size than a census block in the U.S. There are 917 neighborhoods in the city proper (i.e., the urban center of the city) and 589 neighborhoods in surrounding suburbs (see Figure 2.1 for a section of the city map).[4] The lowest level of geographical unit is a *location*, a geographic position returned by a cellular tower station, which represents a building complex or an establishment within a neighborhood. The median and average number of distinct locations in a

[3]The city consists of an urban core which is divided into eight districts, and fifteen surrounding suburban and rural counties. These eight districts and fifteen counties are all equally part of the city proper and under its administrative authority.

[4]These neighborhoods are constructed by our data provider for billing purposes. The average sizes for an administrative district/county, a neighborhood in the city proper, and a neighborhood in the suburb are 712 km^2, 0.45 km^2, and 25.03 km^2, respectively.

neighborhood is seven and thirteen, respectively. In total there are close to eighteen thousand locations.

Figure 2.1: Neighborhoods and Locations and in the City



Source: the city is divided into 1,406 neighborhoods that are delineated by major roads (polygons in the map that are separated by dark lines) and 17,881 locations (orange dots). The number in each polygon denotes the area size in km^2.

Spatial attributes come from two GIS shape files (maps). The first shape file delineates administrative divisions, roads, highways, railways, parks, as well as points of interests, such as hospitals, schools, shopping mall, parking lots, and restaurants. The second shape file depicts neighborhood boundaries.

Call Data China's cellphone penetration rate is very high. According to the China Family Panel Studies (CFPS), a nationally representative longitudinal survey of individuals' social and economic status since 2010, 85% of correspondents sixteen years and older report possessing a cellphone.

Our anonymized and geocoded call data contain the universe of phone records for all mobile phone subscribers of a major Chinese telecommunications company in a city that cover the period of November 2016 to October 2017. The data provider

(hereafter Company A) serves between 30-65% of all mobile phone users in the city we study.[5]

Cellphone usage records are automatically collected when individuals send a text message, make a call, or browse the internet. These records include identifiers (IDs), location at the time of usage, and the time and duration of usage. Our data are aggregated to the weekly level and contain encrypted IDs of the calling party and the receiving party, call frequency and duration in seconds, whether or not a user is Company A's subscriber, and demographic information about the subscribers, such as age, gender, and place of birth. The birth county enables us to distinguish migrants from local residents. The existing literature has shown that migrants are much more likely to refer and work with other migrants from their birth city and province (Dai et al., 2018).

An important advantage of our data is the geocoded locations whenever the mobile device is used and every 15 minutes when the device is turned on. The serving cellular tower station records a geographical position in longitude and latitude that is accurate up to a 100-200 meter radius, or roughly the size of a large building complex. For each individual and week, we observe the location that has the most frequent phone usage (calls, texts, internet browsing, etc.) between 9am and 6pm during the weekdays (which we call a 'work location') as well as the location that has the most frequent usage between 10pm and 7am for the same week (which we call a 'residential location').[6] In contrast to traditional data sets in social science studies that typically lack fine-grained geographical information about human interactions,

[5]There are three major telecommunications companies in China. We report a range for the market share to keep company A anonymous. For individuals with multiple phones, we observe usage on the most commonly used phone. If they subscribe to services from multiple carriers (which is uncommon), we only observe activities within company A.

[6]Phone usage during 7am-9am and 6pm-10pm is excluded because people are likely on the move during these time intervals.

these geocoded locations trace out individuals' spatial trajectories over time, and allow us to construct diverse types of social ties (including friends, neighbors, past and present coworkers, friends' coworkers).

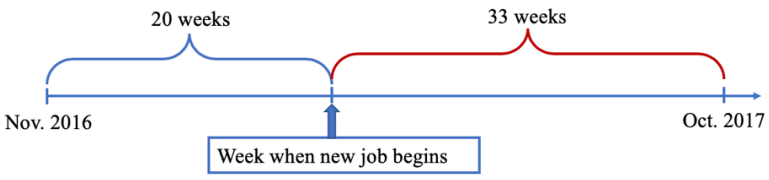
Constructing individuals' workplace history using recorded geocodes is the most crucial step of our analysis. Since we do not directly observe employment status or place of work, we take a very conservative approach in order to mitigate measurement errors in work-related variables. There are 1.6 million individuals in the raw data. We focus on those with valid work locations for at least forty-five weeks – a period long enough to precisely identify workplaces. Locations that are visited during the working hours on a daily basis for weeks in a row are likely to be a workplace rather than shopping centers or recreational facilities. This gives us 560k individuals.[7] After further restricting to individuals who have at most two working locations throughout the sample period (which excludes sales persons and individuals with out-of-town business travels and family visits) and for whom we have complete demographic information, our final sample reduces to 456k users. We carry out the core empirical analysis using this sample and conduct robustness checks in Section 2.4.5 using less stringent sample selection criteria.

We identify individual i as a *job switcher* if the following criteria are satisfied. First, as shown in Figure 2.2, a job switcher is someone who worked in two work locations, is observed at least four weeks in either location, and switches location only once. Second, the distance between these two locations must be at least 1 km. We choose the cutoff of 1 km to avoid erroneously identifying someone as a switcher,

[7]Several factors contribute to sample attrition. China's cellphone market is dynamic, with a high fraction of subscribers switching carriers during each month. In addition, the location information is missing for weeks when individuals travel out-of-town, experience frequent location changes (common for unemployed or part-time workers, salesman, etc.), or have limited phone usage (especially toward the end of a billing period for subscribers on prepaid plans).

because individuals' work locations are geocoded up to a radius of 100-200 meters.[8] Among the 456k users in our final sample, 8% (38,102) are identified as job switchers. Though constructed using different data sources, this on-the-job switching rate is similar to that reported in the literature for China's labor market, which is around 7% (Nie and Sousa-Poza, 2017). China's job-to-job mobility is lower than in Western countries (e.g., 15-18% in the European Union as documented in Recchi 2009), partly because of the Hukou system, which imposes significant restrictions on individuals' migrating across provinces or from rural to urban areas (Ngai et al., 2017; Whalley and Zhang, 2007). Our switchers found jobs in a total of 5,800 work locations that are spread in 1,100 neighborhoods, about two-thirds of which are in the city proper; the reminder are in surrounding counties.

Figure 2.2: Job Switch Timeline



Vacancy Data To gauge the dynamics of local labor market conditions, we collect listings from the two largest online job posting websites, zhilian.com and 58.com, from August 2016 to February 2018.[9] These websites hold on average 10,000 job postings per month. We obtained a total of 121,055 postings and merge them to our call data based on locations.

[8]The average distance between neighborhood centroids is 1.4km.

[9]Zhillian.com reported a 27.5% market share in the fourth quarter of 2017 and became the largest online posting platform in the second quarter of 2018 (https://www.analysys.cn/article/detail/20018775). The website 58.com is a close second, accounting for 26.5% of the market in the fourth quarter of 2017 and serving more than four million firms (http://www.ebrun.com/20161230/208984.shtml).

Each posting reports the posting date, job title and description, full time or part time, qualifications (minimum education and years of experience), monthly salary (in a range), firm address, firm size (number of total employees), and firm industry. On the basis of the job title and description, we group these postings into eight occupations using the 2010 U.S. occupation code. Popular occupations include Professionals (26.70%), Service (26.61%), Sales and Office administration (19.24%), and Management (17.47%), followed by Education, Legal, Arts and Media (11.53%), Farming, Fishing, and Construction (6.44%), Production and Transportation (2.29%), and Health related (1.45%). Industries are classified in ten sectors based on the 2012 US census codes (See Appendix A for more details).

Our vacancy postings report a wide salary range (e.g., an annual salary of RMB 25k-40k). Using the mid-point of the reported salary range delivers a rather flat wage profile across industries: jobs in the construction sectors bring a salary that is similar to jobs in professional services. Missing salaries are also common. In addition, a sizable fraction of worker compensation consists of non-wage benefits, including bonuses and commissions, paid vacations, health and unemployment insurance, etc.(Cai et al., 2011) For these reasons, we rely on the payroll information from the firm administrative data in our empirical analysis.

Administrative Firm-Level Records We utilize two firm-level administrative datasets to obtain wages and benefits, local industry composition, and firm attributes. The first is the annual National Enterprise Income Tax Records from 2010 to 2015, which is collected by the State Administration of Taxation and contains firm ID, industry, ownership, balance sheet information (revenue, payroll, employee size, etc), and tax payments. This database over-samples large companies (major tax

payers) and small to medium-sized firms, covering about 85-90% of the city's GDP. Location information is obtained by merging these tax records with the Business Registration Database that is maintained by China's State Administration for Industry and Commerce. Our final data set contains firm location, industry, ownership type (whether or not state owned), employee size, revenue, wage payroll, and capital for a total of between five to ten thousand firms.[10]

In our sample most firms are private (85.6%), followed by state-owned (7.0%), foreign (0.7%), and other ownership types (6.6%). Over 60% of firms belong to the manufacturing sector, which is higher than the national average of 25.4% (National Bureau of Statistics of China, 2014) and reflects the industrial focus of the city. Using the average payroll as a measure of job compensation, jobs in non-manufacturing firms are paid significantly higher than those in manufacturing firms, demanding nearly a fifty-percent premium (the average annual wage being RMB 32,005 vs. RMB 20,609).

Housing Price Our main data source does not contain individuals' socioeconomic measures such as wealth or income. To overcome this data limitation, we scrape housing data from Anjuke.com, a major online real estate brokerage intermediary and rental service provider in China that collects housing information for both residential and commercial properties. For each residential complex, Anjuke.com reports its name, property type and attributes, the monthly average housing price per square meter, year built, total number of units, average size, and street address. We successfully merged 64% of the neighborhoods in the city proper and 20% of neighborhoods in surrounding counties with residential neighborhoods in

[10]The exact number of firms is omitted to keep the city anonymous.

Anjuke.com.

These data sources provide information on a large number of attributes for each location and neighborhood, including the most common occupations among job postings, industry composition, number of employees and vacancies, average wage, and housing price. For each individual in our final sample, we observe his work and residential location, friends, neighbors, friends' workplaces and home locations, as well attributes for each location.

Chinese Labor Market China's labor market has several noticeable features. Relative to other developing countries, China has a high female labor participation rate. In response to the employment pressure generated by its large population, China has instituted a mandated early retirement age, which is 55 for female workers and 60 for male workers.

Established in the 1950s, China's hukou system categorizes individuals as agricultural or non-agricultural on the basis of their birth place, partly to anchor peasants to the countryside. According to Zhang and Wu (2018), China's urban labor market has a two-tier system: urban cities and rural areas. The large divide that separates these two tiers in terms of job opportunities, social benefits, and amenities (education, health care, etc.) has created a high fraction of migrant workers in urban cities who take jobs with low wages and long working hours, and often are denied social benefits.

State owned enterprises (SOEs) account for a small fraction of the total number of firms, but they constitute more than 30 percent of China's GDP and 20 percent of total employment (State Assets Supervision and Administration Commission 2017).

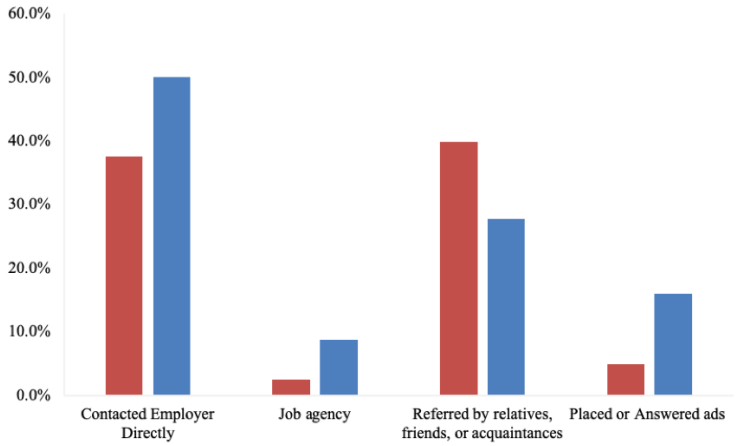
Many SOEs appear in the Fortune Global 500 list and are among the largest conglomerates in the world. Private and foreign companies trail behind SOEs in terms of firm size and revenue. Employment opportunities at SOEs are sought after for their job security, generous benefits, and sometimes higher wages than those in non-state sectors.

Similar to the U.S. and European countries, referrals are common among Chinese workers. Figure 2.3 compares the popularity of different job search methods among Chinese and U.S. workers using data from the 2014 China Family Panel Studies (CFPS) and the 2014 U.S. Current Population Survey, respectively. Workers in China (represented by red bars) are more likely to rely on informal search methods (38% of workers in China find jobs through friends, compared to 30% in the U.S.), while formal search methods, such as ads, job agencies, or contacting employers directly, are more prevalent in the U.S. (blue bars). In addition, referral is more important for young workers in China, with a higher fraction of young correspondents citing referrals as their main channel of landing a job.

Summary Statistics: Demographics and Social Ties Table 2.1a presents descriptive statistics of individuals in our sample. Thirty-six percent of users are female and ninety percent of users are younger than sixty, reflecting the higher mobile-phone penetration among males and the younger population. Three quarters of our sample users were born in the local province; the rest migrated from other provinces. Thirty-nine percent of users were born in the city proper. The last column presents the national average as reported in 2014 CFPS for people who use a cellphone.[11] Our sample exhibits similar demographics as the national average, except it contains

[11]The CFPS sample is restricted to adults with phone-related expenses that exceed 30 RMB per month to ensure proper phone usage, weighted by representative national weights.

Figure 2.3: Job Search Methods in China vs. in the U.S. (2014)



Notes: the horizontal axis reports different job search methods. The vertical axis displays the fraction of each method used among job seekers. Red (blue) bars represent China (U.S.). Source: the 2014 China Family Panel Studies and the 2014 U.S. Current Population Survey.

a smaller fraction of those under age 25, partly because we focus on individuals with stable jobs and exclude students. Our sample also has fewer females than the national average.

The bulk of our analysis focuses on job switchers and their social network. Individual i's social contacts include everyone who makes a phone call to or receives a phone call from individual i at least once during our sample period.[12] As Table 2.1b illustrates, job switchers bear similar demographics as non-switchers, except for age. Job switchers are more likely to be in their thirties and on average are two years younger than non-switchers. They are less likely to be migrants, and have a smaller fraction of friends who use Company A's mobile service, although the magnitude of these differences is modest.

[12]They are also called 'one-way' contacts. An alternative definition requires a contact to make a phone call to *and* receive a phone call from individual i at lease once during the sample period. These two definitions lead to very similar results. Section 2.4.5 conducts robustness analysis on the definition of social contacts.

Table 2.1: Summary Statistics

(a) All users

				CFPS 2014 National	
	Mean	SD	N	Mean	SD
Female	0.36	0.48	435,098	0.45	0.50
Age 25-34	0.29	0.46	455,572	0.23	0.42
Age 35-44	0.26	0.44	455,572	0.24	0.43
Age 45-59	0.27	0.45	455,572	0.27	0.45
Age above 60	0.11	0.32	455,572	0.09	0.29
Age (midpoint)	40.18	11.97	435,194	39.28	14.07
Born in local province	0.75	0.43	455,572	0.76	0.43
Born in local city proper	0.39	0.49	455,572	-	-
Frac of social contacts in Company A	0.50	0.19	455,572	-	-
Job switch	0.08	0.28	455,572	0.07	-

(b) Switchers vs. Non-switchers

	Non-switchers			Switchers				
	Mean	SD	N	Mean	SD	N	Diff.	t-stat
Female	0.36	0.48	398,742	0.36	0.48	36,356	-0.00	-0.45
Age (midpoint)	40.36	12.00	398,817	38.23	11.49	36,377	2.13***	32.49
Born in local province	0.75	0.43	417,470	0.74	0.44	38,102	0.01***	3.62
Born in local city proper	0.39	0.49	417,470	0.38	0.49	38,102	0.00	0.70
Frac of social contacts in A	0.50	0.19	417,470	0.51	0.19	38,102	-0.00	-0.53

Notes: The sample is restricted to individuals with valid work information for at least 45 weeks during the sample period. Number of users = 455,572. 'Age' uses the midpoint of each age range. 'Frac of social contacts in A' is the fraction of individuals' contacts who are company A's subscribers. 'Job switcher' is a dummy for individuals who changed jobs, based on the criteria described in the text. The last column presents the national average as reported in 2014 CFPS among individuals with phone-related expenses that exceed 30 RMB per month, weighted by representative national weights. Job switch rate for the CFPS sample is calculated by Nie and Sousa-Poza (2017). *** p<0.01, ** p<0.05, * p<0.1.

The call data provide rich information on users' social network, but only contain information on work locations for Company A's subscribers. On average, 50% of an individual's friends are Company A's users. One might be concerned about potential sample selection bias if Company A's subscriber network over-represents certain demographic groups. This is unlikely to be a major concern. First, company A's network of users is geographically spread out and covers all street-blocks of the city. Second, pricing and plan offerings are similar across mobile service providers. Nonetheless, to examine the robustness of our results with respect to potential sample selection bias, Section 2.4.2 separates individuals whose friend coverage is

above the median from the rest and documents similar findings.

To help interpret the magnitude of the coefficient estimates in Section 2.4, Appendix Table B.2 tabulates summary statistics for key variables referenced in various regression samples.

2.3 Motivating Evidence: Information Flow and Worker Flow

We are interested in understanding how information exchange affects urban labor markets. We use the number of phone calls between two geographic areas to measure the information flow and relate it to the worker flow, which is the total number of job switches between the same areas in our data sample.[13] To the best of our knowledge, this is the first analysis that examines the empirical relationship between the information flow and worker flow.

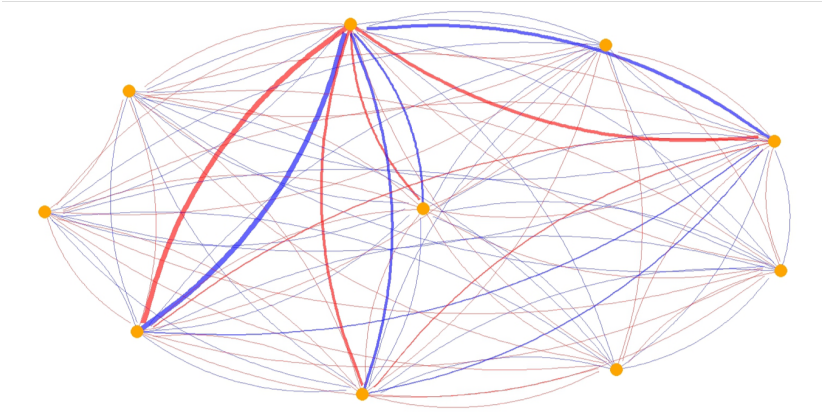
Descriptive Evidence To illustrate the patterns of worker flow and mobile communication, Figure 2.4 plots worker flows against the number of calls between a pair of administrative districts for ten randomly chosen districts within the city proper. A blue non-directional edge corresponds to the number of job switches among a district-pair; its width is scaled proportionately to the number of switches. Red non-directional edges denote the average number of weekly calls, with scaled edge-widths as well.[14] Note the remarkably strong correlation between the two types of edges. City districts with frequent information exchanges (thicker blue lines) also

[13] An alternative measure of information flow, the total call volume in minutes, delivers similar results.

[14] The graph is produced using the Fruchterman & Reingold algorithm which aims to distribute vertices evenly (Fruchterman and Reingold 1991).

have more worker flows (thicker red lines), with the correlation between these two series exceeding 0.94. The two nodes that have the thickest edges are, respectively, the commercial center of the city which has large retail chains, and an urban core with the second highest GDP among all districts in the city. The strong correlation remains when we include districts in suburbs that have less economic activity, with fewer job switchers and lower call volumes.

Figure 2.4: Information Flow and Worker Flow Among Administrative Districts



Notes: Each node is an administrative district in the city. We plot randomly selected ten districts out of a total of 23. Blue (non-directional) edges correspond to the number of job switches among the pairs of nodes, with the width of each edge scaled proportionately to the number of switches. Red edges denote the average number of weekly calls, with a scaled edge-width as well. The graph is produced using the Fruchterman & Reingold algorithm that aims to distribute vertices evenly.

Some correlation arises naturally from heterogeneous spatial and economic attributes, such as the two economic centers in the example described above. To address this, we run a regression analysis and control for origin and destination fixed effects. Regressing the worker flow between a district-pair on the total number of phone calls leads to an economically and statistically significant coefficient: three hundred more calls are associated with one more job switch (Column 1 in Table 2.2a). Using a log-log specification suggests that doubling the number of calls is associated with a 35% increase in worker flows. The R-squared is 0.24 when the number of calls is the only regressor, and jumps to 0.90 when origin and destination

fixed effects are included.

A key premise of our analysis is that call volumes serve as proxies for the amount of information exchanged among individuals. To better measure job-related information that facilitates worker flows, we limit our analysis to calls received or made by job switchers prior to their job change in Column 2. In practice, some calls might be initiated after individuals have decided to move and could reflect communications arising from newly established (work) relationships. In Column 3 and 4, we further exclude calls made within one month (Column 3) and three months (Column 4) before the job switch. When we exclude calls that are unrelated to job-openings, the magnitude strengthens as we move from Column 2 to Column 4, with one additional worker flow following eight more calls.

This strong correlation persists at finer geographical areas. Table 2.2b presents coefficient estimates when we regress worker flow on information flow at the location-pair level. Our data cover eighteen thousand locations and millions of location pairs. Predicting the exact location (a building complex in our example) of job movers is a demanding exercise. Reassuringly, the positive correlation exists even at this fine scale, with one thousand more calls associated with one additional worker flow using the switcher sample (Column 4). We also repeated our analysis at the neighborhood level, where the correlation between information flow and worker flow is 0.75. Neighborhood regressions deliver very similar results, indicate that job-related information flow plays an important role in worker flows, regardless of the level of spatial aggregation.

Out-of-sample Prediction Existing studies have shown that mobile phone usage can predict economic activities (Kreindler and Miyauchi (2019)). In a similar spirit,

we uncover the relationship between the information exchange and worker flow using the first 6 months of the sample, predict the worker flow for the second six months, and compare our prediction with data. We illustrate our results using neighborhood level observations, though results are similar using other geographical areas. We estimate a linear, a linear spline, and a cubic-spline regression model, respectively, with and without neighborhood fixed effects.[15] Based on the estimated model specifications, we predict worker flow in the second 6-month sample, and then regress the observed worker flow on model predictions.

As shown in Table 2.3 (where even columns control for neighborhood fixed effects), the out-of-sample prediction exercise does well. In all cases we have examined, the regression coefficient between our prediction and the observed outcome is close to one, varying between 0.97 to 1.03 depending on specifications. The correlation between the predicted and actual worker flows is 0.55-0.56 across specifications. The R-squared varies from 0.30 to 0.32, which is high for cross-sectional studies with a large sample. These encouraging results suggest that information flow is indeed an important predictor of worker flow.

Diversity and Economic Outcome The results above provide evidence of a strong parallel movement between information flow and worker flow. Both the sociology and economics literature have long emphasized the importance of diversity (Alesina et al., 2016; Ashraf and Galor, 2011; Ottaviano and Peri, 2006). In our setting, the content and value of information might vary over time and across individuals. Economic opportunities are diverse and more likely to come from contacts outside a tightly knit local friendship group. A high volume of information exchange that

[15]We use the default number of spline knots in STATA.

is limited to the same area or social group might not be as beneficial as information from a more diverse setting that taps into different social entities.

Following Eagle et al. (2010), we define three diversity measures using the normalized Shannon entropy: social entropy, spacial entropy, and income entropy.[16] Social entropy measures the diversity of individual i's social ties and is defined as:

$$
\begin{aligned}
D^{\text{social}}(i) &= -\frac{\sum_j P_{ij} * log(P_{ij})}{log(\text{NumFriend}_i)} \\
&= -\frac{\sum_j \frac{v_{ij}}{V_i} log(\frac{v_{ij}}{V_i})}{log(\text{NumFriend}_i)}
\end{aligned}
$$

where P_{ij} is the probability of communication between individuals i and j. It is measured by the ratio of v_{ij}, the number of calls between i and j, and V_i, the total number of calls placed or received by i. The denominator, log of the number of i's friends, is a scaling number that normalizes the Shannon entropy. Normalized entropy measures are guaranteed to vary between zero and one and are comparable across different measures, with higher values representing more diverse outcomes.

Spatial entropy measures the diversity of an individual's social ties in geographic locations:

$$
\begin{aligned}
D^{\text{spatial}}(i) &= -\frac{\sum_l P_{il} * log(P_{il})}{log(\text{NumLocation}_i)} \\
&= -\frac{\sum_l \frac{v_{il}}{V_i} log(\frac{v_{il}}{V_i})}{log(\text{NumLocation}_i)}
\end{aligned}
$$

where P_{il} is the probability of communication between individual i and location l, v_{il} is number of calls between i and location l, and V_i is defined as above. The

[16] Cover and Thomas (2006) is a classic textbook on information theory and entropy measures.

denominator $log(\text{NumLocation}_i)$ is the log number of locations where i has social contacts.

Finally, we define income entropy as:

$$D^{\text{income}}(i) = -\frac{\sum_d P_{id} * log(P_{id})}{log(\text{NumDecile}_i)} = -\frac{\sum_d \frac{v_{id}}{V_i} log(\frac{v_{id}}{V_i})}{log(\text{NumDecile}_i)}$$

where v_{id} is the number of calls between i and all individuals whose housing price falls in the dth decile of the overall housing price distribution. The variable V_i is defined as above. As in the other entropy measures, the normalization is through the number of unique deciles that are spanned by the housing prices of individual i's friends. Income entropy measures socioeconomic diversity among i's social network.

These entropy measures reflect the complexity of an individual's network in terms of socioeconomic status and spatial coverage. We average the diversity measures over all individuals who reside or work in each location. A high value indicates that the working or residential population at a particular location communicates with diverse sources of information. To examine the importance of diversity, we regress the log of worker flow into this location on the average entropy measures at the location level. Our controls include the total call volumes, which, as shown above, is an important predictor of worker flows; the number of individuals (subscribers of our data provider) observed in a location, which captures the scale effect (more populated areas naturally have a higher job inflow); and neighborhood fixed effects. Our key parameters are estimated from within-neighborhood across-location variation. The standard errors are clustered by neighborhood.

Columns 1 to 3 of Table 2.4 include one entropy measure at a time, while Column 4 stacks all three measures together.[17] Both social and income entropy, which reflect the socioeconomic diversity of individuals' information sources, have a sizable and significant impact on job inflow *conditional on* the total number of calls, with a stronger correlation between the income entropy and worker flows. A one standard deviation increase in social and income entropy is associated with a 3% and 10% increase in the worker inflow, respectively. Spatial entropy, on the other hand, is insignificant with a negative sign. This might be because our sample consists of individuals from the same city with limited spatial diversity.

Next we examine the relative importance of information possessed by the *working* vs. *residential* population at each location. Table 2.5 repeats the regressions in Table 2.4 but includes the entropy measures for both the working and residential populations, also with clustered standard errors. As shown in Appendix Table B.1, the entropy measures for these two populations have similar distributions. However, the information diversity of the working population has a much stronger correlation with worker inflows than that of residents in the same location. Conditional on the entropy measure of the working population, the coefficient associated with the residential population's entropy is insignificant and much smaller. While our analysis is descriptive, these results highlight the heterogeneous values of information possessed by different social groups and reflect the fact that information about jobs exists predominantly in the domain of the working population.

It is worth noting that our results are remarkably similar to the findings in Eagle et al. (2010), who examines phone calls in UK in 2005 and relate communication flows to the socioeconomic well-being of communities. While the average number

[17]Here we limit observations to locations that have at least five workers and five residents. Results are similar if we use all locations or limit to those with at least ten observed workers and ten residents.

of monthly contacts is higher in our context (24 vs. 10.1 in the UK, which reflects a denser social network in China), the average minimum number of direct or indirect edges that connect two individuals is very similar (10.4 in our context versus 9.4 in the UK). Moreover, as in our setting, there is a strong correlation (varying from 0.58 to 0.73) between information diversity and the socioeconomic development of communities in the UK. These results reflect common features of the role of information diversity that are at play across different socioeconomic contexts and not limited to specific regions or time periods.

Having illustrated the high correlation between the information exchange and job flows, we turn to the bulk of our empirical analysis that focuses on a specific channel of information at work: information on job openings shared among social contacts. The existing literature has documented that 30 to 60 percent of all jobs are typically found through informal contacts rather than formal search methods (Burks et al., 2015; Topa, 2001). This appears a universal pattern that holds across countries and over time, regardless of the occupation or industry. We next use our calling data to analyze the spatial trajectories of individuals and their social network, and quantify the importance of referral effects in the labor market.

Table 2.2: Information Flow and Worker Flows

(a) At the Administrative District Level

Dependent variable: Worker flow (l, k)	All calls	Calls from/to job switchers ***before*** switch		
		No exclusion	Excluding calls within 1 month of job switch	Excluding calls within 3 months of job switch
	(1)	(2)	(3)	(4)
Information flow (l, k)	0.003***	0.09***	0.10***	0.13***
	(0.0004)	(0.004)	(0.005)	(0.006)
Obs.	253	253	253	253
R-squared	0.90	0.97	0.97	0.97
District l + District k fixed effects	Yes	Yes	Yes	Yes

(b) At the Location Level

Dependent variable: Worker flow (l, k)	All calls	Calls from/to job switchers ***before*** switch		
		No exclusion	Excluding calls within 1 month of job switch	Excluding calls within 3 months of job switch
	(1)	(2)	(3)	(4)
Information flow (l, k)	5.30e-05***	0.0006***	0.0006***	0.0007***
	(1.34e-05)	(7.73e-05)	(1.33e-04)	(1.64e-04)
Observations	159,856,140	159,856,140	159,856,140	159,856,140
R-squared	0.04	0.07	0.03	0.02
Location l + Location k fixed effects	Yes	Yes	Yes	Yes

Notes: In Panel (a), one unit of observation is a pair of administrative districts. In Panel (b), one unit of observation is a pair of locations. There are 23 administrative districts and 17,881 locations in the city. The dependent variable, "Worker flow (l, k)", is the total number of workers moving between area l and area k. "Information flow (l, k)" is the total number of calls between area l and k among all individuals in Column 1, and the total number of calls between switchers and their pre-existing contacts in Columns 2 to 4. Standard errors are reported in parentheses. They are two-way clustered by District l and District k in Panel (a), and by Location l and Location k in Panel (b). *** p<0.01, ** p<0.05, * p<0.1.

Table 2.3: Out-of-Sample Prediction for Worker Flows at the Neighborhood Level

	Dependent variable: actual worker flow between (l, k)					
	(1)	(2)	(3)	(4)	(5)	(6)
Predicted worker flow (linear regression)	1.02***	1.02***				
	(0.18)	(0.002)				
Predicted worker flow (linear spline)			0.97***	0.97***		
			(0.16)	(0.001)		
Predicted worker flow (cubic spline)					1.02***	1.03***
					(0.17)	(0.002)
Constant	0.01***		0.01***		0.004**	
	(0.001)		(0.001)		(0.002)	
Observations	987,713	987,713	987,713	987,713	987,713	987,713
R-squared	0.30	0.31	0.31	0.32	0.31	0.31
Num. Knots			5	5	6	6
Neighborhood l + Neighborhood k FE	No	Yes	No	Yes	No	Yes

Notes: One unit of observation is pair of neighborhoods. The training data consist of switches in the first six months. The prediction is conducted for switches in the second six months. The table reports OLS regressions of the actual worker flow between neighborhood l and k on the predicted worker flow. All three prediction models include neighborhood l and neighborhood k fixed effects. The key predictor is the number of calls by switchers prior to job changes. The linear spline model uses five knots, and the cubic spline model uses six knots. Both are default options from STATA. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.4: Information Diversity and Worker Flows

Dependent variable: log inflow	(1)	(2)	(3)	(4)
Social entropy	0.82**			0.95**
	(0.36)			(0.41)
Spatial entropy		-0.19		-0.58
		(0.32)		(0.36)
Income entropy			0.81***	0.70***
			(0.24)	(0.23)
Total call volume	0.001***	0.001***	0.001***	0.001***
	(0.00)	(0.00)	(0.00)	(0.00)
Observations	6,161	6,161	6,161	6,161
R-squared	0.64	0.64	0.64	0.64
Neighborhood FE	Yes	Yes	Yes	Yes
Num. of Neighborhood FE	1,183	1,183	1,183	1,183

Notes: One unit of observation is a location with at least five workers and five residents. "Log inflow" is the log of the number of people moving to a given location. Social entropy, spatial entropy, and income entropy are normalized Shannon entropies as defined in the text. Total call volume is the total number of calls in thousand from or to a given location. Number of carrier A users in each location is controlled in all specifications. Standard errors are clustered by neighborhood and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.5: Information Diversity and Worker Flows: Working vs. Residential Population

Dependent variable: log inflow	(1)	(2)	(3)
Working population's			
Social entropy	0.84**		
	(0.37)		
Spatial entropy		-0.11	
		(0.32)	
Income entropy			0.75***
			(0.23)
Residential population's			
Social entropy	-0.10		
	(0.28)		
Spatial entropy		-0.32	
		(0.29)	
Wealth entropy			0.27
			(0.18)
Total call volume	0.001***	0.001***	0.001***
	(0.00)	(0.00)	(0.00)
Observations	6,161	6,161	6,161
R-squared	0.64	0.64	0.64
Neighborhood FE	Yes	Yes	Yes
Num. of Neighborhood FE	1,183	1,183	1,183

Notes: One unit of observation is a location with at least five workers and five residents. "Log inflow" is the log of the number of people moving to a given location. Social entropy, spatial entropy, and income entropy are normalized Shannon entropies as defined in the text and constructed separately for the working vs. residential population. Total call volume is the total number of calls in thousand from or to a given location. Number of carrier A users in each location is controlled in all specifications. Standard errors are clustered by neighborhood and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

2.4 Empirical Analysis: Referral-Based Worker Flow

Throughout this section (except when noted otherwise), we limit an individual's network to the one formed *three months* prior to his job switch.[18] This avoids endogenous links formed surrounding the job switch.

Among the 38,102 job switchers observed in our sample, 4,703 (12%) of them have missing friend locations (Panel A of Table 2.6). Among the switchers with non-missing locations for at least one friend, 25% find a job through a referral. Note that this should be interpreted as a lower bound as we only observe friends' locations if they have forty-five weeks of non-missing work locations.[19] As discussed in Section 2.2, forty-five weeks ensures the accuracy of identified job switches, but it may underreport the fraction of referred job moves. In Panel B, we relax the friend sample to include all social contacts with at least four weeks of non-missing work locations. Among switchers with friend location information, 43% move to a friend. In light of this difference, Sections 2.4.2 to 2.4.4 present estimates with our preferred friend definition, while Section 2.4.5 repeats these analyses using friends for whom there is at least four weeks of work information. our results are robust to this alternative friend definition.

In this section, we first conduct an event study on the time series variation of the information flow between job seekers and their referral as well as non-referral friends in Section 2.4.1. Then we perform a battery of regression analyses to illustrate that our estimated referral effect is not driven by confounding factors, in particular, sorting or homophily, in Section 2.4.2. Finally, we evaluate the benefits of referrals

[18] Excluding social contacts formed within one months prior to the job switch delivers quantitatively similar results.

[19] Neither do we observe relatives, classmates, social contacts via WeChat, etc.. Hence this measure for jobs through referrals is necessarily a lower bound.

Table 2.6: Percentage of Job Switchers Switching to a Friend's Workplace

Panel A: including friends with at least 45 weeks of location information			
	Percent	N. Individuals.	N. dyads
Switching to a friend	0.22	8,518	135,866
Switching to somewhere else	0.65	24,881	265,571
Missing all friends' locations	0.12	4,703	
All job switchers		38,102	
Panel B: including friends with at least 4 weeks of location information			
	Percent	N. Individuals.	N. dyads
Switching to a friend	0.40	15,374	487,678
Switching to somewhere else	0.54	20,417	487,126
Missing all friends' locations	0.06	2,311	
All job switchers		38,102	

Notes: Job switchers are identified based on the criteria described in the text. Panel A includes all friends with non-missing work locations for at least 45 weeks. Panel B includes all friends with non-missing work locations for at least 4 weeks. "Switching to a friend" takes value one if a switcher moves to a pre-existing friend's workplace. "Missing all friends' locations" reports the number of switchers with no valid location information for any pre-existing friend. "N. dyads" is the number of switcher-friend pairs.

to workers in Section 2.4.3 and benefits to firms in Section 2.4.4.

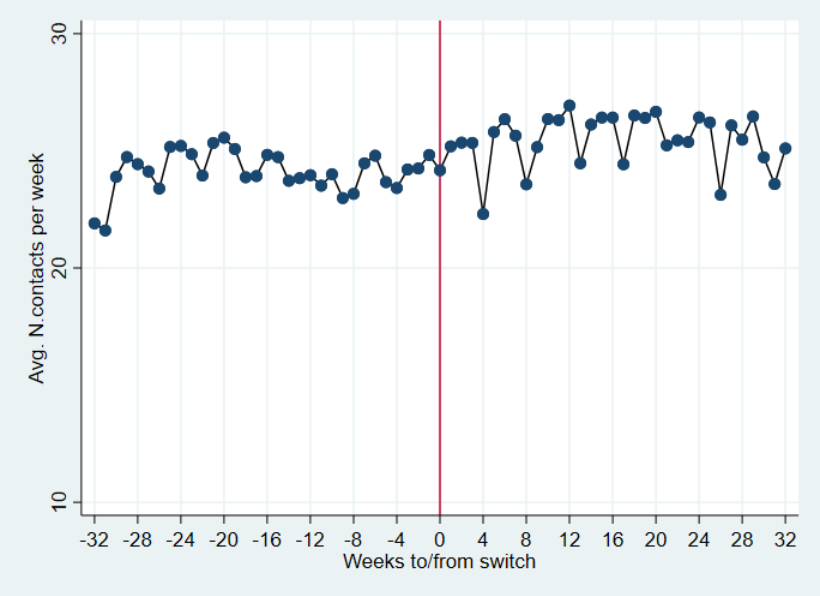
2.4.1 Event Study

The detailed calling records allow us to examine communication patterns between a job seeker i and his referral vs. non-referral friends over time. To the best of our knowledge, this is the first empirical analysis that directly measures information exchange between job seekers and referrals.

We first show that individuals' number of contacts is stable prior to their job change. Figure 2.5 illustrates that there are no spikes in the number of friends communicated with during the weeks leading to the job switch. The average number varies between twenty-three and twenty-five for most weeks, with a modest decrease just prior to the switch. The weekly number of contacts communicated with after the job switch is moderately higher. These patterns suggest that social links established

prior to job switches are likely exogenous; otherwise we should expect a spike in weeks approaching the job switch. This is consistent with the observation that Chinese employers' recruitment decisions are usually made one to two months before the expected start date, due to China's Labor Contract Law that requires a 30-day notice before the termination of an employment contract. Nonetheless, to mitigate concerns of potentially endogenous links, we use social contacts established three months prior to the job switch in our empirical analysis.

Figure 2.5: Number of Social Contacts Per Week: Job Switchers



Notes: The figure plots the average number of social contacts (regardless of carriers) per week who communicated with a switcher. The vertical line indicates the week of job switch.

In contrast to the stable number of contacts, the calling frequency between switchers and their contacts exhibit interesting dynamics. Using an event analysis, we regress the phone call frequency between individual i and his friends during the event window of eleven months before and ten months after the job switch, with a

rich set of fixed effects:[20]

$$\text{Freq}_{ijt} = \alpha \text{Referral}_{ij} + \sum_{s=-11}^{10} \gamma_s \text{Referral}_{ij} \cdot 1[t = s] + \sum_{s=-11}^{10} b_s \text{Non-Referral}_{ij} \cdot 1[t = s] + \lambda_i + \tau_t + \epsilon_{ijt}, \tag{2.1}$$

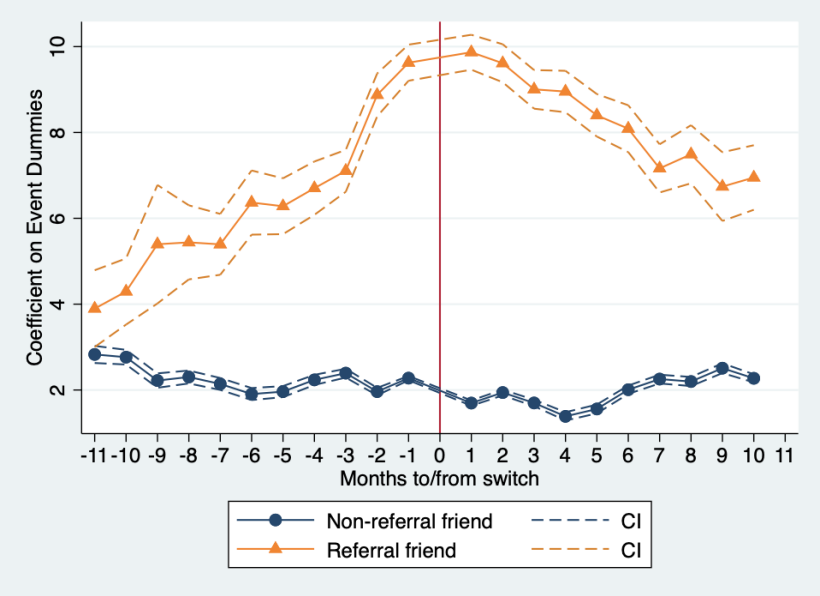
where Freq_{ijt} is the number of calls between caller i and his friend j in month t, Referral_{ij} (Non-Referral_{ij}) takes value one (zero) if switcher i's moves to friend j's workplace during the sample period and zero (one) otherwise, λ_i is an individual fixed effect, and τ_t is a month fixed effect.[21] The key coefficients $\{\gamma_s, b_s\}$ vary by event month s relative to when the job switch occurs ($s = 0$ for the month of job switch). Figure 2.6 plots the regression coefficients and their confidence intervals for referral pairs ($\alpha + \gamma_s$) and non-referral pairs (b_s) separately. Note that the confidence intervals are much tighter for b_s because switcher-non-referral pairs are more common: there are 4.9 million switcher-non-referral-month observations relative to 253k switcher-referral-month observations.

The communication patterns between referral and non-referral pairs are distinct, even after controlling for a rich set of fixed effects. First, switchers have more frequent calls with referral friends than non-referral ones. Second, the intensity of information flow between switchers and their referrals exhibits an inverted-U shape that peaks at the time of job change. In contrast, the information flow between non-referral pairs remains stable throughout the sample period, with no noticeable change in the months prior to the job switch. Lastly, communication intensity between referrals and referees remains elevated post job switch and is noticeably larger than that between non-referral pairs. Information flow increases with the

[20] We use a monthly event window instead of a weekly window to average out noises in time trends. Repeating the event study separately for referral-pairs and non-referral-pairs delivers similar patterns.

[21] Note that Referral_{ij} does not vary with t by construction.

Figure 2.6: Event Study – Number of Calls to Referrals vs. Non-referrals



Notes: The figure plots the coefficient estimates and their standard errors for each event month dummy. Orange line represents calls between switchers and the referrals (Obs = 252,852). Blue line represents calls between switchers and non-referrals (Obs = 4,915,656). Switcher fixed effects and month fixed effects are included in the regression.

dimensions of social interaction, as referrals and referees are friends before the job switch and become friends and colleagues afterwards.[22]

One might be concerned that individuals share with friends news about their job offer, which would also lead to intensified communication before they move to the new work place. However, if this were true, we should expect to observe a spike in the communication volume with *both* referral *and* non-referral friends. The fact that we do not see such an increase with non-referral friends indicates that the communication between workers and referrals is unlikely to be mainly driven by workers' informing friends of their job change. Finally, some phone calls between the referral pairs could be inquiries about workplace amenities (instead of job openings

[22]Calling patterns between job switchers and their social contacts who live but not work in the new workplace are noisy but similar to those between job switchers and non-referral friends without any pronounced hump-shape.

per say). We regard all such calls as information-related communication via referrals that facilitates a job switch.

2.4.2 Referrals and Work Location Choices

We turn to a regression framework to quantify the referral effect that shapes job seekers' location choices. Specifically, we compare the propensity for an individual to find a job at a friend's workplace versus a nearby location, using the following model:

$$M_{il} = \beta \text{Friend}_{il} + \boldsymbol{X_i Z_l \gamma} + \lambda_c + \varepsilon_{il}, \tag{2.2}$$

where M_{il} is one if i moves to location l. $Friend_{il}$ is a dummy variable for having at least one friend working in location l, while λ_c denotes neighborhood fixed effects that control for unobserved location attributes (number of job vacancies, industrial composition, number of locations, etc.). $\boldsymbol{X_i}$ is a vector of demographic controls which consists of gender, migrant status, and age group categories (age 25-34, age 35-44, age 45-59, and above 60). We also include i's total number of social contacts (irrespective of carriers) to capture differences in personality and social outreach. $\boldsymbol{Z_l}$ measures amenities at each work location, including the number of restaurants, the number of roads and parking lots, as well as the number of schools within a 500-meter radius. To allow for differential preference towards local amenities, we interact gender with schools and parking lots within a 500-meter radius, age group dummies with number of restaurants within 500-meter radius, migrant dummy with the number of roads, and number of social contacts with all location attributes.[23]

[23] In unreported robustness analyses, we have controlled for flexible interactions among all demographic attributes and neighborhood characteristics. Results barely change.

Note that we only consider job switchers (people who have found a job). Analyzing how referrals affect the probability of looking for a job (the extensive margin) is interesting but it lies outside the scope of our analysis. In addition, we restrict individual i's choices to locations *within* the neighborhood c that contains his new workplace. This is done purposefully. Job location choices are influenced by many factors, including industry composition and labor demand, commuting distance and local amenities, and intra-household bargaining for married couples, many of which cannot be directly controlled in our framework. Limiting an individual's choices to locations within the neighborhood of his new workplace greatly reduces the extent of heterogeneity across choices and allows us to better isolate the effect of referrals from competing explanations of location choice.

The coefficient of interest is β, which captures the referral effect. There are two main threats to a causal interpretation. First, a positive correlation can arise in a scenario with exogenous worker flows from one area to another. For example, firms sometimes relocate, consolidate, or open new plants at different locations. If employees are relocated in different time periods, the estimated β could capture flows of workers who move to the work location of pre-existing contacts (colleagues). In addition, there are unobserved location attributes that affect worker flows across areas. We tackle this problem by adding the *interaction* of the origin and destination neighborhood fixed effects:

$$M_{il} = \beta \text{Friend}_{il} + \boldsymbol{X_i Z_l \gamma} + \lambda_{\tilde{c},c} + \varepsilon_{il}$$

where $\lambda_{\tilde{c},c}$ is a dummy for the pair of neighborhoods $(\tilde{c},c)$ that contains individual i's previous and current workplace. This is a demanding specification wherein the key

coefficient β is estimated from the within origin-destination variation. We essentially compare individuals who have the same origin-destination neighborhood pair but different friend networks and examine their choice of locations.

The second long-standing challenge in the literature using observational data is the difficulty in distinguishing a referral effect from homophily and sorting. If individuals share similar preferences and skills with their friends, then a positive β could be driven by sorting rather than referrals. In addition, not all locations have desirable openings. An individual might move to location l not because of referrals but because other locations lack appropriate job opportunities. In other words, the friend dummy might simply proxy for locations specializing in jobs that require similar skills shared by individual i and his friends.

Leveraging the richness and structure of our data, we propose the following battery of tests. First, we limit our analysis to workers for whom there is at least one other location within the same neighborhood that has vacancy listings in the same occupation as the one that he takes.[24]. This mitigates the concern that individuals sort into friends' locations that provide the only employment opportunity in the area.

Second, we distinguish between friends who are currently working in location l and friends who used to work there but moved away prior to the job switch. Given that sorting by unobserved preferences or skills should happen regardless of a friend's *current* location, we would expect to find similar β estimates for both types of friends if our finding is driven by sorting. Third, we distinguish between friends who work vs. friends who live at location l. Larger estimates for friends who work in

[24]The occupation of location l is the most common occupation among all postings. It is coded as missing if the most common occupation accounts for less than 33% of all postings at the same location.

Table 2.7: Referral Effects on Job Switches

Dependent variable Probability i switches to location l	(1)	(2)	(3)	(4)	(5)	(6)
Friend	0.36*** (0.003)	0.36*** (0.004)	0.34*** (0.02)	0.34*** (0.02)	0.35*** (0.01)	0.35*** (0.01)
Controls	No	Yes	No	Yes	No	Yes
Observations	1,151,676	1,120,797	1,151,676	1,120,797	1,151,676	1,120,797
R-squared	0.08	0.08	0.14	0.13	0.14	0.14
New work Neighborhood FE	No	No	Yes	Yes	No	No
Old x New Neighborhood FE	No	No	No	No	Yes	Yes
Num. of Neighborhood FE	NA	NA	1,111	1,107	21,250	20,811

Notes: One unit of observation is a switcher-location pair. "Friend" is a dummy variable that equals one if there is at least one friend working at a given location. The controls include gender interacted with schools and parking lots within a 500-meter radius, age group dummies interacted with number of restaurants within 500-meter radius, migrant dummy interacted with the number of roads, and the number of social contacts interacted with all location attributes mentioned above (the number of restaurants, roads and parking lots, and schools within a 500-meter radius). Standard errors are reported in parentheses, and are clustered by neighborhood in Columns 3 and 4 and by neighborhood-pair in Columns 5 and 6. *** p<0.01, ** p<0.05, * p<0.1.

location l would be consistent with the referral effect: affiliation with the workplace enables friends who work there to have an information advantage of jobs openings.

Results Table 2.7 reports the coefficient estimates for model (2.2). Column 1 only controls whether there is a friend in a given work location. Column 2 adds the interaction between demographic variables and location attributes. Columns 3 and 4 repeat the first two columns with neighborhood fixed effects for the new workplace. Columns 5 and 6 further include about 21,000 fixed effects for the pair of old and new work neighborhoods. The standard errors are clustered by neighborhood in Columns 3 and 4 and by neighborhood-pair Columns 5 and 6. We report clustered standard errors by neighborhood or neighborhood pairs in all regression tables, and two-way clustered standard errors when we control for old and new neighborhoods separately. The statistic significance of key parameter estimates is robust to the choice of clusters.

The mean propensity to choose a location within a neighborhood is 0.09. The coefficient for the referral effect, which ranges from 0.34 to 0.36, is economically large, precisely estimated, and stable across all columns in Table 2.7. The probability of moving to location l increases by four times with a friend working there. Adding demographic controls and interaction of origin-destination neighborhood fixed effects has little impact on the key parameter estimate.

In Table 2.8, we conduct a goodness-of-fit exercise similar to that performed in an independent study by Buchel et al. (2019), and we report the percentage of correct predictions (the second to the last row). A correct prediction is one in which the observed location choice has the highest fitted linear probability. Column 1 only includes pair fixed effects. Column 2 adds Friend_{il}. Column 3 further controls for the number of calls between individual i and location l prior to the job change, echoing results documented in Section 2.3. Adding the friend dummy in Column 2 boosts the R-squared by 2.5 times, from 0.07 to 0.14 for a sample of nearly one million observations. Correspondingly, the fraction of correct predictions is 8.9% in Column 1, and jumps to 23.9% in Column 2 with the friend dummy, before further increasing to 30% in Column 3.

One might be concerned about sample selection bias given that information about work location is missing for friends outside Company A's subscriber network. Table 2.9, which splits the sample based on whether the friend coverage is above or below the median (the cutoff is 48%), replicates Columns 2, 4, and 6 in Table 2.7. The difference in the friend coefficient between the two subsamples is modest (smaller than 0.02) and insignificant.

To evaluate whether our finding is driven by sorting, we conduct in Table 2.10 the

Table 2.8: Referral Effects on Job Switches: % of Correct Predictions

Dependent variable Probability i switches to location l	(1)	(2)	(3)
Friend		0.35***	0.30***
		(0.01)	(0.01)
Num. calls to l			0.39***
			(0.06)
Controls	Yes	Yes	Yes
Observations	1,120,797	1,120,797	1,120,797
R-squared	0.07	0.14	0.15
Old x New Neighborhood FE	Yes	Yes	Yes
Num. of Neighborhood FE	20,811	20,811	20,811
Correct predictions at location level	8.9%	23.9%	30.0%
Percent increase w.r.t previous column		170%	25.5%

Notes: This table replicates Column 6 of Table 2.7 using model (2.2). Column 1 excludes the "Friend" dummy and Column 3 adds "Num. calls to l", the number of calls in thousand between switcher i and location l prior to the job switch. A correct prediction is one where the chosen location has the highest fitted linear probability. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.9: The Referral Effect: by Friend Coverage

Dependent variable Probability i switches to location l	(1) Above	(2) Below	(3) Above	(4) Below	(5) Above	(6) Below
Friend	0.35***	0.37***	0.33***	0.35***	0.34***	0.36***
	(0.00)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)
Observations	612,230	508,567	612,230	508,567	612,230	508,567
R-squared	0.09	0.06	0.15	0.12	0.15	0.12
Controls	Yes	Yes	Yes	Yes	Yes	Yes
New work Neighborhood FE	No	No	Yes	Yes	No	No
Old x New Neighborhood FE	No	No	No	No	Yes	Yes
Num. of Neighborhood FE	NA	NA	1,050	1,033	11,889	10,787

Notes: This table replicates Columns 2, 4, and 6 of Table 2.7. Odd columns use job switchers whose fraction of social contacts in carrier A exceeds the median (the median is 48%). Even columns use job switchers whose fraction of social contacts in carrier A is below the median. Standard errors are reported in parentheses, and are clustered by neighborhood in Columns 3 and 4 and by neighborhood-pair in Columns 5 and 6. *** p<0.01, ** p<0.05, * p<0.1.

three tests described above. All columns include the old and new neighborhood-pair fixed effects and demographic controls, with clustered standard errors. Columns 1 and 2 are limited to the subset of switchers who have at least one alternative work location within the same neighborhood that has openings in the same occupation as the one the switchers take. This modestly affects the estimate: the coefficient of

Friend_{ij} changes from 0.35 to 0.34. Columns 3 to 6 use the same sample as that in Columns 1 and 2. Columns 3 and 4 contrast friends currently in the new work location with friends who recently moved away, while Columns 5 and 6 compare friends working vs. friends living there. In both cases, friends currently working in the new location have a much larger impact on the choice probability: they are four times more influential than friends who recently moved away and 150% more effective than friends who live in the same location. The differences in parameter estimates are statistically significant at the 1% level. These results cannot be reconciled with sorting and provide evidence that referrals at work carry useful information that facilitates the matching between workers and job openings.

Table 2.10: The Referral Effect – Falsification Tests

Dependent variable	Individuals with similar job opportunities nearby					
Probability i switches to location l	(1)	(2)	(3)	(4)	(5)	(6)
Friend	0.34***	0.34***	0.34***	0.34***	0.30***	0.30***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.013)	(0.013)
Friend moved before the switch			0.08***	0.08***		
			(0.021)	(0.022)		
Friend living but not working there					0.19***	0.19***
					(0.013)	(0.013)
Controls	No	Yes	No	Yes	No	Yes
Observations	1,134,849	1,104,171	1,134,849	1,104,171	1,134,849	1,104,171
R-squared	0.12	0.12	0.12	0.12	0.13	0.13
Old x New Neighborhood FE	Yes	Yes	Yes	Yes	Yes	Yes
Num. of Neighborhood FE	20,062	19,644	20,062	19,644	20,062	19,644

Notes: This table uses the same specifications as in Table 2.7, except it limits to job switchers for whom there is at least one other location within the same neighborhood that has vacancy listings in the same occupation as the one that he takes. Columns 3 and 4 compare friends currently working in the new workplace with friends who moved away prior to the job switch. Columns 5 and 6 contrast friends working there with friends living there. Standard errors are clustered by neighborhood-pair and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Pathway Referrals facilitate the match between job seekers and vacancies in different ways. For example, current employees can share job opportunities with their

social contacts (information to workers). Alternatively, employees can inform their employer of their friends' work attitude and labor market prospects (information to firms). Although we cannot disentangle these different mechanisms, we test their common implication that referrals mitigate information frictions in the hiring process. We thus examine whether referrals are more important when information asymmetry is more severe.

Individuals who live far away from the new work location, have limited work experience, or change industrial sectors are likely to be disadvantaged when it comes to obtaining information about new openings. Similarly, employers are less likely to be knowledgeable about these workers. In Table 2.11, we interact Friend_{il} with the distance between the old and new work place, the distance between home and the new work location, a dummy for young workers (between 25 and 34), moving from rural to urban locations, and changing sectors.[25] Referrals facilitate job transitions in *all* these situations, especially for rural workers migrating to urban areas and for people changing industry sectors. For these two groups of individuals, the point estimate of the referral effect is 0.68 and 0.53, respectively, which is a significant boost above the base estimate of 0.35. In Column 7 of Table 2.11, we interact Friend_{il} with the demeaned number of calls between individual i and friends in location l prior to the job-switch. The referral effect increases with calling intensity: one hundred calls are associated with a two percentage point increase in the probability of moving to a friend's place, which is consistent with findings in Gee et al. (2017) using Facebook friends.

The evidence in Table 2.11 also sheds light on a couple of alternative explanations. One is that our results are simply driven by preferences: individuals enjoy the

[25]The sample size drops in Column 6 because the dominant sector is undefined for a large number of locations whose postings from the most common sector account for fewer than 33% of all postings.

Table 2.11: Referral Effects and Information Asymmetry

Dependent variable Probability i switches to location l	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Friend	0.35***	0.33***	0.33***	0.33***	0.34***	0.26***	0.34***
	(0.014)	(0.015)	(0.016)	(0.015)	(0.014)	(0.015)	(0.014)
Friend×Distance(job1, job2)		0.002***					
		(0.0003)					
Friend×Distance(home, job2)			0.004***				
			(0.001)				
Friend×Young (Age 25-34)				0.04***			
				(0.009)			
Friend×Rural to urban					0.34***		
					(0.039)		
Friend×Change sector						0.27***	
						(0.022)	
Friend×Call intensity							0.0002***
							(3.22e-05)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,120,797	1,120,797	1,041,950	1,120,797	1,120,797	240,435	1,120,797
R-squared	0.14	0.14	0.15	0.14	0.14	0.15	0.14
Old x New Work Neighborhood FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Num. of Neighborhood FE	20,811	20,811	19,595	20,811	20,811	5,684	20,811

Notes: This table uses the same specification as that in Column 6 of Table 2.7, and interacts "Friend" dummy with various measures on the extent of information asymmetry. "Rural to urban" indicates switchers who move from outside the city proper into the city proper. Six percent switchers move from the rural to the urban part of the city. "Change sector" is one if the switcher changes his sector. "Call intensity" is the demeaned number of calls between switcher i and friends working at location l prior to the job switch. See Appendix Table B.2 for summary statistics of key variables. In Columns 2-6, we also control for the baseline level of the interacted variable. Standard errors are clustered by neighborhood-pair and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

company of friends and hence prefer to work at their place. However, the stronger referral effect when information asymmetry is more severe, together with the communication patterns documented in Section 2.4.1, suggests that preference cannot be the full story. Similarly, could our estimates be mainly driven by nepotism, that is, friends and family are hired instead of the best available candidates (Hoffman, 2017)? It is probably not so in our case: such estimates would not predict that the effect is stronger when information asymmetry is more severe. Moreover, as shown below, referrals are more common among people in the same age range, whereas nepotism often involve individuals from different age groups (children of relatives) (Foley, 2014; Wang, 2013).

Comparison with the Literature How do our results compare to the existing literature that examines job referral effects? There are two common approaches of inferring referrals in observational studies. The first, pioneered by Bayer et al. (2008), defines referrals as residential neighbors. Using data from the Boston metropolitan area, they consider as friends individuals who live in the same Census block. The second approach assumes that social interactions are stronger within an ethnic group, and defines friends as co-workers who are members of the same minority group (Bandiera et al., 2009; Dustmann et al., 2016). We re-estimate model (2.2) using these two definitions of friendship and report the results in Table 2.12. 'Residential neighbor' is a dummy variable that takes value one if individual i has a neighbor who shares the same residential location as i and works in location l. Ethnicity, which is inapplicable in China's context, is replaced with birth county as the literature documents strong social ties among individuals from the same birth region (Zhao, 2003). 'Same birth county' takes value 1 if individual i has a co-worker in location l who was born in the same county. Columns 1 and 2 only include these alternative definitions of friends. Column 3 contrasts neighbors with friends who are not neighbors, while Column 4 compares coworkers who share the same birth county with friends who work in the same location but have different birth counties.

The results shown in Table 2.12 confirm the findings in the literature that neighbors and coworkers from the same birth counties are important. The coefficients on neighbors and the same birth county are 0.21 and 0.10, respectively, when they are the only measure of an individual's social network. Given the average moving probability of 0.09, having a social tie of either type more than doubles the probability of switching to location l. On the other hand, friends dominate both types of social ties by a large margin. The difference in magnitude is both statistically significant

Table 2.12: The Referral Effect – Comparison with the Literature

Dependent variable Probability i switches to location l	(1)	(2)	(3)	(4)
Friend Definition				
Residential Neighbor	0.21***		0.18***	
	(0.01)		(0.01)	
Same Birth County		0.10***		0.09***
		(0.00)		(0.00)
Friend, not Neighbor			0.25***	
			(0.01)	
Friend, not Same Birth County				0.35***
				(0.02)
Controls	Yes	Yes	Yes	Yes
Observations	1,120,797	1,120,797	1,120,797	1,120,797
R-squared	0.16	0.11	0.20	0.16
OldxNew Neighborhood FE	Yes	Yes	Yes	Yes
Num. of Neighborhood FE	20,811	20,811	20,811	20,811

Notes: This table uses the same specification as that in Column 6 of Table 2.7. "Residential Neighbor" is a dummy that equals one if there is at least one individual who works in the new work location and shares the same residential location as the job switcher. "Same Birth County" is a dummy that equals one if there is at least one individual who works in the new work location and shares the same birth county as switcher i. Standard errors are clustered by neighborhood-pair and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

and economically sizable, and in the case of 'same birth county', the effect of friends is three and half times as large. Our results confirm findings in the literature but suggest that they constitute a lower bound of the referral effects.

Attributes of Referrals and Referees To examine the characteristics of workers who find a job through referrals and those of friends who provide referral information, we use a dyadic regression framework wherein the probability that individual i moves to friend j's workplace is a function of both referral and referee attributes:

$$M_{ij} = \boldsymbol{X_i}\boldsymbol{\alpha} + \boldsymbol{X_j}\boldsymbol{\beta} + \boldsymbol{X_{ij}}\boldsymbol{\gamma} + \lambda_c + \varepsilon_{ij} \tag{2.3}$$

where M_{ij} is one if i moves to friend j's workplace, and $\boldsymbol{X_i}$ and $\boldsymbol{X_j}$ include gender, age, and birth county dummies for switcher i and friend j. $\boldsymbol{X_{ij}}$ includes dummies for the same gender, the same birth county, and an absolute difference in age.

We limit the regression sample to the subset of switchers (8,518 individuals) who find a job at some friend's workplace. Then we compare dyad $\{i, j\}$, wherein i moves to j's work location, with dyad $\{i, m\}$, where i does not move to m's work location. Column 1 of Table 2.13 includes all eligible dyads that have non-missing demographic information, for a total of 93k observations. Column 2 only includes switchers for whom there is at least another location within the same neighborhood that has vacancy listings in the same occupation as the one that the mover takes; here there are 88k observations. Females and migrant workers are more likely to receive referrals. There are strong assortative patterns in referral provision. Females on average are less likely to provide referrals but they are more likely to provide referrals to other women. Similarly, workers are more likely to refer other workers who are from the same hometown county. This is consistent with recent findings that community networks based on birth county facilitate entry and the growth of private enterprises in China (Dai et al., 2018). Finally, older workers are more likely to provide referrals, whereas individuals of similar age are more likely to refer jobs to each other, although both effects are modest. Given that females and migrant workers are disadvantaged in urban labor markets (Abramitzky and Boustan, 2017; Blau and Kahn, 2017; Gagnon et al., 2014), these results provide suggestive evidence that referrals improve labor market inequality.

Table 2.13: Attributes of Referrals and Referees via a Dyadic Regression

Dependent variable Probability A switches to B	(1)	(2)
Female A	0.01**	0.01**
	(0.01)	(0.01)
Female B	-0.00	-0.00
	(0.00)	(0.00)
Both female	0.03***	0.03***
	(0.01)	(0.01)
Age A	0.00	0.00
	(0.00)	(0.00)
Age B	0.001***	0.001***
	(0.00)	(0.00)
|Age A - Age B|	-0.001***	-0.001***
	(0.00)	(0.00)
Migrant A	0.01**	0.01*
	(0.01)	(0.01)
Migrant B	-0.00	-0.00
	(0.00)	(0.00)
Both migrants with the same birth county	0.03***	0.03***
	(0.01)	(0.01)
Observations	93,196	88,207
R-squared	0.10	0.09
B work Neighborhood FE	Yes	Yes
Num. of Neighborhood FE	1,176	941

Notes: One unit of observation is a switcher-friend pair. A denotes the referred person and B denotes the referral. The dependent variable mean is 0.14. The sample restricts to switchers who eventually switch to some friend's workplace. Column 2 further restricts to switchers facing at least one vacancy in the same occupation in alternative locations of the same neighborhood. Standard errors are clustered by neighborhood and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

2.4.3 Referral Benefits To Workers

In this section, we examine whether referrals improve referees' labor market outcomes, conditional on finding a job. Our framework for analyzing the benefit of referrals is conceptually similar to model (2.2):

$$Y_{ilr} = \beta \text{Friend}_{ilr} + \boldsymbol{X_i}\boldsymbol{\gamma} + \lambda_c + \alpha_r + \varepsilon_{ilr}, \tag{2.4}$$

where Y_{ilr} denotes the labor market outcome of worker i who lives in residential neighborhood r and switches to work location l in neighborhood c. We control for the same set of demographic variables considered in model (2.2). Because we do not observe individuals' socioeconomic background, such as education and wealth, we include in all regressions the residential neighborhood fixed effect (α_r) as a proxy for one's socioeconomic status (luxurious complexes vs. low-income neighborhoods).

We construct five different measures of job quality. Our first measure is the expected wage at the new job, measured by the average annual payroll (in thousand RMB) among firms in the same location weighted by their number of employees.[26]

Wage dispersion is often driven by across-firm rather than within-firm differences (Card et al., 2018). In addition, an individual's housing value is correlated with his labor income. Thus, we use *coworkers'* housing price as a second measure to proxy for monetary compensation. Specifically, we construct the difference between the average housing price of co-workers at the new workplace and that of co-workers at the previous job. Large positive differences are more likely to be associated with increases in wages and other pecuniary benefits.

The other three measures of job amenities include whether the move is from a part-time job to a full-time job, changes in the commuting distance, and whether the move is from a non-SOE firm to a SOE, because SOEs are sought after for their job security and pension benefits (Zhu, 2013). Although none of these measures of job outcomes is perfect, collectively they speak to both the financial and non-financial aspects of job quality.

[26]We assign each firm in the tax data to the nearest location in our sample and cap the distance at 500 meters. Firms that are farther away are dropped. For 79% of job switchers, the wage information is obtained from a firm within 300 meters. The employment-weighted annual average payroll reflects more accurately the average worker's compensation.

Results Since our labor market outcomes are constructed from different data sources, the number of observations across specifications in Table 2.14 varies from 15,881 to 29,117, and reflects the varying extent of missing observations. Referral jobs pay higher expected wages than non-referral jobs. The point estimate of the wage premium is RMB 620, or about 2% of the average wage reported in our sample.[27] Turning to differences in coworkers' home values in the new versus the old workplace, referral jobs are associated with a 0.5% higher housing price per square meter, where the average housing price in the city is RMB 13,000 ($2,000) per square meters.

Table 2.14: Referral Benefits to Workers

	Income Effect		Job Quality		
	(1)	(2)	(3)	(4)	(5)
Dep var.	Wage at new job	ΔCoworker HP	PT to FT	Closer to Home	Non-SOE to SOE
Friend	0.62**	0.07*	0.014**	0.09***	0.012**
	(0.31)	(0.04)	(0.007)	(0.01)	(0.005)
Observations	17,615	23,323	19,431	29,117	15,881
R-squared	0.79	0.53	0.11	0.12	0.56
Residence Neighborhood FE	Yes	Yes	Yes	Yes	Yes
New Work Neighborhood FE	Yes	Yes	Yes	Yes	Yes

Notes: The sample includes all job switchers. Same demographic controls as in Table 2.7. "Wage at new job" is the average annual payroll per worker in thousand RMB, weighted by employee sizes among firms in the new work location. "ΔCoworker HP" is the difference between coworkers' average house price (thousand RMB) in the new workplace and that in the old workplace. "PT to FT" is a dummy that equals one if the switcher works part-time (less than 30 hours per week) before the switch and full-time (more than 30 hours) after the switch. "Closer to Home" is a dummy that equals one if the commuting distance at the new workplace is shorter than before. "Non-SOE to SOE" is a dummy that equals one if the new workplace is an SOE dominant location (with the majority of employees working in SOE firms), while the previous job is not. See Appendix Table B.2 for summary statistics of key variables. Standard errors are reported in parentheses, and are two-way clustered by residence neighborhood and by new work neighborhood. *** p<0.01, ** p<0.05, * p<0.1.

Having at least one friend at the new workplace helps to increase the probability of moving from part-time to full-time jobs by one percentage point, which is a 2% increase in the likelihood of working full-time.[28] Thirty-one percent of job switches

[27]The annual wage is measured in thousands RMB and the mean is 31.

[28]Hours worked is measured by the duration of phone usage during a workday at the workplace.

involve a shorter commute. Referred jobs are associated with a 30% increase in the likelihood of working closer to home. Finally, having a friend at a SOE firm raises the probability of moving there by one percentage point, which is a 9% increase from the mean (0.11).[29] Higher wages are an indication of worker productivity, and shorter commutes and full-time positions reflect better job amenities. Our results provide evidence that referrals lead to better matches between workers and vacancies, and are consistent with the hypothesis that referrals improve workers' labor market outcomes through mitigating information frictions in the hiring process.

2.4.4 Referral Benefits To Firms

With a few exceptions, most empirical studies of job referrals abstract away from analyzing firm outcomes, because comprehensive data on the performance of both employees and employers are hard to obtain.[30] We merge our calling data with administrative firm-level data based on locations and examine variation across a large number of firms in different industries.

We successfully merge between 5k and 10k firms, 67% of which are manufacturing firms that require production facilities.[31] Our main specification focuses on locations matched to large firms that have more than one hundred employees, which represent about 20% of our sample. While this choice significantly reduces the sample size,

For example, if an individual uses his phone at 10am and then at 4pm in the work location, then the hours worked is six. This is an under-estimate of the actual hours worked. Part-time (full-time) is defined as thirty hours or less (more than thirty hours). On average, 57% of the switchers work full-time before the job change, reflecting the conservativeness of our measure for hours worked.

[29] A workplace is classified as 'SOE' if the majority of workers at that location are employed by SOE firms.

[30] A notable exception is Burks et al. (2015), who use data from nine large firms in three industries (call centers, trucking, and high-tech) to analyze whether firms benefit from referrals .

[31] The exact number of successful merges is withheld to keep the city anonymous.

it mitigates measurement errors because there could be multiple firms in the same location and it is difficult to match workers to firms. The average employment for these large firms is 150; thus, it is likely that these firms occupy an entire location, and consequently, reduces the likelihood of erroneously linking workers to unrelated firms. Appendix Table B.3 reports results from replicating the analysis using all firms. Results are similar both statistically and economically, which is reassuring. In the rest of this section, we use "location" and "firm" interchangeably.

We compare the performance of firms that hire through referrals to firms that hire through other channels via the following model:

$$Y_i = \gamma \text{Referral}_i + \boldsymbol{Z_i}\boldsymbol{\beta} + \lambda_c + \varepsilon_i \tag{2.5}$$

where i denotes a firm. We examine three measures of firm performance, Y_i: (1) inflow of workers, or the number of hires; (2) match rate, measured by the number of hires over vacancies; and (3) firm growth rate, measured by the number of hires over total number of employees. We limit our analysis to locations with at least one hiring, otherwise the estimate of γ will be inflated artificially since the number of hires is at least one for locations with referrals by construction.

Referral_i is a dummy variable that takes value one if at least one worker who switches to firm i has a friend working there, while λ_c denotes neighborhood fixed effects – the same as in model (2.4). $\boldsymbol{Z_i}$ denotes firm attributes and employee characteristics. Firm attributes include firm age, the average number of employees (firm size), dummies for eighteen different industries, large firms, and SOEs, and average real capital from 2010 to 2015. To capture pre-existing trends, we also control for the average employment growth rate from 2010 to 2015. In addition, we include a

firm's referral network size, which is defined as the number of unique social contacts owned by employees who work in firm i prior to the arrival of new hires. Worker attributes include the shares of female workers and migrants, the average age of employees, and average housing price of the pre-existing employees.[32]

The dependent variables are in logs, hence the key coefficient γ is directly semi-elasticities: the percentage change in the outcome when firms hire through referrals. There are two measures of worker inflow: gross inflow and net inflow. Results reported below use net inflows, though they are similar when we use gross inflows.

Results The parameter estimate γ captures the effect of using referrals on firms' performance. To the extent that firms that grow quickly are more likely to hire through employee referrals, our estimate could be biased upward. To tackle this problem, we estimate model (2.5) by increasing the set of variables that help to control for firm growth and employee quality. Nonetheless, because we lack suitable instruments, our results in Table 2.15 are largely descriptive.

The Referral_i coefficient estimates are remarkably similar across different sets of controls for firm and worker attributes, indicating that our results are unlikely to be inflated by unobserved firm or worker characteristics. Firms that hire through referrals are associated with more hires, better matching rates, and a higher growth rate: using referrals increases a firm's net labor inflow by 63%, enhances the job matching rate by 86% (the average matching rate for large firms is 0.76), and raises the firm growth rate by 45% (the median growth rate is 4% for large firms). Results in Appendix Table B.3 that use all firms document similar patterns. Although our analysis in this section is descriptive, the fact that the estimates are robust to a rich

[32]The number of Company A's users at each location is included in all regressions.

Table 2.15: Referral Benefits to Large Firms with Positive Hiring

Dependent variable Panel A: Log of Net Inflow	(1)	(2)	(3)	(4)
Referral	0.60***	0.62***	0.62***	0.63***
	(0.13)	(0.14)	(0.14)	(0.14)
Observations	[600,1000]	[600,1000]	[600,1000]	[600,1000]
R-squared	0.64	0.65	0.65	0.66
Panel B: Matching Rate	(5)	(6)	(7)	(8)
Referral	0.91***	0.88***	0.88***	0.86***
	(0.24)	(0.26)	(0.26)	(0.27)
Observations	[400,1000]	[400,1000]	[400,1000]	[400,1000]
R-squared	0.85	0.87	0.87	0.87
Panel C: Firm Growth Rate	(9)	(10)	(11)	(12)
Referral	0.49***	0.45***	0.44***	0.45***
	(0.11)	(0.10)	(0.10)	(0.11)
Observations	[600,1000]	[600,1000]	[600,1000]	[600,1000]
R-squared	0.76	0.83	0.83	0.83
Controls				
Firm Attributes	No	Yes	Yes	Yes
Previous Growth Rate	No	No	Yes	Yes
Employee Attributes	No	No	No	Yes
Neighborhood FE	Yes	Yes	Yes	Yes

Notes: One unit of observation is a location with at least one matched firm that has more than 100 employees and positive hirings. There are 225 neighborhood fixed effects in Panel A, 190 in Panel B, and 271 in Panel C. "Referral" takes value 1 if there is at least one switcher moving to a friend in the firm. "Net inflow" is the number of switchers moving in minus moving out. "Matching rate" is defined as the inflow over the number of vacancies. "Firm growth rate" is measured as the inflow over the employee size. Firm attributes include age, employee size, SOE dummy, average real capital from 2010 to 2015, and the average annual employee growth rate from 2010 to 2015. Employee attributes includes share of female, share of migrants, and the average age of pre-existing employees. Firm network size, measured by the number of distinct contacts of the firm's pre-existing employees, as well as the number of carrier A's users at each location is controlled in all columns. See Appendix Table B.2 for summary statistics of key variables. Standard errors are reported in parentheses and clustered by neighborhood. *** p<0.01, ** p<0.05, * p<0.1.

set of firm and worker controls raises our confidence that these estimates are not simply picking up unobserved characteristics related to firm and employee quality. Instead, firms are likely to benefit from employee-provided referrals, consistent with the fact that referral-based hiring programs are common (Burks et al., 2015).

2.4.5 Alternative Definition of Friends

We conclude our analysis with a few additional robustness checks. Our core analysis limits to friends who have at least forty-five weeks of non-missing work locations. This mitigates measurement errors in job locations for these friends, but omits a large fraction of friends for whom we observe fewer than forty-five weeks of location information. In this section, we examine the robustness of our results to alternative sample selection criteria.

Appendix Table B.4 replicates Table 2.7, and includes all friends who have at least four weeks of non-missing work locations. This enlarges the number of individual-friend pairs from 401,437 to 979,595. The estimated referral effect remains unchanged: having a friend in a location increases the probability of moving there by 35 percentage points.

Using this alternative definition of friend, referral jobs are associated with a 1.3% increase in wage premium, a 0.6% increase in job-related benefits (as proxied by coworkers' housing prices), and a 12% increase in the likelihood of working full-time (Appendix Table B.5). These effects are similar to those found in our base specification. The effect on the likelihood of a shorter commute and transitioning to a SOE firm is nearly identical to that found in the base specification. Turning to the referral benefit for firms, the alternative definition of friend produces slightly more pronounced results than those reported in the base specification (Appendix Table B.6). We have replicated our analysis with various other friend selection criteria (e.g., using all friends with at least three months or six months of work locations) and obtained very similar results.

Finally, Appendix Table B.7 repeats Table 2.7, but defines individual i's friends

as social contacts with two-way communications: social contacts who both make and receive at least one call from individual i. In addition, all friends with at least four weeks of non-missing work locations are included in the analysis. The referral estimate is again very similar to that of our base specification.

2.5 Conclusion

This paper uses geocoded mobile phone records to study how information provided by social contacts mitigates information asymmetry and improves the labor market performance.

Our study provides three broad lessons for future research. First, panel data with fine spatial and temporal variation hold great potential for overcoming the challenges of causal inference with observational data in the context of social networks. For example, the ability to identify different types of social contacts in small geographical areas at overlapping periods helps us to tackle sorting. Second, big data from non-conventional sources complement traditional datasets on socioeconomic measures. In our analysis, tax records and firm registration data are crucial in our analysis on how referrals benefit firms, a topic that is understudied in the existing literature. Third, information exchange, and in particular, social and socioeconomic diversity in communication appears to facilitate worker movement. In the future, studies on the exact mechanism that governs how information exchange through referrals increases labor market efficiency would be extremely valuable.

CHAPTER 3

THE EFFECTS OF PARENTAL RETIREMENT ON ADULT CHILDREN'S LABOR SUPPLY: EVIDENCE FROM CHINA

3.1 Introduction

Aging and increasing retired population are a global challenge. Virtually every country in the world is experiencing growth in both the size and the proportion of older persons in the population. In 2019, there were 703 million persons aged 65 years or over in the global population. This number is projected to double to 1.5 billion in 2050. Globally, the share of the population aged 65 years or over increased from 6 per cent in 1990 to 9 per cent in 2019. This proportion is projected to rise further to 16 per cent in 2050, when it is expected that one in six people worldwide will be aged 65 years or over (UN World Population Ageing 2019).[1] A concurring problem with aging is retirement. Since a retired person either relies on public programs (such as pension system) or assistance from family members, the rising number of older persons can intensify the pressure on their families, especially in countries where public transfers are relatively low.

Research has found that retirement not only affects the retiree but also the economic behaviors of the spouse, for example income and consumption behavior (Battistin et al., 2009; Charles, 2004), leisure activities (Stancanelli and Soest, 2016), home

Joint with Xin Gao

[1] https://www.un.org/en/development/desa/population/publications/pdf/ageing/WorldPopulationAgeing2019-Report.pdf

production (Stancanelli and Soest, 2012), cognitive abilities (Mazzonna and Peracchi, 2012) and health and health behavior (Behncke, 2012; Coe and Lindeboom, 2008; Coe and Zamarro, 2011; Eibich, 2015; Insler, 2014; Johnston and Lee, 2009). Studies on the intergenerational effects of retirement, in particular the effect of parental retirement on adult children's labor supply, remain scarce. Bertrand et al. (2003) examine how the pension transfer paid to parents affects the labor supply of prime-age individuals living with these elderly in extended families in South Africa. They find a sharp drop in the working hours of prime-age individuals in these households when women turn 60 years old or men turn 65, the ages at which they become eligible for pensions. In addition, the oldest son in the household reduces his working hours more than any other prime-age household member. Recent research finds that in Europe parental retirement increases fertility rate and women's retirement leads to an increase in their daughters' employment in countries with low family benefits, while the opposite is true in high family-benefits countries (Eibich and Siedler, 2020; Fenoll, 2020). In the Chinese context, Chen and Zhang (2018) find that maternal retirement decreases female children's childcare time by eight hours per week. At the same time, the retirement of mothers/in-law significantly increases the employment rate of women with children by 12%. There are, however, two major limitations of the existing literature. First of all, it only looks at the extensive margin of female labor supply, namely, the binary definition of whether the woman is working or not, without looking at the actual hours. It is entirely possible that adult female children's labor participation rate increased but the average hours decreased, which is a net negative effect. Second, the study focuses narrowly on the effect of maternal retirement on female children without examining either the effect of paternal retirement or the heterogeneous effects of parental retirement on female and male children.

Another intriguing question to explore, especially for Asian countries like China where gender role is clearly defined and salient, is whether parental retirement affects male and female children's labor supply differently. Pioneering research by Akerlof and Kranton (2000), for example, suggests that men and women are associated with different behavioral prescriptions, such as "men work in the labor force and women work in the home". Such prescriptions may simultaneously affect hours of market labor supply and the division of tasks within households, such as childcare and eldercare. Budig and England (2001) find that the burden of childcare often falls disproportionally on women. Ettner (1995) uses data from the 1986-1988 panels of the Survey of Income and Program Participation (SIPP) and shows that informal care-giving has a significant negative effect on female market labor supply in the U.S.. In particular, coresidence with a disabled parent has a large and significant negative impact on female labor supply, although most of this effect is due to non-participation in the labor force rather than to a reduction in hours among workers. Ettner argues that the asymmetrical effect of eldercare on male and female can be explained by social norms "making decisions to substitute nonmarket for market labor more difficult for men." The effect of gender identity on female labor supply and home production is particularly relevant in the case of China, where the society is in a transition period in terms of social norms. Although the rate of female labor force participation in China is relatively high, this rate is gradually decreasing: the participation rate for the females was 91% in 1990, 87.6% in 2000, and 83.2% in 2010 (Li et al., 2015). This might indicate that Chinese women are finding it more challenging to balance work and family responsibilities.

This paper studies the effects of parental retirement on adult children's labor supply and explores the mechanism of such effects through the change in intergen-

erational time and monetary transfers. We use four waves of the China Family Panel Studies (CFPS), which is representative of 95 percent of the Chinese population and provides detailed information on birth year and month, hours worked, retirement status and other demographic information. To identify causal relationship, we apply a regression discontinuity (RD) design to examine the effect of reaching the mandatory retirement age (60 years old for men, 55 years old for women in SOE, 50 years old for women in private enterprises). We find a sizable increase in actual retirement rate at the mandatory age cut-off, indicating high compliance rate to the retirement policy.

We find that there is a drop of 77 to 82 hours per year in adult children's labor supply at parents' mandatory retirement age, which is equivalent to a 3 to 4 percent drop in annual average hours. This effect is statistically significant and robust to the inclusion of self-employed workers who enjoy more flexible working schedule as well as alternative model and time window specifications. We also find a significant increase in the probability of adult children transferring time and money to retired parents. Adult children are 3.8 percent and 4.5 percent more likely to transfer money and time respectively to parents after they retire. We propose two possible explanations. First, due to social convention and the lack of formal eldercare programs in China, adult children are the primary caregivers for their parents and need to shoulder the majority of time and monetary burden. Second, there seems to be an increasing demand for care from parents post retirement since we find a significant drop in parents' self-rated health level after retirement. This is consistent with findings in the literature that parents tend to believe that they are less healthy and require more attention from caregivers when transitioning into retirement life (Fitzpatrick and Moore, 2018; Müller and Shaikh, 2018). We also find that the drop

in hours is driven by parents who self-rated as less healthy.

In addition, we find that the decrease in working hours is more salient for daughters than for sons. Controlling for children's own age and year fixed effects, daughters annual hours of labor supply decrease by 123.7 hours from an average of 2,138.67 hours (6 percent), which is sizable and significant at the level of 1 percentage point. In comparison, sons only experienced a statistically insignificant decrease of 49.05 hours. This is consistent with the findings in the gender role literature where daughters are more likely to devote time into care-giving within the household. Likewise, although the probability of parents making time and monetary transfer to children also increased after retirement, we find that daughters are less likely to receive such transfers than sons. Such disparate transfer pattern can be explained by social norm and traditional gender role. Chinese parents favor sons over daughters (Ebenstein, 2010) and sons are considered as the "family name bearer". Therefore, parents are more likely to give money to and do housework for sons. Moreover, the disproportional care provided by daughters can also be explained by gender role. Traditionally men and women are associated with specific behavioral prescriptions as "the bread winners" and "the home makers" respectively. The effect of gender identity on female labor supply and home production is particularly relevant a society is in a transition period in terms of social norms like China. Chinese women are found to be spending twice the time fathers do on childcare, indicating that it has become more challenging for Chinese women to balance work and family responsibilities (Chen and Zhang, 2018). Since it is costly to deviate from social norms or gender identity, such prescriptions reinforce daughters to carry more duty on taking care of and providing help to parents (Budig and England, 2001; Ettner, 1995).

Our paper has three major contributions. First, we add to the aging and retirement

literature by examining the spillover effects of retirement. In particular, we study the effects of parental retirement, both paternal and maternal, on both male and female adult children's labor supply. In addition, we further investigate the underlying mechanism behind such changes by examining intergenerational transfers – money and time – both downward (from parents to children) and upward (from children to parents). By doing so, we are able to not only examine the gender difference in terms of the externality of retirement but also identify the gender difference within each channel of such impacts. We provide new evidence on the negative impacts of the drop in parental self-rated health upon retirement on adult children's labor supply, which has not been examined in the past literature. Second, we add to the gender inequality literature by examining the complex interplay among gender identity, care-giving, and female labor supply in China. Using parental retirement as an exogenous shock, we are able to estimate how male and female adult children response differently to the increased needs for eldercare, how male and female adult children give and receive monetary and time support from parents disproportionately, and how such inequalities affect male and female children's hours of market supply differently. Third, our results call for policy reform that addresses the negative effects of parental retirement on adult children, especially on women. An affordable public elderly care system may help reduce the burden of prime age adults, increase overall labor supply and therefore tax base, and boost economic development in the end. In addition, workplace policies designed to help women employees specifically, for example flexible hours, will also help female adult children adjust their schedules, take care of their parents without negatively affecting their job performances.

This paper proceeds in seven sections. In Section 3.2, we give background on

the aging, mandatory retirement and eldercare in China. In Section 3.3, we describe the data and the sample construction. In Section 3.4, we conduct the analysis accessing the change in adult children's labor supply due to parental retirement and heterogeneous effects by gender. Section 3.5 explores the mechanisms. Section 3.6 provides robustness checks using alternative model specification and alternative time window specification. Section 3.7 concludes.

3.2 Aging, Mandatory Retirement and Eldercare in China

China has the largest population and faces the fastest growing aging population (United Nations, 2019). The old-age dependency ratio, defined as the number of people at retirement age per 100 working people, increases from 10 percent in 2000 to 17 percent in 2020. The share of the population over 60 years of age is now projected to rise from 17.4 percent in 2020 to 30 percent in 2040, while the fertility rate will continue to remain low (United Nations, 2019). This is mainly due to decades of falling birth rates and steeply rising life expectancy. At the same time, China has been implementing the mandatory retirement policy since 1978 and is one of the countries with the earliest retirement age. For workers in private enterprises, men and women are supposed to retire at the age of 60 and 50 respectively. In sectors such as public sectors, state-owned enterprises (SOE) and collectively-owned enterprises (COE), the mandatory retirement for men and women are 60 and 55 respectively.

The combination of aging and mandatory retirement placed huge pressure on Chinese families. In China, families have been the main source of financial support and care-giving for the elderly. Nearly 75 percent older adults have children living

with them or residing nearby who can provide care (Lei et al., 2015).[2] Studies suggest that only 3% of the elderly have a commercial pension and 0.2% a private occupational pension issued by a private employer in 2013 (Zhu and Walker, 2018).

Policy makers in China have made several attempts to tackle the many aging and care-giving related issues, including proposing a "three tiers of social services for the aged" – home-based care as the "basis," community-based services as "backing," and institutional care as "support." A series of national policy initiatives over the last decade attempted to develop community-based services. A notable example was the Starlight Program, under which the government invested a total of 13.4 billion yuan (roughly US $2.1 billion) to build urban community-based senior services centers during 2001–04. However, the centers have apparently not served their intended purpose partly because of dwindling financial support from the government, raising questions about the viability of similar initiatives. To date, self-sustaining, community-based long-term care services remain largely nonexistent, except in a few major urban centers like Shanghai (Wu et al., 2005). In addition, policy initiatives to support home or community-based care have been largely limited to urban areas, and even there, the number of beneficiaries is still relatively small. In much of rural China, institutional elder care was rare and limited to state-run institutions exclusively serving childless elderly adults, orphans, the mentally ill, and developmentally disabled adults without families in 2000s (Feng et al., 2012).

With insufficient formal care programs, adult children still act as the primary caregiver to retired parents. That is, adult children have no choice but to provide significant time and monetary transfers to retired parents.

[2] About 41% of older adults live with an adult child, and another 34% have an adult child living nearby.

3.3 Data

Our data comes from four waves of the China Family Panel Studies (CFPS), the representative household survey for year 2010, 2012, 2014 and 2016, respectively. The CFPS sample covers 25 provinces/municipalities/autonomous regions, representing 95% of the Chinese population. At the individual level (i.e. each household member), the survey collects information on demographic information including birth year and month, gender, and an urban/rural indicator. It also includes information on individual's smoking and drinking behavior in recent months, self-rated health status, retirement status, marital status, education level, employment status, annual hours worked, job type (waged/agricultural), sector (state-owned/collectively-owned/private), and annual income for both parents and adult children. At the household level, the survey collects information on total assets, family size, and number of kids under age 16.

Sample Construction We construct a child-centric sample where each observation is an adult child for year 2010, 2012, 2014 and 2016. First, we match each adult child to his or her first-retired or first-to-retire parent. For example, if person X's mother retired in 2011 and her father retired in 2013, we will pair X with her mother. This is because the effect of the second retirement within the same household tend to be attenuated by the first shock of retirement. To avoid such bias, we only look at the first retirement that occurred to each adult child. For consistency, only the first-to-retire parent is included on the left hand side of the cut-off. For example, if in 2012, person X's mother is three years away from retirement and her father is two years away from retirement, we will pair X with her father. To simplify our language when referring to one's first-retired or first-to-retire parent in later sections, we term

it “parent" for convenience. For adult children who are married, we also include his or her spouse in the sample, pairing the spouse with the same retired parent in the household.

Second, we exclude people who do not report working hours and those who only engage in their own agricultural production. About 72.3 percent of non-agricultural workers report working hours, whereas only 26.7 percent of agricultural workers who do temporary paid job and earn wages report working hours. Third, we focus on the sample of individuals with non-missing working hours and whose parents' ages are within a certain window around the mandatory retirement age. This is because the identification strategy we use – RD design – requires a small window around the cutoff to deliver the local treatment effect of an exogenous shock. Details about the RD design are discussed in section 3.4. However, there is no statistical or econometric consensus on the choice of window size. The rule of thumb is to select a window size narrow enough to ensure the local-ness of the estimates, but not so narrow that the sample size becomes too small. Here, we choose five years below and above the cut-off so as to balance the sample size and the local-ness of the estimates. We use seven years below and above the cut-off as a robustness check in section 3.6 and the main results stay valid.

In terms of intergenerational transfers, we observe both monetary and time transfers on the extensive margin. Namely, we observe the occurrence of time and monetary transfers, but not the frequency or amount of such transfers. In the survey, questions such as “Did you give money/care/financial management to parent in the past six months?” and “Did you give monetary support/housework help/financial management to your child [1/2/3/4/...]?” provide us with information on the occurrence of intergenerational transfers. For the latter question for parents, we

observe each parent's response for each of his or her own biological child, who is assigned a unique child ID. For upward transfers from adult child to parent, we match each adult child's answer to the child-centric sample using his or her person ID. For downward transfers from parent to each adult child, we match parent's answer to the child-centric sample using the child ID nominated by the first retired parent.

We code answers to the transfer questions as binary variables to indicate the probability of parent providing (receiving) care and money to (from) any specific adult child. For example, if person X answers "yes" to"Did you give money to parent in the past six months?", then his or her "Ever transfer to parent – Money" variable would be coded as one, zero otherwise. Similarly, if a first-retired or first-to-retire parent answers "yes" to "Did you give monetary support to your child [3]?" and person X's ID matches that of child [3], then his or her "Ever receive from parent – Money" variable would be coded as one, zero otherwise.

Summary Statistics Table 3.1 describes the key features of our constructed sample. 41 percent of adult children are female and the average age is 28.93 years old. 56.3 percent of the adult children reside in urban areas. On average they have more than 10 years of education. Average net asset holding is about 321,000 RMB, which includes house, car, and financial assets minus debts. 61 percent of the adult children in our sample have kids. The majority of the kids are between 6 and 16 years old.

Parental age is re-coded so that 60 is set as the "reference cut-off" for both mother and father. For example, if a mother is 53 years old and works for a private enterprise (meaning that her mandatory retirement age is 50), we re-code her age as 63 using the formula 53+(60-50)=63. Similarly, if a mother is 67 years old and worked for a SOE (meaning that her mandatory retirement age is 55), we re-code her age as 72

Table 3.1: Summary Statistics: Adult Children

	Mean	SD	N
Panel A			
Female	0.41	0.49	7565
Age	28.93	5.11	7565
Urban Area	0.56	0.50	7245
Married	0.70	0.46	7565
Years of schooling	10.54	3.88	7204
Net asset (thousand RMB)	321.36	1176.42	7115
Have kids	0.61	0.49	7565
N. kids	1.49	0.80	4626
N. kids under age 1	0.10	0.30	4626
N. kids age 1-2	0.32	0.51	4626
N. kids age 3-5	0.45	0.60	4626
N. kids age 6-16	0.62	0.75	4626
Parent Age (recode)	58.21	3.53	7565
Engage in Non-agricultural work	0.98	0.15	6525
Parent Retired(a)	0.30	0.46	5,618
Parent Retired(b)	0.26	0.44	15,498
Panel B			
Hours			
Annual hours worked	2583.57	1203.76	7565
Annual hours worked plus self employed	2415.86	1120.00	7799
Transfers			
Whether give [..] in the past 6 months			
Money to parent	0.04	0.19	4773
Housework to parent	0.03	0.16	4773
Care to parent	0.03	0.16	4773
Financial management to parent	0.003	0.06	4773
Money to support children	0.02	0.13	4773
Housework children	0.05	0.21	4773
Financial management to children	0.01	0.07	4773

Notes: One unit of observation is one adult child. This table reports the characteristics of adult children with non-missing working hours and whose parents' re-coded ages are within the ± 5 years window. Parent Retired (a) is the fraction of adult with non-missing working hours. Parent Retired(b) is the fraction of adults with or without missing working hours.

using the formula 67+(60-55)=72. The average re-coded parental age is 58.2 years old, which is just around the 60-year-old "reference cut-off". About 30 percent of adult children in our sample have at least one retired parent. On average, they work for 2,583.57 hours per year, or around 49 hours per week.

Table 3.1 panel B shows the average probabilities of upward and downward transfers. Interestingly, adult children are more likely to transfer money to parents

(3.9 percent) than providing care or housework (2.5 and 2.6 percent). On the contrary, parents are more likely to do housework for adult children (4.7 percent) than providing financial support (1.8 percent).

3.4 Assessing the Change in Adult Children's Labor Supply Due to Parental Retirement

To estimate the change in adult children's labor supply due to parental retirement, we compare the annual hours of labor supply of adult children whose parents' ages are right above the mandatory retirement age to those right below the cut-off. Parental retirement decision could be affected by unobserved factors that could simultaneously affect the adult children's labor supply, for example, valuation for family time, work ethics, etc. Thus, to reach a causal inference, we use a Regression Discountinuity (RD) design to eliminate the potential endogeneity of parental retirement decisions.

The RD design has been used in previous literature studying the causal effect of reaching retirement age on health insurance coverage, mortality, and spousal health outcome (Card et al., 2008; Fitzpatrick and Moore, 2018; Lee and Lemieux, 2010; Shigeoka, 2014). The RD approach aims to compare the average outcomes just below and just above the cut-off. As discussed in Section 3.3, there is no statistical or econometric consensus on the choice of window range. Here, we choose five years as the window to balance both the sample size and the localness of the estimates.[3]

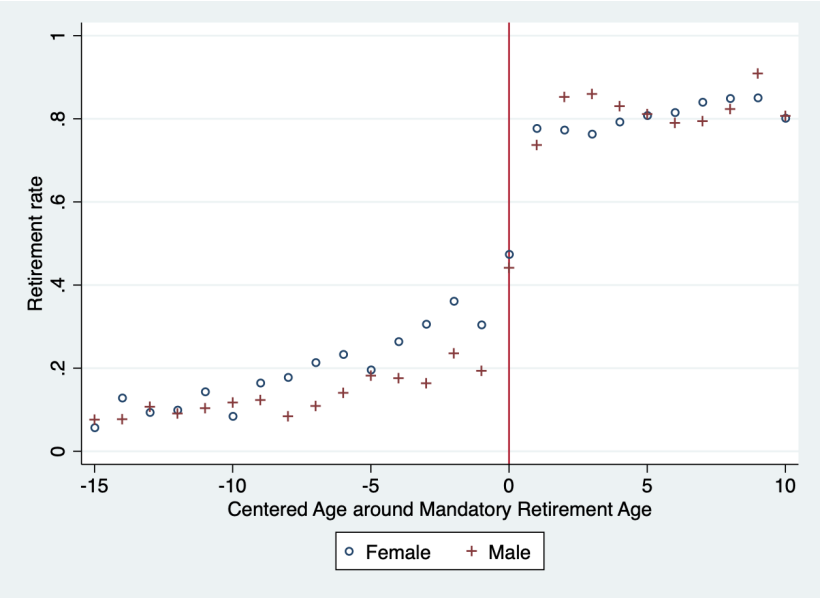
In addition, "age RD design" involves a distinct feature from the standard RD

[3]We also use seven year as robustness checks and the main results stay similar.

design. Since all individuals will eventually pass the retirement age, assignment to treatment is inevitable. Hence, individuals who anticipate the parental retirement may adjust their behavior ahead of time (Lee and Lemieux, 2010). If adult children anticipate potential change in lifestyle, for example a reduction in overall family income, they would increase their labor supply before the parental retirement, which would lead to an upward bias in our estimate of the effects of parental retirement on adult children's labor supply. On the contrary, adult children may predict that their job performance will be negatively affected by parental retirement anyways and start to work fewer hours before the actual retirement, which would lead to a downward bias of our estimate. To test if our results are in fact biased by such anticipation effects, we run a "donut hole" RD as a robustness check in Section 3.6, where we exclude observations within one year above or below the threshold (Mazumder and Miller, 2016; Shigeoka, 2014).

The mandatory age provides an exogenous shock to retirement decisions. Figure 3.1 shows that retirement rate has a sizable jump right around the cut-off. There is a clear increase in the retirement rate for both male and female, suggesting that people are indeed complying to the mandatory retirement policy. Thus, our RD estimates can be interpreted as valid intent-to-treatment effects of parental retirement on adult children's labor supply, as long as other observed factors affecting parental retirement do not change discontinuously right around the cut-off. We test this condition with the validity check in Section 3.4.2.

Figure 3.1: Fraction of Parental Retirement and Parent Age Relative to the Mandatory Cutoff



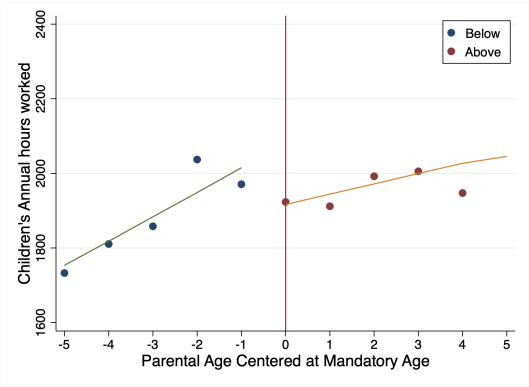
Notes: This figure shows the compliance rate of parents. The x-axis is parental age relative to mandatory retirement age (The mandatory retirement age for is 50 for general female workers and 55 for females who work in public sectors, state-owned enterprises and collectively-owned enterprises; 60 for male workers). The y-axis is the fraction of people whose reported employment status is "retired". The blue circles represent female while the red dots represent male.

3.4.1 Graphical Result

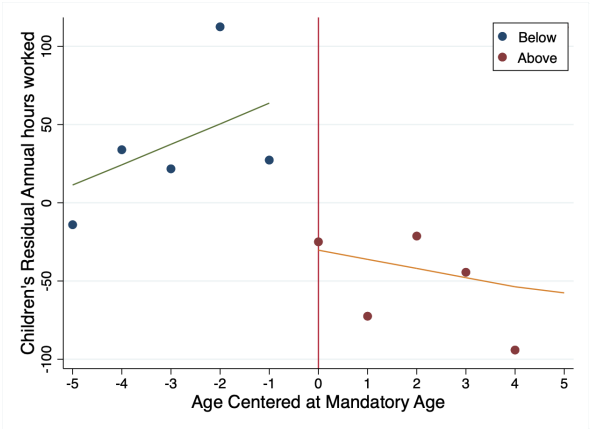
Figure 3.2 shows the scatter plots of adult children's annual working hours overlaid with lines from local linear regressions in a window of ±5 years around the mandatory retirement cut-off. Panel (a) clearly reveals a significant drop in adult children's average annual hours as soon parents reach the mandatory retirement age.

It is possible that people develop different working schedules as they age, and that the reported hours could be systematically different across years. To remove the effects of adult childrens' own age and year fixed effects, we also plot the residual

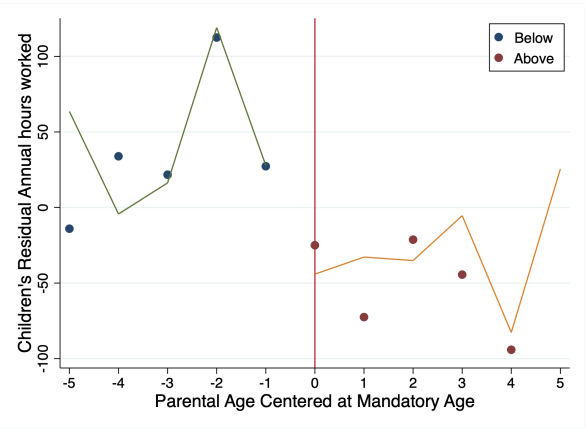
Figure 3.2: Adult Child Annual Hours and Parental Mandatory Retirement



(a) Annual hours: Linear



(b) Residual hours: Linear



(c) Residual hours: Local polynomial

Notes: The x-axis is the re-coded parental, with zero corresponding to the mandatory retirement cut-off. In figure 3.2a, the y-axis is adult children's annual hours of labor supply. Dots are means in 1-year bins. The red and blue lines are fitted from two separate linear regressions, one using data points above the cut-off and the other using data points below the cut-off. In figures 3.2b and 3.2c, the y-axis is the residual annual hours of labor supply predicted from regression $H_{it} = \alpha_1 Age_{it} + \eta_t + \nu_i$. In 3.2b, dots are means in 1-year bins.The red and blue lines are fitted from two separate linear regressions, one using data points above the cut-off and the other using data points below the cut-off. In 3.2c dots are means in 1-year bins. The red and blue lines are fitted from two separate local linear regressions using a triangular kernel with a 0.74-year bandwidth, one using data points above the cut-off and the other using data points below the cut-off.

annual hours predicted from the following model (model 3.1):

$$H_{it} = \alpha_1 Age_{it} + \eta_t + \nu_i \tag{3.1}$$

where H_{it} is adult i's annual hours of labor supply, Age_{it} is adult i's own age, and η_t

is year fixed effects. Residual $\hat{H}_{it}$ from model 3.1 is the residual annual hours which partial out any other potential effects and focus on the impact of parental retirement alone. Figure 3.2 panels 3.2b and 3.2c illustrate the residual annual hours against years from or to the parental retirement. We fit a linear model in Panel 3.2b and use a non-parametric triangular kernel approach in Panel 3.2c.

After removing the effect of own age and year fixed effects as suggested in Panel 3.2b and Panel 3.2c, the drop in annual residual hours around the cut-off becomes even more pronounced. Our graphical results therefore clearly shows a significant discontinuity in adult children's working hours at the threshold.

3.4.2 Regression Results

Following similar study designs in the literature (Card et al., 2008; Lee and Lemieux, 2010; Shigeoka, 2014), we employ a Regression Discontinuity (RD) approach using the mandatory retirement age as the cut-off. Our main regression equation is as follows:

$$H_{it} = \beta_1 Post_{it} + \beta_2 Running_{it} + \beta_3 Post_{it} \times Running_{it} + Age_{it} + \eta_t + \varepsilon_i \tag{3.2}$$

where

$$Post_{it} = \mathbf{1}\{Running_{it} \geq 0\}$$

$$Running_{it} = \max\{R_{it}^{dad}, R_{it}^{mom}\}$$

$$R_{it}^{g} = Age_{pt}^{g} - C^{g}, g = \{dad, mom\}$$

where subscripts i and p denote adult child and parent respectively. Model 3.2 is child-centric, where the dependent variable and regressors are defined from the perspective of each adult child. H_{it} is the outcome variable – adult child i's annual hours of labor supply in year t. C is the mandatory retirement age and varies by gender and occupation, which is individual- and time-invariant. Age^{g}_{pt} is the adult child i's father's age and mother's age, and R^{g}_{it} is the distance between i's mother or father's age and the mandatory retirement age. Our running variable, $Running_{it}$, picks the greater of R^{dad}_{it} and R^{mom}_{it} – namely, only the first-retired or first-to-retire parent's information will be included in our regression. $Post_{it}$ is a dummy variable that takes value one if the individual i's first-retired or first-to-retire parent has reached the cut-off C in year t. We include the interaction term of $Post$ and the running variable $Running_{it}$ to allow for different slopes below and above the cut-off. Age_{it} is adult child i's own age in year t. To capture differential economic condition and measurement discrepancy across survey years, we include year fixed effects η_t. ε_i is an i.i.d distributed error term. The parameter of interest is β_1, which captures the change in adult child's annual hours of labor supply due to parental retirement.

Validity Checks One key assumption of the RD design is that other pre-determined characteristics of the parents are smooth at the cut-off. Pre-determined variables include parents' marital status, gender, years of education, number of kids, whether they are frequent smokers or drinkers, and whether parents are covered by the pension system. Figure C.1 in the Appendix shows the scatter plots of the above variables, overlaid with lines from local linear regressions using data within our ±5 years window. The graphs show no visible discontinuities at the cut-off, indicating that local assignment around the cut-off is random. Table C.2 in the Appendix

shows the corresponding statistical test results. We find no significant changes at the cut-off, which confirms that the pre-determined covariates are smooth[4]. Overall, the RD validity checks support our empirical strategy and provide no evidence of violations of the key identifying assumptions.

Table 3.2: Baseline: Adult Children Hours Worked Around Parental Retirement

	(1)	(2)	(3)
		Hours	
Dep. var.	Parent Retired	Hired jobs	Hired plus Self-employed
Post	0.17***	-81.50***	-77.22***
	(0.01)	(27.63)	(27.37)
Running	0.01***	-2.86	-2.48
	(0.002)	(10.82)	(11.15)
Running x Post	0.038***	14.86	13.43
	(0.004)	(17.67)	(16.05)
Age_c		12.47***	17.05***
		(2.26)	(2.443)
R-squared	0.206	0.103	0.068
Year FE	Yes	Yes	Yes
windows	5	5	5
Observations	15,498	7,573	7,799

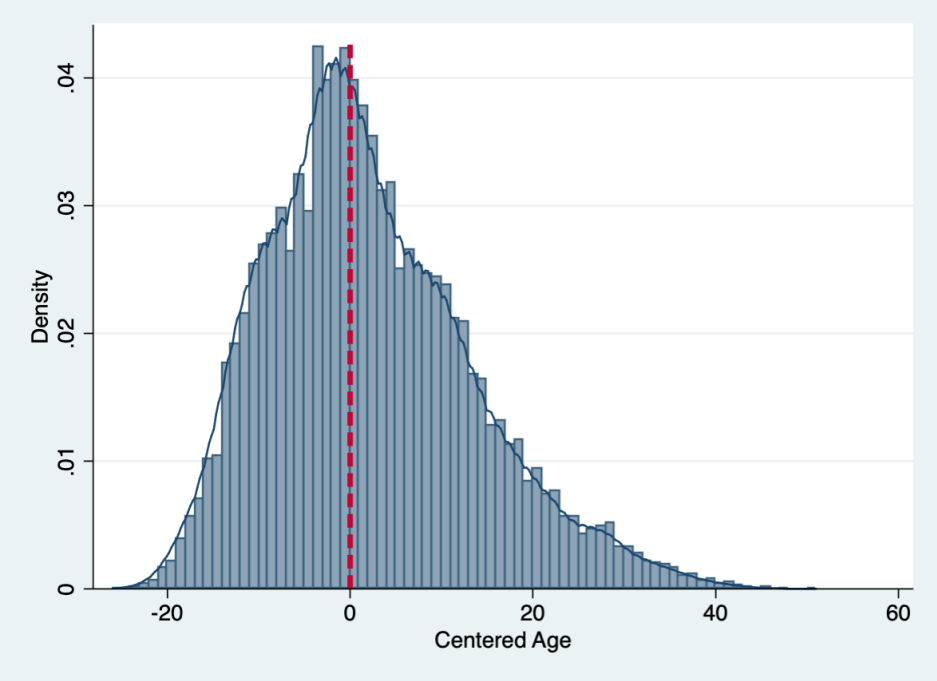
Notes: One unit of observation is a parent in column 1 and an adult child in column 2 and 3. "Parent Retired" is a dummy variable that takes value one if the parent's employment status is "retired" and zero otherwise. The dependent variable in Column 2 is the annual hours for hired jobs. The dependent variable in Column 3 is the annual hours of hired job plus self-employment hours. "Post" is a dummy variable that takes value one if an adult's first-to-retire parent has reached the mandatory retirement age and zero otherwise. "Running" is the first-to-retire parent's age minus the his or her corresponding mandatory retirement age. "Age_{it}" is the age of the adult child i's own age in year t. Year fixed effects are included in all columns. Standard errors are reported in parentheses and clustered by adult child birth year. *** p<0.01, ** p<0.05, * p<0.1.

Results Table 3.2 reports the baseline results from model 3.2. Column 1 suggests that on average parents are significantly more likely to retire once they pass the mandatory age. This correlation corresponds to the clear jump in the retirement rate in Figure 3.3. Columns 2 and 3 suggest that adult children's hours decrease by about 77 to 82 hours in a year, which is equivalent to a 3 to 4 percent drop from

[4]We also check children's pre-determined characteristics including marital status, gender, years of education, number of kids. The results are shown in Figure C.2 and Table C.3 in the Appendix. Again we find no significant discontinuities at the cut-off.

the annual average. The estimates are precise and significant at the level of one percentage point. In Column 2, we measure the dependent variable by considering the annual working hours for people who work in hired jobs. This includes both waged workers in urban areas and seasonal hired workers in rural areas. To make our analysis more general, we include self-employed adult children in our sample and report our regression results in Column 3. It's interesting to note that column 3 shows a smaller drop in hours of labor supply, meaning that self-employed adults experience a less significant negative effect. It is likely that the self-employed have more flexibility in their schedules so they can more easily accommodate the need to take care of parents without reducing total hours.

Figure 3.3: Parental Age Density Distribution Around the Mandatory Cut-off



Notes: This figure shows the age distribution around the mandatory retirement cut-off for individuals' parent without limiting to the ±5 windows. The x-axis is parental age relative to mandatory retirement age.

3.4.3 Heterogeneous Effects on Male and Female Children

In addition to the overall effect of parental retirement on adult children's labor supply, we are also interested in exploring if male children (sons and male spouses) and female children (daughters and female spouses) are affected differently. For this purpose, We re-run the baseline analysis in Model 3.2 for men and women separately.

Table 3.3 reports our gender-specific results. In columns 1 and 2, we include adult children who reported hours in hired jobs, while in columns 3 and 4 we also include self-employed individuals. Column 1 shows that after controlling for own age and year fixed effects, women's annual hours decreases by 123.7 hours from an average of 2,138.67 hours, which is equivalent to a 6 percent drop ($p < 0.01$). Column 2, however, shows a statistically insignificant change in men's annual hours. When we include self-employed individuals in columns 3 and 4, the sample sizes for both men and women slightly increase. As shown in column 3, women's hours decreases by around 89 hours, which is equivalent to 4 percent of the annual average. Men's hours as shown in Column 4, however, has not change significantly.

It is surprising that there is a significant difference in women and men's labor supply responses to parental retirement. However, this finding is consistent with other related findings in the the social norm and gender role literature where women are expected to perform more family duties compared to men. Traditionally men and women are associated with specific behavioral prescriptions as "the bread winners" and "the home makers" respectively (Budig and England, 2001; Ettner, 1995). It is possible that women suffers the burden due to the duty reinforced by social norms or gender identity. If this is the case, we expect to see differential transfer patterns between daughters and sons.

Table 3.3: Adult Children Hours Worked Around Parental Retirement: By Gender

	(1)	(2)	(3)	(4)
	Hours in hired jobs		Hours hired plus SE	
Dep. var.	Women	Men	Women	Men
Post	-123.70**	-49.05	-88.92*	-62.31
	(50.02)	(45.28)	(50.09)	(49.66)
Running	5.45	-9.61	2.93	-5.77
	(16.56)	(10.66)	(17.35)	(11.36)
Running x Post	24.58	12.54	22.45	10.02
	(26.35)	(17.47)	(28.12)	(16.00)
Age_c	7.57*	11.85***	11.67***	15.84***
	(4.13)	(2.87)	(4.07)	(3.01)
R-squared	0.110	0.103	0.069	0.072
Year FE	Yes	Yes	Yes	Yes
windows	5	5	5	5
Observations	3,110	4,455	3,197	4,669

Notes: One unit of observation is an adult child. The dependent variables in Columns 1 and 3 are the annual hours of labor supply in hired jobs. The dependent variables in Columns 2 and 4 are the annual hours in hired job plus the self-employment hours. "Post" is a dummy variable that takes value one if an adult's first-to-retire parent has reached the mandatory retirement age and zero otherwise. "Running" is the first-to-retire parent's age minus the his or her corresponding mandatory retirement age. "Age_{it}" is the age of the adult child i's own age in year t. Year fixed effects are included in all columns. Standard errors are reported in parentheses and clustered by adult child birth year. *** p<0.01, ** p<0.05, * p<0.1.

In sum, we find that overall there is a 3 to 4 percent drop in adult children's annual hours of labor supply when their parents retire. This reduction is more salient for daughters or female spouses than for sons or male spouses. In the next section, we explore possible explanations for the drop and for the gender-specific effects by studying the time and monetary transfer patterns between parents and children.

3.5 Mechanism of the Change in Adult Children's Labor Supply

In this section, we explore the underlying mechanism that helps explain the changes in adult children's labor supply due to parental retirement. We examine changes in monetary and time transfers between adult children and their parents, which could

be caused by changes in living arrangement and changes in parental health.

3.5.1 Time and Money Transfers

Due to retirement, parents would experience a drop in income and consequently consumption. Conforming to social norm, children would have the incentive to transfer money to support their parents(Bertrand et al., 2003). At the same time, due to the lack of formal eldercare system, adult children have to act as primary caregivers and parents may prefer to compensate children with money for their provision of care (Antonucci, 1990; Bernheim et al., 1985; Brandt and Deindl, 2013).

With an increase in parents' leisure time, we may expect that parents would help children more with housework, which would lead to an increase children's labor supply. However, papers in the past have shown that parents experience physical and mental decline when transitioning to retirement (Fitzpatrick and Moore, 2018; Müller and Shaikh, 2018). Therefore, it is possible that parents would need more support from adult children once they retire, especially at the beginning of this transition. To explore the changes in intergenerational transfer patterns, We replace the dependent variable in Model 3.2 with various transfer measures and estimate the following linear probability model:

$$Y_{it} = \gamma_1 Post_{it} + \gamma_2 Running_{it} + \gamma_3 Post_{it} \times Running_{it} + Age_{it} + \tilde{\eta}_t + \tilde{\varepsilon}_c \qquad (3.3)$$

where $Y_{it} = \{CareC_{it}, CareP_{it}\}$ is the set of transfer variables. To be consistent with the baseline Model 3.2, Model 3.3 is also specified as child-centric, where the

dependent variables and regressors are defined from the perspective of each adult child. Dependent variable $CareC_{it}$ is a dummy variable that takes value one if adult child i provided care or monetary transfer to his or her parent in the past six months in year t, and zero otherwise. Similarly, $CareP_{it}$ is a dummy variable that takes value one if adult child i received time or monetary transfer from parents in the past six months in year t, and zero otherwise. Other variables are the same as described in Model 3.2. Here, to be consistent, we adopt the linear specification without further parametric assumption on the error term as in Model 3.2. Given that the dependent variables are binary, we conduct robustenss checks using Probit model in Section 3.6.

γ_1 is the parameter of interest since it captures the change in the probability of upward and downward transfers due to parental retirement. Note that we can only observe transfers between parents and their biological children. So a caveat in interpreting our results is that our regression sample only considers biological daughters and sons while excluding spouses[5].

Table 3.4 reports our regression results from Model 3. Columns 1 and 2 suggest that adult children are 3.8 percent and 4.5 percent more likely to transfer money to and do housework for parents after their parents retire. These findings are consistent with the social norm in China where adult children are expected to take care of their parents. Columns 3 and 4 report changes in the likelihood of transfers from parents to adult children. Parents are 3.9 and 5.6 percent more likely to transfer money to and or do housework for their children after retirement. There are two possible explanations for such increases in downward transfer. On the one hand, since adult children spend more time taking care of parents after the parents retire, it is

[5]It is possible that parents attribute the efforts of children's spouses to their own children, which might lead to an upward or downward bias.

Table 3.4: Transfer: Adult Child Give or Receive Help From Parents

	(1)	(2)	(3)	(4)
	Ever transfer to parent		Ever receive from parents	
Dep. var.	Money	Housework	Money	Housework
Post	0.04**	0.05***	0.04***	0.06***
	(0.02)	(0.01)	(0.01)	(0.02)
Running	-0.003	0.002	0.002	-0.01
	(0.003)	(0.003)	(0.004)	(0.003)
Running x Post	0.02***	-0.004	-0.004	0.02***
	(0.01)	(0.01)	(0.01)	(0.01)
Age_c	0.01***	0.002	-0.01***	0.01***
	(0.001)	(0.001)	(0.002)	(0.001)
Constant	-0.24***	-0.02	0.24***	-0.29***
	(0.04)	(0.03)	(0.07)	(0.04)
R-squared	0.204	0.210	0.207	0.316
Year FE	Yes	Yes	Yes	Yes
windows	5	5	5	5
Observations	4,773	4,773	4,773	4,773

Notes: One unit of observation is an adult child. The dependent variables in column 1 and 2 are dummy variables that take value one if the adult child transferred money to or did housework for the parent in the past six months and zero otherwise. The dependent variables in column 3 and 4 are dummy variables that take value one if the adult child received money or housework help from the parent in the past six months and zero otherwise. "Post" is a dummy variable that takes value one if an adult's parent has reached the mandatory retirement age and zero otherwise. "Running" is the parent's age minus the his or her corresponding mandatory retirement age. "Age_{it}" is the age of the adult child i's own age in year t. Year fixed effects are included in all columns. Standard errors are reported in parentheses and clustered by adult child birth year. *** p<0.01, ** p<0.05, * p<0.1.

possible that parents compensate their children by giving small money in exchange (Antonucci, 1990; Bernheim et al., 1985; Brandt and Deindl, 2013). On the other hand, parents might use their own money when they do housework for their children. For example, anecdotal evidence suggests that parents sometimes cover their children's daily expenses partially or pay for grocery shopping and transportation. One may point out that that the magnitudes of coefficients for downward transfers are greater than those for upward transfers. However, one should be cautious when making such comparisons, because we only observe the extensive margin instead of the amount of transfers.

Table 3.5: Adult Children Transfer: By Gender

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Ever transfer to parent				Ever receive from parents			
	Money		Housework		Money		Housework	
Dep. var.	Daughter	Son	Daughter	Son	Daughter	Son	Daughter	Son
Post	0.04*	0.04	0.07***	0.04*	0.02	0.05***	0.002	0.08***
	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
Running	-0.01*	0.001	0.001	0.003	0.003	0.002	0.001	-0.01**
	(0.01)	(0.004)	(0.004)	(0.003)	(0.01)	(0.003)	(0.01)	(0.004)
Running x Post	0.02**	0.02***	-0.01	-0.002	-0.003	-0.01	0.01	0.03***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Age_c	0.01***	0.01***	-0.003	0.004***	-0.01***	-0.01***	0.01***	0.02***
	(0.002)	(0.001)	(0.002)	(0.001)	(0.003)	(0.002)	(0.002)	(0.002)
R-squared	0.177	0.226	0.217	0.206	0.216	0.210	0.217	0.388
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
windows	5	5	5	5	5	5	5	5
Observations	1,631	3,131	1,631	3,131	1,631	3,131	1,631	3,131

Notes: One unit of observation is an adult child. The dependent variables in column 1 (2) and 3 (4) are dummy variables that take value one if the daughter (son) transferred money to, or did housework for the parent in the past six months, zero otherwise. The dependent variables in column 5 (6) and 7 (8) are dummy variables and take value one if the daughter (son) received money, or received housework help from the parent in the past six months, zero otherwise. "Post" is a dummy variable that takes value one if an adult's parent has reached the mandatory retirement age and zero otherwise. "Running" is the parent's age minus the his or her corresponding mandatory retirement age. "Age_{it}" is the age of the adult child i's own age in year t. Year fixed effects are included in all columns. Standard errors are reported in parentheses and clustered by adult child birth year. *** p<0.01, ** p<0.05, * p<0.1.

Next, we investigate if male and female children experience different changes in terms of intergenerational transfer after parental retirement. Namely, we re-run Model 3.3 for male children and female children separately and report our results in Table 3.5. Columns 1 to 4 correspond to the time and monetary transfers to parents (upward transfers), whereas Columns 5 to 8 correspond to the time and monetary from parents to adult children (downward transfers). Interestingly, we observe very different patterns by gender. In columns 1, we see that daughters are 4.4 percent more likely to transfer money to parents, whereas column 2 suggests that sons are not statistically significantly more likely to provide financial supports to parents. However, in columns 5 and 6, we observe that the probability of sons receiving mon-

etary transfer from parents after parental retirement increased significantly, whereas for daughters this is not the case. The same pattern holds true for time transfers. Daughters are 6.7 percent more likely to do housework for parents after their parents retire, while sons' increase in time transfer is almost statistically insignificant and half in size in terms of magnitude. Meanwhile, sons are 8.4 percent more likely to receive help from parents while daughters are not.

The distinctive transfer patterns between daughters and sons can be explained by the theory of social norm and gender role, as mentioned in Section 3.1. Regarding social norm, it is a cultural tradition that Chinese parents favor sons over daughters (Li and Wu, 2011; Zheng, 2015). Sons are considered as the "family name bearer". Therefore, parents are more likely to provide both time and monetary transfers to sons. The disproportional upward transfer from daughters can be explained by the theory of gender role. Akerlof and Kranton (2000) for example, suggests that the division of tasks within households is self-sustained through gender norms and identity. The traditional role of Chinese women as homemakers, therefore, can be sustained into the modern era. As an empirical evidence, Chen and Zhang (2018) find that Chinese women devote significantly more hours into housework than men, especially in care-giving. Therefore, the burden of care naturally falls on the shoulders of women as parents retire, resulting in the discrepancies by gender both in terms of labor reply responses and changes in intergenerational transfer patterns.

In the next two subsections, we explore the explanations for the changes in intergenerational transfer upon parental retirement and the disparate transfer patterns by gender.

3.5.2 Changes in Living Arrangement

One plausible cause of the increase in upward transfers and consequently the decrease in adult children's labor supply could be the changes in living arrangement after parental retirement. It is possible that parents choose to live with their children after they retire and such change could lead to changes in children's time allocation (Bertrand et al., 2003). We do not find evidence for changes in living arrangement associated with parental retirement. Figure 3.3 shows the age distribution of parents around the mandatory retirement cut-off. It suggests that living arrangement does not suddenly change around the cut-off. If parents do not live with their children before they retire and move in with their children after retirement, we would expect to see fewer observations on the left of the cutoff and more on the right, as the first wave of the sample selection only covered parents that live with their children. Our plot, however, shows that the density is rather smooth around the cut-off. Thus, the changes in adult children's labor supply and intergenerational transfer patterns are not likely to be caused by living arrangement changes.

3.5.3 Changes in Parental Health

To understand why children increase their upward transfers after parental retirement, we examine the changes in parents' lives that might lead to an increased demand for money or care. One key aspect is parental health. We use the the following model to detect significant changes in parents' self-rated health:

$$ParentalHealth_{pt} = \lambda_1 Post_{pt} + \lambda_2 Running_{pt} + \lambda_3 Post_{pt} \times Running_{pt} + X_{pt} + \tilde{\eta}_t + \tilde{\nu}_c \tag{3.4}$$

where $ParentalHealth_{pt}$ includes 3 sets of outcome variables for a parent. The first set includes binary indicators for each of the five levels of smoking intensity in the past month: (1) never, (2) seldom, (3) frequent, (4) more frequent and (5) heavily smoke. The second set is a dummy variable that takes value one if the parent drank more than 3 times in the past week, and zero otherwise. The third set includes binary indicators for each of the five levels of parental self-rated health: (1) very healthy, (2) moderately healthy, (3) neutral, (4) less healthy and (5) very unhealthy. X_{pt} is parents' smoking and drinking behaviors, which are included here as control variables. The other regressors are the same as in model 3.2.

Columns 8 to 13 in Table C.2 suggest no clear increase in parents' risky health behaviors in terms of smoking and alcohol use. We therefore turn to look at changes in parents' subjective health ratings, which are reported in Table 3.6. Table 3.6 columns 1 to 5 report the changes in the likelihood of considering oneself as very healthy to very unhealthy. Column 1 suggests that people are less likely to positively rate themselves as healthy after retirement, after controlling for their drinking and smoking behaviors. The 4.3 percent drop in feeling very healthy (rate as 1) is statistically significant at one percent level. Meanwhile, parents are 4.4 percent more likely to consider themselves as very unhealthy. Thus finding confirms our hypothesis that parents self-rated health are negatively impacted by retirement, which increases their demand for attention and help from adult children.

Table 3.7 reports the change in hours by parental self-rated health. If a parent reports as "Neutral", "Less Healthy" or "Very Unhealthy", he or she is considered as "unhealthy", otherwise "healthy". Columns 1 and 3 show a large and significant drop in adult children's labor supply due to the poor self-rated health. If we look at people working in hired jobs only, column 1 suggests that the average hours go

Table 3.6: Parental Self-rated Health and Retirement

Dep. var.: Healthy	(1) Very	(2) Modest	(3) Neutral	(4) Less	(5) Unhealthy
Post	-0.04**	-0.002	0.001	0.02	0.04**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Running	0.004	0.001	0.001	-0.004	-0.01
	(0.01)	(0.01)	(0.004)	(0.01)	(0.01)
Running x Post	0.01	-0.004	3.91e-06	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Smoke in recent 1 mon	0.02	-0.03	0.06***	-0.001	-0.05***
	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)
Drink more than 3 times a week	-0.01	-0.01	0.003	0.04**	0.02
	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)
R-squared	0.288	0.130	0.190	0.038	0.100
Year FE	yes	yes	yes	yes	yes
windows	5	5	5	5	5
Observations	6,073	6,073	6,073	6,073	6,073

Notes: One unit of observation is a parent. The dependent variables in column 1-5 are dummy variables and take value one if one rates himself/herself as "Very Healthy", "Moderately Healthy", "Neutral", "Less Healthy" and "Very Unhealthy" respectively, and zero otherwise. "Post" is a dummy variable that takes value one if an adult's parent has reached the mandatory retirement age and zero otherwise. "Running" is the parent's age minus the his or her corresponding mandatory retirement age. "Smoke in recent 1 mon" and "Drink more than 3 times a week" are parents' risky healthy behaviors that are included here as control variables. Year fixed effects are included in all columns. Standard errors are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

down by 205.60 hours, which is 2.25 times the effect on people with healthy parents, 91 hours in column 2. At the same time, column 2 suggests that the effect on people with healthy parents is not statistically significant. This finding is also robust when we include self-employed people in columns 3 and 4. This comparison of effects by parental self-rated health suggests that the overall negative impact is driven by self-rated unhealthy parents, who requires more attention from adult children.

Table 3.7: Adult Children Hours Worked By Parental Self-rated Health

	(1)	(2)	(3)	(4)
	Hired jobs		Hired plus Self-employed	
Dep. var.: Hours	Unhealthy	Healthy	Unealthy	Healthy
Post	-205.6**	-91.05	-130.5*	-47.40
	(83.61)	(83.88)	(74.26)	(60.34)
Running	32.21*	20.25	0.83	14.35
	(16.20)	(21.76)	(18.94)	(21.81)
Running x Post	-32.85	-40.48	31.38	-30.34
	(24.28)	(25.27)	(29.57)	(39.83)
Age_c	32.74***	53.54***	6.514	10.80**
	(7.276)	(6.866)	(4.356)	(4.317)
R-squared	0.169	0.133	0.159	0.102
windows	5	5	5	5
Observations	2,468	3,849	1,707	2,281

Notes: One unit of observation is an adult child. The dependent variable in Columns 1 and 2 is the annual hours for hired jobs. The dependent variable in Columns 3 and 4 is the annual hours of hired job plus self-employment hours. "Unhealthy" takes value one if the parent's self-rated health is "Neutral", "Less Healthy" or "Very Unhealthy". "Healthy" takes value one if the parent's self-rated health is "Very Healthy" or "Moderately Healthy". "Post" is a dummy variable that takes value one if an adult's first-to-retire parent has reached the mandatory retirement age and zero otherwise. "Running" is the first-to-retire parent's age minus the his or her corresponding mandatory retirement age. "Age_{it}" is the age of the adult child i's own age in year t. Year fixed effects are included in all columns. Standard errors are reported in parentheses and clustered by adult child birth year. *** p<0.01, ** p<0.05, * p<0.1.

3.6 Robustness Checks

We perform three robustness checks for our baseline estimates from the following perspectives. First, we check if our estimates of the reduction in adult children's labor supply are affected by the anticipation effects described in Section 3.4 by performing a set of "donut-hole" RD regressions. Second, we check if our results are driven by parents who retire early because of health issues. Third, we specify an alternative time window (± 7 years) with respect to the mandatory retirement cutoff in Section 3.6.3 to check if the significant reduction in labor supply still remains. Lastly, given that the dependent variables for transfer and parental self-rated health are binary variables, we replace the linear probability model with a Probit model and check if

the estimated effects are sensitive to model specifications in Section 3.6.4.

3.6.1 Donut-Hole Design

Since retirement is anticipated, it is possible that people adjust their behavior ahead of time. For example, adult children may increase their labor supply ahead of time in anticipation of their parents' retirement and the potential drop in family income. This will result in an upward bias in our estimate of the labor supply reduction at cut-off. On the other hand, family may choose to reallocate the duties among household members in anticipation of changes in family life. For example, knowing that his or her parent is retiring soon, the child may start to ease out of his/her current role at work to prepare for the transition to a more family-centered role. This will lead to a downward bias in our estimate of the labor supply reduction, especially for women, since they are often the ones expected to transition early into a family role.

To check if our RD estimates are sensitive to anticipation effects, we implement a "donut-hole" RD design. The main idea is to exclude the few observations just above or below the cut-off. One drawback of this methodology is that there is no clear consensus regarding the optimal size of the donut hole. We choose to exclude observations one year above and below the cut-off.

Figure 3.4 graphically shows that the sizable drop in labor supply still remains after we exclude adult children whose parents are one year above and below the mandatory retirement age. In panel 3.4a, we plot adult children's annual hours of labor supply against the running variable. To partial out own age effect, we also plot the residuals of annual working hours as in model 3.1. Similar to our main results in

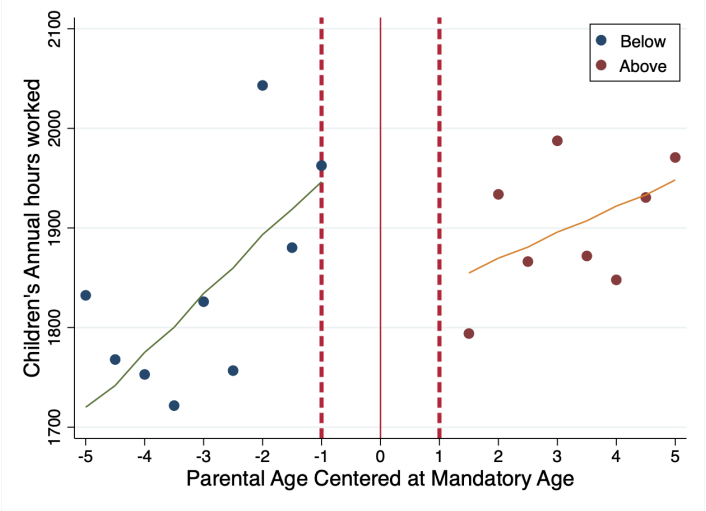
Section 3.4.2, the drop in adult children's working hours remains significant around the cut-off. Table 3.8 reports the regression results using our donut-hole sample. Column 1 suggests that there is on average an 152-hour drop in adult children's annual working hours when we only consider waged jobs. This effect is larger than the corresponding RD estimate (82.72 hours). When we include people who are self-employed in Column 2, the main effect remains significant, and is also larger than the RD estimates (77 hours). This greater effect suggests that anticipation effects lead to an overall downward bias in our estimate, meaning that the household duty reallocation effect dominates the saving-up for retirement effect.

Table 3.8: Robustness: Donut-Hole RD Design

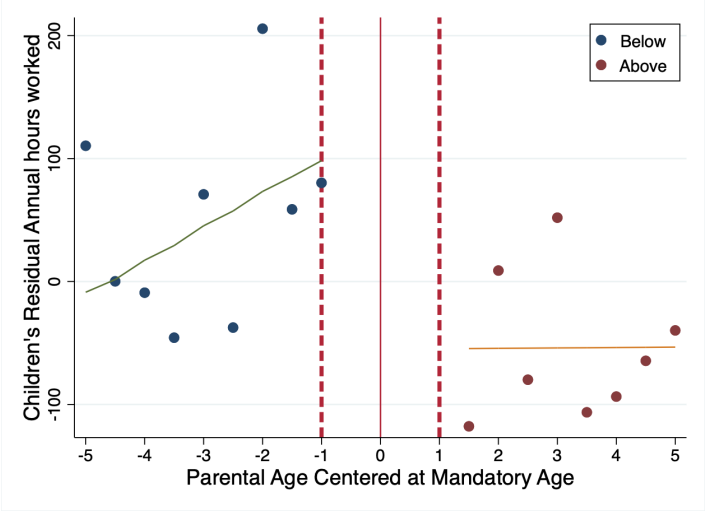
Dep. var.: Hours	(1) Hired jobs	(2) Hired plus SE
Post	-152.1***	-99.32*
	(53.67)	(53.65)
Running	26.31*	7.431
	(13.15)	(15.68)
Running x Post	-34.90	2.768
	(23.76)	(24.87)
Age_c	41.58***	13.56***
	(3.830)	(3.088)
Constant	160.7	2,030***
	(142.3)	(102.2)
R-squared	0.141	0.008
Windows	5-year with 1-year hole	5-year with 1-year hole
Observations	9,209	5,014

Notes: One unit of observation is an adult child. This table reports the robustness check for baseline model excluding observations ±1 year around the cut-off. The dependent variable in Column 1 is adult children's annual hours of labor supply in hired jobs. The dependent variable in Column 2 is adult children's annual hours of labor supply in hired jobs and hours reported as self-employed. 'Post" is a dummy variable that takes value one if an adult's parent has reached the mandatory retirement age and zero otherwise. "Running" is the parent's age minus the his or her corresponding mandatory retirement age. "Age_{it}" is the age of the adult child i's own age in year t. Year fixed effects are included in all columns. Standard errors are reported in parentheses and clustered by adult child birth year. *** p<0.01, ** p<0.05, * p<0.1.

Figure 3.4: Robustness: Donut-Hole Design



(a) Annual hours



(b) Residual annual hours

Notes: The x-axis is the parent's age normalized so that zero represents the mandatory retirement threshold. In panel 3.4a, the y-axis is the adult children's annual hours of labor supply. In panel 3.4b, the y-axis is the residual hours of adult children's annual labor supply after controlling for own age effect. Dots are means in 0.5-year bins. Lines are from separate above- and below-threshold linear regressions.

3.6.2 Early Retirement Due to Health Issues?

If parents are sick and choose to retire early, then the impact of parental retirement on adult's labor supply could be endogenous. To check if our result is driven by the sick parents who retire early, we exclude a) adults whose parents retire earlier than the mandatory age (11 percent of the sample), and b) adults whose parents retire earlier than the mandatory age and in bad objective health condition (smoking and drinking heavily, be in hospital in the year) (1 percent of the sample). Table 3.9 column 1 reports the baseline result in Table 3.2 column 2 using the entire sample. Table 3.9 columns 2 and 3 report the estimates when excluding people whose parents retire early and sick and retire early, respectively. Table 3.9 shows that the negative result on hours still holds. So this relieves the concern of reverse causality due to early (sick) retirement.

3.6.3 Alternative Time Window

As discussed in the sample construction subsection in Section 3.3, there is no statistical or economic consensus on the choice of the window size in RD design. To check the robustness of our results, we consider an alternative time window of ±7 years around the mandatory retirement cut-off for parents.

Table 3.10 reports the baseline estimates for Model 3.2 using our ±7 years sample. The drop in adult children's labor supply at the cut-off remains statistically significant. The magnitudes are also consistent with our baseline results.

Table 3.9: Robustness: Excluding Early Retirement

Dep. var.: Hours	(1) Hired jobs	(2) Excluding early retirement	(3) Excluding early and sick retirement
Post	-81.50***	-91.88***	-77.65**
	(27.63)	(28.17)	(28.73)
Running	-2.86	4.04	-4.94
	(10.82)	(11.06)	(10.96)
Running x Post	14.86	-0.759	17.15
	(17.67)	(20.58)	(17.37)
Age_c	12.47***	15.13***	12.25***
	(2.26)	(2.42)	(2.25)
Constant	2,041***	1,961***	2,040***
	(75.90)	(84.32)	(74.87)
Observations	7,573	5,319	7,512
R-squared	0.103	0.097	0.104
windows	5	5	5

Notes: One unit of observation is an adult child. This table reports the robustness check for baseline model excluding observations. Column 2 excludes adults whose parent retire earlier than the mandatory age. Column 3 excludes adults whose parent retire earlier than the mandatory age and in bad objective health condition (smoking and drinking heavily, be in hospital in the year). The dependent variable in all columns is adult children's annual hours of labor supply in hired jobs. 'Post" is a dummy variable that takes value one if an adult's parent has reached the mandatory retirement age and zero otherwise. "Running" is the parent's age minus the his or her corresponding mandatory retirement age. "Age_{it}" is the age of the adult child i's own age in year t. Year fixed effects are included in all columns. Standard errors are reported in parentheses and clustered by adult child birth year. *** p<0.01, ** p<0.05, * p<0.1.

3.6.4 Alternative Model Specifications

Since the dependent variables in Model 3.3 and Model 3.4 are binary variables, we check if our results are sensitive to model specification. In particular, we re-run the regressions in Model 3.3 and Model 3.4 using Probit model.

Table 3.11 reports the estimated marginal effects of parental retirement on the probability of giving (receiving) transfers to (from) parents with Probit model, using our original ±5 years sample. The estimates are slightly larger in magnitude than the effects in 3.4 and are still significant.

Table 3.12 reports the marginal effect estimates using Probit model for gender-

Table 3.10: Robustness: Seven-year Window

Dep. var.: Hours	(1) Hired jobs	(2) Hired plus SE
Post	-113.20***	-85.05***
	(32.15)	(28.26)
Running	14.98***	1.45
	(5.41)	(6.12)
Running x Pose	-17.64*	9.255
	(9.86)	(8.08)
Age_c	37.09***	11.70***
	(3.58)	(2.36)
Constant	205.80	2,045***
	(135.20)	(78.19)
R-squared	0.156	0.099
windows	7	7
Observations	15,010	10,031

Notes: One unit of observation is an adult child. This table reports the robustness check for baseline model using ±7 years as an alternative window. The dependent variable in Column 1 is adult children's annual hours of labor supply in hired jobs. The dependent variable in Column 2 is adult children's annual hours of labor supply in hired jobs and hours reported as self-employed. 'Post" is a dummy variable that takes value one if an adult's parent has reached the mandatory retirement age and zero otherwise. "Running" is the parent's age minus the his or her corresponding mandatory retirement age. "Age_{it}" is the age of the adult child i's own age in year t. Year fixed effects are included in all columns. Standard errors are reported in parentheses and clustered by adult child birth year. *** p<0.01, ** p<0.05, * p<0.1.

specific transfer probabilities. The coefficients are largely consistent with our estimates in Table 3.5. In columns 5 to 8, we still find disparate transfer patterns for daughters and sons in terms of receiving help from parents. Columns 1 to 4 compare the probabilities of providing help to parents by gender. Although the probability of sons providing upward transfers after parental retirement is higher using Probit model compared to linear probability model, the magnitude is still smaller than that of daughters. Thus, the disparate transfer patterns between male and female children remain robust to alternative model specification.

Table 3.13 reports the estimated marginal effects of retirement on parents' self-rated health using an Ordered Probit model. Column 1 to 4 report changes in the likelihood of considering oneself as "very healthy" to "less healthy", using "very

unhealthy" as the baseline. Column 1 suggests that parents are less likely to positively rate themselves as healthy after retirement after controlling for their risky health behaviors (drinking and smoking). The 4.4 percent drop in feeling very healthy (rate as 1) is very close to the corresponding estimate in column 1 of Table 3.6. Estimates in column 2 and 4 are similar to those in Table 3.6 as well. Thus, our estimates of changes in parents' self-rated health are also robust to alternative model specification.

Table 3.11: Robustness: Transfer

	(1)	(2)	(3)	(4)
	Ever transfer to parent		Ever receive from parents	
Dep. var.	Money	Housework	Money	Housework
Post	0.05***	0.05***	0.05***	0.07***
	(0.02)	(0.01)	(0.01)	(0.02)
Running	0.000	0.003	0.002	-0.003
	(0.004)	(0.003)	(0.004)	(0.004)
Running x Post	0.01***	-0.01	-0.002	0.01***
	(0.01)	(0.01)	(0.01)	(0.01)
Age_c	0.01***	0.003*	-0.01***	0.01***
	(0.001)	(0.001)	(0.002)	(0.001)
Log Likelihood	-2430.647	-2397.704	-2150.9	-2612.681
Year FE	yes	yes	yes	yes
windows	5	5	5	5
Observations	4,773	4,773	4,773	4,773

Notes: This table is the robustness check for Table 3.4 using Probit model. Marginal effects are reported in the first row, and standard errors are reported in parentheses. One unit of observation is an adult child. The dependent variables in column 1 and 2 are dummy variables that take value one if the adult child transferred money to or did housework for the parent in the past six months and zero otherwise. The dependent variables in column 3 and 4 are dummy variables that take value one if the adult child received money or housework help from the parent in the past six months and zero otherwise. "Post" is a dummy variable that takes value one if an adult's parent has reached the mandatory retirement age and zero otherwise. "Running" is the parent's age minus the his or her corresponding mandatory retirement age. "Age_{it}" is the age of the adult child i's own age in year t. Standard errors are reported in parentheses and clustered by adult child birth year. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.12: Robustness: Transfer By Gender

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Ever transfer to parent				Ever receive from parents			
	Money		Housework		Money		Housework	
Dep. var.	Daughter	Son	Daughter	Son	Daughter	Son	Daughter	Son
Post	0.06***	0.05**	0.07***	0.05**	0.03	0.06***	0.02	0.10***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
Running	-0.01	0.01	0.001	0.01	0.002	0.002	0.002	-0.01
	(0.01)	(0.01)	(0.01)	(0.004)	(0.01)	(0.004)	(0.01)	(0.01)
Running x Post	0.02*	0.01	-0.01	-0.01	0.002	-0.003	0.01	0.02**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Age_c	0.01***	0.01***	-0.001***	0.004***	-0.01***	-0.004***	0.01***	0.01***
	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	0.002)	(0.001)	(0.001)
Log Likelihood	-779.439	-1625.484	-849.65	-1535.977	-739.102	-1391.821	-868.843	-1661.932
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
windows	5	5	5	5	5	5	5	5
Observations	1,631	3,131	1,631	3,131	1,631	3,131	1,631	3,131

Note: This table is the robustness check for Table 3.5 using Probit model. Marginal effects are reported in the first row, and standard errors are reported in parentheses. One unit of observation is an adult child. The dependent variables in column 1 (2) and 3 (4) are dummy variables that take value one if the daughter (son) transferred money to, or did housework for the parent in the past six months, zero otherwise. The dependent variables in column 5 (6) and 7 (8) are dummy variables and take value one if the daughter (son) received money, or received housework help from the parent in the past six months, zero otherwise. "Post" is a dummy variable that takes value one if an adult's parent has reached the mandatory retirement age and zero otherwise. "Running" is the parent's age minus the his or her corresponding mandatory retirement age. "Age_{it}" is the age of the adult child i's own age in year t. Standard errors are reported in parentheses and clustered by adult child birth year. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.13: Robustness: Parental Self-rated Health and Retirement using Ordered Probit Model

Dep. var. Self-rate as:	(1) More Healthy	(2) Moderate	(3) Neutral	(4) Less
Post	-0.044***	0.000	0.006**	0.039**
	(0.019)	(0.000)	(0.002)	(0.017)
Running	0.004	0.000	-0.001	-0.004
	(0.005)	(0.000)	(0.001)	(0.004)
Running x Post	0.008	0.000	-0.001	-0.007
	(0.006)	(0.000)	(0.001)	(0.006)
Smoke in recent 1 mon	0.032**	0.000	-0.004**	-0.027**
	(0.014)	(0.001)	(0.002)	(0.012)
Drink more than 3 times a week	-0.011	0.000	0.001	0.010
	(0.012)	(0.000)	(0.002)	(0.011)
	0.360	0.458	0.361	0.360
Log likelihood = -7181.432 N = 6073				

Notes: This table is the robustness check for Table 3.6 with ordered probit model. Marginal effects are reported in the first row, and standard errors are reported in parentheses.
One unit of observation is a parent. The dependent variables in column 1-4 are indicator variables of self-rated health: “Very Healthy” (value 1), “Moderately Healthy”(value 2), “Neutral”(value 3), “Less Healthy” (value 4). “Very Unhealthy” is used as the baseline. “Post” is a dummy variable that takes value one if an adult’s parent has reached the mandatory retirement age and zero otherwise. “Running” is the parent’s age minus the his or her corresponding mandatory retirement age. “Smoke in recent 1 mon" and “Drink more than 3 times a week" are parents’ risky healthy behaviors that are included here as control variables. Standard errors are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

3.7 Concluding Remarks

This paper studies the impact of parental retirement on adult children's labor supply and investigates the mechanisms, namely the changes in time and monetary transfers between parents and adult children due to parental retirement. We exploit the exogenous mandatory retirement age in China and use a regression discontinuity (RD) design to estimate the intent-to-treat effect of parents reaching mandatory retirement age. We find a significant reduction in adult children's annual hours of labor supply by 3 to 4 percent. The negative effect is especially pronounced for female children.

We find that the parents' self-rated health also experience a sizable drop as they pass the mandatory retirement age. The negative effect is driven by self-rated unhealthy parents. With a lack of formal eldercare provision, parents rely more on adult children and demand more care from them when they are transitioning into retired life. Our results indeed suggest that the upward transfer from children to parents, both in terms of money and in terms of time, increased significantly upon parental retirement. In addition, we find that daughters are more likely to provide money and help to parents while receiving less support from parents compared to sons. This showcases the barrier of traditional gender role and social norm imposes on Chinese women's endeavors in balancing market labor supply and home production.

Our study has two major policy implications. First of all, since formal elderly care and assistance from family members are close substitutes, central and local government should devote more resources into building affordable elderly care facilities so as to alleviate the burden on and career costs to adult children with retired parents.

Second, since social norm and traditional gender role dictate Chinese women as the main care-givers, workplace amenities such as flexible working hours and “elderly care days" will help female employees balance the demands from work and family.

Two main limitations exist in our study. First of all, due to the limited scale of the survey and the fact that many respondents failed to report working hours, the number of observations included in our final sample is not large enough for further dissection. For example, with sufficiently large sample size, we could have compared the effects of father’s retirement to mother’s retirement, or parent’s retirement to in-law’s retirement. With our sample size, however, the statistical power will be jeopardized. Second, we only observe the extensive margin of inter-generational transfers, not the number of hours or monetary amounts. This limits our ability to quantify the size of upward and downward transfers and the statistical significance of changes in size. Therefore, more research will be required in order to understand the true career cost of parental retirement to adult children and the details of the underlying mechanisms.

APPENDIX A

APPENDIX TO CHAPTER 1

A.1 Tables and Figures in Appendix

Table A.1: 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 A.2: 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 A.1: Distribution of Share Friend

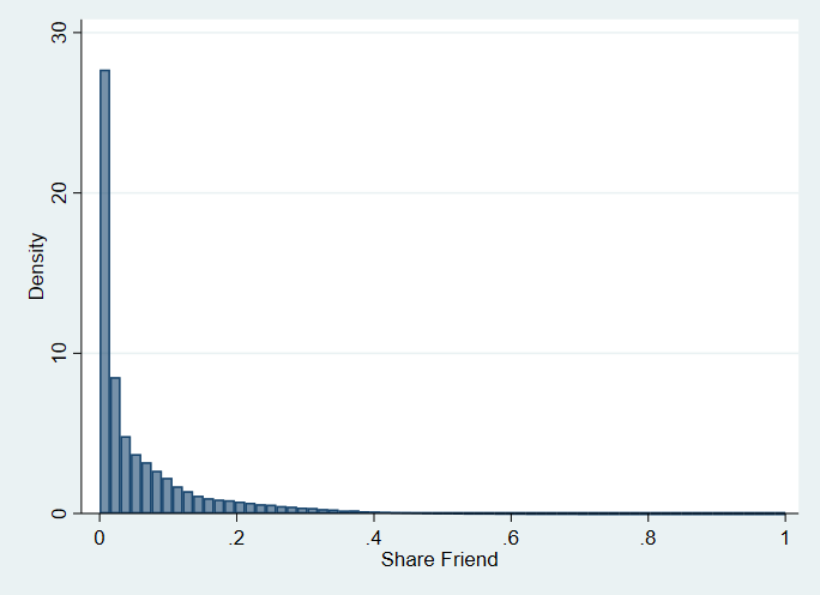


Table A.3: Summary Statistics: Share of Friends by Different Groups

	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 A.4: Balance Test: By Fraction of Same-carrier Friends

	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 A.5: 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 A.6: 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 A.7: 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.

Table A.8: 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.

A.2 Research File for Sample Construction

A.2.1 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:

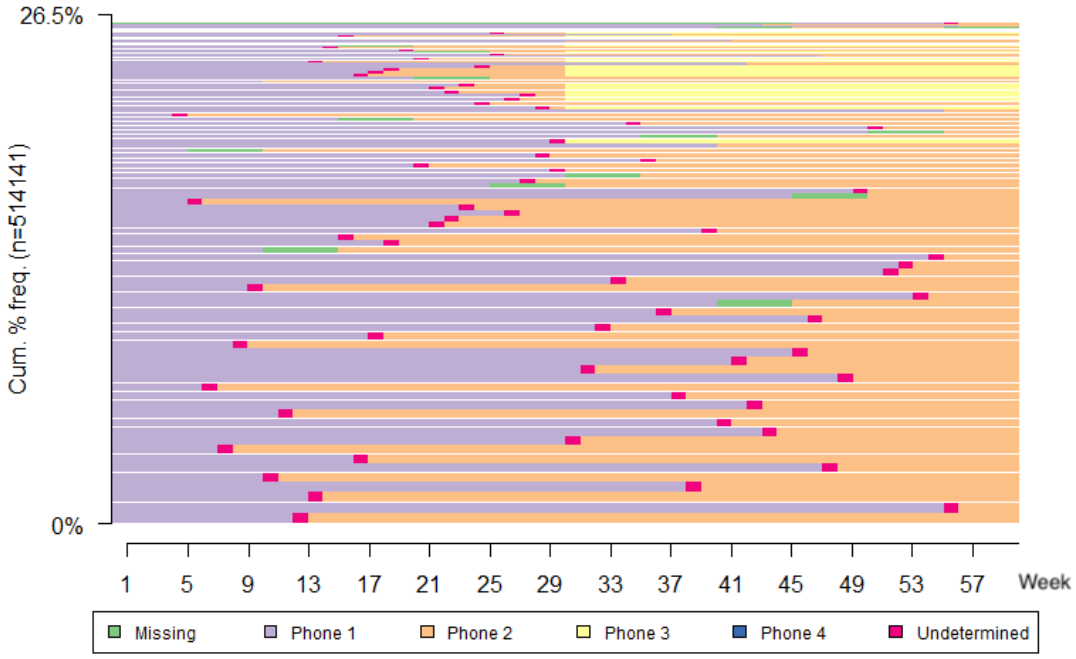
1. One sim card experienced more than one devices (brand + model) in the sample periods
2. There is no re-occurrence of a previously held device
3. Holding the new device for at least one month
4. holding at least one previous device for at least one month

Table A.9: 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 A.2: 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.

A.2.2 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 A.10 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 A.10: 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 A.11: 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 A.11 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 A.12: 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

A.2.3 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) (com-

pared 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 A.13: 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

A.3 Estimation and Counterfactual Simulation Procedures

A.3.1 Demand Estimation Routine

For each individual, R = 1000, fix a set of draws $\{\nu_i^r\}_{r=1}^R$ 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 $\boldsymbol{\delta}^0$.

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

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

   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:

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

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

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

   Iterate over the contraction mapping until $\boldsymbol{\delta}$ 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 $\boldsymbol{\delta}(\boldsymbol{\theta}_2^0)$.

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

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

The simulated likelihood function of the sample becomes

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

4. Choose $\boldsymbol{\theta}_2$ and $\boldsymbol{\delta}(\boldsymbol{\theta}_2)$ 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:

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

A.3.2 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

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

   (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^*

A.4 Prices and Social Influence: Model Prediction Illustration

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

$$W_f = \sum_{j \in J_f} (p_{j1} - mc_{j1}) Q_{j1} + \delta (p_{j2} - mc_{j2}) Q_{j2} \tag{A.2}$$

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

$$mc_{j2} = p_{j2}^* + [\Delta_{f2}^{-1} \times \mathbf{Q}_2]_j \tag{A.3}$$

$$
\begin{aligned}
mc_{j1} &= p^*_{j1} + \left[\Delta^{-1}_{f1} \times \left\{Q_{j1} + \delta \sum_{r \in J_f} (p_{r2} - mc_{r2}) \frac{\partial Q_{r2}}{\partial p_{j1}} + \delta Q_{j2} \frac{\partial p_{j2}}{\partial p_{j1}}\right\}\right] \\
&= p^*_{j1} + \left[\Delta^{-1}_{f1} \times \left\{\mathbf{Q}_1 - \delta \frac{\partial \mathbf{Q}_2}{\partial \mathbf{p}_1}[\Delta^{-1}_{f2} \times \mathbf{Q}_2] - \delta Diag\left(\frac{\partial \mathbf{Q}_2}{\partial \mathbf{p}_1}\right)[\Delta^{-1}_{f2} \times \mathbf{Q}_2]\right\}\right]_j
\end{aligned}
\tag{A.4}
$$

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

$$
p^*_{j2} = mc_{j2} - [\Delta^{-1}_{f2} \times \mathbf{Q}_2]_j
$$

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

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

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

$$\begin{aligned} p^{\theta=0}_{j1} - p^{\theta=\theta^*}_{j1} &= \left[-\Delta^{-1}_{f1}(\boldsymbol{Q}^0_1 - \boldsymbol{Q}^{\theta^*}_1) + \Delta^{-1}_{f1} \times \delta\left\{\frac{\partial \boldsymbol{Q}_2}{\partial \boldsymbol{p}_1}[\Delta^{-1}_{f2} \times \boldsymbol{Q}_2] + Diag\left(\frac{\partial \boldsymbol{Q}_2}{\partial \boldsymbol{p}_1}\right)[\Delta^{-1}_{f2} \times \boldsymbol{Q}_2]\middle|\theta^*\right\}\right]_j \\ &\approx \left[\Delta^{-1}_{f1} \times \delta\left\{\frac{\partial \boldsymbol{Q}_2}{\partial \mathbf{p}_1}[\Delta^{-1}_{f2} \times \boldsymbol{Q}_2] + Diag\left(\frac{\partial \boldsymbol{Q}_2}{\partial \boldsymbol{p}_1}\right)[\Delta^{-1}_{f2} \times \boldsymbol{Q}_2]\middle|\theta^*\right\}\right]_j > 0 \end{aligned} \tag{A.5}$$

Note that the first term $\boldsymbol{Q}^0_1 - \boldsymbol{Q}^{\theta^*}_1$ in first line in Equation A.5 is driven by the price effect $p^{\theta=0}_{j1} - p^{\theta=\theta^*}_{j1}$ 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^{-1}_{f1} 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 $\frac{\partial \boldsymbol{Q}_2}{\partial \boldsymbol{p}_1}$ 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.

APPENDIX B

APPENDIX TO CHAPTER 2

B.0.1 Occupancy Description

We use job descriptions and job titles and the US 2010 occupation codes to classify the occupation for each posting. Here are the occupations that we use:

1. Management – includes customer service manager, warehouse manager, production manager, hospital manager, human resource manager, CEO, retail shop manager and vice manager, sales manager, education administrator, etc.
2. Professionals – includes business operation, finance operation, computer and science, social science and non-training professionals; business related, including wholesale trader, market research analyst, meeting and event planner, cost estimator, risk control worker, customer relation, accountants and auditors; computer and science related, including software developers, computer support specialists, database administrator, web developer, network and computer systems administrators, architects, biomedical engineers, mining and geological engineers, mapping technicians, nutritionists.
3. Education, legal, arts, design, and media – education includes training professionals, preschool and kindergarten teachers, afterschool class teachers, teaching assistants, vocational training instructors, driving coach; legal includes lawyer and paralegals; arts, design, and media include director, model, hosts, actors, writers, photographers, video editors, news reporters, designers, magazine editors, webpage editors.

4. Service – includes cashier, customer service, front desk, fire fighter, nail polisher, cleaner, massage, flight attendants, food server, cooks, laundry workers, counter attendants, security guards, surveillance control workers.
5. Sales and office administration – sales includes retail salesperson, insurance salesperson, real estate sales agents, pharmaceutical sales representatives; office administration includes office secretary, file clerks, curriculum consultants (in private education organizations).
6. Health related – includes therapists, nurses, pharmacists, rehabilitation doctors, and surgeons.
7. Production and transportation – production includes printing press operators, layout workers, general production workers, painting workers, cutting workers; transportation includes sailors, cargo shipping drivers, drivers in general, warehouse workers, and material moving workers.
8. Farming, fishing, and construction – includes related natural resource, installation, maintenance, repair, welder, installation workers, computer repairers, maintenance workers, gardeners, agricultural workers, forest workers, breeding workers, and livestock cultivators.

We combine the three smallest categories (Health related, Production and transportation, and Farming, fishing, and construction) into 'other category' in our empirical analysis.

B.0.2 Tables in the Appendix

Table B.1: Summary Statistics of Diversity Measures

Variable	Mean	SD	Median	Min	Max
Social entropy (working population)	0.67	0.03	0.67	0.40	0.83
Social entropy (residential population)	0.67	0.05	0.67	0	0.95
Spatial entropy (working population)	0.71	0.04	0.70	0.40	0.94
Spatial entropy (residential population)	0.72	0.05	0.72	0	1.00
Income entropy (working population)	0.46	0.11	0.46	0	0.83
Income entropy (residential population)	0.46	0.10	0.46	0	0.92

Notes: Each entropy measure is the normalized Shannon entropy averaged across either the working population or the residential population at a given location. There are 6,161 locations in total.

Table B.2: Summary Statistics of Key Variables in Regression Samples

Panel A: Switcher Attributes	Mean	SD	Median	N
Pr(i switches to l)	0.09	0.16	0.00	33,399
Friend	0.26	0.44	0.00	33,399
Distance(job1, job2) in km	10.45	15.72	3.95	38,102
Distance(home, job2) in km	8.58	12.95	3.29	34,927
Rural to urban	0.06	0.24	0.00	38,102
Young (Age 25-34)	0.36	0.48	0.00	38,102
Change sector	0.61	0.49	1.00	10,116
Panel B: Job Benefits	**Mean**	**SD**	**Median**	**N**
Wage at new job (thousand RMB)	31.47	24.30	25.22	17,615
ΔCoworker HP (thousand RMB)	-0.11	3.40	-0.06	23,323
PT to FT	0.16	0.37	0.00	19,431
Closer to home	0.31	0.46	0.00	29,117
Non-SOE to SOE	0.09	0.29	0.00	15,881
Panel C: Large firms with Positive hirings	**Mean**	**SD**	**Median**	**N**
Net inflow	2.77	6.35	1.00	[600,1000]
Matching rate	0.76	0.38	1.00	[600,1000]
Growth rate	0.04	0.06	0.02	[600,1000]
Firm network size (log)	5.92	1.90	6.13	[600,1000]
Referral	0.57	0.50	1.00	[600,1000]

Notes: Panel A reports summary statistics for key variables in Table 2.11. Panel B reports summary statistics for key variables in Table 2.14. Panel C reports summary statistics for key variables in Table 2.15.

Table B.3: Referral Benefits to All Firms with Positive Hiring

Dependent variable	(1) Net inflow	(2) Matching rate	(3) Growth rate
Referral	0.46***	0.57***	0.49***
	(0.05)	(0.11)	(0.05)
Controls	Yes	Yes	Yes
Observations	[3000,5000]	[2000,5000]	[3000,5000]
R-squared	0.53	0.79	0.70
Neighborhood FE	Yes	Yes	Yes
Num. of Neighborhood FE	631	526	707

Notes: One unit of observation is a location with at least one matched firm and positive hirings. Same specification as in Column 4 of Table 2.15. Standard errors are reported in parentheses and clustered by neighborhood. *** p<0.01, ** p<0.05, * p<0.1.

Table B.4: Referral Effects with an Alternative Friend Definition

Dependent variable Probability i switches to location l	(1)	(2)	(3)	(4)	(5)	(6)
Friend	0.36***	0.36***	0.34***	0.34***	0.35***	0.35***
	(0.00)	(0.00)	(0.02)	(0.02)	(0.02)	(0.01)
Controls	No	Yes	No	Yes	No	Yes
Observations	1,151,676	1,120,797	1,151,676	1,120,797	1,151,676	1,120,797
R-squared	0.15	0.15	0.20	0.20	0.21	0.21
New work Neighborhood FE	No	No	Yes	Yes	No	No
Old x New Neighborhood FE	No	No	No	No	Yes	Yes
N. Neighborhood FE	NA	NA	1,111	1,107	21,250	20,811

Notes: Same specification as in Table 2.7. Friends have at least four weeks of non-missing work locations. Standard errors are reported in parentheses. They are clustered by neighborhood in Columns 3 and 4 and by neighborhood-pair in Columns 5 and 6. *** p<0.01, ** p<0.05, * p<0.1.

Table B.5: Referral Benefits to Workers with an Alternative Friend Definition

	Income Effect		Job Quality		
Dependent variable	(1) Wage at new job	(2) ΔCoworker HP	(3) PT to FT	(4) Closer to home	(5) Non-SOE to SOE
Friend	0.40*	0.08**	0.02***	0.09***	0.0075*
	(0.21)	(0.04)	(0.01)	(0.01)	(0.004)
Observations	18,595	24,835	21,016	31,013	16,789
R-squared	0.79	0.52	0.10	0.12	0.56
Residence Neighborhood FE	Yes	Yes	Yes	Yes	Yes
New Work Neighborhood FE	Yes	Yes	Yes	Yes	Yes

Notes: Same specifications as in Table 2.14. Friends have at least four weeks of non-missing work locations. Standard errors are reported in parentheses and two-way clustered by residence neighborhood and by new work neighborhood. *** p<0.01, ** p<0.05, * p<0.1.

Table B.6: Referral Benefits to All Firms with an Alternative Friend Definition

Dependent variable	(1) Net inflow	(2) Matching rate	(3) Growth rate
Referral	0.81***	0.73***	0.62***
	(0.13)	(0.27)	(0.09)
Controls	Yes	Yes	Yes
Observations	[600,1000]	[400,1000]	[600,1000]
R-squared	0.68	0.87	0.85
Neighborhood FE	Yes	Yes	Yes
Num. of Neighborhood FE	225	190	271

Notes: Same specification as in Column 4 of Table 2.15. One unit of observation is a location with at least one matched firm and positive hirings. Friends have at least four weeks of non-missing work locations. Standard errors are reported in parentheses and clustered by neighborhood. *** p<0.01, ** p<0.05, * p<0.1.

Table B.7: Referral Effect with a Two-way Friend Definition

Dependent variable Probability i switches to location l	(1)	(2)	(3)	(4)	(5)	(6)
Friend	0.39***	0.39***	0.37***	0.37***	0.38***	0.38***
	(0.00)	(0.00)	(0.02)	(0.02)	(0.02)	(0.02)
Controls	No	Yes	No	Yes	No	Yes
Observations	1,151,676	1,120,797	1,151,676	1,120,797	1,151,676	1,120,797
R-squared	0.08	0.08	0.14	0.14	0.14	0.14
New work Neighborhood FE	No	No	Yes	Yes	No	No
Old x New Neighborhood FE	No	No	No	No	Yes	Yes
N. of Neighborhood FE	NA	NA	1,111	1,107	21,250	20,811

Notes: Same specification as in Table 2.7. Friends are social contacts with two-way communications: they both place calls to and receive calls from individual i. Friends have at least four weeks of non-missing work locations. Standard errors are reported in parentheses, and are clustered by neighborhood in Columns 3 and 4 and by neighborhood-pair in Columns 5 and 6. *** p<0.01, ** p<0.05, * p<0.1.

APPENDIX C

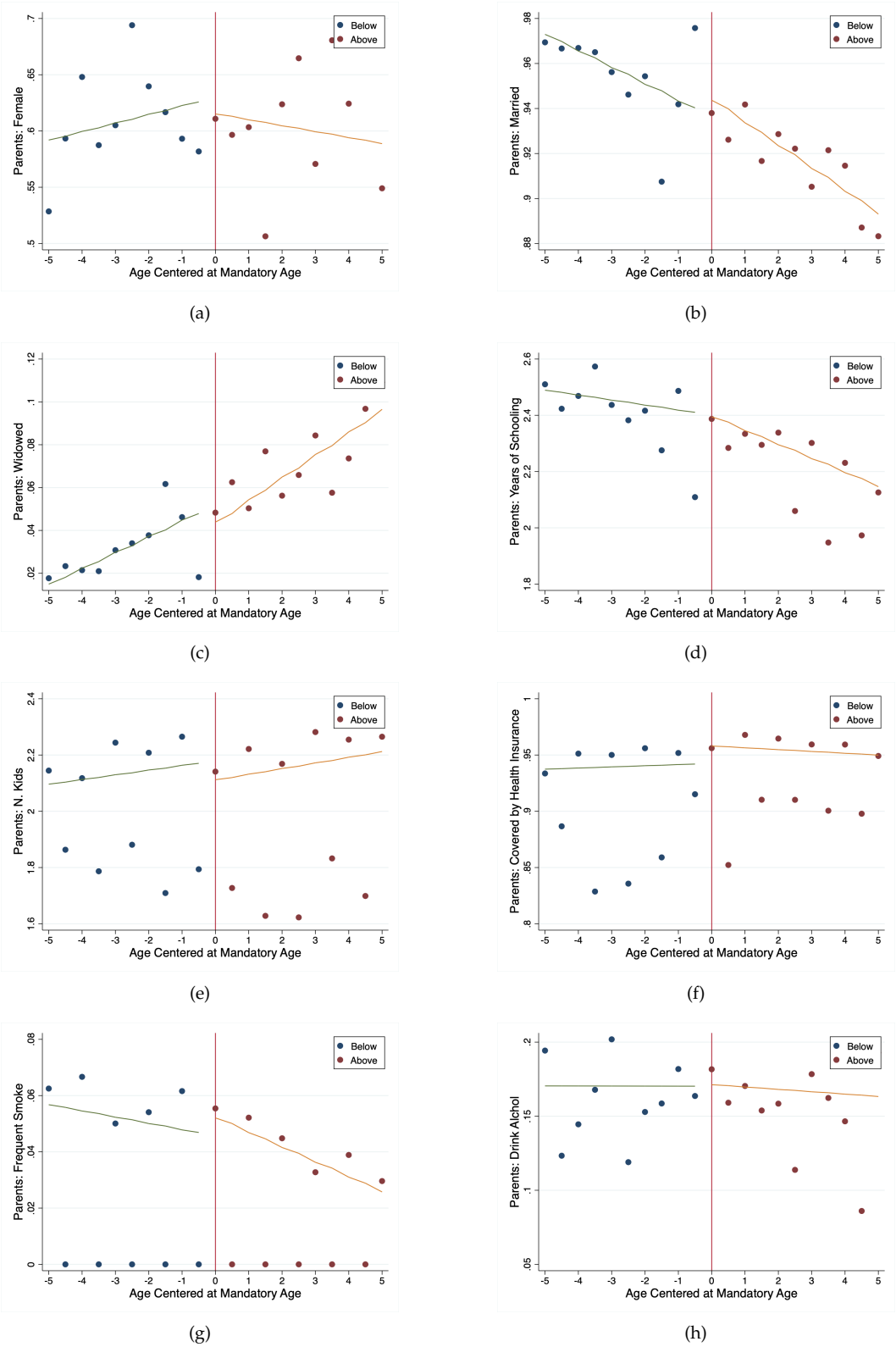
APPENDIX TO CHAPTER 3

Table C.1: Balance Test for Missing Hours

	Non-missing hours			missing Hours				
	Mean	SD	N	Mean	SD	N	Diff.	t-stat
Female	0.32	0.47	9009	0.30	0.46	4065	-0.02*	-1.74
Age	27.59	5.00	9009	29.25	5.51	4065	1.67***	17.07
Urban Area	0.47	0.50	8775	0.37	0.48	4048	-0.10***	-10.60
Married	0.40	0.49	9009	0.32	0.47	4065	-0.07***	-8.19
Years of schooling	10.46	4.00	8623	9.00	4.44	4050	-1.46***	-18.54
Income	19992.77	112110.89	8947	4327.64	13988.20	3998	-15665.13***	-8.80
Asset (thsd yuan)	310.10	531.81	8762	181.78	392.48	3998	-128.32***	-13.65
N.kid under age 1	0.05	0.24	9009	0.06	0.25	4065	0.01	1.18
N.kid age 1-2	0.21	0.45	9009	0.22	0.46	4065	0.02*	1.89
N.kid age 3-5	0.27	0.53	9009	0.38	0.62	4065	0.10***	9.55
N.kid age 6-16	0.36	0.66	9009	0.51	0.77	4065	0.15***	11.34
Parent Age (recode)	59.78	2.95	9009	60.33	3.03	4065	0.55***	9.73
Parent Retired	0.26	0.44	6928	0.31	0.46	2778	0.05***	5.34
Post	0.51	0.50	9009	0.58	0.49	4065	0.07***	7.82

Notes: This table reports the characteristics of adult children with and without missing working hours.

Figure C.1: Validity Test: Parental Covariates



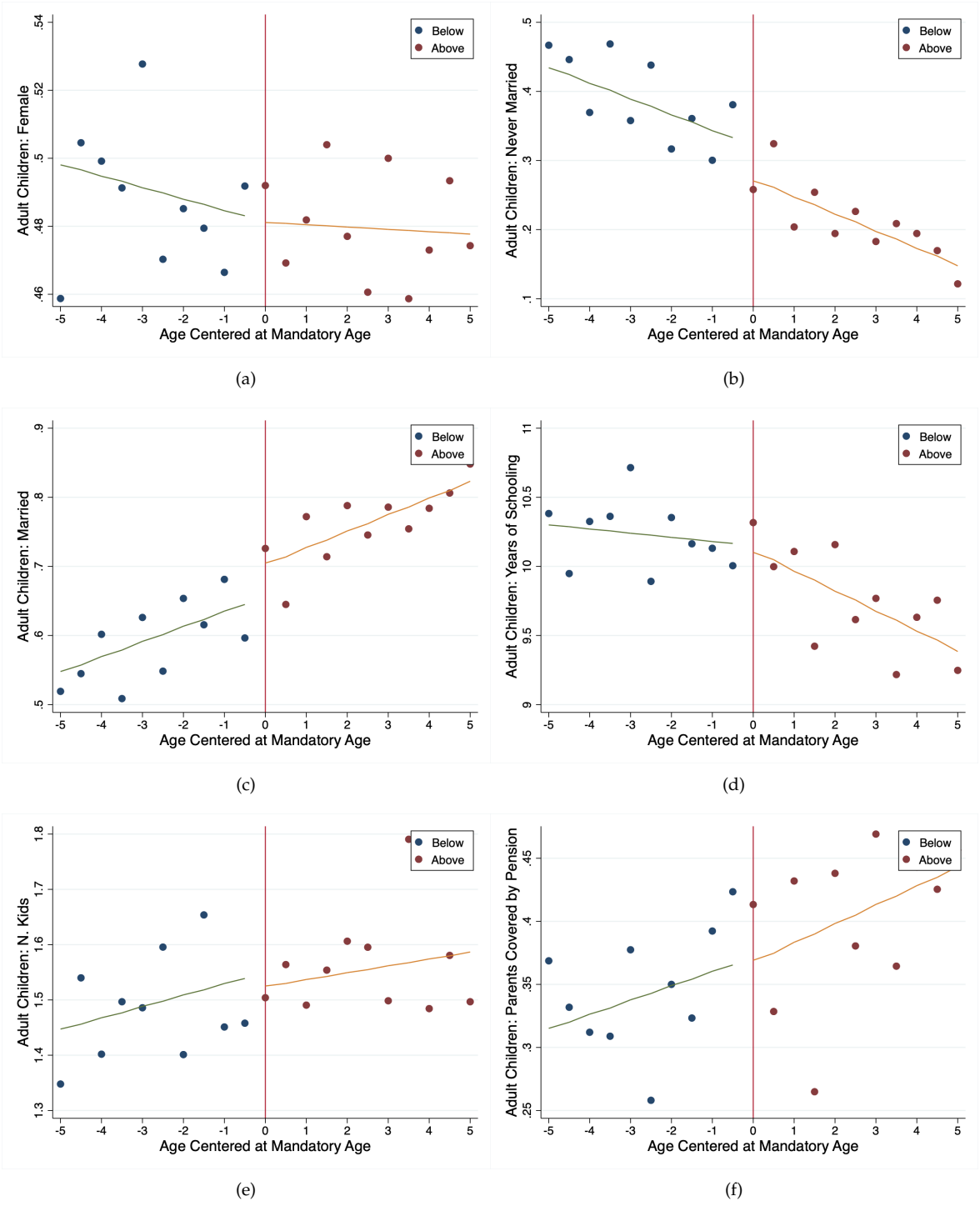
Notes: Validity Test: We plot the change in different covariates of parents below and above the mandatory retirement cutoff. C.1a describes the fraction of female; C.1b and C.1c describe the fraction of individuals who are married and widowed respectively; C.1d describes the years of schooling; C.1e describes the number of adult children in the family; C.1f describes the fraction of individuals whose parents are covered by pension. C.1g and C.1h describe the fraction of individuals who are frequent smokers and alcohol users respectively.

Table C.2: Covariates Smooth at Age Cutoff: Parents

	Marital status							Smoking						
VARIABLES	(1) Never	(2) Married	(3) Cohabitation	(4) Divorced	(5) Widowed	(6) schooling	(7) N. children	(8) No	(9) Seldom	(10) Frequent	(11) More frequent	(12) Heavy	(13) Alcoholic	(20) Pension
Post	-0.001	0.019*	-0.001	-0.0003	-0.017*	-0.112	-0.099*	-0.009	0.003	-0.021**	-0.003	0.015	0.005	0.013
	(0.002)	(0.011)	(0.002)	(0.003)	(0.010)	(0.114)	(0.054)	(0.019)	(0.024)	(0.010)	(0.017)	(0.012)	(0.026)	(0.012)
Running	0.0003	-0.010***	0.001*	-0.0003	0.009***	0.013	0.013	0.009	-0.007	0.001	0.002	-0.004*	-0.008	-0.004
	(0.0003)	(0.003)	(0.001)	(0.001)	(0.002)	(0.032)	(0.016)	(0.006)	(0.006)	(0.004)	(0.005)	(0.002)	(0.008)	(0.003)
Post x Running	-0.0004	0.0003	-0.002**	0.001	0.001	-0.043	-0.008	-0.001	-0.003	-0.004	0.003	0.002	-1.62e-06	0.007
	(0.001)	(0.004)	(0.001)	(0.001)	(0.004)	(0.037)	(0.027)	(0.009)	(0.009)	(0.005)	(0.006)	(0.003)	(0.012)	(0.005)
Constant	0.002**	0.926***	0.004*	0.005*	0.062***	2.643***	1.980***	0.422***	0.203***	0.090***	0.145***	0.067***	0.193***	0.913***
	(0.001)	(0.010)	(0.002)	(0.003)	(0.007)	(0.093)	(0.050)	(0.024)	(0.024)	(0.011)	(0.014)	(0.008)	(0.034)	(0.011)
Observations	6,073	6,073	6,073	6,073	6,073	6,073	3,421	6,073	6,073	6,073	6,073	6,073	6,073	6,073
R-squared	0.002	0.009	0.002	0.000	0.010	0.026	0.075	0.533	0.045	0.112	0.226	0.114	0.018	0.025
Year FE	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
windows	5	5	5	5	5	5	5	5	5	5	5	5	5	5

Notes: This table reports the change in covariates of parents below and above the mandatory retirement cutoff. Here the parent refers to the ”first-to" for “first" retired parent in the household.

Figure C.2: Validity Test: Adult Children Covariates



Notes: We plot the change in different covariates of adult children below and above parental age centered around the mandatory retirement cutoff. C.2a describes the fraction of female adult children; C.2b and C.2c describe the fraction of individuals who are never married and married respectively; C.2d describes the years of schooling; C.2e describes the number of children in the family; and, C.2f describes the fraction of individuals whose parents are covered by pension.

Table C.3: Covariates Smooth at Age Cutoff: Adult Children

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
		Marital status					N. Children					Parents covered by
Dep. Var	Female	Never Married	Married	Cohabitation	Divorced	Years of Schooling	Total	N. Age < 1	Age 1-3	Age 3-5	N. Age ≥ 6	Pension
Post	0.005	-0.017	0.016	0.002	-0.003	-0.082	-0.036	-0.004	-0.017	-0.039	0.025	-0.001
	(0.015)	(0.010)	(0.010)	(0.004)	(0.004)	(0.158)	(0.041)	(0.020)	(0.037)	(0.029)	(0.036)	(0.013)
Running	0.003	0.001	-0.001	0.001	0.000	-0.004	0.012	0.002	0.007	0.010	-0.006	0.003
	(0.005)	(0.003)	(0.003)	(0.001)	(0.001)	(0.037)	(0.011)	(0.005)	(0.007)	(0.007)	(0.011)	(0.004)
Running x Post	0.00418	0.007	-0.007	-0.001	0.000	-0.087*	-0.007	-0.003	-0.015**	-0.010	0.019	-0.008
	(0.006)	(0.006)	(0.006)	(0.001)	(0.002)	(0.045)	(0.013)	(0.006)	(0.007)	(0.010)	(0.015)	(0.006)
Constant	0.784***	1.578***	-0.548***	0.023***	-0.053***	11.37***	1.024***	0.442***	1.032***	0.577***	-1.026***	0.804***
	(0.048)	(0.090)	(0.092)	(0.007)	(0.008)	(0.784)	(0.087)	(0.039)	(0.074)	(0.111)	(0.188)	(0.027)
Observations	11,194	11,193	11,193	11,193	11,193	10,717	6,897	6,897	6,897	6,897	6,897	8,438
R-squared	0.013	0.281	0.248	0.002	0.012	0.014	0.013	0.031	0.050	0.007	0.136	0.441
windows	5	5	5	5	5	5	5	5	5	5	5	5
year FE	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes

Notes: This table reports the change in covariates of adult children below and above parental age centered around the mandatory retirement cutoff. Here the parent refers to the ”first-to" for “first" retired parent in the household.

BIBLIOGRAPHY

Abramitzky, R. and L. Boustan (2017, December). Immigration in american economic history. *Journal of Economic Literature 55*(4), 1311–45.

Akbar, P. A., V. Couture, G. Duranton, and A. Storeygard (2018, November). Mobility and congestion in urban india. Working Paper 25218, National Bureau of Economic Research.

Akerlof, G. A. (1970, 08). The Market for "Lemons": Quality Uncertainty and the Market Mechanism*. *The Quarterly Journal of Economics 84*(3), 488–500.

Akerlof, G. A. and R. E. Kranton (2000, 08). Economics and Identity*. *The Quarterly Journal of Economics 115*(3), 715–753.

Alesina, A., J. Harnoss, and H. Rapoport (2016, Jun). Birthplace diversity and economic prosperity. *Journal of Economic Growth 21*(2), 101–138.

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.

Antonucci, Toni. C., . J. J. S. (1990). *The role of reciprocity in social support*. In B. R. Sarason, I. G. Sarason, G. R. Pierce (Eds.).

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/.

Arrow, K. J. and G. Debreu (1954). Existence of an equilibrium for a competitive economy. *Econometrica 22*(3), 265–290.

Ashraf, Q. and O. Galor (2011, December). Cultural diversity, geographical isolation, and the origin of the wealth of nations. Working Paper 17640, NBER.

Aslund, O., L. Hensvik, and O. Skans (2014). Seeking similarity: How immigrants and natives manage in the labor market. *Journal of Labor Economics 32*(3), 405–41.

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.

Bai, J. (2018). Melons as lemons: Asymmetric information, consumer learning and quality provision. Working paper.

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., R. Cao, T. Kuchler, J. Stroebel, and A. Wong (2018, August). Social

connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives 32*(3), 259–80.

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.

Baker, S. R. (2018). Debt and the response to household income shocks: Validation and application of linked financial account data. *Journal of Political Economy 126*(4), 1504–1557.

Bandiera, O., I. Barankay, and I. Rasul (2009). Social connections and incentives in the workplace: Evidence from personnel data. *Econometrica 77*(4), 1047–1094.

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.

Battistin, E., A. Brugiavini, E. Rettore, and G. Weber (2009). The retirement consumption puzzle: Evidence from a regression discontinuity approach. *The American Economic Review 99*(5), 2209–2226.

Bayer, P., L. R. Stephen, and G. Topa (2008). Place of work and place of residence: Informal hiring networks and labor market outcomes. *Journal of Political Economy 116*(6), 1150–96.

Beaman, L. (2012). Social networks and the dynamics of labour market outcomes: Evidence from refugees resettled in the u.s. *Review of Economic Studies 79*(1), 128–61.

Behncke, S. (2012). Does retirement trigger ill health? *Health Economics 21*(3), 282–300.

Bernheim, B. D., A. Shleifer, and L. H. Summers (1985). The strategic bequest motive. *Journal of Political Economy 93*(6), 1045–1076.

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.

Bertrand, M., S. Mullainathan, and D. Miller (2003). Public policy and extended families: Evidence from pensions in south africa. *The World Bank Economic Review 17*(1), 27–50.

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.

Bjorkegren, D. and D. Grissen (2018). Behavior revealed in mobile phone usage predicts credit repayment. *World Bank Economic Review*. Accepted.

Blau, F. D. and L. M. Kahn (2017, September). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature 55*(3), 789–865.

Blumenstock, J., G. Chi, and X. Tan (2019, March). Migration and the Value of Social Networks. CEPR Discussion Papers 13611, C.E.P.R. Discussion Papers.

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.

Brandt, M. and C. Deindl (2013). Intergenerational transfers to adult children in europe: Do social policies matter? *Journal of Marriage and Family 75*(1), 235–251.

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.

Budig, M. J. and P. England (2001). The wage penalty for motherhood. *American Sociological Review 66*(2), 204–225.

Burks, S. V., B. Cowgill, M. Hoffman, and M. Housman (2015, Feburary). The value of hiring through employee referrals. *Quarterly Journal of Economics*, 805–839.

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

Cai, H., H. Fang, and L. C. Xu (2011). Eat, drink, firms, government: An investigation of corruption from the entertainment and travel costs of chinese firms. *The Journal of Law and Economics 54*(1), 55–78.

Card, D., A. R. Cardoso, J. Heining, and P. Kline (2018). Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics 36*(S1), S13–S70.

Card, D., C. Dobkin, and N. Maestas (2008, December). The impact of nearly universal insurance coverage on health care utilization: Evidence from medicare. *American Economic Review 98*(5), 2242–58.

Charles, K. K. (2004). Is Retirement Depressing?: Labor Force Inactivity and Psychological Well-Being in Later Life. *Research in Labor Economics 23*, 269–299.

Chen, Y. and X. Zhang (2018). When mommies become nannies: The effects of parental retirement across generations. *SSRN Working Paper*.

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

Cingano, F. and A. Rosolia (2012). People i know: Job search and social networks. *Journal of Labor Economics 30*(2), 291–323.

Coe, N. B. and M. Lindeboom (2008, November). Does Retirement Kill You? Evidence from Early Retirement Windows. IZA Discussion Papers 3817, Institute of Labor Economics (IZA).

Coe, N. B. and G. Zamarro (2011). Retirement effects on health in europe. *Journal of Health Economics 30*(1), 77 – 86.

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.

Cover, T. M. and J. A. Thomas (2006). *Elements of Information Theory 2nd Edition (Wiley Series in Telecommunications and Signal Processing)*. Wiley–Interscience.

Dai, R., D. Mookherjee, K. Munshi, and X. Zhang (2018, August). Community networks and the growth of private enterprise in china. *VoxDev.*.

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.

Donaldson, D. and A. Storeygard (2016, November). The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives 30*(4), 171–98.

Dustmann, C., A. Glitz, U. Schönberg, and H. Brücker (2016). Referral-based Job Search Networks. *The Review of Economic Studies 83*(2), 514–546.

Eagle, N., M. Macy, and R. Claxton (2010). Network diversity and economic development. *Science 328*(5981), 1029–1031.

Ebenstein, A. (2010). The "missing girls" of china and the unintended consequences of the one child policy. *Journal of Human Resources 45*(1), 87–115.

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

Edin, P.-A., P. Fredriksson, and O. Aslund (2003). Ethnic enclaves and the economic success of immigrants: Evidence from a natural experiment. *The Quarterly Journal of Economics 118*(1), 329–357.

Eibich, P. (2015). Understanding the effect of retirement on health: Mechanisms and heterogeneity. *Journal of Health Economics 43*, 1 – 12.

Eibich, P. and T. Siedler (2020). Retirement, intergenerational time transfers, and fertility. *European Economic Review 124*, 103392.

Ettner, S. L. (1995). The impact of "parent care" on female labor supply decisions. *Demography 32*(1), 63–80.

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.

Feng, Z., C. Liu, X. Guan, and V. Mor (2012, December). China's rapidly aging population creates policy challenges in shaping a viable long-term care system. *Health Aff (Millwood) 31*(12), 2764–73.

Fenoll, A. A. (2020, September). The uneven impact of women's retirement on their daughters' employment. *Review of Economics of the Household 18*(3), 795–821.

Fitzpatrick, M. D. and T. J. Moore (2018). The mortality effects of retirement: Evidence from social security eligibility at age 62. *Journal of Public Economics 157*, 121 – 137.

Foley, M. (2014). *JPMorgan Gives Regulators Evidence of Nepotism in China*. Retrieved March 13, 2014, from https://www.cheatsheet.com/money-career.

Fruchterman, T. M. J. and E. M. Reingold (1991, November). Graph drawing by force-directed placement. *Software: Practice and Experience 21*(11), 1129–1164.

Gagnon, J., T. Xenogiani, and C. Xing (2014, Oct). Are migrants discriminated against in chinese urban labour markets? *IZA Journal of Labor & Development 3*(1), 17.

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.

Gee, L., J. Jones, and M. Burke (2017). Social networks and labor markets: How strong ties relate to job finding on facebook's social network. *Journal of Labor Economics 35*(2), 485–518.

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

Giltz, A. (2017). Coworker networks in the labor market. *Labour Economics 44*, 218–230.

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

Glaeser, E. L., S. D. Kominers, M. Luca, and N. Naik (2015, December). Big data and big cities: The promises and limitations of improved measures of urban life. Working Paper 21778, National Bureau of Economic Research.

Gonzalez, M. C., C. A. Hidalgo, and A.-L. Barabasi (2008, June). Understanding individual human mobility patterns. *Nature 453*, 779.

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

Hastings, J. S. and J. M. Weinstein (2008, 11). Information, School Choice, and Academic Achievement: Evidence from Two Experiments*. *The Quarterly Journal of Economics 123*(4), 1373–1414.

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

Hellerstein, J., M. McInerney, and D. Neumark (2011). Neighbors and coworkers: The importance of residential labor market networks. *Journal of Labor Economics 29*(4), 659–95.

Hellerstein, J., M. McInerney, and D. Neumark (2014). Do labor market networks have an important spatial dimension? *Journal of Urban Economics 79*(4), 39–58.

Henderson, J. V., A. Storeygard, and D. N. Weil (2012, April). Measuring economic growth from outer space. *American Economic Review 102*(2), 994–1028.

Hoffman, M. (2017, June). The value of hiring through employee referrals in developed countries. *IZA World of Labor*.

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.

Insler, M. (2014). The health consequences of retirement. *Journal of Human Resources 49*(1), 195–233.

Ioannides, Y. M. and L. D. Loury (2004). Job information networks, neighborhood effects, and inequality. *Journal of Economic Literature 42*, 1056–1093.

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.

Jeffers, J. (2018). The impact of restricting labor mobility on corporate investment and entrepreneurship. Working paper, SSRN.

Jin, G. Z. and P. Leslie (2003). The effect of information on product quality: Evidence from restaurant hygiene grade cards. *The Quarterly Journal of Economics 118*(2), 409–451.

Johnston, D. W. and W.-S. Lee (2009). Retiring to the good life? the short-term effects of retirement on health. *Economics Letters 103*(1), 8 – 11.

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.

Kramarz, F. and O. Skans (2014). When strong ties are strong: Networks and youth labour market entry. *Review of Economic Studies 82*(3), 1164–1200.

Kreindler, G. E. and Y. Miyauchi (2019). Measuring commuting and economic activity inside cities with cell phone records. Working paper.

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.

Lee, D. S. and T. Lemieux (2010, June). Regression discontinuity designs in economics. *Journal of Economic Literature 48*(2), 281–355.

Lei, X., J. Strauss, M. Tian, and Y. Zhao (2015). Living arrangements of the elderly in china: Evidence from the charls national baseline. *China Economic Journal 8*(3), 191–214.

Li, H., X. Shi, and B. Wu (2015). The retirement consumption puzzle in china. *The American Economic Review 105*(5), 437–441.

Li, L. and X. Wu (2011). Gender of children, bargaining power, and intrahousehold resource allocation in china. *The Journal of Human Resources 46*(2), 295–316.

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/.

Mazumder, B. and S. Miller (2016, August). The effects of the massachusetts health reform on household financial distress. *American Economic Journal: Economic Policy 8*(3), 284–313.

Mazzonna, F. and F. Peracchi (2012). Ageing, cognitive abilities and retirement. *European Economic Review 56*(4), 691 – 710.

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

Munshi, K. and M. Rosenzweig (2013). Networks, commitment, and competence: Caste in indian local politics. Working Paper 19197, National Bureau of Economic Research.

Müller, T. and M. Shaikh (2018). Your retirement and my health behavior: Evidence on retirement externalities from a fuzzy regression discontinuity design. *Journal of Health Economics 57*, 45 – 59.

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.

National Bureau of Statistics of China (2014). *The Third National Economic Census Report*. Retrieved December, 2014, from http://www.stats.gov.cn.

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

Ngai, L. R., C. A. Pissarides, and J. Wang (2017). Chinass mobility barriers and employment allocations. *Journal of European Economic Association*. Forthcoming.

Nie, P. and A. Sousa-Poza (August 2017). What chinese worker value: An analysis of job satisfaction, job expectations and labor turnover in china. *IZA Discussion Papers No. 10963 108*.

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.

Ottaviano, G. I. and G. Peri (2006, January). The economic value of cultural diversity: evidence from US cities. *Journal of Economic Geography 6*(1), 9–44.

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.

Page, S. E. (2007). *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies (New Edition)*. Princeton University Press.

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

Recchi, E. (2009). Chapter 4: The social mobility of mobile europeans. Volume 1 of *Pioneers of European integration citizenship and mobility in the EU: Citizenship and mobility in the EU*, pp. 72 – 97. Edward Elgar Publishing.

Rothschild, M. and J. Stiglitz (1976). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *The Quarterly Journal of Economics 90*(4), 629–649.

Saygin, P. O., A. Weber, and M. A. Weynandt (2018). Coworkers, networks and job search outcomes. *ILR Review*. Forthcoming.

Schmutte, I. (2015). Job referral networks and the determination of earnings in local labor markets. *Journal of Labor Economics 33*(1), 1–33.

Schmutte, I. (2016). How do social networks affect labor markets? *IZA World of Labor October*, 1–10.

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

Shigeoka, H. (2014, July). The effect of patient cost sharing on utilization, health, and risk protection. *American Economic Review 104*(7), 2152–84.

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.

Smith, V. K. and F. R. Johnson (1988). How do risk perceptions respond to information? the case of radon. *The Review of Economics and Statistics 70*(1), 1–8.

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics 87*(3), 355–374.

Stancanelli, E. and A. V. Soest (2012). Retirement and home production: A regression discontinuity approach. *The American Economic Review 102*(3), 600–605.

Stancanelli, E. and A. V. Soest (2016, December). Partners' leisure time truly together upon retirement. *IZA Journal of Labor Policy 5*(1), 1–19.

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

Topa, G. (2001). Social interactions, local spillovers and unemployment. *Review of Economic Studies 68*(261-295).

Topa, G. (2011). Labor markets and referrals. *Handbook of Social Economics 1B*, 1193–1221.

United Nations (2019). *World Population Prospects 2019*. Retrieved from https://population.un.org/.

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.

Wang, S. Y. (2013). Marriage networks, nepotism, and labor market outcomes in china. *American Economic Journal: Applied Economics 5*(3), 91–112.

Whalley, J. and S. Zhang (2007). A numerical simulation analysis of (hukou) labour mobility restrictions in china. *83*(2), 392–410.

Wu, B., M. W. Carter, R. T. Goins, and C. Cheng (2005). Emerging services for community-based long-term care in urban china: a systematic analysis of shanghai's community-based agencies. *Journal of Aging and Social Policy 4*(17), 37–60.

Zhang, J. and J. Wu (2018). The chinese labor market, 2000–2016. *IZA World of Labor* (437).

Zhao, Y. (2003). The role of migrant networks in labor migration: The case of china. *Contemporary Economic Policy 21*(4), 500–511.

Zheng, L. (2015). Sibling sex composition, intrahousehold resource allocation, and educational attainment in china. *The Journal of Chinese Sociology 2*(2).

Zhu, H. and A. Walker (2018). Pension system reform in china: Who gets what pensions? *Social Policy & Administration 52*(7), 1410–1424.

Zhu, J. (2013). *Chinese SOEs and "invisible benefits"*. Retrieved May 13, 2013, from https://www.ft.com.