Preliminary Version

Online Learning and the Education Gap:

A Digital Footprint Approach*

Yongheng Deng, Xinfei Huang, Teng Li, Shang-Jin Wei

This Draft: December 2, 2022

Abstract

While online learning has become more common including during the Covid-19 period, its effect on the education gap between rich and poor children is still unclear. Using school closures (compulsory online education) across 21 cities in Southern China during the Covid-19 pandemic as a natural experiment, we study how online learning affects the education gap. We propose a digital footprint approach based on smartphone usage to identify (a) families with children who graduate from a primary school before the summer and move to a middle school in the fall semester in each year during 2019-2021, (b) family economic status based on the housing prices of their residential locations and their shopping behavior. While children from better-to-do families are more likely to go to a top middle school (whose entry depends on an exam) even without school closure, we find that their advantage increases with the duration of school closure. Furthermore, we report evidence that different app usage habits (e.g., the relative time spent on education apps versus games) during school closure are likely to be a contributing factor to the widening educational gap.

Keywords**:** School closures, education gap, family wealth, mobile Apps, time allocation

JEL Classification:

* Deng: University of Wisconsin-Madison, (email: yongheng.deng@wisc.edu); Huang: Sun Yat-sen University, (e-mail: huangxf3@mail.sysu.edu.cn); Li: Sun Yat-sen University, (e-mail: liteng27@mail.sysu.edu.cn); Wei: Columbia University, (e-mail: sw2446@gsb.columbia.edu). We are grateful to … for their helpful comments and … for granting us permission to access the data. Errors are ours.

1. Introduction

Rising inequality hurts economic growth. The gap in educational achievement between the rich and the poor is likely to be an important factor in the inequality pattern(Rajan, 2010). A switch to online learning during the Covid-19 pandemic has raised concerns about declining educational achievement. However, the relative efficacy of online versus offline learning is subject to debate. On the one hand, the increased flexibility and comfort of the home environment associated with online learning could enhance learning outcomes. A meta-analysis by Gao et al. (2022) of 21 randomized control trials (RCTs) of undergraduate medical school students from the United States and other countries[1] plus 6 other non-RCT comparative studies that are deemed of high quality, published during 2003-2018, suggests that the students' learning outcomes are generally better for the treatment (online learning) group than the control (offline) group. They also find that the students in the online groups report a higher level of satisfaction with their learning mode. These studies corroborate the idea that acquisition and retention of a certain knowledge (such as those taught about root canals, ears, noses and throats, and immunization system in a typical medical school course) may be better in an online teaching environment. Since undergraduate medical school students are the subjects of these research papers, we do not know to what extent these conclusions are applicable to younger students in primary, middle, or high schools. Furthermore, none of the studies included in the meta-analysis examines the possible gap in the learning outcomes between well-do-to and less-well-to-do students.

On the other hand, a series of articles in the Economist have argued that the school closure during the Covid-19 pandemic, which is generally accompanied by online learning, has been bad

[1] Typical examples include an RCT comparison of the grades and knowledge retention of of second and third year US medical school students from an online versus classroom teaching of an immunization course (Porter, Pitterle, and Hyney, 2014), an RCT comparison of face-to-face teaching versus e-learning of a course on ears, noses, and throat (ENT) for fourth and fifth year medical students (Alnabelsi, Al-Hussaini, and Ownes, 2015), and an RCT evaluation of knowledge acquisition and retention by Iranian dental medical students from a course on root canals in an online vs classroom setting (Moazami, Bahrampour, and Azar, 2014).

for student learning[2]. While the articles are full of colorful anecdotes and opinions, there is a relative dearth of rigorous evidence. A report by McKinsey & Company (2022) shows some evidence that students in US schools generally fall behind in 2020 after a prolonged period of school closure. Moreover, the lag tends to be bigger for African American and Hispanic students than for whites. While race is correlated with income, the correlation is not perfect, and there is no explicit and rigorous evaluation of how the education gap between the rich and the poor is affected by school closure.

In this paper, we exploit Covid-induced school closure and the mandatory online learning to study the effect of online learning on the gap in education outcomes. We propose a digital footprint approach to overcome a lack of systematic data about both the economic status of the students and the learning outcomes. Specifically, we utilize a unique dataset that contains the universe of de-identified and geocoded cell phone usage information from a major Chinese telecommunication service provider over the course of almost 4 years in Guangdong, the most populous province in China, with a GDP larger than all but the top 12 countries in the world. Leveraging the high-frequency and high-resolution mobile phone usage data, we use the digital footprints to identify (a) families with children who graduate from a primary school before the summer and move to a middle school in the fall semester, (b) family wealth status based on the housing prices of their residence and their shopping behavior. Separately, we use administrative data to identify both the quality of middle schools and duration of Covid-induced school closure.

The quality of middle schools is heterogeneous in a given city. Approximately ¼ of them are designated as "provincial-level key middle schools" reflecting their better quality teachers and facilities. Admission to a top middle school is based on an entrance exam rather than residential

[2] The main conclusions are evident from the titles chosen for the articles: "How Covid-19 Is Interrupting Children's Education – Almost A Billion Children Have Seen Their Schools Close" (Economist, March 19th, 2020), "Covid-19 School Closures Are Widening Europe's Class Divisions" (Economist, February 13th, 2021), "Covid Learning Loss Has Been a Global Disaster" (Economist July 7th, 2022), "Governments Are Ignoring the Pandemic's Disastrous Effect on Education - Neglected Pupils Will Suffer for the Rest of Their Lives" (Economist, July 7th, 2022).

location. One place where the education gap could manifest itself is that the probability of a child attending a top middle school rises with the wealth level of the family. From our data, we document that such probability is indeed a positive function of the family wealth, proxied by the average housing value of a family's residential location. In particular, the probability of entering a top middle school for a child in the top tercile of the housing value distribution is almost xx percentage points higher than her counterpart in the bottom tercile of the home value distribution.

Because the Covid-cases and the associated school closure are highly localized, we explore such variations across the 21 cities in our sample. We find education gap becomes bigger in cities with a longer period of school closure during 2020 and 2021. An increase in the duration of school closure by one standard deviation in the last semester in a primary school raises the probability of entering a top middle school between the top and bottom of the family wealth distribution by an additional 6.3 percentage points.

Finally, we explore the role of differential app usage in the widened education gap. Using parents' monthly mobile App data, we examine whether rich parents have systematically different digital footprints from those of poor families when encountering Covid-induced school closure days in ways that broadly allow children from rich families to have a better chance to enter high-quality schools. We find that compared to poor parents, rich parents increased their time on child education apps and decreased time on leisure apps. These results indicate that different app usage habits of parents are likely to be a contributing factor. [In the next version, we will also examine whether the probability of having a parent accompanying a child studying online rises with family wealth.]

The rest of the paper is structured as follows. Section 2 describes the background and data. Section 3 provides the empirical specifications and presents and interprets the empirical results. Section 4 concludes.

2. Background and Data

2.1 Covid-induced school closure and online education

Under China's Covid-19 control policy, local governments often order school closures when there is a Covid outbreak, sometimes involving a relatively small number of cases. During our sample period, a typical primary school student in Guangdong province experienced 79 days of Covid-induced school closure, with a minimum of xx days, a maximum of xx days, and a standard deviation of 26 days (across different parts of the province). On the days that the schools are closed physically, the students are asked to use a digital device to take classes online.

Besides the classes offered by schools, children can take additional commercial online classes on weekends or evenings. Commercial online education is a profitable and growing business in China, even before but especially since the onset of the Covid-19 pandemic. The online education market has grown to almost US$40 billion in 2020.[3] A typical class involves both the children and the parents: Children use tablets to attend online courses, and parents use the relevant Apps on mobile phones to choose lessons, make payments, and check the children's progress in the course. The most popular education apps are Yuanfudao, Zuoyebang, and Xueersi (all appearing in our data set).

2.2 Data sources

The mobile phone penetration rate is high in China. According to the 2018 China Family Panel Studies, a nationally representative longitudinal survey of individuals' social and economic status, 89% of correspondents sixteen years and older reported possessing at least one cellphone. The

[3] See https://china.usc.edu/online-education-china.

number has likely gone up once the pandemic has started with the ubiquitous requirement of showing a health code app for entering apartment blocks, supermarkets, or shopping malls.

Our mobile app data is from Guangdong Province, which is the largest province in China, with a population of 126.8 million in 2021[4] (greater than Japan, Germany, or the United Kingdom), and a provincial GDP of 1.95 trillion US dollars in 2021[5] (similar to Canada or South Korea, but greater than Australia, Brazil, or Mexico). The province has 21 cities, which include Shenzhen and Guangzhou, among the wealthiest and most economically advanced cities in China.[6] The penetration rate of mobile phones in the province is 95.8% in 2020 (up from 69.2% in 2010). Appendix Figure A.1 shows a strong correlation between the number of China Mobile users and the number of residents by city. Cities with a higher GDP per capita (represented by the size of the circles in Figure A.1) tend to have higher mobile phone ownership.

We have access to detailed app usage data (the time spent on each App in a month) for all of its 71 million users of a leading telecom provider in Guangdong Province from January 2018 to October 2021, accounting for 63% of all mobile users in the province. The user count of 71 million is larger than the population of 164 of the 185 World Bank/IMF member economies. In a research lab which is a subsidiary of the company, we observe users' demographic information, such as age, gender, and the place where the phone number is registered. In addition, we have access to the location records every five-minute interval during the same period.

We randomly extracted 430,874 local users from the database who lived in Guangdong province between January and December 2018.[7] We further restrict the sample to users aged 21 to

[4] https://www.statista.com/statistics/1033846/china-population-of-guangdong/#:~:text=In%202021%2C%20the%20population%20of,all%20the%20provinces%20in%20China.

[5] https://english.news.cn/20220120/5b853b68100b40f2a9c1f4faad40958b/c.html

[6] Cities in Guangdong differ substantially in terms of both population and GDP in 2018, as illustrated in Table A.2.
[7] The user characteristics of our random sample are quantitatively similar to that in the full sample. Table A.3 compares the user characteristics of our random sample to that in the full sample.

60. We exclude users younger than 25 and whose home locations are on university campuses.[8] Our final sample consists of 392,087 local users. As shown in Table 1, 42% of the users are females, and the average age is around 39.6.

2.3 Identifying families with children entering middle schools

China's Compulsory Schooling Laws (CSLs) require all age-eligible children (i.e., 13-15 years) to enroll in a middle school. Based on statistics in 2018 from the National Bureau of Statistics of China, more than 99 percent of age-eligible children participate in middle school education.

Each year, the graduates from primary schools take the city-level middle school entrance exams. The “key middle schools” (a large or medium-sized city typically has multiple key middle schools) would allocate available spots to the students with a high exam score residing within the school district (following the principle of “Division by District and Nearby Admission” proposed by the Compulsory Schooling Law). Students who get high scores on middle school entrance exams could generally choose their preferred middle schools to enter. In contrast, students who performed less well on middle school entrance exams would typically be enrolled in schools in the districts where they reside.[9]

We use the following algorithm to identify individuals whose children were admitted to middle schools during our sample period. First, for all the users in our sample, we use metadata on phone usage, including encrypted IDs of the calling party and the receiving party, the date of calls, and the coordinates of the calling and receiving party’s locations. Based on these data, we identify whether a calling or receiving party is located at middle schools.[10] Next, as the first semester of the first year in Chinese middle schools generally starts on September 1st, we expect that parents

[8] A user’s home location is defined as the location she/he stays most frequently at night. See below for details.
[9] School districts are designed by local administrations.
[10] In our sample, all middle schools are equipped with at least one 2G/3G/4G network. Therefore, we can precisely identify whether a calling or receiving party is located at middle schools.

(whose children were newly admitted into middle schools) would call their children after September but would not make frequent calls to the same location before September. Likewise, before the end date (June 30th) of the last semester in primary schools, we expect that parents would call their children often before June but would not make calls located in primary school areas after September.[11]

As China is the world's largest producer of phones, especially in low-price and mid-price ranges, the rate of phone ownership is high (data?). The low-end phones cost about $9-14 per piece,[12] or about 0.1% of the per capita income. Therefore, most parents, including those from low-income families often give their children a phone. In any case, if there is a greater fraction of poor families with no phones for children, we are likely to underestimate the effect of family wealth on the probability of children entering a top middle school.

We identify parents whose children graduated from a primary school in the spring and move to a middle school in 2019 as those who (i) made no calls to people located in middle school areas from January to June 2019; (ii) made at least one call per month to people located in middle school areas from September to December 2019; (iii) made at least one call per month to people located in primary school areas from January to June 2019; and (iv) made no calls to people located in primary school areas from September to December 2019. We select similar sets of parents in 2020 and 2021.

With this set of criteria, we identify 12,406 individuals who are likely parents with a child that has moved from a primary school in the spring to a middle school in the fall. They account for

[11] According to a 2020 report by China Internet Network Information Center, 92% of the teenagers in China own a mobile phone (https://pic.cyol.com/img/20210720/img_960114c132531c521023e29b6c223e438461.pdf). Additionally, according to the survey by Guangzhou Consumers Commission in 2015, more than 90% students in a middle school own mobile phone devices. Since Guangdong is one of the richest provinces, the student ownership of mobile phones is likely to be higher than the national average.

[12] https://detail.zol.com.cn/cell_phone_index/subcate57_0_list_1_0_3_2_0_1.html

2.1% of our sample.[13] Their average age is 43.3, ranging from 35 to 49. Specifically, there were 3,925 (4,354 and 4,127) individuals whose children entered middle schools in September 2019 (2020 and 2021). Additionally, 2,605 individuals' children entered *top* middle schools, occupying 21.0% of our sample. This ratio is comparable to the proportion of key middle schools in Guangdong province (24.2%).

For parents identified this way, we make an educated guess of the phone users who are adult family members (mostly spouses): (i) they made at least one call per month to some individual that we identify above during our sample period; (ii) their age difference is fewer than ten, and they have opposite sex; and (iii) they made at least one call per month to the individual's child during our sample period. Based on the criterion, 8,158 of the 12,406 individuals are matched to their husband or wife.[14] Our final sample contains 8,471 mothers and 9,297 fathers. 12,406 families identify at least one parent, and 8,158 families identify both parents. Using this sample, we examine how parents' time allocation among mobile App categories affects their children's probability of key middle school admission.

2.4 Quality of Middle Schools

Out of the 928 middle schools (also known as "junior high schools") in Guangdong province, 225 are designated as "provincial-level key middle schools" based on a combination of absolute quality and relative quality when compared with other schools in the same city.[15] The list of key and non-key middle schools comes from the Bureau of Education of Guangdong Province. Figure A.6 displays the geographical distribution of all the schools. In almost all cities, there is at least one

[13] According to the data in the China Family Panel Survey, the proportion of families with child entering middle school in 2018 in Guangdong province is around 3%, only slightly larger than that in our sample.
[14] The main reason for individuals that are not successfully matched is that their husband or wife use other telecom providers.
[15] Designated key schools are schools distinguished from ordinary schools by their academic reputation and are generally allocated more resources by the state. See for Ye (2015) details.

key middle school. The only exception is the city of Heyuan, which we drop from our sample. In Appendix xx, we regress the number of key middle schools in logarithm against log population, log per capita income, and their second-order terms across all cities. We find that the log number of key middle schools rises linearly with the log local population. The coefficient 0.86 is not statically different from 1. None of the other terms is significant.

2.5 Approximating the Rank of Family Wealth

For a given family with such a student, we use the digital footprint to estimate the rank of family wealth. Our operating assumption is that a family's total wealth and the value of their home are highly positively correlated. Our first proxy is the average housing prices in 2018 of the neighborhood (or cell tower area) where parents and their children live. China's housing market has experienced phenomenal growth since the early 2000s (Liu and Xiong, 2020). According to the National Bureau of Statistics, the average housing price in China has increased by more than 200% from 2000 to 2015. Even in real terms, China's house prices rose by more than 10% annually (Glaeser et al., 2017). Also, housing occupies a significant proportion of China's typical household's wealth portfolio. The homeownership rate is beyond 80% in urban China (Gan et al., 2013). Moreover, housing equity accounts for two-thirds of a typical Chinese household's wealth (China Household Finance Survey, 2018).

Recent developments and the widespread diffusion of geospatial data acquisition technologies have enabled the creation of highly accurate spatial and temporal data (Gonzalez et al., 2008). We define users' home location as the location where a user spends at least 5 hours a day between 10 pm and 7 am for at least fifteen workdays in a given month and live for at least ten months in 2018. Note that the passive collection of geolocation information – which underlies our data collection procedure – works on all traditional mobile networks (2G, 3G, or 4G). The average coverage radius of 4G networks is 0.2-1.2 km. Typically, one neighborhood (or single office building) is equipped with one or two 4G networks. Next, we use coordinates information of each neighborhood to match the average housing prices in 2018 that are derived from Soufun.com. Soufun is a major online real estate brokerage intermediary and rental service provider in China that collects housing listing and transaction information for residential properties (Deng et al., 2015). Table 1 shows that the average listed housing prices were about 2.66 million RMB (US$0.43 million) in 2018. Figure A.2 displays the distribution of housing prices. The median housing prices are about 1.56 million RMB (US$0.25 million), with an interquartile range of 2.23 million RMB.

An important caveat is that we cannot distinguish between homeowners and renters from our phone usage data. Even though the homeownership rate is high (over 80%) in China, some renters may be mislabeled as "rich" property owners as they reside in rental units in expensive neighborhoods. However, since rent and home values are strongly positively correlated, those families that are renters in a high-priced residential area are likely to have a relatively high level of family wealth as well. As a second measure of family wealth, we compute the frequency of visits to high-end shopping centers, representing consumption behaviors. High-end shopping centers are defined as those having high-end apparel brands.[16] Alternatively, we use high-end local jewelry brands to define high-end shopping centers. More than 95% of high-end shopping centers using the two definitions are overlapped.[17] We identify 1,128 high-end shopping centers across Guangdong province and derive their coordinates from Google Maps. Figure A.3 displays the geographical distribution of high-end shopping centers. Based on location records of every 5-

[16] According to BrandFinance, we select the top 10 sales apparel brands in 2018, including Nike, H&M, Zara, Adidas, Louis Vuitton, Cartier, Hermes, Gucci, Uniqlo, Rolex. See https://brandirectory.com/rankings/apparel/2018/table.
[17] The top 10 local jewellery brands in 2018, including Chow Tai Fook, Chow Sang Sang, Lao Feng Xiang, Lukfook, TSL, CHJ, Chow Tai Seng, Tesiro, I DO, Keerworld. See https://brandirectory.com/rankings/apparel/2018/table.

minute intervals, we define someone as a customer at a shopping center if they have stayed there for more than an hour but fewer than 4 hours.[18]

Panel (A) of Figure A.4 shows the distribution of average monthly visits to high-end shopping centers in 2018. On average, a typical person visits high-end shopping centers 2.2 times per month. Panel B of Figure A.4 presents the correlation between housing prices (property values) and monthly visits to high-end shopping centers. Reassuringly, the higher the property values, the more frequently people visit high-end shopping centers. This evidence demonstrates that property values and frequency of visits to high-end shopping centers are reasonable measures of individual wealth.

As a robustness test of our main analysis, we exclude individuals who display mixed signals on wealth measures, e.g., high property values but low frequency of visits to high-end shopping centers and low property values but high frequency of visits to high-end shopping centers.

To further validate the quality of the wealth proxies, we provide another three measures, i.e., car owners, number of phones, and phone prices. First, we define car owners by whether users installed an administrative App called "*Jiaoguan 12123*" in 2018. This platform governs issues, including applying for driving licenses, paying fines, and other driving-related matters in China. Therefore, with this mobile App installed, users are likely to be car owners. As shown in Table 1, 7% of users installed this App in our sample, comparable to the administrative statistics (11.3%, Guangdong Statistical Yearbook 2019). Second, we have information on the number of phones and phone brands for all users. Based on the prices of each phone brand and model in 2018, we divide phone prices into seven categories, i.e., below 1,000, 1,000-2,000, 2,000-3,000, 3,000-4,000, 4,000-5,000, 6,000-7,000, and above 7,000 RMB. Table 1 shows that a typical user has one mobile phone whose price fell at 2,000-3,000 RMB. Next, we examine the pairwise correlations between

[18] People who stay at shopping centers for fewer than an hour are likely to be food delivery workers. Likewise, people who stay there for more than 4 hours are likely to be workers in shopping centers.

housing prices, shopping center visits, and the other three wealth measures. Table A.4 shows that property values and shopping center visits are significantly positively correlated with the other three wealth measures. This suggests that property values and shopping center visits are reasonable proxies for wealth.

2.6 App usage data

Smartphones have become ubiquitous and integrated into people's daily lives. In 2021, people worldwide consumed more than 3 hours daily on their mobile Apps, according to App monitoring firm App Annie. China has the world's biggest population of smartphone users, rising from less than 100 million in 2010 to around 953 million in 2020.[19] Moreover, the average daily time spent on mobile Apps surged from 22 minutes in 2011 to more than 3 hours in 2021.[20] According to a German internet company survey, smartphone time by Chinese nationals ranked second in the world after Brazilians.[21]

According to the Ministry of Industry and Information Technology, more than 3 million mobile Apps are available in the Android or Apple App market. All Apps can be sorted into eight major categories: social media, news, game, video, shopping, child education, finance, and others. The most popular mobile Apps in 2021 China are led by WeChat (a social media super App similar to Facebook, Snapchat, Twitter, and Apple Pay combined), Taobao (a leading online shopping App similar to eBay), and Alipay (a leading digital payment App that also comes with a convenient money market fund and has a supermarket for mutual fund investments).[22]

The data provide monthly time spent on each mobile App for each individual in our sample from January 2018 to October 2021. One point worth noting is that the data only include the Apps

[19] See https://newzoo.com/insights/rankings/top-countries-by-smartphone-penetration-and-users/.
[20] See https://www.statista.com/statistics/467518/china-average-daily-time-spent-using-smartphones/.
[21] See https://www.chinadaily.com.cn/china/2017-06/28/content_29916889.htm.
[22] See https://www.chinainternetwatch.com/30778/top-mobile-apps/. WeChat is a Chinese social media App developed by Tencent. Taobao and Alipay are Chinese online shopping and finance Apps developed by Alibaba.

whose users are beyond 500 in Guangdong province during our sample period. For the purpose of analysis, we aggregate time spent on all mobile Apps into 8 categories, including child education, games, video, social media, and others.[23] Additionally, for individuals with multiple phones, we aggregate their App usage at the individual level by their unique identity numbers. When individuals travel out of Guangdong province, our data have no mobile App usage records until they return. The average sample length for all individuals in our sample is around 44.2 months. In other words, each individual on average visits outside Guangdong province for 1.8 months during our sample period.[24]

Table 2 summarizes the monthly total time (hours) spent on each App category of all parents during our sample period. Panel A shows that a typical parent spends about 93.7 hours per month (about 3 hours per day) on all mobile Apps, including 3.1 hours on child education, 14.7 hours on games, 6.4 hours on video, 59.1 hours on social media, and 10.4 hours on the rest.[25] Panel B indicates that mothers spend more time on child education Apps, while fathers consume more time on games and social media Apps. Panel C divides parents into three wealth categories, i.e., housing prices below 1, 1-3, and above 3 million RMB.[26] It demonstrates that the time spent on all App categories monotonically increases with housing wealth.

3. Results

We will document the education gap between rich and poor children in terms of their chance of attending a top middle school. We will then examine whether school closures and online leaning

[23] The classification of mobile Apps follows those provided by Android and Apple App store. Child-education Apps refer to early childhood and K-12 education Apps. The "others" category includes finance, music, photos, tools, travel, health, and food.
[24] As a robustness test, we drop the users who stayed in Gaungdong province fewer than 40 months during our sample period. Our main results remain stable. See section xx for details.
[25] These statistics are consistent with those in other reports. For instance, according to the report by CIW team, in 2021, Chinese adults spend an average of 3 hours and 16 minutes a day checking their smartphones, excluding phone calls (See https://www.chinainternetwatch.com/32047/media-time-spend/).
[26] These cutoffs roughly correspond to 50th and 80th quantiles.

contribute to a widening of the gap. We will explore the role of differential app usage in contributing to the widening of the gap.

3.1 Education gap between the rich and the poor

We use admission to key middle schools to measure academic performance. Using the algorithm documented in section 2.2, we identified 12,406 individuals with children entering middle schools during our sample period. Specifically, there were 3,925 (4,354 and 4,127) individuals whose children joined middle schools in September 2019 (2020 and 2021). 2,605 individuals' children were admitted into *key* middle schools.

We use the following empirical framework to examine the correlation between family wealth and children's educational outcomes:

$$E_{ict} = \alpha H_{i,2018} + \delta_c + \eta_t + \varepsilon_{ict}, \qquad (1)$$

where *c* denotes the city where the mobile user's property is located in 2018, and *t* is the year. E_{ict} is an indicator that equals one if individual i's children were admitted into key middle schools. $H_{i,2018}$ is individual i's housing prices in 2018. δ_c and η_t denote city and year fixed effects, respectively. ε_{ict} represents the idiosyncratic error term. We use the wild-bootstrapping method (Wu, 1986; Davison and Flachaire, 2008) to cluster standard errors at the city level.

Column (1) in Table 3 demonstrates that housing prices positively affect the probability of key middle school admission. An increase in home value by one million RMBs increases the likelihood of top middle school admission by 0.7 percent. In column (2), we divide housing prices into five groups, i.e., below 0.5, 0.5-1, 1-3, 3-6, and above 6 million RMB. Using below 0.5 million RMB as the reference group, the estimates in column (2) display a monotonically increasing pattern. The education gap between rich and poor families is striking. Compared to the poorest group, the probability of key middle school admission of children from the top wealth group increases by almost 50 percentage points.

3.2 School closures and the education gap

We now examine whether Covid-induced school closures exacerbate the education gap between rich and poor children. The specification is as follows:

$$E_{ict} = \alpha_1 H_{i,2018} \times SC_{ict} + \alpha_2 SC_{it} + \alpha_3 H_{i,2018} + \delta_c + \eta_t + \varepsilon_{ict}, \qquad (2)$$

where SC_{ict} denotes the number of days of Covid-induced school closures during the last semester in primary schools (from January 1st to June 30th) in year *t* and city c. Other notations are the same as those in Equation (1). The coefficient α_1 measures how the effect of school closure days during the last semester in primary schools on the probability of entering top middle schools varies by family wealth.

The results are summarized in columns (3) to (5) in Table 3. First, in column (3), the number of Covid-induced school closure days in the last semester in primary schools does not affect the overall probability of key middle school admission. This is reasonable because even if all students perform worse due to Covid-induced school closure, top middle schools would still admit relatively better students. In column (4), the coefficient of the interaction term is statistically significant at the 1% level. It demonstrates that conditional on the number of school closure days in the last semester in primary schools, the wealthier families, the higher the probability of their children being admitted into key middle schools.

The housing prices for the top and bottom 10% of families in our sample are 5.3 and 0.5 million RMB, respectively. The coefficient in column (4) indicates that a standard deviation increase in school closure days (26.1) in the last semester in primary schools enlarges the probability gap between the richest and the poorest families by 6.3 percentage points. This effect is both statistically and economically significant.

Considering the potential nonlinearity of the heterogeneous wealth effect of school closures, column (5) in Table 3 replaces the linear measure of housing wealth with a set of indicators. These estimates demonstrate that compared to the poorest group (below 0.5 million RMB), the probability of key middle school admission of children from the top wealth group (above 6 million RMB) increases by 47.9 percentage points. More importantly, the coefficients on interaction terms suggest that a 10-day increase in school closure days would enlarge the admission probability gap by 0.7 percentage points. Since 2020, the average number of school closure days in the last semester of primary schools reached 78.6. That is to say, the school closure policy in Guangdong province may enlarge the admission probability gap between the poorest and the richest families by 5.5 percentage points, around 11.5% of the initial probability gap.

3.3 The role of Digital App Behavior

Next, we examine whether rich parents have systematically different digital footprints from those of poor families when encountering Covid-induced school closure days in ways that broadly allow children from rich families to have a better chance to enter high-quality schools. We restrict our sample to each parent during the last semester for their children in primary schools (from January 1st to June 30th of each year). The specification is as follows:

$$P_{ict} = \beta_1 H_{i,2018} \times SC_{ict} + \beta_2 SC_{it} + \beta_3 H_{i,2018} + \theta_i + \lambda_t + \varepsilon_{ict}, \quad (3)$$

where P_{ict} is the hours spent on mobile Apps in month t. SC_{ict} denotes the number of days of Covid-induced school closures in each month t. θ_i represents individual fixed effect, addressing the concern that the level of time spent on specific mobile app categories may be correlated with demographic characteristics (e.g., education). λ_t denotes year-by-month fixed effects, controlling the year-specific seasonal shocks that are common to all individuals. Other notations are the same as those in Equation (1). The coefficient β_1 measures the wealth heterogeneous effect of school closure days during the last semester in primary schools on time allocation among mobile Apps.

Panel (A) of Table 4 reveals how the App usage patterns of mothers changed when experiencing school closure days. The estimates demonstrate that conditional on the number of school closure days, the higher housing prices, the more time mothers spend on child education Apps. Specifically, the sign of the coefficient on *number of school closure days* indicates that the school closure effect on time spent on child education Apps is positive for mothers whose housing prices were zero in 2018. On top of that, the coefficient on the interaction term is statistically significant and positive. It demonstrates that the wealthier mothers, the larger the school closure effect on time spent on child education Apps. Specifically, 10 more school closure days in one month widen the education App time gap between mothers from the richest (top 10%) and the poorest (bottom 10%) families by 1.8 hours, approximately 45% of the sample mean. However, there are no significant differences between rich and poor mothers in the use of other App categories.

In contrast, panel (B) conducts the same analysis for the father sample. The results indicate that conditional on the number of school closure days, the higher housing prices, the more time fathers spend on child education Apps. Specifically, 10 more school closure days in one month widen the education App time gap between fathers from the richest (top 10%) and the poorest (bottom 10%) families by 0.3 hours, approximately 13% of the sample mean. The difference is smaller than that in the mother sample. Another interesting finding is that during school closure days fathers in rich families also reduced time spent on mobile leisure activities, including games and social media Apps. By doing this, fathers may set an example for their children and help them focus on study.

These results suggest that the difference in digital footprints between the rich and the poor may play a critical role in explaining the wealth-educational outcome relationship.

4. Extensions and Additional Sensitivity Checks (to do)

In the next few months, we will investigate the following:

a. Sensitivity to ways of identifying families with children moving from a primary to a middle school

b. Sensitivity checks on measures of wealth

c. Check for cumulative effects of school closure over two years (2020-2021) the graduating cohort of 2021

d. Possible non-linear effects (in duration of school closure and in wealth)

e. Check whether likelihood of having parental company with the child during online lessons rises with family income

5. Concluding Remarks

There is no definitive conclusion in the literature on the effect of online learning on learning outcomes. We study how prolonged online learning, triggered by Covid-19 pandemic-related school closures, affects the educational gap between rich and poor children. Our methodological innovation is to use a digital footprint approach – location and usage of smartphones – to identify education outcomes (i.e., children that graduate from a primary school in early summer and enter a middle school in the fall) and the rank of family wealth.

While the probability of attending a top middle school is always higher for children from better-to-do families without school closure, we find the education gap – measured by the difference in the probability of attending a top middle school between rich and poor children – widens as the number of school closure days increases. As the length of the school closure in a city is driven by Covid infection cases in a city, which is arguably random and outside the control of the local population, the result can be read as causal.

References

Alnabelsi, T., A. Al-Hussaini, and D. Owens, 2015, “Comparison of Traditional Face-to-Face Teaching with Sunchronous e-Learning in Otolaryngolory Emergencies Teaching to Medical Undergraduates: A Randomized Controlled Trial,” *European Archives of Oto-Rhino-Laryngology*, 272: 759-763.

Aguiar, M., Bils, M., Charles, K. K., & Hurst, E. (2021). Leisure luxuries and the labor supply of young men. *Journal of Political Economy*, 129(2), 337-382.

Angrist, J., & Kolesár, M. (2021). One instrument to rule them all: The bias and coverage of just-ID IV (No. w29417). National Bureau of Economic Research.

Chor, D., & Li, B. (2021). Illuminating the Effects of the US-China Tariff War on China’s Economy (No. w29349). National Bureau of Economic Research.

Connolly, M. (2008). Here comes the rain again: Weather and the intertemporal substitution of leisure. *Journal of Labor Economics*, 26(1), 73-100.

Deaton, A., (1997). The analysis of household surveys: A microeconometric approach to development policy. John Hopkins University Press: Baltimore and London.

Deng, Y., Morck, R., Wu, J., & Yeung, B. (2015). China’s pseudo-monetary policy. *Review of Finance*, 19(1), 55-93.

Deschenes, O., & Moretti, E. (2009). Extreme weather events, mortality, and migration. *The Review of Economics and Statistics*, 91(4), 659-681.

Gan, L., Z. Yin, N. Jia, S. Xu, S. Ma, & L. Zheng. (2013). Data You Need to Know About China: Research Report of China Household Finance Survey 2012, Springer Berlin Heidelberg.

Gao, M., Y. Cui, H. Chen, H. Zeng, Z. Zhu, and X. Zu, 2022, “The Efficacy and Acceptance of Online Learning vs. Offline Learning in Medical Student Education: A Systematic Review and Meta-Analysis,” *Journal of Xiangya Medicine*, 7:13.

Glaeser, E., Huang, W., Ma, Y., & Shleifer, A. (2017). A real estate boom with Chinese characteristics. *Journal of Economic Perspectives*, 31(1), 93-116.

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A. L. (2008). Understanding individual human mobility patterns. *Nature*, 453(7196), 779-782.

Graff Zivin, J., & Neidell, M. (2014). Temperature and the allocation of time: Implications for climate change. *Journal of Labor Economics*, 32(1), 1-26.

Hong, C. Y., Lu, X., & Pan, J. (2020). Fintech adoption and household risk-taking (No. w28063). National Bureau of Economic Research.

Keen, M. (1986). Zero expenditures and the estimation of Engel curves. *Journal of Applied Econometrics*, 1(3), 277-286.

Liu, C. & Xiong, W. (2020). China's Real Estate Market. The Handbook of China's Financial System, Princeton: Princeton University Press, 183-207.

Moazami, F., E. Bahrampour, and M. Azar, 2014, "Comparing Two Methods of Education (Virtual versus Traditional) on Learning of Iranian Medical Students: A Post-Test Only Design Study," BMC Medical Education, 14:45.

Porter A, M. Pitterle, and M. Haynes, 2014, "Comparison of Online versus Classroom Delivery of an Immunization Elective Course," *American Journal of Pharmaceutical Education*, 78:96

Rajan, Raghuram G. 2010, *Fault Lines: How Hidden Fractures Still Threaten the World Economy*, Princeton University Press.

Sigmundová, D., Sigmund, E., Badura, P., Vokáčová, J., Trhlíková, L., & Bucksch, J. (2016). Weekday-weekend patterns of physical activity and screen time in parents and their pre-schoolers. *BMC Public Health*, 16(1), 1-9.

Xiang, M., Zhang, Z., & Kuwahara, K. (2020). Impact of COVID-19 pandemic on children and adolescents' lifestyle behavior larger than expected. *Progress in Cardiovascular Diseases*, 63(4), 531.

Ye, H. (2015). Key-point schools and entry into tertiary education in China. *Chinese Sociological Review*, 47(2), 128-153.

Table 1. Family characteristics of our analysis sample

Variables	Observations	Mean	Std. Dev.	Minimum	Maximum
Average parent age (years)	12,406	39.16	12.40	21	60
Housing prices (million RMB)	12,406	2.69	5.79	0.1	111.67
Monthly high-end mall visits	12,406	2.12	1.96	0	8
Car owner (=1)	12,406	0.07	0.26	0	1
Average number of phones of parents	12,406	1.06	0.24	1	4
Average phone price category of parents	12,406	2.72	1.39	1	7

Notes: This table provides descriptive statistics of family characteristics for our analysis sample. Each observation denotes a unique family. *Average parent age* is the age in years of the average parent in each family. *Housing prices* are the average listed housing prices of the neighborhood in 2018 where families lived at. *Monthly high-end mall visits* denotes the monthly number of visits to high-end shopping centers of both parents. *Car owner* is a dummy variable equal to 1 if any parents installed the "Jiaoguan 12333" App in 2018, zero otherwise. *Number of phones* is the number of phones that are actively used simultaneously in 2018 of an average parent in one family. Phone prices are sorted into 7 categories, i.e., below 1,000, 1,000-2,000, 2,000-3,000, 3,000-4,000, 4,000-5,000, 6,000-7,000, and above 7,000 RMB, respectively.

Table 2. Monthly mobile time (hours) allocation by parent characteristics

	Panel A	**Panel B: By gender**		**Panel C: By housing prices (1 million RMB)**		
	Full sample	Mothers	Fathers	Below 1	1-3	Above 3
Observations	568,571	314,441	254,130	260,895	224,650	83,026
No. of users	17,768	8,471	9,297	8,692	7,019	2,057
Mobile time	93.7	88.4	96.5	87.5	93.3	104.9
Child-education	3.1	4.2	2.2	2.5	3.0	4.0
	(70.4)	(85.8)	(55.8)	(66.0)	(67.8)	(82.8)
Game	14.7	13.2	15.8	14.1	14.7	15.7
	(53.5)	(50.5)	(55.7)	(54.6)	(53.5)	(51.7)
Video	6.4	6.3	6.4	5.8	6.3	7.4
	(21.9)	(22.0)	(21.8)	(20.9)	(21.9)	(23.8)
Social media	59.1	54.0	63.1	55.6	59.2	66.0
	(61.5)	(59.2)	(63.2)	(59.6)	(61.7)	(64.4)
Others	10.4	10.7	9.0	9.5	10.1	11.8
	(10.6)	(7.6)	(12.4)	(10.4)	(10.4)	(11.0)

Notes: This table summarizes monthly mobile time spent on apps per category by parent characteristics. The classification of mobile apps is the same as those provided by Android and Apple app stores. Child education apps refer to early childhood and K-12 education apps. The "others" category includes finance, shopping, music, photos, tools, travel, health, food, and unclassified apps. Panel A shows the statistics of our analysis sample. Panel B compares installed apps by gender. Panel C displays the pattern by housing wealth in 2018. Standard deviations are in parentheses.

Table 3. Effect of Covid-induced school closure days on the probability of key middle school admission by family wealth

	Entering key middle schools (=1)				
	(1)	(2)	(3)	(4)	(5)
Housing prices (1 million RMB)	0.008***		0.009***	0.007***	
	(0.001)		(0.001)	(0.001)	
Housing prices (0.5-1 million RMB)		0.101***			0.091***
		(0.010)			(0.025)
Housing prices (1-3 million RMB)		0.267***			0.234***
		(0.012)			(0.035)
Housing prices (3-6 million RMB)		0.391***			0.327***
		(0.014)			(0.052)
Housing prices (6 million RMB above)		0.504***			0.479***
		(0.016)			(0.068)
No. of school-closure days (10 days)			0.006	-0.009***	-0.003***
			(0.012)	(0.002)	(0.001)
No. of school closure days (10 days)				0.005***	
* Housing prices (1 million RMB)				(0.001)	
No. of school closure days (10 days)					-0.000
* Housing prices (0.5-1 million RMB)					(0.001)
No. of school closure days (10 days)					0.001
* Housing prices (1-3 million RMB)					(0.001)
No. of school closure days (10 days)					0.003**
* Housing prices (3-6 million RMB)					(0.001)
No. of school closure days (10 days)					0.004***
* Housing prices (6 million RMB above)					(0.001)
Sample mean	0.210	0.210	0.210	0.210	0.210
City and year FEs	√	√	√	√	√
Observations	12,406	12,406	12,406	12,406	12,406

Notes: This table presents the effect of school closure days on the probability of key middle school admission by family wealth. Each observation represents one family. The dependent variable is an indicator that equals one if individuals satisfy the selection criterion for key middle schools. *No. of school closure days* refers to the number of Covid-induced days in the six months before the middle school entrance examination. All coefficients are OLS estimates. In all regressions, we control the city- and year-fixed effects. All standard errors are clustered at the city level. Standard errors are reported in parentheses under the coefficient estimates, and ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels.

Table 4. Effect of Covid-induced school closure days on App usage

	(1) Child education Apps	(2) Game Apps	(3) Video Apps	(4) Social media Apps	(5) Other Apps
Panel A: Mother sample					
No. of school closure days (10 days)	0.15**	0.04	0.10	0.04	0.07
	(0.07)	(0.05)	(0.11)	(0.05)	(0.06)
No. of school closure days (10 days)	0.39***	-0.12	0.03	-0.12	0.06
* Housing prices (1 million RMB)	(0.12)	(0.10)	(0.07)	(0.10)	(0.06)
Sample mean	4.0	13.1	6.4	54.3	11.2
Individual and year-month FEs	√	√	√	√	√
Observations	138,752	138,752	138,752	138,752	138,752
Panel B: Father sample					
No. of school closure days (10 days)	0.04	0.26***	0.19	0.26***	0.19**
	(0.07)	(0.09)	(0.14)	(0.09)	(0.09)
No. of school closure days (10 days)	0.06*	-0.12**	0.09	-0.12**	0.05
* Housing prices (1 million RMB)	(0.04)	(0.06)	(0.07)	(0.06)	(0.05)
Sample mean	2.3	16.1	6.4	64.2	8.8
Individual and year-month FEs	√	√	√	√	√
Observations	135,219	135,219	135,219	135,219	135,219

Notes: This table presents the effect of school closure days on App usage by mothers and fathers whose children were in the last semester of primary school, respectively. Each observation denotes a parent-year-month cell. All coefficients are OLS estimates. In all regressions, we control individual and the year-month fixed effects. All standard errors are clustered at the city level. Standard errors are reported in parentheses under the coefficient estimates, and ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels.

Internet Appendix for

“Online learning and the education gap: a digital footprint approach”

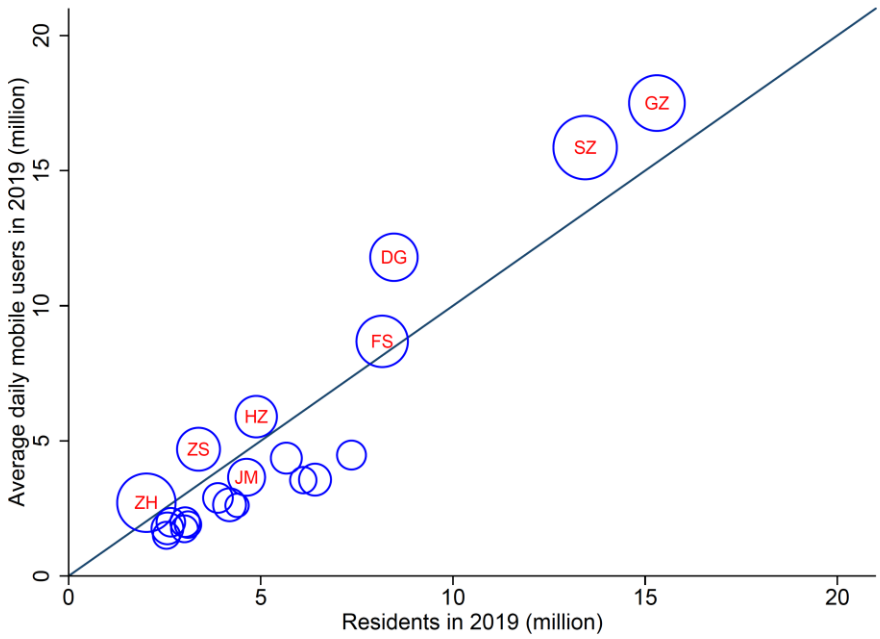
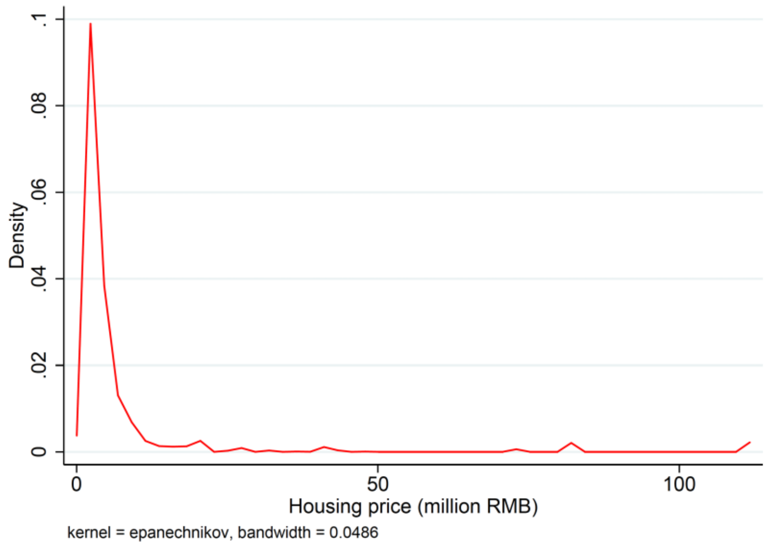
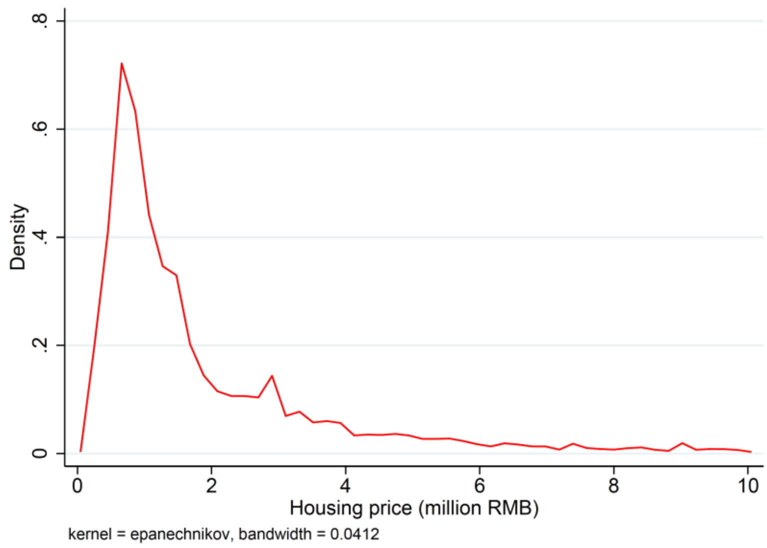


Figure A.1. Mobile users vs. residents in 2019

Notes: This graph presents the relationship between average daily mobile users and residents of cities in Guangdong in 2019. The solid line is a 45-degree line. The size of each marker denotes GDP per capita in 2019. We label the city names for those whose GDP per capita in 2019 is above 60,000 RMB (around 8,500 USD). The abbreviated city names are listed in Table A.2.



(A) Full sample



(B) Housing prices below 10 million RMB

Figure A.2. Distribution of housing prices of users' home locations

Notes: Panel (A) in this graph displays the distribution of housing prices of users' home locations in 2018. Panel (B) shows a censored distribution in which housing prices are below 10 million RMB in 2018.

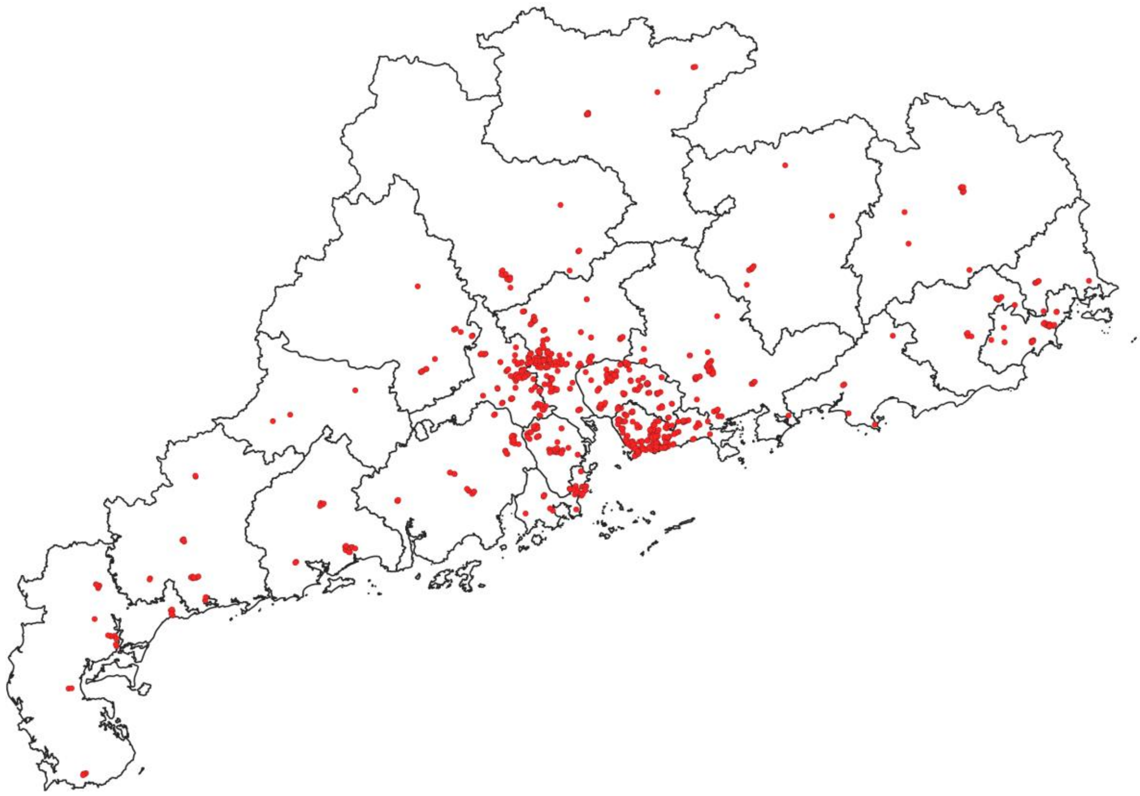
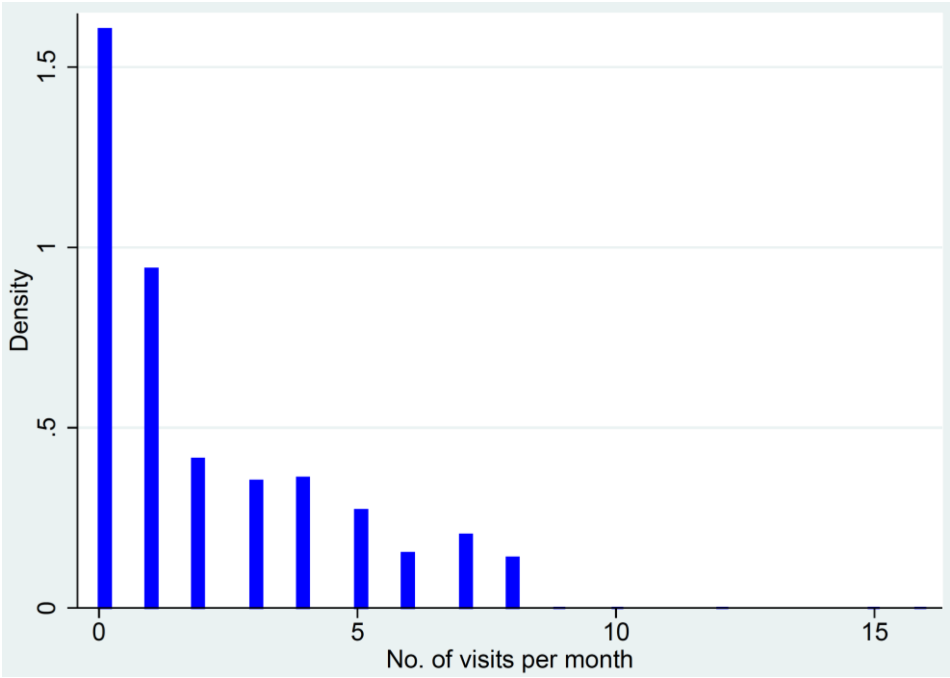
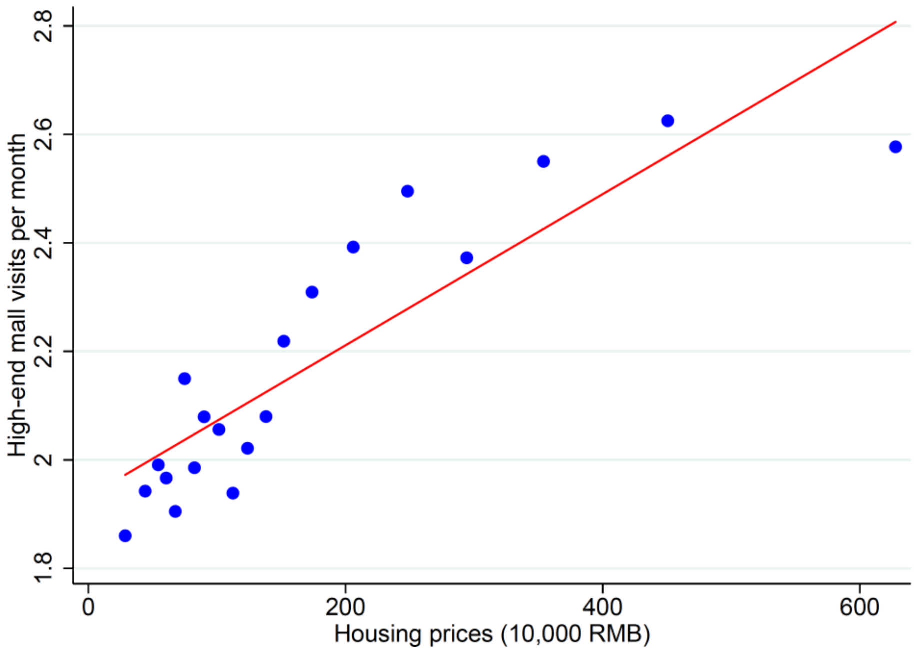


Figure A.3. Geographical distribution of high-end shopping centers in Guangdong

Notes: This figure displays the geographical distribution of 1,128 high-end shopping centers in Guangdong. High-end shopping centers are defined as those having high-end apparel brands. According to BrandFinance, we select the top 10 sales apparel brands in 2018, including Nike, H&M, Zara, Adidas, Louis Vuitton, Cartier, Hermes, Gucci, Uniqlo, and Rolex. See https://brandirectory.com/rankings/apparel/2018/table.



(A) Distribution of monthly visits to high-end shopping centers



(B) Correlation between housing prices and monthly visits to high-end shopping centers

Figure A.4. Distribution of visits to high-end shopping centers and its correlation with housing prices

Notes: Panel (A) shows the distribution of average monthly visits to high-end shopping centers in 2018. On average, a typical person visits high-end shopping centers 2.2 times per month in 2018. Panel (B) presents a bin scatter plot that shows the correlation between housing prices and monthly visits to high-end shopping centers in 2018. We divide the x-variable into twenty equal-sized groups and plot the means of the y-variable within each bin against the mean value of the x-variable within each bin. The solid red line shows the linear fit estimated on the underlying microdata using OLS.

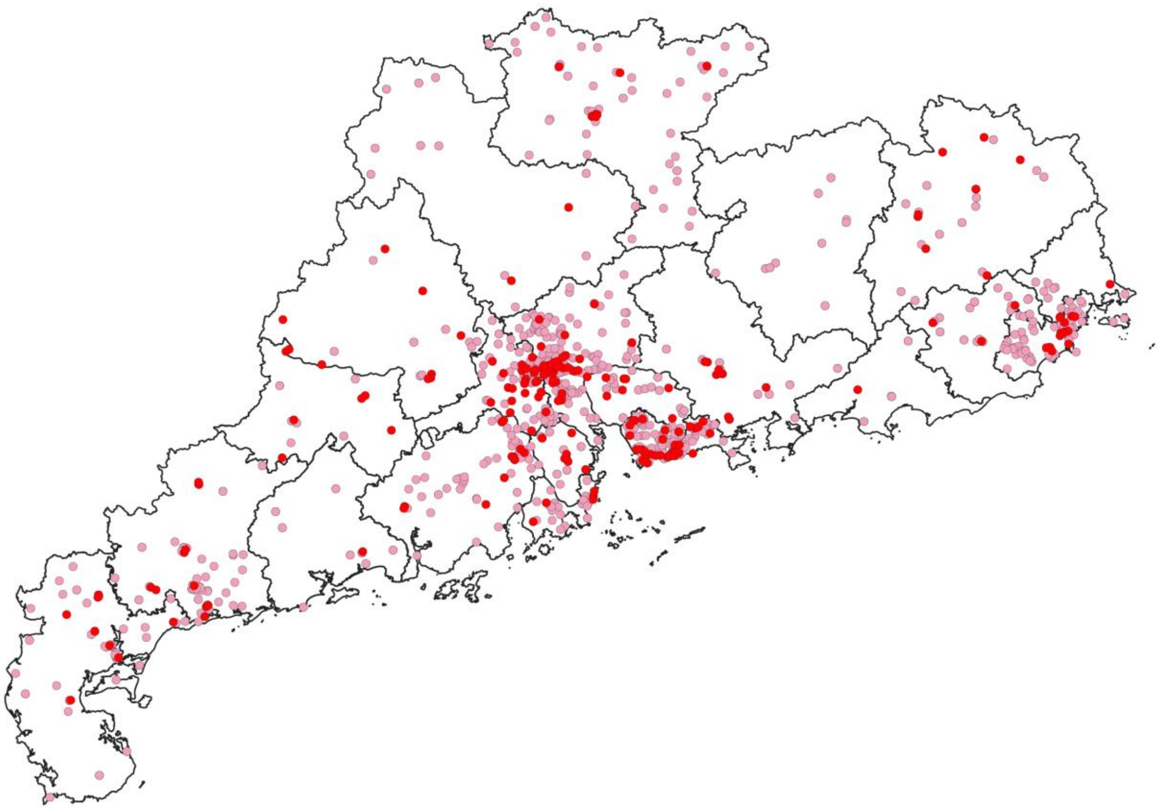


Figure A.5. Geographical distribution of key and non-key middle schools in Guangdong

Notes: This graph shows the geographical distribution of all the 928 middle schools in Guangdong during our sample period. Among the schools, 225 are provincial-level key middle schools. The light red and dark red dots in the graph denote key and non-key middle schools, respectively. The list of key and non-key middle schools is derived from the Bureau of Education of Guangdong province.

Table A.1. Pairwise correlations between wealth measures

	Housing prices	High-end mall visits	Car owner	Number of Phones	Phone price category
Housing values	-				
	-				
High-end mall visits	0.051***	-			
	(0.000)	-			
Car ownership	0.001	0.160***	-		
	(0.544)	(0.000)	-		
Number of phones	0.003*	0.002	0.032***	-	
	(0.070)	(0.302)	(0.000)	-	
Phone price category	0.006***	0.002	0.096***	0.153***	-
	(0.002)	(0.202)	(0.000)	(0.000)	-

Notes: This table presents the pairwise correlations across wealth measures. The variable definitions are the same as in Table 1. *P-values* in parentheses. *** p<0.01, ** p<0.05, * p<0.1.