The Value of Big Data in a Pandemic*

Kairong Xiao†

January 11, 2022

Abstract

Although digital contact tracing apps have been widely used to combat the COVID-19 pandemic, little empirical evidence regarding their effectiveness is available. This paper exploits the staggered implementation of a digital contact tracing app across 322 Chinese cities and finds that this big data technology significantly reduced virus transmission and facilitated economic recovery during the pandemic. A macroeconomic Susceptible-Infectious-Recovered (SIR) model calibrated to the micro-level estimates shows that this technology reduced the economic loss by 0.5% of GDP and saved more than 200,000 lives. The trade-off between economic benefits and privacy costs is also discussed.

JEL Classification Codes: D80, I10, E60
Keywords: Big data, informational frictions, COVID-19, lockdown

*I would like to thank Kent Daniel, Hanming Fang, Zhiguo He, Chang-Tai Hsieh, Wei Jiang, Stijn van Nieuwerburgh, Tommaso Porzio, Tarun Ramadorai, Tianyue Ruan, Laura Veldkamp, Yiyao Wang, Shang-jin Wei, Bernard Yeung, and Stephen Zeldes for their helpful comments and discussions. I also thank seminar participants at the American Finance Association Annual Meeting (AFA), the Becker Friedman Institute China (BFI-China) at the University of Chicago, the Columbia COVID-19 Virtual Symposium, the Columbia Macro-Lunch Workshop, the Columbia Finance Brownbag, and the Peak Initiative conference on FinTech for their helpful suggestions. I thank Meichen Qian for her excellent research assistance.

†Assistant Professor of Business, Columbia Business School. kx2139@gsb.columbia.edu

1 Introduction

Informational frictions are a key reason the COVID-19 pandemic has developed into a global economic and health crisis. Unlike its close sibling, Severe Acute Respiratory Syndrome (SARS), COVID-19 often causes no symptoms for an extended period during which the virus can be unknowingly transmitted.[1] As a result, even if the virus only infects a small fraction of the population at any given point in time, it can cause large-scale and persistent disruptions to the economy.[2] If it were possible to identify all the infected individuals and quarantine them for 14 days, the virus would die out, and the economic damages would be quite minimal.

Traditional technologies to produce information, such as diagnostic tests and manual contact tracing, are costly on the scale of the whole population. For instance, if we were to test the whole American population once a day for a year, it would cost more than $1 trillion. Moreover, diagnostic tests are often conducted after the symptoms appear, by which time the virus may have already been transmitted.

Big data may offer a possible solution to such informational frictions. Our digital age generates enormous data as byproduct of our economic and social activity. It is conceivable that we can use these data to generate low-cost and real-time information about virus transmission. Indeed, governments and private institutions worldwide have experimented with numerous digital contact tracing apps, resulting in many technological advancements and rich data resources.[3] However, there is little consensus on whether these technologies actually work. Braithwaite, Callender, Bullock, and Aldridge (2020) reviewed research published in the past twenty years and concluded "no empirical evidence of the effectiveness of automated contact tracing (regarding contacts identified or transmission reduction) was identified", and that "large-scale manual contact tracing is therefore still key in most contexts." Furthermore, much of the existing discussion focuses on public health outcomes. However, the COVID-19 pandemic is an economic crisis as much as a public health one. Therefore, it is also crucial to understand the economic impacts of these big data technologies.

This paper sheds light on these questions by studying the "Health Code app" intro-

[1]See World Health Organization Coronavirus disease 2019 (COVID-19) situation report-73.

[2]At the peak of the COVID-19 pandemic in the U.S. in 2021, the daily cases are less than 0.1% of the U.S. population.

[3]Notable examples include the "Exposure Notification API" developed by Apple and Google, the "Check-In" app developed by PwC, the "IBM Digital Health Pass" developed by IBM, and the "CommonPass" developed by the Commons Project.

duced during the COVID-19 outbreak in China. This app uses mobile location and digital transaction records to produce a colored QR code indicating the health status of the holder. If the app holder has not been in contact with any COVID-19 patients in the past 14 days, the app will generate a green QR code as shown in Figure 1a. However, if a potential contact is detected, the code will turn yellow or red, and the holder must self-quarantine for 7 to 14 days until the code turns green again.

The information produced by the Health Code app can alleviate the informational frictions in important ways. First, the Health Code app notifies infected individuals of the potential exposure so that they can take diagnostic tests sooner. Second, public and private institutions can use the colored QR code to monitor health conditions in the public space. Healthy individuals can conveniently certify their health status at the checkpoints of airports, railway stations, and restaurants. Infected individuals, in contrast, face an "effective quarantine" because they cannot enter public spaces without a green code. This "effective quarantine" alleviates the public concern about being infected by hidden carriers of the virus, which facilitates the resumption of normal work and life.

A key advantage of this big data technology over traditional technologies is its timeliness. Once a COVID-19 patient is confirmed, individuals in direct contact with the patient can be instantly identified so that the transmission chains can be cut off as soon as possible. Another advantage of this big data technology is the low cost. Because the contact history data are readily available, the economic cost of producing another QR code is virtually zero.[4] Therefore, it can be implemented on a large scale. While this big data technology has many advantages, it is still unclear how accurate the digital contact history can predict infections. Furthermore, it is also unclear how effectively the information produced can facilitate economic recovery.

A key challenge to evaluating the effects of big data technologies is that they are often introduced as a nationwide policy, so their effects are confounded with other disease control policies, people's attitudes, and culture. A unique feature of the Health Code app is that it was introduced in a staggered manner across 322 Chinese cities. To better understand the empirical setting, consider the example of Hangzhou and Nanjing, which are two similar coastal cities. The Health Code app was first introduced in Hangzhou because Hangzhou is the headquarter of Ant Financial, the FinTech company that developed this app.[5] After the

[4]The use of big data technology could lead to a potential privacy cost, which will be discussed later.

[5]The Hangzhou Municipal government approached Ant Financial to develop a solution based on big data to replace manual contact tracing and physical health records on February 4, 2020, 12 days after the Wuhan lockdown. On February 9, 2020, 17 days after Wuhan's lockdown, "Hangzhou Health Code" was launched

launch of the Health Code app, Hangzhou's economic activity, measured by daily greenhouse gas emissions, appeared to recover significantly. Economic recovery did not appear to bring a resurgence of infections. Instead, new COVID-19 cases dropped in Hangzhou. In comparison, a neighboring city, Nanjing, experienced little economic recovery but more infections during the same period.[6]

By exploiting the staggered implementation of the Health Code app as a natural experiment, I show that Hangzhou's experience can be generalized to other places. Specifically, I investigate high-frequency changes in cities' economy and public health around the Health Code launch dates using the event study design. I find that cities displayed no pre-trends in economic activity or infections before the Health Code launch. However, four weeks after the launch, the treated cities experienced a 24% increase in greenhouse gas emissions relative to the control cities. Interestingly, the resumption of economic activity did not bring a resurgence of infections. Instead, local infection growth dropped significantly after the city launched Health Code. Population inflows from outbreak epicenters also led to fewer local cases after the introduction of the Health Code app. The results are robust if the treated cities are matched with control cities with similar economic activity levels and outbreak severity.

One may worry that the Health Code launch timing may be endogenous to other policy changes. For instance, local governments may have launched the Health Code app when they reopen the economy. Thus, the resumption in economic activity could be mechanically driven by the relaxation of the lockdown policy rather than reducing informational frictions. To alleviate this concern, I use the daily intensity of each city's population movement constructed from mobile location data to control lockdown policies' strictness. Furthermore, I also use local governments' public health emergency levels to control other disease control measures such as mask mandate and social distancing rules. In other words, the identification is obtained by comparing two cities with similar lockdown and social distancing rules. More importantly, this alternative explanation cannot explain the decline in local infections after the introduction of the Health Code app. If the Health Code app had no effects on reducing informational frictions, lifting the lockdown would have led to a resurgence of infections.

A related concern is that city officials may have decided to launch the Health Code app when they observed signals that the local transmission was going down. This concern is

in Hangzhou.

[6]Both Hangzhou and Nanjing raised the public health emergency level to the highest level on January 26. By February 9, Hangzhou and Nanjing had 169 and 71 cases, respectively. In the following three weeks, new cases rose by 4 and 22 for these two cities, respectively.

partly addressed by the absence of pre-trends in economic activity and inflections before the introduction of the Health Code ap. To further address this concern, I employ an instrumental variable approach by exploiting the fact that the Health Code app was developed by FinTech firms and distributed through popular payment apps such as Alipay. Therefore, cities with higher FinTech penetration were likely to launch the Health Code app earlier than cities with low FinTech penetration. Indeed, I find that FinTech penetration in 2018 significantly predicts how early a city launched the Health Code app during 2020. The instrument is exogenous to the simultaneity concern because FinTech penetration is determined using the 2018 data and is unlikely to be correlated with private information that local officials had or actions they took during the pandemic. This alternative empirical strategy yields similar results: cities that launched the Health Code app early achieved greater economic recovery and lower infection growth.

While the micro-level evidence provides cleanly identified effects, it is difficult to draw welfare implications for the aggregate economy because common aggregate effects are differenced out.[7] To this end, I construct a macroeconomic Susceptible-Infectious-Recovered (SIR) model following Alvarez, Argente, and Lippi (2020). In the model, big data technology generates a publicly verifiable signal on agents' health status. This information allows society to impose targeted quarantines on those who received bad signals. I calibrate the model using parameters identified from the micro-level data. The model's key parameters include the accuracy of the signal, the adoption rate of this technology in the population, and the effectiveness of other disease control policies. The calibrated baseline model closely matches the path of COVID-19 cases in China.

I use the estimated model to compare economies with and without the Health Code app. I find that this big data technology saved the economy from suffering a new wave of infections after lockdowns were lifted, saving around 200,000 lives and 0.5% of the GDP in 2020. This result is striking because the signals produced by Health Code appear to be far less accurate than typical diagnostic tests. The Health Code app is estimated to have a false negative rate of 47% while typical viral and antibody tests need to have a below 5% false-negative rate to qualify for U.S. FDA approval.[8] However, because these signals can be produced in realtime, this technology can significantly impact the aggregate economy even if the signal is only modestly accurate.

[7] Nakamura and Steinsson (2018) review the growing literature that combines micro-level data and macro models to achieve both causal identification and welfare analysis.

[8] See "Policy for Coronavirus Disease–2019 Tests During the Public Health Emergency," https://www.fda.gov/media/135659/download.

I compare the effects of the Health Code app with those of the lockdown policies. To this end, I simulate a counterfactual economy in which lockdown policies were not implemented at the end of January 2020. In that case, the infections would have kept rising in the first two months of 2020 until the Health Code app and other disease control measures were rolled out. Overall, the strict lockdown policies saved 74,000 lives in 2020. However, because the massive lockdown constrained the economic activity of the whole population, it caused 1.1% of GDP loss in 2020. Comparing the effects of the lockdown policies with those of the Health Code app, I find that the Health Code app had a significant contribution in containing the outbreak without inflicting steep costs on the economy.

Although the above evidence shows that big data technologies can reduce informational frictions and create economic value, a necessary condition to realize this potential is that people are willing to adopt such technologies. The baseline estimate suggests that the average adoption rate among citizens is around 90% following the official launch in a city. This estimate is consistent with the disclosed adoptions rates in Zhejiang Province. The high adoption rate is essential for big data technology to be effective. In a counterfactual simulation in which the adoption rate is decreased to 20%, the death toll would have been 64,000 higher than the baseline simulation. The economic value created by the Health Code app would decrease by around 67%.

Big data technologies often raise concerns about privacy infringement. Due to privacy concerns, many COVID apps, such as the ones developed by Apple and Google, avoid linking the contact history to the user's identity. Instead, they use a "private notification" model in which an anonymous message is sent to the holder when the user is exposed to the virus. Under this private notification model, public health authorities or private businesses cannot use the notification, or the absence thereof, to monitor human flows in public spaces. App users have full discretion over whether to self-quarantine after receiving the notification. Because people may not strictly follow the self-quarantine rule, especially when they are asymptomatic, the effectiveness of the big data technology may be compromised. Indeed, the counterfactual simulation shows that making the signal privately observable can significantly change the outcomes, even the signals' accuracy remains constant. Under the assumption that 40% of the agents who receive bad signals decide to self-quarantine, the death toll would have been 16,000 higher than the baseline simulation. The economic value created by the Health Code app would decrease by 65%. This counterfactual exercise suggests a trade-off between protecting the privacy and resolving informational frictions. Whether the value created by the big data technology can justify the potential privacy cost is an open question and worth future research.

This paper contributes to the interdisciplinary literature on the potential of big data technologies to help control infectious diseases. Existing literature has argued that digital contact tracing could be an effective method to stop the COVID-19 pandemic based on evidence from simulated models (Ferretti, Wymant, Kendall, Zhao, Nurtay, Abeler-Dörner, Parker, Bonsall, and Fraser, 2020). However, real-world evidence for this approach is still scarce (Braithwaite, Callender, Bullock, and Aldridge, 2020). This paper is one of the first that provide such evidence.[9] Furthermore, the existing literature is often silent about these technologies' economic impacts, while this paper suggests the economic value created by these technologies can be substantial. The results can inform policy responses for the current pandemic and future ones, which have become increasingly common due to the increased human intrusion into natural habitats of wild animals (Jones, Patel, Levy, Storeygard, Balk, Gittleman, and Daszak, 2008).

This paper also contributes to the fast-growing literature on the economic implications of the COVID-19 pandemic.[10] In particular, Alvarez, Argente, and Lippi (2020) and Farboodi, Jarosch, and Shimer (2020) study the optimal lockdown policy of the government that faces a trade-off between saving lives and saving the economy. A key conclusion from these studies is that, in the absence of better technology, massive and sustained lockdowns are necessary to control an infectious disease like COVID-19. However, such lockdowns could lead to steep economic costs. My paper shows that big data technology can alleviate the information frictions faced by the society, which can save both lives and livelihoods.

This paper also contributes to the literature on the economics of big data (see Veldkamp and Chung (2019) and Goldstein, Jiang, and Karolyi (2019) for reviews). Existing literature shows that big data can alleviate informational frictions in the mortgage market (Buchak,

[9] See Wymant, Ferretti, Tsallis, Charalambides, Abeler-Dörner, Bonsall, Hinch, Kendall, Milsom, Ayres, Holmes, Briers, and Fraser (2021) and Rodríguez, Graña, Alvarez-León, Battaglini, Darias, Hernán, López, Llaneza, Martín, Ramirez-Rubio, Romaní, Suárez-Rodríguez, Sánchez-Monedero, Arenas, and Lacasa (2021) for evidence in the U.K. and Spain.

[10] An incomplete list of this fast-growing literature includes: Acemoglu, Makhdoumi, Malekian, and Ozdaglar (2019), Alvarez, Argente, and Lippi (2020), Atkeson (2020), Baker, Farrokhnia, Meyer, Pagel, and Yannelis (2020), Barro, Ursúa, and Weng (2020), Barrios and Hochberg (2020), Berger, Herkenhoff, and Mongey (2020), Bethune and Korinek (2020), Bian, Li, Xu, and Foutz (2020), Bordalo, Coffman, Gennaioli, and Shleifer (2020), Budish (2020), Charoenwong, Kwan, and Pursiainen (2020), Chen, Qian, and Wen (2020), Correia, Luck, and Verner (2020), Coven and Gupta (2020), Dewatripont, Goldman, Muraille, and Platteau (2020), Eichenbaum, Rebelo, and Trabandt (2020), Elenev, Landvoigt, and Van Nieuwerburgh (2020), Fang, Wang, and Yang (2020), Farboodi, Jarosch, and Shimer (2020), Fernández-Villaverde and Jones (2020), Gallant, Kroft, Lange, and Notowidigdo (2020), Goolsbee and Syverson (2020), Guerrieri, Lorenzoni, Straub, and Werning (2020), Hassan, Hollander, van Lent, and Tahoun (2020), Hall, Jones, and Klenow (2020), Hong, Wang, and Yang (2020), Hortaçsu, Liu, and Schwieg (2020), Jones, Philippon, and Venkateswaran (2020), Kozlowski, Veldkamp, and Venkateswaran (2020), Piguillem and Shi (2020), and Redding, Glaeser, and Gorback (2020).

Matvos, Piskorski, and Seru, 2018) and equity market (Begenau, Farboodi, and Veldkamp, 2018; Zhu, 2019). My paper adds to this growing body of research by showing that big data technologies can help tackle the twin economic and health crises of COVID-19. A related strand of literature studies the implications of big data on privacy concerns (Argente, Hsieh, and Lee, 2020; Athey, Catalini, and Tucker, 2017; Ramadorai, Uettwiller, and Walther, 2019; Tang, 2019; Agarwal, Ghosh, Ruan, and Zhang, 2020). This paper adds to this line of inquiry by discussing the trade-off between making health information publicly verifiable or keeping it private to the users. This paper also broadly contributes to the vast economic literature on informational frictions.[11] The key insight of this literature is that imperfect information constitutes important economic friction, which can lead to market failures and suboptimal equilibrium outcomes. My paper contributes to this literature by showing that big data technologies can alleviate informational frictions and create economic value.

2 Background and Data

Health Code. Health Code is a big data technology that uses location history and mobile transactions to predict an individual's risk of being infected by COVID-19. It was developed by several technology companies in China, including Alipay and Tencent, during the height of the COVID-19 outbreak. Although various versions of Health Code apps are introduced in different cities, these apps are functionally similar. As a result, I use the "Health Code app" as an umbrella term to refer to this technology throughout the paper.

The Health Code app generates colored QR codes as shown in Figure 1a. Holders of a green code can freely travel in the city; holders of yellow or red codes are supposed to quarantine for 7 or 14 days, respectively. The yellow or red codes turn back to green after the quarantine. These colored codes not only notify users of possible exposure but also serve as digital health certificates. As shown in Figure 1b, public health authorities usually set up checkpoints in public spaces, such as subways, to check passengers' codes. Only green code holders can board public transportation. Many private businesses, such as supermarkets and shopping malls, also require employees and customers to scan their codes. These businesses do so in their own best interests because customers are more comfortable entering business sites when they know that other people who work or shop there are likely to be healthy. The

[11] See Stiglitz (2002) for a review of the theoretical literature on informational frictions, see also Veldkamp (2011) and Angeletos and Lian (2016) for reviews on the applications of information theory in macroeconomics and finance.

numerous checkpoints encourage adoption and generate more data for contact tracing. At the same time, these checkpoints also impose an effective quarantine on potentially infected individuals.

The Health Code app can potentially mitigate the adverse impact of pandemics in two ways. First, this technology can accelerate the virus's detection and curb the transmission through digital contact tracing. Second, by generating a credible real-time health certificate, this technology allows healthy people to resume normal work and life, reducing the pandemic's economic damages. This function was particularly valuable in the early stage of the COVID-19 pandemic when testing capacity was limited.

Because many checkpoints are set up in public spaces, people have a strong incentive to adopt this technology. For instance, around 90% of the provincial population obtained Health Code within 15 days after the app's introduction in Zhejiang Province. Among all the Health Codes issued in Zhejiang, 98.2% were green, and 1.8% were yellow or red as of February 24, 2020.[12] It is worth noting that people do not need smartphones to use their code because they can print the QR code for scanning at checkpoints.[13] This feature allows Health Code to cover people who may not have smartphones, such as the elderly.

The first Health Code app was developed by Alipay and implemented in its headquarters city, Hangzhou, on February 9, 2020, 17 days after Wuhan's lockdown. Following Hangzhou, many provinces and cities launched their versions of the Health Code. It is worth noting that local governments launched their Health Code apps without the central government's coordination, which created a patchwork of policies in the early stage.[14] Cities did not recognize each other's Health Code; different versions of Health Code apps sometimes showed inconsistent results; some people were required to scan multiple versions of Health Code at a single location.[15] The uncoordinated implementation of the Health Code created an inconvenience for people who travel across cities. Health Code was eventually homogenized with the national government guideline published on May 2, 2020. This study will focus on February 9, 2020-March 31, 2020, because the staggered implementation across cities provides a natural experiment to study this technology's effects.

[12]See March 1, 2020, *New York Times* article: "In Coronavirus Fight, China Gives Citizens a Color Code, With Red Flags", by Paul Mozur, Raymond Zhong, and Aaron Krolik.

[13]When the QR code is printed out, the colored coding associated with this QR code can be shown in the device which scans the QR code. See June 24, 2020 *Xinhua News* article, "How to show Health Code without smartphones", http://www.xinhuanet.com/2020-06/24/c_1126157347.htm.

[14]Across different public spaces within the same city, the adoption of the Health Code is highly coordinated by the local government.

[15]See March 9, 2020, *South China Morning Post* article, "National version of China's controversial Health Code isn't ready."

Summary statistics. Table 1 provides the summary statistics of the sample. The sample period starts from January 1, 2020, and ends on March 31, 2020. I describe the data sources and variables definitions below.

Health Code Launch Dates. I collect the Health Code's launch dates for 322 Chinese cities from local government websites and news media. The launch dates are determined based on local governments' announcements that local Health Codes are available to download and checked in certain public spaces. Figure 2 shows the number of cities that launch the Health Code app over time. The implementation process started 17 days after the Wuhan lockdown and lasted for two months until the end of March when all 322 cities had launched Health Code. Figure 3 shows the geographical distribution of the Health Code launch dates. The launch dates are not related to cities' geographical proximity to the epicenter of the virus outbreak, Wuhan.

FinTech penetration. I use the Peking University Digital Financial Inclusion Index to measure FinTech penetration.[16] This index is constructed by using Ant Financial's massive dataset on digital financial inclusion, using the information on the number of Alipay accounts per 10,000 people, the number of transactions conducted through mobile banking, and the number of loans obtained from FinTech leaders, and so on. This index is available at the city level from 2011–2018. I use the 2018 value as an instrument for the timing of Health Code implementation.

Greenhouse gas emissions. I use the daily level of nitrogen dioxide (NO2) and sulfur dioxide (SO2) from the China National Environmental Monitoring Center (CNEMC) as high-frequency measures of economic activity.[17] NO2 and SO2 are greenhouse gases created by factories and automobiles burning fossil fuels. Because the Chinese economy heavily relies on fossil fuels as a source of energy, the amounts of NO2 and SO2 are good economic activity measures in China.[18]

Figure 4a plots the average greenhouse gas emissions of the sample cities. The values are normalized by the average of the first two weeks in 2020. The first vertical line indicates January 23, 2020, the day of Wuhan's lockdown. The second vertical line indicates February 9, 2020, when the Hangzhou Health Code was introduced. A 40% drop in greenhouse gas emissions occurred shortly after Wuhan's lockdown and did not recover until late February. The rapid reduction in NO2 emissions and the later recovery are visible in the satellite images

[16] The data can be accessed from Https://tech.antfin.com/research/data.

[17] The source of the data can be found on the website of CNEMC: Http://www.cnemc.cn/.

[18] See Morris and Zhang (2019) who use NO2 emissions to verify China's output data.

produced by Google Earth, as shown in Figure 4b.

Virus outbreak. I collect the daily count of confirmed, dead, and recovered COVID-19 cases for 322 cities from the Centers for Disease Control and Prevention of China (CDC).[19] Figure 5a plots the time series of COVID-19 cases in the sample. From January 11, 2020, to April 3, 2020, the data cover 81,198 confirmed COVID-19 cases, 3,302 deaths, and 75,887 recoveries. It is worth noting that the pandemic had not reached its peak when the Health Code app was first introduced.

Public health emergency level. I collect information on local governments' public health emergency levels from their websites and local news reports. The lowest level is coded as 0 and the highest as 3. A higher level of emergency allows local governments to impose stricter lockdown and social distancing rules. Figure 5b shows the average emergency levels over time. The emergency level was raised to the highest level shortly after the Wuhan lockdown on January 23, 2020. The emergency level was gradually lowered starting from March 2020 and reached around 1.5 at the end of the sample period.

Population movements. The population movement data cover 322 Chinese cities between January 1, 2020, and April 10, 2020. The data are created using real-time phone location data from the largest Chinese search engine in China, Baidu.[20] The Baidu data also provide between-city migration flows, which are used to construct infection inflows from epicenter cities to other cities.

Figure 5b plots the national average within-city movement in the sample period.[21] I report the value as a percentage of the average value in the first week of 2020. Figure 5b shows a steep drop in within-city movement after January 23, 2020, the day of Wuhan's lockdown. The Within-city movement slowly recovered in mid-February as the public health emergency levels were lowered. By the end of the sample period, the within-city movement rebounded to about 95% of the pre-COVID-19 level.

[19] The source of the data can be found on the website of CDC: Http://2019nCoV.chinacdc.cn/2019-nCoV/.

[20] The source of the data can be found on the website of CNEMC: Http://https://qianxi.baidu.com/.

[21] Note that the sample period contains the Lunar New Year holiday, during which the population movements in cities would naturally decrease. To control for the Lunar New Year's effect, I normalize the within-city movement using the same-day value of the 2019 lunar calendar.

3 Empirical Results

In this section, I test two main hypotheses: (1) the introduction of the Health Code app increased local economic activity, and (2) the introduction of the Health Code app reduced COVID-19 infections.

3.1 Evidence from the Event-Study Approach

I first use an event-study approach following Borusyak and Jaravel (2017) to examine changes in high-frequency economic and public health indicators around the launch dates. Consider a panel of $i = 1, ..., N$ cities in which the outcome $Y_{i,t}$ is observed for $t = 1, ..., T$ periods, where t is calendar time. City i launches the Health Code app in some time E_i. Let $K_{i,t} = t - E_i$ denote the "relative time"—the number of periods relative to the event. The regression model is the following:

$$Y_{i,t} = \alpha_i + \beta_t + \sum_k \gamma_k \mathbb{1}\{K_{i,t} = k\} + \eta X_{i,t} + \epsilon_{i,t},$$

where $\{\gamma_k\}$ for $k < 0$ corresponds to pre-trends, and for $k \geq 0$ corresponds to dynamic effects k periods relative to the event. In the baseline estimation, I define the unit of periods as a week so γ_k represents the treatment effects k weeks from the event date. I use a event window of 14 weeks around the introduction of Health Code. α_i and β_t are city fixed effects and time fixed effects, which absorb time-invariant city characteristics and aggregate shocks. $X_{i,t}$ includes within-city population movement and local public health emergency level, which control for the intensity of lockdown policies and other disease control measures. The empirical design effectively compares changes in treated cities' economic activity and public health condition with those of control cities before and after the introduction of the Health Code app.

3.1.1 Economic activity

I first examine the effect of the introduction of the Health Code on economic activity measured by daily NO2 emissions. Table 2 presents the results and Figure 6a plots the point estimates and 95% confidence intervals. Note that Table 2 only presents the pre-treatment effects for two weeks due to space constraints. The full set of estimates is plotted in Figure 6a.

A key identification assumption of the event study approach is parallel trends, which I can verify by examining $\{\gamma_k\}$ for $k < 0$. Indeed, before the introduction of the Health Code app, there is no significant difference between treated and control cities. Two weeks after the introduction of the Health Code app, the NO2 emissions of the treated cities start to rise significantly compared with the control cities. The difference in NO2 emissions rises to 24% four weeks after the introduction of the Health Code app. The results suggest that the introduction of the Health Code app significantly facilitates economic recovery.

3.1.2 COVID-19 infections

So far, the evidence suggests that the introduction of the Health Code app allowed the economy to return to normal. However, one important question is whether the reopening of the economy led to a resurgence of virus infections. To answer this question, I estimate the same regression (3.1) on local infection growth. Table 3 presents the results and Figure 6b plots the point estimates and 95% confidence intervals. Again, there is no pre-trend before the introduction of the Health Code, which suggests the introduction does not seem to correlate with the severity of the outbreaks. About two weeks after the introduction of the Health Code app, the infection growth dropped by around 70% and stayed low afterward.

Next, I estimate the sensitivity of local cases to the inflows of infected individuals from epicenter cities following Fang, Wang, and Yang (2020). Specifically, I construct a new variable, Infection inflow$_{i,t}$, to measure the potential outside infections brought in by population inflows:

$$\text{Infection inflow}_{i,t} = \sum_j \text{Existing cases}_{j,t} \times \text{Inflow}_{i,j,t}, \tag{1}$$

where Inflow$_{i,j,t}$ is the population flow from source city j to destination city i, and Existing cases$_{j,t}$ is the number of existing cases in a source city j. I then follow Fang, Wang, and Yang (2020) to estimate impulse-response functions of local cases to the infection inflow. The key difference is that I allow the impulse-response functions to be different after the introduction of the Health Code app. Specifically, the regression model is the following:

$$\log(\text{New cases})_{i,t+h} = (\beta_{1,h} + \beta_{2,h}\mathbb{1}\{t > E_i\}) \times \log(\text{Infection inflow})_{i,t} + \gamma X_{i,t} + \epsilon_{i,t}, \tag{2}$$

where $\log(\text{New cases})_{i,t+h}$ is the log number of new cases in city i at date $t + h$; E_i is the introduction date of the Health Code app in city i; $X_{i,t}$ is a vector of control variables which include the emergency level, the log number of confirmed cases, city fixed effects and time-fixed effects. $\beta_{1,h}$ measures the response of local cases to the past infection inflows without Health Code in place, while $\beta_{1,h} + \beta_{2,h}$ measures the response of local cases to the past infection inflows with Health Code in place.

Figure 7a plots the impulse response of local cases to past infection inflows without the Health Code. A 1% increase in infection inflows leads to a 0.08% increase in local cases in 7-10 days after the inflow occurred. In contrast, with the Health Code, the sensitivity of local cases to past infection decreases to around 0.04%, as shown in Figure 7b. This result suggests that the Health Code helps screen out around 50% of the potential sources of infection.

3.1.3 Alternative hypotheses and robustness

One may worry that cities that launched the Health Code app early may have had a less severe outbreak to start with. To address this concern, I match the treated cities to control cities with a similar number of active cases when the Health Code was introduced using propensity score. The result is presented in column 2 of Tables 2–3. The results are again robust in the matching sample. A related concern is that cities that launched Health Code early could have higher economic importance. Thus they were forced to reopen before other cities. I address this concern by matching treated cities to control cities with similar pre-COVID-19 economic activity. The result is presented in column 3 of Tables 2–3. The results are also robust to this alternative matching scheme.

A related concern is that the introduction of the Health Code app could be correlated with changes in other disease control measures. For instance, cities may have relaxed their lockdown policy upon the introduction. Even if Health Code did not alleviate the outbreak, we would observe an increased economic activity mechanically.

This alternative hypothesis does not seem to be consistent with the reduction in infections after the introduction of the Health Code app. In other words, a mechanical change in lockdown policy can only move the economy along a given output-infection frontier, and more infections will accompany the economic recovery. To achieve high economic activity and low infections, one has to change the underlying technology to expand the output-infection frontier. Therefore, a shift from a blanket lockdown to more-targeted quarantines could be interpreted as a consequence of the introduction of the Health Code app in the sense that the information generated by this technology makes a targeted quarantine policy feasible. Furthermore, there is no significant correlation between the app's introduction and lockdown easing in the data. Regressing the Health Code introduction dummy on the contemporaneous changes in the public emergency levels leads to a precisely estimated zero. The 95% confidence interval of the regression coefficient is -.045 to .022. This result is consistent with Fisman, Lin, Sun, Wang, and Zhao (2021) who show that the lift in lockdown policy by Chinese local governments was likely prompted by citizen discontent instead of the innovations in disease control technologies.

Recent literature on event study design shows that linear regressions implicitly assume the treatment effects are homogeneous (see Borusyak, Jaravel, and Spiess (2021) and the references therein). If this assumption is violated, the estimator may place distorted weights on treatment effects. To address this concern, Table 4 uses the imputation estimator proposed by Borusyak, Jaravel, and Spiess (2021), which are robust to treatment-effect heterogeneity. The results are robust to this alternative estimator.

3.2 Evidence from Instrumental Variable Regressions

I use an instrumental variable approach to obtain exogenous variations in the Health Code launch timing to sharpen the identification further. I exploit the fact that the Health Code apps are developed by FinTech firms and are distributed through popular payment apps. Therefore, cities with higher FinTech penetration are more likely to introduce Health Code early. Specifically, I construct a new variable "Health Code Earliness$_i$," defined as the fraction of days that the Health Code is in place between February 9, 2020, and March 31, 2020, which corresponds to the date when the Health Code app was first introduced in Hangzhou and the end of the sample period, respectively. The earlier a city launched the Health Code app, the higher the Health Code Earliness$_i$ value. I use FinTech penetration measured in 2018 to predict the earliness of the Health Code launch during the 2020 pandemic. Because a city's FinTech penetration was determined using 2018 data and was largely driven by its geographical distance from Alipay's headquarter, it is unlikely to be correlated with private information that local officials had during the pandemic or actions taken around the Health Code app's launch date. Formally, the first-stage regression is the following:

$$\text{Health Code Earliness}_i = \gamma \text{FinTech penetration}_i + u_i.$$

The sample is a cross-section of cities with non-zero infections as of February 9, 2020, and non-missing FinTech penetration. I find that FinTech penetration in 2018 significantly predicts how early a city launched the Health Code app during the COVID-19 pandemic. The F-statistic of the first-stage regression is 21, suggesting that this instrument is relevant for the timing of the Health Code launch. Then, I relate the predicted earliness of the Health Code launch to the change in economic activity during the period when the Health Code app was rolled out across different cities:

$$\Delta Y_i = \beta \text{Health } \widehat{\text{Code}} \text{ Earliness}_i + \eta X_i + \epsilon_i,$$

where ΔY_i is the change in economic or public health conditions between February 9, 2020 and March 31, 2020. $X_{i,t}$ is a vector of control variables which include within-city movement, public health emergency level, and the log number of confirmed cases as of February 9, 2020.

Table 5 presents the results. Cities that launched the Health Code app early achieved greater economic recovery. Specifically, daily NO2 emissions were 3.3 times higher if a city launched the Health Code app at the beginning of the sample period instead of never launching it. Cities that launched the Health Code app early also experienced lower COVID-19 infections and deaths during the sample period. The results are consistent with the event-study approach.

4 Model

Although the micro-level data provide clear evidence that big data technologies can mitigate the economic and human costs of the COVID-19 pandemic, it is difficult to draw welfare implications. To evaluate Health Code's aggregate welfare effects, I calibrate a macroeconomic SIR model following Alvarez, Argente, and Lippi (2020). My model differs from theirs in that I introduce a big data technology that can produce a signal on the agents' health status. I discipline the model using parameters identified from the micro-data. I then use the model to conduct counterfactual simulations to analyze the effects of big data technology on mitigating the economic and human costs of the COVID-19 crisis.

4.1 Setting

The total living population $N(t)$ is divided between those susceptible $S(t)$, infected $I(t)$, and recovered $R(t)$, that is,

$$N(t) = S(t) + I(t) + R(t). \tag{3}$$

Note that the total living population $N(t)$ will change through time, as deceased agents are subtracted from the total. I normalize the initial population to $N(0) = 1$. The recovered agents are assumed to be permanently immune to the virus.

The law of motion of the infected agents is:

$$\dot{I}(t) = \beta S(t)(1 - L_S) \times I(t)(1 - L_I) - \gamma I(t), \tag{4}$$

where β is the number of susceptible agents per unit of time to whom an infected agent can transmit the virus, and γ is the rate that infected agents recover or die. L_S and L_I are the reduction in activities of susceptible and infected agents, respectively.

An infected agent is removed from the susceptible group, so the law of motion of susceptible agents is:

$$\dot{S}(t) = -\beta S(t)(1 - L_S) \times I(t)(1 - L_I). \tag{5}$$

The infected agents die with a rate of κ, so the law of motion of the total living population is:

$$\dot{N}(t) = -\kappa I(t). \tag{6}$$

The infected agents recover at a rate of $\gamma - \kappa$, so the law of motion of recovered agents is:

$$\dot{R}(t) = (\gamma - \kappa)I(t). \tag{7}$$

The welfare loss in the pandemic equals the sum of economic output loss and the economic value of human life loss:

$$\int_0^T e^{-rt} \left(\underbrace{\lambda(S(t)L_S + I(t)L_I)}_{\text{Economic costs}} + \underbrace{\chi D(t)}_{\text{Human toll}} \right) dt, \tag{8}$$

where r is the discount rate, T is the time that a vaccine is developed, $S(t)L_S + I(t)L_I$ is the reduction of in-person economic activity due to lockdown or fear of infections, λ is the elasticity of output loss due to the reduction of in-person activity, $D(t) = \kappa I(t)$ is the death toll, and χ is the economic value of human life.

A big data technology generates a signal y on the agents' type according to the following distribution:

$$\begin{aligned} p(y = s|S) &= p_s \\ p(y = i|I) &= p_i, \end{aligned} \tag{9}$$

where p_s is the probability that a healthy agent is predicted as low-risk, and p_i is the probability that an infected agent is predicted as high-risk. This technology allows society to impose a targeted quarantine policy: if an agent receives high-risk signal i, he or she will be quarantined with probability ℓ_i; if an agent receives low-risk signal s, he or she will be quarantined with probability ℓ_s. In the absence of the big data technology, the government can only impose an untargeted lockdown, ℓ, on the whole population.

4.2 Parameterization

To map the model to the data, I need to parameterize the reduction in activities, L_s and L_i, to functions of observable quantities in the data. I first measure the fraction of the population that can potentially adopt the big data technology as $\sum_i w_i \mathbb{1}\{t > E_i\}$, where E_i is the launch date of Health Code by city i, and w_i is the population weight of city i. I then define α as the adoption rate of citizens after the Health Code app is launched in a city. So the fraction of population covered by the Health Code is given by

$$h(t) = \alpha \sum_i w_i \mathbb{1}\{t > E_i\} \tag{10}$$

Given $h(t)$, the fraction of susceptible agents and the fraction of infected agents under quarantine are the following:

$$\begin{aligned} L_S &= h\left(\ell_s p_s + \ell_i(1 - p_s)\right) + (1 - h)\ell \\ L_I &= h\left(\ell_s(1 - p_i) + \ell_i p_i\right) + (1 - h)\ell. \end{aligned} \tag{11}$$

Before the Health Code is introduced, the reductions in activity by both susceptible and infected agents is parametrized as the following function:

$$\ell(t) = \phi x(t) + \psi d(t). \tag{12}$$

$x(t)$ is the reduction in population movements, which is observable in the data. ϕ is the effectiveness of mobility restrictions in reducing transmission. $d(t)$ is a dummy variable that takes a value of 1 after January 23, 2020, Wuhan's lockdown date. ψ captures other disease control measures introduced since Wuhan's lockdown, including expanding testing capacity, requiring masks, and increasing awareness of personal hygiene.

The reduction in population movements can be further decomposed into two components:

$$x(t) = g(t) + u(t), \tag{13}$$

where $g(t)$ is the reduction in population movements due to a government lockdown, and $u(t)$ is the reduction in population movements to the fear of infections (Goolsbee and Syverson, 2020). I assume this variable is a function of the number of infected agents that are not under quarantine: $u(t) = \theta \ln(I(1 - L_I))$.

After the Health Code is introduced, agents who receive high-risk signals will be quarantined, $\ell_i = 1$. Agents who receive low-risk signals will face the same lockdown policy before the introduction of Health Code: $\ell_s = \ell$.

4.3 Calibrated Parameters

For COVID-19-specific parameters, I directly take the values from Alvarez, Argente, and Lippi (2020). Specifically, I set β, the rate at which infected agents meet other agents and transmit the virus, at a value of 20%. I set γ, the rate at which infected agents either recover or die to 1/18, reflecting an estimated duration of illness of 18 days. The fatality rate per day is set to 0.05γ, which implies a 5% fatality rate for those infected. The economic value of human life, χ, is set to 150 times GDP per capita following Kniesner, Viscusi, Woock, and Ziliak (2012). I calibrate the elasticity of output loss due to the reduction of in-person activity, λ, to 0.12 to match the 4% reduction in Chinese GDP compared to its long-run trend. Figure 7a and 7b show that the introduction of the Health Code app leads to a 50% reduction in infection rates. This moment implies a restriction on the adoption rate and the accuracy of detecting infections, $(1-\alpha)p_i = 0.5$.[22] I calibrate the sensitivity of population movements to the number of undetected infections θ to 0.025 to match the relative magnitude of economic loss and death toll.[23]

I estimate the remaining parameters by minimizing the differences between the simulated cases and the actual cases in the data:

$$\min_{p_i,\alpha,\phi,\psi,I_0} \sum_t (\hat{C}(t) - C(t))^2 \tag{14}$$

where $\hat{C}(t)$ and $C(t)$ are the daily confirmed cases in the model and the data, respectively.[24] The remaining parameters include: (1) the probability that the Health Code app detects infections p_i; (2) the adoption rate α; (3) the effect of mobility restriction on infections ϕ; (4) the effect of other disease control measures on infections, ψ; (5) the initial number of infected cases, I_0. The data input includes the observed reduction in population movement, $x(t)$, and the fraction of the population that can potentially adopt the Health Code, $\sum_i w_i \mathbb{1}\{t > E_i\}$. The sample period is from January 16, 2020, to March 31, 2020.

Table 6 shows the parameters of the model. The false-positive rate of the Health Code app is estimated to be 47%. In other words, around half of the patients who carry the virus is undetected by the Health Code. The false-positive rate is significantly higher than conventional viral or antibody tests, which requires the false-negative rate to be below 5%.[25] Nevertheless, the Health Code's comparative advantage is that it can track millions of users in realtime.

The false-positive rate of the Health Code is estimated to be 2%.[26] Although the 2%

[22] Specifically, from equation (11), the change in the infection rates can be derived as $\frac{C}{C^-} = \frac{\beta S(1-L_s)I(1-L_I)}{\beta S(1-L_s^-)I(1-L_I^-)} \approx \frac{1-L_I}{1-L_I^-} = (1-\alpha)p_i$ where C^- and C are the infection rates before and after the Health Code app was introduced, respectively.

[23] Intuitively, a larger θ reduces the death toll but increases economic losses because there is a larger reduction in economic activity for a given level of infections.

[24] Note that I allow a 14-day lag between the time an infection occurs and a case is confirmed in the model: $\hat{C}(t) = \beta S(t-\tau)(1-L_S(t-\tau)) \times I(t-\tau)(1-L_I(t-\tau))$. I drop two observations on February 13 and 14, 2020, because there was a change in diagnosis criteria, which led to an abnormal spike on these two dates.

[25] See "Policy for Coronavirus Disease-2019 Tests During the Public Health Emergency," which states: "For this guidance, FDA defines the acceptance criteria for the performance as 95% agreement at 1x-2x LoD, and 100% agreement at all other concentrations and for negative specimen." https://www.fda.gov/media/135659/download.

[26] This parameter is calibrated by matching the ratio of number of yellow and red code holders and the actual COVID patients in Zhejiang Province. As of February 24, 2020, the fraction of yellow or red code holders in Zhejiang Province was 1.8%, and the fraction of the infected population was 0.002%.

false positive rate appears to be small, it is significantly higher than the effective zero false-positive rate required for the conventional viral or antibody tests.[27] If we multiply the 2% false positive rate with the size of the population, it implies that millions of healthy people will receive yellow or red code and have to face quarantine. In the model, agents who receive a bad signal cannot work. Therefore, the 2% false positive rate creates a high hurdle for the Health Code app to have an overall positive impact on the economy.

The adoption rate in the population is estimated to be 95%, consistent with the data in Zhejiang Province, where around 90% of the provincial population obtained the Health Code app within 15 days after it was introduced.

Figure 8 shows the simulated path of confirmed cases in the data and the model. The model fits the data quite well: the predicted peak of the infection took place around two weeks after the Wuhan's lockdown on January 23, 2020, which matches closely with the data. Furthermore, the number of new cases remained low after lockdowns were gradually loosened in March 2020.

4.4 Mechanism

Before I conduct counterfactual simulations, it is useful to analyze the basic reproduction number in epidemiology, or R_0, implied in the model. R_0 represents the expected number of cases directly generated by one case in a population where all individuals are susceptible to infection. In the absence of interventions, the basic reproduction number is $R_0 = \beta/\gamma$. When $R_0 > 1$, each infected agent is expected to infect more than one agent before this infected agent recovers or dies. Therefore, in the absence of a cure or vaccine, the virus will infect all the agents in the long-run steady state. In contrast, if $R_0 < 1$, each infected agent is expected to infect less than one agent. The virus will disappear in the long run, even without a cure or vaccine. In the model, the basic reproduction number is the following:

$$R_0 = \beta/\gamma \times \underbrace{\left(1 - h\left(\ell p_s + (1-p_s)\right) - (1-h)\ell\right)}_{1-L_S} \times \underbrace{\left(1 - h\left(\ell(1-p_i) + p_i\right) - (1-h)\ell\right)}_{1-L_I}, \quad (15)$$

where the first term is the basic reproduction number without any interventions, the second and third terms are the fractions of susceptible and infected agents that are not under quarantine.

[27] See "Policy for Coronavirus Disease-2019 Tests During the Public Health Emergency," https://www.fda.gov/media/135659/download.

I plot the resulting basic reproduction number for each level of ℓ, the fraction of susceptible agents under lockdown. The resulting curve is the infection-possibilities frontier, as shown in Figure 9, which shows the possible combinations of lockdown intensity and the infection rate that can be achieved using available technology. I consider two scenarios. In the first scenario, the Heath Code app is not adopted, $h = 0$. The basic reproduction number of COVID-19 in this scenario is 3.6 if no lockdown is imposed. In this case, the government needs to lock down 52.7% of the susceptible population to bring R_0 below 1. In the second scenario, the Heath Code app is adopted by the whole population, $h = 1$. In this case, the infection-possibilities frontier is expanded, and only 20% of the susceptible population needs to be locked down to bring R_0 below 1. It is worth noting that, even when the whole population adopts the Heath Code app, $h = 1$, the R_0 is still above 1 without any lockdown. In other words, the Health Code app alone cannot contain the COVID-19 pandemic without other disease control policies in place.

4.5 Counterfactual Simulations

So far, I have analyzed the Health Code's effect through the lens of the basic reproduction number of COVID-19. This section will use the structural model to analyze the welfare implications of this technology. First, I simulate the path of the baseline economy until the end of 2020. I also assume the behavioral and technological changes since the outbreak are still in place for the rest of the year. The blue line of Figure 10 shows the fraction of infected agents in the population in this baseline scenario. I find that infections will not rebound later in the year, even if the economy reopens. Row 1 of Table 7 shows the economic losses and human toll of the baseline economy. The economic loss amounts to 4.26% of the annual GDP, while the death toll amounts to around 4,650, consistent with the data. The small number of deaths has a limited impact on social welfare under the assumption that the economic value of human life is 150 times GDP per capita, so the loss of welfare is around 4.31% of the annual GDP.

Next, I consider a counterfactual economy without the Health Code app. Other disease control measures represented by ψ remain in this counterfactual scenario. The red line of Figure 10 shows the fraction of infected agents in the population in this counterfactual scenario. In this case, infections start to rebound as the lockdown was eased in early March and keep rising throughout 2020. Row 2 of Table 7 shows the economic and human losses of this counterfactual economy. Compared with the baseline case, this simulation shows that the Health Code app reduced economic loss by 0.5% of GDP and saved more than 200,000 lives.

To evaluate the Health Code app's relative contribution to the lockdown policies, I consider a counterfactual economy in which the government did not impose any lockdown in January 2020. As shown by the yellow line of Figure 10, infections kept rising until the end of March. The infections only leveled off when the Health Code app and other disease control measures were implemented. This counterfactual suggests that 70,000 more lives would have been lost if strict lockdowns were not imposed during the outbreak's peak. This result is consistent with Fang, Wang, and Yang (2020), who show that the lockdown significantly reduced future infections. However, the lockdown policies also imposed a steep economic price: the economy suffered a 1.1% loss in GDP in 2020 due to the strict lockdown policies.

In summary, these two counterfactual simulations show that the Health Code app had a similar if not larger effect in containing the outbreak compared with the strict lockdown policies. Furthermore, unlike the strict lockdown policies that inflicted a steep cost on the economy, the Health Code app facilitated economic recovery and reduced the pandemic's economic impact.

4.6 Discussion on Privacy

Big data technology often raises concerns about privacy infringement. Because of this concern, many contact tracing apps avoid linking the contact history to the user's identity. Instead, an anonymous message is sent to the app holder if he or she is exposed to the virus. A possible trade-off is that targeted quarantine may not be enforced as people who receive exposure may not strictly follow self-isolation guidelines.[28]

To examine the private notification model's effectiveness, I consider a counterfactual scenario in which the signal is privately observable, and only 40% of individuals choose to quarantine themselves upon receiving a bad signal. Under this counterfactual scenario, the economy will experience a second wave of infections when the lockdown is lifted after March, as shown in the purple line of Figure 10. The economy would suffer an additional

[28]For example, a July 22, 2020, *Washington Post* article entitled, "Australians are ignoring self-isolation guidelines. Coronavirus cases are climbing" reports that between July 7 and July 21, nearly 90% of people who tested positive in the state of Victoria did not self-isolate between when their symptoms began and when they were tested. More than 50% of people who tested positive in that same period did not self-isolate between the time they were tested for the virus and the time they got their results.

loss of 0.33% GDP compared with the baseline scenario, suggesting that the economic value created by the Health Code app would decrease by around 65%. In addition, the death toll would have been 16,000 higher than the baseline simulation with publicly verifiable signals. Overall, the private notification model reduces welfare by 0.5% GDP relative to the baseline simulation. This counterfactual exercise suggests a trade-off between the economic value of big data technology and privacy concern.

Does the value of privacy justify the efficiency loss due to privacy protection? This question is difficult to answer because the value of privacy is hard to measure. Nevertheless, I conduct a back-of-the-envelope comparison based on the estimated value of privacy in the literature. The most related estimate of how much Chinese people value privacy is from Tang (2019), who ran a field experiment in China and finds that the subjects are willing to give up their personal information for $33. This value amounts to 0.33% GDP per capita, which seems to be slightly smaller than the 0.50% GDP welfare loss resulting from the private notification model.[29]

This back-of-envelope calculation is subject to the caveat that people in different countries may value privacy differently. Furthermore, one's stated preference for privacy may differ from real preferences inferred from actual actions. Athey, Catalini, and Tucker (2017) conduct experiments in the United States and find that even people who claim to value privacy are willing to relinquish their private data for small benefits. It is also worth noting that using big data technologies may not necessarily lead to a loss of privacy because data anonymization technology can reduce the risk of privacy violations. For instance, many digital health certificate apps, such as the "Travel Pass" app from the International Air Transport Association (IATA) and the "CommonPass" app developed by the Commons Project, are designed to deliver a simple yes or no answer as to whether the individual meets the current entry criteria without transmitting the underlying data to a centralized database.

Overall, this counterfactual exercise suggests a trade-off between protecting the privacy and resolving informational frictions. This paper quantifies the value of the big data technology in the COVID-19 pandemic. Whether this value can justify the potential privacy cost requires a better understanding of how people value privacy, which is left for future research.

[29] The GDP per capita in China was around $10,000 as of 2018. See the World Bank data: https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?locations=CN.

4.7 Discussion on Adoption

A key necessary condition for the effectiveness of this big data technology is a high adoption rate. I consider a counterfactual situation in which the adoption rate is 20%. I find that the economy would have experienced a second wave after the lockdown was lifted, as shown in the green line in Figure 10. The economy would have incurred an additional 0.34% loss in GDP compared with the baseline simulation with high adoption rates as shown by Table 7. In other words, the economic value created by the Health Code app would have decreased by around 67%. The death toll would have been 64,000 higher than the baseline simulation with high adoption rates. The overall welfare loss would be 1% GDP higher than the baseline simulation.

An important contributing factor to the high adoption rates could be the app's design. The app combines contact tracing and health certification, which creates important synergies. On the one hand, digital contact tracing reduces the costs of digital health certification because users do not have to conduct regular diagnostic tests to certify their health. On the other hand, because it is easy to obtain such health certificates, private institutions and businesses can easily monitor the health condition in their business sites by setting up checkpoints. The prevalence of the Health Code checkpoints, in turn, encourages more adoption by individuals. Higher adoption also makes contact tracing more accurate. This mechanism has important implications in the ongoing debate on the design of contact tracing and health certificate apps.[30]

In addition to the app design, government coordination could have played a role. Instead of having different apps used by different private businesses, one app was introduced for the whole city, creating large network effects. Furthermore, people's familiarity with FinTech may have also contributed. China is the global leader in mobile payment adoption: over 81% of smartphone users had made a mobile payment in the past six months as of the end of 2019.[31] The Health Code app was developed by highly trusted FinTech firms in China and was integrated with popular mobile payment apps, which makes the adoption much easier.

[30] For instance, on March 4, 2021, France's Health Minister raised issues around whether the digital contact tracing app in France should be repurposed to a digital pass for public venues. See a March 4, 2021, *Politico* article, "France mulls digital COVID pass for public venues".

[31] See "Global proximity mobile payment usage penetration 2019, by country" survey by eMarketer.

5 Conclusion

The COVID-19 pandemic has featured one of the largest ever experiments with big data technologies. Governments and private institutions introduced numerous contact tracing and digital health certificate apps to combat the dual economic and public health crisis caused by the COVID-19 pandemic. Using the staggered introduction of the Health Code app across Chinese cities, this paper shows that such technologies can effectively alleviate economic and human losses caused by the COVID-19 pandemic. This large-scale experiment demonstrates big data's potential to solve large-scale social and economic problems caused by informational frictions. It also provides valuable lessons for the ongoing debate on the trade-off between big data technologies' economic potential and privacy concerns.

References

Acemoglu, D., A. Makhdoumi, A. Malekian, and A. Ozdaglar, 2019, "Too much data: Prices and inefficiencies in data markets," working paper, National Bureau of Economic Research.

Agarwal, S., P. Ghosh, T. Ruan, and Y. Zhang, 2020, "Privacy versus Convenience: Customer Response to Data Breaches of Their Information," *Available at SSRN 3729730*.

Alvarez, F., D. Argente, and F. Lippi, 2020, "A Simple Planning Problem for COVID-19 Lockdown," *University of Chicago, Becker Friedman Institute for Economics Working Paper*, (2020-34).

Angeletos, G.-M., and C. Lian, 2016, "Incomplete information in macroeconomics: Accommodating frictions in coordination," in *Handbook of macroeconomics*. Elsevier, vol. 2, pp. 1065–1240.

Argente, D. O., C.-T. Hsieh, and M. Lee, 2020, "The Cost of Privacy: Welfare Effect of the Disclosure of COVID-19 Cases," working paper, National Bureau of Economic Research.

Athey, S., C. Catalini, and C. Tucker, 2017, "The digital privacy paradox: Small money, small costs, small talk," working paper, National Bureau of Economic Research.

Atkeson, A., 2020, "What will be the economic impact of covid-19 in the us? rough estimates of disease scenarios," working paper, National Bureau of Economic Research.

Baker, S. R., R. A. Farrokhnia, S. Meyer, M. Pagel, and C. Yannelis, 2020, "How does household spending respond to an epidemic? consumption during the 2020 covid-19 pandemic," working paper, National Bureau of Economic Research.

Barrios, J. M., and Y. Hochberg, 2020, "Risk perception through the lens of politics in the time of the covid-19 pandemic," working paper, National Bureau of Economic Research.

Barro, R. J., J. F. Ursúa, and J. Weng, 2020, "The coronavirus and the great influenza pandemic: Lessons from the spanish flu for the coronavirus's potential effects on mortality and economic activity," working paper, National Bureau of Economic Research.

Begenau, J., M. Farboodi, and L. Veldkamp, 2018, "Big data in finance and the growth of large firms," *Journal of Monetary Economics*, 97, 71–87.

Berger, D. W., K. F. Herkenhoff, and S. Mongey, 2020, "An seir infectious disease model with testing and conditional quarantine," working paper, National Bureau of Economic Research.

Bethune, Z. A., and A. Korinek, 2020, "Covid-19 infection externalities: Trading off lives vs. livelihoods," working paper, National Bureau of Economic Research.

Bian, B., J. Li, T. Xu, and N. Foutz, 2020, "Individualism During Crises: Big Data Analytics of Collective Actions amid COVID-19," *Available at SSRN 3620364*.

Bordalo, P., K. B. Coffman, N. Gennaioli, and A. Shleifer, 2020, "Older people are less pessimistic about the health risks of COVID-19," working paper, National Bureau of Economic Research.

Borusyak, K., and X. Jaravel, 2017, "Revisiting event study designs," *Available at SSRN 2826228*.

Borusyak, K., X. Jaravel, and J. Spiess, 2021, "Revisiting event study designs: Robust and efficient estimation," *Unpublished working paper, version May*, 19, 2021.

Braithwaite, I., T. Callender, M. Bullock, and R. W. Aldridge, 2020, "Automated and partly automated contact tracing: a systematic review to inform the control of COVID-19," *The Lancet Digital Health*.

Buchak, G., G. Matvos, T. Piskorski, and A. Seru, 2018, "Fintech, regulatory arbitrage, and the rise of shadow banks," *Journal of Financial Economics*, 130(3), 453–483.

Budish, E. B., 2020, "R<1 as an Economic Constraint: Can We Expand the Frontier in the Fight Against Covid-19?," *University of Chicago, Becker Friedman Institute for Economics Working Paper*, (2020-31).

Charoenwong, B., A. Kwan, and V. Pursiainen, 2020, "Social connections to COVID-19-affected areas increase compliance with mobility restrictions," *Available at SSRN*.

Chen, H., W. Qian, and Q. Wen, 2020, "The impact of the COVID-19 pandemic on consumption: Learning from high frequency transaction data," *Available at SSRN 3568574*.

Correia, S., S. Luck, and E. Verner, 2020, "Pandemics Depress the Economy, Public Health Interventions Do Not: Evidence from the 1918 Flu," .

Coven, J., and A. Gupta, 2020, "Disparities in mobility responses to covid-19," working paper, NYU Stern Working Paper.

Dewatripont, M., M. Goldman, E. Muraille, and J.-P. Platteau, 2020, "Rapid identification of workers immune to COVID-19 and virus-free: A priority to restart the economy," working paper, Discussion paper, Universit Libre de Bruxelles.

Eichenbaum, M. S., S. Rebelo, and M. Trabandt, 2020, "The macroeconomics of epidemics," working paper, National Bureau of Economic Research.

Elenev, V., T. Landvoigt, and S. Van Nieuwerburgh, 2020, "Can the Covid Bailouts Save the Economy?," working paper, National Bureau of Economic Research.

Fang, H., L. Wang, and Y. Yang, 2020, "Human mobility restrictions and the spread of the novel coronavirus (2019-ncov) in china," working paper, National Bureau of Economic Research.

Farboodi, M., G. Jarosch, and R. Shimer, 2020, "Internal and external effects of social distancing in a pandemic," working paper, National Bureau of Economic Research.

Fernández-Villaverde, J., and C. I. Jones, 2020, "Estimating and Simulating a SIRD Model of COVID-19 for Many Countries, States, and Cities," working paper, National Bureau of Economic Research.

Ferretti, L., C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, and C. Fraser, 2020, "Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing," *Science*.

Fisman, R., H. Lin, C. Sun, Y. Wang, and D. Zhao, 2021, "What motivates non-democratic leadership: Evidence from COVID-19 reopenings in China," *Journal of Public Economics*, 196, 104389.

Gallant, J., K. Kroft, F. Lange, and M. J. Notowidigdo, 2020, "Temporary unemployment and labor market dynamics during the COVID-19 recession," working paper, National Bureau of Economic Research.

Goldstein, I., W. Jiang, and G. A. Karolyi, 2019, "To FinTech and beyond," *The Review of Financial Studies*, 32(5), 1647–1661.

Goolsbee, A., and C. Syverson, 2020, "Fear, lockdown, and diversion: comparing drivers of pandemic economic decline 2020," working paper, National Bureau of Economic Research.

Guerrieri, V., G. Lorenzoni, L. Straub, and I. Werning, 2020, "Macroeconomic implications of COVID-19: Can negative supply shocks cause demand shortages?," working paper, National Bureau of Economic Research.

Hall, R. E., C. I. Jones, and P. J. Klenow, 2020, "Trading off consumption and covid-19 deaths," working paper, National Bureau of Economic Research.

Hassan, T. A., S. Hollander, L. van Lent, and A. Tahoun, 2020, "Firm-level exposure to epidemic diseases: Covid-19, SARS, and H1N1," working paper, National Bureau of Economic Research.

Hong, H., N. Wang, and J. Yang, 2020, "Mitigating Disaster Risks to Sustain Growth," working paper, National Bureau of Economic Research.

Hortaçsu, A., J. Liu, and T. Schwieg, 2020, "Estimating the fraction of unreported infections in epidemics with a known epicenter: an application to covid-19," working paper, National Bureau of Economic Research.

Jones, C., T. Philippon, and V. Venkateswaran, 2020, "Optimal Mitigation Policies in a Pandemic," working paper, Covid Economics Issue 4.

Jones, K. E., N. G. Patel, M. A. Levy, A. Storeygard, D. Balk, J. L. Gittleman, and P. Daszak, 2008, "Global trends in emerging infectious diseases," *Nature*, 451(7181), 990–993.

Keshta, A. S., S. I. Mallah, K. Al Zubaidi, O. K. Ghorab, M. S. Keshta, D. Alarabi, M. A. Abousaleh, M. T. Salman, O. E. Taha, A. A. Zeidan, et al., 2021, "COVID-19 versus SARS: A comparative review," *Journal of Infection and Public Health*.

Kniesner, T. J., W. K. Viscusi, C. Woock, and J. P. Ziliak, 2012, "The value of a statistical life: Evidence from panel data," *Review of Economics and Statistics*, 94(1), 74–87.

Kozlowski, J., L. Veldkamp, and V. Venkateswaran, 2020, "Scarring body and mind: the long-term belief-scarring effects of Covid-19," *FRB St. Louis Working Paper*, (2020-009).

Morris, S. D., and J. Zhang, 2019, "Validating China's Output Data Using Satellite Observations," *Macroeconomic Dynamics*, 23(8), 3327–3354.

Nakamura, E., and J. Steinsson, 2018, "Identification in macroeconomics," *Journal of Economic Perspectives*, 32(3), 59–86.

Piguillem, F., and L. Shi, 2020, "Optimal COVID-19 quarantine and testing policies," .

Ramadorai, T., A. Uettwiller, and A. Walther, 2019, "The market for data privacy," *Available at SSRN 3352175*.

Redding, S. J., E. L. Glaeser, and C. Gorback, 2020, "How Much Does COVID-19 Increase with Mobility? Evidence from New York and Four Other US Cities," .

Rodríguez, P., S. Graña, E. E. Alvarez-León, M. Battaglini, F. J. Darias, M. A. Hernán, R. López, P. Llaneza, M. C. Martín, O. Ramirez-Rubio, A. Romaní, B. Suárez-Rodríguez, J. Sánchez-Monedero, A. Arenas, and L. Lacasa, 2021, "A population-based controlled experiment assessing the epidemiological impact of digital contact tracing," *Nature communications*, 12(1), 1–6.

Stiglitz, J. E., 2002, "Information and the Change in the Paradigm in Economics," *American economic review*, 92(3), 460–501.

Tang, H., 2019, "The value of privacy: Evidence from online borrowers," working paper, Working Paper, HEC Paris.

Veldkamp, L., and C. Chung, 2019, "Data and the aggregate economy," in *Annual Meeting Plenary*, no. 2019-1. Society for Economic Dynamics.

Veldkamp, L. L., 2011, *Information choice in macroeconomics and finance*. Princeton University Press.

Wymant, C., L. Ferretti, D. Tsallis, M. Charalambides, L. Abeler-Dörner, D. Bonsall, R. Hinch, M. Kendall, L. Milsom, M. Ayres, C. Holmes, M. Briers, and C. Fraser, 2021, "The epidemiological impact of the NHS COVID-19 App," .

Zhu, C., 2019, "Big data as a governance mechanism," *The Review of Financial Studies*, 32(5), 2021–2061.

Table 1: **Summary Statistics**

	N	mean	sd	p5	p25	p50	p75	p95
NO2	24742.0	63.0	30.3	22.9	40.1	57.5	81.4	119.5
SO2	24742.0	82.8	36.7	33.9	58.8	79.1	100.0	145.3
Confirmed cases	28658.0	144.3	1985.5	0.0	0.0	8.0	31.0	213.0
Cured cases	28658.0	82.6	1251.6	0.0	0.0	3.0	18.0	140.0
Deaths	28658.0	5.0	91.1	0.0	0.0	0.0	0.0	3.0
Emergency level	28658.0	1.7	1.2	0.0	0.0	2.0	3.0	3.0
Within-city movements	28658.0	77.5	25.6	32.3	56.3	84.1	99.1	109.3
FinTech Penetration Index	26522.0	0.8	0.1	0.7	0.7	0.7	0.8	0.9

Note: This table reports summary statistics of the regression sample. The sample is a panel of 322 cities from January 1, 2020, to March 31, 2020. Data sources: Baidu Migration, China National Environmental Monitoring Center, Chinese Center for Disease Control and Prevention.

Table 2: **Effects of the Health Code on NO2 Emissions: Event Study**

	(1) NO2	(2) NO2	(3) NO2
Week -2	0.008	0.021	0.010
	[0.024]	[0.028]	[0.024]
Week -1	0.034	0.041	0.035
	[0.025]	[0.030]	[0.025]
Week 0	0.036	0.043	0.038
	[0.026]	[0.032]	[0.026]
Week 1	0.081***	0.091**	0.082***
	[0.030]	[0.036]	[0.030]
Week 2	0.139***	0.137***	0.140***
	[0.034]	[0.039]	[0.035]
Week 3	0.206***	0.207***	0.205***
	[0.040]	[0.046]	[0.040]
Week 4	0.244***	0.245***	0.239***
	[0.047]	[0.054]	[0.047]
Week 5	0.325***	0.320***	0.316***
	[0.054]	[0.062]	[0.054]
Week 6+	0.445***	0.464***	0.432***
	[0.070]	[0.077]	[0.067]
Constant	0.268***	0.253***	0.266***
	[0.040]	[0.044]	[0.040]
Control	Yes	Yes	Yes
City F.E.	Yes	Yes	Yes
Time F.E.	Yes	Yes	Yes
Emergency F.E.	Yes	Yes	Yes
Sample	Full sample	Match by cases	Match by act.
Observations	24,742	24,742	24,742
Adj. R-squared	0.562	0.573	0.562

Note: This table reports the effects of the Health Code on economic activity measured by daily NO2 emissions. The dependent variable is NO2 emission normalized by the average of the first two weeks in 2020. Week k represents the effect k weeks from the launch date. Week 6+ represents the effects 6 weeks or more after the launch date. The control variable is within-city movement. The sample is a panel of 322 cities from January 1, 2020 to March 31, 2020. Standard errors are clustered at both the date and the city level. Data sources: Baidu, Chinese Center for Disease Control and Prevention.

Table 3: **Effects of the Health Code on Infections: Event Study**

	(1) Infection	(2) Infection	(3) Infection
Week -2	0.013	0.034	0.021
	[0.076]	[0.084]	[0.074]
Week -1	0.017	0.031	0.021
	[0.059]	[0.060]	[0.058]
Week 0	0.010	0.037	0.026
	[0.068]	[0.075]	[0.073]
Week 1	-0.228***	-0.195**	-0.219**
	[0.083]	[0.089]	[0.082]
Week 2	-0.717***	-0.682***	-0.704***
	[0.078]	[0.090]	[0.083]
Week 3	-1.048***	-0.992***	-1.046***
	[0.149]	[0.171]	[0.139]
Week 4	-0.723***	-0.676***	-0.699***
	[0.121]	[0.149]	[0.108]
Week 5	-0.881***	-0.846***	-0.849***
	[0.177]	[0.203]	[0.149]
Week 6+	-0.823***	-0.781***	-0.759***
	[0.226]	[0.255]	[0.202]
Constant	-0.078	-0.064	-0.086
	[0.067]	[0.082]	[0.064]
Control	Yes	Yes	Yes
City F.E.	Yes	Yes	Yes
Time F.E.	Yes	Yes	Yes
Emergency F.E.	Yes	Yes	Yes
Sample	Full sample	Match by cases	Match by act.
Observations	1,899	1,899	1,899
Adj. R-squared	0.145	0.161	0.144

Note: This table reports the effects of the Health Code on infection growth. The dependent variable is infection growth rate measured by three-day moving average of changes in log new cases. Week k represents the effect k weeks from the launch date. Week 6+ represents the effects 6 weeks or more after the launch date. The control variable is within-city movement. The sample is a panel of 322 cities from January 1, 2020 to March 31, 2020. Standard errors are clustered at both the date and the city level. Data sources: Baidu, Chinese Center for Disease Control and Prevention.

Table 4: **Effects of the Health Code on NO2 Emissions: Robustness**

	(1) NO2	(2) NO2	(3) NO2
tau	0.084***	0.065***	0.084***
	[0.026]	[0.020]	[0.028]
Control	Yes	Yes	Yes
City F.E.	Yes	Yes	Yes
Time F.E.	Yes	Yes	Yes
Emergency F.E.	Yes	Yes	Yes
Sample	Full sample	Match by cases	Match by act.
Observations	22,796	22,796	22,796

Note: This table reports the effects of the Health Code on economic activity measured by daily NO2 emissions using the imputation estimator proposed by Borusyak, Jaravel, and Spiess (2021). The dependent variable is NO2 emission normalized by the average of the first two weeks in 2020. "tau" represents the average treatment effect of the Health Code over the treated cities. The sample is a panel of 322 cities from January 1, 2020 to March 31, 2020. Standard errors are clustered at both the date and the city level. Data sources: Baidu, Chinese Center for Disease Control and Prevention.

Table 5: **Health Code, Economic Activity, and Public Health: Instrumental Variable Approach**

	(1) NO2	(2) SO2	(3) Infection	(4) Death
Health Code Earliness	3.336***	1.487**	-1.588***	-2.851***
	[0.505]	[0.616]	[0.509]	[0.629]
Within-city movements	0.007***	0.010***	-0.005***	-0.006***
	[0.002]	[0.002]	[0.002]	[0.002]
Confirmed cases	0.014	0.040**	-0.025*	0.200***
	[0.013]	[0.016]	[0.013]	[0.016]
Emergency F.E.	Yes	Yes	Yes	Yes
Observations	220	220	235	235
Adj. R-squared	0.196	0.100	0.077	0.446

Note: This table reports the results of the following two-stage regression model

$$\text{Health Code Earliness}_i = \gamma \text{FinTech penetration}_i + u_i,$$
$$\Delta Y_i = \beta \text{Health} \widehat{\text{Code Earliness}}_i + \eta X_i + \epsilon_i,$$

where ΔY_i is the changes in economic activity or public health condition from February 9, 2020, to March 31, 2020 of city i. Health Code Earliness$_i$ is the fraction of time that the Heath Code app is in place from February 9, 2020, to March 31, 2020. Health Code$_i$ is instrumented by FinTech penetration$_i$ measured in 2018. The vector of the control variables, X_i, includes within-city movement, public health emergency levels, and the log number of confirmed case as of February 9, 2020. Data sources: Baidu, Chinese Center for Disease Control and Prevention.

Table 6: **Parameter Values for SIR Model with Health Code**

Parameter	Value	Definition
$1 - p_i$	0.47	False negative rate
$1 - p_s$	0.02	False positive rate
α	0.95	Adoption rate
ϕ	0.40	Effectiveness of population movement restrictions
ψ	0.19	Other disease control measures
θ	0.03	Elasticity of population movement reduction due to undetected infections
I_0	1250	Initial infection number
λ	0.12	Economic costs of lockdown
β	0.20	Daily increase of active cases without intervention
κ	0.05γ	Fatalities per active case (per day)
γ	1/18	Daily rate of infected recovery (includes those that die)
χ	150	Value of statistical life

Table 7: **Counterfactual Simulations**

Baseline	Economic loss	Death	Welfare loss
	4.26	4650	4.31
Counterfactuals		Relative to baseline	
No Health Code	+0.51	+216385	+3.01
No lockdown	-1.10	+74755	-0.23
Private signal	+0.33	+15525	+0.50
Low adoption	+0.34	+64147	+1.08

(a) Hangzhou Health Code

(b) Subway entrance in Hangzhou

Figure 1: **Health Code and Checkpoints**
Figure 1a shows the first Health Code introduced in China, the Hangzhou Health Code. Individuals with a green code can freely travel in the city. Individuals with a yellow code must quarantine for 7 days. Individuals with a red code must quarantine for 14 days. The code turns back to green after the corresponding quarantine period. Figure 1b shows a subway guard in Hangzhou checking phones while helping a man set up the Alipay Health Code software. Everyone must have a green code to pass. Source: *New York Times*.

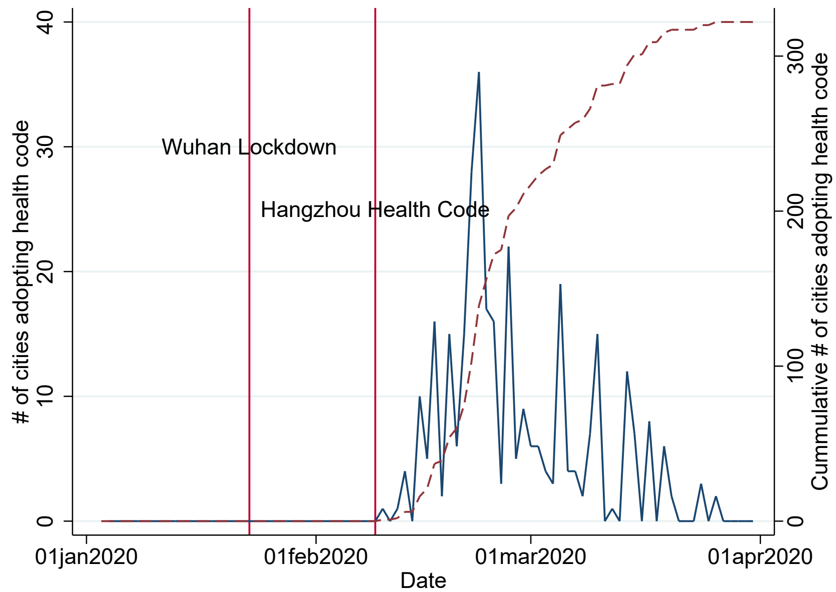


Figure 2: **Implementation of Health Code in Chinese Cities**
This figure plots the number of cities that launch Health Code on each date. The first vertical line indicates January 23, 2020, the date of Wuhan's lockdown. The second vertical line indicates February 9, 2020, when the first Health Code was introduced in Hangzhou.

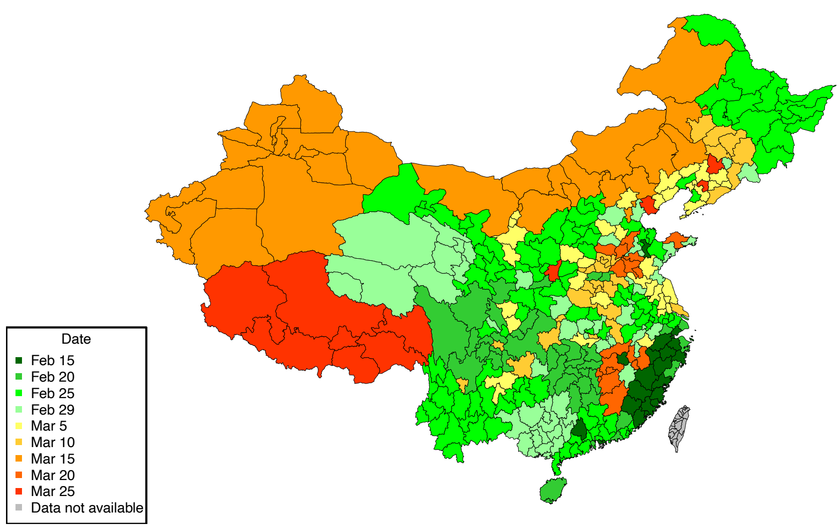
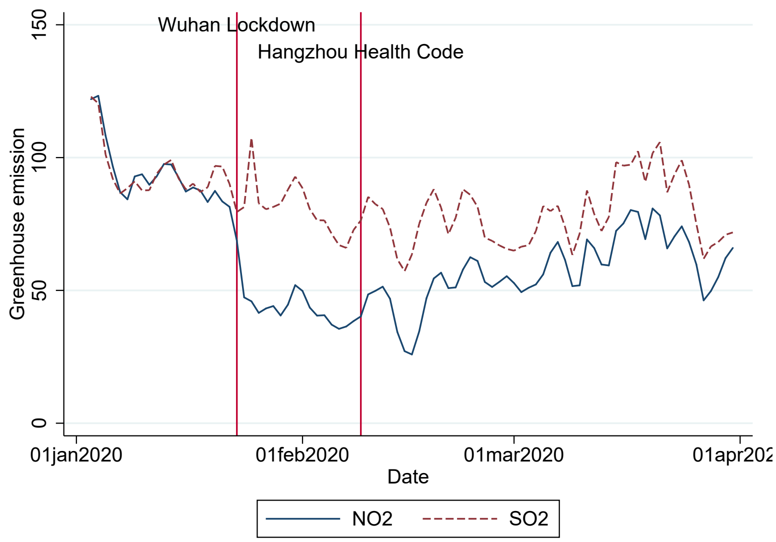
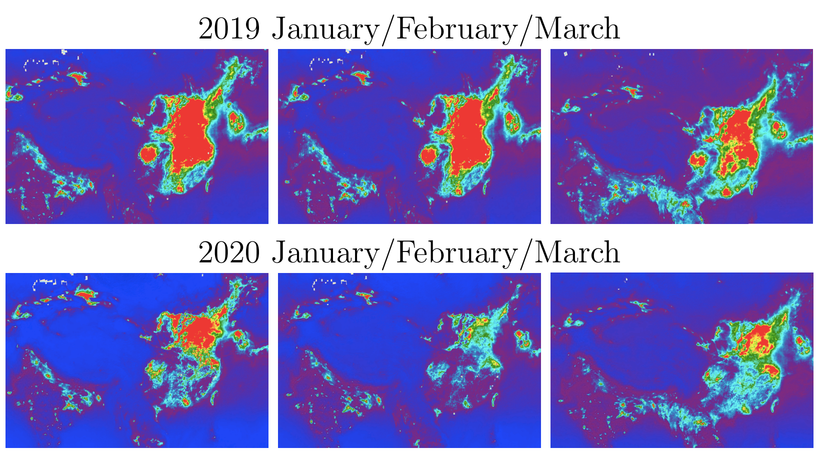


Figure 3: **Implementation of Health Code in Chinese Cities**
This figure shows the launch dates of the Health Code app in five-day intervals. Data source: local government websites and news reports.



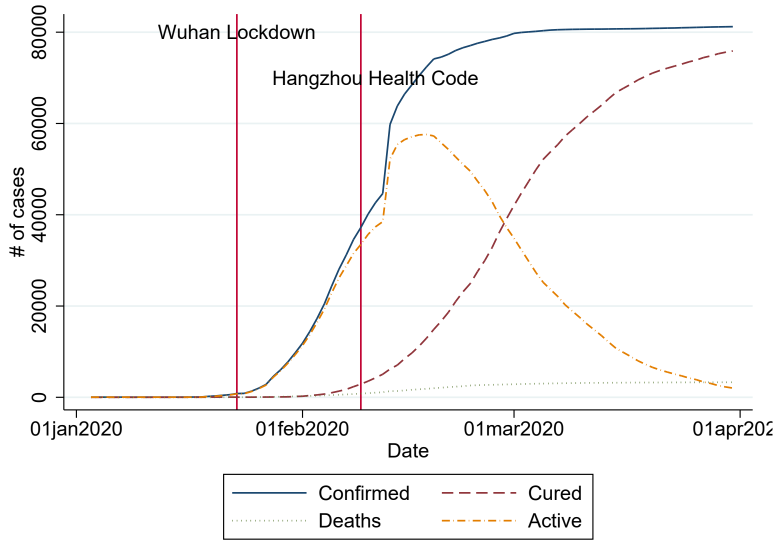
(a) Average emissions over time



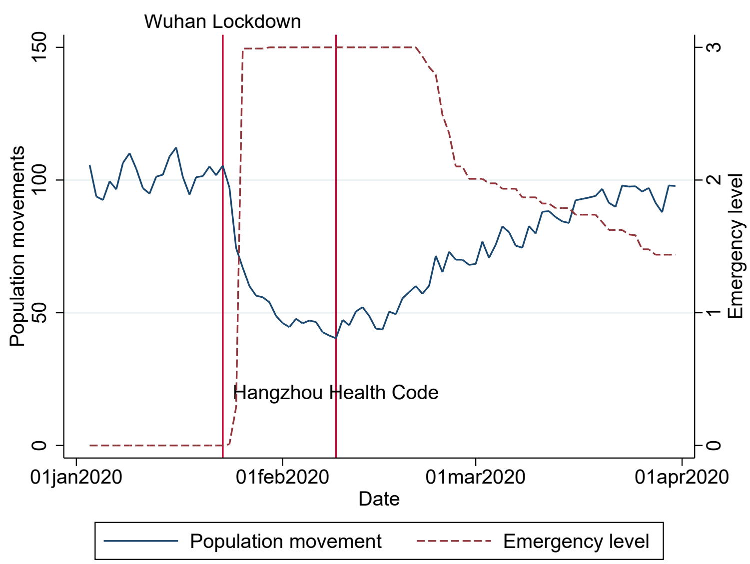
(b) Satellite Images of NO2 Levels over China

Figure 4: **Greenhouse Gas Emission**

Figure 4a shows the average economic activity of Chinese cities measured by greenhouse gas emissions. Data source: the China National Environmental Monitoring Center (CNEMC). Figure 4b shows the heat maps of nitrogen dioxide (NO2) levels over China. Data source: Google Earth.



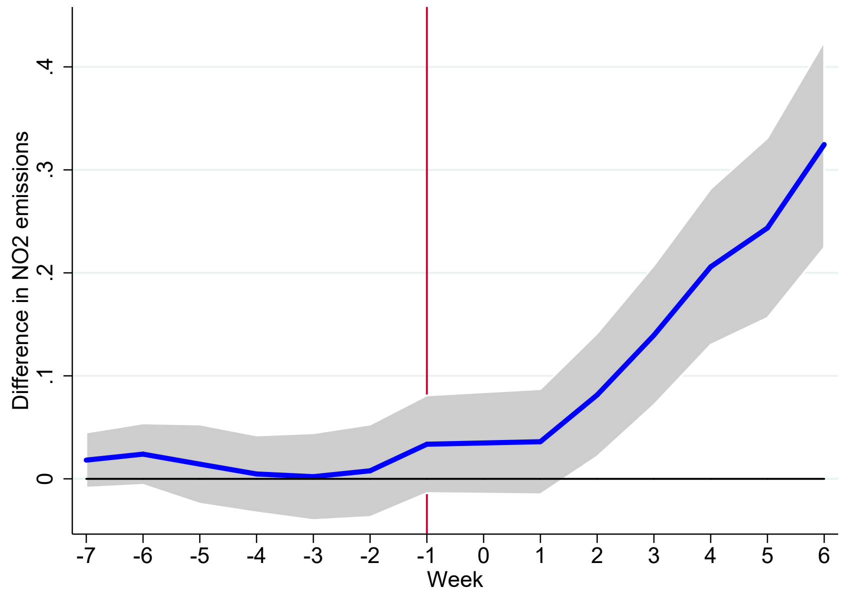
(a) Infections and Deaths



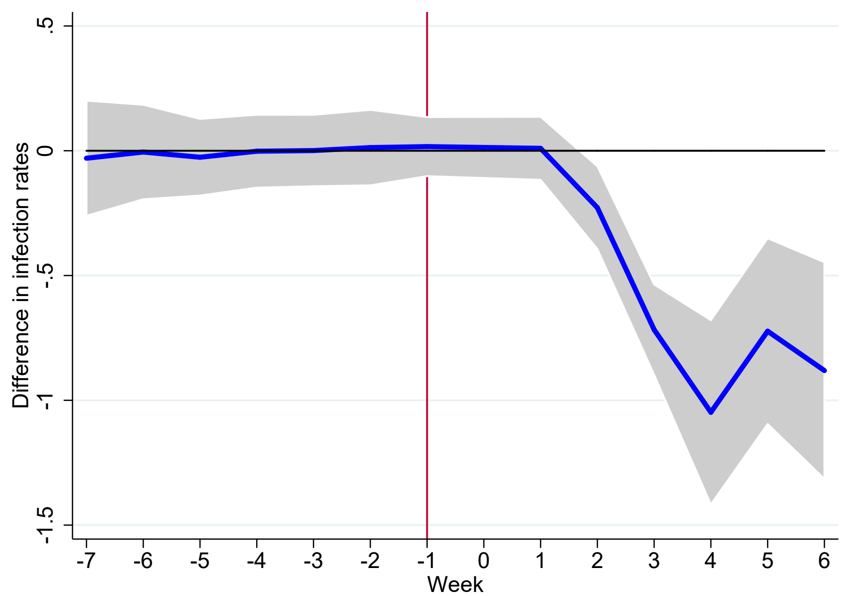
(b) Population Movements and Emergency Level

Figure 5: **COVID Infections, Population Movements, and Emergency Level**

Figure 5a plots the cumulative confirmed, cured, dead, and current cases of COVID-19 in the sample. Figure 5b plots the average within-city population movements and public health emergency level. Data source: Chinese Center for Disease Control and Prevention, Baidu Migration.



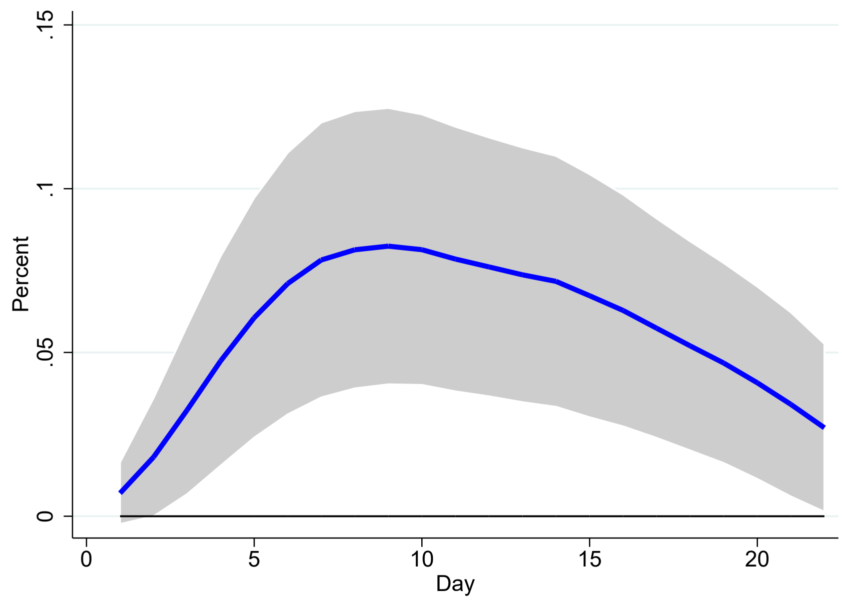
(a) NO2 emissions



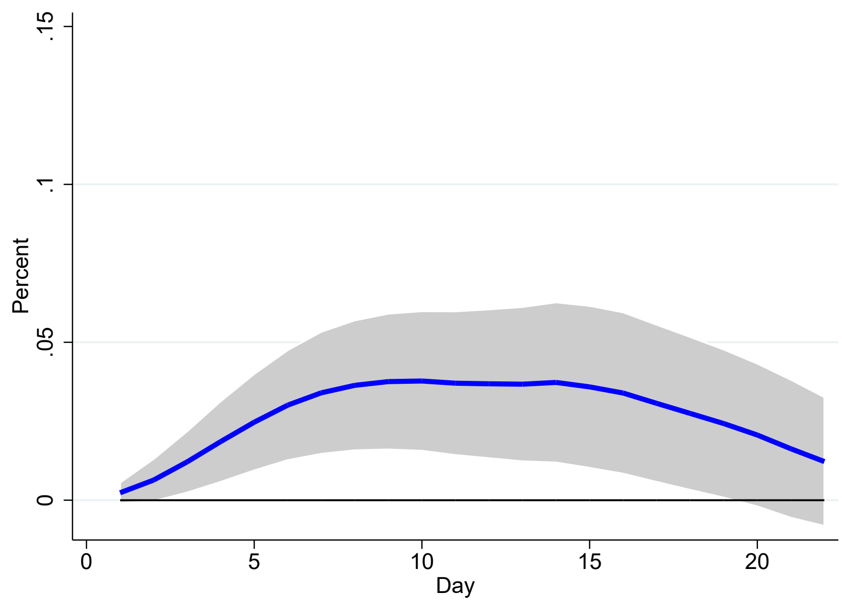
(b) Infection rates

Figure 6: **Dynamic Effects of the Health Code Introduction**

This figure plots the dynamic effects of Health Code 14 weeks around the launch date of the Health Code app. The horizontal axis is the week relative to the launch date of the Health Code app. Data source: the China National Environmental Monitoring Center (CNEMC) and Chinese Center for Disease Control and Prevention.



(a) Without Health Code



(b) With Health Code

Figure 7: **Impulse Response of Local Cases to Infection Inflows**
This figure plots the impulse response of locally confirmed cases to a 1% increase in infection inflows. The upper and lower panels show the effect without and with Health Code, respectively. The horizontal axis is the date since the infected inflows have occurred. Standard errors are clustered at both the date and the city level. Data source: Baidu, Chinese Center for Disease Control and Prevention.

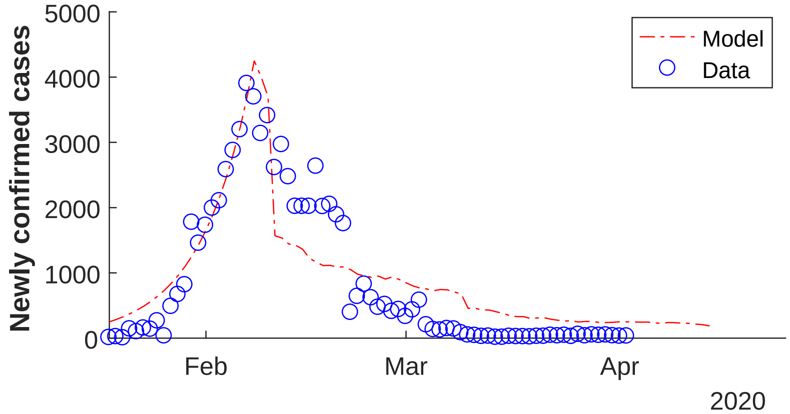


Figure 8: **New Daily Cases in Simulation and Data**
This figure plots shows new daily cases in simulation and in data. The sample period is from January 16th, 2020, to March 31, 2020. The parameters used for the simulation are reported in Table 6.

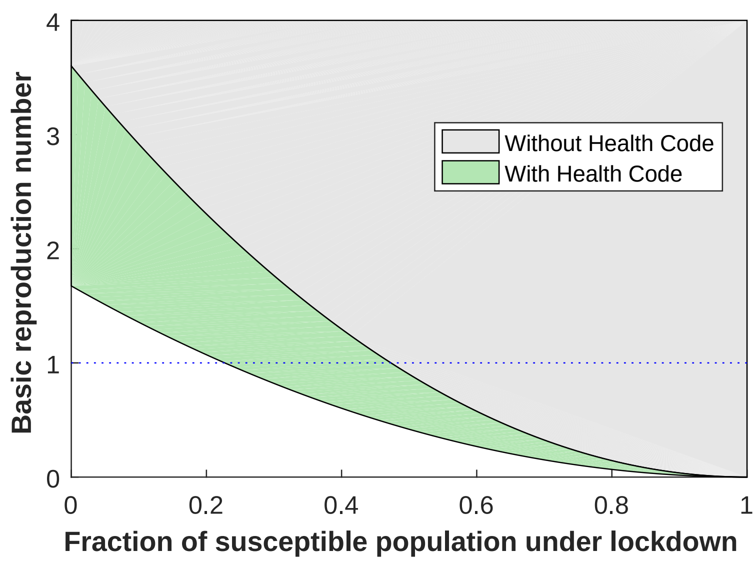


Figure 9: **Infection-Possibilities Frontier**
This figure plots the infection-possibilities frontier, which shows the possible combinations of reproduction number of a pandemic and the fraction of population under lockdown for a given set of technology. The gray region is the set of possible combinations without Health Code. The green region is the additional possibilities created by Health Code. The parameters used to construct these two curves are reported in Table 6.

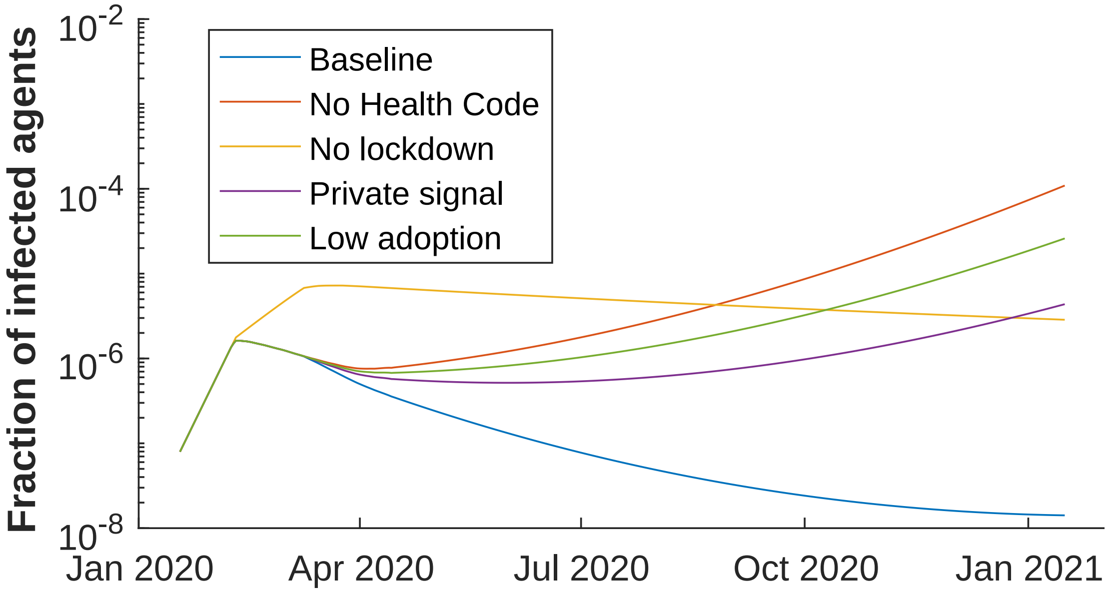


Figure 10: **Counterfactual Paths of Infections**

This figure shows the counterfactual paths of infections. The blue line is the baseline case with both Health Code and lockdown. The red line is a counterfactual case in which Health Code is absent. The yellow line is a counterfactual case in which a lockdown is absent. The purple line is a counterfactual case in which the signal is private and 40% of agents with bad signals choose to self-quarantine. The green line is a counterfactual case in which the adoption rate is 20%. The parameters used for the simulation are reported in Table 6.