The COVID-19 Pandemic and Unemployment: Evidence from Mobile Phone Data from China*

Teng Li   Panle Jia Barwick   Yongheng Deng   Xinfei Huang   Shanjun Li

January 2023

Abstract

Based on mobile phone records for 71 million users and location tracking information for one million users over almost three years, this study examines the labor market impacts of the COVID-19 pandemic in China's Guangdong province, whose GDP is larger than that of all but the top 12 countries in the world. Using a difference-in-differences framework, our analysis shows dramatic and protracted effects of the pandemic on the labor market: it increased unemployment by 72% and unemployment benefits claims by 57% after the full reopening in 2020 relative to their levels during the same period in 2019. The impact was highly heterogeneous, with women, workers older than 40, and migrants being particularly affected. Cities that rely more on export or have a higher share of the hospitality industry in GDP but a lower share of the finance and health care industries experienced a more pronounced increase in unemployment. The lingering impact likely reflects the global nature of the pandemic, the impact of which has propagated through the supply chain and trade channels.

JEL Classification: E24, H12
Keywords: COVID-19, Mobile Phone, Recovery, Unemployment

*Teng Li: Sun Yat-sen University (e-mail: liteng27@mail.sysu.edu.cn); Panle Jia Barwick: University of Wisconsin–Madison and NBER (e-mail: pbarwick@wisc.edu); Yongheng Deng (corresponding author): University of Wisconsin–Madison (e-mail: yongheng.deng@wisc.edu); Xinfei Huang: Sun Yat-sen University (e-mail: huangxf3@mail.sysu.edu.cn); Shanjun Li: Cornell University and NBER (e-mail: sl2448@cornell.edu).
We thank Lin Ma, participants of the 9th Annual Conference of the ABFER, and seminar participants at the University of Wisconsin–Madison, for helpful comments and Chris Chen for excellent research assistance. Panle Barwick and Shanjun Li gratefully acknowledge support from the Cornell Center for Social Sciences and Cornell Atkinson Center for Sustainability. Xinfei Huang gratefully acknowledges support from the Major Project of the National Social Sciences Fund (16ZDA042). Teng Li gratefully acknowledges support from the Youth Program of the National Natural Science Foundation of China (72003202).

1 Introduction

Effective and targeted policies to address the adverse consequences of the COVID-19 pandemic for the economy rely on prompt and accurate measures of the labor market effects across different demographic groups and geographic regions. Traditional measures of labor market outcomes, in particular unemployment rates, are based on surveys. However, in addition to the substantial time lag and limited availability for small geographic areas associated with survey data, statistics inferred from surveys suffer from considerable uncertainty and are routinely revised.[1]

In China, information on unemployment is derived from the number of individuals who registered with unemployment benefit agencies prior to 2018 and is supplemented by household survey data for the years afterward.[2] Measuring unemployment accurately is particularly challenging in the Chinese context due to the large fraction of the population who do not have local household registrations (*Hukou*) and hence are excluded from the unemployment surveys. In addition, reporting and aggregation errors, as well as potential data manipulations, have been documented (Giles et al., 2005; Liu, 2012; Cai et al., 2013). China's official national unemployment rate has varied within a tight range of 3.1%–4.3% over the past two decades, leading to questions about its reliability (Feng et al., 2017), especially in the face of the rapid and unprecedented social and economic changes brought about by the pandemic.[3]

This study leverages high-frequency and high-resolution mobile phone usage data from Guangdong, the most populous province in China, with a GDP larger than that of all but the world's top 12 countries. Our primary data source consists of location tracking information

[1] In addition to the delay, it is technically challenging to measure unemployment accurately from survey data. Their quality also varies considerably over time due to changes in participation rates, modifications to survey methodologies, inconsistencies and measurement errors in sample responses, and rotation group bias (Poterba and Summers, 1986; Jones and Riddell, 1999; Card, 2011; Feng et al., 2017; Meyer et al., 2015; Krueger et al., 2017; Heffetz and Reeves, 2019).

[2] See, for example, the report at https://www.marketwatch.com/story/china-unveils-urban-survey-unemployment-rate-2018-04-17-1485030.

[3] Please refer to the *Wall Street Journal* article at https://www.wsj.com/articles/chinas-jobs-rebound-doesnt-appear-as-robust-as-the-government-claims-11591551390.

for one million randomly selected users and mobile phone records for 71 million users from January 2018 to September 2020. We examine the pandemic's labor market impacts for various demographic groups and across cities with different industrial structures by employing the standard difference-in-differences (DID) framework. We use the observations from the year 2020 as the treatment group and those from the year 2019 as the control group. The key identification assumption is that the labor market outcomes would have tracked between the two groups in the absence of the pandemic and hence that the observed differences can be attributed to the pandemic rather than to time-varying unobservables. Results from event studies provide strong support for the assumption of common trends between the two groups prior to the event date.

We leverage two unique data features to estimate the impact on the labor market: a) the number of individuals who stopped commuting to work for an extended period of time (noncommuters) as a measure of unemployment, and b) the number of unique individuals who contacted the unemployment benefit agencies via the designated hotline (12333) as a measure of unemployment benefit claims. We first validate these two measures and then provide several pieces of evidence to show that the identified impact on unemployment is unlikely to be driven by the shift to work-from-home (WFH), a key confounder in interpreting the results based on commuting patterns. We also conduct a host of robustness checks and find that our results are robust to alternative variable definitions, data selection and model specifications.

Several key findings emerge from our analysis. First, the pandemic increased unemployment in Guangdong by 72% and unemployment benefit claims by 57% after the full reopening relative to those during the same period (from May to September) in 2019. The effect does not show a diminishing trend within the five-month window before September 2020, the end of our data period. The sharp rise in unemployment is much higher than that reported in the government statistics, which registered an increase of 13.3% in Guangdong province's unemployment rate (from 2.26 percentage points in January–March to 2.56 percentage points in

July–September) for the same period.

Second, the pandemic's impact on unemployment was highly uneven across demographic groups and more pronounced among women, people over 40, and especially migrants. The escalating increase in unemployment among migrants shows no sign of abatement during our sample period. This echoes a massive reduction in the number of migrant workers reported by the National Bureau of Statistics (NBS) and indicates the possibility of large-scale layoffs among this group.[4]

Third, the pandemic's impact was more substantial in cities with a high labor share of the hospitality, real estate, or transportation industries but less severe in cities where employment is concentrated in the finance, health care, or education industries. In addition, the impact was more pronounced in cities that rely heavily on export, reflecting the global nature of the shock in the interconnected world economy. Industry composition accounts for 39% of the heterogeneity in the pandemic's unemployment impact across cities, while trade exposure contributes 29% of the heterogeneity.

Last, our results reveal the severity and uneven character of the pandemic's labor market impacts, which speak to the importance of conducting analysis at granular levels. In addition, these results illustrate the ripple effect of the pandemic across cities within the country and across countries worldwide through the supply chain and trade channels. A city's (or country's) industry composition, its exposure to trade, and the nature of the supply chain are crucial determinants of the pandemic's effects on its economy. Our measures help us understand the pandemic's labor market impact at a granular level and can inform targeted policies to help the most severely affected groups and regions.

The key contribution of our study is twofold. First, our study adds to the emerging literature that leverages granular and high-frequency mobile phone data to better measure economic and social activities. Examples include studies that use mobile phone data to

[4]According to the NBS statistics, the number of migrant workers decreased by 3.8 million in September 2020 from that in 2019. Please refer to the NBS article at http://www.stats.gov.cn/tjsj/sjjd/202010/t20201019_1794729.html.

improve labor market measurements (Toole et al., 2015; Barwick et al., 2019), track human movement in real time and at a fine spatial scale, and understand mobility, knowledge spillovers, or racial disparities in voting wait times (González et al., 2008; Ahas et al., 2010; Couronné et al., 2013; Chen et al., 2018; Kreindler and Miyauchi, 2021; Chen et al., 2020; Atkin et al., 2020; Chen and Pope, 2020). Our study illustrates how this type of data can also be used to understand the labor market impact of the pandemic in near real time.

Second, our study contributes to the burgeoning literature on the impacts of the pandemic. This literature is now too large to discuss in detail, but our paper is closely related to studies that use mobile phone data to measure mobility and social distancing after the pandemic (Couture et al., 2020; Gupta et al., 2020) and those that focus on the labor market impacts of the pandemic (Adams-Prassl et al., 2020; Ahn and Hamilton, 2020; Alon et al., 2020; Bick and Blandin, 2020; Coibion et al., 2020; Cajner et al., 2020). Compared to the sample data in these studies, our mobile phone data benefit from a large sample size and fine resolution in both temporal and spatial dimensions, and our study is the only one on China.

The rest of this paper is organized as follows: Section 2 discusses the context of the study and provides descriptive evidence. Section 3 lays out the empirical framework. Sections 4 and 5 present the event-study and regression results, and Section 6 concludes.

2 Background and Data

2.1 Background and Data Sources

Exploiting the increasing availability of high-frequency and high-resolution mobile phone data is particularly advantageous in the context of China, as its cellphone penetration rate is high among developing countries. In the 2018 wave of the China Family Panel Studies, a nationally representative longitudinal survey of individuals' social and economic status, 89% of correspondents sixteen years and older reported possessing a cellphone. In addition, each household owns 2.5 cell phones on average according to data for 2018 from the National Bureau of Statistics. Appendix Figure A.1 shows a strong correlation between the number of mobile phone 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.

The context of our analysis is Guangdong, the most populous province with the largest provincial GDP in China. Guangdong contributes 11% of China's GDP and approximately a quarter of China's foreign trade (China Statistical Yearbook 2020). Its major cities include Shenzhen and Guangzhou, which rank among the wealthiest and most economically advanced cities in China. Within Guangdong province, cities differ substantially in terms of both population and GDP, as Table A.1 illustrates.[5] The economy of Guangdong is widely recognized as the most dynamic and resilient among all provinces in China (World Bank, 2010; Gong et al., 2020). Another reason for Guangdong's relevance is that its number of daily confirmed COVID-19 cases—like that of most other provinces in China—has been under a few dozen since the full reopening (Appendix Figure A.2). Our measures of the pandemic's consequences could thus apply to other regions as well.

Our data come from one dominant cellular service provider in China. We have access to detailed phone usage data (encrypted IDs of the calling party and the receiving party, date of calls, and call duration in seconds) from January 2018 to September 2020 for all of the provider's 71 million users in Guangdong province, who account for 63% of all mobile users in the province. We observe some user demographic information, such as age, gender, and the city where the phone number is registered. In addition, we observe geocoded phone locations at five-minute intervals for one million randomly selected users during the same period.

The cellular service provider delineates Guangdong province into 787 cell tower areas

[5]Among the 21 cities, Guangzhou had the largest population (15.31 million) in 2019, while Yunfu had a population of only 2.55 million. The economic scale of the largest city, Shenzhen, at $390 billion in 2019, is almost 30 times that of the smallest city, Yunfu.

(similar to zip codes in the U.S.) for billing purposes. We use "cell tower areas" and "neighborhoods" interchangeably in this analysis.

Guangdong's lockdown Guangdong's provincial government acted swiftly and adopted vigilant procedures at the onset of the pandemic. Guangdong was one of the first provinces to release detailed information on newly confirmed cases (daily new cases, location, gender, etc.), starting from as early as February 3, 2020. These procedures proved successful and have kept the number of daily confirmed cases under a few dozen since the full reopening. As shown in Figure A.2, the daily confirmed new cases peaked at 254 on January 31 and quickly declined to under 50 three weeks into the lockdown period. The number of cases has been modest since then and varied between 0 and 34 over the Phase I and Phase II reopening.

The lockdown in Guangdong lasted 32 days, from January 23 to February 24, 2020. The provincial government issued an order on February 6, 2020, and encouraged workers in some industries to return to work after February 24. It is worth noting that the lockdown procedures in Guangdong were not as strict as the lockdown procedures implemented in the epicenter Wuhan. On February 24, 2020, Guangdong province entered its Phase I reopening, which lasted 76 days. During the Phase I reopening, people were allowed (and encouraged in certain industries) to return to work and visit outdoor public places. The Phase II reopening, or full reopening, officially started on May 9, 2020, when all businesses, including shopping malls, supermarkets, and restaurants, were allowed to open fully. The only exception was movie theaters, which remained closed till mid-July of 2020.

While Guangdong's COVID-19 case numbers are low, this does not imply that the pandemic had little or only a modest effect on the local economy. On the contrary, the measures implemented to alleviate the public health impact of the pandemic significantly affected the economy. As shown in our analysis in the main text, the pandemic inflicted sizeable damages to Guangdong's labor market, leading to a 72% increase in the number of unemployed and a 57% increase in unemployment benefit claims after the full reopening during May–September

2020 relative to those during the same period in 2019. As Guangdong's economy is among the most resilient among the economies of China's provinces, the aggregate labor market implications of the pandemic could be much more severe than those suggested by the national statistics.

2.2 Unemployment Measures

We leverage two features of the mobile phone data to understand the impact of the pandemic on Guangdong province's labor market outcomes. Specifically, individuals' commuting patterns observed over a long period of time help us monitor their employment status. We then use changes in these commuting patterns to construct unemployment measures. In addition, we take advantage of the designated unemployment benefits hotline – (12333) – and use the number of people who contacted the hotline (combined with the changes in commuting patterns) to construct measures of unemployment benefit claims.

Work commute The first feature of the mobile phone data that we leverage is the geocoded location information (in longitude and latitude) collected by mobile devices at 5-minute intervals when they are powered on.[6] We randomly select one million mobile users and use their location information at 5-minute intervals from January 2018 to September 2020 to construct their job and home locations. We define the work location as the location where a user spends at least 5 hours a day between 9 am and 6 pm for at least fifteen workdays in a given month. The home location is constructed similarly, except that we use the location where the user spends the most time between 10 pm and 7 am each month.[7] These geocoded locations trace out individuals' spatial trajectories over time and allow us

[6] Recent developments and the widespread diffusion of geospatial data acquisition technologies have enabled the creation of highly accurate spatial and temporal datasets. Passive collection of geolocation information—which underlies our data collection procedure—works on all traditional mobile networks (2G, 3G, or 4G). Researchers have used such mobile positioning data to study urban and transportation issues (González et al., 2008; Ahas et al., 2010), though few studies have exploited long panels of location data to examine labor market dynamics (Barwick et al., 2019).

[7] The location information from 7 am–9 am and 6 pm–10 pm is discarded because people are likely on the move during these time intervals.

to record the time of arrival and departure at job locations.

We provide two pieces of evidence that our assignment of home and work locations captures an intuitive spatial distribution of users in our sample. First, we use the coordinates of work and residential locations to compute the commuting distance for users with valid job location information. The distribution of the commuting distance decays exponentially (Figure A.3), consistent with evidence from other studies using both cell phone data and household surveys (Miyauchi et al., 2020; Rao, 2021). Additionally, the average commuting distance in our sample period is around 6.6 km, close to the average commuting distance of 8.7 km reported in the 2020 travel survey conducted by the Guangzhou Municipal Transportation Bureau (GMTB, 2020). Second, for the city of Guangzhou (the provincial capital), Appendix Figure A.4 plots the log difference between the number of users at 11 am and the number of users at 11 pm, averaged separately for weekdays and weekends in 2019. The figure includes all geographic locations recorded in the data. On both weekdays and weekends, the city center gains population, and the suburbs lose population during the daytime relative to their nighttime population. However, these differences are much more pronounced on weekdays than on weekends, especially in the busiest parts of the city center. The enlarged area in Appendix Figure A.4 illustrates this for a famous industrial park in Guangzhou. These spatial and temporal patterns of population density are remarkably consistent with the GMTB reports (GMTB, 2020).

We use the changes in the number of commuters before and after the lockdown and the changes relative to the number during the same period in 2019 as our measure of pandemic-induced unemployment. Changes in commuting patterns on a continuing basis can provide a valuable barometer of employment status, especially when participation in unemployment benefit programs is low, as is the case in China. To the extent that some of these changes reflect a post-lockdown shift toward more flexible work modes, such as work-from-home (WFH), they should be interpreted as an upper bound estimate on pandemic-induced unemployment. However, we provide multiple pieces of evidence below that our measure of unemployment based on commuting patterns over an extended period of time is unlikely to be driven by WFH. Panel A of Table A.2 presents descriptive statistics of key variables used in the commuting sample.

Calls to unemployment benefit agencies The second useful feature of our data is detailed records of calls to the designated government hotline (12333) for unemployment benefit agencies (with each call's time and duration). The hotline offers the public comprehensive one-stop service, provides eligibility information, helps with unemployment registration, and facilitates applications for unemployment benefits. Relative to filing online or visiting local social security bureaus, calling the designated hotline (12333) is the preferred choice for many due to its simplicity and provision of comprehensive help from customer services. Figure 1 shows the weekly Baidu index for the keywords "12333" and "unemployment insurance" in Guangdong province from 2019 to 2020.[8] The correlation of the Baidu index for the two keywords is 0.83 during the sample period. The comovement of these two indices offers additional support for our decision to use calls to the 12333 hotline as a proxy for the number of individuals claiming unemployment benefits.

The number of individuals making calls to 12333 provides an estimate of the level of unemployment benefit claims. During our sample period, 6,208,225 individuals contacted the unemployment benefit agencies via the designated hotline. However, despite the popularity of the hotline, lifetime unemployment benefits in China are capped at 24 months, thus limiting choices for people who have already exhausted their benefits. Therefore, instead of focusing on the level of unemployment calls, our analysis below exploits its changes over time. We show that changes in unemployment calls can provide useful information on unemployment benefit claim rates and short-run labor market dynamics that is otherwise unavailable through official statistics.

As people might reach out to the hotline multiple times to claim unemployment bene-

[8] The Baidu index, which is similar to Google Trends, is a keyword-analysis tool launched by Baidu, the largest search-engine company in China. It reflects the search frequency of certain keywords on the Baidu website.

fits, we treat multiple calls from the same user as one claim incident and therefore use the number of individuals calling the unemployment hotline, instead of the number of calls to 12333, to construct our unemployment benefit claim measure. In addition, calls that failed to go through to the receiving party and calls shorter than 30 seconds are excluded from the analysis. For brevity, the terms "number of individuals calling the unemployment hotline" and "number of unemployment calls" are used interchangeably throughout the analysis. Appendix Figure A.5 plots the number of individuals calling the unemployment hotline across cities in 2019. The correlation between city-level unemployment calls and the official unemployment rate released by the NBS, which is available only annually at the city level, is reasonably high at 0.7 for 2019.

Our analysis based on unemployment calls counts only the first time when a user reaches out to the unemployment benefit hotline. We aggregate the duration of all subsequent calls in calculating "call duration to the hotline". Our main analysis excludes users under age 18, as they are unlikely to be working due to the Law on Protection of Minors. The results excluding users under age 25 (to eliminate those still in school) are almost identical. Panel B of Table A.2 presents descriptive statistics of key variables used in the unemployment-call sample.

Some of our analyses examine migrants and nonmigrants separately. It is important to note that migrants who had been working in Guangdong but did not have Guangdong *Hukou* became eligible for unemployment benefits from 2014.[9] The new regulation was designed to attract migrants and help improve labor relations. One data limitation is that we do not observe whether an individual has nonlocal *Hukou* status—the official definition of migrants. Instead, we define migrants as individuals who registered their phone numbers outside Guangdong province. This is an imperfect measure of migrant status, as workers from outside Guangdong who bought and registered their mobile phones in Guangdong are treated as nonmigrants in our analysis. Consequently, the actual unemployment gap between

[9]See the announcement by Guangdong's Human Resources and Social Security Department: https://www.gdhrss.gov.cn/sy/20140801/10101.html.

migrants and residents might be even larger than our estimates.

3 Empirical Framework

Our analysis employs the DID approach by comparing labor market outcomes in 2020 before and after the event date (when Guangdong implemented the lockdown) with those before and after the same (lunar) calendar dates in 2019. As Guangdong's lockdown occurred two days before the 2020 Chinese New Year, we use the lunar calendar instead of the standard almanac calendar to define the event date. Specifically, the event date is January 23, 2020, for the year 2020 and February 3, 2019, for 2019, two days before Chinese New Year in the lunar calendar. We use observations from the year 2020 as the treatment group and those from the year 2019 as the control group. In other words, our analyses compare changes in labor market outcomes before and after the event date in 2020 with changes in labor market measures before and after the exact event date in 2019. We divide the interval from 60 days before the lockdown to 252 days after the lockdown into four periods: before lockdown (60 days), during the lockdown (32 days), Phase I reopening (76 days), and Phase II (full) reopening (144 days).

To control for potential differences in time-varying unobservables, we include a rich set of fixed effects such as day-of-week, event-day, holiday, and treatment group fixed effects. The identification assumption is that conditional on inclusion of these controls, there would have been no systemic differences in time-varying unobservables between the two groups in the absence of the pandemic. Results from event studies that are discussed below support this assumption of common trends between the two groups prior to the event date. We use the following DID framework and ten-day intervals to trace out the dynamic impact of the pandemic over time:

$$y_{cit} = \sum_{q=-5}^{24} \beta_q \cdot d_i \cdot \mathbb{1}\Big(t \in [q * 10 + 1, (q+1) * 10]\Big) + \alpha_c + \gamma_i + \eta_t + \xi_{it} + \varepsilon_{cit}, \qquad (1)$$

where c denotes a neighborhood (area covered by the nearest cell tower), i denotes the treatment group (year 2020) and control group (year 2019), and t denotes the event day (t=0 stands for January 23 in 2020 and February 3 in 2019). The event window is sixty days before the lockdown and 252 days after the lockdown.

y_{cit} is the outcome variable, such as the log number of noncommuters. We report results based on log(outcome+1) to avoid taking the logarithm over zero. However, results based on the inverse hyperbolic sine function (which is very similar to the log function and can handle zero values) are very similar. β_q are event-study coefficients capturing differences between the treatment and control groups. Variable d_i is a dummy equal to one for the treatment group, and $\mathbb{1}(.)$ is an indicator variable for each 10-day interval of the sample. We include neighborhood fixed effects γ_c, group fixed effects γ_i, and 312 event-day fixed effects η_t. We also include day-of-week fixed effects and holiday fixed effects ξ_{it} that vary by group and time (e.g., the International Labor Day holiday falls on different lunar calendar days in 2019 and 2020). Standard errors are clustered at the event-day level.

Since the key regressors are dummy variables, $\hat{\beta}$ is not a consistent estimator of the percentage change in unemployment, with a larger bias when $\hat{\beta}$ is further from zero. While we report $\hat{\beta}$ in all figures and tables, we calculate the percentage changes using $100 * \Big(\exp[\hat{\beta} - \widehat{var(\beta)}/2] - 1\Big)$, a consistent estimator of the percentage impact, throughout the paper. The second component in the bracket reduces the finite-sample bias.[10]

To further explore the heterogeneity across cities and the importance of industrial composition and trade exposure, we employ the following specification:

$$\begin{aligned} y_{cit} \quad = \quad & d_i \cdot \mathbb{1}\Big(t \in [0, 252]\Big) \cdot \boldsymbol{Z}'\tau + d_i \cdot \boldsymbol{Z}'\mu + \mathbb{1}\Big(t \in [0, 252]\Big) \cdot \boldsymbol{Z}'\rho \\ & + \beta \cdot d_i \cdot \mathbb{1}\Big(t \in [0, 252]\Big) + \alpha_c + \gamma_i + \eta_t + \xi_{it} + \varepsilon_{cit}, \end{aligned} \tag{2}$$

where $\boldsymbol{Z}$ is a vector of city attributes in 2019 and η, μ, and ρ are the corresponding coef-

[10]This adjustment method was proposed by Kennedy (1981).

ficients. For example, $\boldsymbol{Z}$ could be a city's labor share in each of the 13 major industries, dummies for the 21 cities, or a city's export-over-GDP ratio. In addition to the interaction between the pandemic treatment and city attributes, we control for all lower-level interactions in the regression. The variables d_i and $\mathbb{1}(.)$ and the fixed effects α_c,γ_i, η_t, and ξ_{it} are the same as in Equation (1). The key coefficient is τ, which measures the heterogeneous impact by city characteristics $\boldsymbol{Z}$ based on their values in 2019. Unlike in Equation (1), where we estimate the pandemic's impact for each ten-day interval, here, we estimate the average effect over all periods and focus on heterogeneity across industries and cities.

4 Event Studies

As discussed in Section 2.2, we use the number of individuals who used to commute to work regularly but stopped commuting altogether in a given period as a measure of unemployment. To mitigate potential measurement errors (see a detailed discussion of this below in Section 4.4), we refine our analysis by limiting our sample to individuals who stop commuting altogether *and* who do not use any email/virtual meeting apps when their commuting patterns change.

Our second measure is related to unemployment benefits claims, where we use the number of people who contacted the designated unemployment benefits hotline 12333 as a proxy for people claiming unemployment benefits. We also refine our analysis by limiting the sample to individuals who both contacted the hotline and stopped commuting altogether, though the results based on this sample are nearly identical.

This section first presents event-study figures for both measures to illustrate their time series patterns, especially the pronounced increases after the onset of COVID-19. Then, we devote an entire subsection to addressing threats to our empirical analysis, especially the issue of flexible work arrangements (such as WFH) that also affect commuting patterns and the concern that the 12333 hotline also provides support for accessing other services, such as social security or assistance with labor disputes.

4.1 Noncommuters

We exploit the variation in commuting patterns based on the location tracking data for one million randomly selected users. We treat an individual as commuting to work for a given time window (e.g., two weeks) if she visits her work location at least once during that time window. To accommodate the possibility of (partial) WFH modes during and after the lockdown, we construct three commuter measures using different time windows: a week, two weeks, and a month. For example, under the definition that uses a month as the relevant time window, an individual is classified as a commuter for a given month if she visits a work location at least once in that month. Noncommuters during a certain time window are individuals who used to commute to work but no longer commute during that period.

Figure 2 shows the event study where we use noncommuters as our measure of unemployment, with the observations for the same period in 2019 as the control group. Panel (a) depicts the event-study coefficients that measure changes in the number of noncommuters. Panel (b) repeats the analysis but limits the sample to noncommuters who do not use any of the email/virtual meeting apps available to mobile phone users during our sample period (see the discussion in Section 4.4 for more details). This helps exclude people who work from home. For 2020, the event date (or day zero) is the lockdown, January 23, 2020, two days before Chinese New Year. Correspondingly, the event date in 2019 is February 3, 2019, also two days before the 2019 Chinese New Year. The Phase I reopening started on February 24, 2020, 32 days after the lockdown, when people were allowed to return to work and visit outdoor public places. The Phase II reopening, or full reopening, started on May 9, 2020, 108 days after the lockdown, when shopping malls, supermarkets, and restaurants were allowed to fully reopen.

There are three salient patterns from both panels. First, for the prelockdown period, there is virtually no difference in the number of noncommuters between 2019 and 2020, lending support to the assumption of parallel trends between the treatment and control groups, the key identification assumption of our analysis. Second, during the lockdown in 2020, the number of noncommuters increased severalfold relative to that in 2019. The increase reflects not only changes in unemployment but, more importantly, temporary leaves with/without pay and work from home due to the strict nature of the lockdown. Third, as the economy opened up, the increase in non-commuters gradually came down to about 70% by the end of Phase I reopening and remained stable till the end of September, four months after Phase II (or full) reopening. The increase in panel (b) is slightly smaller but the pattern stays the same. While these changes are unprecedented, it is milder relative to that observed in the U.S. during the same period: the unemployment rate in the U.S. increased from about 3.6% in the second half of 2019 to 13% and 8.8% in the second and third quarter of 2020, respectively, according to the U.S. Bureau of Labor Statistics.

4.2 Calls to unemployment benefit hotline

We then examine the second outcome: the pandemic's impact on unemployment benefit claims based on calls to 12333. Figure 3 depicts the differences in the log daily number of individuals calling the unemployment hotline between 2019 and 2020. Similar to Figure 2, there are no differential trends in the calls to 12333 between the two years before the lockdown period, leading credence to the parallel trend assumption.

In contrast to the sharp increase in non-commuters during the lockdown as shown in Figure 2, the number of people contacting the unemployment benefits hotline dropped significantly during this period. This is likely due to the uncertainty about the severity and duration of the pandemic during the initial stage. In addition, the increase in the number of non-commuters during the lockdown was likely driven by changes in work arrangement instead of unemployment.

As the severity of the pandemic unfolded in China, the number of individuals calling the unemployment benefit hotline increased sharply during Phase I. The jump in the number of people calling 12333 began in April 2020, when the mobility restrictions were beginning to ease in the Hubei province (the epic center of the pandemic) as well as other provinces.[11] The increase stabilized at about 50% by the end of the Phase I reopening and remained so till the end of our data period. Both the pattern and magnitude are consistent with those in Figure 2.[12]

We have repeated the event study limiting to individuals who both called the unemployment benefits hotline and stopped commuting altogether. The patterns are nearly identical to Figure 3.

4.3 Migration post unemployment

While migration is not the focus of our empirical analysis, examining whether people migrate after experiencing negative labor market shocks could provide insights on our understanding of how people adjust and adapt to changes in employment status. Since it is straight forward to identify the date of contacting unemployment benefits agencies, we conduct an event study on whether individuals migrate to other cities after they call 12333.

To do so, we keep the IDs for all individuals who reached out to the unemployment benefits hotline and track their residential cities during the period between 150 days before the call and 360 days after. The period [-150, -121] before the call serves as the reference group. We drop observations during the lockdown period from January 23rd to February 24th, 2020. The dependent variable is a cumulative measure of whether individuals have migrated to a different city by period t. To increase the precision of the residential city measure, all valid observations must have resided in a given city for at least two months. Migrated individuals are those who lived in one city for at least two months and moved and

[11] This is consistent with the evidence that migrants were allowed to return to their place of work and some filed unemployment benefit claims upon job losses. See https://baijiahao.baidu.com/s?id=1663364477907792823&wfr=spider&for=pc and https://baijiahao.baidu.com/s?id=1663996321334096972&wfr=spider&for=pc.

[12] As a comparison, the initial claims of unemployment benefits in the U.S. skyrocketed from 0.2 million in February 2020 to 6.1 million in April 2020, and gradually decreased to 0.8 million in the end of September according to the U.S. Bureau of Labor Statistics.

lived in other cities for at least two months.

Appendix Figure A.6 presents the event study. About 10% of workers migrated to other cities two months after contacting the unemployment benefit agencies. The fraction increases modestly over time to 13% one year after the call. Hence, the majority of individuals who experienced negative employment shocks appeared to have stayed in the city where they lived, with a small fraction migrating to other cities for employment opportunities.

4.4 Threats to the empirical strategy

There are two major threats to our empirical strategy. First, flexible work arrangements, such as work-from-home (WFH), might pose a challenge to our analysis using commuting patterns. Second, the 12333 hotline provides other services such as social security benefits. This could inflate our measure of the number of people claiming unemployment benefits. We next examine each threat in turn.

Work from home One could argue that the increase in non-commuters post the lockdown may be partially driven by the increase in WFH. We present several pieces of evidence that our non-commuter definition – not visiting the workplace at all during an extended period of time (such as two weeks or a month) – accurately reflects individuals' unemployment status and that changes in non-commuters are unlikely to be primarily driven by WFH.

First, we examine each individual's usage of all virtual meeting apps and email apps that are available in the Apple store and Android Google Play store.[13] Using the two-week window to define commuters, the share of commuters using these apps at least once during a two-week window was 30.3%, 26.1%, 24.7% during the lockdown, Phase I, and Phase II reopening

[13]To gauge the prevalence of WFH, we obtain an exhaustive list of all 21 virtual meeting apps and 29 email apps in the Apple store and Android Google Play store. The virtual meeting apps include: Ailiao, Alitong online, Chubao, DingDing, Feige, Feiyin, Laidian, Shangqitong, Shuoba, SKYPE, Tencent meeting, Tongtong, uu online, Weihui, Weiwei, Yiliao, Youhuatong, Youliao, Youxin, Zhangshangbao, Zoom. The email apps include: 139 light mail, 139 mail, 189 mail, 21CN light mail, 21CN mail, 263 mail, Ali mail, Baidu mail, China Mobile mail, Coremail, Foxmail, Gmail, Hotmail, Ke space, Live mail, Mail master, Mi mail, Microsoft Outlook, Qixinbao, QQ mail, Sina mail, Sohu mail, Tencent mail, TOM, Wangyi mail, Woo mail, Yahoo mail, Youqia, Yun home.

periods, respectively. In contrast, the share of non-commuters using any virtual meetings and email apps at least once during a two-week window was 5.2%, 0.6%, and 0.06% during the lockdown, Phase I, and Phase II reopening periods, respectively. The patterns are very similar when we limit to virtual meeting apps or use one week or one month as the relevant time window to define commuters and non-commuters. Furthermore, the sharp contrast in usage patterns between commuters and non-commuters is very similar in 2019: the share of commuters using these apps at least once during a two-week window was 21.2%, relative to 1.0% among non-commuters. If our measure of non-commuters in 2020 is primarily driven by a significant increase in the fraction of workers who telecommute in 2020 relative to 2019, we would expect the virtual meeting app usage patterns to be very different over these two years. We would also anticipate a much higher usage of virtual apps among non-commuters in 2020. Neither prediction is supported by data.

Second, our three commuter (non-commuter) measures are highly correlated (their correlations exceed 0.92). More than 94% of individuals who are non-commuters over a two-week window remain non-commuters over the entire month. These patterns hold in both 2019 and 2020. If non-commuters in 2020 mostly consisted of people who work from home and visit offices once in a while, we should anticipate the persistence in non-commuting (or telecommuting) patterns to be significantly lower in 2020 than that in 2019 when the lockdown restrictions were gradually lifted. This is not what we observe. These patterns provide evidence that our commuter measures accommodate flexible work modes (e.g., WFH from time to time). When individuals stop visiting their workplace altogether over an extended period, as defined in our analysis, they are most likely not working (unemployed) rather than WFH.

Third, an important indicator of whether the economic activities have resumed normal is time spent away from home (including both outdoor and indoor activities). We compare the time spent on non-work activities away from home in 2020 to that in 2019. We define non-work activities as those happened at places other than home and the workplace and lasted for at least half an hour. Appendix Figure A.7 presents the event study plot. The pattern is consistent with that in Figure 2. Before the lockdown period, there was no difference in the time spent on non-work activities between 2019 and 2020, but there was a sharp drop during the lockdown period in 2020. Time spent on non-work activities gradually recovered during the Phase I reopening. It fully recovered and even showed a slight increase during the Phase II reopening. This presents strong evidence that economic activities have returned to the pre-pandemic level and people are free to spend time away from home, which further lends support to the limited role of WFH by the end of Phase II reopening.

Purpose of calling 12333 While we do not directly observe individuals' purpose of contacting the unemployment benefits agencies, it helps to examine their commuting patterns before and after individuals reach out to the hotline. If their phone calls are motivated by fear of unemployment that has not actually realized, or if people are calling for information on social services other than unemployment benefits, then we should not observe any changes in their commuting patterns.

To do so, we extract a sample that includes all individuals who have made calls to the unemployment hotline. Then we examine their commuting patterns, especially any changes in commuting patterns (whether they ever stop commuting altogether), from four months before the call to one year after the call.[14] Panel (a) of Figure A.8 depicts the cumulative probability of ever stopping-commuting for at least two weeks among these callers with respect to the event of reaching out to the unemployment benefit agencies. The probability of stop commuting is practically zero three months before the call, increases to about 20% one month before the call, and quickly jumps to 80% two to three months after the call. In other words, the majority of individuals experience changes in commuting patterns (which we interpret as changes in the employment status) surrounding the time when they contact the hotline 12333. This provides evidence that people reach out to the government unemployment benefits agencies primarily because they either have already lost or are about to lose

[14] Extending the event window to six months before the call or earlier makes a difference, as the probability of non-commuting before the call is low.

their jobs, rather than collecting information on other social services. We have repeated this analysis with a more stringent definition, where stopping commuting altogether is defined using the one-month window (not visiting one's workplace for an entire month). Results are nearly identical (see panel (b) of Figure A.8).

One might be concerned that the pandemic has greatly enhanced people's awareness of government-provided services, including unemployment benefits and the designated hotline 12333. If this is the case, then our measure of changes in unemployment benefits claims in 2020 relative to 2019 could be inflated, because people are more likely to call the hotline in 2020 to inquire about potential benefits. To examine whether this is the case, we repeat the analysis in panel (a) of Figure A.8 separately for 2019 and 2020 and plot the patterns in Figure A.9, where the red line with diamonds represents 2019 and the blue line represents 2020. If people become more aware of the existence of the hotline in 2020 and are simply reaching out for benefit-related information, then we should expect a much lower probability of non-commuting in 2020 relative to that in 2019. This is not what we see. Rather, the pattern for 2020 is remarkably similar to that in 2019. In fact, the two lines are hardly distinguishable.

The fact that the calling-and-stopping-commuting pattern appears rather indistinguishable between 2019 and 2020 provides strong evidence that the nature of reaching out to government agencies upon job losses has been stable over the years and not affected by the pandemic. It lends credence to our DID empirical strategy of comparing changes in 2020 pre- and post the lockdown to changes in 2019 during the same period. Finally, it also helps address the concern of WFH. If the majority of non-commuters in 2020 are people working from home, rather than unemployed, we should expect rather different patterns between the two years. The fact that there were no changes in the relationship between calling unemployment benefit hotline and commuting patterns from 2019 to 2020 suggests that the increase in unemployment in 2020 relative to 2019 were unlikely to be driven by work from home, a phenomenon that is much less common in 2019.

5 Regression Results

The specification for the regression analyses is analogous to that for the event study, expect that we report coefficients for each of the four periods instead of coefficients at ten-day intervals: 1-30 days before the lockdown, during the lockdown, Phase I, and Phase II. The reference group is 31-60 days before the lockdown. Regressions for unemployment calls is at the neighborhood and day level, with a total of 489,514 observations. Regressions for commuting patterns aggregate to the neighborhood-week, neighborhood-fortnight, and neighborhood-month level when appropriate.[15]

5.1 Effect on unemployment

Table 1 reports parameter estimates for the percentage increase in non-commuters measured by a two-week window, grouping the ten-day intervals into four periods.[16] During the lockdown period, the number of non-commuters increased by nearly 43 folds, reflecting the draconian nature of the lockdown. The number of non-commuters increased by 163% during the Phase I reopening and by 72% during the Phase II full reopening.[17] The effect size is robust to alternative window lengths of one-week or one-month in defining non-commuters (See Section 5.4). Since economic activities have largely returned to the pre-pandemic level by Phase II and WFH is unlikely to play a major role during Phase II as shown in Section 4.4, we therefore interpret the 72% increase in non-commuters as the impact of the pandemic on unemployment. In 2020, the average number of non-commuters was 38,729, or 7.4 percent of all workers (commuters plus non-commuters). The 72% increase during the Phase II reopening relative to the baseline corresponds to a 5.3 percentage point increase in non-commuters over the same period in 2019.

[15] We also adjust the fixed effects accordingly. For example, We drop the day-of-week fixed effects and replace event-day fixed effects with event-fortnight fixed effects and cluster the standard errors at the fortnight level when we measure non-commuters using a two-week window.

[16] The omitted group is 31-60 days before lockdown.

[17] The coefficient of Phase I and Phase II in Table 1 is 1.03 and 0.59, respectively. The effect sizes reported throughout the main text are calculated using $\left(\exp[\hat{\beta} - \widehat{var(\beta)}/2] - 1\right)$, as discussed in Section 3.

Using commuting patterns to measure unemployment has a significant advantage over measures derived from applications for unemployment benefits: it is not subject to the participation bias (eligible people do not participate). Due to the inertia, lack of information, stigma, time and "hassle" costs associated with applications, it is estimated that 66% of eligible households do not participate in major social programs in the U.S. (Ribar, 2020). Non-participation is much more severe in developing countries due to limited program benefits. In comparison, commuting patterns, observed over an extended period of time, provide a real-time and likely more accurate indicator of the underlying labor market dynamics.

Column (2) of Table 1 examines changes in work duration among individuals working on-site. Hours on-site dropped by 19% and 8% during the lockdown and Phase I, respectively, but returned to the pre-pandemic level in Phase II. The pandemic does not seem to have brought about dramatic changes in the nature of working on-site during the Phase II reopening, lending further support to our strategy of measuring unemployment status based on changes in commuting patterns.

5.2 Effect on unemployment benefit claims

Table 2 examines the pandemic's impact on the log number of individuals reaching out to the unemployment benefit hotline (column (1)) and call duration (column (2)). The table presents the coefficient estimates of β_q in Equation (1), except that the ten-day intervals are grouped into four periods: 1-30 days before lockdown, during the lockdown, Phase I reopening, and Phase II full reopening. Echoing results in Figure 3, the number of individuals calling the unemployment benefit hotline decreased by 31% during the lockdown and increased by 28% during the Phase I reopening and nearly 57% during the Phase II full reopening. As discussed above, the level of unemployment calls is a lower bound estimate of the number of unemployed, as not all individuals who have lost jobs file for unemployment benefits. However, in light of the remarkable similarity in unemployment calls between 2020 and 2019 prior to the pandemic (Figure 3), the percentage change in unemployment calls in the Phase II full reopening period estimated in Table 2 (57%) is likely a reliable measure of the percentage change in unemployment benefit claims as a result of the pandemic.

The call duration displayed a similar pattern: the average call time dropped during the lockdown but increased after the reopening. Call duration increases partly because more migrants applied for unemployment benefits after the reopening (as we show below in the heterogeneity analysis) and that they generally need to provide more information than local residents.

5.3 Heterogeneity analysis

Unemployment To examine the pandemic's differential impacts across demographic groups, we repeat the baseline event-study analysis shown in panel (a) of Figure 2 by gender, age, and migrant status and plot coefficient estimates in panels (a)-(c) of Figure 4. Specifically, the dependent variable is the difference in the logarithm number of non-commuters between females and males in panel (a), between individuals 40 years old and above and those under 40 in panel (b), and between migrants and non-migrants in panel (c). Females are more affected by the pandemic. The percentage increase in female non-commuters is 10-20 percentage points higher than that in male non-commuters during lockdown and Phase I reopening. The gap became smaller and statistically insignificant during the Phase II reopening.

Older workers fared worse than younger cohorts: the percentage increase in the number of non-commuters among workers 40 and above was about 20-60 percentage points higher than that for workers under 40 during the lockdown and Phase I reopening periods. Even during the Phase II reopening, the gap between the two age groups still remained at about 20 percentage points. Lastly, migrants are most severely affected by the pandemic. The number of non-commuters among migrants increased more during lockdown but especially the Phase I reopening relative to that among non-migrants. By the end of the Phase II reopening, the percentage increase in the number of migrant non-commuters was still about 40 percentage points higher than that among non-migrants, highlighting the disproportionate large and lingering burden on migrants. According to the NBS, the number of migrant workers in 2019 reduced by 3.8 million nationwide by September 2020. The escalating number of migrants who stop commuting to work in our sample is consistent with the massive reduction in migrant workers reported by the NBS and suggested a large-scale layoff among migrant workers.

To further examine the heterogeneous impact across demographic groups, we conduct an analysis at the neighborhood level in Guangzhou where we have data on micro-level social economic variables. Specifically, we regress the percentage change in the number of non-commuters between 2019 and 2020 on the quadratic forms of the 2018 average housing price and migrant share in each neighborhood (i.e., cell-tower-area). Figure 5 plots the predicted percentage changes against the average housing price and migrant share. Neighborhoods with a lower housing price and a higher share of migrants experienced a higher percentage increase in non-commuters. Specifically, a one-standard-deviation increase in migrant share and housing price would induce 22.1 percentage points increase and 2.7 percentage points decrease in non-commuters, respectively. These results corroborate the existing literature documenting that the least advantaged social groups, including migrants, are most vulnerable to adverse shocks and risks (Banerjee and Duflo, 2012). In addition, we test whether the number of migrants can explain the underestimation of the official unemployment rate. We compute the differences between the official unemployment rate in 2020 and those we derived based on the commuting sample at the city level. The correlation coefficient between the differences and the ratio of migrants across cities is 0.63, suggesting that migrants are one major contributor to the underestimation of the official unemployment rate.

In addition, the disproportionately harsher impacts on females, older workers, and migrants likely reflect the pandemic's heterogeneous shocks across industries, as we show below in Figure 6. These groups are more likely to work in hospitality industries, including restaurants and hotels, which have been hard hit by the pandemic, and less likely to work in the less affected education and high-tech industries.

There is considerable variation across cities in Guangdong in terms of population and GDP (Appendix Table A.1). Panel (a) of Figure 6 examines the impact of heterogeneity across cities. The figure reports coefficients on the interactions of the pandemic treatment variable (which is one from January 23, 2020, to September 30, 2020) and city dummies, following Equation (2). Heterogeneity across cities is sizeable: the number of non-commuters increased for all but one city and the range varied from 10% to as high as 150%.[18] The economically more developed cities such as Guangzhou and Zhuhai experienced the largest increase while the less developed cities such as Yangjiang seemed to be unscathed. At least two factors drive the differential effects across cities. First, cities have different industry compositions. Among the seven cities that experienced the most significant increase in unemployment calls, the average share of the workforce in hotel and catering, real estate, and transportation was 13.9% in 2019, while the average share was less than 3% among the seven least affected cities. To illustrate the heterogeneous impact across industries directly, we run a separate regression following Equation (2), where we interact the pandemic treatment variable with city-level labor shares by industry for all thirteen major industries.

Panel (b) of Figure 6 reports coefficients for all industries. The hotel and catering, real estate, and leasing and business experienced the largest increase in non-commuters, ceteris paribus. In comparison, the finance, health care, and education sectors witnessed reductions in non-commuters, consistent with findings using data from other countries (Adams-Prassl et al., 2020; Alon et al., 2020). To evaluate the importance of industry compositions, we predict the number of non-commuters in logs (the dependent variable) using coefficient estimates and each city's observed labor share across industries and compare the range of predicted values with the observed range of the dependent variable. Variation in industry composition across cities contributes to 38.7% of changes in non-commuters.

In addition to differences in industrial composition, cities also have differential trade exposure measured by the total export relative to local GDP in 2019. For the 21 cities

[18] The effect size for Yangjiang is -5%, but statistically insignificant at the 10% level.

in Guangdong, the median export-to-GDP share in 2019 is 14.7%. Shantou city has the least exposure to international trade, whose export-to-GDP ratio is only 2.5%. At the other extreme is Dongguan, whose export-to-GDP ratio is 91.0%. As shown in panel (a) of Figure 6, the pandemic's impact on Shantou's unemployment is much milder relative to that on Dongguan. In Table 3, we interact the pandemic treatment variable with a city's export-to-GDP share. As expected, the interaction coefficient is statistically significant and positive. A ten percentage-point increase in 2019's export-to-GDP ratio is associated with a 4.4% increase in the number of non-commuters for a given city. Similar to industry compositions, variation in the export-to-GDP ratio is also critical and explains 28.5% of the heterogeneity in the pandemic's unemployment impact across cities.

The sizeable estimates in Table 1 (a 72% increase in the unemployment rate) and the significant heterogeneity across cities and industries as highlighted in Figure 6 speak to the severity of the pandemic's labor market implications, the uneven impact of the pandemic, and the importance of conducting analysis at granular levels. In addition, these results illustrate the rippling effect of the pandemic across cities within a country and across countries around the globe through the supply chain and trade channels, where the industry composition of a city (or a country), the nature of the supply chain, and exposure to trade are crucial factors in determining the effect that the pandemic has on its economy (Forsythe, 2020; Goldberg, 2020; von Gaudecker et al., 2020; World Trade Organization, 2020a,b).

Unemployment benefit claims We repeat the heterogeneity analysis on unemployment benefit claims based on call data to the unemployment benefit hotline. Figures 7, 8 and A.10 present heterogeneous impacts across demographic groups, cities and industries, as well as household income and migrant shares. The results are qualitatively similar to those that are based on non-commuters: females, workers over 40, and migrant workers experienced a large increase in unemployment benefit claims since the pandemic. The same is true for areas with a low income and a high migrant share as shown in Figure A.10. There is also a significant amount of heterogeneity across cities and industries, closely mirroring patterns reported in Figure 6. Finally, as shown in Table 4, a ten percentage-point increase in the export-to-GDP ratio is associated with a 2.7% increase in the number of unemployment benefit claims for a given city, consistent with the result from Table 3 based on commuting data. The magnitude at 2.7% is slightly lower the increase in unemployment, which is expected due to the limited participation in unemployment benefit programs.

One might be concerned that increases in the number of individuals calling the unemployment benefit hotline are merely driven by a higher awareness of unemployment benefits post the pandemic. However, as shown above, females, workers above 40, and especially migrants have reported a significantly higher increase in hotline callings than other groups of workers. They are precisely the sub-population that experienced the worst hit in employment during the pandemic. In addition, changes in the number of hotline calls are highly uneven across industries. There is also a great deal of heterogeneity across cities. The number varies from -8% in Yangjiang to 99% in Guangzhou, which closely mirrors the industry and worker composition across cities. These patterns are unlikely to be purely driven by a significant increase in the awareness of the unemployment hotline post the pandemic, as information on these government services is primarily disseminated at the national and provincial level, instead of at the sub-population, industry, or city level.

5.4 Robustness Checks

The analysis of commuting patterns discussed above uses the two-week window to define a commuter. We have constructed two alternative measures of commuters using the one-week and one-month window. Reassuringly, these three variables are highly correlated: the correlation is 0.95 between the one-week and two-week measure, 0.92 between the one-week and one-month measure, and 0.97 between the two-week and one-month measure, respectively. In addition, more than 94% of individuals who are non-commuters over a two-week window remain non-commuters over the entire month. These patterns suggest that our commuter measures accommodate flexible work modes (e.g., working at home from time to time). When individuals stop visiting their workplace altogether over an extended period as defined in our analysis, they are essentially not working rather than working at home as we discussed in detail in Section 4.4.

In panels A and B of Table A.3, we repeat the analysis using non-commuters defined over the one-week and one-month window. The estimated effect size is 75% and 71% using these alternative measures, similar to the baseline estimate of 72% in Table 1. Table A.4 replicates Table 1 but limits to non-commuters who do not use any email or virtual meeting Apps when they stop commuting. This is a demanding set of criteria. The coefficients and the implied effect sizes remain robust at about 70%. These patterns corroborate the evidence above that WFH is unlikely to be driving our results and that our measures of commuting accommodate hybrid work modes.

Our main analysis includes users who are 18 years or older. As some users between the age of 18 and 25 might still be in schools, we exclude users under the age of 25 as a robustness analysis. Results on non-commuters (Table A.5) and calls to the unemployment benefit hotline (Table A.6) barely change when we limit to users aged 25 and above.

Table 5 replicates the analysis in Table 2 but is limited to individuals who both called the unemployment benefit hotline and stopped commuting. These individuals are less likely to be mis-classified as being unemployed. The results are similar to those reported in Table 2. For example, the effect size on unemployment claims during the Phase II reopening is 0.57 in Table 2 compared to 0.49 in Table 5. The effect on call duration is 0.75 for the full sample (all individuals reaching out to the hotline) and 0.78 for the restricted sample (individuals who both reach out to the hotline and stop commuting altogether), respectively.[19]

The last two robustness checks replicate the baseline analysis but weigh each observation with the average number of non-commuters per day in each neighborhood in 2018. The

[19] In Table 2, the coefficient for Phase II is 0.45 for calls to unemployment benefit hotline and 0.56 for call duration. In Table 5, the coefficient is 0.4 and 0.58, respectively. The effect sizes reported in the main text are calculated using $\left(\exp\left[\hat{\beta} - \widehat{var(\beta)}/2\right] - 1\right)$ as discussed in Section 3.

results are reported in Tables A.7 and A.8 for non-commuters and unemployment calls, respectively. The results are slightly larger in the weighted regressions but qualitatively the same.

6 Conclusion

Based on granular and high-frequency mobile phone data in China's most populous province, our analysis found that the pandemic led to a 72% increase in unemployment and a 57% increase in unemployment benefit claims even after full reopening, relative to those during the same period from May to September in 2019.

While dramatic, these effects are smaller than those in the United States. This is partly due to the differences in the composition of the economy between these two countries: the service sector, which has been hard hit by the pandemic, employs 79% of the workforce and produces 68% of the GDP in the United States, compared to 47% and 50% in China in 2018. In addition, the draconian measures adopted in China to control the pandemic have reduced the spread of the virus more effectively (Hsiang et al., 2020; Kraemer et al., 2020; Zhang et al., 2020) and likely mitigated the impact on the economy during our data period.

Our analysis shows uneven labor market impacts across demographic groups and industries. The heterogeneous impact is consistent with recent studies in other countries (Adams-Prassl et al., 2020; Alon et al., 2020). Our research adds to the literature by showing that the pandemic's adverse impact on the labor market is more severe in areas that rely more heavily on export and hence more exposed to external shocks through global trade channels. Future research can use our approach to study the longer-term impacts as well as the impacts of the most recent lockdowns across many cities in China.

References

Adams-Prassl, Abigail, Teodora Boneva, Marta Golin, and Christopher Rauh, "Inequality in the Impact of the Coronavirus Shock: Evidence from Real Time Surveys," 2020. Working Paper.

Ahas, Rein, Anto Aasa, Siiri Silm, and Margus Tiru, "Daily rhythms of suburban commuters' movements in the Tallinn metropolitan area: Case study with mobile positioning data," *Transportation Research Part C: Emerging Technologies*, 02 2010, *18*, 45–54.

Ahn, Hie Joo and James D Hamilton, "Measuring Labor-Force Participation and the Incidence and Duration of Unemployment," Working Paper 27394, National Bureau of Economic Research June 2020.

Alon, Titan M, Matthias Doepke, Jane Olmstead-Rumsey, and Michèle Tertilt, "The Impact of COVID-19 on Gender Equality," Working Paper 26947, National Bureau of Economic Research April 2020.

Atkin, David, Keith Chen, and Anton Popov, "The Returns to Serendipity: Knowledge Spillovers in Silicon Valley," 2020. Working Paper.

Banerjee, Abhijit and Esther Duflo, *Poor Economics*, PublicAffairs, 2012.

Barwick, Panle Jia, Yanyan Liu, Eleonora Patacchini, and Qi Wu, "Information, Mobile Communication, and Referral Effects," Working Paper 25873, National Bureau of Economic Research May 2019.

Bick, Alexander and Adam Blandin, "Real Time Labor Market Estimates During the 2020 Coronavirus Outbreak," 2020. Working Paper.

Cai, Fang, Yang Du, and Meiyan Wang, "Demystify the labor statistics in China," *China Economic Journal*, 2013, *6* (2-3), 123–133.

Cajner, Tomaz, Leland D Crane, Ryan A Decker, John Grigsby, Adrian Hamins-Puertolas, Erik Hurst, Christopher Kurz, and Ahu Yildirmaz, "The U.S. Labor Market during the Beginning of the Pandemic Recession," Working Paper 27159, National Bureau of Economic Research May 2020.

Card, David, "Origins of the Unemployment Rate: The Lasting Legacy of Measurement without Theory," *American Economic Review*, May 2011, *101* (3), 552–57.

Chen, Bi Yu, Yafei Wang, Donggen Wang, Qingquan Li, William H. K. Lam, and Shih-Lung Shaw, "Understanding the Impacts of Human Mobility on Accessibility Using Massive Mobile Phone Tracking Data," *Annals of the American Association of Geographers*, 2018, *108* (4), 1115–1133.

Chen, M. Keith and Devin G Pope, "Geographic Mobility in America: Evidence from Cell Phone Data," Working Paper 27072, National Bureau of Economic Research May 2020.

— , **Kareem Haggag, Devin G. Pope, and Ryne Rohla**, "Racial Disparities in Voting Wait Times: Evidence from Smartphone Data," *The Review of Economics and Statistics*, 12 2020, pp. 1–27.

Coibion, Olivier, Yuriy Gorodnichenko, and Michael Weber, "Labor Markets During the Covid 19 Crisis: A Preliminary View," 2020. Working Paper.

Couronné, Thomas, Zbigniew Smoreda, and Ana-Maria Olteanu Raimond, "Chatty Mobiles:Individual mobility and communication patterns," *CoRR*, 2013, *abs/1301.6553.*

Couture, Victor, Jonathan I Dingel, Allison E Green, Jessie Handbury, and Kevin R Williams, "Measuring Movement and Social Contact with Smartphone Data: A Real-Time Application to COVID-19," Working Paper 27560, National Bureau of Economic Research July 2020.

Feng, Shuaizhang, Yingyao Hu, and Robert Moffitt, "Long run trends in unemployment and labor force participation in urban China," *Journal of Comparative Economics*, 2017, *45* (2), 304 – 324.

Forsythe, Eliza, "Labor market flows and the Covid-19 economy," Technical Report, Available at SSRN 3586074 2020.

Giles, John, Albert Park, and Juwei Zhang, "What is China's true unemployment rate?," *China Economic Review*, 2005, *16* (2), 149–170.

GMTB, "2020 Travel Survey by Guangzhou Municipal Transportation Bureau," Technical Report 2020.

Goldberg, Penny, "Global value chains, COVID-19, and the future of trade," Technical Report, Paper presented at Princeton University 2020.

Gong, Huiwen, Robert Hassink, Juntao Tan, and Dacang Huang, "Regional Resilience in Times of a Pandemic Crisis: The Case of COVID-19 in China," *Tijdschrift voor Economische en Sociale Geografie*, 2020, *111* (3), 497–512.

González, Marta C., César A. Hidalgo, and Albert-László Barabási, "Understanding individual human mobility patterns," *Nature*, 2008, *453*, 779–782.

Gupta, Sumedha, Thuy D Nguyen, Felipe Lozano Rojas, Shyam Raman, Byungkyu Lee, Ana Bento, Kosali I Simon, and Coady Wing, "Tracking Public and Private Responses to the COVID-19 Epidemic: Evidence from State and Local Government Actions," Working Paper 27027, National Bureau of Economic Research April 2020.

Heffetz, Ori and Daniel B. Reeves, "Difficulty of Reaching Respondents and Nonresponse Bias: Evidence from Large Government Surveys," *The Review of Economics and Statistics*, 03 2019, *101* (1), 176–191.

Hsiang, Solomon, Daniel Allen, Sébastien Annan-Phan, Kendon Bell, Ian Bolliger, Trinetta Chong, Hannah Druckenmiller, Luna Yue Huang, Andrew Hultgren, Emma Krasovich, Peiley Lau, Jaecheol Lee, Esther Rolf, Jeanette Tseng, and Tiffany Wu, "The effect of large-scale anti-contagion policies on the COVID-19 pandemic," *Nature*, 2020.

Jones, Stephen R. G. and W. Craig Riddell, "The Measurement of Unemployment: An Empirical Approach," *Econometrica*, 1999, *67* (1), 147–161.

Kennedy, Peter E., "Estimation with Correctly Interpreted Dummy Variables in Semilogarithmic Equations," *The American Economic Review*, 1981, *71* (4), 801–801.

Kraemer, Moritz, Chia-Hung Yang, Bernardo Gutierrez, Chieh-Hsi Wu, Brennan Klein, David M. Pigott, , Louis du Plessis, Nuno R. Faria, Ruoran Li, William P. Hanage, John S. Brownstein, Maylis Layan, Alessandro Vespignani, Huaiyu Tian, Christopher Dye, Oliver G. Pybus, and Samuel V. Scarpino, "The effect of human mobility and control measures on the COVID-19 epidemic in China," *Science*, 2020, *368* (6490), 493–497.

Kreindler, Gabriel E and Yuhei Miyauchi, "Measuring Commuting and Economic Activity inside Cities with Cell Phone Records," Working Paper 28516, National Bureau of Economic Research February 2021.

Krueger, Alan B., Alexandre Mas, and Xiaotong Niu, "The Evolution of Rotation Group Bias: Will the Real Unemployment Rate Please Stand Up?," *The Review of Economics and Statistics*, 05 2017, *99* (2), 258–264.

Liu, Qian, "Unemployment and labor force participation in urban China," *China Economic Review*, 2012, *23* (1), 18–33.

Meyer, Bruce D., Wallace K. C. Mok, and James X. Sullivan, "Household Surveys in Crisis," *Journal of Economic Perspectives*, November 2015, *29* (4), 199–226.

Miyauchi, Yuhei, Kentaro Nakajima, and Stephen J Redding, "Consumption access and agglomeration: evidence from smartphone data," Technical Report, National Bureau of Economic Research 2020.

Poterba, James M. and Lawrence H. Summers, "Reporting Errors and Labor Market Dynamics," *Econometrica*, 1986, *54* (6), 1319–1338.

Rao, Deyu, "The Role of Environmental Amenities in the Urban Economy: Evidence From a Spatial General Equilibrium Approach," 2021. Working Paper.

Ribar, David, "How to improve participation in social assistance programs," 2020. IZA World of Labor.

Toole, Jameson L., Yu-Ru Lin, Erich Muehlegger, Daniel Shoag, Marta C. González, and David Lazer, "Tracking employment shocks using mobile phone data," *Journal of The Royal Society Interface*, 2015, *12* (107), 20150185.

von Gaudecker, Hans-Martin, Radost Holler, Lena Janys, Bettina Siflinger, and Christian Zimpelmann, "Labour supply in the early stages of the COVID-19 pandemic: Empirical evidence on hours, home office, and expectations," Technical Report, IZA working paper 2020.

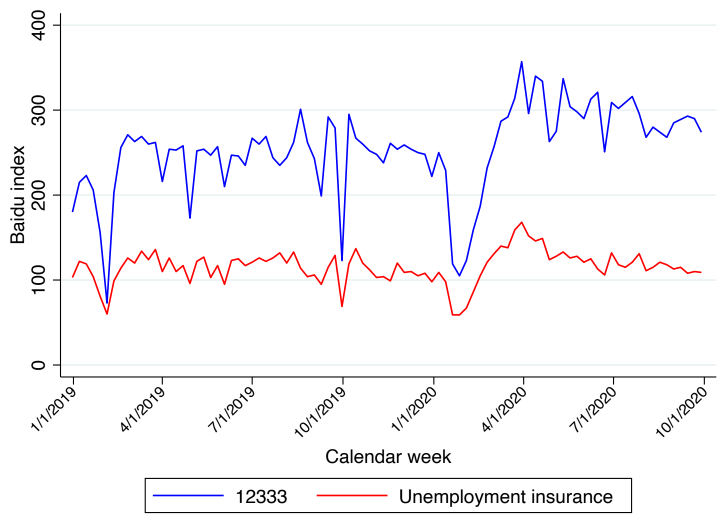
World Bank, "Reducing inequality for shared growth in China: strategy and policy options for Guangdong province.," 2010. The World Bank report.

World Trade Organization, "Trade shows signs of rebound from COVID-19, recovery still uncertain," Technical Report, WTO Technical Report 2020a.

_ , "The economic impact of COVID-19 on women in vulnerable sectors and economies," Technical Report, WTO Technical Report 2020b.

Zhang, Renyi, Yixin Li, Annie L. Zhang, Yuan Wang, and Mario J. Molina, "Identifying airborne transmission as the dominant route for the spread of COVID-19," *Proceedings of the National Academy of Sciences*, 2020, *117* (26), 14857–14863.

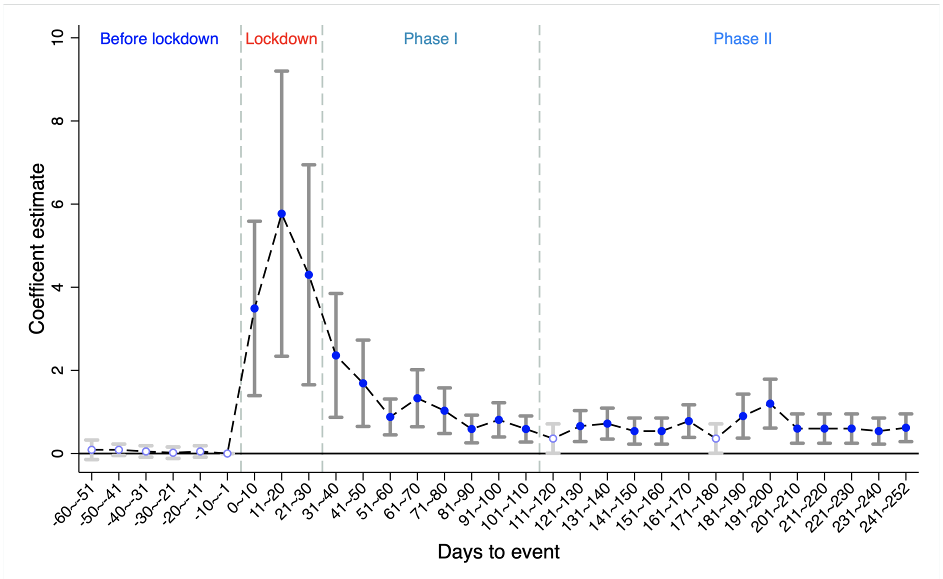
Figure 1: Correlation of Baidu search indices for keywords "12333" and "unemployment insurance" in Guangdong province



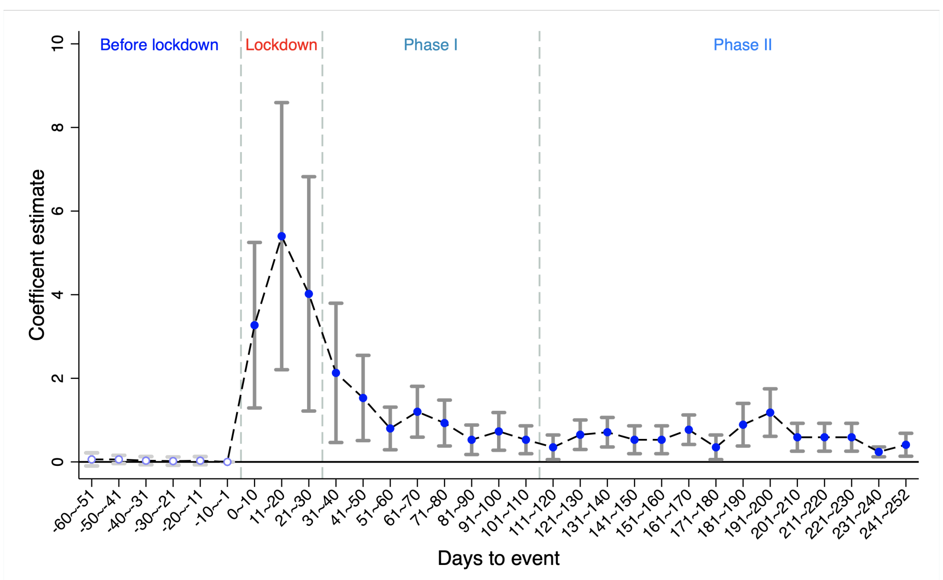
Notes: This graph shows the weekly Baidu Search Index for the keywords – "12333" and "unemployment insurance" – in Guangdong province from 2019 to 2020. Baidu Index, which is similar to Google Trends, is a keyword-analysis tool launched by Baidu, the largest search engine company in China. It reflects the search frequency of certain keywords on the Baidu website by users in a specified region (Guangdong in this example). The correlation between the two keywords during the sample period is 0.83.

Figure 2: Event study on differences in non-commuters between 2019 and 2020

(a) Non-commuters

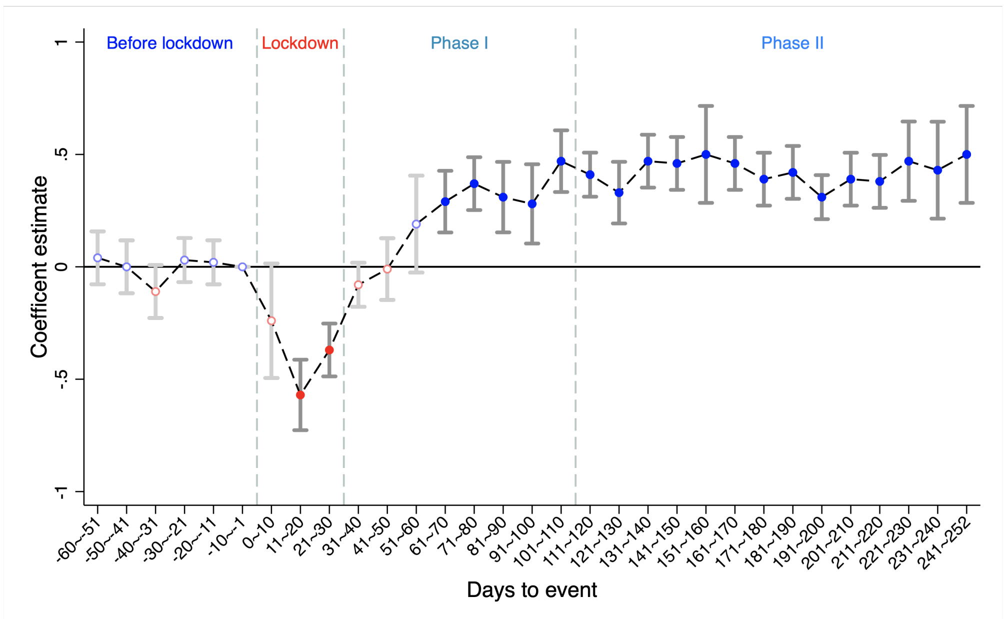


(b) Non-commuters who do not use emails/virtual meeting apps



Notes: Both panels report the event study coefficients for non-commuters (which we use as a measure of unemployment), with 2019 as the control group. The dependent variable is the number of non-commuters (in log) in panel (a) and the number of non-commuters who stopped using email/virtual meeting apps (in log) in panel (b). Panel (a) depicts changes in the number of non-commuters in 2020 relative to that in 2019, and panel (b) is based on the number of non-commuters who also stopped using email/virtual meeting apps. The event days are based on the lunar calendar. For 2020, the event date (or day zero) is the lockdown, January 23, 2020, two days before the Chinese New Year. Correspondingly, the event date in 2019 is February 3, 2019, also two days before the 2019 Chinese New Year. Phase I reopening started on February 24, 2020, 32 days after the lockdown, when people were allowed to go back to work and visit outdoor public places. Phase II reopening, or full reopening, started on May 9, 2020, 108 days after the lockdown. Shopping malls, supermarkets, restaurants were allowed to fully reopen.

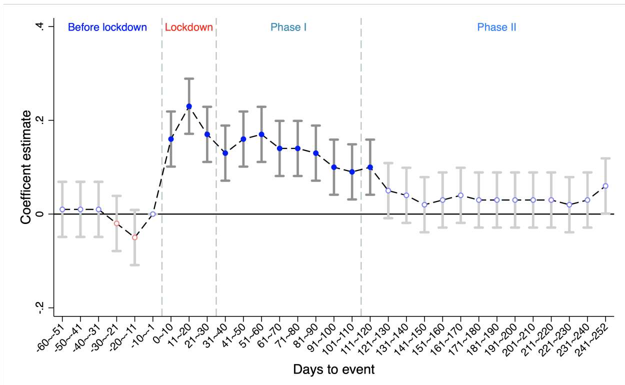
Figure 3: Event study on differences in calls to the unemployment benefit hotline between 2019 and 2020



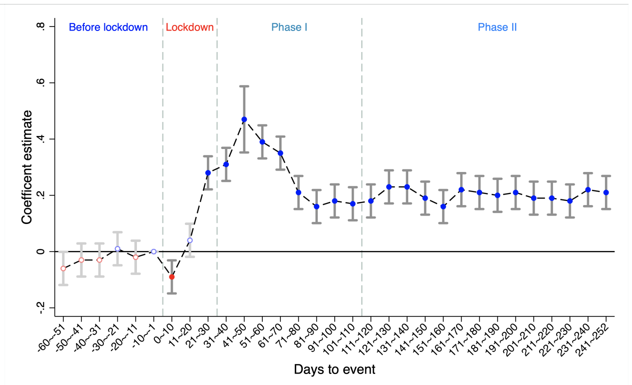
Notes: This graph depicts the event study coefficients for the daily number of individuals calling the unemployment hotline 12333 (in thousands) in 2020 relative to that in 2019. The dependent variable is the number of calls to the unemployment benefit hotline (in log). The event days are based on the lunar calendar. For 2020, the event date (or day zero) is the lockdown, January 23, 2020, two days before the Chinese New Year. Correspondingly, the event date in 2019 is February 3, 2019, also two days before the 2019 Chinese New Year. Phase I reopening started on February 24, 2020, 32 days after the lockdown, when people were allowed to go back to work and visit outdoor public places. Phase II reopening, or full reopening, started on May 9, 2020, 108 days after the lockdown. Shopping malls, supermarkets, restaurants were allowed to fully reopen.

Figure 4: Heterogeneity in the pandemic's impact on unemployment across demographic groups

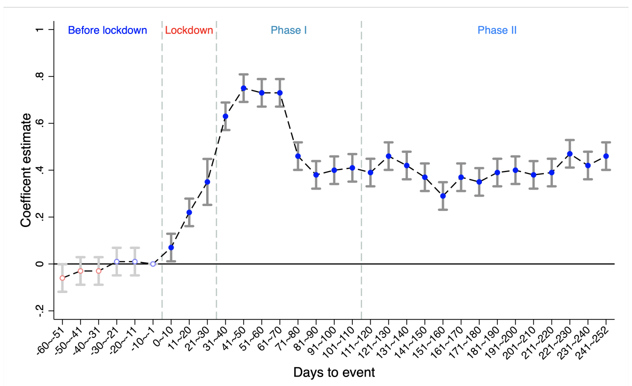
(a) Female v.s. male



(b) 40 and above v.s. under 40 years old

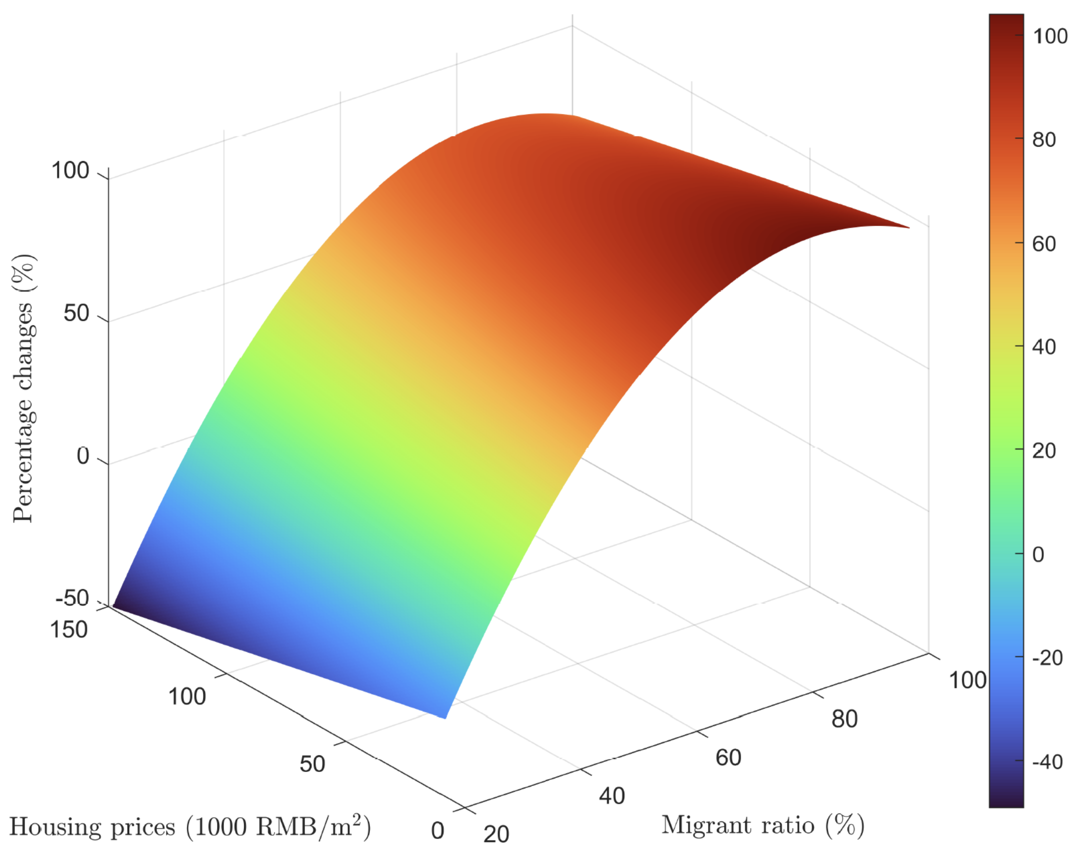


(c) Migrants v.s. non-migrants



Notes: All event-study graphs plot coefficient estimates of β_q from Equation (1). The dependent variable is the difference between log(female non-commuters) and log (male non-commuters) in panel (a), the difference between log non-commuters who are 40 and above and log non-commuters who are below 40-year-old in panel (b), and the difference between log non-commuters who are migrants and log non-commuters who are non-migrants in panel (c). This analysis uses one week to define non-commuters. All regressions include neighborhood, event-week, and the treatment group fixed effects. The standard errors are clustered at the event-week level. Using two weeks or one month to define non-commuters delivers similar results.

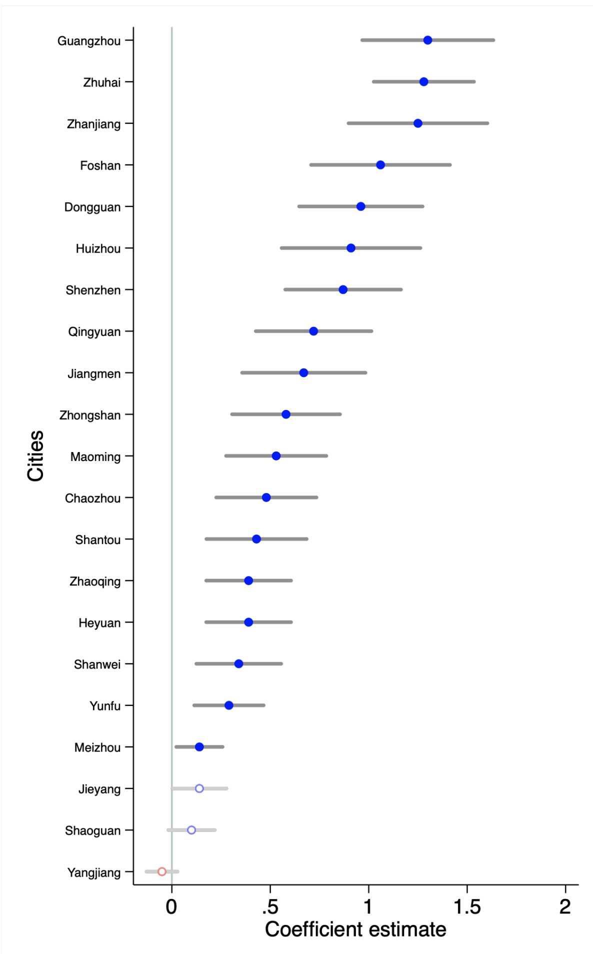
Figure 5: Changes in unemployment by income and migrant share



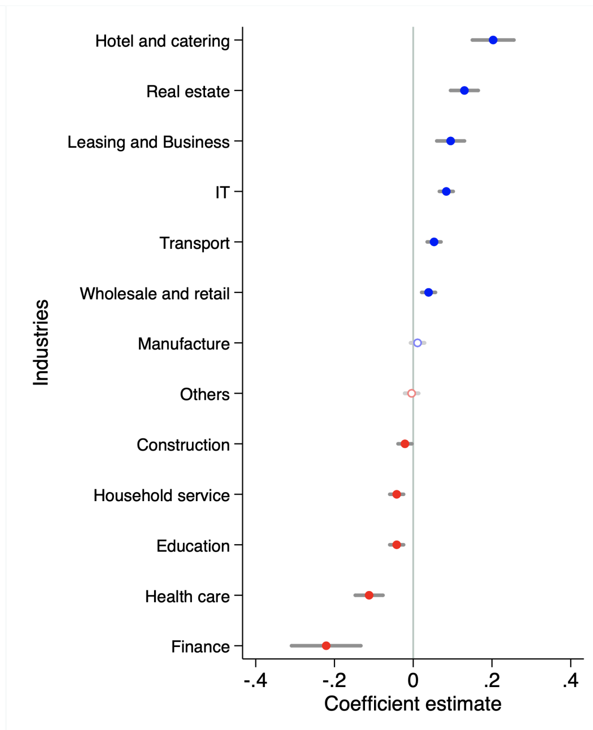
Notes: This graph depicts the percentage changes in the number of non-commuters between 2019 and 2020 at the neighborhood level (i.e., cell-tower-area) based on a regression that includes quadratic forms of the average housing price and migrant share for each neighborhood in Guangzhou. The housing prices come from Soufang.com. Migrant shares are based on our phone data in 2018.

Figure 6: Heterogeneity in the pandemic's impact on unemployment across industries and cities

(a) By city



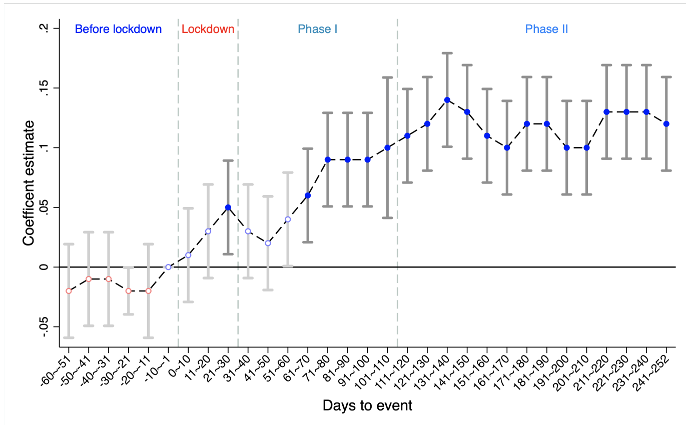
(b) By industry



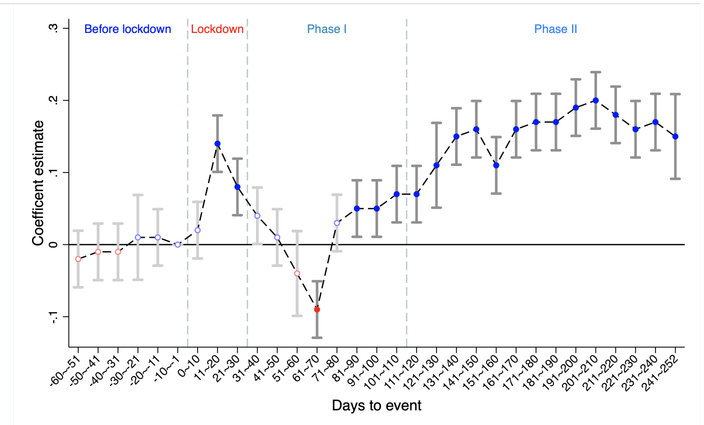
Notes: This figure illustrates heterogeneity in unemployment (i.e., non-commuters) across cities (panel (a)) and industries (panel (b)) following Equation (2). In panel (a), we add interactions between the after-lockdown dummy and city fixed effects. In panel (b), we add interactions between the after-lockdown dummy and a city's share of employment in each of the 13 industries. A positive change indicates an increase in non-commuters relative to 2019. This analysis uses one week to define non-commuters. Both regressions include neighborhood, event-week, and the treatment group fixed effects. The standard errors are clustered at the event-week level. Results are similar when using two weeks or one month to define non-commuters.

Figure 7: Heterogeneity in unemployment benefit claims across demographic groups

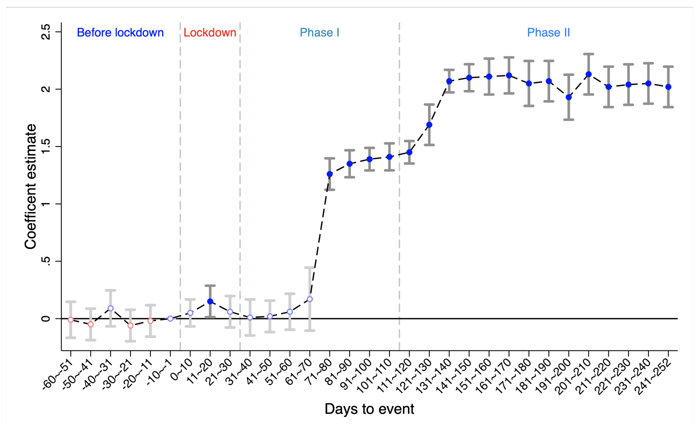
(a) Female v.s. male



(b) 40 and above v.s. under 40 years old



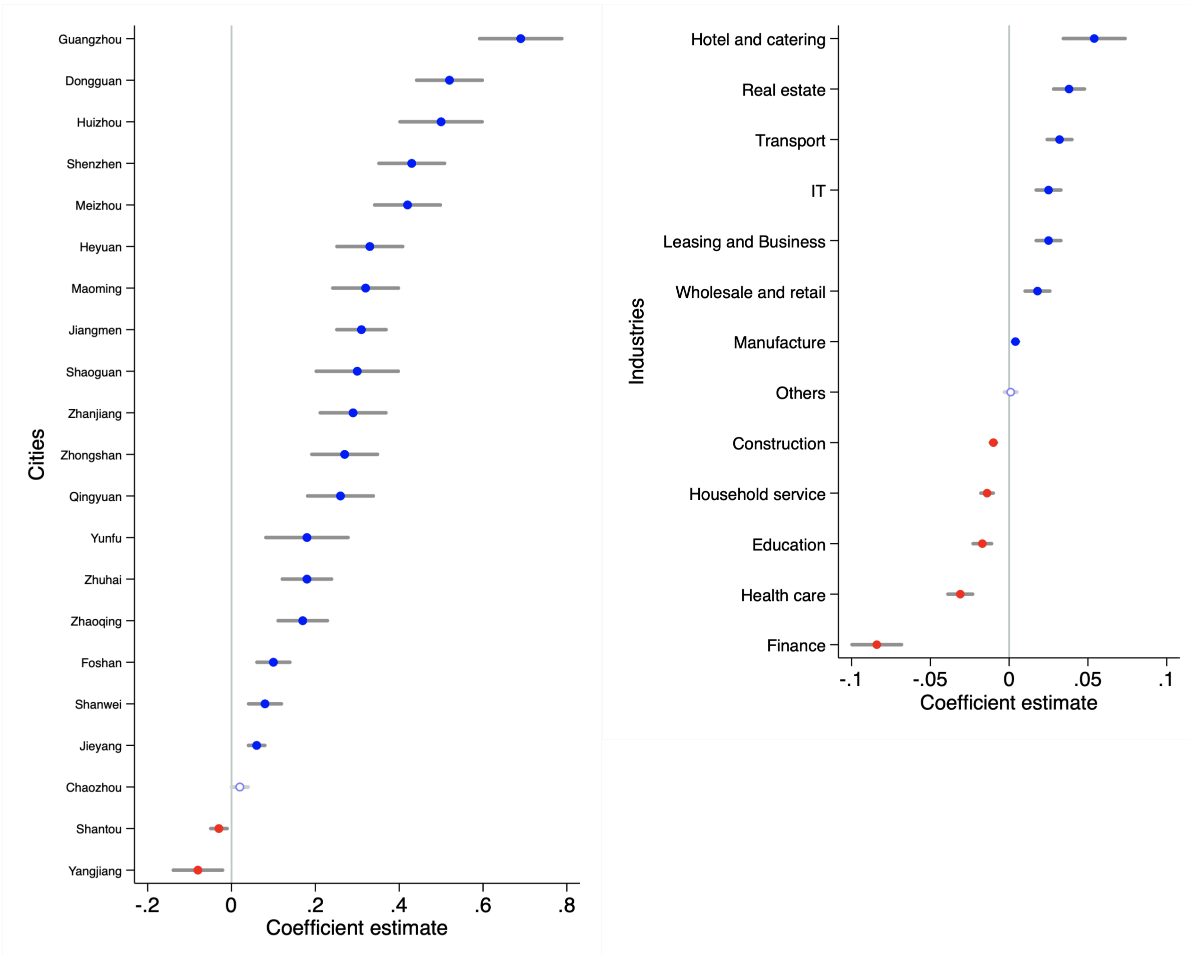
(c) Migrants v.s. non-migrants



Notes: All event-study graphs plot coefficient estimates of β_q from Equation (1). The dependent variable is the difference between log number of female calling the hotline and log number of male calling the hotline in panel (a), the difference between log number of individuals who call the hotline and are 40 and above and log number of those who call the hotline and are under 40 in panel (b), and the difference between log number of migrants calling the hotline and log number of non-migrants calling the hotline in panel (c). All regressions include neighborhood, day-of-week, event-day, holiday, and the treatment group fixed effects. The standard errors are clustered at the event-day level.

Figure 8: Heterogeneity in unemployment benefit claims across industries and cities

(a) By city (b) By industry



Notes: This figure illustrates heterogeneity across cities (panel (a)) and industries (panel (b)) following Equation (2). In panel (a), we add interactions between the after-lockdown dummy and city fixed effects. In panel (b), we add interactions between the after-lockdown dummy and a city's share of employment in each of the 13 industries. A positive change indicates an increase in the number of individuals reaching out to the unemployment benefit hotline relative to 2019. Both regressions include neighborhood, day-of-week, event-day, holiday, and the treatment group fixed effects. The standard errors are clustered at the event-day level.

Table 1: The pandemic's impact on non-commuters and working hours on-site

Variable	(1) No. of non-commuters (two-week window, in log)	(2) Working hours (in log)
1-30 days before lockdown	0.07	0.01
	(0.05)	(0.01)
Lockdown period	4.51***	-0.21***
	(1.26)	(0.02)
Phase I re-opening	1.03***	-0.08***
	(0.36)	(0.01)
Phase II re-opening	0.59**	-0.02
	(0.30)	(0.02)
Observations	34,965	34,965
R-squared	0.92	0.95
Neighborhood FE	Yes	Yes
Event-fortnight FE	Yes	Yes
Treatment group FE	Yes	Yes

Notes: This table examines the pandemic's effect on the number of non-commuters defined over a two-week window and duration of on-site working hours. It is similar to Equation (1), except that the ten-day intervals are grouped into four periods: before the lockdown, during the lockdown, Phase I reopening, and Phase II full reopening. The observations are at the neighborhood by fortnight level. The dependent variable is the log number of non-commuters in column (1) and log number of average working hours for commuters in column (2), respectively. A non-commuter is an individual who visits his work location at least 15 days in the previous 30 days and stops commuting altogether in the current two-week window. Both columns include neighborhood, event-fortnight, holiday, and the treatment group fixed effects. Standard errors are reported in parentheses and clustered at the event-fortnight level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 2: The pandemic's impact on calls to unemployment benefit hotline and call duration

Variable	(1) No. of individuals making calls (in log)	(2) Average call duration (in log)
1-30 days before lockdown	0.03	0.03
	(0.03)	(0.04)
Lockdown period	-0.37***	-0.36***
	(0.06)	(0.08)
Phase I re-opening	0.25***	0.24***
	(0.03)	(0.05)
Phase II re-opening	0.45***	0.56***
	(0.02)	(0.04)
Observations	489,514	489,514
R-squared	0.81	0.57
Neighborhood FE	Yes	Yes
Event-day FE	Yes	Yes
Day-of-week FE	Yes	Yes
Holidays FE	Yes	Yes
Treatment group FE	Yes	Yes

Notes: The dependent variable is the log number of individuals calling unemployment benefit hotline in column (1) and the log of average call duration in seconds in column (2), respectively. The table examines the pandemic's effect on the number of individuals making unemployment calls and call duration. It is similar to Equation (1), except that the ten-day intervals are grouped into four periods: before lockdown, during the lockdown, Phase I reopening, and Phase II full reopening. The observations are at the neighborhood by day level. Both columns include neighborhood, day-of-week, event-day, holiday, and the treatment group fixed effects. Standard errors are reported in parentheses and clustered at the event day: *p<0.10; **p<0.05; ***p<0.01.

Table 3: Heterogeneity in the pandemic's impact on non-commuters by export-to-GDP ratio

Variable	(1) No. of non-commuters (two-week window, in log)	(2) Working hours (in log)
Phase II re-opening * (Export/GDP in 2019)	0.37***	-0.02*
	(0.07)	(0.01)
Phase II re-opening (=1)	0.03***	-0.06***
	(0.01)	(0.01)
Observations	34,965	34,965
R-squared	0.93	0.81
Neighborhood FE	Yes	Yes
Event-fortnight FE	Yes	Yes
Treatment group FE	Yes	Yes

Notes: This table examines whether the pandemic's effect differs across cities with varying exposure to international trade following Equation (2), where we interact the phase II re-opening dummy with a city's 2019 export-to-GDP ratio (in percentage). The dependent variable is the log number of non-commuters in column (1) and log of the average working hours for commuters in column (2), respectively. A non-commuter is someone who visits his work location at least 15 days in last 30 days and stops commuting in next two weeks. Both columns include neighborhood, event-fortnight, and the treatment group fixed effects. Standard errors are reported in parentheses and clustered at event-fortnight. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Results are similar using one week or one month to define non-commuters.

Table 4: Heterogeneity in the pandemic's impact on calls to unemployment benefit hotline by export-to-GDP ratio

Variable	(1) No. of individuals making calls (in log)	(2) Average call duration (in log)
Phase II re-opening * (Export/GDP in 2019)	0.24***	0.25***
	(0.05)	(0.06)
Phase II re-opening (=1)	0.43***	0.47***
	(0.06)	(0.09)
Observations	489,514	489,514
R-squared	0.76	0.54
Neighborhood FE	Yes	Yes
Event-day FE	Yes	Yes
Day-of-week FE	Yes	Yes
Holidays FE	Yes	Yes
Treatment group FE	Yes	Yes

Notes: This table examines whether the pandemic's effect differs across cities with varying exposure to international trade following Equation (2), where we interact the phase II re-opening dummy with a city's 2019 export-to-GDP ratio (in percentage). The dependent variables in columns (1)-(2) are the number of individuals who made unemployment calls and stopped commuting for at least fortnight in the current month and average duration of unemployment calls in seconds (in logarithm), respectively. The observations are at the neighborhood and day level. Both columns include neighborhood, day-of-week, event-day, holiday, and the treatment group fixed effects. Standard errors are reported in parentheses and clustered at the event-day. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 5: The pandemic's impact on calls to unemployment benefit hotline by non-commuters

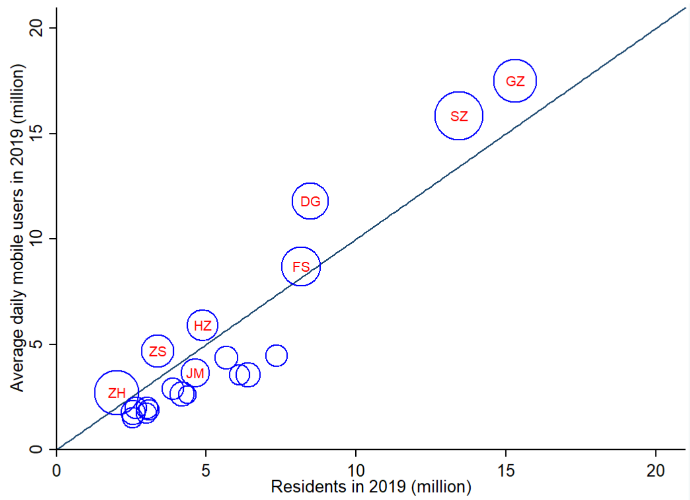
Variable	(1) No. of non-commuters making calls (in log)	(2) Average call duration (in log)
1-30 days before lockdown	0.02	0.03
	(0.04)	(0.04)
Lockdown period	-0.42***	-0.40***
	(0.07)	(0.07)
Phase I re-opening	0.23***	0.26***
	(0.03)	(0.04)
Phase II re-opening	0.40***	0.58***
	(0.03)	(0.05)
Observations	489,514	489,514
R-squared	0.81	0.57
Neighborhood FE	Yes	Yes
Event-day FE	Yes	Yes
Day-of-week FE	Yes	Yes
Holidays FE	Yes	Yes
Treatment group FE	Yes	Yes

Notes: This table replicates results in Table 2, but limits the sample that reaches out to the unemployment benefit hotline to non-commuters. Both columns include neighborhood, day-of-week, event-day, holiday, and the treatment group fixed effects. Standard errors are reported in parentheses and clustered at the event day. *p<0.10; **p<0.05; ***p<0.01.

Appendices. For Online Publication Only

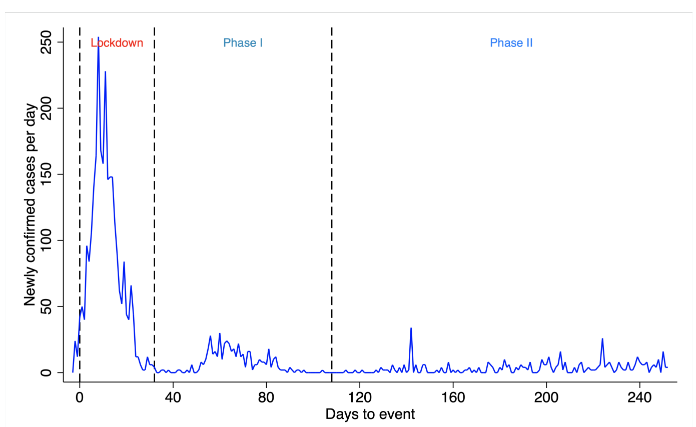
Appendix A: Figures and Tables

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



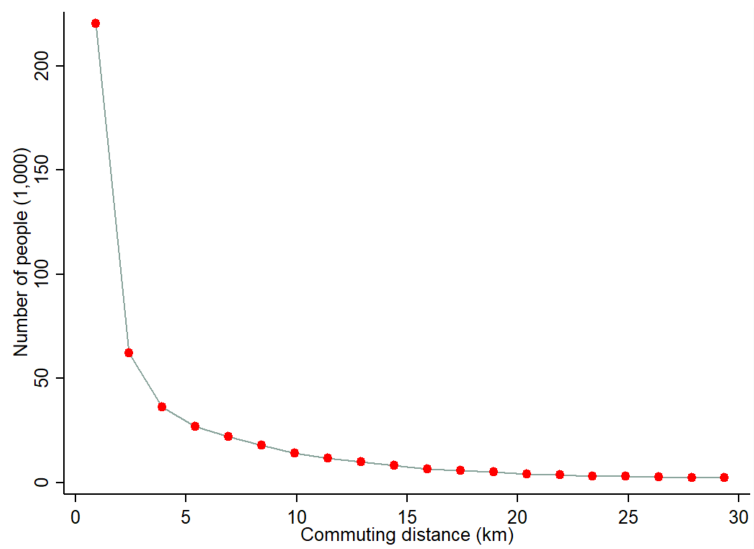
Notes: This graph presents the relationship between average daily mobile users and residents for cities in Guangdong in 2019. The solid line is a 45-degree line. The size of each marker denotes GDP per capita in 2019. We label cities whose 2019 GDP per capita exceeds 60,000 RMB (around $8,500). The abbreviated city names are listed in Table A.1.

Figure A.2: Newly confirmed COVID-19 cases in Guangdong province



Notes: This graph shows the daily number of newly confirmed COVID-19 cases in Guangdong province. The provincial government started to report the number of COVID-19 cases on January 10, 2020, 13 days before its lockdown.

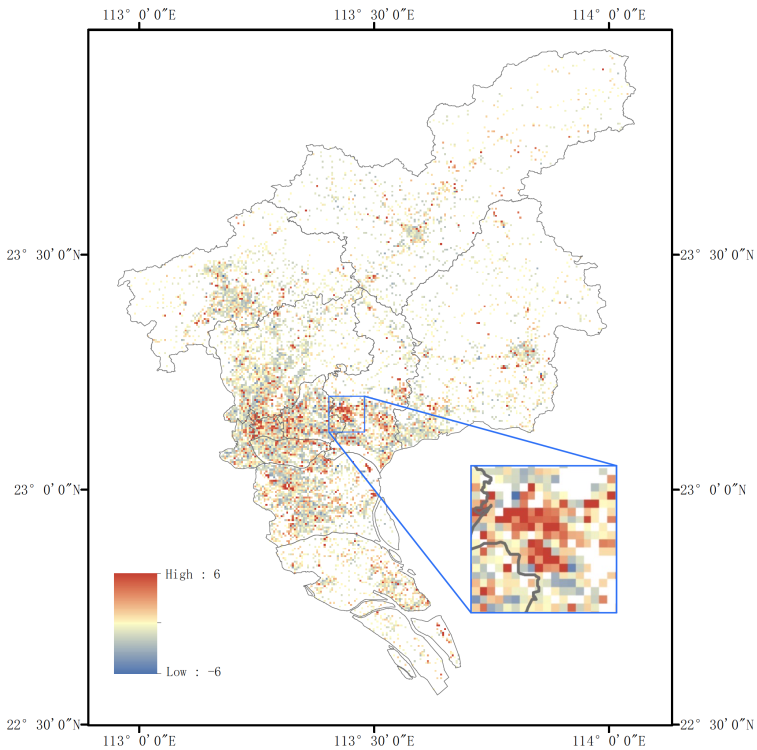
Figure A.3: The number of commuters by commuting distance



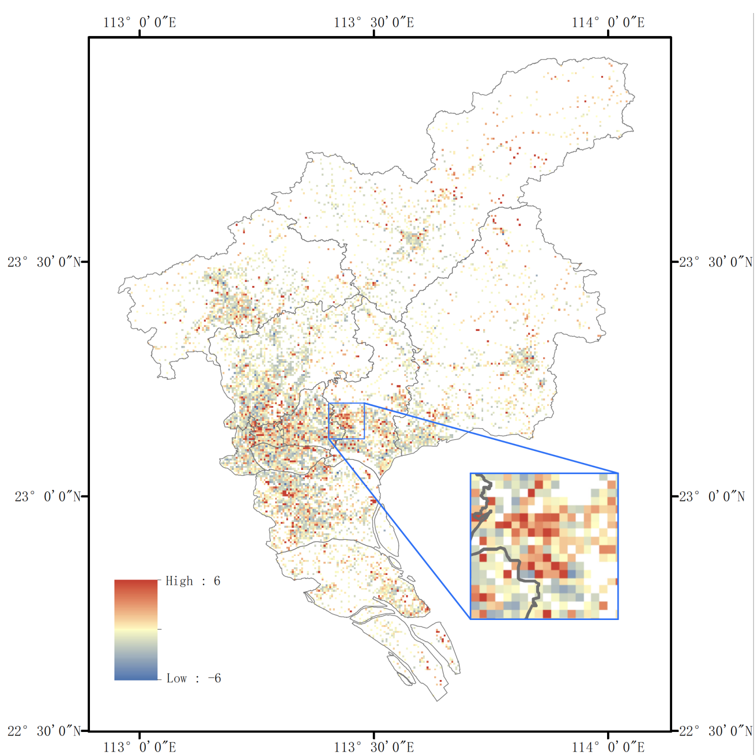
Notes: This graph shows the number of people commuting at each distance for our commuter sample. We use the coordinates of job and home locations to compute the spherical commuting distance.

Figure A.4: Differences in (log) population during daytime and nighttime in Guangzhou

(a) Weekdays

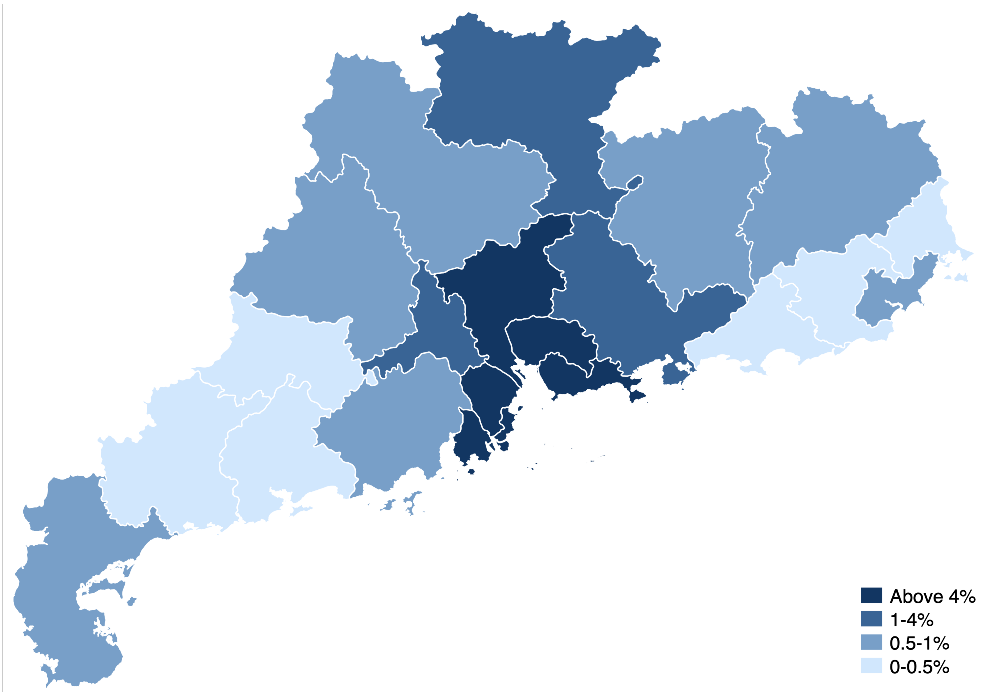


(b) Weekends



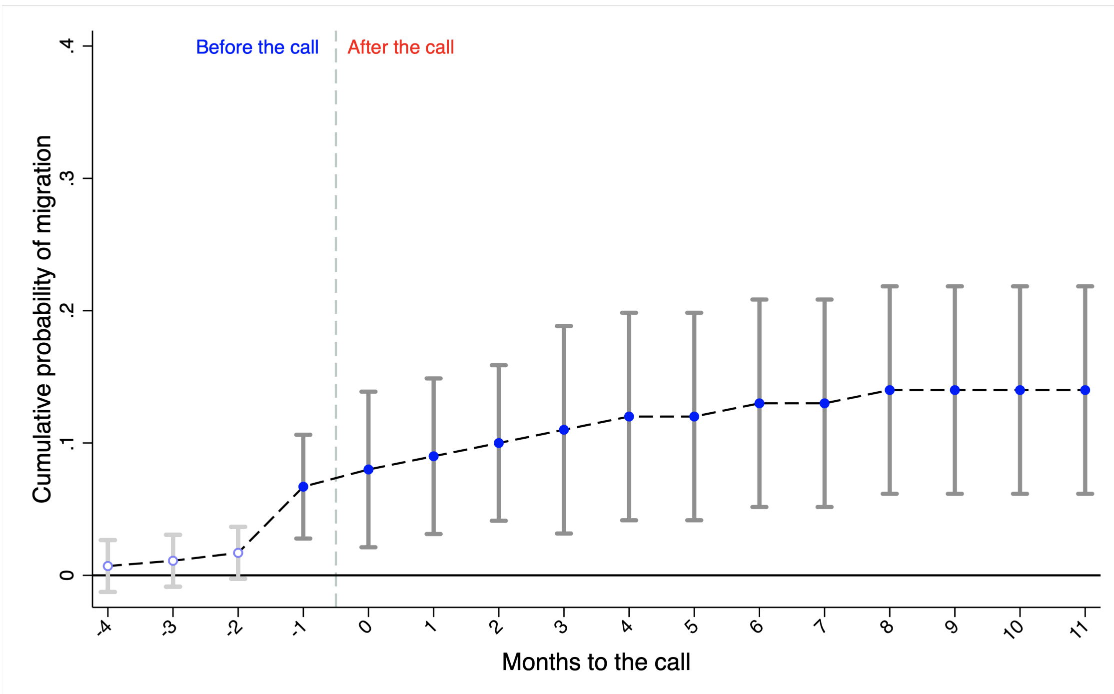
Notes: Panel (a) displays the average daily log difference of users during daytime (11 am) and nighttime (11 pm) at the centroid of every phone tower on the weekdays of 2019. In contrast, panel (b) shows the same measure on the weekends of 2019.

Figure A.5: Estimated unemployment rate based on individuals calling unemployment benefit hotline



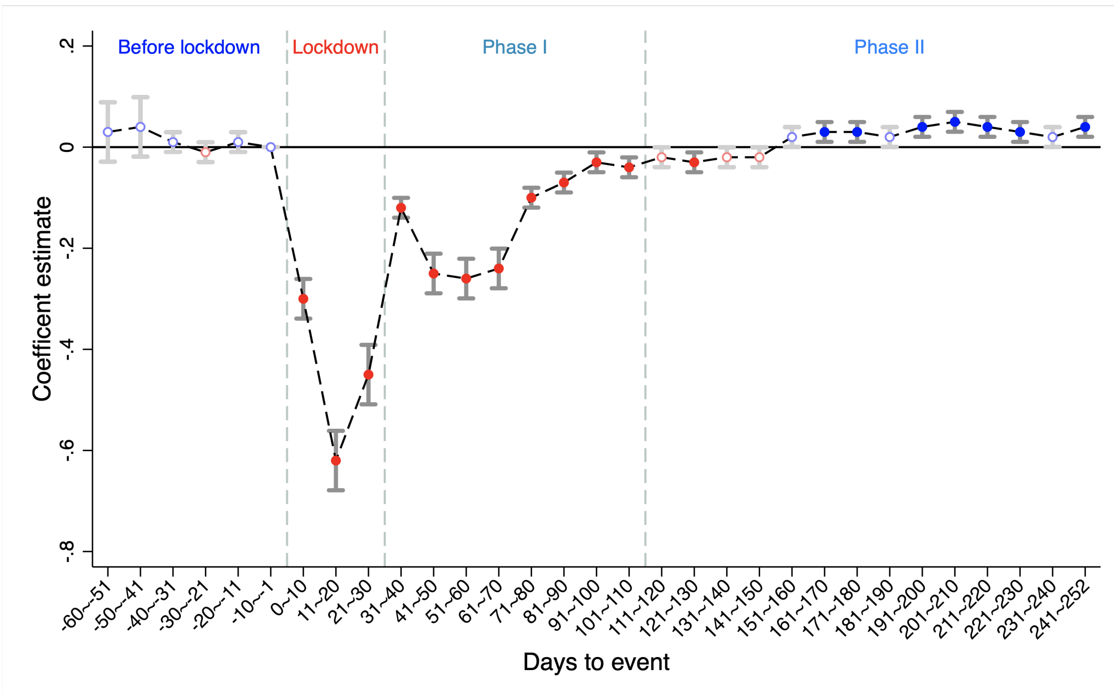
Notes: This graph shows our estimated unemployment rate, which is the ratio of individuals making unemployment calls over the size of the labor force by city in 2019. The correlation between our unemployment measure and the official unemployment measure at the city level is 0.7 in 2019.

Figure A.6: Cumulative probability of migration for individuals calling unemployment benefit hotline



Notes: This figure shows the cumulative probability of migration among individuals reaching out to the unemployment benefit hotline. To increase the precision of the residential city measure, all valid observations must have resided in a given city for at least two months. Migrated individuals are those who lived in one city for at least two months and moved and lived in other cities for at least two months. The x-axis denotes the months to making the calls. The standard errors are denoted by the length of the vertical bars. Slightly more than 10% of users migrated to other cities two months after contacting the unemployment benefit agencies.

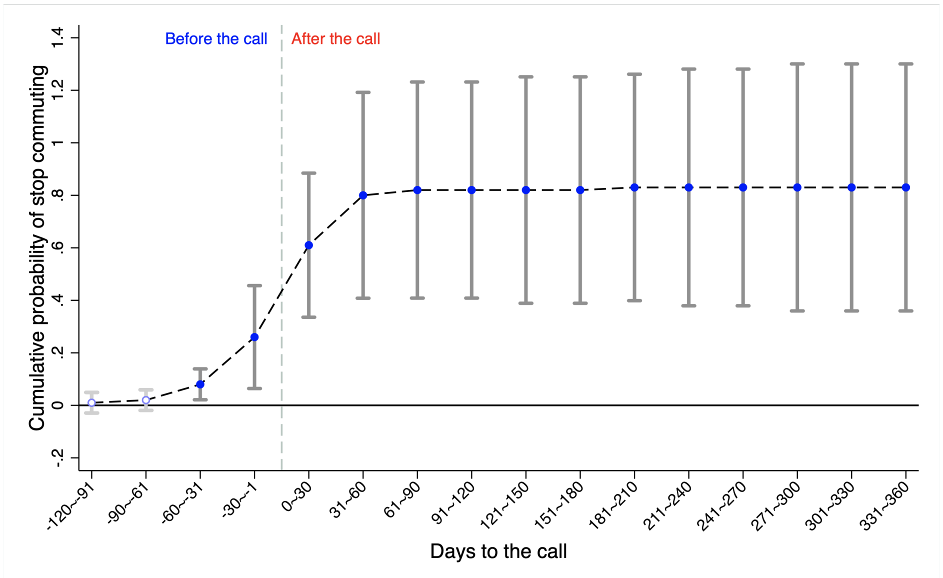
Figure A.7: Event study on hours spent on non-work activities away from home



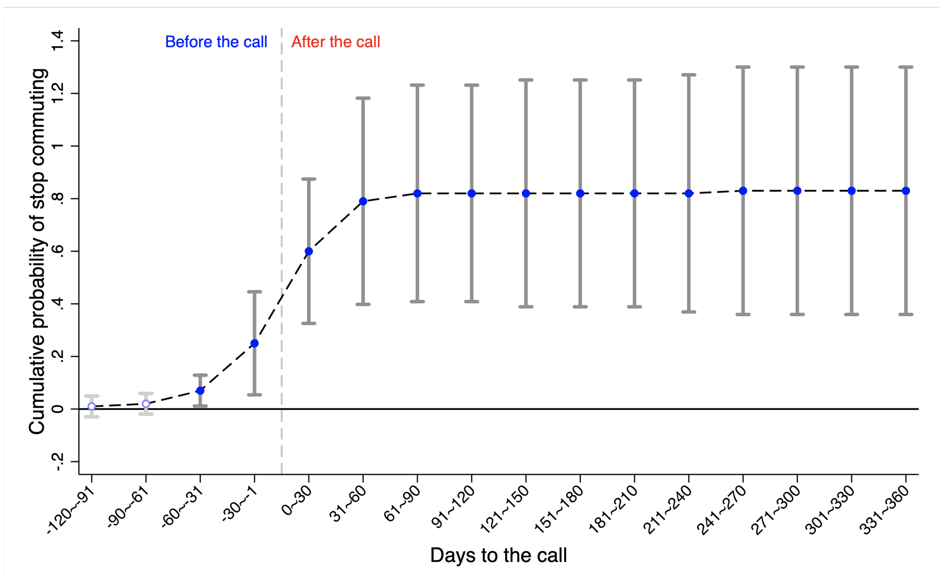
Notes: This figure depicts the changes in the hours of non-work activities away from home in 2020 relative to that in 2019. See Figure 2 for the explanation of various event days. Non-work activities as those that happened at places other than home and the workplace and lasted for at least half an hour.

Figure A.8: Cumulative probability of stopping commuting for individuals calling unemployment benefit hotline: different non-commuter definitions

(a) Non-commuting defined by two weeks

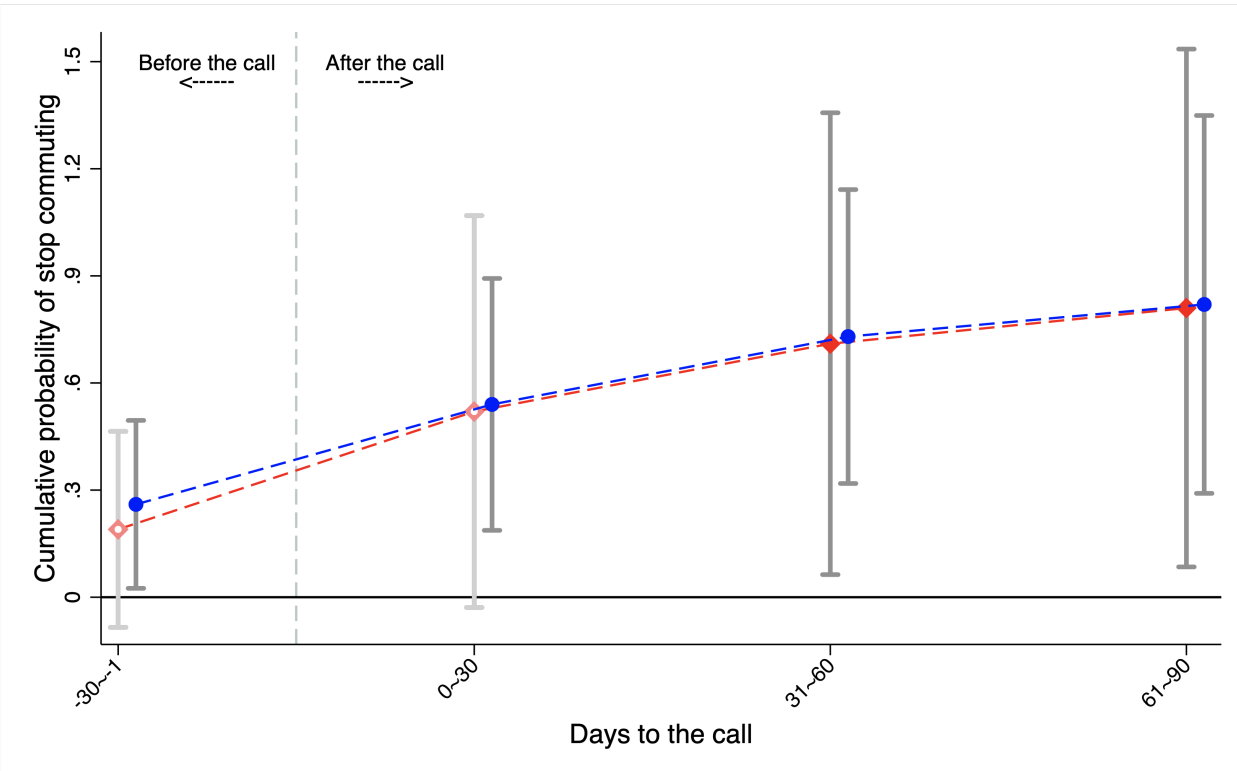


(b) Non-commuting defined by one month



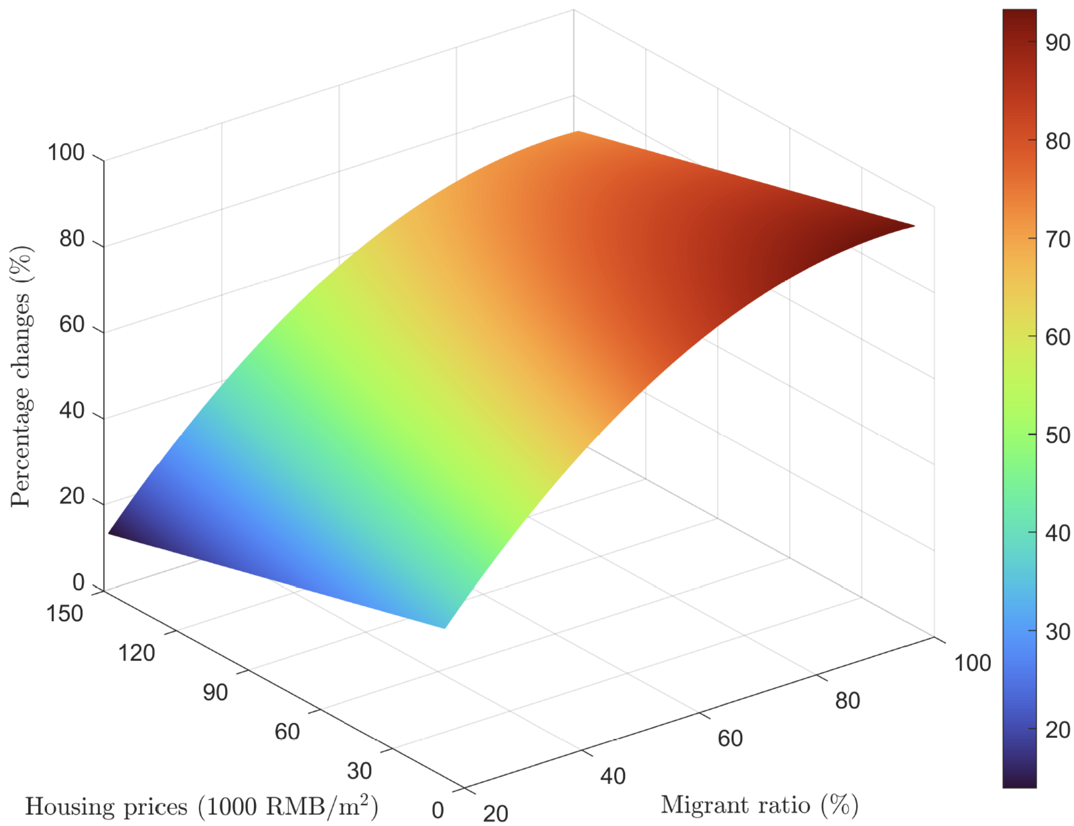
Notes: This figure shows the cumulative probability of stopping commuting for at least two weeks (panel a) and one month (panel b) among individuals reaching out to the unemployment benefit hotline. The vertical line denotes the event day of calling 12333. The x-axis denotes the days to the call. The standard errors are denoted by the length of the vertical bars. About 20% of the callers had already stopped commuting one month prior to the call. The cumulative probability grows to 80% by three months after the call. These cumulative probabilities are remarkably robust to different non-commuter definitions (whether by week or by month).

Figure A.9: Cumulative probability of stopping commuting for individuals calling unemployment benefit hotline: 2019 vs. 2020



Notes: This figure shows the cumulative probability of stopping commuting for at least two weeks among individuals reaching out to the unemployment benefit hotline. The x-axis denotes the days to making the calls. The line with the red diamond (blue dot) represents 2019 (2020). The standard errors (denoted by the length of the vertical bars) are larger for 2019 because there were fewer people calling the unemployment benefit hotline in 2019 than in 2020. About 20% of the callers had already stopped commuting one month prior to the call. The cumulative probability grows to 80% by three months after the call. These cumulative probabilities are remarkably similar between 2019 and 2020, suggesting no change in the relationship between calling unemployment benefit hotline and commuting patterns.

Figure A.10: Changes in unemployment benefit claims by income and migrant share



Notes: This graph depicts the percentage changes in the number of individuals calling the unemployment hotline between 2019 and 2020 at the neighborhood (i.e., cell-tower-area) level based on a regression including quadratic forms of the average housing price and migrant share for each neighborhood in Guangzhou. The housing prices from Soufang.com and the migrant shares are based on our phone data in 2018.

Table A.1: Cities in Guangdong province

Cities	Population in 2019 (million)	GDP in 2019 ($ billion)
Shenzhen (SZ)	13.44	390.25
Guangzhou (GZ)	15.31	342.44
Foshan (FS)	8.16	155.81
Dongguan (DG)	8.46	137.43
Huizhou (HZ)	4.88	60.54
Zhuhai (ZH)	2.02	49.80
Maoming (MM)	6.41	47.13
Jiangmen (JM)	4.63	45.60
Zhongshan (ZS)	3.38	44.94
Zhanjiang (ZJ)	7.36	44.42
Shantou (ST)	5.66	39.04
Zhaoqing (ZQ)	4.19	32.59
Jieyang (JY)	6.11	30.46
Qingyuan (QY)	3.89	24.61
Shaoguan (SG)	3.03	19.11
Yangjiang (YJ)	2.57	18.73
Meizhou (MZ)	4.38	17.20
Chaozhou (CZ)	2.66	15.67
Shanwei (SW)	3.02	15.66
Heyuan (HY)	3.11	15.65
Yunfu (YF)	2.55	13.36

Notes: This table shows the 2019 population and GDP for each city in Guangdong province. Data source: The Bureau of Statistics in Guangdong.

Table A.2: Summary statistics

Variable	N	Mean	Std. Dev.
Panel A: Commuting sample			
Num. of non-commuters per two weeks	34,965	37,624	14,327
Working hours	34,965	8.7	1.0
Female (=1)	1,000,000	0.4	0.5
Age	1,000,000	33.6	8.4
Migrant (=1)	1,000,000	0.4	0.4
Panel B: Unemployment-call sample			
Daily num. of individuals calling 12333	489,514	14.6	58.8
Call duration (seconds)	489,514	169.5	141.7
Female (=1)	6,208,225	0.5	0.5
Age	6,208,225	36.5	9.7
Migrant (=1)	6,208,225	0.6	0.5

Notes: Panel A presents the summary statistics for the commuting sample, which consists of one million users who are randomly extracted from all mobile phone users. The regression analysis aggregate the non-commuter sample to the neighborhood-fortnight level (34,965 observations). Panel B shows the summary statistics for individuals calling 12333. During our sample period, 6,208,225 individuals contacted the unemployment benefit agencies via the designated hotline. The regression analysis aggregate the caller data to the neighborhood-day level (489,514 observations).

Table A.3: The pandemic's impact on non-commuters and working hours: alternative non-commuter definitions

Variable	(1) No. of non-commuters (in log)	(2) Working hours (in log)
Panel A: One-week window		
1-30 days before lockdown	0.06	-0.00
	(0.05)	(0.01)
Lockdown period	4.95***	-0.20***
	(1.23)	(0.02)
Phase I re-opening	1.09***	-0.08***
	(0.37)	(0.02)
Phase II re-opening	0.61**	-0.02
	(0.31)	(0.02)
Observations	69,930	69,930
R-squared	0.95	0.94
Panel B: One-month window		
1-30 days before lockdown	0.02	0.01
	(0.03)	(0.01)
Lockdown period	4.32***	-0.24***
	(1.05)	(0.02)
Phase I re-opening	1.01***	-0.08***
	(0.38)	(0.01)
Phase II re-opening	0.58**	-0.01
	(0.29)	(0.02)
Observations	16,650	16,650
R-squared	0.96	0.95
Neighborhood FE	Yes	Yes
Event-time FE	Yes	Yes
Treatment group FE	Yes	Yes

Notes: This table replicates the analysis in Table 1, except that the data are aggregated at the neighborhood-week (panel A) and neighborhood-month (panel B) level. The dependent variables in columns (1)-(2) are the number of non-commuters (in logarithm) and average working hours for commuters (in logarithm), respectively. A non-commuter is someone who visits his work location at least 15 days in last 30 days and stops commuting in next week (panel A) and next month (panel B). Both columns include neighborhood, event-time, and the treatment group fixed effects. Standard errors are reported in parentheses and clustered at event-time. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.4: Effects on non-commuters who stopped using email or virtual meeting Apps

Variable	(1) No. of non-commuters (two-week window, in log)
1-30 days before lockdown	0.06
	(0.04)
Lockdown period	4.19***
	(1.17)
Phase I re-opening	0.98***
	(0.30)
Phase II re-opening	0.57**
	(0.27)
Observations	34,965
R-squared	0.91
Neighborhood FE	Yes
Event-fortnight FE	Yes
Treatment group FE	Yes

Notes: This table replicates column (1) in Table 1, except that a non-commuter is defined as who visits his work location at least 15 days in last 30 days and stops commuting and do not using email or virtual meeting Apps in next two weeks. We control neighborhood, event-fortnight, holiday, and the treatment group fixed effects. Standard errors are reported in parentheses and clustered at the event-fortnight level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.5: The pandemic's impact on non-commuters: excluding people under 25

Variable	(1) No. of non-commuters (two-week window, in log)	(2) Working hours (in log)
1-30 days before lockdown	0.06	0.01
	(0.05)	(0.01)
Lockdown period	4.63***	-0.21***
	(1.29)	(0.02)
Phase I re-opening	1.10***	-0.09***
	(0.39)	(0.02)
Phase II re-opening	0.60**	-0.02
	(0.31)	(0.02)
Observations	34,965	34,965
R-squared	0.93	0.88
Neighborhood FE	Yes	Yes
Event-fortnight FE	Yes	Yes
Treatment group FE	Yes	Yes

Notes: This table replicates the analysis in Table 1, except that individuals under 25 years old are excluded. The dependent variable is the log number of non-commuters in column (1) and log number of average working hours for commuters in column (2), respectively. A non-commuter is someone who visits his work location at least 15 days in last 30 days and stops commuting in next two weeks. Both columns include neighborhood, event-fortnight, and the treatment group fixed effects. Standard errors are reported in parentheses and clustered at event-fortnight. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Results are similar using one week or one month to define non-commuters.

Table A.6: The pandemic's impact on calls to unemployment benefit hotline: excluding people under 25

Variable	(1) No. of individuals making calls (in log)	(2) Average call duration (in log)
1-30 days before lockdown	0.03	0.03
	(0.04)	(0.05)
Lockdown period	-0.41***	-0.39***
	(0.07)	(0.06)
Phase I re-opening	0.25***	0.27***
	(0.04)	(0.04)
Phase II re-opening	0.41***	0.58***
	(0.04)	(0.05)
Observations	489,514	489,514
R-squared	0.81	0.57
Neighborhood FE	Yes	Yes
Event-day FE	Yes	Yes
Day-of-week FE	Yes	Yes
Holidays FE	Yes	Yes
Treatment group FE	Yes	Yes

Notes: This table replicates the analysis in Table 2, except that individuals under 25 years old are excluded. The dependent variable is log (number of individuals calling unemployment benefit hotline) in column (1) and log(average call duration in seconds) in column (2), respectively. Both columns include neighborhood, day-of-week, event-day, holiday, and the treatment group fixed effects. Standard errors are reported in parentheses and clustered at the event-day. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.7: The pandemic's impact on non-commuters: weighted regressions

Variable	(1) No. of non-commuters (two-week window, in log)	(2) Working hours (in log)
1-30 days before lockdown	0.08	-0.02
	(0.07)	(0.02)
Lockdown period	6.92***	-0.29***
	(2.22)	(0.02)
Phase I re-opening	1.76***	-0.12***
	(0.53)	(0.02)
Phase II re-opening	0.68**	-0.03
	(0.32)	(0.03)
Observations	34,965	34,965
R-squared	0.92	0.95
Neighborhood FE	Yes	Yes
Event-fortnight FE	Yes	Yes
Treatment group FE	Yes	Yes

Notes: This table replicates the analysis in Table 1, except that the regressions are weighted by the average number of commuters per day in 2019 in each neighborhood. Both columns include neighborhood, event-fortnight, holiday, and the treatment group fixed effects. Standard errors are reported in parentheses and clustered at the event-fortnight level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Results are similar using one week or one month to define non-commuters.

Table A.8: The pandemic's impact on calls to unemployment benefit hotline: weighted regressions

Variable	(1) No. of individuals making calls (in log)	(2) Average call duration (in log)
1-30 days before lockdown	0.02	0.05
	(0.06)	(0.06)
Lockdown period	-0.47***	-0.43***
	(0.08)	(0.08)
Phase I re-opening	0.29***	0.29***
	(0.05)	(0.05)
Phase II re-opening	0.50***	0.60***
	(0.05)	(0.05)
Observations	489,514	489,514
R-squared	0.81	0.57
Neighborhood FE	Yes	Yes
Event-day FE	Yes	Yes
Day-of-week FE	Yes	Yes
Holidays FE	Yes	Yes
Treatment group FE	Yes	Yes

Notes: This table replicates the analysis in Table 2, except that the regressions are weighted by the average number of commuters per day in 2019 in each neighborhood. Both columns include neighborhood, day-of-week, event-day, holiday, and the treatment group fixed effects. Standard errors are reported in parentheses and clustered at the event-day. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.