Title: COVID-19 Pandemic and Unemployment: Evidence from Mobile Phone Data in China

Short title: COVID-19 Pandemic and Unemployment

Authors: Panle Jia Barwick,[1]*† Yongheng Deng,[2]† Xinfei Huang,[3]† Shanjun Li,[1]† Teng Li[3]†

Affiliations:

[1] Cornell University and NBER, Ithaca, NY 14853, USA.

[2] University of Wisconsin-Madison, WI 53706, USA.

[3] Sun Yat-sen University, Guangzhou, China.

* Correspondence to: panle.barwick@cornell.edu

† These authors contributed equally to this work.

Abstract: Based on mobile phone records for 71 million users and location tracking information for one million users over two years, this study examines labor market impacts of the COVID-19 pandemic in China's Guangdong province. Using a standard difference-in-differences framework, our analysis finds that the pandemic increased unemployment by 27-62% from late January 2020 to the end of September 2020, much higher than the official statistics. Females, workers older than 40, and migrants are more affected. Cities that rely more on export or have a higher share of GDP in the hospitality industry but a lower share in the finance and healthcare industries experienced a more pronounced increase in unemployment. Five months after the full reopening, employment has not recovered to the pre-pandemic level. This lingering impact reflects the global nature of the pandemic and the interconnectedness of the world economy.

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. In addition to the substantial time-lag and limited availability for small geographic areas, statistics inferred from surveys suffer from considerable uncertainty and are routinely

revised. Their quality also varies considerably over time due to changes in participation rates, modifications in the survey methodology, inconsistencies and measurement errors in sample responses, or rotation group bias (*1-7*).

In China, information on unemployment is derived from the number of individuals who registered with the unemployment benefit agencies prior to 2018 and supplemented by household surveys afterward (*8*). Measuring unemployment accurately is particularly challenging due to a large fraction of the population who do not have local household registrations (*Hukou*) and hence excluded from the unemployment surveys. Besides, reporting and aggregation errors, as well as potential data manipulations, have also been documented (*9-12*). China's national unemployment rate varies between a tight range of 3.1%-4.3% over the past two decades, and the official unemployment rate inched up from 5.3% in January to 6% in June 2020, leading to questions about its reliability (*12, 13*) especially in the face of rapid and unprecedented social and economic changes brought about by the pandemic.

This study leverages high-frequency and high-resolution mobile phone usage data in Guangdong, the most populous province in China, with a GDP larger than all but the top 12 countries in the world. Our primary data source consists of mobile phone records for 71 million users and location tracking information for one million randomly-selected 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 leverage two unique data features to bound the impact on unemployment: a) the number of individuals who contact the designated unemployment benefits agencies via phone, and b) the number of individuals working on-site.

Several key findings emerge from our analysis. First, the pandemic has increased unemployment by 27-62% as of September 30, 2020, based on two strategies to measure the unemployment impact described in detail below. The sharp rise in unemployment is much higher than the government statistics that report an increase of 13.3% in the unemployment rate (from 2.26 percentage points in January-March to 2.56 percentage points in July-September) in Guangdong during the same period. In comparison, the national unemployment rate

released by the National Bureau of Statistics (henceforth NBS) that is based on the monthly survey of urban unemployment increased from 5.3% in January 2020 to a peak of 6.2% in February and came down to 5.4% in September 2020, close to the pre-pandemic level.

Second, nearly five months after Guangdong's full reopening, the number of individuals who reach out to unemployment benefits agencies to claim unemployment benefits continues to rise, and employment has not returned to the pre-pandemic level. The shock of the pandemic on unemployment is highly uneven across demographic groups and more pronounced among females, 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 NBS' reported migrant workers and indicates the possibility of a large-scale layoff among this group.

Third, the pandemic's impact is more extensive in cities with a high labor share of hospitality, real estate, or transportation industries but smaller in cities where employments are concentrated in finance, health care, or education industries. In addition, the impact is more pronounced in cities that rely heavily on export, reflecting the global nature of the shock in an interconnected world economy. Industry compositions account for 36% of the heterogeneity in the pandemic's unemployment impact across cities, while trade exposure contributes to 30% of the heterogeneity.

Lastly, our results 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 countries across the world through the supply chain and the trade channels. The industry composition of a city (or a country), exposure to trade, and nature of the supply chain are crucial factors in determining the effect that the pandemic has on its economy. Our measures help us understand the pandemic's labor market impact at a granular level and inform targeted policies to help the most severely affected groups and regions.

Results

Data sources and unemployment measures

Researchers in recent studies have used mobile phone data to improve labor market measurements (*14, 15*), track human movement in real-time and at a fine spatial scale, and quantify the pandemic's impact on voting behaviors, mobility, and social contacts (*16-25*). Exploiting the increasingly available high-frequency and high-resolution mobile phone data is particularly advantageous for China, as its cellphone penetration rate is high among developing countries. According to the 2018 China Family Panel Studies, a nationally representative longitudinal survey of individuals' social and economic status, 89% of correspondents sixteen years and older reported possessing a cellphone. In addition, each household owns 2.5 cellphones on average (National Bureau of Statistics 2018). Fig. S1 in the Supplementary Material shows a strong correlation between the number of China Mobile users and the number of residents by cities. Cities with a higher GDP per capita (represented by the size of the dots in Fig. S1) 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 to 11% of China's GDP and around a quarter of China's foreign trade (China Statistical Yearbook 2020). Its major cities include Shenzhen and Guangzhou, among the wealthiest and economically most advanced cities in China. Its economy is widely believed to be the most dynamic and resilient among all provinces in China (*26, 27*). Another reason that makes Guangdong particularly relevant is that Guangdong suffered severely from the 2003 SARS epidemic. In the early stage of the COVID-19, the provincial government acted swiftly. It adopted vigilant procedures, including locking down cities, quarantining affected individuals, and releasing the number and location of confirmed cases. These procedures proved successful and have kept the number of daily confirmed cases under a few handfuls since the full reopening (Fig. S2 in the Supplementary Material). While helping to reduce the death toll from the pandemic in Guangdong, these measures could have had significant consequences on the labor market and the economy in general.

Our data come from China Mobile, the dominant cellular service provider in China, with 58% of national mobile service subscribers. We have access to detailed phone usage (encrypted IDs of the calling party and the receiving party, date of calls, and call duration in seconds) for all of its 71 million users in Guangdong

Province from January 2018 to September 2020, accounting for 63% of all mobile users in the province. We observe some user demographic information, such as age, gender, and the place where the phone number is registered.

We leverage two features of the mobile phone data to construct unemployment measures. The first feature is detailed call records (time and duration of each call) to the designated government hotline (12333) for unemployment benefits. The hotline offers comprehensive one-stop social insurance public 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 the all-inclusive help from customer services. According to a 2016 study on China's special economic zones, 96% of individuals use 12333 to claim unemployment office, and only 4% visit local social security bureaus or file online (*28*).

Despite the popularity of the hotline, the number of individuals making calls to 12333 provides a *lower bound* estimate for the *level* of unemployment, for the same reasons that make the government official statistics conservative. Not all unemployed workers reach out to the government agency to claim unemployment benefits, especially those who are optimistic about finding a new job soon. In addition, the lifetime unemployment benefits in China are capped at 24 months, thus limiting choices for people who have exhausted benefits in the past. Therefore, instead of focusing on the level of unemployment calls, our analysis below exploits its *changes*. We show that the measure of changes in unemployment calls can provide useful information on short-run labor market dynamics otherwise unavailable through official statistics.

As people might reach out to the hotline multiple times to claim their unemployment benefits, we use the number of individuals calling 12333, instead of the raw number of calls to 12333, to construct our first unemployment measure. For brevity, the term "number of individuals calling 12333" and "number of unemployment calls" are used interchangeably throughout the analysis. Fig. S3 in the Supplementary Material plots the number of unemployment calls across cities in 2019. The correlation between city-level

unemployment calls and the official unemployment rate released by the NBS, which is only available annually for city-level statistics, is reasonably high at 0.7 in 2019.

The second feature of the mobile phone data is the location tracking information (in longitude and latitude) collected by mobile devices during 5-minute intervals except when they are powered off. Researchers have used such mobile positioning data to study urban and transportation issues (*16, 17*), though few studies exploited long panels of location data to examine labor market dynamics (*15*). 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 similarly constructed, except that we use the location with the longest duration between 10 pm and 7 am each month. These geocoded locations trace out individuals' spatial trajectories over time and allow us to record the time of arrival and departure at job locations.

We use reductions in the number of people working on-site before and after the lockdown and relative to 2019 as an upper bound estimate on pandemic-induced unemployment. These changes are upper bounds because some individuals may switch to work remotely after the full reopening. However, we provide evidence below that the fraction of individuals working remotely is modest post the lockdown. The Supplemental Material contains more details on our sample size and how we construct various measures for the labor market outcome.

Data patterns

The lockdown in Guangdong lasted 32 days from January 23 to February 24, 2020. On February 24, 2020, Guangdong province entered Phase I reopening, which lasted 76 days. During Phase I reopening, people were allowed (and in certain industries, people were encouraged) to go back to work and visit outdoor public places. Phase II reopening, or full reopening, officially started on May 9, 2020. All businesses, including shopping malls, supermarkets, and restaurants, were allowed to reopen fully. The only exception was movie theaters that remained closed till mid-July. We delineate the interval from the 60 days before the lockdown

to the 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).

Our analyses use a standard DID framework and exploit differences in our two key measures of labor market outcomes between 2020 (the treatment group) and 2019 (the control group). As Guangdong's lockdown occurred on January 23, two days before the Chinese New Year on January 25, the most important holiday in China, it is critical to control this holiday effect directly. Hence, we use the lunar calendar instead of the standard almanac calendar to define the event date. Specifically, the event date for 2020 is January 23, 2020, while the event date for 2019 is February 5, 2019, two days before the Chinese New Year in 2019. In other words, our analyses compare changes in labor market outcomes before and after the event date in 2020 vs. changes in labor market measures before and after the event date in 2019.

Fig. 1 depicts the differences in the daily number of unemployment calls between 2019 and 2020. Red bars indicate more calls in 2020 than in 2019, and blue bars indicate fewer calls in 2020 than in 2019. The daily number of unemployment calls in 2020 was similar to the 2019 baseline before the lockdown, which supports the parallel trend assumption underlying the DID strategy. However, it dropped significantly during the lockdown. The unexpected drop may be attributed to the uncertainty about the severity and duration of the epidemic during the pandemic's initial stage. As the severity of the pandemic unfolded in China, the number of unemployment calls increased sharply one month after the beginning of Phase I reopening, and the trend continued to the end of our data period. Some of the time-series variations in the bar chart reflect differences between weekdays and weekends and during holidays. For example, few people reached out to the unemployment hotline right before the Phase II reopening, which coincides with the International Labor Day, a 5-day public holiday from May 1 to May 5 in 2020 (event day 100 to event day 104). The dotted black line shows the cumulative differences in unemployment calls between 2019 and 2020. The cumulative differences during the first 60 days prior to the event hover around zero, again supporting the view that the labor market prospects are comparable across the two years before the pandemic.

In terms of magnitude, the average number of daily unemployment calls in Guangdong province is 11,828, with an interquartile range of 11,371. There are a total of 3.31 million unemployment calls from the lockdown to September 30, 2020, while the number of calls during the same period in 2019 is 2.58 million. Taken together, these numbers suggest that the pandemic has led to a cumulative 28% increase in unemployment in our sample period.

Our second set of labor market measures includes the number of people commuting to work and the amount of time spent at the workplace conditional on working on-site among one million randomly chosen mobile users. The average number of daily commuters is 0.48 million individuals in 2019, which is consistent with the working population ratio in Guangdong (Guangdong Provincial Statistical Yearbook 2020). Fig. 2 presents the daily differences between 2019 and 2020. Similar to Fig. 1, the commuting patterns between 2019 and 2020 displayed few differences before the lockdown, both in terms of daily commuters and hours worked on-site. The average number of daily commuters is 484,750 (with an interquartile range of 124,690) and the average hours worked are 8.8 (with an interquartile range of 0.6) pre the pandemic. During the lockdown in 2020, both the number of commuters and hours worked on-site registered a sharp reduction relative to those during the same lunar calendar days in 2019, as expected. The drops reached their trough toward the end of the lockdown, with the number of commuters decreasing by 0.19 million (about 40%) and the number of hours worked on-site falling by two hours per day. Both series had gradually improved since Phase I reopening. By September 2020, 144 days after Phase II full reopening, the number of hours working on-site had returned to the 2019 level. However, the number of commuters has not fully recovered to the pre-pandemic level. The gap between 2020 and 2019 persists till the end of our sample, with 6.7% fewer people commuting to work in September 2020 relative to September 2019.

Regression Analysis

Fig. 1 and Fig. 2 provide compelling evidence that the pandemic significantly hurt the labor market in 2020. We turn to a regression framework to control for confounding factors and quantify the pandemic's effect on unemployment. All regressions include days-to-event (i.e., the number of days from the event), day-of-week, holiday, treatment group, and neighborhood fixed effects (see the Materials and Methods section below for

more details). Similar to the discussion on data patterns above, our regression analysis also consists of two parts. The first part exploits variation in the number of individuals making unemployment calls (or "unemployment calls" for brevity), while the second part leverages changes in commuting patterns.

To illustrate the dynamic effect of the pandemic, we present the regression coefficient estimates for all ten-day intervals over the entire event window (β_q in Eq. (1)). Panel (a) of Fig. 3 depicts the percentage change in unemployment calls in 2020 relative to those in 2019. The pre-event coefficient estimates center around zero and are statistically insignificant, supporting our identification assumption of the parallel trend between 2019 and 2020. Consistent with the descriptive evidence in Fig. 1, unemployment calls dropped during the lockdown but increased sharply one-month after the Phase I reopening (panel (a)). The trend continued till the end of September, nearly five months into the Phase II reopening. The increase in the number of unemployment calls ranges from 20-50% from mid-March to September.

Panel A of Table 1 presents the coefficient estimates of β_q in Eq. (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. Column (1) shows the pandemic's impact on the number of unemployment calls, while column (2) examines the length of call duration. Echoing results in Fig. 3, the number of calls decreased by 37% during the lockdown and increased by 25% during Phase I reopening and 45% during Phase II reopening. The call duration displayed a similar pattern: the average call time dropped during the lockdown but increased after the reopening. This increase in call duration could be driven by two factors: a) migrant workers generally need to provide more information to claim unemployment benefits, b) more migrants applied for unemployment benefits after the reopening. Around 25% of the workforce in Guangdong are skilled and unskilled workers who migrated from other provinces (Human Resources and Social Security Department of Guangdong province, 2018). During the Phase I reopening, migrants returned to Guangdong after being away during the Chinese New Year national holiday. With a dim job prospect, some migrants were unable to find a job and began to register for unemployment benefits. In our sample period, the average duration per call for migrant and non-migrant workers are about 6 and 4 minutes, respectively. In addition,

as we show below, the number of migrants reaching out to the unemployment hotline skyrocketed after the reopening.

Panel B of Table 1 further groups the three periods during and post the lockdown into one group. The coefficient estimates directly measure the pandemic's average labor market effect. Overall, the pandemic has increased the number of unemployment calls by 27%. This is very similar to the magnitude discussed above when we compare the raw cumulative number of calls between 2020 and 2019, suggesting that the role of confounders (as captured by our rich set of fixed effects and controls) is limited. 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, the percentage *change* in unemployment calls estimated in Table 1 (27%) is likely a reliable measure of the percentage *change* in the unemployment rate as a result of the pandemic.

To examine the pandemic's differential impacts across demographic groups, we repeat the event study analysis by gender, age, and migrant status. Coefficients for these event studies are plotted in panels (b)-(d) of Fig. 3. Specifically, we redefine the dependent variable in Eq. (1) as the difference in the (logarithm) outcome variable between female and male (panel (b)), between individuals 40 years old and above and those under 40 (panel (c)), and between migrants and non-migrants (panel (d)). Females are more affected by the pandemic. The number of unemployment calls by females is 10-15 percentage points higher than that by males from April to September. Note that this is not driven by females claiming unemployment benefits on behalf of their family members, which is prohibited in China. As many childcare facilities were closed during (part of) the pandemic, the increasing demand for home childcare might have played a role in the more pronounced unemployment impact among females.

Older workers fared worse than younger cohorts: the number of unemployment calls by workers 40 and above is around 20 percentage points higher than those under 40. Lastly, migrants are most severely affected by the pandemic. The unemployment calls by migrants skyrocketed and increased by 120-220 percentage

points more than calls by non-migrants between April and September, with no signs of abating at the end of our sample period. According to NBS, there were 291 million migrant workers in 2019, constituting 36% of the total workforce nationwide. By September 2020, the number dwindled to 179 million. The escalating number of migrants reaching out to the unemployment benefit office in our sample is consistent with the massive reduction in NBS' reported migrant workers and suggests the possibility of a large-scale layoff among migrant workers. Our results corroborate the existing literature documenting that the least advantaged social groups, including migrants, are most vulnerable to adverse shocks and risks (*29*). Finally, the disproportionately more harsh impacts on females, older workers, and migrants likely reflect the pandemic's heterogeneous shocks across industries. 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 (Table S1. In Supplemental Material). Panel (a) of Fig. 4 examines the impact 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 Eq. (2). Heterogeneity across cities is sizeable: while some cities saw unemployment calls increasing by nearly 70%, a few cities were unscathed or even experienced reductions in unemployment calls. 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 Eq. (2), where we interact the pandemic treatment variable with city-level labor shares by industry for all thirteen major industries. Panel (b) of Fig. 4 reports coefficients for all industries. The hotel and catering, real estate, and transportation sectors experienced the largest increase in unemployment calls, *ceteris paribus*. In comparison, the finance, health care, and education sectors witnessed *reductions* in unemployment calls after the lockdown in January, consistent with findings using data from other countries (*30, 31*). To evaluate the importance of industry compositions, we predict the number of unemployment calls

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 36.3% of changes in the unemployment rate.

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 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 Fig. 4, the pandemic's impact on Shantou's unemployment calls is not statistically significant from zero, while Dongguan is among the worst-hit cities, with its unemployment calls increasing 56% since the onset of the pandemic. In Table S2 in the Supplemental Material, 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 one percentage point increase in 2019's export-to-GDP ratio is associated with a 0.19% increase in the number of unemployment calls for a given city. Similar to industry compositions, variation in the export-to-GDP ratio is also critical and explains 30% of the heterogeneity in the pandemic's unemployment impact across cities.

The sizeable estimates in Table 1 (a 27% increase in the unemployment rate) and the significant heterogeneity across cities and industries as highlighted in Fig. 4 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 countries across the world through the supply chain and the trade channels, where the industry composition of a city (or a country), nature of the supply chain, and exposure to trade are crucial factors in determining the effect that the pandemic has on its economy (*32-36*).

We now turn to the second set of our analysis that exploits variation in commuting patterns, using the location tracking data for one million randomly selected users. As some people might switch to working remotely during and after the lockdown, we treat an individual as commuting to work for a given week if he

visits his work location at least once that week in the following regression analysis. We experimented with different criteria of commuting. As shown in the Supplemental Material, our coefficient estimates are remarkably stable whether we use a two-week window (i.e., counting someone as commuting to work if he visits his work location at least once in the two-week window), a one-week window, or count commuters on a daily basis. As most individuals commute within a city boundary, we aggregate the number of commuters to the city by day level, with a total of 13,052 observations, though results are similar if we aggregate commuters to the neighborhood level (see Supplemental Material for more detail).

Fig. 5 uses the same DID framework following Eq. (1) as does Fig 3., except that it presents the percentage changes for the number of workers who commute to job locations and the work duration on-site. Panel (a) shows that the number of people who work on-site dropped as much as 40% during the lockdown, as businesses were temporarily shut down and non-essential workers were prevented from commuting to work.

During Phase I reopening, the number of people working on-site was 5-25% lower than the 2019 level. When all businesses were allowed to open during Phase II reopening, the number of people working on-site is still about 5-10% lower than the 2019 level. This reduction could be attributed to unemployment, or it could be driven by individuals switching from working on-site to working remotely. To gauge the possibility of telecommuting, we obtain an exhaustive list of all 21 virtual meeting Apps (including Tencent meeting, DingDing, Zoom, etc.) in the Apple store and Andriod gallery and examine the usage of these Apps among individuals during weeks when they do not commute to work. The shares of these individuals using any of the virtual meeting Apps at least once in weeks when they do not commute to work were 3.1%, 0.9%, and 0.05%, during the lockdown, Phase I, and Phase II reopening periods, respectively. The extremely low frequency of virtual meeting App usage suggests that these individuals are unlikely to be working remotely when they stop commuting to their work locations on a continuous basis, especially during Phase II reopening period.

Table 2 reports the parameter estimates for the percentage reduction in commuters following Eq. (1). The number of commuters reduced by 29%, 11%, and 5%, during the lockdown, Phase I reopening, and Phase

II reopening periods. Using the magnitude for Phase II reopening when all firms are allowed to resume operation, the five percentage point reduction amounts to 24,238 fewer commuters, or 484,750*5%, per day relative to the pre-pandemic level. In 2019, the average number of individuals who stopped commuting to their work location for at least one week and had no new job location was 39,094. The reduction of 24,238 in Phase II of 2020 implies a 62% increase in the number of individuals who cease commuting relative to the 2019 level. Granted that some of these individuals might be working remotely, and that some firms might take longer to resume operation post the lockdown, we treat the 62% increase as an upper bound for the pandemic's impact on unemployment in Guangdong province.

Measuring unemployment at any given point in time is challenging, partly because it depends on each worker's unemployment duration. In our commuter analysis above, we use a one-week window and distinguish between individuals who visit their work locations at least once a week (commuters) and those with a valid work location but not visit their work location at all in that week. For robustness, we repeat this entire analysis using a two-week window. Specifically, the average number of individuals who stop commuting to work for at least two weeks in 2019 is 37,153. At the same time, the pandemic has led to 23,268 (which is 484,750*4.8%) more individuals who stopped commuting in 2020 using the two-week threshold. Together, these numbers imply a 63% increase in the unemployment rate.

Note that the number of non-commuters (which is a proxy for unemployment) is similar whether we use the one-week window or the two-week window. This is consistent with the evidence above that telecommuting on a full-time basis is limited after the full reopening.

Panel (b) of Fig. 5 presents changes in the work duration for individuals working on-site. Hours on-site were reduced by 30% during the lockdown period but had returned to the 2019 level in phase II reopening. These results suggest that while the extensive margin of the number of commuters is still being impacted by the pandemic, the intensive margin had returned to the pre-pandemic level. The pandemic does not seem to have brought about dramatic changes in the nature of working on-site, lending further support to our strategy of measuring unemployment based on changes in individuals commuting to work.

Fig. 6 presents the heterogeneous impacts across demographic groups. The results are qualitatively the same as those in Fig. 3 that is based on unemployment calls: females, workers over 40, and migrant workers experienced a more pronounced reduction in commuting since the pandemic.

Discussion

Our analysis contributes to the recent studies on the broad labor market impacts of the pandemic. To address the significant delay in official labor market indicators compiled by the Bureau of Labor Statistics, recent studies on the United States have used (near) real-time surveys to understand the labor market impacts. They have shown that the COVID-19 pandemic has generated unprecedented impacts on unemployment in terms of its scale and speed in modern history (*37, 38*). (*39*) argues that the BLS unemployment rate, which was 4.5% in March, peaked at 14% in April, and came down at 11% in June, suffers from a downward bias. (*37*) documents an unemployment rate of over 20% in May based on an online labor-market survey. (*40*) use weekly administrative payroll data and find the aggregate employment fell by 21 percent during the first four months of 2020.

Compared to these studies, our mobile phone data benefit from a large sample size and fine resolutions in both temporal and spatial dimensions. The pandemic's estimated impact on unemployment (27-62% increase) in China is much smaller than that in the United States, 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 (*41-43*) and likely mitigated the impact on the economy as a result.

Our findings on the uneven labor market impacts across demographic groups and industries are consistent with recent studies in other countries, though we conduct the first quantitative analysis on China using large-scale datasets. Based on a real-time survey in the U.S., U.K., and Germany, (*30*) find that the share of tasks

that can be done at home is an important determinant of job loss during the pandemic and that less-educated workers and women are more affected. (*31*) argues that different from normal economic downturns, the pandemic could lead to more job losses for women than men due to the more severe impact on the hospitality industry with high female employment shares and the increased need for childcare from school closures.

Our analysis further 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 trade channels. The evidence from Guangdong shows that the detrimental labor market impact lingers on even five months after the full reopening of the economy. The global nature of the pandemic and interconnectedness of the world economy could weaken the prospect of recovery as the economic impact in one country propagates to other parts of the world.

Materials and Methods

Econometric models

Our analysis employs the standard difference-in-differences (DID) approach by comparing changes in 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 for 2020 is January 23, 2020, while the event date for 2019 is February 5, 2019, two days before the Chinese New Year in the lunar calendar.

We use the year 2020 as the treatment group and the year 2019 as the control group. To control for potential differences in time-varying unobservables, we include a rich set of fixed effects such as day-of-week, days-to-event, holiday, and the treatment group fixed effects. The identification assumption is that after including these controls, there are no systemic differences in time-varying unobservables between the two groups in the absence of the pandemic. Results from event studies above support this common trend assumption 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:

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

where c denotes a neighborhood (a cell-tower area in the analysis on unemployment calls or a city in the analysis on commuting patterns), i denotes the treatment group (the year 2020) or the control group (the year 2019), and t denotes the day from the event/treatment (January 23 in 2020 and February 5 in 2019). The event window is sixty days before the event and 252 days post the event.

y_{cit} is the outcome variable such as the log number of unemployment calls or commuters. We report results based on log(outcome+1) to avoid taking logarithm over zero, though 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 group and the control group. Variable d_i is a dummy that is equal to one for the treatment group. $\mathbf{1}(.)$ is an indicator variable for each 10-day interval of the sample. We control for neighborhood fixed effects α_c, group fixed effects γ_i and the 312 days-to-event fixed effects η_t. We also control for the 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) as well as day-of-week fixed effects. Standard errors are clustered at the days-to-event level.

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

$$y_{cit} = d_i \cdot \mathbf{1}(t \in [0, 252]) \cdot \boldsymbol{Z}'\boldsymbol{\tau} + d_i \cdot \boldsymbol{Z}'\boldsymbol{\mu} + \mathbf{1}(t \in [0, 252]) \cdot \boldsymbol{Z}'\boldsymbol{\rho} +$$

$$\beta \cdot d_i \cdot \mathbf{1}(t \in [0, 252]) + \alpha_c + \gamma_i + \eta_t + \xi_{it} + \varepsilon_{cit}, \qquad (2)$$

Where $\boldsymbol{Z}$ is a vector of city attributes in 2019 and $\boldsymbol{\tau}, \boldsymbol{\mu}, \boldsymbol{\rho}$ are corresponding coefficients. 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. Variables d_i and $\mathbf{1}(.)$ and the fixed effects $\alpha_c, \gamma_i, \eta_t, \xi_{it}$ are

the same as in Eq. (1). The key coefficient is $\boldsymbol{\tau}$, which measures the heterogeneous impact by city characteristics $\boldsymbol{Z}$ based on their values in 2019. Unlike in Eq. (1) where we estimate the pandemic's impact for each ten-day interval, here we estimate the average effect $\boldsymbol{\tau}$ across the three periods (the lockdown, Phase I, and Phase II) and focus on heterogeneity across industries and cities.

Sample construction

Our sample consists of a) call records (encrypted IDs of the calling party and the receiving party, date of calls, and call duration in seconds) for all of China Mobile's 71 million users in Guangdong Province from January 2018 to September 2020, and b) location records every five-minute interval for one million randomly selected users for the same period.

In our analysis using unemployment calls, we use the number of individuals making calls to the 12333 hotline, instead of the number of calls to 12333, to address the issue that people might reach out to the hotline multiple times to claim their unemployment benefits. In other words, we only count the first time when a user reaches out to the unemployment benefit hotline. We aggregate the duration of all subsequent calls when we measure the duration of calls to the hotline.

Our main analysis excludes users under the age of 18, as individuals under the age of 18 are subject to the Law on the Protection of Minors and are unlikely to be working. Results excluding users under the age of 25 (to eliminate those still in school) are almost identical (see Supplementary Material for more detail).

China Mobile delineates Guangdong province into 787 mobile phone cell-tower areas (similar to zip codes in the U.S.) for billing purposes. We aggregate the number of unemployment calls by cell-tower-area and day, with a total of 489,514 observations. All of the regressions using unemployment calls control for cell-tower-area (or neighborhood) fixed effects.

Some of our analyses examine heterogeneous labor market prospects between migrants and non-migrants. It is important to note that migrants without Guangdong *Hukou* who had been working in Guangdong became

eligible for unemployment benefits since 2014 (*44*). This was largely designed by the Guangdong government to attract migrants and to help improve labor relations. As we do not observe whether an individual has local *Hukou* status – the official definition of migrants, we use the city where the mobile phone number is registered as a proxy. Specifically, we define migrants as individuals who registered their phone numbers outside Guangdong province. This is an imperfect measure of local *Hukou* status, as workers from outside Guangdong can buy and register their mobile phones in Guangdong and will be treated as non-migrants in our analysis. Consequently, the actual gap in unemployment between migrants and residents might be even larger than our estimates.

To construct work locations for the one-million users, we first identify the location where the user spent the most time between 9 am and 6 pm during weekdays for each month. If an individual spends at least 5 hours a day between 9 am and 6 pm for at least fifteen workdays at his primary location for a given month, we code this location as this individual's work location for that month. Among the one million mobile users, the average daily number of commuters is 0.48 million, which is consistent with the working population ratio in Guangdong (Guangdong Provincial Statistical Yearbook 2020). Home locations are constructed similarly, except that we use the location with the longest duration between 10 pm and 7 am each month. 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. All users have home 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 commuting distance decays exponentially (Fig. S4 in the Supplemental Material), which is intuitive. Additionally, the average commuting distance in our sample period is around 6.6km, close to the average commuting distance of 6.9km reported in the 2017 travel survey by the Guangzhou Municipal Transportation Bureau (GMTB) (*45*). Second, for the city of Guangzhou (the provincial capital city), Fig. S5 in the Supplementary Material 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 center of the city gains population, and the suburbs lose population during the daytime relative to the nighttime. But these differences are much more pronounced on weekdays than weekends, especially for the center of the city. For instance, the enlarged area is a famous industrial park in Guangzhou, which clearly shows that the daytime population is much larger than the nighttime population. The differences are amplified on weekdays. The spatial pattern of population density is consistent with those from the report by GMTB (*45*).

Finally, for regressions that exploit commuting patterns, we aggregate the number of commuters to the city by day level, with a total of 13,052 observations.

References

1. J. M. Poterba, L. H. Summers, Reporting errors and labor market dynamics. *Econometrica*. **54**, 1319–1338 (1986).

2. S. R. Jones, & W. C. Riddell, The measurement of unemployment: An empirical approach. *Econometrica*. **67**, 147-161 (1999).

3. D. Card, Origins of the unemployment rate: The lasting legacy of measurement without theory. *American Economic Review*. **101**, 552–57 (2011).

4. S. Feng, Y. Hu, Misclassification errors and the underestimation of the us unemployment rate. *American Economic Review*. **103**, 1054–70 (2013).

5. B. D. Meyer, W. K. Mok, & J. X. Sullivan, Household surveys in crisis. *Journal of Economic Perspectives*. **29**, 199-226 (2015).

6. A. B. Krueger, A. Mas, & X. Niu, The evolution of rotation group bias: Will the real unemployment rate please stand up? *Review of Economics and Statistics*. **99**, 258-264 (2017).

7. O. Heffetz, & D. B. Reeves, Difficulty of reaching respondents and nonresponse Bias: Evidence from large government surveys. *Review of Economics and Statistics*. **101**, 176-191 (2019).

8. *China unveils urban survey unemployment rate*. (MarketWatch, https://www.marketwatch.com/story/china-unveils-urban-survey-unemployment-rate-2018-04-17-1485030).

9. J. Giles, A. Park, and J. Zhang, What is China's true unemployment rate? *China Economic Review*. **16**, 149–170 (2005).

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

11. F. Cai, Y. Du, & M. Wang, Demystify the labor statistics in China. *China Economic Journal*. **6**, 123-133 (2013).

12. S. Feng, Y. Hu, R. Moffitt, Long run trends in unemployment and labor force participation in urban China. *Journal of Comparative Economics*. **45**, 304–324 (2017).

13. *Some Economists Question the Strength of China's Labor Market*. (Wall Street Journal, https://www.wsj.com/articles/chinas-jobs-rebound-doesnt-appear-as-robust-as-the-government-claims-11591551390

14. J. L. Toole, Y. R. Lin, E. Muehlegger, D. Shoag, M. C. González, & D. Lazer, Tracking employment shocks using mobile phone data. *Journal of The Royal Society Interface*. **12**, 20150185 (2015).

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

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

17. R. Ahas, A. Aasa, S. Silm, & M. Tiru, Daily rhythms of suburban commuters' movements in the Tallinn metropolitan area: Case study with mobile positioning data. *Transportation Research Part C: Emerging Technologies*. **18**, 45-54 (2010).

18. T. Couronné, Z. Smoreda, & A. M. Olteanu, Chatty mobiles: Individual mobility and communication patterns. *arXiv preprint arXiv*. **1301**, 6553 (2013).

19. B. Y. Chen, Y. Wang, D. Wang, Q. Li, W. H. Lam, & S. L. Shaw, Understanding the impacts of human mobility on accessibility using massive mobile phone tracking data. *Annals of the American Association of Geographers*. **108**, 1115-1133 (2018).

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

21. M. K. Chen, K. Haggag, D. G. Pope, R. Rohla, Racial disparities in voting wait times: Evidence from smartphone data. National Bureau of Economic Research, Working Paper. 26487 (2019).

22. D. Atkin, K. Chen, A. Popov, The returns to serendipity: Knowledge spillovers in Silicon Valley. Working Paper. (2020).

23. M. K. Chen, D. G. Pope, Geographic mobility in America: Evidence from cell phone data. National Bureau of Economic Research, Working Paper. 27072 (2020).

24. V. Couture, J. I. Dingel, A.E. Green, J. Handbury, K. R. Williams, Measuring movement and social contact with smartphone data: A real-time application to COVID-19. National Bureau of Economic Research, Working Paper. 27560 (2020).

25. S. Gupta, T. D. Nguyen, F. L. Rojas, S. Raman, B. Lee, A. Bento, K. I. Simon, and C. Wing, Tracking public and private responses to the COVID-19 epidemic: Evidence from state and local government actions. National Bureau of Economic Research, Working Paper. 27027 (2020).

26. X. Yu, & C. Zhang, Reducing inequality for shared growth in China: strategy and policy options for Guangdong province. *Technical Report by the World Bank.* 58846, 1-285 (2011).

27. H. Gong, R. Hassink, J. Tan, & D. Huang, Regional Resilience in Times of a Pandemic Crisis: The Case of COVID-19 in China. *Tijdschrift voor economische en sociale geografie*. **11**, 497-512 (2020).

28. T. Yitao, & L. Meng, *Annual report on the development of China's special economic zones. Technical Report*. 3-20 (2016).

29. E. Duflo, & A. Banerjee, *Poor economics*. (PublicAffairs, 2011)

30. A. Adams-Prassl, T. Boneva, M. Golin, C. Rauh, Inequality in the impact of the coronavirus shock: Evidence from real time surveys. *Technical Report Available at SSRN.* 3594297 (2020).

31. T. M. Alon, M. Doepke, J. Olmstead-Rumsey, M. Tertilt, The impact of COVID-19 on gender equality. National Bureau of Economic Research, Working Paper. 26947 (2020).

32. P. K. Goldberg, Global value chains, COVID-19, and the future of trade. paper presented at Princeton University, 17 April 2020.

33. *Trade shows signs of rebound from COVID-19, recovery still uncertain*. (WTO Technical Report Series No. 862, World Trade Organization, 2020).

34. *The economic impact of COVID-19 on women in vulnerable sectors and economies*. (WTO Technical Report, World Trade Organization, 3 August 2020).

35. E. Forsythe, Labor Market Flows and The Covid-19 Economy. *Technical Report Available at SSRN.* 3586074 (2020).

36. H. M. Von Gaudecker, R. Holler, L. Janys, B. Siflinger, & C. Zimpelmann, Labour supply in the early stages of the CoViD-19 Pandemic: Empirical Evidence on hours, home office, and expectations. *Technical Report Available at SSRN.* 3579251 (2020).

37. A. Bick, A. Blandin, Real time labor market estimates during the 2020 coronavirus outbreak. *Technical Report Available at SSRN.* 3692425 (2020).

38. O. Coibion, Y. Gorodnichenko, M. Weber, Labor markets during the COVID-19 crisis: A preliminary view. National Bureau of Economic Research, Working Paper. 27017 (2020).

39. H. J. Ahn, J. D. Hamilton, Measuring labor-force participation and the incidence and duration of unemployment. National Bureau of Economic Research, Working Paper. 27394 (2020).

40. T. Cajner, L. D. Crane, R. A. Decker, J. Grigsby, A. Hamins-Puertolas, E. Hurst, C. Kurz, and A. Yildirmaz, The U.S. labor market during the beginning of the pandemic recession. National Bureau of Economic Research, Working Paper. 27159 (2020).

41. S. Hsiang, D. Allen, S. Annan-Phan, K. Bell, I. Bolliger, T. Chong, H. Druckenmiller, L. Y. Huang, A. Hultgren, E. Krasovich, and P. Lau, The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature*. **584**, 262-267 (2020).

42. M. U. Kraemer, C. H. Yang, B. Gutierrez, C. H. Wu, B. Klein, D. M. Pigott, L. Du Plessis, N. R. Faria, R. Li, W. P. Hanage, and J. S. Brownstein. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*. **368**, 493–497 (2020).

43. R. Zhang, Y. Li, A. L. Zhang, Y. Wang, M. J. Molina, Identifying airborne transmission as the dominant route for the spread of COVID-19. *Proceedings of the National Academy of Sciences*. **117**, 14857–14863 (2020).

44.*Regulations on Unemployment Benefits in Guangdong Province*, (Human Resources and Social Security Department in Guangdong, https://www.gdhrss.gov.cn/sy/20140801/10101.html)

45. *Comprehensive Report of New Round Traffic Survey of Guangzhou*. (Guangzhou Transportation Planning Research Institute, 2017).

Acknowledgments

We thank Lihua Deng and Jiazhu Pan for generous help with data and Chris Chen for the literature review. **Funding**: Barwick and Li gratefully acknowledge the support by Cornell Center for Social Sciences and Cornell Atkinson Center for Sustainability. Huang gratefully acknowledges the support of the Major Project of the National Social Sciences Fund (16ZDA042). Li gratefully acknowledges the support of the Youth Program of the National Natural Science Foundation of China (72003202). **Author contributions**: All authors contributed equally to the paper. **Competing interests**: The authors declare no competing interests. **Data and materials availability**: Individual-level phone calls and location information are confidential and stored at China Mobile's data center in Guangdong Province. We are unable to share individual phone calls and location data but can provide all other auxiliary data used in the analysis.

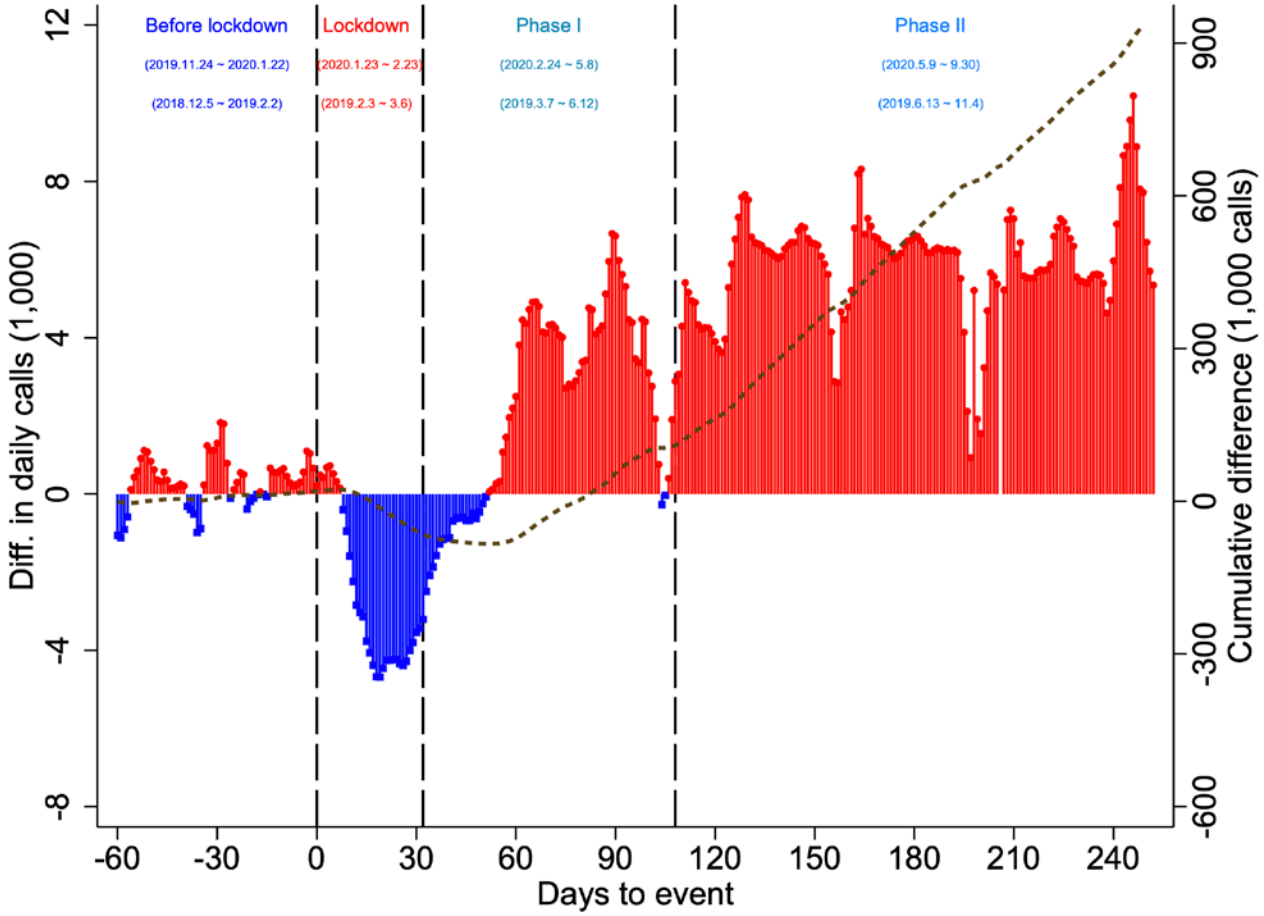
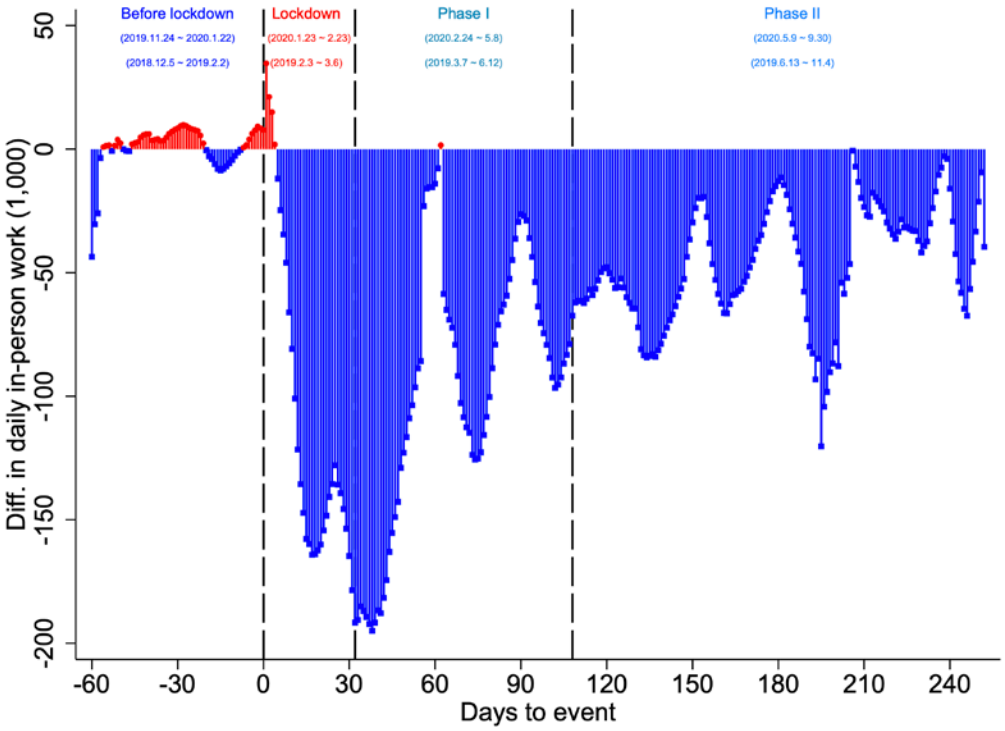
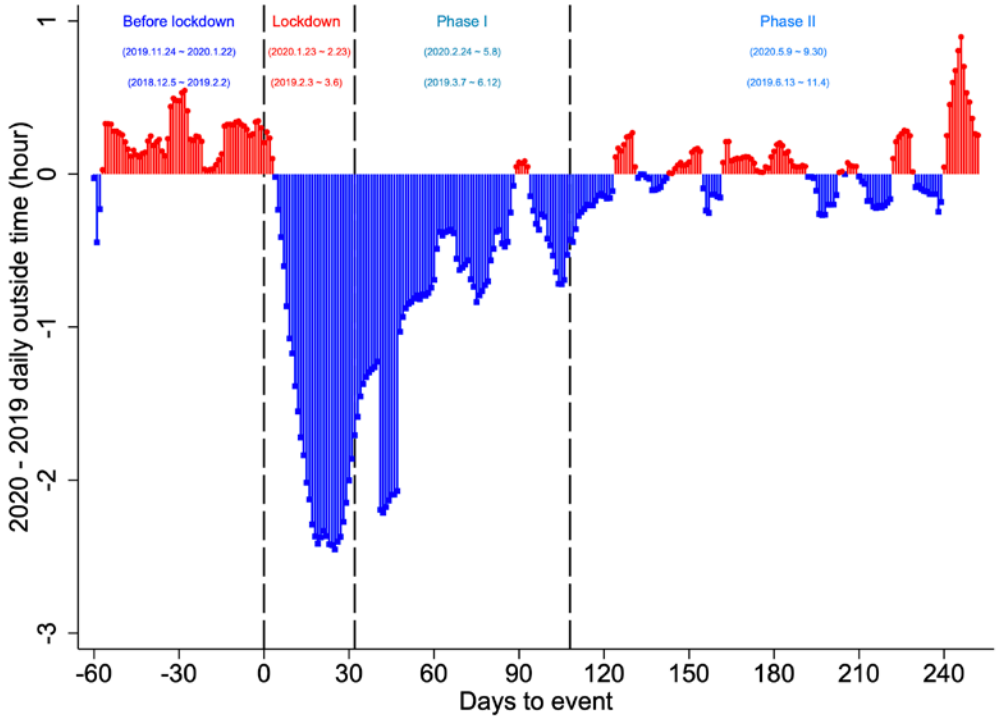


Fig. 1. Differences in the number of individuals calling 12333 between 2020 and 2019. This graph depicts the daily number of individuals calling the unemployment-benefit hotline 12333 (in thousands) in 2020 minus that in 2019. The red and blue bars (left y-axis) denote daily differences in the number of individuals calling 12333 between 2020 and 2019. The dark dash curve (the right y-axis) illustrates the cumulative difference in the number of individuals calling 12333 since 60 days before the event date. 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.



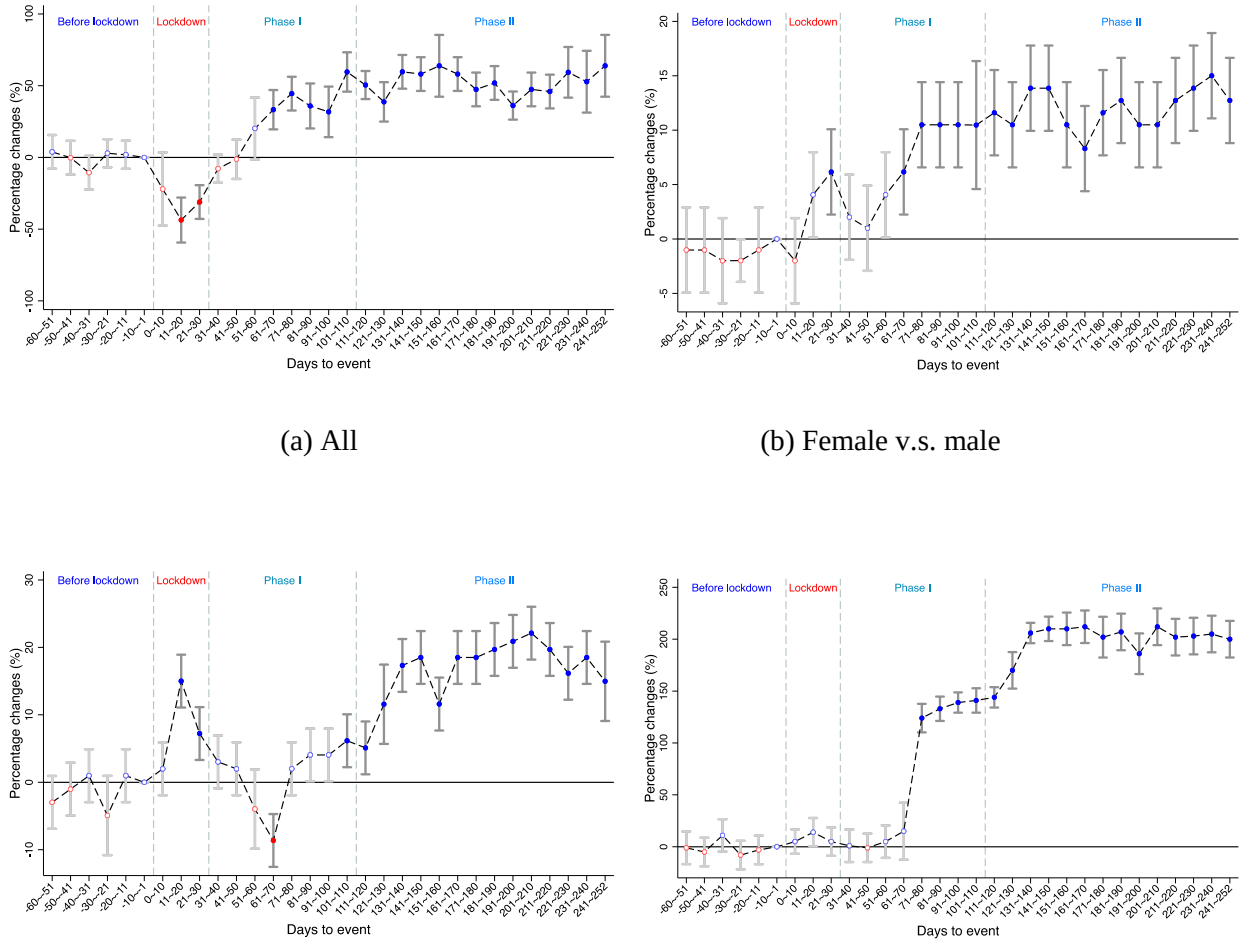
(a) Differences in daily workers on-site between 2020 and 2019



(b) Differences in daily work hours among individuals working on-site between 2020 and 2019

Fig. 2. Differences in daily workers on-site and hours on-site between 2020 and 2019. Panel (a) depicts the daily number of individuals working on-site in 2020 minus that in 2019, and panel (b) depicts the difference in daily hours working on-site, conditioning on being on-site. 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.



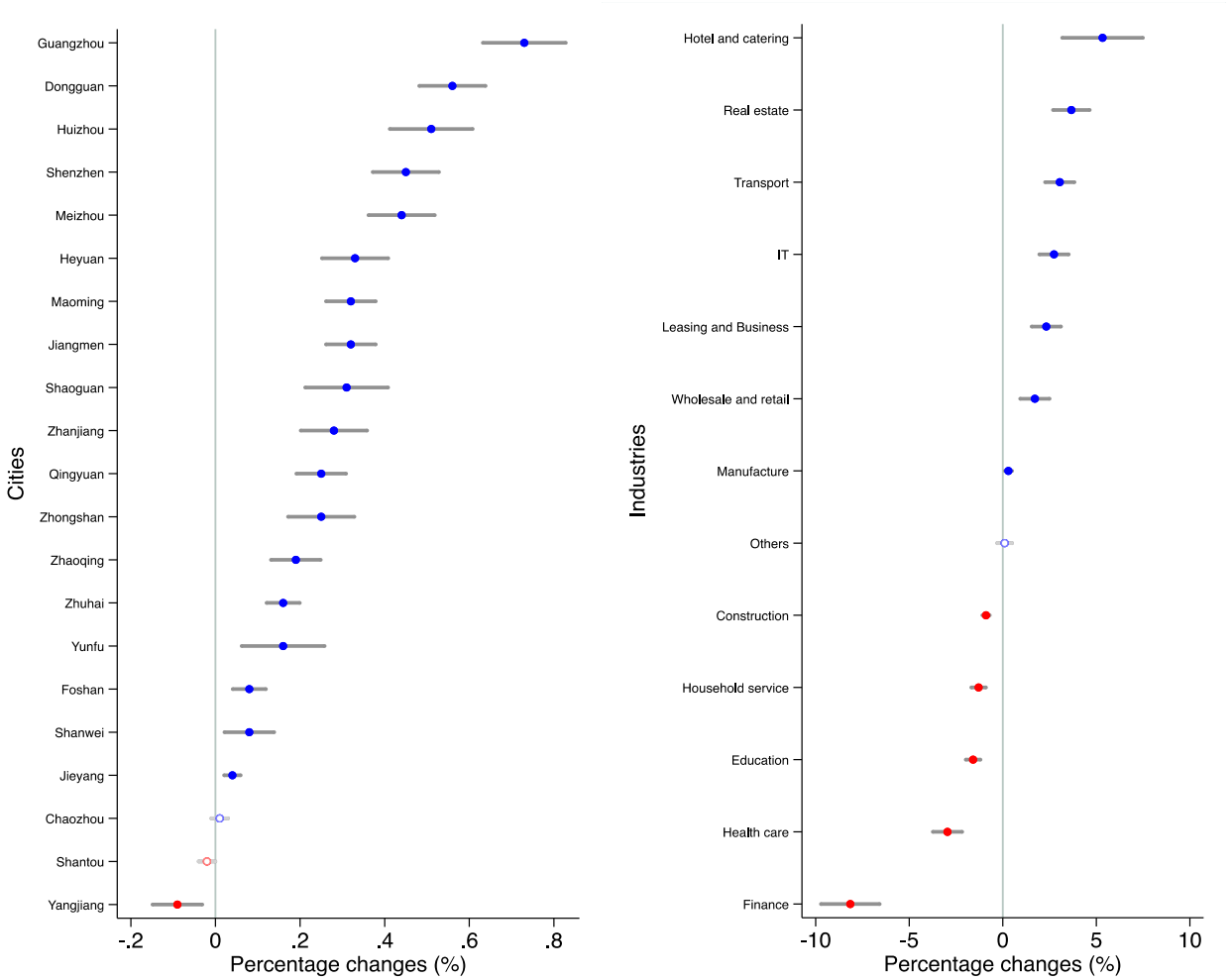
(a) All (b) Female v.s. male

(c) 40 and above v.s. under 40 years old (d) Migrants v.s. non-migrants

Fig. 3. Event study using unemployment calls. All event-study graphs plot coefficients β_q from Eq. (1), which are percentage changes in individuals making unemployment calls in 2020 relative to 2019. The dependent variable is the log number of individuals calling the unemployment-benefit hotline 12333 in panel (a), the difference between female and male calling the hotline (in logarithm) in panel (b), the difference between individuals age 40 and above and those below 40-year-old making unemployment calls (in logarithm) in panel (c), and the difference between non-migrants and migrants making unemployment calls (in logarithm) in panel (d). All regressions include neighborhood, day-of-week, days-to-event, holiday, and the treatment group fixed effects. The standard errors are clustered at the days-to-event level.

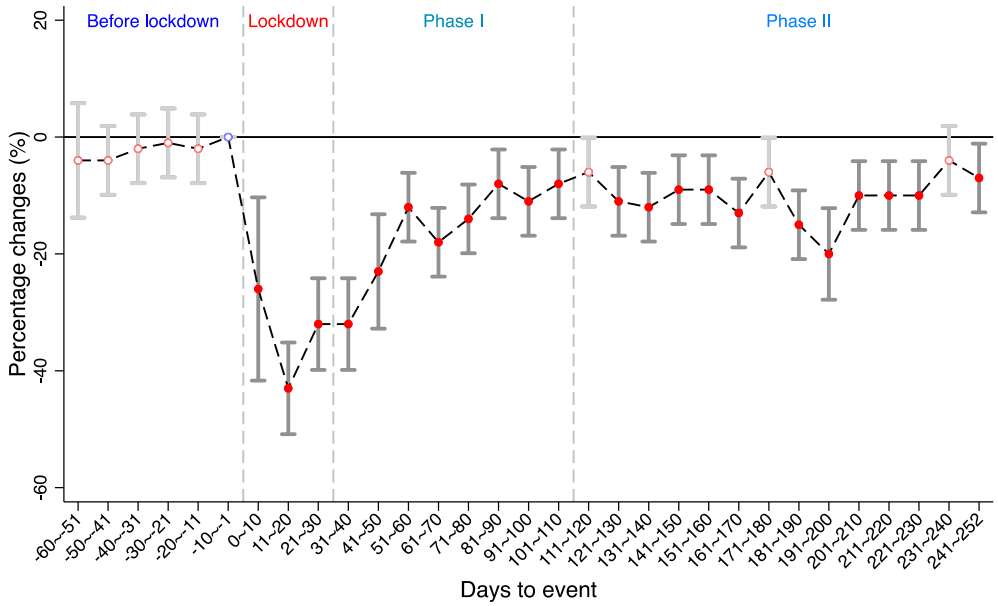
The event days are based on the lunar calendar. For 2020, the event date (or day zero) is the lockdown, January 23, 2020, two days before the Chinese New Year. Correspondingly, the event date in 2019 is February 3, 2019, also two days before the 2019 Chinese New Year. Phase I reopening started on February

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.

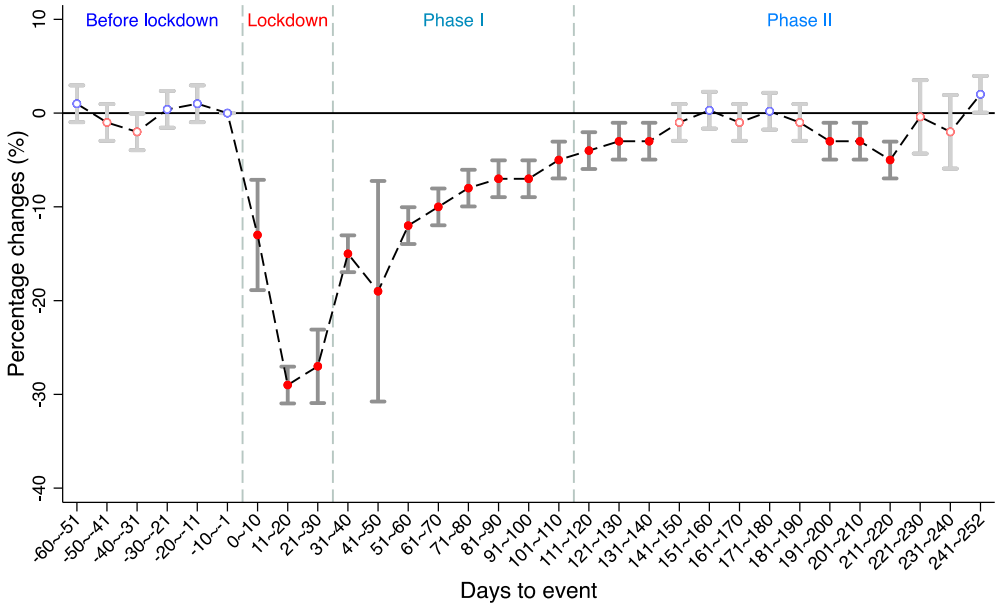


(a) By city (b) By industry

Fig. 4. Heterogeneity across industries and cities. This figure illustrates heterogeneity across cities (panel (a)) and industries (Panel (b)) following Eq. (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 unemployment relative to 2019. All regressions include neighborhood, day-of-week, days-to-event, holiday, and the treatment group fixed effects. The standard errors are clustered at the days-to-event level.



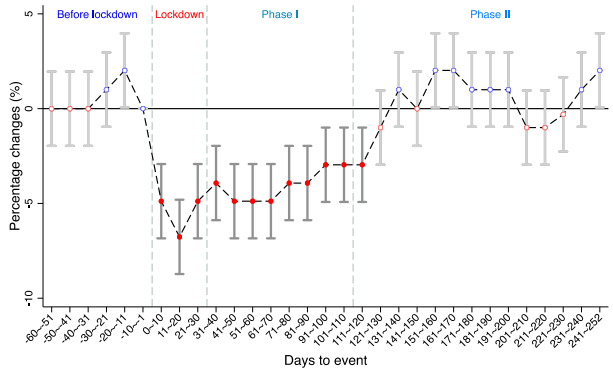
(a) No. of people commuting to work



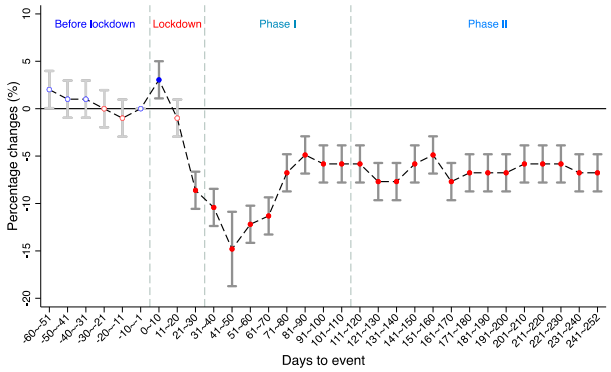
(b) Hours working on-site

Fig. 5. Event study on the number of individuals commuting to work and hours working on-site. The event-study graphs plot coefficients β_q from Eq. (1), which are percentage changes in the number of individuals commuting to work (panel (a)) and hours on-site conditioning on commuting (panel (b)). All regressions include city, day-of-week, days-to-event, holiday, and the treatment group fixed effects. The standard errors are clustered at the days-to-event level.

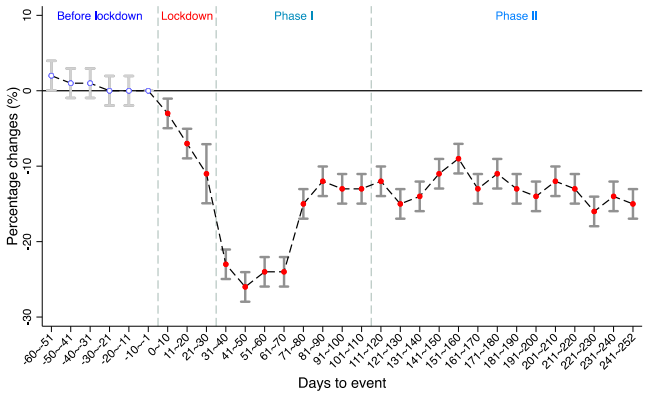
The event days are based on the lunar calendar. For 2020, the event date (or day zero) is the lockdown, January 23, 2020, two days before the Chinese New Year. Correspondingly, the event date in 2019 is February 3, 2019, also two days before the 2019 Chinese New Year. Phase I reopening started on February 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.



(a) Female v.s. male



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



(c) Migrants v.s. non-migrants

Fig. 6. Heterogeneity in commuting patterns across demographic groups. All event-study graphs plot coefficients β_q from Eq. (1), which are percentage changes in individuals commuting to work in 2020 relative to 2019. The dependent variable is the difference between female and male commuting to work (in logarithm) in panel (a), the difference between individuals age 40 and above and those below 40-year-old commuting to work (in logarithm) in panel (b), and the difference between migrants and non-migrants commuting to work (in logarithm) in panel (c). All regressions include city, day-of-week, days-to-event, holiday, and the treatment group fixed effects. The standard errors are clustered at the days-to-event level.

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.

Table 1. Effects on the number of individuals making unemployment calls and call duration

VARIABLES	(1) No. of individuals making calls (in log)	(2) Duration per call (in log)
Panel A:		
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)
Panel B:		
Pandemic period	0.27***	0.33***
(Lockdown + Phase I + Phase II)	(0.03)	(0.03)
Observations	489,514	489,514
R-squared	0.81	0.57
Neighborhood FE	Yes	Yes
Days-to-event FE	Yes	Yes
Day-of-week FE	Yes	Yes
Holidays FE	Yes	Yes
Treatment group FE	Yes	Yes

Note: Panel A in this table examines the percentage change in the number of individuals making unemployment calls and call duration as a result of the pandemic following Eq. (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. In panel B, we combine the three periods during the lockdown, Phase I reopening, and Phase II full reopening into one dummy variable, named the 'Pandemic period,' to examine the average impact of the pandemic. The observations are at the neighborhood by day level. The dependent variables in columns (1)-(2) are the number of individuals making unemployment calls (in logarithm) and average duration of unemployment calls (in seconds, in logarithm), respectively. All columns include neighborhood, day-of-week, days-to-event, holiday, and the treatment group fixed effects. Standard errors are reported in parentheses and clustered at the days-to-event. $*$ p <0.1, $**$ p < 0.05, $***$ p < 0.01.

Table 2. Regression results on commuters and working hours

VARIABLES	(1) No. of commuters (in log)	(2) Working hours (in log)
1-30 days before lockdown (=1)	0.04	0.01
	(0.02)	(0.01)
Lockdown period (=1)	-0.29***	-0.22***
	(0.04)	(0.02)
Phase I re-opening (=1)	-0.11***	-0.09***
	(0.02)	(0.01)
Phase II re-opening (=1)	-0.05**	-0.01
	(0.02)	(0.01)
Observations	13,052	13,052
R-squared	0.98	0.89
City FE	Yes	Yes
Days-to-event FE	Yes	Yes
Day-of-week FE	Yes	Yes
Holidays FE	Yes	Yes
Treatment group FE	Yes	Yes

Note: This table examines the percentage change in the number of commuters and duration of on-site working hours as a result of the pandemic, following Eq. (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 city by day level. The dependent variables in columns (1)-(2) are the number of commuters (in logarithm) and average working hours for commuters (in logarithm), respectively. A user that commutes to his work location at least once in a given week is coded as a commuter, and those who do not visit their work location at any time within a given week is coded as a non-commuter. All columns include city, day-of-week, days-to-event, holiday, and the treatment group fixed effects. Standard errors are reported in parentheses and clustered at the days-to-event. $*$ p <0.1, $**$ p < 0.05, $***$ p < 0.01.

Supplementary Materials for

"COVID-19 Pandemic and Unemployment: Evidence from Mobile Phone Data in China"

Panle Jia Barwick, Yongheng Deng, Xinfei Huang, Shanjun Li, Teng Li

Supplementary Text

Data description

China's mobile phone penetration is high among developing countries. Fig. S1 shows a strong correlation between the number of China Mobile users and the number of residents by cities. Cities with a higher GDP per capita (represented by the size of the dots in Fig. S1) tend to have higher mobile phone ownership.

Our data come from China Mobile, the dominant cellular service provider in China, with 58% of national mobile service subscribers. We have access to detailed phone usage and demographic information for all of its 71 million users in Guangdong Province from January 2018 to September 2020. We use the number of individuals calling 12333 instead of the number of calls to 12333 to measure the number of individuals claiming unemployment benefits, as people might make multiple calls to claim their benefits. Calls that failed to go through to the receiving party and calls shorter than 30 seconds are excluded from our data.

An important feature of our mobile phone data is the location tracking information (in longitude and latitude) that is collected by one million randomly selected mobile devices during 5-minute intervals from January 2018 to September 2020. Recent developments and the widespread diffusion of geospatial data acquisition technologies have enabled the creation of highly accurate spatial and temporal data (*16*). Note that passive collection of geolocation information – which underlies our data collection procedure -- works on all traditional mobile networks (2G, 3G, or 4G).

To construct users' work locations, we first identify the location where the user spent the most time between 9 am and 6 pm during weekdays for each month. If an individual spends at least 5 hours a day between 9 am and 6 pm for at least fifteen workdays at his primary location for a given month, we code this location as

this individual's work location for that month. For each individual and a given month, we define his home location as the location that has the most frequent usage between 10 pm and 7 am. Phone usage during 7 am-9 am and 6 pm-10 pm is excluded because people are likely on the move during these time intervals.

To validate our approach of using the number of unemployment calls to proxy for unemployment, Fig. S3 plots the number of unemployment calls across cities in 2019. The correlation between city-level unemployment calls and the official unemployment rate released by the NBS is reasonably high at 0.7 in 2019.

To validate our home and work locations, we provide two pieces of evidence. First, we compute the commuting distance for users with valid job location information. The distribution of commuting distance decays exponentially (Fig. S4 in the Supplemental Material), which is intuitive. Additionally, the average commuting distance in our sample period is around 6.6km, which is close to the administrative statistics (6.9km) by Guangzhou Municipal Transportation Bureau (GMTB) (*45*). Second, for the city of Guangzhou (the provincial capital city), Fig. S5 in the Supplementary Material 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 center of the city gains population, and the suburbs lose population during the daytime relative to the nighttime. But these differences are much more pronounced on weekdays than weekends, especially for the center of the city. To better illustrate the contrast between the day time and night time, we plot the industrial park in Guangzhou in the enlarged area. It is clear from the enlarged plot that the daytime population is much larger than the nighttime population, and the differences are amplified on weekdays.

Guangdong's lock down

Guangdong's provincial government acted swiftly and adopted vigilant procedures since the onset of the pandemic. Guangdong was one of the first provinces to release detailed information (frequency, location, gender, etc.) on the newly confirmed cases, starting from as early as February 3, 2020. These procedures proved successful and have kept the number of daily confirmed cases under a few handfuls since the full

reopening (Fig. S2 in the Supplementary Material). As shown in Fig. S2, the daily confirmed new cases reached a peak of 254 on January 31 and quickly reduced to under 50 three weeks into the lockdown period. The number of cases has been modest since then and varies between 0 and 34 throughout 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 get back to work after February 24. It is worth noting that the lockdown procedures in Guangdong are not as strict as the lockdown procedures implemented in the epicenter Wuhan. On February 24, 2020, Guangdong province entered Phase I reopening, which lasted 76 days. During Phase I reopening, people were allowed (and encouraged in certain industries) to go back to work and visit outdoor public places. Phase II reopening, or full reopening, officially started on May 9, 2020, when all businesses, including shopping malls, supermarkets, and restaurants, were allowed to fully open. The only exception was movie theaters that remained closed till mid-July.

While Guangdong's number of COVID cases is low, it does not imply that the pandemic has a modest or no effect on the local economy: the measures implemented to reduce the health impact of the pandemic could significantly affect the economy. As shown in our analysis in the main text, the pandemic has inflicted sizeable damage to Guangdong's labor market, leading to a 27%-62% increase in the unemployment rate. As Guangdong's economy is among the most vigilant across provinces in China, the aggregate labor market implications could be much more severe than those suggested by the national statistics.

Regression Analysis

As Table S1 illustrates, cities in Guangdong differ substantially, in terms of both population and GDP in 2019. Among the 21 cities, Guangzhou has the largest population (15.31 million), while Yunfu only has 2.55 million population. The economy scale of the largest city, Shenzhen, at $390 billion in 2019, is almost 30 times as large as that of the least city, Yunfu.

Table S2 uses Eq. (2) to explore the heterogeneous effects by trade exposure, which is measured by the ratio of a city's total export to its local GDP in 2019. The interactions between pandemic period dummy and export-to-GDP share are statistically significant and positive, indicating that cities with larger export-to-GDP shares experienced larger increases in unemployment calls.

The analysis of commuters in the main text uses the one-week window as a cutoff. Hence a user that commutes to his work location at least once in a given week is coded as a commuter, and those that do not visit their work location at any time within a given week are coded as a non-commuter. In Table S3 we repeat the entire analysis using a two-week window as a cutoff. The estimates are remarkably close to those in Table 2. Repeating the analysis using daily commuters again in Table S4 leads to very similar estimates as those in Table 2, consistent with Fig. 2 in the main text.

As most individuals commute within a city boundary, we aggregate the number of commuters to the city by day level in the main analysis. Table S5 shows the results are similar to those in Table 2 if we aggregate commuters to the neighborhood level.

Our main analysis excludes all users under the age of 18. 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 the percentage changes in calls (Table S6) and commuters (Table S7) are nearly identical when we limit to users 25 and above.

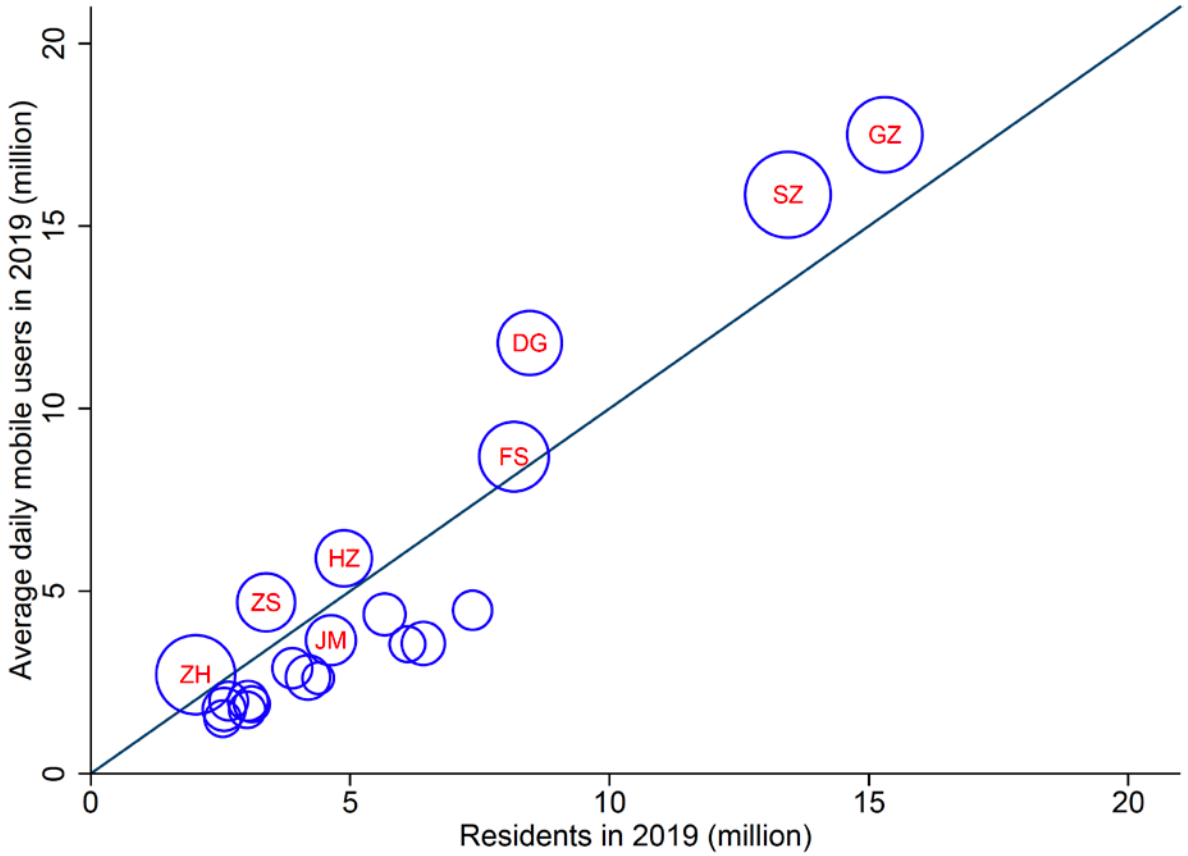


Fig. S1. Mobile users vs. residents in 2019. This graph presents the relationship between average daily mobile users and residents for cities in Guangdong in 2019. The solid line is a 45-degree line. The size of each marker denotes GDP per capita in 2019. We label the city names for those whose GDP per capita in 2019 is above 60,000 RMB (around 8,500 USD). The abbreviated city names are listed in Table S1.

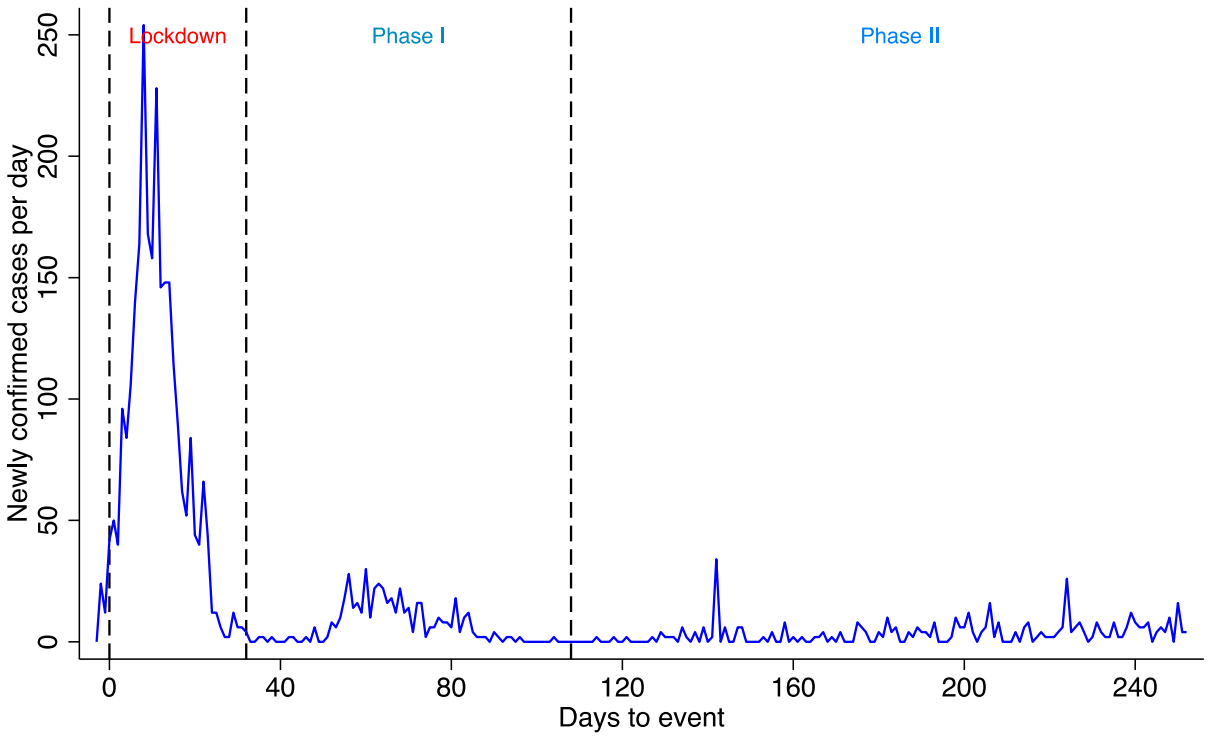


Fig. S2. Newly confirmed COVID-19 cases in Guangdong province. This graph shows the number of daily newly confirmed COVID-19 cases in Guangdong province. The provincial government started to report the number of COVID-19 cases on January 10, 2020, 13 days before the lockdown in Guangdong.

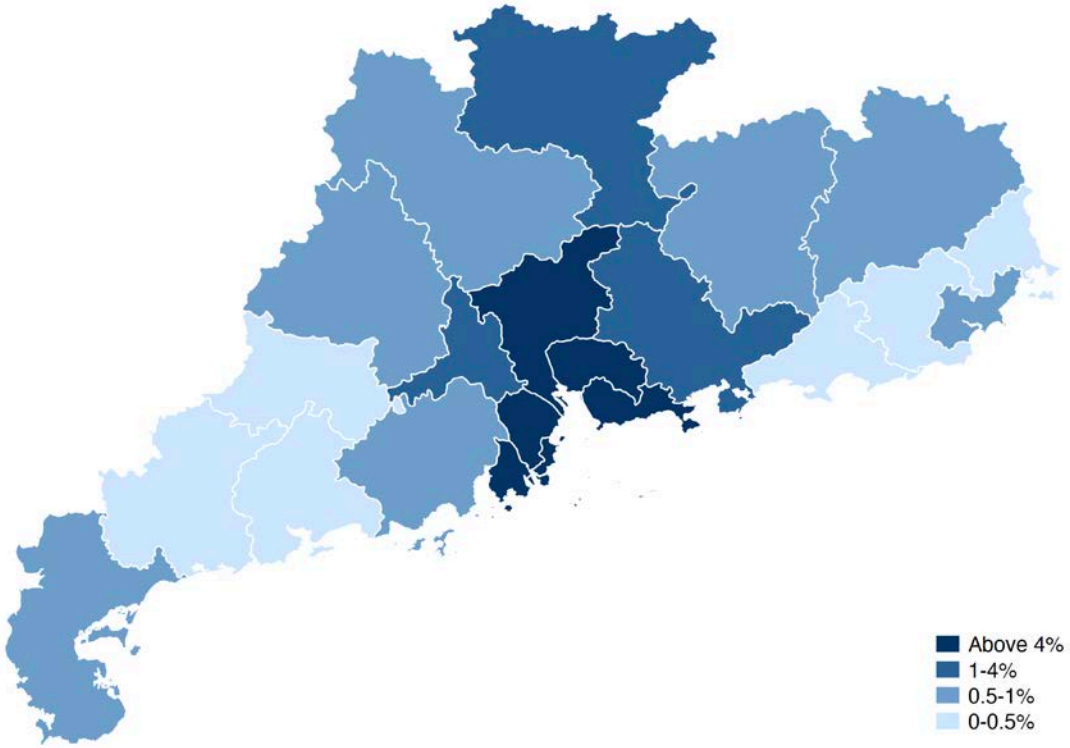


Fig. S3. Unemployment rate by city in 2019 based on the number of individuals making unemployment calls. This graph shows our estimated unemployment rate, which is the ratio of individuals making unemployment calls over the size of the labor force by city in 2019. The correlation between our unemployment measure and the official unemployment measure at the city level is 0.7 in 2019.

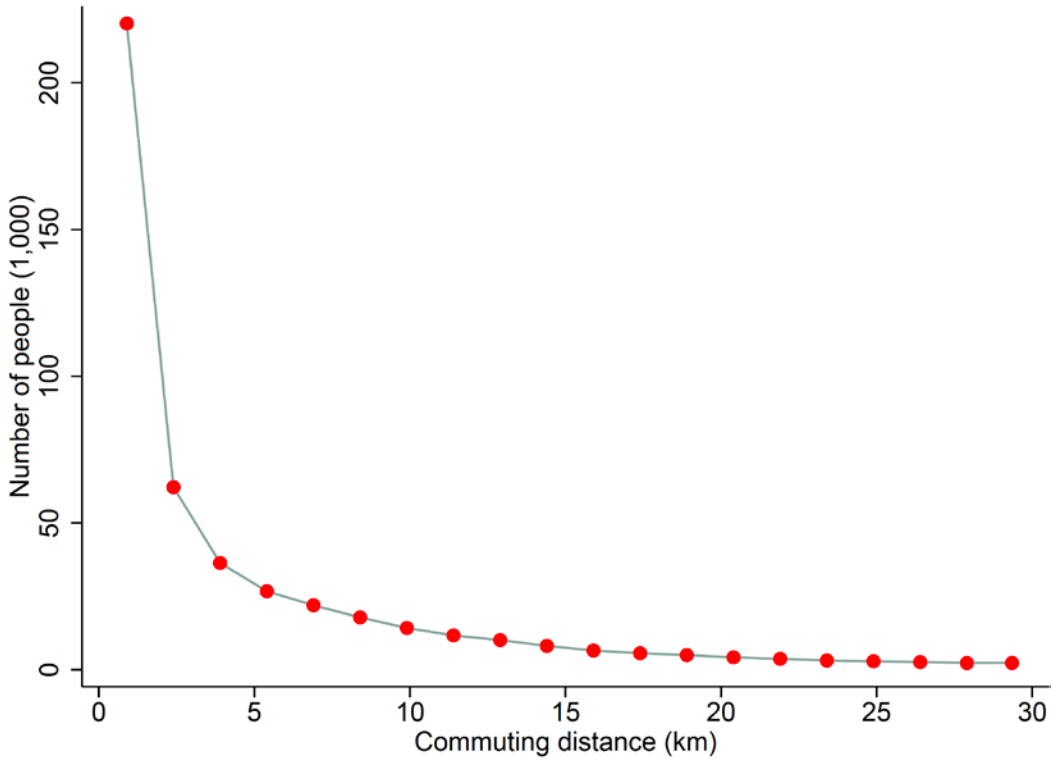
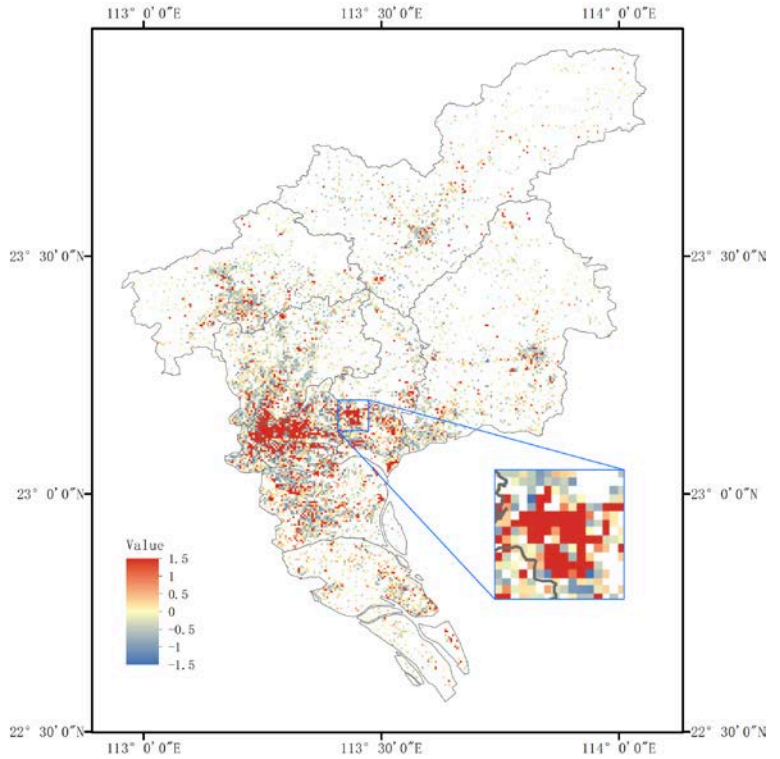
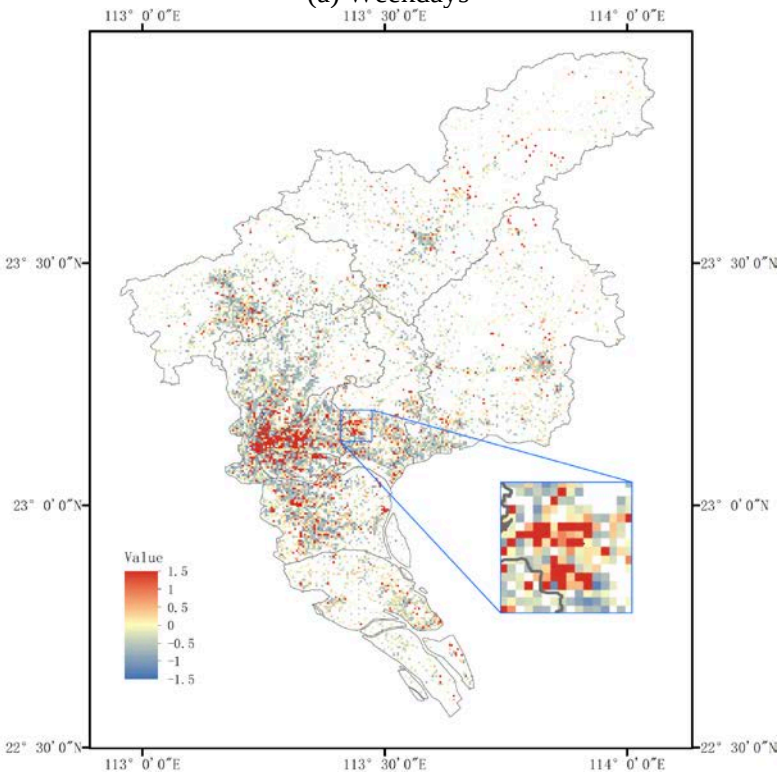


Fig. S4. The number of people commuting by commuting distance. This graph shows the number of people commuting at each distance. We use the coordinates of job and home locations to compute the spherical commuting distance.



(a) Weekdays



(b) Weekends

Fig. S5. Differences in (log) population during daytime and nighttime in Guangzhou. 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.

Table S1. Cities in Guangdong province

Cities	Population in 2019 (million)	GDP in 2019 (billion USD)
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	49.8
Maoming (MM)	6.41	47.13
Jiangmen (JM)	4.63	45.6
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	17.2
Chaozhou (CZ)	2.66	15.67
Shanwei (SW)	3.02	15.66
Heyuan (HY)	3.11	15.65
Yunfu (YF)	2.55	13.36

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

Table S2. Heterogeneity by Export-to-GDP ratio

VARIABLES	(1) No. of individuals making calls (in log)	(2) Duration per call (in log)
Pandemic period * Export/GDP in 2019 (%)	0.0019***	0.0019***
	(0.03)	(0.04)
Pandemic period (=1)	0.2086***	0.2688***
	(0.0232)	(0.0381)
Observations	489,514	489,514
R-squared	0.80	0.57
Neighborhood FE	Yes	Yes
Days-to-event FE	Yes	Yes
Day-of-week FE	Yes	Yes
Holidays FE	Yes	Yes
Treatment group FE	Yes	Yes

Note: This table examines whether the pandemic's effect differs across cities with varying exposure to international trade following Eq. (2), where we interact the pandemic treatment with a city's 2019 export-to-GDP ratio. The dependent variables in columns (1)-(2) are the number of unemployment calls (in logarithm) and average duration of unemployment calls in seconds (in logarithm), respectively. The observations are at the neighborhood and day level. All columns include neighborhood, day-of-week, days-to-event, holiday, and the treatment group fixed effects. Standard errors are reported in parentheses and clustered at the days-to-event. $*$ p <0.1, $**$ p <0.05, $***$ p < 0.01.

Table S3. Regression results on commuters and working hours (two-week-window)

VARIABLES	(1) No. of commuters (in log)	(2) Working hours (in log)
1-30 days before lockdown (=1)	0.03	0.01
	(0.02)	(0.01)
Lockdown period (=1)	-0.30***	-0.21***
	(0.03)	(0.02)
Phase I re-opening (=1)	-0.10***	-0.08***
	(0.03)	(0.01)
Phase II re-opening (=1)	-0.048**	-0.02
	(0.02)	(0.02)
Observations	13,052	13,052
R-squared	0.98	0.90
City FE	Yes	Yes
Days-to-event FE	Yes	Yes
Day-of-week FE	Yes	Yes
Holidays FE	Yes	Yes
Treatment group FE	Yes	Yes

Note: This table replicates the analysis in Table 2, but uses a two-week window instead of a one-week window to define commuters. The dependent variables in columns (1)-(2) are the number of commuters (in logarithm) and average working hours for commuters (in logarithm), respectively. A user that commutes to his work location at least once in two weeks is coded as a commuter during the two-week window, and those who do not visit their work location at any time within that two weeks is coded as a non-commuter in those two weeks. All columns include city, day-of-week, days-to-event, holiday, and the treatment group fixed effects. Standard errors are reported in parentheses and clustered at the days-to-event. $*$ $p < 0.1$, $**$ $p < 0.05$, $***$ $p < 0.01$.

Table S4. Regression results on commuters and working hours (one-day-window)

VARIABLES	(1) No. of commuters (in log)	(2) Working hours (in log)
1-30 days before lockdown (=1)	0.02	0.02
	(0.02)	(0.03)
Lockdown period (=1)	-0.35***	-0.26***
	(0.06)	(0.03)
Phase I re-opening (=1)	-0.15***	-0.11***
	(0.03)	(0.02)
Phase II re-opening (=1)	-0.07***	-0.02
	(0.02)	(0.03)
Observations	13,052	13,052
R-squared	0.97	0.90
City FE	Yes	Yes
Days-to-event FE	Yes	Yes
Day-of-week FE	Yes	Yes
Holidays FE	Yes	Yes
Treatment group FE	Yes	Yes

Note: This table replicates the analysis in Table 2, but uses a one-day window instead of a one-week window. The dependent variables in columns (1)-(2) are the number of commuters (in logarithm) and average working hours for commuters (in logarithm), respectively. A user who commutes to his work location on a given day is coded as a commuter on that day, and those who do not visit their work location on that day is coded as a non-commuter on that day. All columns include city, day-of-week, days-to-event, holiday, and the treatment group fixed effects. Standard errors are reported in parentheses and clustered at the days-to-event. $*$ $p < 0.1$, $**$ $p < 0.05$, $***$ $p < 0.01$.

Table S5. Regression results on commuters and working hours (neighborhood level)

VARIABLES	(1) No. of commuters (in log)	(2) Working hours (in log)
1-30 days before lockdown (=1)	0.01	0.01
	(0.02)	(0.03)
Lockdown period (=1)	-0.26***	-0.19***
	(0.02)	(0.02)
Phase I re-opening (=1)	-0.08***	-0.06***
	(0.01)	(0.02)
Phase II re-opening (=1)	-0.03***	-0.02
	(0.01)	(0.02)
Observations	489,514	489,514
R-squared	0.93	0.85
Neighborhood FE	Yes	Yes
Days-to-event FE	Yes	Yes
Day-of-week FE	Yes	Yes
Holidays FE	Yes	Yes
Treatment group FE	Yes	Yes

Note: This table replicates the analysis in Table 2, except that the data are aggregated to the cell-tower-date level (i.e., neighborhood by day). The dependent variables in columns (1)-(2) are the number of commuters (in logarithm) and average working hours for commuters (in logarithm), respectively. A user who commutes to his work location at least once in a given week is coded as a commuter during that week, and those who do not visit their work location at any time within a given week is coded as a non-commuter for that week. All columns include neighborhood, day-of-week, days-to-event, holiday, and the treatment group fixed effects. Standard errors are reported in parentheses and clustered at the days-to-event. $*$ $p < 0.1$, $**$ $p < 0.05$, $***$ $p < 0.01$.

Table S6. Effects on number of individuals making unemployment calls and call duration (excluding people under 25 years old)

VARIABLES	(1) No. of individuals making calls (in log)	(2) Duration per call (in log)
Panel A:		
1-30 days before lockdown	0.03	0.03
	(0.03)	(0.04)
Lockdown period	-0.38***	-0.35***
	(0.06)	(0.07)
Phase I re-opening	0.26***	0.22***
	(0.02)	(0.05)
Phase II re-opening	0.43***	0.58***
	(0.02)	(0.05)
Panel B:		
Pandemic period	0.27***	0.32***
(Lockdown + Phase I + Phase II)	(0.04)	(0.03)
Observations	489,514	489,514
R-squared	0.82	0.59
Neighborhood FE	Yes	Yes
Days-to-event FE	Yes	Yes
Day-of-week FE	Yes	Yes
Holidays FE	Yes	Yes
Treatment group FE	Yes	Yes

Note: This table replicates the analysis in Table 1, except that individuals under 25 years old are excluded from the analysis. The dependent variables in columns (1)-(2) are the number of individuals making unemployment calls (in logarithm) and average duration of unemployment calls (in seconds, in logarithm), respectively. All columns include neighborhood, day-of-week, days-to-event, holiday, and the treatment group fixed effects. Standard errors are reported in parentheses and clustered at the days-to-event. $*$ p <0.1, $**$ p < 0.05, $***$ p < 0.01.

Table S7. Regression results on commuters and working hours

(excluding people under 25 years old)

VARIABLES	(1) No. of commuters (in log)	(2) Working hours (in log)
1-30 days before lockdown (=1)	0.04	0.01
	(0.02)	(0.01)
Lockdown period (=1)	-0.27***	-0.23***
	(0.03)	(0.02)
Phase I re-opening (=1)	-0.12***	-0.10***
	(0.03)	(0.02)
Phase II re-opening (=1)	-0.06***	-0.02
	(0.02)	(0.02)
Observations	13,052	13,052
R-squared	0.98	0.89
City FE	Yes	Yes
Days-to-event FE	Yes	Yes
Day-of-week FE	Yes	Yes
Holidays FE	Yes	Yes
Treatment group FE	Yes	Yes

Note: This table replicates the analysis in Table 2, except that individuals under 25 years old are excluded from the analysis. The dependent variables in columns (1)-(2) are the number of commuters (in logarithm) and average working hours for commuters (in logarithm), respectively. A user that commutes to his work location at least once in a given week is coded as a commuter, and those who do not visit their work location at any time within a given week is coded as a non-commuter. All columns include city, day-of-week, days-to-event, holiday, and the treatment group fixed effects. Standard errors are reported in parentheses and clustered at the days-to-event. $*$ p <0.1, $**$ p < 0.05, $***$ p < 0.01.