The informal economy at times of COVID-19 pandemic

Feng Guo [a,d], Yiping Huang [b], Jingyi Wang [c], Xue Wang [b,*]

[a] School of Public Economics and Administration, Shanghai University of Finance and Economics, PR China
[b] National School of Development/Institute of Digital Finance, Peking University, PR China
[c] School of Finance, Central University of Finance and Economics; Institute of Digital Finance, Peking University, PR China
[d] Institute of Digital Finance, Peking University, PR China

ARTICLE INFO

ABSTRACT

We provide a first view of vulnerable informal economy after the blows from COVID-19, using transaction-level business data of around 80 million offline micro businesses (OMBs) owners from the largest Fintech company in China and employing machine learning method for causal inference. We find that the OMBs activities in China experienced an immediate and dramatic drop of 50% during the trough. The businesses had rebounded to around 80% of where they should be seven weeks after the COVID-19 outbreak, but had remained at this level until the end of our time window. We find a larger disruption to the OMBs in urban areas, the female merchants and the merchants who were not grown up in the places where they conducted businesses. We discuss the implications for policy support to the most vulnerable, and highlight the importance to take full advantage of digital development to follow up the informal economy.

1. Introduction

Obtaining accurate statistics about the informal economy has become one of the most daunting challenges in most developing countries (Blackburn, Bose, & Capasso, 2012; Capasso & Jappelli, 2013), especially examining the pain they felt at the times of big crisis such as COVID-19 pandemic. Even if the pandemic caused heavy blows to society as a whole, it would hit the informal workers disproportionately. For example, the street vendors mainly work in the services sector, are usually self-employed or informally employed without social insurance, and are mainly in micro and family enterprises. Measuring the size of the informal economy accurately and examining the impact that they felt is important for making effective economic policy decisions. In China, it is becoming possible due to the rapid digitalization of the economy, as paying with digital payment tools has become a daily occurrence. According to People's Bank of China,[1] the proportion of adults using digital payments was 82.39% in 2018. Domestic users of Alipay and WeChat Pay, China's two largest digital payment service providers, have exceeded 900 million. People use digital payment tools for online and offline transactions, and even small businesses such as street shops and peddlers have adopted digital payments such as Alipay, WeChat Pay, UnionPay, and others. The accumulation of digital information enables us to largely approximate the size of informal sector and gauge to what extent they have been hit by the COVID-19 pandemic.

In China, mobility restrictions like lockdowns and social distancing measures have greatly contributed to the containment of the spread of the disease (Chinazzi et al., 2020; Tian et al., 2020), and the numbers of domestic new cases came under control in March

* Corresponding author at: National School of Development/Institute of Digital Finance, Peking University, No.5 Yiheyuan Road Haidian District, Beijing 100871, PR China.
E-mail address: xwang1122@126.com (X. Wang).

[1] http://www.pbc.gov.cn/goutongjiaoliu/113456/113469/3905926/index.html, China's Financial Inclusion in 2018; accessed April 11, 2020.

2020. However, aggressive countermeasures, such as stringent lockdowns, have also imposed tremendous economic costs on the country in the short run. China's economy shrank by 6.8% year-on-year in the first quarter of 2020, the first contraction in the past four decades. As a result, vulnerable groups are likely experiencing deterioration of their income and livelihood. China emerged from a two-month containment phase and moved into the mitigation stage by early April 2020.[2] In this article, we explore the impacts felt by China's offline micro businesses (OMBs) in the informal sector, which are mainly self-employed in the services sector and not able to work from home.

Tens of millions of OMBs have been disproportionately affected by the pandemic and lockdown measures. First, OMBs operate largely in the informal services sector of the economy, and they are usually self-employed or informally employed without social insurance. Most OMBs employ low-skilled workers who could not work from home during the pandemic. Second, most OMBs survive hand-to-mouth, with limited savings and lack of access to unemployment benefits. They rely heavily on cash flows and short-term loans, due to limited savings. Those employed in the gig economy are particularly vulnerable to dramatic collapses of income and loss of livelihood. The negative effects on this group are likely to have a long-lasting impact and are a leading cause of persistent inequalities and low mobility (Lustig, Stone, & Tommasi, 2020).

Using weekly data on around 80 million "QR code merchants" (*Mashang* in Chinese) from Ant Group (hereinafter "Ant"), the largest Fintech company in China, this article studies the initial impacts of COVID-19 on OMBs and their recovery since the virus has been largely under control. The QR code merchants are OMBs that collect payments via Alipay, a digital payment tool of Ant. One of the officially public statistic is that the registered individual business is around 100 million,[3] which is comparable to our sample including around 80 million OMBs. Our sample spans from December 31, 2019 to April 2, 2020 and the corresponding lunar calendar dates in 2018 and 2019. We use the lunar calendar dates to account for the seasonality in economic activities during the Lunar New Year. A notable day in COVID-19 control and prevention was January 20, 2020 (five days before the Lunar New Year), when human-to-human transmission of the coronavirus was confirmed and reported to the public. Thus, we use that day as the start of the outbreak and the event date in the following analysis. We define the periods before and after January 20 (December 26 in the lunar calendar) as the pre- and post-virus periods, respectively. We use the corresponding lunar calendar dates in 2018 and 2019 to define the pre- and post-virus periods.

A simple year-on-year change in OMB activities would lead to mismeasurement of the real economic impacts of the pandemic, because it is likely that the businesses were on a growing or decreasing path relative to the same period last year, if there were no COVID-19 pandemic.[4] We first predict the counterfactuals using a machine learning technique and further interpret the difference between the realized and counterfactual results as the causal impact of the pandemic (Athey, 2017). Specifically, we merge OMB data with other economic, population, and geographic characteristics at the Thiessen-polygon level.[5] We predict the counterfactual activities of OMBs in the post-virus period in 2020 by modelling the relationship between the activities of OMBs in the post-virus period and the other feature variables. The feature variables include OMBs activities around the same period in the previous year, OMBs in the weeks before the COVID-19 outbreak in 2020, cross-section data as of December 2019 such as economic environment, population, and geographics, and panel data like meteorological characteristics in 2020. The parameters are from the trained gradient boosting decision tree (GBDT) model using the data from 2018 and 2019, which reveals the relationship between the activities of OMBs in the post-virus period in 2019 and the feature variables. The difference between the predicted counterfactuals and the actual values in 2020 gives an estimate of the real impact of the pandemic on OMBs.

We find a massive decline in the number and sales turnover of active OMBs in the post-virus period in China. The number and sales turnover of active merchants bottomed out in the second week of the post-virus period, January 31 – February 6, and remained in a downturn for the following three weeks. Relative to the counterfactual level estimated from the machine learning technique, the average weekly drops in the number and sales turnover of active merchants were around 50%, between January 31 and February 20, after which OMB activity started to rebound. As of early April, one month after the trough, OMB activities bounced back to around 80% of their counterfactual levels. In addition, we find that the announcements of government lockdown polies could explain a limited portion of overall decline of OMBs activities, and what matters most may plausibly be the voluntary containment measures.

OMBs in urban areas bore the hardest hit. The largest weekly decline in the number of active merchants was about 54% in urban areas, compared with a 41% contraction in rural areas. In addition, we see a simultaneous contraction of business activity during the entire month of February, followed by a nationwide synchronous recovery starting at the end of February, although there were regional variations in the spread of the virus. Female merchants saw drops of around 53 and 57% in the number and sales turnover of active OMBs during the trough, respectively, and the drops were about 5 and 9% larger than the average for the male merchants. The decrease in economic activities were larger for the outsiders, the owners who were not born in the province where they conducted businesses, with a larger drop of 7% in the number of active merchants than the natives who were managing businesses in their birth provinces. Generally speaking, the outsiders do not enjoy the same social benefits as the natives and are more vulnerable to shocks.

Although there is fast-growing research on COVID-19, as far as we know, our paper is the first to study the impact on OMBs that would be disproportionately hard hit by the pandemic and lockdown measures. Other studies focus on the spread, containment, and economic and political consequences of COVID-19 and previous pandemics, given its significant damage (Atkeson, 2020; Baker,

Farrokhnia, Meyer, Pagel, & Yannelis, 2020; Barro, Ursúa, & Weng, 2020; Chen, Qian, & Wen, 2020; Eichenbaum, Rebelo, & Trabandt, 2020; Fang, Wang, & Yang, 2020). Dai, Mookherjee, Quan, and Zhang (2020) examine how exposure of Chinese registered firms to the Covid-19 shock varied with a cluster index (measuring spatial agglomeration of firms in related industries) at the county level. However, few studies estimate the real impact on hard-hit informal workers that were particularly vulnerable to dramatic collapse during the COVID-19 pandemic.

In addition, we contribute to the broader literature on informal economy. Informal businesses are inherently difficult to identify, because most of them are usually small-scale, frequently family-based, and perhaps low-productivity businesses with much informality and nearly no records of social security. However, informal economy such as OMBs contribute significantly to employment, especially in developing countries (La Porta & Shleifer, 2014; Maloney, 2004). Informal workers in the Asia-Pacific region account for nearly 60% of nonfarm employment, ranging from around 20% in Japan to over 80% in Myanmar and Cambodia.[6] OMBs in China operate largely in the informal services sector, and most of them could not work from home during the pandemic. We identify informal business in the gig economy by taking advantage of their digital footprints, using data from the world's largest Fintech company.

2. Data description

Although research on small and medium enterprises (SMEs) is growing rapidly, there are very few national datasets that provide information on the economic activities of micro businesses, especially of those that are not registered with the Industry and Commerce Department. We employ data on the weekly number and sales turnover of around 80 million QR code merchants from December 31, 2019 to April 2, 2020 and for the corresponding lunar calendar dates in 2018 and 2019 from Ant Group. On January 20, human-to-human transmission of the coronavirus was confirmed by Chinese authorities, which marked a dramatic change in how the evolving pandemic was managed and contained. Thus, we define the periods before and after January 20 (December 26, lunar calendar) as the pre- and post-virus periods, respectively, and the corresponding lunar calendar dates in the same period in 2018 and 2019 accordingly.[7] Lunar New Year's Eve in 2020 fell on January 24, four days after the event date, and there is often a seven-day Lunar New Year holiday.[8] Thus, the first timeslot in the post-virus period includes the three days before Lunar New Year's Eve and seven days afterward, taking the potential impact of the holiday into account.

We propose to aggregate the OMB data at the Thiessen-polygon level. Thiessen polygon, otherwise known as Voronoi diagram, is an essential method for the analysis of proximity and neighborhood (Blumenstock, Eagle, & Fafchamps, 2016; Jia, Khadka, & Kim, 2018; Piza & Gilchrist, 2018).[9] The method defines an area around a center point, where every location is nearer to this point than to all the others. In our analysis, each OMB belonging to a specific Thiessen polygon is the closest to its own center point, compared with the distance to any other center point. We merge the bank branches within each 500-meter grid cell into one as the center point, by taking the average of their geographic coordinates, and finally establish 138,629 polygons across mainland China.

In addition, we collected raster data on variables that may affect OMB activities, such as economic development, population, and geographic characteristics. The customers of OMBs are mainly nearby, as OMBs are offline and small in scale, and the surroundings of OMBs play a significant role in their daily business. First, given the spread of the virus and the sensitivity of OMB activity to the weather, we include meteorological conditions in the analysis, such as temperature, wind speed, air pressure, humidity, and precipitation. Second, we obtained data on around 35 million "points of interest"[10] from AutoNavi (*Gaode* in Chinese) Application Programming Interface in December 2019, to proxy for local general conditions. Third, we have the cross-section data as of December 2019 including raster economic development, population, and geographic variables, specifically, nighttime lights data with a 500-meter spatial resolution, population data at the 1000-meter grid cell level, and elevation at the 30-meter grid cell level. We further calculate the driving distance from the center of the polygon to that of the county, prefecture-level city, and capital city of the province, reflecting transportation convenience at the polygon level.[11]

3. Model specification

To estimate the real impact of the pandemic on OMBs, we need to predict the counterfactual level of OMBs economic activities without the COVID-19 outbreak. The difference between the actual transactions and the counterfactuals would tell the real drop. Most studies that focus on the economic impact of COVID-19 use linear regression in a difference in difference (DID) specification (Chen et al., 2020; Fang et al., 2020), but in our setting and data frame, we document that two issues related to the linear DID specification would lead to a biased estimation. First, it is not clear that the explanatory variables, including economic, population, geographic characteristics, and the activities of OMBs in the previous year and in the pre-virus period, are linearly related to the changes in OMBs activities in the post-virus period. And it is also not clear *ex ante* what factors among the dozens of variables would be most relevant.

---

[6] https://blogs.imf.org/2020/04/30/a-new-deal-for-informal-workers-in-asia/, A "New Deal" for Informal Workers in Asia, accessed May 27, 2020.

[7] Please see Appendix B.1 for more details on the time window.

[8] We ignore the holiday extension to contain the virus spread, to make the time window comparable with the previous years.

[9] Please see Appendix A for more details on Thiessen polygon.

[10] A point of interest (POI) is a specific location that someone may find useful or interesting. Simply, a hotel, campsite, fuel station, store, or any other specific entity on a map or navigation system can be called a POI.

[11] Please see Appendix B.2 for summary statistics and data sources.

The result of running a kitchen sink regression is that we would likely run the risk of overfitting the data and estimate spurious relations between regressors and regressand (Rossi & Utkus, 2020). Instead, we rely on a machine learning method known as GBDT. GBDT not only allows large conditioning information sets, but it also allows for non-linearities without overfitting or falling prey of the so-called curse of dimensionality. Second, the core assumption to identify the treatment effect in DID estimators is the Parallel Paths assumption, namely that the average change in outcome for the treated in the absence of treatment equals the average change in outcome for the non-treated. However, the QR code merchants in China just have experienced rapid growth since 2018, and it is still on a growing path these two years. The average growth of OMBs activities in 2020 in the absence of COVID-19 pandemic is hard to equal that in 2019, which invalidates the Parallel Paths assumption. In fact, we tested paths of the number and sales turnover of active OMBs in the pre-virus period in 2019 and 2020 and did find a non-parallel pre-trend. Our setting employing GBDT and nonparametric estimation allows for more flexible dynamic responses of OMBs activities to the feature variables. The basic assumption in our model is that the "relationship" between inputs and response variables rather than the growth path of OMBs in 2019 still holds in 2020, which relaxes the restrictions on the average changes of outcome variables.

To be more specific, in our setting, we predict the counterfactual level of OMBs activities in the post virus period by leveraging the relationship between the activities of OMBs in the post-virus period and the feature variables including OMBs around the same virus-period in the previous year, OMBs in the weeks before the COVID-19 outbreak in 2020, economic conditions, population, and geographic and meteorological characteristics in 2020. The parameters are from the trained GBDT model using the same data frame from 2018 and 2019, which reveals the relationship between the activities of OMBs in the post-virus period in 2019 and the feature variables. The assumption behind this estimation is that the relationship between the business activities of the OMBs in the post-virus period in 2019 and the feature variables would still exist in 2020 if there were no COVID-19 outbreak. Eq. (1) shows the prediction model for counterfactual OMBs activities in 2020:

$$\widehat{OMB_{i,2020+k}} = F\left(OMB_{i,2019+k}, OMB_{i,2019+(k-1)}, OMB_{i,2019-h}, OMB_{i,2020-h}, X_{i,2020+k}, Z_i\right) \tag{1}$$

where $\widehat{OMB_{i,2020+k}}$ indicates the predicted number or sales turnover of active OMBs at the Thiessen polygon level in polygon i in the kth week in the post-virus period in 2020 if there were no pandemic. $OMB_{i,\ 2020-h}$ ($OMB_{i,\ 2019-h}$) is a vector of OMB activities in polygon i in the three weeks (i.e., h = 1,2,3) before the outbreak in 2020 (2019), including the number of merchants and associated sales turnover. $OMB_{i,\ 2019+k}$ ($OMB_{i,\ 2019+(k-1)}$) include OMB activities in polygon i in the kth ($k-1$th) week in the post-virus period in 2019. $X_{i,\ 2020+k}$ is a vector including the meteorological variables, and Z_i includes the economic, population, and geographic characteristics.

The parameters in model (1) are borrowed from model (2) which models the relationship between OMB activities in the post-virus period in 2019 and the feature variables using the same data structure in 2018 and 2019. The basic assumption is that the conditional distribution of the business activities of OMBs would be stable if there were no exogenous shock in a relatively short run, two years in our setting. Given that the COVID-19 outbreak coincided with the Lunar New Year, during which economic activity shows obvious periodicity, we suggest that our estimation is reliable to a great extent. Here is the specification of model (2):

$$OMB_{i,2019+k} = F\left(OMB_{i,2018+k}, OMB_{i,2018+(k-1)}, OMB_{i,2018-h}, OMB_{i,2019-h}, X_{i,2019+k}, Z_i\right) \tag{2}$$

where $OMB_{i,\ 2019+k}$ is the number or sales turnover of active OMBs in polygon i in the kth week in the post-virus period in 2019. The feature variables used for estimating the weekly number and sales of active OMBs in the post-virus period in 2019 include OMBs activities around the same period in 2018, OMBs in the weeks before the COVID-19 outbreak in 2019, cross-section data as of 2019 such as economic environment, population, and geographics, and panel data like meteorological characteristics in 2019. We train a gradient boosting decision tree (GBDT) model at the Thiessen polygon level.[12]

4. Results

4.1. Direct COVID-19 impacts

Fig. 1 provides a graphical illustration of the impacts of COVID-19 by displaying the peaks and troughs of the number and sales turnover of active OMBs. The left panel shows the number of active OMBs each week in 2020 and the same period in 2019, while the right panel shows the sales turnover, both scaled by the average of the first week in the time window in 2018. We check the validity of our predicted counterfactuals in two ways. First, we predict the number and sales of active OMBs in the corresponding lunar calendar weeks in 2019, namely, the predicted results spanning 10 weeks after January 31, 2019 in model (1). The black dashed lines in Fig. 1 show the predicted paths in 2019, while the black solid lines show the actual paths. It may not be easy to distinguish the black dashed and solid lines because they largely overlap, indicating that the predicted number and sales of OMBs from our model are very close to their actual values in 2019. Second, we predict the number and sales turnover of OMBs in the pre-virus period in 2020 following similar setting in model (1) and (2) but shifting the event date three weeks backwards. In another word, we assume that the pseudo-event date is 30 December 2019, and follow the basic logic of the models in Section 3 to predict the OMBs activities between 30 December 2019 and 20 January 2020. In this way, the weeks between 30 December 2019 and 20 January 2020 are pseudo post-virus periods. And the parameters used to predict OMBs activities in these three weeks are those estimated using data in the previous two years. The red

---

[12] Please see Appendix C for the hyperparameters and performance of the model.

Number of Active OMBs
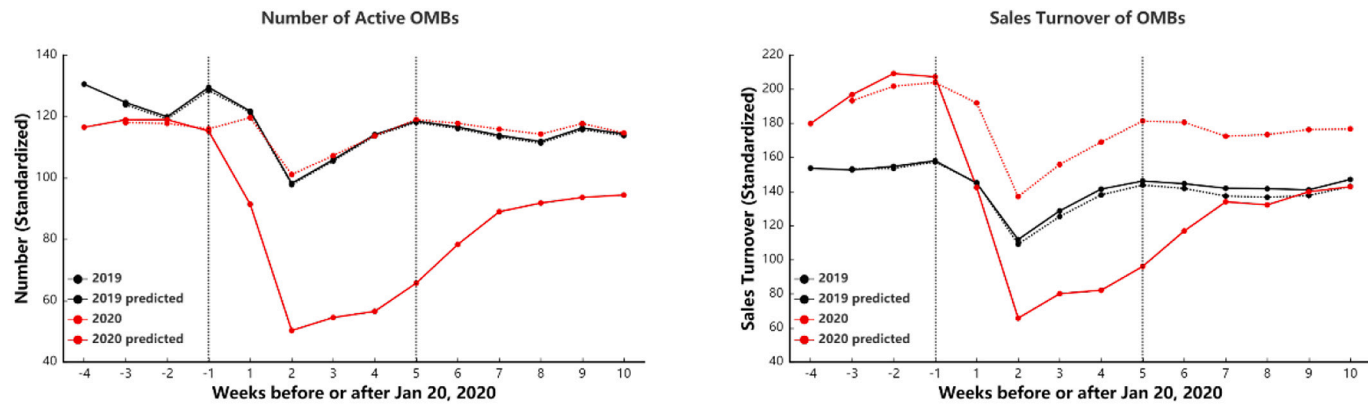
Sales Turnover of OMBs

Fig. 1. Actual and predicted OMB activities over time.

Note: 1. The horizontal axis indicates weeks after January 20, and negative numbers are for weeks before January 20. 2. The first week includes the 10 days from January 21 to 30, and the other weeks are seven-day time windows. 3. The vertical axes are the number or sales turnover of active OMBs, both scaled by the average of week -4 in 2018 (January 15–21, 2018). 4. *2019* and *2020* show the actual number and sales turnover of OMBs in the two years, and *2019 predicted* and *2020 predicted* trace the counterfactual paths. 5. The first vertical dashed line marks the first turning point of OMB activities, and the second vertical dashed line marks the start of the bounce of OMBs activities relative to their counterfactuals.

dashed lines in Fig. 1 show the predicted paths in 2020, while the red solid lines show the actual paths. We see the predicted paths in the pre-virus period in 2020 largely coincides with the actual paths (week -1, -2 and -3), implying that the relationship between feature variables and OMBs activities in the pre-virus period in 2019 does hold in the same period in 2020. Therefore, this at least seems to be more reasonable to believe that the relationship in the true post-virus period (weeks after 20 January 2020) is also stable in 2020 if there were no COVID-19 pandemic.

We first explore how business operations were affected by the pandemic. Fig. 1 shows that the worst was underway at the end of the Lunar New Year holiday. The gap between the series in 2020 and that in 2019 became wider in the second week in the post-virus period, and the gap was widest in the fourth week (February 14–20, 2020). The negative economic impact on OMBs is already visible, simply comparing with 2019. However, we should provide a better answer to the question: how large was the drop of OMB activity? The answer relies on the counterfactual path in the post-outbreak period if there were no pandemic.

We then provide a detailed exploration of the impacts on OMBs activities. Fig. 2 shows the time series of the ratio of actual values to their counterfactuals in terms of the number and sales turnover of OMBs. A ratio of 1 indicates that the actual economic activity of OMBs was exactly at the level of what it should have been, and the smaller the ratio is, the sharper was the decline in business activities. The difference between the ratio and 1 indicates the percentage change in the weekly number and sales turnover of active OMBs relative to their counterfactuals.

Fig. 2 shows that OMB activities did not bottom out until the second week in the post-virus period. There are two main reasons for the smaller drop in the first 10 days. First, there was an initial seasonal decrease in the post-virus period, as most merchants closed their doors and enjoyed their holidays, and the impact was yet to unravel. Second, the smaller drop is consistent with the stockpiling behavior of consumers, as it increasingly became clear that there would be a significant number of virus cases in China (Baker et al., 2020). Our headline estimate is that the number of active OMB merchants dropped by around 50% (black line, $0.5-1 = -0.5$) in response to the pandemic and shutdown during the worst period. The sales turnover saw a drop of 52% in the second week (red line), and it remained in a downturn in the following two weeks before a rebound in the fifth week. As the virus became under control, mobility restrictions were gradually eased and the government facilitated the resumption of work. The number and sales turnover of active OMBs rebounded to around 77 and 78%, respectively, in March (the seventh week in the post-virus period, March 6–12, 2020), relative to their counterfactuals for 2020. The estimated number of OMBs in China in 2018 was about 97.76 million,[13] employing about 230 million people. If all the OMBs experienced a similar contraction in the number of active merchants as the QR merchants did, there were about 115 million ($= 0.5 * 230$ million) previously self-employed workers in China who were out of work during the worst period, accounting for around 14.3% ($= 115/806$) of the total labor force in 2018. This is a very rough estimation, as we are trying to provide a sense of the overall contraction of OMB activities.

4.2. Disentangling lockdown effects from the overall impacts

We have seen an immediate drop and partial recovery in the number and sales turnover of active OMBs during the following ten weeks after the virus outbreak. It is not clear, however, whether the economic decline came from the lockdown orders or the panic and spontaneous willingness to stay at home brought by the virus. The question of how much of this collapse can be explained by the government regulations is not immediately obvious in the figures. In this section, we estimate how lockdown affected economic activities of OMBs. The severity of COVID-19 prevalence varied across different regions, and the specific terms and requirements of the lockdown also differed across provinces and cities, which brings a great challenge to define the effective day of the lockdown policy. However, the mismeasurement due to difference between various measurements of policies effective date are not likely to exceed one week, the challenge is largely mitigated when we define the effective week. Therefore, we follow the information from He, Pan, and Tanaka (2020) on local government's lockdown policies at city level obtained from media news and government announcements. The information covered 324 cities, and 95 out of them were totally locked down. To assess the impacts of government regulations on OMBs activities, we employ a staggered DID specification. The DID model leverages variations in the effective week of lockdown policies across cities, and thus estimates how city lockdown depresses OMBs activities relative to non-locked-down cities in 2020. This empirical strategy allows us to control for various confounding factors that potentially affect OMBs activities, and to disentangle a plausible clean impact of virus containment measures. It is worth noting that the possible problems related to DID estimation in Section 3 come from comparing how COVID-19 outbreak in 2020 affect OMBs relative to trends in the previous year, which is different from identifying the impacts of local government regulations here. We first test the pre-treatment parallel trends assumptions and investigate the dynamic evolution of the impacts with the following event study approach:

$$OMB_{i,c,k} = \beta_0 + \sum_{j=-3}^{-1} \beta_j Lockdown_{c,j} + \sum_{l=1}^{10} \beta_l Lockdown_{c,l} + \delta X_{i,c,t} + u_c + v_k + \varepsilon_{i,c,k} \qquad (3)$$

where $OMB_{i,c,k}$ is the logarithms of the number or sales turnover of OMBs in polygon i and city c in the kth week. $Lockdown_{c,j}$ takes a value of one for OMBs in each polygon in city c in the jth week before lockdown policies taking into effect, and zero otherwise. While $Lockdown_{c,l}$ takes a value of one for OMBs in polygons in city c in the lth week after lockdown, and zero otherwise. $X_{i,c,t}$ is a vector of time-varying control variables including newly confirmed cases, new deaths, and meteorological variables. The terms u_c and v_k capture the polygon and week fixed effects, respectively, to absorb time-invariant factors at the polygon level and time trends of OMBs

---

[13] http://www.xinhuanet.com/tech/2020-02/28/c_1125639522.htm; accessed May 31, 2020.

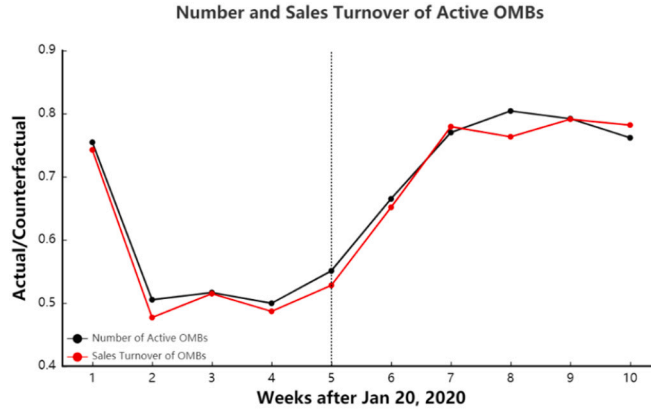**Number and Sales Turnover of Active OMBs**

Fig. 2. Changes in OMB activities.

Note: 1&2. The same as in Fig. 1. 3. The vertical axis is the ratio of the actual number (sales turnover) of active OMBs to their counterfactuals. 4. The vertical dashed line marks the week from which the provincial governments started to revise down the public health emergency response level.

activities.

Fig. 3 plots the estimated coefficients of lockdown variables in the event study regression. We find that the estimated coefficients for the lead terms ($j = -1, -2, -3$) are not significantly different from zero, which implies that there is no systematic difference in the trends between the treatment and control groups before the city lockdown. Therefore, we assume that the parallel trends across two groups would hold in the absence of the lockdown. In addition, we could see that the lagged terms ($l > 0$) are negative and statistically significant. The number of active OMBs had seen drops varying from 4 to 8% during the first seven weeks after lockdown, and the OMBs had seen sales fall by 6 to 11 percentage points in the same period. The average impacts of lockdown became moderate since the seventh week after lockdown, which is in line with the fact that some cities started to lift the lockdown policies then and hopped this would restart the economy.

We then estimate the average impacts of lockdown policies by comparing the average changes of OMBs activities in the treatment group (lockdown cities) relative to the control group (non-lockdown cities). We have seen most cities had repealed their lockdown regulations starting from the seventh week after January 20, 2020, therefore, we keep the dataset spanning from week -3 to week 6, i. e., December 31, 2019 to March 5, 2020, to prevent the impacts of lifting local government regulations. We evaluate the aggregate impacts of lockdown policies following a standard DID regression:

$$OMB_{i,c,k} = \beta_0 + \beta_1 Lockdown_{c,k} + \delta X_{i,c,k} + u_c + v_k + \varepsilon_{i,c,k} \tag{4}$$

where $OMB_{i,c,k}$ is the logarithms of the number or sales turnover of OMBs in polygon i and city c in the kth week. $Lockdown_{c,k}$ takes a value of one for OMBs in each polygon in city c in the kth week after lockdown policies taking into effect, and zero otherwise. $X_{i,c,k}$ is a vector of time-varying control variables the same as that in Eq. (3). The terms u_c and v_k capture the polygon and week fixed effects, respectively.

Table 1 presents the estimated results of Eq. (4). Column (1) and (2) show the estimated impacts on the number of active OMBs. The coefficient of the term *Lockdown* is -0.08 and statistically significant at 1% level. The results indicate that compared with OMBs in cities without formal lockdown policies, the number of active OMBs declined by 8 percentage points when including weather controls, polygon and week fixed effects. We further include the weekly new confirmed cases and deaths in each city to control for the COVID-19 prevalence, and find that the number of active OMBs had seen a drop of 5.9%, suggesting that the changes in the number of active OMBs caused by city lockdown cannot be fully explained by virus itself. Column (3) and (4) report the estimated impacts on the sales turnover of active OMBs. On average, the active OMBs experienced a loss of 11.2% in sales turnover due to lockdown policies. Further, controlling for new confirmed cases and deaths reduces the estimated impact to 7.7%.

The evidence indicates that the estimated impact of government announcements of lockdown policies on OMBs activities is relatively modest, compared with the aggregate collapse of around 50% caused by the COVID-19 outbreak as we estimated in Section 4.1. On the one hand, there are many factors leading to the OMBs' struggle, for example, anxious individuals may engage in physical distancing on their own accord and avoid interacting with others, and the virus containment measures of other cities may also affect the owners' businesses even though their own cities are not locked down. On the other hand, we note that there are caveats in our analysis and we may underestimate the impacts of lockdown policies. In fact, many villages and communities have already imposed strict virus containment measures before the local governments officially announced the lockdown policies, due to anxiety for being held accountable for community transmissions. Then the control groups (non-lockdown cities) actually were locked down partly, although to a very limited degree, however, there is no available dataset that could accurately measure such degree of lockdown policies. To be more conservative, we can treat the estimated effects as the impacts owing to the official lockdown announcements and we thus document that the announcements of government lockdown polies could explain a limited portion of overall decline of OMBs, and what matters most may plausibly be the de facto voluntary containment measures.
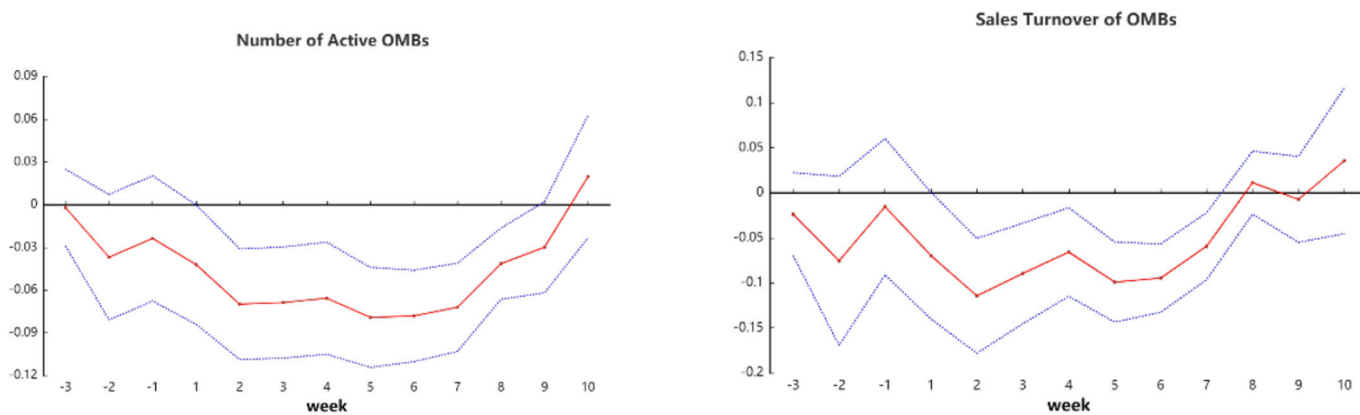
Number of Active OMBs

Sales Turnover of OMBs

Fig. 3. The dynamic evolution of the lockdown effects.
Note: 1&2. The same as in Fig. 1. 3. The OMBs activities between the lockdown cities are compared with the non-lockdown cities, and week 1 is the effective week of lock down policies. 4. The solid lines plot the estimated coefficients in Eq. (3), and the dashed lines plot their 95% confidence intervals.

Table 1
Impacts of lockdown policies.

Dependent variables	Log (number of OMBs)		Log (sales turnover of OMBs)	
	(1)	(2)	(3)	(4)
lockdown	−0.080***	−0.059***	−0.112***	−0.077***
	(0.016)	(0.012)	(0.026)	(0.021)
lagged.case		0.000***		0.000*
		(0.000)		(0.000)
lagged.death		−0.001***		−0.001***
		(0.000)		(0.000)
temperature	−0.008***	−0.009***	−0.000	−0.005
	(0.002)	(0.002)	(0.004)	(0.003)
air pressure	0.006***	0.007***	0.006*	0.010***
	(0.002)	(0.002)	(0.003)	(0.003)
precipitation	−0.000	−0.000	0.001**	0.001**
	(0.000)	(0.000)	(0.000)	(0.000)
humidity	0.001**	0.001*	0.002***	0.002***
	(0.000)	(0.000)	(0.001)	(0.001)
wind speed	0.001	−0.003	0.008	0.009
	(0.005)	(0.005)	(0.012)	(0.011)
Thiessen polygon FE	YES	YES	YES	YES
Week FE	YES	YES	YES	YES
Observations	1,100,007	962,391	1,100,007	962,391
R-squared	0.43	0.42	0.35	0.35

Note: 1. The regression leverages variations in the effective week of lockdown policies across cities. 2. The estimation compares how city lockdown depresses OMBs activities at the polygon level relative to OMBs in the non-locked-down cities in 2020.

4.3. Impacts across urban and rural areas

We move on to study the economic effects on OMBs across urban and rural areas. We classify the polygons into two types of regions based on the classification using raster nighttime lights data in 2019, reflecting the variations in the level of economic activity, with urban areas being the more active. To match the granularity of our data, we rely on nighttime lights and population data to classify the Thiessen polygons into urban and rural areas. The National Oceanic and Atmospheric Administration provides the Suomi National Polar-orbiting Partnership Visible Infrared Imaging Radiometer Suite (VIIRS) nighttime lights data at the 500-meter pixel level, and the Earth Observing System Data and Information System provides the gridded population data at the 1-kilometer pixel level. The VIIRS nighttime lights data have been widely used for the timely and accurate extraction of urban land area in the study of urban economics (Dou, Liu, He, & Yue, 2017; Gibson, Olivia, Boe-Gibson, & Li, 2021). Following the methods suggested by Dou et al. (2017), Henderson, Yeh, Gong, Elvidge, and Baugh (2003), and Shi et al. (2014), we use local-optimized thresholding (LOT) to extract the urban lands. Generally, LOT determines an optimal threshold according to ancillary data (e.g., socioeconomic data, medium- to high-resolution remote sensing data, and so forth) and extracts areas with nighttime light brightness greater than the optimal threshold as urban areas.

We adopt a similar approach to determine the optimal threshold value of built-up urban area extraction as Dou et al. (2017). First, two types of nighttime light images for each city were extracted from the global data sets by using a mask polygon of the administrative boundary. Then, a threshold of the minimum digital number value was used to segment the images into urban areas and non-urban areas. The absolute difference between the extracted area using the VIIRS nighttime lights data and the reference data was recorded. Such processes were iterated by increasing the threshold values until reaching the maximum pixel value of the image. The threshold value that produced minimum difference was selected as the threshold for urban built-up area extraction of the city. We evenly allocate the population raster data at 1-kilometer resolution into 500-meter raster cells, namely, each 1-kilometer raster cell is split into four 500-meter ones. In this way, we accordingly classify the population into two groups living in the above-defined subregions based on the nighttime lights data. In this paper, a polygon may contain many 500-meter grid cells, which could be different subregions. Thus, we suggest classifying the polygons according to the subregion in which most of its population (over 50%) lives. For example, if over 50% of the population in a polygon lives in the urban area, we would classify this polygon as urban area.

Fig. 4 shows that urban areas were hit harder, compared to OMBs in rural areas. The left (right) panel shows the ratio of the number (sales turnover) of active OMBs to its counterfactual for the post-virus period. In the second week, the number of active OMBs fell the most, by 54 and 41% in urban and rural areas, respectively. For sales turnover, the sales of OMBs in rural areas fell by 43%, compared with much sharper contractions in urban areas at 54%. Sales turnover rebounded to 66% in the third week in rural areas, an increase of 9% from the previous week, while sales turnover in urban areas did not start to bounce back until the fifth week. Urban areas bore the hardest hit of the shock, indicating that the great lockdown following the outbreak created greater disruptions in the previously more economically active areas. However, OMB activities relative to their counterfactuals started to converge to a similar level between urban and rural areas in the seventh week, that is, in early March.

There are three possible explanations of the larger disruption to urban areas. First, during the worst period, it is plausible that the urban community exacted more stringent restrictions than the rural villages due to the higher population density and thus higher transmission risks, which directly led to a dramatic drop of the OMBs activities. Second, residential and commercial areas are relatively
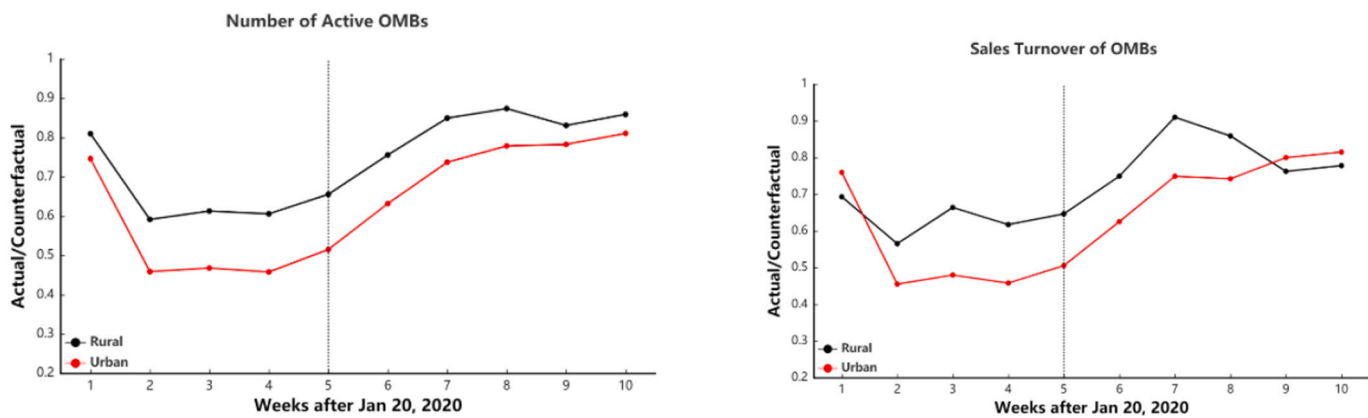
Number of Active OMBs

Sales Turnover of OMBs

Fig. 4. Changes in OMB activities: urban versus rural.
Note: 1–4. The same as in Fig. 2. 5. The two types of regions are classified based on the raster nighttime lights data in 2018, reflecting the variation in the levels of economic activity, with urban areas being the more active.
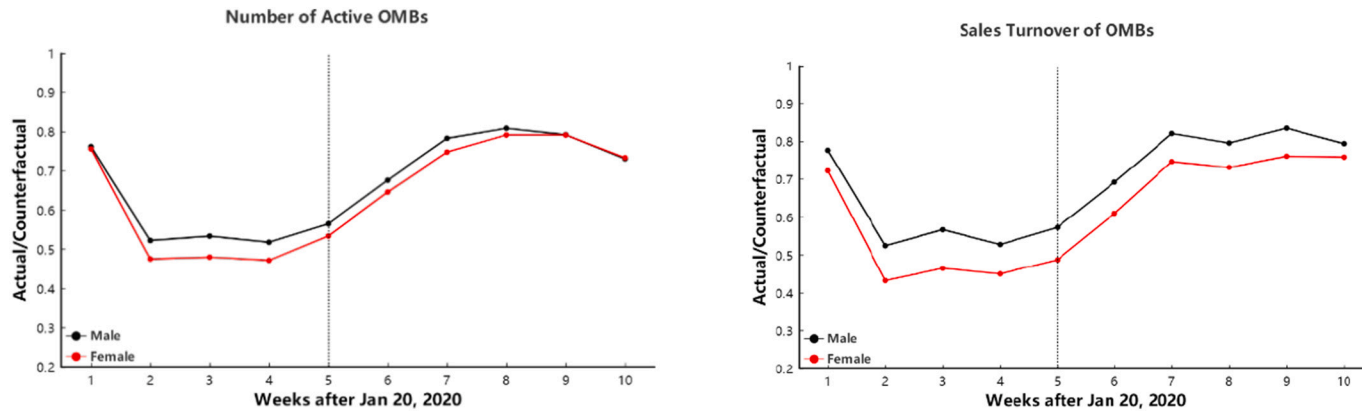
Number of Active OMBs

Sales Turnover of OMBs

Fig. 5. Changes in OMB activities: male versus female.

Note: 1–4. The same as in Fig. 2. 5. The OMBs are classified into two groups by the gender of their owners.
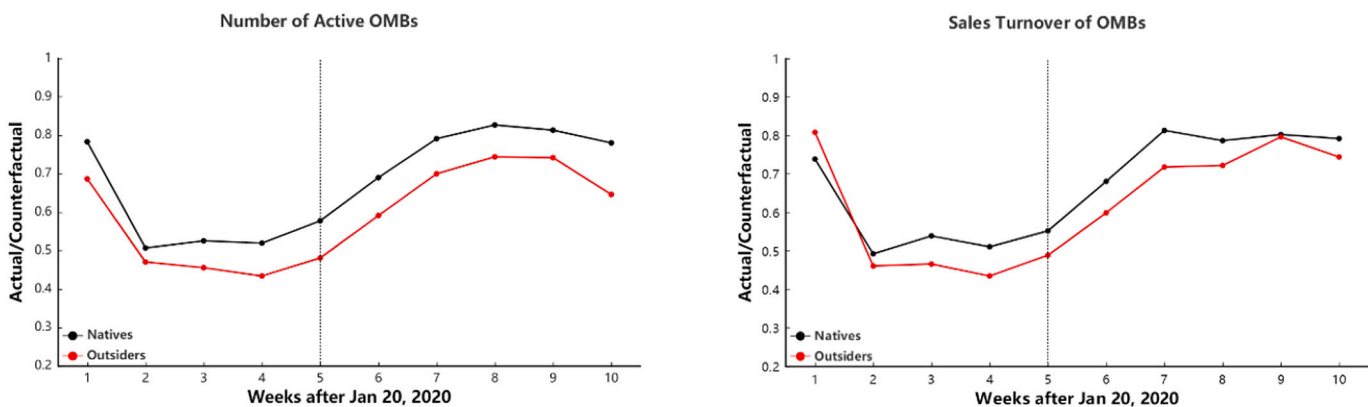
Fig. 6. Changes in OMB activities: outsiders versus natives.
Note: 1–4. The same as in Fig. 2. 5. OMBs are classified into two groups by differentiating whether the owner was born in the place of residence. The owner who was born in the same city where he or she conducted businesses is labelled as native. While the other owners who were not born in the cities where they were managing businesses are outsiders.

separated in most urban regions, while the mobility restrictions within community limited household offline consumption. However, most rural OMBs are in the villages, and households could go out for purchases without violating mobility restrictions. Third, in China, most migrants go back to their hometowns that are mainly rural areas during Lunar New Year Holiday. The consumption needs of migrants that delayed return to work partially mitigates the COVID-19 impact on rural OMBs, compared with the counterfactual consumption need if there were no pandemic. In addition, rural OMBs would have been more willing to start operation due to a relative lower population density in rural villages.

4.4. Impacts by owners demographics

We further estimate the impact for groups of different genders in the subsamples, and study the heterogeneities of impact by differentiating whether the owner was born in the place of residence. We predict the counterfactual number of active OMBs and amount of sales turnover, for male and female owners in each polygon, respectively. We show a graphical display of the ratio of actual activities of OMBs to their counterfactuals, and a ratio of 1 indicates that the actual number and sales turnover of OMBs was at the same level as what it should have been if there were no COVID-19 pandemic, and the difference between the ratio and 1 shows the percentage decline in business activities.

Fig. 5 shows that the female business owners experienced a sharper decline than the male owners, with a weekly average decrease during the trough of about 53 and 57% in the number of merchants and amount of sales turnover, respectively, while the decline for male owners was 48% in both the number of merchants and transaction values during the worst week. This finding suggests that female OMB owners were disproportionately hit by the COVID-19 pandemic, and may live in harder conditions during the weeks when containing the virus spread was obviously the priority. We find that the red line in the right panel of Fig. 5 always lay below the dark line, indicating that the drop of sales turnover of female merchants were always larger than that of male owners. In the end of our sample period, the tenth week after the COVID-19 outbreak, the sales turnover of male merchants rebounded to around 80%, respectively, relative to their counterfactual levels. However, the female merchants recovered 75% of their businesses. There is possibility that the harder hit for female OMB owners may have a long-lasting effect.

We explore the heterogenous impact on different OMBs by differentiating whether the owner was born in the place of residence. For the owner who was born in the same province where he or she conducted businesses, we classify them into a group and label them as *natives*. We label the other owners who were not born in the provinces where they were managing businesses as *outsiders*. For example, if an OMB owner A was born in province B and had businesses in province C, then he or she was labelled as an outsider, but he or she would be a native if having businesses in province B. Given that the OMBs in our sample are all very small in scale and have no branches, every owner conducted business only in one province. Fig. 6 illustrates the estimated decline in the economic activities of OMBs for natives and outsiders. The decrease in economic activities were larger for the outsiders, with the number of active merchants at 43% and sales turnover at 51% of their counterfactual levels during the trough period. In the worst week, the drop in number of active outsiders was 7 percentage points larger than that of active natives. We find that the OMB activities relative to their counterfactuals were nearly at a similar level between outsiders and natives in the tenth week after COVID-19 break, one month after the work resumption, although the number of active outsiders was still far from the normal level, which reflect a positive effect of lifted restrictions but a relatively lagged return of migrant workers.

5. Conclusions and implications

We suggest that obtaining the accurate statistics about the informal economy can be possible by taking full advantage of digitalization of economy and the widespread of mobile payments. Using transaction-level business data of around 80 million OMBs from the largest Fintech company in China and the machine learning technique to predict the counterfactual path of OMBs activities in 2020, we have provided a first view of vulnerable informal workers after the blows from COVID-19. The tens of millions of OMBs in China work largely in the informal services sector of the economy, mainly in micro and family-based enterprises, and employ low-skilled workers who mostly could not work from home during the COVID-19 pandemic. Many OMBs rely heavily on cash flows and short-term loans due to limited savings. If they cannot work for an extended period of time, the whole family would be at risk. Effective policies are needed to ease the pain felt by vulnerable OMBs in the gig economy and mitigate the potential poverty and inequality impacts.

We find that the number and sales of OMBs experienced immediate and dramatic collapse, with the biggest weekly contraction of around 50%, while the decline due to lockdown policy was modest and negligible. OMBs in urban areas experienced a sharper contraction during the trough, with a weekly average decrease of around 54% in the number and sales turnover of active merchants in the worst week, compared with the drops of 41 and 43% in the number and sales turnover of active merchants for rural OMBs, respectively. Female merchants were hit harder than the male merchants, with drops of 5 and 9% greater in the number of active merchants and sales turnover during the trough, respectively, than the drops of male owners. We have also seen that the business owners who were not born in the province where they conducted businesses were disproportionately disrupted, and they are often migrants in the city without social insurance. In short, we find that the most vulnerable workers in the gig economy were hit hard by the COVID-19 pandemic. Therefore, we suggest an inclusive policy response to mitigate the impact of the crisis and support this vulnerable group. We should pay special attention to the potential urban poor that may be hard to make a living.

The economic activities of OMBs plummeted during the whole month of February, before starting a sharp rebound at the end of February. The businesses had rebounded to around 80% of where they should be seven weeks after the COVID-19 outbreak, hovering at this level for additional three weeks until the end of our sample period. The quick recovery of OMBs since the nationwide encouragement of work resumption provides evidence of the necessity of prioritizing containment of the virus and the importance of

government support in reopening the economy. Although the informal businesses are small and vulnerable, a sharp recovery is likely once the spread of the virus is largely or totally under control. However, the pain felt by informal workers is real, and we see a bottleneck for further rebound and an unbalanced recovery for different groups. The negative effects on the vulnerable informal workers as a whole, especially on the female and migrants group, are likely to be long-lasting, thus we suggest a more continuous policy response to ensure adequate support for the most vulnerable at a relatively longer-term amid the new normal of epidemic prevention and control.

OMBs' quick recovery would not have been possible without significant policy support. For example, as of February 12, 2020, at least 25 provinces in China rolled out up to 90 measures to support small and medium-size enterprises, among which the most frequently mentioned measures included support for reopening, delay in fee collection (electricity, water, and gas), tax payment delay, rent or tax deduction, delay or refund of social security contributions, lower financing costs, and strengthened financing support. We caution that these are short-term responses and partial analysis, only months after the outbreak of the virus; however, we highlight the importance of taking full advantage of digital development to measure the size of and follow up the economic activities of the informal sector.

Appendix A. Data aggregation method

There are around 80 million offline micro business (OMB) merchants in our data set. With the data privacy issue in mind, we would aggregate the OMB data at an appropriate geographical level. The widely-used delineation based on administrative units, census tracks, and other established areas leads to a great loss of information in our large sample. For example, it would be unfortunate if we assigned 80 million OMB merchants into 2800 counties. A principle of the proposed aggregation method is to start at the lowest possible geographical level and the area could adjust adaptively according to local business activities, namely, the tracts should be densely distributed in the areas with higher population density and level of economic activity, and vice versa. This avoids a substantial increase in the number of initial territorial units, like unified small grid cells, and ensures a closer relationship between the number of OMBs and the characteristics of economic development, population, and geography in every single tract. This ensures keeping information at the granular data to a large extent.

We propose to create Thiessen polygons and assign the OMBs into each polygon. Thiessen polygons, otherwise known as Voronoi diagrams, are an essential method for the analysis of proximity and neighborhood, for example, airtime transfers and mobile communications (Blumenstock et al., 2016), police enforcement and spatial distribution of crime (Piza & Gilchrist, 2018), determinants of HIV infection (Bertocchi & Dimico, 2019), and traffic crashes (Jia, Khadka, & Kim, 2018). The method defines an area around a center point, where every location is nearer to this point than to all the others. In our analysis, each OMB belonging to a specific Thiessen polygon is the closest to its own center point, compared with the distance to any other center point. A key to the delineation of Thiessen polygons is to determine the center point of each polygon. We propose to use bank branches (including self-service branches) as the center of the polygons. On the one hand, the distribution of bank branches is in line with the proposed principle of aggregation, namely, the branches are densely dispersed in areas with active businesses and economies. On the other hand, the bank branches almost cover all the places across the country[14] and, given the importance of financial services in business activities, it is reasonable to use bank branches as the center points of the polygons for the data aggregation.

We merge the bank branches within each 500-meter grid cell into one as the center point, by taking the average of their geographic coordinates,[15] and establish 138,629 polygons around these points across mainland China. Fig. A.1 shows an illustration for Chaoyang district in Beijing after being partitioned into 437 polygons, with each red dot representing a bank branch and the lines the boundaries of the polygons. Chaoyang district provides a great visualization of the flexibility of the area of the Thiessen polygon. Specifically, the west of Chaoyang district is a commercial area, with the well-known Sanlitun Street and Central Business District (CBD) being here. CBD is one of the most dynamic business locations in Beijing, where many headquarters of Chinese and foreign companies chose to settle. We see a very dense distribution of the polygons on the west side of the schematic. The east of Chaoyang district largely lies outside the 3rd Ring Road in Beijing, includes many green areas and parks, and is less business-active. The polygons on the east of the schematic are sparsely scattered. The comparison clearly illustrates the advantage of the Thiessen polygons in adjusting the area for

---

[14] There are no bank branches in only two counties in the 31 provinces in mainland China; they are Shuanghu and Shenzha in Tibet, which are also known as Changtang, whose average elevation is of 5000 m. The location information of the branches is obtained from AutoNavi Application Programming Interface, accessed in December 2019.

[15] According to Senechal (1993), it is meaningless and unreasonable to establish Thiessen polygons based on spatial distances within tens of meters, and the bank branches would be too densely distributed in some bustling urban areas. We merge the branches within 500-m grids for accuracy and reasonability.

aggregating the individual samples adapting to local business activities. We later aggregate the millions of OMBs at the polygon level, and match the other economic, population, and geographic characteristics accordingly.
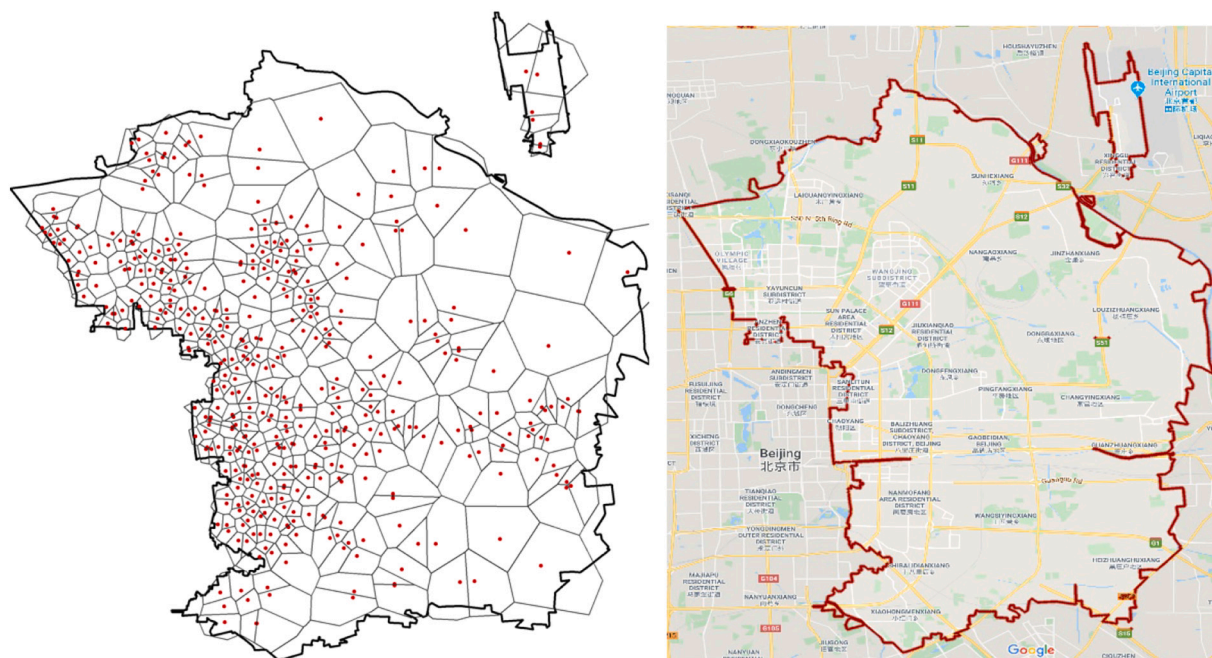

Fig. A.1. Thiessen polygons: Chaoyang District, Beijing.
Note: 1. The left panel shows an illustration of Chaoyang district, Beijing, after being partitioned into 437 polygons, with each red dot representing a bank branch and the lines boundaries for the polygons. 2. The right panel is taken from Google Maps.

Appendix B. Data description

B.1. Offline micro businesses data

We employ weekly data on around 80 million QR code merchants (or *Mashang* in Chinese), aggregated at the Thiessen polygon level over December 31, 2019 to April 2, 2020 and for the corresponding lunar calendar dates in 2018 and 2019 from Ant Financial, the largest Fintech company in China. Table B.1 shows a full picture of the time window in our analysis.

We show four critical facts about the QR code merchants, the offline micro businesses (OMBs) that adopt the digital payment solution. First, they are offline micro businesses that accept customer payments via an Alipay QR code specialized for merchants, including businesses that are registered or unregistered with the Industry and Commerce Department. Second, they are mostly self-employed and family-based businesses. Only individuals but not legal entities may apply for the QR code for merchants, which coincides with the characteristics of self-employed OMBs that hardly distinguish between the owner and business entity. This excludes the large enterprises that operate and settle in the name of a company, such as large supermarkets, chain stores, restaurants, and so forth, from our sample. Third, they are largely very micro businesses. The average annual revenue was less than 400,000 yuan (about US$56,000) for 99% of these merchants in 2019. According to the classification of the Ministry of Industry and Information Technology, retail businesses with annual revenue less than 1 million yuan are small enterprises. Fourth, they are quite representative for the OMBs in China. The Alipay QR code of merchants is specialized for the OMBs, because accounts for merchants have lower withdrawal fees than personal accounts, and it takes only several minutes to apply for the account and associated QR code with easy taps. Thus, OMBs have strong incentives to use the QR code solution to accept payments, especially since mobile payment has become the first choice in daily transactions for most Chinese people.

Table B.1
Time window.

Week	Start	End	Start	End	Start	End
-3	2019-12-31	2020-01-06	2019-01-11	2019-01-17	2018-01-22	2018-01-28
-2	2020-01-07	2020-01-13	2019-01-18	2019-01-24	2018-01-29	2018-02-04
-1	2020-01-14	2020-01-20	2019-01-25	2019-01-31	2018-02-05	2018-02-11
1	2020-01-21	2020-01-30	2019-02-01	2019-02-10	2018-02-12	2018-02-21

<div align="right">(continued on next page)</div>

transportation convenience at the polygon level. To merge these variables with the OMBs' activities at the polygon level on a weekly basis, we first calculate the average of the variables each week in the window of analysis for each polygon, and then match the variables to the weekly average of OMBs' activities based on the polygon code. Table B.2 presents the weekly summary statistics of these variables. We divide the sample in each year into two periods: the pre- and the post-outbreak periods. Panel A reports the summary statistics for the OMBs' activities and meteorological variables at the polygon level. Relative to the pre-period, during the 10 weeks after the event data, the number of active OMBs fell by 11.28% in 2018, 10.38% in 2019, and 33.20% in 2020, and the sales turnover dropped by 1.41% in 2018, 1.31% in 2019, and 3.35% in 2020. The overall decrease in the post-virus period reflects the seasonality around the Lunar New Year; however, we see a relatively sharper decline in OMBs' activities in 2020. The average contraction of OMBs' activities relative to the pre-virus period in the 10-week period since the outbreak does not directly tell us the causal impact of the pandemic; for this, we would explore the causal collapse of the OMBs' activities in the post-virus period on a weekly basis.

Table B.2
Summary statistics.

Panel A: Weekly OMB activity and meteorological variables

	2018			2019			2020		
	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.
Number of merchants									
Pre-outbreak	689,960	100.00	174.05	691,174	135.01	224.91	691,134	120.96	208.98
Post-outbreak	1,240,279	88.72	163.77	1,243,484	121.00	216.12	1,243,370	80.80	151.51
Log (Sales turnover of merchants)									
Pre-outbreak	689,960	100.00	11.82	691,174	104.34	10.23	691,134	105.85	9.76
Post-outbreak	1,240,279	98.59	12.67	1,243,484	103.03	10.73	1,243,370	102.50	9.53
Weekly average humidity (%)									
Pre-outbreak	693,145	62.31	17.59	693,145	66.16	18.50	693,145	74.29	11.01
Post-outbreak	1,247,661	64.32	15.31	1,247,661	64.63	19.94	1,247,661	66.74	15.17
Weekly accumulative precipitation (mm)									
Pre-outbreak	693,145	5.10	9.83	693,145	6.67	12.63	693,145	11.91	20.69
Post-outbreak	1,247,661	14.39	20.02	1,247,661	14.98	25.89	1,247,661	11.24	17.43
Weekly average air pressure (hPa)									
Pre-outbreak	693,145	985.75	60.25	693,145	986.21	60.04	693,145	986.78	58.49
Post-outbreak	1,247,661	978.81	58.30	1,247,661	980.41	58.98	1,247,661	983.48	57.70
Weekly average temperature (°C)									
Pre-outbreak	693,145	1.66	8.83	693,145	3.31	7.96	693,145	3.68	8.52
Post-outbreak	1,247,661	13.00	6.65	1,247,661	10.07	6.90	1,247,661	8.98	7.47
Weekly average wind speeds (m/s)									
Pre-outbreak	693,145	2.90	1.18	693,145	2.44	1.05	693,145	2.57	1.00
Post-outbreak	1,247,661	3.32	1.23	1,247,661	2.61	1.08	1,247,661	3.02	1.08

Panel B POI and other time-invariant variables in the window of analysis

	N	Mean	Std. Dev.
Number of POI related to			
Auto services	138,629	7.41	12.32
Auto dealers	138,629	1.37	5.05
Auto repair	138,629	3.23	5.91
Motorcycle service	138,629	1.10	2.18
Food and beverages	138,629	32.86	38.69
Shopping	138,629	79.56	97.57
Daily life service	138,629	47.96	53.92
Sports and recreation	138,629	4.69	7.21
Medical service	138,629	11.08	13.52
Accommodation service	138,629	5.98	15.84
Tourist attraction	138,629	2.07	4.77
Commercial house	138,629	5.61	11.96
Governmental organization and social group	138,629	14.44	13.36
Science/culture and education services	138,629	9.88	11.36
Transportation service	138,629	13.88	17.01
Finance and insurance service	138,629	4.06	3.95
Enterprises	138,629	6.34	9.06
Public facility	138,629	2.98	4.82

Other economic, demographic, and geographic characteristics

Table B.2 (*continued*)

Panel B POI and other time-invariant variables in the window of analysis

	N	Mean	Std. Dev.
Sum of the nighttime lights	138,629	123.92	212.57
Population	138,629	10,143.55	11,899.89
Proportion of population in downtown areas	138,629	0.28	0.42
Proportion of population in urban areas	138,629	0.16	0.30
Proportion of population in suburbs and rural areas	138,629	0.15	0.27
Driving distance from the center of the polygon to that of the county (meters, take logarithm)	138,629	8.78	1.32
Driving distance from the center of the polygon to that of the prefecture-level city (meters, take logarithm)	138,629	10.24	1.33
Driving distance from the center of the polygon to that of the capital city of the province (meters, take logarithm)	138,629	10.24	1.33
Average altitude (meters)	138,629	338.35	563.93

Note: 1. The number and sales turnover of OMBs are scaled by the average of the first week in the time window (week -4) of 2018. 2. The pre-outbreak period covers the three weeks before January 20, 2020, while the post-outbreak period is the 10 weeks after, and the pre- and post-outbreak periods in 2018 and 2019 are defined corresponding to the lunar calendar. 3. The meteorological variables are merged with OMB activities at the polygon level on a weekly basis. 4. The POI data were collected from the AutoNavi Application Programming Interface (API) in December 2019 and merged with OMBs at the polygon level. 5. We aggregated the nighttime lights, demographic, and geographic characteristics at the 500-meter resolution to match the Thiessen polygons of which the center point was chosen at 500-meter grid cells.

Data sources: Ant Group; The National Meteorological Information Center of China; AutoNavi API; The National Oceanic and Atmospheric Administration; The Earth Observing System Data and Information System.

Appendix C. Prediction based on a gradient boosting decision tree

In this section, we clarify the details of prediction on the counterfactual path. The gradient boosting decision tree (GBDT), first proposed in Friedman (2001), is one of the typical algorithms of machine learning (Chen & Guestrin, 2016) and has a wide range of academic applications. It has been widely used recently mainly due to its high accuracy and fast training and prediction speed. Here are three specifics worth noting in the process of predicting the OMBs' activity. First, we train the GBDT model at the Thiessen polygon level in 2019. We predict the weekly number and sales of active OMBs in a 10-week time window in the post-virus period; thus, we need to train a total of 20 (=10 * 2) models for each response variable in each week, with the feature variable list including 56 variables. Second, 10% of the data set is selected as the test set, then we split the remaining data into five sets, and the model was optimized by five-fold cross validation before the optimized hyperparameters were selected. Next, according to the optimized parameters, we establish the prediction model on OMBs' activity and carry out the prediction analysis on the test set. We evaluate the generalization ability of the model by calculating the coefficient R^2 on the test set, which has never been seen by the model, and the prediction value of the algorithm is explained to the target number and sales turnover of OMBs. The larger is the R^2, the higher is the predicted accuracy of OMBs' activity, and the stronger is the generalization ability. The R^2 for the models in predicting OMBs' activity are above 0.90. Finally, we update the selected hyperparameter sets using the whole data set for 2018 and 2019 and establish the prediction model for the post-virus counterfactual path in 2020 based on the final updated parameters. See the Tables C.1 and C.2 for more details on the hyperparameters and performance of the model.

Table C.1
Hyperparameter sets of the GBDT model.

Hyperparameters	Model: Number of OMBs	Model: Sales turnover of OMBs
N_estimator	400	300
Maximum number of terminal nodes or leaves in a tree	32	32
Minimum samples required in a terminal node or leaf	10	40
Learning rate	0.1	0.05
Maximum depth of a tree	10	10
Subsample	0.6	0.6

Table C.2
Performance of the GBDT model.

The # week since Jan. 20	Model: Number of OMBs		Model: Sales turnover of OMBs	
	R^2	MAE	R^2	MAE
1	0.967	11.87	0.934	0.287
2	0.979	11.56	0.927	0.300
3	0.903	13.72	0.927	0.308
4	0.933	12.68	0.926	0.313
5	0.905	13.70	0.933	0.295
6	0.934	12.86	0.929	0.306
7	0.954	12.55	0.927	0.315
8	0.935	13.44	0.924	0.321

Table C.2 (*continued*)

The # week since Jan. 20	Model: Number of OMBs		Model: Sales turnover of OMBs	
	R^2	MAE	R^2	MAE
9	0.931	14.37	0.929	0.293
10	0.917	13.56	0.926	0.298

Note: 1. Performance is measured on a test data set, which has never been used in the process of training. 2. We establish a model for each week in the post-virus period, but they share the same super-parameter setting. MAE = mean absolute error.

References

Athey, S. (2017). Beyond prediction: Using big data for policy problems. *Science, 355*, 483–485.

Atkeson, A. (2020). *What will be the economic impact of Covid-19 in the US? Rough estimates of disease scenarios*. National Bureau of Economic Research.

Baker, S. R., Farrokhnia, R. A., Meyer, S., Pagel, M., & Yannelis, C. (2020). *How does household spending respond to an epidemic? Consumption during the 2020 Covid-19 pandemic*. National Bureau of Economic Research.

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

Bertocchi, G., & Dimico, A. (2019). The long-term determinants of female HIV infection in Africa: The slave trade, polygyny, and sexual behavior. *Journal of Development Economics, 140*, 90–105.

Blackburn, K., Bose, N., & Capasso, S. (2012). Tax evasion, the underground economy and financial development. *Journal of Economic Behavior & Organization, 83*, 243–253.

Blumenstock, J. E., Eagle, N., & Fafchamps, M. (2016). Airtime transfers and mobile communications: Evidence in the aftermath of natural disasters. *Journal of Development Economics, 120*, 157–181.

Capasso, S., & Jappelli, T. (2013). Financial development and the underground economy. *Journal of Development Economics, 101*, 167–178.

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

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Presented at the Proceedings of the 22Nd ACM SIGKDD international conference on knowledge discovery and data mining, San Francisco, California, USA* (pp. 785–794). https://doi.org/10.1145/2939672.2939785

Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., Piontti, A. P., Mu, K., Rossi, L., Sun, K., et al. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science, 368*, 395–400.

Dai, R., Mookherjee, D., Quan, Y., & Zhang, X. (2020). *Industrial clusters, networks and resilience to the Covid-19 shock in China*. National Bureau of Economic Research.

Dou, Y., Liu, Z., He, C., & Yue, H. (2017). Urban land extraction using VIIRS nighttime light data: An evaluation of three popular methods. *Remote Sensing, 9*, 175.

Eichenbaum, M. S., Rebelo, S., & Trabandt, M. (2020). *The macroeconomics of epidemics*. National Bureau of Economic Research.

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

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 1189–1232.

Gibson, J., Olivia, S., Boe-Gibson, G., & Li, C. (2021). Which night lights data should we use in economics, and where? *Journal of Development Economics, 102602*.

He, G., Pan, Y., & Tanaka, T. (2020). The short-term impacts of COVID-19 lockdown on urban air pollution in China. *Nature Sustainability, 3*, 1005–1011.

Henderson, M., Yeh, E. T., Gong, P., Elvidge, C., & Baugh, K. (2003). Validation of urban boundaries derived from global night-time satellite imagery. *International Journal of Remote Sensing, 24*, 595–609.

Jia, R., Khadka, A., & Kim, I. (2018). Traffic crash analysis with point-of-interest spatial clustering. *Accident Analysis & Prevention, 121*, 223–230.

La Porta, R., & Shleifer, A. (2014). Informality and development. *Journal of Economic Perspectives, 28*, 109–126.

Lustig, N., Stone, S., & Tommasi, M. (2020). *COVID-19 and social protection of poor and vulnerable groups in lain America: A conceptual framework* (UNDP Policy Documents Series).

Maloney, W. F. (2004). Informality revisited. *World Development, 32*, 1159–1178.

Piza, E. L., & Gilchrist, A. M. (2018). Measuring the effect heterogeneity of police enforcement actions across spatial contexts. *Journal of Criminal Justice, 54*, 76–87.

Rossi, A. G., & Utkus, S. P. (2020). *Who benefits from robo-advising? Evidence from machine learning (No. 3552671)*. SSRN scholarly paper. Rochester, NY: Social Science Research Network.

Senechal, M. (1993). Spatial tessellations: Concepts and applications of voronoi diagrams. *Science, 260*, 1170–1173.

Shi, K., Huang, C., Yu, B., Yin, B., Huang, Y., & Wu, J. (2014). Evaluation of NPP-VIIRS night-time light composite data for extracting built-up urban areas. *Remote Sensing Letters, 5*, 358–366.

Tian, H., Liu, Y., Li, Y., Wu, C.-H., Chen, B., Kraemer, M. U., Li, B., Cai, J., Xu, B., Yang, Q., et al. (2020). An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science, 368*, 638–642.