INDUSTRIAL CLUSTERS, NETWORKS AND RESILIENCE TO THE COVID-19 SHOCK IN CHINA*

Ruochen Dai† Dilip Mookherjee‡ Yingyue Quan§ Xiaobo Zhang¶

October 21, 2020

Abstract

We examine how exposure of Chinese firms to the Covid-19 shock varied with a cluster index (measuring spatial agglomeration of firms in related industries) at the county level. Two data sources are used: entry flows of newly registered firms in the entire country, and an entrepreneur survey regarding operation of existing firms. Both show greater resilience in counties with a higher cluster index, after controlling for industry dummies and local infection rates, besides county and time dummies in the entry data. Reliance of clusters on informal entrepreneur hometown networks and closer proximity to suppliers and customers help explain these findings.

Keywords. Clusters, Covid-19, China, Firms, Social Networks.

JEL. J12. J16. D31. I3.

*We are grateful to seminar participants and an EDI referee for their constructive comments. Research support from the Economic Development and Institutions (EDI) and from China Natural Science Foundation (#71874008) is gratefully acknowledged. We are responsible for any errors that may remain.

†Central University of Finance and Economics, r.dai@cufe.edu.cn

‡Boston University, dilipm@bu.edu

§Peking University, yyquan2018@nsd.pku.edu.cn

¶Peking University and IFPRI, x.zhang@nsd.pku.edu.cn

1 Introduction

The importance of firm clusters in industrial organization has been noted by many scholars, going back to Marshall (1920).The standard definition refers to spatial agglomeration of firms in a common industry to realize inter-firm spillovers in sharing of technology, inputs and customers. Clusters have played an important role in industrial development in the early 20th century in both the UK and the USA, and continue to play an important role (e.g. in the Michigan auto industry and California IT sector). They also play a prominent role in many less developed countries (LDCs) in Asia and Africa, though with some distinctive characteristics from developed country counterparts: small firm size, low capital intensity, a high degree of vertical disintegration and specialization in different stages of production, strong buyer-seller networks across stages of production, prevalence of trade credit, sharing of tools and information. Inter-firm exchanges within the cluster are governed by informal relational contracts rather than formal, legally enforced contracts. Entrepreneurs often belong to a common social network or community, defined by ethnicity or birthplace, allowing informal agreements to be enforced via community norms. The importance of many of these features such as prevalence of relational contracts, trade credit and community enforcement in LDCs have been studied by many authors (Kali (1999), Kranton (1996), McMillan and Woodruff (1999a,b), Banerjee and Munshi (2004), Machiavello and Morjaria (2015, 2020), Dai et al (2020b)). Other papers in the theoretical networks literature have provided interesting insights into the role of buyer-seller networks in coping with demand and supply shocks (Kranton and Minehart (2000, 2001)).

In China, most major industrial clusters are located close to marketplaces for final products and intermediate goods where customers from far and near come to make purchases, reducing the need for firms to carry large inventories e.g., the Zhili childrens garment cluster (Fleisher et al., 2010), or the Puyuan cashmere sweater cluster (Ruan and Zhang, 2009). The fine division of labor within clusters reduce capital barriers to entry (Ruan and Zhang, 2009). In addition, the prevalence of trade credit and flexible payment arrangements ameliorate working capital constraints facing SMEs (Long and Zhang, 2011; Merima, Peerlings, and Zhang, 2014). As shown by Dai et al (2020b) using Chinese firm registration data from 1990-2009, entrepreneurs from common hometown networks with high levels of informal trust and local cooperation (proxied by population density), have succeeded in achieving higher rates of firm entry, concentrated in fewer sectors and locations, entered with smaller firms which subsequently grew at faster rates. Besides China, clusters are widespread in other Asian countries (Sonobe and Otsuka, 2006) and in Africa (World Bank, 2011) and share similar features. Other examples include the Tirupur garment industry cluster in South India (Banerjee and Munshi 2004), aquaculture clusters in Bangladesh (Zhang, Chen, and Fang, 2019) and handloom clusters in Ethiopia (Zhang, Moorman, and Ayele, 2011).

As mentioned above, LDC firm clusters are contrasted to forms of industrial organization more common in high income countries or in multinational corporations (MNCs) — characterized by larger firm size, greater vertical integration, capital intensity, distance from suppliers/customers, and reliance on formal market contracts rather than informal networks. However, many LDCs exhibit dualism, or prevalence of the two polar forms of industrial organization. Firm size distributions tend to feature a thick bottom tail representing large numbers of small firms (including clusters) mainly serving the domestic market, and medium to large size firms featuring greater vertical integration, capital and export intensity. Comparisons of productivity and growth across different categories of firms has been the topic of a large recent literature, stemming from the findings of Hsieh and Klenow (2009) of high misallocation (dispersion of marginal revenue products) across firms in China and India compared with the US. This literature has focused mainly on comparisons of firm size, productivity and growth. Less is known about comparative assessments of vulnerability to external risk which comprise an additional dimension of performance, or mechanisms explaining the origin and persistence of small firms with low productivity. This constitutes the primary motivation of this paper.

We use firm data from China to assess the relative resilience of clusters with respect to the recent Covid-19 shock. The shock arrived on the eve of the Chinese New Year in the middle of February 2020, resulting in a severe lockdown in some parts of the country with high infection rates (Fang et al 2020), and restrictions on mobility between other parts and with the outside world. The pandemic eased by early April, resulting in a gradual lifting of the mobility restrictions thereafter. While the Covid-19 infection rates may have directly impacted some entrepreneurs and workers who fell ill, the mobility restrictions imposed on the rest of the population and on the movement of goods were of potentially greater significance. Chinese firms rely to a considerable extent on entrepreneurs and workers who have migrated from their hometowns to the place where the firm is located. Many of them had gone back to their hometowns for the New Year celebrations and were unable to return to their place of work until the lockdown restrictions were eased. Moreover, the movement of inputs supplies and goods to the market was impeded, as well as the volume of imports and exports, resulting in significant supply and demand shocks faced by firms. As we show in the paper, in the latter half of February there was a sharp (approximately 70%) reduction in entry flows of new firms compared to entry rates at the same time in previous years, and an effect of the same order of magnitude on the number of incumbent firms reopening after the New Year.

Our primary empirical finding is that the Covid-19 impact was significantly lower in counties/industries exhibiting a higher degree of clustering, after controlling for industry and time (month and year) dummies, as well as county dummies in the case of new entry rates. Cluster measures change very little over time, hence the results are robust to cluster measures based on firm censuses in 2008, 2004 and 1995. Counties with an above median cluster index featured a 67% reduction in entry rate during the month immediately following the Chinese New Year compared to previous years, compared to a 74% reduction in counties with below median cluster index. A one s.d. increase in the cluster index was associated with a 12% rise in the entry rate. These results are robust to alternative specifications (at the weekly rather than monthly level) and controls for local infection rates. The results concerning impacts on new entry also appear on functioning of incumbent firms: a 1% rise in the cluster index was associated with a .05-.07% higher likelihood of reopening in February after the New Year, and a .03-.04% higher likelihood in May, after controlling for local infection rates and industry dummies. We find evidence of a direct adverse impact of Covid-19 infections: higher local infection rates were associated with lower entry rates and reopening likelihood among incumbents. But despite the higher infection rates in counties with higher clustering, they were less adversely impacted overall.

The remainder of the paper seeks to disentangle the role of different attributes of clusters in affecting resilience. In particular we test hypotheses suggested by findings of Dai et al (2020b) concerning the role of entrepreneur hometown networks in the emergence and functioning of production clusters. Broadly speaking, their theory is based on the notion that poorly functioning markets for credit and technology and weak institutions for formal contract enforcement in LDCs inhibit private entrepreneurship. These barriers are overcome partly by informal cooperation among social networks of entrepreneurs. In the Chinese context the networks are based on a common birthplace or hometown.

Those who have become entrepreneurs in a particular industry and location help provide information and assistance to others in their hometown to enable them to enter the same sector and location. Entrepreneurs functioning in a given sector-location from the same hometown share credit, risks, infrastructure, technology, supplier and customer lists based on informal agreements. Owing to credit market imperfections, they enter with firms of small size in a common location, specialize in different stages of production of a particular end-product, and engage in intensive buyer-seller relationships. Firms at a given stage of production share customer orders and fixed capital overheads flexibly. Such forms of mutual assistance raise productivity and lower costs of capital of network member firms. Entry barriers are lowered, thereby attracting greater inflows of entrepreneurs from the hometown. Higher density networks tend to concentrate more in specific industries and locations, owing to the greater lock-in induced by stronger spillovers. The network spillovers also induce a form of adverse selection of entrepreneurial quality, manifested by greater dispersion (and possibly lower average) of quality, productivity and size of entering firms.[1]

Dai et al (2020b) confirm these predictions empirically in the SAIC firm registration data over 1990-2009, using 1982 population density of the hometown as a proxy for network quality. They justify use of this proxy measure by showing that for rural counties, informal trust, social interactions, and patterns of cross-participation of entrepreneurs in each others firms, are all rising in local population density (after controlling for population size, education and occupational patterns). Moreover, population density changes little over time: the 1982 density is highly correlated with density in later decades. Since significant restrictions on movement of people were still in place in 1982 when the market based economy was just beginning to emerge, the 1982 population density can be reasonably treated as a predetermined parameter for any given hometown.

Section 2 extends the Dai et al (2020b) model to incorporate the effect of an unanticipated Covid-19 shock which raises factor costs owing to factor shortages arising from mobility restrictions and

[1]This may not be obvious at first glance; see their paper or Section 2 here for more details. While the entry threshold for entrepreneurial quality would be lower in a higher quality network, this would be compensated by a higher productivity spillover from other firms in the network. However, the net effect would be negative, since the marginal entrepreneur in the network must be indifferent between entering and not, and the outside option would be lower owing to the lower entry threshold. Hence the net effect on the marginal entrepreneur's productivity would be negative. Dai et al (2020b) show that with a log uniform distribution over entrepreneurial quality in the population, the average productivity of incumbents would also be lower in a higher quality network. This is what they actually find in the data, using observed firm size as a proxy for productivity.

infections. We assume that the severity of the factor shortages is lower in a high density network owing to superior supply reliability and sharing of capital and risks. There are two principal predictions. First, entry flows in higher density networks will be less adversely affected. On the other hand, it is possible that a larger fraction of incumbent firms in a higher density network shutdown following the Covid-19 shock. This owes to the adverse selection effect: productivity is more widely dispersed among incumbents in a higher density network, with a larger lower tail that is more vulnerable to a sudden rise in factor prices.

These predictions turn out to be upheld by the data. We first verify that clusters are positively correlated with hometown density. We then show evidence that counties with higher hometown density experienced associated a smaller Covid-19 impact on rates of registration of new firms. On the other hand, they also experienced a lower likelihood of reopening among incumbents. These results are robust to controls for clustering, hometown heterogeneity and local infection rates.

However, higher hometown density alone cannot explain the observed differences between the high and low cluster areas. Neither can variations in the Covid-19 infection rates: areas with a higher degree of clustering featured higher infection rates, lowering both entry and the reopening likelihood of incumbents. Using the ESIEC survey data, we subsequently examine the role of other firm attributes associated with clusters. Controlling for hometown density and heterogeneity, greater spatial proximity to suppliers and buyers (including online sales) was associated with a higher re-opening rate among incumbents. Overall, the results therefore suggest that both network quality and spatial agglomeration played some part in explaining the superior resilience of clusters.

Section 2 presents the theoretical model. Section 3 then provides details of the data, the cluster index, firm attributes correlated with clustering, and measures of entrepreneur network quality, along with relevant descriptive statistics. Section 4 presents the main result concerning resilience of areas with greater clustering. Section 5 then examines the role of different attributes of clusters: quality of entrepreneur hometown networks, spatial agglomeration and other related firm characteristics. Section 6 describes how the results varied across four main industry groups, while Section 7 concludes. The Data Appendix provides details of variables used in the analysis.

2 Theory: Network Density and Covid-19 Impact

We use the same notation and assumptions as Section 3 in Dai et al (2020b). Consider a hometown with population density p which is a proxy for the level of trust and informal cooperation among its residents. There are different cohorts of new agents $t = 1, 2, ..$ of equal size, in each of which latent entrepreneurial talent ω is drawn independently from a log-uniform distribution on the unit interval. Each agent makes a once-and-for-all occupational decision on whether to select a traditional occupation and earn ω^σ for ever, where $\sigma \in (0, 1)$. A fixed fraction $k \in (0, 1)$ receive an opportunity to join a network of older incumbents from the same hometown who are operating in some destination. A fraction $s_{i,t-1}$ of incumbent entrepreneurs from this town are distributed across different destinations i (sector-location pairs), cumulating upto cohort $t - 1$. $s_{i,t-1}$ also represents the fraction of destination i offers arriving among the new cohort t agents, reflecting a social process of contacts and formation of aspirations via mutual association.

A cohort t agent receiving an entrepreneurship offer decides once and for all whether to accept it. Entry decisions are made myopically, comparing prospective profits from the two occupational options at date t.[2] Consequent on entering the entrepreneur selects a scale $K \geq 0$ of operation, where factor needs are proportional to K and the factor price (in a pre-Covid year) is r which does not change with t. The production function equals $A\omega^{1-\alpha}K^\alpha$, where A denotes the 'community-TFP' given by of the entrepreneur depends both on the entrepreneur's own talent ω and on the size $n_{i,t-1}$ and quality $\theta(p)$ of the incumbent network at destination i, according to

$$A_{it} = A_0 \exp(\theta(p) n_{i,t-1}) \tag{1}$$

where $\theta(p)$ denotes (exogenous) network quality, and $n_{i,t-1}$ (endogenous) size of the incumbent network at destination i. A micro-foundation for the community TFP specification is provided in Dai et al (2020b), in terms of provision of mutual help among incumbents based on informal cooperation. Network quality θ is rising in density p, which helps sustain higher levels of help within the community.

Consequent on entering, a date t cohort agent of quality ω would select scale K to maximize profits

[2]The model becomes more complicated if agents are more far-sighted, but the results continue to extend. See the Appendix in Dai et al (2020b).

$A_{it}\omega^{1-\alpha}K^{\alpha} - rK$, resulting in:

$$\log K(\omega, A_{it}) = \log \omega + \log \phi + \frac{1}{1-\alpha} \log A_{it} - \frac{1}{1-\alpha} \log r \tag{2}$$

(where $\phi \equiv \alpha^{\frac{1}{1-\alpha}}$). The resulting profit is

$$\log \Pi(\omega, A_{it}) = \log \omega + \log \psi + \frac{1}{1-\alpha} \log A_{it} - \frac{\alpha}{1-\alpha} \log r \tag{3}$$

(where $\psi \equiv \phi^{\alpha} - \phi$).

The agent with an option to enter accepts it if (3) exceeds the earnings ω^{σ} in the traditional occupation. These agents will be endowed with a level of ability that exceeds a threshold $\underline{\omega}_{it}$ satisfying:

$$\log \underline{\omega}_{it} \equiv \frac{1}{1-\sigma}[\log \frac{1}{\psi} - \frac{1}{1-\alpha} \log A_{it} + \frac{\alpha}{1-\alpha} \log r] \tag{4}$$

The threshold is assumed to lie in the interior of the support of the ability distribution at the beginning of the process for each destination, and attention is restricted to 'early phases of industrialization' when this continues to be true in later cohorts.

This defines the dynamics of entry and firm size of entrants, thereby determining the evolution of the network at different destinations across successive cohorts. Entry flows are given by

$$e_{it} = ks_{i,t-1}[B + C\theta(p)n_{i,t-1}] \tag{5}$$

where $B \equiv 1 - \frac{1}{1-\sigma} \log \frac{1}{\psi} - \frac{\alpha}{(1-\sigma)(1-\alpha)} \log r + \frac{1}{(1-\sigma)(1-\alpha)} \log A_0$ and $C \equiv \frac{1}{(1-\sigma)(1-\alpha)}$. Dai et al (2020b) show that entry and sectoral concentration (measured by the Herfindahl index) are rising over time. Moreover, both levels and changes of entry and concentration are rising in network density p when there are two destinations. Hence community TFP and network size are increasing in p. Entering firm size and quality of the marginal entrant is decreasing in community TFP, because a higher community TFP implies higher profits consequent upon entry by any given entrepreneur, and the marginal entrant must be indifferent between entering and going to his alternative occupation. Hence the net profit of the marginal entrant must be lower, implying that the marginal entrant must have lower TFP overall. This represents a form of adverse selection: the lower individual quality of the marginal entrant must outweigh the higher community TFP of a higher quality network. If $\sigma > \frac{1}{2}$ it turns out that the average entrant also enters with a lower productivity, firm size and profit.

Fix a particular destination i, and drop the notation for i in what follows. Also abstract from the sectoral distribution and set the sector shares to unity. Now suppose in the year $T = 2020$, there is a sudden unanticipated rise in the factor price r at this destination to $r(1 + \Delta)\xi(p)$, where Δ represents the size of the Covid-19 shock, and $\xi(p)$ represents its relative intensity for an incumbent network of density p. Here $(1 + \Delta)\xi(0) > 1$ ensuring that factor prices have increased for every network. Higher quality networks manage to buffer the shock through mutual help and sharing of capital and risks, so assume that the intensity ξ is decreasing in p. Using (5), entry flows in year T then fall relative to the counterfactual of no Covid-19 shock, by

$$k\frac{\alpha}{(1-\alpha)(1-\sigma)}[\log(1+\Delta) + \log\xi(p)] \tag{6}$$

which is decreasing in p. This is the first, obvious prediction of the model: higher density networks will experience a smaller contraction in flow of new firms entering.

Turn now to how incumbent firms are affected. The rise in factor price will reduce operating profits, and some of the less productive entrepreneurs may not be able to break even so they will not reopen. We investigate how the shut-down probability will vary with network density p. In general this is ambiguous, but in the following Proposition we give a sufficient condition for it to be increasing in p between any two values.

PROPOSITION **1** *Consider incumbents belonging to cohort $t < T$ that have already entered. If $n_{t'}(p)$ denotes the total size of the incumbent network that has entered by date t', i.e., cumulating across all cohorts prior to t', suppose that*

$$\frac{n_t^{'}(p)}{n_T^{'}(p)} > 1 - \sigma \tag{7}$$

a condition which will hold as long as t is not too distant from T. Suppose also that $\xi(p) = \exp(-\eta p)$ where $\eta > 0$. Take any two networks with densities p_L, p_H respectively where $p_L < p_H$. Then there exists η sufficiently small and a range of shock intensities Δ for which a larger fraction of cohort t incumbents in network with density p_H will shut down in year T following the Covid shock.

Proof: The operating profit of an incumbent entrepreneur with ability ω in network with density p that remains open in year T would equal

$$\log \Pi_T(\omega; p; \Delta) = \log \omega + \frac{1}{1-\sigma}[\log A_0 + \theta(p) n_T(p)] - \frac{\alpha}{1-\alpha}\{\log r + \log(1+\Delta) + \log \xi(p)\} + \log \psi \tag{8}$$

which is increasing in p and positive if and only if the entrepreneur's ability exceeds the threshold $\omega_B(p, \Delta)$ given by

$$\log \omega_B(p, \Delta) = \frac{\alpha}{1-\alpha}[\log(1+\Delta) + \log \xi(p)] - \theta(p) n_T(p) - \hat{c} \tag{9}$$

where $\hat{c} \equiv \frac{1}{1-\alpha}\log A_0 - \frac{\alpha}{1-\alpha}\log r + \log \psi$. Clearly in the absence of any shock ($\Delta = 0$) all incumbents will make positive profits, because the network size and hence community TFP has grown since they entered. Therefore with $\Delta = 0$, the breakeven threshold $\omega_B(p, 0)$ for any network density p is below the entry threshold $\underline{\omega}_t(p)$. But for Δ large enough this inequality will be reversed and some low ability incumbents will shut down.

Now observe that the range of operating profits across cohort t incumbents will correspond to (8) over a range of ability exceeding the threshold $\underline{\omega}_t(p)$ given by (3). So the minimum operating profit for incumbents of this cohort at date T is $\underline{\Pi}_t(p; \Delta)$ given by (8) evaluated at the entry threshold $\underline{\omega}_t(p)$, which equals

$$\log \underline{\Pi}_t(p; \Delta) = c_0 - \frac{1}{1-\alpha}[\theta(p)\{\frac{n_t(p)}{1-\sigma} - n_T(p)\} - \alpha\{\log(1+\Delta) + \log \xi(p)\}] \tag{10}$$

where $c_0 \equiv \frac{\alpha}{1-\alpha}\log r - \frac{\sigma}{(1-\alpha)(1-\sigma)}\log A_0 - \frac{\sigma}{1-\sigma}\log \psi$. Condition (7) implies the lowest post-shock operating profit (10) among cohort t incumbents is decreasing in p.

Given $p_L < p_H$ observe that (7) implies that $Q \equiv \inf_{p \in (p_L, p_H)} \frac{\partial[\theta(p)\{n_t(p) - (1-\alpha)(1-\sigma) n_T(p)]}{\partial p}$ is strictly positive. Then $\eta < Q\frac{1}{\alpha(1-\sigma)}$ implies that $\log \omega_B(p_L; 0) - \log \omega_B(p_H; 0)$ is smaller than $\log \underline{\omega}_t(p_L) - \log \underline{\omega}_t(p_H)$. Hence there exist a range of values of Δ for which $\log \underline{\omega}_t(p_H) < \log \omega_B(p_H; \Delta) < \log \omega_B(p_L; \Delta) < \log \underline{\omega}_t(p_L)$, i.e., where a positive fraction of the incumbents in network p_H will shut down, but none among the incumbents in network p_L would shut down. ■

Condition (7) ensures the adverse selection effect persists beyond the entry date t until the date T when the shock hits.[3] Then the lower bound of operating profits in the high density network is smaller,

[3] This condition holds if $T = t$. However with $t < T$, the size of the higher density network grows faster, so it may not continue to hold at later dates if T is sufficiently distant from t.

implying that it has a longer lower tail. The fraction of incumbents shutting down would be higher in the more dense network if the shock is of a magnitude that threatens only the excess lower tail. Proposition 1 shows that for any pair of densities, there exist parameter values for which a higher fraction of high density network incumbents would shut down. In general, the comparison is ambiguous: examples can be constructed where the opposite result obtains. For instance, for sufficiently high shock, all incumbents in the low density network shut down, while a positive fraction of those in the high density network (the extreme upper tail) continues to operate. This is again a consequence of the wider dispersion of productivity in the higher density network.

3 Cluster Index, Data and Descriptive Statistics

3.1 Cluster Index

Standard measures of industrial clusters in the IO and urban economics literatures are based on indices of regional specialization in specific industries, such as concentration ratio, relative concentration or spatial Hirschman-Herfindahl Index (HHI) of firms located in any given region across different industries. Examples are the Krugman index or the Ellison-Glaeser index. However, as argued by Ruan and Zhang (2015), these indices do not adequately measure presence of clusters in LDCs. This owes to the distinctive features of clusters in LDCs compared to DCs, involving co-existence of firms in many different but related industries, resulting from a high degree of vertical disintegration. Consequently LDC clusters frequently include firms in different upstream and downstream industries connected via trade links, or firms producing diverse products but sharing common inputs. The diversity of industries within the cluster is then reflected in a low measure of regional specialization.

The Puyuan cashmere sweater cluster in Tongxiang county provides a ready illustration. It contains seven different 3-digit industries with an employment share exceeding 1% for the entire country, each of which corresponds to different stages of sweater production (with the 3-digit industry code in parentheses): (1) silk spinning/printing/dyeing (174); (2) wool spinning/printing/dyeing (172); (3) manufacturing of knitted fabrics (176); (4) leather tanning/processing (191); (5) fur tanning/processing (193); (6) synthetic fiber manufacturing (282); (7) financial information services (694)). In a more vertically integrated firm, these different stages would have been represented by different divisions within the firm, resulting in a greater measure of specialization (i.e. classification as a single industry rather than seven different ones).

To deal with this problem, Ruan and Zhang (2015) develop a cluster index better suited to LDC context, based on a measure of inter-industry proximity or 'related industries', based on similarity of 'revealed comparative advantage' (RCA). The measure of **proximity** of industries i, j (based on employment E_{ri}, E_{rj} across regions r) is given by

$$\phi_{ij}^e = \min\{P(LQ_{ri}^e > 1 | LQ_{rj}^e > 1), P(LQ_{rj}^e > 1 | LQ_{ri}^e > 1)\}$$

where P denotes conditional probability and $LQ_{rj}^e \equiv \frac{E_{rj}/E_r}{E_j/E}$. Say that region r exhibits RCA in industry j if the employment share of the industry j in region r exceeds that for the country as a whole. The proximity measure between industries i, j corresponds to the fraction of regions in the country that exhibit RCA in both industries — i.e., the extent to which the two industries tend to co-locate in the same regions.

Given this proximity measure, the **region r cluster index (employment-based)** ϕ_r^e is defined as the weighted average of ϕ_{ij}^e, using employment weights $[E_{rj}/E_{r-i}] * [E_{ri}/E_i]$. It represents the extent to which the region involves co-location of proximate industries. Using alternative output or capital weights in the averaging procedure provides an alternative measure of clustering. The overall Ruan-Zhang (RZ) cluster index takes the average across employment, output, and capital based cluster indices.

Ruan and Zhang (2015) calculate the RZ index for China using a SIC3 classification of industries at the county level, and firm data from the 1995 China Industrial Census, and the 2004, 2008 China Economic Censuses. It successfully predicts 53 out of top 100 clusters identified by Chinese industry and government experts, compared with maximum of 3 predicted by various regional specialization indices such as CR3, Gini, HHI, Krugman or Ellison-Glaeser indices. The latter measures predict the extent of clustering to be the highest in regions with fewer firms and fewer industries located inland. In contrast the RZ cluster measure is the highest along the South-East China coast (Guangdong, Shanghai, Zhejian, Jiangsu provinces) which accords with common wisdom. This is shown in Figure 1 which provides variations in the RZ cluster index across different regions of China. Hence this measure seems more appropriate in the Chinese context than the conventional measures of regional specialization, and we shall use it for the rest of this paper.

Figure 2 shows that the cluster index and its relative magnitude across counties changes relatively little over time. It plots the log cluster index in 2004 and 2008 on the vertical axes, against values of the same index in 1995. Both are highly positively correlated with the 1995 index, with a slight tendency for clustering to rise over time. Hence it is reasonable to treat the extent of clustering as pre-determined by pre-1995 entry patterns, alleviating concerns about possible reverse causality.

3.2 Data and Descriptive Statistics

The first data set we employ is the State Administration of Industry and Commerce (SAIC) database that covers the universe of registered firms in China. This provides details of each firm registered, its location, capital, industry classification and principal business personnel such as shareholders and top managers (with identifiers for their birthplace). We use this to measure the flow of new firm registrations at the monthly level in each county-industry pair for the period between 2017 and June 2020. The data also permits us to identify the birth county of the principal representative of each firm, which we use to measure the quality of hometown entrepreneur networks, as explained further below.

There were approximately 21 million registered firms in 2018 in China. Since we will be using the SAIC data to estimate entry flows at a disaggregated (county-industry) level, we group firms into four main industries: Agriculture, Manufacturing, Business Services and Residential Services. Figures 3 and 4 respectively show the number of registered firms and employment (units of thousand) in the four industry groups. It is evident that the service sector accounts for the largest share of firms and employment, followed by manufacturing.

To examine effects on operation of incumbent firms, we use a second data set: the Enterprise Survey for Innovation and Entrepreneurship in China (ESIEC) led by Peking University. Starting in 2017, the ESIEC survey originally covered 16 counties in Henan Province, and expanded to 117 counties in six provinces in 2018. Although the sample is only representative at the provincial level, the industrial distribution of our 2017-2019 sample largely resembles the national distribution at the SIC-1 industry level. The surveys in 2017-2019 includes questions on total asset, employment, besides a large range of firm attributes. Figure 5 compares different attributes between counties with high (above median) and low (below median) cluster index. The attributes are the proportion of firms in the county whose primary supplier is located in the same country (MLocalSup), whose primary customer is located in the same county (MLocalCon), who have stable suppliers (StableSup), have stable clients (StableCon), who sell on credit to their largest client (MSellCredit), have undertaken a process innovation (NewProcess), and have positive online sales (Online). Moreover, we show inventory as a proportion of working capital (Stock), and percentage of employees who are local residents (LocalEmpRa). It is evident that high cluster regions have a significantly larger proportion of firms with local suppliers and clients, have stable customers, have online sales, sell on credit, and have undertaken process innovations. They carry smaller inventories and rely less on local workers. Table 1 provides a firm level regression of these various attributes on the log of the cluster index (in the county of location), controlling for industry dummies. We see clusters are associated with significantly greater spatial proximity to suppliers and customers, have more stable demand, more likely to sell on credit, to have online customers, process innovations, and more reliant on non-local workers.

After the outbreak of the Covid-19 pandemic, the ESIEC project alliance (comprising Peking University, Central University of Finance and Economics, Harbin Institute of Technology at Shenzhen, Guangdong University of Foreign Studies, and Shanghai University of International Business and Economics) conducted rapid phone surveys with previously interviewed entrepreneurs in the months of February and May. The completion rate was about 50% for those with valid contact information. The firm size distribution from the phone surveys match closely with the national firm size distribution based on the China Economic Census 2018 (Dai et al, 2020a). The phone surveys in February and May 2020 included a question on whether the firm had re-opened since the New Year, and various aspects of its operations. We use these two rounds of phone surveys to assess the likelihood of reopening, besides various details of their operations such as problems with suppliers, customers, and labor shortages.

Figure 6 displays the average proportion of firms that re-opened after the New Year in February and May respectively, across different industry groups. The manufacturing and residential service sectors were particularly hard-hit, with less than 20% of firms that succeeded in re-opening in February, while the other sectors had a re-opening rate of 27-28%. By May between 77-85% of firms had re-opened, with little variation across sectors. Part of the reason that firms were adversely affected was the rate of Covid infections in the local area. Figure 7 shows a scatterplot of the log of the cluster index in the county against the infection rate in the prefecture. It is evident that areas with a higher cluster index experienced a higher infection rate. Figure 8 shows that the local infection rate was also positively correlated with the (average) infection rate in entrepreneur hometowns. Despite experiencing higher covid infection rates, we shall see below that higher cluster regions experienced a lower contraction in new firm registrations and reopening rates.

3.3 Hometown network quality measures

As explained in the Introduction, Dai et al (2020b) show population density of entrepreneur home counties is a suitable proxy of of their social connectedness. However, they found that this was only true for rural county birthplaces; urban birthplaces feature higher population densities and markedly lower levels of trust and cooperation owing to greater social heterogeneity. Hence for urban county birthplaces we replace the true population density with zero, and then construct the weighted average of population density of birth counties of listed entrepreneurs. We use the 1982 Census to construct population density, so that the measure is pre-determined and not subject to any reverse causality.

This variable alone cannot serve as a suitable measure of relevant social connectedness of entrepreneurs operating in any given county, since a large fraction of entrepreneurs in China (approximately 60%) from rural counties set up their enterprises outside their birth county. Hence the area where the enterprise is located (the destination) is frequently different from the birth county (the entrepreneurs origin). If at a given destination the entrepreneurs come from many different origins, their connectedness would be considerably lower than if they all came from the same origin. Therefore we need to supplement average home county density with a measure of homogeneity or spatial concentration of their origins. We measure the latter by the Herfindahl-Hirschman index (HHI) of concentration across different birth home counties, excluding the destination county.

If the story in Dai et al (2020b) is correct in explaining the origins of the clusters, we would expect counties with a higher cluster index to be associated with a higher average hometown density and a higher hometown concentration, since either of the latter two attributes would increase network-based entry of firms from the respective hometowns thereby raising the number of cluster firms. Figures 9 and 10 bear out this prediction. This suggests that average hometown density alone is a good measure of network quality, and corrections for dispersion are unlikely to be important. Nevertheless in the regressions below we shall include controls for hometown concentration when we use average density as a measure of network quality.

4 Empirical Results: Covid-Resilience and Clusters

We use monthly firm registration data at the county-industry level from 2015 to 2020. Similar results obtain when we analyze the weekly data, but these results are less reliable owing to greater frequencies of zeroes in the data. The sample excludes a few provinces (Xinjiang, Qinghai, Tibet, and Inner Mongolia), which have large pastural areas, are sparsely populated and register very few firms at the county level. In addition, our sample does not include Hubei Province, the epicenter of Covid-19 pandemic as it was under complete lockdown and businesses ground to a halt for about two months.

With log of new per capita firm entries at the county-industry-month-year as the dependent variable, Figure 11 shows estimated regression interaction coefficients (along with 95% confidence bands) between month dummies and a 2020 year dummy, when the sample is split into a high (above median) and low (below median) cluster index. The regressions include dummies for month, 2020, county and industry, thus controlling for common unobserved sector and location characteristics that do not vary over time. We see a significantly smaller drop in February 2020 compared to February of previous years for the high cluster counties: entry rates in February 2020 declined by 67% compared to February in previous years in the high cluster regions, compared to 74% in the other regions.

Figure 12 and Table 2 present results from a more demanding specification using a continuous cluster index interacted with month and 2020, controlling for per capita infection rates in the county and in the entrepreneurs' hometown, and includes dummies for county-industry-month, county-industry-year and month-year (i : county, j : industry, t : year, m : month, $i'(i)$: prefecture that i belongs to):

$$Perfirm_{ijtm} = \alpha + \sum_{m} \beta_m D_m * D_{2020} * LnCluster_i + \gamma I_{i'(i)tm} + \delta BI_{ijtm} + \lambda_{ijt} + \mu_{tm} + \pi_{ijm} + \epsilon_{ijtm} \tag{11}$$

where $Perfirm$ denotes the log of (per capita entry of new firms $+0.001$), D denotes dummy, $LnCluster_i$ denotes log of the cluster index in county i, I denotes covid infection rate, and BI denotes infection rate in the birthplace of the entrepreneurs in the county-industry pair (averaged using hometown shares as weights). Interaction coefficients between deviations of each month of 2020 from New Year's Eve and $LnCluster$ are plotted in Figure 12, along with 95% confidence intervals. We see a significant positive coefficient of the cluster index in the month immediately following New Year. Moreover, Table 2 shows a significant negative impact of the local infection rate.

Next we turn to the ESIEC entrepreneur phone survey data and examine covid impacts on the performance of incumbent firms in the February and May 2020 rounds, and how it varied with the cluster index. Table 3 shows regression coefficients of $LnCluster$ on a firm dummy for reopening in February and May respectively, controlling for an offseason dummy and industry dummies (both the 4-sector classification as well as SIC1 classification). We see a significant $LnCluster$ coefficient ranging between 3-3.5 % in February. Controlling for the local infection rate, this rises to 4.6-5.3%, which implies a 4-4.5% greater likelihood in counties with a 1 s.d. higher cluster index. The direct impact of the infection rate is again negative and significant. Similar to the case of the entry data, the superior resilience of clusters obtains irrespective of whether or not we control for the infecton rate. In May we continue to see a significant higher likelihood of over 2.5% of being open with 1 s.d. higher cluster index.[4] Hence differences between high and low cluster regions persisted even after four months, despite the substantial easing of the pandemic and related restrictions.

In summary, both entry of new firms and incumbent performance were less adversely affected in counties with higher clustering.

5 Disentangling Role of Different Attributes of Clusters

5.1 Hometown Network Density

We have already seen that counties with high clustering also featured higher quality (population density) entrepreneur hometown networks. To what extent could this explain their lower vulnerability to the covid shock?

We first add interactions of month and 2020 with average (log) population density of entrepreneur hometowns and with (log) HHI of hometowns to regression (11) for new firm entries. The resulting interaction coefficients are shown in Figure 13. The interaction coefficients of hometown concentration are not significant. Table 4 shows the estimated regression coefficients for infection rates, and

[4]For the reopening rate in May, we do not control for the infection rate since the pandemic had eased substantially by that time.

interactions with months following New Year's Eve of logs of hometown density and the cluster index. Higher population density has a significant, positive interaction coefficient one month after New Year's Eve, while that of the cluster index also remains positive and significant (though somewhat attenuated compared to Figure 12 when we did not control for hometown density and concentration). The coefficients of density and cluster happen to have almost the same magnitude and significance. The s.d. of hometown density is 0.69 compared to 0.88 for the cluster index.Therefore we see that reliance on higher quality hometown networks helps explain some of the benefits of clustering, but not entirely.

Even after controlling for entrepreneurial network quality, a 1 s.d. increase in the cluster index was associated with a 12% higher entry rate between Feb 10 and March 6, 2020, significant at the 1% level. In the preceding five months and subsequent three months the estimated interactions are statistically indistinguishable from zero.[5] The regression coefficient of the county infection rate continued to be negative and significant, while that of the hometown infection rate was insignificant — indicating a strong adverse direct impact of the covid shock.

Table 5 shows the corresponding results for the reopening likelihood of surveyed firms in February and May 2020, when we add hometown density and concentration (in logs) to the regression reported in Table 3. Consistent with the case depicted in Proposition 1, we find a significant negative effect of higher density in both February and May (though for the former this happens when the cluster index is also included in the regression). The network-based adverse selection effect therefore provides a possible explanation of this result. The effect of hometown concentration is throughout insignificant.

These results also imply that the effect of clustering is even larger when we control for hometown network quality (i.e., compared to Table 2). Therefore as in the case of the entry results, the benign effects of clustering on vulnerability to the covid shock survive even despite controlling for the network effects. In other words, superior network quality alone cannot account for the greater buffering capacity of clusters. This calls for an exploration of *other* benefits of clustering.

[5]Four months prior to New Years Eve, however, we see effects of cluster and network concentration were significant, while that of density was negative and significant. This corresponded to October 2019, with a large countrywide weekly holiday.

5.2 The Role of Other Attributes of Clustering

As shown in Table 1, areas with higher clustering are located closer to their suppliers and customers, are more likely to sell online and on credit, and have more stable customers. They also rely less on local workers. The closer proximity to suppliers and customers could have helped clusters buffer the covid shock which imposed severe limits on the movement of goods (input supplies, movement of sold goods outside the local area) and people (e.g., customers who visited personally). On the other hand, their greater reliance on migrant workers would have rendered them more vulnerable, as workers would have gone home during the New Year and may not have been able to return owing to lockdown restrictions or necessary quarantine procedures.

Tables 6 and 7 show how the regression results in Table 5 are modified when we replace the cluster index with related firm attributes, for the February and May reopening likelihoods respectively. We continue to control for hometown network quality. Firms with more stable suppliers and customers were more likely to reopen. The same is true for firms with local suppliers and customers, and for those selling online (significant in May). Firms relying more on local workers were less likely to remain open, which is somewhat surprising in view of the mobility restrictions associated with the pandemic. It is possible that higher productivity firms tend to rely more on migrant workers, so this result may be driven by endogenous selection (particularly in May after the lockdown restrictions had eased). In summary, greater spatial (both physical and online) proximity to suppliers and customers partly accounted for the superior resilience of clusters, besides their reliance on higher quality entrepreneur networks.

6 Variation Across Industry Groups

How did the preceding results vary across sectors? We re-run the analysis for each of the four industry groups separately. Table 8 shows the regression coefficients on network density and cluster index on entry rates for each month following January 2020. Neither matters much in agriculture. The benign effects of network density appear in the two service sectors, while those of clustering appear to be largest in manufacturing, but also significant (though the magnitude of the coefficient is nearly half of that of higher network density). This is roughly consistent with the notion that benefits of spatial proximity are greatest in manufacturing which involves movement of bulky goods. Tables 9 and 10 show how effects of clustering and network quality on reopening rates varied across industry groups. They are consistent with the results on entry: network density mattered only in the service sector, while spatial proximity mattered in both manufacturing and services.

7 Concluding Comments

In summary, we find that counties with greater presence of clustering were less adversely affected by the covid shock in terms of both entry of new firms and performance of incumbents. Part of the explanation of the entry result could be provided by higher entrepreneur network density of such areas in which incumbents shared risks better with one another and provided greater assistance to new entrants from the same hometown in overcoming entry barriers. But superior network quality also tends to co-exist with lower productivity on average owing to the adverse selection it induces by lowering entry thresholds, which lowered incumbent performance. Hence the superior ability of incumbents in clusters to adapt to the covid shock arose despite, rather than because of, superior network quality. The entrepreneur survey results suggest the role of closer proximity to suppliers and customers in stabilizing supply chains, reducing vulnerability to transport bottlenecks and market demand fluctuations.

Our measure of network density was based on 1982 Population Census data on population density, while the results are robust to using cluster measures based on 2004 Economic Census, rather than the 2008 data. Hence they are unlikely to be susceptible to problems of reverse causality. Our entry results are robust to industry dummies as well as the use of time-varying county dummies, thereby controlling for local infrastructure and governance; there are no discernible pre-trend differences between high and low cluster regions. These reduce concerns of omitted variable bias, thereby suggesting the results can be given a causal interpretation.

The results of the paper are useful in two different ways. First, it provides evidence of and insight into possible reasons for the superior capacity of production clusters to withstand external shocks in a volatile environment with underdeveloped formal markets and institutions — resulting from a combination of informal network-based cooperation, risk-sharing and spatial proximity among buyers and sellers. These risk-coping advantages may account for their survival and growth, despite lower productivity (on average) compared to other forms of industrial organization based on high vertical integration, capital intensity and spatial separation from suppliers and buyers. Second, it can help predict relative vulnerabilty of different regions or industries to possible recurrence of shocks that impair the movement of goods and people, thus providing a useful tool for direction of assistance by governments or international aid agencies.

References

Banerjee A. and K. Munshi, 2004. ”How Efficiently is Capital Allocated? Evidence from the Knitted Garment Industry in Tirupur.” Review of Economic Studies, 71(1), 19-42.

Dai Ruochen, Hao Feng, Junpeng Hu, Quan Jin, Huiwen Li, Ranran Wang, Ruixin Wang, Lihe Xu, and Xiaobo Zhang. 2020a. ”The Impact of COVID-19 on Small and Medium-sized Enterprises: Evidence from Two-wave Phone Surveys in China.” Center for Global Development Working Paper No. 546.

Dai Ruochen, Dilip Mookherjee, Kaivan Munshi and Xiaobo Zhang, 2020b, ”Community networks and the growth of private enterprise in China”, Working Paper, Economic Development and Institutions Network.

Fang Hanming, Wang Long and Yang Yang, 2020. ”Human mobility restrictions and the spread of the Novel Coronavirus (2019-nCoV) in China”, Journal of Public Economics, Volume 191.

Ellison, Glenn, and Edward L. Glaeser, 1997. ”Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach.” Journal of Political Economy 105 (5): 889-927.

Fleisher, Belton, Dinghuan Hu, William McGuire, and Xiaobo Zhang, 2010. ”The Evolution of an Industrial Cluster in China.” China Economic Review, 21(3): 456-469.

Hsieh, C.T and Klenow, P., 2009. ”Misallocation and manufacturing TFP in China and India”. Quarterly Journal of Economics 124 (4), 1403–1448.

Kali R., 1999. ”Endogenous Business Networks,” Journal of Law Economics and Organization, 15(3), 615-636.

Kranton R., 1996. ”Reciprocal Exchange: A Self-Sustaining System,” American Economic Review, 830-851.

Kranton R. and D. Minehart, 2000. ”Networks versus Vertical Integration,” Rand Journal of Economics, 31(3), 570-601.

——————, 2001. ”A Theory of Buyer-Seller Networks,” American Economic Review, 91(3), 485-508.

Long, Cheryl and Xiaobo Zhang, 2011. ”Cluster-based industrialization in China: Financing and performance”, Journal of International Economics 84(1): 112–123.

Machiavello R. and A. Morjaria, 2015. ”The Value of Relationships: Evidence from a Supply Shock to Kenyan Rose Exports,” American Economic Review, 105(9), 2911-2945.

———— 2020. ”Competition and Relational Contracts in the Rwanda Coffee Chain,” working paper, London School of Economics.

Marshall, A., 1920. Principles of Economics, 8th ed. Macmillan, London. (Original work published 1890).

McMillan J. and C. Woodruff, 1999a. ”Interfirm Relationships and Informal Credit in Vietnam,” Quarterly Journal of Economics, 114(4), 1285-1320.

———— 1999b. ”Dispute Prevention Without Courts in Vietnam,” Journal of Law Economics and Organization, 15(3), 637-658.

Merima, Ali, Jack Peerlings, and Xiaobo Zhang, 2014. ”Clustering as an Organizational Response to Capital Market Inefficiency: Evidence from Microenterprises in Ethiopia.” Small Business Economics, 43:697-709.

Ruan, Jianqing and Xiaobo Zhang, 2009. ”Finance and Cluster-Based Industrial Development in China,” Economic Development and Cultural Change, 58:143-164.

Ruan, Jianqing and Xiaobo Zhang, 2015. ”A Proximity-Based Measure of Industrial Clustering.” International Food Policy Research Institute (IFPRI) Discussion Paper 1468.

Sonobe, Tetsushi and Keijiro Otsuka, 2006. Cluster-Based Industrial Development: An East Asian Model. Palgrave MacMillan.

World Bank, 2011. Industrial Clusters and Micro and Small Enterprises in Africa. Yutaka Yoshiko (Editor), Washington DC: World Bank Publications.

Zhang, Xiaobo, Qingqing Chen, and Peixun Fang, 2019. ”Cluster-based Aquaculture Growth” in The Making of a Blue Revolution in Bangladesh: Enablers, Impacts, and the Path Ahead for Aquaculture, Edited by Shahidur Rashid and Xiaobo Zhang, page 57-76. International Food Policy Research Institute.

Zhang, Xiaobo, Lisa Moorman, and Gazagegn Ayele, 2011. ”Infrastructure and Cluster Development: A Case Study of Handloom Weavers in Ethiopia,” Journal of Development Studies, 47:12, 1869-1886.

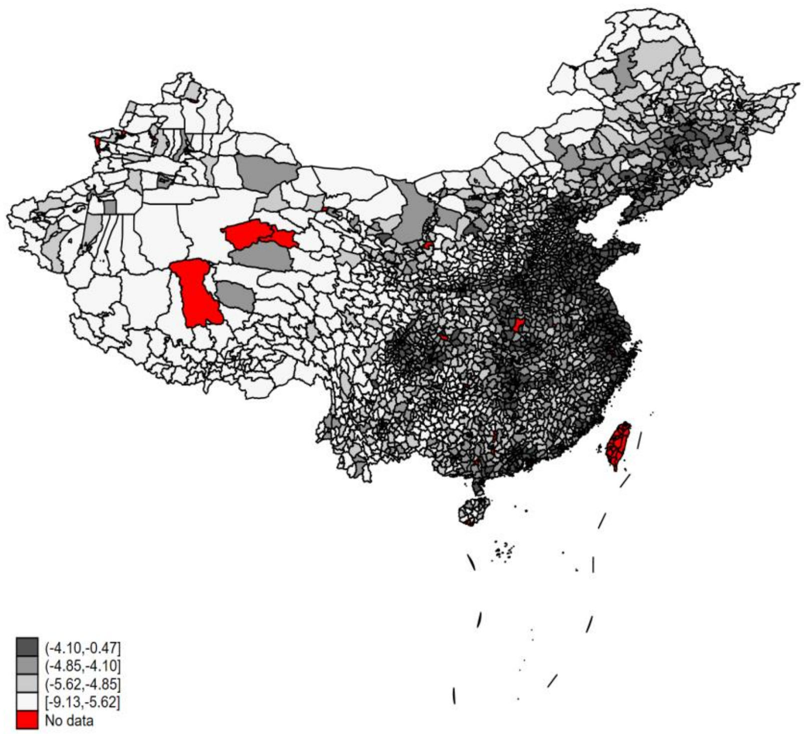

FIGURE 1: MAP OF CLUSTER INDEX ACROSS CHINA

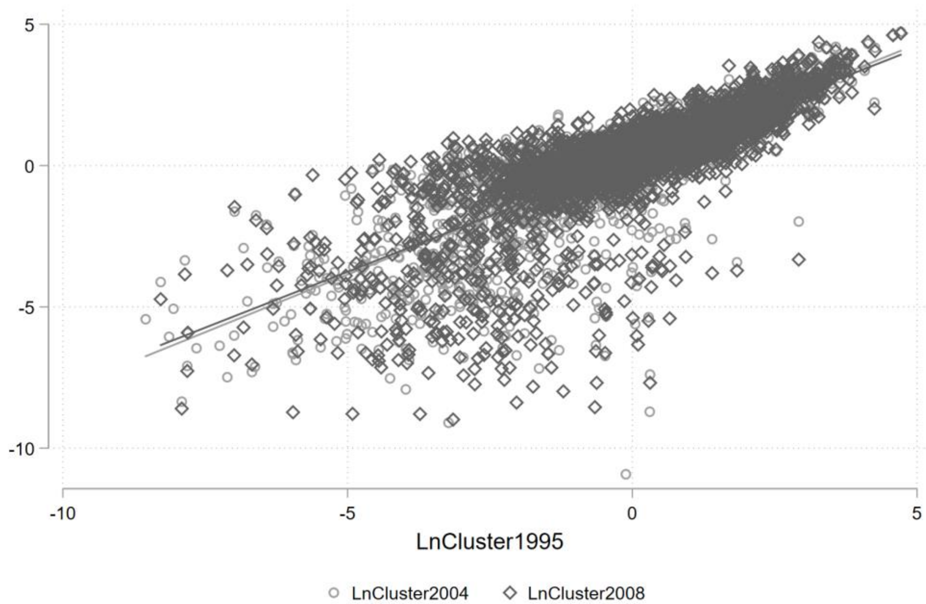

FIGURE 2: LOG CLUSTER INDEX 2004, 2008 vs 1995 (SAIC registration data)

Note: We use SAIC registration data to compute the cluster index in this figure for all the three years because the China Industry Census 1995 does not include firms in the service and agricultural sectors.

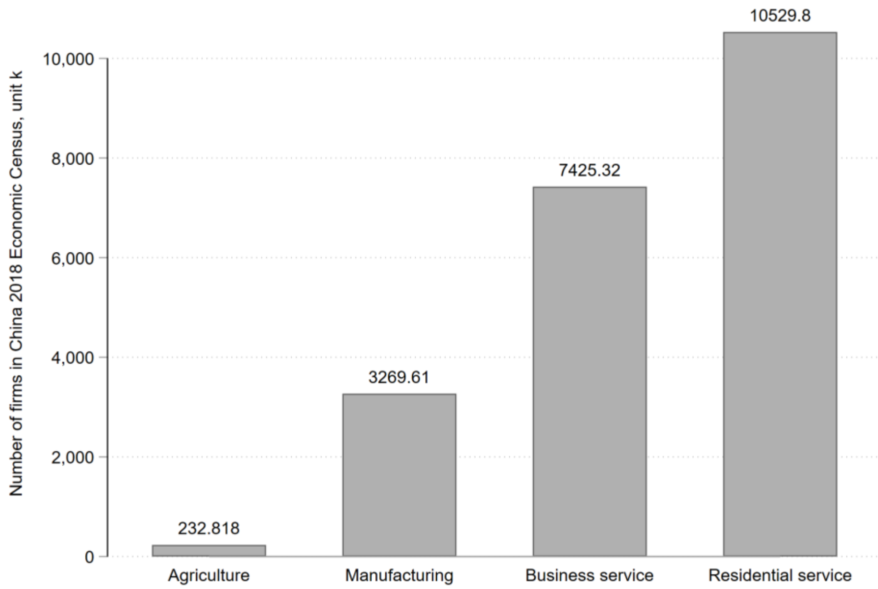

FIGURE 3: NUMBER OF REGISTERED FIRMS IN 2018 ('000) BY INDUSTRY GROUP

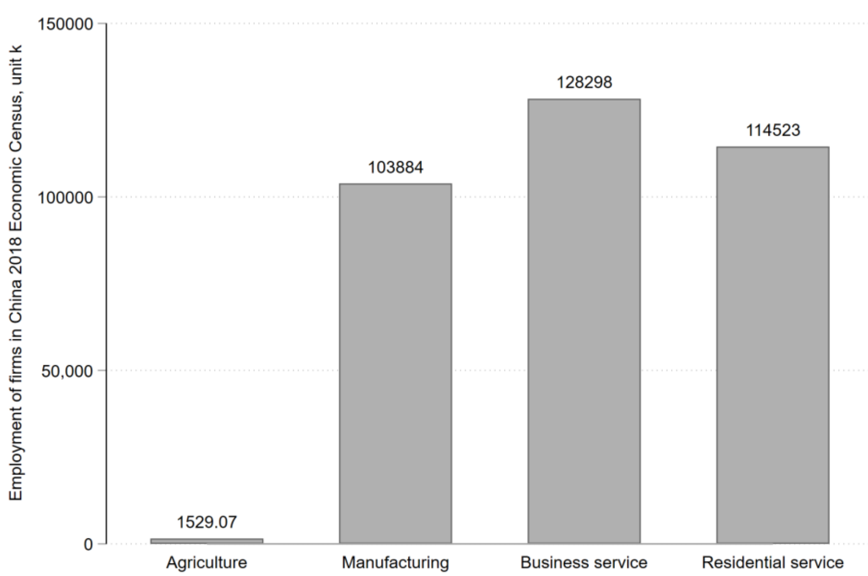

FIGURE 4: EMPLOYMENT IN 2018 ('000) BY INDUSTRY GROUP

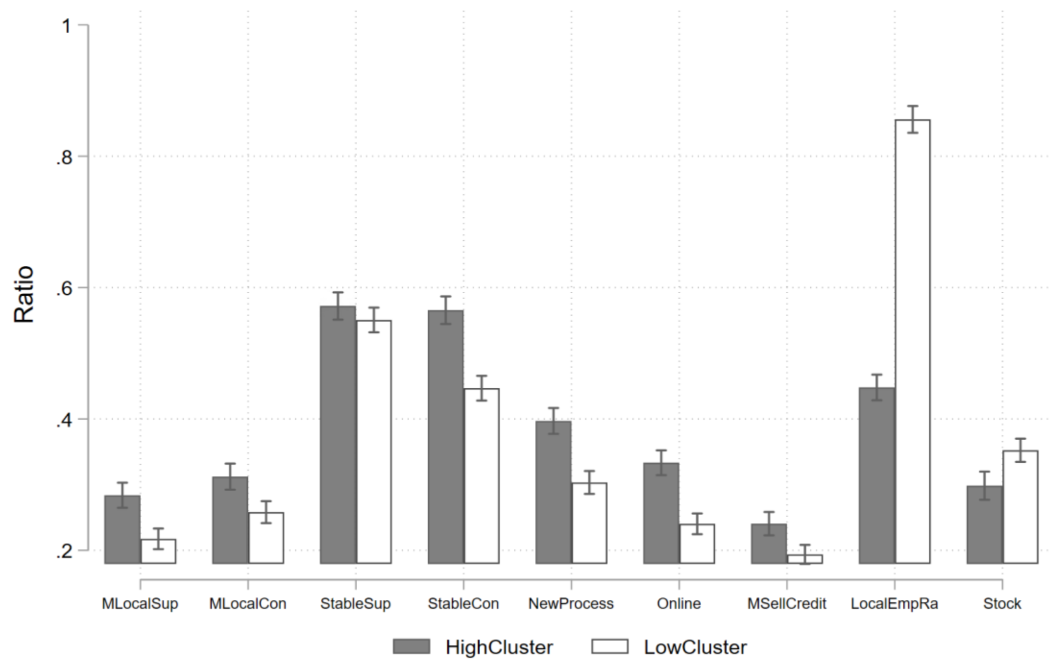

FIGURE 5: FIRM ATTRIBUTES, HIGH VS LOW CLUSTER COUNTIES

Note: Computed by authors based on ESIEC survey. The attributes are the proportion of firms in the county whose primary supplier is located in the same country (MLocalSup), whose primary customer is located in the same county (MLocalCon), who have stable suppliers (StableSup), have stable clients (StableCon), have undertaken a process innovation (New-Process), have positive online sales (Online) and who sell on credit to their largest client (MSellCredit). Moreover, we show inventory as a proportion of working capital (Stock), and percentage of employees who are local residents (LocalEmpRa).

VARIABLES	(1) StableSup	(2) StableCon	(3) MLocalSup	(4) MLocalCon	(5) NewProcess	(6) Online	(7) MSellCredit	(8) LocalEmpRa
LnCluster	0.004	0.049	0.027	0.014	0.032	0.035	0.013	-0.142
	(0.009)	(0.009)	(0.007)	(0.009)	(0.007)	(0.007)	(0.008)	(0.019)
Constant	0.576	0.680	0.347	0.336	0.461	0.393	0.266	0.146
	(0.034)	(0.032)	(0.026)	(0.033)	(0.024)	(0.028)	(0.030)	(0.067)
Observations	4,795	4,739	4,708	4,657	4,967	5,857	5,000	4,109
Adjusted R-squared	0.024	0.045	0.007	0.011	0.014	0.016	0.018	0.223
Indus FE	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses

TABLE 1: CLUSTER INDEX AND FIRM ATTRIBUTES

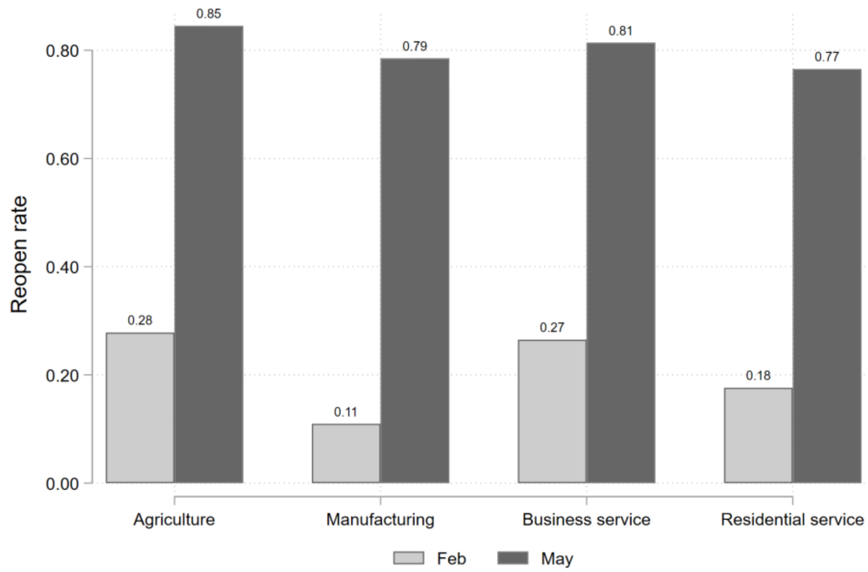

FIGURE 6: PERCENT FIRMS REOPENING AFTER NEW YEAR IN FEBRUARY & MAY 2020, BY INDUSTRY GROUP

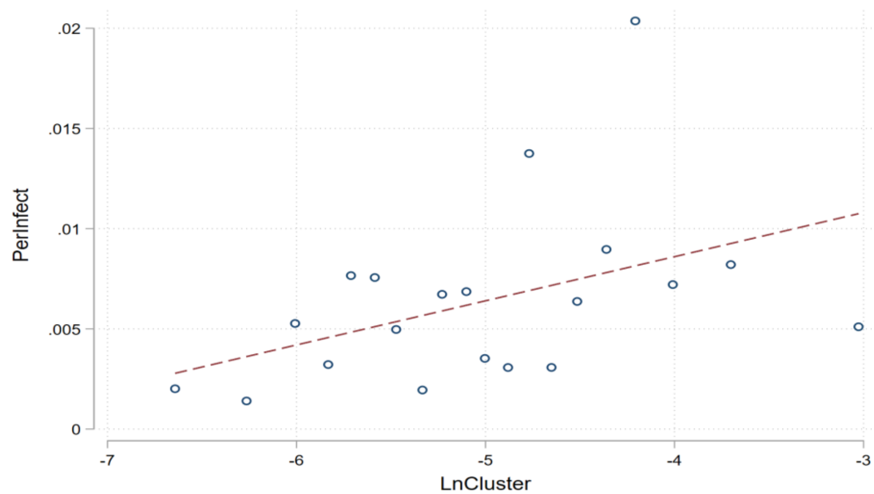

FIGURE 7: SCATTERPLOT OF (LOG) CLUSTER INDEX VS LOCAL COVID INFECTION RATE

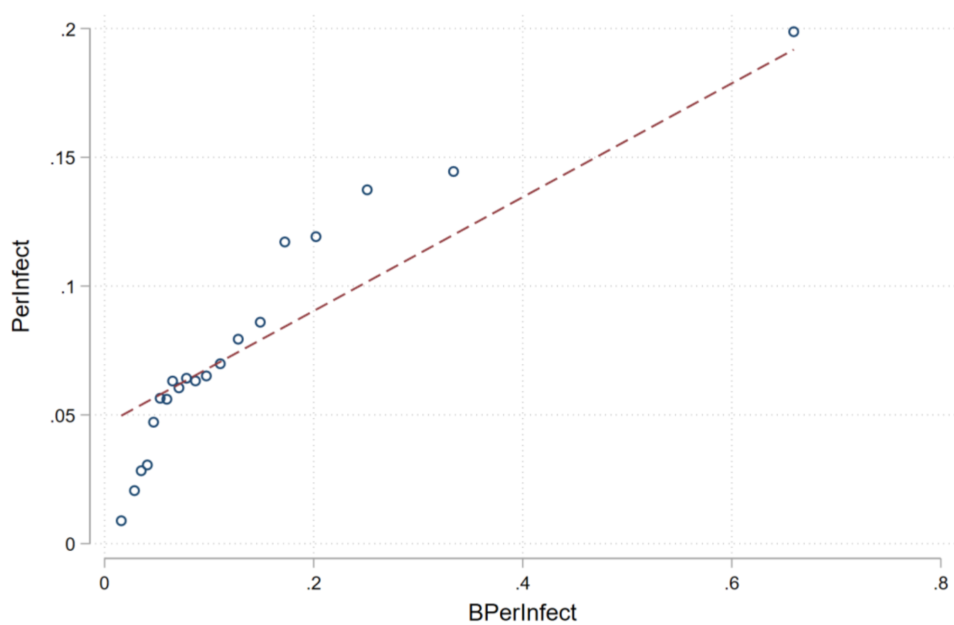

FIGURE 8: SCATTERPLOT OF COVID INFECTION RATE: LOCAL VERSUS ENTREPRENEUR HOMETOWN

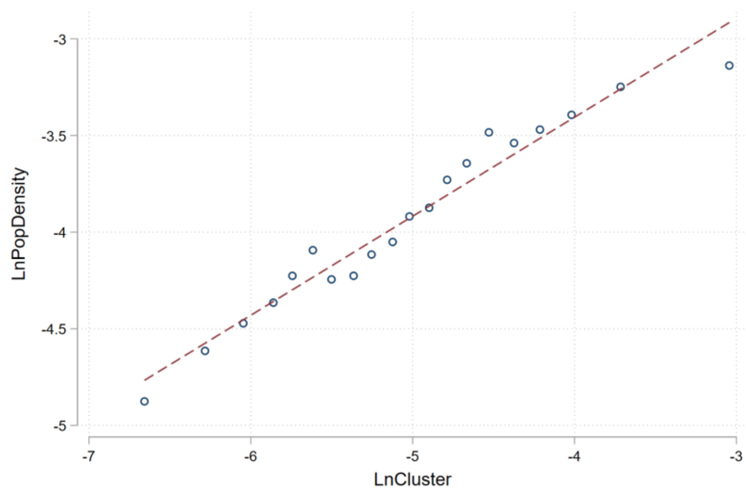

FIGURE 9: SCATTERPLOT OF (LOG) CLUSTER INDEX VS (LOG) POPULATION DENSITY OF ENTREPRENEUR HOMETOWN

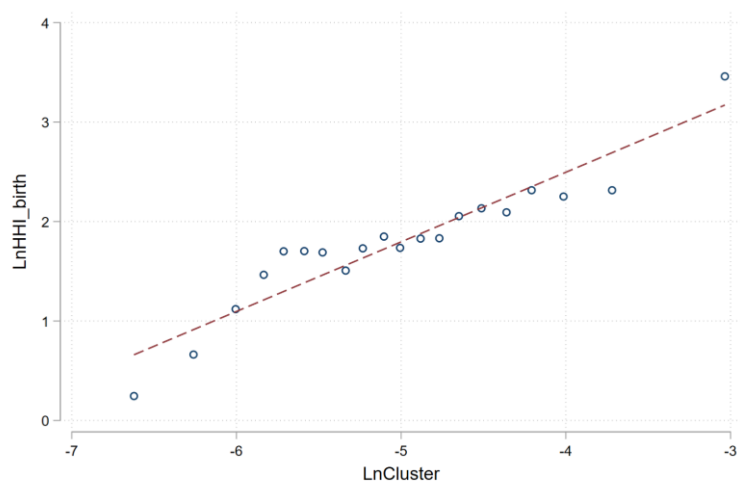

FIGURE 10: SCATTERPLOT OF (LOG) CLUSTER INDEX VS SPATIAL CONCENTRATION (LOG HHI) OF ENTREPRENEUR HOMETOWN

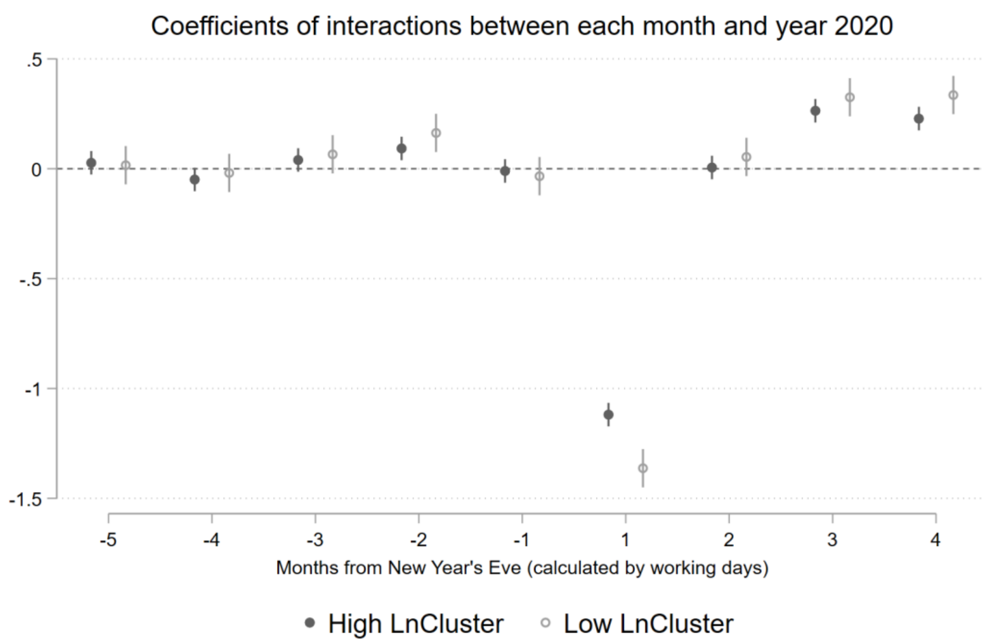

FIGURE 11: PER CAPITA ENTRY REGRESSION INTERACTIONS BETWEEN MONTH DUMMIES AND 2020, SEPARATELY BY HIGH AND LOW CLUSTER COUNTIES

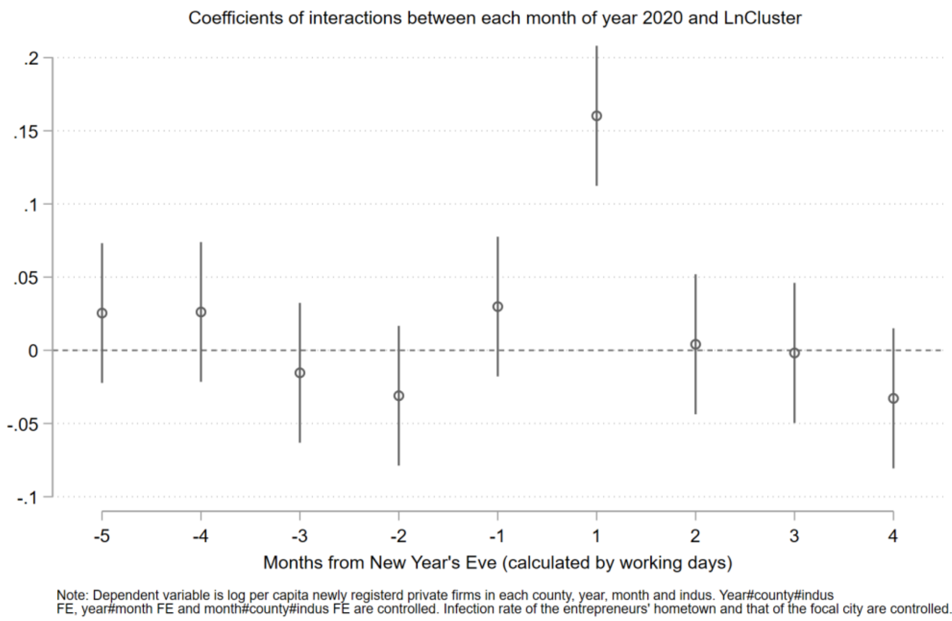

FIGURE 12: PER CAPITA ENTRY REGRESSION INTERACTIONS BETWEEN LNCLUSTER AND NUMBER OF MONTHS FROM NEW YEAR

	(1)
VARIABLES	Perfirm
LnCluster#Month -5	0.025
	(0.024)
LnCluster#Month -4	0.026
	(0.024)
LnCluster#Month -3	-0.015
	(0.024)
LnCluster#Month -2	-0.031
	(0.024)
LnCluster#Month -1	0.030
	(0.024)
LnCluster#Month 1	0.160
	(0.024)
LnCluster#Month 2	0.004
	(0.024)
LnCluster#Month 3	-0.002
	(0.024)
LnCluster#Month 4	-0.033
	(0.024)
BPerInfect	0.100
	(0.066)
PerInfect	-0.791
	(0.118)
Constant	-1.613
	(0.021)
Observations	231,080
Adjusted R-squared	0.673
Year-month FE	YES
Year-cnty-indus FE	YES
Month-cnty-indus FE	YES

Standard errors in parentheses

TABLE 2: PER CAPITA ENTRY REGRESSION COEFFICIENTS: INFECTION RATES, INTERACTIONS BETWEEN LNCLUSTER AND NUMBER OF MONTHS FROM NEW YEAR

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	RunWell2	RunWell2	RunWell2	RunWell2	RunWell5	RunWell5
LnCluster	0.035	0.053	0.030	0.046	0.026	0.027
	(0.012)	(0.012)	(0.011)	(0.011)	(0.010)	(0.010)
PerInfect		-0.167		-0.150		
		(0.066)		(0.065)		
OffSeason	-0.085	-0.085	-0.071	-0.071	0.010	-0.000
	(0.026)	(0.026)	(0.025)	(0.025)	(0.024)	(0.024)
Constant	0.348	0.445	0.328	0.415	0.894	0.899
	(0.047)	(0.054)	(0.044)	(0.054)	(0.035)	(0.035)
Observations	1,715	1,715	1,715	1,715	1,825	1,825
Adjusted R-squared	0.037	0.043	0.047	0.051	0.008	0.037
Indus FE	YES	YES			YES	
SIC-1 indus FE			YES	YES		YES

Robust standard errors in parentheses

TABLE 3: FIRM REOPENING (RUNWELL2: FEB 2020, RUNWELL5: MAY 2020) LIKELIHOOD REGRESSION ON INFECTION RATE AND CLUSTER INDEX

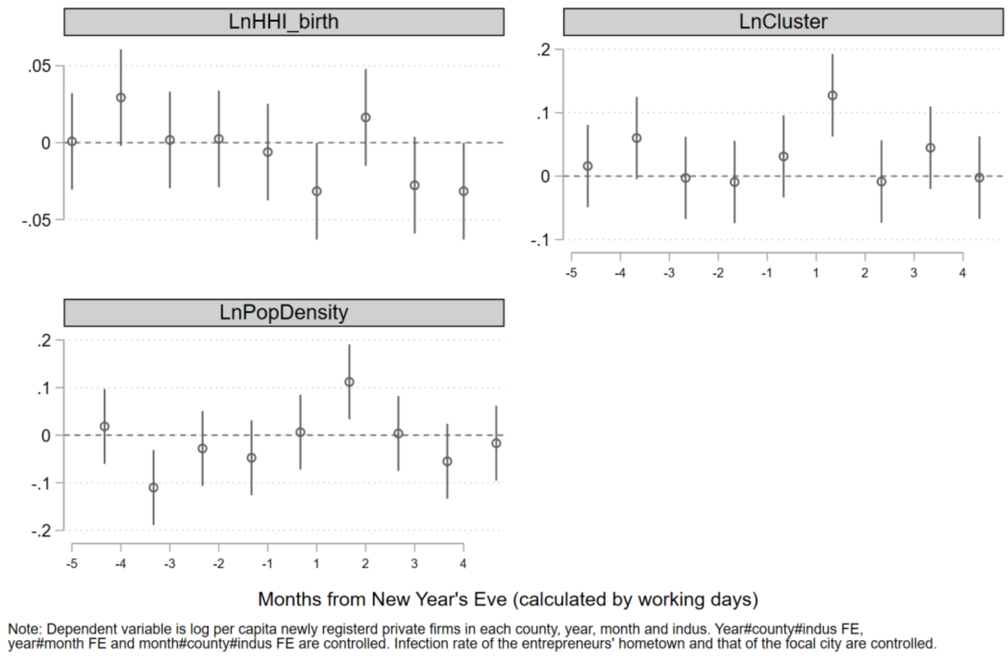

FIGURE 13: PER CAPITA ENTRY REGRESSION INTERACTIONS BETWEEN NUMBER OF MONTHS FROM NEW YEAR AND (LOGS OF) HOMETOWN SPATIAL CONCENTRATION, POPULATION DENSITY AND CLUSTER INDEX

VARIABLES	(1) Perfirm
LnPopDensity#Month 1	0.112
	(0.040)
LnPopDensity#Month 2	0.004
	(0.040)
LnPopDensity#Month 3	-0.055
	(0.040)
LnPopDensity#Month 4	-0.017
	(0.040)
LnCluster#Month 1	0.127
	(0.033)
LnCluster#Month 2	-0.009
	(0.033)
LnCluster#Month 3	0.045
	(0.033)
LnCluster#Month 4	-0.002
	(0.033)
BPerInfect	0.093
	(0.066)
PerInfect	-0.806
	(0.118)
Constant	-1.610
	(0.027)
Observations	231,080
Adjusted R-squared	0.673
Year-month FE	YES
Year-cnty-indus FE	YES
Month-cnty-indus FE	YES

Standard errors in parentheses

TABLE 4: PER CAPITA ENTRY REGRESSION COEFFICIENTS: INFECTION RATES, INTERACTIONS BETWEEN LNCLUSTER, LN HOMETOWN DENSITY AND NUMBER OF MONTHS FROM NEW YEAR

VARIABLES	(1) RunWell2	(2) RunWell2	(3) RunWell2	(4) RunWell2	(5) RunWell5	(6) RunWell5	(7) RunWell5	(8) RunWell5
LnCluster		0.072		0.065		0.044		0.047
		(0.013)		(0.013)		(0.010)		(0.010)
LnPopDensity	0.005	-0.040	0.002	-0.038	-0.063	-0.091	-0.067	-0.096
	(0.016)	(0.018)	(0.016)	(0.017)	(0.017)	(0.016)	(0.017)	(0.016)
LnHHI_birth	0.010	-0.007	0.007	-0.007	0.013	0.000	0.013	0.000
	(0.010)	(0.012)	(0.010)	(0.011)	(0.011)	(0.011)	(0.010)	(0.010)
PerInfect	-0.017	-0.210	-0.024	-0.192				
	(0.054)	(0.068)	(0.051)	(0.067)				
OffSeason	-0.101	-0.094	-0.083	-0.080	-0.015	-0.008	-0.026	-0.021
	(0.026)	(0.026)	(0.025)	(0.025)	(0.024)	(0.025)	(0.024)	(0.025)
Constant	0.222	0.407	0.219	0.383	0.563	0.662	0.553	0.657
	(0.076)	(0.090)	(0.074)	(0.089)	(0.069)	(0.070)	(0.069)	(0.069)
Observations	1,699	1,699	1,699	1,699	1,816	1,816	1,816	1,816
Adjusted R-squared	0.026	0.046	0.039	0.055	0.010	0.024	0.041	0.055
Indus FE	YES	YES			YES	YES		
SIC-1 indus FE			YES	YES			YES	YES

Robust standard errors in parentheses

TABLE 5: FIRM REOPENING (RUNWELL2: FEB 2020, RUNWELL5: MAY 2020) LIKELIHOOD REGRESSION ON INFECTION RATE, CLUSTER INDEX, HOMETOWN DENSITY AND CONCENTRATION

VARIABLES	(1) RunWell2	(2) RunWell2	(3) RunWell2	(4) RunWell2	(5) RunWell2	(6) RunWell2
LnPopDensity	0.006	0.005	0.002	0.007	0.004	0.011
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.017)
LnHHI_birth	0.005	0.001	0.002	0.005	0.007	0.005
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
MeanStableSup	0.420					
	(0.088)					
MeanStableCon		0.426				
		(0.093)				
MeanMLocalSup			0.782			
			(0.149)			
MeanMLocalCon				0.560		
				(0.147)		
MeanOnline					0.105	
					(0.119)	
MeanLocalEmpRa						-0.033
						(0.077)
Constant	0.140	0.151	0.148	0.167	0.196	0.245
	(0.066)	(0.066)	(0.066)	(0.066)	(0.066)	(0.069)
Observations	1,732	1,732	1,732	1,732	1,731	1,532
Adjusted R-squared	0.051	0.049	0.053	0.045	0.038	0.031
SIC-1 indus FE	YES	YES	YES	YES	YES	YES

Standard errors in parentheses

TABLE 6: FIRM REOPENING (FEB 2020) LIKELIHOOD REGRESSION ON FIRM ATTRIBUTES, HOMETOWN DENSITY AND CONCENTRATION

VARIABLES	(1) RunWell5	(2) RunWell5	(3) RunWell5	(4) RunWell5	(5) RunWell5	(6) RunWell5
LnPopDensity	-0.062	-0.058	-0.063	-0.057	-0.060	-0.053
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.016)
LnHHI_birth	0.011	0.007	0.010	0.009	0.008	0.008
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
MeanStableSup	0.225					
	(0.090)					
MeanStableCon		0.361				
		(0.094)				
MeanMLocalSup			0.324			
			(0.156)			
MeanMLocalCon				0.544		
				(0.148)		
MeanOnline					0.415	
					(0.117)	
MeanLocalEmpRa						-0.185
						(0.073)
Constant	0.513	0.517	0.526	0.522	0.519	0.648
	(0.060)	(0.059)	(0.059)	(0.059)	(0.059)	(0.062)
Observations	1,870	1,869	1,870	1,869	1,870	1,591
Adjusted R-squared	0.041	0.046	0.040	0.045	0.045	0.037
SIC-1 indus FE	YES	YES	YES	YES	YES	YES

Standard errors in parentheses

TABLE 7: FIRM REOPENING (MAY 2020) LIKELIHOOD REGRESSION ON FIRM ATTRIBUTES, HOMETOWN DENSITY AND CONCENTRATION

VARIABLES	(1) Perfirm (agriculture)	(2) Perfirm (manufacturing)	(3) Perfirm (business service)	(4) Perfirm (residential service)
LnPopDensity#Month 1	-0.069	0.057	0.259	0.225
	(0.099)	(0.102)	(0.050)	(0.055)
LnPopDensity#Month 2	-0.071	0.061	0.028	0.068
	(0.099)	(0.102)	(0.050)	(0.055)
LnPopDensity#Month 3	-0.116	-0.096	0.019	0.085
	(0.099)	(0.102)	(0.050)	(0.055)
LnPopDensity#Month 4	-0.004	-0.150	0.050	0.136
	(0.099)	(0.102)	(0.050)	(0.055)
LnCluster#Month 1	0.004	0.220	0.108	0.191
	(0.086)	(0.085)	(0.041)	(0.044)
LnCluster#Month 2	0.004	-0.025	-0.078	-0.023
	(0.086)	(0.085)	(0.041)	(0.044)
LnCluster#Month 3	0.022	0.145	-0.071	-0.039
	(0.086)	(0.085)	(0.041)	(0.044)
LnCluster#Month 4	-0.065	0.079	-0.087	-0.061
	(0.086)	(0.085)	(0.041)	(0.044)
BPerInfect	0.038	0.023	-0.055	0.670
	(0.183)	(0.142)	(0.080)	(0.102)
PerInfect	-0.920	-0.197	-0.217	-1.987
	(0.314)	(0.285)	(0.145)	(0.157)
Constant	-2.713	-2.475	-0.693	-0.599
	(0.060)	(0.074)	(0.035)	(0.037)
Observations	57,520	57,920	57,880	57,760
Adjusted R-squared	0.435	0.581	0.643	0.622
Year-month FE	YES	YES	YES	YES
Year-cnty-indus FE	YES	YES	YES	YES
Month-cnty-indus FE	YES	YES	YES	YES

Standard errors in parentheses

TABLE 8: ENTRY REGRESSION COEFFICIENTS ACROSS INDUSTRY GROUPS

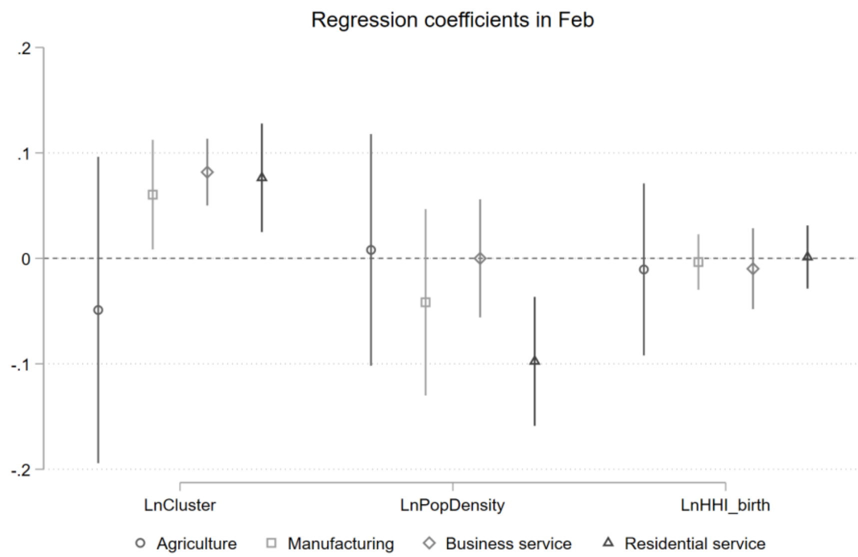

TABLE 9: FIRM REOPENING (FEB 2020) LIKELIHOOD REGRESSION COEFFICIENTS BY INDUSTRY GROUP

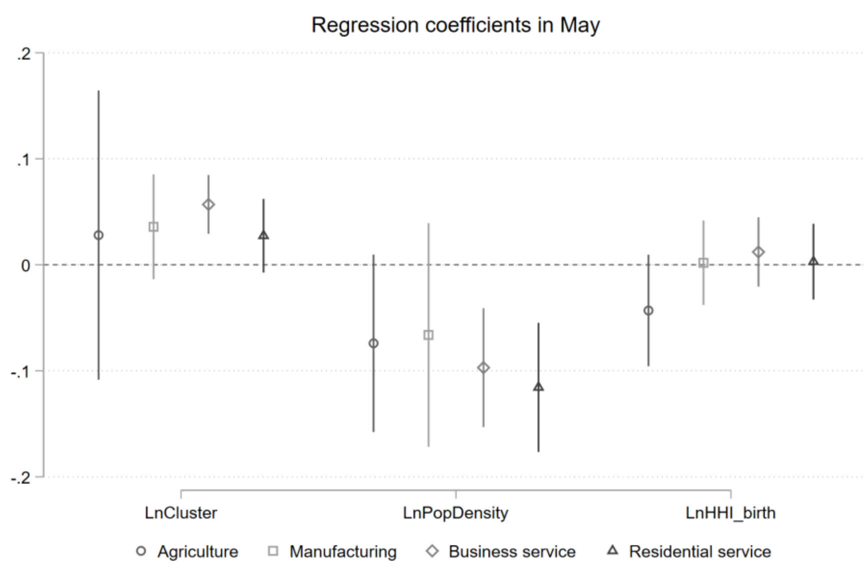

TABLE 10: FIRM REOPENING (MAY 2020) LIKELIHOOD REGRESSION COEFFICIENTS BY INDUSTRY GROUP

DATA APPENDIX

Table A1. Variables

i: $county$, j: $industry$, t: $year$, m: $month$, $i'(i)$: $city\ that\ county\ i\ belongs\ to$.

Dataset	Variable	Meaning
Registration data	$Perfirm_{ijtm}$	log(per capita entry + 0.001)
	$LnPopDensity_{ij}$	averaged population density.
	$LnHHI\_birth_{ij}$	HHI of entrepreneurs' birth place.
China 2008 Economic Census	$LnCluster_i$	clustering index.
ESIEC (firm level)	$RunWell2$	dummy, = 1 if the interviewed enterprise was in normal operation in Feb. (from ESIEC 2020)
	$RunWell5$	dummy, = 1 if the interviewed enterprise was in normal operation in May. (from ESIEC 2020)
	$StableSup$	dummy, = 1 if firm has stable suppliers. (from ESIEC 2017-2019.)
	$StableCon$	dummy, = 1 if firm has stable clients. (from ESIEC 2017-2019.)
	$MLocalSup$	dummy, = 1 if firm's largest supplier is local. (from ESIEC 2017-2019.)
	$MLocalCon$	dummy, = 1 if firm's largest client is local. (from ESIEC 2017-2019.)
	$MSellCredit$	dummy, = 1 if firm sales on credit with its largest client. (from ESIEC 2017-2019.)
	$Online$	dummy, = 1 if firm has online sales. (from ESIEC 2017-2019.)
	$LocalEmpRa$	percentage of local employees. (from ESIEC 2018.)
	$OffSeason$	dummy, =1 if firm is off season in Feb/May. (from ESIEC 2018)
Public dataset	$PerInfect_{i'(i)tm}$	infected cases of Covid-19 among 10 thousand people.
	$BPerInfect_{ijtm}$	average infected cases of Covid-19 of entrepreneurs' birth place.

Table 1b. MEANS AND STD DEVIATIONS OF KEY VARIABLES

Variable	Obs.	Mean	Std
$Perfirm$	231,080	-1.64	1.75
$LnPopDensity$	5,777	-3.76	0.69
$LnHHI\_birth$	5,777	1.99	1.47
$LnCluster$	1,438	-5.02	0.88
$BPerInfect$	5,777	0.13	0.16
$PerInfect$	265	0.08	0.10
$RunWell2$	1,768	0.22	0.41
$RunWell5$	1,891	0.88	0.41

NOTES:

1. Only private firms are included.
2. For CAIS firm registration dataset, we only consider rural counties.
3. For CAIS registration dataset, we consider 6 months before the lunar New Year and 4 months after the lunar New Year. Years included: 2017 to 2020.
4. Four industries considered: agriculture, manufacturing, business service, residential service: includes following SIC1 and SIC2 industries:

Industry	**one-digit code**	**two-digit code**
Agriculture	A	01-05
Manufacturing	C	13-43
Business service	E, G, I, J, K, L, M	47-50,53-60, 63-75
Residential service	F, H, O, P, Q, R, S, T	51-52, 61-62, 80-97