Journal of Economic Behavior and Organization 183 (2021) 433–455

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Industrial clusters, networks and resilience to the Covid-19 shock in China☆

Ruochen Dai[a], Dilip Mookherjee[b,*], Yingyue Quan[c], Xiaobo Zhang[d]

[a] Central University of Finance and Economics, China
[b] Boston University, United States
[c] Peking University, China
[d] Peking University, China and IFPRI, United States

ARTICLE INFO

Article history:
Received 8 December 2020
Accepted 13 January 2021
Available online 26 January 2021

JEL classification:
J12
J16
D31
I3

Keywords:
Clusters
Covid-19
China
Firms
Social networks

ABSTRACT

We examine how Covid-19 resilience of Chinese firms 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 sample 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 high density informal entrepreneur hometown networks and closer proximity to suppliers and customers help explain these findings.

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

☆ We are grateful to seminar participants and an EDI referee for their constructive comments. Research support from the Economic Development and Institutions (EDI #A0014), China Natural Science Foundation (#71874008 and #71950011), Ministry of Education in China (#14JJD790027), and Peking University is gratefully acknowledged. We would also like to thank the Enterprise Survey for Innovation and Entrepreneurship in China (ESIEC) Project Alliance (constituting 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) for conducting surveys on which this paper is partly based. We are responsible for any errors that may remain.

* Corresponding author.
E-mail addresses: r.dai@cufe.edu.cn (R. Dai), dilipm@bu.edu (D. Mookherjee), yyquan2018@nsd.pku.edu.cn (Y. Quan), x.zhang@nsd.pku.edu.cn (X. Zhang).

https://doi.org/10.1016/j.jebo.2021.01.017
0167-2681/

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; 1999b; Banerjee and Munshi, 2004; Macchiavello and Morjaria, 2015; 2020; Dai et al., 2020b). Other papers in the theoretical networks literature have provided interesting insights into how structural network features of clusters such as degree dependence (hierarchy) and correlation (homophily) affect their resilience to shocks (Crespo et al., 2014) and 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; Ali et al., 2014). Long and Zhang (2011) use data from Chinese Industrial Censuses to show that clustering was positively associated with greater entry of new privately owned firms, lower capital size, and higher value added. Dai et al. (2020b) use Chinese firm registration data from 1990 to 2009 to examine the role of social networks. They show entrepreneurs from common hometown networks with high levels of informal trust and local cooperation (proxied by population density) achieved higher rates of firm entry, were 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 (Yoshino, 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 et al., 2019) and handloom clusters in Ethiopia (Zhang et al., 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 coexistence 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, while medium to large size firms feature 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. With few exceptions in high income countries (Behrens et al., 2020; Martin et al., 2013), little is known about comparative assessments of vulnerability to external risk which comprise an additional dimension of firm performance, or about mechanisms accounting for persistence of small firms in a volatile environment. 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. We use two distinct data sets. The first concerns registration of new firms in the entire country, including small and medium enterprises. The second is a longitudinal entrepreneur sample survey, including two phone surveys conducted in February and May 2020.[1] The shock arrived on the eve of the Chinese New Year in late January 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 also potentially significant. 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 shown in this paper, there was a sharp (approximately 70%) reduction in entry flows of new firms in the latter half of February 2020 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. Dai et al. (2020a) find 79.2% of incumbent firms were shut at the end of February. These rates were substantially higher than temporary shutdown rates (43% on average) in the US between late March and early April (Bartik et al., 2020).

Our primary empirical finding is that the Covid-19 impact on Chinese firms 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 our results are robust to cluster measures based on firm registration data in 2008, 2004 and 1995.[2] Counties with an above median cluster index featured

[1] See Dai et al. (2020a,b) and for further details of these two datasets.

[2] We do not report these robustness results in this paper; details are available on request.

a 67% reduction in entry (i.e., new firm registration) 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 from the entrepreneur surveys: a 1% rise in the cluster index was associated with a 0.05–0.07% higher likelihood of reopening in February after the New Year, and a 0.03–0.04% 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 attempts to disentangle the role of two specific attributes of clusters in affecting resilience: reliance on high density hometown entrepreneur networks, and spatial agglomeration. The role of the former attribute on entry, location, concentration and firm size is the focus of earlier papers by Peng (2004) and Dai et al. (2020b). Here we extend their analysis by examining network effects on resilience to the Covid-19 shock. The theoretical analysis of Dai et al. (2020b) (which forms the basis of our model) is set in a context where entrepreneurship is limited by poorly functioning markets for credit and technology and weak institutions for formal contract enforcement in LDCs. These barriers are overcome by informal cooperation among social networks of entrepreneurs in a common cluster. 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 join the cluster, and share credit, risks, infrastructure, technology, supplier and customer lists based on informal agreements. 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 spillovers also induce a form of adverse selection of entrepreneurial quality, wherein the cluster attracts lower ability entrepreneurs, resulting in greater dispersion of (and possibly lower average) quality, productivity and size of entering firms.

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 infections. The severity of the factor shortages is plausibly lower in a high density network owing to superior supply reliability and greater sharing of capital and risks. The model generates two main 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 shut down 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 that counties with entrepreneurs originating from hometowns with higher density experienced a smaller contraction in rates of registration of new firms following Covid-19. On the other hand, incumbent firms in such locations were less likely to reopen following the shock. These results are robust to controls for clustering, hometown heterogeneity and local infection rates.

However, the observed differences in impacts on rates of registration of new firms between the high and low cluster areas is only partially explained by the higher average hometown network density: despite controlling for the latter, high cluster areas still experienced significantly lower contraction. And since higher network density was associated with a lower likelihood of reopening among incumbents, it obviously cannot explain why high cluster counties experienced a higher likelihood of reopening (as clustering and network density were positively correlated). On both counts, this suggests that attributes other than network density also played a role in explaining the superior resilience of clusters. 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.[3]

As previously mentioned, the literature on resilience of firms to shocks is relatively sparse. The most closely related studies are of resilience of Canadian textile industries to external shocks between 2001 and 2013 (Behrens et al., 2020), and of French exporting firms to the 2008–2009 crisis (Martin et al., 2013). Their findings are different from ours: in the former

[3] They also suggest that the benefits of proximity to infrastructure or other networks at the same location were not anticipated by entrepreneurs prior to their decision to enter, while they did anticipate the benefits of membership in their own network. Otherwise spatial agglomeration would have given rise to a similar form of adverse selection as the own-network spillovers, which would have been manifested in a lower (rather than higher) reopening rate among incumbents after controlling for average hometown density.

paper there was little evidence that clusters were more resilient, while the latter paper found firms in clusters were **less** resilient, possibly because of their dependence on the fate of their 'leader' firm. Bartik et al. (2020) examine variations in shutdown rates in a survey of 5800 US businesses carried out between March 28 and April 4, 2020. They find an average 43% rate of temporary shutdown and 2% permanent shutdown, that small businesses were more adversely affected, and that closure rates varied with beliefs about the duration of Covid related disruptions, particularly demand reductions and employee health concerns. Crane et al. (2020) investigate permanent shutdown rates in the US based on later data, and how these varied across different industries. These papers on US industries did not investigate how resilience varied between clusters and other firms.

Section 2 presents the theoretical model. Section 3 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 $t = 1, 2, \ldots$ 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 at a given location i 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.[4] 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 ω denotes the entrepreneur's own talent, and A denotes the 'community-TFP'. At any date t and destination i, community TFP A_{it} depends on the size $n_{i,t-1}$ and quality $\theta(p)$ of the network, according to

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

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.

Underlying this specification is the implicit assumption of no interaction with other networks at the same location. Most clusters feature co-existence of firms from multiple networks, who may cooperate with one another, which the Dai et al. (2020b) model abstracts from. Their model could be extended to incorporate complementarity among multiple networks, representing positive spillovers across networks (possibly weaker in intensity compared with within-network spillovers). A_0 could include the benefits of spatial agglomeration, or spillovers from the presence of other networks at the same location. Counties with higher A_0 would then be areas with a higher cluster index, which could be host to many different networks of varying density. In such an extension, entry rates would increase both in clustering (i.e., A_0) and average density of networks located there. A_0 would depend on t as well, reflecting growth of incumbents from other networks, which would complicate the expressions derived below.[5] However, the extension is conceptually straightforward, and involves replacing A_0 in the expressions below by a time-varying parameter whose size and growth depends on the extent of spatial proximity to other networks.

Consequent on entering, a date t cohort agent of quality ω would select scale K to maximize profits $A_{it}\omega^{1-\alpha}K^\alpha - rK$, resulting in:

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

[4] 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).

[5] To fit the empirical patterns on reopening rate of incumbents following the Covid-19 shock, the extended model would allow these location-specific benefits to be not (or less well) anticipated by newly born agents prior to making their entry decisions. In other words, newly born agents correctly anticipate own-network benefits of becoming an entrepreneur in a given location, but could undervalue the advantages of spatial proximity to other networks in the cluster. Then areas with high A_0's will be associated with a high cluster index, but would be less subject to adverse selection associated with high network density.

(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:

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

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 increasing in t. 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. On the other hand, 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 $(1+\Delta)\xi(1) > 1$ 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 shock 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 locally increasing in p.

Proposition 1. *Consider any combination of a specific destination and hometown origin, and let $n_t(p)$ denotes the total size of the incumbent network from the hometown that has entered this destination by date t, i.e., cumulating across all cohorts prior to t. Suppose that t is not too distant from T in the sense that*

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

Suppose also that $\xi(p) = Z\exp(-\eta p)$ where $\eta > 0$ and $Z(1+\Delta) > 1$. Take any two networks with densities p_L, p_H respectively where $p_L < p_H$. Then for a range of values of η sufficiently small and a range of shock intensities Δ, 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 $\underline{\omega}_t(p)$ given by (4). But for Δ large enough this inequality will be reversed and some low ability incumbents will shut down.

The range of operating profits across cohort t incumbents (if they all remained open at T) will correspond to (8) over a range of ability exceeding the threshold $\underline{\omega}_t(p)$. So the minimum operating profit for incumbents of this cohort at date T is $\underline{\Pi}_t(p;\Delta)$ (equal to (8) evaluated at the entry threshold $\underline{\omega}_t(p)$):

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

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) implies the adverse selection effect persists beyond the entry date t until the date T when the shock hits.[6] Then the lower bound of operating profits in the high density network is smaller, 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 a sufficiently large shock, all incumbents in the low density network would shut down, while a positive fraction of those in the high density network (the extreme upper tail) continue 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 (Ellison and Glaeser, 1997). 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, 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.

Tongxiang county which the Puyuan cashmere sweater cluster belongs to 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)). A vertically integrated firm would include all of these different stages in the form of different divisions within the firm. The result would be 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

$$\phi^e_{ij} = \min\{P(LQ^e_{ri} > 1|LQ^e_{rj} > 1), P(LQ^e_{rj} > 1|LQ^e_{ri} > 1)\}$$

where P denotes conditional probability and $LQ^e_{rj} \equiv \frac{E_{rj}/E_r}{E_j/E}$. 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.[7]

[6] 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. The ESIEC survey we rely on included firms established only since 2010, so this assumption seems reasonable.

[7] Same proximity indices between two industries could be based either on output, capital or number of firms. In this paper,we take an simple average of proximity indices based on employment, output, capital and number of firms respectively.

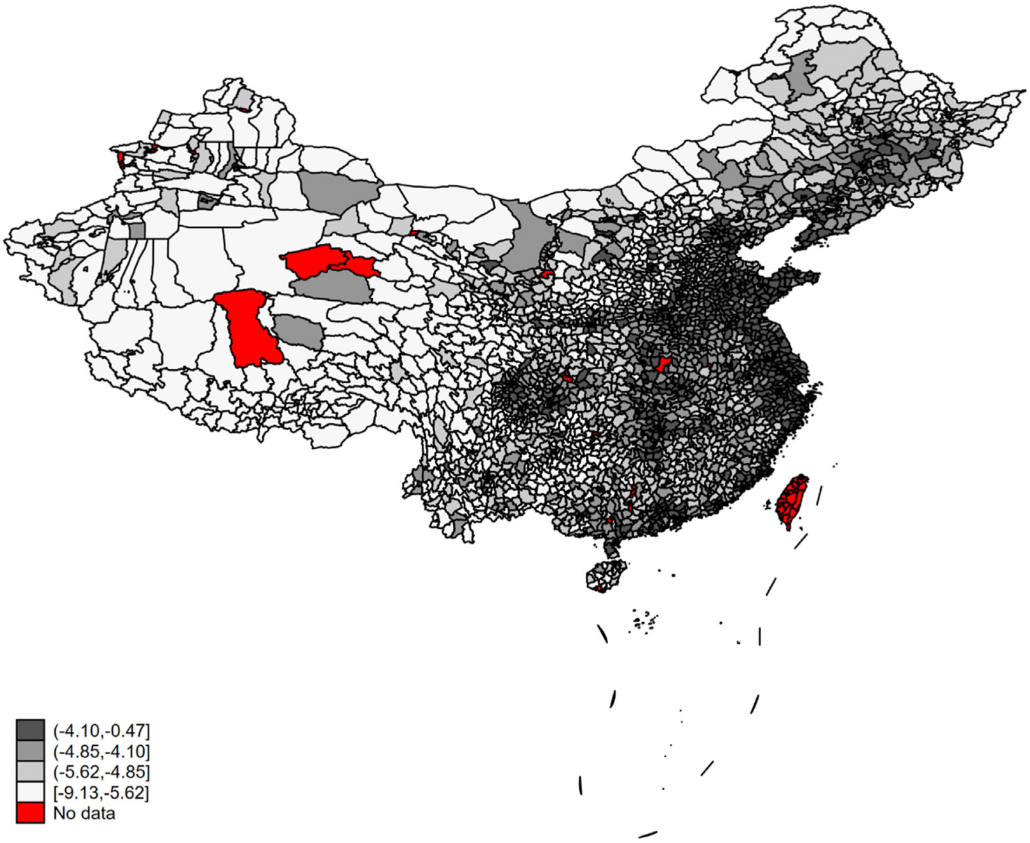

Fig. 1. Map of cluster index across China.
Source: author calculation based on the 2008 China Economic Census.

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, capital or number of firms weights in the averaging procedure provides an alternative measure of clustering. The overall Ruan–Zhang (RZ) cluster index takes the simple average across employment, output, capital and number of firms 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 and HHI. 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 Fig. 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.

Fig. 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.[8] 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 (besides dependence on different data sources).

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

[8] In Fig. 2 the cluster index is based on the firm registration database, in order to facilitate comparisons across different points of time. In the empirical analysis in the rest of the paper, we use the 2008 index based on Census Data in Ruan and Zhang (2015). The cluster index based on Census Data in 1995 and 2004 are not comparable with 2008, since the latter includes all sectors, while the former exclude agriculture and service sector firms.

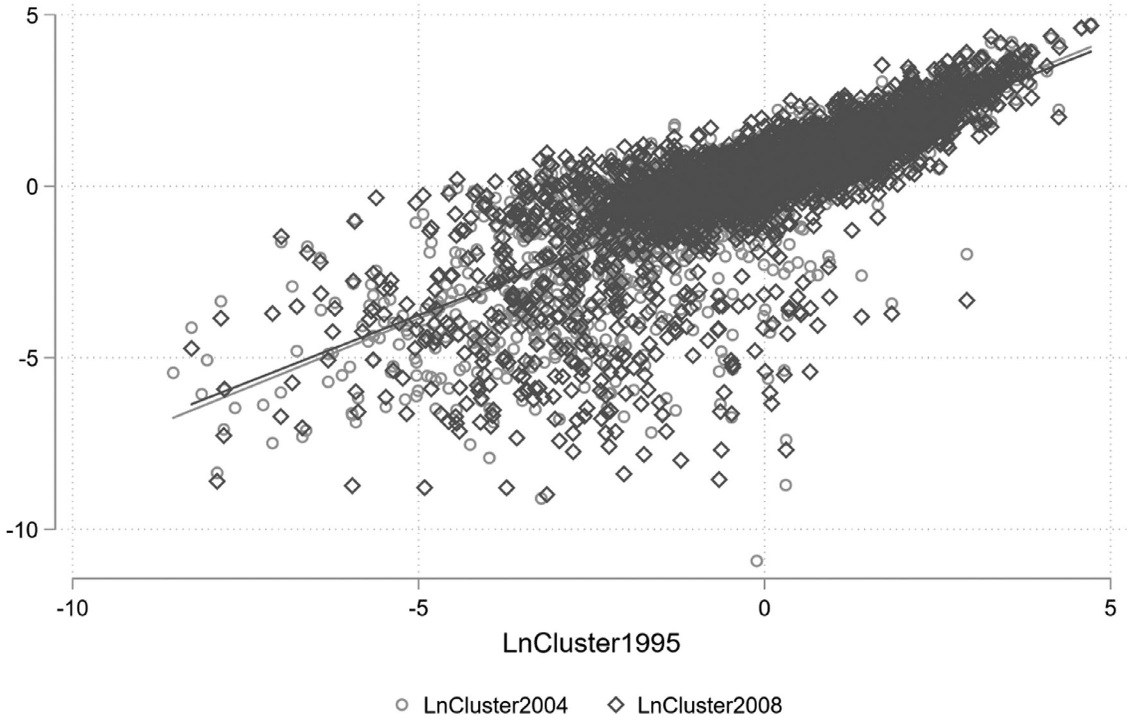

Fig. 2. Log cluster index 2004, 2008 vs 1995 (SAIC registration sata).
Source: author calculation based on the firm registration database in 1995, 2004 and 2008.

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 January 2017 and June 2020. The data also permits us to identify the birth county of the principal legal 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. Figs. 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.[9] 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. Fig. 5 compares different attributes between counties with high (above median) and low (below median) cluster index. The attributes are dummy variables indicating whether a firm has stable suppliers (StableSup), whether it has stable clients (StableCon), whether it sells on credit to their largest client (MSellCredit), whether its primary supplier is located in the same country (MLocalSup), whether its primary customer is located in the same county (MLocalCon), whether it has undertaken a process innovation (NewProcess), and whether it has positive online sales (Online). Moreover, we show percentage of employees who are local residents (LocalEmpRa), and inventory as a proportion of working capital (Stock).

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. This is verified in Table 1 which 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.

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

[9] Both urban and rural counties are included.

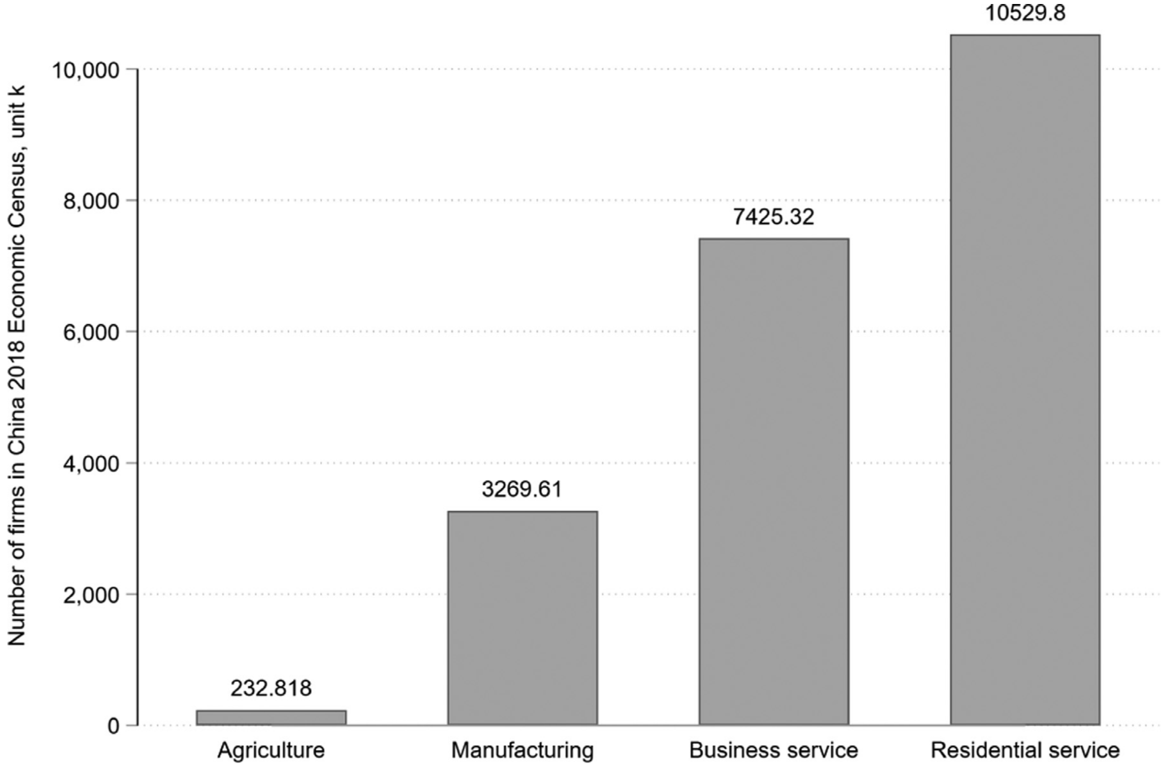

Fig. 3. Number of registered firms in 2018 by industry group.
Source: the 2018 China Economic Census.

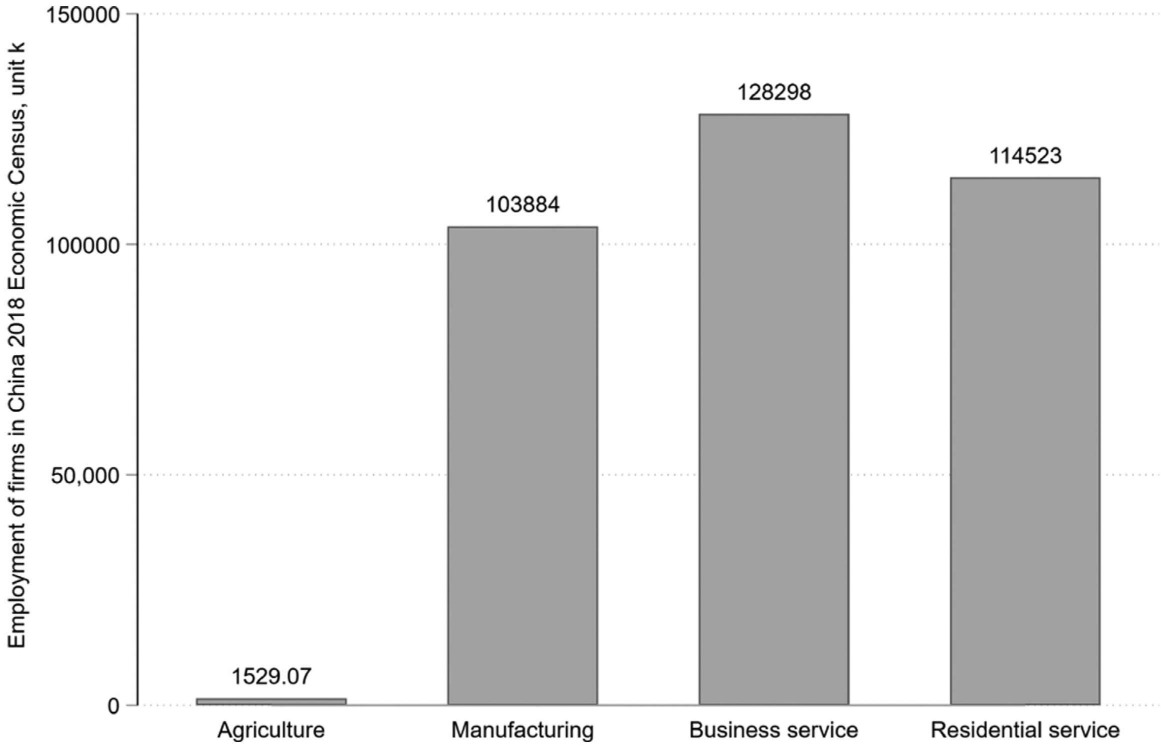

Fig. 4. Employment in 2018 by industry group.
Source: the 2018 China Economic Census.

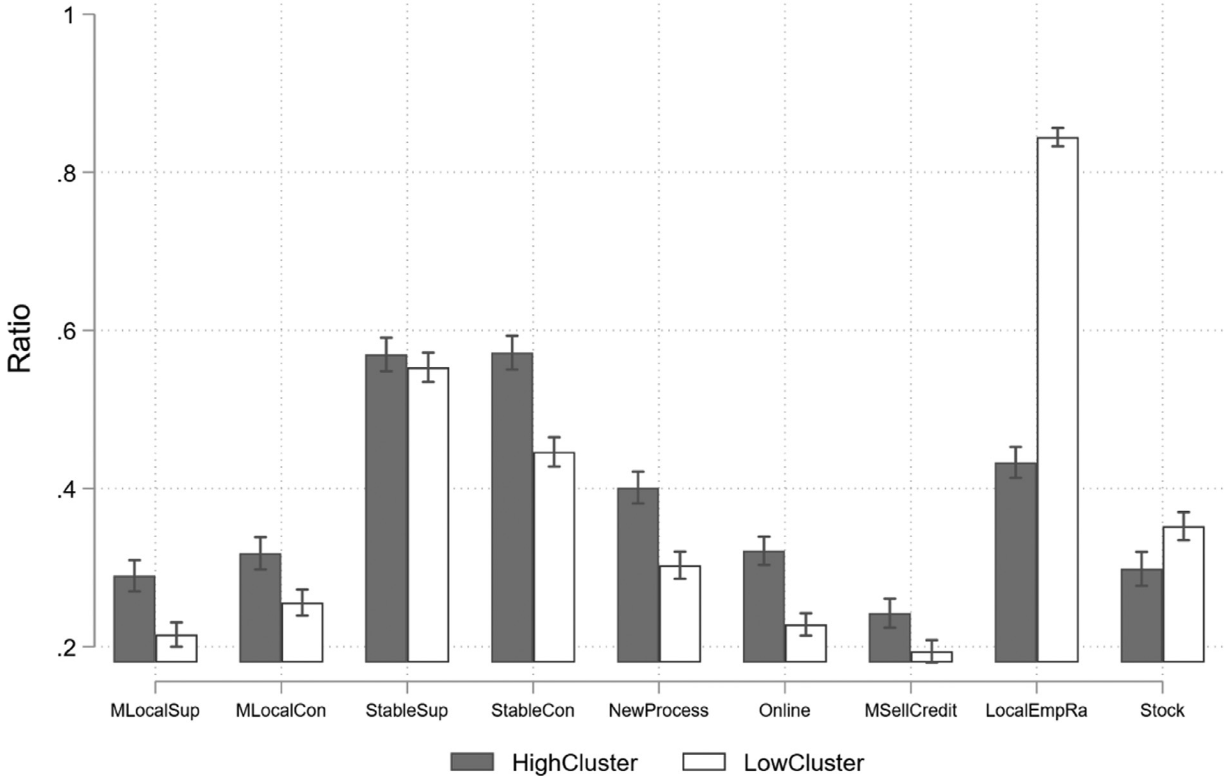

Fig. 5. Firm attributes, high vs low cluster counties.
Source: ESIEC 2017–2020. Counties are grouped to HighCluster (with cluster index above the median) and LowCluster (with cluster index below the median). Variables in *x*-axis indicate whether a firm's primary supplier is located in the same country (MLocalSup), whether its primary customer is located in the same county (MLocalCon), whether it has stable suppliers (StableSup), whether it has stable clients (StableCon), whether it has undertaken a process innovation (NewProcess), whether it has positive online sales (Online), and whether it sells on credit to their largest client (MSellCredit). LocalEmpRa is the firm-level percentage of employees who are local residents. Stock is the share of inventory of working capital, from ESIEC 2020. Mean of each variable of two groups is reported. The vertical line corresponding to the bar represents 95% confidence interval.

Table 1
Cluster index and firm attributes.

Dependent variables	(1) StableSup	(2) StableCon	(3) MLocalSup	(4) MLocalCon	(5) NewProcess	(6) Online	(7) MSellCredit	(8) LocalEmpRa
LnCluster	0.005	0.049	0.026	0.016	0.033	0.030	0.016	−0.142
	(0.008)	(0.008)	(0.006)	(0.009)	(0.007)	(0.008)	(0.008)	(0.019)
Constant	0.578	0.678	0.343	0.342	0.461	0.375	0.275	0.146
	(0.032)	(0.030)	(0.025)	(0.033)	(0.022)	(0.031)	(0.029)	(0.064)
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4795	4739	4708	4657	4967	5857	5000	4109
Adjusted R-squared	0.024	0.046	0.007	0.011	0.015	0.014	0.019	0.229

Note: dependent variables are dummy variables at the firm level, indicating whether a firm has stable clients (StableCon), whether it sells on credit to their largest client (MSellCredit), whether its primary supplier is located in the same country (MLocalSup), whether its primary customer is located in the same county (MLocalCon), whether it has stable suppliers (StableSup), whether it has undertaken a process innovation (NewProcess), and whether it has positive online sales (Online). LocalEmpRa is the firm-level percentage of employees who are local residents. The dependent variables calculated from ESIEC 2017–2019. LnCluster is the county level clustering index, calculated from 2008 Census Data. Industry FE includes Agriculture, Manufacturing, Business Services and Residential Services fixed effects. Standard errors clustered in county level reported in parentheses.

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.

Fig. 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

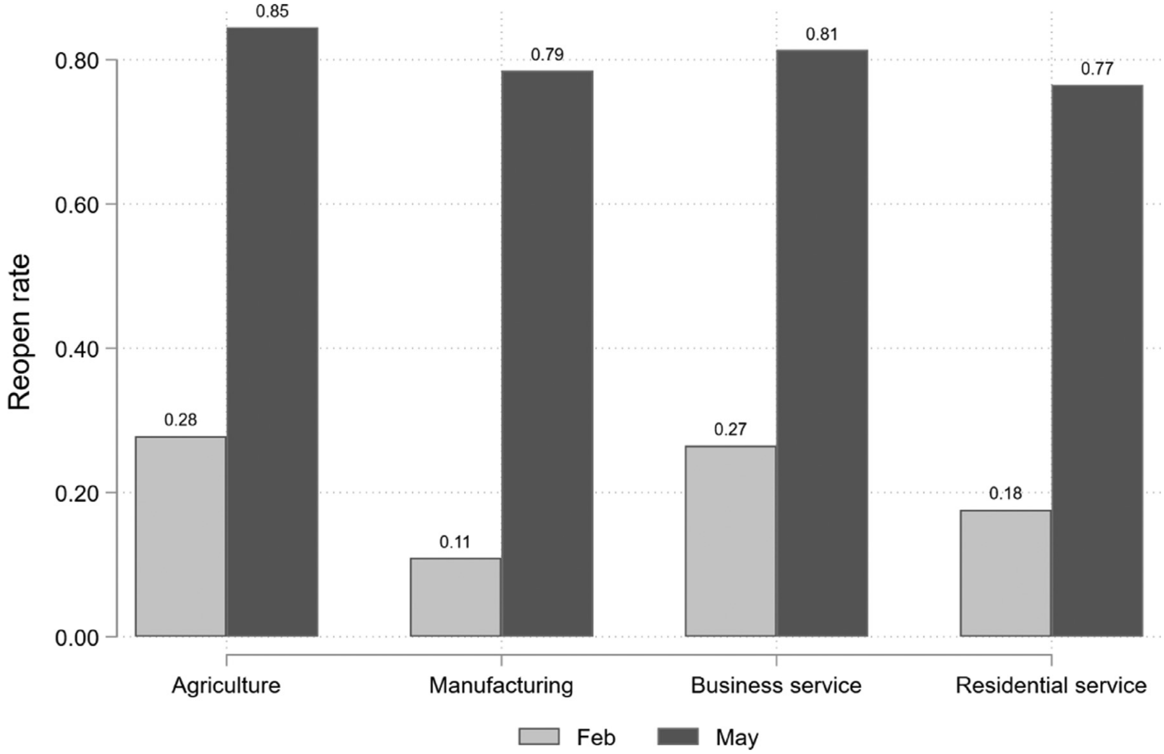

Fig. 6. Percent firms reopening after new year in February & May 2020, by industry group.
Source: author calculation based on ESIEC 2020.

77% and 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. Fig. 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. Fig. 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 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. Since most clusters are located in rural counties, we restrict our sample to these counties. To measure the connectedness among entrepreneurs at a given rural county, we use the weighted average of population density of birth county of the entrepreneurs (i.e., listed legal representative).[10] Since most entrepreneurs operating in rural counties were themselves born in rural areas, entrepreneurs born in urban counties are not included while computing the average population density. Population density is based on the 1982 Census, 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.[11]

[10] The weight is the share of entrepreneurs from each birth place at the given location in 2015. Entrepreneurs from the local area are included.

[11] As explained below, we focus on entry into rural county locations, where most firm entrepreneurs came from rural county hometowns. Hence the concentration index pertains mainly to rural county hometowns.

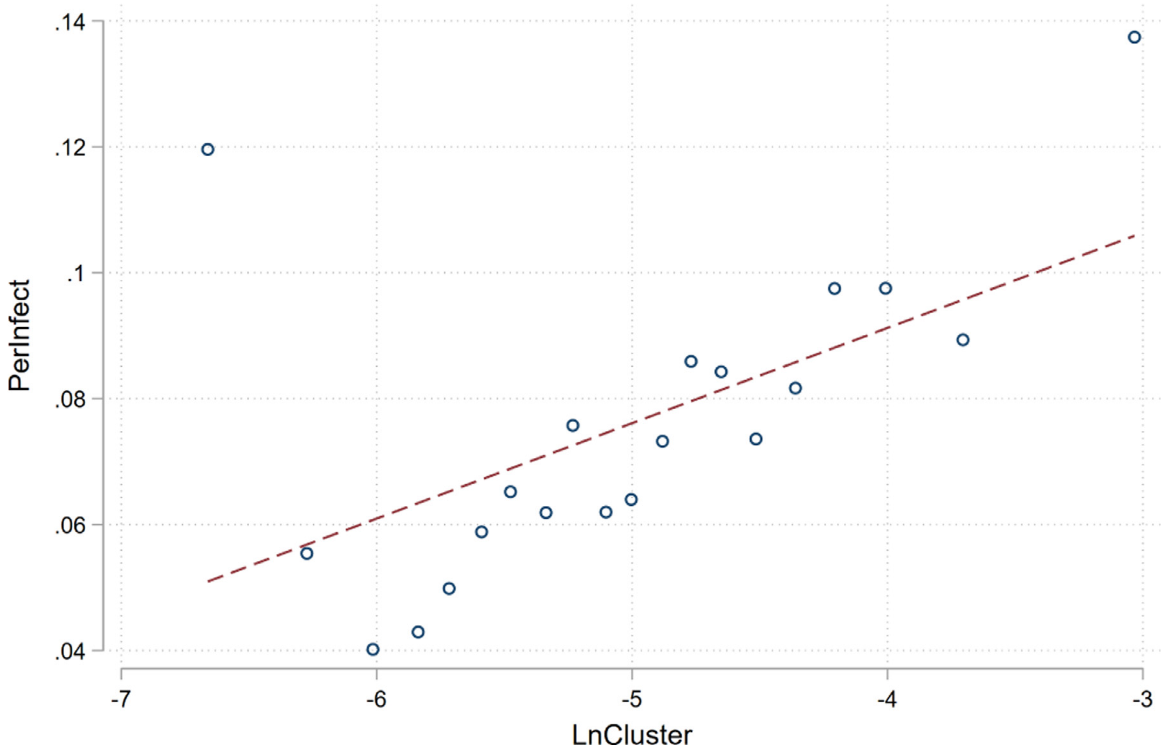

Fig. 7. Scatterplot of (log) cluster index vs local COVID-19 infection rate.
Source: author calculation. LnCluster is the log value of clustering index. PerInfect is each city's number of cumulative infected cases of Covid-19 among 10 thousand people till 17th Feb.

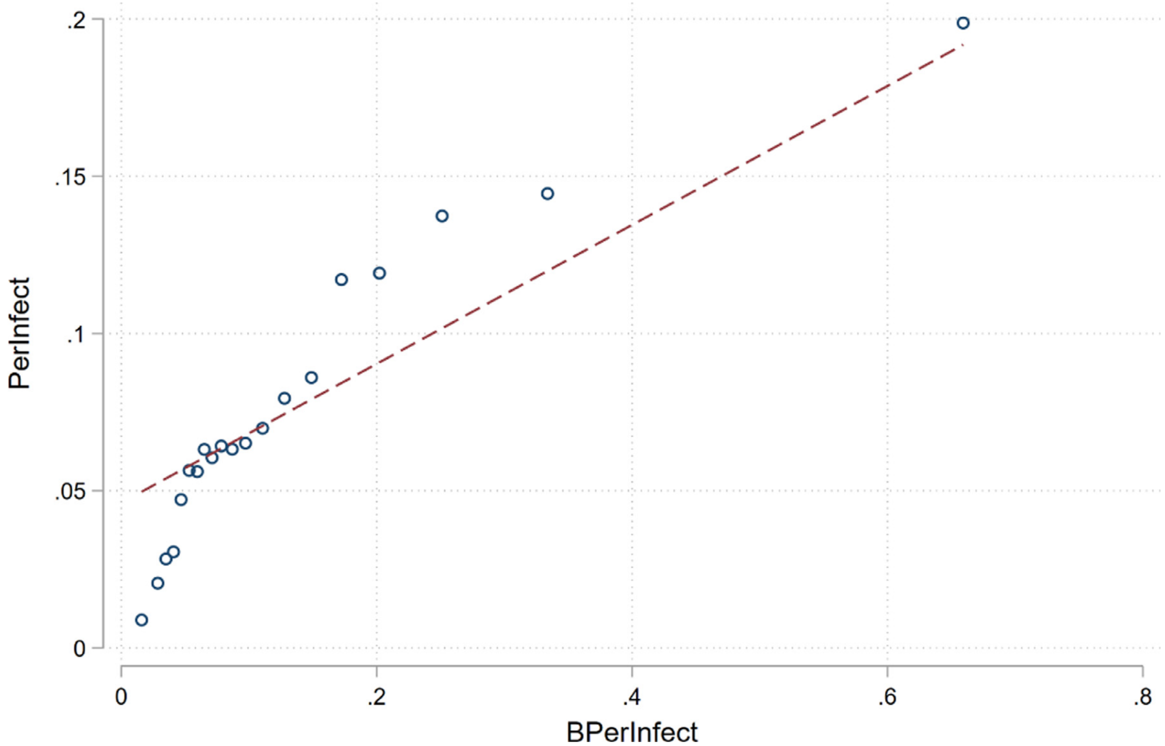

Fig. 8. Scatterplot of COVID-19 infection rate: local vs entrepreneur hometown.
Source: author calculation. PerInfect is each city's number of cumulative infected cases of Covid-19 among 10 thousand people till 17th Feb. BPerIfect is the weighted average of infection rates of entrepreneurs' birth places.

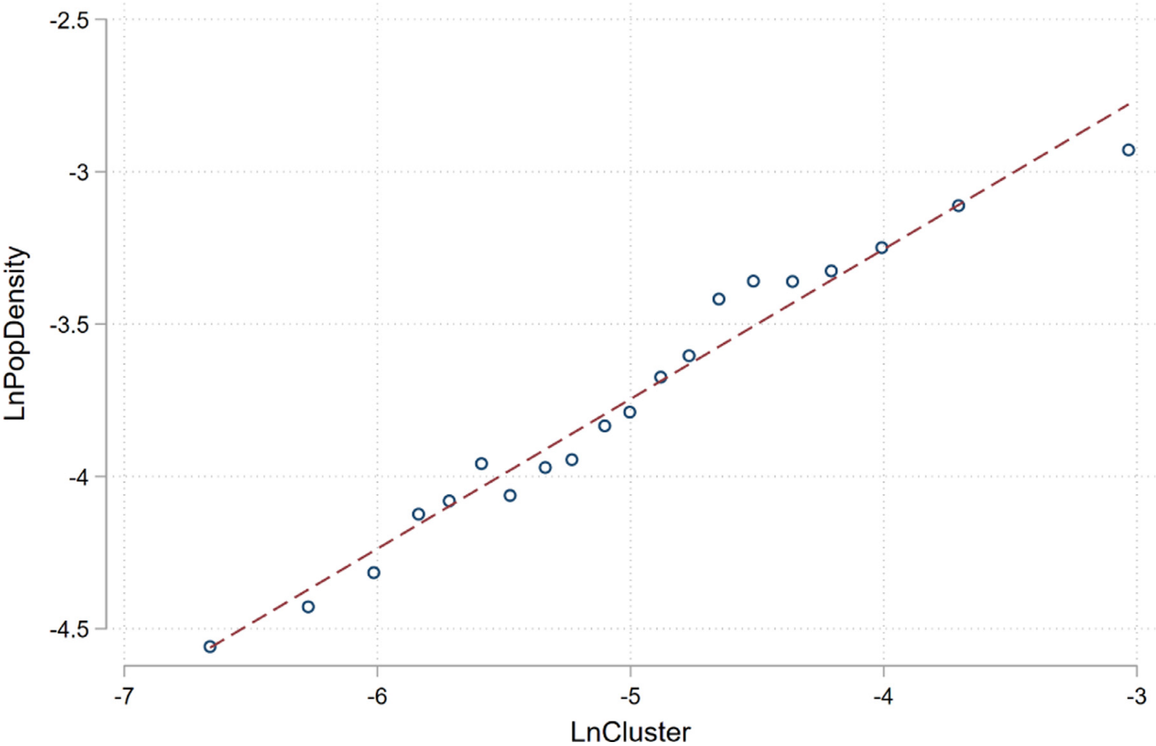

Fig. 9. Scatterplot of (log) cluster index vs (log) population density of entrepreneur hometown.
Source: author calculation. LnCluster is the log value of clustering index. LnPopDensity is the log value of weighted average population density of entrepreneurs' birth counties.

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. Figs. 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

For the entry analysis, we use monthly firm registration data at the county-industry level from 2017 to 2020. We include only private firms located in rural counties. Similar results obtain when we analyze weekly rather than monthly data, but do not show these results as they are less reliable owing to a greater frequency 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. The ESIEC survey data has already been described above: we use responses from the February and May 2020 phone surveys, in conjunction with firm characteristics based on 2017–2019 surveys. The cluster index is calculated based on the 2008 China Economic Census, while Covid-19 infection rates are based on public domain data. See the Data Appendix for further details.

With log of new per capita firm entries at the county-industry-month-year as the dependent variable, Fig. 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.

Fig. 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 :

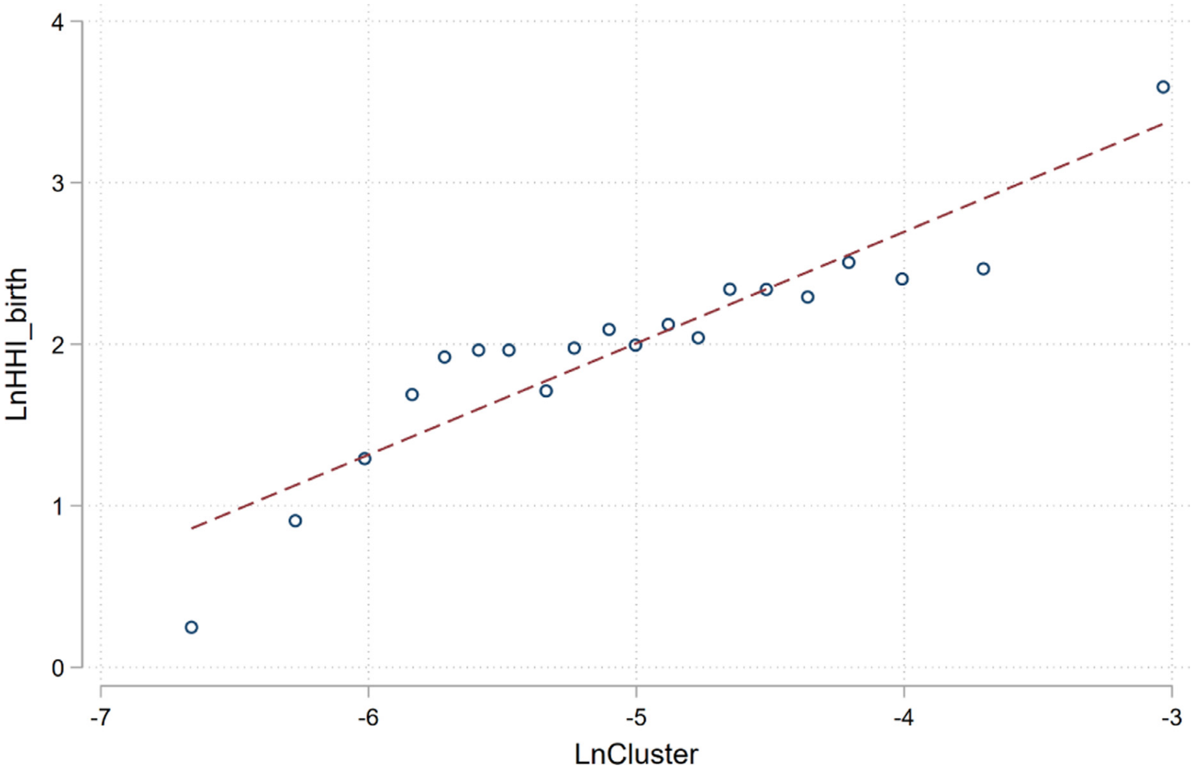

Fig. 10. Scatterplot of (log) cluster index vs spatial concentration (log HHI) of entrepreneur hometown.
Source: author calculation. LnCluster is the log value of clustering index. LnHHI_birth is the log value of the Hirschman–Herfindahl Index (HHI) of entrepreneurs' birth places.

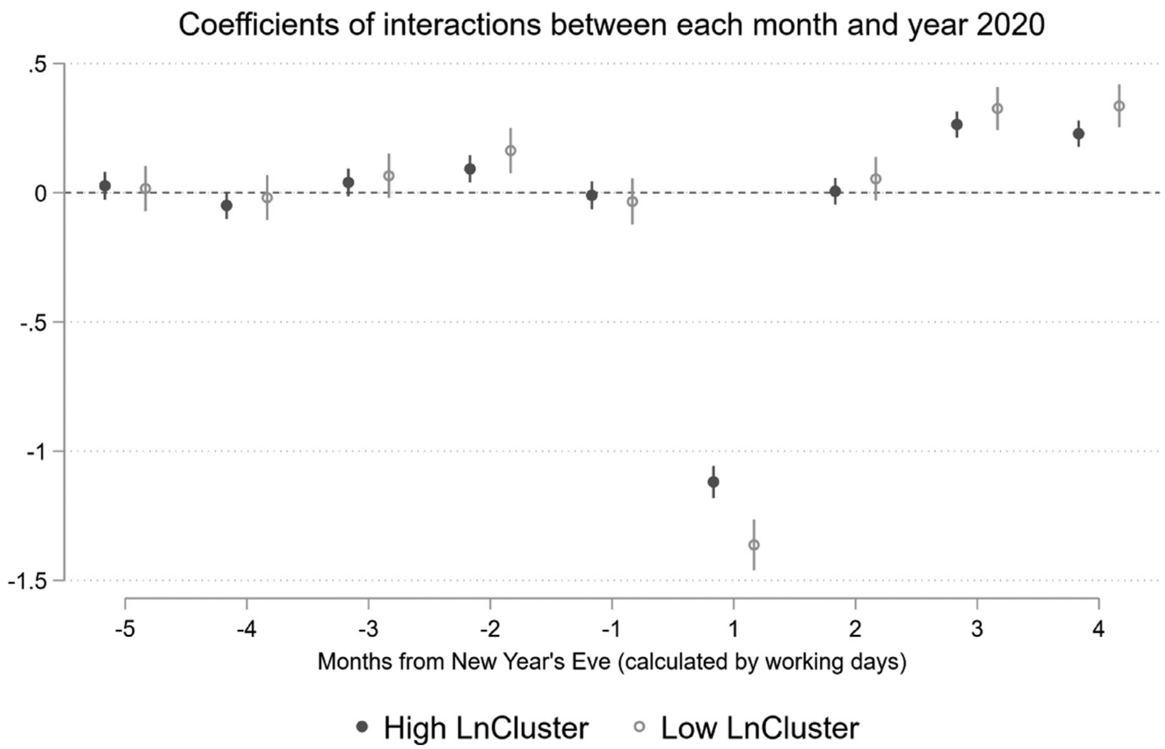

Fig. 11. Per capita entry regression interactions between month dummies and 2020, separately by high and low cluster counties. Note: Coefficients and their 95% confidence intervals plotted at the figure. Dependent variable is the natural logarithm number of new firms per capita at the county-industry-year-month level. Independent variables are interaction terms between the 2020 year dummy and each month dummy. Six months before the Chinese New Year and four months after included. The year 2020 dummy is controlled. Month FE, county FE, industry FE included.

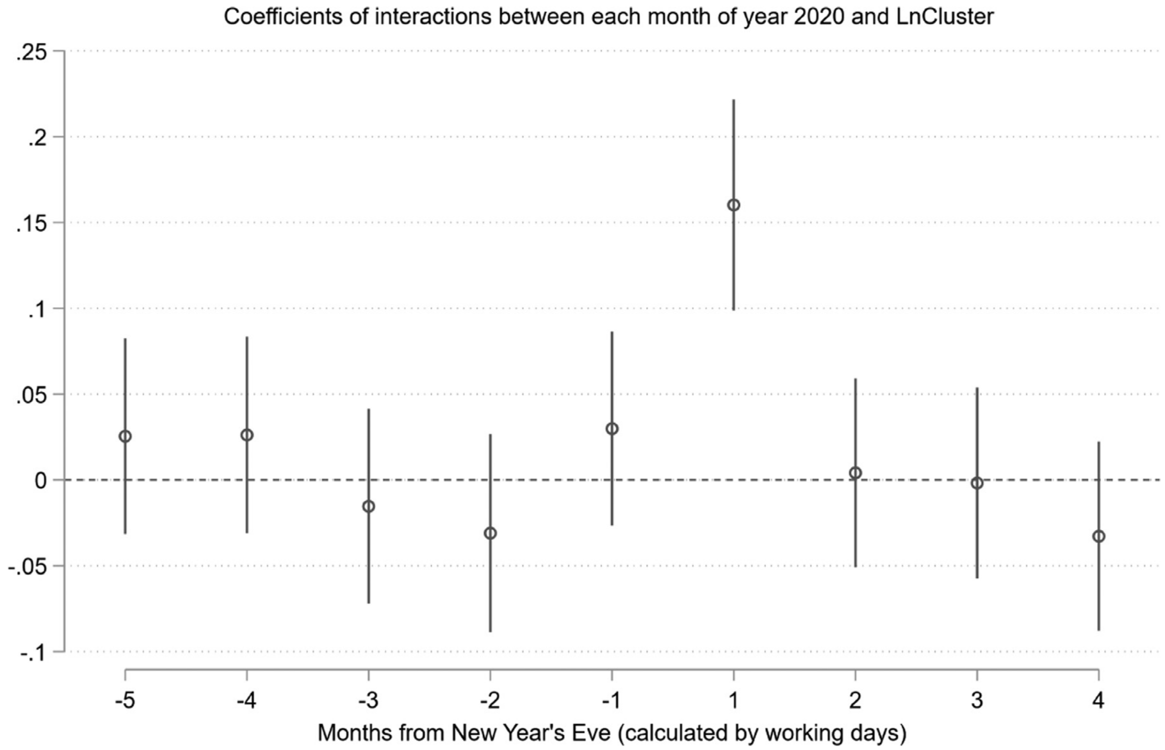

Fig. 12. Per capita entry regression interactions between LnCluster and number of months from new year. Note: Coefficients and their 95% confidence intervals plotted at the figure. Dependent variable is the natural logarithm number of new firms per capita at county-industry-year-month level. Independent variables are interaction terms between the LnCluster and each month dummy before and after the Chinese New Year. Average Covid-19 infection rates of the entrepreneurs' hometown and local Covid-19 infection rate are controlled. Year-county-industry FE, year-month FE and month-county-industry FE included.

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)) where 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 Fig. 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% and 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 infection 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.[12] 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.

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

Table 2
Per capita entry regression coefficients: infection rates, interactions between cluster index and number of months from new year.

Dependent variable	(1) Per capita entry
LnCluster#Month -5	0.025
	(0.029)
LnCluster#Month -4	0.026
	(0.029)
LnCluster#Month -3	−0.015
	(0.029)
LnCluster#Month -2	−0.031
	(0.029)
LnCluster#Month -1	0.030
	(0.029)
LnCluster#Month 1	0.160
	(0.031)
LnCluster#Month 2	0.004
	(0.028)
LnCluster#Month 3	−0.002
	(0.028)
LnCluster#Month 4	−0.033
	(0.028)
BPerInfect	0.100
	(0.101)
PerInfect	−0.791
	(0.191)
Constant	−1.613
	(0.025)
Year-month FE	Yes
Year-county-industry FE	Yes
Month-county-industry FE	Yes
Observations	231,080
Adjusted R-squared	0.673

Note: dependent variable is the natural logarithm number of new firms entry per capita at the county-industry-year-month level. LnCluster is the county level clustering index, calculated from 2008 Census Data. LnCluster#Month x is the interaction term of LnCluster and year 2020 dummy and month x (x indicating the number of months before the New Year if x is less than zero, x indicating the number of months after the New Year if x is larger than zero). PerInfect is each city's number of cumulative infected cases of Covid-19 among 10 thousand people till 17th Feb. BPerIfect is weighted average infected rates of entrepreneurs' birth place. A few provinces (Xinjiang, Qinghai, Tibet, and Inner Mongolia), which have large pastural areas, are excluded. Hubei Province, the epicenter of Covid-19 pandemic, is excluded also. To be consistent with Dai et al. (2020b), only rural counties are included. Robust standard errors reported in parentheses.

Table 3
Firm reopening likelihood regression on infection rate and cluster index.

Dependent variables	(1) RunWell2	(2)	(3)	(4)	(5) RunWell5	(6)
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)
Industry FE	Yes	Yes	No	No	Yes	No
SIC-1 FE	No	No	Yes	Yes	No	Yes
Observations	1715	1715	1715	1715	1825	1825
Adjusted R-squared	0.037	0.043	0.047	0.051	0.008	0.037

Note: RunWell2 is a dummy variable indicating whether firm has reopened on 10th February. RunWell5 is a dummy variable indicating whether firm is in operation in May. The two dependent variables are obtained through the two waves of ESIEC phone survey in Feb and May 2020. LnCluster is the county level clustering index, calculated from 2008 Census Data. PerInfect is each city's number of cumulative infected cases of Covid-19 among 10 thousand people till 17th Feb. OffSeason is a dummy variable indicating whether firm is in the off season in Feb (if dependent variable is RunWell2) or in May (if dependent variable is RunWell5), obtained from ESIEC 2018. Industry FE includes Agriculture, Manufacturing, Business Services and Residential Services fixed effects. SIC-1 FE includes one-digit industry classification fixed effects. Standard errors clustered in county level reported in parentheses.

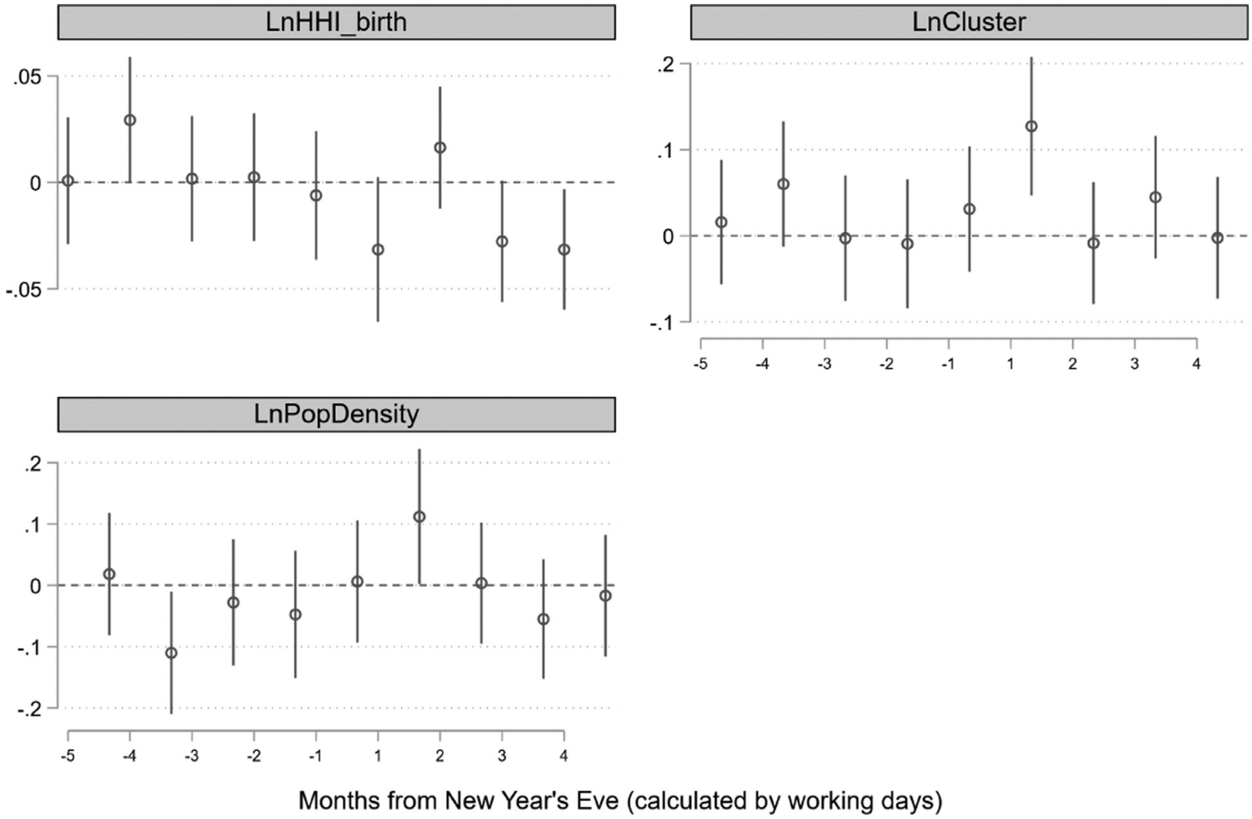

Fig. 13. Per capita entry regression interactions between number of months from new year and (logs of) hometown spatial concentration, population density and cluster index. Note: Coefficients and their 95% confidence intervals plotted at the figure. Dependent variable is the natural logarithm number of new firms per capita at the county-industry-year-month level. Independent variables are interaction terms between the LnCluster/LnHHI_birth/nPopDensity and each month dummy before and after the Chinese New Year. Average Covid-19 infection rates of the entrepreneurs' hometown and local Covid-19 infection rate are controlled. Year-county-industry FE, year-month FE and month-county-industry FE included.

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 Fig. 13. The interaction coefficients of hometown concentration are not significant. Table 4 shows the estimated regression coefficients for infection rates, and 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 Fig. 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.[13] 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 de-

[13] 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 celebration of the 70th year anniversary of the founding of People's Republic of China.

Table 4
Per capita entry regression coefficients: infection rates, interactions between cluster index/hometown population density and number of months from new year.

Dependent variable	(1) Per capita entry
LnPopDensity#Month 1	0.112
	(0.056)
LnPopDensity#Month 2	0.004
	(0.050)
LnPopDensity#Month 3	−0.055
	(0.050)
LnPopDensity#Month 4	−0.017
	(0.051)
LnCluster#Month 1	0.127
	(0.041)
LnCluster#Month 2	−0.009
	(0.033)
LnCluster#Month 3	0.045
	(0.036)
LnCluster#Month 4	−0.002
	(0.036)
BPerInfect	0.093
	(0.101)
PerInfect	−0.806
	(0.191)
Constant	−1.610
	(0.028)
Year-month FE	Yes
Year-county-industry FE	Yes
Month-county-industry FE	Yes
Observations	231,080
Adjusted R-squared	0.673

Note: dependent variable is the natural logarithm number of new firms entry per capita at the county-industry-year-month level. LnCluster is the county level clustering index, calculated from 2008 Census Data. LnPopDensity is weighted average population density of rural-born entrepreneurs' birth county. LnCluster (LnPopDensity) # Month x is the interaction term of LnCluster (LnPopDensity) and year 2020 dummy and month x (x indicating the number of months after the New Year). PerInfect is each city's number of cumulative infected cases of Covid-19 among 10 thousand people till 17th Feb. BPerIfect is weighted average infected rates of entrepreneurs' birth place. A few provinces (Xinjiang, Qinghai, Tibet, and Inner Mongolia), which have large pastural areas, are excluded. Hubei Province, the epicenter of Covid-19 pandemic, is excluded also. To be consistent with Dai et al. (2020b), only rural counties are included. Robust standard errors reported in parentheses.

Table 5
Firm reopening likelihood regression on infection rate, cluster index, hometown population density and hometown concentration.

Dependent variables	(1) RunWell2	(2)	(3)	(4)	(5) RunWell5	(6)	(7)	(8)
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)
Industry FE	Yes	Yes	No	No	Yes	Yes	No	No
SIC-1 FE	No	No	Yes	Yes	No	No	Yes	Yes
Observations	1699	1699	1699	1699	1816	1816	1816	1816
Adjusted R-squared	0.026	0.046	0.039	0.055	0.010	0.024	0.041	0.055

Note: RunWell2 is a dummy variable indicating whether firm has reopened on 10th February. RunWell5 is a dummy variable indicating whether firm is in operation in May. The two dependent variables are obtained through the two waves of ESIEC phone survey in Feb and May 2020. LnCluster is the county level clustering index, calculated from 2008 Census Data. LnPopDensity is weighted average population density of entrepreneurs' birth county for each county-industry. LnHHI_birth is adjusted Hirschman–Herfindahl Index(HHI) of entrepreneurs' birth home counties, excluding the destination county. PerInfect is each city's number of cumulative infected cases of Covid-19 among 10 thousand people till 17th Feb. OffSeason is a dummy variable indicating whether firm is in the off season in Feb (if dependent variable is RunWell2) or in May (if dependent variable is RunWell5), obtained from ESIEC 2018. Industry FE includes Agriculture, Manufacturing, Business Services and Residential Services fixed effects. SIC-1 FE includes one-digit industry classification fixed effects. Standard errors clustered in county level reported in parentheses.

Table 6
Firm ReOpening (Feb 2020) likelihood regression on average firm attributes, hometown population density and hometown concentration.

Dependent variables	(1) RunWell2	(2)	(3)	(4)	(5)	(6)
LnPopDensity	0.006	0.005	0.002	0.007	0.004	0.011
	(0.015)	(0.016)	(0.015)	(0.016)	(0.015)	(0.018)
LnHHI_birth	0.005	0.001	0.002	0.005	0.007	0.005
	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)	(0.011)
MeanStableSup	0.420	–	–	–	–	–
	(0.103)					
MeanStableCon	–	0.426	–	–	–	–
		(0.103)				
MeanMLocalSup	–	–	0.782	–	–	–
			(0.197)			
MeanMLocalCon	–	–	–	0.560	–	–
				(0.167)		
MeanOnline	–	–	–	–	0.105	–
					(0.163)	
MeanLocalEmpRa	–	–	–	–	–	−0.033
						(0.107)
Constant	0.140	0.151	0.148	0.167	0.196	0.245
	(0.066)	(0.070)	(0.067)	(0.073)	(0.072)	(0.081)
SIC-1 FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1732	1732	1732	1732	1731	1532
Adjusted R-squared	0.051	0.049	0.053	0.045	0.038	0.031

Note: RunWell2 is a dummy variable indicating whether firm has reopened on 10th February. Explanation variables are at the county level, including the proportion of firms in the county whose primary supplier is located in the same country (MeanMLocalSup), whose primary customer is located in the same county (MeanMLocalCon), which have stable suppliers (MeanStableSup), which have stable clients (MeanStableCon), which have positive online sales (MeanOnline). MeanLocalEmpRa is the county-level average of each firm's percentage of employees who are local residents (MeanLocalEmpRa). LnHHI_birth is adjusted Hirschman–Herfindahl Index (HHI) of entrepreneurs' birth home counties, excluding the destination county. SIC-1 FE includes one-digit industry classification fixed effects. Standard errors clustered in county level reported in parentheses.

picted 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.

5.2. 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 the average of local firms' attributes. These attributes directly measure the spatial proximity between firms within clusters from multiple perspectives, as suggested in Table 1. We continue to control for hometown network quality. Firms in counties with larger share of firms which have stable suppliers and customers were more likely to reopen. The same is true for counties with higher ratio for local suppliers and customers, and for those selling online (significant in May). Counties 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 county with higher firm productivity tends 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,

Table 7
Firm ReOpening (May 2020) likelihood regression on average firm attributes, hometown population density and hometown concentration.

Dependent variables	(1) RunWell5	(2)	(3)	(4)	(5)	(6)
LnPopDensity	−0.062	−0.058	−0.063	−0.057	−0.060	−0.053
	(0.019)	(0.018)	(0.019)	(0.017)	(0.018)	(0.020)
LnHHI_birth	0.011	0.007	0.010	0.009	0.008	0.008
	(0.012)	(0.011)	(0.012)	(0.011)	(0.011)	(0.013)
MeanStableSup	0.225	–	–	–	–	–
	(0.095)					
MeanStableCon	–	0.361	–	–	–	–
		(0.096)				
MeanMLocalSup	–	–	0.324	–	–	–
			(0.158)			
MeanMLocalCon	–	–	–	0.544	–	–
				(0.167)		
MeanOnline	–	–	–	–	0.415	–
					(0.157)	
MeanLocalEmpRa	–	–	–	–	–	−0.185
						(0.080)
Constant	0.513	0.517	0.526	0.522	0.519	0.648
	(0.080)	(0.077)	(0.080)	(0.074)	(0.075)	(0.092)
SIC-1 FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1870	1869	1870	1869	1870	1591
Adjusted R-squared	0.041	0.046	0.040	0.045	0.045	0.037

Note: RunWell5 is a dummy variable indicating whether firm has reopened in May. Explanation variables are at the county level, including the proportion of firms in the county whose primary supplier is located in the same country (MeanMLocalSup), whose primary customer is located in the same county (MeanMLocalCon), who have stable suppliers (MeanStableSup), who have stable clients (MeanStableCon), who have positive online sales (MeanOnline). MeanLocalEmpRa is the county-level average of each firm's percentage of employees who are local residents (MeanLocalEmpRa). LnHHI_birth is adjusted Hirschman–Herfindahl Index(HHI) of entrepreneurs' birth home counties, excluding the destination county. SIC-1 FE includes one-digit industry classification fixed effects. Standard errors clustered in county level reported in parentheses.

Table 8
Entry regression coefficients across industry groups.

Dependent variable	(1) Per capita entry	(2)	(3)	(4)
Industry	Agriculture	Manufacturing	Business service	Residential service
LnPopDensity#Month 1	−0.069	0.057	0.259	0.225
	(0.125)	(0.123)	(0.085)	(0.104)
LnPopDensity#Month 2	−0.071	0.061	0.028	0.068
	(0.112)	(0.117)	(0.080)	(0.082)
LnPopDensity#Month 3	−0.116	−0.096	0.019	0.085
	(0.109)	(0.126)	(0.066)	(0.081)
LnPopDensity#Month 4	−0.004	−0.150	0.050	0.136
	(0.113)	(0.124)	(0.069)	(0.083)
LnCluster#Month 1	0.004	0.220	0.108	0.191
	(0.100)	(0.086)	(0.064)	(0.072)
LnCluster#Month 2	0.004	−0.025	−0.078	−0.023
	(0.087)	(0.083)	(0.056)	(0.057)
LnCluster#Month 3	0.022	0.145	−0.071	−0.039
	(0.088)	(0.087)	(0.048)	(0.058)
LnCluster#Month 4	−0.065	0.079	−0.087	−0.061
	(0.089)	(0.082)	(0.051)	(0.059)
BPerInfect	0.038	0.023	−0.055	0.670
	(0.223)	(0.209)	(0.111)	(0.245)
PerInfect	−0.920	−0.197	−0.217	−1.987
	(0.412)	(0.372)	(0.341)	(0.396)
Constant	−2.713	−2.475	−0.693	−0.599
	(0.066)	(0.078)	(0.033)	(0.036)
Year-month FE	Yes	Yes	Yes	Yes
Year-county FE	Yes	Yes	Yes	Yes
Month-county FE	Yes	Yes	Yes	Yes
Observations	57,520	57,920	57,880	57,760
Adjusted R-squared	0.435	0.581	0.643	0.622

Note: dependent variable is the natural logarithm number of new firms entry per capita at the county-industry-year-month level. LnCluster is the county level clustering index, calculated from 2008 Census Data. LnPopDensity is weighted average population density of rural-born entrepreneurs' birth county. LnCluster (LnPopDensity) # Month x is the interaction term of LnCluster (LnPopDensity) and year 2020 dummy and month x (x indicating the number of months after the New Year). PerInfect is each city's number of cumulative infected cases of Covid-19 among 10 thousand people till 17th Feb. BPerIfect is the weighted average infected rates of entrepreneurs' birth place. A few provinces (Xinjiang, Qinghai, Tibet, and Inner Mongolia), which have large pastural areas, are excluded. Hubei Province, the epicenter of Covid-19 pandemic, is excluded also. To be consistent with Dai et al. (2020b), only rural counties are included. Robust standard errors reported in parentheses.

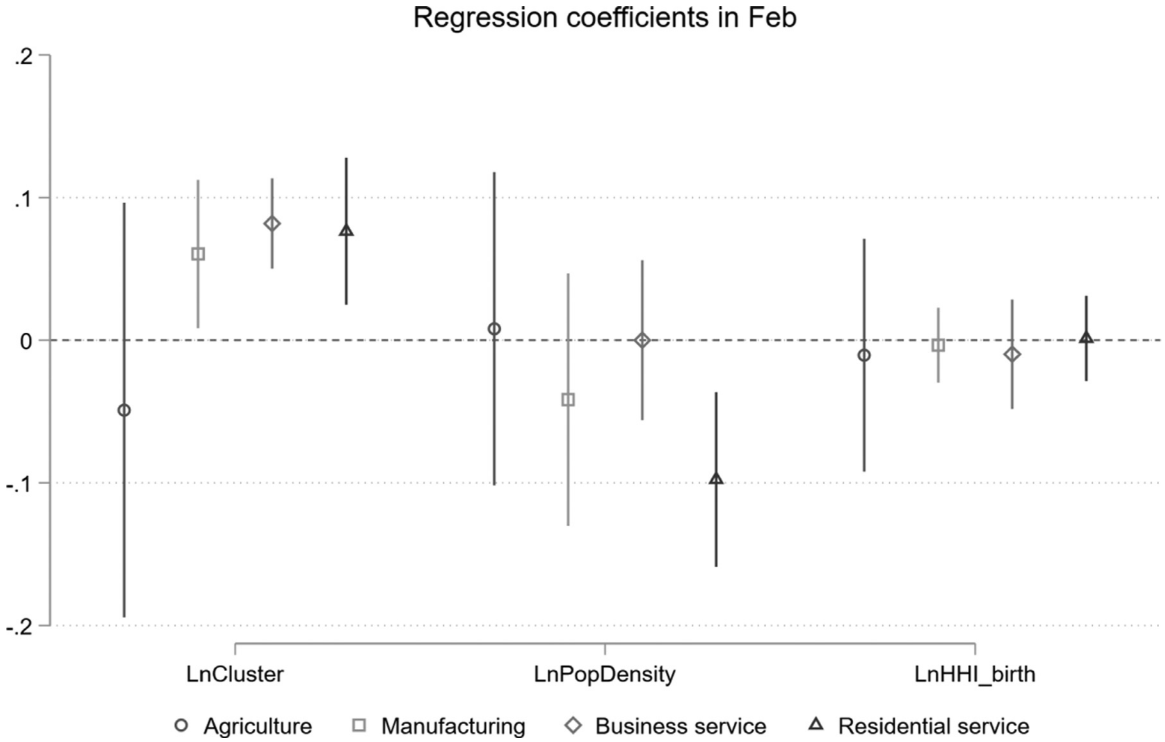

Fig. 14. Firm reopening (Feb 2020) likelihood regression coefficients by industry group. Note: Coefficients and their 95% confidence intervals plotted at the figure. Dependent variable is a dummy variable indicating whether a firm has reopened on 10th February. Coefficients are estimated by industry. Standard errors are clustered in county level.

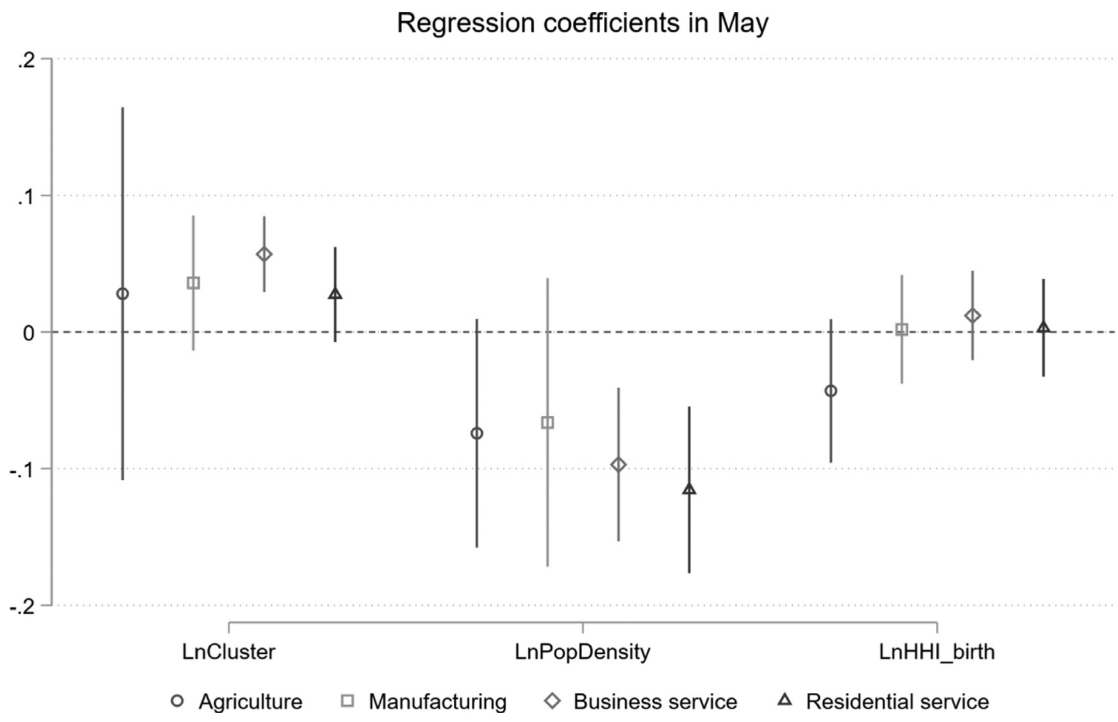

Fig. 15. Firm reopening (May 2020) likelihood regression coefficients by industry group. Note: Coefficients and their 95% confidence intervals plotted at the figure. Dependent variable is a dummy variable indicating whether a firm has reopened in May. Coefficients are estimated by industry. Standard errors are clustered in county level.

while those of clustering appear to be largest in manufacturing, but also significant at 90% level (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. Figs. 14 and 15 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 rural 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 firm registration data, rather than the 2008 Economic Census 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. Nevertheless, we cannot rule out the possibility that the differential resilience of firms in clusters owed to some unobserved attribute of these firms or their locations.

The results of the paper could be useful in two different ways. First, they provide 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 vulnerability 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.

Declaration of Competing Interest

The authors have nothing to disclose.

References

Ali, M., Peerlings, J., Zhang, X., 2014. Clustering as an organizational response to capital market inefficiency: evidence from microenterprises in Ethiopia. Small Bus. Econ. 43 (3), 697–709.
Banerjee, A., Munshi, K., 2004. How efficiently is capital allocated? Evidence from the knitted garment industry in Tirupur. Rev. Econ. Stud. 71 (1), 19–42.
Bartik, A.W., Bertrand, M., Cullen, Z., Glaeser, E.L., Luca, M., Stanton, C., 2020. The impact of Covid-19 on small business outcomes and expectations. Proc. Natl. Acad. Sci. 117 (30), 17656–17666.
Behrens, K., Boualam, B., Martin, J., 2020. Are clusters resilient? Evidence from canadian textile industries. J. Econ. Geogr. 20 (1), 1–36.
Crane, L.D., Decker, R.A., Flaaen, A., Hamins-Puertolas, A., Kurz, C., 2020. Business Exit During the COVID-19 Pandemic: Non-Traditional Measures in Historical Context. Finance and Economics Discussion Paper. Federal Reserve Board, Washington DC.
Crespo, J., Suire, R., Vicente, J., 2014. Lock-in or lock-out? How structural properties of knowledge networks affect regional resilience. J. Econ. Geogr. 14 (1), 199–219.
Dai, R., Feng, H., Hu, J., Jin, Q., Li, H., Ranran, W., Wang, R., Xu, L., Zhang, X., 2020a. The Impact of Covid-19 on Small and Medium-Sized Enterprises: Evidence From Two-Wave Phone Surveys in China. Working paper. Center for Global Development.
Dai, R., Mookherjee, D., Munshi, K., Zhang, X., 2020b. The Community Origins of Private Enterprise in China. Working paper. Economic Development and Institutions Network.
Ellison, G., Glaeser, E.L., 1997. Geographic concentration in us manufacturing industries: a dartboard approach. J. Polit. Econ. 105 (5), 889–927.
Fang, H., Wang, L., Yang, Y., 2020. Human mobility restrictions and the spread of the novel coronavirus (2019-nCOV) in China. J. Publ. Econ. 191.
Fleisher, B., Hu, D., McGuire, W., Zhang, X., 2010. The evolution of an industrial cluster in China. China Econ. Rev. 21 (3), 456–469.
Hsieh, C.-T., Klenow, P.J., 2009. Misallocation and manufacturing TFP in China and India. Q. J. Econ. 124 (4), 1403–1448.
Kali, R., 1999. Endogenous business networks. J. Law Econ. Organ. 15 (3), 615–636.
Kranton, R.E., 1996. Reciprocal exchange: a self-sustaining system. Am. Econ. Rev. 86 (4), 830–851.
Kranton, R.E., Minehart, D.F., 2000. Networks versus vertical integration. Rand. J. Econ. 31 (3), 570–601.
Kranton, R.E., Minehart, D.F., 2001. A theory of buyer-seller networks. Am. Econ. Rev. 91 (3), 485–508.
Long, C., Zhang, X., 2011. Cluster-based industrialization in China: financing and performance. J. Int. Econ. 84 (1), 112–123.
Macchiavello, R., Morjaria, A., 2015. The value of relationships: evidence from a supply shock to kenyan rose exports. Am. Econ. Rev. 105 (9), 2911–2945.
Macchiavello, R., Morjaria, A., 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)
Martin, P., Mayer, T., Mayneris, F., 2013. Are Clusters More Resilient in Crises? Evidence From French Exporters in 2008–2009. CEPR Discussion Paper. Center for Economic Policy Research, London and Washington DC https://cepr.org/.

McMillan, J., Woodruff, C., 1999a. Dispute prevention without courts in Vietnam. J. Law Econ. Organ. 15 (3), 637–658.
McMillan, J., Woodruff, C., 1999b. Interfirm relationships and informal credit in Vietnam. Q. J. Econ. 114 (4), 1285–1320.
Peng, Y., 2004. Kinship networks and entrepreneurs in China's transitional economy. Am. J. Sociol. 109 (5), 1045–1074.
Ruan, J., Zhang, X., 2009. Finance and cluster-based industrial development in China. Econ. Dev. Cult. Change 58 (1), 143–164.
Ruan, J., Zhang, X., 2015. A Proximity-Based Measure of Industrial Clustering. International Food Policy Research Institute (IFPRI) Discussion Paper. International Food Policy Research Institute, Washington DC https://ifpri.org/.
Sonobe, T., Otsuka, K., 2006. Cluster-Based Industrial Development: An East Asian Model. Palgrave MacMillan.
Yoshino, Y., 2011. Industrial Clusters and Micro and Small Enterprises in Africa: From Survival to Growth. World Bank.
Zhang, X., Chen, Q., Fang, P., 2019. Cluster-based aquaculture growth. In: Rashid, S., Zhang, X. (Eds.), The Making of a Blue Revolution in Bangladesh: Enablers, Impacts, and the Path Ahead for Aquaculture. International Food Policy Research Institute, pp. 57–76.
Zhang, X., Moorman, L., Ayele, G., 2011. Infrastructure and cluster development: a case study of handloom weavers in rural Ethiopia. J. Dev. Stud. 47 (12), 1869–1886.