To Follow the Crowd? Benefits and Costs of Migrant Networks*

Yulu Tang†

September 1, 2024
Click here for an updated version

Abstract

Globally, migrant workers often cluster with hometown peers in the same sectors and locations. This paper quantifies two countervailing effects of migrant networks: learning benefits and congestion costs, using data on millions of migrant workers from a food delivery platform in China. First, I provide direct evidence of knowledge spillovers. New workers learn from their peers' delivery experiences to save search time by 10%. Using quasi-random variation in workers' location choices induced by entry bonuses, I show that having one hometown peer nearby increases new workers' productivity by 2%. Second, I quantify the congestion cost of clustering due to correlated shocks. Since migrant workers tend to work longer during adverse hometown shocks, clustering results in hometown workers competing for deliveries when their labor supply surges simultaneously, decreasing real wages by 10%. Third, I build an equilibrium location choice model to quantify the trade-off between the learning benefits and congestion costs. Counterfactuals show that providing insurance for hometown shocks doubles clustering levels and increases productivity.

*I am deeply grateful to Emily Breza, Edward Glaeser, Gabriel Kreindler, and David Atkin for their support throughout this project. I also thank Pol Antras, Jie Bai, Adrien Bilal, Long Chen, Shresth Garg, Oliver Hart, Gordon Hanson, Elhanan Helpman, Lawrence Katz, Asim Khwaja, Michael Kremer, Anders Jensen, Diego Maria, Marc Melitz, Oluchi Mbonu, Ben Olken, Gautam Rao, Jesse Shapiro, Ambra Seck, Patricia Sun, Elie Tamer, Jaya Wen, Daniel Xu, David Yang, Jie Zhou for many helpful suggestions. I also thank participants of the Harvard development, trade, and labor workshops, and conference and seminar participants at World Bank DECRG, Berkeley Hass, Upenn, LSE, Michigan, Northwestern, CUHK, HKU, NUS, NBER Summer Institute, MIDWEST for many helpful suggestions. All errors are my own.

†Dartmouth College. Email: yulu.tang@dartmouth.edu

1 Introduction

Throughout the world, migrant workers often cluster in the same sectors and locations with peers from the same region. For example, 14% of taxi drivers in New York City came from Pakistan (Schaller, 2006). Indian migrants operated over 60% of motels in Los Angeles (Light et al., 1994), and 24% of hairdressers in Houston were Vietnamese migrants (Patel and Vella, 2013). In China, around 30% of migrant workers from the same county clustered in the same sector and province (Chen et al., 2010).

What are the benefits and costs of such clustering? On the one hand, clustering with same-origin peers may provide new workers with better information, job opportunities, and social support (see Munshi (2020) for a comprehensive review).[1] On the other hand, a large cluster can heighten competition within the network (e.g., Beaman, 2012). While both agglomeration and dispersion forces of clustering are important for policy implications, few studies have jointly examined them in the same context and analyzed how they interact to impact worker welfare and location choices.

In this paper, I provide a unified theoretical and empirical analysis of two countervailing effects of clustering through several quasi-experiments and a model. First, I show how clustering facilitates knowledge spillovers, leading to higher productivity and average wages. Second, I explore a new channel of cost: clustering results in hometown workers competing with one another during adverse hometown shocks when their labor supply surges simultaneously, leading to higher income risks. Third, I analyze how the two forces interact to impact workers' location choices through an equilibrium model.

I study the effects of migrant clustering by using data on millions of migrant workers from a large food delivery platform in China.[2] Delivery workers earn a piece rate by delivering meals from restaurants to consumers, and 98% of delivery workers in large cities are domestic seasonal migrants. There are several advantages of studying migrant networks in this context. First, the gig economy has grown rapidly and emerged as the NO.1 occupational choice among migrant workers (Pew, 2021; McKinsey, 2022), yet we know little about their choices and trade-offs under this new economic form. Furthermore, with exceptionally granular GPS data tracking workers' movements, I am able to construct very precise measurements of worker productivity and explore the mechanism of knowledge spillovers, which are substantial challenges in prior research.

[1]For example, migrant networks can increase employment rates (e.g., Edin et al., 2003; Munshi, 2003), wages (e.g., Damm, 2009; Egger et al., 2021), and provide social support (e.g., Biavaschi et al., 2021; Blumenstock et al., 2021).

[2]The platform operates similarly to most food delivery companies, such as *Uber Eats* and *DoorDash*.

Similar to other contexts, I find a high level of clustering among migrant workers. First, the delivery platform divides each large city into approximately 150 districts, each covering a $5km \times 5km$ area. Delivery workers typically operate within a single district in the destination city.[3] I thus compute the clustering level based on the share of hometown workers who choose to work in the same district.[4] If workers were randomly allocated across districts, fewer than 1% of same-origin workers would work in the same district. Instead, the observed clustering level is 30%. As a result, the largest cluster formed by same-origin workers within a district makes up 15% of the district's total labor force on the platform.

The main analysis consists of three steps. In Step 1, I analyze how clustering enables knowledge spillovers among hometown peers and increases productivity. In Step 2, I provide evidence for the labor market congestion costs of clustering using hometown floods and pandemic lockdowns as exogenous variations. In Step 3, I build an equilibrium location choice model to quantify the trade-off between the learning benefits and congestion costs, and how they interaction to affect workers' location choices and welfare.

In Step 1, I conduct two main analyses: first, I open the black box of peer learning and show what knowledge workers share with each other; second, I use quasi-experiments based on entry bonuses to quantify the overall impacts of clustering on productivity.

In the first analysis, I examine whether workers can deliver faster (1) when they themselves have visited a location before, which highlights learning by doing and (2) when their peers have visited a location before, which indicates learning from peers. The identification builds on the platform's order allocation algorithm, which exogenously assigns workers to visit different locations.I also exploit the fact that 60% of new workers have a referrer upon entry.[5] Referrers are incumbent delivery workers on the platform. This referral information reveals social connections between workers and enables me to directly estimate knowledge transmission within each new worker-referrer pair. Furthermore, using granular GPS data, I decompose each delivery into three parts: (a) time spent on search for restaurants, (b) driving speed on the road, and (c) time spent on searching for consumer locations. This decomposition provides rich insight into where learning

[3]The GPS data shows that workers complete 85% of their daily deliveries within a $3km \times 3km$ area on average. In addition, most workers select a single district as their primary work location, where they complete 70% of their daily deliveries.

[4]Hometown is defined as a worker's county of birth. There are 2,843 counties in China, of which 2,805 are represented by at least one active delivery worker in the analysis sample.

[5]For new workers in Shanghai in 2021, around 41% of new workers had referrers, and around 54% of new workers came from the same county as their referrers. Restricting the sample to delivery workers who finished at least 50 deliveries, around 60% of the new workers had referrers, and around 65% of the pairs came from the same county.

occurs across different steps of a delivery.

The results show that both learning by doing and peer learning play important roles. A worker's prior visit to a location reduces search time by 20% when they travel to the same restaurant or consumer location the second time.[6] Peers' knowledge is around half-effective. Conversely, travel speed on the main roads does not display the same patterns. These results align with field insights that GPS tools can help workers travel on the main roads, but knowledge accumulated from past delivery experiences can help workers navigate local neighborhoods better. Additional analysis also show that order and more crowded neighborhoods require more knowledge.

In the second analysis, I quantity the overall impact of clustering on worker productivity. Estimating this has been challenging in prior studies: a simple regression of productivity on clustering can be biased since workers' location choices are endogenous.[7][8] To deal with this, I construct an instrumental variable, which exploits the quasi-random variation in new workers' location choices induced by entry bonuses. These entry bonuses are fixed monetary transfers awarded to new workers who complete a certain number of deliveries in targeted districts.[9] They thus function as natural experiments that exogenously nudge new workers to enter different districts. In summary, the endogenous variable is the share of hometown workers in a worker's work district. The instrument variable is the hometown share in the bonus-target district.[10] In addition, I only use bonuses active the following week after new workers' entries. This timeline helps avoid bonuses abstracting different types of new workers joining the platform.

The IV regressions of new workers' labor market performances on clustering levels indicate that clustering significantly increases new workers' productivity. On average, one additional hometown peer increases productivity by 0.4% in the first three months. Working in a district with an average clustering level increases new workers' productivity

[6]A further heterogeneity analysis shows that the estimated search time decreases are most pronounced in old, dense, and crowded neighborhoods, where it is hard to find exact locations. These findings highlight the importance of learning in the city despite the availability of advanced GPS tools.

[7]For example, more competent workers can learn faster based on their own experiences and thus cluster less.

[8]Another endogeneity issue I deal with is that new workers are more productive in districts with more hometown peers, but for reasons unrelated to migrant networks. For example, Cantonese workers work in districts with more restaurants speaking Cantonese, and their productivity is higher due to better communication with restaurants instead of more Cantonese workers around.

[9]The platform occasionally imposes entry bonuses in different districts and weeks to balance labor demand and supply. In Shanghai in 2021, 72% of entry bonuses were active for only one week, and 87% of districts offered at least one week of entry bonuses each year.

[10]Specifically, I find that 90% of new workers choose to work in districts within 6km of their referrers' districts. I thus define the set of districts adjacent to the referrer as a new worker's choice set. The IV is the hometown-worker share in a district that offers entry bonuses within this choice set.

by 5%. While I do not find that clustering impacts working hours, workers' total earnings rise with higher clustering due to higher productivity.

More interestingly, I also estimate the entire shape of the clustering-productivity relationship by running the IV regression with indicators for different clustering levels. The results indicate an increasing but concave curve, highlighting the diminishing marginal productivity gains from clustering.

In Step 2, I examine the congestion cost of clustering. The intuition is that clustering leads to many migrant workers increasing labor supply simultaneously during during adverse hometown shocks. When labor demand is inelastic, clustering increases competition and decreases real wages, precisely when workers have the greatest desire to earn more. To document the congestion cost empirically, I construct two sets of shocks: (1) severe floods that occurred between June and August 2020, the biggest since 1998, and (2) the pandemic lockdowns across counties in 2021[11].

Both shocks exhibit similar results. First, delivery workers from affected counties increase their weekly labor supply by six hours on average during these hometown crises. This increased labor supply allows them to earn an extra 200 RMB weekly, which can be remitted to support affected family members back home.

Second, for each district, I construct aggregate shock indicators based on the predicted share of workers facing hometown shocks in each district and week from June to August 2020.[12] By regressing district-level outcomes on aggregate shock indicators, I find that total labor supply surges when many workers in one district experience hometown shocks simulanatesouly, For example, in districts with over 15% of workers experiencing hometown shocks, total working hours increase by 10% that week. However, I find no significant change in consumer demand, as measured by the total order volume.[13]

Finally, I regress workers performances on both the individual hometown-shock indicator, aggregate hometown-shock indicator, and their interactions. The coefficient in front the interaction terms capture the congestion costs. I find that for clustered workers, even though they work longer during their hometown shocks, they finish significant

[11] I collect both the numbers of daily COVID cases from government websites and identify lockdowns in each county and week by consumer orders below half of normal medians.

[12] I compute this predicted share using the hometown-district composition in May 2020 as the baseline. I use the predicted shock share instead of the actual share of shock-affected workers in each district and week to avoid the endogenous changes in hometown-district composition due to hometown shocks. Therefore, the variation in the district-level aggregate shock share comes from the geographic and time differences in hometown shocks.

[13] Though more active delivery workers potentially lead to faster delivery, I do not find consumers respond, at least in the short run.

fewer number of deliveries relative to non-cluster worker. On average, clustered workers experience a 10% decrease in their real wages compared to non-clustered workers. This congestion result is driven by clustered workers competing for deliveries and fewer deliveries being assigned to each worker per unit of time.[14]

In Step 3, I build an equilibrium model to quantify the trade-off of these two impacts of clustering: first-moment learning benefits and second-moment congestion costs.[15] The model proceeds in four stages. In stage 1, conditional on migrating to the destination city and becoming a delivery worker, new migrant workers form expectations about working in different districts and choose to work in a district with the highest expected utility. In stage 2, the hometown shocks will be realized. In stage 3, workers decide on working hours and remittances to maximize utility, where they care about their own consumption,left-behind family's consumption, and leisure time. Finally, the equilibrium wage will be determined through the market clearing condition in each district. The model also explicitly assumes that workers are aware of the learning and congestion effects of clustering and consider both when choosing work districts.

I estimate the model using empirical results from Steps 1 and 2, as well as additional variation from model predictions and quasi-experiments for key parameters. Specifically, the productivity increase as a function of clustering comes directly from the IV regression of delivery speed on clustering in Step 1. Hometown income levels during the regular and shock periods are identified from workers' labor supply responses to hometown shocks in Step 2. I then leverage a quasi-experiment based on semi-annual platform festivals to estimate workers' labor supply elasticity to income. Specifically, the platform randomly selects workers to receive 50-1000 RMB cash prizes during the festival. I thus regress workers' working hours on prize amounts to estimate the elasticity.

Two other key parameters are workers' risk aversion coefficient and the scale of workers' idiosyncratic preferences across districts. First, I recover risk aversion by comparing the clustering levels of workers from high (risky) versus low (safe) shock-probability hometowns. The model predicts that workers from risky hometowns will cluster less due to greater exposure to the congestion cost of clustering. Furthermore, the more risk-averse

[14] I also estimate the effect of clustering on workers' relocation rate during adverse hometown shocks. I find weakly significant but small effects. This suggests that some workers may relocate to another district after the aggregate shock, but not all workers relocate. One explanation is the location-specific knowledge as identified in Step 1. I show that workers' delivery speed decreases and the timeout rate increases after relocation.

[15] A model is essential here since trading off the two effects requires a credible estimation of workers' risk aversion coefficient and other parameters governing the utility function. Moreover, the model enables analyzing the whole shape of worker utility with clustering, estimating externalities, and running counterfactuals for policy implications.

workers are, the larger the gap is. I thus exploit the gaps in clustering levels between risky and safe hometowns to estimate the risk aversion coefficient. Second, I run a gravity regression of new workers' entry choices on the entry bonuses to estimate the scale of workers' idiosyncratic preferences across work districts.

Using the estimated parameters, I show that workers' utility displays an inverted-U shape: utility first increases with clustering levels due to the learning benefit and then fall when the congestion cost dominates. Moreover, the model predicts two sources of externalities: (1) a positive learning externality as workers do not internalize their role as teachers for same-origin workers, and (2) a negative congestion externality as the congestion leads to longer idle waiting time during aggregate hometown shocks. I show that the cumulative externality is negative since the congestion costs impact all workers in the district, while the learning benefits decay quickly.

Finally, I run counterfactuals examining various policies by accounting for the two clustering impacts. In the first counterfactual, I simulate the equilibrium where workers are charged for the social externality of their entry decision. I find that when accounting for externality, the equilibrium predicts significantly lower clustering levels and encourages the formation of many small clusters. In the second counterfactual, I explore the provision of insurance to mitigate adverse hometown shocks. By eliminating congestion costs of clustering through insurance, the equilibrium clustering level nearly doubles, and workers' average productivity increases by 30%. Finally, I show that if the platform can provide better learning tools to all workers, the clustering level decreases significantly, and worker welfare increases by 4% due to less exposure to the congestion costs.

This paper joins a large body of literature analyzing the value of migrant networks. Existing studies provide extensive evidence that migrant networks can impact migrant workers' performances in the destination's labor market, such as by increasing employment rates and income (e.g., Borjas, 1995; Munshi, 2003; Antoninis, 2006; Woodruff and Zenteno, 2007; Bayer et al., 2008; McKenzie and Rapoport, 2010; Beaman, 2012; Patel and Vella, 2013; Munshi and Rosenzweig, 2016; Giulietti et al., 2018; Egger et al., 2021).

This project contributes to this body of work in several ways. First, the granular GPS data enables me to provide direct evidence of knowledge spillovers among hometown workers. These results establish a vital micro-foundation for understanding the positive network effects on migrant workers' labor-market performance. Second, I highlight a new cost of clustering: clustering results in labor market congestion during hometown shocks and leads to higher income risks. Third, I propose a new framework to estimate workers' utility as a function of clustering and measure its externality to other non-clustered workers.

The findings on migrant workers' responses to hometown shocks also relate to extensive studies analyzing the role of migration as insurance for rural shocks.[16] These studies analyze the impacts of adverse shocks on migrants' labor supply and remittances (e.g., Lucas, 2004; Yang, 2006; Yang, 2011; McKenzie et al., 2014; Gröger and Zylberberg, 2016; Joseph et al., 2018), general equilibrium effects of agricultural shocks (e.g. Jayachandran, 2006; Akram et al., 2017), and how migrants function as insurance (e.g., Munshi and Rosenzweig, 2016; Morten, 2019). A closely related study is Michuda (2021), which uses Uber driver's data from Uganda and finds that adverse agricultural shocks are associated with longer working hours.

This project adds consistent results that migrant workers adapt their labor supply in response to shocks in their hometowns.[17] I also expand on existing studies by quantifying this new general equilibrium channel through which clustered migrant workers compete with one another during adverse hometown shocks. This GE effect reduces the effectiveness of their role as rural insurance.

Third, examining how agglomeration enhances learning and its associated costs aligns with a comprehensive urban literature that analyzes the determinants and impacts of urban agglomeration.[18] Learning emerges as a key mechanism to explain urban agglomeration (e.g. Marshall, 1890; Jaffe et al., 1993; Glaeser, 1999; Graham and Marvin, 2002; Peri, 2002; Roca and Puga, 2017; Davis and Dingel, 2019; Atkin et al., 2022). This study contributes by providing crucial empirical support for learning from peers in urban settings. Furthermore, I show that workers share location-specific knowledge with each other, which confines knowledge spillovers within specific areas. [19]

Finally, my analysis of new workers learning from past delivery experiences adds

[16] Rural-urban migration has significantly shaped China's economy over the past three decades, with 300 million rural migrant workers employed in urban areas by 2022. This study contributes to the examination of this phenomenon by focusing on the unique migration patterns in the emerging era of the platform economy alongside existing research (e.g., Kinnan et al., 2018; Chen et al., 2010; Dai et al., 2019).

[17] Given the rapid growth of gig economy, an expanding body of literature is investigating the flexibility it offers (e.g., Mas and Pallais, 2017; Chen et al., 2019; Angrist et al., 2021; Michuda, 2021). This study highlights that the flexibility inherent in gig economy jobs enables workers to adjust their labor supply during hometown shocks, thereby contributing to mitigating geographical aggregate shocks.

[18] Urban economics literature has documented various factors leading to the agglomeration of workers, firms, and industries in urban contexts (e.g., Rosenthal and Strange, 2001; Duranton and Puga, 2004; Giuliano et al., 2007; Glaeser, 2010; Glaeser et al., 2010; Kukalis, 2010; Krugman, 2011; Moretti, 2012; Giulietti et al., 2018; Duranton and Puga, 2020; Fajgelbaum and Gaubert, 2020).

[19] Moreover, the labor market congestion cost of clustering also relates to studies highlighting labor pooling in urban contexts (Rauch, 1993; Overman and Puga, 2010; Kerr and Robert-Nicoud, 2020). While labor pooling can mitigate sector-specific volatility by creating a larger labor market, this project shows that when workers face friction of moving across locations, they may still face correlated shocks.

to studies estimating the impacts of on-the-job learning on worker productivity (e.g., Thompson, 2010; Levitt et al., 2013; Haggag et al., 2017; Mao et al., 2019; Papay et al., 2020; Cook et al., 2021). Moreover, the finding that new workers learn from referrers' experiences highlights additional values of referrers in the labor market: they can transmit productive knowledge and increase new worker productivity (e.g., Antoninis, 2006; Beaman and Magruder, 2012; Burks et al., 2015; Dustmann et al., 2016; Friebel et al., 2023).

The paper proceeds as follows. Section II describes the context. Section III builds a district choice model indicating how learning benefits and congestion costs of clustering interact to impact location choices. Section IV shows clustering enables knowledge spillovers and increases new worker productivity. Section V presents the general equilibrium effects of clustering during adverse hometown shocks. In Sections VI and VII, I estimate the model, measure worker utility with clustering, and discuss counterfactuals. Section VII presents conclusions.

2 Context

2.1 Migrant Workers in the Gig Economy

This project focuses on rural-urban migrant workers in the gig economy in urban China. This is a particularly good setting to study the impacts of migrant clustering. First, the gig economy has grown rapidly and is estimated to have 435 million workers worldwide in 2023 (Datta et al., 2023). Furthermore, the gig economy has become the NO.1 occupational choice among migrant workers due to lower entry barriers (Pew, 2021; McKinsey, 2022). Since this is a new economic form, where we may expect technology to have replaced lots of human interactions, we know much less about how workers make choices and the value of migrant networks in the gig economy.

Specifically, I use data from a large food delivery platform in China with over 1 million delivery workers and operating in more than 100 cities nationwide. The platform operates similarly to other platforms, such as *UberEats* and *DoorDash* in the United States.[20] Workers earn a per-delivery commission fee ranging from \$0.5 to \$1.5. Their monthly earnings typically fall between \$300 and \$1,000, depending on the number of

[20] Once consumers place orders on the platform, an algorithm matches each order with nearby delivery workers. To complete a delivery, workers first travel to the restaurant to pick up the food. They then drive to the consumer's address to drop it off, as illustrated in Figure B.1. A major difference is that delivery workers drive motorcycles to deliver food instead of driving cars, which is also very common in many other developing countries.

deliveries completed.

Most delivery workers in large cities are domestic seasonal migrant workers. For example, in major cities like Beijing and Shanghai, around 98% of delivery workers are domestic migrant workers. As shown in Figure B.2 and Table B.2, most workers come from central and western parts of China (with lower GDP) and work in several large cities (with higher GDP). Table B.1 provides details on workers' demographic characteristics.[21] For the primary analysis, I focus on delivery workers in the five largest cities in China. More details on data construction and sample selection are provided in Appendix C.

2.2 Clustering of Migrant Workers

Similar to other contexts, I find a high level of clustering when migrant workers choose where to work. First, the food delivery platform divides each large city into around 100$\sim$200 comparable districts.[22] Each district spans roughly $5km \times 5km$ and hosts about 100$\sim$200 workers. GPS data show that most workers operate within a single district and complete around 70% of their daily deliveries there.[23]

I thus compute the clustering level based on the share of workers from the same hometown who work in the same district.[24] If workers were allocated to these districts randomly, fewer than 1% of people from the same hometown would work in the same district. However, the observed clustering level is 30%. Figure B.3 provides examples of clusters formed by workers from different hometowns. I plot the histogram of the clustering levels in Figure 1. In turn, the largest cluster of same-origin workers in each district comprises 15% of the labor force in each district on the platform (figure 2).[25]

[21] The entry barriers to becoming a delivery worker are low. Workers first download the platform's app and register by verifying their identity cards. They rent or purchase a motorcycle for around \$400, and a helmet and work clothes for around \$100.

[22] data show that 85% of deliveries are within 3 km.

[23] The platform typically defines districts by drawing circles around shopping malls with many restaurants to minimize the number of cross-district deliveries.

[24] Hometown is defined at the county level. There are 2,844 counties in China. I use GPS data to define each worker's work district in a given week. Specifically, I identify the district where each worker has the most GPS coordinates that week as their primary work district.

[25] I have also calculated clustering levels based on the share of deliveries completed by workers from each hometown across districts, as shown in Figure B.4. These delivery-based clustering levels are very similar to the worker share-based levels described here.

Separate Work and Living Locations

The GPS data also allows me to distinguish between individuals' work and living locations. Taking workers in Shanghai as an example (figure B.5), I find most migrant workers live on the outskirts or cluster in lower-rent compounds.[26] Typically, same-origin migrants cluster residentially to share rents. However, living together shall not prohibit these migrant workers from working in different districts.[27] By separating workers' choices of work locations from living locations, I avoid making strong assumptions to shut down other reasons for clustering. Instead, migrant workers can enjoy various benefits of clustering by living together, such as sharing rent and food. However, the two forces studied in this project, the learning benefits and congestion costs, directly affect workers' labor income and are salient in influencing their work location choices.

3 The Impact of Clustering on Learning and Productivity

In this section, I analyze how clustering enables knowledge spillovers through the migrant networks, leading to higher productivity and average earnings. I conduct two main analyses: first, I open the black box to show what knowledge workers share with each other; second, I use quasi-experiments based on entry bonuses to quantify the overall productivity return of clustering.

3.1 Descriptive Findings

In the food delivery sector, faster delivery speed means workers can finish more deliveries and earn more per hour, leading to higher real wages, making it a good measure of worker productivity.[28] I first show that new workers' productivity increases over time. In Figure 3(a), the x-axis displays the number of weeks since new workers joined the platform, and the y-axis is the new workers' average speed. The concave shape indicates that workers' productivity increases fast first and stabilizes after the first three months.

Furthermore, new workers who cluster with hometown peers appear to improve faster.

[26] I infer residential locations from GPS coordinates before 7 am or after 11 pm daily. Work locations are based on average latitudes/longitudes of daily deliveries.

[27] These migrant workers rarely own property in their work cities. Based on a platform-wide survey in 2021, around 0% of migrant workers owned an apartment in their destination city, and less than 3% planned a future purchase. Thus, their residential locations are not tied to owned properties.

[28] Delivery speed is computed as the delivery distance (meter) to be divided by the delivery duration (minute).

In Figure 3(b), I classify new workers into two groups: those working in a district with same-origin workers and those without. The average delivery speed of new workers with hometown peers (red line) increases faster than those without (blue line).[29] This pattern suggests that new workers may learn faster by working alongside hometown peers in the same district. However, directly comparing clustered and non-clustered workers can be biased due to the potential inherent differences between the two groups. Next, I will use quasi-experiments to quantify the magnitude of knowledge spillovers and explore the mechanisms.

3.2 Direct Evidence of Learning

3.2.1 What Do Workers Learn from Peers?

In the first analysis, I aim to open the box and explore what knowledge delivery workers learn from their own and hometown peers' experiences. Specifically, I analyze whether workers can travel and deliver faster when (1) workers themselves have visited a location before, which represents learning by doing, and (2) workers' close peers have visited a location before, which indicates learning from peers. To define each worker's close peer, I exploit the fact that 60% of new workers have a referrer upon entry.[30] Referrers are also existing delivery workers on the platform. This new worker-referrer information highlights the social connections between the two individuals. Thus, I can identify new workers' peers as their referrers from whom new workers learn and estimate knowledge transmission within these pairs.

Furthermore, using granular GPS data, I decompose each delivery process into three parts: (1) time spent searching for restaurants, (2) time spent driving on the road, and (3) time spent searching for consumers.[31] This decomposition enables me to analyze further where learning occurs across different delivery steps.

I thus regress workers' productivity on indicators of whether workers themselves or

[29] I also plot the trend of new workers' hourly wage, as shown in Figure B.7. It exhibits a similar trend that new workers can earn higher wages per hour, mainly driven by finishing a higher number of deliveries per hour. Furthermore, I plot the productivity trends for migrant workers versus local residents, as shown in Figure B.8. Local residents present a flatter trend, whose initial productivity is significantly higher than migrant workers.

[30] Among new workers in Shanghai in 2021, 41% had referrers, and 54% of new worker-referrer pairs were from the same county of birth. Restricting the sample to workers who finished more than 50 deliveries, 60% had referrers, and 65% of those pairs came from the same counties.

[31] I calculate search times by counting GPS coordinates within 150-meter radii of restaurants and consumers. Road travel time is the rest of the delivery time divided by distances.

their referrers have ever visited the same location before. The identification builds on the platform's order allocation algorithm, which exogenously assigns workers to visit different locations and results in workers possessing varying prior knowledge across locations and time. The corresponding regression function is as follows.

$$\begin{aligned} Y_{itab} = \alpha_1 \text{ OwnVisit}_{ita} + \alpha_2 \text{ RefVisit}_{ita} + \beta_1 \text{ OwnVisit}_{itb} + \beta_2 \text{ RefVisit}_{itb} + \\ \delta_a + \delta_b + \gamma_{it} + \text{AveVisit}^{\text{Own}}_{ita} + \text{AveVisit}^{\text{Own}}_{itb} + \text{AveVisit}^{\text{Ref}}_{ita} + \text{AveVisit}^{\text{Ref}}_{itb} + \epsilon_{itab} \quad (1) \end{aligned}$$

The regressions are at the delivery level. For each delivery from restaurant a to consumer location b, I construct three dependent variables on worker productivity as outlined above. I also have the four independent variables: (1) whether a worker i has ever visited the same restaurant in the past month before time t, $OwnVisit_{ita}$, (2) whether worker i has ever visited the same consumer building, $OwnVisit_{itb}$, (3) whether i's referrer has ever visited either the restaurant, $RefVisit_{ita}$, and (4) the consumer building, and $RefVisit_{itb}$. I also include the location fixed effect, δ_a and δ_b, to control for inherent location characteristics. The worker $\times$ date $\times$ hour fixed effect, γ_{it}, takes care of worker and time heterogeneity, and enables me to compare the productivity across deliveries that a worker completes in the same hour but with differing prior knowledge.

Another concern that we may have is that the order allocation is not random. Discussions with the algorithm team revealed that the most important input for the algorithm is workers' real-time GPS so that the platform can match orders with workers nearby. Thus, the endogeneity concern is that workers may locate somewhere and wait for certain types of orders strategically. To deal with this, I follow the intuition from Borusyak and Hull (2023) and construct an average delivery probability for each worker and referrer. Specifically, I randomly shuffle order allocation within each hour, each $2km \times 2km$ grid, and each tenure group. I repeat this exercise 100 times and calculate the average likelihood of a worker and a referrer visiting a location based on these counterfactual order assignments.[32] By controlling for the delivery probability, the actual identification assumption is that workers who are at the same location at the same time are very close substitutes in the algorithm's eyes. When the algorithm assigns orders among them, who gets which order is almost random.

I also conduct balance checks, where I compare the characteristics of deliveries: (1)

[32] As a robustness check, I try another method to calculate the probability of visiting a location given workers' characteristics. Specifically, I run a logit regression of visiting a location on a worker's tenure, age, gender, historical performances, whether they have a referrer, and the referrers' characteristics. The results are similar.

allocated to workers who visit a location for the first time versus those on repeat visits, (2) allocated to workers with varying tenures, and (3) accepted versus rejected by workers. I find no significant differences for these comparisons in Figure B.9 and Figure B.10, suggesting that inherent worker attributes do not seem to influence order allocation significantly.

Table 1 presents the regression findings, showing that both learning by doing and learning from peers play essential roles. A worker's own prior visit to a location reduces search time by approximately 20% when they visit the same restaurant or consumer building again. A referrer's prior visit decreases a new worker's search time by around 10%, indicating knowledge transmission within this pair of workers. Conversely, travel speed on the road does not display the same patterns. These results are consistent with field insights: GPS tools help workers navigate main roads, but learning is crucial when workers navigate in narrow neighborhoods and search for specific addresses.

3.2.2 In which part of the city do workers learn the most?

The coefficients from this delivery-level regression also reveal valuable information: the amount of learning needed in each location. I thus divide Shanghai into approximately 200 grids that are a $2km \times 2km$ each. I run the delivery-level regressions for each grid separately and compute a grid learning index based on the coefficients in front of the OwnVisit_{ita} and RefVisit_{ita} variables.

In Figure 4 and Table 2, I examine the correlation between these learning indexes and grid-level attributes, including average building age and housing density. I find that the estimated learning magnitudes are most pronounced in old, dense, and crowded neighborhoods. This highlights the importance of learning in areas where exact locations are difficult to find.

3.3 IV Estimates of Clustering on Productivity

The preceding analysis provides direct evidence of what knowledge workers learn from peers. Quantifying the overall impact of clustering on new workers' productivity is also important. However, estimating it has been challenging in the literature as a simple regression of productivity on clustering is subject to a major endogeneity concern: workers' location choices are endogenous. For example, more competent workers may learn faster based on their own experiences and thus cluster less. In this section, I construct an instrumental variable based on entry bonuses to address it.

3.3.1 IV: bonus-predicted clustering level

Consider a new worker i, who comes from hometown h, enters the platform in week T^e. The endogenous variable is the share of hometown workers in a worker's actual work district, Share_{hdt}.

$$\text{Share}_{hdt} = \frac{N_{hdt}}{N_{dt}}$$

where N_{hdt} is the number of workers from hometown h in district d in week t, and N_{dt} is the total number of workers in district d in week t.

To construct the instrumental variable for Share_{hdt}, I explore the quasi-random variation in new workers' district choices induced by entry bonuses. These entry bonuses are fixed monetary incentives awarded to new workers who complete a certain number of deliveries in bonus-targeted districts.[33] They thus function as natural experiments that exogenously nudge new workers to enter different districts.[34] In addition, I only use bonuses active in the subsequent week following a new worker's entry. This timeline (as shown in figure B.13) avoids the situation that bonuses attract different types of workers to join the platform.

To formalize the IV, as mentioned in the previous section, 60% of new workers have a referrer j who works in district d_j. I find that 90% of new workers choose to work in districts within a 6-km radius of their referrers' districts (Table B.3). Therefore, I define districts adjacent to a new worker's referrer's district as their potential choice set, d^P. New workers are exogenously nudged to join one of the districts within d^P based on which districts offer entry bonuses in week $T^e + 1$.

In summary, the instrument variable is the share of hometown workers in a district that provides entry bonuses within each new workers' choice set, denoted as $\text{PredictShare}_{ihd^Pt}$. The main variation of the IV comes from the quasi-random assignment of entry bonuses across time and districts.

$$\text{PredictShare}_{ihd^Pt} = \max_{d \in d^P}\{\frac{N_{hdT^e}}{N_{dT^e}} \times \mathbb{1}\{bonus_{d,T^e+1}\}\}$$

[33]The platform introduces entry bonuses targeted at various districts in different weeks. These bonuses serve to balance the labor supply and demand across districts. In Shanghai in 2021, 72% of entry bonuses are active for only one week. 87% of districts experience at least one week of entry bonuses each year.

[34]The specific terms can vary across different bonuses. In most cases, new workers are required to operate within the targeted districts for approximately a month or complete around 200 deliveries within the district.

3.3.2 Results: Average Effect of Clustering on Productivity

I run the IV regression at the worker-week level as below.

$$Y_{ihdt} = \alpha \text{ Share}_{hdt} + \gamma_{dt} + \gamma_{ht} + \gamma_{T^e} + \text{AveShare}_{ihd^pt} + \epsilon_{ihdt} \tag{2}$$

As explained above, the independent variable is the share of hometown workers in each new worker's actual work district, Share_{hdt}, and the instrumental variable is the bonus-predicted share, $\text{PredictShare}_{ihd^pt}$. I also include (1) the district-week fixed effect to control for any district-level labor market fluctuations, including the heterogeneity due to the allocation of entry bonuses, (2) the hometown-week fixed effect to adjust for any hometown shocks that might influence workers' labor supply, and (3) entry cohort fixed effects to only compare the productivity of workers who enter around the same time.

Furthermore, while the allocations of entry bonuses are exogeneous, Share_{hdt} and $\text{PredictShare}_{ihd^pt}$ also depend on the sizes of existing hometown networks, N_{hdt}. This means that new workers with more hometown peers nearby (higher N_{hdt}) are likely to have a larger clustering level (higher $\text{PredictShare}_{ihd^pt}$) regardless of which district offers entry bonuses. To address the endogeneity of existing hometown networks, I follow the intuition from Borusyak and Hull (2023) and construct an average predicted share based on bonus counterfactuals. Specifically, I shuffle the allocation of entry bonuses within three months before and after each worker's entry week. Each bonus allocation counterfactual generates a bonus-predicted clustering level. I then calculate an average clustering level, AveShare_{ihd^pt}, across all bonus allocation counterfactuals. By controlling for this average clustering level, I compare new workers with similar hometown networks in the destination city. Still, their realized clustering levels can differ since which district offers entry bonuses in which week is exogenous. I am thus able to extract the exogenous variation from entry bonus allocation across districts and time for identification.

The IV results are presented in Table 3. Column 5 reports the first-stage results, where the predicted clustering level strongly correlates with the actual clustering level. Columns 6-9 present the IV results. They indicate that new workers operating in districts with a higher share of hometown peers tend to deliver food more rapidly, complete more deliveries per hour, and have higher weekly earnings. Conversely, the clustering level does not significantly affect new workers' hours. In terms of magnitude, compared to the average delivery speed, the presence of one additional same-origin peer in the same district corresponds to a 0.4% increase in new workers' productivity. Working in a district with an average clustering level results in a 5% productivity increase for new workers.

I also present a set of robustness checks. Table B.5 reports additional measurements of workers' labor market performances. Workers are less likely to be late for deliveries and slightly less likely to relocate to other districts when they work in districts with more hometown peers. Table B.6 reports the effects of clustering on delivery speed under different IV specifications. Column (1) uses entry bonuses with a take-up time window of less than one week; column (2) uses all entry bonuses in the destination city without restricting the entry bonuses near the referrer's district; column (3) focuses on workers with only one bonus-predicted district; column (4) constructs the instrument based on time-varying clustering levels across districts, and column (5) construct the predicted clustering level to be weighed by the size of the entry bonuses. All columns report similar results: a higher clustering leads to higher productivity.

Table B.7 reports results based on different independent variable specifications. Column (1) uses the number of hometown workers instead of the hometown-worker share in the baseline specification; column (2) reports the effect on productivity of locating in the same district as one's referrer; column (3) constructs the clustering level adjusted by tenure; column (4) explore the share of deliveries completed by hometown peers as the proxy for clustering levels; and column (5) expands hometown peers to workers from the same birth city. These different measurement of clustering levels report patterns similar to those of the baseline specification.

3.3.3 Diminishing Marginal Gain from Clustering

More interestingly, I am able to recover the full shape of productivity improvements across clustering levels by running the IV regression with a set of indicators, $\mathbb{1}\{\text{Share}_{ihdt} = c\}_{c=1,..,20}$ and corresponding instrumental variables, $\mathbb{1}\{\text{PredictShare}_{ihdt} = c\}_{c=1,..,20}$, each representing a distinct clustering level.

$$Y_{ihdt} = \sum_{c=1,..,20} \alpha_c \mathbb{1}\{\text{Share}_{ihdt} = c\} + \gamma_{dt} + \gamma_{ht} + \gamma_{T^e} + \text{AveShare}_{ihd^p t} + \epsilon_{ihdt} \quad (3)$$

I visualize these coefficients $\alpha_{cc=1,..,20}$ in Figure 5. It shows that new workers achieve higher productivity as the clustering level rises, but the marginal gain decreases. Specifically, when the clustering level reaches around 10%, adding one more same-origin peer does not lead to a further increase in productivity.

3.4 Additional Validity Test: Restaurant Clusters and Worker Clusters

The IV regression establishes that higher clustering increases new workers' productivity. We may also wonder while new workers are more productive in high-clustering districts, it may be due to factors unrelated to migrant networks. For example, suppose Cantonese restaurants are all located in one district. Cantonese delivery workers may also work in that district, and they are more productive than other workers. However, it might be because they communicate more efficiently with Cantonese restaurant owners, but not because they have a larger migrant cluster. While this endogeneity issue does not directly bias the magnitude of estimates, it could influence the interpretation of the results.

To address this concern, I directly compare the geographic distributions of restaurants and workers from the same origins. Specifically, I infer the origins of restaurants based on their brands and cuisines, such as "Cantonese food" or "Sichuan food." As shown in Figure 6, I find no significant correlations between restaurant clusters and workers clusters from the same origins. This test provides additional support for results that the migrant network increases productivity through the knowledge spillover channel.

3.5 Other Channels of Learning

The analysis above has quantified the overall impacts of clustering on productivity and highlighted one crucial mechanism: new workers can learn from peers' delivery experiences, which saves time searching for restaurants and consumer locations. In reality, delivery workers can learn more from each other. Here, I provide suggestive evidence on some additional mechanisms of peer learning.

First, I explore a quasi-experiment based on unexpected entrance closure and analyze how information spreads through the social network. In Figure B.12(a), I provide one example of workers deviating from the fastest delivery routes. The black line represents a worker who initially went to a closed gate of a residential compound and then took a detour to another open gate, resulting in a four-minute delay. Following this, I identified 20 compounds in Shanghai that unexpectedly closed one of their entrances in 2021. As Figure B.12(b) shows, the following day, nearly 40% of workers still tried the locked entrance. Interestingly, workers with referrers who previously visited the locked entrance were much less likely to make this error (figure B.12). This suggests that workers share their daily experiences about the ever-changing urban landscape with each other.

Second, Figure B.16 shows that new workers tend to turn out for work around the same time as their referrers. Specifically, I find that hourly wages usually peak at lunch (11:30 a.m.-1 p.m.) and dinner (5:30-6:30 p.m.). Locating in the same district with one's referrer significantly increases new workers' probability of working during these high-earning hours versus afternoons.

These findings highlight two main mechanisms of peer learning: (1) learning from peers' location-specific experiences and (2) emulating peers' behaviors. Thus, having more peers nearby expands new workers' accessible knowledge pool and increases average productivity.

4 The Impact of Clustering on Income Risks

In this section, I analyze the labor market congestion costs of, which increase migrant workers' income volatility.

4.1 Individual Labor Supply Responses to Shocks

Existing studies have extensively documented that migrant workers tend to work longer and function as insurance against adverse shocks in rural areas (Yang, 2011). The reason is that migrant workers care about left-behind family members and aim to send remittances when adverse shocks occur in their hometowns. In this context, delivery workers are usually the primary income source for their household. Furthermore, the flexibility of the gig economy allows delivery workers to quickly adapt their labor supply in response to hometown shocks (Michuda, 2021). To estimate how delivery workers respond to adverse hometown shocks, I construct two types of shocks: floods and pandemic lockdowns.

First, I use floods that occurred between June and August 2020, the worst since 1998 (Wall Street Journal, 2020).[35] I identify floods based on daily precipitation data. Specifically, I create a proxy for floods by identifying instances where accumulated rainfall over two days exceeds 160*mm* in each county and week, doubling the threshold for heavy rain.[36] Of 2,423 counties with precipitation data, 8% experienced flood shocks at least once during the period. Among counties with at least one active delivery worker in the

[35]According to the Ministry of Emergency Management, the floods have affected 63.46 million people and caused a direct economic loss of 178.96 billion RMB, which are 12.7% and 15.5% higher than the 2015-2019 average (Post, 2020).

[36]In most countries, the threshold for heavy rain and above is experiencing 80 millimeters per 24 hours. I also run robustness checks using different rainfall thresholds.

sample, 11% experienced flood shocks at least once. Figure B.17 presents the geographic distribution of flood shocks.

Second, between 2020 and 2022, many places in China experienced occasional strict pandemic lockdowns. Lockdowns in workers' hometowns meant their left-behind family members could not work outside, which might also incentivize migrant workers to work longer and send remittances to support their families. I compile two data sources: (1) daily records of confirmed COVID-19 cases by city and (2) lockdowns in each county and week, indicated by the number of consumer orders on the platform falling below the half of normal medians.

I run a regression at the worker-week level to compare workers' labor supply before and after they experience adverse hometown shocks. The specification is as follows.

$$Y_{ihdt} = \beta \, \text{Shock}_{ht} + \gamma_i + \gamma_{pt} + \gamma_d + \epsilon_{ihdt} \tag{4}$$

where Shock_{ht} indicates if worker i's hometown h experienced any adverse shock in the past month.[37] I include workers' origin province-week fixed effect, γ_{pt}, to absorb origin region-related time fluctuations, individual fixed effect, γ_i, to control for worker heterogeneity and district fixed effect γ_d, to control for district heterogeneity.[38]

Table 4 presents the hometown flood results. Migrant delivery workers increase weekly labor supply by around six hours. By working longer, workers complete 16 additional deliveries, raising weekly income by around 150 RMB. The extensive margin also shifts, where the share of workers completing at least one delivery increases slightly among workers from affected counties.[39] Regressions examining the impacts of hometown pandemic lockdowns show similar patterns. As Table B.9 shows, both COVID cases and pandemic lockdowns in workers' hometowns increase migrant workers' labor supply. Magnitudes are slightly smaller but more precise since I have more accurate identification of the lockdown shocks.

I also plot the event study analysis of workers' labor supply dynamics around hometown shocks in Figure 7. The x-axis displays the number of weeks before and after the hometown shock, and the y-axis shows the changes in weekly working hours. Labor

[37]Changes over time can be found in the event study analysis in Figure 7. It shows that workers still work longer, around four weeks after the shock, compared to the normal period.

[38]Specifically, I add origin province-week fixed effects to capture different work habits for workers from different regions across the year. For example, workers from northern China may prefer to work longer during summer.

[39]Based on conversations with workers, this is partly due to the widely available digital cash transfer system in China. Workers can send money back without traveling back in person.

supply increases substantially in the first week after the shock and diminishes over time.[40] Additional robustness checks in which I vary the flood threshold can be found in Table B.8, suggesting similar results.

4.2 District Responses to Aggregate Shocks

I next estimate the aggregate impacts of migrant workers' hometown shocks in the destination labor market. The intuition is that consider a district with a large share of delivery workers coming from the same county of birth, h_c. If an adverse shock occurs in hometown h_c, these migrant workers may increase working hours around the same time, resulting in the total labor supply surging in the destination market.

I first construct a set of district shocks: the predicted share of workers experiencing adverse hometown shocks in each district and week, as below.

$$\text{ShockShare}_{dt} = \sum_h \frac{N_{hdt^0}}{N_{dt^0}} \text{Shock}_{ht}$$

where $\frac{N_{hdt^0}}{N_{dt^0}}$ is the hometown composition in each district in May 2020. I set it as fixed when computing the shock share in order to control for the endogenous changes in hometown compositions due to hometown shocks. The variation of ShockShare_{dt} thus comes from two sources: (1) different hometown compositions across districts and (2) hometown shocks that occur during different weeks. Figure B.19(a) shows the histogram of district shocks, and figure B.19(b) plots the correlation between hometown clustering levels and sizes of district shocks: a higher clustering level is correlated with a larger district shock.[41]

I next construct the indicator for aggregate district shocks as follows.

$$\text{AggShock}_{dt} = \mathbb{1}\{\text{ShockShare}_{dt} > 15\%\}$$

where I identify an aggregate shock when the predicted share of workers facing adverse hometown shocks exceeds 15% in a district. I choose this threshold for two reasons: (1) 15% is the average clustering level across districts. As discussed above, with i.i.d. hometown shocks, the shock share is highly correlated with the highest clustering level in

[40]This extended impact is likely due to the sector's relatively low wages, requiring long-term efforts to accumulate savings.

[41]By assuming that shocks are i.i.d. across hometowns, the upper bound of Shock_{dt} depends on the highest level of clustering level in the district: $\max_t \ \text{Shock}_{dt} \simeq \max_h \ \frac{N_{hdt^0}}{dt^0}$.

each district; (2) the model predicts concave impacts of aggregate shocks on labor supply. Adopting a binary shock indicator facilitates interpretation and make it easier to compare with previous estimates of individual hometown shock impacts.

I run the regression at the district-week level to estimate the impacts of aggregate shocks on district outcomes.

$$Y_{dt} = \delta \text{ AggShock}_{dt} + \gamma_d + \gamma_t + \epsilon_{dt} \tag{5}$$

where I add district fixed effect, γ_d, and week fixed effect, γ_t, to control for district heterogeneity and labor supply fluctuations across time. For the dependent variables, I calculate the total number of orders placed by consumers each week, the total number of active workers, and the total number of working hours.[42]

Table 5 presents the results. On the labor supply side, the total working hours rise substantially during aggregate district shocks, though the number of active workers remains stable. This implies that adverse hometown shocks cause existing migrant workers to work longer instead of attracting more workers to come and work in the short run. On the demand side, I do not find the total number of consumer orders responds significantly.[43] The results indicate large aggregate shocks result in labor supply surging without the increases of corresponding consumer demand. As a result, each worker is assigned fewer deliveries per unit of time.

As a robustness check, I present the results using continuous shock shares in Table B.10, which displays the same pattern. Furthermore, column (4) in Table 5 shows the per-delivery commission fee does not change significantly during aggregate shocks. This is because the platform does not frequently adjust commission fees for stability reasons.[44] With invariant per-delivery fees, the congestion effect on workers' real wages and income stems from fewer deliveries being assigned to each worker per hour during an aggregate shock, not lower per-delivery fees.

[42] Working hours add up all hours workers engage with the platform's application by opening it and taking some actions without necessarily delivering orders at that moment.

[43] This outcome is partly due to the low delivery fee in the context. Though the increases in labor supply potentially lead to faster delivery, I do not find consumers respond, at least in the short run.

[44] Based on conversations with the wage setting team, they only adjust per-delivery commission fee, especially lowering the commission fee, at most twice a year to avoid angering workers and damaging reputation.

4.3 Effect of Clustering on Income during Hometown Shocks

Finally, I analyze how clustering affect workers' income during adverse hometown shocks. As noted, during aggregate hometown shocks, labor supply rises, but consumer demand remains constant. This results in workers being assigned fewer deliveries per hour. In equilibrium, though the per-delivery commission fee paid to workers stays constant (as shown in Table 5), workers' real wage decreases due to the labor market congestion.

This effect can be more pronounced for clustered workers who face individual and aggregate shocks simultaneously. Labor supply surges precisely when they desire to work longer. I use the following specification to analyze this general equilibrium effect of clustering.

$$Y_{ihdt} = \beta \text{ Shock}_{ht} + \delta \text{ AggShock}_{dt} + \alpha \text{ Shock}_{ht} \times \text{ AggShock}_{dt} + \gamma_i + \gamma_{pt} + \gamma_d + \epsilon_{ihdt} \quad (6)$$

where the interaction terms, $\text{Shock}_{ht} \times \text{AggShock}_{dt}$, captures the general equilibrium impacts of experiencing the individual and aggregate shocks at the same time. Similar to the previous two regressions, I include workers' origin province-week fixed effect, γ_{pt}, to absorb origin region-related time fluctuations, individual fixed effect, γ_i, to control for worker heterogeneity, and district fixed effect γ_d, to control for district heterogeneity.

Table 6 presents the findings. In column (3), workers increase weekly hours by around six hours on average after adverse hometown shocks, regardless of whether they experience any aggregate district shocks. The positive coefficient on Shock_{ht} in column (2) shows that affected workers complete more deliveries by working longer. However, the interaction term in column (2) reveals that when experiencing individual hometown shocks and aggregate shocks simultaneously, clustered workers complete much fewer deliveries despite working similar additional hours as non-clustered workers.

In terms of magnitudes, during adverse hometown shocks, non-clustered workers complete 20 additional deliveries by working longer, earning 200 RMB extra. On the contrary, clustered workers only complete around ten more deliveries, earning 100 RMB extra, even though they work similar hours as non-clustered workers. A back-of-the-envelope calculation shows that during adverse hometown shocks, clustered workers' real wages are 10% lower than the normal period.

4.4 Negative Externality of Clustering: Peak v.s. Off-peak hour

The congestion cost of clustering is relevant for both clustered workers and non-clustered workers in the district, albeit to varying degrees. Evaluating the negative externality of clustering requires examining delivery patterns.

Specifically, due to the food delivery sector's characteristics, 80% of workers are active during lunch and dinner hours, as shown in Figure B.20. As a result, there is limited room to expand their labor supply further during these periods. Instead, workers facing hometown shocks are more likely to work in the afternoon and late evening (Figure B.21). With fewer consumer orders during these non-peak hours, the labor supply surge intensifies competition further.

In Table B.11, I separate worker performance into peak (11 am-2 pm, 6 pm-8 pm) versus non-peak hours. Aligning with illustrated work patterns, the congestion effects occur mainly during non-peak times. On the one hand, workers facing hometown shocks receive lower wages precisely when they have the greatest desire to earn more. On the other hand, workers who do not experience hometown shocks themselves but operate in districts with aggregate shocks receive lower real wages during these periods. As a result, they work for fewer hours, complete fewer deliveries, and earn less income weekly.

In summary, I analyze the congestion cost of clustering in this section. Adverse hometown shocks increase migrant worker's weekly labor supply by six hours on average. For clustered workers, individual shocks translate into aggregate shocks. This intensifies competition for consumer orders and lowers real wages in the equilibrium. The congestion cost becomes more pronounced with higher clustering. Since each hometown has some probability of experiencing adverse shocks, clustering raises income volatility and risks.

5 Theoretical Framework

Drawing on empirical results from sections 3 and 4, I establish a model to quantify the trade-off between the two countervailing effects of migrant networks: learning benefits and congestion costs. A model is essential here because the learning benefits increase workers' average productivity and income, while the congestion costs increase workers' income volatility. Thus, trading off the two forces requires a credible estimation of workers' risk aversion coefficient and additional parameters governing the utility function. The model also guides discussions on the market externality of clustering and allows counterfactual policy analysis.

The model has four stages. In stage 1, conditional on migrating to the destination city and becoming a delivery worker, new migrant workers form expectations about working in different districts and choose to work in a district with the highest expected utility. In stage 2, the hometown shocks are realized. In stage 3, workers decide on working hours and remittances to maximize utility, where they care about their own consumption,left-behind family consumption, and own leisure time. Finally, the equilibrium wage is determined through the market clearing condition in each district. The model also explicitly assumes that workers are aware of the learning and congestion effects of clustering and consider both when choosing work districts.[45]

5.1 Choices of Work Districts

Consider a new worker i from hometown h. The utility of worker i working in district d, as denoted by V_{ihd}, is expressed as

$$V_{ihd} = \mathbb{E}U_{hd} + \varepsilon_{ihd} \tag{7}$$

where $\mathbb{E}U_{hd}$ is the expected utility of working in district d for any worker from hometown h, and ε_{ihd} is the idiosyncratic preference shock, following a type-1 extreme value distribution with scale, $1/\sigma$. The probability of workers from hometown h working in district d, $\pi_{hd|h}$, is as follows.

$$\pi_{hd|h} = \frac{\exp(\sigma \mathbb{E}U_{hd})}{\sum_{d'} \exp(\sigma \mathbb{E}U_{hd'})} \tag{8}$$

5.2 Labor Supply Decision

After worker i chooses to work in a district d, the worker decides on labor supply, L_i, and the amount of remittances sent back to their families, T, to maximize utility as follows.

[45] In Section 6, I provide empirical evidence that workers consider congestion costs by exploring hometown heterogeneity. Other studies have also shown the existence of dispersion forces in migrant workers' location choices(e.g., Card and Lewis, 2007; Monras, 2020)

$$
\begin{aligned}
\max_{L,T} \quad & (C_i^{\text{own}})^{\alpha}(C_i^{\text{family}})^{\beta}(\bar{L}-L_i)^{1-\alpha-\beta} \\
\text{s.t.} \quad & C_i^{\text{own}} + T_i = \tilde{w}_i L_i \\
& C_i^{\text{family}} = S_h + T_i
\end{aligned}
$$

where C^{own} is the worker's own consumption, C^{family} is the left-behind family's consumption, $\bar{L} - L_i$ is the worker's leisure time.

The two budget constraints are (1) workers allocate their total earnings, $\tilde{w}_i L$, between their own consumption and remittances sent back to families, as denoted by T; and (2) the left-behind family's consumption depends on the remittance received and the agricultural income, S, which is assumed to be stochastic with the following distribution:

$$
S = \begin{cases} S^{\text{low}} & \text{with probability } P_{\text{shock}} \\ S^{\text{high}} & \text{with probability } 1 - P_{\text{shock}} \end{cases}
$$

where $S^{\text{high}} > S^{\text{low}}$. Workers from the same origin face the same level of hometown shock S_h. Shocks for workers from different origins are independent.

Market Clearing Condition

Assume that each delivery worker completes B_i deliveries per unit of time. In the equilibrium, the number of deliveries workers complete shall equal the number of orders put by consuemrs. The market clearing condition is expressed as:

$$
\sum_i B_i L_i = D_d
$$

where D_d is the total consumer demand.

Learning from Peers

New workers can learn how to navigate local neighborhoods from hometown workers in the same district. This enables new workers to deliver faster and complete more deliveries per hour. I thus denote each worker's delivery speed as a function of the number of same-origin peers in the district, $\delta(N_{hd})$.

Specifically, I assume that a worker without peers ($N_{hd} = 0$) can finish B_d number of deliveries per unit of time. B_d can also be treated as the average productivity. Second, a worker i with N_{hd} peers nearby can finishes $B_i = \delta(N_{hd})B_d$ number of deliveries per unit of time. I further assume that $\delta(N_{hd})$ has the following properties.

$$\delta(N_{hd} = 0) = 1 \; ; \delta'(N_{hd}) \geq 0 \; ; \delta^{''}(N_{hd}) \leq 0$$

which ensure that workers' productivity increases with the clustering level, but the marginal gain from clustering decreases.

Real Wage

With $B_i = \delta(N_{hd})B_d$, the market clearing condition becomes as follows.

$$\sum_h \delta(N_{hd})B_d L_{hd} N_{hd} = D_d$$

A delivery worker's real wage equals the number of deliveries completed per hour times the per-delivery commission fee, A.[46] I can solve the real wage as follows.

$$\tilde{w}_i = \tilde{w}_{hd} = A \times \delta(N_{hd})B_d = A \times \delta(N_{hd}) \frac{D_d}{\sum_{h'} \delta(N_{h'j}) L_{ih'j} N_{h'j}}$$

5.3 Model Predictions

From the utility maximization problem and market clearing conditions, I obtain the real wage and optimal labor supply as below.

$$\begin{cases} \tilde{w}^*_{hd} = \frac{1}{\alpha+\beta} \frac{\delta(N_{hd})}{\sum_{h'} \delta(N_{h'j}) N_{h'j}} \frac{AD_d + (1-\alpha-\beta)\sum_{h'} S_{h'j} N_{h'j}}{\bar{L}} \\ \\ \frac{L_i^*}{\bar{L}} = (\alpha+\beta)\left(1 - \frac{\sum_{h'} \delta(N_{h'j}) N_{h'j}}{\delta(N_{hd})} \frac{(1-\alpha-\beta)S_{ihd}}{AD_d + (1-\alpha-\beta)\sum_{h'} S_{h'j} N_{h'j}}\right) \end{cases} \tag{9}$$

The indirect utility for worker i from hometown h to work in the district d, as denoted by u_{hd}, depends on the hometown composition in the district and realizations of hometown shocks, $(\{N_{hd}\}, \{S_{hd}\})$.

[46] The platform does not adjust this fixed fee frequently for stability reasons. I assume this commission fee as fixed.

$$u_{hd} = \text{Constant} \times \left(\left(\delta(N_{hd}) w_d^*\right)^{\alpha+\beta} + \frac{S_{hd}}{\bar{L}} \left(\delta(N_{hd}) w_d^*\right)^{\alpha+\beta-1} \right) \tag{10}$$
$$\text{where } w_d^* = \frac{AD_d + (1-\alpha-\beta) \sum_{h'} S_{h'j} N_{h'j}}{(\alpha+\beta) \bar{L} \sum_{h'} \delta(N_{h'j}) N_{h'j}}$$

From the solutions above, I derive the following comparative statistics.

1. $\partial \tilde{w}^*_{hd}(S_h = S^{\text{high}})/\partial N_{hd} > 0$: when there is no hometown shock, a higher clustering level leads to higher real wage due to the learning effect.

2. $\partial L_i^*/\partial S_{ihd} < 0$: workers who face hometown shocks, S^{low}, increase labor supply.

3. $\partial \tilde{w}^*_{hd}/\partial(\sum_{h'} S_{h'j} N_{h'j}) > 0$: When more workers experience adverse hometown shocks, the real wage in the district decreases due to the congestion effect.

4. $\partial u_{ihd}/\partial S_{ihd} > 0$: workers' utility decreases during adverse hometown shocks.

5. $\partial u_{ihd}/\partial(\sum_{h'} S_{h'j} N_{h'j}) > 0$: workers' utility decreases as more workers experience adverse hometown shocks in their work districts.

Clustering Effect

Clustering can significantly impact workers' utility for two main reasons: (1) a worker's productivity, $\delta(N_{hd})$, directly depends on the number of same-origin workers in their district; (2) due to i.i.d. hometown shocks, the aggregate shock in a district, $\tilde{S}_d$, is a function of the clustering level, N_{hd}, as follows.

$$\tilde{S}_d = \sum_{h'} S_{h'j} N_{h'j} = N_{hd} S_h + \sum_{h' \neq h} S_{h'j} N_{h'j} \simeq N_{hd} S_h + (N_d - N_{hd}) \bar{S}$$
$$\text{where } \bar{S} = P_{\text{shock}} S^{\text{low}} + (1 - P_{\text{shock}}) S^{\text{high}}$$

The covariance between individual and aggregate shocks, $Cov(S_{hd}, \tilde{S}_d) = N_{hd} Var(S_h)$, increases as the clustering level, N_{hd}, increases.

Proposition 1. *When $\delta(N_d) < \eta_1$ and $\delta'(N_d) < \eta_2$[47], $u_{hd}(S_{hd} = S^{high})$ increases with N_{hd}. $u_{hd}(S_{hd} = S^{low})$ first increases but then decreases with N_{hd}.*
The proof and the choice of the threshold, η_1 and η_2, can be found in the appendix.[48]

[47] N_d is the total number of workers in the district

[48] In a later empirical analysis, I estimate $\delta(N_d) \simeq 5\%$ and $\delta'(N_d) \simeq 0$. Along with other estimated parameters, such as estimated S^{high} and S^{low}, I find these two thresholds are satisfied.

This proposition highlights the main trade-off of clustering. On the one hand, clustering with same-origin workers enables new workers to learn from peers and achieve higher productivity: $\mathrm{d}\delta(N_{hd})/\mathrm{d}N_{hd} > 0$. However, during bad times, a high clustering level (1) leads to larger aggregate shocks, $\tilde{S}_d$; (2) drives down the real wage, $\tilde{w}_d^*$; and (3) offsets the learning benefits of clustering.

Finally, since workers choose work districts before their hometown shocks are realized, new workers form expectations based on the probability of hometown shocks. I also assume that workers are risk averse, with a relative risk-aversion ratio γ. I derive the expected utility of working in district d, as denoted by $\mathbb{E}U_{ihd}^{\text{shock}}(c_h)$, as follows:

$$\mathbb{E}U_{hd} \quad = \tfrac{1}{1-\gamma}\left((1 - P_{\text{shock}})(u_{hd}^{\text{noshock}})^{1-\gamma} + P_{\text{shock}}(u_{hd}^{\text{shock}})^{1-\gamma} \right) \tag{11}$$

Proposition 2. *When $\delta'(N_d) < \eta_3$, the expected utility, $\mathbb{E}U_{hd}$, first increases with N_{hd}, and then decreases. There exists an optimal clustering level, $N_{hd}^*(\gamma) = \max \mathbb{E}U_{hd}$*
The proof and the choice of the threshold, η_3, can be found in the appendix.[49]

This proposition shows that for a risk-averse worker, the productivity effect of clustering dominates the congestion effect under a low level of clustering. However, under the high level of clustering, the congestion effect dominates. In addition, the risk aversion coefficient governs the exact magnitude of this trade-off between the higher average utility and higher utility variance.

Equilibrium

I define the equilibrium as the fixed point where each worker's choice of work district is optimal given their randomly drawn preference and other workers' choices. In other words, the probability of new workers entering different districts will be consistent with the distribution of existing workers across districts, which new workers use to form expectations. This is formulated as a fixed point of the best response mapping.[50]

$$\mathbb{E}[\frac{N_{hd}}{N_h}] = \frac{\exp(\sigma \mathbb{E}U(\{N_{h'd}\}_{h'})}{\sum_{d'} \exp(\sigma \mathbb{E}U(\{N_{h'd'}\}_{h'})} \tag{12}$$

[49]In a later empirical analysis, I estimate $\delta'(N_d) \simeq 0$. Along with other estimated parameters, I find this threshold is satisfied.

[50]$\{N_{h'd}\}_{h'})$ represents the entire hometown composition in the district d since worker i's utility not only depends on one's hometown peers but is also affected by other hometown clusters.

5.4 Market Externality

Workers' choices of work location change clustering levels across districts and can inherently impact the utilities of other workers. For a worker i from hometown h, the entry into district d increases the clustering level by $\triangle N_{hd} > 0$. I decompose externalities associated with this entry decision into two types.

1. Learning externality: worker i can share knowledge with other workers from hometown h, leading to higher productivity.

$$\Phi^{\text{learn}}(N_{hd}) = \triangle \mathbb{E}U^{\text{learn}}(N_{hd}) \times N_{hd}$$

   where $\mathbb{E}U^{\text{learn}}(N_{hd})$, which is the expected utility with only the learning benefits.

2. Congestion externality: worker i increases every worker's income variance because of the possibility of a bigger aggregate shock.

$$\Phi^{\text{cost}}(N_{hd}) = \triangle \mathbb{E}U^{\text{cost}}(N_{hd}) \times N_{hd} + \triangle \mathbb{E}U^{\text{cost-other}}(N_{hd}) \times (N_d - N_{hd})$$

   where $\mathbb{E}U^{\text{cost}}(N_{hd})$ is the expected utility with only the congestion costs for workers from hometown h, and $\mathbb{E}U^{\text{cost}}(N_{hd})$ is the one for the rest of workers in district d.

Proposition 3. *When $\delta'(N_{hd} = 0) < \eta_4$, the overall externality, $\Phi(N_{hd}) = \Phi^{learn}(N_{hd}) + \Phi^{cost}(N_{hd})$ is negative for any clustering level N_{hd}.*
The proof and the choice of the threshold, η_4, can be found in the appendix.[51]

Proposition 3 highlights that when workers choose work districts to maximize individual utilities, the equilibrium clustering may not be optimal if the central planner accounts for the two types of externalities as above. I discuss these magnitudes in more detail in section 7 after estimating key parameters.

6 Estimation

I structurally estimate the model using empirical results in sections 3 and 4, as well as additional variation from model predictions and quasi-experiments for key parameters.

[51] In a later empirical analysis, I estimate $\delta'(N_{hd} = 0) \simeq 2$. Along with other estimated parameters, I find this threshold is satisfied.

6.1 Estimation

The main parameters are as follows.

1. $\delta(N)$: productivity as a function of the clustering level
2. P_{shock}: probability of hometown shocks
3. $\alpha + \beta$: Cobb-Douglas utility function parameter
4. $(S^{\text{high}}, S^{\text{low}})$: household agricultural income with and without adverse shocks
5. γ: workers' risk aversion coefficient
6. σ: the scale of idiosyncratic preference for working in different districts

I discuss the estimation approach in three steps. First, I outline the identification strategy for parameters $(\delta(N), P_{\text{shock}}, \alpha + \beta, S^{\text{high}}, S^{\text{low}})$, which bring learning benefits and congestion costs of clustering into the utility function and quantifies worker's utility under both scenarios. Second, I turn to γ, which governs the concavity of workers' utility function and the relative importance of learning benefits versus congestion costs. Third, I discuss σ, affecting workers' location choices.

Part 1: Utility Function

$\boldsymbol{\alpha + \beta}$. This Cobb-Douglas parameter represents worker's labor supply elasticity to income. I estimate it through a quasi-experiment: semi-annual worker festivals organized by the food delivery platform. During the festival, the platform randomly selects workers to receive cash prizes ranging from 50 to 1000 RMB.[52] Following the comparative statistics from the mode, I regress workers' working hours on prize amounts won in these lotteries to estimate $\alpha + \beta$. Sepcficially, I divide each worker's prize amount by the average hourly wage in their work district as the independent variable. The coefficient of this adjusted prize amount then maps directly to the parameter $(1 - \alpha - \beta)$. able 7 reports the results.

$\boldsymbol{(S^{\text{high}}, S^{\text{low}})}$. As the model predicts, hometown income levels affect migrant workers' labor supply. Specifically, during adverse hometown shocks (S^{low}), workers tend to work longer to send more remittances. Therefore, I can identify the two hometown

[52]All delivery workers can participate by submitting their phone numbers associated with their accounts on the platform. The number of workers winning any prize is usually around 500-1000 people. The process of drawing the lottery prize is completely random.

income levels by regressing workers' labor supply on shock indicators, as shown in Table 4 in section 5. The identification intuition is that once (α, β) are estimated, I can recover these two hometown income levels by comparing workers' labor supply during normal periods versus adverse hometown shocks and using workers' labor supply responses to lottery prizes as the benchmark.

$\boldsymbol{(\delta(c), P_{\text{shock}})}$. The productivity function, $\delta(c)$, is directly identified from the IV regression in section 3 (figure 5). I calibrate P_{shock} based on the annual flood probability for each county based on 2000-2020 precipitation data.

Part 2: Risk Aversion Coefficient

I identify γ following the model prediction that workers with a higher probability of adverse hometown shocks will cluster less since they are more exposed to the congestion cost of clustering. Table B.13 confirms this pattern, where I regress the average clustering level on the flood probability for each hometown. It shows that workers from hometowns with a higher shock probability cluster less than workers from safe hometowns.[53] Furthermore, the more risk-averse workers there are, the bigger the difference is. In other words, the extent to which risky-hometown workers cluster less relative to safe-hometown workers reveals workers' degree of risk aversion.

I thus divide workers into two groups based on whether their hometowns' shock probabilities are above (risky) or below (safe) the median. I compute clustering levels separately for workers from risky versus safe hometowns. With log-normalization and triple differences, I construct a moment to recover the concavity of workers' expected utility function as below.

$$\frac{\triangle log(\pi(N_1)) - \triangle log(\pi(N_2))}{\triangle log(\pi(N_3)) - \triangle log(\pi(N_2))} = \frac{\triangle u(N_1) - \triangle u(N_2)}{\triangle u(N_3) - \triangle u(N_2)} \tag{13}$$

where $\pi(N)$ is the new worker's entry probability at each clustering level, N, and $\triangle u(N)$ is the utility difference between periods with and without hometown shocks at clustering level N, $\triangle u(N) = u(N, S^{\text{high}})^{1-\gamma} - u(N, S^{\text{low}})^{1-\gamma}$.

It can be proven that this moment only depends on the risk aversion coefficient and

[53]This result also provides evidence that workers are aware of this labor market congestion cost of clustering when choosing work locations, as the model assumes. I conduct robustness checks on whether the shock probability affects workers' behaviors in other dimensions, such as delivery speed. Table B.14 shows no significant differences, except for weakly positive effects on the number of active workers.

utility-related parameters in part 1, where I cancel out the idiosyncratic preference scale, σ, in the process. The intuition is that while the difference of entry probability between two clustering levels, $\triangle log(\pi(N_1)) - \triangle log(\pi(N_2))$, depends on the joint product of utility difference ($\triangle u$) and σ, adding a third clustering level and comparing the relative difference across all three levels cancel out σ and recover the concavity of the utility function. Calculation details are in the appendix A.5. I thus compute the entry probability across clustering levels from 0% to 15% for both risky and safe hometowns as the empirical moments.

Part 3: Scale of Idiosyncratic Preferences

Following the model in section 3 (equation 7), for a new worker i who comes from hometown h joining the platform at time $t-1$, the utility of worker i working in district d, V_{ihdt}, consists of (1) the estimated expected utility, $\mathbb{E}U(N_{hdt})$, which depends on the clustering level of hometown h in district d, N_{hdt}, and (2) workers' idiosyncratic preferences ε_{ihdt}. Building on this baseline model, I add a third utility component, entry bonuses in the district, B_{dt}, if any. With ε_{ihdt} following a type-1 extreme value distribution with scale, $1/\sigma$, I derive the gravity equation as follows:

$$\mathbb{E}[\frac{N^{\text{new}}_{hdt-1}}{N^{\text{new}}_{ht-1}}] = \frac{\exp(\sigma \mathbb{E}U(N_{hdt}\ , B_{dt}))}{\sum_{d'} \exp(\sigma \mathbb{E}U(N_{hd't}\ , B_{dt}))} \tag{14}$$

Furthermore, I split the utility into two parts, quantify the increases in the utility driven by entry bonuses, and rewrite the gravity equation.[54]

$$\mathbb{E}U(N_{hdt}, B_{dt}) = \mathbb{E}U(N_{hdt}) + \triangle \mathbb{E}U(B_{dt}) \tag{15}$$

$$\mathbb{E}[\frac{N^{\text{new}}_{hdt-1}}{N^{\text{new}}_{ht-1}}] = \frac{\exp(\sigma \mathbb{E}U(N_{hdt}) + \sigma \triangle \mathbb{E}U(B_{dt}))}{\sum_{d'} \exp(\sigma \mathbb{E}U(N_{hd't}) + \sigma \triangle \mathbb{E}U(B_{d't}))} \tag{16}$$

Since entry bonuses are exogenous, I run the gravity regression with the entry bonuses as the independent variables.[55] As in Table 8, the coefficient in front of $\triangle \mathbb{E}U(B_{dt})$, is the estimate for the scale of the idiosyncratic preferences, σ.

Implementation. I estimate parameters, $\{S^{\text{high}}, S^{\text{low}}, \gamma\}$ together using two-step GMM. The values $\alpha + \beta$ and σ are estimated by separate regressions, as discussed above.

[54] Simulates confirm that $\triangle \mathbb{E}U(N_{hdt}, B_{dt})$ at any level of N_{hdt} is very similar to $\triangle \mathbb{E}U(B_{dt})$ where I set $N_{hdt} = 0$.

[55] I run the gravity regression using Poisson Pseudo MLE (Sotelo, 2019; Dingel and Tintelnot, 2020).

6.2 Estimates and Model Fit

Table 9 presents the estimation results. The estimated parameters are in the expected range. The Cobb-Douglas parameter falls between 0 and 1 as restricted. The risk aversion coefficient aligns with estimates from most literature with a range between 1 and 10 (e.g., Swanson, 2012). The hometown income during adverse shocks is significantly lower than in normal periods, consistent with the empirical findings.

I validate the model estimates by plotting the percentage of new workers who choose districts at different clustering levels from the data and the estimated model in Figure 8. Overall, the two distributions follow the same pattern, where the entry probability first increases with the clustering level and then decreases. This is consistent with the learning-congestion trade-off predicted by the model. The two curves also peak around the same clustering level.

The exact magnitude of entry probability differs slightly between the two distributions for several potential reasons. For example, some new workers may face information friction and do not know where their hometown peers work. Following this intuition, I assume that some new workers do not know anyone in the city and choose districts solely based on their idiosyncratic preferences. I infer this number by the share of new workers who do not have a referrer. After adding this information friction to the model, I find that the existing estimation does not need to be changed much. Figure 8(c) shows that model predictions align even closer to the data distribution after adding this fix.

7 Utility, Externality, and Counterfactuals

In this section, I quantify a migrant worker's utility under different clustering levels, given the estimated parameters. I also discuss the market externality of clustering and simulate equilibria with and without accounting for externality. Lastly, I evaluate counterfactuals where workers receive insurance or better learning tools and analyze how workers' location choice change.

7.1 Inverted-U-shaped Utility

I first quantify workers' expected utility with clustering following the specification 11. To illustrate different forces clearly, I consider a simple case where only one same-origin cluster at level N_c exists in a district. The rest of the workers in the district come from different hometowns. I plot the utility of workers from this clustered hometown, $\mathbb{E}U(N_c)$, in Figure 9. It indicates a clear inverted-U shape where the expected utility first increases with the clustering level, reaches its peak around 10% clustering level, and decreases afterward.

This inverted U shape arises because worker utility first increases due to the learning benefits and falls due to the congestion costs as the clustering level increases. To quantify the two effects separately, I decompose $\mathbb{E}U(N_c) \sim \mathbb{E}U^{\text{learn}}(N_c) + \mathbb{E}U^{\text{cost}}(N_c)$.

$\mathbb{E}U^{\text{learn}}(N_c)$ is defined as the utility function with only the learning effect, without congestion effect during hometown shocks.

$$\mathbb{E}U^{\text{learn}}(N_c) = \frac{1}{1-\gamma}\Big((1-P_{\text{shock}})(u(N_c, S^{\text{high}}))^{1-\gamma} + P_{\text{shock}}(u(0, S^{\text{low}}))^{1-\gamma}\Big) \tag{17}$$

$\mathbb{E}U^{\text{cost}}(N_c)$ is the utility function with only the congestion effect, without learning effect during normal periods.

$$\mathbb{E}U^{\text{cost}}(N_c) = \frac{1}{1-\gamma}\Big((1-P_{\text{shock}})(u(0, S^{\text{high}}))^{1-\gamma} + P_{\text{shock}}(u(N_c, S^{\text{low}}))^{1-\gamma}\Big) \tag{18}$$

I plot both functions in Figure 9. $\mathbb{E}U^{\text{learn}}(N_c)$ is an increasing and concave curve, following the same pattern as the productivity function, $\delta(N)$. The intuition is that the additional productivity gain from higher clustering fades away quickly after the clustering level reaches 10%. $\mathbb{E}U^{\text{cost}}(N)$ strictly decreases as the clustering level increases since a bigger cluster results in stronger congestion effects during hometown shocks.

I also quantify the utility curve for the rest of the non-clustered workers in the district.

$$\mathbb{E}U^{\text{other}}(N_c) = \frac{1}{1-\gamma}\Big((1-P_{\text{shock}})^2(u(0, S^{\text{high}}))^{1-\gamma} + (1-P_{\text{shock}})P_{\text{shock}}(u(N_c, S^{\text{high}}))^{1-\gamma} \tag{19}$$

$$+P_{\text{shock}}(1-P_{\text{shock}})(u(0, S^{\text{low}}))^{1-\gamma} + P_{\text{shock}}^2(u(N_c, S^{\text{low}}))^{1-\gamma}\Big) \tag{20}$$

I plot the curve in Figure 9, which also strictly decreases. These non-clustered workers do not enjoy any knowledge spillovers but only experience the congestion effect. Furthermore, with i.i.d. hometown shocks, non-clustered workers usually do not experience their hometown shocks and the aggregate district shocks around the same time. As a result, their utility decreases less due to the congestion costs $|\mathbb{E}U^{\text{other}}(N_c)| <$

$|\mathbb{E}U^{\text{cost}}(N_c)|$

7.2 Externality of Clustering

Both the learning and congestion effects of clustering generate market externalities. On the one hand, knowledge spillovers have a positive externality as workers do not internalize teaching value. On the other hand, congestion effects result in workers waiting and wasting time during shocks, producing a negative externality. A central planner may improve the total welfare by realizing the full potential of knowledge spillovers and more efficiently allocating labor supply to avoid wasted time.

Following the definition in section 3, I first quantify the positive learning and negative congestion externality across clustering levels.

1. Learning externality: $\Phi^{\text{learn}}(N_c) = \triangle\mathbb{E}U^{\text{learn}}(N_c) \times N_c$
2. Congestion externality: $\Phi^{\text{cost}}(N_c) = \triangle\mathbb{E}U^{\text{cost}}(N_c) \times N_c + \triangle\mathbb{E}U^{\text{other}}(N_c) \times (N - N_c)$

I plot these curves in Figure 10. The learning externality is positive but decreases rapidly as the clustering level increases, consistent with the concavity of the productivity function $\delta(N)$ where marginal teaching values diminish quickly. The congestion externality is negative and decreases since a larger cluster leads to more severe congestion during hometown shocks.

I also plot the total externality, $\Phi(N_c) = \Phi^{\text{learn}}(N_c) + \Phi^{\text{cost}}(N_c)$, in Figure 10. It is negative and strictly decreases, differing from the inverted U-shape of individual expected utility, $\mathbb{E}U(N)$. These results suggest that accounting for the externalities may lead to an equilibrium with lower clustering levels than observed.[56]

7.3 Counterfactual: Equilibrium with Externality

In the first counterfactual, I simulate the equilibrium to incorporate externalities into workers' district choice. I compute the total externality associated with each entry decision follow the definition of externality in section 7.3 and without allowing other workers to adjust. I solve for the equilibrium where workers receive the Pigouvian tax equal to the externality of their choices when they deice where to work. I compare it with the

[56]If multiple clusters exist in one district, interactions exist across different clusters. The externality analysis follows the same logic but is more complicated to derive an analytical solution.

decentralized equilibrium, where workers choose where to work to maximize individual expected utility, $\mathbb{E}U(N_{hdt})$.

Implementation. I first simulate the number of hometowns and their size distributions based on the data. I start with a random allocation of workers across districts for each simulation. At each step, a small share of new workers arrives (as proportional to the distribution), and they choose districts following specification 7. I use an iterative procedure to solve for the fixed point, where the distributions of existing workers and new entries converge. Given the estimated parameters, the procedure is robust to varying starting conditions and identifies the same equilibrium. Notice that the equilibrium is defined as the distribution of clustering levels instead of specific hometown-district mappings, which are contingent on starting conditions. This equilibrium can also be interpreted as a stable distribution in the long run.

The first two bars in figure 11 compare the baseline decentralized equilibrium with the equilibrium with externliaty. Accounting for externalities significantly reduces the maximum clustering level compared to the decentralized equilibrium. Workers also create many more small clusters in the equilibrium with externalities. Though the highest clustering level decreases, I find workers' average productivity increases by around 1% instead of decreasing. This is mainly due to the diminishing marginal productivity return of clustering. Creating small clusters instead of a large cluster allows more workers to benefit from the knowledge spillovers without big losses of productivities due to the geographic dispersion of workers.

7.4 Counterfactual: Provide Insurance for Hometown Shocks

In the second counterfactual, I consider providing migrant workers with insurance to mitigate hometown shocks. Providing insurance eliminates the correlated shocks and labor supply responses among same-origin workers, which turns off the labor market congestion impacts of clustering. The third bar in figure 11 exhibites the counterfactual. The average clustering level almost doubles relative to the decentralized equilibrium. As a result, the average productivity of new workers increases by around 30% due to the higher clustering level.

7.5 Counterfactual: Replace Network-based Learning with Technology

In the last two counterfactuals, I simulate equilibrium where the platform can provide workers with better technologies to substitute network-based learning. For example, the platform can collect local knowledge from experienced workers and distribute the information to all workers through a centralized system. I consider two scenarios where technology can substitute 50% of the current network-based learning or 100% of peer learning in Figure 11. As expected, workers cluster much less in both scenarios, and the average productivity increases substantially since non-clustered workers can also benefit from the technology-based learning. By clustering less, migrant workers also experience smaller congestion costs, which increases workers' utility by 3-4% on average.

8 Discussion

Through several quasi-experiments and a model, I provide novel evidence of two countervailing effects of clustering: learning benefits and labor market congestion costs. I further examine how the two forces interact to affect workers' location choices and utility. I show that a worker's utility displays an inverted U shape with clustering, where utility first increases (due to learning) and then falls (due to congestion).

First, these results establish a crucial micro-foundation for understanding knowledge spillovers through social networks. Second, the congestion cost highlights a new general equilibrium channel in which clustered migrants compete with each other during adverse hometown shocks, reducing the effectiveness of their role as rural insurance. Third, quantifying the trade-off between the learning benefits and congestion costs offers valuable insights into the i mpacts of migrant networks.

References

Akram, Agha Ali, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak, "Effects of emigration on rural labor markets," Technical Report, National Bureau of Economic Research 2017.

Angrist, Joshua D, Sydnee Caldwell, and Jonathan V Hall, "Uber versus taxi: A driver's eye view," *American Economic Journal: Applied Economics*, 2021, *13* (3), 272–308.

Antoninis, Manos, "The wage effects from the use of personal contacts as hiring channels," *Journal of Economic Behavior & Organization*, 2006, *59* (1), 133–146.

Atkin, David, M Keith Chen, and Anton Popov, "The returns to face-to-face interactions: Knowledge spillovers in Silicon Valley," Technical Report, National Bureau of Economic Research 2022.

Bayer, Patrick, Stephen L Ross, and Giorgio Topa, "Place of work and place of residence: Informal hiring networks and labor market outcomes," *Journal of political Economy*, 2008, *116* (6), 1150–1196.

Beaman, Lori A, "Social networks and the dynamics of labour market outcomes: Evidence from refugees resettled in the US," *The Review of Economic Studies*, 2012, *79* (1), 128–161.

Beaman, Lori and Jeremy Magruder, "Who gets the job referral? Evidence from a social networks experiment," *American Economic Review*, 2012, *102* (7), 3574–93.

Biavaschi, Costanza, Corrado Giulietti, and Yves Zenou, "Social networks and (political) assimilation in the Age of Mass Migration," 2021.

Blumenstock, Joshua Evan, Guanghua Chi, and Xu Tan, "Migration and the value of social networks," 2021.

Borjas, George J, "Ethnicity, Neighborhoods, and Human-Capital Externalities," *The American Economic Review*, 1995, pp. 365–390.

Borusyak, Kirill and Peter Hull, "Non-random exposure to exogenous shocks: Theory and applications," Technical Report 2023.

Bureau, National Statistics, "Migrant Worker Monitoring and Survey Report," *The Government Report*, April 2020.

Burks, Stephen V, Bo Cowgill, Mitchell Hoffman, and Michael Housman, "The value of hiring through employee referrals," *The Quarterly Journal of Economics*, 2015, *130* (2), 805–839.

Card, David and Ethan G Lewis, "The diffusion of Mexican immigrants during the 1990s: Explanations and impacts," in "Mexican immigration to the United States" University of Chicago Press 2007, pp. 193–228.

Chen, M Keith, Peter E Rossi, Judith A Chevalier, and Emily Oehlsen, "The value of flexible work: Evidence from Uber drivers," *Journal of political economy*, 2019, *127* (6), 2735–2794.

Chen, Yuyu, Ginger Zhe Jin, and Yang Yue, "Peer migration in China," Technical Report, National Bureau of Economic Research 2010.

Cook, Cody, Rebecca Diamond, Jonathan V Hall, John A List, and Paul Oyer, "The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers," *The Review of Economic Studies*, 2021, *88* (5), 2210–2238.

Dai, Ruochen, Dilip Mookherjee, Kaivan Munshi, Xiaobo Zhang et al., "The community origins of private enterprise in China," Technical Report, Boston University-Department of Economics 2019.

Damm, Anna Piil, "Ethnic enclaves and immigrant labor market outcomes: Quasi-experimental evidence," *Journal of Labor Economics*, 2009, *27* (2), 281–314.

Datta, Namita, Chen Rong, Sunamika Singh, Clara Stinshoff, Nadina Iacob, Natnael Simachew Nigatu, Mpumelelo Nxumalo, and Luka Klimaviciute, "Working Without Borders: The Promise and Peril of Online Gig Work," 2023.

Davis, Donald R and Jonathan I Dingel, "A spatial knowledge economy," *American Economic Review*, 2019, *109* (1), 153–170.

Dingel, Jonathan I and Felix Tintelnot, "Spatial economics for granular settings," Technical Report, National Bureau of Economic Research 2020.

Duranton, Gilles and Diego Puga, "Micro-foundations of urban agglomeration economies," in "Handbook of regional and urban economics," Vol. 4, Elsevier, 2004, pp. 2063–2117.

_ and _ , "The economics of urban density," *Journal of economic perspectives*, 2020, *34* (3), 3–26.

Dustmann, Christian, Albrecht Glitz, Uta Schönberg, and Herbert Brücker, "Referral-based job search networks," *The Review of Economic Studies*, 2016, *83* (2), 514–546.

Edin, Per-Anders, Peter Fredriksson, and Olof Åslund, "Ethnic enclaves and the economic success of immigrants—Evidence from a natural experiment," *The quarterly journal of economics*, 2003, *118* (1), 329–357.

Egger, Dennis, Daniel Auer, and Johannes Kunz, "Effects of Migrant Networks on Labor Market Integration, Local Firms and Employees," 2021.

Fajgelbaum, Pablo D and Cecile Gaubert, "Optimal spatial policies, geography, and sorting," *The Quarterly Journal of Economics*, 2020, *135* (2), 959–1036.

Friebel, Guido, Matthias Heinz, Mitchell Hoffman, and Nick Zubanov, "What do employee referral programs do? Measuring the direct and overall effects of a management practice," *Journal of Political Economy*, 2023, *131* (3), 633–686.

Giuliano, Genevieve, Christian Redfearn, Ajay Agarwal, Chen Li, and Duan Zhuang, "Employment concentrations in Los Angeles, 1980–2000," *Environment and planning A*, 2007, *39* (12), 2935–2957.

Giulietti, Corrado, Jackline Wahba, and Yves Zenou, "Strong versus weak ties in migration," *European Economic Review*, 2018, *104*, 111–137.

Glaeser, Edward L, "Learning in cities," *Journal of urban Economics*, 1999, *46* (2), 254–277.

_ , *Agglomeration economics*, University of Chicago Press, 2010.

_ **, William R Kerr, and Giacomo AM Ponzetto**, "Clusters of entrepreneurship," *Journal of urban economics*, 2010, *67* (1), 150–168.

Graham, Steve and Simon Marvin, *Telecommunications and the city: Electronic spaces, urban places*, Routledge, 2002.

Gröger, André and Yanos Zylberberg, "Internal labor migration as a shock coping strategy: Evidence from a typhoon," *American Economic Journal: Applied Economics*, 2016, *8* (2), 123–53.

Haggag, Kareem, Brian McManus, and Giovanni Paci, "Learning by driving: Productivity improvements by new york city taxi drivers," *American Economic Journal: Applied Economics*, 2017, *9* (1), 70–95.

Jaffe, Adam B, Manuel Trajtenberg, and Rebecca Henderson, "Geographic localization of knowledge spillovers as evidenced by patent citations," *the Quarterly journal of Economics*, 1993, *108* (3), 577–598.

Jayachandran, Seema, "Selling labor low: Wage responses to productivity shocks in developing countries," *Journal of political Economy*, 2006, *114* (3), 538–575.

Joseph, Thomas, Yaw Nyarko, and Shing-Yi Wang, "Asymmetric information and remittances: Evidence from matched administrative data," *American Economic Journal: Applied Economics*, 2018, *10* (2), 58–100.

Journal, The Wall Street, "China's Mighty Yangtze Is Heaving From Rain and the Three Gorges Will Be Tested," *The Wall Street Journal Magazine*, July 2020.

Jr, Robert E Lucas, "Life earnings and rural-urban migration," *Journal of political economy*, 2004, *112* (S1), S29–S59.

Kerr, William R and Frederic Robert-Nicoud, "Tech clusters," *Journal of Economic Perspectives*, 2020, *34* (3), 50–76.

Kinnan, Cynthia, Shing-Yi Wang, and Yongxiang Wang, "Access to migration for rural households," *American Economic Journal: Applied Economics*, 2018, *10* (4), 79–119.

Krugman, Paul, "The new economic geography, now middle-aged," *Regional studies*, 2011, *45* (1), 1–7.

Kukalis, Sal, "Agglomeration economies and firm performance: the case of industry clusters," *Journal of Management*, 2010, *36* (2), 453–481.

Levitt, Steven D, John A List, and Chad Syverson, "Toward an understanding of learning by doing: Evidence from an automobile assembly plant," *Journal of political Economy*, 2013, *121* (4), 643–681.

Light, Ivan, Georges Sabagh, Mehdi Bozorgmehr, and Claudia Der-Martirosian, "Beyond the ethnic enclave economy," *Social Problems*, 1994, *41* (1), 65–80.

Mao, Wenzheng, Liu Ming, Ying Rong, Christopher S Tang, and Huan Zheng, "Faster deliveries and smarter order assignments for an on-demand meal delivery platform," *Available at SSRN 3469015*, 2019.

Marshall, Alfred, *Principles of economics, by Alfred Marshall*, Macmillan and Company, 1890.

Mas, Alexandre and Amanda Pallais, "Valuing alternative work arrangements," *American Economic Review*, 2017, *107* (12), 3722–59.

McKenzie, David and Hillel Rapoport, "Self-selection patterns in Mexico-US migration: the role of migration networks," *the Review of Economics and Statistics*, 2010, *92* (4), 811–821.

_ , Caroline Theoharides, and Dean Yang, "Distortions in the international migrant labor market: evidence from Filipino migration and wage responses to destination country economic shocks," *American Economic Journal: Applied Economics*, 2014, *6* (2), 49–75.

McKinsey, "Independent work is booming," 2022.

Michuda, Aleksandr, "Urban Labor Supply Responses to Adverse Weather Shocks on Rideshare Platforms." PhD dissertation, University of California, Davis 2021.

Monras, Joan, "Immigration and wage dynamics: Evidence from the mexican peso crisis," *Journal of Political Economy*, 2020, *128* (8), 3017–3089.

Moretti, Enrico, *The new geography of jobs*, Houghton Mifflin Harcourt, 2012.

Morten, Melanie, "Temporary migration and endogenous risk sharing in village india," *Journal of Political Economy*, 2019, *127* (1), 1–46.

Munshi, Kaivan, "Networks in the modern economy: Mexican migrants in the US labor market," *The Quarterly Journal of Economics*, 2003, *118* (2), 549–599.

— , "Social networks and migration," *Annual Review of Economics*, 2020, *12*, 503–524.

— **and Mark Rosenzweig**, "Networks and misallocation: Insurance, migration, and the rural-urban wage gap," *American Economic Review*, 2016, *106* (1), 46–98.

Overman, Henry G and Diego Puga, "Labor pooling as a source of agglomeration: An empirical investigation," in "Agglomeration economics," University of Chicago Press, 2010, pp. 133–150.

Papay, John P, Eric S Taylor, John H Tyler, and Mary E Laski, "Learning job skills from colleagues at work: Evidence from a field experiment using teacher performance data," *American Economic Journal: Economic Policy*, 2020, *12* (1), 359–388.

Patel, Krishna and Francis Vella, "Immigrant networks and their implications for occupational choice and wages," *Review of Economics and Statistics*, 2013, *95* (4), 1249–1277.

Peri, Giovanni, "Young workers, learning, and agglomerations," *Journal of urban Economics*, 2002, *52* (3), 582–607.

Pew, "The State of Gig Work in 2021," 2021.

Post, South China Morning, "After coronavirus, flooding hits southern China with 14 million affected," *South China Morning Post*, July 2020.

Rauch, James E, "Productivity gains from geographic concentration of human capital: evidence from the cities," *Journal of urban economics*, 1993, *34* (3), 380–400.

Roca, Jorge De La and Diego Puga, "Learning by working in big cities," *The Review of Economic Studies*, 2017, *84* (1), 106–142.

Rosenthal, Stuart S and William C Strange, "The determinants of agglomeration," *Journal of urban economics*, 2001, *50* (2), 191–229.

Schaller, Bruce, "The New York City taxicab fact book," *Schaller Consulting*, March 2006.

Sotelo, Sebastian, "Practical aspects of implementing the multinomial pml estimator," *Ann Arbor: University of Michigan, mimeo*, 2019.

Swanson, Eric T, "Risk aversion and the labor margin in dynamic equilibrium models," *American Economic Review*, 2012, *102* (4), 1663–1691.

Thompson, Peter, "Learning by doing," *Handbook of the Economics of Innovation*, 2010, *1*, 429–476.

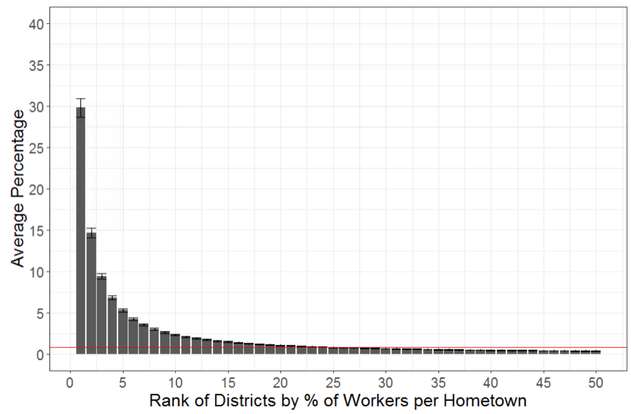
Woodruff, Christopher and Rene Zenteno, "Migration networks and microenterprises in Mexico," *Journal of development economics*, 2007, *82* (2), 509–528.

Yang, Dean, "Why do migrants return to poor countries? Evidence from Philippine migrants' responses to exchange rate shocks," *The Review of Economics and Statistics*, 2006, *88* (4), 715–735.

— , "Migrant remittances," *Journal of Economic perspectives*, 2011, *25* (3), 129–52.

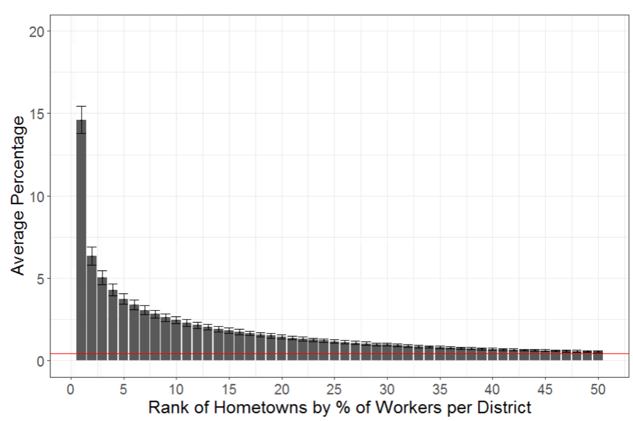
9 Figures

Figure 1: *Average Clustering Level by Sending Hometowns*



Note: This Figure shows the distribution of clustering levels by workers' hometowns. I include all active workers between 2020 and 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.1. For each hometown h, I calculate the share of workers in each district d: $P_{hd} = \frac{N_{hd}}{N_h} = \frac{N_{hd}}{\sum_{d \in \mathbf{D}} N_{hd}}$. I rank districts for each hometown based on this worker share, P_{hd}. I plot the average $P_{h,d}$ by the rank of districts for each hometown. The first bar shows that 30% of workers from the same hometown cluster in one district. The red line is the simulated average share of workers in each district if workers were randomly allocated across districts.

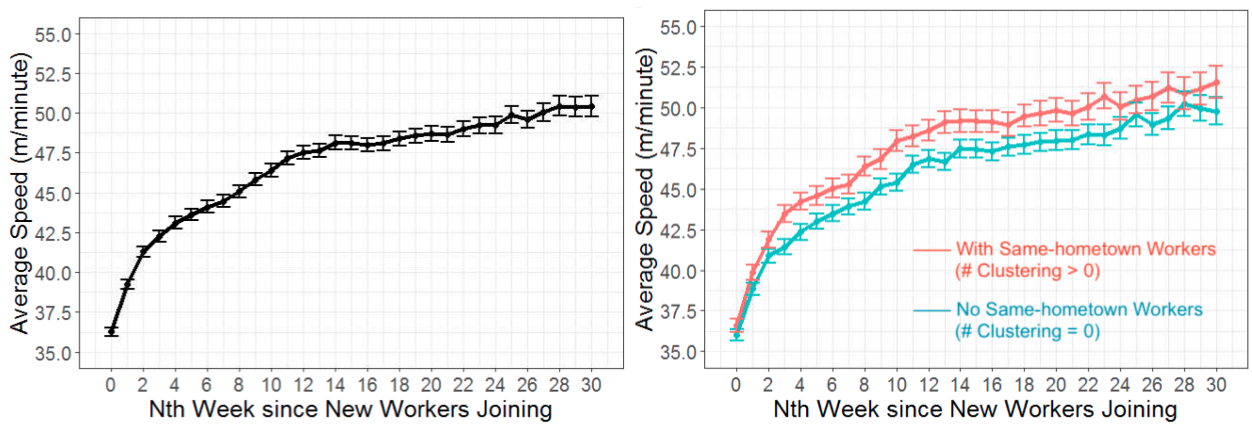
Figure 2: *Average Clustering Level by Receiving District*



Note: This Figure plots the distribution of clustering levels by receiving districts. I include all active workers between 2020 and 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.1. For each district d, I calculate the share of workers from each hometown d: $S_{hd} = \frac{N_{hd}}{N_d} = \frac{N_{hd}}{\sum_{h \in \mathbf{H}} N_{hd}}$. I rank hometowns for each district based on this share, $S_{h,d}$. I plot the average $S_{h,d}$ by the rank of hometowns for each district. The first bar shows that 15% of workers in one district come from the same hometown. The red line is the simulated average percentage of workers from each hometown if workers are randomly allocated across districts.

Figure 3: *Productivity Trend of New Workers*

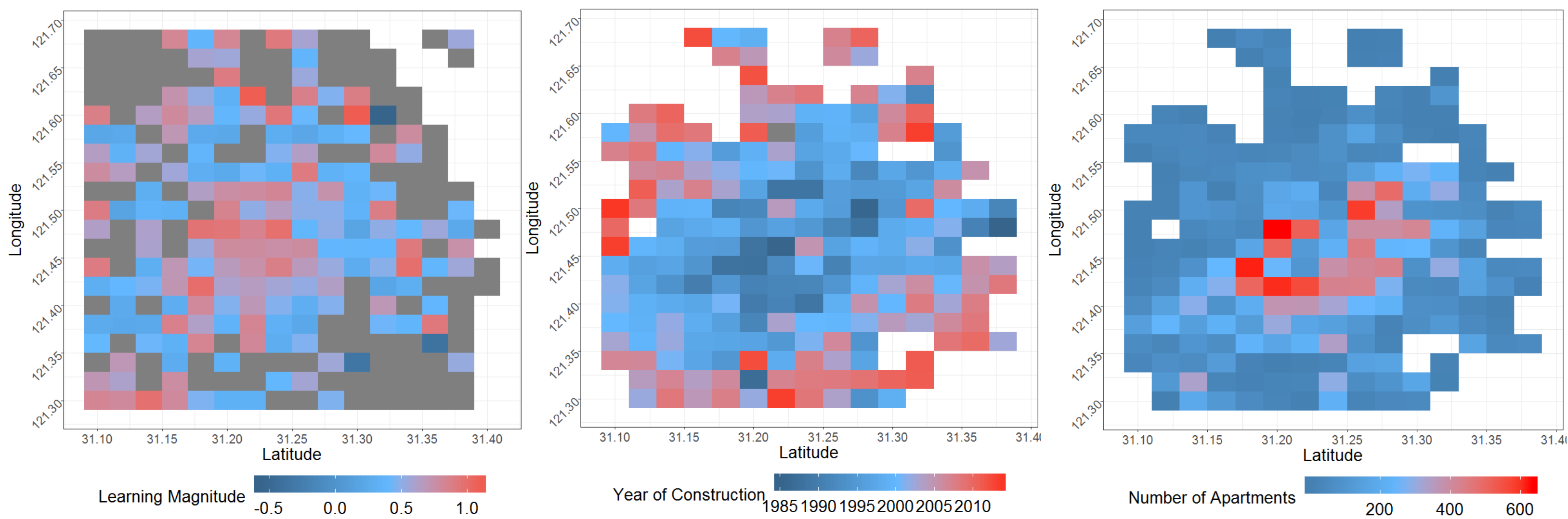
(a) *All New Workers* **(b)** *With Hometown Workers v.s. No Hometown*



Note: This Figure plots the productivity curve of new workers. I include all new active workers in 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.3. Productivity is measured by delivery speed, where I divide traveling distances (meters) by the duration (minutes). In Figure (a), I plot the productivity curve for all new workers. In Figure (b), I classify these new workers into two groups based on whether they work in a district with same-origin workers: the red line represents those with at least one same-origin worker in the same work district, and the blue line represents those without. The x-axis is the number of weeks since new workers joined the platform, and the y-axis is the average delivery speed.

Figure 4: *Heterogeneous Analysis: Learning Magnitudes across Grids in Shanghai*

(a) *Learning Magnitude* **(b)** *Average Year of Construction* **(c)** *Number of Apartments*



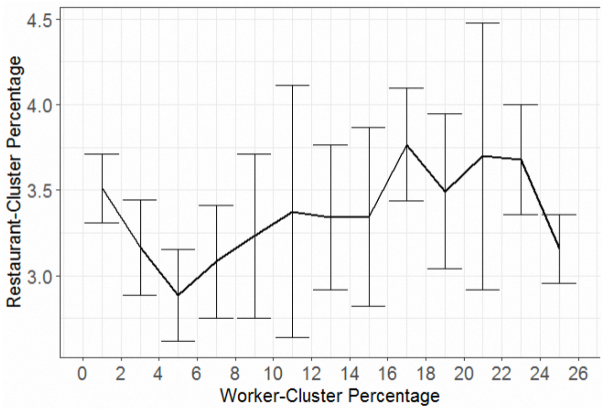
Note: This Figure plots the estimated learning magnitude in each grid and additional grid characteristics to understand where learning occurs in Shanghai. I first divide Shanghai into around 200 grids, where a grid represents a $2km \times 2km$ area. I include all new active workers between March and June 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.2. Figure (a) plots the coefficient from regressing new workers' search time for a location on an indicator of whether they have visited the exact location before or not, following the specification in section 4.2. I run these regressions for each grid separately and plot the coefficient in front of the indicator in Figure (a). I also obtain additional data from Lianjia.com on the average year of construction (figure (b)) and the number of available apartments (figure (c)) in each grid in Shanghai.

Figure 5: *The Productivity Gain as a Function of Clustering*

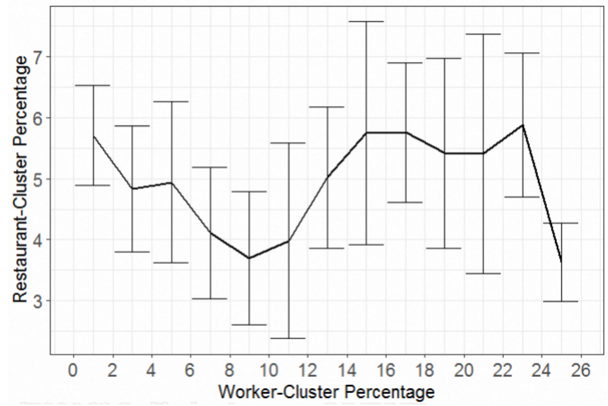
Note: This Figure plots new workers' productivity increases as a function of clustering. I include all new active workers in 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.3. I run an IV regression following the specification in section 4.3, where the dependent variable is new workers' average delivery speed, and the independent variables are a set of indicators for each different clustering level in the work districts. IVs are indicators for each different clustering level in the bonus-predicted districts. I plot the coefficients of these indicators in this Figure. The y-axis is the delivery speed (meter/minute). The x-axis represents clustering levels from 0% to 20%, which means the share of hometown workers in a district.

Figure 6: *Correlation between Clustering of Restaurant and Workers from Same Origins*

(a) *Number of Restaurant and Workers*



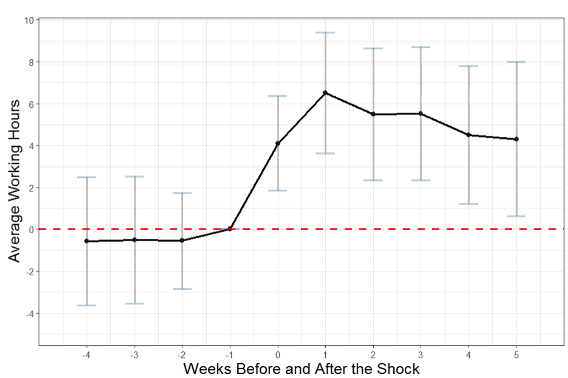
(b) *Market Share of Restaurant and Workers*



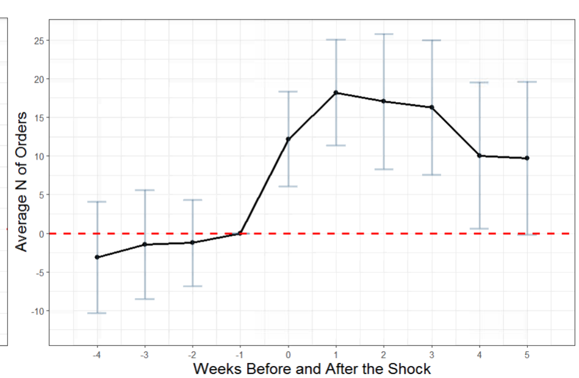
Note: This Figure plots the correlation between the share of restaurants and workers from the same origins in each district. I include all active workers between 2020 and 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.1. I first infer the hometown origins of restaurants based on their brands and flavors, such as Shanghai food, Sichuan food, and so on. In Figure (a), the x-axis is the share of hometown workers in each district. The y-axis is the share of restaurants from that same hometown in each district. In Figure (b), I plot the same correlation but construct the share of hometown workers and restaurants based on their sales or number of deliveries completed instead of absolute numbers in Figure (a).

Figure 7: *The Impact of Hometown Floods*

(a) *Labor Supply*



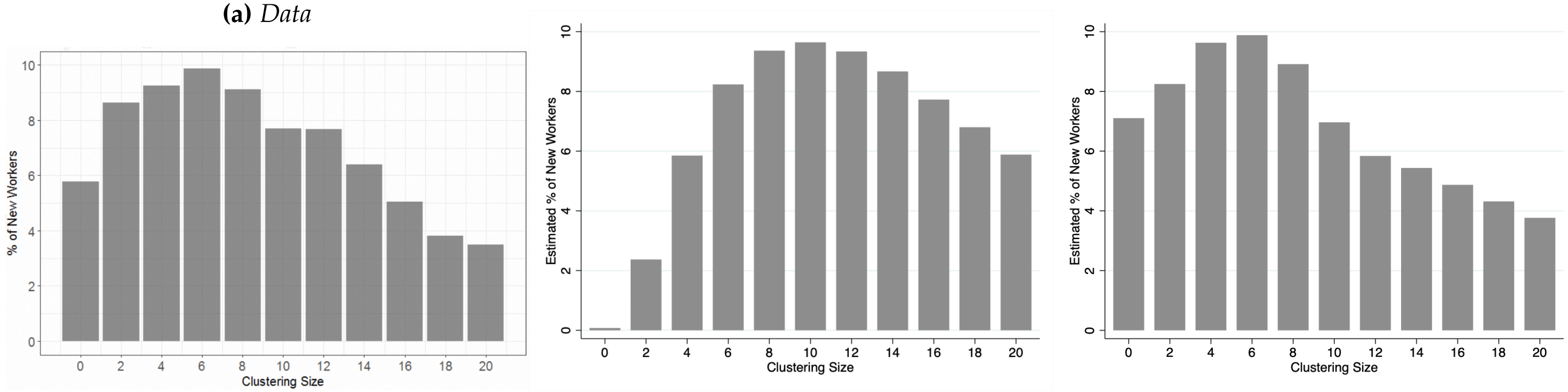
(b) *Earnings (Number of Orders)*



Note: This Figure plots the impact of hometown floods on workers' labor market performances. I include all active workers between May and August 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.4. I conduct an event study analysis. The x-axis is the number of weeks before and after hometown shocks. In Figure (a), the y-axis is the workers' weekly working hours. In Figure (b), the y-axis is the number of deliveries workers completed weekly.

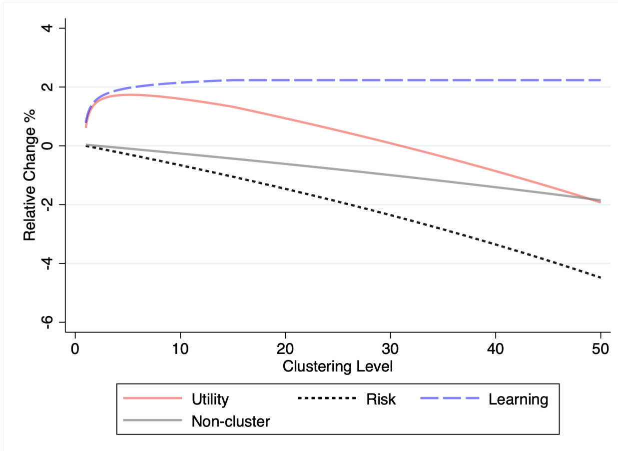
Figure 8: *Model Fit*

(a) *Data* **(b)** *Model Estimates* **(c)** *Model Estimates with Information Friction*



Note: This Figure plots the model fit. I plot new workers' entry probabilities across clustering levels. The x-axis is the clustering level from 0% to 20%, which is measured by the share of hometown workers in each district. The y-axis is the percentage of new workers entering districts at different clustering levels. In Figure (a), I compute entry probabilities from the data, which uses the same sample as Figure 1. In Figure (b), I simulate the entry probabilities from the estimated model. Figure (c) plots the simulated probabilities after adding information frictions to the model. Specifically, I assume 40% of new workers who do not have a referrer do not have any information on existing clustering levels in the city. They thus choose districts based on their idiosyncratic preferences.

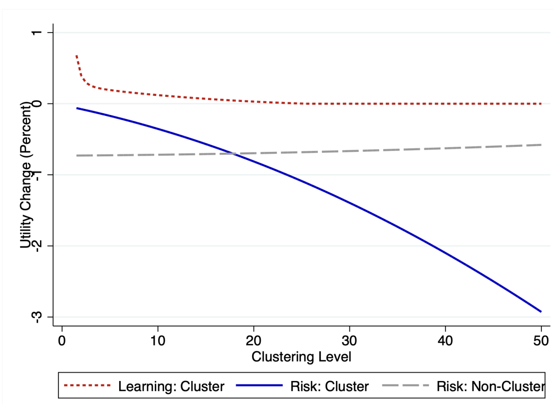
Figure 9: *Utility as a Function of Clustering*



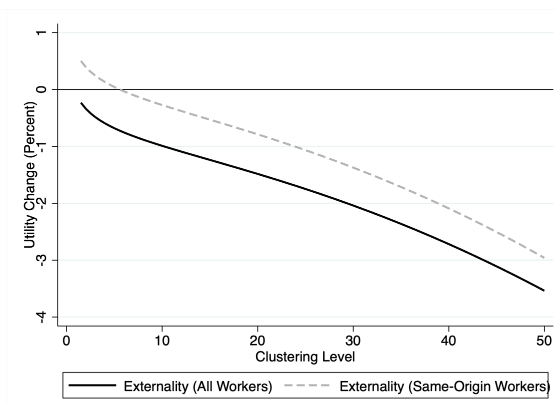
Note: This Figure plots the estimated utility with respect to clustering. The x-axis is the clustering level from 0% to 50%, meaning the share of hometown workers in each district. The y-axis is the change in utility relative to the benchmark where there is no migrant clustering. The red line represents a clustered worker's overall utility. I also decompose it into two parts: (1) the blue dashed line represents utility with only the learning benefits from clustering; (2) the black dotted line represents utility with only the congestion cost of clustering. The gray line represents the utility of non-clustered workers, who are exposed to the congestion cost of clustering.

Figure 10: *Externality of Clustering*

(a) *Decompose Externality*



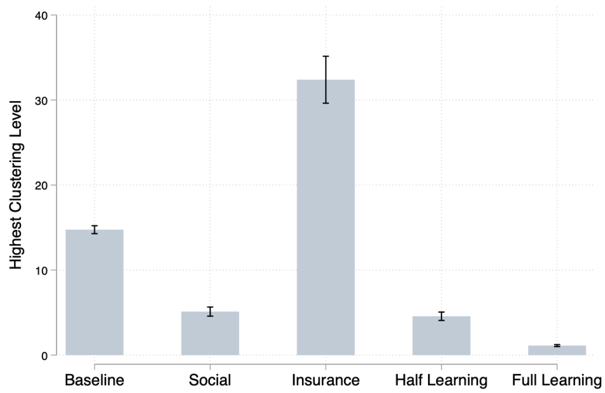
(b) *Total Externality*



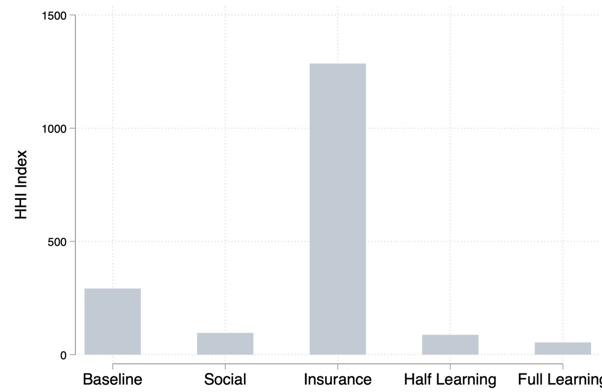
Note: This Figure plots the externality with respect to clustering. The x-axis is the clustering level from 0% to 50%, meaning the share of hometown workers in each district. The y-axis is the externality, measured by the sum of other workers' utility changes induced by a worker entry at each clustering level. In Figure (a), the red line represents the learning externality since a worker can share knowledge with other same-origin workers; (2) the dark blue line represents the congestion externality that a worker imposes on other same-origin workers; (3) the light blue line represents the congestion externality that a worker imposes on the rest of workers from other origins. I add all three externalities together, as shown by the black line in Figure (b). The gray represents the total externality of an entry for same-origin workers.

Figure 11: *Counterfactuals*

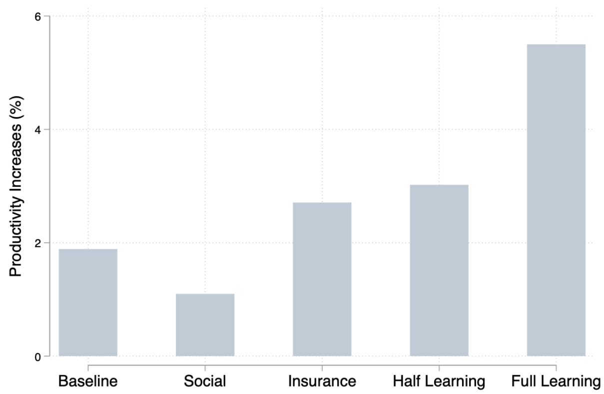
(a) *Maximum Clustering Level*



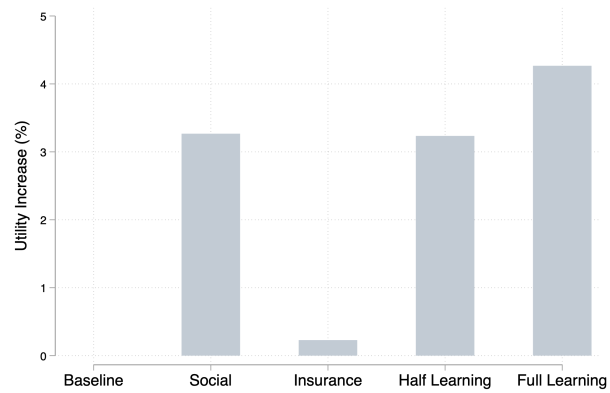
(b) *HHI of Clustering Level*



(c) *Average Productivity*



(d) *Average Exepcted Utility*



Note: I consider five scenarios: (1) "baseline" where workers choose work districts to maximize individual utility, (2) "social" where workers receive the Pigouvian tax equal to the externality of their choices when they deice where to work, (3) "insurance" where workers are provided insurance to mitigate hometown shocks, (4) "half learning" where the platform provides better technology to substitute half of network-based learning, and (5) "full learning" where the platform provides better technology to substitute the entire network-based learning. Figure (a) presents the highest clustering level across the five scenarios, figure (b) plots the HHI of clustering levels, figure (c) exhibits the average worker productivity, and figure (d) highlights the changes of worker expected utility.

10 Tables

Table 1: *Effects of Own and Peers' Experiences on Productivity*

	Restaurant Search Time (min)	Road Driving Speed (min/km)	Consumer Search Time (min)
	(1)	(2)	(3)
Own Visit to Restaurant	-1.43***	-0.08	-0.12
	(0.09)	(0.16)	(0.10)
Referrer's Visit to Restaurant	-0.51***	-0.06	-0.10
	(0.14)	(0.15)	(0.09)
Own Visit to Consumer	-0.21	-0.15	-1.25***
	(0.15)	(0.10)	(0.05)
Referrer's Visit to Consumer	-0.10	0.03	-0.58***
	(0.16)	(0.12)	(0.09)
Date X Hour X Worker FE	Y	Y	Y
Location FE	Y	Y	Y
BH Control	Y	Y	Y
Ave. Dep. Var.	6.22	10.86	5.93
Observations	5,837,304	5,837,304	5,837,304
R^2	0.43	0.39	0.44

Notes: This Table shows regressions of productivity on four indicators: (1) whether the new worker has visited the restaurant or consumer building before; (2) whether the new worker's referrer has visited the restaurant or consumer building in the last month, following the specification in section 4.2. I include all new active workers between March and June 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.2. Dependent variables in columns (1) and (3) are searching time (minutes) for the restaurant or the consumer building. I measure them by counting the number of GPS coordinates within a location's 150-meter radius. The dependent variable in column (2) is the average driving speed (minute/min) between the restaurant and the consumer building. These are proxies for workers' productivity for different parts of a delivery. Expected for fixed effects illustrated in the Table, I control for the expected probability of a worker visiting a location by randomly re-allocating orders that are put in within the same hour, grid ($2km \times 2km$), and worker tenure group (every six months), following the intuition of Borusyak and Hull (2023). Standard errors are clustered two-way at the district level and the date level.

Table 2: *Learning Magnitudes across Grids in Shanghai*

Dependent Variable: Learning Index α_d			
	(1)	(2)	(3)
Average Construction Year	-0.010**		-0.008**
	(0.004)		(0.004)
Average Apartment Density		0.081***	0.076***
		(0.023)	(0.027)
Ave. Dep. Var.	0.82	0.82	0.82
Observations	204	204	204
R^2	0.117	0.136	0.156

Notes: This Table shows regressions of the estimated learning magnitude on grid characteristics to understand where learning occurs in Shanghai. I include all new active workers between March and June 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.2. I first divide Shanghai into around 200 grids, where a grid represents a $2km \times 2km$ area. I then regress new workers' search time for a location on indicators of whether they or their referrers have visited the exact location in each grid separately, following the specification in section 4.2. The dependent variables are coefficients from these regressions. Specifically, I calculate the learning magnitudes, $\alpha_d = (\alpha_{\text{Ownvisit}} + \alpha_{\text{Refvisit}})/2$. The independent variables are scraped from Lianjia.com, including the average year of construction across all buildings in each grid and the total number of available apartments in each grid in Shanghai. Standard errors are clustered at the district level.

Table 3: *Effect of Clustering on Worker Productivity*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	OLS				First Stage	IV			
	Ave. Speed	Deliveries Per Hour	Total Income	Working Hours	Actual Clustering Level	Ave. Speed	Deliveries Per Hour	Total Income	Working Hours
Predicted Clustering Level					0.353***				
					(0.003)				
Actual Clustering Level	16.291***	1.610***	343.692***	−8.458		17.371***	1.984***	389.124***	−32.752
	(0.694)	(0.272)	(51.582)	(18.591)		(1.121)	(0.390)	(82.448)	(25.648)
District X Week FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Hometown X Week FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Entry Cohort FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
BH Control					Y	Y	Y	Y	Y
Ave. Dep. Var.	46.18	4.58	682.56	24.18	2.18	46.18	4.58	682.56	24.18
Observations	417,684	417,684	417,684	417,684	417,684	417,684	417,684	417,684	417,684
R^2	0.415	0.387	0.320	0.287	0.610				
First-Stage F					1.36e+04				

Notes: This Table shows regressions of new workers' productivity on clustering levels. I include all new active workers in 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.3. I report the OLS results in columns (1)-(4), where the independent variable is the share of hometown workers in each new worker's work district. I report first-stage in column (5) and IV results in (6)-(8). The instrumental variable is the predicted clustering level induced by entry bonuses. Dependent variables are (1) average speed, as measured by delivery distance to be divided by delivery duration, (2) average number of deliveries completed per hour, (3) total weekly earnings, and (4) total weekly working hours. All dependent variables are calculated at the worker-week level. Standard errors are clustered two-way at the district level and the week level.

Table 4: *Delivery Workers' Labor Supply Responses to Hometown Shocks*

Dep. Var.: Average Weekly Outcomes			
	$\mathbb{1}\{$ Active $\}$	Ave. Deliveries	Ave. Hours
$\mathbb{1}\{$Hometown Shock $\}$	0.057***	16.381***	5.931***
	(0.010)	(2.582)	(1.069)
Province × Week FE	Y	Y	Y
Worker FE	Y	Y	Y
District FE	Y	Y	Y
Ave. Dep. Var.	0.76	78.10	22.08
Observations	638,810	638,810	638,810
R^2	0.521	0.530	0.557

Notes: This Table shows regressions of workers' weekly labor supply on hometown shocks. I include all active workers between May and August 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.4. The dependent variables are workers' weekly performances on the platform: (1) being active (# of deliveries > 0),(2) number of deliveries completed, and (3) number of working hours. The independent variable indicates whether a worker's hometown has had flood shock in the last month. I include origin province × week fixed effect, district fixed effect, and worker fixed effects. Standard errors are clustered two-way at the district level and the week level.

Table 5: *District-level Response to Aggregate Shocks*

Dep. Var.: Weekly Outcomes at District Level				
	Number of Workers	Number of Orders	Total Working Hours	Per-delivery Commission Fee
$\mathbb{1}$\{Predicted Shock Share > 15%\}	4.925	95.720	148.502***	-0.226
	(3.259)	(104.525)	(33.291)	(0.309)
Week FE	Y	Y	Y	Y
District FE	Y	Y	Y	Y
Ave. Dep. Var	132.59	7,233.97	2,408.26	6.41
Observations	14,391	14,391	14,391	14,391
R^2	0.884	0.886	0.886	0.903

Notes: This Table shows regressions of district-level market performances on aggregate hometown shocks. I include all active workers between May and August 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.4. Dependent variables are (1) the total number of active workers in the district each week,(2) the total number of deliveries completed, (3) the total number of working hours, and (4) the average per-delivery commission fee. The independent variable indicates whether the predicted shock share is higher than 15%. The shock share is the share of workers experiencing floods, given the hometown composition in the district in May 2020. I include week and district fixed effects. Standard errors are clustered at the week level.

Table 6: *Congestion: Impacts of Clustering on Delivery Workers' Real Wages during Shocks*

Dep. Var.: Average Weekly Outcomes			
	$\mathbb{1}\{$ Active $\}$	Ave. Deliveries	Ave. Hours
$\mathbb{1}\{$ Hometown Shock$\}$	0.072^{***}	19.939^{***}	6.112^{***}
	(0.014)	(3.912)	(1.108)
$\mathbb{1}\{$District Shock Share > 15%$\}$	-0.003	-4.870	-0.648
	(0.022)	(2.997)	(0.455)
Interaction Term	-0.009	-8.997^{***}	-0.302
	(0.010)	(2.080)	(0.378)
Province × Week FE	Y	Y	Y
Worker FE	Y	Y	Y
District FE	Y	Y	Y
Ave. Dep. Var.	0.76	78.10	22.08
Observations	638,810	638,810	638,810
R^2	0.542	0.573	0.582

Notes: This Table shows regressions of workers' weekly labor supply on hometown shocks and district shocks. I include all active workers between May and August 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.4. Dependent variables are workers' weekly performances on the platform: (1) being active (# of deliveries > 0),(2) number of deliveries completed, and (3) number of working hours. Independent variables are (1) the indicator of whether a worker's hometown has any flood shock in the last month, (2) an indicator of district shock based on whether the predicted shock share in the district is higher than 15%, and (3) their interaction term. I include origin province × week fixed effects, worker fixed effects, and district fixed effects. Standard errors are clustered two-way at the district level and the week level.

Table 7: *Estimate $\alpha + \beta$: Labor Supply to Lottery Prizes*

Dep. Var.: Weekly Working Hours			
	(1)	(2)	(3)
Lottery/Average Hour Wage	-0.57***	-0.63***	-0.61***
	(0.01)	(0.05)	(0.07)
Week FE	Y	Y	
Worker FE		Y	Y
Week X District FE			Y
Week X Hometown FE			Y
Observations	48,126	48,126	48,126
R^2	0.18	0.41	0.54

Notes: This Table shows regressions of workers' labor supply on the amount of prizes they won from the lotteries. The sample includes three lotteries offered by the platform between 2020 and 2021. I include delivery workers who completed at least 50 deliveries within seven days before the lottery, and worked in a district with at least one worker winning the lottery. Dependent variables are workers' weekly working hours. The Independent variable is the prize amount each worker received. Columns (1)-(3) include different fixed effects. Standard errors are clustered two-way at the district level and the week level.

Table 8: *Gravity Regression*

Dep. Var.: $\mathbb{1}\{$ Worker i from Hometown h Choose District d $\}$				
	(1)	(2)	(3)	(4)
$\triangle U(B_{dt})$	2.95***	2.49***	2.53***	2.18***
(%)	(0.34)	(0.43)	(0.47)	(0.38)
Month FE		Y		Y
Hometown FE		Y		
District FE		Y	Y	
Month X Hometown FE			Y	
District X Hometown FE				Y
Observations	241,190	241,190	241,190	241,190

Notes: This Table shows gravity regressions of new workers' district choices on entry bonuses. I include all new active workers in 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.3. The dependent variable is the share of new workers from each hometown joining different districts across weeks. The Independent variable is the utility change caused by the entry bonus in each district across weeks. Columns (1)-(4) include different fixed effects.

Table 9: *Parameter Estimates*

	Parameter	Estimate	Conf. Interval
Labor Supply Elasticity	$\alpha + \beta$	0.39	[0.26,0.52]
High Hometown Incom	S^{high}	279.61	[256.21, 340.88]
Low Hometown Income	S^{low}	55.24	[4.83, 99.35]
Risk Aversion Coefficient	γ	2.31	[1.87, 3.30]
Scale of Idiosyncratic Preferences	σ	2.18	[1.43, 2.93]

Notes: This Table reports the estimated parameters. Coefficients, $\alpha + \beta$ and σ are estimated by regressions as specified in section 6. Coefficnets, S^{high}, S^{low}, and γ, are estimated by two-step GMM. The last column reports the 95% confidence interval estimated via either regression results or a 200-round bootstrap.

A Appendix: Proofs and Derivations

A.1 Worker Utility Maximization Problem

For worker i who come from hometown h and work in district d:

$$\begin{aligned} \max_{L,T} \quad & (C_i^{\text{own}})^{\alpha}(C_i^{\text{family}})^{\beta}(\bar{L}-L_i)^{1-\alpha-\beta} \\ \text{s.t.} \quad & C_i^{\text{own}} + T_i = w_{hd}L_i \\ & C_i^{\text{family}} = S_h + T_i \end{aligned}$$

Take F.O.C for L and T

$$L_i = \frac{\alpha w\bar{L} + (1-\beta)T_i}{(1-\alpha-\beta)w_{hd}}$$

$$T_i = \frac{\beta w_{hd}L_i - \alpha S_h}{\alpha+\beta}$$

Plug T in the first F.O.C to solve for L:

$$L_i = (\alpha+\beta)\bar{L} - (1-\alpha-\beta)\frac{S_h}{w_{hd}}$$

From the delivery market clearing condition

$$w_{hd} = \frac{AD_d\delta(N_{hd})}{\sum_{i'}\delta(N_{i'hj})L_{i'}}$$

Labor supply and wage in the equilibrium are

$$\begin{cases} w_{hd}^* = \frac{1}{\alpha+\beta}\frac{\delta(N_{hd})}{\sum_{i'}\delta(N_{i'hj})}\frac{AD_d+(1-\alpha-\beta)(\sum_{i'}\delta(N_{i'hj})S_{i'hj})/\delta(N_{ihd})}{\bar{L}} \\ \\ \frac{L_i^*}{\bar{L}} = (\alpha+\beta)\left(1-\frac{\sum_{o\in j}\delta(N_{ohj})}{\delta(N_{ihd})}\frac{(1-\alpha-\beta)S_{hd}}{AD_d+(1-\alpha-\beta)(\sum_{i'}\delta(N_{i'hj})S_{i'hj})/\delta(N_{ihd})}\right) \end{cases}$$

Take these solutions into the utility:

$$u_{ihd} = \alpha^{\alpha}(1-\alpha)^{\beta}(\alpha+\beta)^{\beta}\left(\frac{1-\alpha-\beta}{w_{hd}^*}\right)^{1-\alpha-\beta}(w_{hd}^*\bar{L}+S_h)$$

Set $\alpha^\alpha(1-\alpha)^\beta(\alpha+\beta)^\beta(1-\alpha-\beta)^1-\alpha-\beta\bar{L} = Cons$

$$u_{ihd} = Cons \times \left({w^*_{hd}}^{\alpha+\beta} + \frac{S_h}{\bar{L}}(w^*_{hd})^{\alpha+\beta-1} \right)$$

A.2 Proposition 1

Take the derivative of u_{ihd} with respect to w^*_{hd}

$$\frac{\partial u_{ihd}}{\partial w^*_{hd}} = Const \times (w^*_{hd})^{\alpha+\beta-2}\left((\alpha+\beta)w^*_{hd} - (1-\alpha-\beta)\frac{S_h}{\bar{L}} \right)$$

Since $L_i = (\alpha+\beta)\bar{L} - (1-\alpha-\beta)\frac{S_h}{w_{hd}} > 0$, $\frac{\partial u_{ihd}}{\partial w^*_{hd}} > 0$. The core is to derive $\frac{\partial w^*_{hd}}{\partial N_{hd}}$

$$\frac{\partial w^*_{hd}}{\partial N_{hd}} = \frac{\partial}{\partial N_{hd}}\left(\delta(N_{hd}) \times (AD_d + (1-\alpha-\beta)\sum_{i'} S_{i'hj}) \right)\left(\sum_{i'} \delta(N_{i'hj}) \right)^{-1}$$

Assume that there are N_d workers in district j in total. Consider the case where there exists only one cluster of workers from the same hometown, h. Without loss of generality,

$$\delta(N_{hd}) \in [\delta(0), \delta(N_d)]$$

$$\sum_{i'} \delta(N_{i'hj}) = N_{hd}\delta(N_{hd}) + (N_d - N_{hd})$$

$$\textstyle\sum_{i'} S_{i'hj} \quad = N_{hd}S_h + (N_d - N_{hd})\bar{S}$$

$$\text{where } \bar{S} = P_{\text{shock}}S^{\text{low}} + (1-P_{\text{shock}})S^{\text{high}}$$

Given that $S^{\text{low}} < \bar{S} < S^{\text{high}}$, When $S_h = S^{\text{high}}$,

$$\begin{aligned} w^{high}_{hd} &= \delta(N_{hd}) \times \frac{AD_d + (1-\alpha-\beta)(\sum_{i'} S_{i'hj})}{(\alpha+\beta)\bar{L}\sum_{i'}\delta(N_{i'hj})} \\ &= \delta(N_{hd}) \times \frac{AD_d + (1-\alpha-\beta)(N_{hd}(S^{\text{high}} - \bar{S}) + N_d\bar{S})}{(\alpha+\beta)\bar{L}(N_{hd}(\delta(N_{hd}) - 1) + N_d)} \end{aligned}$$

It is clear that the numerator increases with N_{hd}, while the denominator also increases with N_{hd}.

$$
\begin{aligned}
w_{hd}^{high} &= \frac{\delta(N_{hd})}{(\alpha+\beta)\bar{L}} \times \frac{AD_d + (1-\alpha-\beta)(S^{\text{high}} - \bar{S})N_{hd} + (1-\alpha-\beta)N_d\bar{S}}{N_{hd}(\delta(N_{hd})-1)+N_d} \\
&\text{Set } AD_d = K_1 \text{ , } (1-\alpha-\beta)(S^{\text{high}}-\bar{S}) = K_2 \text{ , } (1-\alpha-\beta)N_d\bar{S} = K_3 \text{ and } N_d = K_4 \\
&\sim \delta(N_{hd})\frac{K_1 + K_2 N_{hd} + K_3}{N_{hd}(\delta(N_{hd}-1)) + K_4}
\end{aligned}
$$

$$
\frac{\partial w_{hd}^{high}}{\partial N_{hd}} \sim K_2K_4\delta(N_{hd}) + (K_1 + K_2N + K_3)\delta'(N_{hd})(K_4 - N) - (K_1 + K3)\delta(N_{hd})(\delta(N_{hd}) - 1)
$$

$$
\begin{aligned}
\frac{\partial^2 w_{hd}^{high}}{\partial^2 N_{hd}} = &\Big(K_2K_4 + K_2(K_4 - N_{hd}) - K_2N - 2(K_1+K_3)\delta(N_{hd})\Big)\delta'(N_{hd}) \\
&+ (K_1 + K_2N + K_3)(K_4 - N_{hd})\delta''(N_{hd})
\end{aligned}
$$

Since $\delta''(N_{hd}) < 0$ and $K_2K_4 + K_2(K_4 - N_{hd}) < K_2N + 2(K_1+K_3)\delta(N_{hd})$, we have $\partial^2 w_{hd}^{high}/\partial^2 N_{hd} < 0$. The sufficient condition to ensure$\partial w_{hd}^{high}/\partial N_{hd} > 0$ is

$$
\begin{aligned}
&\frac{\partial w_{hd}^{high}}{\partial N_{hd}}(N_{hd} = N_d) > 0 \\
\rightarrow &K_2K_4\delta(N_{hd}) + (K_1 + K_2N + K_3)\delta'(N_{hd})(K_4 - N) > (K_1 + K3)\delta(N_{hd})(\delta(N_{hd}) - 1)
\end{aligned}
$$

Since $\delta'(N_{hd})$>0, I further relax the condition to be

$$
\begin{aligned}
\rightarrow &K_2K_4\delta(N_{hd}) > (K_1 + K3)\delta(N_{hd})(\delta(N_{hd}) - 1) \\
\rightarrow &\delta(N_{hd}) < \frac{K_2K_4}{K_1+K_3} + 1
\end{aligned}
$$

On the other hand, when $S_{hd} = S^{\text{low}}$,

$$
w_{hd}^{low} = \frac{AD_d\delta(N_{hd}) + (1-\alpha-\beta)(-N_{hd}(\bar{S} - S^{\text{low}}) + N_d\bar{S})}{(\alpha+\beta)\bar{L}(N_{hd}\delta(N_{hd}) + (N_d - N_{hd}))}
$$

Given the concavity of $\delta(N_{hd})$, the numerator first increase and then decrease with N_{hd}, while the denominator always increases with N_{hd}. I consider the upper bound of the function,

$$
\begin{aligned}
w_{hd}^{low-upperlimit} &= \frac{AD_d\delta(N_{hd}) + (1-\alpha-\beta)(-N_{hd}(\bar{S}-S^{\text{low}}) + N_d\bar{S})}{(\alpha+\beta)\bar{L}N_d} \\
&= \frac{AD_d}{(\alpha+\beta)\bar{L}N_d}\delta(N_{hd}) - \frac{(1-\alpha-\beta)(\bar{S}-S^{\text{low}})}{(\alpha+\beta)\bar{L}N_d}N_{hd} + \frac{(1-\alpha-\beta)N_d\bar{S}}{(\alpha+\beta)\bar{L}N_d}
\end{aligned}
$$

$$
\frac{\partial w_{hd}^{low-upperlimit}}{\partial N_{hd}} = \frac{AD_d}{(\alpha+\beta)\bar{L}N_d}\delta'(N_{hd}) - \frac{(1-\alpha-\beta)(\bar{S}-S^{\text{low}})}{(\alpha+\beta)\bar{L}N_d}
$$

Since$\partial^2 w_{hd}^{low-upperlimit}/\partial^2 N_{hd} < 0$, the sufficient condition for w_{hd}^{low} decrease with N_{hd} when N_{hd} is large is

$$
\begin{aligned}
&\frac{\partial w_{hd}^{low-upperlimit}}{\partial N_{hd}}(N_{hd} = N_d) < 0 \\
\rightarrow &\delta'(N_d) < \frac{1-\alpha-\beta}{AD_d}(\bar{S}-S^{\text{low}})
\end{aligned}
$$

Addition notes: Endogeneity of N_d.

For this proof, N_d is independent of N_{hd}. However, when workers have complete information on the hometown composition $\{N_{hd}\}$, N_d may be a function of N_{hd} because of the negative externality of clustering on other workers. In addition, the degree to which N_{hd} influences N_d depends on the labor market clearing conditions. I discuss two cases here.

Case 1: non-clustering workers enter the districts until the real wage without aggregate shocks equalizes across districts.

$$
\begin{aligned}
w^c &= \frac{AD_d + (1-\alpha-\beta)(N_{hd}(S^{\text{high}}-\bar{S}) + N_d\bar{S})}{(\alpha+\beta)\bar{L}(N_{hd}(\delta(N_{hd})-1) + N_d)} \\
N_d &= \frac{AD_d + (1-\alpha-\beta)(S^{\text{high}}-\bar{S})N_{hd} - (\alpha+\beta)\bar{L}w^c N_{hd}(\delta(N_{hd})-1)}{(\alpha+\beta)\bar{L}w^c - (1-\alpha-\beta)\bar{S}}
\end{aligned}
$$

$$
\frac{\partial N_d}{\partial N_{hd}} \sim (1-\alpha-\beta)(S^{\text{high}}-\bar{S}) - (\alpha+\beta)\bar{L}w^c(N_{hd}\delta'(N_{hd}) + \delta(N_{hd}) - 1)
$$

When $(\alpha+\beta)\bar{L}w^c(N_d\delta'(N_d) + \delta(N_d) - 1) < (1-\alpha-\beta)(S^{\text{high}}-\bar{S})$, N_d increases with N_{hd}, which makes the competition effect stronger.

Case 2: non-clustering workers enter the districts until the real wages with and without aggregate shocks both equalize across districts. This condition assumes that the non-clustering workers are fully flexible. Their entry and exit decisions are made after the hometown shocks are realized. In this case, we have

$$w^c = \frac{AD_d + (1-\alpha-\beta)(N_{hd}(S^{\text{high}} - \bar{S}) + N_d^{noshock}\bar{S})}{(\alpha+\beta)\bar{L}(N_{hd}(\delta(N_{hd}) - 1) + N_d^{noshock})} \text{and}$$
$$w^c = \frac{AD_d + (1-\alpha-\beta)(N_{hd}(S^{\text{high}} - \bar{S}) + N_d^{withshock}\bar{S})}{(\alpha+\beta)\bar{L}(N_{hd}(\delta(N_{hd}) - 1) + N_d^{withshock})}$$

This means there is no congestion effect since the changes in the labor supply of clustered workers are balanced by the entry and exit of non-clustering workers. With only productivity gain from clustering, workers' utility increases with the clustering level.

Considering these two cases, the extent to which N_{hd} influences N_d becomes more of an empirical question. From the empirical analysis, I do not find significant effects of clustering level on the total number of active workers in the districts, both with and without shocks.

A.3 Proposition 2

$$\mathbb{E}U_{ihd} = \frac{1}{1-\gamma}\left((1-P_{\text{shock}})(u_{ihd}^{\text{noshock}})^{1-\gamma} + P_{\text{shock}}(u_{ihd}^{\text{shock}})^{1-\gamma}\right)$$

$$\begin{aligned}
\frac{\partial \mathbb{E}U_{ihd}}{\partial N_{hd}} &= (1-P_{\text{shock}})(u_{ihd}^{\text{noshock}})^{-\gamma}\frac{\partial u_{ihd}^{\text{noshock}}}{\partial w_{hd}^{high}}\frac{\partial w_{hd}^{high}}{\partial N_{hd}} + P_{\text{shock}}(u_{ihd}^{\text{shock}})^{-\gamma}\frac{\partial u_{ihd}^{\text{shock}}}{\partial w_{hd}^{low}}\frac{\partial w_{hd}^{low}}{\partial N_{hd}} \\
&= (1-P_{\text{shock}})(u_{ihd}^{\text{noshock}})^{-\gamma}Const(w_{hd}^{high})^{\alpha+\beta-2}\left((\alpha+\beta)w_{hd}^{high} - (1-\alpha-\beta)\frac{S_h}{\bar{L}}\right)\frac{\partial w_{hd}^{high}}{\partial N_{hd}} \\
&+ P_{\text{shock}}(u_{ihd}^{\text{shock}})^{-\gamma}Const(w_{hd}^{low})^{\alpha+\beta-2}\left((\alpha+\beta)w_{hd}^{low} - (1-\alpha-\beta)\frac{S_h}{\bar{L}}\right)\frac{\partial w_{hd}^{low}}{\partial N_{hd}}
\end{aligned}$$

Set $O^{high} = (u_{ihd}^{\text{noshock}})^{-\gamma}Const(w_{hd}^{high})^{\alpha+\beta-2}\left((\alpha+\beta)w_{hd}^{high} - (1-\alpha-\beta)\frac{S_h}{\bar{L}}\right)$
and $O^{low} = (u_{ihd}^{\text{shock}})^{-\gamma}Const(w_{hd}^{low})^{\alpha+\beta-2}\left((\alpha+\beta)w_{hd}^{low} - (1-\alpha-\beta)\frac{S_h}{\bar{L}}\right)$
The sufficient condition is

$$\begin{aligned}
&\frac{\partial \mathbb{E}U_{ihd}}{\partial N_{hd}}(N_{hd} = N_d) < 0 \\
\rightarrow &(1-P_{\text{shock}})O^{high}\frac{\partial w_{hd}^{high}}{\partial N_{hd}} < (1-P_{\text{shock}})O^{high}\frac{\partial w_{hd}^{high-upperlimit}}{\partial N_{hd}} \\
&< -P_{\text{shock}}O^{low}\frac{\partial w_{hd}^{low-upperlimit}}{\partial N_{hd}} < -P_{\text{shock}}O^{low}\frac{\partial w_{hd}^{low}}{\partial N_{hd}} \\
\rightarrow &\delta'(M) < \frac{1-\alpha-\beta}{AD_d((1-P_{\text{shock}})O^{high} + P_{\text{shock}}O^{low})} \\
&\times \left(\bar{S}((1-P_{\text{shock}})O^{high} + P_{\text{shock}}O^{low}) - (1-P_{\text{shock}})O^{high}\delta(N_d)S^{\text{high}} + P_{\text{shock}}O^{low}\delta(N_d)S^{\text{low}}\right)
\end{aligned}$$

A.4 Proposition 3

Following the definition of externality in section 3,

$$\begin{aligned}
\Phi(N_{hd}) &= \Phi^{\text{learn}}(N_{hd}) + \Phi^{\text{cost}}(N_{hd}) \\
&= \triangle\mathbb{E}U^{\text{learn}}(N_{hd}) \times N_{hd} + \triangle\mathbb{E}U^{\text{cost}}(N_{hd}) \times N_{hd} + \triangle\mathbb{E}U^{\text{other}}(N_{hd}) \times (N_d - N_{hd})
\end{aligned}$$

where the three utility functions are defined as below.

- $\mathbb{E}U^{\text{learn}}(N_{hd}) = \frac{1}{1-\gamma}\Big((1-P_{\text{shock}})(u(N_{hd}, S^{\text{high}}))^{1-\gamma} + P_{\text{shock}}(u(0, S^{\text{low}}))^{1-\gamma}\Big)$
- $\mathbb{E}U^{\text{cost}}(N_{hd}) = \frac{1}{1-\gamma}\Big((1-P_{\text{shock}})(u(0, S^{\text{high}}))^{1-\gamma} + P_{\text{shock}}(u(N_{hd}, S^{\text{low}}))^{1-\gamma}\Big)$
- $\mathbb{E}U^{\text{other}}(N_{hd}) = \frac{1}{1-\gamma}\Big((1-P_{\text{shock}})^2(u(0, S^{\text{high}}))^{1-\gamma} + (1-P_{\text{shock}})P_{\text{shock}}(u(N_{hd}, S^{\text{high}}))^{1-\gamma}$
  $+ P_{\text{shock}}(1-P_{\text{shock}})(u(0, S^{\text{low}}))^{1-\gamma} + P_{\text{shock}}^2(u(N_{hd}, S^{\text{low}}))^{1-\gamma}\Big)$

Following the calculation of the utility function as above, these three parts of externality are as follows:

$$\frac{\partial \mathbb{E}U^{\text{learn}}(N_{hd})}{N_{hd}} N_{hd} \simeq C_1 \times (u(N_{hd}, S^{\text{high}}))^{-\gamma} \frac{\partial u}{w} \delta'(N_{hd}) N_{hd} > 0$$

$$\frac{\partial \mathbb{E}U^{\text{cost}}(N_{hd})}{N_{hd}} N_{hd} \simeq C_2 \times (u(N_{hd}, S^{\text{low}}))^{-\gamma} \frac{\partial u}{w} \Big(\delta(N_{hd}) S^{\text{low}} - S^{\text{high}}\Big) N_{hd} < 0$$

$$\frac{\partial \mathbb{E}U^{\text{other}}(N_{hd})}{N_{hd}} (N_d - N_{hd}) \simeq C_3 \times \Big(P_{\text{shock}}(u(N_{hd}, S^{\text{low}}))^{-\gamma} + (1-P_{\text{shock}})(u(N_{hd}, S^{\text{high}}))^{-\gamma}\Big) \frac{\partial u}{w} \times \Big(\delta(N_{hd}) S^{\text{low}} - S^{\text{high}}\Big)(N_d - N_{hd}) < 0$$

where C_1, C_2, and C_3 are constant terms containing the constant terms from $u = Constant \times \left({w_{hd}^*}^{\alpha+\beta} + \frac{S_h}{\bar{L}}(w_{hd}^*)^{\alpha+\beta-1}\right)$. Rearrange these three equations; the total externality is as follows.

$$\Phi(N_{hd}) \simeq A_1 \delta'(N_{hd}) N_{hd} + A_2 N_{hd} + A_3(N_d - N_{hd})$$

where $A_1 > 0$, $A_2 < A_3 < 0$. Thus, $\Phi(N_{hd} = 0) = A_3 N_d < 0$

$$\frac{\partial \Phi(N_{hd})}{N_{hd}} \simeq A_1 \delta'(N_{hd}) + A_1 \delta''(N_{hd}) N_{hd} + (A_2 - A_3)$$

With the condition $\max(\delta'(N_{hd}) + \delta''(N_{hd}) N_{hd}) < \delta'(N_{hd} = 0) < \frac{A_3 - A_2}{A_1}$, we have $\Phi(N_{hd}) < 0$ and $\Phi(N_{hd})$ decreases with N_{hd}.

A.5 Use hometown heterogeneity to identify risk aversion

For worker i from hometown h, the utility of working in district d at time t:

$$\tilde{U}(c) = \mathbb{E}U(c) + \beta(c)$$

where$c$ is the clustering level. $\mathbb{E}U(c)$ is the expected utility derived from the Cobb-Douglas utility maximization, $\mathbb{E}U(c) = \frac{1}{1-\gamma}\Bigg((1-P_{\text{shock}})(u^{\text{good}}(c))^{1-\gamma} + P_{\text{shock}}(u^{\text{bad}}(c))^{1-\gamma}\Bigg)$. I further relax assumptions by allowing workers to hold unobserved utility with respect to the clustering level, $\beta(c)$. The restriction is that $\beta(c)$ depends only on the clustering level but not on workers' risk aversion coefficients.

Consider workers from two types of hometowns. The probability of shocks is lower in one hometown, P^{safe}. Conversely, the shock probability is high in the other hometown, P^{risky}.

At the clustering level, c, the difference in workers' utility from the two hometowns is below.

$$\begin{aligned}
\triangle\tilde{U}(c) = \tilde{U}^{safe}(c) - \tilde{U}^{risky}(c) &= (\mathbb{E}U^{safe}(c) + \beta(c)) - (\mathbb{E}U^{risky}(c) + \beta(c)) \\
&= \frac{(P^{risky} - P^{safe})}{1-\gamma}((u^{\text{good}}(c))^{1-\gamma} - (u^{\text{bad}}(c))^{1-\gamma}) \\
&= \frac{(P^{risky} - P^{safe})}{1-\gamma}\triangle u^{*}(c)
\end{aligned}$$

This difference in the utility, $\triangle u^{*}(c)$, depends on (1) u^{good} and u^{bad}, which are estimated from the Cobb-Douglas utility maximization, and (2) the worker's risk aversion coefficient, γ. I then estimate γ from new workers' choices of work districts. $\pi^{risky}(c)$ and $\pi^{safe}(c)$ represent the fraction of new workers, either from a risky hometown or a safe hometown, choosing to enter a district with a clustering level c. From the gravity model, there is a mapping between the fraction and workers' utility.

$$\pi^{risky}(c) = \frac{N_c^{risky}}{\sum_{c'} N_{c'}^{risky}} = \frac{\exp(\sigma\tilde{U}^{risky}(c))}{\sum_{c'}\exp(\sigma\tilde{U}^{risky}(c'))}$$

$$\pi^{safe}(c) = \frac{N_c^{safe}}{\sum_{c'} N_{c'}^{safe}} = \frac{\exp(\sigma\tilde{U}^{safe}(c))}{\sum_{c'}\exp(\sigma\tilde{U}^{safe}(c'))}$$

$$\frac{\pi^{risky}(c)}{\pi^{safe}(c)} = \frac{\exp(\sigma \tilde{U}^{risky}(c))}{\exp(\sigma \tilde{U}^{safe}(c))} \times \frac{\sum_{c'} \exp(\sigma \tilde{U}^{risky}(c'))}{\sum_{c'} \exp(\sigma \tilde{U}^{safe}(c'))}$$
$$= \exp\left(\sigma(\tilde{U}^{risky}(c) - \tilde{U}^{safe}(c))\right) \times \frac{\sum_{c'} \exp(\sigma \tilde{U}^{risky}(c'))}{\sum_{c'} \exp(\sigma \tilde{U}^{safe}(c'))}$$

Set $\frac{\sum_{c'} \exp(\sigma \tilde{U}^{risky}(c'))}{\sum_{c'} \exp(\sigma \tilde{U}^{safe}(c'))} = A$

$$\triangle \tilde{U}(c) = \tilde{U}^{risky}(c) - \tilde{U}^{safe}(c) = \frac{1}{\sigma}(log(\pi^{risky}(c)) - log(\pi^{peack}(c)) - log(A))$$
$$= \frac{1}{\sigma}(\triangle log(\pi)(c) - log(A))$$

Combining all equations, new workers' entry decisions can reflect the utility difference across the two types of hometowns.

$$\frac{\triangle \tilde{U}(c_1) - \triangle \tilde{U}(c_2)}{\triangle \tilde{U}(c_2) - \triangle \tilde{U}(c_3)} = \frac{\triangle log(\pi)(c_1) - \triangle log(\pi)(c_2)}{\triangle log(\pi)(c_2) - \triangle log(\pi)(c_3)} = \frac{\triangle u^*(c_1) - \triangle u^*(c_2)}{\triangle u^*(c_2) - \triangle u^*(c_3)}$$

B Appendix: Figures and Tables

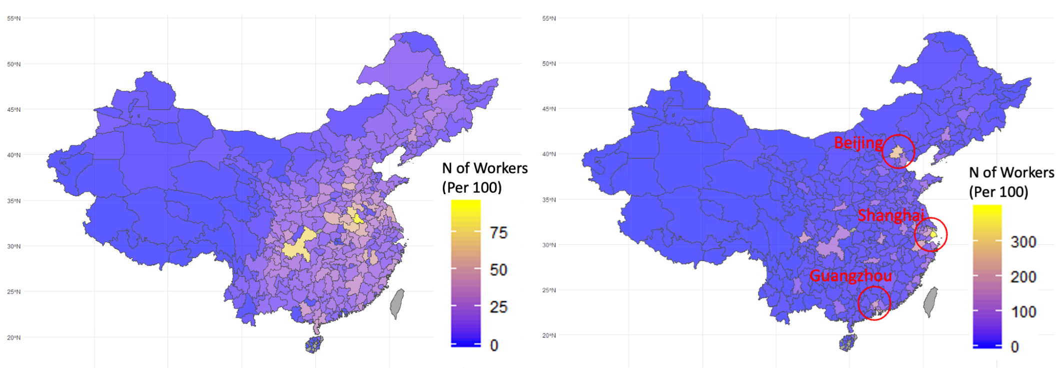
B.1 Figures

Figure B.1: *Delivery Process*

Note: This Figure plots an example of a delivery. It involves three steps: (1) workers first travel to the restaurant to pick up the food, (2) workers drive on the road, and (3) workers deliver the meal to consumers.

Figure B.2: *Geographic Distribution of Food Delivery Workers*

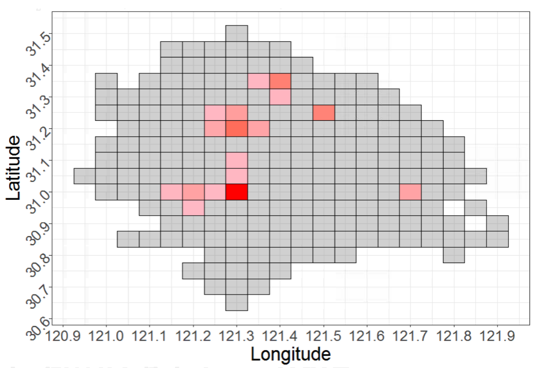
(a) *Hometown* **(b)** *Destination*



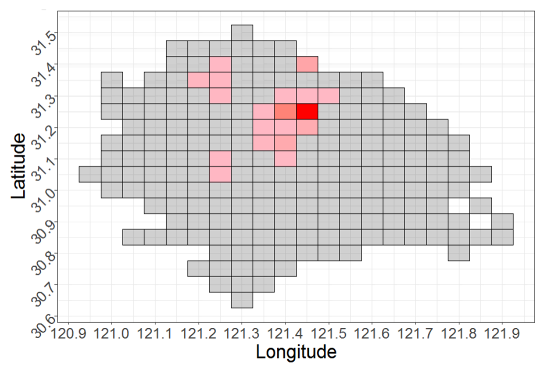
Note: This Figure shows the geographic distribution of delivery workers at the city level in China. I include all active workers between 2020 and 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.1. Figure (a) classifies workers based on their county of birth, and Figure (b) classifies workers based on their current active working cities. Color from blue to yellow represents the increase in the number of workers.

Figure B.3: *Distribution of Workers from* ***one Hometown*** *in Shanghai: Examples*

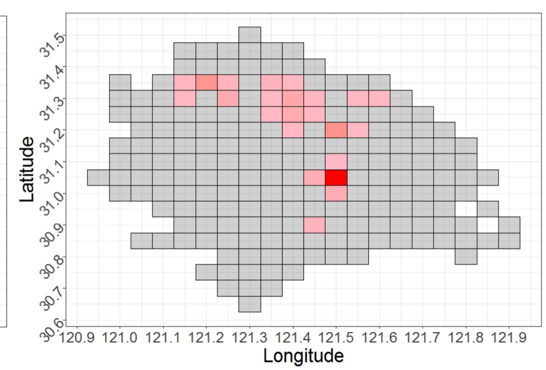
(a) *Hometown 1*



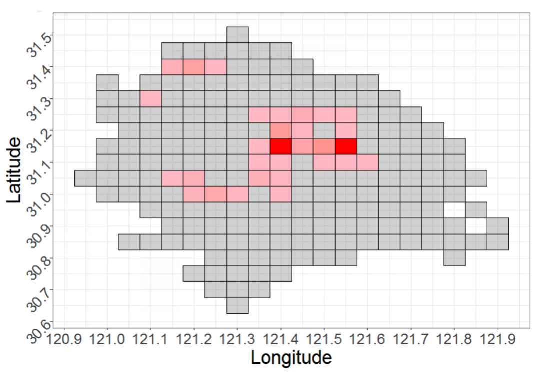
(b) *Hometown 2*



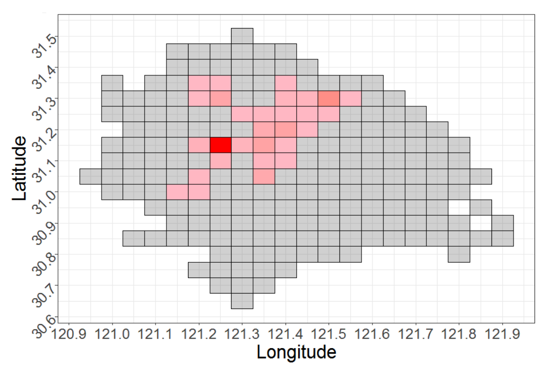
(c) *Hometown 3*



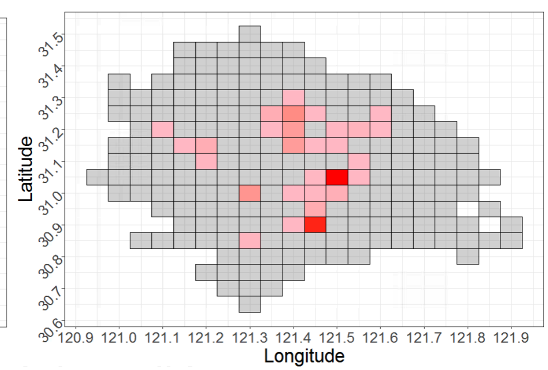
(d) *Hometown 4*



(e) *Hometown 5*



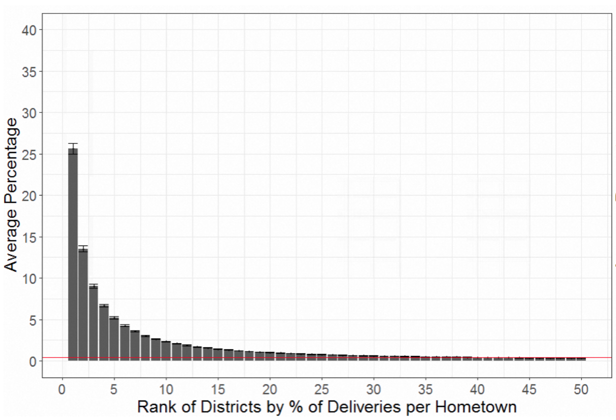
(f) *Hometown 6*



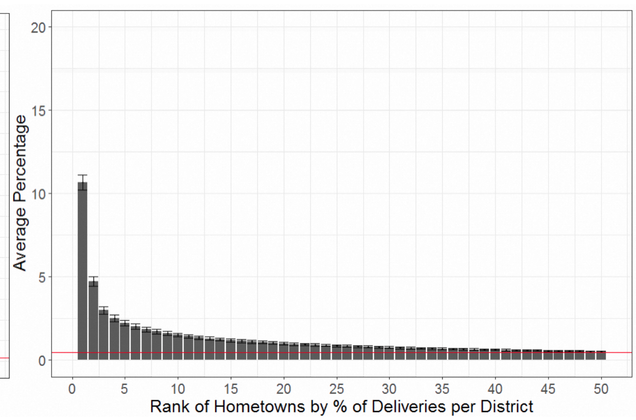
Note: This Figure provides examples of hometown clusters in Shanghai. I include all active workers between 2020 and 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.1. Figure (a)-(f) each represent a different hometown (defined as a county in China). I first find the average work location for each worker using GPS data and calculate the number of same-origin workers in each location. The color shift from light red to dark red represents the increase in the number of workers. Gray areas mean no workers from the hometown work there.

Figure B.4: *Clustering Levels by % of Deliveries*

(a) *Clustering by Sending Hometowns*

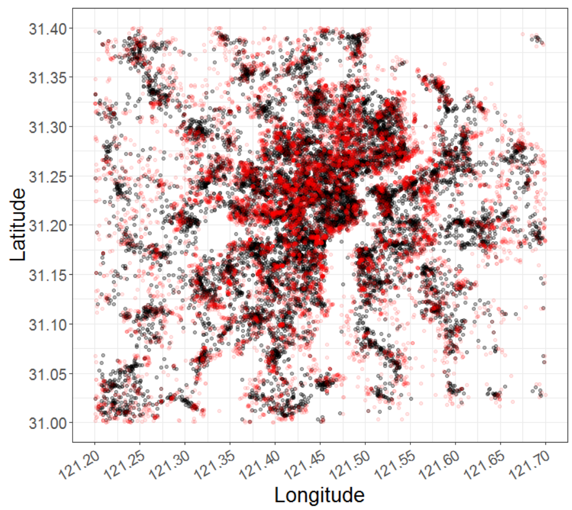


(b) *Clustering by Receiving Districts*



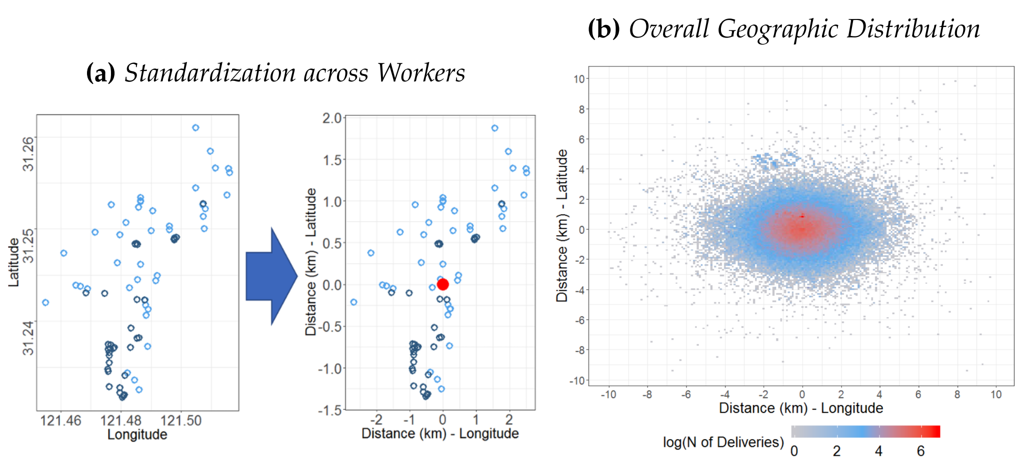
Note: This Figure plots the distribution of clustering levels. I include all active workers between 2020 and 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.1. I calculate the clustering levels across hometowns and districts following the same process as Figure 1 and 2. Calculations of clustering levels in Figure 1 and 2 are based on the number of workers. In this Figure, I compute the number of deliveries completed by workers from each hometown across districts. Figure (a) plots the share of deliveries completed in the same district by workers from each hometown. Figure (b) plots the share of deliveries completed by workers from the same origin in each district.

Figure B.5: *Distribution of Workers' Residential versus Work Location: Shanghai*



Note: This figure compares the workers' residential location versus work location for active delivery workers in Shanghai in 2021. Residential locations are red dots, inferred based on workers' daily GPS coordinates observed before 7 a.m. or after 11 p.m. Work locations are black dots calculated based on the average locations of daily deliveries. The red dots are further away from the city center and are more concentrated.

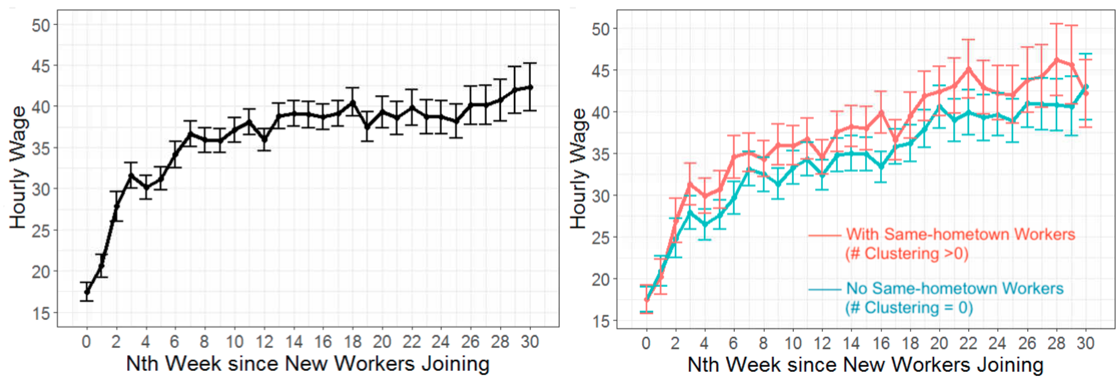
Figure B.6: *Geographic Distribution of Deliveries*



Note: This Figure shows the overall geographic distribution of deliveries for each worker. I include all new active workers in 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.3. Figure (a) shows the calculation steps. I first calculate a central point by their daily deliveries for each worker. I then calculate the distance of each delivery relative to the central point. Figure (b) plots the overall distribution of the traveling distances. Colors from gray to blue and to red represent the log of the number of deliveries. It shows most workers only travel within a $4km \times 4km$ area.

Figure B.7: *Wage Trend of New Workers*

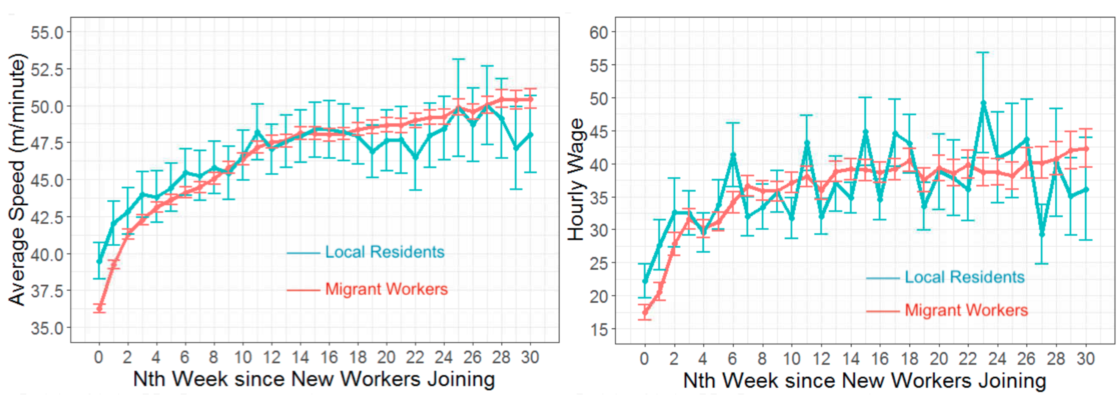
(a) *All New Workers* **(b)** *With Hometown Workers v.s. Without*



Note: This Figure plots the average wage of new workers. I include all new active workers in 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.3. In Figure (a), I plot the productivity curve for all new workers. In Figure (b), I classify these new workers into two groups based on whether they work in a district with same-origin workers: the red line represents those with at least one same-origin worker in the district, and the blue line represents those without. The x-axis is the number of weeks since new workers joined the platform, and the y-axis is the average hourly wage.

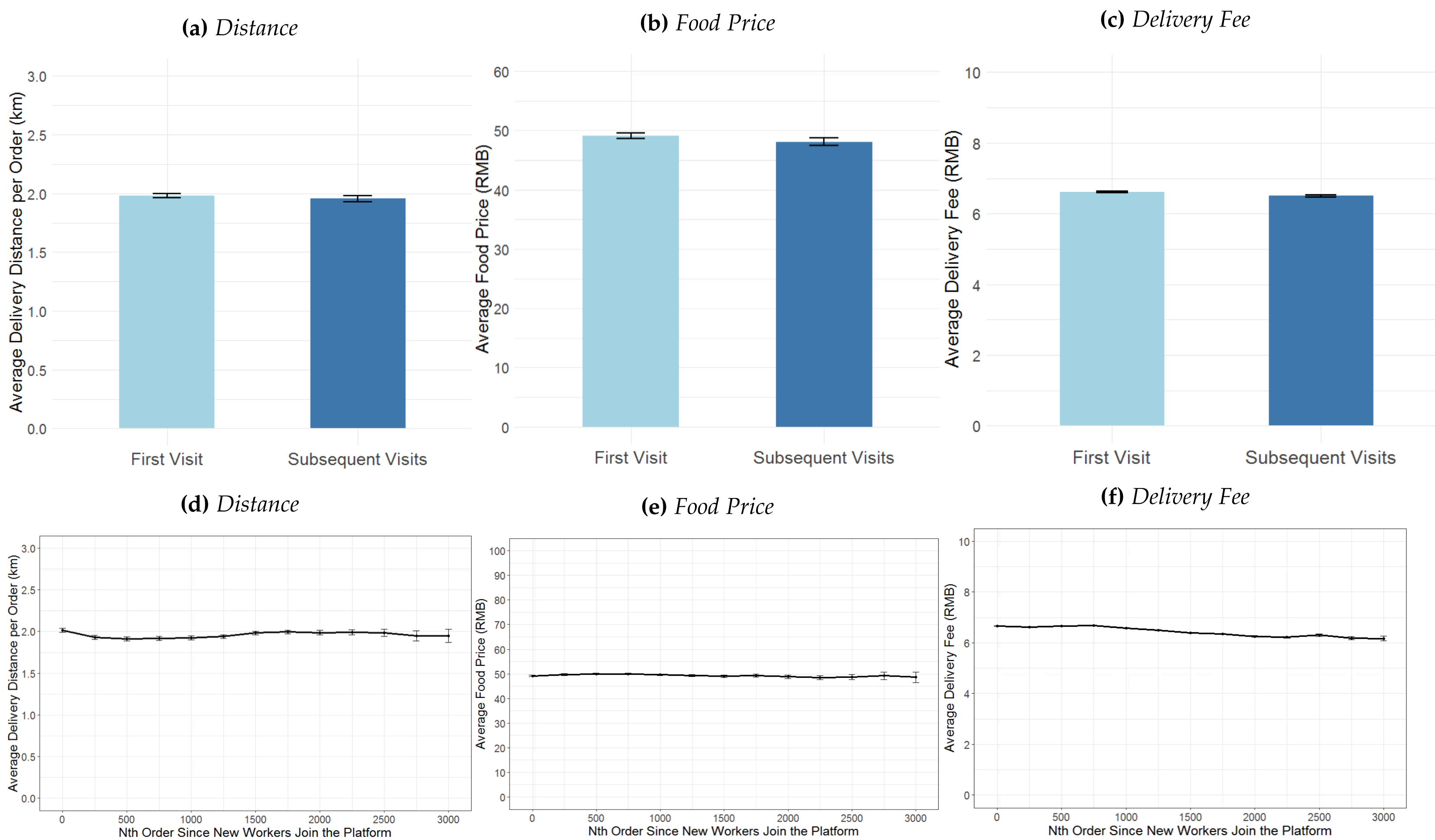
Figure B.8: *Performances of New Workers: Locals versus Migrants*

(a) *Average Travel Speed* **(b)** *Average Hourly Wage*



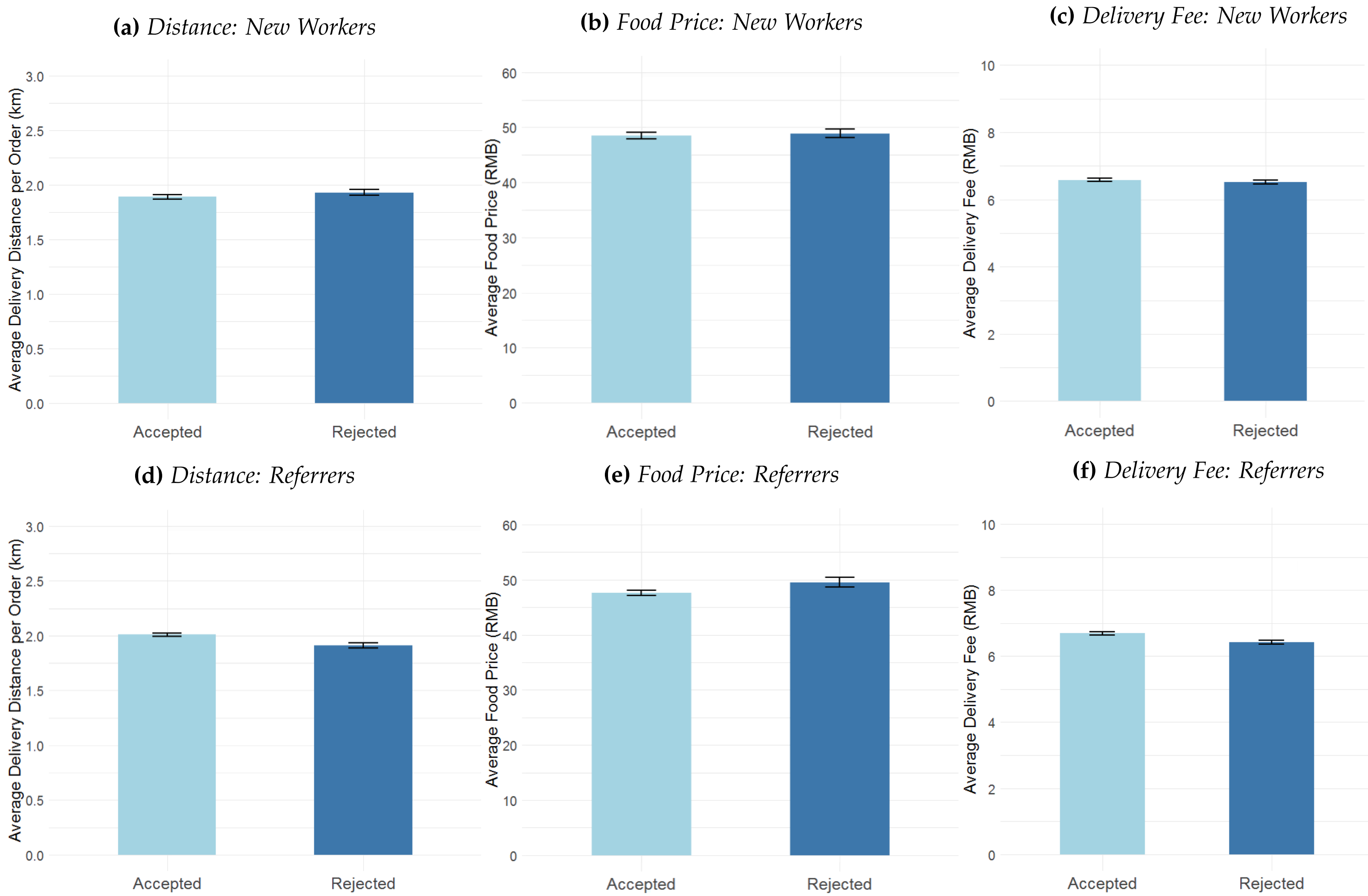
Note: This Figure plots the productivity and average wage of new workers. I include all new active workers in 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.3. I classify new workers into two groups based on whether they are local residents: the blue line represents workers born in the destination city, and the red line represents migrant workers. In Figure (a), I plot the productivity curve, which is measured by delivery speed - traveling distances (meters) to be divided by the duration (minutes). In Figure (b), I plot the average wage. The x-axis is the number of weeks since new workers joined the platform.

Figure B.9: *Balance Check: Order Characteristics across Experience and Tenure*



Note: These Figures plot characteristics of orders assigned to workers on first-time versus repeat visits to a location and workers with varying tenure. I include all new active workers between March and June 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.2. In Figures (a)-(c), I classify orders based on whether the assigned worker has visited the location before. The light bar represents the first visits, and the dark bar represents subsequent visits. In Figures (d)-(f), I examine how assigned order characteristics change as workers gain experience. I plot three order characteristics: consumer-restaurant distance, food price, and delivery fee.

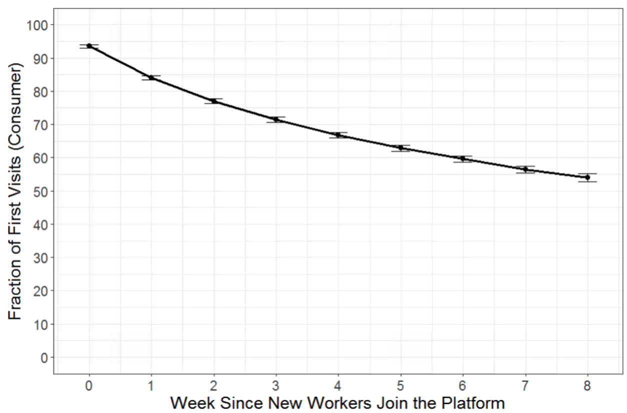
Figure B.10: *Balance Check: Order Characteristics by Rejected versus Accepcted*



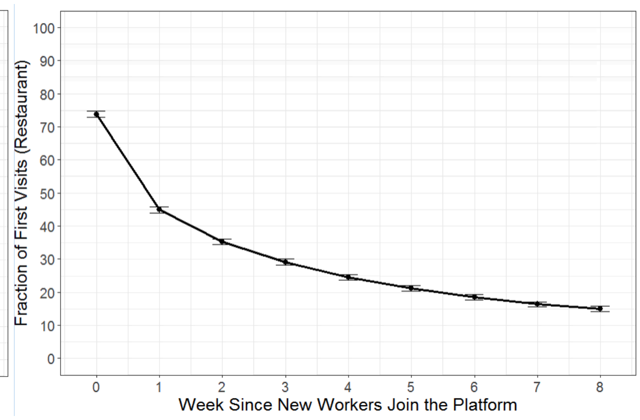
Note: These Figures plot the characteristics of orders accepted versus rejected by workers. I include all new active workers between March and June 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.2. In Figures (a)-(c), I plot orders accepted or rejected by new workers. In Figures (d)-(f), I examine orders accepted or rejected by these new workers' referrers. I plot three order characteristics: consumer-restaurant distance, food price, and delivery fee.

Figure B.11: *Distribution of Deliveries by Own and Referrers' Experience*

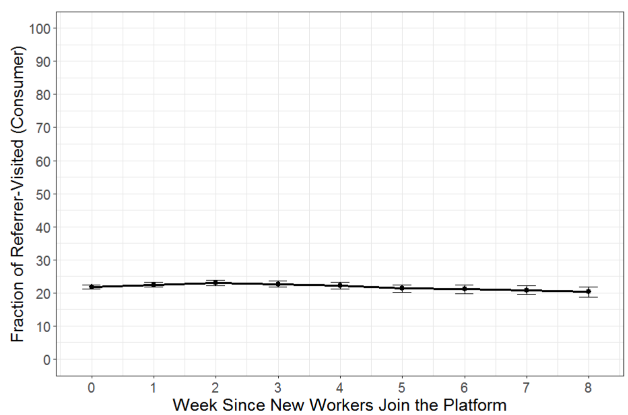
(a) *Own Experience: Consumer Address*



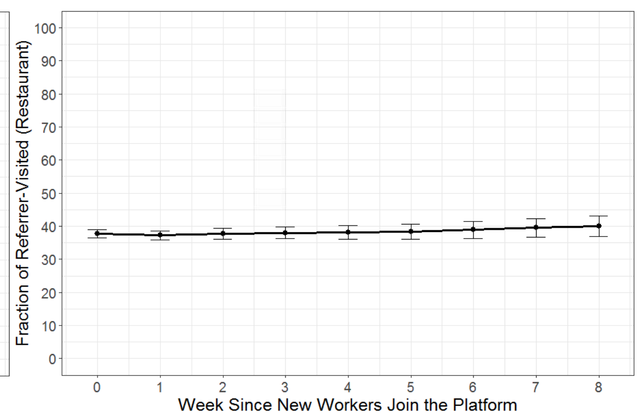
(b) *Own Experience: Restaurant*



(c) *Referrer's Experience: Consumer Address*



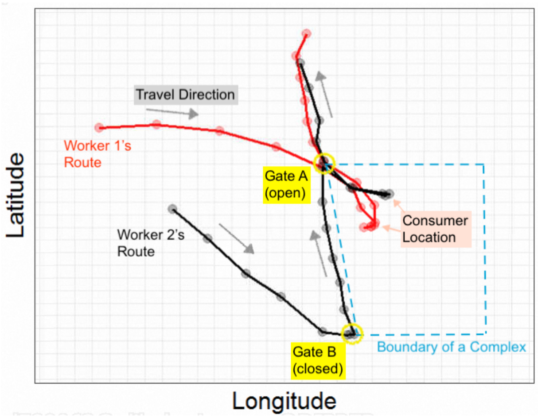
(d) *Referrer's Experience: Restaurant*



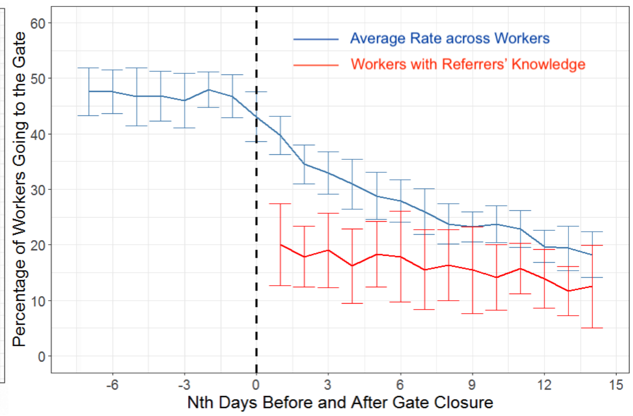
Note: This Figure plots the share of deliveries that a new worker or his referrer has visited the corresponding locations in the past month. I include all new active workers between March and June 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.2. Figure (a) plots the share of deliveries that new workers have not been to the corresponding locations before. Figure (b) represents the share of deliveries where new workers' referrers have visited the corresponding location in the past month.

Figure B.12: *Example of Deviating from Optimal Delivery Routes*

(a) *Examples of Going to the Wrong Entrance*



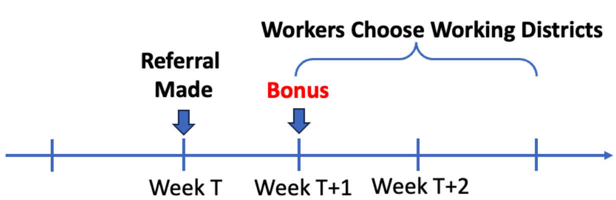
(b) *% of Workers Going to Wrong Entrances*



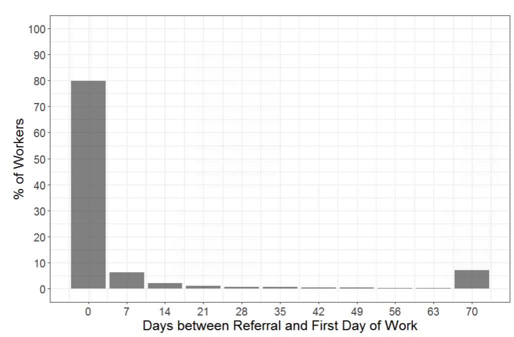
Note: This Figure provides one specific example of workers deviating from optimal delivery routes. In Figure (a), the red line represents the delivery route of a worker who knew the correct gate of the compound. The black line represents the route of a worker who first went to a locked gate and took the detour. In Figure (b), I identify twenty compounds that suddenly closed one of their gates in Shanghai in 2021. I calculate the percentage of delivery workers who still went to that locked gate the following days. I classify all workers who are assigned to deliver food to these compounds into two groups by whether their referrers have ever been to the locked gate during the period. The red line represents those with referrers visiting the locked gate before, and the blue line represents those without.

Figure B.13: *Timeline of instrumental variable*

(a) *Timeline*



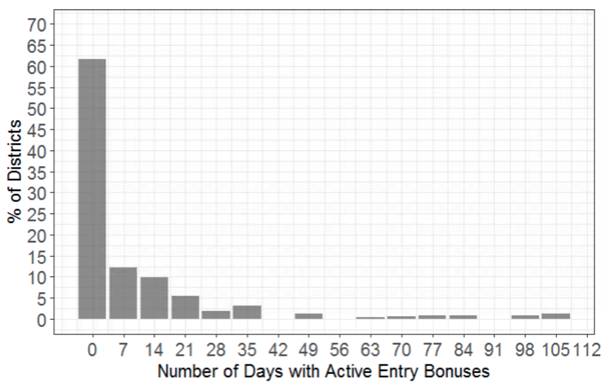
(b) *Time Difference between Referral being Made and First Day of Work*



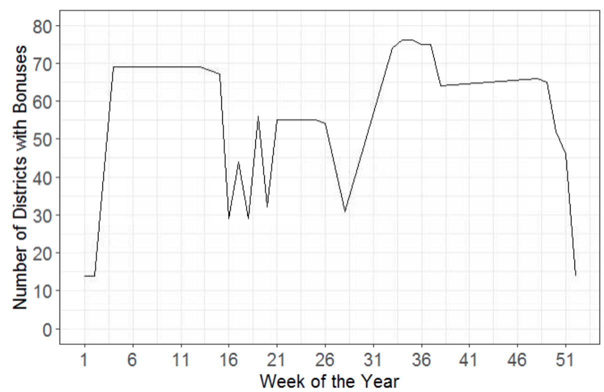
Note: This Figure shows the timeline of the instrumental variable. In Figure (b), I plot the distribution of the number of days between referrals being made and new workers' first day of work, using the same sample as 3.

Figure B.14: *Details of Entry Bonuses*

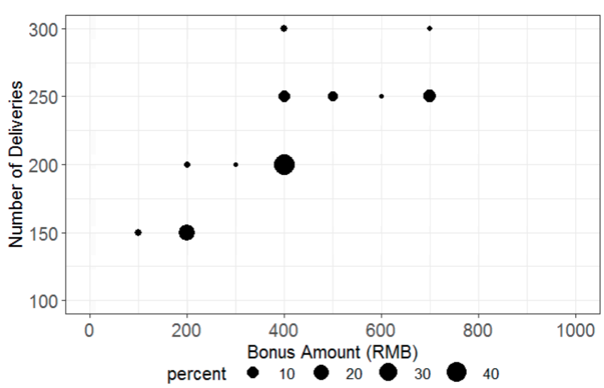
(a) *Number of Days with Active Bonuses*



(b) *Number of Districts with Active Bonuses*



(c) *Bonus Amount versus Requirement*



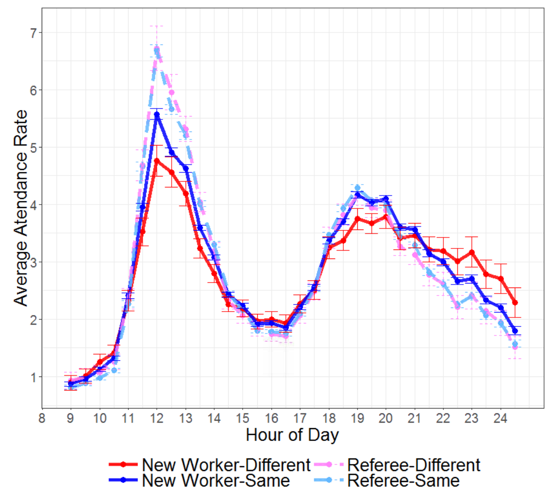
Note: This Figure plots the details of entry bonuses in Shanghai between 2020 and 2022. Figure (a) plots the average time window of entry bonuses. Figure (b) plots the number of active entry bonuses each week. Figure (c) plots the bonus requirement. The y-axis is the delivery target, and the x-axis is the bonus size.

Figure B.15: *Event Study: Weekly District-level Outcomes Before and After Bonus Increases*

(a) *# of Deliveries* **(b)** *Earnings* **(c)** *# of Active Workers*

Note: This Figure plots changes in districts with entry bonuses. I include all new active workers in 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.3. Independent variables indicate the number of weeks before or after the district offers entry bonuses. Figure (a) plots the total number of deliveries completed at the district-week level. Figure (b) plots the total delivery fee paid to workers. Figure (c) plots the total number of active workers.

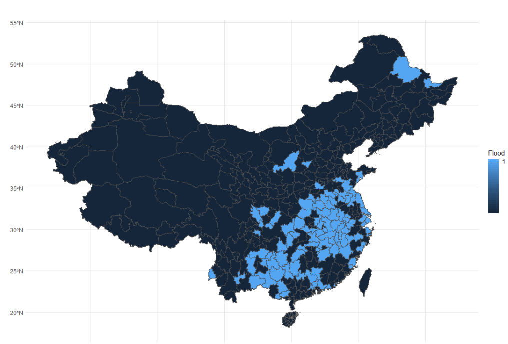
Figure B.16: *New Workers Choices of Working Hours*



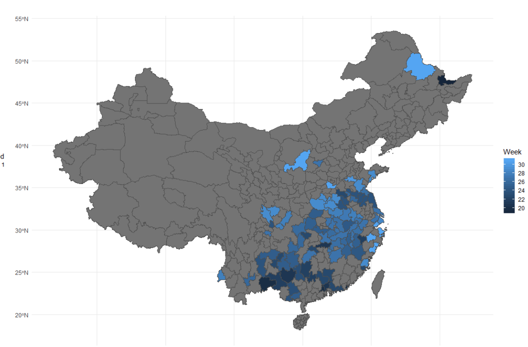
Note: This Figure plots the distribution of workers' attendance across hours. I include all new active workers between March and June 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.2. New workers are classified into two groups based on whether they work in the same district as their referral: those who cluster with referrers are represented by the blue line, and the red line represents those who do not. I also plot the hourly attendance of these new workers' referrers, as shown by the dashed line.

Figure B.17: *Geographic Distribution of Floods*

(a) *Geographic Distribution of Floods*



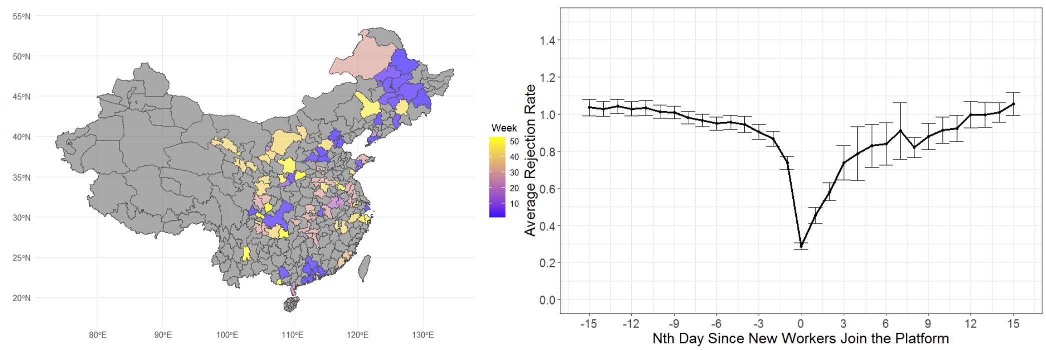
(b) *Weeks with Floods*



Note: This Figure plots the distribution of floods across cities and weeks. I use the daily rainfall data from 2423 stations (counties) in China. I identify a flood by whether two-day accumulated rainfall is over 160*mm* in each county and week. (The threshold for heavy rain and above is 80 millimeters per 24 hours.) Among counties with at least one active delivery worker in the sample during the period, 11% counties experienced floods at least once. Figure (a) plots the geographic distribution of counties with floods (plotted at the city level). Figure (b) plots the time variation of floods at the week level. The color from dark to light represents the time from June to August.

Figure B.18: *Distribution of Pandemic Lockdowns*

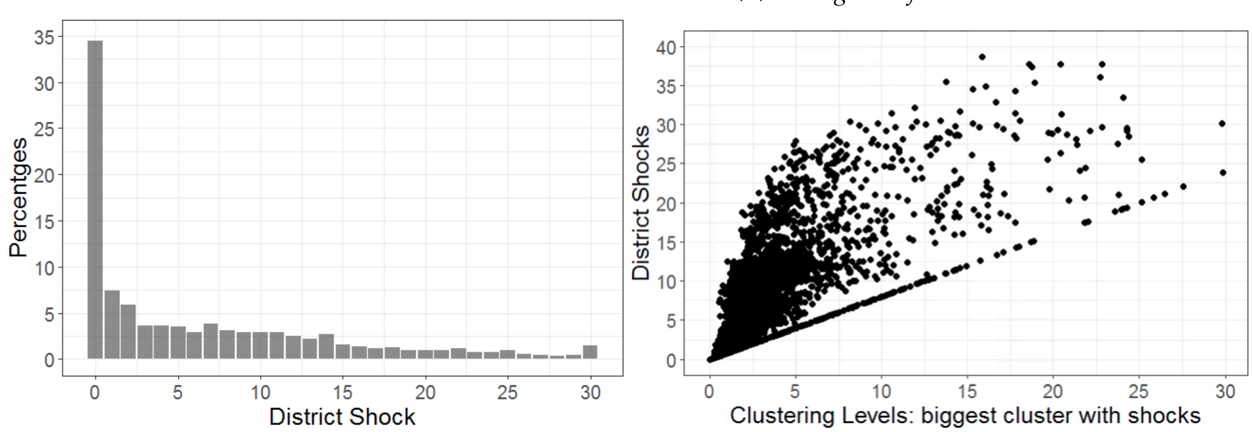
(a) *Geographic Distribution of Covid Cases* **(b)** *Identification of Lockdowns*



Note: This figure plots pandemic lockdown distributions across cities and weeks in 2021. Figure (a) shows the distribution of cities with COVID-19 cases in 2021. Colors from blue to yellow represent the first week each city had cases. Figure (b) plots consumer order responses to lockdowns in each city and week. The x-axis displays weeks before and after lockdowns. The y-axis shows the number of consumer orders.

Figure B.19: *Histogram of District-level Shocks*

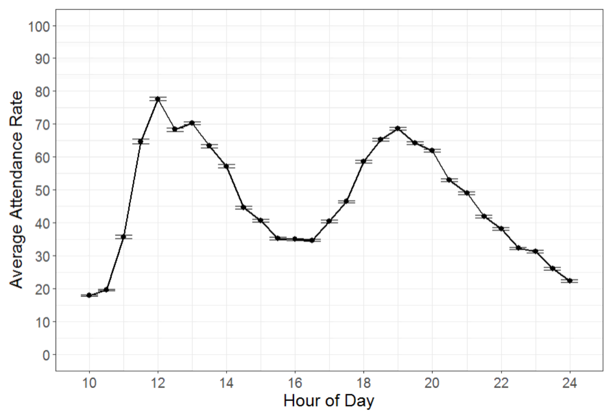
(a) *Histogram of District-level Shocks* **(b)** *Histogram of District-level Shocks*



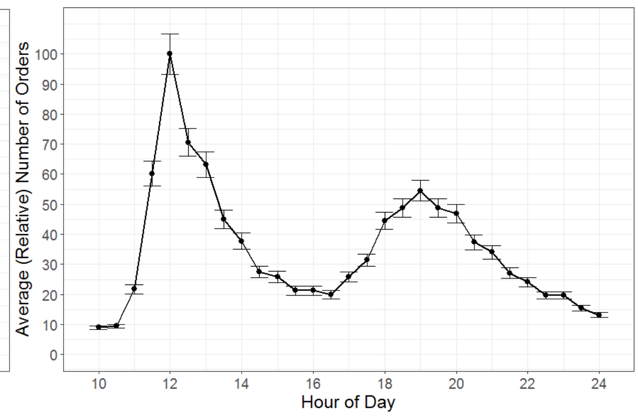
Note: This Figure plots the distribution of aggregate hometown shocks at the district level. I include all active workers between May and August 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.4. Figure (a) plots the histogram of the predicted shock share, which is the share of workers experiencing hometown floods based on the hometown composition in each district in May 2020. Figure (b) plots the correlation between the shock share and the clustering level across districts. The x-axis is the share of workers from the same origin in each district, and the y-axis is the predicted shock share.

Figure B.20: *Attendance Rate across Hours*

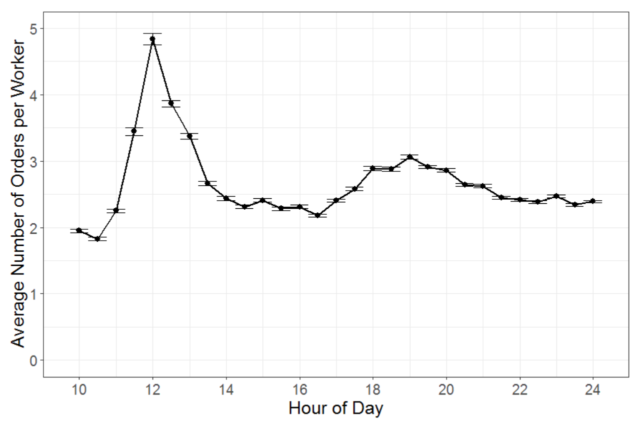
(a) *Average Attendance Rate*



(b) *Average Standardized # of Orders*



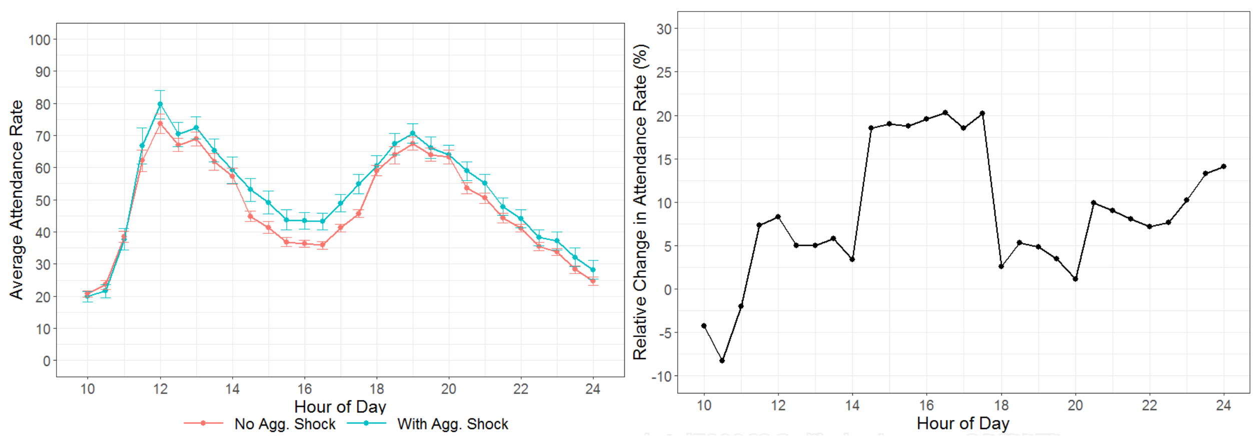
(c) *Average # of Orders per Worker*



Note: This Figure plots labor market fluctuations across hours. I include all active workers between May and August 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.4. The x-axis is the hour within a day. Figure (a) plots workers' average attendance rate across hours. Figure (b) plots the (standardized) number of orders put by consumers across hours. Figure (c) plots the average number of deliveries assigned to each active worker across hours.

Figure B.21: *Attendance Rate across Hours during District Shocks*

(a) *Average Attendance Rate* **(b)** *Relative Changes in Attendance Rates*



Note: This Figure plots the attendance rate during normal periods and shocks. I include all active workers between May and August 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.4. I classify districts into two groups by whether the district experiences aggregate shock in the week (defined as shock share $> 15\%$). The blue line represents those with shocks, and the red line represents those without. Figure (a) plots the share of workers attending to work among daily active workers in the district. Figure (b) plots the relative changes in the number of active workers with and without shocks across hours.

B.2 Tables

Table B.1: *Workers Demographic Characteristics*

	Mean	SD
Age	36.22	9.13
Gender (1=Male,0=Female)	0.972	
Number of Delivery Finished per Week	98.43	100.59
Working Hours per Week	24.14	22.26
Income per Week (RMB)	968.17	1071.32
Migrant Worker (1=Local Resident,0=The Rest)	0.983	

Notes: This Table reports delivery workers' demographic characteristics. I include all active workers between 2020 and 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.1.

Table B.2: *Determinants of the Flows of Migrant Workers*

Dep. Variable: ln(Number of Delivery Workers)				
	(1)	(2)	(3)	(4)
Home: ln(Population)	0.042***			0.241***
	(0.003)			(0.005)
Destination: ln(Population)	0.274***			0.744***
	(0.003)			(0.005)
Home: ln(GDPP)		-0.144***		-0.370***
		(0.006)		(0.007)
Destination: ln(GDPP)		1.051***		1.234***
		(0.006)		(0.007)
ln(Distance)			-0.452***	-0.409***
			(0.004)	(0.003)
Observations	115,811	115,811	115,811	115,811
R^2	0.060	0.201	0.124	0.513

Notes: This Table shows regressions of the number of delivery workers from origin cities on destination cities' characteristics. I include all active workers between 2020 and 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.1. Population and GDP are from the 2019 economic annual books at the city level. Distances are calculated based on latitude and longitude in kilometers. Standard errors are clustered at the province level.

Table B.3: *Factors Affecting Workers' Choices of Districts*

Dep. Var.: $\mathbb{1}$ { New Worker Working in the District }		
	OLS	Logit
$\mathbb{1}$ { Referrer's Distrct }	0.407***	2.431***
	(0.040)	(0.038)
$\mathbb{1}$ { Having Entry Bonus }	0.294***	1.895***
	(0.015)	(0.103)
Distance to Referrer's District		-0.305***
		(0.004)
Clustering Level in the District		0.179***
		(0.005)
District FE	Y	Y
Hometown FE	Y	Y
Entry Cohort FE	Y	Y
Observations	8,589,336	8,589,336
R^2	0.325	

Notes: This Table shows regressions of new worker's entry probabilities on district characteristics. I include all new active workers in 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.3. The dependent variable is a dummy indicating whether a new worker works in the district. Three independent variables are (1) whether the new worker's referrer works in the district; (2) whether there is an entry bonus in the district in the subsequent week after the new worker joins the platform; (3) the clustering level in the districts, which is measured as the share of same-origin workers. I run a linear regression in column (1) and a logit regression in column (2). Standard errors are clustered two-way at the hometown level and entry-week level.

Table B.4: *Decompose Local Knowwldge*

Robustness Check (1): $\mathbb{E}(P_{ia})$ *and* $\mathbb{E}(P_{i^ja})$ *as controls*

	Restaurant Search Time (minute)	Road Driving Speed (meter/minute)	Consumer Search Time (minute)
	(1)	(2)	(3)
Own Visit	-1.45***	-0.13	-1.28***
	(0.07)	(0.10)	(0.05)
Referrer Visit	-0.64***	-0.20	-0.70***
	(0.10)	(0.17)	(0.08)
Date X Hour X Worker FE	Y	Y	Y
Location FE	Y	Y	Y
Referrer's Expected Probability of Visiting a Location	Y	Y	Y
Ave. Dep. Var.	6.22	10.86	5.93
Observations	5,837,304	5,837,304	5,837,304
R^2	0.43	0.38	0.43

Notes: This Table shows regressions of productivity on four indicators: (1) whether the new worker has visited the restaurant or consumer building before; (2) whether the new worker's referrer has visited the restaurant or consumer building in the last month, following the specification in section 4.2. I include all new active workers between March and June 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.2. Dependent variables in columns (1) and (3) are searching time (minutes) for the restaurant or the consumer building. I measure by counting GPS coordinates within a location's 150-meter radius. The dependent variable in column (2) is the average driving speed (minute/min) between the restaurant and the consumer building. These are proxies for workers' productivity. Difference from Table 1, I construct the expected probability of a worker visiting a location by running a logit regression of the probability of visiting a location on workers' characteristics, including tenure, age, gender ratings, and so on (Borusyak and Hull, 2023). Standard errors are clustered two-way at the worker level and the date level.

Table B.5: *Effect of Clustering on Worker Performance*

Robustness Check (1): additional dependent variables

	Retention Rate (%)	Relocation Rate (%)	Timeout Rate (%)
	(1)	(2)	(3)
Actual Clustering Level	0.261	-0.377**	-0.038***
	(0.229)	(0.186)	(0.009)
District X Week FE	Y	Y	Y
Hometown X Week FE	Y	Y	Y
Entry Cohort FE	Y	Y	Y
BH Control	Y	Y	Y
Ave. Dep. Var.	76.29	9.51	8.45
Observations	417,684	417,684	417,684

Notes: This Table shows regressions of new workers' productivity on clustering levels. I include all new active workers in 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.3. I run IV regressions where the independent variable is the share of hometown workers in each new worker's work district, and the instrumental variable is the predicted clustering level induced by entry bonuses. Dependent variables are (1) whether a new worker is active on the platform the following week, (2) whether a new worker relocates to a different district the following week, and (2) the share of late deliveries. Standard errors are clustered two-way at the district level and the week level.

Table B.6: *Effect of Clustering on Worker Performance*

Robustness Check (2): Various Methods of Instrumental Variable Construction

Dep. Var.: Average Delivery Speed (meter/min)	One-week Bonus	All Districts	Sole Prediction	Dynamic	Bonus Size
	(1)	(2)	(3)	(4)	(5)
Actual Clustering Level	19.147***	22.070***	16.501***	17.819***	19.733***
	(2.549)	(6.824)	(0.918)	(1.241)	(1.940)
District X Week FE	Y	Y	Y	Y	Y
Hometown X Week FE	Y	Y	Y	Y	Y
Entry Cohort FE	Y	Y	Y	Y	Y
BH Control	Y	Y	Y	Y	Y
Observations	417,684	417,684	257,293	417,684	417,684
First-Stage F	5820	962	1.51e+04	1.03e+04	8393

Notes: This Table shows regressions of new workers' productivity on clustering levels. I include all new active workers in 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.3. The dependent variable is each worker's average weekly delivery speed, as measured by delivery distance, divided by delivery duration. I run IV regressions where the independent variable is the share of hometown workers in each new worker's work district, and the instrumental variable is the predicted clustering level induced by entry bonuses. Columns (1)-(5) report results under different IV specifications. Column (1) uses entry bonuses with a take-up time window of less than one week; column (2) uses all entry bonuses in the destination city without restriction to the referrers' district; column (3) focuses on workers with only one bonus-predicted district; column (4) constructs the instrument based on time-varying clustering levels across districts, and column (5) construct the predicted clustering level to be weighed by the size of entry bonuses. Standard errors are clustered two-way at the district level and the week level.

Table B.7: *Robustness Check: Effect of Clustering on Worker Performance*

Robustness Check (3): Various Methods of Independent Variable Construction

Dep. Var.: Average Delivery Speed (meter/min)					
Indep. Var.:	Number of Hometown Worker	$\mathbb{1}$ { Working in Referrer's District }	Clustering Level Adjusted by Tenure	Share of Deliveries by Hometown	City as Hometown
	(1)	(2)	(3)	(4)	(5)
Indep. Var.	0.218***	1.003***	15.362***	21.155***	12.813**
	(0.014)	(0.092)	(2.936)	(1.539)	(5.194)
District X Week FE	Y	Y	Y	Y	Y
Hometown X Week FE	Y	Y	Y	Y	Y
Entry Cohort FE	Y	Y	Y	Y	Y
BH Control	Y	Y	Y	Y	Y
Observations			417,684		
First-Stage F	1.04e+04	1.83e+04	1.00e+04	1.13e+04	690

Notes: This Table shows regressions of new workers' productivity on clustering levels. I include all new active workers in 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.3. The dependent variable is each worker's average delivery speed in each week, as measured by delivery distance, to be divided by delivery duration. I run IV regressions where the independent variable is the share of hometown workers in each new worker's work district, and the instrumental variable is the predicted clustering level induced by entry bonuses. Columns (1)-(5) report results under different independent variable specifications. Column (1) uses the number of hometown peers instead of the share in the baseline specification; column (2) reports the effect of locating in the same district as one's referrer on productivity; column (3) constructs the clustering level adjusted by tenure; column (4) explore the share of deliveries completed by hometown peers as the proxy for clustering levels; and column (5) expands hometown peers to workers from the same birth city. Standard errors are clustered two-way at the district level and the week level.

Table B.8: *Delivery workers' Responses to Shocks*

Robustness Check (1): varying the threshold for hometown floods

Dep. Var.: Average Weekly Outcomes			
	$\mathbb{1}$\{ Active \}	Ave. Deliveries	Ave. Hours
Using 200 millimeters as threshold for floods			
$\mathbb{1}$\{Hometown Flood \}	0.062**	19.038***	6.812**
	(0.029)	(5.274)	(3.092)
Province × Week FE	Y	Y	Y
Worker FE	Y	Y	Y
District FE	Y	Y	Y
Ave. Dep. Var.	0.76	78.10	22.08
Observations	638,810	638,810	638,810
R^2	0.519	0.488	0.523

Notes: This Table shows regressions of workers' weekly labor supply on hometown shocks. I include all active workers between May and August 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.4. Dependent variables are workers' weekly performances on the platform: (1) being active (# of deliveries > 0),(2) number of deliveries completed, and (3) number of working hours. The independent variable indicates whether a worker's hometown has had flood shock in the last month. Different from Table 4, which uses 160*mm* as the threshold for floods, this Table uses 200*mm* as the threshold for floods as the independent variable. I include origin province × week fixed effect, district fixed effect, and worker fixed effects. Standard errors are clustered two-way at the district level and the week level.

Table B.9: *Delivery workers' Responses to Shocks*

Robustness Check (2): hometown pandemic lockdowns

Dep. Var.: Average Weekly Outcomes			
	$\mathbb{1}\{$ Active $\}$	Ave. Deliveries	Ave. Hours
Panel A: Covid Cases			
Number of weekly COVID cases	0.001***	0.173***	0.085***
	(0.000)	(0.041)	(0.018)
Province × Week FE	Y	Y	Y
Worker FE	Y	Y	Y
District FE	Y	Y	Y
Ave. Dep. Var.	0.73	77.54	21.40
Observations	2,387,219	2,387,219	2,387,219
R^2	0.489	0.498	0.512
Panel B: Identified Pandemic Lockdown			
Experiencing Hometown Lockdown	0.076***	12.810***	4.184***
	(0.018)	(2.319)	(0.822)
Province × Week FE	Y	Y	Y
Worker FE	Y	Y	Y
District FE	Y	Y	Y
Ave. Dep. Var.	0.76	78.10	22.08
Observations	2,387,219	2,387,219	2,387,219
R^2	0.502	0.523	0.527

Notes: This Table shows regressions of workers' weekly labor supply on hometown shocks. I include all active workers in 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.5. Dependent variables are workers' weekly performances on the platform: (1) being active (# of deliveries > 0),(2) number of deliveries completed, and (3) number of working hours. The independent variable indicates whether a worker's hometown has any pandemic lockdowns. Panel A uses the number of weekly COVID cases at the city level, as scraped from government websites. Panel B identifies hometown pandemic lockdown by whether the number of consumer orders dropped below half of the median of normal periods in workers' hometowns. I include origin province × week fixed effect, district fixed effect, and worker fixed effects. Standard errors are clustered two-way at the district level and the week level.

Table B.10: *delivery workers' Responses to Shocks*

Robustness Check (3): Continuous District-level Shocks

Dep. Var.: Average Weekly Outcomes			
	Number of Workers	Number of Orders	Total Working Hours
Predicted Fraction with Shocks (%)	0.288	3.127	9.683***
	(0.177)	(2.490)	(3.018)
Week FE	Y	Y	Y
District FE	Y	Y	Y
Ave. Dep. Var	132.59	7,233.97	2,408.26
Observations	14,391	14,391	14,391
R^2	0.781	0.781	0.802

Notes: This Table shows regressions of district-level market performances on aggregate hometown shocks. I include all active workers between May and August 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.4. Dependent variables are (1) the total number of active workers in the district each week,(2) the total number of deliveries completed, (3) the total number of working hours, and (4) the average per-delivery commission fee. Different from Table 5, where I use an indicator for the aggregate shock, the independent variable in this Table is the predicted shock share in the district. The shock share is the share of workers experiencing floods, given the hometown composition in the district in May 2020. I include week and district fixed effects. Standard errors are clustered at the week level.

Table B.11: *delivery workers' Responses to Shocks*

Robustness Check (4): Peak Hour versus Off-peak Hours

Dep. Var.: Average Weekly Outcomes			
	Attendance Rate	Ave. Deliveries	Ave. Hours
Panel A: Peak Hours			
$\mathbb{1}$\{ Hometown Shock\}	0.052***	2.559***	0.577***
	(0.010)	(0.511)	(0.103)
$\mathbb{1}$\{District Shock Share > 15%\}	0.012	1.661	0.411
	(0.009)	(1.307)	(0.381)
Interaction Term	0.009	0.418	0.134
	(0.006)	(0.388)	(0.095)
Panel B: Off-peak Hours			
$\mathbb{1}$\{ Hometown Shock\}	0.272***	17.102***	5.898***
	(0.051)	(3.190)	(1.121)
$\mathbb{1}$\{District Shock Share > 15%\}	-0.044**	-3.127**	-1.093**
	(0.018)	(1.546)	(0.541)
Interaction Term	-0.013	-7.037***	-0.700
	(0.010)	(1.618)	(0.567)
Province × Week FE	Y	Y	Y
Worker FE	Y	Y	Y
District FE	Y	Y	Y
Observations	638,810	638,810	638,810

Notes: This Table shows regressions of workers' weekly labor supply on hometown shocks and district shocks. I include all active workers between May and August 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.4. Dependent variables are workers' weekly performances on the platform: (1) being active (# of deliveries > 0),(2) number of deliveries completed, and (3) number of working hours. Independent variables are (1) the indicator of whether a worker's hometown has any flood shock in the last month, (2) an indicator of district shock based on whether the predicted shock share in the district is higher than 15%, and (3) their interaction term. Different from Table 6, which uses each worker's total working hours per week, I separate workers' performance during peak versus non-peak hours in this Table. Peak hours are defined as between 11 am and 2 pm or between 6 pm and 8 pm. I include week fixed effects, worker fixed effects, and district fixed effects. Standard errors are clustered two-way at the district level and the week level.

Table B.12: *Effect of Relocation on Productivity*

Dep. Var.: Average Weekly Outcomes				
	Relocation Rate	Effect of Relocation		
		Ave. Speed	Deliveries Per Hour	Timeout Rate (%)
$\mathbb{1}$\{ Hometown Shock\}	-0.007			
	(0.012)			
$\mathbb{1}$\{District Shock Share > 15%\}	0.008			
	(0.016)			
Interaction Term	0.033^{**}			
	(0.014)			
Relocation: $\mathbb{1}\{j_{t-1} \neq j_t\}$		-1.299^{**}	-0.158^{**}	0.361^{***}
		(0.635)	(0.070)	(0.048)
Province × Week FE	Y	Y	Y	Y
Worker FE	Y	Y	Y	Y
District FE	Y	Y	Y	Y
Ave. Dep. Var.	0.09	46.18	4.58	9.36
Observations	638,810	638,810	638,810	638,810
R^2	0.498	0.348	0.339	0.342

Notes: This Table shows regressions of workers' relocation probability on hometown shocks. I include all active workers between May and August 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.4. The dependent variable in column (1) is whether a worker relocates to a different district the following week. Independent variables are (1) the indicator of whether a worker's hometown has any flood shock in the last month, (2) an indicator of district shock based on whether the predicted shock share in the district is higher than 15%, and (3) their interaction term. Columns (2)-(4) report regressions of workers' labor market performances on the relocation indicator. Dependent variables are (1) workers' average delivery speed, (2) the number of deliveries completed per hour, and (3) the share of late deliveries. The independent variable is whether a worker has just relocated to the district this week. I include week fixed effects, worker fixed effects, and district fixed effects. Standard errors are clustered two-way at the district level and the week level.

Table B.13: *Shock Probability and Clustering Levels*

Dep/ Var.: Clustering Level for Each Hometown			
	(1)	(2)	(3)
Flood Prob. $(0 \sim 1)$	-6.59***		
	(1.37)		
$\{Prob. > 10\%\}$		-0.56***	
		(0.15)	
$\{Prob. > 15\%\}$			-0.76***
			(0.22)
Week FE	Y	Y	Y
Origin Province FE	Y	Y	Y
Destination City FE	Y	Y	Y
Ave. Dep. Var.	11.36	11.36	11.36
Observations	78,776	78,776	78,776
R^2	0.26	0.23	0.21

Notes: This Table shows regressions of clustering levels on flood probability for each hometown. I include all active workers between 2020 and 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.1. The dependent variable is the average clustering level for each hometown in each destination city. The clustering level is defined as the highest share of hometown workers in each district within each city. The independent variable is the probability of flood shock. I use the rainfall data between 2000 and 2020 to compute each county's probability of experiencing floods. In column (1), I use the continuous flood probability. Columns (2) and (3) use indicators for probabilities over 10% or 15%. Standard errors are clustered two-way at the origin province level and the destination city level.

Table B.14: *Hometown Heterogeneity by Shock Probability*

	Number of Workers	Distance to City Center (km)	Tenure (month)	Delivery Speed (m/min)
	(1)	(2)	(3)	(4)
Flood Prob. $(0 \sim 1)$	5.18*	-1.49	0.59	0.81
	(3.11)	(2.50)	(1.75)	(0.66)
Week FE	Y	Y	Y	Y
Origin Province FE	Y	Y	Y	Y
Destination City FE	Y	Y	Y	Y
Ave. Dep. Var.	40.01	6.29	6.97	47.03
Observations	78,776	78,776	78,776	78,776
R^2	0.23	0.29	0.27	0.36

Notes: This Table shows regressions of hometown characteristics on flood probability for each hometown. I include all active workers between 2020 and 2021 in five large cities. Details of the sample selection can be found in Appendix C.2.1. The variables are (1) the number of active workers from each hometown, (2) workers' average distance to city centers, (3) average tenure, and (4) average delivery speed. The independent variable is the probability of flood shock. I use the rainfall data between 2000 and 2020 to compute the probability of each county. Standard errors are clustered two-way at the origin province level and the destination city level.

C Appendix: Data Construction

C.1 Data Overview

C.1.1 Data Sources

I use data from four main sources: (1) worker performance metrics on the platform, (2) information on the number of buildings and year of construction across Shanghai, as scraped from Lianjia.com Website, (3) daily rainfall data from 2,423 stations (counties) in China, and (4) daily records of confirmed COVID-19 cases by city, as scraped from government websites.

I use the building characteristics in Shanghai to understand where learning occurs and the latter two data sources (floods and COVID cases) to identify adverse shocks in migrant workers' hometowns. Next, I describe the platform data in detail.

C.1.2 Platform Data

GPS Data. Workers' GPS coordinates are recorded every twenty seconds when workers are active on the platform. Each GPS data coordinate includes the latitude, longitude, and exact timestamp.

Order Information. For each order/delivery, the data set includes the restaurant location (restaurant name and GPS coordinates), consumer location (building name and GPS coordinates), the delivery worker ID, total price, delivery fee, and delivery distance. It also contains six timestamps: (1) when the consumer places the order, (2) when the delivery worker accepts the order, (3) when the worker arrives at the restaurant, (4) when the food is ready as confirmed by the restaurant, (5) when the worker arrives at the consumer drop-off location, and (6) when the consumer confirms receiving the food. All consumer information and worker IDs are anonymized.

Network Data. The network data records each referral pair of a new worker and the referrer. Referrers are existing delivery workers on the platform. The data set includes the anonymized IDs of the two workers and the timestamp when the referral was made.

Platform Bonuses and Lotteries. This data covers all entry bonuses and platform lotteries conducted through the platform. It includes the timestamp when each bonus was active, the bonus amount, the requirements, and the geographic eligibility. For workers, the data shows which anonymized worker ID received a bonus or lottery prize and when.

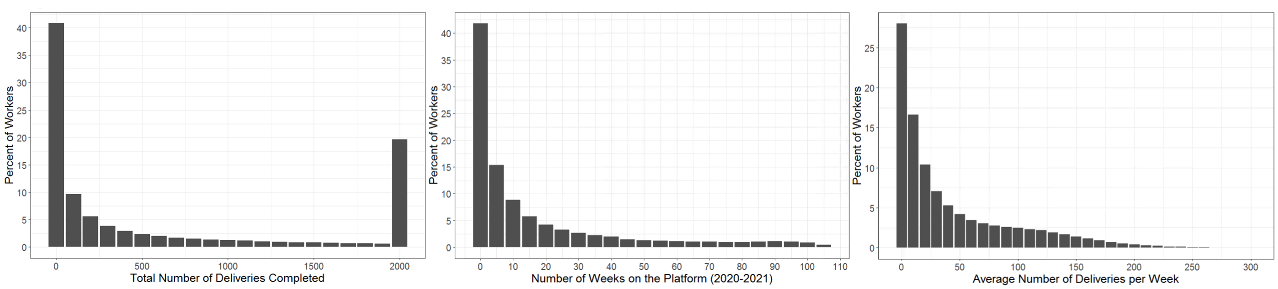
Worker Characteristics. Worker characteristics data includes each worker's county of birth, gender, age, current work city, start date, and total number of deliveries completed.

C.1.3 Overall Sample Selections

The analysis uses data on all delivery workers on the food delivery platform across five major Chinese cities (Shanghai, Beijing, Guangzhou, Shenzhen, and Hangzhou) between 2020 and 2021. Approximately 1 million people completed at least one delivery in these five cities during this period. However, as with most gig-economy sectors, this food delivery industry experiences high turnover. Many workers stay on the platform for less than a week and complete fewer than ten deliveries. Figure C.1 plots the distribution of total deliveries completed per worker, the number of weeks worked, and average weekly deliveries completed per worker. All histograms skew right with peaks near zero. Specifically, 41% of workers completed fewer than 100 deliveries by the end of 2021, 42% worked less than five weeks, and 72% completed fewer than 50 deliveries weekly on average.

Figure C.1: *Workers' Performances on the Platform*

(a) *All Deliveries* **(b)** *Number of Weeks* **(c)** *Number of Weekly Deliveries*



Note: This Figure plots the histogram of total deliveries completed per worker, the number of weeks worked, and average weekly deliveries completed per worker. The figure builds on all delivery workers on the food delivery platform across five major Chinese cities (Shanghai, Beijing, Guangzhou, Shenzhen, and Hangzhou) between 2020 and 2021.

Using the entire sample of workers for analysis may introduce noise. For instance, the 42% of workers who stayed less than five weeks may not have had enough time to learn. Also, the 72% who completed fewer than 50 deliveries per week likely did not rely on this job as their primary income source. Their decisions and choices could differ from typical migrant workers, who are the primary earners for families back home and rely on this delivery work as their major income source.

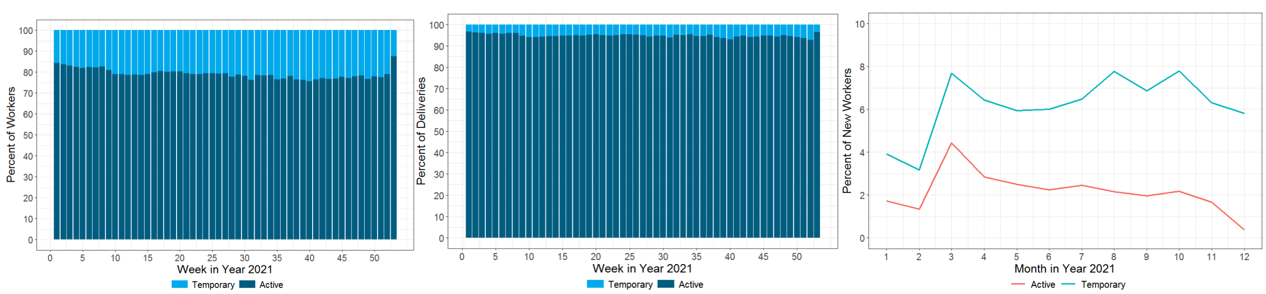
I thus classify workers into two types: active workers and temporary workers. Active workers are defined as those who worked for at least four weeks on the platform and completed at least 50 deliveries per week on average, by the end of 2021.

I choose these thresholds because the average monthly salary for domestic migrant workers is 4,000 RMB (around $570 USD) in China in 2020 (Bureau, 2020). Completing 50 deliveries weekly over four weeks can generate approximately 2,000 RMB per month as a delivery worker - about half of the average migrant worker's salary.

Among all delivery workers across the five cities in 2020-2021, 37% met the criteria to be classified as active workers. However, these active workers completed 95% of total deliveries on the platform. Figure C.2 plots the percentage of workers and deliveries completed by active versus temporary workers each week in 2021. On average, active workers comprised 75% of the workforce but completed 95% of weekly deliveries.[57] Focusing on this active worker sample still captures the majority of platform orders.

Figure C.2: *Workers' Performances on the Platform*

(a) *Weekly Worker Share* **(b)** *Weekly Delivery Share* **(c)** *New Worker Share*



Note: This Figure shows the percentage of workers and the percentage of deliveries completed by active vs. temporary workers each week in 2021.

Figure C.2(c) also plots the entry of active and temporary workers by week in 2021. It shows that the largest entry of active workers occurred in March, the month following the Chinese New Year. This aligns with typical seasonal migration patterns, where migrant workers travel to work cities after the festival and remain there working for approximately one year before returning home for the next New Year celebration.

For most analyses, I use the sample of active workers as defined above rather than the full worker sample. However, when calculating metrics such as total deliveries or work hours by district or computing clustering levels by delivery share, I use the full sample.

[57] The percentage of active workers is 75% weekly but only 37% over the full 2020-2021 sample. This difference is due to the much higher turnover rate among temporary workers, resulting in a larger share of the total sample.

C.2 Sample Selection per Analysis

C.2.1 Distribution of Workers by Hometown and Work District

Sample Selection

The analysis covers active delivery workers on the food delivery platform across five major Chinese cities (Shanghai, Beijing, Guangzhou, Shenzhen, and Hangzhou) in 2020 and 2021. Active workers are defined as those who worked at least four weeks and completed a minimum of 50 deliveries per week by the end of 2021, as discussed in Section C.1.3. The final sample contains 402,394 active workers.

Variable Definition

Hometown. A worker's hometown is defined as their county of birth. There are 2,843 counties in China, of which 2,805 are represented by at least one active delivery worker in the five cities between 2020 and 2021.

Work District. A worker's work district is defined weekly based on their GPS coordinates. For each worker-week, I identify the district with the maximum number of GPS data points for that worker.

C.2.2 Direct Evidence of Knowledge Spillovers

Sample Selection

The analysis covers all deliveries on the food delivery platform in the five major Chinese cities from March to June 2021. I include deliveries that were completed by a new active worker or their referrer. I define new active delivery workers as those who entered after the Chinese New Year festival (February 12, 2021) and before June 2021, had a referrer upon entry, and met the active worker criteria defined in Section C.1.3. This comprises 23,846 new workers and 5,837,304 analyzed deliveries.

Variable Definition

Search Time. A worker's search time for a restaurant or consumer building is measured based on GPS coordinates within a 150-meter radius of the location. I have the exact timestamp for each GPS data point. The search time is calculated as the last timestamp when the worker is within the 150-meter radius minus the first timestamp.

Driving Speed. A worker's driving speed is measured as the total driving distance divided by duration, excluding time within 150 meters of restaurants or buildings. When a worker picks up food from restaurant A and delivers it to consumer B, they may drive to other restaurants or consumers along the route. In such cases, I also exclude search time at those additional stops. Thus, the delivery speed aims to capture the actual driving speed on main roads between the pickup and drop-off points.

C.2.3 IV Regression of Productivity on Clustering

Sample Selection

The analysis covers new active delivery workers on the food delivery platform across the five major Chinese cities in 2021. New active workers are defined as those who joined in 2021, had a referrer at entry, and met the active worker criteria in Section C.1.3. This comprises 53,498 new workers. I create a panel data set tracking each worker from platform join date to permanent exit, capped at the first twelve weeks if a worker stays beyond twelve weeks. The final sample contains 417,684 observations at the worker-week level.

Variable Definition

Delivery Speed. A worker's delivery speed per order is measured as the total driving distance divided by duration. The duration starts when the worker accepts the order and starts moving. The duration ends at delivery to the consumer. The whole process includes search time within 150 meters of restaurants and buildings, contrary to the previous analysis. As mentioned previously, if additional stops are made between pickup at restaurant A and drop-off at consumer B, I exclude search time at those stops. Therefore, the delivery speed aims to capture average productivity per delivery.

Total Work Hours A worker's total weekly work hours are measured by the total time logged in and active on the platform. This does not require the worker to be actively delivering food at all times.

Hourly Earning I measure a worker's average hourly earnings per week by the total earnings divided by the total work hours in each week.

C.2.4 Hometown Shock Analysis: Floods

Sample Selection

The analysis covers active delivery workers on the food delivery platform across five major Chinese cities (Shanghai, Beijing, Guangzhou, Shenzhen, and Hangzhou) between May and August 2020. Active workers are defined as those who worked for four weeks and completed at least 50 deliveries per week in May 2020, as discussed in Section C.1.3. This comprises 80,677 active workers. I create a panel data set tracking each worker from the first week of June to permanent exit, capped at the end of August if a worker stays beyond that. The final sample contains 638,810 observations at the worker-week level.

Variable Definition

Floods I use daily rainfall data from 2,423 stations (counties) across China. Floods are identified when the accumulated rainfall over two days exceeds 160mm in a each county and week, doubling the threshold for heavy rain. In most countries, heavy rain is defined as experiencing over 80 millimeters within 24 hours.

Total Work Hours A worker's total weekly work hours are measured by the total time logged in and active on the platform. This does not require the worker to be actively delivering food at all times.

C.2.5 Hometown Shock Analysis: Pandemic Lockdowns

Sample Selection

The analysis covers active delivery workers on the food delivery platform across five major Chinese cities (Shanghai, Beijing, Guangzhou, Shenzhen, and Hangzhou) in 2021. Active workers are defined as those who worked for at least four weeks and completed at least 50 deliveries per week between January and March 2021, as discussed in Section C.1.3. This comprises 139,508 active workers. I create a panel data set tracking each worker from the first week of April to permanent exit, capped at the end of 2021 if a worker stays beyond that. The final sample contains 2,387,219 observations at the worker-week level.

Variable Definition

Covid Cases I collect daily records of confirmed COVID-19 cases by city from the government websites.

Pandemic Lockdowns Pandemic lockdowns are identified at the county-week level based on whether the number of consumer orders dropped below half of the median order volume during regular periods.

Total Work Hours A worker's total weekly work hours are measured by the total time logged in and active on the platform. This does not require the worker to be actively delivering food at all times.