Can the Invisible Hand Correct Misallocations?

Evidence from Labor Recruitment Agencies in China

Linchuan Xu *

(Job Market Paper)

September 5, 2025

Abstract

Can market forces correct misallocations through endogenous institutional change? Using a newly constructed dataset covering the near universe of labor recruitment agencies in China, I find that these profit-driven intermediaries established branches to arbitrage the largest pre-existing inter-regional wage disparities. The expansion of agency networks then facilitated temporary mobility of unskilled workers and contributed to 1.35% of aggregate output gains, of which agencies appropriated 28% as profits. I develop a simple neoclassical growth model with mobility barriers and endogenous agency dynamics, where initial misallocations create pecuniary incentives for agencies who in turn promote aggregate efficiency. Quantitative results highlight the role of competition among market intermediaries in shaping the transition path of regional convergence: a larger share of the surplus gains accruing to entrepreneurs accelerates capital accumulation, generating persistent efficiency gains even after agencies exit.

*Department of Economics, London School of Economics and Political Science. Email: l.xu25@lse.ac.uk. I am grateful for Timothy Besley, Maitreesh Ghatak, and Alwyn Young for their continuous guidance and encouragement on this paper. I also thank Britta Augsburg, Clare Balboni, Robin Burgess, Mike Callen, Dave Donaldson, Alan Manning, Andrew Newman, Rachel Ngai, Imran Rasul, Christiane Szerman, and seminar participants at LSE and IFS for helpful comments and suggestions. Lingyun Zhang, Liuxian Li, and Tianyu Zhang provided excellent research assistance. I acknowledge financial support from STICERD at LSE.

1 Introduction

A central question in development economics, as famously raised by Lucas (1990), is why factors of production do not automatically flow to their most productive uses. Recent works following Hsieh and Klenow (2009) have shown that promoting factor reallocation by removing the *wedges* can, at least potentially, generate substantial productivity and output gains in many empirical contexts. But a natural Schumpeterian question follows[1]: if the potential gains are so large, shouldn't one expect profit driven economic agents, i.e., entrepreneurs or intermediaries, to arise endogenously to deliver them?

This paper explores this question by studying the expansion of labor recruitment agencies in a developing economy with well-known spatial misallocations of labor. Since the 1970s, China has undertaken gradual and selective reforms, spanning rural land markets, enterprise privatization, trade openness, and banking expansion; these are also well documented in the literature (for example, Lin, 1992; Lau et al., 2000; Young, 2000). Yet labor market frictions have remained persistent, most notably through the *hukou* restrictions[2] as well as the informational and administrative costs associated with inter-regional mobility. It is the latter type of barriers that labor recruitment agencies are well positioned to address[3]: through different local branches, they channel surplus labor from across the country into auxiliary positions in various workplaces. This in turn helps meet the growing demand for unskilled workers in e.g., coastal cities, a task that became increasingly pertinent as China entered the labor-intensive and export-led growth episode in the early 2000s. In this paper, I construct a comprehensive dataset that covers the near-universe of labor recruitment agencies in China[4]. I document how, provided with the pre-existing regional wage disparities, these labor market intermediaries expanded across space following the 2008 Labor Contract Law that recognized their legal status. I then examine what this implies for factor reallocation and aggregate efficiency along the path of economic transition.

The central argument of this paper is that initial misallocations of labor may also become arbitrage opportunities that provide pecuniary incentives for intermediaries to

[1]By definition in Schumpeter (1934), entrepreneurs "have not accumulated any kinds of goods, they have created no original means of production, but have employed existing means of production differently, more appropriately, more advantageously. They have 'carried out new combinations.' " He also argued that the surplus that is generated should fall onto the entrepreneurs as profits since they are the one "whose achievement is to introduce the looms" (page 132).

[2]The *Hukou* registration system requires migrants to seek approval from local government to change their officially registered location (Ngai et al., 2019). Furthermore, there used to be official requirements for migrants to work across regions, but these have been largely relaxed (Ebenstein and Zhao, 2015): for example, the process of obtaining a temporary residence permit has been streamlined in 2003. The major obstacle for workers to live in a city without *hukou* is associated with access to public services, healthcare, and educational resources (Meng, 2012; Tombe and Zhu, 2019).

[3]Agencies are less likely to help workers to overcome the *hukou* barrier: labor agencies might send workers to multiple places without changing their *Hukou* status. Second, agency workers are mostly less educated and unskilled, and *Hukou* relaxation for this segment of the workforce is rare. In general it is rare for temporary unskilled migrants to obtain *hukou* in richer cities (Meng, 2012).

[4]They are formally known as Labor Dispatch Agencies. See Yan (2023) for a related paper.

expand, who can in turn promote aggregate efficiency gains. To establish this argument, I first show that the expansion decision of labor recruitment agencies exhibits an U-shaped relationship with ex-ante average local wage. In other words, they establish branches to connect cities at the two extremes of the ex-ante regional wage distribution, organizing labor from the least developed areas to high wage places where they are needed the most. Furthermore, drawing on a sub-sample of labor agencies from the administrative tax survey dataset (2007-2016), I find strong positive correlation between ex-ante wage disparity and their annual profits, and the number of competing agency networks reduces profits. In other words, among cities where labor misallocations are the greatest, there are also greater (pecuniary) incentives for intermediaries to establish branches there to exploit the arbitrage opportunities.

I then provide empirical evidence on how the expansion of agencies networks affects factor reallocation and regional convergence. Using population census data in 2005 and 2010, I find that when agencies established branches between cities, the temporary flow of workers (without changing their *hukou* status) also increased. There is a particularly larger effect on increasing labor flow when the origin and destination city have higher ex-ante wage difference. Furthermore, I find the expanding agency networks have a stronger effect on temporary flow of below college, unskilled workers than for workers with at least a college degree, which is consistent with the role of labor agencies in organizing predominantly unskilled workers. These patterns of labor reallocation, in turn, contribute to regional convergence in wages and marginal product of labor (MPL). With a sample of manufacturing firms in the tax survey dataset, I calculate average wage and MPL for each city and wage (and MPL) gap for all pairs of cities. I find that 1% larger initial wage disparity is associated with 0.24% additional subsequent reduction of the wage gap, as well as 0.03% additional reduction of the MPL gap once cities are connected by agency branches. Furthermore, the extensive margin of agency connection among cities matters quantitatively more for reducing wage and MPL gaps than the intensive margins, i.e., the number of agency groups that connect the same pairs of cities. I interpret this as evidence that while competition among agency networks may bring down their profits, it may be less important for the arbitrage forces to make the law of one price to hold.

A few additional pieces of evidence support the wage convergence results. First, the expansion of labor agencies leads to similar results on wage convergence in sectors such as construction, hospitality, and logistics, which are similarly likely to employ agency workers. Second, using China Labor Dynamics Survey (2012-2018), I calculate average city wages separately for workers with and without a college degree. I find a stronger effect on regional wage convergence among less-skilled workers, which resonates with our earlier finding that agency connections promote larger temporary flows for this segment of the workforce.

The next step to establish my proposition is to use back-of-the-envelope calculations to quantify the aggregate output gains from inter-regional labor intermediation via re-

cruitment agencies. Using the labor reallocation results from population census data, I calculate predicted labor flows in the entire city system in China with and without agencies. The difference between these two scenarios then translates into a 1.35% gain in aggregate output attributable to agency expansion by 2010. Furthermore, the reduced form results on agency profits allow me to calculate how such aggregate gains from reducing misallocations are distributed between intermediaries and the real economy (i.e., firms and workers). Strikingly, I estimate that 28.7% of the output gains can be captured by agencies as profits. This likely reflects the limited competition among different agency groups, since for many of the city pairs, there is usually only one agency group that connects them.

Lastly, I take a long run perspective to explore how labor market intermediaries may endogenously emerge as the economy transitions from a highly misallocated stage towards the steady state of regional convergence. I write down a simple neoclassical model of regional growth which features mobility barriers as in Gennaioli et al. (2014). The steady state will always be efficient under this set up, which allows me to concentrate on how endogenous agency dynamics affect the transition path. This also resonates well with the classic themes in e.g., Barro et al. (1995) and Banerjee and Moll (2010), where market frictions alone may not be a handicap to long run efficiency, but the speed of convergence may be affected by e.g., access to markets or intermediaries. The core question I then explore is to what extent the private incentives of the labor market intermediaries can align with improvements in aggregate surplus during economic transition.

Calibrating the model to regional factor stock and productivities in China, I find that agencies will be active in the initial periods where regional wage disparities are the highest and hence there are larger opportunities to arbitrage. While agencies bring unskilled workers from poor to rich regions, there are also higher incentives for capital accumulation in rich regions, and lower incentives for unskilled worker migration on their own. Reallocation of unskilled workers by agencies generates substantial surplus gains compared to the no agencies counterfactual. The largest gain, i.e., 1.4%, is observed at the initial period, the magnitude of which is consistent with my back-of-the-envelope calculations. Agencies then exit as wage gap quickly shrinks, but the surplus gains remain persistent. The main idea is labor reallocations by agencies, in spite of its temporary nature, can speed up capital accumulation. Hence aggregate surplus is always on a higher path than the no agencies counterfactual until it reaches the steady state.

I then study how alternative scenarios of the agency sector lead to different transition dynamics to the steady state. I find that increasing their fixed costs and reducing competition among agencies (i.e., increasing their bargaining power) can both lead to slower transition, which mirrors the amendment to the Labor Contract Law introduced in 2013 that raises the entry barriers of agencies. Importantly, although in this model increasing competition does not affect the intensive margins of labor reallocation as well as the timing of agency exit, it still generates higher aggregate surplus gains. The core

insight is that since entrepreneurs can only capture a higher share of the surplus from labor intermediation, they will accelerate capital accumulation with higher firm profits, which in turn leads to faster convergence to the steady state. This result highlights an additional channel where distributional impact from increasing competition among market intermediaries can then have efficiency consequences through the dynamic path of factor accumulation. Furthermore, I demonstrate that facilitating temporary agency workers to join the local labor force in rich regions can also generate higher surplus gains while agencies operate for a shorter horizon. Taken together, these results reveal how agencies dynamics interact with the transition path of regional convergence, as well as how alternative structures of the agency sector determine who benefits from reducing misallocations.

Related Literature This paper is related to the following strands of literature: first, following the seminal contribution by Hsieh and Klenow (2009) and Restuccia and Rogerson (2008), recent papers in the macro-development literature on misallocations start to (i) utilize natural experiments to estimate misallocations (Sraer and Thesmar, 2023; Bau and Matray, 2023), and (ii) look at their welfare and distributional effects (Baqaee and Farhi, 2019; Davila and Schaab, 2024; Atkin et al., 2024; Ghatak and Mookherjee, 2024). However, most of these papers focus on capital market liberalization with less documentation on the labor market (with one exception being Garicano et al. (2016)). Also related is Banerjee and Moll (2010) and Moll (2014) who argue that capital accumulation by entrepreneurs themselves may overcome financial constraints over time. In this paper, I demonstrate the potential for profit-driven organizations to emerge endogenously and arbitrage away wedges. I also show its implications on the functional distribution of gains from reducing misallocations.

Secondly, this paper is also related to the ongoing discussion on the macroeconomics of alternative working arrangements. Among the current papers on agencies, productivity, and welfare, the most related paper is Bilal and Lhuillier (2025) who argue that labor outsourcing in France leads to a 1.8% increase of aggregate output but 3.1% reduction in earnings and welfare of the low skill service workers. In another related paper, Bertrand et al. (2024) show that the increasing adoption of contract labor reduces the marginal cost of labor faced by large plants because of greater unionization, therefore reduces misallocation and can account for 7.6% of the manufacturing TFP growth in India. Estefan et al. (2024) find that domestic outsourcing of core manufacturing workers leads to lower labor compensation in Mexico, and the recent ban on such outsourcing practices increases labor income share, while reducing capital investment and increasing small firm exits. This paper features the endogenous expansion of agencies, as well as imperfect competition of the agency sector, and discusses what this implies for productivity gains via factor reallocations *across* firms and regions.

This paper also resonates with the earlier theories of economic development (Lewis, 1954; Banerjee and Newman, 1998), and growth theories of conditional convergence start-

ing from Solow (1956). I demonstrate how labor agencies serve as a bridge between the modern and traditional sector and speed up regional convergence of wages; and I highlight the role of competition among market intermediaries in shaping such dynamics.

The remainder of this paper is organized as follows: the next section provides the institutional background and presents a few motivating facts constructed from different sources of datasets. Section 3 shows the core empirical results on the expansion of labor agencies, and its implications on factor reallocation and aggregate efficiency. In Section 4, I use a simple neoclassical growth model to illustrate how endogenous agency dynamics interacts with the transition path of regional convergence. Section 5 concludes.

2 Background, Data, and Motivating Facts

The classic path of development as described in Lewis (1954) often entails rising wages, as the modern sector accumulate capital and absorb surplus labor from the traditional sector, a trajectory also observed in China. Figure 1 illustrates this process. Average manufacturing wages increased fivefold between 1998 and 2016, from under 10k to over 50k RMB. At the same time, the dispersions of wages across firms narrowed: the coefficient of variation fell from 0.75 to 0.4, indicating wage convergence across producers. A decomposition of the national variance of firm wages shows that the spatial wage disparities are persistent. Roughly 20% of the overall dispersion can be explained by between-city variations, and such share has remained stable. These patterns, i.e., rising wage levels, gradual wage convergence across firms yet persistent spatial wage disparities, consist of the empirical background from which labor recruitment agencies have emerged.

2.1 The Rise of Labor Agencies in China

Labor dispatch and outsourcing, where firms hire workers through recruitment and staffing agencies, has two distinct origins in China. In the late 1970s, foreign companies entering the Chinese market were required to hire workers through government-established agency platforms (Wang, 2019). Secondly, the large scale reform of state-owned enterprises (SOEs) in the late 1990s laid off millions of workers. Many of the SOEs therefore established labor agencies to facilitate the transition of these workers into new jobs. Later on, with the rising labor costs after 2000 (as shown above), domestic enterprises, both private and state-owned, began to adopt agency workers because of the flexibility and reduced costs in hiring and firing[5]. However, labor outsourcing practices were not legally recognized until the implementation of Labor Contract Law in 2008[6], which formalized

[5]This is especially true for state-owned enterprises which are often subject to mandated quota for its formal employment. The use of agency workers enables them to bypass these restrictions.

[6]Therefore if a dispatched worker made a lawsuit case before 2008 of the company it worked for because of e.g, violation of contracts and delay of compensation payment, such cases might be identified as workers de facto directly working for the company as labor dispatch was not legally recognized by that time. In some ways, the introduction of Labor Contract Law helped to shield companies from such legal claims from the (dispatched) workers (Huang (2017)).

the registrations of Labor Dispatch Agencies (LDA) in China.

According to the 2008 Labor Contract Law, agencies were required to have registered capital of at least 0.5 million RMB, and sign fixed term contracts of more than two years with their workers, and provide monthly wages. In principle, they were allowed to send their workers to *temporary*, *auxiliary*, or *substitutable* positions. In practice, firms and agencies often exploited the ambiguity of these classifications, resulting in a much broader use of agency workers than the law originally intended[7](Liu, 2014). Firms sign contracts with agencies to obtain workers for auxiliary positions, which typically specify the number of workers and the total payment conditional on task completion. This arrangement substantially lowers firing costs, since firms can simply return workers when necessary and request replacements from the agencies, an aspect commonly emphasized in the labor economics literature on staffing agencies (e.g. Autor (2003)). However, one important feature of China's labor outsourcing system is that agencies often send workers across cities, especially from surplus labor regions to more developed areas with higher demand[8]. Such inter-regional mobility constitutes the primary channel through which agency networks can reduce spatial labor misallocations, a mechanism that I emphasize in this paper. Notably, such temporary labor mobility for work purposes are not restricted by the *hukou* system, particularly for the recent decades that I study here.

Figure 3 plot the time series of monthly registrations of new LDAs, based on the agencies dataset described below. The formalization of labor agencies in 2008 triggered a surge in registrations nationwide, with monthly entries peaking at around 750 in early 2013. It is also interesting to note that most of the post-2008 entrants are privately owned, which is the core distinction from the agencies established by the government of SOEs before 2008. Agencies are also perceived to employ a large fraction of the workforce: according to All-China Federation of Trade Unions (ACFTU), there are 37 million agency workers in 2011, which takes up 13.1% of the total workforce in China.

The rapid rise of agencies also brought numerous labor disputes that expose firms' abusive use of agency workers, which include unequal pay for equal work, overtime work, limited career advancement channels and social insurance coverage[9] (Wang, 2019). These ultimately led to an amendment of the Labor Contract Law, which took effect on July 1, 2013. Focusing specially on labor agencies, it imposed several important restrictions:

[7]The subsequent Labor Contract Law Amendment in 2013 and the No.22 Order of the Ministry of Human Resources and Social Security (MOHRSS) in China clarified these categories: " *temporary positions* refer to the positions that exist for not more than six months; *auxiliary positions* are non-main business that provide services for the positions of main businesses; and *substitutable positions* are those that can be held by substitute workers during a certain period of time when the employees of the employers who originally hold such positions are unable to work because they are undergoing full-time study, or on leave or for any other reason."

[8]This practice has sometimes triggered legal disputes due to regional differences in social benefits and compensation standards. In 2014, MOHRSS issued an additional regulation that agency workers' social benefits should be paid by either agency or firm according to local standards of the workplace.

[9]ACFTU issued a regulation in 2009 that required all LDAs to establish Trade Union and encourage agency workers to join. However, labor unions in China remain largely symbolic institutions, as they are largely politically controlled and often align with the employers in labor disputes (Tu (2017)).

for examples, agencies were required to obtain an additional certificate from the Human Resources Department of the local government, and the minimum registered capital was raised to 2 million RMB. In addition, agencies were required to apply for separate certificates if they sought to establish new subsidiaries in different cities. The amendment also clarified the boundaries of "temporary, auxiliary, or substitutable" positions, narrowing the scope for ambiguity. Later in 2014, the MOHRSS imposed a further regulation that capped the share of agency workers at 10% of total employees for all firms. In practice, however, this has been hardly enforced: firms may circumvent the cap by signing "outsourcing" contracts, which is still not legally well defined, instead of dispatch contracts, even though workers are supplied by agencies in much the same way. Therefore it is hard to distinguish between labor outsourcing, labor dispatch, and temporary employment in the Chinese context. For this reason, my analysis focus on the broader spatial expansion of agencies and its aggregate impact on wages and efficiency.

As in Figure 3, the 2013 amendment resulted in a sharp drop of new LDA registrations, which then recovered again especially in late 2017 due to e.g., the rising demand for delivery workers from e-commerce platforms. Most of the post-amendment decline of new registrations stem from local agencies that are not connected to other cities via branches or headquarters; the inter-regional agency groups are largely unaffected. Figure 2 illustrates the expansion of agency networks from 2005 to 2015: they began in relatively developed regions such as Beijing, Shanghai, and Guangdong before spreading to many other cities. For the remainder of this paper, I will focus on the inter-regional agency groups that establish branches or subsidiaries across multiple cities.

2.2 Data

In this paper, I construct a novel comprehensive dataset that covers the near universe of LDAs in China that registered between 2000 and 2019. I obtain the agencies information on entry and exit, location, ownership information, registered capital as well as business scope from one of the core business information platforms in China, *aiqicha*. I identify an enterprise as a LDA if its main business scope contains labor dispatch and or labor outsourcing, and their relevant characteristics of e.g., registered capital are consistent with the legal requirements mentioned above. I then construct the spatial network of labor agency groups where I identify headquarter as the LDA at the top of the ownership chain where the rests are local branches. I then manually classify the specializations of agencies by reading through their business scope descriptions.

I then combine the administrative tax survey dataset of firms, population census, as well as labor survey to study the impacts of agencies' spatial expansion. Following the current literature on worker migration in China (for example, Tombe and Zhu (2019), Fang and Huang (2022)), I use the 1% population *mini*-census in 2005 as well as the decennial census in 2010 to construct city level flow of workers (without changing initial *Hukou* registration location). I focus on the spatial reallocation of such *floating* popula-

tion, which according to Chen et al. (2022), takes up 70% of all migration in 2000. The administrative tax survey of firms from 2007 to 2016 by the State Taxation Administration contains a repeated cross section of firms, covering various sectors. I use this to calculate average wages (weighted by employment) and manufacturing marginal product of labor (using a Cobb-Douglas production function as in Hsieh and Klenow (2009)) for all cities. I also checked the consistency of the tax survey dataset with the more frequently used dataset of ASIE and the correlation of my main variables between the two datasets is around 0.8-0.9, suggesting that the firm data is reasonably reliable. I also identify a sub-sample of labor agencies in the tax survey dataset which provides richer information on profits and employment. Lastly, I use China Labor Dynamics Survey (CLDS) in 2012, 2014, 2016, and 2018 to study how agencies expansion affect wages by educational background. The China City Statistical Yearbook provide information on average wages, share of national GDP for each city and total *hukou* population for my counterfactual analysis. It also contains data on number of high schools and their student-teacher ratios for each city, which are used to control for the local availability of educational resources.

2.3 Motivating Facts

Even though they are perceived to employ a large fraction of the workforce, our knowledge of labor agencies in China remains scarce. Drawing on the above mentioned dataset, I am able to construct a few stylized facts consistent with the existing findings from similar contexts: for example, labor agencies in China pay on average lower wages than comparable business service firms, which is consistent with the general notion that agency jobs are often seen as "bad" jobs (Kalleberg (2011), Hatton (2011)). The dispersion of average wage among agencies is substantially higher than that in the manufacturing sector, with the left tail of the agency workers often identifying themselves as "second-class" workers. However, I now highlight the following facts that provide more information on how the agencies sector operates and motivate our further empirical investigation.

Fact 1: Agencies mainly covers business management, security, property management (including cleaning and security), as well as construction. Agency workers are less likely to have a college degree.

The bottom panel of Figure 3 classifies agencies by task specialization. Around 75% of the labor agencies operate as general purpose human resources firms, and 10% are business management firms, which may encompass a broad range of services ranging from accounting and translation to cleaning and security. I also find more specialized labor agencies: 6.5% of the labor agencies focus exclusively on security services, and 3.9% of them are providing property management services (which include security and cleaning). These patterns suggest that agency workers are more likely to be concentrated in relatively low-skill occupations. I use the CLDS dataset to compare the characteristics of agency workers with those of directly hired workers, although the sample size of the former is

fairly small due to missing observations in worker contract information. Nevertheless, only 13.2% of the surveyed agency workers obtained a college or above degree, which is significantly lower than the 21.3% of college graduates among direct hires. The fact that most of the agency workers are below college and working in low skill occupations foreshadows my later findings that the effect of agencies expansion on factor reallocation and wage convergence is stronger for below-college workers.

Fact 2: The agencies sector is characterized by a few large agency groups alongside many small ones. The largest agency group has over 80 branches across multiple cities in China, but the majority of agency groups maintain only two branches. Many of city pairs are connected via only one agency group's spatial network.

In the top panel of Figure 4, I plot the distribution of number of branches among agency groups: the majority of agency groups only have two branches, and the largest agency group has over 80 branches spanning a number of cities in China. Then in the bottom panel, I plot the frequencies of city pairs that are connected (i.e., agency group have branches in both cities) by each given number of agency groups in respectively 2005 and 2015. In 2005, only 234 pairs of cities are connected by agency groups; and by 2015, around 10% of all pairs of cities (around 50k) are connected by the spatial branch network of labor agencies. However, around 65% of the city pairs are connected by only one agency group, while there are 6 pairs of cities being connected by over 50 agency groups. This implies there is potential specialization among agencies in connecting different pairs of cities, and the substantial bargaining power these specialized agencies have when they bring workers across the cities. This is an important fact for interpreting my later results on the extent to which agencies can capture profits from reallocating workers across cities by arbitraging the wage differences.

Now I present my core empirical results on the expansion of labor agencies and its implications on labor misallocations in China.

3 Empirical Evidence

In this section, I will present empirical evidence on (i) where do labor agencies expand into given the pre-existing regional wage disparities, and (ii) how the spatial expansion of labor agencies affect inter-regional wage convergence and labor misallocations. Then I will use back-of-the-envelope calculations to translate the labor reallocation results into estimates of aggregate output gains.

3.1 The Expansion of Labor Agencies

My first task is to explore where agencies choose to establish branches to recruit (or place) workers, following the 2008 Labor Contract Law which formally recognized their legal status. The central hypothesis is that agency groups can exploit the largest arbitrage opportunities by connecting cities at opposite ends of the ex-ante regional wage

distribution, namely, low wage cities with abundant surplus labor and high wage cities with strong demand for unskilled or auxiliary workers. To test this U-shaped relationship between branch location and ex-ante city wages, I run the following specification for $branched_{rg}$ on a set of dummy variables characterizing the decile group that city r belongs to in the ex-ante (i.e., 2007) distribution of city average wage from the Chinese City Statistical Yearbook:

$$branched_{rg} = \sum_{i=1}^{i=10} \beta_i \mathbb{1}\{r \text{ in decile } i\} + distance_{rg} + \delta_g + \mathbf{X_r} + \varepsilon_{rg}, \tag{1}$$

where $branched_{rg}$ is a binary variable which equals 1 if an agency group g has ever established a branch in city r during the ten years (i.e., 2008 - 2018) following the formalization of labor agencies. δ_g is the agency group fixed effects. $distance_{rg}$ is the distance between the headquarter of group g and city r. $\mathbf{X_r}$ is a set of city characteristics such as number of high schools, and student-teacher ratio in high school. These are included to deal with potential concerns that agencies might simply expand to places with limited quantity and/or quality of educational resources.

In column 1 of Table 1, I report the coefficient estimates from a similar specification as equation (1) but only keeping dummy variables for the top and bottom 10% groups. Note that this result is conditional whether the agency group had a branch in the city before 2008 (by including the variable Branched Before 2007). I find that cities in these two extreme decile groups of ex-ante wage have a higher chance of being branched by an agency group. On the other hand, a larger geographical distance from agency headquarter city is associated with lower probability of being branched. Column 2 and 3 reports the coefficient estimates on the subsample such that no branch by the given agency group was established in the city before 2008, where I use a conditional logit model in column 3. Both columns report consistent results as in column 1 that agency groups are more likely to establish branches in cities at the two extremes of the ex-ante wage distribution.

I then report the coefficients estimates β_i for all decile groups of equation (1) in Figure 5, which uses the linear probability model with the subsample that had never been branched by the agency groups before 2007. This result visualizes the U-shaped relationship between ex-ante wage and agency expansion decision, with the first and the tenth decile group again having higher estimates than the rest of the decile groups.

We can also explore the relationship between ex-ante wage disparity and the expansion decision of agency networks at the city pairs level. I estimate the following specification:

$$connected_{mn} = \beta_1 \left|\Delta Wage_{mn}^{pre}\right| + \beta_2 distance_{mn} + \varepsilon_{mn}, \tag{2}$$

where $connected_{mn}$ is a binary variable which equals 1 if there exists an agency group that has active branches in both cities m and n during 2008 - 2018, and $|\Delta Wage_{mn}^{pre}|$ is the absolute value of the ex-ante difference of (log) average manufacturing wage between

the two cities in 2007. $distance_{mn}$ is the (log) distance between cities m and n. Applying the same idea as above, we expect $\beta_1 > 0$ because larger wage disparities across cities should incentivize agencies to establish branches to connect these cities.

Column 1 in Table 2 confirms this idea, which says one percentage increase in ex-ante wage disparity is associated with around 0.2% increase in the likelihood that the two cities are connected via any agency network in this linear probability model. It also says that a longer distance between the two cities discourage agency network to cover them. In column 2 - 3, I report the estimates of β_1 and β_2 respectively using OLS, logit and probit models, conditional on that the city pair mn was not connected by an agency group before 2007. These also provide similar results that a higher ex-ante wage disparity is associated with higher chance of the two cities being connected by an agency group. We can in principle specify $connected_{mn}$ by agency groups, and include group fixed effects into the specification: I have also studied the top 50 agency groups' expansion decision and find consistent results[10]. Taken together these results suggest that labor agencies establish their networks to connect cities at the both extremes of the ex-ante wage distribution, and arbitrage the wage disparities caused by spatial barriers across different local labor markets at different development stages.

My last task in this section is to demonstrate that such ex-ante wage disparities can be translated into pecuniary benefits for agencies to capture. To explore this, I use the subsample of labor agencies from the administrative tax survey dataset and run average profits of agency headquarters on the maximum ex-ante wage disparity among the cities that the agency group later expand into, and report the results in Table 3. In column 1, I find that on average 1% increase of the ex-ante wage disparity is associated with 1.65% increase in agency profits. Column 2 includes the average size of the agency headquarters over 2008-2016 (when it is not missing from the tax survey dataset), which brings down the coefficient between ex-ante wage disparity and agency profits to 1.09%. Interestingly, in column 3, I further include the number of agency groups that have expanded into the same city pairs with the maximum ex-ante wage disparity, which is then negatively correlated with agency profits. This is consistent with the potential competition effect among agency groups in bringing down profits. Now 1% increase in ex-ante wage disparity is correlated with 1.03% increase in agency profits. Column 4 further includes an interaction term between the number of (competing) agency groups and the ex-ante wage gap. It shows that the competition effects are predominantly among city pairs with higher ex-ante wage disparities, i.e., the marginal effect of number of competing groups on profits will reduce by 0.047% if there is 1% increase in the ex-ante wage gap. In other words, the higher pre-existing wage disparities, the more (monetary) incentives for agency groups to expand and arbitrage; therefore there will be

[10]This is because we have identified altogether around 4000 agency groups, many of these groups only establish branches in two or three cities, reflecting the limited informational or organizational capacity for these agencies to expand. Also including all agency groups will explode the size of the dataset, rendering estimation into a challenging task given the current computational capacity.

stronger competition among agency groups, which, as we will show below, will determine the distribution of efficiency gains from reducing misallocations. In the last column, I interact $|\max \Delta Wage^{pre}|$ with a binary variable which equals to 1 if the agency group is the first one who connects the two cities that exhibit such ex-ante wage gap. I find that the marginal effect of ex-ante wage disparity on profits will marginally increase by 0.2% if the agency group is the earliest connector between the cities.

3.2 The Impact of Agencies Expansion

The spatial expansion patterns reflect how ex-ante wage disparities create incentives for agencies to capture. Building on this, I now examine how agencies expansion affects factor reallocation and wage convergence, and therefore contributes to efficiency improvement. I begin by presenting the core results on temporary labor flows across cities.

3.2.1 Factor Reallocation

Utilizing the population census data in 2005 and 2010, I calculate $flow_{mnt}$, i.e., total headcount of individuals (then weighted by city population shares) who had left their *hukou* registration location m and stayed in city n for more than one year (i.e., the "floating migrants" as discussed in the migration literature). I also calculate such total flow of by their education background (below or above college), as well as the net flow among all pairs of cities. Then combined with my agency network dataset, I can explore how agency expansion facilitated worker reallocation across regions (without changing their *hukou* registration) using the following specification:

$$\begin{aligned} Flow_{mnt} &= \beta_1 Connected_{mnt} + \beta_2 Connected_{mn} \times \Delta Wage^{pre}_{nm} \\ &\quad + \beta_3 Connection_{mnt} + \beta_2 Connection_{mn} \times \Delta Wage^{pre}_{nm} + \lambda_{mn} + \gamma_t + \varepsilon_{mnt}, \end{aligned}$$

where $\Delta Wage^{pre}_{nm}$ is the ex-ante (i.e., 2005) log wage difference between city n and city m, i.e., $\Delta Wage^{pre}_{nm} = Wage^{pre}_{n} - Wage^{pre}_{m}$. We obtain the city average wage from the Chinese City Statistical Yearbook data. λ_{mn} is the *directional* city pair fixed effects, which allows us to control for time in-variant factors specific to the directions of labor flow, and γ_t is year fixed effects.

Table 4 reports the regression results. The dependent variable is gross flow for first three columns, and net flow for the last three columns. In column 1, I find that one unit increase in the ex-ante wage gap between cities n and m is associated with around 10 additional increase in the headcount of labor flow from m to n due to agency connection. If such ex-ante wage gap is zero, agency connection can also leads to 4.2 increase in such labor flow, indicating that agencies might also facilitate temporary migration due to non-economic reasons. Furthermore, the intensive margins of agency connections, i.e., the number of agency groups connecting both cities, also contribute to increased labor flow, although with a smaller magnitude than the extensive margins.

In columns 2 and 3, I report the estimation results separately for above and below college labor flow. I find that in general, there is a stronger reallocation effect among labor without a college degree: the higher is ex-ante wage difference $\Delta Wage^{pre}_{nm}$, the expansion of agency network is found to generate a larger impact on less-educated worker flows from city m to city n. For example, 1% increase in the ex-ante wage gap contributes to 6.76 headcount increase in below college worker flows from low wage to high wage city, contrasting with the 2.85 increase for above college labor. Correspondingly, I find again that the intensive margins of agency connections has a smaller impact on labor reallocation than the extensive margins, and such impact is also stronger among below college labor flow. This confirms the primary role of labor agencies in organizing unskilled surplus labor from ex-ante low to high wage cities.

In columns 4 - 6, I estimate the similar specification by replacing the dependent variable with net flow among city pairs and the city pair fixed effects now become *non-directional*. Since we calculate net flow simply by taking $Flow_{mnt} - Flow_{nmt}$, one can show mathematically that $\hat{\beta}^{net}_1$ and $\hat{\beta}^{net}_3$ will be close to zero, $\hat{\beta}^{net}_2$ will be close to $2\hat{\beta}^{gross}_2$, and $\hat{\beta}^{net}_4$ will be close to $2\hat{\beta}^{gross}_4$. Our results confirm this. Among columns 4 - 6, all my coefficients on $Connected_{mnt}$ and $Connection_{mnt}$ are not statistically different from zero, and the coefficients on the interaction terms are roughly twice as large as those obtained using gross flow data.

Furthermore, I have also find similar results for the movement of individuals due to work reasons, as well as the total flows of workers in related business service or industrial sector. To summarize, these results provide a strong indication that the expansion of agency network has facilitated labor mobility, especially among those less-educated workers, from ex-ante low wage cities to ex-ante high wage cities. The next step is to show how the reallocations of factors can be translated into changes in their prices.

3.2.2 Wage Convergence

If agencies reallocate labor from low wage to high wage cities, this will leads to wage convergence between the cities. I first use the firm level tax survey dataset to calculate city average wage among manufacturing firms, and demonstrate if cities being connected by agency groups can see their average wage gaps reduced. In fact, one should also expect that such wage disparity reduction effects are stronger among city pairs that exhibit larger ex-ante wage gap, i.e., the expansion of agency network can have a larger impact on city pairs that exhibit higher pre-existing wage gaps.

Manufacturing Wages The following two-way fixed effects specification allows us to explore this:

$$|\Delta Wage_{mnt}| = \beta_1 Connected_{mnt} + \beta_2 Connected_{mnt} \times |\Delta Wage^{pre}_{mn}| + \lambda_{mn} + \gamma_t + \varepsilon_{mnt}, \quad (3)$$

where $|\Delta Wage_{mnt}|$ is the absolute value of the (log) difference of manufacturing average wage between city m and city n at year t. $Connected_{mnt}$ is a binary variable which equals 1 if there exists an agency group that has active branches in both cities m and n in year t. λ_{mn} and γ_t are respectively city pair and year fixed effects. For wage convergence, we should expect $\beta_2 < 0$ such that the wage gap reduction effects are stronger among city pairs that exhibit larger wage gap before the formalization of labor agencies in 2008. I report the results in Table 5. Note that I use the ex-ante manufacturing average wage here and the dataset does not allow me to distinguish wages by agency and permanent workers. However, according to the Chinese accounting standards, agency workers and their compensations will also be included in their employer's accounting statement. Therefore the average wage among firms within the same city can be a reasonable representation of the workforce's average compensation.

In column 1, I include $Connected_{mnt}$ only and find that cities being connected by an agency group is associated with on average 1.5% reduction of their wage gap. I then include the interaction term in column 2, which supports the notion that the wage disparity reduction effects are stronger if the ex-ante wage gap is larger: one percent increase in $|\Delta Wage^{pre}_{mn}|$ is associated with 0.24% additional reduction of wage disparity when there exists an agency group which include both cities into its spatial network. In column 3, I further include the number of agency groups that have branches in both cities, i.e., $Connections_{mnt}$, as well as its interaction with pre-existing wage gap, and find such intensive margin of agency connections are quantitatively less important than the extensive margin of $Connected_{mnt}$ in facilitating wage convergence. Here, 1% increase in the ex-ante wage gap is only associated with 0.03% marginal reduction of the regional wage disparity when an additional agency group comes in and connects both cities. This is intuitive because arbitraging the wage differences requires only one agency linkage, and additional agency linkages are less important for the law of one price to hold. Where such intensive margin of agency connection does matter, as we mentioned earlier, is the profits that agencies can appropriate when arbitraging such wage disparities.

Event Study Specification I have shown before that agencies selectively choose cities to establish branches and form their spatial networks, but such selection problem poses challenges for our empirical strategies to causally explore the consequences of agency expansion. In parallel with the previous two-way fixed effects regressions, I now estimate the event study specification, which allows me to be more explicit on the causal interpretation of the wage convergence results:

$$|\Delta Wage_{mnt}| = \sum_{h=-\tau}^{T} \beta_h \mathbb{1}\left[D_{mnt} = h\right] + \lambda_{mn} + \gamma_t + \varepsilon_{mnt}, \tag{4}$$

where $\mathbb{1}\left[D_{mnt} = h\right]$ is the lead/lad indicator for the h-th year before/after there exists an agency network that connect the pair of cities m and n, where $D_{mnt} = t - A_{mn}$ (A_{mn}

is the time period when the city pair mn was first connected). Furthermore, to explore the regional convergence effects, we run this specification separately for city pairs that have ex-ante above median and below median wage disparities. This implies a *conditional parallel trend assumption* such that for treated and control city pairs that exhibit similar ex-ante wage disparity, they should evolve similarly over time. Furthermore, given our setting of staggered adoption and potential heterogeneity in treatment effects, I estimate the above specification using the estimator recently proposed by Callaway and Sant'Anna (2021).

I report the estimates of β_h in Figure 6, which shows that the wage disparities among pairs of cities with ex-ante above median wage disparities start to decline when cities are connected via agency network, and there are minimal changes for the below median counterparts. I find a 2% reduction after the fourth year when agency connection is established between the cities with ex-ante large wage disparities. In contrast, I find among the city pairs with ex-ante below median wage disparities, there is no significant reductions of wage disparities. Furthermore, the pre-trend coefficients are not statistically significantly from zero, which increases our confidence that the conditional parallel trend assumption is likely to hold among city pairs that exhibit similar ex-ante wage disparities.

Comparable Sectors In Table 6, I report estimates of specification (3) but the outcome variables are now the average wage disparities a few for comparable sectors that are also likely to heavily rely on agency workers, namely construction, hospitality, and logistics/warehousing. This is still based on the administrative tax survey dataset.

I find that the expansion of agency networks also bring regional wage convergence for these sectors as well: for example, 1% increase in the ex-ante wage difference is associated with respectively 0.07%, 0.07%, and 0.11% reduction in the sector specific wage gaps when agency networks connect the pairs of cities together. Furthermore, consistent with our previous results on the manufacturing sector, the intensive margins of agency connections also play a quantitatively less important role in bring down the wage disparities. Taken together these results underscore the generality of the wage convergence patterns brought by the expansion of agency networks.

Wage by Education So far I have shown that the expansion of agency network has facilitated regional convergence of *average* wage by sectors. But this may obscure important heterogeneities, and the most relevant to our analysis is how such convergence patterns differ by workers' human capital levels. I utilize the CLDS dataset in 2012, 2014, 2016, and 2018 to calculate average annual wage for workers with or without a college degree for each city, as well as the corresponding wage gaps for city pairs. This allows me to run similar specifications as in (3) but now for skill specific wage.

The results are summarized in Table 7. Similar to my findings before, I find strong wage convergence results for workers without a college degree: one percentage increase in ex-ante wage gap is associated with 0.22% additional reduction upon up agency con-

nection. In contrast, there are no statistically significant changes to the below college wage gap for city pairs that have the same ex-ante average wage. The magnitudes of the effects here are similar to what I found before in Table 5, despite that Table 5 is based on different datasets and the manufacturing sector only. Furthermore, I also find that the intensive margins of agency connections have a quantitatively less important (in terms of magnitude) and statistically insignificant role in promoting regional convergence in unskilled wages, again consistent with the previous results. However, when we move to column 3 and 4, I find smaller effects on wage convergence for the more educated workers with at least a college degree[11]: 1% increase in the ex-ante wage gap is associated with around 0.14% additional reduction of the above college wage gap. These results suggest the wage convergence, as facilitated by the expanding labor agencies who were primarily organizing the unskilled workers, might be *partial*, in the sense that labor wages at the lower spectrum of the skill distribution are converging, but less so for the more educated workers.

3.2.3 Marginal Products of Factor Inputs

I have so far demonstrated that the expansion of agency networks leads to factor reallocation across regions, which in turn translate into factor input prices, namely labor wages. My next results find similar convergence patterns for marginal product of labor among manufacturing firms.

By calculating average marginal product of labor across manufacturing firms for each city as well as the gaps of such average marginal product of labor $|\Delta MPL|$ between cities, I run the same specification as in (3) and the results are summarized in Table 8. As before, I find that the expansion of agency network leads to stronger effects of MPL convergence among those city pairs that have larger ex-ante wage gaps. 1% increase in the ex-ante wage gap is associated with 0.02% additional reduction of MPL gap across cities when they are connected by the agency groups. Furthermore, including the intensive margins of agency connections in column 3 again implies that this is quantitatively less important than the extensive margins to promote regional convergence of marginal product of labor. In column 4 and 5, I include the interaction terms of agency connections with ex-ante MPL gap, and find similar result for MPL convergence across cities: here, 1% increase in the ex-ante MPL gap is associated with around 0.05% additional reduction of the MPL gap when agency groups connect the city pairs with local branches.

Turning to the regional gaps in MPK, I follow Hsieh and Klenow (2009) to calculate the marginal product of capital (i.e., $MPK = \frac{\alpha Y}{K}$ if the production function is Cobb-Douglas) for each manufacturing firm in the tax survey dataset, and then obtain the city average (weighted by capital input) MPK as well as the MPK gap for each city pair

[11]The drop in the number of observations is due to the lack of above college observations in the CLDS dataset. We obtain similar results if we restrict the sample to city pairs where both below and above college wage gap is none-missing.

ΔMPK_{mnt}. I then estimate the similar specification as in (3) but replace the regional wage gaps by regional gaps in MPK, and obtain the results in Table 9. Unlike our previous results on regional convergence in manufacturing wages and MPL, I now find that 1% increase in the ex-ante wage gap is associated with 0.09% increase in regional MPK gap, and the estimates are statistically significant. To understand this, notice first that with Cobb-Douglas production functions, firms' marginal product of labor is simply an increasing function of capital-labor ratio $\frac{K}{L}$, and their marginal product of capital is an increasing function of the inverse of $\frac{K}{L}$. Hence assuming regional capital input doesn't change, if agency connection is reducing the regional gaps in MPL (and hence $\frac{K}{L}$), it should increase the gap in the inverse of $\frac{K}{L}$ at the same time. Intuitively, our results here can be rationalized by the idea that if there are additional frictions on capital flows across regions (so that regional capital stock hardly changes), then reallocating labor from low wage to high wage regions will mechanically enlarge the MPK gap.

3.3 Aggregate Output Gains

In this section, I use the empirical results on labor reallocation across cities and some back-of-the-envelope calculations to quantify the aggregate output gains from the expansion of labor agencies.

The core task is to first establish inter-regional labor reallocation with labor agencies, as well as the counterfactual labor flow without them. Given the results in Table 4, it then follows that when setting $Connected_{mnt} = Connection_{mnt} = 0$, one can obtain the counterfactual labor flow from city m to city n at year t simply as $\hat{Flow}_{mnt}^{counterfactual} = \hat{\lambda}_{mn} + \hat{\gamma}_t$, and the fitted value $\hat{Flow}_{mnt}$ by plugging in the agency network $\{Connection_{mn}\}_t$ for predicted labor flows with labor agencies.

I contrast the total labor flow with agencies in 2010 with the counterfactual labor flow in the $N \times N$ city system (where N is the total number of cities) in Figure 7. The horizontal (vertical) axis shows origin (destination) cities ranked from the highest to the lowest in the ex-ante average wage distribution. Hence the i-th column and j-th row pixel shows the total number of labor from from the i-th rank city to the $N + 1 - j$-th rank city in 2010 respectively with labor agencies (right) and in the counterfactual scenario (left). This also implies all pixels below the diagonal show labor flow from lower wage cities to higher wage cities and vice versa for the above diagonal pixels. The core message from this comparison is that labor agencies reinforce the underlying worker flows from low wage to high wage cities. First, one can see substantial flows of workers from cities across the ex-ante wage distribution to the highest wage cities, and all the negative flows from low to high wage cities in the counterfactual scenarios get eliminated in the scenario with labor agencies. Second, there are negative labor flows in the top left corner of city system with agencies, which implies that with labor agencies, it is even less likely for workers to flow reversely, i.e., from ex-ante high wage to low wage cities.

In Figure 8, I show such fitted and counterfactual labor flows separately for college

labor and below college labor. Similar patterns emerge: for both college and below college population, I find that labor agencies reinforce the flows to the highest wage cities, and all the (counterfactual) negative flows from low to high wage cities get eliminated under labor agencies. Furthermore, compared to college labor, I find that labor agencies lead to a stronger negative effects among below college labor for the reverse labor flow (i.e., from high to low wage cities). This is consistent with the predominant role of labor agencies in organizing unskilled labor across regions.

Figure 9 uses the regression results on net flows and construct the fitted and counterfactual flows in the $\frac{N\times(N-1)}{2}$ city system. Again I find that agencies promote (positive net) labor flows from cities across the ex-ante wage distribution to the highest wage cities (i.e., the bottom rows in the diagram). There is also stronger outflow of workers from the lowest wage cities to higher wage cities (i.e., the right side of the diagram).

With these results on labor reallocation across cities, I can now quantify the aggregate output gains from the expansion of labor agencies. I assume the regional aggregate production function to be the standard constant returns to scale Cobb-Douglas function: (i) $Y = AK^{\alpha}L^{1-\alpha}$ where L is the total labor in the region, and I also allow for (ii) $Y = AK^{\alpha}(L_U)^{\beta}(L_S)^{1-\alpha-\beta}$ where L_U and L_S are respectively unskilled (below college) and skilled (college) labor. I use the above fitted and counterfactual labor flows across cities and the population distribution from the census data to calculate the total labor supply in each city in 2010. Suppose that regional capital stock and productivity are fixed, then using $Y = AK^{\alpha}L^{1-\alpha}$, I obtain[12]

$$\frac{\sum_n Y_n - \sum_n Y_n^c}{\sum_n Y_n^c} = \sum_n \frac{Y_n^c}{\sum_i Y_i^c}\left(\frac{Y_n - Y_n^c}{Y_n^c}\right) = \sum_n \frac{Y_n^c}{\sum_i Y_i^c}\left[\left(\frac{L_n}{L_n^c}\right)^{1-\alpha} - 1\right], \tag{5}$$

where Y_n^c is the counterfactual output level of city n without agencies, L_n^c is the counterfactual labor stock without agencies and L_n is the labor stock with agencies. The above equation says that assuming fixed capital stock and productivity, aggregate output gains due to agencies can simply be calculated as a weighted average of the labor gains, weighted by the counterfactual output share for each city $\frac{Y_n^c}{\sum_i Y_i^c}$.

To measure the city output share, I calculate each city's share of national GDP in 2005 using the Chinese City Statistical Yearbook. I take $\alpha = 0.3$ which follows from (i) the capital output ratio of manufacturing firms in the administrative tax survey dataset and (ii) running a regression of city level industrial output on industrial fixed assets and total labor input and obtaining the coefficient of 0.29 for the capital input[13]. To determine the unskilled labor share β, I find that the payment for managers takes up around 25% of total labor payment among manufacturing firms in the tax survey dataset. Using this as a proxy for the share of unskilled labor over total labor share $\beta/(1-\alpha)$, I obtain a

[12]It is easy to obtain similar result using $Y = AK^{\alpha}(L_U)^{\beta}(L_S)^{1-\alpha-\beta}$.

[13]Specifically, I obtain $\beta_1 = 0.29$ from the two-way fixed effects model $y_{nt} = \beta_1 K_{nt} + \beta_2 L_{nt} + \nu_n + \gamma_t + \varepsilon_{nt}$.

benchmark value of $\beta = 0.2$, but will also consider alternative levels of β and explore how they affect our aggregate gains calculations.

I can now calculate the aggregate output gains with labor agencies using equation (5). I report the aggregate output gains by labor agencies *relative* to the no agency counterfactual in Table 10. Using total labor flows, I find that the existence of labor agencies across cities in 2010 can contribute to 1.35% aggregate output gains in China relative to the counterfactual scenario without agencies. I also find a 1.21-1.30% aggregate gains using the skill specific production function with different values of β, where higher values of β yields higher estimates for the relative aggregate gains. This is intuitive because I find a stronger effect of labor agencies to move unskilled workers across cities, and as unskilled labor becomes more important in the production function (i.e., higher β), one can expect agencies to generate a larger aggregate impact.

Moreover, one can have a back-of-the-envelope estimate of how much of such aggregate output gains are captured by labor agencies as profits. To calculate agency profits in reduced-form, I first use the subsample of agencies from the administrative tax survey dataset to run a regression of agency headquarter profits per worker on ex-ante maximum wage difference between cities that each agency group expands into[14]. I then obtain that one unit[15] increase in the ex-ante wage gap translates into 0.133 unit increase in agency profits per worker, but one extra competing agency group between the corresponding city pair reduces profits by 0.009 unit. We can then calculate the per worker profits for a given agency group operating between cities m and n as $0.133 \times \Delta Wage^{pre}_{mn} - 0.009 Connection_{mn}$. Multiplying this by the total number of additional labor flow due to agencies $\hat{Flow}_{mn} - \hat{Flow}^{counterfactual}_{mn}$, as well as the number of agency groups that connect the cities $Connection_{mn}$, I can obtain total profits of the agency sector when reallocating labor between cities m and n. Aggregating such agency profits over the entire $N \times N$ city system, I find that total agency profits takes up around 30% of aggregate output gains[16], leaving the remaining 70% of them to firms and workers. The fact that almost a third of the surplus gains from labor reallocation are captured by the intermediaries is however consistent with the previous empirical facts that there might be limited competition among agencies (i.e., many of the city pairs are connected via only one agency group, and the number of competing agencies is associated with lower profits).

I can also run counterfactual analysis for aggregate output gains by incorporating the expansion of labor agencies after 2010. We keep the no agency counterfactuals the same

[14]Specifically, I run

$$Profits_g = \beta_1 \left|\max \Delta Wage^{pre}_g\right| + \beta_2 Number\ of\ Agency\ Groups_g + \epsilon_g,$$

where $|\max \Delta Wage^{pre}|$ is the maximum ex-ante (i.e., 2007) wage difference among cities that the agency group later expanded into during 2008-2018. Number of Agency Groups is the total number of agency groups that had established branches in the same cities that exhibit such maximum wage gap.

[15]The units for both profits and city wage gap are 1k RMB.

[16]We take the total GDP in 2005 from the Chinese City Statistical Yearbook as the no agency output, hence the aggregate output gain is simply 1.5% of such total GDP.

as in 2010, i.e., $\hat{Flow}_{mnt}^{counterfactual} = \hat{\lambda}_{mn} + \hat{\gamma}_t$, but I calculate the predicted values of flows with agencies by taking the future expansion of agency network into account. Therefore the core question I ask in this exercise is how much additional aggregate output gains can be delivered in 2010, if one is to have more agencies across cities as in 2011 - 2018. I report these relative output gains calculations in Table 11. Labor agencies continue to expand to more cities after 2010, and as one would expect, there will be larger gains in aggregate output, which ranges from around 8% to 20% given the agencies network respectively in 2011 and 2018. With the skill specific aggregate production function, I find similar levels of output gains: for example, using the benchmark value of $\beta = 0.2$, I find that the agencies network as in 2018 can deliver 28% aggregate output gains, and these aggregate gains are all increasing as agencies continue to expand over time.

The fact that labor market intermediaries can appropriate a third of the aggregate gains as profits speaks to the importance of agencies competition in determining the functional distribution of surplus gains among economic groups. This also implies that one need to account for the endogenous agency dynamics into the process of reducing misallocations, and, importantly, how it interacts with the dynamic path of regional convergence. A series of interesting questions emerge: for example, how will agencies affect the dynamics of labor transition and capital accumulation in the long run, how much can agencies benefit along the transition path, and how persistent the aggregate gains can be even after agencies exit. I turn to these issues now.

4 A Neoclassical Growth Model with Labor Agencies

In this section, I develop a neoclassical growth model with spatial mobility frictions, and endogenous entry and exit of labor market intermediaries that arbitrage regional wage disparities. With a longer run perspective, the core question I ask is (1) to what extent can the private incentives of labor agencies align with aggregate efficiency improvement, i.e., how far labor agencies are able to correct misallocations, speed up regional convergence and generate dynamic surplus gains; (2) what determines the distribution of surplus gains along the transition path of regional convergence.

4.1 The Model

Let us consider two regions $r \in \{R, P\}$ which differ in their productivity parameters A_r (with $A_R > A_P$). There are two types of agents: workers and landlords/entrepreneurs. Worker's skill level can be $s \in \{U, S\}$. The total stock of type s labor in the economy is $\overline{H}_s$. The fixed amount of land owned by the entrepreneurs in each region is $\overline{L}_r$[17]. Each

[17]We introduce this to guarantee an interior solution for the steady state with constant capital stock in both regions.

region has a Cobb-Douglas production technology given by

$$y_r = A_r \left(K_r\right)^{\alpha} \left(H_U\right)^{\beta} \left(H_S\right)^{\gamma} \overline{L}_r^{1-\alpha-\beta-\gamma}, \tag{6}$$

where K_r is physical capital, H_s is labor input of skill type s, and the technology has constant returns to scale. Let us assume for now that physical capital is immobile, as we interpret physical capital as buildings and structures as in Kleinman et al. (2023). This implies that entrepreneurs rely on their own capital stock for production. We also assume that firms can only hire labor locally, so there will be region specific wage for each skill level, w_r^s. Therefore for each period given their capital stock, entrepreneurs will choose labor inputs to maximize profits. Entrepreneurial profit can therefore be written as

$$\begin{aligned} \pi_r\left(K_r\right) &= \max_{H_U, H_S} A_r K_r^{\alpha} \left(H_U\right)^{\beta} \left(H_S\right)^{\gamma} \overline{L}_r^{1-\alpha-\beta-\gamma} - w_r^U H_U - w_r^S H_S \\ &= (1-\beta-\gamma) A_r K_r^{\alpha} \left(H_U^*\right)^{\beta} \left(H_S^*\right)^{\gamma} \overline{L}_r^{1-\alpha-\beta-\gamma} \end{aligned}$$

Capital Accumulation The inter-temporal consumption-investment decision of entrepreneurs can be formulated as follows: taking initial capital stock K_{0r} as given, entrepreneurs in region r choose the consumption path $\{c_{rt}\}_{t\geq 0}$ to maximize lifetime utility

$$\sum_{t=\infty}^{\infty} \rho^t \frac{c_{rt}^{1-\theta} - 1}{1-\theta} \tag{7}$$

subject to the budget constraint:

$$K_{r,t+1} = \pi_r\left(K_{rt}\right) + (1-\delta) K_{rt} - c_{rt} \tag{8}$$

where δ is capital depreciation rate. One can therefore derive the Euler equation as

$$\begin{aligned} \left(\frac{c_{r,t+1}}{c_{rt}}\right)^{\theta} &= \rho\left[\pi_{rK}(K_{r,t+1}) + 1 - \delta\right] \\ &= \rho\left[\alpha A_r K_{r,t+1}^{\alpha-1} \left(H_{Ur,t+1}^*\right)^{\beta} \left(H_{Sr,t+1}^*\right)^{\gamma} \overline{L}_r^{1-\alpha-\beta-\gamma} + 1 - \delta\right] \end{aligned}$$

where the second equality follows from envelop theorem.

Imperfect Labor Mobility For simplicity I assume workers cannot save; therefore for each period they will migrate to whichever region that gives them higher wages and immediately consume their earnings, subject to the spatial mobility barriers implicitly modeled below. I follow Gennaioli et al. (2014) for a tractable way to introduce mobility barriers into the neoclassical growth model: if there is infinite mobility costs, then each region's labor stock at period t, H_{srt}, simply equals to the level at the previous period, i.e., H_{srt-1}, i.e., workers just stay at where they were in the previous period. On the other hand, if there is perfect mobility, workers will flow freely so that the skill specific

wage rate at period t will equalize across regions. If this is the case, the supply of type s workers at period t will simply be proportional to regional capital stock and productivity parameters, i.e.,

$$\frac{\hat{H}_{URt}}{\hat{H}_{UPt}} = \frac{\hat{H}_{SRt}}{\hat{H}_{SPt}} = \left(\frac{A_R K_{Rt}^{\alpha} \overline{L}_R^{1-\alpha-\beta-\gamma}}{A_P K_{Pt}^{\alpha} \overline{L}_P^{1-\alpha-\beta-\gamma}}\right)^{\frac{1}{1-\beta-\gamma}} \tag{9}$$

Therefore the (hypothetical) allocation of type s labor for region r will simply become

$$\hat{H}_{srt} = \frac{\left(A_r K_{rt}^{\alpha} \overline{L}_r^{1-\alpha-\beta-\gamma}\right)^{\frac{1}{1-\beta-\gamma}}}{\sum_q \left(A_q K_{q,t}^{\alpha} \overline{L}_q^{1-\alpha-\beta-\gamma}\right)^{\frac{1}{1-\beta-\gamma}}} \overline{H}_s \equiv Q_{rt} \overline{H}_s. \tag{10}$$

Then for intermediate level of mobility barriers, I assume that the total stock of type s worker in region r is

$$H_{srt} = \nu_{st} \left(H_{srt-1}\right)^{\tau_s} \left(\hat{H}_{srt}\right)^{1-\tau_s}, \tag{11}$$

where the normalization factor is $\nu_{st} \equiv \frac{\overline{H}_s}{\sum_q (H_{sqt-1})^{\tau_s} \left(\hat{H}_{sqt}\right)^{1-\tau_s}}$, which makes sure that $\sum_q H_{sqt} = \overline{H}_s$ for all t. One can think of the role of spatial mobility barriers in this model as delaying the adjustment to the most efficient allocations. Furthermore, I allow for skill specific degree of mobility barriers τ_s across the two regions, which captures the fact that better educated workers may have more information and/or connections with job opportunities in rich regions.

Before I introduce labor agencies, I characterize the steady state of the economy and summarize the results here:

Proposition 1 *The steady state of the economy is unique and efficient such that despite the mobility barriers, the long run capital and labor allocations are proportional to regional productivity parameters:*

$$\frac{K_R^*}{K_P^*} = \frac{H_{UR}^*}{H_{UP}^*} = \frac{H_{SR}^*}{H_{SP}^*} = \left(\frac{A_R}{A_P}\right)^{\frac{1}{1-\alpha-\beta-\gamma}} \frac{L_R}{L_P}. \tag{12}$$

This result parallels the previous arguments such that market frictions may not be handicap to long run outcomes, as agents can gradually accumulate physical or human capital overtime to overcome these constraints and achieve efficient outcomes (see, for example, Banerjee and Moll (2010), Ghatak (2015)). This allows me to concentrate only on the role of market intermediation in speeding up the convergence process (Barro et al. (1995)). I now model how labor recruitment agencies can impact the convergence process.

Labor Agencies I introduce the technology of labor agencies such that they can **temporarily** move unskilled workers across regions: suppose at period t, labor agencies decide to hire H_{at} of type U workers in region P at the competitive local wage w_{Pt}^U, and then move them to region R so that they will temporarily join the workforce at region R;

then at the end of period t, with probability q, these workers can successfully settle down in region R and become a permanent worker there; with probability $1-q$, agency workers will return back to region P. One can think of q capturing the *Hukou* quota for migrant workers, or more generally the barriers for agency workers to migrate to new regions even though they have temporarily worked there. The assumption of competitive local labor market for the unskilled workers is plausible because I consider the type of environment as in Lewis (1954) where there is a large amount of surplus labor in regions at early stages of economic development. Moreover, as I mentioned in Section 2.1, the "right-to-manage" bargaining models of labor unions and/or agencies (Harris (1982), Holmes and Snider (2011), Felix and Wong (2024)) seem less plausible given the dis-functioning of unions in the empirical context of this paper.

This technology set up implies that the regional allocation of workers under the (hypothetical) frictionless mobility case, i.e., equation (9) will become

$$\frac{\hat{H}_{UR,t}+H_{a,t}}{\hat{H}_{UP,t}-H_{a,t}}=\frac{\hat{H}_{SR,t}}{\hat{H}_{SP,t}}=\left(\frac{A_R K_{R,t}^{\alpha}\overline{L}_R^{1-\alpha-\beta-\gamma}}{A_P K_{P,t}^{\alpha}\overline{L}_P^{1-\alpha-\beta-\gamma}}\right)^{\frac{1}{1-\beta-\gamma}} \tag{13}$$

This says that the labor **available** to the two regions will be proportional to capital stock and productivity parameter. It then follows that $\hat{H}_{UR,t}=Q_{R,t}\overline{H}_s-H_{a,t}$ and $\hat{H}_{UP,t}=Q_{P,t}\overline{H}_s+H_{a,t}$. Intuitively, if labor agencies can reallocate unskilled workers across regions for production, this reduces the demand for workers who migrate by themselves. Furthermore, the temporariness of such labor allocation via agencies implies that when the agency ceases to be active, the $(1-q)H_a$ amount of agency workers will have to immediately go back to their original region and are again subject to the same mobility dynamics specified before. The dynamics of labor mobility now becomes

$$H_{srt}=\nu_{st}\left(H_{srt-1}+qH_{at-1}\right)^{\tau_s}\left(\hat{H}_{srt}\right)^{1-\tau_s}, \tag{14}$$

Labor agencies supply unskilled workers to entrepreneurs in region R at price p_{at}. I assume that labor agencies incur a fixed cost F every period in order to operate. For now let us temporarily drop the t subscript here to model the static profit optimization of entrepreneurs and agencies. The profit of labor agencies can be written as

$$W(p_a,H_a)=\left(p_a-w_P^U\right)H_a-\phi\left(H_a\right)-F, \tag{15}$$

where $\phi\left(H_a\right)$ is a convex cost function of training and transporting H_a unskilled workers from region P to region R. Entrepreneurs in region R's static profit now becomes

$$\tilde{\pi}_R\left(K,p_a,H_a\right)=\max_{H_U,H_S}A_R K_R^{\alpha}\left(H_U+H_a\right)^{\beta}\left(H_S\right)^{\gamma}L_R^{1-\alpha-\beta-\gamma}-w_R^U H_U-w_R^S H_S-p_a H_a \tag{16}$$

Note that because agency workers and low skilled local workers are perfect substitute,

for entrepreneurs to employ all agency workers, we need

$$p_a \leq w_R^U. \tag{17}$$

The firm and agency then bargain over a labor contract which specifies $\{p_a, H_a\}$:

$$\max_{p_a \leq w_R^U, H_a \leq H_{UP}} W(p_a, H_a)^{\chi} \left[\tilde{\pi}_R\left(K, p_a, H_a\right) - \pi_R\left(K\right)\right]^{1-\chi}, \tag{18}$$

where χ represents the bargaining power of agencies relative to entrepreneurs, $\pi_R\left(K\right)$ is the entrepreneurial profits without agencies, and I set the outside option of agencies to be zero. One can also find similar bargaining structure between firms and agencies in recent models of labor outsourcing (for example, Felix and Wong (2024)). I can then solve the optimal $\{p_a^*, H_a^*\}$ and summarize the results as follows:

Proposition 2 *The optimal amount of agency workers is characterized by*

$$w_R^U - w_P^U = \phi'(H_a^*). \tag{19}$$

The price that firm pay to agency is $p_a^* = \min\{w_R^U, \hat{p}\}$*, where* $\hat{p}$ *is characterized by*

$$(1-\chi)W(p_a, H_a^*) = \chi\left[\tilde{\pi}_R\left(K, p_a, H_a^*\right) - \pi_R\left(K\right)\right]. \tag{20}$$

I provide the proof for this proposition in the appendix. Equation (19) suggests that the optimal amount of agency workers is entirely driven by the regional wage gap, which follows from the fact that agency and local unskilled workers are perfect substitutes, so that the marginal product of unskilled labor should be equal to w_R^U (if the firm is going to hire any local unskilled labor). Hence the bargaining power of agencies doesn't affect the intensive margins of labor reallocation here. However it does affect how the price of agency workers that firms pay, and in turn the distribution of surplus among agencies and firms. Nash bargaining in (20) implies that the profits that agencies obtain will take up χ share of the total surplus generated from labor market intermediation: $\tilde{\pi}_R\left(K, p_a^*\right) - \pi_R\left(K\right) + W(p_a^*)$.

Lastly, for the inter-temporal consumption-investment decision, the budget constraint of the entrepreneurs will be the same as equation (8) except that we replace $\pi_r(K)$ by $\tilde{\pi}_R(K, p_a, H_a^*)$ for region R and $\tilde{\pi}_P(K)$ for region P where

$$\tilde{\pi}_P(K) = (1-\beta-\gamma)A_P K^{\alpha}\left(H_P^U - H_a\right)^{\beta}\left(H_P^S\right)^{\gamma} L_P^{1-\alpha-\beta-\gamma},$$

i.e., the unskilled labor supply now becomes $H_P^U - H_a^*$. Note that $\tilde{\pi}_r = \pi_r$ when labor agencies are not active. The Euler equation is

$$\left(\frac{c_{r,t+1}}{c_{rt}}\right)^{\theta} = \rho\left[\tilde{\pi}_{rK}(K_{r,t+1}) + 1 - \delta\right]$$

Moreover one can intuitively see how intermediation affects the speed of convergence: they bring labor across regions to arbitrage the wage disparities more quickly than workers' mobility by themselves; they also brings efficiency gains to entrepreneurs and indirectly speed up the capital accumulation process by raising marginal products of capital (since capital and labor are complements). Importantly, this implies that even if agencies exit because of limited arbitrage opportunities, we still expect the aggregate output gains along the transition path to be persistent because the rich region now has a higher stock of capital compared to the no agencies counterfactual. Furthermore, although market competition among agencies do not affect the optimal choice of agency workers (under the assumption of perfect substitutability between agency and local workers), it might affect the extensive margin of labor reallocation, i.e., whether agencies will be active or not. It is easy to see from equation (20) that p_a^* is (at least weakly) increasing in χ, i.e., higher bargaining power of agencies implies higher price that they can charge, which in turn raises their profits and the incentives to operate. Furthermore, when agencies exit because of limited arbitrage opportunities, agency workers temporarily working in region R will have to go back to region P with probability $1-q$, leading to a sharp drop in unskilled labor available to region R, which will then slowly recover subject to mobility barriers.

Also one can show that the economy will converge to the same efficient steady state as characterized in Proposition 1. The core logic is that since marginal product of capital and labor input will be eventually equalized across regions, there is no room for agencies to arbitrage so in steady state we must have $W < 0$:

Proposition 3 *Labor agencies will not be active in steady state.*

I provide the proof of this proposition in the appendix. This result suggests that labor agencies will only affect the transition dynamics to the steady state, and I will now turn to this with the quantitative results. I will first describe how parameter values are calibrated to empirical moments.

4.2 Calibration

For the factor share parameters in the production function, I take capital share α to be 0.3 to match the capital output ratio of manufacturing firms in the administrative tax survey dataset. Consistent with Hsieh and Klenow (2009), I take the labor share $\beta+\gamma$ to be 0.5. Then to determine the skilled labor share γ, I find that the payment for managers takes up around 75% of total labor payment among manufacturing firms in the tax survey dataset. Using this as a proxy for the share of skilled labor over total labor share $\gamma/(\beta+\gamma)$, I obtain $\gamma = 0.37$ and $\beta = 0.13$.

For the regional productivity parameters A_R and A_P, I run a regression of output of manufacturing firms on their capital and labor payment (including taxes) using the tax survey dataset and collect residuals to proxy firm productivities. I then take the average

of these residuals among firms respectively in cities with above and below median average manufacturing wages. This therefore gives average productivities for cities with above and below median wage, which correspond to region R and region P in the model. I normalize $A_R = 1$, and obtain $A_P = 0.85$.

I take the total stock of labor $\overline{H}_s = 1$ for both skill types. To determine the initial regional labor stock, I use the population census in 2005 and find that the above median wage cities (specified before) takes up around two third of workers with a college or above degree in the country. I therefore take $H_{SR0} = 0.7$ and $H_{SP0} = 0.3$. I then assume that region P takes up a larger share of unskilled workers in the initial period $H_{UP0} = 0.3$ than region R. I assume that the initial capital stock in both regions are 50% of their steady state levels. I will simulate the model over time periods $t \in \{0, 1, ..., 150\}$, where I assume that the economy will converge to the steady state at the 150th period.

I assume the following simple quadratic form for the cost function of agencies:

$$\phi(H) = \frac{1}{2}\phi H^2. \tag{21}$$

To calibrate ϕ, I use the previous results that conditional on the number of competing agency groups, one percentage increase in ex-ante wage difference is associated with around 1% increase in agency profits. Since in the model I have for agency profits $\frac{\partial W}{\partial p_a - w_P^U} = H_a^* = \frac{p_a - w_P^U}{\phi}$. If the agency has no competitor, the highest level of price p_a will simply be w_R^U; then wage difference $w_R^U - w_P^U$ will directly translate into profits with weight $\frac{1}{\phi}$. We can therefore take $\phi = 1$.

I assume capital depreciation rate $\delta = 5\%$, take $\theta = 2$ to capture a reasonable level of risk aversion in preferences, and take $\rho = 0.98$ for the patience parameter. For the results below, I assume the level of mobility barriers such that $\tau_U = 0.95$ and $\tau_S = 0.85$ to guarantee that it takes times for regions to converge to the steady state. This also implies that skilled workers face lower mobility barriers than the unskilled workers, as they tend to have better information or they are subject to relaxed *hukou* barriers of movement to richer cities. I take $q = 0.06$ as a benchmark value to match the fact that the probability of successful applications among (floating) migrants for obtaining *hukou* in Beijing is around 5.6% in 2019. In the policy counterfactual exercises, I also consider a higher level of q to explore how this affects the transition dynamics.

Lastly, in the benchmark result below, I consider the bargaining power parameter $\chi = 0.5$, which says that agency will capture 50% of the total surplus from intermediation. But I will later consider alternative cases of $\chi \in \{0.2, 0.5, 0.8\}$ to capture different degrees of competition among agencies where the quantitative results may vary. I take the benchmark fixed cost $F = 10^{-3}$ and will raise it to 10^{-2} in counterfactual scenarios.

4.3 Benchmark Results

Focusing on the intermediate level of bargaining power $\chi = 0.5$, Figure 10 presents the transition dynamics to the steady state for some of the core variables. The top panel respectively shows capital accumulation and labor transition among the two regions. Region R is absorbing unskilled workers from region P while accumulating capital, which is the core process of development as in Lewis (1954). Along the transition path, skilled workers show quantitatively minimal movements across regions. Furthermore, in the initial periods, capital stock in region P also increases because of its relatively higher marginal returns as workers have not yet moved to region P. It falls later on because workers are transitioning to region P and converges to the steady state level proportional to its productivity parameter.

The bottom panel of Figure 10 shows how labor agencies behave in the transition dynamics. According to the left diagram, agencies are active for the first 47 periods and shut down thereafter. As I will show below, there is larger regional (unskilled) wage disparity that they can arbitrage for the initial periods, but such arbitrage opportunities shrink over time so agencies will eventually shut down. The right diagram shows how the number of agency workers H_a and price that firms pay to agencies p_a evolves along the dynamic path. In the first 20 periods, agencies moves around 10-15% of the unskilled workers from region P to region R, and the intermediate level of bargaining power $\chi = 0.5$ allows the agency to charge a higher price with p_a being around 0.3 for this. Importantly, in this case the constraint $p_a \leq w_R^U$ is binding, i.e., the agency is charging the highest price it can so that the firm is indifferent between hiring local and agency workers. However, both H_a and p_a gradually decreases as unskilled wage in region R (P) decreases (rises) and arbitrage opportunities fade away.

I show how unskilled wages evolves along the transition path in the top left panel of Figure 11, and highlight the role of labor agencies by making a contrast with a counterfactual dynamic path where no agencies are allowed to operate. The dashed curves are such counterfactual path for the two regions: there is large wage disparity in the initial periods, but unskilled wages gradually converge to each other as labor slowly moves across regions subject to the mobility barriers. When one allows labor agencies to operate (which I show using the solid curves), they drastically bring down unskilled wages in region R by introducing H_a amount of agency workers, and they also raises region P's unskilled wages. This is consistent with the empirical results that the expansion of agency network is associated with reduction in unskilled wage disparities among pairs of cities. The unskilled wage gap between the two regions continues to decrease while agencies are active and workers move across regions subject to mobility barriers. However, when agencies shut down at the 47th period because of limited profits they can capture from the arbitrage opportunities, $1 - q$ of agency workers who are brought to region R temporarily by agencies will now have to go back to region P. Such temporariness of labor allocation via agencies leads to a sharp spike (drop) in region R (P)'s unskilled

wage. In other words, the exit of agencies will then temporarily enlarge the unskilled wage gap, which will then shrink over time as we are back to the case where workers cannot rely on agencies to transport them across regions. This highlights the core tradeoff in how agencies shape the dynamic path of regional wage convergence: profit-driven agencies can promote a drastic drop in regional wage disparities by exploiting the arbitrage opportunities in the initial periods, but the temporary labor reallocation is no perfect substitute for costly migration, which then implies later widening of such wage gap once the arbitrage opportunities diminish and agencies exit.

I also contrast the dynamic paths of capital accumulation and labor transition in the bottom panels of Figure 11 with those in the counterfactual scenario when agencies are never allowed to operate. The left diagram suggests that compared to the counterfactual scenario, the additional agency workers H_a available to region R raises the marginal returns of capital and therefore entrepreneurs' incentives to accumulate capital, consistent with our qualitative analysis before. This speaks to the wider development strategies that stress on keeping labor wages in the advanced sector low to speed up capital accumulation in the early stages of economic transition (see, for example, a recent argument by Itskhoki and Moll (2019)). On the other hand, there are reduced incentives for saving among entrepreneurs in region P as they lose H_a of unskilled workers. In the right diagram, I show the dynamics of labor stock in region R, where the dashed blue curve represents unskilled labor in the counterfactual scenario, and the dotted blue curve is the stock of unskilled labor H_{UR} in our benchmark case where agencies are allowed to operate. The solid blue curve represents the total supply of unskilled workers *available* to region R, i.e., $H_{UR} + H_a$. Again, agencies generate a drastic increase in the supply of unskilled labor to region R in the initial periods; but the exit of labor agencies in the 47th period leads to a sudden drop in the total supply of unskilled workers in R, which then gradually converge to the same steady state as in the no agency counterfactual. On the other hand, labor allocations via agencies also raises the marginal product of skilled workers in the initial periods, so H_{SR} (solid green curve) is higher than the no agency counterfactual (dashed green curve), but such gap diminishes later on as agency exit and the economy converges to the steady state. It is also useful to note that the overall impact of agencies on skilled workers' mobility is less pronounced that that for the unskilled workers, consistent with the empirical results that the expansion of agency networks has a larger impact on movement of for workers without a college degree.

More importantly, the results on how the entry and exit of agencies shapes labor reallocation and the dynamic path of wage convergence suggest one needs to carefully evaluate the overall surplus gains from labor market intermediaries. I define aggregate surplus of the economy as

$$\sum_r y_r - \frac{1}{2}\phi H_a^2, \tag{22}$$

i.e., the aggregate output minus the cost of moving agency workers across regions. I plot

the dynamics of aggregate surplus gains in the top right panel of Figure 11. Agencies promote surplus gains (compared to the no agency counterfactual) in the initial periods, i.e., an 1.4% surplus gains, and the largest gain can be found in the first period when agencies reallocate the highest amount of workers across regions because of the largest arbitrage opportunities. But such surplus gains diminish over time with reduced wage disparities and workers gradually move across regions. Importantly, the aggregate surplus gains are persistent even after agencies exit at the 47th period: because of temporary labor reallocations by agencies, region R gets on faster track of capital accumulation and permanent labor transition, which makes the path of aggregate surplus always higher than that without agencies.

Taken together, these results demonstrate that initial misallocations can have dynamic implications: the arbitrage opportunities from large wage disparities due to labor mobility barriers incentivize agencies to bring workers from low wage to high wage regions and earn profits. This immediately brings down regional wage gap and generate large surplus gains in the initial periods. Because of the impact of temporary labor reallocations on capital accumulation and permanent labor transition, such surplus gains are also likely to be persistent even after agencies exit. I will now discuss how different scenarios of the agency sector affect the dynamic path of regional convergence and the distribution of surplus gains.

4.4 Policy Counterfactuals

I explore how different levels of fixed cost of operating agencies (F), bargaining power of agencies (χ), as well as the probability of transition of agency workers to permanent local worker (q) affect the gains from labor reallocation via agencies. The empirical background for the policy counterfactuals in this section is the amendment to Labor Contract Law in 2013 that increases the entry barriers of labor dispatch agencies in China, which includes increased minimum capital requirements, additional certificates agencies need to obtain from local government, etc. This corresponds to two parameters in the model: first, a direct increase in the fix costs F for agencies to operate; second, the Law amendment leads to drop in agencies entries (as well as exit of some of the agencies with smaller scale), leading to potentially reduced competition among agencies and therefore higher bargaining power χ. Furthermore, the discussion on q will shed light on how policies such as lowering *hukou* barriers interacts with agency activities along the convergence path.

4.4.1 Higher Fixed Costs

I begin with the more straightforward case of higher fixed costs for agencies. I keep $\chi = 0.5$ but now raises F from 10^{-3} to 10^{-2}. Figure 12 shows again the path of capital and labor transition, as well as agency variables. In this case, agencies will only be active for the first 4 periods: since agency profits decreases over time (because of diminished arbitrage opportunities), higher fixed costs means earlier exit of agencies. The top panel

of Figure 13 shows the dynamics of unskilled wages again contrasted with the no agency scenario, where one can see an even more drastic increase in region R's unskilled wages due to agency exit: the rich region now loses its temporary unskilled labor, leading to higher jump in unskilled wages than in the top left panel of Figure 11 where agencies exit at a much later stage and the amount of agency workers also diminishes over time (so the immediate loss of unskilled workers is also less pronounced).

I also plot the dynamics of relative aggregate surplus gains with higher fixed costs in the bottom panel of Figure 13, contrasted with the surplus gains when $F = 10^{-3}$. Again labor agencies can generate substantial surplus gains in the initial periods when they are active. When they exit in the 4th period, the surplus gains are persistent as well but, unsurprisingly, always lower than when $F = 10^{-3}$.

4.4.2 Agency Competition

Now I consider how different degrees of competition among labor agencies affect the path of regional growth and aggregate surplus. Figure 14 shows the dynamics of factor accumulation as well as agency variables in two counterfactual scenarios, i.e., a low competition case $\chi = 0.8$ (top panel) and a high competition case $\chi = 0.2$ (bottom panel). The core observation is that when agencies have higher bargaining power ($\chi = 0.8$), agencies are always charging the highest price $p_a = w_R^U$ such that firms are indifferent between hiring local and agency workers. In contrast, when $\chi = 0.2$, the constraint $p_a \leq w_R^U$ is slack for the initial periods but becomes binding again as regional wage gap shrinks. Furthermore, in this exercise, competition (and the profits margins of agencies) does not affect the extensive margins of agency activities, i.e., agencies exit at the same time period regardless of χ. This is due to the fact that when $p_a \leq w_R^U$ is binding (because w_R^U becomes lower in later periods), χ does not affect agency profits at all.

Interestingly however, despite the fact that the amount of agency worker H_a does not depend on χ, as well as the quantitative result that agencies exit at the same time, competition does affect the aggregate surplus gains. In the top panel of Figure 15, I plot the relative aggregate surplus gains respectively for $\chi = 0.8$ and $\chi = 0.2$, and find that the lower bargaining power of agencies leads to a higher path of aggregate gains. The core insight is that when $\chi = 0.2$, entrepreneurs obtain higher share of the surplus from labor intermediation, which then speeds up capital accumulation and hence generates larger aggregate gains along the convergence path.

Competition among agencies also affect how the surplus gains are distributed. In the bottom panel of Figure 15, I show the share of agency profits in aggregate surplus gains. When $\tau = 0.8$, agencies charge the highest price $p_a = w_R^U$ and can capture around 60% of the surplus gains at the initial period, but such profit shares decline as regional wage gap converges and arbitrage opportunities shrink. By comparison, $\chi = 0.2$ yields an overall lower share of aggregate gains being captured by agencies as profits: for example, it brings down such profit share in the initial period from around 60% to 35%.

To summarize, a higher level of competition among agencies implies that a higher share of surplus from labor intermediation going to entrepreneurs, which has both efficiency and distributional implications: it speeds up capital accumulation and regional convergence while not affecting the timing of agency exit in our quantitative exercises; it also implies that a larger share of the aggregate efficiency gains being distributed to firms and workers.

4.4.3 Agency Workers Transition

Lastly, I explore how the probability of agency workers in transitioning into a permanent local residents in region R affects the regional convergence path. In the top panel of Figure 16, I contrast the dynamics of unskilled wages when $q = 10\%$ (the dashed curve) with those with our benchmark value $q = 6\%$ (the solid curve). With higher q, more of the agency workers who temporarily stay in region R joins the region as a permanent resident, which speeds up labor transition across regions. This implies a sharper reduction of the initial regional wage gap, and agencies exit at an earlier date because of quicker shrinkage of arbitrage opportunities. The exit of agencies again leads to a temporary widening of the wage gap, but such wage gap is always lower than that with $q = 6\%$ when agencies exit in both scenarios.

Raising q can also contribute to higher aggregate surplus gains along the convergence path. In the bottom panel of Figure 16, I find that the path of aggregate surplus gains relative to the no agencies counterfactual is always higher when $q = 10\%$ than when $q = 6\%$. This is intuitive because a higher q implies that the labor reallocation promoted by agencies is more persistent, i.e., some of the agency workers will successfully stay in the rich region, without having to return to region P and become subject to the mobility barriers again when agencies reduce their employment or exit.

Taken together, these counterfactual simulations reveal how different structures of the agency sector affect the transition dynamics of regional convergence. Lowering agency fixed costs may encourage agencies to operate, therefore generating larger surplus gains along the path. Increasing the transition probability of agency workers to permanent (local) workers implies that labor reallocations via agencies become more persistent, which then speed up frictional labor transition across regions. Lastly, although in this model, competition does not affect the intensive margins of labor reallocation (i.e., the amount of agency workers being hired) and the timing of agencies exit, it does raise the share of surplus going to firms and workers. Interestingly, such distributional implications then translate into persistent efficiency gains via faster capital accumulation.

5 Conclusion

This paper shows that market intermediaries may arise endogenously to arbitrage disparities in prices and marginal product of factor inputs and at least partially correct aggregate misallocations. It also discusses the extent to which the private interests of

intermediaries may align with aggregate efficiency gains along the dynamic path of economic transition. By looking at the empirical context of labor recruitment agencies in China, I document the spatial patterns of agencies expansion, as well as its implications on factor reallocation and regional convergence in wages and MPL. I then quantify the aggregate output gains brought by labor agencies using some back-of-the-envelope calculations, and find that almost one third of such aggregate gains are captured by agencies as profits throughout the labor reallocation process. Then with a calibrated neoclassical model of regional growth, I show how labor agencies may initially mitigate the inter-regional wage disparities and speed up the convergence process to the efficient steady states. The quantitative results demonstrate the importance of increasing competition among labor market intermediaries, which then implies a larger share of surplus accruing to entrepreneurs. Such distributional impact has efficiency consequences as it speeds up capital accumulation and bring surplus gains.

In my ongoing work, I find that the increasing prevalence of labor agencies across the country has enlarged pre-existing regional gaps in educational investment among the youth. This is consistent with the idea that providing job opportunities in richer areas to young people from poorer areas increases their opportunity cost of pursuing higher education, a theme explored as in Atkin (2016) and Rozelle and Hell (2020). This also implies that agencies does not only affect the transition dynamics of capital and labor input, but also shapes human capital accumulation which then have richer implications on aggregate welfare and long run outcomes. This requires incorporating human capital investment into the dynamic framework to evaluate the welfare impacts of factor reallocations via agencies.

Table 1: Ex-Ante Wage and Agency Expansion

	(1) Branched	(2) Branched	(3) Branched
Bottom 10%	0.00372***	0.00377***	0.195***
	(0.000270)	(0.000269)	(0.0582)
Top 10%	0.0112***	0.0113***	0.288***
	(0.000576)	(0.000577)	(0.0420)
Distance	-0.0447***	-0.0452***	-1.437***
	(0.000382)	(0.000376)	(0.0245)
Branched Before 2007	0.712***		
	(0.00561)		
High School Student Teacher Ratio	-0.000465***	-0.000478***	0.0628***
	(0.0000395)	(0.0000395)	(0.00657)
Number of High Schools	-0.00583***	-0.00586***	-0.0798**
	(0.000206)	(0.000206)	(0.0321)
Library Collections	0.00683***	0.00684***	0.647***
	(0.000209)	(0.000208)	(0.0211)
Observations	906449	905467	888023
Methods	OLS	OLS	Conditional Logit
Fixed Effects			
Agency Group	Yes	Yes	Yes

Notes: This table shows the regression results for the relationship between ex-ante city average wage and agency expansion decision, where Bottom 10% and Top 10% are binary indicator for whether the ex-ante (i.e., 2007) city average wage in the Chinese City Statistical Yearbook belongs to the bottom 10% and top 10% decile group. Branched is a binary variable which equals 1 if an agency group has ever established a branch in the city during 2008-2018. Branched Before 2007 is a binary variable which equals to 1 if the agency group had branch in the city before 2007. The first and the second column is estimated using the linear probability model, and the third column is obtained from a conditional logit model. The second and the third columns are based on observations that don't have branch in the city before 2007. Standard errors reported in parentheses are clustered at agency group level: *$p < 0.10$, *$p < 0.05$, *$p < 0.01$.

Table 2: Ex-Ante Wage Disparity and Agency Network

	(1) Connected	(2) Connected	(3) Connected	(4) Connected
$\lvert\Delta Wage^{pre}\rvert$	0.192***	0.193***	1.496***	0.807***
	(0.00899)	(0.00904)	(0.0632)	(0.0351)
Distance	-0.0889***	-0.0897***	-0.684***	-0.376***
	(0.00285)	(0.00288)	(0.0202)	(0.0113)
Connected Before 2007	0.749***			
	(0.00722)			
Observations	40186	39977	39977	39977
Methods	OLS	OLS	Logit	Probit

Notes: This table shows the regression results for the relationship between ex-ante wage disparity between cities and expansion decision of agency network, where $\lvert\Delta Wage^{pre}\rvert$ is the absolute value of ex-ante (i.e., 2007) logged average wage from the Chinese City Statistical Yearbook for each pair of cities. Connected is a binary variable which equals 1 if both cities have ever been branched by any agency group during 2008-2018. The first two columns are estimated using the linear probability model, the third column is obtained from the logit model, and the fourth column is from a probit model. Robust standard errors reported in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Ex-Ante Wage Disparity and Agency Profits

	(1) Profits	(2) Profits	(3) Profits	(4) Profits	(5) Profits
$\lvert\max \Delta Wage^{pre}\rvert$	1.657***	1.097***	1.035***	1.304***	0.924**
	(0.385)	(0.313)	(0.311)	(0.337)	(0.376)
Size		0.491***	0.472***	0.461***	0.469***
		(0.0471)	(0.0458)	(0.0463)	(0.0465)
Agency Groups			-0.0208***	0.00218	-0.0198***
			(0.00622)	(0.0133)	(0.00647)
Agency Groups× $\lvert\max \Delta Wage^{pre}\rvert$				-0.0475**	
				(0.0226)	
Earliest× $\lvert\max \Delta Wage^{pre}\rvert$					0.201
					(0.344)
Observations	231	231	231	231	231

Notes: This table shows the regression results for the relationship between ex-ante wage disparity of agency network and agency profits. Profits is the (log) average annual profits of headquarter of agency group over 2008-2016, based on the administrative tax survey dataset. $\lvert\max \Delta Wage^{pre}\rvert$ is the maximum ex-ante (i.e., 2007) wage gap among cities that the agency group later expanded into during 2008-2018, where the city average wage is from the Chinese City Statistical Yearbook. Size is the (log) average total employment of headquarter of agency group over 2008-2016. Agency Groups is the number of agency groups that had established branches in the same cities that exhibit such maximum wage gap. Earliest is a binary variable which equals to 1 if the agency group is the earliest group that connect the two cities with $\lvert\max \Delta Wage^{pre}\rvert$. Robust standard errors reported in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Agency Network and Labor Mobility Across Cities: 2005 and 2010

	(1) Gross Flow	(2) Gross Flow (College)	(3) Gross Flow (Below College)	(4) Net Flow	(5) Net Flow (College)	(6) Net Flow (Below College)
Connected	4.246***	1.543***	2.568**	-1.447	0.289	-1.681
	(1.309)	(0.362)	(0.997)	(2.130)	(0.697)	(1.536)
Connected×$\Delta Wage^{pre}$	10.01*	2.851**	6.763	18.83*	5.839**	12.24
	(5.474)	(1.195)	(4.275)	(9.971)	(2.387)	(7.588)
Connection	2.214***	0.719***	1.432***	1.398	0.0497	1.294
	(0.637)	(0.194)	(0.492)	(1.334)	(0.476)	(0.918)
Connection×$\Delta Wage^{pre}$	6.628***	1.354**	5.137***	14.31***	2.755**	11.24***
	(2.473)	(0.542)	(1.990)	(4.264)	(1.270)	(3.107)
Observations	158484	158484	158484	79242	79242	79242
Fixed Effects						
City Pair	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the OLS estimates of the effect of expansion of agency network on worker mobility across cities. The results are obtained based on the sub-sample of population census in 2005 and 2010. $\Delta Wage^{pre}$ is the ex-ante (i.e., 2007) log average wage difference between city n and m. Gross flow from city m to city n (where $n \neq m$) is calculated as the total headcount (weighted by city population shares) for which the respondents reported to have lived in city n (for more than one year) but their hukou registration is in city m. Gross flow of college (below college) labor is measured by such flow for which the respondents have a college (below college) degree. Net flow is calculated by $NetFlow_{mnt} = Flow_{mnt} - Flow_{nmt}$. Connected is a binary variable which equals to 1 if there exist one agency group that has active branches in both cities. Standard errors reported in parentheses are clustered at city pair level: *$p < 0.10$, *$p < 0.05$, *$p < 0.01$.

Table 5: Agency Network and Regional Wage Gap (Manufacturing Sectors): 2008 - 2016

	(1) $\lvert\Delta Wage\rvert$	(2) $\lvert\Delta Wage\rvert$	(3) $\lvert\Delta Wage\rvert$
Connected	-0.0147***	0.0442***	0.0369***
	(0.00118)	(0.00163)	(0.00243)
Connected×$\lvert\Delta Wage^{pre}\rvert$		-0.244***	-0.207***
		(0.00636)	(0.0135)
Connections			0.00542***
			(0.00133)
Connections×$\lvert\Delta Wage^{pre}\rvert$			-0.0282***
			(0.00914)
Observations	436759	415359	415359
Fixed Effects			
City Pair	Yes	Yes	Yes
Year	Yes	Yes	Yes

Notes: This table reports the OLS estimates of the effect of LDA expansion on (manufacturing) wage dispersion across cities during 2008-2016. The results are obtained based on the administrative dataset of tax survey of firms. $\lvert\Delta Wage\rvert$ is the absolute value of difference of (log) average wage between two cities. Connected is a binary variable which equals to 1 if there exist one agency group that has active branches in both cities. Connection is the number of such agency groups. $\lvert\Delta Wage^{pre}\rvert$ is the absolute value of ex-ante (i.e., 2007) log average (weighted by employment) wage difference. Standard errors reported in parentheses are clustered at city pair level: *$p < 0.10$, ** $p < 0.05$, ***$p < 0.01$.

Table 6: Agency Network and Regional Wage Gap (Comparable Sectors): 2008 - 2016

	(1)	(2)	(3)	(4)	(5)	(6)
	Construction	Construction	Hospitality	Hospitality	Logistics& Warehousing	Logistics& Warehousing
Connected	0.0434***	0.0362***	0.0178***	0.0152***	0.0398***	0.0362***
	(0.00596)	(0.00633)	(0.00488)	(0.00514)	(0.00568)	(0.00592)
Connected×$\lvert\Delta Wage^{pre}\rvert$	-0.0931***	-0.0738***	-0.0798***	-0.0677***	-0.125***	-0.111***
	(0.0112)	(0.0123)	(0.00969)	(0.0104)	(0.0106)	(0.0115)
Connections		0.00554***		0.00153		0.00263**
		(0.00144)		(0.00115)		(0.00117)
Connections×$\lvert\Delta Wage^{pre}\rvert$		-0.0153***		-0.00820***		-0.0111***
		(0.00367)		(0.00245)		(0.00339)
Observations	265589	265589	89979	89979	210956	210956
Fixed Effects						
City Pair	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the OLS estimates of the effect of LDA expansion on wage dispersion across cities in correspondingly construction, hospitality, and logistics/warehousing sectors during 2008-2016. The results are obtained based on the administrative dataset of tax survey of firms. Outcome variables are the absolute value of difference of (log) average wage for the corresponding sector between two cities. Connected is a binary variable which equals to 1 if there exist one agency group that has active branches in both cities. Connection is the number of such agency groups. $\lvert\Delta Wage^{pre}\rvert$ is the absolute value of ex-ante (i.e., 2007) log average (weighted by employment) wage difference for the corresponding sector. Standard errors reported in parentheses are clustered at city pair level: *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table 7: Agency Network and Regional Wage Gap (by Education): 2011, 2013, 2015, and 2017

	(1) $\lvert\Delta Wage_0\rvert$	(2) $\lvert\Delta Wage_0\rvert$	(3) $\lvert\Delta Wage_1\rvert$	(4) $\lvert\Delta Wage_1\rvert$
Connected	0.0240	0.0224	-0.0107	-0.00474
	(0.0171)	(0.0174)	(0.0326)	(0.0330)
Connected×$\lvert\Delta Wage^{pre}\rvert$	-0.224***	-0.212***	-0.134*	-0.143*
	(0.0404)	(0.0428)	(0.0759)	(0.0780)
Connections		0.000228		0.00365
		(0.00253)		(0.00254)
Connections×$\lvert\Delta Wage^{pre}\rvert$		-0.00904		0.0118
		(0.00949)		(0.0126)
Observations	20119	20119	3895	3895
Fixed Effects				
City Pair	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes

Notes: This table reports the OLS estimates of the effect of agency network expansion on wage (by education) dispersion across cities in 2011, 2013, 2015, and 2017. The results are obtained based on the CLDS dataset (2012, 2014, 2016 and 2018). $\lvert\Delta Wage_e\rvert$ is the absolute value of difference of log average annual wage specific to educational background $e \in \{0, 1\}$ between two cities, where $e = 0$ stands for below college degree, and $e = 1$ is college degree. Connected is a binary variable which equals to 1 if there exist one agency group that has active branches in both cities. Connection is the number of such agency groups. $\lvert\Delta Wage^{pre}\rvert$ is the absolute value of ex-ante (i.e., 2007) log average wage difference from the Chinese City Statistical Yearbook. Standard errors reported in parentheses are clustered at city pair level: *$p < 0.10$, *$p < 0.05$, *$p < 0.01$.

Table 8: Agency Network and Regional MPL Gap: 2008 - 2016

	(1) $\lvert\Delta MPL\rvert$	(2) $\lvert\Delta MPL\rvert$	(3) $\lvert\Delta MPL\rvert$	(4) $\lvert\Delta MPL\rvert$	(5) $\lvert\Delta MPL\rvert$
Connected	0.00710***	0.00854**	0.00653	0.0181***	0.0162***
	(0.00241)	(0.00382)	(0.00413)	(0.00403)	(0.00426)
Connected×$\lvert\Delta Wage^{pre}\rvert$		-0.0239*	-0.0209		
		(0.0136)	(0.0149)		
Connections			0.00153		0.00143
			(0.00121)		(0.00110)
Connections×$\lvert\Delta Wage^{pre}\rvert$			-0.00210		
			(0.00441)		
Connected×$\lvert\Delta MPL^{pre}\rvert$				-0.0493***	-0.0461***
				(0.0121)	(0.0136)
Connections×$\lvert\Delta MPL^{pre}\rvert$					-0.00236
					(0.00471)
Observations	436759	415359	415359	415359	415359
Fixed Effects					
City Pair	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the OLS estimates of the effect of agency network expansion on manufacturing MPL dispersion across cities during 2008-2016. The results are obtained based on the administrative dataset of tax survey of firms. $|\Delta MPL|$ is the absolute value of difference of log average (weighted by employment) MPL between two cities. Connected is a binary variable which equals to 1 if there exist one agency group that has active branches in both cities. Connection is the number of such agency groups. $|\Delta Wage^{pre}|$ is the absolute value of ex-ante (i.e., 2007) log average (weighted by employment) wage difference, and $|\Delta MPL^{pre}|$ is the absolute value of ex-ante (i.e., 2007) log average (weighted by employment) MPL difference. Standard errors reported in parentheses are clustered at city pair level: *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table 9: Agency Network and Regional MPK Gap: 2008 - 2016

	(1) $\lvert\Delta MPK\rvert$	(2) $\lvert\Delta MPK\rvert$	(3) $\lvert\Delta MPK\rvert$
Connected	0.00439	-0.0239***	-0.0200***
	(0.00310)	(0.00492)	(0.00514)
Connected×$\lvert\Delta Wage^{pre}\rvert$		0.0910***	0.0816***
		(0.0170)	(0.0177)
Connections			-0.00297**
			(0.00138)
Connections×$\lvert\Delta Wage^{pre}\rvert$			0.00692
			(0.00445)
Observations	436759	415359	415359
Fixed Effects			
City Pair	Yes	Yes	Yes
Year	Yes	Yes	Yes

Notes: This table reports the OLS estimates of the effect of LDA expansion on (manufacturing) MPK dispersion across cities during 2008-2016. The results are obtained based on the administrative dataset of tax survey of firms. $\lvert\Delta MPK\rvert$ is the absolute value of difference of (log) average MPK (weighted by capital input) between two cities. Connected is a binary variable which equals to 1 if there exist one agency group that has active branches in both cities. Connection is the number of such agency groups. $\lvert\Delta MPK^{pre}\rvert$ is the absolute value of ex-ante (i.e., 2007) log average (weighted by employment) wage difference. Standard errors reported in parentheses are clustered at city pair level: *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table 10: Relative Output Gains in 2010

	$Y = AK^{\alpha}L^{1-\alpha}$	$Y = AK^{\alpha}(L_U)^{\beta}(L_S)^{1-\alpha-\beta}$		
		$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$
Relative Output Gains	1.35%	1.21%	1.24%	1.31%
Agency Profit Share	28.72%			

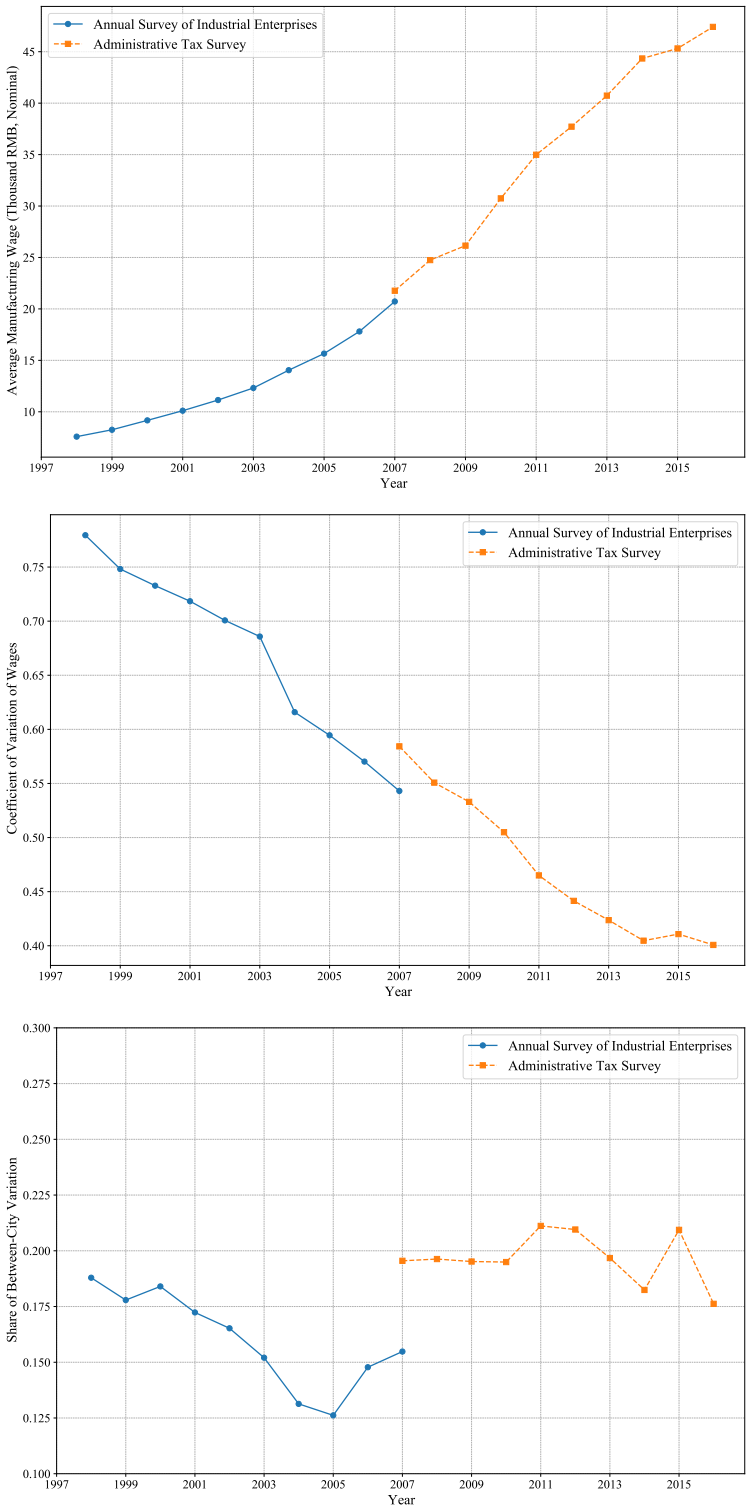
Notes: This table reports the aggregate output gains *relative* to the no agency counterfactual, as well as the share of agency profits in aggregate output gains in 2010.

Table 11: Relative Output Gains: Linear Projections

	$y = AK^{\alpha}L^{1-\alpha}$	$y = AK^{\alpha}(L_U)^{\beta}(L_S)^{1-\alpha-\beta}$		
		$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$
$\{\text{Connections}\}_{2011}$	8.85%	5.42%	7.06%	8.85%
$\{\text{Connections}\}_{2012}$	9.97%	7.30%	8.79%	10.45%
$\{\text{Connections}\}_{2013}$	10.86%	8.47%	9.88%	11.53%
$\{\text{Connections}\}_{2014}$	12.43%	9.95%	11.36%	13.09%
$\{\text{Connections}\}_{2015}$	13.86%	12.15%	13.44%	15.12%
$\{\text{Connections}\}_{2016}$	16.40%	15.55%	16.61%	18.24%
$\{\text{Connections}\}_{2017}$	19.90%	21.28%	21.76%	22.98%
$\{\text{Connections}\}_{2018}$	23.47%	28.43%	28.14%	28.63%

Notes: This table reports the aggregate output gains *relative* to the no agency counterfactual where we use the predicted labor flows with the agency network as in 2011-2018.

FIGURE 1: WAGES IN MANUFACTURING SECTOR IN CHINA: 1998-2016

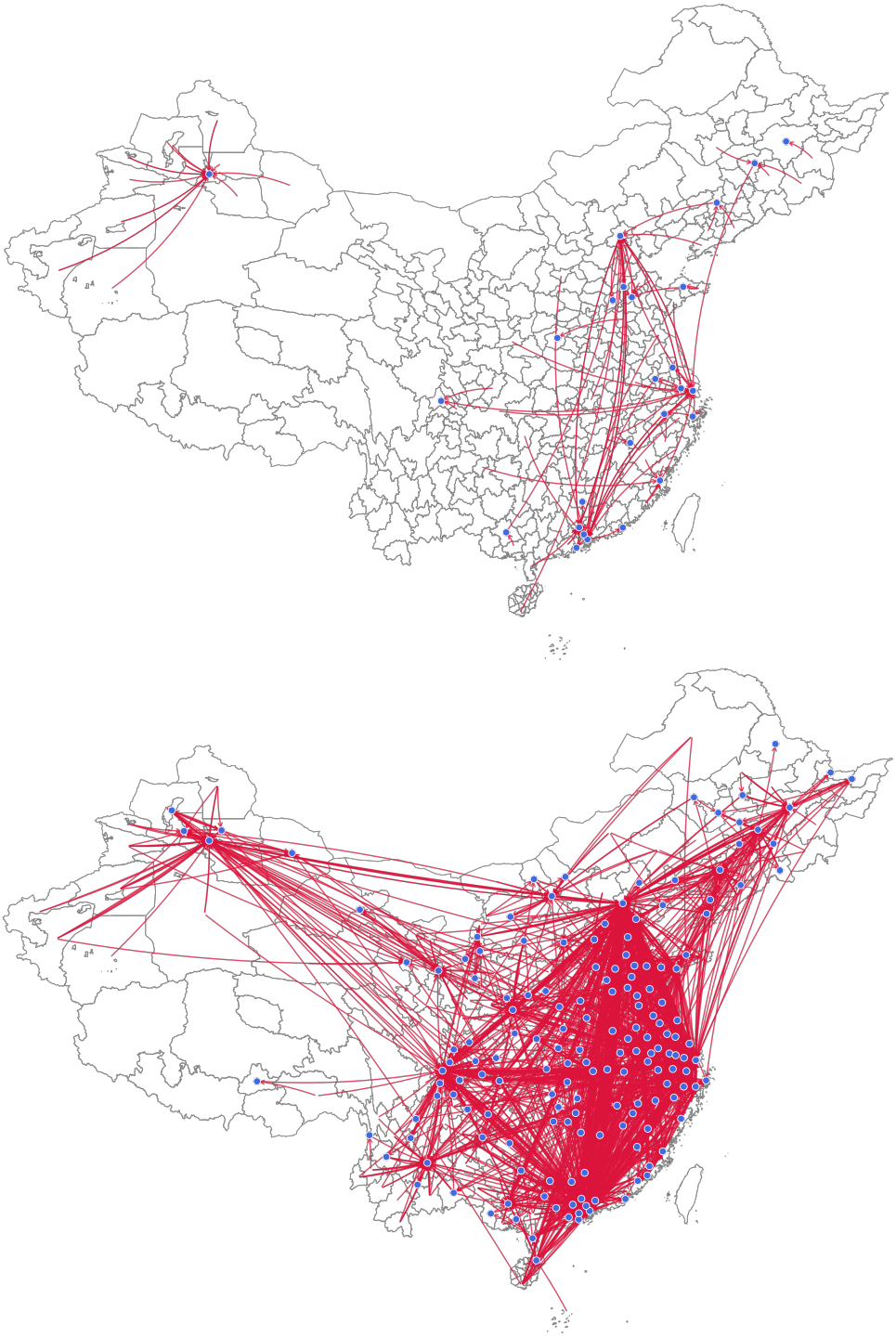


Notes: This figure shows average wages and the dispersion of wages in manufacturing sector in China over 1998-2016, based on the Annual Survey of Industrial Enterprises and Administrative Tax Survey. The top panel shows the annual average wage of manufacturing firms and the middle panel shows the coefficient of variation of firm wages. The bottom panel shows the share of between-city variation of firm wages in the national total variance, following on the variance decomposition exercise

$$\text{var}(w_{ic}) = \underbrace{\frac{1}{N}\sum_{c} N_c\left(\overline{w}_c - \overline{w}\right)^2}_{\text{between-city variance}} + \underbrace{\frac{1}{N}\sum_{c}\sum_{i\in c}\left(w_{ic} - \overline{w}_c\right)^2}_{\text{within-city variance}} .$$

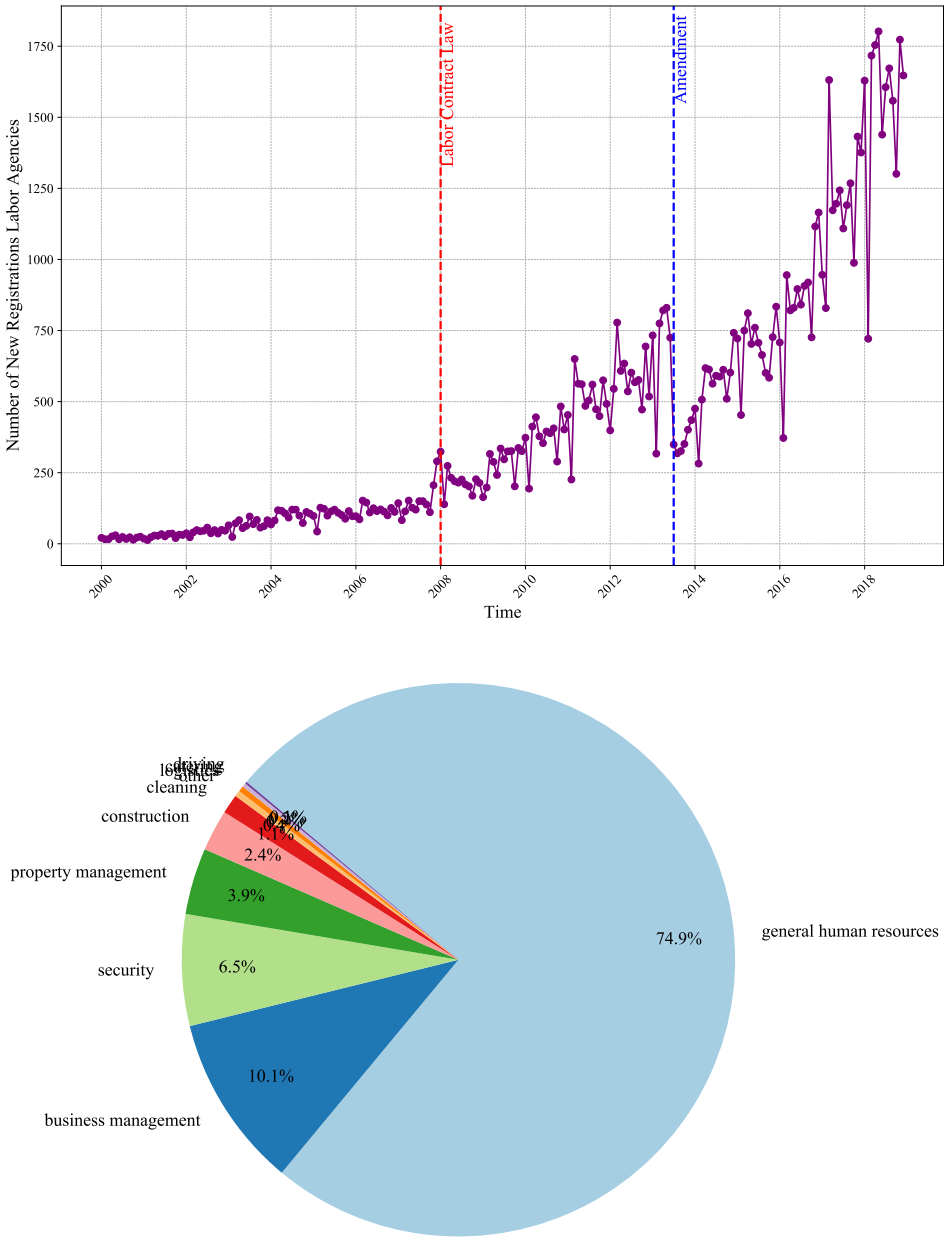
where w_{ic} is the average wage of firm i in city c, N is the total number of firms, N_c is the total number of firms in city c, $\overline{w}_c$ is the average wage in city c, and $\overline{c}$ is the national average wage.

FIGURE 2: AGENCY NETWORK: 2005 (TOP) VERSUS 2015 (BOTTOM)



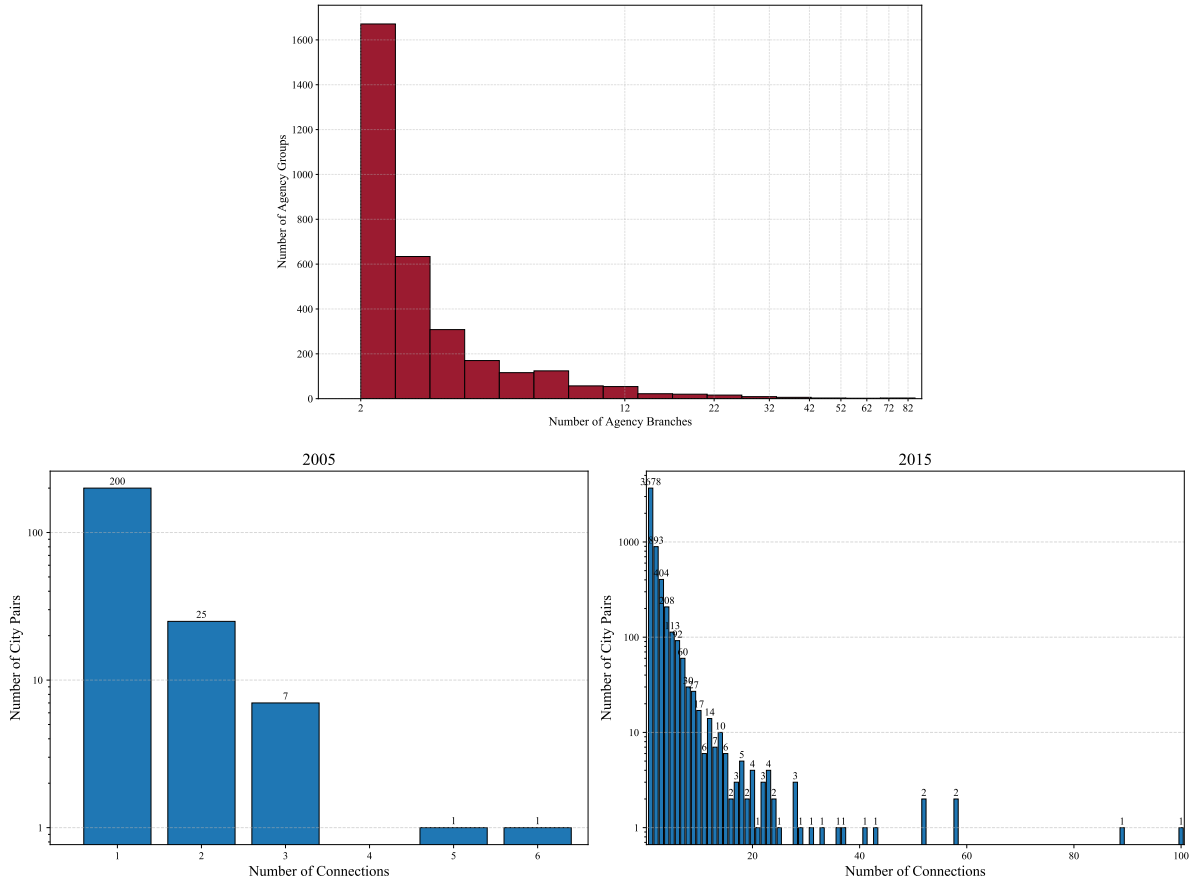
Notes: This figure shows expansion of agencies network from 2005 to 2015, based on my labor agencies dataset. The blue dots represent the headquarter cities for different agency groups, and the red lines connect headquarters with local branches.

FIGURE 3: LABOR AGENCIES: REGISTRATION AND SPECIALIZATION



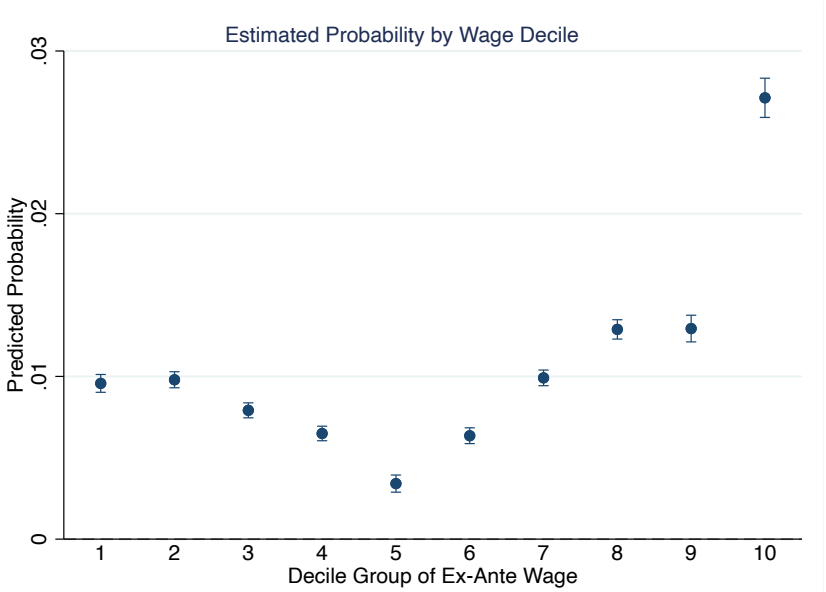
Notes: The top panel shows the monthly time series of LDA registrations in China over 2000-2019. The red dashed line captures the timing of the introduction of Labor Contract Law in January 2008, and the blue dashed line captures the Amendment to Labor Contract Law in July 2013. The bottom panel presents the categorization of agency groups and their local branches, based on their business range information on *aiqicha*.

Figure 4: Distribution of Agency Groups



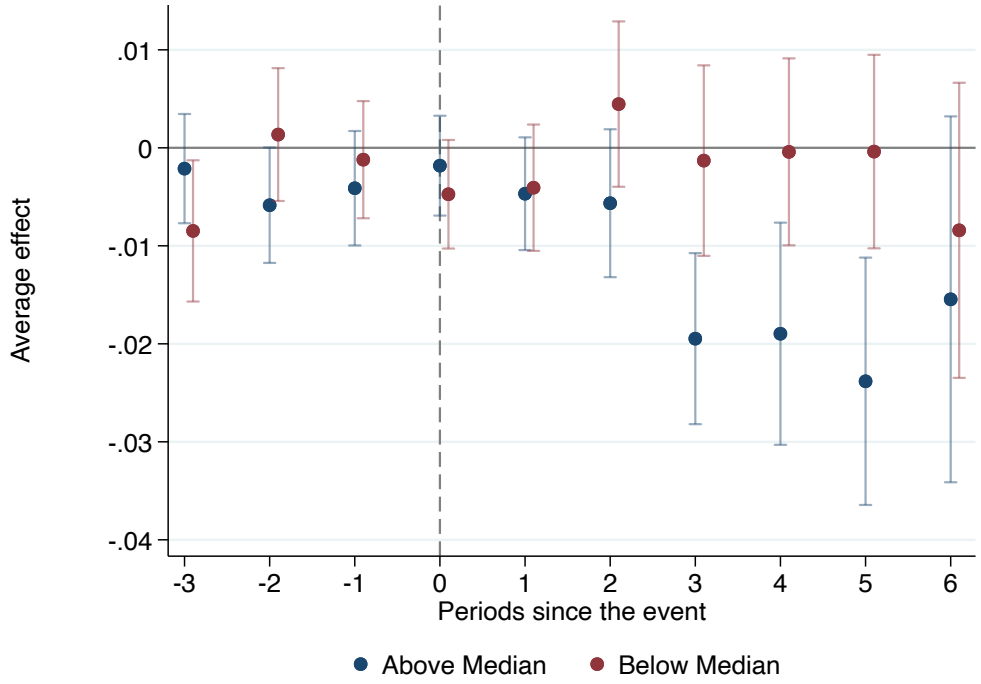
Notes: The top panel shows the distribution of size of agency groups (i.e., the number of branches within an agency group). The bottom panel shows the frequency of city pairs that are connected (i.e., agency group have branches in both cities) by the corresponding number of agency groups in 2005 (left) and 2015 (right).

Figure 5: Ex-ante Wage Decile and Agency Expansion



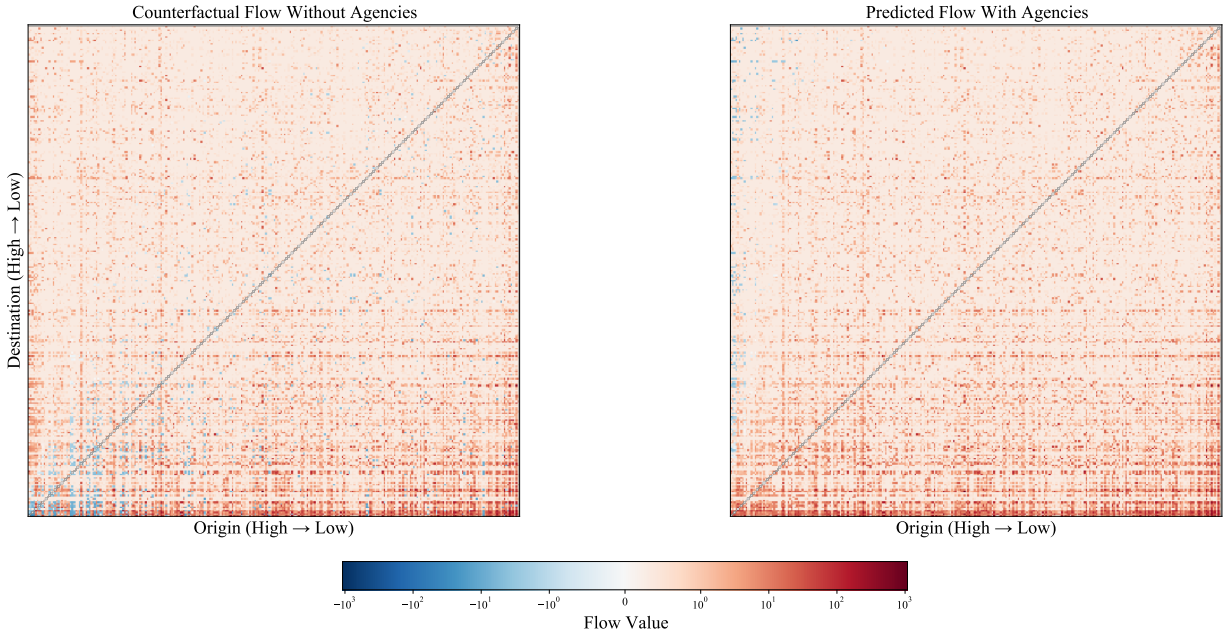
Notes: The figure shows estimates of the linear probability model for an agency group to establish branches in a city during 2008-2018 on decile group of ex-ante (i.e., 2007) city average wage (from the Chinese City Statistical Yearbook), controlling for distance between the city and the agency headquarter, the ex-ante number of high schools in the city, as well as the ex-ante city average student-teacher ratio in high school.

Figure 6: Event Study: Agency Network and Wage Dispersion



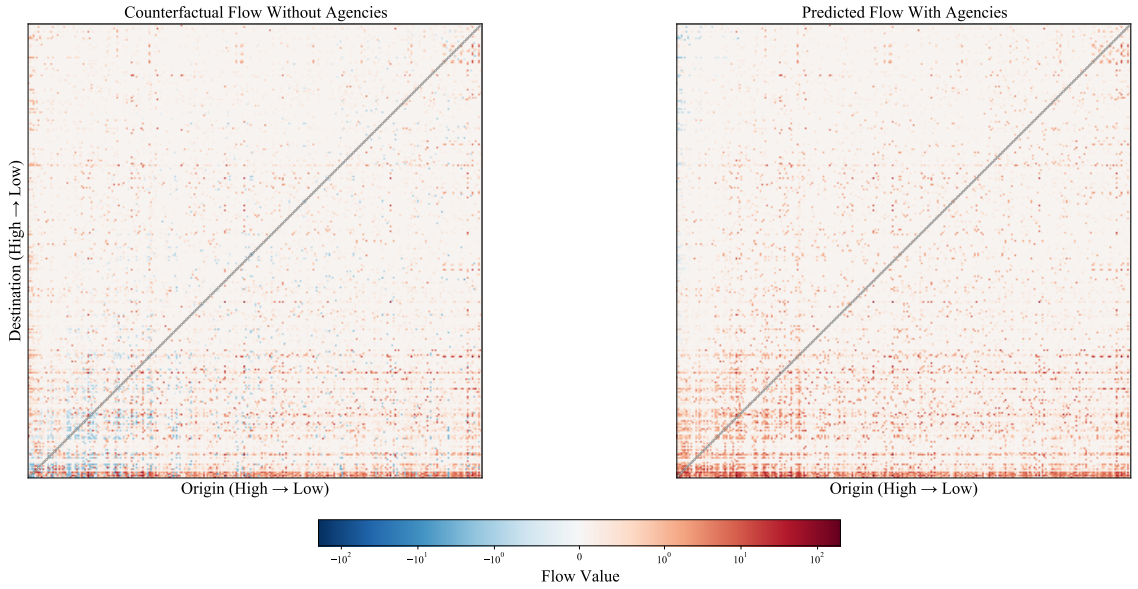
Notes: The figure shows estimates from the event study specification of difference of (log) average (manufacturing) wage for city pairs on Connected (i.e., if there exist an agency group that has active branches in both cities), respectively for city pairs with ex-ante above median and below median wage disparity, based on the methodology of Callaway and Sant'Anna (2021).

Figure 7: Inter-City Labor Reallocation

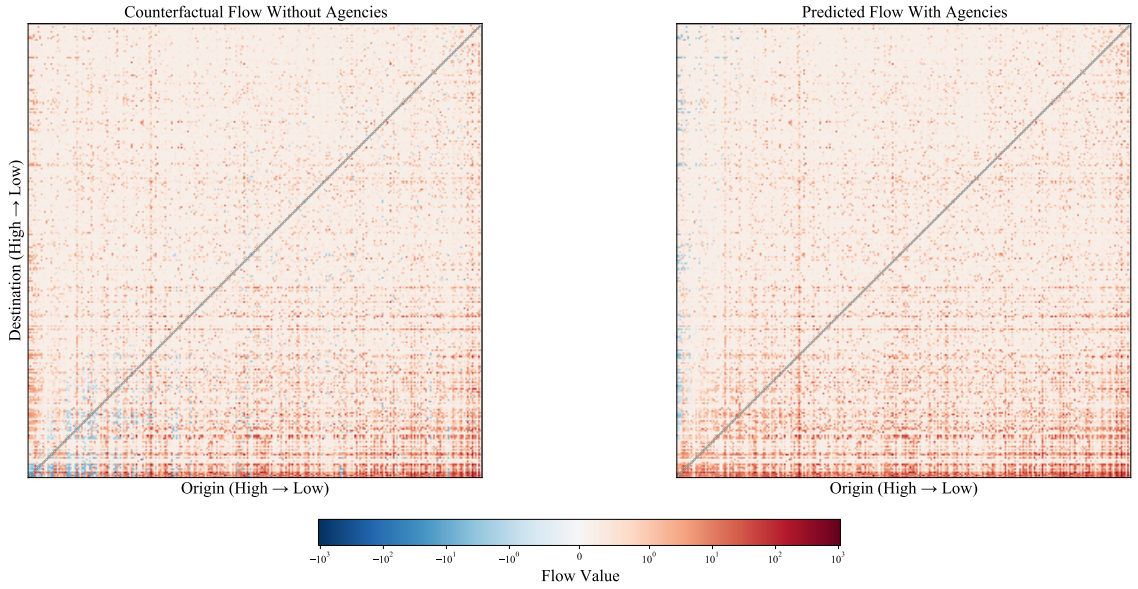


Notes: This figure shows the counterfactual total labor flow (left), $\hat{Flow}_{mn}^{counterfactual}$, and the fitted value of total labor flow with agencies (right), $\hat{Flow}_{mn}$ in the $N \times N$ city system in 2010, all weighted by city population shares. The horizontal (vertical) axis shows the origin (destination) city, ordered from left (bottom) to right (top) according to their ex-ante (2005) average wage. Therefore the i-th column and j-th row pixel shows the total number of labor flow from the city with i-th rank in the ex-ante average wage distribution to the city with $N+1-j$-th rank in the ex-ante average wage distribution.

FIGURE 8: INTER-CITY LABOR REALLOCATION: COLLEGE VERSUS BELOW COLLEGE WORKERS



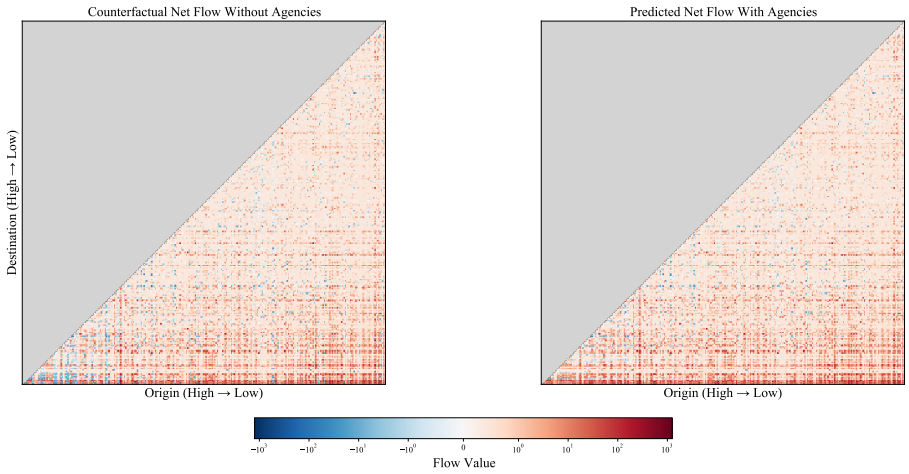
(A) COLLEGE



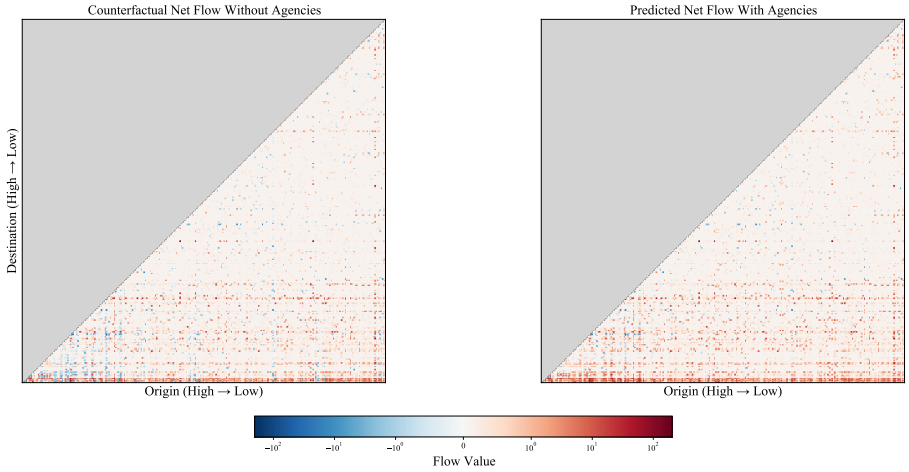
(B) BELOW COLLEGE

Notes: This figure shows the counterfactual labor flow (left), $\hat{Flow}_{mn}^{counterfactual}$, and the fitted value of labor flow with agencies (right), $\hat{Flow}_{mn}$ in the $N \times N$ city system in 2010, respectively for individuals with a college or above degree (top panel) and those without a college degree (bottom panel). The horizontal (vertical) axis shows the origin (destination) city, ordered from left (bottom) to right (top) according to their ex-ante (2005) average wage. Therefore the i-th column and j-th row pixel shows the total number of labor flow from the city with i-th rank in the ex-ante average wage distribution to the city with $N+1-j$-th rank in the ex-ante average wage distribution.

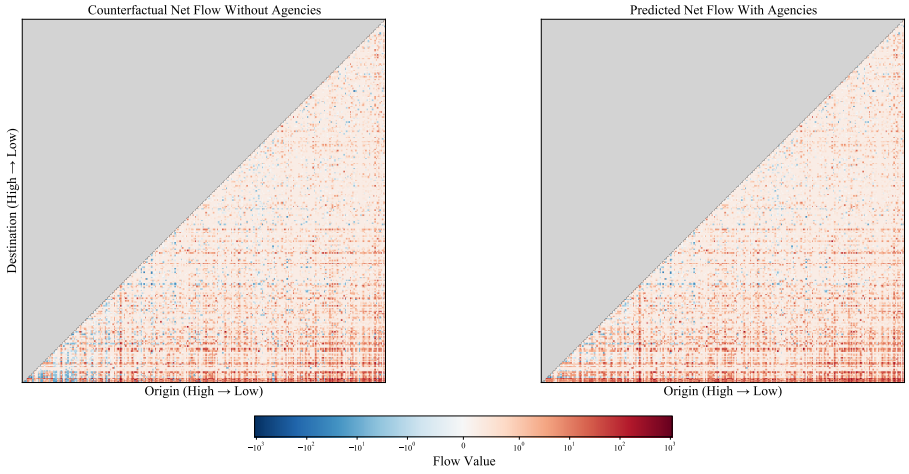
FIGURE 9: INTER-CITY LABOR REALLOCATION: NET FLOWS



(A) ALL



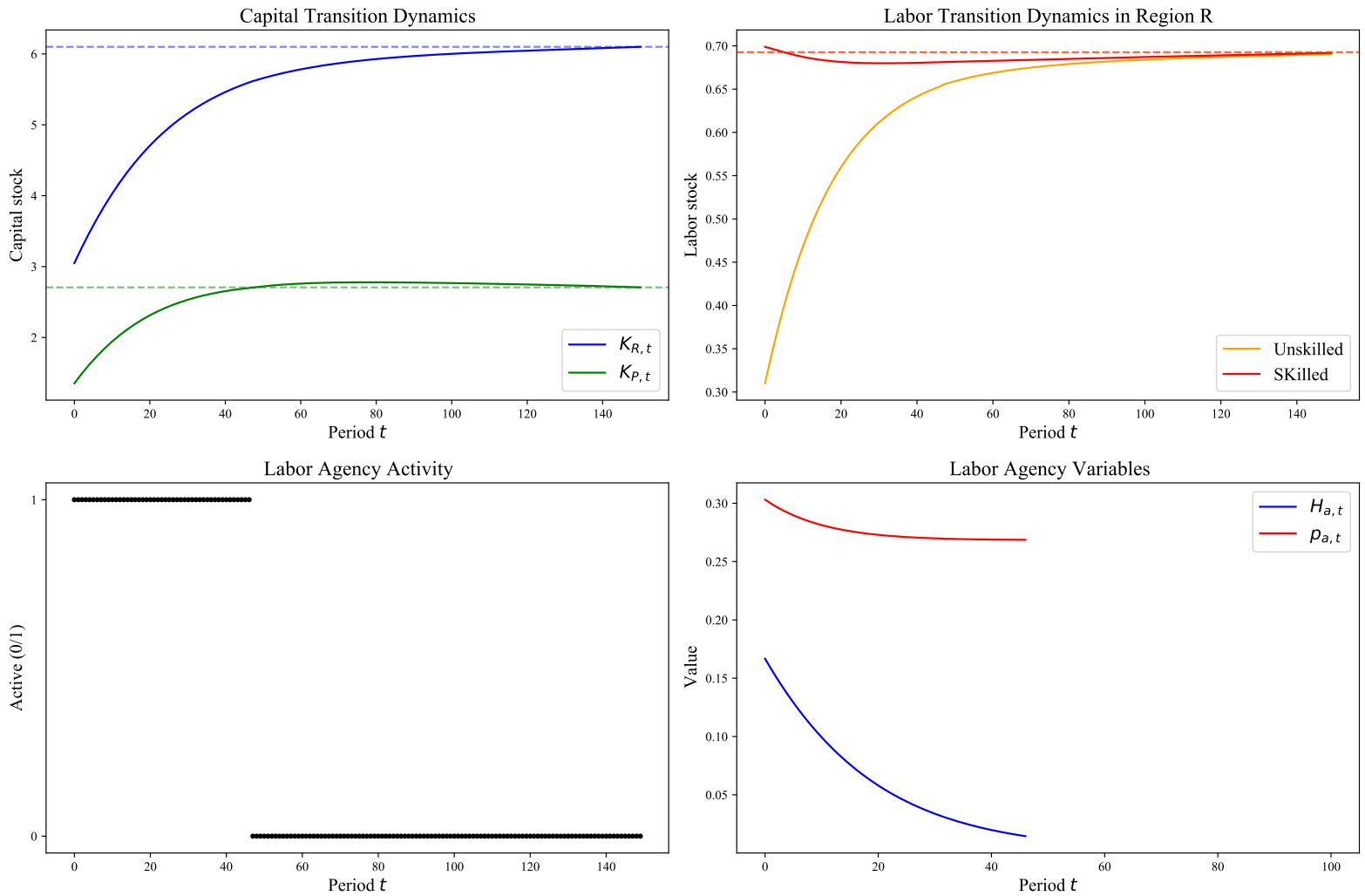
(B) COLLEGE



(C) BELOW COLLEGE

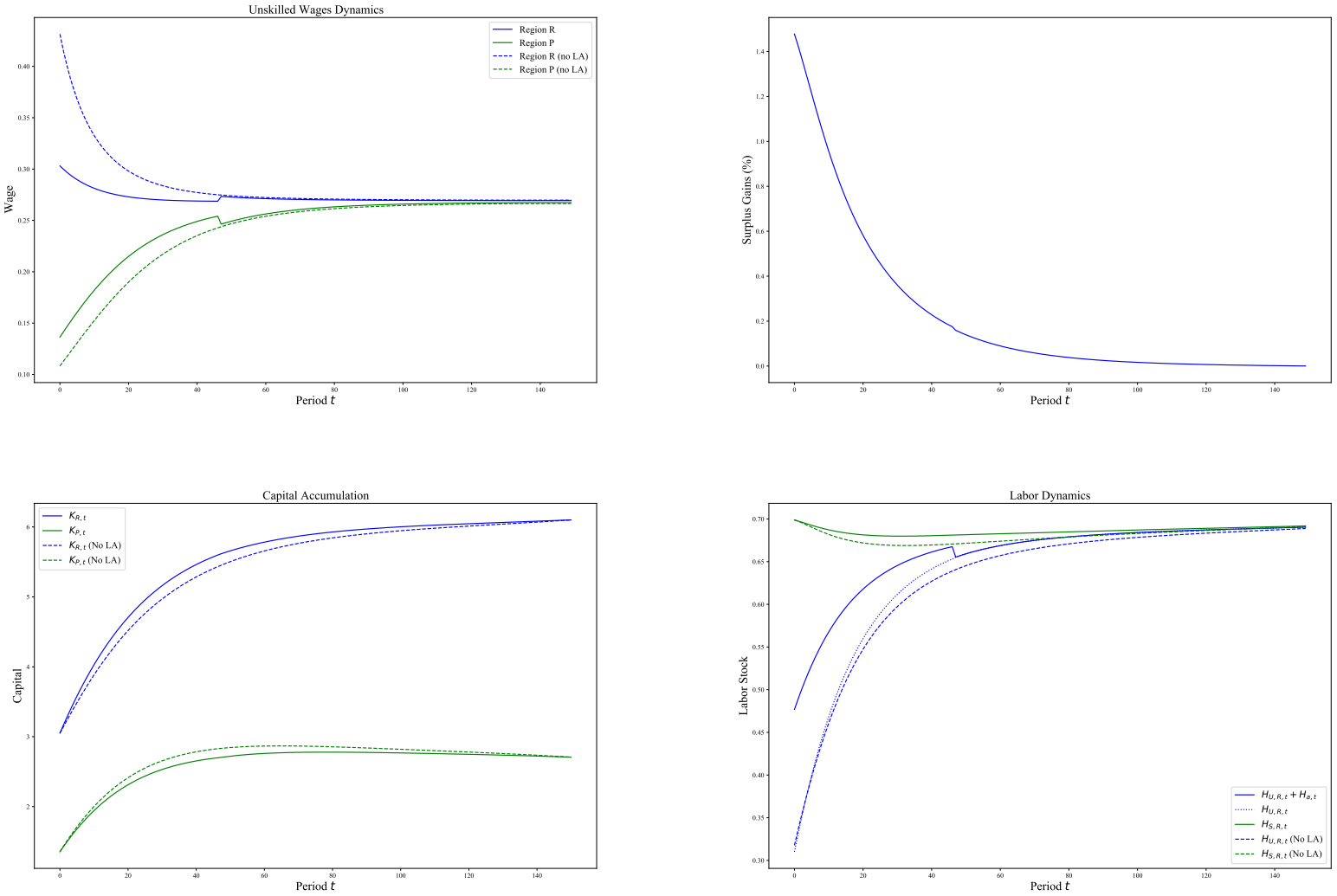
Notes: This figure shows the counterfactual labor net flow (left), $\hat{Flow}_{mn}^{counterfactual}$, and the fitted value of labor net flow with agencies (right), $\hat{Flow}_{mn}$ in the $\frac{N\times(N-1)}{2}$ city system in 2010, respectively for all individuals (top), individuals with a college or above degree (middle panel) and those without a college degree (bottom panel). The horizontal (vertical) axis shows the origin (destination) city, ordered from left (bottom) to right (top) according to their ex-ante (2005) average wage. Therefore the i-th column and j-th row pixel shows the net flow from the city with i-th rank in the ex-ante average wage distribution to the city with $N+1-j$-th rank in the ex-ante average wage distribution.

FIGURE 10: BENCHMARK SCENARIO



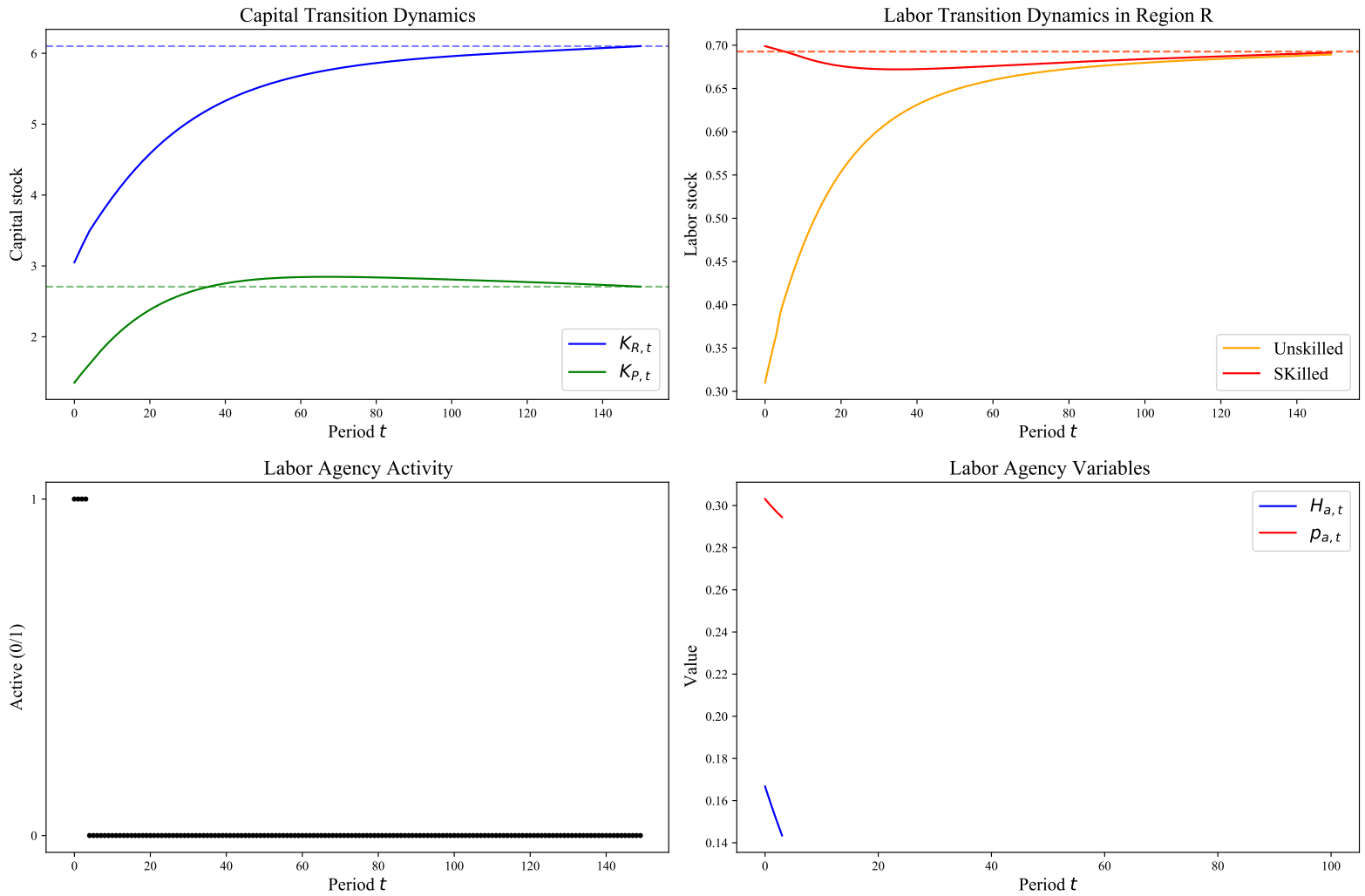
Notes: The figure shows the transition dynamics of capital accumulation, labor mobility, as well as agency variables by taking $\chi = 0.5$. Active is a binary variable which equals to one if agencies operate. In the bottom right diagram I show how the amount of agency workers H_a (blue curve) and price that firms pay to agencies p_a (red curve) evolves over time (while agencies are active).

Figure 11: Benchmark Scenario versus No Agencies Counterfactual



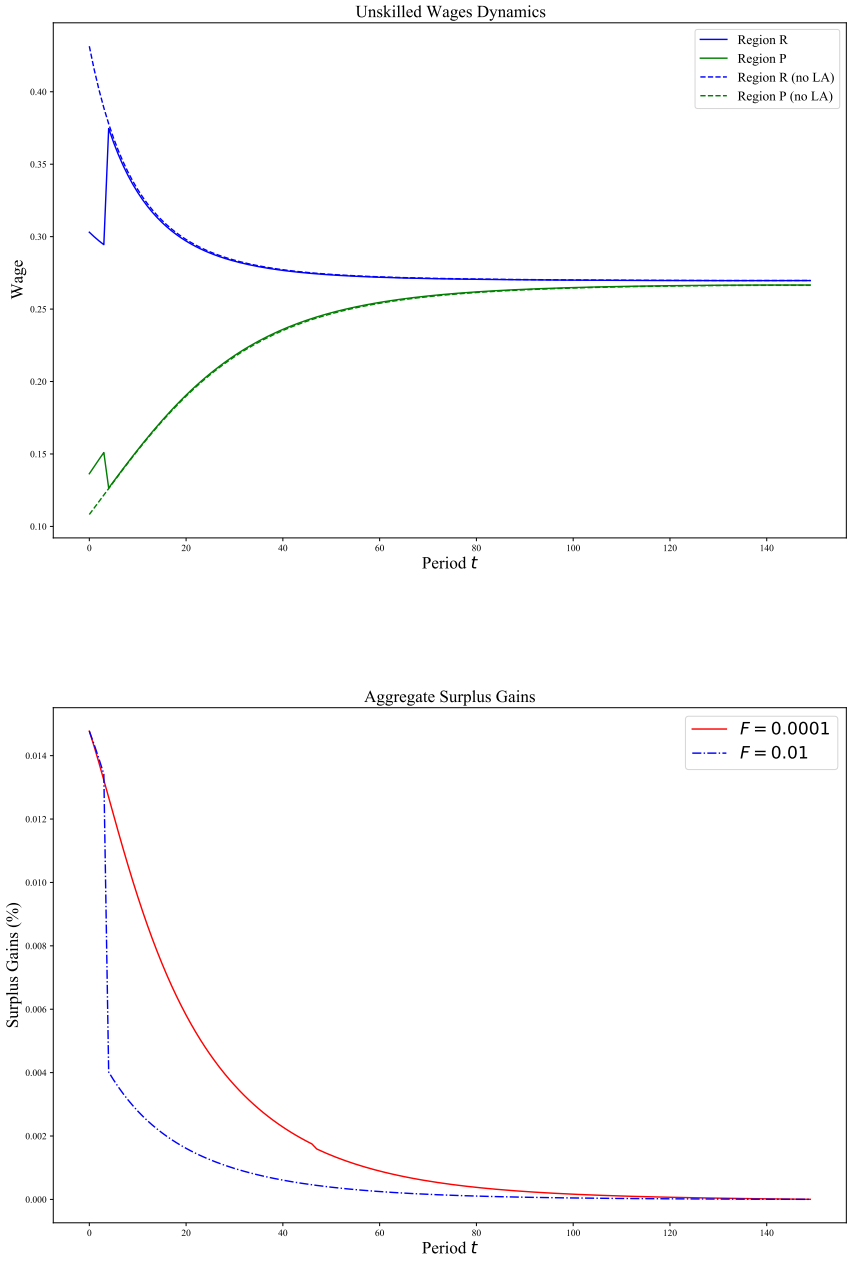
Notes: This figure shows the dynamics of unskilled wages (top left), aggregate surplus gains (top right), as well as capital (bottom left) and labor (bottom right) stock where $\chi = 0.5$. In the top left and the two bottom figures, the solid curves are for the benchmark scenario where $\chi = 0.5$, and the dashed curves are the counterfactual scenarios where no agencies are allowed to operate. The blue curves are for region R and green curves are for region P. In the top right figure, aggregate surplus is calculated as $\sum_r y_r - \frac{1}{2}\phi H_a^2$, and we then report $(Surplus - Surplus^{no\ LA})/Surplus^{no\ LA}$. In the bottom left figure, the blue curve represents capital accumulation in region R and the green curve is for region P. In the bottom right diagram, the green curve is for skilled labor and blue curve is for unskilled workers, both in region R. Note that the dotted blue curve is for the stock of unskilled labor in region R, i.e.,H_{UR} whereas the solid blue curve is for the total unskilled labor *available* to region R, i.e., $H_{UR} + H_a$.

FIGURE 12: HIGHER FIXED COSTS



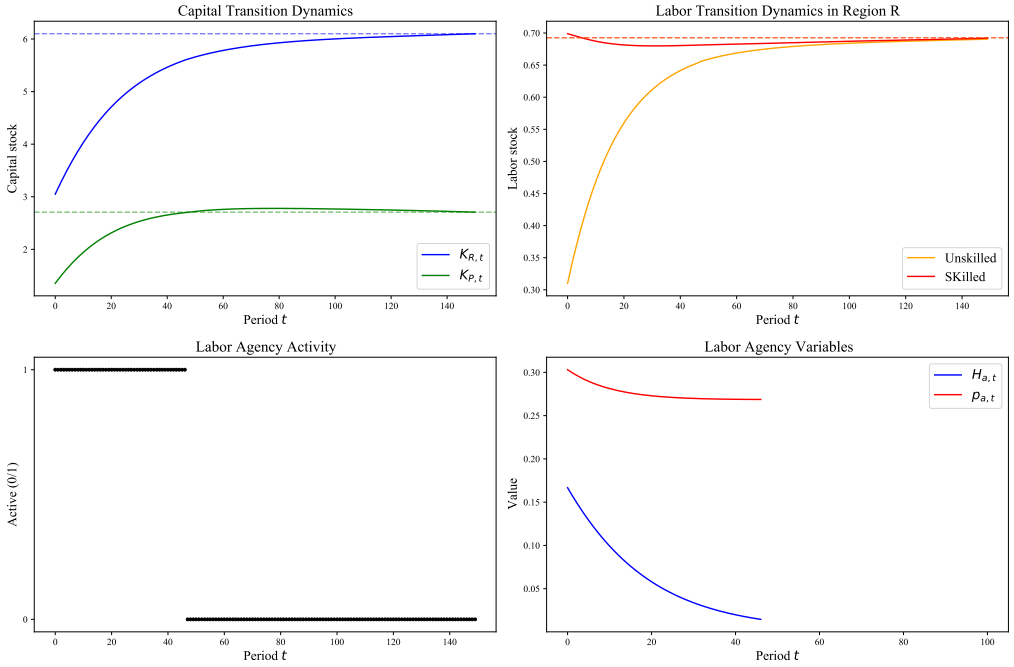
Notes: The figure shows the transition dynamics of capital accumulation, labor mobility, as well as agency variables by taking $\chi = 0.5$ and $F = 10^{-2}$. Active is a binary variable which equals to one if agencies operate. In the bottom right diagram I show how the amount of agency workers H_a (blue curve) and price that firms pay to agencies p_a (red curve) evolves over time (while agencies are active).

FIGURE 13: HIGHER FIXED COSTS: UNSKILLED WAGES AND SURPLUS GAINS

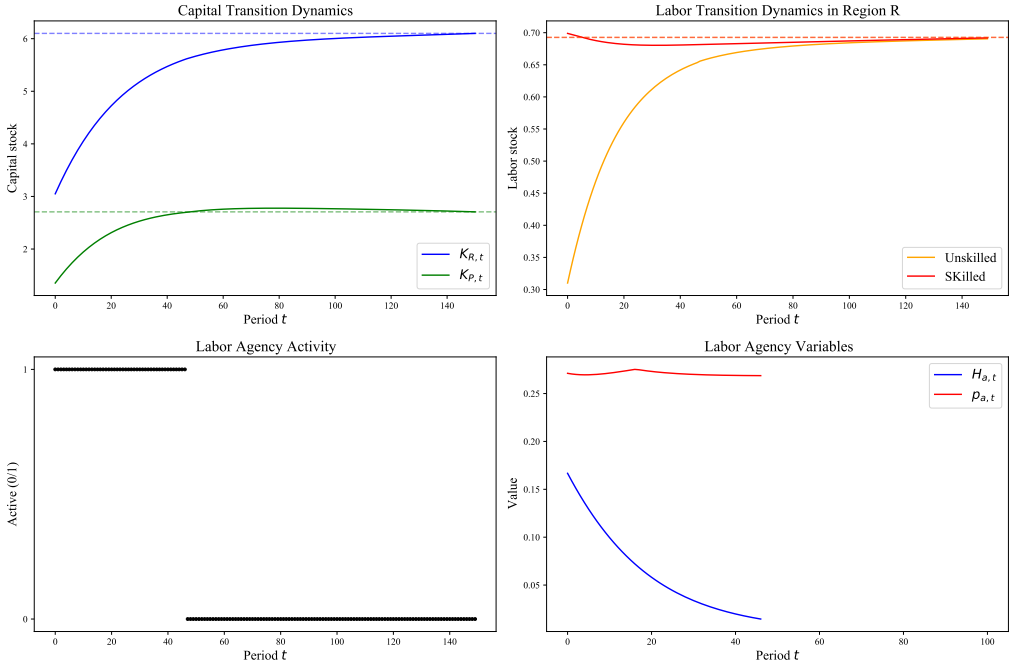


Notes: The top panel shows the dynamics of unskilled wages where $\chi = 0.5$ and $F = 10^{-2}$. The solid curves are for the benchmark scenario where $\chi = 0.5$, and the dashed curves are the counterfactual scenarios where no agencies are allowed to operate. The blue curves are for region R and green curves are for region P. The bottom panel shows aggregate surplus gains with $\chi = 0.5$ and $F = 10^{-2}$ contrasted with the case where $F = 10^{-3}$. The aggregate surplus is calculated as $\sum_r y_r - \frac{1}{2}\phi H_a^2$. The red solid curve is for $F = 10^{-3}$ and the blue dashed curve is for $F = 10^{-2}$.

Figure 14: Agency Competition

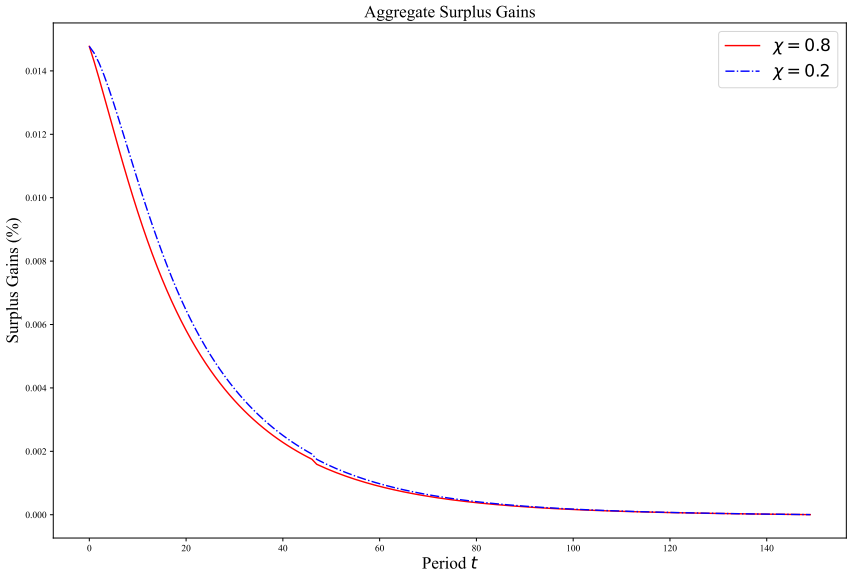


(i) $\chi = 0.8$

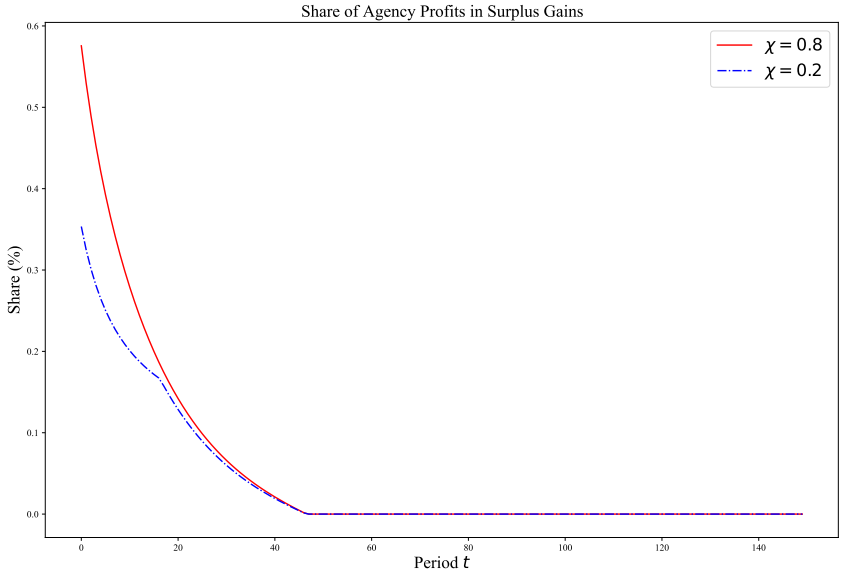


(ii) $\chi = 0.2$

Notes: The figure shows the transition dynamics of capital accumulation, labor mobility, and agency variables for $\chi = 0.8$ (top) and $\chi = 0.2$ (bottom). "Active" equals 1 when agencies operate. In the bottom-right diagram of each panel I show how the amount of agency workers H_a (blue) and the price firms pay to agencies p_a (red) evolve over time (while agencies are active).



(i) Surplus Gains

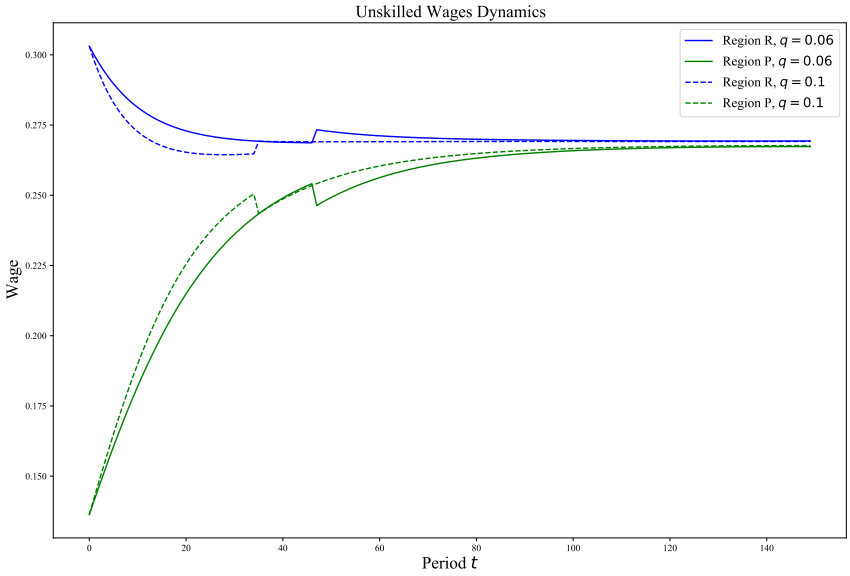


(ii) Share of Agency Profits

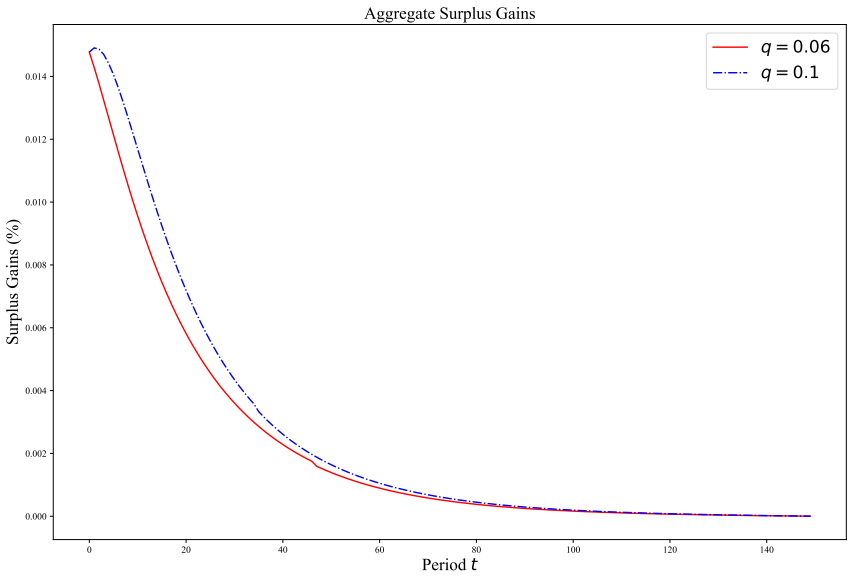
Notes: The figure shows the dynamics of aggregate surplus gains (top) and the share of agency profits in surplus gains (bottom). The aggregate surplus is calculated as $\sum_r y_r - \frac{1}{2}\phi H_a^2$. The red solid curve is for $\chi = 0.8$ and the blue dashed curve is for $\chi = 0.2$.

Figure 15: Agency Competition: Aggregate Surplus Gains

FIGURE 16: AGENCY WORKER TRANSITION: AGGREGATE SURPLUS GAINS



(i) UNSKILLED WAGES



(ii) AGGREGATE GAINS

Notes: The figure shows the dynamics of unskilled wages (top) and aggregate surplus gains (bottom) with $q = 0.1$ (contrasted with $q = 0.06$). The aggregate surplus is calculated as $\sum_r y_r - \frac{1}{2}\phi H_a^2$. In the top panel, the solid curves are for the benchmark scenario where $q = 0.06$ (the benchmark value), and the dashed curves are for $q = 0.1$. In the bottom panel, the red solid curve is for $q = 0.06$ and the blue dashed curve is for $q = 0.1$.

References

Atkin, D. (2016). Endogenous skill acquisition and export manufacturing in mexico. *American Economic Review*, 106(8):2046–85.

Atkin, D., Bernadac, B., Donaldson, D., Garg, T., and Huneeus, F. (2024). The incidence of distortions. Working paper.

Autor, D. H. (2003). Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing. *Journal of Labor Economics*, 21(1):1–266.

Banerjee, A. V. and Moll, B. (2010). Why does misallocation persist? *American Economic Journal: Macroeconomics*, 2(1):189–206.

Banerjee, A. V. and Newman, A. F. (1998). Information, the dual economy, and development. *Review of Economic Studies*, 65(4):631–653.

Baqaee, D. R. and Farhi, E. (2019). Productivity and misallocation in general equilibrium. *Quarterly Journal of Economics*, 135(1):105–163.

Barro, R. J., Mankiw, N. G., and Sala-I-Martin, X. (1995). Capital mobility in neoclassical models of growth. *American Economic Review*, 85(1):103–115.

Bau, N. and Matray, A. (2023). Misallocation and capital market integration: Evidence from india. *Econometrica*, 91(1):67–106.

Bertrand, M., Hsieh, C.-T., and Tsivanidis, N. (2024). Contract labor and firm growth in india. Working paper.

Bilal, A. and Lhuillier, H. (2025). Outsourcing, inequality, and aggregate output. Working paper.

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225:200–230.

Chen, S., Oliva, P., and Zhang, P. (2022). The effect of air pollution on migration: Evidence from china. *Journal of Development Economics*, 156:102833.

Davila, E. and Schaab, A. (2024). Welfare accounting. Working paper.

Ebenstein, A. and Zhao, Y. (2015). Tracking rural-to-urban migration in china: Lessons from the 2005 inter-census population survey. *Population Studies*, 69(3):337–353.

Estefan, A., Gerhard, R., Kaboski, J. P., Kondo, I. O., and Qian, W. (2024). Outsourcing policy and worker outcomes. Working paper.

Fang, M. and Huang, Z. (2022). Migration, housing constraints, and inequality: A quantitative analysis of china. *Labour Economics*, 78:102200.

Felix, M. and Wong, M. B. (2024). The reallocation effects of domestic outsourcing. Working paper.

Garicano, L., Lelarge, C., and Reenen, J. V. (2016). Firm size distortions and the productivity distribution: Evidence from france. *American Economic Review*, 106(11):3439–79.

Gennaioli, N., Porta, R. L., Lopez-De-Silanes, F., and Shleifer, A. (2014). Growth in regions. *Journal of Economic Growth*, 19:259–309.

Ghatak, M. (2015). Theories of poverty traps and anti-poverty policies. *World Bank Economic Review*, 29:s77–s105.

Ghatak, M. and Mookherjee, D. (2024). Misallocating misallocation? Working paper.

Harris, H. J. (1982). *The Right to Manage: Industrial Relations Policies of American Business in the 1940s*. University of Wisconsin Press.

Hatton, E. (2011). *The Temp Economy: From Kelly Girls to Permatemps in Postwar America*. Temple University Press.

Holmes, T. J. and Snider, J. T. (2011). A theory of outsourcing and wage decline. *American Economic Journal: Microeconomics*, 3:38–59.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly Journal of Economics*, 124(4):1403–1448.

Huang, P. C. C. (2017). Dispatch work in china: A study from case records, part i. *Modern China*, 43(3):247–287.

Itskhoki, O. and Moll, B. (2019). Optimal development policies with financial frictions. *Econometrica*, 87(1):139–173.

Kalleberg, A. L. (2011). *Good Jobs, Bad Jobs: The Rise of Polarized and Precarious Employment Systems in the United States, 1970s-2000s*. Russell Sage Foundation.

Kleinman, B., Liu, E., and Redding, S. J. (2023). Dynamic spatial general equilibrium. *Econometrica*, 91(2):385–424.

Lau, L. J., Qian, Y., and Roland, G. (2000). Reform without losers: An interpretation of china's dual-track approach to transition. *Journal of Political Economy*, 108(1):120–143.

Lewis, W. A. (1954). Economic development with unlimited supplies of labour. *The Manchester School*, 22(2):139–191.

Lin, J. Y. (1992). Rural reforms and agricultural growth in china. *American Economic Review*, 82(1):34–51.

Liu, G. (2014). Private employment agencies and labor dispatch in china. ILO Working paper.

Lucas, R. E. (1990). Why doesn't capital flow from rich to poor countries. *American Economic Review: Papers and Proceedings*, 80(2):92–93.

Meng, X. (2012). Labor market outcomes and reforms in china. *Journal of Economic Perspectives*, 26(4):75–102.

Moll, B. (2014). Productivity losses from financial frictions: Can self-financing undo capital misallocation? *American Economic Review*, 104(10):3186–3221.

Ngai, L. R., Pissarides, C. A., and Wang, J. (2019). China's mobility barriers and employment allocations. *Journal of the European Economic Association*, 17(5):1617–1653.

Restuccia, D. and Rogerson, R. (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics*, 11:707–720.

Rozelle, S. and Hell, N. (2020). *Invisible China: How the Urban-Rural Divide Threatens China's Rise*. The University of Chicago Press.

Schumpeter, J. A. (1934). *The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle*. Harvard University Press, Cambridge, MA.

Solow, R. M. (1956). A contribution to the theory of economic growth. *Quarterly Journal of Economics*, 70(1):65–94.

Sraer, D. and Thesmar, D. (2023). How to use natural experiments to estimate misallocation. *American Economic Review*, 113(4):906–938.

Tombe, T. and Zhu, X. (2019). Trade, migration, and productivity: A quantitative analysis of china. *American Economic Review*, 109(5):1843–1872.

Tu, W. (2017). The changing role of government towards labour. In Brown, W. and Kai, C., editors, *The Emerging Industrial Relations in China*, pages 112–140. Cambridge University Press, Cambridge.

Wang, F. (2019). *Labor Dispatch in China: Three Essays*. PhD thesis, Rutgers, The State University of New Jersey.

Yan, W. (2023). Evolution and the impetus mechanisms of the labour dispatch in china. *World Economy*, 11:128–154.

Young, A. (2000). The razor's edge: Distortions and incremental reform in the people's republic of china. *Quarterly Journal of Economics*, 115(4):1091–1135.

A Proofs

A.1 Proof of Proposition 1

By imposing $H_{srt} = H_{srt-1} = H^*_{sr}$ in steady state, equation (11) becomes

$$H^*_{sr} = (\nu^*_s)^{\frac{1}{1-\tau_s}} \hat{H}^*_{sr}. \tag{23}$$

Then we have

$$\frac{H^*_{sR}}{H^*_{sP}} = \frac{\hat{H}^*_{sR}}{\hat{H}^*_{sP}} = \left(\frac{A_R (K^*_R)^{\alpha} \overline{L}_R^{1-\alpha-\beta-\gamma}}{A_P (K^*_P)^{\alpha} \overline{L}_P^{1-\alpha-\beta-\gamma}} \right)^{\frac{1}{1-\beta-\gamma}} \tag{24}$$

for both $s \in \{R, P\}$. Then imposing $c_{rt+1} = c_t$ in the Euler equation, we obtain

$$\frac{A_R (K^*_R)^{\alpha-1} (H^*_{UR})^{\beta} (H^*_{SR})^{\gamma} \overline{L}_R^{1-\alpha-\beta-\gamma}}{A_P (K^*_P)^{\alpha-1} (H^*_{UP})^{\beta} (H^*_{SP})^{\gamma} \overline{L}_P^{1-\alpha-\beta-\gamma}} = 1 \tag{25}$$

Using the above two equations, we then have

$$\frac{K^*_R}{K^*_P} = \frac{H^*_{UR}}{H^*_{UP}} = \frac{H^*_{SR}}{H^*_{SP}} = \left(\frac{A_R}{A_P} \right)^{\frac{1}{1-\alpha-\beta-\gamma}} \frac{L_R}{L_P}. \tag{26}$$

A.2 Proof of Proposition 2

Suppose that H_{UP} is large enough so that the constraint $H_a \leq H_{UP}$ is never binding. Then we have two cases to consider: first if $p_a < w^U_R$, then we can first log-transform the objective function in (18) and take the first order conditions to obtain

$$\begin{aligned} \chi \frac{H_a}{W(p_a, H_a)} - (1-\chi) \frac{H_a}{[\tilde{\pi}_R(K, p_a, H_a) - \pi_R(K)]} &= 0; \\ \chi \frac{(p_a - w^U_P) - \phi'(H_a)}{W(p_a, H_a)} + (1-\chi) \frac{A_R K_A^{\alpha} \beta (H_U + H_a)^{\beta-1} (H_S)^{\gamma} L_R^{1-\alpha-\beta-\gamma} - p_a}{[\tilde{\pi}_R(K, p_a, H_a) - \pi_R(K)]} &= 0. \end{aligned}$$

Note also from the firm's first order condition, we obtain $A_R K_A^{\alpha} \beta (H_U + H_a)^{\beta-1} (H_S)^{\gamma} L_R^{1-\alpha-\beta-\gamma} = w^U_R$. Then these two conditions then yield

$$w^U_R - w^U_P = \phi'(H^*_a). \tag{27}$$

The price p_a will determined through the first equation which then gives

$$(1-\chi) W(p_a, H^*_a) = \chi \left[\tilde{\pi}_R(K, p_a, H^*_a) - \pi_R(K) \right], \tag{28}$$

conditional on $p_a < w^U_R$. If instead $p_a = w^U_R$, we can determine optimal H_a by taking the first order condition with respect to H_a and obtain the second equation in the above equation system. Plugging in $p_a = w^U_R$ still yields $w^U_R - w^U_P = \phi'(H^*_a)$.

Lastly, for firms to hire local workers as well, we need to impose

$$A_R K_R^{\alpha} \overline{L}_R^{1-\alpha-\beta-\gamma} \frac{\beta}{w_R^U} \left(\frac{w_R^U}{w_R^S}\right)^{\gamma} \left(\frac{\gamma}{\beta}\right)^{\gamma} > (\phi')^{-1}(w_R^U - w_P^U) \equiv H_a^*. \tag{29}$$

B Gross Flow versus Net Flow

Denote the total number of cities as M. First consider the following two regression models: first, based on all $M(M-1)$ directional city pairs,

$$y_{mnt}^G = \beta_1 c_{mnt} + \beta_2 c_{mnt} \times d_{nm} + r_{mnt}, \tag{30}$$

where y_{mnt}^G is the gross flow from city m to city n at year t, $c_{mnt} = c_{nmt}$ is a binary variable, and $d_{nm} = -d_{mn}$ is the ex-ante wage difference. Then based on all $\frac{M(M-1)}{2}$ non-directional city pairs (with $m < n$),

$$y_{mnt}^N = \gamma_1 c_{mnt} + \gamma_2 c_{mnt} \times d_{nm} + u_{mnt}, \tag{31}$$

where y_{mnt}^N is the net flow from city m to city n, i.e., $y_{mnt}^N = y_{mnt}^G - y_{nmt}^G$. For notational purposes, let us define $x_{mnt} \equiv c_{mnt} \times d_{nm}$.

Let us examine the gross flow model (30), where our estimates can be obtained from:

$$\begin{bmatrix} \sum c_{mnt}^2 & \sum c_{mnt} x_{mnt} \\ \sum c_{mnt} x_{mnt} & \sum x_{mnt}^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} \sum c_{mnt} y_{mnt}^G \\ \sum x_{mnt} y_{mnt}^G \end{bmatrix} \tag{32}$$

It is useful to note that $\sum c_{mnt} x_{mnt} = 0$, then it follows that

$$\hat{\beta}_1 = \frac{\sum c_{mnt} y_{mnt}^G}{\sum c_{mnt}^2}, \hat{\beta}_2 = \frac{\sum x_{mnt} y_{mnt}^G}{\sum x_{mnt}^2}.$$

Then for the net flow model (31), our estimates can be computed by

$$\begin{bmatrix} \sum_{m<n} c_{mnt}^2 & \sum_{m<n} c_{mnt} x_{mnt} \\ \sum_{m<n} c_{mnt} x_{mnt} & \sum_{m<n} x_{mnt}^2 \end{bmatrix} \begin{bmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{bmatrix} = \begin{bmatrix} \sum_{m<n} c_{mnt} y_{mnt}^N \\ \sum_{m<n} x_{mnt} y_{mnt}^N \end{bmatrix} \tag{33}$$

Note first that $\sum_{m<n} c_{mnt}^2 = \frac{1}{2} \sum c_{mnt}^2$ and $\sum_{m<n} x_{mnt}^2 = \frac{1}{2} \sum x_{mnt}^2$. Furthermore,

$$\sum_{m<n} x_{mnt} y_{mnt}^N = \sum_{m<n} x_{mnt} \left[y_{mnt}^G - y_{nmt}^G\right] = \sum_{m<n} x_{mnt} y_{mnt}^G + \sum_{m<n} x_{nmt} y_{nmt}^G = \sum x_{mnt} y_{mnt}^G.$$

From (33), we have

$$\begin{aligned}
\hat{\gamma}_1 &= \frac{\left(\sum_{m<n} c_{mnt} y^N_{mnt}\right)\left(\sum_{m<n} x^2_{mnt}\right) - \left(\sum_{m<n} x_{mnt} y^N_{mnt}\right)\left(\sum_{m<n} c_{mnt} x_{mnt}\right)}{\left(\sum_{m<n} c^2_{mnt}\right)\left(\sum_{m<n} x^2_{mnt}\right) - \left(\sum_{m<n} c_{mnt} x_{mnt}\right)^2} \\
&= \frac{\left(\sum_{m<n} c_{mnt} y^N_{mnt}\right)\left(\sum_{m<n} x^2_{mnt}\right) - \left(\sum x_{mnt} y^N_{mnt}\right)\left(\sum_{m<n} c_{mnt} x_{mnt}\right)}{\left(\sum_{m<n} c^2_{mnt}\right)\left(\sum_{m<n} x^2_{mnt}\right) - \left(\sum_{m<n} c_{mnt} x_{mnt}\right)^2} \\
&= \frac{\left(\sum_{m<n} c_{mnt} y^N_{mnt}\right)\left(\sum_{m<n} x^2_{mnt}\right) - \hat{\beta}_2 \left(\sum x^2_{mnt}\right)\left(\sum_{m<n} c_{mnt} x_{mnt}\right)}{\left(\sum_{m<n} c^2_{mnt}\right)\left(\sum_{m<n} x^2_{mnt}\right) - \left(\sum_{m<n} c_{mnt} x_{mnt}\right)^2} \\
&= \frac{\left(\sum_{m<n} c_{mnt} y^N_{mnt}\right) - 2\hat{\beta}_2 \left(\sum_{m<n} c_{mnt} x_{mnt}\right)}{\left(\sum_{m<n} c^2_{mnt}\right) - \frac{\left(\sum_{m<n} c_{mnt} x_{mnt}\right)^2}{\sum_{m<n} x^2_{mnt}}}
\end{aligned}$$

Furthermore, note that

$$\sum_{m<n} c_{mnt} y^N_{mnt} = \sum_{m<n} c_{mnt}\left[y^G_{mnt} - y^G_{nmt}\right] = \sum s_{mn} c_{mnt} y^G_{mnt},$$

where we define $s_{mn} = 1$ if $m > n$, and $s_{mn} = -1$ if $m < n$. Using OLS of (30) we have

$$\sum s_{mn} c_{mnt} y^G_{mnt} = \sum s_{mn} c^2_{mnt} \hat{\beta}_1 + 2\hat{\beta}_2 \sum_{m<n} c_{mnt} x_{mnt} + \sum s_{mn} c_{mnt} r_{mnt}.$$

Noting $\sum s_{mn} c^2_{mnt} = 0$, we therefore have

$$\hat{\gamma}_1 = \frac{\sum_{m<n} c_{mnt}\left[r_{mnt} - r_{nmt}\right]}{\left(\sum_{m<n} c^2_{mnt}\right) - \frac{\left(\sum_{m<n} c_{mnt} x_{mnt}\right)^2}{\sum_{m<n} x^2_{mnt}}}$$

We can the plug this back into (33) and obtain

$$\hat{\gamma}_2 = 2\hat{\beta}_2 - \frac{\sum_{m<n} c_{mnt} x_{mnt}}{\left(\sum_{m<n} c^2_{mnt}\right)\left(\sum_{m<n} x^2_{mnt}\right) - \left(\sum_{m<n} c_{mnt} x_{mnt}\right)^2} \sum_{m<n} c_{mnt}\left[r_{mnt} - r_{nmt}\right]. \quad (34)$$

If we can assume additionally,

$$\mathbb{E}\left[r_{mnt} | c_{mnt}, d_{nm}\right] = \mathbb{E}\left[r_{nmt} | c_{mnt}, d_{nm}\right], \quad (35)$$

then we have

$$\gamma_1 = 0, \gamma_2 = 2\beta_2.$$

Adding Fixed Effects Let us now consider adding city pair fixed effects and time fixed effects into the model. Specifically, (30) becomes

$$y^G_{mnt} = \beta_1 c_{mnt} + \beta_2 x_{mnt} + \lambda_{mn} + \delta_t + r_{mnt},$$

where λ_{mn} are the directional city pair fixed effects. (31) now becomes

$$y^N_{mnt} = \gamma_1 c_{mnt} + \gamma_2 x_{mnt} + \lambda_{mn} + \delta_t + u_{mnt}.$$

By Frisch-Waugh-Lovell theorem, we know that our OLS estimate of $\beta_1, \beta_2, \gamma_1, \gamma_2$ can be obtained from the following regression:

$$\begin{aligned} \tilde{y}^G_{mnt} &= \beta_1 \tilde{c}_{mnt} + \beta_2 \tilde{x}_{mnt} + \tilde{r}_{mnt}; \\ \check{y}^N_{mnt} &= \gamma_1 \check{c}_{mnt} + \gamma_2 \check{x}_{mnt} + \check{u}_{mnt}, \end{aligned}$$

where we take the within transformation for variables on city pair and time fixed effects.

Let us first derive the relationships between the demeaned variables. The first observation is that

$$\tilde{c}_{mnt} = \tilde{c}_{nmt} = \check{c}_{mnt}, \tag{36}$$

which follows from the fact that $c_{mnt} = c_{nmt}$ (therefore when subtracting the city pair mean and time mean, the demeaned c_{mnt} is still invariant for direction). We also have

$$\tilde{x}_{mnt} = -\tilde{x}_{nmt}. \tag{37}$$

To see this, note that

$$\tilde{x}_{mnt} = x_{mnt} - \bar{x}_{mn.} - \bar{x}_{.t} + \bar{x} = -(x_{nmt} - \bar{x}_{nm.}) = -\tilde{x}_{nmt},$$

which follows from the fact that $\bar{x}_{.t} = 0$ and $\bar{x} = 0$. It then follows that we have

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum \tilde{c}_{mnt} \tilde{y}^G_{mnt}}{\sum \tilde{c}^2_{mnt}}; \\ \hat{\beta}_2 &= \frac{\sum \tilde{x}_{mnt} \tilde{y}^G_{mnt}}{\sum \tilde{x}^2_{mnt}}. \end{aligned}$$

It is also useful to establish the relationship between demeaned variables in the gross and net flow regressions: first note that

$$\check{x}_{mnt} = x_{mnt} - \bar{x}^N_{mn.} - \bar{x}^N_{.t} + \bar{x}^N = x_{mnt} - \bar{x}_{mn.} - \bar{x}^N_{.t} + \bar{x}^N = \tilde{x}_{mnt} - \theta_t, \tag{38}$$

where I denote $\theta_t \equiv \bar{x}^N_{.t} - \bar{x}^N$. Also by definition,

$$\begin{aligned} \check{y}^N_{mnt} &= y^N_{mnt} - \bar{y}^N_{mn.} - \bar{y}^N_{.t} + \bar{y}^N \\ &= y^G_{mnt} - y^G_{nmt} - \left(\bar{y}^G_{mn.} - \bar{y}^G_{nm.}\right) - \psi_t, \end{aligned}$$

which follows from the fact that $\bar{y}^N_{mn.} = \bar{y}^G_{mn.} - \bar{y}^G_{nm.}$ and denote $\psi_t \equiv \bar{y}^N_{.t} - \bar{y}^N$. Furthermore,

$$\tilde{y}^G_{mnt} - \tilde{y}^G_{nmt} = y^G_{mnt} - y^G_{nmt} - \left(\bar{y}^G_{mn.} - \bar{y}^G_{nm.}\right) - \left(\bar{y}^G_{.t} - \bar{y}^G_{.t}\right) + \left(\bar{y}^G - \bar{y}^G\right).$$

Then we establish:

$$\check{y}^N_{mnt} = \tilde{y}^G_{mnt} - \tilde{y}^G_{nmt} - \psi_t. \tag{39}$$

Then a few intermediate step follows. First,

$$\begin{aligned}
\sum_{m<n} \check{x}_{mnt}\check{y}^N_{mnt} &= \sum_{m<n} \left[\tilde{x}_{mnt} - \theta_t\right] \check{y}^N_{mnt} = \sum_{m<n} \tilde{x}_{mnt}\check{y}^N_{mnt} - \sum_{m<n} \theta_t \check{y}^N_{mnt} = \sum_{m<n} \tilde{x}_{mnt}\check{y}^N_{mnt} \\
&= \sum_{m<n} \tilde{x}_{mnt} \left[\tilde{y}^G_{mnt} - \tilde{y}^G_{nmt} - \psi_t\right] = \sum_{m<n} \tilde{x}_{mnt} \left[\tilde{y}^G_{mnt} - \tilde{y}^G_{nmt}\right] - \sum_{m<n} \tilde{x}_{mnt}\psi_t \\
&= 2\hat{\beta}_2 \sum_{m<n} \tilde{x}^2_{mnt} - \sum_{m<n} \tilde{x}_{mnt}\psi_t = 2\hat{\beta}_2 \sum_{m<n} \left(\check{x}^2_{mnt} + \theta^2_t\right) - \sum_{m<n} \theta_t\psi_t.
\end{aligned}$$

where the first line follows from (38) and the fact that $\sum_{m<n} \check{y}^N_{mnt} = 0$ for any given t, and the second line follows from (39). Second, similar to our no fixed effects model before,

$$\sum_{m<n} \check{c}_{mnt}\check{y}^N_{mnt} = \sum_{m<n} \check{c}_{mnt} \left[\tilde{y}^G_{mnt} - \tilde{y}^G_{nmt}\right] = \sum_{m<n} \tilde{c}_{mnt}[\tilde{y}^G_{mnt} - \tilde{y}^G_{nmt}] = \sum s_{mn}\tilde{c}_{mnt}\tilde{y}^G_{mnt}.$$

Then plugging in the within regression for gross flows, we have

$$\begin{aligned}
\sum s_{mn}\tilde{c}_{mnt}\tilde{y}^G_{mnt} &= \hat{\beta}_1 \sum s_{mn}\tilde{c}^2_{mnt} + \hat{\beta}_2 \sum s_{mn}\tilde{c}_{mnt}\tilde{x}_{mnt} + \sum s_{mn}\tilde{c}_{mnt}\tilde{r}_{mnt} \\
&= 2\hat{\beta}_2 \sum_{m<n} \tilde{c}_{mnt}\tilde{x}_{mnt} + \sum_{m<n} \tilde{c}_{mnt} \left[\tilde{r}_{mnt} - \tilde{r}_{nmt}\right].
\end{aligned}$$

Furthermore, given $\sum_{m<n} \check{c}_{mnt} = 0$ for any given t, we can see that

$$\sum_{m<n} \tilde{c}_{mnt}\tilde{x}_{mnt} = \sum_{m<n} \check{c}_{mnt}[\check{x}_{mnt} + \theta_t] = \sum_{m<n} \check{c}_{mnt}\check{x}_{mnt}.$$

The OLS estimate of γ_1 is

$$\hat{\gamma}_1 = \frac{\left(\sum_{m<n} \check{c}_{mnt}\check{y}^N_{mnt}\right)\left(\sum_{m<n} \check{x}^2_{mnt}\right) - \left(\sum_{m<n} \check{x}_{mnt}\check{y}^N_{mnt}\right)\left(\sum_{m<n} \check{c}_{mnt}\check{x}_{mnt}\right)}{\left(\sum_{m<n} \check{c}^2_{mnt}\right)\left(\sum_{m<n} \check{x}^2_{mnt}\right) - \left(\sum_{m<n} \check{c}_{mnt}\check{x}_{mnt}\right)^2}$$

Denote $R \equiv \sum_{m<n} \tilde{c}_{mnt} \left[\tilde{r}_{mnt} - \tilde{r}_{nmt}\right]$. Then the numerator is

$$\begin{aligned}
&\left[2\hat{\beta}_2 \sum_{m<n} \tilde{c}_{mnt}\tilde{x}_{mnt} + R\right] \sum_{m<n} \check{x}_{mnt}^2 - \left[2\hat{\beta}_2 \sum_{m<n} \tilde{x}_{mnt}^2 - \sum_{m<n} \tilde{x}_{mnt}\psi_t\right] \sum_{m<n} \check{c}_{mnt}\check{x}_{mnt} \\
= &\sum_{m<n} \check{c}_{mnt}\check{x}_{mnt} \left[2\hat{\beta}_2 \sum_{m<n} \left(\check{x}_{mnt}^2 - \tilde{x}_{mnt}^2\right) + \sum_{m<n} \tilde{x}_{mnt}\psi_t\right] + R \sum_{m<n} \check{x}_{mnt}^2 \\
= &\sum_{m<n} \check{c}_{mnt}\check{x}_{mnt} \left[2\hat{\beta}_2 \sum_{m<n} \left(\check{x}_{mnt}^2 - \tilde{x}_{mnt}^2\right) + \sum_{m<n} \theta_t\psi_t\right] + R \sum_{m<n} \check{x}_{mnt}^2 \\
= &\sum_{m<n} \check{c}_{mnt}\check{x}_{mnt} \left[\sum_{m<n} \theta_t\psi_t - 2\hat{\beta}_2 \sum_{m<n} \theta_t^2\right] + R \sum_{m<n} \check{x}_{mnt}^2.
\end{aligned}$$

Hence

$$\hat{\gamma}_1 = \frac{\sum_{m<n} \check{c}_{mnt}\check{x}_{mnt} \left[\sum_{m<n} \theta_t\psi_t - 2\hat{\beta}_2 \sum_{m<n} \theta_t^2\right] + R \sum_{m<n} \check{x}_{mnt}^2}{\left(\sum_{m<n} \check{c}_{mnt}^2\right)\left(\sum_{m<n} \check{x}_{mnt}^2\right) - \left(\sum_{m<n} \check{c}_{mnt}\check{x}_{mnt}\right)^2} \tag{40}$$

The OLS estimate of γ_2 is

$$\hat{\gamma}_2 = \frac{\left(\sum_{m<n} \check{c}_{mnt}^2\right)\left(\sum_{m<n} \check{x}_{mnt}\check{y}_{mnt}^N\right) - \left(\sum_{m<n} \check{c}_{mnt}\check{x}_{mnt}\right)\left(\sum_{m<n} \check{c}_{mnt}\check{y}_{mnt}^N\right)}{\left(\sum_{m<n} \check{c}_{mnt}^2\right)\left(\sum_{m<n} \check{x}_{mnt}^2\right) - \left(\sum_{m<n} \check{c}_{mnt}\check{x}_{mnt}\right)^2}$$

Plugging in the above relationships, the numerator can be written as

$$\left(\sum_{m<n} \check{c}_{mnt}^2\right)\left(2\hat{\beta}_2 \sum_{m<n} \left(\check{x}_{mnt}^2 + \theta_t^2\right) - \sum_{m<n} \theta_t\psi_t\right) - \left(\sum_{m<n} \check{c}_{mnt}\check{x}_{mnt}\right)\left(2\hat{\beta}_2 \sum_{m<n} \check{c}_{mnt}\check{x}_{mnt} + R\right).$$

Then we have

$$\hat{\gamma}_2 = 2\hat{\beta}_2 + \frac{\left(\sum_{m<n} \check{c}_{mnt}^2\right)\left(2\hat{\beta}_2 \sum_{m<n} \theta_t^2 - \sum_{m<n} \theta_t\psi_t\right) - R \sum_{m<n} \check{c}_{mnt}\check{x}_{mnt}}{\left(\sum_{m<n} \check{c}_{mnt}^2\right)\left(\sum_{m<n} \check{x}_{mnt}^2\right) - \left(\sum_{m<n} \check{c}_{mnt}\check{x}_{mnt}\right)^2}.$$

Therefore, for $\hat{\gamma}_1 = 0$ and $\hat{\gamma}_2 = 2\hat{\beta}_2$, we require (i)

$$R \equiv \sum_{m<n} \tilde{c}_{mnt} \left[\tilde{r}_{mnt} - \tilde{r}_{nmt}\right] = 0, \tag{41}$$

which is similar to our no fixed effects case; but additionally (ii)

$$\sum_{m<n} \theta_t\psi_t = 2\hat{\beta}_2 \sum_{m<n} \theta_t^2 \tag{42}$$