Moving Firms or Moving People? The Welfare Effects of Place-based Policies*

Nghiêm Q. Huỳnh†

June 18, 2023

Preliminary and Incomplete

Abstract

Do policies encouraging business development in impoverished regions improve welfare more effectively than those promoting migration out of such areas? This question is explored in the context of Vietnam, where the government has implemented tax incentives to attract firms to disadvantaged districts and relaxed the Ho Khau policy, which restricts internal migration. To quantify the welfare effects of these two policies, I develop a dynamic spatial general equilibrium model and identify its key parameters through triple difference-in-differences, difference-in-differences, and shift-share instrumental variable approaches. The findings from this analysis could provide valuable insights to policymakers worldwide grappling with the dilemma of moving firms or moving people.

Keywords: Place-based policy, migration, inequality, Vietnam

*I am deeply grateful to Sam Kortum, Lorenzo Caliendo, and Mark Rosenzweig for their invaluable guidance and unwavering support throughout this project. I extend my heartfelt thanks to Brian McCaig, Nguyen Viet Phong, Ho Van Bao, Michael Epprecht, and Nguyen Thi Hue for generously sharing their data and providing crucial insights to understand them. I would also like to express my appreciation to Costas Arkolakis, Charles Cai, Ana Cecilia Fieler, John Finlay, Tim Guinnane, Amit Khandelwal, Ryungha Oh, Michael Peters, and Bryan Stuart for their constructive feedback. Furthermore, I am grateful for the comments from Yale's Trade and Development Lunch participants. I sincerely appreciate the financial assistance from the Sylff EGC grants and the Yale Council on South East Asian Studies.

†Yale University, nghiem.huynh@yale.edu

1 Introduction

I explore whether policies promoting business establishment in underprivileged areas yield greater welfare benefits than those favoring migration out of such regions. This question resides at the nexus of spatial economics and development policy. While numerous studies have examined these policies separately, few have conducted comparative analyses within the same context and periods.

This paper aims to fill the existing gap by quantifying the welfare effects of two place-based policies within the unique context of Vietnam. In 2003, Vietnam revised its uniform enterprise tax schedules, providing preferential treatment to young establishments in underprivileged regions. Then, in 2005, the government relaxed the Ho Khau restrictions that previously limited migration across regions. This easing was less significant for those intending to establish permanent residence in the major cities.

To quantify the welfare effects of these policies, I proceed in three steps. First, I build a dynamic spatial general equilibrium with entrepreneurs and households whose welfares depend on policies. Next, I apply this model to the establishment and household data spanning two decades, encapsulating periods before and after the policy alterations. Lastly, I perform counterfactual exercises, disabling each policy change and both together, to estimate each's welfare effects, as well as their combined impact.

The model identifies three critical factors determining the welfare effects of spatial policies: firm entry elasticity, migration elasticity, and migration costs associated with the Ho Khau policy change. These elements respectively measure entrepreneurial response to tax changes, worker response to migration cost shifts, and the conversion of the Ho Khau policy to numerical migration costs for workers, which necessitates counterfactual analysis.

I exploit policy change variations to estimate each key parameter separately through derived reduced-form equations from the model. A triple diff-in-diff design exploiting variations across time, space, and firm age facilitates the identification of firm entry elasticity. A similar approach estimates the Ho Khau policy's associated migration costs using destination, origin, and time variations.

Since workers can react to policy changes by altering migration patterns or switching sectors within the same location, estimating migration elasticity requires an additional parameter – the sectoral elasticity governing sectoral choice. To estimate this parameter, I use the 2001 US-VN bilateral trade agreement (BTA) as an exogenous labor demand shock in a diff-in-diff design to instrument for sectoral wages. I then employ its shift-share, or Bartik version, as an instrument for local real wages to estimate migration elasticity.

Related literature This research contributes to the ongoing discourse on whether to relocate firms to impoverished regions or incentivize out-migration, an issue raised in Austin et al. (2018). In quantifying the effects of both policies in a single context, this paper extends three major literature branches, examining the impacts of place-based, migration, and trade policies.

First, this study explores place-based policies for firms targeting backward districts and special economic zones (SEZs). The growth in studies exploring the effects of place-based policies on firms has been consistent since Neumark and Simpson (2015)'s review. Significant contributions include Busso et al. (2013) and Kline and Moretti (2014) in the United States, and particularly relevant to this research are studies from developing countries, including Wang (2013) and Lu et al. (2019) using SEZs in China, Chaurey (2017) and Hasan et al. (2021) using backward districts in India, and Vu and Yamada (2022) using SEZs in Vietnam. This paper expands on these studies by conducting counterfactual analyses in general equilibrium (GE) using a dynamic model and introduces new estimation strategies to identify structural parameters amidst complex GE effects.

The study's unique aspect is its combination of diverse policy variations in a dynamic model, contributing to quantitative studies on place-based policies. The dynamic model here extends the work of Caliendo and Parro (2020) and Kleinman et al. (2023)[1] to link the intricate policy structures to welfare. On the estimation side, the study diverges from state-tax policy studies by Fajgelbaum et al. (2018) through a dynamic model and exploiting variations from one-time policy changes.[2]

Migration subsidy, a possible alternative to place-based policy, has shown promising results, especially in the development literature. Lagakos et al. (2023) estimates the welfare effects of migration subsidy to rural-urban migration using experimental variation conducted in Bryan et al. (2014). Tombe and Zhu (2019) estimate the effects of trade and migration cost reductions in China, some of which are related to the Hukou policy.

This study estimates migration costs exclusively associated with changes in Ho Khau policy, which is a version of the Hukou policy[3]. Additionally, Caliendo et al. (2021) examines the effects of trade and migration costs due to the EU expansion in a dynamic GE context. Nonetheless, current literature has yet to investigate the GE effects of migration

[1] These papers build on Caliendo et al. (2019) and Artuç et al. (2010). Fajgelbaum and Gaubert (2020) study optimal place-based policies under sorting of heterogenous workers.

[2] Recent studies on place-based policy, such as Heblich et al. (2022) in China and Atalay et al. (2023) in Turkey, underscore the importance of supply chains or input-output linkages. However, this study shifts focus instead to studying the place-based policy effects in the presence of another policy, the Ho Khau migration policy.

[3] Other empirical studies include Wang et al. (2021), Kinnan et al. (2018) on Hukou and China. Imbert et al. (2022) examine the effects of migrants on firms in China.

policy on workers and firms, and part of this study fills in this gap.

Lastly, this paper employs the US-VN bilateral trade agreement to estimate sectoral and migration elasticities. Previous literature, such as works by McCaig (2011), Fukase (2013), McCaig and Pavcnik (2018), and McCaig et al. (2022), argue the shock's exogenous nature and document its effects on local markets, workers, and firms. This study uses this shock to construct instrumental variables for labor demand at the sectoral and local levels, providing evidence of this policy's effects on sectoral choice and migration through specific channels.

I begin by providing the details of the context in Section 2. Section 3 discusses the key data sources while Section 4 lays out the model. Section 5 takes the model to the data, enabling counterfactual analysis in Section 6.

2 Context: Vietnam's policies in early 2000s

Vietnam provides a unique context for examining the effects of place-based and migration policies due to its significant growth and structural transformation since the early 2000s. In 2000, the average income of Vietnamese individuals was about 8% of that of Americans. However, by 2019, this figure had risen to 15%, as indicated in Figure A1.[4] Not only did Vietnam undergo remarkable growth during this period, but the government also implemented policies that aimed to influence the spatial distribution of economic activities within the country. This section outlines three such policies: a trade agreement with the US that boosts Vietnam's exports to the US, profit tax incentives to encourage firms to operate in disadvantaged districts, and a reform of the Ho Khau system that restricts internal mobility.

2.1 Trade policy: 2001 US-VN Bilateral Trade Aggrement

In 2001, Vietnam's status was upgraded from Column 2 U.S. tariffs to Most Favored Nation (MFN) status through the US-VN Bilateral Trade Agreement. This change is unlikely to result from endogenous industry efforts on either side, as both tariff schedules had existed long before 2001 (McCaig, 2011). Moreover, the tariff reductions were substantial and varied across industries, ranging from 0 to 63 percentage points, as illustrated at the 2-digit VSIC levels in Figure A3.

[4] McCaig and Pavcnik (2013) report that the share of employment in manufacturing grew from 8 to 14 percent between 2000 and 2009, while the share of agriculture drops from 69 to 53 percent.

The immediate and permanent reduction in tariffs resulted in a surge in manufacturing exports from Vietnam to the US, from about $0.7 Billion in 2000 to $10 Billion in 2008, as depicted in Figure A2. According to Fukase (2013), McCaig (2011), and McCaig and Pavcnik (2018), this export shock raised wages, reduced regional poverty, and increased formality. Consequently, this policy can serve as an exogenous variation to labor demand across industries and regions, allowing for estimating labor supply elasticity in later sections.

2.2 Place-based tax policy: 2003-2004 Enterprise Income Tax Law

The 2003 revisions to the Enterprise Income Tax Law by the National Assembly, as illustrated in Figure 1,[5] constitute a place-based policy that can influence economic activities across regions in Vietnam. The previous version of the law, effective from 1997 to 2003, imposed a 32% profit tax on all establishments, regardless of their location, age, or cohorts. Given that enterprise income taxes accounted for almost 40% of Vietnamese government tax revenue in 2000, surpassing all other taxes, such as value-added tax (22%) and personal income taxes (2%), any changes in the tax rate are highly significant (Shukla et al. (2011)).[6]

This study focuses on three major revisions to the law in 2003. First, the government reduced the tax rate for all operating establishments to 28%.

Second, the government defined a list of districts from among Vietnam's 600 districts that would receive tax incentives for post-2003 cohorts. Decree 164/2003/ND-CP classified districts based on their socioeconomic difficulties as Challenged (C), Burdened (B), or Advantaged (A) districts.

As demonstrated in Figure 1a, establishments that begin in or relocate to C or B districts receive lower tax rates starting from the first year they make profits in those areas. For instance, if a business enters a C district in 2005, it is indefinitely taxed at a 28% rate. Conversely, a counterpart in a B district pays no profit tax for the first two years, then 10% from three to eight years old, 20% from eight to ten years old, and 28% thereafter. Later, these variations in the tax schedule across cohorts, locations, and ages are used to identify key structural parameters.

Third, besides backward districts, the government amended the tax law in 2004 to include Special Economic Zones (SEZs) among the regions that receive preferential tax rates[7]. A manufacturing entrant in a cohort after 2004 receives the same tax schedule as

[5]For more details, see Law on Enterprise Income Tax No. 09/2003/QH11

[6]In 2010, enterprise income tax and VAT were the main sources of tax revenue (both at 32.4%).

[7]Decree 88/2004/TT-BTC

if it entered a C district, while a service entrant receives the schedule of B districts.

I compiled a list of all SEZs until 2021, including industrial parks and high-tech zones and their respective locations. While the size of SEZs varies, most are smaller than a district. SEZ analysis is limited to those established before 2004, as those established after could have been in response to favorable tax policies, particularly after 2008, when local governments, not the prime minister, were permitted to establish SEZs. Moreover, the majority of SEZs are located in A districts. Out of the 85 communes with at least one economic zone before 2004, only eight are located in C districts, while 17 are in B districts.

Due to the granular nature of SEZs, the district-zone level is chosen as the unit of analysis of this policy, as depicted in Figure 1b. The map displays each polygon as a commune, an administrative level below the district. The color of each commune corresponds to its socioeconomic status, such as C, B, A, or Zone. A commune is classified as Zone if it had at least one economic zone before 2004. For example, consider Cu Chi, a district in Ho Chi Minh City. Based on the 2003 tax law, it is classified as A, but one of its communes is labeled as Zone and colored Green in Figure 1b at the top of the bottom right corner. Since subsequent analysis aggregates data to the district-zone level, Cu Chi is treated as two distinct units: Cu Chi with Zone and Cu Chi without Zone.

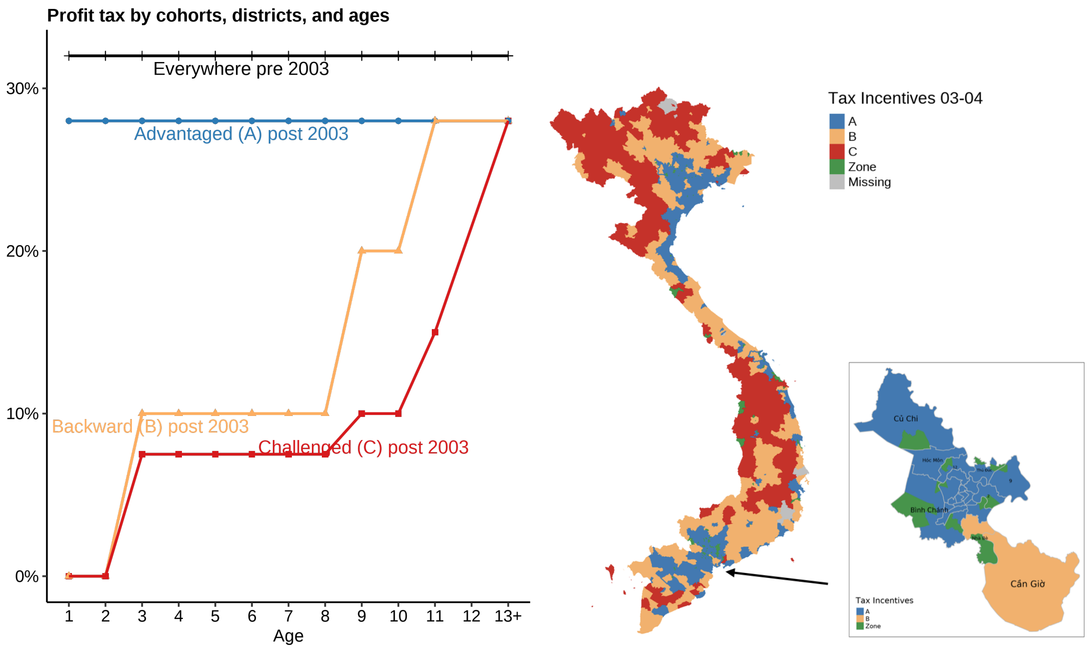
While the place-based policy offers the potential to uplift backward districts, it is susceptible to firms "cheating" by closing and reopening the same business in locations with favorable tax breaks. However, such behavior is unlikely to be practical due to the high cost of closing a business, extensive government audits, and regulations designed to prevent tax evasion and capital flight out of the country. Furthermore, as seen in Figure A7, which tracks exit rates across cohorts over time, there is no significant increase in exit rates in 2003, particularly among the youngest cohort, which entered in 2002 and has a strong incentive to avoid the 28% tax schedule.

The effects of the policy on the reallocation of activities within multi-entity firms may also be complicated. However, according to Table A1, 99% of firms in Vietnam in the early 2000s had only one establishment. If firms concentrate their operations in B or C establishments, the policy remains effective in stimulating local development and reducing spatial inequality.

2.3 Migration policy: 2005 Ho Khau relaxation

In 2005, the government began to relax the Ho Khau requirement, which regulates household registration and internal migration across provinces in Vietnam, similar to the Hukou

Figure 1: Profit tax on establishments varies across time, districts, and ages



(a) Tax schedule before and after 2003

(b) Map of district labels

Notes: District categories are from Decree 164/2003/ND-CP. The map is at the commune level in 2010. The zoomed-in map is Ho Chi Minh City.

in China[8]. Before this policy change, households living in a province other than their birthplace found it difficult to obtain permanent residency in their destination province, especially in centrally administered cities such as Ho Chi Minh City or Ha Noi. To obtain permanent residency, migrants were required to meet various prerequisites, including a moving certificate from the origin province and a land-use or house-ownership certificate, which ironically required a Ho Khau permanent status at the destination location (Liu and Meng (2019)).

Distinguishing temporary and permanent status primarily depends on access to amenities in the destination province. Temporary migrants often struggle to access healthcare and education for their children (World Bank Group and Vietnam Academy of Social Sciences, 2016). In 2005, the government eliminated numerous prerequisites and established a path to permanent residency based solely on proof of residence and income. As a result, migrants with stable employment can immediately apply for permanent residency

[8] The Ho Khau policy is less strict than the Hukhou policy because, unlike the Hukhou, it does not restrict individuals to the birth sector.

in all destination provinces except for the five centrally administered cities of Ha Noi, Hai Phong, Da Nang, Ho Chi Minh City, and Can Tho.[9]. In these major cities, temporary migrants must provide proof of a one-year minimum stay[10] although in practice, people typically stayed longer than one year before acquiring permanent status due to the lengthy bureaucratic process.

In short, the Ho Khau policy change lowers migration costs for migrants across all origin and destination locations, although the reduction is slightly less pronounced for the five major cities. To simplify the language, I will refer to these central cities as "Urban" and the rest of Vietnam as "Rural" while acknowledging that both categories encompass various urbanized regions. Unlike tariffs or taxes, the exact magnitude of this policy change remains unknown. Therefore, I will use a dynamic model and policy change to estimate the magnitude of changes in migration costs associated with the Ho Khau policy over time and their spatial variations.

3 Data and Facts

The following data and empirical facts provide insights into the economic trends in Vietnam from 2000 to 2019 and serve as the foundation for the coming model and its estimation.

3.1 Establishment-level data

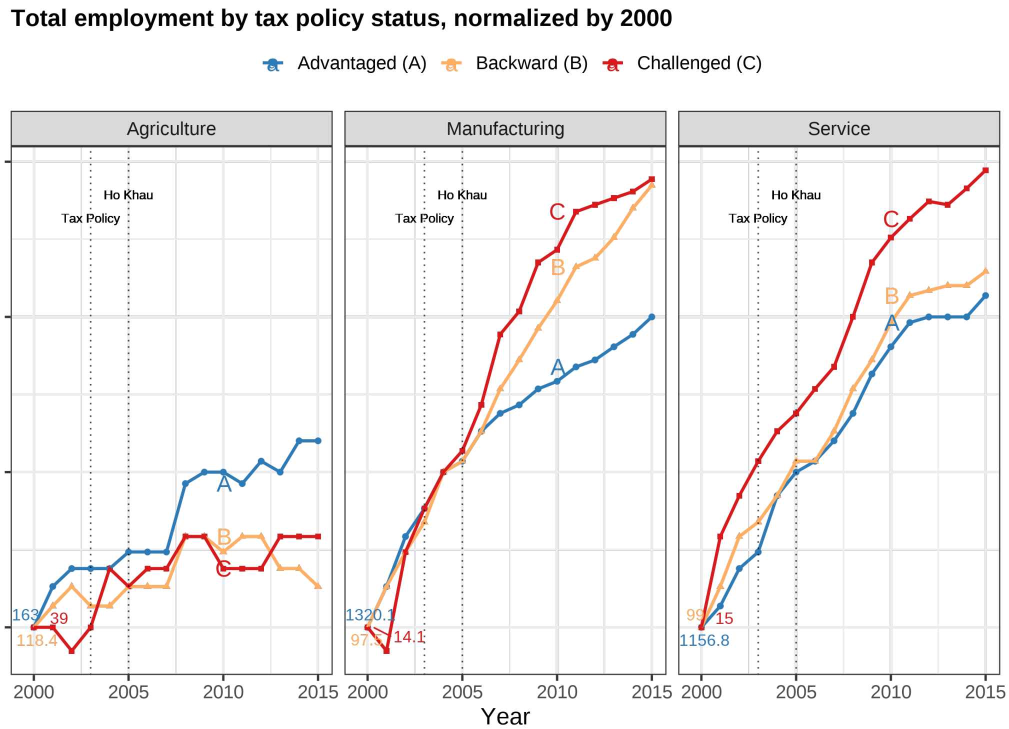
The establishment-level data covers the years 2000 to 2015 and is obtained from annual enterprise surveys conducted by the General Statistics Office (GSO). The survey is mandatory for all registered firms, except for private businesses with fewer than 10 workers operating in only one location. However, many private firms with fewer than 10 workers voluntarily report, as shown in Figure A4 of the firm size distribution[11]. The data includes unique firm identifiers, 4-digit ISIC industry codes, location information, and variables such as employment, revenue, wage bill, and fixed assets. To ensure consistency, I follow the approach used by McCaig et al. (2022) to create firm identifiers, location identifiers at the commune level, and 4-digit industry codes over time.

[9] Ha Tay was merged into Ha Noi in 2008 and thus is considered part of Ha Noi in this study

[10] While Decree 108/2005/ND-CP requires a continuous stay of at least three years, the Residence Law of 2006 No. 81/2006/QH11 reduces the requirement to at least one year.

[11] McCaig et al. (2022) states that the employment threshold was introduced in 2004, as before that, there was no such requirement. It is important to note that this threshold is not based on the average size of informal firms, which typically have 1.6 workers and are much smaller than the cutoff.

Figure 2: Employment by tax policy status, 2000-2015



Notes: Data come from the annual establishment surveys (2000-2015). I aggregate the firm-level employment to the district statuses in Figure 1 and broad sectors. Each data point represents the ratio of employment in a specific type of district-sector combination for a given year compared to its employment level in 2000.

By combining the firm location and district label in the 2003 tax policy, I verify Vietnam's overall growth trend and identify differences in growth rates across industries and locations. Figure 2 displays the temporal evolution of total employment categorized by district tax labels and the three main sectors: Agriculture, Manufacturing, and Services. The employment values are normalized based on their respective 2000 levels. The figure illustrates the notable expansion of formal employment, approximately doubling every five years between 2000 and 2015. Furthermore, the manufacturing and service sectors experience rapid growth, while the agriculture sector remains stagnant, indicating a structural transformation in the economy.

In 2000, Advantaged (A) districts exhibited the highest employment levels among the three categories, with B and C districts being more agriculture-oriented. However, from 2004 onwards, the growth rates of manufacturing employment in B and C districts surpassed those in A districts, highlighting the potential impacts of tax and Ho Khau policy reforms. This shift is particularly noteworthy considering that employment in these regions previously followed almost identical trends. The observed pattern persists even when considering the Zone category as the fourth location type in Figure A5.

During this period, the entry and exit margins play a crucial role in shaping the economy and are likely the driving force behind the substantial growth in non-agricultural employment observed in Figure 2, particularly in poorer districts. Table A2 presents data on entrants and exiters in 2000 and 2015. In 2015, firms that were newly established (not present in 2000) accounted for 98% of all firms, contributing to 85% of employment and 83% of revenue. On the other hand, exiters, referring to firms present in 2000 but no longer in operation by 2015, tended to be smaller in size, comprising only 47% of employment and 49% of revenue in 2000. These figures highlight the significant degree of establishment turnover in the economy and underscore the importance of new establishments in driving employment growth.

At the local level, entry is not limited to completely new firms. Firms can also enter a location by relocating from another place between $t-1$ and t, or "in-migrants". Additionally, firms can re-enter after a period of inactivity or for other reasons that make them unobserved. Among re-entrants, there are "re-entrant incumbents" that return to the same location and "re-entrant migrants" that relocate from elsewhere.

Nonetheless, most entrants to a district are completely new, as shown in Figure A6. The average annual entry rate is 27%, consistent with the rate reported by McCaig and Pavcnik (2021). Relocation across districts for incumbents is 6%, while the re-entrant in-migrant rate is less than 1%[12].

Although the establishment-level data used in this study covers a large and extended period, it fails to capture the economy's informal sector. It does not provide information on whether an entrant was informal before. However, given that formalization of private domestic firms are rare, it should not significantly affect the measurement of firm entry. According to McCaig et al. (2022) and Malesky and Taussig (2009), most private domestic firms did not start as informal, and only a small percentage of informal firms transitioned to formality within two years[13]. Furthermore, although the informal sector remains a significant share of employment, it constitutes only 20% of GDP (Cling et al., 2011).[14]

[12] The fraction of firms that re-enter after one exit period is extremely rare, at 0.1% among all firms that exit on average across exit cohorts.

[13] McCaig et al. (2022) writes "most of the private domestic firms in the enterprise sector did not start as informal firms (depending on the method/survey, estimates range from 16 to 30% started as informal) and that most of those that did were only in the informal sector for 1 to 3 months". Furthermore, McCaig and Pavcnik (2021) estimates that the share of informal firms transitioning into formality within two years is no more than 2% using Vietnam Household Living Standard Surveys from 2004 to 2018.

[14] 86 percent of Vietnam's workforce is in the informal sector, according to the population census data in

3.2 Household data

In addition to examining entrepreneur behaviors from establishment surveys, I also analyze the choices made by workers using the Population and Housing Censuses. Specifically, I use the 3% samples from 1999, 2009, and 2019 to assess the effects of policies on household sectoral and location choices. The sample is restricted to the working-age population, aged 15 to 65.

In this study, a migrant is defined as someone who reports the current location as different from their province of residence five years ago, consistent with the GSO definition. However, this definition may underestimate actual migration since it does not consider seasonal migration or migrants who have returned to their origin within the past five years. Nonetheless, it is likely to reflect permanent migration more than temporary migration.

In Figure 3, I present the distribution of migrants across destinations, categorized by origin types, before and after the implementation of the Ho Khau policy. In the left panel, each dot represents the share of migrants from all Rural origins to a specific province in 1999 (x-axis) and 2009 (y-axis).

Migrant shares across provinces range from less than 1% to over 10%. In 1999, the average migrant share across all provinces is approximately 3%, which is lower than the average out-of-state migrant share of 25% in the US between 1880 and 1940 but comparable to the inter-provincial migrant share in China in 2000 (Morris-Levenson and Prato, 2021; Tombe and Zhu, 2019). Between 1999 and 2009, migrant shares witness an average increase of 50%, with greater increases observed in provinces with lower initial shares, as indicated by the smaller slopes of the best-fit lines in comparison to the 45-degree lines.

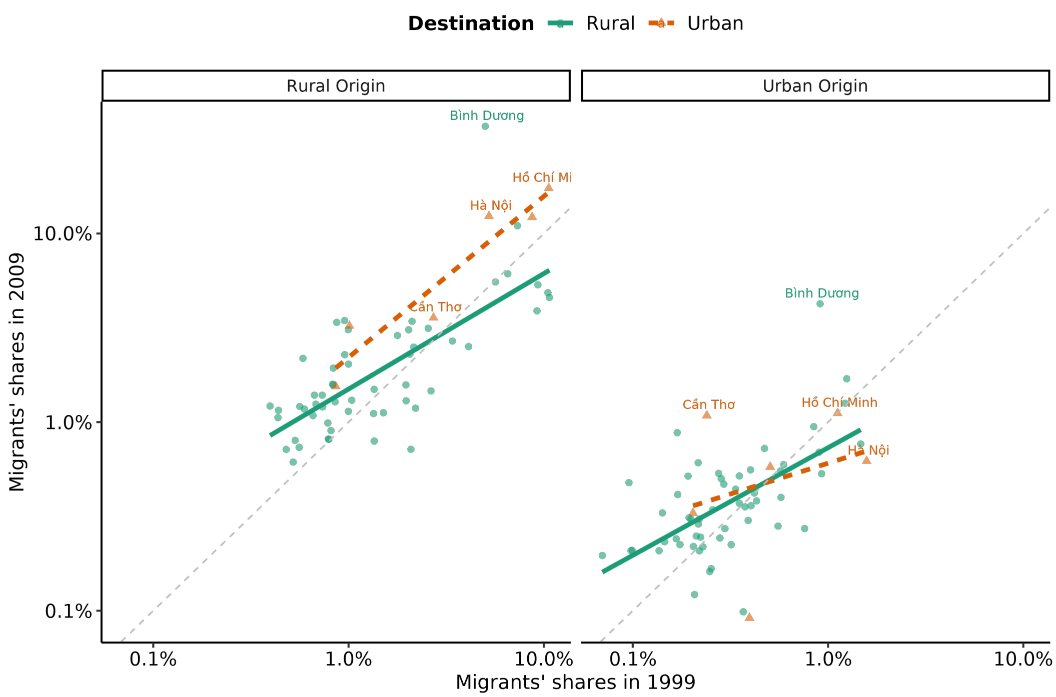
While the Ho Khau policy never imposes restrictions on the origin of migrants, there could be a stronger response observed among Urban origins compared to Rural origins. Among Rural origins, the best-fit line for Urban destinations lies above the best-fit line for Rural destinations, suggesting an increase in migration from Rural to Urban destinations, after the policy change. In contrast, Urban origins exhibit a weak preference for Urban destinations over Rural destinations after the Ho Khau reform.

3.3 Other data

To complement the micro-level data, various additional data sources have been used for estimation and identification purposes.

1999 (McCaig and Pavcnik (2015). This share dropped to 79 percent in 2009, and within manufacturing, the share of informal jobs dropped from 58 to 43 percent between 1999 and 2009.

Figure 3: Migrants' shares 1999 and 2009 across provinces, by origin types



Notes: Migration data come from population censuses 1999 and 2009, which I aggregate to the province level. Urban refers to the centrally administered cities, and Rural means the rest of Vietnam. The dashed lines are the 45-degree lines.

One such source is the Special Economic Zones (SEZs) mentioned in Section 2. The cumulative count of SEZs per commune has been computed annually, and communes with at least one SEZ prior to 2004 are categorized as "Zone". It is important to note that 87% of these communes have only one zone, allowing the measurement of zones at the commune level to serve as a close approximation for the entire zone. As the firm data does not specify whether a firm is located within a zone, all firms within a commune with SEZs are attributed to the SEZ, capturing some spillover effects from the zones to nearby regions within the commune.[15]

Changes in industry tariffs at the 3-digit level resulting from the US-VN BTA are derived from McCaig (2011). For measuring internal trade, volume data from JICA (2000) is employed, which reports inter-provincial freight traffic by transport distance for 13 industries in 1999, which I explain further in section 5.

[15]The median size of communes is 5.7 square miles, and the median population within a commune is approximately 6000 people, which is slightly larger than a median Census Track in the US.

3.4 Empirical evidence: Diff-in-Diff design

In this subsection, I employ a Difference-in-Differences (DiD) design to examine the impact of profit tax incentives on firm location choices and employment. Given that the policy provides subsidies for firm entry into C and B districts, as well as Zone communes, I anticipate that the policy will lead to an increase in new entries and employment in these regions compared to A districts.

The subsequent analyses are based on aggregating the establishment data into a balanced panel at the district-zone-sector level, which comprises about 700 district zones and 60 sectors at the two-digit level.

The time and spatial variation of the policies suggest the following regression

$$Y_{dst} = \sum_{t'=2000,t'\neq 2003}^{2015} \beta^{P}_{t'} \cdot \text{Treat}_d \cdot \mathbf{1}_{t'} + \alpha_d + \delta_t + \theta_s \cdot \delta_t + \varepsilon_{d,s,t}. \tag{1}$$

Here, $Y_{d,s,t}$ is the outcome variable, which measures the share of entrants located in district-zone d among all entrants in sector s at year t, as well as the total labor employed by all firms in district-zone d and sector s in year t. The variable Treat_d takes four different values (A, B, C, and Zone) with A being the control group. The commune d is classified as Zone if it has an economic zone prior to 2004.

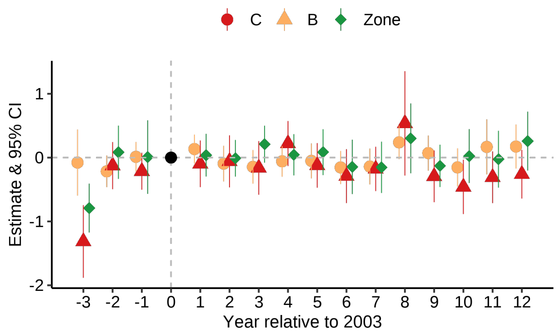
The regression also includes district-zone α_d, δ_t year, and $\theta_s \cdot \delta_t$ sector-year fixed effects to control for time-invariant differences across space, macroeconomic shocks, and time-varying sectoral shocks, respectively. To account for unobserved correlation within district-zone, I cluster the standard errors at the district-zone level.

The regression model also incorporates district-zone fixed effects α_d, year fixed effects δ_t, and sector-year fixed effects $\theta_s \cdot \delta_t$ to control for time-invariant differences across regions, macroeconomic shocks, and time-varying sectoral shocks, respectively. In order to address potential unobserved correlations within locations, standard errors are clustered at the district-zone level.

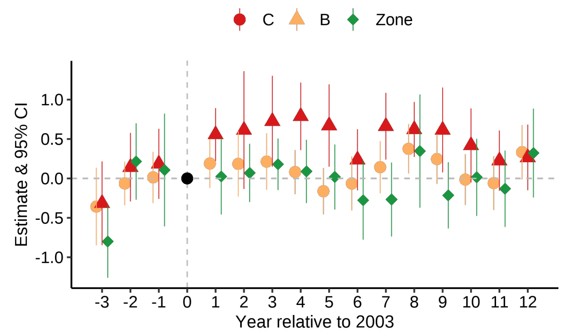
To estimate Equation 1 and account for zeros in the outcome variables, I employ Poisson regression[16] I expect that the estimated coefficients $\beta^B_t, \beta^C_t, \beta^{\text{Zone}}_t$ will be positive for both outcomes after 2003, assuming a parallel trend in the log of outcomes in the absence of the policy change. The estimated coefficients for years prior to 2003 are used to examine this identification assumption before the policy was implemented.

[16]For more details, refer to Mullahy and Norton (2022) and Chen and Roth (2023).

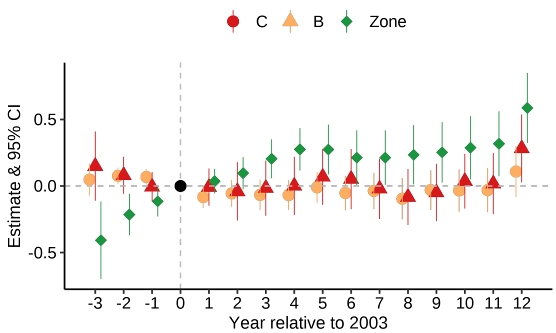
Figure 4: DiD estimates of entry and employment



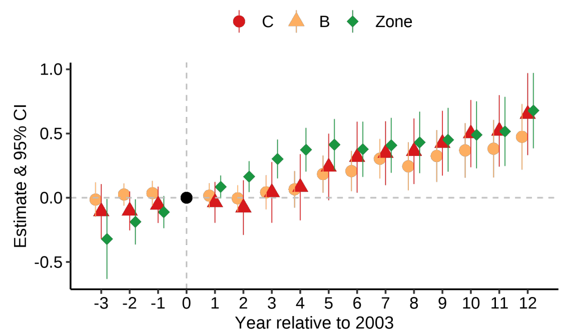
(a) Entry - All



(b) Entry - Manufacturing



(c) Employment - All



(d) Employment - Manufacturing

Notes: The graphs depict entry share and employment, with (a) and (c) representing all sectors and (b) and (d) representing manufacturing. The estimates are relative to type A district and year 2003, and the data is aggregated at the district-sector-year level from annual establishment surveys.

DiD Results and Discussion: The tax incentives do not demonstrate significant effects on firm entry or employment in the B and C districts when considering all firms, as illustrated in Figure 4a and Figure 4c.

Manufacturing, being a tradable sector, is more likely to respond to tax incentives in poorer districts since its market is less reliant on local demand. However, considering that the share of manufacturing firms among all firms is relatively small, at approximately 24%, the overall results may not fully capture the response from the manufacturing sector. Consequently, when analyzing the data by breaking it down into major sectors such as agriculture, manufacturing, and services, I observe a notable increase in both entry rates and employment specifically within the manufacturing sector.

Table 1: DiD Estimates of tax policy by sector

Sector	Agriculture		Manufacturing		Service	
	Entry	Labor	Entry	Labor	Entry	Labor
	(1)	(2)	(3)	(4)	(5)	(6)
B x Post	-0.45**	-0.26***	0.23**	0.23***	0.31***	-0.04
	(0.20)	(0.08)	(0.11)	(0.09)	(0.12)	(0.09)
C x Post	-0.30	-0.12	0.54***	0.38***	0.37*	0.22**
	(0.20)	(0.11)	(0.15)	(0.12)	(0.19)	(0.09)
Zone x Post	-0.04	-0.01	0.09	0.55***	0.59***	0.41**
	(0.25)	(0.27)	(0.18)	(0.13)	(0.19)	(0.18)
Observations	36,276	36,512	149,772	150,064	141,070	141,622

* *Notes:* Unit of observation is district-zone-sector-year, which aggregates from establishment surveys 2000-2015. The model is estimated by PPML. Robust standard errors are clustered at the district-zone level and reported in parenthesis. ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$.

Tax incentives result in an approximately 65% increase in the share of entrants in C districts compared to A districts, as indicated by $\exp(0.5) - 1$. Moreover, these effects are relatively persistent over time, as illustrated in Figure 4b. These findings align with research conducted in the Indian context, where manufacturing entry experienced a 60% increase in backward districts (Chaurey, 2017; Hasan et al., 2021). Additionally, the rise in firm entry potentially leads to a boost in employment within manufacturing firms in B and C districts, as demonstrated in Figure 4d. It is worth noting that the significant increase in firm entries cannot be solely attributed to the formalization of existing informal firms, as discussed in Section 3.

The Zone results also suggest that tax incentives may not be the main driver of firm entry into SEZs. The absence of a significant increase in firm entry in zones may be due to other favorable conditions in SEZs beyond profit tax cuts.

In summary, the Difference-in-Differences (DiD) analysis reveals that tax policy has the potential to stimulate structural transformation in less affluent districts. Table 1 presents the results of pooled regressions based on (1), where I aggregate the data into pre- and post-periods and analyze them separately for major sectors. Following the implementation of tax incentives, both entry and employment in the manufacturing sector experience a surge in B and C districts compared to A districts. Conversely, the agriculture sector witnesses a decline. There is a modest increase in the service sector, although not statistically significant, possibly due to a higher likelihood of informal operations with smaller-scale businesses in the service sector.

Finally, the DiD analysis presented here serves as a descriptive overview of the impact of tax policy. I intentionally exclude the Ho Khau reform from this analysis because, even when I include dummies for major cities in addition to the treatment variable, the results remain similar. However, it is important to note that this DiD analysis is of a partial equilibrium nature. In the subsequent section, I will introduce a general equilibrium model and delve into the estimation process in Section 5, where I will discuss the complexities associated with examining the policy effects in the presence of general equilibrium and the simultaneous timing of the two policies. Fortunately, I can leverage the model and the rich variations in both policies to address these identification challenges.

4 A Dynamic Spatial GE Model

The simultaneous implementation of multiple place-based policies presents a challenge for the previous reduced-form analysis. For example, many A districts, which serve as the control group in regression (1) for the tax policy, is also affected by the Ho Khau policy. The close timing of these policies makes it difficult to isolate the employment effects solely attributed to the tax policy. Additionally, large-scale policies can have complex effects on employment and entry outcomes through their influence on wages and prices in the general equilibrium. To address these issues and provide clarity on the policy effects, this section develops a dynamic spatial general equilibrium model.

The model not only allows for a distributional examination across space and time of each policy separately as well as both policies collectively but also enables an aggregate welfare analysis of the overall economy. These responses encompass various dimensions of heterogeneity and interactions among local markets, including trade, migration choices of workers, and location choices of entrepreneurs.

To begin, I introduce the notations and provide an overview of the economy's basic structure before examining the individual decisions of each agent. The model extends earlier works such as Caliendo and Parro (2020) and Kleinman et al. (2023) by incorporating heterogeneous entrepreneurs categorized by age and workers with correlated preference shocks.

4.1 Environment

The economy evolves in discrete time, indexed by $t \geq 0$ and consists of N locations indexed by $i, n \in \mathcal{R} = \{1, \ldots, N\}$, and S sectors indexed by $j, k \in \mathcal{S} = \{1, \ldots, S\}$. Interactions between location-sector pairs are captured using superscripts ij, nk to indicate movement from origin ij to destination nk for shipment, migration, and sectoral decisions. For example, the trade cost of shipping goods from location-sector ij to nk is denoted as $d^{ij,nk}$.

Each location-sector pair contains six infinitely-lived households: workers, landlords, informal (I), young (Y), middle-aged (M), and old (O) entrepreneurs. All agents derive utility from consumption c_t and have the following preferences:

$$\mathbb{E}_t \left[\sum_{t'=t}^{\infty} \beta^{t'-t} \ln \left(c_{t'} \right) \right],$$

where β is the discount factor and consumption c_t can differ across local markets. Each agent consumes a bundle of goods from all sectors with the same share, α^j, for goods in each sector j.

In each period t and in each location-sector ij, there are six state variables: L_t^{ij}, which denotes the measure of workers, K_t^{ij}, which represents the amount of capital stock, and $E_{I,t}^{ij}$, $E_{M,t}^{ij}$, $E_{Y,t}^{ij}$, and $E_{O,t}^{ij}$, which are the measures of informal, young, middle-aged and old entrepreneurs, respectively. At the end of each period, a measure of one newborn household chooses their occupation between workers and entrepreneurs, resulting in the measure of informal entrepreneurs $E_{I,t}^{ij}$. The remaining newborns, stayers, and immigrants into ij between $t-1$ and t determine the measure of workers L_t^{ij}. The informal entrepreneurs can then make location decisions to produce. Old entrepreneurs consist of young entrepreneurs who remain between $t-1$ and t and old ones who do not exit. Landlords make investment and consumption decisions determining the amount of capital stock K_t^{ij} at each period t.

The model accounts for policies imposed at time t denoted by $\mathcal{P}_t$, which include the Ho Khau policy between location i and n represented by $\{mpol_t^{i,n}\}_{i,n \in \mathcal{R}}$ and the profit taxes applying to non-informal entrepreneurs of different ages $a \in \mathscr{A} = \{Y, M, O\}$ in location i denoted by $\{\tau_{a,t}^i\}_{i \in \mathcal{R}, a \in \mathscr{A}}$.

In addition to policies, the model also includes the exogenous local amenities $\mathcal{A}^i$ and location-sector productivities A_t^{ij}, which can evolve over time. These factors capture the fundamental differences across local markets, such as the natural beauty of a location or its proximity to ports, which are not captured by policies or market size.

4.2 Workers

Each location-sector has a continuum of workers of measure L_t^{ij}, who supply labor inelastically. At the beginning of each period t, each worker who works in a sector-location pair ij receives a wage w_t^{ij}, consumes the final good in location i for a price P_t^i, and enjoys amenity $\mathcal{A}_t^i$. Then, they draw a vector of idiosyncratic shocks across location-sector pairs $\boldsymbol{\epsilon}_t = \{\epsilon_t^{nk}\}_{n=1,k=1}^{N,S}$ and select the location-sector to work in for the next period based on the expected value of being in location-sector nk, denoted by $U_{t+1}^{nk} = \mathbb{E}\left[u_{t+1}^{nk}\right]$, subject to switching costs $m_t^{ij,nk}$.

In particular, the value of working in location-sector ij for each worker is given by[17]

$$u_t^{ij} = \ln\left(\mathcal{A}_t^i w_t^{ij} / P_t^i\right) + \max_{n,k}\left\{\beta U_{t+1}^{nk} - m_t^{ij,nk} + \epsilon_t^{nk}\right\} \tag{2}$$

where the vector of idiosyncratic shocks $\boldsymbol{\epsilon}_t$ is i.i.d over time and has the following joint cumulative distribution function (c.d.f.)

$$F(\boldsymbol{\epsilon}_t) = \exp\left[-\sum_{n=1}^{N}\left(\sum_{k=1}^{S}\exp\left(-\frac{\epsilon_t^{nk}}{\rho\nu} - \frac{\bar{\gamma}}{\rho}\right)\right)^{\rho}\right],$$

where $\bar{\gamma}$ is the Euler's constant.

This nested logit distributional assumption generalizes the standard logit assumption in the literature and allows sectoral choices to be correlated within each location, captured by the correlation parameter $\rho \in (0,1]$. When $\rho = 1$, the nested logit reduces to a logit model, implying independence between sectoral and location choices. A higher value of ρ indicates lower similarity between sectors within the same location.

Additionally, the parameter $\nu > 0$ controls the dispersion in shocks across locations. When ν is large, there is greater heterogeneity in worker preferences across locations, resulting in less responsiveness to policy changes.

Workers spend their wages w_t^{ij} on goods from all sectors, given by $c_t^{ij} = \prod_k (c_t^{ij,k})^{\alpha^k}$, where α^k is the consumption share of sector-k goods , and $c_t^{ij,k}$ is the amount of sector-k goods consumed by a worker working in location-sector ij. The region price index is calculated using the logarithmic preference of the workers and the fact that the sector of a worker does not affect the price of goods produced in the industry j. Specifically, I can

[17] This specification is common in the migration literature (Jia et al., 2023).

express the regional price index as

$$P_t^i = \prod_{j=1}^{S} \left(P_t^{ij} / \alpha^j \right)^{\alpha^j}, \quad 0 < \alpha_j < 1 \text{ and } \sum_j \alpha^j = 1, \tag{3}$$

where P_t^{ij} is the region-sector price index, determined by the decisions of the entrepreneurs.

Migration Flows and Sectoral Shares

Workers choose the location-sector that maximizes their expected utility. Their utility depends on the wage, local amenities, and migration costs. The migration costs reflect the monetary and non-monetary costs of moving from one location to another. The expected utility is a function of the worker's current location and sector and the future location and sector pairs they could choose from. The expectation is taken with respect to the distribution of shocks and reflects the uncertainty workers face in future periods.

To determine equilibrium sectoral and migration flows, I derive the option value of being in location ij in (2) and denoted it by Ξ_t^{ij}

$$\Xi_t^{ij} \equiv \mathbb{E}\left[\max_{\{n,k\}} \beta U_{t+1}^{nk} - m_t^{ij,nk} + \epsilon_t^{nk} \right] = \nu \ln \left[\sum_{n=1}^{N} \left(\sum_{k=1}^{S} e^{(\beta U_{t+1}^{nk} - m_t^{ij,nk})/\rho\nu} \right)^{\rho} \right], \tag{4}$$

where the expected lifetime utility of a worker in nk at time t is denoted as $U_t^{nk} \equiv \mathbb{E}_t[u_t^{nk}]$. Thus, this expected value U_t^{nk} is given by

$$U_t^{nk} = \ln \left(\mathcal{A}_t^n w_t^{nk} / P_t^n \right) + \Xi_t^{nk}. \tag{5}$$

Based on the distributional assumption, the share of workers who migrate from ij to work in nk between time t and $t+1$ is given by

$$\mu_t^{ij,nk} = \frac{e^{(\beta U_{t+1}^{nk} - m_t^{ij,nk})/\rho\nu}}{\sum_{s=1}^{S} e^{\left(\beta U_{t+1}^{ns} - m_t^{ij,ns}\right)/\rho\nu}} \frac{\left(\sum_{s=1}^{S} e^{\left(\beta U_{t+1}^{ns} - m_t^{ij,ns}\right)/\rho\nu} \right)^{\rho}}{\sum_{c=1}^{N} \left(\sum_{s=1}^{S} e^{\left(\beta U_{t+1}^{cs} - m_t^{ij,cs}\right)/\rho\nu} \right)^{\rho}}. \tag{6}$$

4.3 Entrepreneurs

In this model, young, middle-aged, and old entrepreneurs face monopolistic competition when producing varieties for consumption. Each entrepreneur aims to maximize their expected utility by choosing their firm's price and output levels and making dynamic entry and exit decisions across different location-sectors.

4.3.1 Static decisions

An entrepreneur of type a, where a is Y, M, or O, decides how much to produce and at what price to charge for each period t. The entrepreneur's production function is represented by q_t^{ij} and depends on local productivity (A_t^{ij}), labor (L_t^{ij}), capital (K_t^{ij}), and the share of labor in value-added (ξ). The production function is given by

$$q_t^{ij} = A_t^{ij} \left(L_t^{ij}\right)^{\xi} \left(K_t^{ij}\right)^{1-\xi}.$$

I assume that entrepreneurs are not concerned with local amenities ($\mathcal{A}_t^i$) but rather the productivity of the location (A_t^{ij}).

The firm's cost minimization problem determines the unit cost bundle as

$$x_t^{ij} = B \left(w_t^{ij}\right)^{\xi} \left(r_t^{ij}\right)^{1-\xi} \tag{7}$$

where w_t^{ij} denote wages, r_t^{ij} rental prices, and B is a constant.[18]

Due to perfect competition in the input markets, cost minimization also implies the following capital market clearing condition

$$r_t^{ij} K_t^{ij} = \frac{1-\xi}{\xi} w_t^{ij} L_t^{ij}. \tag{8}$$

In addition to input costs, the firm also faces iceberg trading costs $d^{ij,nj}$ to sell its products from location i to location n. Hence, the optimal price chosen by entrepreneurs is given by

$$p_t^{ij,nj} = \frac{\sigma}{\sigma - 1} \frac{d^{ij,nj} x_t^{ij}}{A_t^{ij}} \tag{9}$$

where σ represents the elasticity of substitution between different varieties.

The entrepreneur sells her products everywhere to earn pre-tax profits, denoted by π_t^{ij} and given by

$$\pi_t^{ij} = \bar{K} \left(\frac{A_t^{ij}}{x_t^{ij}}\right)^{\sigma-1} \left(\sum_{n \in \mathcal{R}} (d^{ij,nj})^{1-\sigma} X_t^{nj} (P_t^{nj})^{\sigma-1}\right) \tag{10}$$

[18] In particular, $B = \xi^{-\xi}(1-\xi)^{-(1-\xi)}$

where $\bar{K} = (\frac{\sigma}{\sigma-1})^{-\sigma}$ is a constant, and X_t^{nj} represents the expenditure on goods j in location n. When goods are substitutes (i.e., when $\sigma > 1$), the entrepreneur earns more profits from lower effective input costs (x_t^{ij}/A_t^{ij}) at her production site and higher market access. Higher market access arises from lower trade costs from the entrepreneur's production site to her consumers and higher demand and prices for her goods elsewhere.

4.3.2 Dynamic decisions

In each period, informal I, young Y and middle-aged M entrepreneurs make forward-looking entry and exit decisions, while O entrepreneurs make no dynamic decisions. Y and M entrepreneurs can choose whether to stay at the current location and become older or exit, whereas I entrepreneurs decide where to produce when they become young. Let $v_{a,t}^{ij}$ denote the value of a type a firm located in location-sector ij at time t. The expected value of being a in ij next period is defined as $V_{a,t+1}^{ij} \equiv \mathbb{E}_t[v_{a,t+1}^{ij}]$.

Old (O) entrepreneurs: In each period, a mass of old entrepreneurs $E_{O,t}^{ij}$ collect after-tax profits and consume local goods. Each firm exits with an exogenous probability $1 - \varrho$. Thus, the expected value of being an old entrepreneur in ij at time t is given by

$$V_{O,t}^{ij} = \ln\left((1 - \tau_{O,t}^{i})\frac{\pi_t^{ij}}{P_t^i}\right) + \varrho\beta V_{O,t+1}^{ij}. \tag{11}$$

The evolution of the mass of O entrepreneurs is given by

$$E_{O,t+1}^{ij} = \varrho E_{O,t}^{ij} + \varsigma_t^{ij} E_{M,t}^{ij}, \tag{12}$$

where $E_{M,t}^{ij}$ represents a mass of M entrepreneurs in location-sector ij at time t and ς_t^{ij} the share of those entrepreneurs that decide to stay in location-sector ij between t and $t+1$.

Young (Y) and Middle-aged (M) entrepreneurs: The value of being an a entrepreneur in location-sector ij at time t consists of the flow utility of real after-tax profits and the option value of whether to continue operating in ij or shutting the firm down, which is given by

$$v_{a,t}^{ij} = \ln\left((1 - \tau_{a,t}^{i})\frac{\pi_t^{ij}}{P_t^i}\right) + \max\left\{\beta V_{a+1,t+1}^{ij} + \chi\epsilon_{a,t}^{ij}; 0 + \chi\epsilon_{a,t}^{ij}\right\},$$

where $\tau_{a,t}^{i}$ is the profit tax on a entrepreneurs in region i and P_t^i is the region price index. Each future decision is subject to two productivity shocks, one for continuing operations and one for exiting, $\epsilon^{ij}_{Y,t}$. These shocks follow a Type-I extreme value distribution whose variance parameter is denoted as χ. A lower variance implies that firms are more responsive to tax changes. A future value of a young entrepreneur is denoted by $V^{ij}_{a+1,t+1}$, which corresponds to the future value of being a middle-aged entrepreneur. This notation applies similarly to middle-aged ones.

Using the Type-I extreme value distribution assumption, the expected value for an a entrepreneur at time t in ij is given by

$$V^{ij}_{a,t} = \ln\left((1-\tau^{i}_{a,t})\frac{\pi^{ij}_t}{P^i_t}\right) + \chi \ln\left[\exp\left(V^{ij}_{a+1,t+1}\right)^{\beta/\chi} + 1\right], \tag{13}$$

and the fraction of a entrepreneurs who choose to stay in location-sector ij between time t and $t+1$ is given by

$$\varsigma^{ij}_{a,t} = \frac{\exp\left(V^{ij}_{a+1,t+1}\right)^{\beta/\chi}}{\exp\left(V^{ij}_{a+1,t+1}\right)^{\beta/\chi} + 1}. \tag{14}$$

Finally, the evolution of a entrepreneurs is given by

$$E^{ij}_{M,t+1} = \varsigma^{ij}_{Y,t} E^{ij}_{Y,t} \tag{15}$$

$$E^{ij}_{Y,t+1} = \psi^{ij}_t E^{j}_{I,t} \tag{16}$$

where ψ^{ij}_t is the probability of an I entrepreneur chooses to enter location-sector ij between t and $t+1$, while $E^j_{I,t}$ is the mass of I-entrepreneurs in sector j.

Informal (I) entrepreneurs: Each informal entrepreneur is endowed with talent ϕ. The value of being an informal (I) entrepreneur with talent ϕ at time t in sector j, $v^j_{I,t}$, is given by

$$v^j_{I,t} = \max_{n\in\mathcal{R}} \left\{\beta V^{nj}_{Y,t+1} - f^{nj}_t(\phi) + \chi\epsilon^{nj}_t\right\},$$

where ϵ^{nj}_t represents idiosyncratic entry cost shocks that follow the same distribution as the Y entrepreneurs. The entry cost, $f^{nj}_t(\phi)$, summarizes the efforts required by entrepreneurs to begin operating in a location sector and is given by

$$f^{nj}_t(\phi) = \underbrace{\ln(f^{nj}_E) - \zeta\ln(E^{nj}_t)}_{f^{nj}_t} - \ln(\phi).$$

The entry cost depends on two components: the entrepreneur's own talent ϕ and the location-sector-specific f_t^{nj}. A higher talent ϕ implies a lower entry cost for the entrepreneur across all location-sectors. The term f_E^{nj} accounts for any time-invariant entry cost, such as geography. The entry cost also depends on the entry market size as in Peters (2022), where ζ represents the inter-temporal knowledge externality, and E_t^{nj} represents the measure of producing entrepreneurs in nj, i.e., $E_t^{nj} = E_{Y,t}^{nj} + E_{O,t}^{nj}$. The larger the number of operating entrepreneurs in a market, the easier for an informal entrepreneur to enter.

The expected value of an I entrepreneur with talent ϕ, denoted by $V_{I,t}^{ij}(\phi)$, is given by

$$V_{I,t}^{j}(\phi) = \ln(\phi) + \chi \ln\left[\sum_{n\in\mathcal{R}} \exp\left(\beta V_{Y,t+1}^{nj} - f_t^{nj}\right)^{1/\chi}\right], \tag{17}$$

and the fraction of I entrepreneurs that enter location-sector ij between t and $t+1$ is given by

$$\psi_t^{ij} = \frac{\exp\left(\beta V_{Y,t+1}^{ij} - f_t^{ij}\right)^{1/\chi}}{\sum_{n\in\mathcal{R}} \exp\left(\beta V_{Y,t+1}^{nj} - f_t^{nj}\right)^{1/\chi}}, \tag{18}$$

which is independent of the entrepreneur's talent ϕ.

Finally, the measure of I entrepreneurs in sector j at time t, $E_{I,t}^{ij}$, is determined by the occupation choice at the end of each period t.

4.4 Occupation Choice

At the end of each period, a measure one of households that differ in talent ϕ, drawn from a distribution with CDF F, emerges in each location-sector ij. Each household can choose to be a worker or an informal entrepreneur. If a household chooses to be a worker, their talent is irrelevant, and their expected value is the option value of being a worker anywhere in the next period, which is given by (4). On the other hand, if a household chooses to be an entrepreneur, they can use their talent to help reduce the entry cost and obtain the expected value of being an I entrepreneur according to (17).

The household's talent cutoff $\underline{\phi}_t^{ij}$ is determined by the indifference household between being an entrepreneur and a worker and must satisfy

$$V_{I,t}^{j}(\underline{\phi}_t^{ij}) = \Xi_t^{ij}. \tag{19}$$

As a result, the measure of I entrepreneurs in location-sector ij, denoted as $E_{I,t}^{ij}$, is given by $E_t^{ij} = 1 - F(\underline{\phi}_t^{ij})$. Summing these measures across regions yields the measure of I entrepreneurs in sector j, $E_{I,t}^j = \sum_{i=1}^N E_t^{ij}$.

Finally, the measure of households that choose to be workers, $F(\underline{\phi}_t^{ij})$, and the measure of immigrants from everywhere, $\mu_t^{n,ij} L_t^n$, determine the evolution of workers in location-sector ij as

$$L_{t+1}^{ij} = F(\underline{\phi}_t^{ij}) + \sum_{n \in \mathcal{R}} \sum_{k \in \mathcal{S}} \mu_t^{nk,ij} L_t^{nk} \tag{20}$$

4.5 Landlords

In each location-sector, a measure one of immobile landlords chooses their local consumption and investment sequences to maximize their expected utility, subject to a budget constraint. The landlord's objective is to maximize:

$$W_t^{ij} = \max_{C_t^{ij}, K_{t+1}^{ij}} \{\ln(C_t^{ij}) + \beta W_{t+1}^{ij}\}.$$

The intertemporal budget constraint for landlords in each location requires that the total income from the existing capital stock equals the total value of goods consumption and net investment. Capital depreciates at a rate of δ. The constraint is given by:

$$r_t^{ij} K_t^{ij} = P_t^i C_t^{ij} + P_t^i \left(K_{t+1}^{ij} - (1-\delta) K_t^{ij} \right).$$

Kleinman et al. (2023) solve the optimal consumption-saving decision for the landlord's problem, which is given by

$$P_t^i C_t^{ij} = (1-\beta)(r_t^{ij} + P_t^i (1-\delta)) K_t^{ij},$$

and the law of motion of capital

$$K_{t+1}^{ij} = \beta (r_t^{ij} / P_t^i + (1-\delta)) K_t^{ij}. \tag{21}$$

4.6 Local government

A local government in location i collects profit tax revenue from Y and O entrepreneurs across all sectors. Thus, the local revenue is given by

$$\Theta_t^i = \sum_{j \in \mathcal{S}} \left(E_{Y,t}^{ij} \tau_{Y,t}^i + E_{O,t}^{ij} \tau_{O,t}^i \right) \pi_t^{ij}. \tag{22}$$

The local government is myopic, has log utility, and spends its revenue Θ_t^i on consuming the local goods with the same share α^j of each sector as other agents.

4.7 Aggregation

In each period t, a final goods producer in each location-sector ij purchases intermediate goods produced by Y, M, and O entrepreneurs. These intermediate goods are aggregated into a CES final good, and the price of this final good is given by

$$P_t^{ij} = \left(\sum_{n \in \mathcal{R}} E_t^{nj} (p_t^{nj,ij})^{1-\sigma} \right)^{1/(1-\sigma)}. \tag{23}$$

The bilateral trade shares of goods bought by location-sector nj and produced by ij is given by

$$\lambda_t^{ij,nj} = E_t^{ij} \frac{p_t^{ij,nj} q_t^{ij,nj}}{X_t^{nj}} = E_t^{ij} \left(\frac{p_t^{ij,nj}}{P_t^{nj}} \right)^{1-\sigma}.$$

Substituting (9) for prices yields

$$\lambda_t^{ij,nj} = \frac{E_t^{ij} (d^{ij,nj} x_t^{ij})^{1-\sigma} (A_t^{ij})^{\sigma-1}}{\sum_{c \in \mathcal{R}} E_t^{cj} (d^{cj,nj} x_t^{cj})^{1-\sigma} (A_t^{cj})^{\sigma-1}}. \tag{24}$$

Income of location-sector ij, denoted by Π_t^{ij}, is the sum of four components: workers' income, landlords' rent income, tax revenue, and total after-tax profits of Y and O entrepreneurs. However, the last two components are the total pre-tax profits of the entrepreneurs. Thus, income can be expressed as

$$\Pi_t^{ij} = w_t^{ij} L_t^{ij} + r_t^{ij} K_t^{ij} + \sum_{a \in \{Y,M,O\}} E_{a,t}^{ij} \pi_{a,t}^{ij}.' \tag{25}$$

Since each agent spends the same share, α^j, on goods from sector j, the total expenditure on goods in sector j in region i is given by

$$X_t^{ij} = \alpha^j \Pi_t^i, \text{ where } \Pi_t^i \equiv \sum_{j \in \mathcal{S}} \Pi_t^{ij}. \tag{26}$$

Since labor receives a share ξ of value-added, the labor market clearing condition requires that payment to labor equals share ξ of net-markup revenues

$$w_t^{ij} L_t^{ij} = \xi \sum_{n=1}^{N} \frac{\sigma - 1}{\sigma} \lambda_t^{ij,nj} X_t^{nj}. \tag{27}$$

4.8 Equilibrium

At the beginning of each period, the exogenous state variables $\mathcal{F}_t$ are realized, which includes trade costs, mobility costs, amenities, and productivity. The set of tax and migration policies $\mathcal{P}_t$, which includes $\tau_{a,t}^{i}$ and $mpol_t^{i,n}$, are then implemented.

All agents in the economy observe their idiosyncratic shocks and make optimal decisions based on the value function of entrepreneurs ($V_t = \{V_{a,t}^{ij}\}$) and the value function of workers ($U_t = \{U_t^{ij}\}$), resulting in migration flows ($\mu_t = \{\mu_t^{ij,nk}\}$), talent cutoff ($\underline{\phi}_t = \{\underline{\phi}_t^{ij}\}$), entry rates ($\psi_t = \{\psi_t^{ij}\}$), stay rates ($\varsigma_t = \{\varsigma_t^{ij}\}$), wages ($w_t = \{w_t^{ij}\}$), rental rates ($r_t = \{r_t^{i}\}$), bilateral expenditure ($X_t = \{X_t^{ij,nj}\}$), and local market prices ($P_t = \{P_t^{ij}\}$).

These decisions discipline the evolution of the endogenous state variables, including the distribution of labor across locations and industries ($L_t = \{L_t^{ij}\}$), the distribution of entrepreneurs ($E_t = \{E_{I,t}^{ij}, E_{Y,t}^{ij}, E_{O,t}^{ij}\}$), and the distribution of capital ($K_t = \{K_t^{ij}\}$).

Definition 1. Given initial conditions $\{L_0^{ij}, E_{a,0}^{ij}\}$, a sequence of exogenous fundamentals $\{\mathcal{F}_t\}_{t=0}^{T}$ and policies $\{\mathcal{P}_t\}_{t=0}^{T}$, a sequential competitive equilibrium consists of sequences of prices and allocations $\{L_t, \mu_t, K_t, E_{a,t}, \varsigma_t, \psi_t, \underline{\phi}_t, V_t, U_t, w_t, r_t, P_t\}_t^{\infty}$ such that

1. Final good producers optimize, setting their prices according to (23), and the local price index follows (3).

2. Workers make optimal consumption, labor supply, and location decisions, as described by (5), (6), (20), and subject to migration policy (29).

3. Entrepreneurs make dynamic production, pricing, and location decisions optimally, as described by (7), (9), (13), (17), (14), (16), (18).

4. Households choose the best occupations according to their talent and give rise to the equilibrium cutoff (19).

5. Landlords make optimal saving and investment decisions according to (21).

6. Local governments consume optimally using all their revenue (22).

7. Final good markets are clear in every market, according to (24), (25), and (26).

8. Labor and capital markets are clear according to (27) and (8).

5 Estimation

This section takes the model to data and lays out identification strategies to estimate the model's parameters and Ho Khau policy magnitude.

5.1 Taking the model to data

In the model, each period t represents a year. The earliest year for which data is available, as explained in Section 3, is the household census of 1999. However, since the first firm survey data is from 2000, I designate the model's first period as $t = 0$ corresponding to 2000 and use observations in 2000 in the firm data for 1999 census. Similarly, since the last period I have firm data is 2015 while the census goes up to 2019, I will use firm information in 2014 and 2015 for the 2019 census.

The endogenous state variables of local markets are captured in the data as follows: I observe the worker size L_t^{ij} as the sum of employment reported in location-sector ij, and the capital stock as the sum of fixed assets in ij from the annual enterprise surveys. Although I do not directly observe the measure of informal entrepreneurs $E_{I,t}^{ij}$, I make the assumption, based on occupation choices, that $E_{I,t}^{ij}$ corresponds to the fraction of firms among the sum of firms and workers in ij.

Additionally, I have data on the total number of entrants in sector j, $E_{I,t}^{j}$, and the number of young entrepreneurs $E_{Y,t}^{ij}$ correspond to the number of one- to five-year-old establishments, middle-aged to six- to ten-year-old, and old to above ten-year-old establishments.

Parameterization I parameterize trade costs as a function of distance following the trade literature such as Monte et al. (2018):

$$d^{ij,nj} = (\text{distance}^{i,n})^{\kappa^j}$$

where κ^j is the sectoral-specific elasticity of trade costs with respect to distance.

The talent distribution is Pareto with a cumulative distribution function $F(\phi) = 1 -$

$\left(\frac{H}{\phi}\right)^{\vartheta}$, where $\phi \geq H$ and ϑ is the tail index. In short, the structural parameters include

$$\left\{\chi, \rho, \nu, \xi, \{\kappa^j\}_j, \{\alpha^j\}_j, \sigma, \beta, \delta, \varrho, \zeta, \vartheta, H\right\}.$$

I leverage the changes in tax policy as quasi-random experiments to identify the inverse of the firm entry elasticity χ in Subsection 5.2. I explain how to back out the changes in migration costs associated with the Ho Khau policy in 5.3. To get the magnitude of the Ho Khau policy change, I need an estimate of ν which I go over in detail in 5.4. This part also covers how to estimate κ^j, α^j.

I externally set the discount factor $\beta = 0.95$, depreciation rate $\delta = 1 - 0.95^5$, $\zeta = 0.71$ from Peters (2022), the elasticity of substitution $\sigma = 5$ which is common in the trade literature, and the labor share $\xi = 0.6$.

The initial challenge in taking the model to the data arises from the fact that I only observe the origin location of workers, not their origin sectors. Specifically, I have access to $\mu_t^{i,nk}$ but not $\mu_t^{ij,nk}$. Therefore, I need to make an assumption regarding the migration cost $m_t^{ij,nk}$ in order to simplify the migration expression from $\mu_t^{ij,nk}$ to $\mu_t^{i,nk}$. Specifically, the migration cost $m_t^{ij,nk}$ comprises industry switching costs and relocation costs:

$$m_t^{ij,nk} = m_t^k + m_t^{i,n} \tag{28}$$

where the relocation cost $m_t^{i,n}$ between origin i and destination n can be split into two components, one of which is policy-relevant, following Caliendo et al. (2021). I denote the policy component as $mpol_t^{i,n}$, which represents the Ho Khau costs for migrants born in a different province than their desired destination n

$$m_t^{i,n} = m^{i,n} + mpol_t^{i,n}, \tag{29}$$

where the term $m^{i,n}$ represents time-invariant costs for the origin-destination pair (i, n), such as distance. I assume that $m_t^{i,i} = 0$ and $m_t^{i,n} > 0$ for $n \neq i$.

$$\mu_t^{i,nk} = \underbrace{\frac{e^{\left(\beta U_{t+1}^{nk} - m_t^k\right)/\rho\nu}}{\sum_{s=1}^{S} e^{\left(\beta U_{t+1}^{ns} - m_t^s\right)/\rho\nu}}}_{\mu_t^{nk}} \underbrace{\frac{\left(\sum_{s=1}^{S} e^{\left(\beta U_{t+1}^{ns} - m_t^{i,ns}\right)/\rho\nu}\right)^{\rho}}{\sum_{c=1}^{N}\left(\sum_{s=1}^{S} e^{\left(\beta U_{t+1}^{cs} - m_t^{i,cs}\right)/\rho\nu}\right)^{\rho}}}_{\mu_t^{i,n}}. \tag{30}$$

Furthermore, I can use the nested logit structure to decompose the migration and sectoral share into two components. In particular, $\mu_t^{ij,nk} = \mu_t^{nk}\mu_t^{ij,n}$ where the sectoral share within a location is given by

$$\mu_t^{nk} = \frac{e^{\frac{\beta}{\rho\nu}\ln w_{t+1}^{nk} - \frac{1}{\rho\nu} m_t^k}}{\sum_{s\in\mathcal{S}} e^{\frac{\beta}{\rho\nu}\ln w_{t+1}^{ns} - \frac{1}{\rho\nu} m_t^s}}, \tag{31}$$

and the migration flow is given by

$$\mu_t^{i,n} = \frac{\left(\sum_{k=1}^{S} e^{\left(\beta U_{t+1}^{nk} - m_t^{i,nk}\right)/\rho\nu}\right)^{\rho}}{\sum_{c=1}^{N}\left(\sum_{k=1}^{S} e^{\left(\beta U_{t+1}^{ck} - m_t^{i,ck}\right)/\rho\nu}\right)^{\rho}}. \tag{32}$$

5.2 Leveraging Tax Policy Variations to Estimate χ

This subsection aims to exploit the tax policy changes in timing, location, and firm age to estimate the structural parameter χ that governs the spatial firm entry elasticity with respect to profit tax. Specifically, it elucidates the effect of the place-based profit tax change described in Section 2 on the entry and exit dynamics of both young and old firms.

By applying the natural logarithm to the entry equation (18), the model implies how the entry share responds to tax cuts aimed at young firms. The resulting expression is given by:

$$\ln\psi_t^{ij} = \frac{\beta}{\chi} V_{Y,t+1}^{ij} - \frac{1}{\chi} f_t^{ij} - \ln\sum_{n=1}^{N}\exp\left(\beta V_{Y,t+1}^{nj} - f_t^{ij}\right)^{1/\chi}.$$

I observe that $\ln\sum_{n=1}^{N}\exp\left(\beta V_{Y,t+1}^{nj} - f^{nj}t\right)^{1/\chi}$ can be captured by a sector-age-time fixed effect, $\theta_{Y,t}^{j}$. Substituting (13) into $V_{Y,t+1}^{ij}$, the expected value of young firms, I arrive at:

$$\ln\psi_t^{ij} = \frac{\beta}{\chi}\ln(1-\tau_{Y,t+1}^{i})\frac{\pi_{t+1}^{ij}}{P_{t+1}^{i}} + \beta\ln\left(\exp\left(V_{O,t+2}^{ij}\right)^{\frac{\beta}{\chi}} + 1\right) - \frac{1}{\chi} f_t^{ij} - \theta_{Y,t}^{j}. \tag{33}$$

To isolate the impact of profit tax changes, $\tau_{Y,t+1}^{i}$, on firm entry, it's essential to account for the forward-looking nature of the firm's decision-making process, particularly the value of an old firm in ij, $V_{O,t+2}^{ij}$. I can derive sufficient statistics for this option value by formulating the future values of staying in ij as a function of the fraction of young firms that remain, based on equation (14):

$$\underbrace{1-\varsigma_{1,t+1}^{ij}}_{\text{Exit rate}} = \frac{1}{\exp(V_{O,t+2}^{ij})^{\frac{\beta}{\chi}} + 1}.$$

Rearranging terms and taking the natural logarithm on both sides, I get

$$\ln\left(\exp\left(V^{ij}_{O,t+2}\right)^{\frac{\beta}{\chi}}+1\right)=\ln(1-\varsigma^{ij}_{t+1})^{-1},$$

indicating that the share of young firms that exit location-sector ij, $1-\varsigma^{ij}_{t+1}$, encapsulates the future value of being an old firm in ij. Substituting this expression into equation (33) results in

$$\ln\left(\psi^{ij}_t(1-\varsigma^{ij}_{t+1})^{\beta}\right)=\frac{\beta}{\chi}\ln(1-\tau^i_{Y,t+1})\frac{\pi^{ij}_{t+1}}{P^i_{t+1}}-\frac{1}{\chi}f^{ij}_t-\theta^j_{Y,t}, \tag{34}$$

which relates local entry and exit to the firms' instantaneous values.

A reduction in the tax rate for young firms in a particular location i, denoted as $\tau^i_{1,t+1}$, increases the entry rate $\ln\psi^{ij}_t$ in the current period as the location's attractiveness improves. However, the exit rate of firms from this location in the subsequent period also rises due to the increased competition for inputs brought about by the surge in the number of firms.

The main challenge in running the regression derived from equation (34) lies in the potential correlation between changes in taxes on young entrepreneurs, $\tau^i_{Y,t+1}$, and other local factors such as local revenues, π^{ij}_{t+1}, and fixed costs, f^{ij}_t, which correlate with entrepreneurial activity. If the policy is intended to stimulate economic activity in areas with high market access, such as SEZs, then changes in taxes are likely to be correlated with changes in fixed costs, rendering a simple difference-in-difference design ineffective for identifying β/χ. However, the policy varies across districts, locations, and ages, so I can exploit these variations.

In particular, I use tax variation across ages to cancel out real pre-tax profits, $\frac{\beta}{\chi}\ln\frac{\pi^{ij}_{t+1}}{P^i_{t+1}}$, with data. Based on equation (14), I take the ratio between the fraction of young firms that stay and those that exit:

$$\frac{\varsigma^{ij}_t}{1-\varsigma^{ij}_t}=\exp\left(V^{ij}_{O,t+1}\right)^{\frac{\beta}{\chi}}.$$

Taking the natural logarithm on both sides and substituting the value of an old firm (11), I get

$$\ln\frac{\varsigma^{ij}_t}{1-\varsigma^{ij}_t}=\frac{\beta}{\chi}\left(\ln\left[(1-\tau^i_{O,t+1})\pi^{ij}_{t+1}/P^i_{t+1}\right]\right)+\beta\varrho\ln\frac{\varsigma^{ij}_{t+1}}{1-\varsigma^{ij}_{t+1}},$$

or

$$\ln \frac{\varsigma_t^{ij}}{1-\varsigma_t^{ij}} \left(\frac{1-\varsigma_{t+1}^{ij}}{\varsigma_{t+1}^{ij}} \right)^{\beta\varrho} = \frac{\beta}{\chi} \ln \left((1-\tau_{O,t+1}^{i}) \frac{\pi_{t+1}^{ij}}{P_{t+1}^{i}} \right),$$

An increase in the profit tax for old firms makes young firms more likely to stay than exit, as indicated by the relative tendency to stay, $\frac{\varsigma_t^{ij}}{1-\varsigma_t^{ij}}$. However, the likelihood of exit in the next period also increases for young firms due to a surge in market competition, captured in $\left(\frac{1-\varsigma_{t+1}^{ij}}{\varsigma_{t+1}^{ij}}\right)^{\beta\varrho}$.

Subtracting each side of this equation from the corresponding side of (34), I obtain

$$\ln \frac{\psi_t^{ij}(1-\varsigma_{t+1}^{ij})^{\beta}}{\frac{\varsigma_t^{ij}}{1-\varsigma_t^{ij}} \left(\frac{1-\varsigma_{t+1}^{ij}}{\varsigma_{t+1}^{ij}}\right)^{\beta\varrho}} = \frac{\beta}{\chi} \ln \frac{(1-\tau_{Y,t+1}^{i})}{(1-\tau_{O,t+1}^{i})} - \frac{1}{\chi} f_t^{ij} - \theta_{Y,t}^{j} \tag{35}$$

where f_t^{ij} can be encapsulated by a location-sector-time fixed effect. Meanwhile, taxes vary at the age-location-time level. This equation, by comparing the logarithms of the ratios of staying-to-exiting firms for young and old firms, allows us to isolate the impact of tax variations on the behavior of firms.

In Appendix B.2, the model is extended to encompass continuous ages where $a \in \{1, \ldots, A\}$ denotes firms' ages. This extension avoids defining which entrepreneurs are classified as "Young" or "Old" in the data and precludes the need for estimating the exogenous exit rate ϱ. Consequently, the equation for estimation features a triple difference-in-difference design and is given by

$$\ln \frac{\psi_t^{ij}(1-\varsigma_{1,t+1}^{ij})^{\beta}}{\frac{\varsigma_{a-1,t}^{ij}}{1-\varsigma_{a-1,t}^{ij}}(1-\varsigma_{a,t+1}^{ij})^{\beta}} = \gamma_F \ln \frac{1-\tau_{1,t+1}^{i}}{1-\tau_{a,t+1}^{i}} + \Gamma_t^{ij} + \theta_{\widetilde{A},t}^{j} + \varphi_{\widetilde{A}}^{i} + \varepsilon_{a,t}^{ij}. \tag{36}$$

The dependent variable in this equation, called the "Local Age-Specific Turnover Rate" (LAST), includes four components. The first ψ_t^{ij} is the share of entrants in location i between periods t and $t+1$ relative to all entrants in location-sector ij at time t. Next, the share of 1-year-old establishments exiting location-sector ij between period $t+1$ and $t+2$ is represented by $(1-\varsigma_{1,t+1}^{ij})$. Third, $\frac{\varsigma_{a-1,t}^{ij}}{1-\varsigma_{a-1,t}^{ij}}$ signifies the ratio of the share of $a-1$-year-old establishments that remain in location-sector ij between t and $t+1$ to the share of those that exit over the same period. Fourth, $(1-\varsigma_{a,t+1}^{ij})$ is the share of a-year-old establishments that exit between $t+1$ and $t+2$ from location-sector ij.

The main independent variable, $\ln \frac{1-\tau^i_{1,t+1}}{1-\tau^i_{a,t+1}}$, reflects the logarithmic difference in net profit rates between 1-year-old and a-year-old establishments. For example, for $t < 2003$, this variable is 0 for all location i, age a. After 2003, for $a = 3$ and $i \in$ Backward (B) districts, $\ln(1-\tau^i_{1,t+1}) - \ln(1-\tau^i_{3,t+1}) = \ln(1-0) - \ln(1-0.1) = -\ln(0.9) > 0$. Besides the qualitative prediction that the coefficient of interest $\gamma_F > 0$, the estimated γ_F identifies $\frac{\beta}{\chi}$. Since β is simply the discount rate which is 0.95, I can uncover χ.

The unit of observation in regression (36) comprises district-zone, 2-digit sector level, age, and year and incorporates fixed effects for district-sector-time (Γ^{ij}_t), sector-age group-time ($\theta^j_{\widetilde{A},t}$), and district-age group ($\varphi^i_{\widetilde{A}}$). Age groups $\widetilde{A}$ is a partition of the set of all ages A based on tax policy classifications: 1-2 years old, 3-8 years old, 8-10 years old, 10-12 years old, and 13 years old and above. The error term ($\varepsilon^{ij}_{a,t}$) accounts for potential measurement errors in entry and exit rates, such as formalizing existing informal entrepreneurs. I estimate (36) using the Poisson Pseudo Maximum Likelihood (PPML) estimator and cluster the standard errors at the district and age group levels.

The assumption underlying the identification of γ_F is that tax changes across time, locations, and age cohorts do not correlate with the error term. To invalidate this assumption, an unobserved variable, such as a new technology or infrastructure project, would need to disproportionately favor 1-year-old establishments over older ones and concurrently shift with the district tax labels, age groups, and the year of the tax policy.

Table 2 presents the firm entry elasticity of 0.82, which is statistically significant at the 1 percent level. Given that $\beta = 0.95$, $\chi = 1.16$.

5.3 Identification strategy for changes in Ho Khau cost

As Section 2.3 explains that the Ho Khau policy features a change over time and space, this subsection aims to estimate both of these changes. Define the time difference in Ho Khau policy for any pairs of locations i and n as

$$\Delta^{i,n} = mpol^{i,n}_{post} - mpol^{i,n}_{pre},$$

where I recall $mpol^{i,n}_t$ is a component of the migration cost in (29). After the policy change in 2005, Ho Khau cost should drop everywhere, i.e. $\Delta^{i,n} < 0$ for all i, n.

Not only did the Ho Khau cost drop over time, but it also varied across locations. Two estimands are the average change over time due to Ho Khau $\Delta_T \equiv \mathbb{E}[\Delta^{i,n}]$ and the spatial variation magnitude $\Delta_L \equiv \mathbb{E}[\Delta^{\prime Urban} - \Delta^{\prime Rural}]$, which should be positive as the drop in ho khau cost in Rural is more than in that of Urban.

Table 2: Triple Difference-in-Differences estimate of the firm entry elasticity

	(1)
Log(Net Profit Rate Ratio)	0.82**
	(0.36)
Observations	45,824

* *Notes:* Unit of observation is district-zone, 2-digit sector level, age, and year. The dependent variable is the Local Age-Specific Turnover Rate given in (36), while the independent variable is the log difference in net profit rates between 1-year-old and a-year-old for $a > 1$. Standard errors, clustered two ways at the district and age group level, are reported in parenthesis.

To identify these changes, I rely on migration data and the relationship between migration flows, and migration costs in (32). Applying this equation and taking the natural log of the ratio between the migration shares from i to n and the shares of stayers in i yields

$$\ln \frac{\mu_t^{i,n}}{\mu_t^{i,i}} = -\frac{1}{\nu} m_t^{i,n} + \rho \ln \sum_{j=1}^{S} e^{\left(\beta U_{t+1}^{nj} - m_t^j\right)/\rho\nu} - \rho \ln \sum_{j=1}^{S} e^{\left(\beta U_{t+1}^{ij} - m_t^j\right)/\rho\nu}. \tag{37}$$

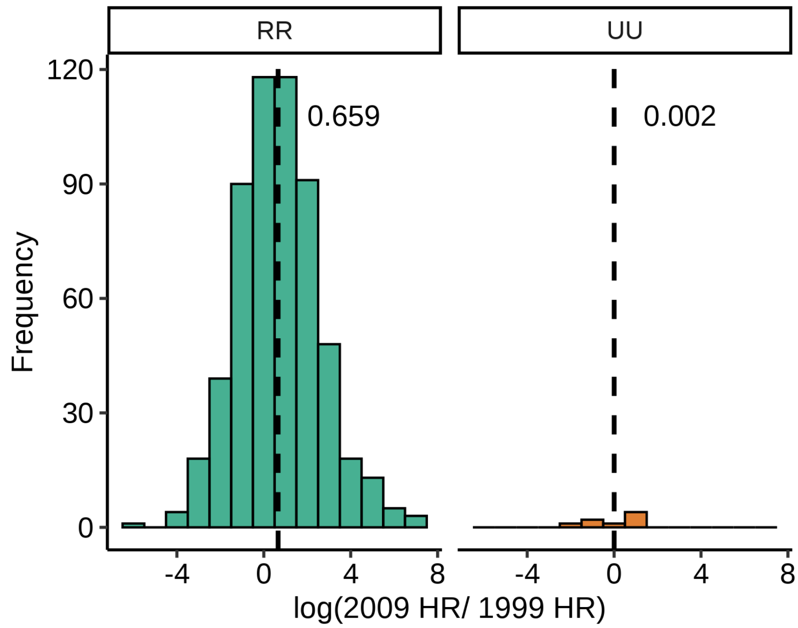
I am interested in isolating the effect of changes in migration costs $m_t^{i,n}$ on migration flows. Upon inspecting equation (37), any changes in migration cost $m_t^{i,n}$ can alter the future value of being in n through the second term on the right-hand side, which captures the general equilibrium effects on wages and prices. Thus, a simple Diff-in-Diff design with a dummy for Urban and Post is insufficient to account for changes in the option values.

A solution is to use the Head and Ries (2001) trick, given by

$$y_t^{i,n} \equiv \ln \left(\frac{\mu_t^{i,n}}{\mu_t^{i,i}} \frac{\mu_t^{n,i}}{\mu_t^{n,n}} \right) = -\frac{1}{\nu} (m_t^{i,n} + m_t^{n,i})$$

where I call the LHS $y_t^{i,n}$ the Head Ries (HR) Index. Since the Head Ries Index is symmetric for i, n, and n, i, I consider only the locations such as the index on $i < n$ to avoid

Figure 5: Distributions of Log(HR Ratio)



Notes:

duplicating observations. Taking the time difference of the HR $y_t^{i,n}$ yields

$$y_{post}^{i,n} - y_{pre}^{i,n} = -\frac{1}{\nu}\left(\Delta^{i,n} + \Delta^{n,i}\right).$$

Changes of Ho Khau over time, scaled by ν**:** To estimate Δ_T, rely on the changes in the Head Ries (HR) over time for Rural-to-Rural flows and Urban-to-Urban flows, assuming symmetric costs between them ($\Delta^{i,n} = \Delta^{n,i}$) when i and n belong to the same flow type. The main identifying assumption is that other time-varying changes are negligible compared to the changes in the Ho Khau policy.

In Figure 5, I constructed the HR for each pair of locations before and after the policy change (specifically, in 1999 and 2009), and plotted the logarithm of the ratio between the HR in 2009 and the HR in 1999 for Rural-Rural (RR) and Urban-Urban (UU) flows. The mean of each distribution is indicated next to the dashed line.

The log of the ratio of the HRs corresponds to $\frac{-2}{\nu}\Delta^{i,n}$ in the model. If $\nu = 1$, then the means on the plots reflect the declines in migration costs over time for both types, which is consistent with the decline in migration costs over time of the policy.

In addition, the decline in RR is much larger than in UU, which is consistent with the spatial variation of the policy.

5.4 BTA to separately identify ρ and ν

This subsection uses the exogenous demand shock from the trade BTA shock to identify the worker's preferences based on two structural parameters ρ and ν. To this end, I need to unpack the general equilibrium term the migration flows equation by substituting the expected value (5) and migration cost (29) into (37) to get

$$\ln \frac{\mu_t^{i,n}}{\mu_t^{i,i}} = \rho \ln \frac{\sum_j (w_{t+1}^{nj})^{\frac{\beta}{\rho\nu}} e^{-\frac{m_t^j}{\rho\nu}} / P_{t+1}^n}{\sum_j (w_{t+1}^{ij})^{\frac{\beta}{\rho\nu}} e^{-\frac{m_t^j}{\rho\nu}} / P_{t+1}^i} + \frac{\beta}{\nu}\left(\Xi_{t+1}^n - \Xi_{t+1}^i\right) - \frac{1}{\nu} m_t^{i,n} + \frac{\beta}{\nu} \ln\left(\frac{A_{t+1}^n}{A_{t+1}^i}\right) \tag{38}$$

where the future value of being location i, Ξ_t^i, is the location version of the location-sector value Ξ_t^{ij} given in (4), i.e. $\Xi_t^i = \sum_{n=1}^N \left(\sum_{k=1}^J e^{\left(\beta U_{t+1}^{nk} - m_t^{i,n}\right)/\rho\nu}\right)^\rho$. Similar to using the exit rate as sufficient statistics for the firm estimation in subsection 5.2, the sufficient statistics for the difference in future values is given by

$$\ln\left(\frac{\mu_t^{i,n}}{\mu_t^{n,n}}\right) = -\frac{1}{\nu} m_t^{i,n} + \frac{1}{\nu}\left(\Xi_t^n - \Xi_t^i\right).$$

Hence, substituting this expression to (38) yields

$$\ln \frac{\mu_t^{i,n}}{\mu_t^{i,i}} \left(\frac{\mu_{t+1}^{i,n}}{\mu_{t+1}^{n,n}}\right)^{-\beta} = \rho \ln\left(\mathcal{W}_t^n\right) - \frac{1}{\nu} mpol^{i,n} + \alpha^{i,n} + \Omega_t^i + \varepsilon_t^{i,n}$$

where Ω_t^i captures the origin-time fixed effects, $\mathcal{W}_t^n = \frac{\sum_j (w_{t+1}^{nj})^{\frac{\beta}{\rho\nu}} e^{-\frac{m_t^j}{\rho\nu}}}{(P_{t+1}^n)^{\frac{\beta}{\rho\nu}}}$ captures the real wage in the destination location, and the error term $\varepsilon_t^{i,n} = \frac{\beta}{\nu} \ln\left(\mathcal{A}_{t+1}^n\right)$ since the ho khau change is a one-time policy change. In other words, the resulting estimating regression is given by

$$\ln \frac{\mu_t^{i,n}}{\mu_t^{i,i}} \left(\frac{\mu_{t+1}^{i,n}}{\mu_{t+1}^{n,n}}\right)^{-\beta} = \rho \ln\left(\mathcal{W}_t^n\right) + \mathbf{1}(n \in Urban, t > 2005) + \alpha^{i,n} + \Omega_t^i + \theta_t + \gamma_Z Z_t^n + \varepsilon_t^{i,n}. \tag{39}$$

This equation suggests the following estimation steps to identify ρ and ν separately. First, I need to compute real wages $\mathcal{W}_t^n$. While I have wage data at the location-sector level, I do not observe the switching costs m_t^j, the local price indexes P_t^n, or the product $\rho\nu$. To overcome these challenges, I first estimate $\rho\nu$ based on time and sectoral variations of the BTA in 5.4.1, then I will use more data and the local price index equation (3) to compute the local price index in 5.4.2. I can then construct a real wage net of switching costs $\widetilde{\mathcal{W}}_t^n = \frac{\sum_j (w_{t+1}^{nj})^{\frac{\beta}{\rho\nu}}}{(P_{t+1}^n)^{\frac{\beta}{\rho\nu}}}$. Finally, to account for the measurement error and the potential correlation between changes in the real wages and amenities in the error terms, I will use an instrument for $\widetilde{\mathcal{W}}_t^n$ described more in 5.4.3.

5.4.1 Identification of $\rho\nu$

To identify $\rho\nu$, taking the natural log of the migration-sector share equation (30) yields

$$\ln \mu_t^{i,nk} = \frac{\beta}{\rho\nu} U_{t+1}^{nk} - \frac{1}{\rho\nu} m_t^{i,nk} + \ln \left(\sum_{s=1}^{S} e^{\left(\beta U_{t+1}^{ns} - m_t^{i,ns}\right)/\rho\nu} \right)^{\rho-1} - \ln \sum_{c=1}^{N} \left(\sum_{s=1}^{S} e^{\left(\beta U_{t+1}^{cs} - m_t^{i,cs}\right)/\rho\nu} \right)^{\rho}.$$

The last term is an origin-time fixed effect, denoted by θ_t^i. Next, substituting U_{t+1}^{nk} from (5) yields

$$\ln \mu_t^{i,nk} = \frac{\beta}{\rho\nu} \ln w_{t+1}^{nk} - \frac{1}{\rho\nu} m_t^{i,nk} + \ln \frac{\mathcal{A}_{t+1}^n}{P_{t+1}^n} \left(\sum_{s=1}^{S} e^{\left(\beta U_{t+1}^{ns} - m_t^{i,ns}\right)/\rho\nu} \right)^{\rho-1} + \Xi_{t+1}^n + \theta_t^i.$$

This relationship suggests the following regression:

$$\ln \mu_t^{i,nk} = \gamma_S \ln w_{t+1}^{nk} + \Omega_i^{nk} + \Phi_t^n + \theta_t^i + \varepsilon_t^{i,nk}, \tag{40}$$

where Ω_i^{nk} is the sector-origin-destination fixed effects to capture any time-invariant differences across locations and sectors, such as some particular origin-destination flows of migrants that work in particular sectors; Φ_t^n and θ_t^i is the destination-time and origin-time fixed effects, respectively that absorbs the changes in location option values.

To address the issue of endogeneity between switching costs in the error terms and wages, I introduce an instrument for wages

$$\tilde{z}_t^k \equiv \Delta\tau^k \times \mathbf{1}(t > 2001)$$

where $\Delta\tau^k$ is the difference between Column 2 and the Most Favored Nation (MFN) tariff at the time of the Bilateral Trade Agreement (BTA). A larger difference between the column 2 tariff and the MFN US tariff indicates a larger demand shock, which in turn predicts a larger wage change.

In addition to the endogeneity challenge, the available data do not capture actual switching between sectors as the origin sector is not observed. To mitigate this limitation, I include location-time fixed effects in the analysis. These fixed effects help capture the potential bias arising from switching cost behaviors by accounting for the correlation between location and sectoral characteristics. For example, certain locations may specialize in specific industries. By including destination-time fixed effects at a finer geographic level, I can better absorb variations in wages that are not driven by the BTA, such as sectoral trends.

Thus, my observation unit for this regression is the migration flows from province i to district n and working in sector k at the 3-digit level for three periods 1995-1999, 2005-2009, and 2015-2019. The local wage w_{t+1}^{nk} information comes from the wage rate in location-sector nk, computed from wage bill and employment data from firm surveys in 2000, 2006-2009, and 2014-2015 for 1999, 2009, and 2019, respectively. Standard errors are clustered at the sector level.

Table 3 reports the OLS and TSLS estimates of $\widehat{\gamma}_S$ from (40), which reflects estimates of our estimand $\beta/(\rho\nu)$. Since $\beta = 0.95$, I can back out $\rho\nu = 0.95/2.1 = 0.45$.

5.4.2 Local Price Index

I gather data and estimates to compute the local price index (3), which comprises sector-location price index (23). Substituting prices (9) into (23) yields

$$P_t^{ij} = \left(\sum_{n=1}^{N} E_t^{nj} \left(\frac{\sigma}{\sigma - 1} \frac{d^{nj,ij} x_t^{nj}}{A_t^{nj}} \right)^{1-\sigma} \right)^{1/(1-\sigma)} .$$

Variable cost bundle x_t^{nj} I compute the variable cost bundle based on a rearrangement of (7)

$$x_t^{nj} = B w_t^{nj} \left(\frac{1-\xi}{\xi} \frac{L_t^{nj}}{K_t^{nj}} \right)^{1-\xi},$$

where I use data from the aggregate capital-to-labor ratio by year-sector-location from the firm data and multiply it with the local wage rate. To map these migration data from the two censuses 1999 and 2019 in (39). I take the 2000 survey for 1999 while the mean across 2006 and 2009 for the 2009 census.

Table 3: OLS and TSLS estimates of $\frac{\beta}{\rho\nu}$

	ln(Flows)		ln(Wage)
	OLS	IV	1stStage
	(1)	(2)	(3)
ln(Wage)	0.122**	2.109**	
	(0.051)	(0.996)	
Δ Tariff $\times$ Post			0.434***
			(0.103)
Observations	33,779	33,779	33,779
Wald (IV only)			17.83

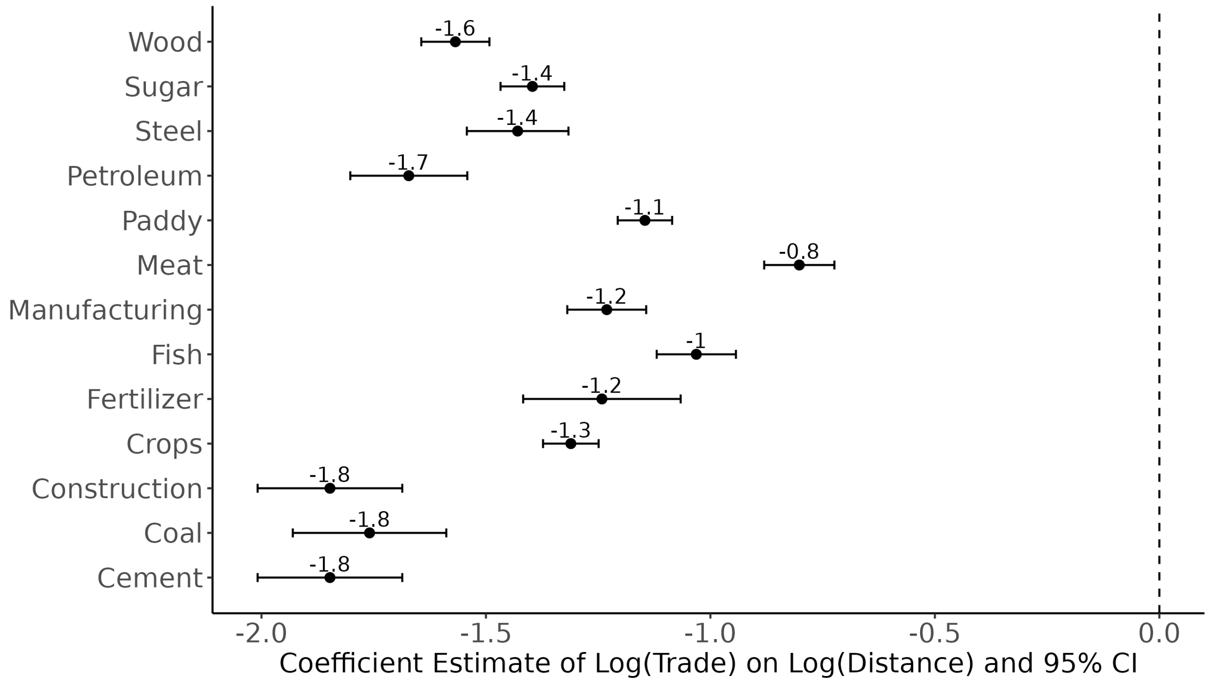
* *Notes:* The migration flows data was obtained from 1999, 2009, and 2019 population censuses. The wage information reflects the wage in location-sector nk, computed from wage bill and employment data from firm surveys in 2000, 2006-2009, and 2014-2015 for 1999, 2009, and 2019, respectively. All regressions include sector-origin-destination, origin-year, and destination-year fixed effects. Robust standard errors clustered at the sector level are reported in parentheses. ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$.

Trade costs $d^{ij,nj}$ Taking the natural log of the trade share equation (24) and considering the cross-sectional variation of the trade flows from location-sector ij to nj yields

$$\ln(\lambda^{ij,nj}) = (1-\sigma)\kappa^j \ln(\text{distance}^{i,n}) + \text{Origin FE } + \text{Destination FE } + \varepsilon^{ij,nj}.$$

I digitized interprovincial trade data in 2000 from JICA (2000) and got truck distance for each pair of provinces using the 1999 map and applied the network analysis tools from ArcGIS. In particular, I first get the province map GIS 1999 from IPUMS, then get the feature to point tool to get centroids. Finally, I use network analysis to get O-D matrix between centroids (truck distance). I match the O-D distance to the reported distances with trade flows. Given these data, I estimate $(1-\sigma)\kappa^j$ using PPML. Figure 6 reports their estimates similar to -1.29 estimated across all sectors in the US from Monte et al. (2018).

Figure 6: Estimates of distance elasticities $(1-\sigma)\kappa^j$



Notes: Inter-provincial trade flows data come from JICA (2000). Standard errors are clustered at the origin-destination level.

Local TFP A_t^{ij} I estimate local TFP A_t^{ij} from the revenue equation, which is the product of σ and the pre-tax profits (10). Taking the natural log of this expression yields

$$\ln(\text{Revenue})_t^{f,ij} = \gamma^f + \theta_t^{ij} + \varepsilon_t^{f,ij}$$

where I observe $(\text{Revenue})_t^{f,ij}$ of firm f in sector j and district i in year t. The terms γ^f and γ_t^{ij} are firm fixed effects and district-sector-year fixed effects, respectively. I extract the residuals of this regression, divide them by $\sigma - 1$, and interpret them as the firm-specific TFP $A_t^{f,ij}$ after taking the exponential. I then aggregate these TFPs to take the mean of TFP for each location-sector market. Nonetheless, this residual approach is prone to have measurement errors for TFP measure A_t^{ij}, which requires an instrument when using it to estimate ρ.

Expendituare shares α^j I need the expenditure shares to compute the local price index (3). I use three equations (26), (25), and (10). First, I substitute (26) into (10) to get

$$\pi_t^{ij} = \alpha^j \bar{K} \left(\frac{x_t^{ij}}{A_t^{ij}} \right)^{1-\sigma} \left(\sum_{n=1}^{N} (d^{ij,nj})^{1-\sigma} \Pi_t^n (P_t^{nj})^{\sigma-1} \right)$$

Then taking the sum across all locations, I get

$$\alpha^j = \frac{\sum_i \pi_t^{ij}}{\sum_i \bar{K} \left(\frac{x_t^{ij}}{A_t^{ij}} \right)^{1-\sigma} \left(\sum_{n=1}^{N} (d^{ij,nj})^{1-\sigma} \Pi_t^n (P_t^{nj})^{\sigma-1} \right)}.$$

5.4.3 Estimation details for ρ and result

I have constructed $\widetilde{W}_t^n$ for regression (39). However, as explained earlier, this object has measurement errors and can correlate with the error terms. Thus, I use the following instrument for it:

$$z_t^n = \sum_{s=1}^{S} \frac{L_0^{is}}{L_0^i} \Delta\tau^s \times \mathbf{1}(t > 2001).$$

The expression $\sum_{s=1}^{S} \frac{L_0^{is}}{L_0^i} \Delta\tau^s$ is the Bartik or shift-share variable of the BTA shocks. In particular, $\Delta\tau^s$ is the change in U.S. tariff for industry s calculated as the difference between the Column 2 and MFN tariff at the time of the BTA, the employment share L_{1999}^{is}/L_{1999}^i uses the 1999 population census where L_0^{is} is the number of workers in sector s in province i while L_0^i is the number of workers in traded sectors in province i, following Kovak (2013).

Z_t^n is a vector of time-varying province controls, including the linear time trends of 1999 shares of workers in agriculture, manufacturing, and Kinh (the major ethnic group) population to address concerns regarding existing trends at the locations, which is similar to McCaig (2011).

To account for changes in the profit taxes, I include the shares of the C districts population in province n as another instrument for real wages. Standard errors are clustered two-way at the origin and destination.

Table 4 reports the estimate of $rho = 0.72$. From $\rho\nu = 0.45$, I obtain $\nu = 0.625$.

Table 4: OLS and TSLS estimations of ρ

	OLS	IV	1st Stage
	(1)	(2)	(3)
Log(Real Wage)	0.010	0.719**	
	(0.047)	(0.341)	
Bartik x Post			6.243***
			(1.364)
C share x Post			0.349***
			(0.094)
Major x Post	-0.176	0.259	-0.688***
	(0.138)	(0.260)	(0.058)
Observations	4,608	4,608	4,608
F-test (1st stage)			46.66
Sargan, p-value		0.93	

* *Notes:* All models include origin-destination, origin-year, and year-fixed effects, and controls for linear time trends of agriculture, manufacturing, and Kinh (the major ethnic group) population share in 1999. Standard errors are clustered two-way at the origin and destination (in parentheses).

6 Counterfactuals

To solve for counterfactuals, I extend the dynamic hat algebra developed by Caliendo et al. (2019). The results are working in progress.

7 Conclusion

References

Aguirregabiria, V. (2021). Empirical Industrial Organization: Models, Methods, and Applications. Retrieved April 17, 2023, from http://aguirregabiria.net/wpapers/book_dynamic_io.pdf

Artuç, E., Chaudhuri, S., & McLaren, J. (2010). Trade Shocks and Labor Adjustment: A Structural Empirical Approach. *American Economic Review*, *100*(3), 1008–45.

Atalay, E., Hortaçsu, A., Runyun, M., Syverson, C., & Ulu, M. F. (2023). Micro- and Macroeconomic Impacts of a Place-Based Industrial Policy.

Austin, B., Glaeser, E., & Summers, L. (2018). Jobs for the Heartland: Place-Based Policies in 21st-Century America. *Brookings Papers on Economic Activity*, 151–232.

Bryan, G., Mobarak, A. M., & Chowdhury, S. (2014). Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh. *Econometrica*, *82*(5), 1671–1748.

Busso, M., Gregory, J., & Kline, P. (2013). Assessing the Incidence and Efficiency of a Prominent Place Based Policy. *American Economic Review*, *103*(2), 897–947.

Caliendo, L., Dvorkin, M., & Parro, F. (2019). Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock. *Econometrica*, *87*(3), 741–835.

Caliendo, L., Opromolla, L. D., Parro, F., & Sforza, A. (2021). Goods and Factor Market Integration: A Quantitative Assessment of the EU Enlargement. *Journal of Political Economy*.

Caliendo, L., & Parro, F. (2020). The Quantitative Effects of Trade Policy on Industrial and Labor Location. *Mimeo*, 1–83.

Chaurey, R. (2017). Location-based tax incentives: Evidence from India. *Journal of Public Economics*, *156*, 101–120.

Chen, J., & Roth, J. (2023). *Log-like? Identified ATEs defined with zero-valued outcomes are (arbitrarily) scale-dependent*.

Cling, J., Razafindrakoto, M., & Roubaud, F. (2011). *The informal economy in Vietnam Hanoi: ILO* (tech. rep.).

Dunne, T., Roberts, M. J., & Samuelson, L. (1988). Patterns of Firm Entry and Exit in U.S. Manufacturing Industries. *The RAND Journal of Economics*, *19*(4), 495.

Fajgelbaum, P., & Gaubert, C. (2020). Optimal spatial policies, geography, and sorting. *Quarterly Journal of Economics*, *135*(2), 959–1036.

Fajgelbaum, P., Morales, E., Serrato, J. C. S., & Zidar, O. (2018). State Taxes and Spatial Misallocation. *The Review of Economic Studies*, *86*(1), 333–376.

Fukase, E. (2013). Export Liberalization, Job Creation, and the Skill Premium: Evidence from the US-Vietnam Bilateral Trade Agreement (BTA). *World Development*, *41*(1), 317–337.

Hasan, R., Jiang, Y., & Rafols, R. M. (2021). Place-based preferential tax policy and industrial development: Evidence from India's program on industrially backward districts. *Journal of Development Economics*, *150*, 102621.

Head, K., & Ries, J. (2001). Increasing Returns Versus National Product Differentiation as an Explanation for the Pattern of U.S.–Canada Trade. *American Economic Review*, *91*(4), 858–876.

Heblich, S., Seror, M., Xu, H., & Zylberberg, Y. (2022). Industrial Clusters in the Long Run: Evidence from Million-Rouble Plants in China.

Imbert, C., Seror, M., Zhang, Y., & Zylberberg, Y. (2022). Migrants and Firms: Evidence from China. *American Economic Review*, *112*(6), 1885–1914.

Jia, N., Molloy, R., Smith, C., & Wozniak, A. (2023). The Economics of Internal Migration: Advances and Policy Questions. *Journal of Economic Literature*, *61*(1), 144–180.

JICA. (2000). *The study on the national transport development strategy in the Socialist Republic of Vietnam (VITRANSS) final report* (tech. rep. Vol.2). Retrieved August 17, 2022, from https://openjicareport.jica.go.jp/pdf/11596764.pdf

Kinnan, C., Wang, S.-Y., & Wang, Y. (2018). Access to Migration for Rural Households. *American Economic Journal: Applied Economics*, *10*(4), 79–119.

Kleinman, B., Liu, E., & Redding, S. J. (2023). Dynamic Spatial General Equilibrium. *Econometrica*, *91*(2), 385–424.

Kline, P., & Moretti, E. (2014). Local economic development, agglomeration economies, and the big push: 100 years of evidence from the Tennessee Valley Authority. *Quarterly Journal of Economics*, *129*(1), 275–331.

Kovak, B. K. (2013). Regional effects of trade reform: What is the correct measure of liberalization? *American Economic Review*, *103*(5), 1960–1976.

Lagakos, D., Mobarak, A. M., & Waugh, M. E. (2023). The Welfare Effects of Encouraging Rural–Urban Migration. *Econometrica*, *91*(3), 803–837.

Liu, A. Y. C., & Meng, X. (Eds.). (2019). *Rural-Urban Migration in Vietnam*. Springer International Publishing.

Lu, Y., Wang, J., & Zhu, L. (2019). Place-Based Policies, Creation, and Agglomeration Economies: Evidence from China's Economic Zone Program. *American Economic Journal: Economic Policy*, *11*(3), 325–360.

Malesky, E., & Taussig, M. (2009). Out of the Gray: The Impact of Provincial Institutions on Business Formalization in Vietnam. *Journal of East Asian Studies*, *9*(2), 249–290.

McCaig, B. (2011). Exporting out of poverty: Provincial poverty in Vietnam and U.S. market access. *Journal of International Economics*, *85*(1), 102–113.

McCaig, B., & Pavcnik, N. (2013). Moving out of Agriculture: Structural Change in Vietnam.

McCaig, B., & Pavcnik, N. (2015). Informal Employment in a Growing and Globalizing Low-Income Country. *American Economic Review*, *105*(5), 545–550.

McCaig, B., & Pavcnik, N. (2018). Export markets and labor allocation in a low-income country. *American Economic Review*, *108*(7), 1899–1941.

McCaig, B., & Pavcnik, N. (2021). Entry and Exit of Informal Firms and Development. *IMF Economic Review*, *69*(3), 540–575.

McCaig, B., Pavcnik, N., & Wong, W. F. (2022). FDI Inflows and Domestic Firms: Adjustments to New Export Opportunities.

Monte, F., Redding, S. J., & Rossi-Hansberg, E. (2018). Commuting, Migration, and Local Employment Elasticities. *American Economic Review*, *108*(12), 3855–3890.

Morris-Levenson, J., & Prato, M. (2021). *The Origins of Regional Specialization*.

Mullahy, J., & Norton, E. (2022). *Why Transform Y? A Critical Assessment of Dependent-Variable Transformations in Regression Models for Skewed and Sometimes-Zero Outcomes* (tech. rep. w30735). National Bureau of Economic Research. Cambridge, MA.

Neumark, D., & Simpson, H. (2015). Chapter 18 - Place-Based Policies. In G. Duranton, J. V. Henderson, & W. C. Strange (Eds.), *Handbook of Regional and Urban Economics* (pp. 1197–1287). Elsevier.

Peters, M. (2022). Market Size and Spatial Growth—Evidence From Germany's Post-War Population Expulsions. *Econometrica*, *90*(5), 2357–2396.

Shukla, G. P., Phạm, M. Đ., Engelschalk, M., & Lê, M. T. (2011). Cải cách Thuế ở Việt Nam: Hướng tới một hệ thống Hiệu quả và Công bằng hơn. *Ban Quản lý Kinh tế và Xóa đói Giảm nghèo Khu vực Châu Á và Thái Binh Dương*, 316.

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

Vu, T. M., & Yamada, H. (2022). Do enterprise zones promote local business development? Evidence from Vietnam. *Journal of Regional Science*, *62*(4), 1206–1224.

Wang, F., Milner, C., & Scheffel, J. (2021). Labour market reform and firm-level employment adjustment: Evidence from the hukou reform in China. *Journal of Development Economics*, *149*, 102584.

Wang, J. (2013). The economic impact of Special Economic Zones: Evidence from Chinese municipalities. *Journal of Development Economics*, *101*(1), 133–147.

World Bank Group & Vietnam Academy of Social Sciences. (2016). *Vietnam's Household Registration System*. Ha Noi: Hong Duc Publishing House.

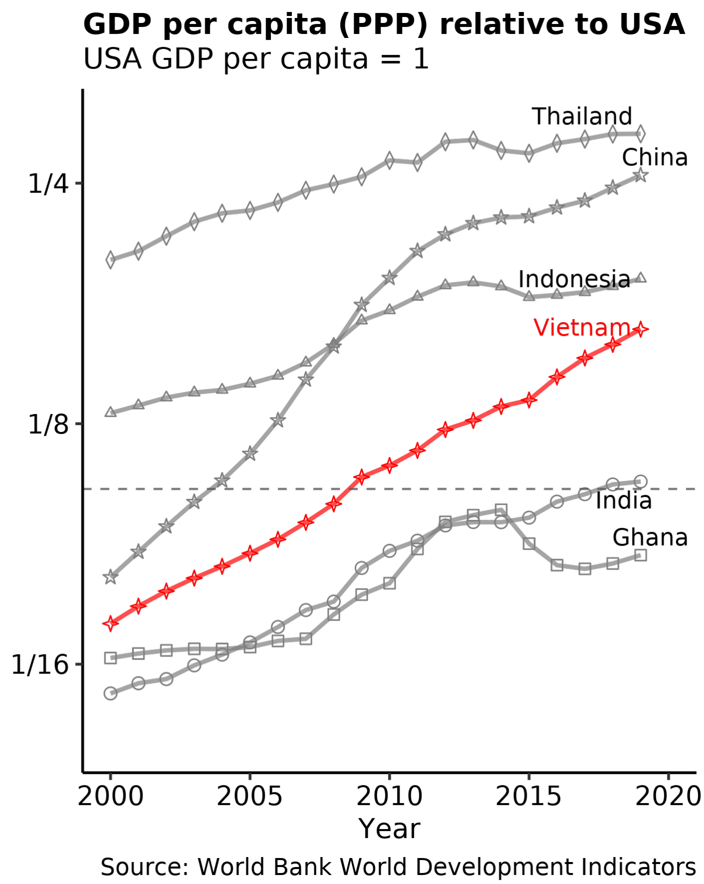


Figure A1: GDP per capita (PPP) relative to the USA, 2000-2019

Notes: The dashed line represents the average GDP per capita (PPP) for over 150 countries, all of which are below 1/4 of the United States' GDP per capita in 2000.

A Appendix

A.1 Appendix Context

A.2 Firm data appendix

The data cover almost all private and foreign plants. Most private and foreign firms are single-plant, while it is common for state-owned enterprises to have multiple plants.

When looking at the entry rates of around 27%, I see that they are high relative to those in other countries like the US where the entry rate is about 10% (Dunne et al., 1988). This high entrant rate reflects the high growth rate of Vietnam during this period of about 7%. Second, the government actively privates SOEs . Third, McCaig and Pavcnik (2021)

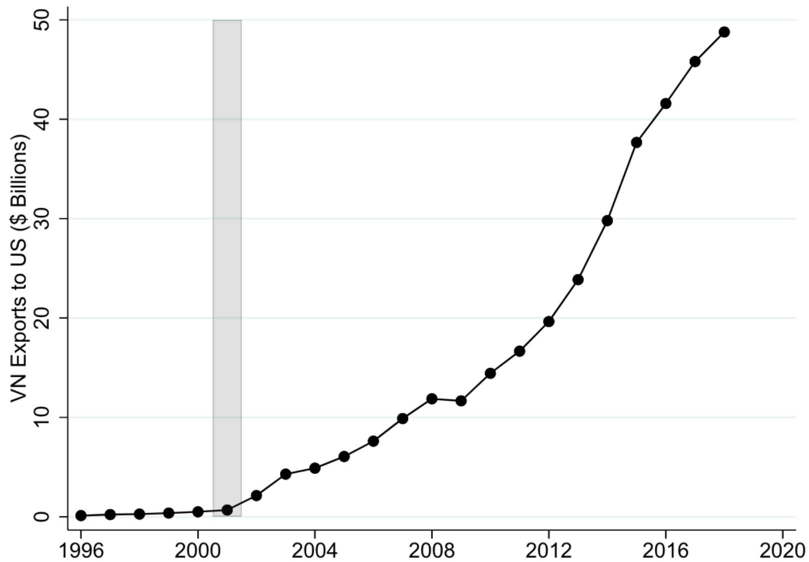


Figure A2: Vietnamese manufacturing exports to the US in billions of USD, 1996-2018. Source: McCaig et al. (2022)

observe that this high entry rate is comparable to other countries that move away from centrally-planed.

A.3 Household Data Appendix

A.4 More DiD results

B Model appendix

In this appendix, I present the steps to solve for the value functions of each agent

B.1 Sectoral and migration shares

To derive results in Section 4.2, I follow Appendix 1, section 11.3, in Aguirregabiria (2021). First, I derive the distribution of the maximum utility. Denote by $\bar{\delta}_t^{ij,nk}$ for the value of working in nk for an individual being in location-sector ij at time t, i.e.,

$$\bar{\delta}_t^{ij,nk} \equiv \beta U_{t+1}^{nk} - m_t^{ij,nk}.$$

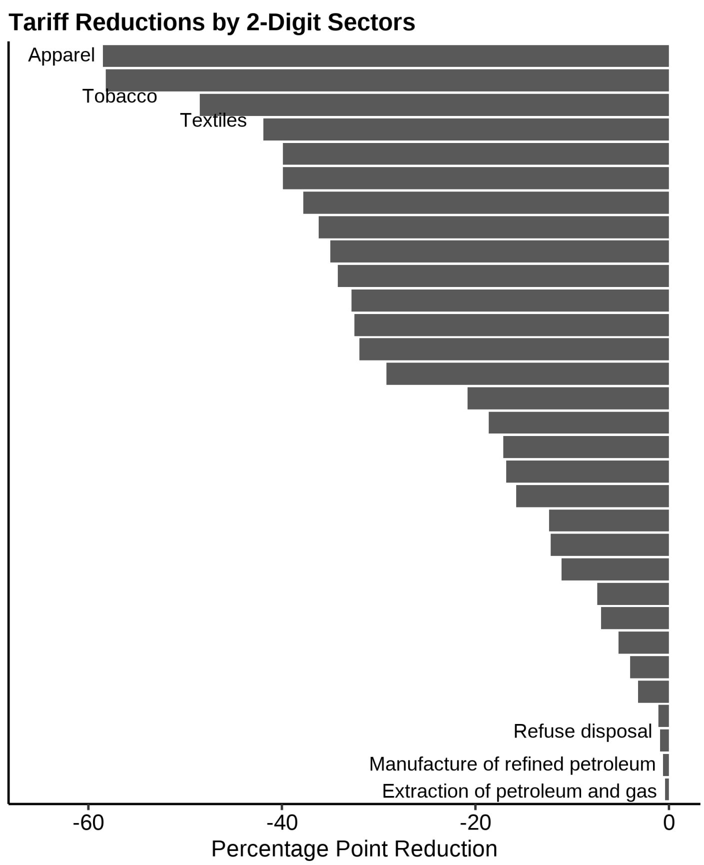


Figure A3: U.S. Tariff reductions for Vietnamese goods. Source: McCaig (2011)

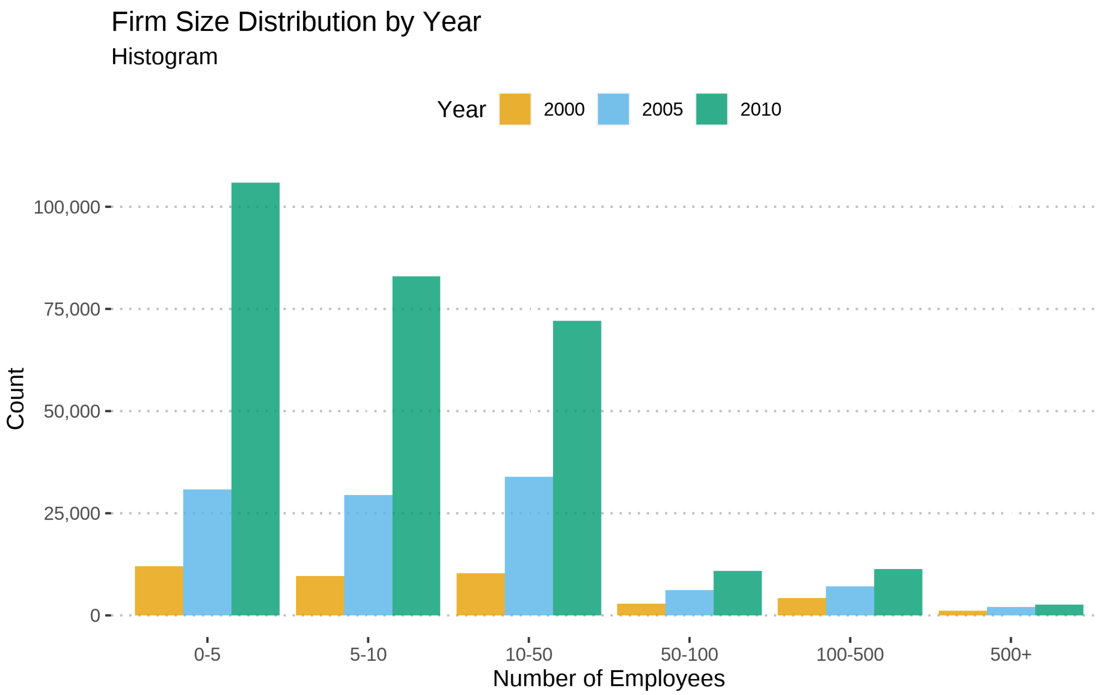


Figure A4: Firm Size Distribution 2000, 2005, and 2010

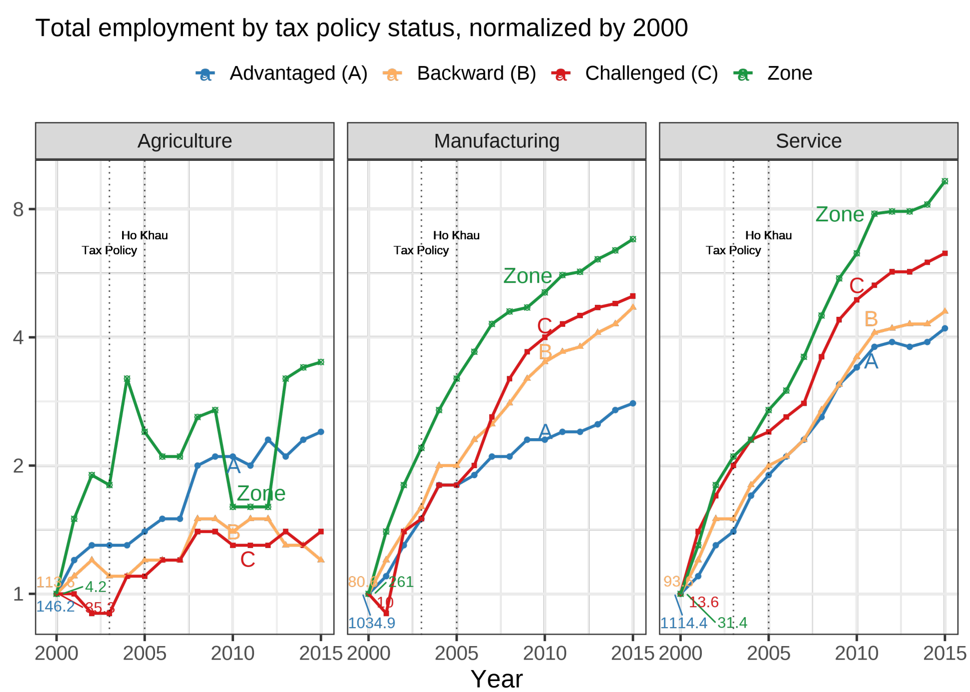


Figure A5: Employment by tax policy status including Zone

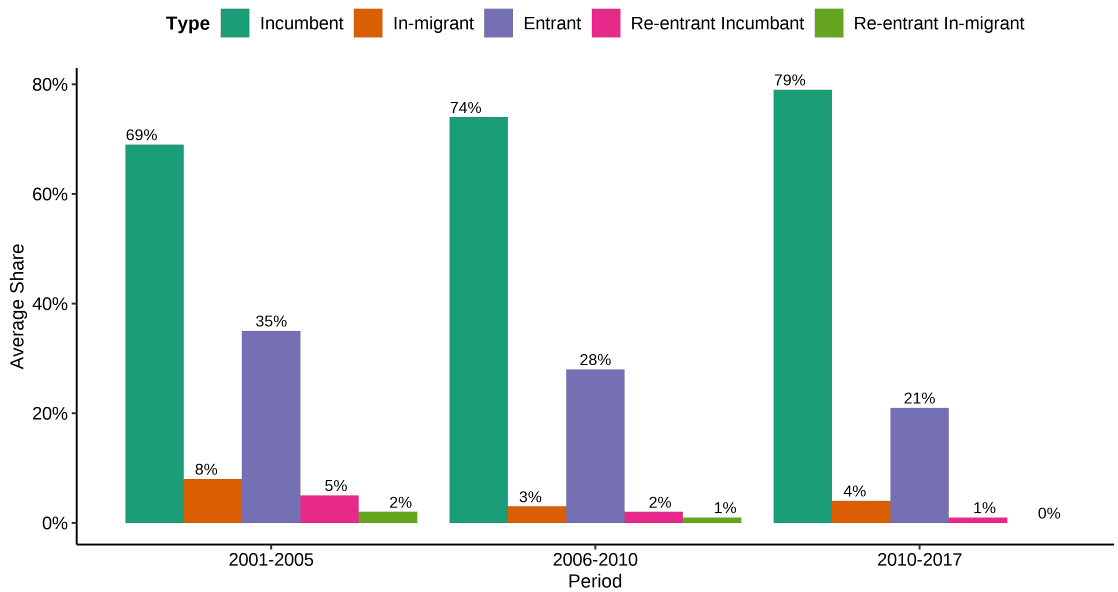


Figure A6: Mean shares of different types of entry

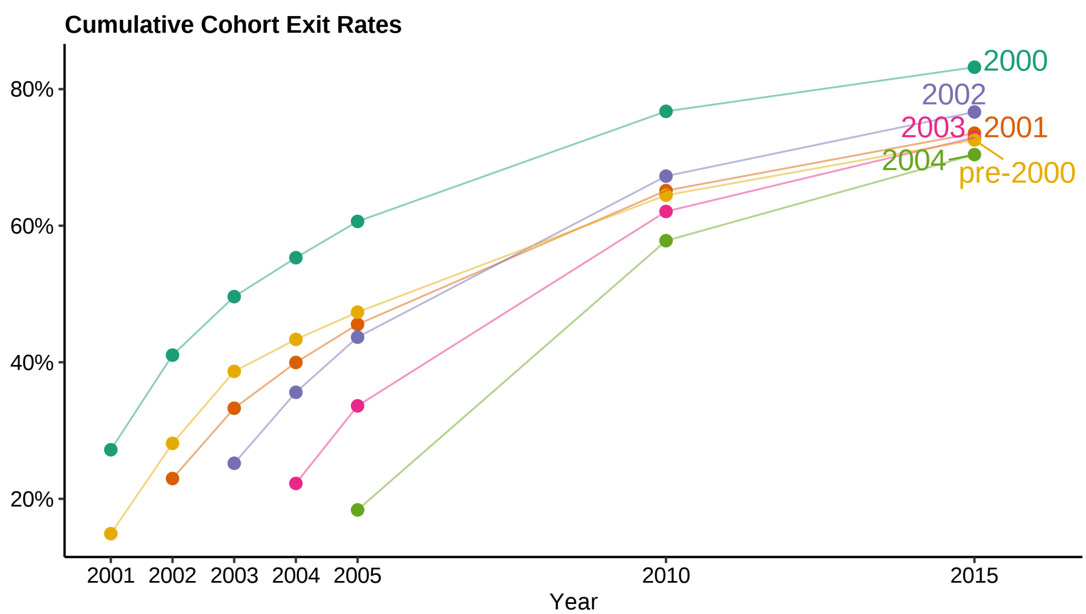
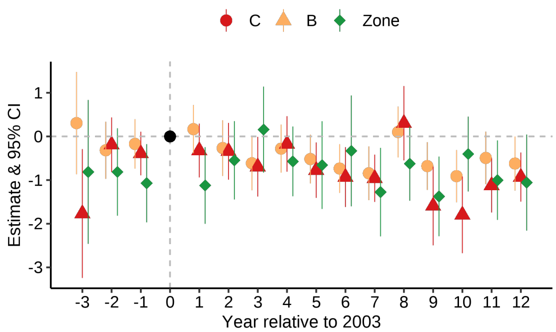
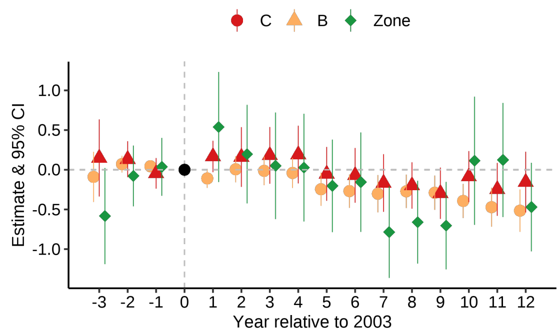


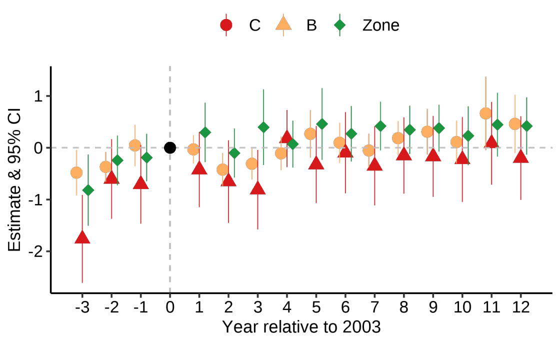
Figure A7: Exit rates by cohorts over time



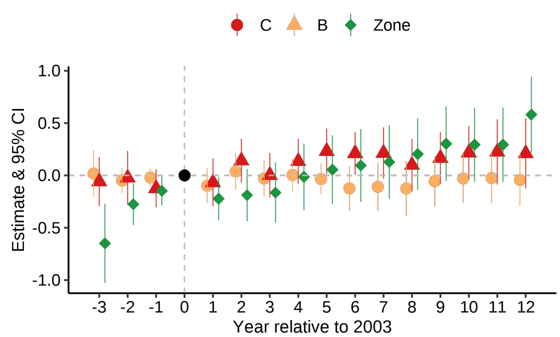
(a) Agriculture - Entry



(b) Agriculture - Labor



(c) Service - Entry



(d) Service - Labor

Figure A8: Effects of tax policy on entry share and labor in other sectors

Table A1: Shares of multi-plant firms in 2000

Shares of Multi-plant	Firms	Sales	Employment
All	0.011	0.078	0.077
SOE	0.068	0.136	0.119
Private	0.002	0.006	0.018
Foreign	0.006	0.006	0.013

*

Let $\bar{\delta}_t^{ij*}$ be the random variable that represents the maximum utility from choosing a market, that is, $\bar{\delta}_t^{ij*} \equiv \max_{n\in\mathcal{R},k\in\mathcal{S}} \left\{\bar{\delta}_t^{ij,nk} + \epsilon_t^{nk}\right\}$.

I want to derive the distribution of the maximum utility $\bar{\delta}_t^{ij*}$, denoted by $\bar{H}^{ij}(\bar{\delta})$.

$$\begin{aligned}\bar{H}^{ij}(\bar{\delta}) \equiv \Pr\left(\bar{\delta}_t^{ij*} \le \bar{\delta}\right) &= \Pr\left(\epsilon_t^{nk} \le \bar{\delta} - \bar{\delta}_t^{ij,nk}, \forall nk\right) \\ &= \exp\left\{-\sum_{n=1}^{N}\left[\sum_{k=1}^{S}\exp\left(-\frac{\bar{\delta}-\bar{\delta}_t^{ij,nk}}{\rho\nu} - \frac{\bar{\gamma}}{\rho}\right)\right]^{\rho}\right\} \\ &= \exp\left\{-\exp\left(-\frac{\bar{\delta}}{\nu} - \bar{\gamma}\right)\sum_{n=1}^{N}\left[\sum_{k=1}^{S}\exp\left(\frac{\bar{\delta}_t^{ij,nk}}{\rho\nu}\right)\right]^{\rho}\right\} \\ &= \exp\left\{-\exp\left(-\frac{\bar{\delta}}{\nu} - \bar{\gamma}\right)\mathcal{U}_t\right\},\end{aligned}$$

where the second equation follows from the joint CDF of ϵ_t and

$$\mathcal{U}_t \equiv \sum_{n=1}^{N}\left[\sum_{k=1}^{S}\exp\left(\frac{\bar{\delta}_t^{ij,nk}}{\rho\nu}\right)\right]^{\rho}.$$

Thus, the density function of $\bar{\delta}_t^{ij*}$ is given by

$$\bar{h}^{ij}(\bar{\delta}) = \bar{H}^{ij\prime}(\bar{\delta}) = \exp\left\{-\exp\left(-\frac{\bar{\delta}}{\nu} - \bar{\gamma}\right)\mathcal{U}_t\right\}\frac{\mathcal{U}_t}{\nu}\exp\left(-\frac{\bar{\delta}}{\nu} - \bar{\gamma}\right)$$

Shares of	Firms	Employment	Revenue
All			
Entrants	0.98	0.85	0.83
Exiters	0.74	0.47	0.49
A			
Entrants	0.96	0.88	0.90
Exiters	0.76	0.49	0.56
B			
Entrants	0.98	0.88	0.97
Exiters	0.82	0.62	0.74
C			
Entrants	0.98	0.85	0.84
Exiters	0.74	0.48	0.50

* *Notes:* This table summarizes establishment turnover between 2000 and 2015 in Vietnam. An entrant is defined as an establishment that is present in year t but not in year $t-1$, while an exiter is present in year $t-1$ but not in year t. For example, the first cell indicates that 98% of establishments present in 2015 did not exist in 2000. The first row of the second column shows that these entrants accounted for 85% of total employment in 2015. The second row of the second column reports that 74% of establishments in 2000 were no longer operating in 2015. This group of exiters accounted for 47% of employment in 2000.

Table A2: Entry and exit of firms in 2000 and 2015

The expected maximum value is therefore given by

$$\begin{aligned}
\Xi_t^{ij} &= \int_{-\infty}^{+\infty} \bar{\delta}_t^{ij*} h\left(\bar{\delta}_t^{ij*}\right) d\bar{\delta}_t^{ij*} \\
&= \int_{-\infty}^{+\infty} \bar{\delta}_t^{ij*} \exp\left\{-\exp\left(-\frac{\bar{\delta}_t^{ij*}}{\nu} - \bar{\gamma}\right) \mathcal{U}_t\right\} \frac{\mathcal{U}_t}{\nu} \exp\left(-\frac{\bar{\delta}_t^{ij*}}{\nu} - \bar{\gamma}\right) d\bar{\delta}_t^{i*}.
\end{aligned}$$

Applying the following change in variable: $\bar{z} = \exp\left(-\frac{\bar{\delta}_t^{ij*}}{\nu} - \bar{\gamma}\right)$, such that $\bar{\delta}_t^{ij*} =$

$-\nu\left(\ln(\bar{z})+\bar{\gamma}\right)$, and $d\bar{\delta}_t^{ij*} = -\nu(d\bar{z}/\bar{z})$. Then,

$$\begin{aligned}
\Xi_t^{ij} &= \int_{+\infty}^{0} -\nu(\ln(\bar{z})+\bar{\gamma})\exp\{-\bar{z}\mathcal{U}_t\}\frac{\mathcal{U}_t}{\nu}\bar{z}\left(-\nu\frac{d\bar{z}}{\bar{z}}\right) \\
&= -\nu\mathcal{U}_t\int_0^{+\infty}\ln(\bar{z})\exp\{-\bar{z}\mathcal{U}_t\}d\bar{z} - \nu\bar{\gamma}\mathcal{U}_t\int_0^{+\infty}\exp\{-\bar{z}\mathcal{U}_t\}d\bar{z}
\end{aligned}$$

And using Laplace transformation where $\int_0^{+\infty}\ln(\bar{z})\exp\{-\bar{z}\mathcal{U}_t\}d\bar{z} = -\frac{\ln(\mathcal{U}_t)+\bar{\gamma}}{\mathcal{U}_t}$

$$\begin{aligned}
\Xi_t^{ij} &= \nu\mathcal{U}_t\left(\frac{\ln(\mathcal{U}_t)+\bar{\gamma}}{\mathcal{U}_t}\right) - \nu\bar{\gamma} \\
&= \nu\ln(\mathcal{U}_t),
\end{aligned}$$

which is similar to (4).

The choice probability $\mu_t^{ij,nk}$ follows from Williams-Daly-Zachary (WDZ) theorem by differentiating Ξ_t^{ij} w.r.t $\bar{\delta}_t^{nk}$, that is,

$$\begin{aligned}
\mu_t^{ij,nk} &= \nu\frac{1}{\mathcal{U}_t}\frac{\partial\mathcal{U}_t}{\partial\bar{\delta}_t^{nk}} \\
&= \frac{e^{(\beta U_{t+1}^{nk}-m_t^{ij,nk})/\rho\nu}\left(\sum_{s\in\mathcal{S}} e^{\left(\beta U_{t+1}^{ns}-m_t^{ij,ns}\right)/\rho\nu}\right)^{\rho-1}}{\sum_{c\in\mathcal{R}}\left(\sum_{s\in\mathcal{S}} e^{\left(\beta U_{t+1}^{cs}-m_t^{ij,cs}\right)/\rho\nu}\right)^{\rho}},
\end{aligned}$$

which is (6).

Derivations of μ_t^{nk} and $\mu_t^{i,n}$

$$\begin{aligned}
\mu_t^{i,nk} &= \frac{e^{(\beta U_{t+1}^{nk}-m_t^{i,nk})/\rho\nu}\left(\sum_{s=1}^{J} e^{\left(\beta U_{t+1}^{ns}-m_t^{i,ns}\right)/\rho\nu}\right)^{\rho-1}}{\sum_{c=1}^{N}\left(\sum_{s=1}^{J} e^{\left(\beta U_{t+1}^{cs}-m_t^{i,cs}\right)/\rho\nu}\right)^{\rho}} \\
&= \frac{e^{(\beta U_{t+1}^{nk}-m_t^{i,nk})/\rho\nu}}{\left(\sum_{s=1}^{J} e^{\left(\beta U_{t+1}^{ns}-m_t^{i,ns}\right)/\rho\nu}\right)}\frac{\left(\sum_{s=1}^{J} e^{\left(\beta U_{t+1}^{ns}-m_t^{i,ns}\right)/\rho\nu}\right)^{\rho}}{\sum_{c=1}^{N}\left(\sum_{s=1}^{J} e^{\left(\beta U_{t+1}^{cs}-m_t^{i,cs}\right)/\rho\nu}\right)^{\rho}}
\end{aligned}$$

The second part of the product in the second equation is $\mu_t^{i,n}$ as in (32). To simplify the first part, I can write $\beta U_{t+1}^{nk} - m_t^{i,nk}$ as a sum of two components based on (5) and (28):

$$\Omega_t^{i,n} \equiv \beta\ln\left(A_{t+1}^{n}/P_{t+1}^{n}\right) + \Xi_{t+1}^{n} - m_t^{i,n},$$

and

$$\Omega_t^{ij} \equiv \beta \ln(w_{t+1}^{nk}) - m_t^k.$$

Substituting these expressions into $\mu_t^{i,nk}$ yields

$$\mu_t^{i,nk} = \frac{e^{\Omega_t^{i,n}/\rho\nu} e^{\Omega_t^{nk}/\rho\nu}}{e^{\Omega_t^{i,n}/\rho\nu} \left(\sum_{s=1}^J e^{\Omega_t^{ns}/\rho\nu}\right)} \mu_t^{i,n} = \frac{e^{\Omega_{t+1}^{nk}/\rho\nu}}{\sum_{s=1}^J e^{\Omega_{t+1}^{ns}/\rho\nu}} \mu_t^{i,n}$$

where the last ratio is μ_t^{nk} as in (31).

B.2 Continuous age entrepreneurs

This subsection extends the entrepreneur age from Young and Old to $a \in \{1, \ldots, A\}$. In particular, we abstract from the exogenous exit of Old entrepreneurs and assume that each a-year-old entrepreneur faces similar choices as a Young entrepreneur in Section 4.3. Thus, the value functions of $a-$year-old entrepreneurs are given by

$$V_{a,t}^{ij} = \ln\left((1 - \tau_{a,t}^i)\frac{\pi_t^{ij}}{P_t^i}\right) + \chi \ln\left[\exp(V_{a+1,t+1}^{ij})^{\frac{\beta}{\chi}} + 1\right] \tag{B.1}$$

$$\varsigma_{a,t}^{ij} = \frac{\exp\left(V_{a+1,t+1}^{ij}\right)^{\frac{\beta}{\chi}}}{\exp\left(V_{a+1,t+1}^{ij}\right)^{\frac{\beta}{\chi}} + 1} \tag{B.2}$$

Again, ψ_t^{ij} denotes fraction of informal entrepreneurs that choose to locate in i among all informal entrepreneurs in sector j between t and $t+1$

$$\psi_t^{ij} = \frac{\exp\left(\beta V_{1,t+1}^{ij} - f_t^{ij}\right)^{1/\chi}}{\sum_{n=1}^N \exp\left(\beta V_{1,t+1}^{nj} - f_t^{nj}\right)^{1/\chi}} \tag{B.3}$$

From here, I follow the same steps as in Section 5.2 by first taking ln of the entry equation ψ_t^{ij}

$$\ln \psi_t^{ij} = -\frac{1}{\chi} f_t^{ij} + \frac{\beta}{\chi} V_{1,t+1}^{ij} - \ln \sum_{n=1}^N \exp\left(V_{1,t+1}^{nj} - f_t^{nj}\right)^{1/\chi}.$$

Next, I substitute the expected value of 1-year-old entrepreneurs $V^{ij}_{1,t+1}$ to get

$$\ln \psi^{ij}_t = \frac{\beta}{\chi} \ln(1-\tau^i_{1,t+1}) \frac{\pi^{ij}_{t+1}}{P^i_{t+1}} + \beta \ln \left[\exp \left(V^{ij}_{2,t+2} \right)^{\frac{\beta}{\chi}} + 1 \right] - \frac{1}{\chi} f^{ij}_t - \ln \sum_{n=1}^{N} \exp \left(\beta V^{nj}_{1,t+1} - f^{nj}_t \right)^{1/\chi}. \tag{B.4}$$

Rearranging terms of and taking the ln of (B.2) yield

$$\ln \left(\exp \left(V^{ij}_{2,t+2} \right)^{\frac{\beta}{\chi}} + 1 \right) = \ln(1-\varsigma^{ij}_{1,t+1})^{-1} \tag{B.5}$$

Substituting this expression into (B.4) yields

$$\ln \left(\psi^{ij}_t (1-\varsigma^{ij}_{1,t+1})^{\beta} \right) = \frac{\beta}{\chi} \ln(1-\tau^i_{1,t+1}) \frac{\pi^{ij}_{t+1}}{P^i_{t+1}} - \frac{1}{\chi} f^{ij}_t - \ln \sum_{n=1}^{N} \exp \left(\beta V^{nj}_{1,t+1} - f^{nj}_t \right)^{1/\chi} \tag{B.6}$$

To further exploit the variation across age groups, consider the ratio between the fraction of a-year-old entrepreneurs that stay and the fraction of them that exit for age $a > 1$ based on (B.2)

$$\frac{\varsigma^{ij}_{a-1,t}}{1-\varsigma^{ij}_{a-1,t}} = \exp \left(V^{ij}_{a,t+1} \right)^{\frac{\beta}{\chi}}.$$

Taking ln both sides and substituting the value (B.1) yields

$$\ln \frac{\varsigma^{ij}_{a-1,t}}{1-\varsigma^{ij}_{a-1,t}} = \frac{\beta}{\chi} \left(\ln \left[(1-\tau^i_{a,t+1}) \pi^{ij}_{t+1} / P^i_{t+1} \right] \right) + \beta \ln \left(\exp(V^{ij}_{a+1,t+2})^{\beta/\chi} + 1 \right).$$

Applying (B.5) yields

$$\ln \frac{\varsigma^{ij}_{a-1,t}}{1-\varsigma^{ij}_{a-1,t}} (1-\varsigma^{ij}_{a,t+1})^{\beta} = \frac{\beta}{\chi} \ln \left((1-\tau^i_{a,t+1}) \frac{\pi^{ij}_{t+1}}{P^i_{t+1}} \right).$$

If an old firm's profit tax increases, the young firms are more likely to stay than exit, which is captured by the relative tendency to stay. Still, the young ones are also likely to exit in the next period also increases due to an increase in market competition for marginal firms.

Finally, by subtracting each side of this equation from the corresponding side of the entry equation (B.6), I obtain

$$\ln \frac{\psi_t^{ij}(1-\varsigma_{1,t+1}^{ij})^{\beta}}{\frac{\varsigma_{a-1,t}^{ij}}{1-\varsigma_{a-1,t}^{ij}}(1-\varsigma_{a,t+1}^{ij})^{\beta}} = \frac{\beta}{\chi}\ln\frac{(1-\tau_{1,t+1}^{i})}{(1-\tau_{a,t+1}^{i})} - \frac{1}{\chi}f_t^{ij} - \ln\sum_{n=1}^{N}\exp\left(\beta V_{1,t+1}^{nj} - f_t^{ij}\right)^{1/\chi} \quad \text{(B.7)}$$

which is (36).