Taxing Production Networks: Trade, Misallocation, and Vertical Integration

Cihang Wang*

University of Illinois†

December 9, 2025

[Click Here for the Latest Version].

Abstract

Many countries tax business revenue, effectively taxing every intermediate transaction in the supply chain. In production networks, such cascading taxes act like within-country tariffs on intermediate goods, inflating input costs, suppressing cross-sector trade, and pushing firms toward inefficient vertical integration. This could potentially lead to large general-equilibrium consequences for entry, wages, and welfare—yet the magnitudes of these measures haven been unclear. Exogenous shifts in tax structure are rare, spillovers propagate across sectors and regions, vertical integration is poorly observed, and welfare effects require a structural counterfactual. To tackle this problem, I exploit China's staggered Business-to-Value Added Tax("B2V") reform, which replaced the business tax with value added tax for service sectors. Combining enterprise income tax records, registry-based entry data, and interregional input–output tables, I estimate event-study difference-in-differences and Poisson pseudo-maximum-likelihood gravity equations, then build a spatial general-equilibrium model with endogenous entry and an explicit make-or-buy margin. I identify an overall trade elasticity of 4.47 using a cross-product gravity estimator that nets out unobservables, solve the model in exact hat algebra, and discipline in-house shares by matching observed entry responses. I find that entry in treated service sectors increases by 21 percent; revenues of smaller and newer firms increase by 22 percent; and service-involving trade increases by 35.6 percent overall (47.9 percent within provinces; 20.7 percent across provinces). Accounting for the endogenous choice of vertical integration has quantitatively important welfare implication: eliminating cascading taxes raises welfare by 4.3 percent and real wages by 5.3 percent nationwide. Importantly, if this choice is ignored, about 6% of the welfare improvement will be missed; in a revenue-neutral counterfactual, 94 percent of the welfare gain remains. In production networks, tax structure beats tax level: removing domestic "tariffs" inside the supply chain delivers most of the gains.

Keywords: Macroeconomic Policy; Misallocation; Firm Growth; Trade; Vertical Integration; Spatial General Equilibrium.

JEL Classification: C23, D22, E60, R12, L60, O25.

*Department of Economics, University of Illinois at Urbana-Champaign. 214 David Kinley Hall, 1407 W. Gregory Dr, Urbana, IL 61801. Email: **cwang153@illinois.edu**. Website: **sites.google.com/view/cihangwang**

†I am grateful to Greg Howard, Dan Bernhardt, Shihan Xie, Cynthia Jing Wu, Flavio Rodrigues, Humberto Martinez-Garcia, Abby Stocker, Vinicios Sant'Anna, Zebang Xu, Lourenço Paz, Jian Zou, Marcos Sora, Jingnan Liu, Zi Wang and all participants at UIUC graduate seminars, macro lunch, Midwest Macroeconomics Meeting and Southern Economics Association Annual Meeting. All errors are mine.

1 Introduction

The past two centuries have witnessed the increasing liberalization in trade of goods and services across locations (Donaldson, 2015). Nonetheless, significant barriers and frictions to trade across locations and sectors remain between and within countries. The literature has documented the impacts of trade barriers, including tariff and non-tariff barriers. This paper applies insights from this literature to within-country markets, examining how a business tax on intermediate inputs distorts sector–region transactions. The business tax functions as a within-country tariff, where firms pay taxes on revenue (therefore intermediate inputs). Not only does this tax hinder trade between sectors and regions by raising input costs, it incentivizes firms to engage in excessive (inefficient) vertical integration in order to evade taxes on intermediate goods. In turn, this further reduces cross-sector trade, particularly depressing demand from upstream suppliers (Coase, 1937; Williamson, 1971). Despite the distortions and inefficiencies created by this tax system, it is still used or has been used by most developing and developed[1] countries (Best et al., 2015; Phillips and Ibaid, 2019; Hansen et al., 2022a).

This paper studies how taxes on intermediate inputs shapes the production networks and contributes to the literature by being the first paper (to my knowledge) to integrate the endogenous choice of vertical integration into a quantitative spatial general equilibrium model. Embedding the intermediate input tax into a production network captures the distortions of the business tax and incorporating the endogenous choice of vertical integration matters for quantitative exercise. To see why this modeling choice is essential, it is useful to see how the business tax operates and why its cascading nature distorts production networks. When taxes are levied on revenue, they cascade through production networks: every intermediate sale embeds taxes paid at earlier stages. As a result, market-sourced goods appear artificially expensive, nudging firms toward the inefficient vertical integration. This mechanism resembles a tariff but is more pervasive because it operates within the domestic supply chain, compounding across sectors and regions more frequently. In complex production networks, cascading taxes distort trade flows and firm organi-

[1]For example, in the United States in fiscal year 2017, \$157 billion of taxes were imposed on business-to-business sales of products, services and equipment, which comprised 42% of total state and local sales tax revenue (Phillips and Ibaid, 2019). Examples from developing countries include Brazil's complex turnover tax system, India's pre-GST regime, Argentina's provincial gross receipts taxes, Colombia's former VAT structure, the Philippines' percentage tax system, Russia's pre-reform tax policies, and Vietnam's Special Consumption Tax. These cases demonstrate that the challenge of efficiently taxing intermediate transactions is a common concern in diverse economies worldwide.

zation: a firm facing intermediates with multiple layers of embedded tax may resort to cheaper but lower-quality inputs or integrate production in-house to avoid the tax altogether. Neither choice is efficient: the first sacrifices access to better-priced inputs, while the second diverts firms into activities where they could lack a comparative advantage.

In response to concerns over excessive and uneven tax burdens under China's business tax system, the central government launched a major reform that effectively restructured the production network. The stated goal was to eliminate cascading taxes that accumulated along supply chains. Before the reform, service and manufacturing firms were taxed under different regimes: service firms paid a business tax on their total revenue—including intermediate inputs—while manufacturing firms paid a value-added tax only on their value added. This asymmetry meant that any transaction involving a service firm, either as a buyer or a seller, was taxed multiple times. As a result, the production network became fragmented: links passing through service sectors faced higher effective costs than links between manufacturing firms. The reform replaced the business tax with a value-added tax for services, thereby removing taxes on intermediate transactions, reconnecting previously distorted transactions and reshaping the entire production network.

The reform was initially implemented for nine pilot provinces in 2012, covering seven service sectors. A year later, it was expanded to the entire nation for those seven sectors. Over the following three years, the remaining service sectors were gradually reformed nationwide, with the reform completed in 2016. By leveraging the staggered timeline of the reform implementation, I provide empirical evidence that service firms in the reformed provinces experience faster growth relative to both manufacturing firms and service firms in provinces yet to implement the reform. This differential growth is primarily driven by small firms and new entrants, supported by a notable increase of approximately 21% in firm entry within affected service sectors. Service firms also experience faster revenue growth of 22% compared to their manufacturing counterparts. At the regional-sector level, gravity-style regressions using inter-region-sector input-output tables before and after the reform demonstrate that trade involving service sectors—both purchases from and sales to service firms—increases by 35.6% relative to trade among manufacturing firms.

Despite these positive outcomes, reduced-form analysis alone is insufficient to evaluate the welfare implications of the reform. The policy's primary objective – lowering the tax burden on service firms – also led to unintended consequences. Local governments experiences tighter budget constraints due to declining tax revenues, potentially affecting public service provision. Additionally, sectoral inter-linkages imply that shocks in one sector can propagate throughout the economy, complicating efforts to isolate the reform's effects. A further challenge is that one of the key pre-reform distortions, excessive vertical integration, is difficult to assess with aggregate data due to limited firm-level coverage.

To quantify broader general equilibrium effects—including firms' vertical integration decisions, welfare consequences, and distributional impacts—I develop a spatial general equilibrium model. The model features endogenous firm entry, sectoral trade linkages, and the vertical integration decision for manufacturing firms, all embedded within a multi-region production network. Through this quantitative analysis, I demonstrate that removing the business tax, thus eliminating cross-sector-region distortions, enhances aggregate welfare by reducing the effective costs associated with purchasing intermediate inputs and inefficient vertical integration. The consequent improvement in allocative efficiency promotes the formation of more productive input-output linkages both within and across regions. Importantly, the reform generates uneven gains, disproportionately benefiting regions and sectors more reliant on services or previously experiencing higher tax-induced distortions, resulting in significant redistributional effects across space.

This paper contributes to the literature on trade, taxation, and firm behavior by being the first study to incorporate the endogenous choice of vertical integration (Bernard et al., 2019) into a quantitative trade model. The model is a multi-sector-region version of the Melitz model (Melitz, 2003) with the choice of vertical integration (Bernard et al., 2019). The model incorporates interregional trade at the sector-region-firm pair level, capturing both firm-to-firm trade in intermediate goods and firm-to-consumer trade for final consumption goods. Manufacturing firms in the model have the option to either produce a service in-house or outsource these inputs from the market, depending on their specific cost considerations and needs. All in-house production is performed exclusively by labor directly employed by the firm. Firms' decision is modeled using a constant elasticity of substitution (CES) production function with two layers, allowing for endogenous determination of trade shares between different intermediate inputs as well as between labor and intermediate inputs. I model firm entry as an endogenous process and assume firms operate under monopolistic competition with heterogeneous productivity following Krugman (1980) and Melitz (2003). This approach enables me to capture the interactions between vertical integration decisions, firm entry, and input-output linkages—a multidimensional relationship in firm boundary, trade and taxation policy. Specifically, I find that economies with higher initial levels of vertical integration experience disproportionately larger gains in firm entry and output growth following the reform. This occurs through an extensive margin where substitution from in-house to market production creates new business opportunities for specialized service providers, which in turn improves the efficiency of the production network and further reduces the incentive for vertical integration.

The key parameter governing the welfare change is the trade elasticity, which I estimate using the input-output table before and after the reform. Following Caliendo and Parro (2015) and adapting their methods by adding a time dimension, I exploit variation in cross-region-sector tax rate before and after the reform to identify the elasticity while avoiding the need to estimate unobservable parameters such as bilateral trade costs or productivity distributions that vary across regions, sectors, and time. The model is solved using exact hat algebra, which expresses all variables in relative changes from the base equilibrium, allowing me to hold unobservable structural parameters constant while isolating the effects of the tax policy change.

To calibrate the model, I must address the challenge that the baseline vertical integration levels are unobservable in standard input-output data. I assume that expenditures on in-house service production constitute a fraction of the value-added component reported in input-output tables, which provides an observable upper bound for this unobserved margin. In the baseline calibration, I allow manufacturing firms to engage in vertical integration while restricting service firms to specialize in their core activities, consistent with empirical evidence on firm boundaries across sectors (Antràs, 2003). The model exhibits a systematic relationship between initial integration levels and reform outcomes: higher baseline vertical integration is associated with larger predicted gains in output, real wages, and firm entry following the tax reform. This occurs because firms with greater initial reliance on in-house production have more scope to reallocate resources toward market transactions when distortions are removed, generating more efficiency gains through improved specialization and deeper production networks.

However, the interpretation of these results requires careful consideration of what different integration levels reveal about underlying economic fundamentals. A higher initial in-house share suggests that in-house production is relatively cost-effective compared to market alternatives in the baseline economy. This typically indicates distortionary tax policies that penalize market transactions. Conversely, lower initial integration indicates a more productive service sector capable of providing intermediate inputs at a competitive prices despite existing tax distortions, representing a more efficient baseline equilibrium. To identify the relevant level of baseline vertical integration, I employ a moment-matching strategy, calibrating the average in-house production share at the industry-region level to replicate the firm entry patterns observed in the reduced-form empirical analysis. This approach ensures consistency between the structural model predictions and causal evidence from the reform.

Having established the empirical relevance of the baseline vertical integration calibration and outlined the qualitative mechanisms through which the reform operates, I next calculate the quantitative magnitudes of these effects. The quantitative analysis reveals that the reform generates substantial economic gains: an aggregate welfare increase of 4.3% and an average rise of 5.3% in real wages nationwide. These magnitudes reflect the combined impact of the multiple channels identified earlier: the reallocation from in-house to market production, enhanced firm entry in service sectors, and strengthened input-output linkages across sectors and regions. The distribution of welfare gains exhibit significant regional heterogeneity. Provinces with large service sectors, such as Beijing and Shanghai, experience the most substantial benefits, with real wage growth reaching the highest levels nationwide, primarily driven by the disproportionately larger expansion of their service sectors. However, these same provinces also bear the greatest impact of reduced tax revenues due to their initial reliance on the business tax. The counterfactual analysis predicts that the aggregate tax revenue decreases by 6.6% following the reform.

This welfare gain comes from two channels: lowering the tax burden for firms and reducing the cross-region-sector trade frictions. To disentangle the importance of the two mechanisms, I conduct a tax revenue-neutral counterfactual analysis. This exercise eliminates cross-sector trade frictions while proportionately raising tax rates across all sectors and regions to maintain the aggregate baseline tax revenue level. Even under this revenue-neutral approach, the model predicts an aggregate welfare increase of 4.06% and a real wage rise of 4.25% — representing 94% of the total welfare gains and 81% of the real wage increase from the actual reform. The decomposition demonstrates that the efficiency gains from eliminating the trade distortions constitute the primary driver of the reform's benefits, far exceeding the direct effects of tax burden reduction.

This study has policy implications for the role of tax system in shaping the economic outcomes, particularly in the context of developing economies. The Chinese experience of the "Business to Value-added Tax" reform shed light on the potential for changes in a tax scheme to remove cross-region-sector distortions, improve firm efficiency and stimulate economic growth. As many developing countries grapple with the challenges of designing tax systems that balance the generation of tax revenue and economic efficiency, this lesson from China can provide valuable insights: how we tax the economy or the production network is very important. Moreover, the quantitative framework developed in this paper, which incorporates firms' strategic decisions on vertical integration, offers a perspective for policymakers and researchers to quantify the impact of vertical integration and/or outsourcing behaviors of firms, which is an important margin in analyzing relevant research questions.

The remainder of the paper is organized as follows. Section 2 discusses the relevant literature. Section 3 and 4 documents the policy background and empirical evidence. Through section 5 to 7, I outline the quantitative trade model and its implications. Section 8 concludes.

2 Relation to the Literature

2.1 Misallocation and Production Networks

This paper contributes to the broader literature of misallocation and how production networks transmit the shocks or the frictions. The literature on misallocation has significantly expanded our understanding of productivity differences across countries and regions. Seminal work by Hsieh and Klenow (2009) demonstrated that resource misallocation could account for substantial total factor productivity (TFP) differences between China, India, and the United States. Building on this, Restuccia and Rogerson (2008) showed how policy distortions leading to misallocation across heterogeneous plants can substantially reduce aggregate productivity. The scope of misallocation research has since broadened to encompass various sources of inefficiency. For instance, Gopinath et al. (2017) examined how capital misallocation contributed to low productivity growth in Southern European countries, while Bau and Matray (2023) demonstrated significant productivity gains from improved capital allocation in India following financial liberalization. The closet work related to my paper in terms of the policy setting is Xing et al. (2024), where they use the same policy variation and a panel data of listed firms to study the impact on the growth of Chinese listed firms in the manufacturing sector and in particular their research and development investment and they argue that firms outsourcing services are the main reason why service firms grow. Works including Hansen et al. (2022b) and Gadenne et al. (2019) provide different perspectives on business tax and value-added tax on firm-level activities and trade. Different from their focus, this paper focuses on the spatial input-output linkage and incorporates the vertical integration choice in the quantitative trade model with endogenous firm entry. Fajgelbaum et al. (2019) argues that the variation in state tax is a source of misallocation while my paper claims that the tax on business inputs creates a highly distortionary economy through the input-output linkages.

Following the seminal work of Gabaix (2011), a large body of literature studies the aggregate effects of "granular" productivity shocks (see (Foerster et al., 2011; Acemoglu et al., 2012; Carvalho and Gabaix, 2013; di Giovanni et al., 2014, 2018), among others). My paper is closely related to the transmission of distortions in economies with intermediate good trade. Previous work studies the effects of distortions within specific input-output structures; see Basu (1995), Ciccone (2002), Yi (2003), Jones (2011), Asker et al. (2014), Bigio and La'O (2020).

2.2 Trade and economic geography

The intersection of trade and economic geography has emerged as a fertile ground for research, particularly with the development of quantitative spatial models. Building on the foundational work of Eaton and Kortum (2002), recent studies have incorporated rich spatial structures to analyze the impact of trade on economic outcomes. Caliendo et al. (2022a) developed a framework modeling the world economy as an input-output table, forming an endogenous economy with trade frictions and allowing for the evaluation of the welfare implications. In particular, I incorporate the possibility of in-house production as Bernard et al. (2019) does to incorporate the vertical integration into the quantitative trade model. Bernard et al. (2019) includes the choice of in-house production in a simple model to guide their empirical findings, while I incorporate the choice of vertical integration to derive quantitative implications. The importance of market access and infrastructure in shaping trade patterns and economic growth has been emphasized in several influential studies. Donaldson (2015) and Donaldson and Hornbeck (2016) demonstrated the significant impact of railroads on economic growth through improved market access. In the context of developing economies, Faber

(2014) showed how infrastructure investments in China's National Trunk Highway System affected industrial development across regions. These studies highlight the critical role of reducing trade frictions in promoting economic integration and growth. Recent literature has also focused on the interplay between trade liberalization, internal migration, and productivity. Tombe and Zhu (2019) quantified the substantial contribution of reduced internal trade and migration costs to China's recent economic growth. Their findings underscore the importance of domestic market integration in driving aggregate productivity gains. Furthermore, the work of Fan (2019) on the distributional impacts of trade liberalization has shed light on the complex interactions between internal geography, labor mobility, and trade openness. The evolving literature on trade and economic geography, as summarized by Redding and Rossi-Hansberg (2017), has increasingly employed quantitative models to analyze the spatial implications of economic policies. These approaches, combined with rich data and natural experiments, have enhanced our understanding of how trade frictions and geographic factors shape economic outcomes across space and time.

3 Policy Background

3.1 Dual-tax System in China

The manufacturing sector in China has long been the traditional economic driver. However, in recent years, the service sector has emerged as an increasingly critical component of the economy, playing a pivotal role in innovation-driven growth and contributing significantly to GDP and tax revenue (Zilibotti, 2017). Despite its growing importance, the service sector has historically been subject to a different and more burdensome tax treatment compared to manufacturing firms, facing heavy taxation on intermediate inputs. China's dual tax system created significant economic distortions. Under this system, firms in the service sector were required to pay business taxes on their total revenue — including intermediate inputs — while manufacturing firms operated under the value-added tax (VAT) system that only taxed value added.

The distortions operated through multiple channels. First, service firms faced heavy taxation because business taxes applied to their gross revenue rather than just value added. The revenue at the same time included the after-tax intermediate input. This is just one layer inside the production network showing the "taxes on taxes". Second, when manufacturing firms purchased in-

Table 1. Timing and Industries of B2V Reform

Reformed industries	Regions	Implementation date
Transportation,	Shanghai	2012.01.01
R&D and technical services, IT services,	Beijing	2012.09.01
cultural and innovation services,	Jiangsu	2012.10.01
logistics auxiliary services,	Anhui	2012.10.01
attestation and consulting services,	Fujian	2012.11.01
and tangible assets leasing services	Guangdong	2012.11.01
	Hubei	2012.12.01
	Tianjin	2012.12.01
	Zhejiang	2012.12.01
	Nationwide	2013.08.01
Postal service, rail transportation	Nationwide	2014.01.01
Telecommunication	Nationwide	2014.06.01
Real estate, construction, finance, and other services	Nationwide	2016.05.01

Notes: This table outlines the waves of the B2V reform across different industries, regions and time.

termediate inputs from service providers, there are forced to pass on the multiple layers of taxation incorporated in the prices of the service goods. The same logic applied to service-to-service transactions as well. This tax cascading made cross-sector trade artificially expensive and incentivized manufacturing firms to engage in excessive vertical integration.

This tax system discouraged the development of efficient production networks by making it costly to source inputs from the most productive suppliers across sectors. It particularly hampered the service firm's growth by increasing operational costs and reducing demand from downstream clients. The resulting misallocation of resources reduced overall economic efficiency. Recognizing these problems, policymakers sought to alleviate the tax burdens on service firms and harmonize the tax structure.

3.2 Business-to-Value Added ("B2V") Tax Reform

To promote the growth of service firms, the government started to implement the "Business to Value-added Tax" reform by transitioning the service sector to a VAT system. Starting in 2012, this reform was carried out progressively across sectors and provinces, and the timeline is displayed in Table 1. It replaced the business tax with a value-added tax for service sector firms (see

Table 2), and at the same time lifting taxes on any intermediate input purchase for all sectors. By extending the VAT system to service firms, the reform enabled these firms to pay taxes on intermediate inputs, thus being taxed only on the value added.

To see how the reform impacted firms in the service sector, I take the Railway Transportation sector in Table 2 for instance. Before the "B2V" reform, Railway Transportation sector would pay 3% on the total revenue of its output (including both intermediate inputs and value added). After the reform, it pays 11% only on the total value added out of its output. The numeric tax rate before and after the reform is not necessarily the same for all sectors. In fact, as seen in Table 2, the tax rate increases for all reformed sectors. Therefore, merely seeing from the perspective of tax payment, whether a sector is subject to lower tax burden also depends on its value added margin. [2]

Table 2. Change of Tax Scheme and Rate

Industry Name	Industry Code	BT Rate	VAT Rate
Railway transportation	G53	3%	11%
Road transportation	G54	3%	11%
Water transportation	G55	3%	11%
Air transportation	G56	3%	11%
Portage and transportation agency	G58	3%	6%
Warehousing	G59	5%	6%
Telecomms, broadcast TV and satellite transmission services	I63	5%	6%
Internet services	I64	5%	6%
Software and information technology services	I65	5%	6%
Leasing	L71	5%	11% or 17%
Business services	L72	5%	6%
Research and experimental development	M73	5%	6%
Professional technical services	M74	5%	6%
News and publication	R85	5%	6%
Radio, television, film and recording production	R86	5%	6%
Culture and art	R87	5%	6%
Construction	E47	3%	11%
Real Estate	K70	5%	11%

Notes: This table illustrates the changes in tax schemes and rates for various service industries following the Business to Value-Added Tax (B2V) reform in China. The "Industry Code" column provides the official classification codes for each sector. The "BT Rate" column shows the business tax rate applied before the reform, while the "VAT Rate" column indicates the new value-added tax rate implemented after the reform. It's important to note that despite the generally higher nominal rates in the VAT column, the actual tax burden may not necessarily increase due to the fundamental differences between business tax and value-added tax systems. The business tax was levied on gross revenue, including intermediate inputs, whereas the VAT allows for the deduction of input costs, potentially resulting in a lower effective tax burden for many firms. The variation in tax rates across different service sectors reflects the government's tailored approach to the reform, considering the unique characteristics and economic importance of each industry. This shift from BT to VAT aimed to reduce tax cascading effects and promote the growth of the service sector by allowing input tax deductions.

Using total output and value-added data from 2012, I calculated the potential change in tax payments for service sector firms if they transitioned from business tax to value-added tax. Figure 1 illustrates that 85% of service sectors would experience a reduction or the same burden in tax payments under this scenario. While this suggests that the majority of service sectors could

[2]For instance, I talked to entrepreneurs in construction sectors in Shanghai and an accountant in Guangdong, they highlight the fact that right after the reform, some firms in the construction and transportation sectors experience an increase in the tax burden initially.

benefit from reduced tax burdens due to the reform, it's important to note that some sectors may experience an increase in their tax obligations.

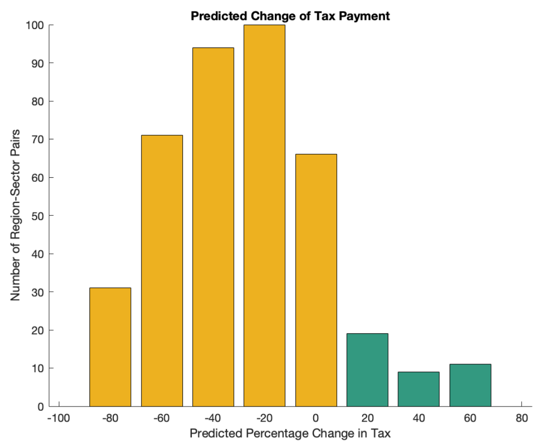


Figure 1. Distribution of Predicted Tax Payment Change

Notes: This figure illustrates the predicted changes in tax payments for service sectors following a hypothetical transition from business tax to value-added tax (VAT), based on 2012 total output and value-added data. The x-axis represents the magnitude of change in tax payments, ranging from -1 (100% reduction) to 0.8 (80% increase). The y-axis indicates the number of sectors experiencing each level of change. The distribution shows that approximately 85% of service sectors would experience a reduction in tax payments (represented by yellow bars), with the majority clustered around 20-60% reductions. A smaller proportion of sectors (represented by green bars) would see no change or an increase in tax burden, with increases generally below 40%. These results suggest that while the majority of service sectors could benefit from reduced tax burdens under the reform, the impact is not uniformly positive across all sectors. Some may face unchanged or increased tax obligations, highlighting the varied effects of the tax system transition on different parts of the service industry.

4 Empirical Evidence

In this section, I present the empirical evidence on how the "B2V" tax reform affects firm dynamics (i.e., firm entry and growth) and interregional trade. The empirical investigation motivates my development of the quantitative general equilibrium model in section 5. I unfold this section with the dataset I use, empirical strategy, and main findings.

4.1 Data

Firm Entry. To document how "B2V" tax reform affects firm entry, I obtain the annual total number of new entry of firms at the prefectural level spanning from 2008 to 2021. I obtain the number of firm entry at the province-by-sector-by-year level through aggregating the individual firm records of the administrative firm registry using the wepscraped data from Qichacha.[3]. The firm registry includes the universal firms that ever registered in China.

Firm Activities. To investigate the impact on firm activities following the "B2V" reform, I use the National Tax Survey Database of the Chinese State Administration of Tax (SAT) from 2008 to 2016. The database includes information on the firms' income, expenditure, and all tax-related outcomes. The sampling of this dataset ensures that large firms are included every year and that smaller firms are sampled as repeated cross-sectional data. The data surveys around 600,000 firms per year, making up 65% percentage of the total enterprise tax revenue.

Trade. To study the impact of the tax reform on interregional trade and cross sector-regional intermediate input use, I use the sector-by-region level trade value generated from the China inter-regional Input-Output Tables (IRIO). The IRIO is available every 5 years, I use data from 2012 and 2017 (Li et al., 2023) that cover the B2V reform period. The IRIO reports bilateral trade flows between 31 provinces and 41 sectors.

Table 3. Summary Statistics of Repeated Cross Sections and Balanced Panel

Year	Balanced Panel				Repeated Cross Sections			
	Avg Revenue	Avg Asset	Obs.	#Service Firms	Avg Revenue	Avg Asset	Obs.	#Service Firms
2010	947,805	4,336,031	8,286	2,402	145,804	424,514	746,772	155,387
2011	1,151,317	5,419,824	8,286	2,402	191,562	553,581	691,353	176,515
2012	1,149,617	6,119,088	8,286	2,402	210,576	635,929	689,395	191,414
2013	1,247,852	5,555,035	8,286	2,402	215,808	661,738	699,369	218,240
2014	1,369,586	7,813,689	8,286	2,402	197,115	697,199	716,765	236,810
2015	1,292,554	9,346,182	8,286	2,402	200,727	740,333	700,525	221,667
2016	1,208,935	10,919,697	8,286	2,402	159,509	766,051	608,799	205,182

Notes: This table presents summary statistics for the balanced panel and repeated cross-sections datasets used in the analysis. All monetary values are reported in thousands of Chinese Yuan (CNY). The balanced panel consists of 8,286 large firms (with 2,402 in the service sector) observed continuously from 2010 to 2016. The repeated cross sections include a broader range of firms each year, including smaller and newer entries, with sample sizes ranging from approximately 600,000 to over 740,000 and service firm counts increasing from 155,387 in 2010 to 205,182 in 2016. The stark difference in average revenue and assets between the two datasets underscores the importance of considering both large, established firms and broader market dynamics when analyzing the effects of the B2V reform.

[3]Qichacha https://www.qcc.com/ is a comprehensive business information platform based in China, providing up-to-date detailed and reliable data about companies.

4.2 Empirical Strategy

The "B2V" reform was implemented gradually across regions (for a subset of sectors) between 2012 and 2013, and gradually across sectors (nationwide) between 2014 and 2016. I exploit this policy variation using an event-study DID framework. While the staggered implementation of the reform across provinces and sectors provides a valuable source of variation for my empirical specification, potential identification concerns remain. These include the possibility of anticipatory effects in provinces that implemented the reform later, endogenous timing of reform implementation based on unobserved local economic conditions, and the challenge of establishing a clear control group given the interconnected nature of the economy. Furthermore, the parallel trends assumption crucial for my identification strategy may be violated if pre-reform trends in outcomes differed systematically between early and late-adopting provinces or sectors, potentially due to unobserved factors correlated with both the timing of reform implementation and our outcomes of interest.

To alleviate these concerns, I add a rich set of sector, province and their interacted fixed effects to control for the selection problem. The parallel trend assumption is satisfied and shown in the next section. However, since the problem of imperfect control groups due to economic linkages between sectors is hard to be fully solved empirically, the empirical results, especially the comparison between service and manufacturing firms, be interpreted as relative effects: comparing more heavily impacted sectors to less impacted sectors rather than absolute treatment effects. To quantify the general equilibrium impacts and obtain welfare estimates that account for these cross-sector linkages over space, I develop a spatial general equilibrium model that captures the interconnected nature of the economy that allows me to evaluate the reform's total effects in Section 5.

4.2.1 Firm Entry

I begin by investigating the impact of the reform on extensive firm entry using the event study difference-in-differences as follows:

$$\log\left(\mathrm{M}_{s,p,t}\right) = \alpha + \sum_{k\neq -1} \beta_k \times \mathrm{Treat}_{s,p,k} + \eta^1_{s,p} + \eta^2_{p,t} + \epsilon_{f,s,p,t}, \tag{4.1}$$

where $M_{s,p,t}$ is the total number of firms in sector s, province p and year t. $\mathrm{Treat}_{s,p,k}$ is equal to 1 if the sector s in province p were reformed at $t-k$."; 0 otherwise (and 0 for all manufacturing sectors). $\eta^1_{s,p}$ and $\eta^2_{p,t}$ are sector-by-province and province-by-year fixed effects, respectively. I cluster standard errors $\epsilon_{s,p,t}$ at the sector-by-province level, as this represents the level of the tax reform variation. First, I did not include the sector-by-year fixed effects as they are directly correlated with the treatment variable, which varies by sector and year. Second, the empirical analysis faces an additional identification challenge from a concurrent policy implemented in 2013 that aimed to stimulate small firm entry nationwide.[4] Since service sectors are characterized by a higher proportion of small firms compared to manufacturing sectors, the policy would disproportionately impact service firms, potentially confounding the estimated effects of the tax reform.

To address this endogeneity concern, I employ two complementary strategies. First, I include province-by-year fixed effects to account for time-varying provincial-level variables, which could account for the impact of the 2013 small firm policy across provinces. Second, I use the methodology of Callaway and Sant'Anna (2021), using the not-yet-treated service sectors as control units rather than manufacturing sectors, to help identify the absolute treatment effect of the reform more accurately.

4.2.2 Firm Revenue

I next examine the impact of the tax reform on firm-level activities at the intensive margin using the administrative tax survey data that provides establishment-level information on firms' total revenue. To capture heterogeneous effects across different types of firms, I construct two complementary datasets. The first is a balanced panel consisting of large firms that are consistently sampled across all years, allowing me to track the same firms over time. The second is a repeated cross-section that includes a broader range of firms each year, with more representation of small and newly-entered firms that may not appear consistently in the survey. Table 3 highlights the substantial differences between these two datasets: firms in the balanced panel are approximately 5-7 times larger than those in the repeated cross-sections in terms of both revenue and assets. This size differential is important because it lets me examine whether the reform's effects vary between established large firms and the broader population of smaller, more dynamic

[4] For more information on the policy, see https://www.gov.cn/zwgk/2013-07/25/content_2455259.htm.

firms. I employ different specifications for the two datasets:

$$\log\left(\text{Revenue}_{f,s,p,t}\right) = \alpha + \sum_{k\neq -1} \beta_k \times \text{Treat}_{s,p,k} + \eta^1_{s,p} + \eta^2_f + \eta^3_{p,t} + \psi_t + \epsilon_{f,s,p,t}, \tag{4.2a}$$

$$\log\left(\text{Revenue}_{f,s,p,t}\right) = \alpha + \sum_{k\neq -1} \beta_k \times \text{Treat}_{s,p,k} + \eta^1_{s,p} + +\eta^2_{p,t}\psi_t + \epsilon_{f,s,p,t}, \tag{4.2b}$$

where $\log\left(\text{Revenue}_{f,s,p,t}\right)$ presents the log value of sales at firm f in sector s, province p, and year t. $\text{Treat}_{s,p,k}$ is defined as the same way above, η^2_f,$\eta_{s,p}$, and ψ_t are the firm, sector-by-province, and year fixed effects, respectively. I drop the firm fixed effect for the specification using the repeated cross-sections dataset to include as many small and new firms as possible. Again, I cluster standard errors at the sector-by-province level.

The investment variable is largely missing in the tax survey dataset. To provide additional evidence of the reform's impact on firm activities and complement the revenue analysis, I examine land transactions as an alternative measure of firm expansion and investment. Land transactions are particularly informative in the Chinese context because land use rights represent a major input for business operations, and land purchases often signal firm growth and expansion plans. Moreover, since land sales constitute an important source of fiscal revenue for local governments[5], analyzing land transactions gives insights about both firm behavior and local government finances. Using comprehensive data that includes every land transaction record in China's primary land market, I find that firms in reformed sectors and regions pay higher land prices and acquire larger land areas compared to their counterparts in non-reformed regions or sectors (see Appendix B). These results corroborate the firm-level revenue analysis by providing independent evidence that the tax reform stimulated firm expansion and investment activities.

4.2.3 Cross-region-sector Trade

The firm-level evidence demonstrates that replacing the business tax with the value added tax stimulated individual firm growth and expansion. However, a key insight from the reform is that it should also affect how firms trade with each other across sectors and regions by reduc-

[5]Land transaction data is obtained from the Ministry of Natural Resources (formerly the Ministry of Land and Resources), which is available at `www.landchina.com`.

ing the frictions of cross-sector trade. To examine this mechanism directly, I analyze changes in interregional trade patterns using input-output data.

I estimate the impact of the tax reform on interregional trade flows using the Inter-Regional Input-Output (IRIO) tables for 2012 and 2017, which captures the trade flow before and after reform's completion. This analysis allows me to test whether the removal of the artificial tax burdens increased trade flow involving service sectors. I employ a gravity-style specification to test this hypothesis:

$$\text{IntUse}_{ok,dj,t} = \alpha + \boldsymbol{\beta} \cdot \text{Treat}_{k,j,t} + \eta^1_{o,t} + \eta^2_{d,t} + \eta^3_{o,k,d,j} + \psi_t + \epsilon_{o,k,d,jt}, \tag{4.3}$$

where $\text{IntUse}_{o,k,d,j,t}$ is the intermediate trade flows from sector k of origin province o to sector j of destination province d in year t if it's the intensive margin; or equal to 1 if it's positive and 0 otherwise for the extensive margin. $\text{Treat}_{k,j,t}$ is defined as one if the year is in 2017 and the trade flow is imported from the service sector or exported to the service sector.[6] $\eta^1_{o,t}$, $\eta^2_{d,t}$, $\eta^3_{o,k,d,j}$, and ψ_t are the origin-province-by-year, destination-province-by-year, origin-by-destination, and year fixed effects, respectively. I use the two-way clustered standard errors at origin-sector and destination-sector level. I use the Poisson pseudo-maximum-likelihood (PPML) estimator for estimations to deal with the issue of having many zeros and since the estimation is based on the levels of the dependent variable rather than its logarithm, both small and large values contribute to the estimation process in a balanced manner (Silva and Tenreyro, 2006).

4.3 Main Findings

4.3.1 Firm Entry

The empirical analysis reveals that removing business taxes has a substantial positive impact on firm entry in affected service sectors. Figure 2 presents the results using two different identification strategies. Panel (A) shows the results from a standard Fixed-Effect difference-in-differences estimation using manufacturing sectors as the control group. The result indicates an average differential increase in firm entry of approximately 25% compared to manufacturing firms. Panel (B) displays the result using Callaway and Sant'Anna (2021) difference-in-differences estimation, which

[6] Shanghai is dropped to avoid the contamination because it was treated in 2012, January.

uses not-yet-treated service sectors as the control units. This approach addresses potential biases in the Fixed-Effect estimation for using manufacturing firms as the control group. It shows an average differential increase of about 21% compared to not-yet-treated service sectors.

The consistency between the two specifications despite using different control groups provides robust evidence that the tax reform significantly stimulated new firm creation in the service sector.

4.3.2 Firm Revenue

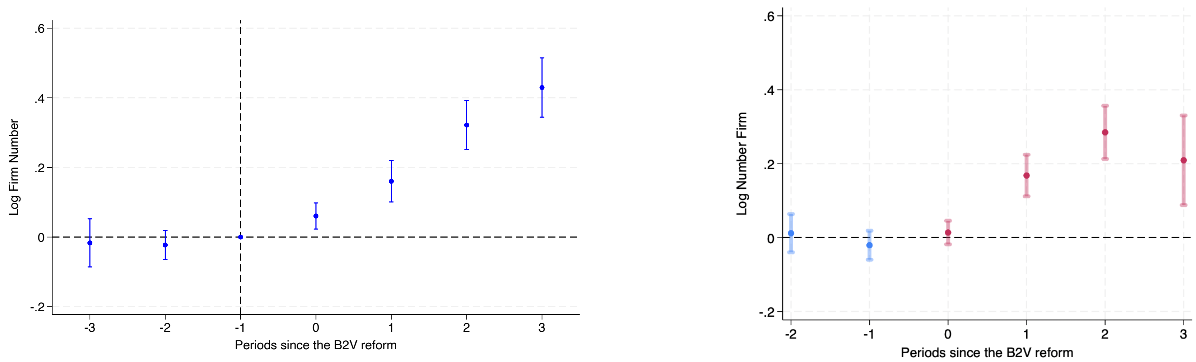
The firm-level revenue analysis indicates striking heterogeneity in how different types of firms responded to the tax reform. Figure 3 compares the revenue dynamics for two distinct samples that capture different segments of the firm distribution.

Panel (A) presents the result for the balanced panel, which consists of large, established firms that are consistently sampled across all years. The pre-reform period shows a clear upward trend in the differential revenue growth between service and manufacturing sectors, which could reflect the rapid expansion of China's service sector during this period. Notably, this pre-trend indicates that large service firms were already experiencing faster growth than their manufacturing counterparts before the reform implementation. Following the reform, this differential growth trajectory flattens. This slowdown may reflect increased competitive pressure as new firms enter the market following the tax reform.

In stark contrast, Panel (B) shows a markedly different pattern for the repeated cross-section sample, which captures a broader range of firms including many smaller and newer enterprises. Crucially, the sample exhibits no significant pre-reform trends, lending credibility to the parallel trends assumption for causal identification. After the reform implementation, these firms experience a substantial and statistically significant increase in revenue growth of approximately 22% relative to the manufacturing firms.

This contrast between between these two panels highlights important insights reform's heterogeneous impacts. While large, established service firms were already on a strong growth trajectory, smaller and newer firms appear to have been disproportionately constrained by the business tax burden. The reform's removal of these tax and trade frictions particularly benefited the smaller firms.

Figure 2. Effect of the Reform on Firm Entry in Service Sectors.

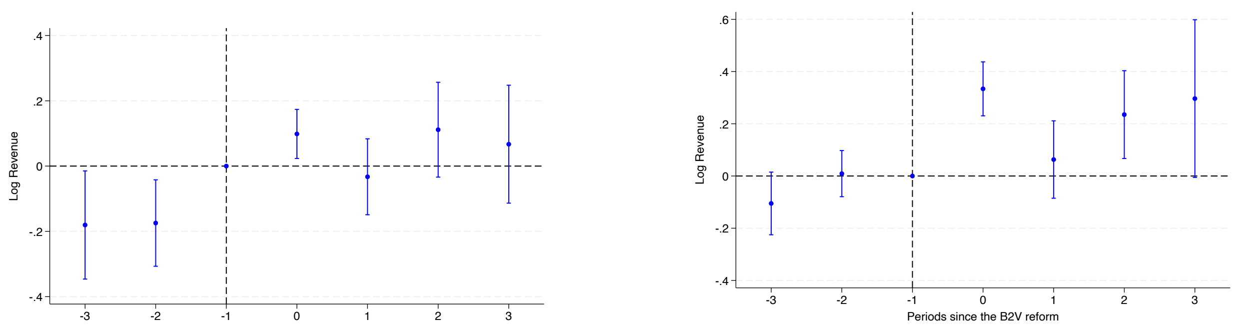


(A) TWFE, Manufacturing as Control

(B) CSDID - Not Yet Treated as Control

Notes: This figure illustrates the impact of the Business to Value-Added Tax (B2V) reform on firm entry dynamics in China's service sectors. panel (A) shows the estimation result of a Two-Way Fixed Effects (TWFE) method using manufacturing sectors as the control group. The x-axis represents time periods relative to the reform implementation, with 0 marking the reform year. The y-axis shows the log point change in the number of new firms entering the market. The graph shows a significant increase in firm entry for service sectors post-reform, with an average differential increase of about 25% compared to manufacturing sectors. Panel (B) displays the result using Callaway and Sant'Anna (2021) Difference-in-Differences (CSDID) estimation, which uses not-yet-treated service sectors as the control group. This approach addresses potential biases in the TWFE estimation when treatment timing varies. The x-axis and y-axis are similar to Panel A. The graph shows a significant increase in firm entry for service sectors post-reform, with an average differential increase of about 22% compared to not-yet-treated service sectors.

Figure 3. Effect of the Reform on Firm Growth in Service Sectors



(A) Growth of Firm Revenue, Balanced Panel

(B) Growth of Firm Revenue, Repeated Cross Section

Notes: This figure presents the differential growth in firm revenue following the B2V tax reform. Panel A shows results for a balanced panel of large, repeatedly sampled firms, while Panel B displays results for a repeated cross-section including smaller and new firms. The x-axis represents time periods relative to the reform implementation, with 0 marking the reform year. The y-axis shows the log point change in revenue. The stark difference between panels highlights the reform's heterogeneous impacts, with smaller and newer firms (Panel B) experiencing substantially higher growth post-reform compared to larger, established firms (Panel A).

Table 4. Effects of the Reform on Intermediate Trade Flow

	All (1)	Intra-Province (2)	Inter-Province (3)
B2V	0.356***	0.479***	0.207***
	(0.030)	(0.041)	(0.076)
fromPfromS × toPtoS FE	✓	✓	✓
fromP×Year FE	✓	✓	✓
toP×Year FE	✓	✓	✓
Year FE	✓	✓	✓
Mean of dep var	10,617	178,436	3,831
Observations	2,582,828	100,380	2,482,448

Notes: This table presents the estimated impacts of the B2V tax reform on intermediate input use across different geographic scales. Column (1) shows the overall effect, while columns (2) and (3) break down the impact into intra-province and inter-province effects, respectively. The coefficient on 'B2V' represents the percentage change in directly impacted intermediate input use attributable to the reform compared to the manufacture-to-manufactur usage. Standard errors are in parentheses. *** indicates significance at the 1% level. The results suggest that the reform led to substantial increases in intermediate input use, with particularly strong effects within provinces.

4.3.3 Cross-region-sector Trade

The trade flow analysis provides direct evidence that the tax reform enhanced economic integration by reducing trade frictions in intermediate goods transactions. I find substantial increases in trade involving service sectors both within and across regions following the reform.

Table 4 presents the estimated effects on intermediate input trade flows. The reform led to a 35.6% increase in trade flow involving service sectors — either as suppliers or purchasers of intermediate inputs — relative to purely manufacturing-to-manufacturing transactions. Trade within provinces increased by 47.9% while trade across regions grew by 20.7%. The larger within-region effects likely reflect the fact that geographic proximity facilitates the new connections of trade. The smaller but positive between-region effects demonstrate that reform's benefits extended beyond the local markets, contributing to broader economic integration across China's regions.

These results indicate that the tax reform facilitated more efficient allocation of resources across sectors and regions, allowing firms to source inputs from potentially more productive suppliers and enhanced the interconnectedness of the production networks. The empirical analysis demonstrates that eliminating the tax on intermediate inputs generated substantial positive effects on firm entry, growth and trade flows. Building on these empirical insights, a quantitative general equilibrium model can extend the analysis in several important dimensions to provide a more comprehensive understanding of the reform's economic impact.

First, while the empirical estimates effectively identify the relative effects between service and manufacturing sectors, a structural model can quantify the absolute welfare changes by accounting for general equilibrium effects that propagate through the interconnected production network. This allows for a complete assessment of the reform's net benefits.

Second, a quantitative model can explicitly identify mechanisms that reduced form analysis suggest but cannot directly examine. These includes: firms' vertical integration decisions, a key channel through which the business tax distorted production choices; the distributional consequences across regions, particularly how provinces with different economic structures were affected; and the fiscal implications for local governments that experienced a varying degree of tax revenue changes.

Third, a structural model can decompose the sources of observed improvements by distinguishing two potential channels: simple tax burden reduction versus the elimination of cross-sector trade frictions that distorted production networks in every stage. This decomposition can provide valuable insights for optimal tax policy design by identifying which mechanisms drive the largest welfare gains.

5 The Model

I develop a quantitative spatial general equilibrium model that complements the quantitative model with input-output linkages from Caliendo et al. (2022b) with a firm-entry margin, as in Melitz (2003) and Krugman (1980). Moreover, firms face the decision of buying service inputs from the market or producing them in-house (Bernard et al., 2019). There are N = 32 (31 Provinces + 1 Rest of the World) regions and J = 41 sectors (1 agricultural, 25 manufacturing and 15 service sectors), with regions denoted by o and d and sectors by k and j. Labor is the sole factor of production supplied inelastically and mobile across sectors within a region but immobile across regions.

5.1 Preference

In each region, L_d measure of representative households seek to maximize utility from consuming a vareity of final goods $Y_{ok}^C(\omega)$. Their utility is specified by:

$$u\left(Y_d^C\right) = \left[\sum_{o=1}^{N}\sum_{k=1}^{J} s_{ok,d}^{\frac{1}{1+\theta}} \left(\int_{\Omega_{ok}} Y_{ok,d}^C(\omega)^{\frac{\theta}{1+\theta}} d\omega\right)\right]^{\frac{1+\theta}{\theta}}, \tag{5.1}$$

subject to the budget constraint $\sum_{o=1}^{N}\sum_{k=1}^{J}\int_{\Omega_{ok}} p_{ok,d}(\omega)Y_{ok,d}^{C}(\omega) \leq I_d$. I denote households' income by $I_d \equiv \left(\sum_{j=1}^{J} w_d L_{dj}\right) + D_d + R_d$, where income is the sum of labor wages, transfer payments, and tax revenue generated during production in region d. $s_{ok,d}$ is the preference parameter residents in region d place on goods from region o and sector k. $\theta+1$ is the elasticity of substitution between any two individual varieties ω and ω'across origins and sectors.

The utility-maximization problem results in the standard Dixit-Stiglitz price index:

$$P_d = \left(\sum_{o=1}^{N}\sum_{k=1}^{J} s_{ok,d}\int_{\Omega_{ok}} p_{ok,d}(\omega)^{-\theta}d\omega\right)^{-\frac{1}{\theta}}, \tag{5.2}$$

where $U\left(Y_d^C\right) = I_d/P_d$.

5.2 Production

5.2.1 Final Goods

There is a continuum Ω of possible varieties that all regions can produce, and suppose that every firm in the world produces a distinct variety $\omega \in \Omega$. Because there is a one-to-one mapping between firms and varieties, I can label each firm by the variety it produces: $y_{dj}(\omega)$. Let the set of varieties produced by firms located in region d, sector j be denoted by $\Omega_{dj} \subseteq \Omega$. , and produced with a Cobb-Douglas technology:

$$y_{dj}(\omega) = \varphi\left[\ell_{dj}(\omega)\right]^{\alpha_{dj}}\left[Y_{dj}^{M}(\omega)\right]^{1-\alpha_{dj}} \tag{5.3}$$

where φ is the firm-specific productivity, ℓ_{dj} is labor and the intermediate input Y_{dj}^{M} is a CES aggregate of intermediate goods from all sectors and regions, including *in-house production goods* (vertical integration). The parameter χ_{dj} captures the relative importance of labor versus intermediate inputs in production and can vary across regions and sectors.

The CES functional form generates endogenous responses instead of constant factor shares in total the expenditure. This flexibility is crucial for modeling how tax changes affect firms' input choices: when taxes make market-purchased intermediate inputs more expensive, firms can substitute towards in-house provision of services or vice versa.

The marginal cost with the value-added tax or business tax paid during the production implies the unit cost bundle for the good j to be (for $\varphi = 1$):

$$c_{dj} = \mathbf{C_1}(t_j w_d)^{\alpha_{dj}} \left\{ [1 + [\mathbf{S} = \mathbf{1}][\mathbf{R} = \mathbf{0}](t_j - 1)] \, P_{dj}^M \right\}^{1-\alpha_{dj}}, \tag{5.4}$$

where $t_j \in \{1 + v_j, 1 + b_j\}$, depending on whether j is a service or manufacturing sector, and whether it's before or after the reform. α_{dj} is the factor share of labor that can vary across regions and sector. [7]

This cost equation captures the key taxation distortions mentioned in previous texts. All intermediate input prices already incorporate taxes imposed. However, the additional tax treatment differs by sector and reform period: if j is a service sector ($\mathbf{S} = \mathbf{1}$) and before the reform ($\mathbf{R} = \mathbf{0}$), firms must pay business taxes on these already-taxed intermediate inputs, creating a tax cascading effect. In contrast, even though manufacturing firms still pay the after-tax price when they buy from service firms, they do not need to pay additional taxes on that.

5.2.2 (Composite) Intermediate Input

$Y_{dj}^M(\omega)$ is the (composite) intermediate input used in region d and sector j by firm ω, a CES aggregate of intermediate goods from all sectors and locations, including the possibility of producing a good in house:

$$Y_{dj}^M(\omega) = \left[\sum_{o=1}^{N} \sum_{k=1}^{J} \iota_{ok,dj}^{\frac{1}{1+\theta}} \int_{\Omega_{ok}} [Y_{ok,dj}^m(\omega', \omega)]^{\frac{\theta}{1+\theta}} d\omega' + \sum_{k=1}^{J} \iota_{Hk,dj}^{\frac{1}{1+\theta}} \left[Y_{Hk,dj}^m(\omega) \right]^{\frac{\theta}{1+\theta}} \right]^{\frac{1+\theta}{\theta}}, \tag{5.5}$$

The input weight $\iota_{ok,dj}$ captures the relative importance of different intermediate goods from sector k and region o in the production of intermediate inputs in sector j and region d.

In-house Production. I assume the production of an in-house service task only employs labor and has the following production function: where z_{dj}^H is the average productivity of in-house production in region d, sector j. This parsimonious formulation abstracts from firm-level hetero-

[7] The resulting constant associated with the marginal cost is $\mathbf{C1} = \left(\alpha_{dj}^{\alpha_{dj}} (1 - \alpha_{dj})^{1-\alpha_{dj}} \right)^{-1}$.

geneity in in-house tasks and avoids the needs to track the destination-sector pair d, j firm-level price index within for the composite intermediate input [8].

The solution to the intermediate input cost minimization problem in (5.5) indicates that the price index of the intermediate input is:

$$P_{dj}^{M}(\omega) = \left\{ \left[\sum_{o=1}^{N} \sum_{k=1}^{J} \iota_{ok,dj} \int_{\Omega_{ok}} \left[p_{ok,dj}(\omega') \right]^{-\theta} d\omega' \right] + \sum_{k=1}^{J} \iota_{Hk,dj} \left(\frac{w_d}{z_{dj}} \right)^{-\theta} \right\}^{-\frac{1}{\theta}}, \tag{5.6}$$

Therefore, I distinguish two types of labor a firm in region d and sector j employs: labor for the primary product and for in-house service task, where the expenditure on the latter is relabeled as "intermediate goods" and therefore has different elasticity of substitution in the production function.

The share of labor for in-house production is ζ_{dj}^{H} and $\alpha_{dj} + \zeta_{dj}^{H}$ should be equal to the value-added share ζ_{dj}^{0} observed in the input-output table.

Profit Maximization and Price. Under the monopolistic market structure, a firm in region o, sector k with productivity φ sets its own price to other producers and consumers by solving the following profit maximization problem, *conditional* on exporting to the pair d, j, sets its the price at

$$p_{ok,dj}(\varphi) = \frac{1+\theta}{\theta} \frac{c_{ok} t_{ok,dj} \kappa_{ok,dj}}{\varphi} \tag{5.7}$$

and *conditional on* exporting to consumers in d, it sets the price:

$$p_{ok,d}(\varphi) = \frac{1+\theta}{\theta} \frac{c_{ok} t_{ok,d} \kappa_{ok,d}}{\varphi} \tag{5.8}$$

[8] In principle, one could allow in-house productivity to be correlated with firm-level productivity, so that more productive firms are also more efficient in their in-house service provision. The price index would be $P_{dj}^{M}(\omega) = \left\{ \left[\sum_{o=1}^{N} \sum_{k=1}^{J} \iota_{ok,dj} \int_{\Omega_{ok}} \left[p_{ok,dj}(\omega') \right]^{-\theta} d\omega' \right] + \sum_{k=1}^{J} \iota_{Hk,dj} \left[\frac{w_d}{z_{dj}(\omega)} \right]^{-\theta} \right\}^{-\frac{1}{\theta}}$. This assumption would be consistent with the idea that larger, more productive firms are more likely to internalize intermediate tasks. However, doing so would imply that the in-house productivity term scales proportionally with φ, requiring integration over a joint distribution of market and in-house productivities. This introduces considerable additional complexity into the aggregation step and makes it difficult to retain the tractable Pareto-based closed forms for average productivity. For tractability, I therefore assume that in-house productivity is region–sector specific but independent of firm-level productivity, which preserves analytical clarity while still capturing the key trade-off between market purchases and in-house production.

The cross-sector tax $t_{ok,dj}$ parameter captures the VAT credit mechanism that creates differential pricing across transaction types. Under the VAT system, manufacturing firms can claim input tax credit when selling to other VAT-registered businesses, but not when selling to final consumers or business tax-paying service firms. Specifically,

$$t_{ok,dj} = \left[1 + \frac{\zeta^0_{ok}(1-t_k)}{t_k}\right],$$

if k and j are both manufacturing sectors (avoid paying taxes) and $t_{ok,dj} = 1$ if it is a transaction selling to or buying from a service sector (cannot avoid paying taxes) before the reform. Recall that ζ^0_{ok} is the value-added share for region-sector pair o, k. After the reform, when all sectors operate under VAT, $t_{ok,dj} = \left[1 + \frac{\zeta^0_{ok}(1-t_k)}{t_k}\right]$ applies universally to all sectors, eliminating cross-sector pricing distortion. $t_{ok,d} = 1$ regardless any sector or reform status, because the consumers always pay the taxes.

5.3 Aggregation

Let $\mu_{ok,dj}(\varphi)$ and $\mu_{ok,d}$ be the (equilibrium) probability density functions of the productivity of firms from region o sector k that sell to producers in region d sector j or consumers in region d. Let $M_{ok,dj}$ and $M_{ok,d}$ be the (equilibrium) measure of firms exporting from pair o, k to pair d, j or to d. Aggregate composite intermediate input price in (A9) can be written as:

$$P^M_{dj} = \left\{ \left[\sum_{o=1}^{N}\sum_{k=1}^{J} \iota_{ok,dj} \int_0^\infty [p_{ok,dj}(\varphi)]^{-\theta} M_{ok,dj}\mu_{ok,dj} d\varphi\right] + \sum_{k=1}^{J} \iota_{Hk,dj} \left(\frac{w_d}{z_{dj}}\right)^{-\theta} \right\}^{-\frac{1}{\theta}}, \tag{5.9}$$

Using the pricing rule (5.7), this can be written as

$$P^M_{dj} = \left[\sum_{o=1}^{N}\sum_{k=1}^{J} \iota_{ok,dj} \left(\frac{1+\theta}{\theta}\frac{c_{ok}t_{ok,dj}\kappa_{ok,dj}}{\tilde{\varphi}_{ok,dj}}\right)^{-\theta} M^e_{ok}\left(1 - G_{ok}(\varphi^*_{ok,dj})\right) + \sum_{k=1}^{J} \iota_{Hk,dj} \left(\frac{w_d}{z_{dj}}\right)^{-\theta}\right]^{-\frac{1}{\theta}}, \tag{5.10}$$

where

$$\tilde{\varphi}_{ok,dj} = \left(\int_0^\infty (\varphi)^\theta \mu_{ok,dj}(\varphi) d\varphi\right)^{\frac{1}{\theta}} \tag{5.11}$$

Similarly, aggregate consumption index defined in (5.2) can be written as:

$$P_d = \left[\sum_{o=1}^{N}\sum_{k=1}^{J} s_{ok,d}\int_0^{\infty} p_{ok,d}(\varphi)^{-\theta} M_{ok,d}\mu_{ok,d}(\varphi)d\varphi\right]^{-\frac{1}{\theta}} \tag{5.12}$$

Using the pricing rule (5.8), this can be written as

$$P_d = \left[\sum_{o=1}^{N}\sum_{k=1}^{J} s_{ok,d}\left(\frac{1+\theta}{\theta}\frac{c_{ok}\kappa_{ok,dj}}{\tilde{\varphi}_{ok,d}}\right)^{-\theta} M^e_{ok}\left(1-G_{ok}(\varphi^*_{ok,d})\right) + \sum_{k=1}^{J}\iota_{Hk,dj}\left(\frac{w_d}{z_{dj}}\right)^{-\theta}\right]^{-\frac{1}{\theta}}, \tag{5.13}$$

where

$$\tilde{\varphi}_{ok,d} = \left(\int_0^{\infty}(\varphi)^{\theta}\mu_{ok,d}(\varphi)d\varphi\right)^{\frac{1}{\theta}} \tag{5.14}$$

Then the aggregate gravity equation for producers in d, j and consumers in d can be written as:

$$X^P_{ok,dj} = M_{ok,dj}\left(\frac{1+\theta}{\theta}t_{ok,dj}c_{ok}\kappa_{ok,dj}\right)^{-\theta}\tilde{\varphi}^{\theta}_{ok,dj}\iota^{(1+\theta)}_{ok,dj}\left(P^M_{dj}\right)^{1+\theta}Y^M_{dj} \tag{5.15}$$

$$X^C_{ok,d} = M_{ok,d}\left(\frac{1+\theta}{\theta}\kappa_{ok,d}c_{ok}\right)^{-\theta}\tilde{\varphi}^{\theta}_{ok,d}s_{ok,d}P^{\theta}_d I_d. \tag{5.16}$$

5.4 Selection into Exporting

To determine the equilibrium number of entrants $M_{ok,dj}$ and the average productivity of entrants $\tilde{\varphi}_{ok,dj}$, I have to consider the export decisions of firms. A firm from region o and sector k with productivity φ conditional on producing will export to producers in region d, sector j and consumers in region d if and only if:

$$\pi^P_{ok,dj}(\varphi) \geq f_{ok,dj}c_{ok}, \;\; \pi^C_{ok,dj}(\varphi) \geq f_{ok,d}c_{ok}$$

This implies the cut-off productivity for exporting from o, k to d, j is:

$$\varphi \geq \varphi^*_{ok,dj} \equiv \left[\frac{(1+\theta)c_{ok}f_{ok,dj}(\frac{1+\theta}{\theta}c_{ok}t_{ok,dj}\kappa_{ok,dj})^{\theta}}{(\iota_{ok,dj}P^M_{dj})^{1+\theta}Y^M_{dj}}\right]^{\frac{1}{\theta}} \tag{5.17}$$

Similarly, the cut-off productivity for exporting from o, k to d to consumers is:

$$\varphi \geq \varphi^*_{ok,d} \equiv \left[\frac{(1+\theta)c_{ok}f_{ok,d}(\frac{1+\theta}{\theta}c_{ok}\kappa_{ok,d})^{\theta}}{s_{ok,d}P_d^{\theta}I_d}\right]^{\frac{1}{\theta}} \tag{5.18}$$

It further allow us to determine the "average" productivity of producers selling from o, k to producers in d, j:

$$\tilde{\varphi}_{ok,dj} = \left(\frac{1}{1 - G_{ok}(\varphi^*_{ok,dj})}\int_{\varphi^*_{ok,dj}}^{\infty} \varphi^{\theta} dG_{ok}(\varphi)\right)^{\frac{1}{\theta}}$$

and "average" productivity of producers selling from o, k to consumers in d:

$$\tilde{\varphi}_{ok,d} = \left(\frac{1}{1 - G_{ok}(\varphi^*_{ok,d})}\int_{\varphi^*_{ok,d}}^{\infty} \varphi^{\theta} dG_{ok}(\varphi)\right)^{\frac{1}{\theta}}$$

5.5 Gravity

General CES structure. Aggregating firm-level CES demands, the bilateral expenditure on intermediate inputs from origin (o, k) to destination (d, j) is

$$X^P_{ok,dj} = \iota_{ok,dj}\, Y^M_{dj}\, \left(P^M_{dj}\right)^{1+\theta} \int_{\Omega_{ok}} p_{ok,dj}(\omega')^{-\theta}\, d\omega', \tag{5.19}$$

where $p_{ok,dj}(\omega')$ is the delivered price of variety ω' including trade frictions. Similarly, the bilateral expenditure on final goods from (o, k) to destination d is

$$X^C_{ok,d} = s_{ok,d}\, I_d\, P_d^{1+\theta} \int_{\Omega_{ok}} p_{ok,d}(\omega')^{-\theta}\, d\omega', \tag{5.20}$$

where I_d is total final expenditure, P_d the CES final-good price index, and $s_{ok,d}$ the expenditure share on composite (o, k) in d.

With heterogeneous firms. Suppose now that sector (o, k) hosts M_{ok} potential entrants drawing productivities φ from distribution $G_{ok}(\varphi)$, and only firms with $\varphi \geq \varphi^*_{ok,\cdot}$ serve each market. Then the integrals in (5.19)–(5.20) evaluate to moments of the productivity distribution. For intermediate inputs:

$$X^P_{ok,dj} = M^e_{ok} \left(\frac{1+\theta}{\theta} \, c_{ok} t_{ok,dj} \kappa_{ok,dj} \right)^{-\theta} \iota_{ok,dj} \left(P^M_{dj} \right)^{1+\theta} \int_{\varphi^*_{ok,dj}}^{\infty} \varphi^\theta \, dG_{ok}(\varphi) \;\; Y^M_{dj}, \tag{5.21}$$

and the corresponding bilateral share is

$$\gamma^P_{ok,dj} \equiv \frac{X^P_{ok,dj}}{P^M_{dj} Y^M_{dj}} = M^e_{ok} \left(\frac{1+\theta}{\theta} \, c_{ok} t_{ok,dj} \kappa_{ok,dj} \right)^{-\theta} \iota_{ok,dj} \left(P^M_{dj} \right)^{\theta} \int_{\varphi^*_{ok,dj}}^{\infty} \varphi^\theta \, dG_{ok}(\varphi). \tag{5.22}$$

For final goods:

$$X^C_{ok,d} = M^e_{ok} \left(\frac{1+\theta}{\theta} \, c_{ok} \kappa_{ok,d} \right)^{-\theta} s_{ok,d} \, P^\theta_d I_d \int_{\varphi^*_{ok,d}}^{\infty} \varphi^\theta \, dG_{ok}(\varphi), \tag{5.23}$$

$$\gamma^C_{ok,d} \equiv \frac{X^C_{ok,d}}{I_d} = M^e_{ok} \left(\frac{1+\theta}{\theta} \, c_{ok} \kappa_{ok,d} \right)^{-\theta} s_{ok,d} \, P^\theta_d \int_{\varphi^*_{ok,d}}^{\infty} \varphi^\theta \, dG_{ok}(\varphi). \tag{5.24}$$

Factor shares. The implied factor share of intermediate input from origin (o, k) used in (d, j) production is

$$\xi_{ok,dj} \equiv (1 - \alpha_{dj}) \, \gamma^P_{ok,dj}. \tag{5.25}$$

5.6 Free entry and the allocation of factors across firms

Firm entry is endogenously determined, as in Melitz (2003). Suppose that firms have to incur an entry cost $f_{ok} > 0$ prior to learning their productivity. Then, the free entry condition requires that the expected profits equal the entry cost.

$$c_{ok} f^e_{ok} = E_\varphi \left[\sum_{d=1}^{N} \sum_{j=1}^{J} \max \left\{ \pi^P_{ok,dj}(\varphi) - f_{ok,dj} c_{ok}, 0 \right\} + \sum_{d=1}^{N} \max \left\{ \pi^C_{ok,d}(\varphi) - f_{ok,d} c_{ok}, 0 \right\} \right].$$

5.7 The Distribution of Firm Productivity

Following Chaney (2008), I assume the productivity of firms in each region-sector follows the Pareto distribution:

$$G_{ok}(\varphi) = G(\varphi) = 1 - \varphi^{-\rho}$$

I assume $\rho > \theta$ in order for the trade flow to be finite. I then can write the "average productivity" and substituting them into above into the two gravity equations yields:[9]

$$X^P_{ok,dj} = C_2(t_{ok,dj}c_{ok}\kappa_{ok,dj})^{-\rho} f_{ok,dj}^{\frac{\theta-\rho}{\theta}} M_{ok}[Y^M_{dj}(P^M_{dj})^{1+\theta}]^{\frac{\rho}{\theta}} (\iota_{ok,dj})^{\frac{\rho}{\theta}} \tag{5.26}$$

$$X^C_{ok,d} = C_2(c_{ok}\kappa_{ok,d})^{-\rho} f_{ok,d}^{\frac{\theta-\rho}{\theta}} M_{ok}(I_d P^\theta_d)^{\frac{\rho}{\theta}} s_{ok,d}^{\frac{\rho}{\theta}}. \tag{5.27}$$

The average productivity can be written as a function of the cutoff productivity:

$$\tilde{\varphi}^\theta_{ok,dj}(\varphi^*_{ok,dj}) = \frac{1}{1-G(\varphi^*_{ok,dj})} \int_{\varphi^*_{ok,dj}}^{\infty} \varphi^\theta dG_{ok}(\varphi) = \frac{1}{1-G(\varphi^*_{ok,dj})} \frac{\rho}{\rho-\theta} (\varphi^*_{ok,dj})^{\theta-\rho} \tag{5.28}$$

$$\tilde{\varphi}^\theta_{ok,d}(\varphi^*_{ok,d}) = \frac{1}{1-G(\varphi^*_{ok,dj})} \int_{\varphi^*_{ok,d}}^{\infty} \varphi^\theta dG_{ok}(\varphi) = \frac{1}{1-G(\varphi^*_{ok,dj})} \frac{\rho}{\rho-\theta} (\varphi^*_{ok,d})^{\theta-\rho} \tag{5.29}$$

5.8 Market Clearing and Trade (Im)balance

Let X_{ok} denote the aggregate expenditure on intermediate goods by region o sector k. Denote $X_{o'k',ok}$ to be the expenditure on intermediate goods from origin o' sector k' by region o sector k. At equilibrium, the total output of the good from region o sector k, Y_{ok} equals to the demand on intermediate and final goods from region o sector k by all other regions $\sum_{o'=1}^{N}\left[\sum_{k'=1}^{J}(X_{o'k'} \times \gamma_{ok,o'k'}) + Y^C_{ok,o'}\right]$, where $\gamma_{ok,o'k'}$ is the share of intermediate input from pair (o,k) in the total intermediate input used by pair (o',k'). In addition, I denote the factor share on intermediate goods from region o' sector k' as $\xi_{o'k',ok}$ (different from $\gamma_{o'k',ok}$, $\xi_{o'k',ok}$ is the share of intermediate input in total expenditure, including labor income). Therefore, the market clearing condition is:

Labor market clearing.

$$t_k w_o L_o = \left(1 - \sum_{o'=1}^{\mathbf{N}} \sum_{k'=1}^{J} \xi_{o'k',ok}\right) \sum_{o'=1}^{N} \left[\sum_{k'=1}^{J} X_{o'k'}\gamma^P_{ok,o'k'} + \gamma^C_{ok,o'} I_{o'}\right] \tag{5.30}$$

Goods market clearing.

$$X_{o'k',ok} = X_{ok}\gamma^P_{o'k',ok} = \xi_{o'k',ok} \sum_{o'=1}^{N} \left[\sum_{k'=1}^{J} \left(X_{o'k'} \times \gamma^P_{ok,o'k'}\right) + \gamma^C_{ok,o'} I_{o'}\right], \tag{5.31}$$

[9] $C_2 = (1+\theta)^{\frac{\theta-\rho}{\theta}} \left(\frac{1+\theta}{\theta}\right)^{-\rho} \frac{\rho}{\rho-\theta}$

where $\xi_{o'k',ok} = \gamma^P_{o'k',ok}(1-\alpha_{ok})$. Recall $I'_o = \sum_{k'} w_{o'k} L_{o'k'} + D_{o'} + R_{o'}$ is the total income of workers in region o' and $\gamma^C_{ok,o'}$ is the expenditure share on final goods from region o sector k such that $Y^C_{ok,o'} = \gamma_{ok,o'} \times I_{o'}$. In particular, the trade deficit $D_{o'}$ is defined using the trade (im)balance condition in next section.

Trade (Im)balance Condition. The trade (im)balance condition is:

$$\sum_{o'=1}^{N}\sum_{k'=1}^{J}\left[\sum_{k=1}^{J} X_{ok}\gamma^P_{o'k',ok} + Y^C_{o'k',o}\right] - D_o = \sum_{o'=1}^{J}\sum_{k'=1}^{N}\left[\sum_{k=1}^{J} X_{o'k'}\gamma^P_{ok,o'k'} + Y^C_{ok,o'}\right] \tag{5.32}$$

5.9 Equilibrium

I now formally define the equilibrium under policies $\{t_j, t_{k,j}\}$ in this model.

Equilibrium in Levels: Definition 1. Given tax structure $(t_j, t_{ok,dj})$, L_d, a distribution of productivity $\mu_{ok}(\varphi)$, trade costs $(\kappa_{ok,dj}, \kappa_{ok,d})$ and a numeraire sector-region pair $(o,k) = (1,1)$, an equilibrium is a wage vector $\mathbf{w}$, price matrix $\mathbf{P}$ that satisfy firms' optimization problems, zero profit cutoff and free entry condition; and household's optimization problems, market clearing condition and trade (im)balance condition.

Equilibrium in Relative Changes: Definition 2 . Instead of solving for an equilibrium under the policy before the tax reform, I solve for changes in prices and wages after the tax reform is completed, which I define as an equilibrium in relative changes. In trade literature, "exact hat algebra" is used as an approach to analyze equilibrium adjustments in response to policy changes, such as tax reforms, by focusing on relative changes from a baseline. Instead of recalculating absolute values, variables are expressed in terms of percentage changes or "hats" denoted as $\hat{x} = x'/x$, where x' represents the variable post-reform and x the baseline. This simplifies complex model computations and isolates the impact of specific policy changes on trade flows, costs, and productivity. Specifically, I can match exactly the model to the data in a base year. Then I can identify the effect on equilibrium outcomes from a pure change in cross-sector tax change t_{kj} and value-added tax change t_j and can solve for the general equilibrium of the model without needing to estimate parameters that are difficult to identify in the data, as productivities z_{dj}, z^H_{dj} , and iceberg trade costs $\kappa_{ok,dj}$, $\kappa_{ok,d}$.

5.10 Counterfactual Equilibrium under the New Tax Structure

I now define the equilibrium of the model in the new tax structure relative to the old one.

Final Consumer Price Index. The consumer price index, combining (5.2), (5.13) and (5.14) becomes:

$$\hat{P}_d = \left[\sum_{o=1}^{N}\sum_{k=1}^{J}\gamma_{ok,d}^{C}\hat{c}_{ok}^{-\theta}\left(\hat{\varphi}_{ok,d}^{*}\right)^{\theta-\rho}\right]^{-\frac{1}{\theta}}. \tag{5.33}$$

Cost Bundle of Inputs. The cost bundle of inputs becomes:

$$\hat{c}_{dj} = (\hat{t}_j\hat{w}_d)^{\alpha_{dj}}\left[\hat{\hat{\mathbf{t}}}_{\mathbf{j}}\hat{P}_{dj}^{M}\right]^{1-\alpha_{dj}},$$

where

$$\hat{\hat{\mathbf{t}}}_{\mathbf{j}} = \begin{cases} 1 & \text{if } j \text{ is in a manufacturing sector} \\ \hat{t}_j & \text{if } j \text{ is in a service sector.} \end{cases}$$

The Average Productivity.

$$\hat{\varphi}_{ok,dj}^{*} = \hat{c}_{ok}^{\frac{1+\theta}{\theta}}\hat{t}_{ok,dj}\left(\hat{P}_{dj}^{M}\right)^{-1}\left(\hat{P}_{dj}^{M}\hat{Y}_{dj}^{M}\right)^{-\frac{1}{\theta}} \tag{5.34}$$

$$\hat{\varphi}_{ok,d}^{*} = \hat{c}_{ok}^{\frac{1+\theta}{\theta}}\hat{P}_d^{-1}\hat{I}_d^{-\frac{1}{\theta}} \tag{5.35}$$

Price Index of Intermediate Inputs. The price index of the intermediate input (based on Equation 5.6) becomes:

$$\hat{P}_{dj}^{M} = \left[\sum_{o=1}^{N}\sum_{k=1}^{J}\gamma_{ok,dj}^{P}\left(\hat{c}_{ok}\hat{t}_{ok,dj}\right)^{-\theta}\left(\hat{\varphi}_{ok,dj}^{*}\right)^{\theta-\rho} + \sum_{k=1}^{J}\gamma_{Hk,dj}^{P}(\hat{t}_j\hat{w}_d)^{-\theta}\right]^{-\frac{1}{\theta}}. \tag{5.36}$$

Assuming that in-house productivities remain unchanged before and after the reform, I only need to take a stance on the total expenditure on the various service inputs a firm produces in-house. Since $\hat{t}_j\hat{w}_d$ comes outside the summation, I only need to know the total weight $\sum_{k=1}^{J}\gamma_{Hk,dj}^{P}$.

Trade Shares. Bilateral trade share in intermediate good expenditure:

$$\hat{\gamma}^P_{ok,dj} = \left(\hat{\varphi}^*_{ok,dj}\right)^{\theta-\rho} \left[\frac{\hat{t}_{ok,dj}\hat{c}_{ok}}{\hat{P}^M_{dj}}\right]^{-\theta}. \tag{5.37}$$

Bilateral trade share in final consumption expenditure:

$$\hat{\gamma}^C_{ok,d} = \left(\hat{\varphi}^*_{ok,d}\right)^{\theta-\rho} \left[\frac{\hat{c}_{ok}}{\hat{P}_d}\right]^{-\theta}. \tag{5.38}$$

Market Clearing. Let X_{dj} denote intermediate expenditure by region d, sector j. The market clearing condition becomes:

$$(X_{o'k',ok})' = (X_{ok})'(\gamma^P_{o'k',ok})' = (\xi_{o'k',ok})' \sum_{o'} \left[\sum_{k'} (X_{o'k'})'(\gamma^P_{ok,o'k'})' + (\gamma^C_{ok,o'})'(I_{ok,o'})'\right], \tag{5.39}$$

where

$$(\xi_{o'k',ok})' = \xi_{o'k',ok} \cdot \hat{\gamma}^P_{ok,dj} \qquad (I_{o'})' = D_{o'} + \sum_{k=1}^{J} (t_k)'\hat{w}_{ok}w_{ok}L'_{ok}.$$

Trade Imbalance. The trade imbalance condition (based on Equation 5.32) becomes:

$$\sum_{o'=1}^{N}\sum_{k'=1}^{J}\left[\sum_{k=1}^{J}(X_{ok})'(\gamma^P_{o'k',ok})' + (Y^C_{o'k',o})'\right] - D_o = \sum_{o'=1}^{J}\sum_{k'=1}^{N}\left[\sum_{k=1}^{J}(X_{o'k'})'(\gamma^P_{ok,o'k'})' + (Y^C_{ok,o'})'\right]. \tag{5.40}$$

Firm Entry. The counterfactual number of incumbent firms in relative changes is given by:

$$\hat{M}_{ok} = \frac{\hat{r}_{ok}}{\hat{c}_{ok}}, \tag{5.41}$$

which is consistent with the idea that a region-sector will have more firms if there is higher demand or if firms become more efficient (i.e., lower marginal costs).

6 Estimation and Calibration of Model Parameters

6.1 Estimation of Trade Elasticity ρ

The trade elasticity parameters θ and ρ jointly play a central role in evaluating the impact of the tax reform on interregional trade and production networks. Following Krugman (1980), θ governs the intensive margin of trade by shaping how existing exporters adjust their sales in response to changes in trade frictions. In contrast, following Chaney (2008), ρ captures the overall trade elasticity, incorporating both the intensive margin (θ) and the extensive margin arising from firm selection into exporting. The difference $\rho - \theta$ reflects the strength of the extensive margin: the responsiveness of the number of exporting firms to changes in trade costs.

In this framework, a higher θ implies that existing exporters respond strongly to cost changes, while a higher $\rho - \theta$ implies that reductions in trade frictions trigger additional firm entry into export markets. While ρ determines the aggregate responsiveness of trade flows to tax-induced frictions, θ pins down relative markups and the distribution of sales across firms. Thus, welfare gains reflect the joint interaction of θ and ρ, with θ governing within-firm reallocation and $\rho - \theta$ governing across-firm reallocation.

The composite intermediate input price in equation (5.10) illustrates this point: price index composition depends on intensive margin θ, but the distribution of firm productivity above the cutoff—and therefore the extent of firm selection into exporting—depends on the gap between θ and ρ. From an empirical perspective, separately identifying ρ and θ is challenging because both parameters shape the responsiveness of trade flows to changes in trade costs, and they enter the gravity equations in complementary ways. To address this difficulty, I follow the literature in calibrating θ to values commonly estimated in firm-level studies, while directly estimating ρ using the policy-induced variation in cross-region-sector frictions generated by the B2V reform. This strategy allows me to pin down the overall trade elasticity relevant for welfare analysis while remaining consistent with the broader empirical trade literature.

To estimate ρ, I exploit the variation in tax rates on intermediate inputs induced by the tax reform. Adapting the new methodology by Caliendo and Parro (2015), I estimate the trade elasticity in the following way. Recall the gravity equation for producers:

$$X^P_{ok,dj} = C_2(t_{ok,dj}c_{ok}\kappa_{ok,dj})^{-\rho} f_{ok,dj}^{\frac{\theta-\rho}{\theta}} M^e_{ok}[Y^M_{dj}(P^M_{dj})^{1+\theta}]^{\frac{\rho}{\theta}} (\iota_{ok,dj})^{\frac{\rho(1+\theta)}{\theta}}$$

Consider three region-sector pairs indexed by ok, dj, and $d'j'$. Take the cross-product of trade flow from ok to dj, from dj to $d'dj'$, and from $d'j'$ to ok, and then the cross-product of the trade flow in the other direction, from ok to $d'j'$, from $d'j'$ to dj, and from dj to ok. Using the gravity equation above, I can calculate each expression and then take the ratio:

$$\frac{\hat{X}_{ok,dj}\hat{X}_{dj,d'j'}\hat{X}_{d'j',ok}}{\hat{X}_{ok,d'j'}\hat{X}_{d'j',dj}\hat{X}_{dj,ok}} = \left[\frac{\hat{t}_{k,j}\hat{t}_{j,j'}\hat{t}_{j',k}}{\hat{t}_{j,k}\hat{t}_{j',j}\hat{t}_{k,j'}}\right]^{-\rho} \tag{6.1}$$

Hatted variables represent ratios between 2012 and 2017. All the terms involving unobservable trade costs, prices, number of firms, fixed costs, total output, and other parameters (across space and time) are canceled out and I end up with a relation between bilateral trade and cross-sector tax friction. The estimation equation therefore is:

$$\frac{\hat{X}_{ok,dj}\hat{X}_{dj,d'j'}\hat{X}_{d'j',ok}}{\hat{X}_{ok,d'j'}\hat{X}_{d'j',dj}\hat{X}_{dj,ok}} = \exp\left[-\rho\log\left(\frac{\hat{t}_{k,j}\hat{t}_{j,j'}\hat{t}_{j',k}}{\hat{t}_{j,k}\hat{t}_{j',j}\hat{t}_{k,j'}}\right)\right] \times \epsilon, \tag{6.2}$$

where ϵ is the error term with $\mathbb{E}[\epsilon|\hat{t}_{k,j}\hat{t}_{j,j'}\hat{t}_{j',k}\hat{t}_{j,k}\hat{t}_{j',j}\hat{t}_{k,j'}] = 1$. The only identification restriction is that ϵ is assumed to be orthogonal to cross-sector tax rate.

My results yield an estimate of $\rho = 4.47$. This implies that a 1% increase in the trade frictions leads to approximately a 4.47% decrease in the trade share of intermediate goods, holding other factors constant. Table 5 shows the results for different sub-samples. This elasticity is in line with estimates from the international trade literature and suggests that trade in intermediate goods is quite responsive to changes in tax rates. In the following sections, I use this estimate in our general equilibrium model to quantify the effects of the tax reform on welfare, wages, and economic activity across Chinese provinces.

6.2 Calibration of Other Model Parameters

I use data from the newly published Inter-regional Input-Output Table (IRIO) in 2012 by Li et al. (2023) to calibrate the model to the base year. 2012 is the closest and best year for this study, as

Table 5. Trade Elasticity Estimates with Sub-samples

Sample	1/1271	2/1271	3/1271	4/1271	5/1271	6/1271	7/1271
ρ	-2.388***	-38.53*	-8.647*	-4.728*	-4.692*	-4.670*	-4.467*
	(0.383)	(20.03)	(4.414)	(2.453)	(2.445)	(2.443)	(2.395)
Observations	224,417	441,877	659,416	877,125	1,093,922	1,311,274	1,528,349

Notes: Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The exhaustive combination of the three cross-product of the trade flow would involve $[(N-1) \times J]^3 = 1271^3 \approx 2$ billion observations, which is computationally infeasible. Here, 1271 is the total number of combinations and I report a series of random samples of from 1 out of 1271 to 7 out of 1271 here.

the IRIO table is only available every five years, the year available before and after is 2007 and 2017. Table 6 provides a summary of the parameters calibrated, their definitions and sources in the model.

Table 6. Summary of Model Parameter Calibration

Parameter	Definition	Source
$\gamma^P_{ok,dj}$	Intermediate good trade share in 2012	IRIO 2012
$\gamma^C_{ok,dj}$	Final consumption good trade share in 2012	IRIO 2012
$\xi_{ok,dj}$	Intermediate input expenditure share in 2012	IRIO 2012
ζ^0_{dj}	Region-sector specific value-added share in 2012	IRIO 2012
$X^P_{ok,dj}, X^C_{ok,d}$	Expenditure on intermediate good and final good	IRIO 2012
$w_d L_{dj}$	Region-sector specific value-added in 2012	IRIO 2012
$t_{k,j}$	Cross-sector tax rate in 2012 and 2017	`www.gov.cn`
θ	Intensive Margin Elasticity of Substitution	e.g. Benguria et al. (2023)

Notes: This table summarizes the model parameters calibrated using data from the 2012 Inter-regional Input-Output Table (IRIO) and additional sources.

The intermediate good trade share ($\gamma^P_{ok,dj}$) and final consumption good trade share ($\gamma^C_{ok,dj}$) are based on bilateral trade flows reported in the IRIO 2012. The intermediate input expenditure share ($\xi_{ok,dj}$) and the region-sector specific value-added share (ζ^0_{dj}) are derived from the IRIO 2012, capturing the share of expenditure on intermediate goods and the value-added component in each region-sector pair. The cross-sector tax rates ($t_{k,j}$) reflect the differential tax rates observed before and after the Business to Value-Added Tax (B2V) reform, obtained from official government data. The expenditure on intermediate ($X^P_{ok,dj}$) and final goods ($X^C_{ok,d}$) in the base year are used to match the observed trade flows in the model. The wage and labor input ($w_d L_{dj}$) are calibrated based on region-sector specific data, accounting for value-added contributions from labor across different regions and sectors. These parameters are crucial for replicating the baseline economic conditions prior to the B2V reform and serve as the foundation for evaluating counterfactual scenarios in the quantitative model. θ is the intensive margin for the trade elasticity.

7 Quantifying the Welfare and Growth Effect from the Reform

In this section, I evaluate the welfare effects of the business-to-value-added tax reform using the calibrated spatial general equilibrium model. The calibration strategy ensures that the model exactly replicates baseline conditions, including existing trade imbalances across provinces. These trade deficits are treated as exogenous in the counterfactual analysis, meaning that changes in tax structure affect the composition and efficiency of trade flows but not aggregate regional trade balances. I use 2012 as the base year, corresponding to the initial implementation of the reform, and conduct two complementary counterfactual exercises that allow me to decompose the reform's total impact into distinct economic channels.

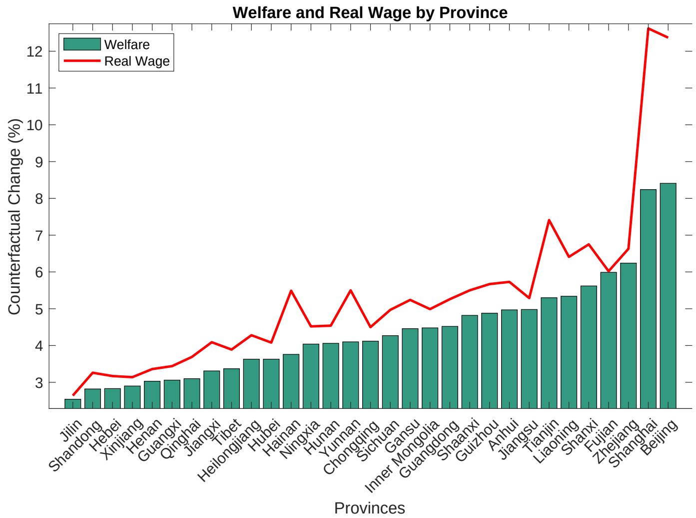
Counterfactual 1: Total Reform Effect. The first exercise measures the complete impact of the tax reform by implementation the actual changes in production structure that occurred between 2012 and 2017 across China's 31 provinces, while holding any other variables constant for the rest of the world. This counterfactual captures both major effects of the reform: (1) the removal of cross-sector tax distortions that previously impeded intermediate transactions, and (2) the reduction in overall tax burden, as aggregate tax revenue decreased following the reform.

Counterfactual 2: Revenue-Neutral Reform. The second exercise isolates the efficiency gains from eliminating tax distortions by implementing a revenue-neutral version of the reform. I first remove the cross-sector tax frictions as in the main reform, but then proportionally adjust the tax rates across all region-sector pairs to maintain the aggregate tax revenue at the baseline level. This approach allows me to separate the welfare gains from the improved efficiency from those arising from simple tax burden reduction.

7.1 Welfare effects from the Tax Reform

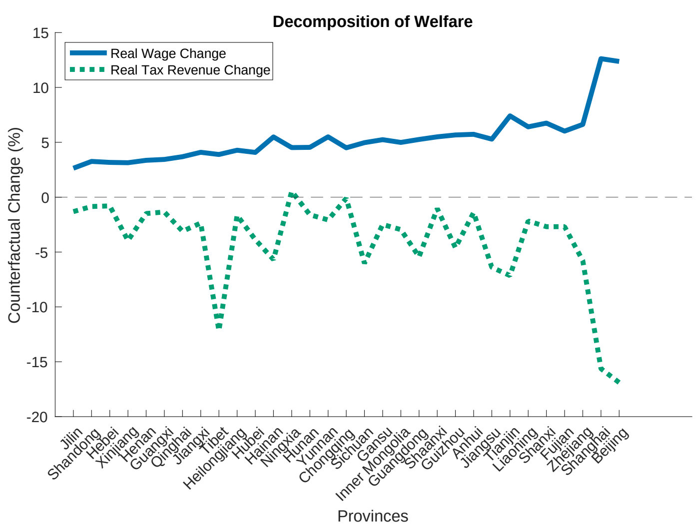
To quantitatively account for the choice of vertical integration, I adopt a realistic assumption that expenditures on in-house services are included as part of the labor value-added in the input-output table. In my counterfactual analysis, I simulate a range of values for the share of in-house production

Figure 4. Counterfactual Welfare and Real Wage Across Space



Notes: This figure presents the estimated average welfare effects and real wage changes across 31 mainland Chinese provinces following the Business to Value-Added Tax (B2V) reform. The figure demonstrates an average welfare increase of 4.3%, with a corresponding average real wage increase of 5.3%. The results highlight the regional heterogeneity in the impact of the reform, with provinces such as Beijing and Shanghai, which have a high share of service sectors, experiencing significant real wage growth but a considerable decline in tax revenue. The model assumes that all variables outside China remain constant at their 2012 levels. The differences in regional impacts are attributed to variations in the economic structure and the relative size of the service sectors in each province. The model also incorporates the effects of varying in-house service shares, illustrating that higher in-house shares in manufacturing sectors lead to greater gains in real wages and firm entry rates. The main result shown here uses the level of in-house share that matches the empirical result of firm entry.

Figure 5. Counterfactual Welfare Decomposition



Notes: This figure decomposes the welfare changes across Chinese provinces resulting from the B2V reform into two key components: changes in real wages and real tax revenue. The decomposition assumes that regional trade imbalances remain exogenous. The figure shows that the wage effect is consistently positive across all provinces, indicating increased labor income. At the same time, the effect on tax revenue decreases across provinces with different magnitudes, reflecting differences in their economic structures and the relative importance of service sectors. The overall welfare change in each province reveals substantial regional disparities in the reform's impact. Provinces with a larger share of service sectors, like Beijing and Shanghai, show more welfare increase compared to other regions, aligning with discussions in Sections 5 and 7 of the paper.

in total expenditures, denoted as $\xi_{Hk,dj}$. This value varies from 0% to 80% of the labor value added: for example, 80% means that 80% of value added are used for in-house production; if the value added represents 30% of the total expenditure, this means that $24\% = 80\% \times 30\%$ of the total expenditure is spent for in-house production. I assume that all manufacturing sectors spent the same percentage of their value added on in-house production, and since the labor value added shares vary across sectors and space this modeling choice still preserves certain flexibility and heterogeneity.

This also approach ensures that the total value-added share from both in-house service labor (as defined in Equation (5.5)) and regular production labor (as defined in Equation (A2)) sum up to the actual value-added share in the input-output table observed in the base year. Notably, there is no need to differentiate between types of services ($k \in J$) used for in-house production, as it is assumed that all in-house workers receive a uniform wage within each location, regardless of service type. Therefore, I only need to take a stance on the aggregate in-house share of all services used by a region-sector pair and reduce the additional ad-hoc assumptions required if we were to distinguish the type of services they produce.

I now report the result of main counterfactual analysis using calibrated parameters that match the empirical evidence. The baseline vertical integration assumptions are as follows: manufacturing sectors allocate 18% of their value added to in-house service production, while service sectors are restricted from vertical integration. This asymmetric treatment reflects the empirical reality that service sectors typically consist of smaller, more specialized firms that lack both the scale economies and diversification incentive that drive vertical integration in manufacturing.

Since value added shares vary across region-sector pairs due to differences in production technologies and local conditions, the actual expenditure on in-house production varies corresponding across sectors and space, even though the percentage of value added allocated to vertical integration remains constant across manufacturing industries.

This calibration is derived through a moment-matching procedure that ensures consistency between model predictions and empirical evidence. Specifically, I select the in-house share that replicates the firm entry patterns observed in the data using the method of Callaway and Sant'Anna (2021). Figure 6 displays the model prediction of firm entry for the service and manufacturing sectors across space. Figure 4 presents the welfare effects resulting from the change in tax structure due to the B2V reform, holding all other external variables constant at their 2012 levels. The reform leads to an average welfare increase of 4.3% across all 31 mainland Chinese provinces, accompanied by an average real wage rise of 5.3%. The aggregate tax revenue decreases by 7%. Notably, regions such as Beijing and Shanghai, which have a substantial share of the service sector, experience significant growth in real wages, albeit with a marked reduction in tax revenue. Provinces with a higher share of service sectors on average experience a higher real wage growth and a bigger drop in tax revenue, and a joint effect of higher welfare increase.

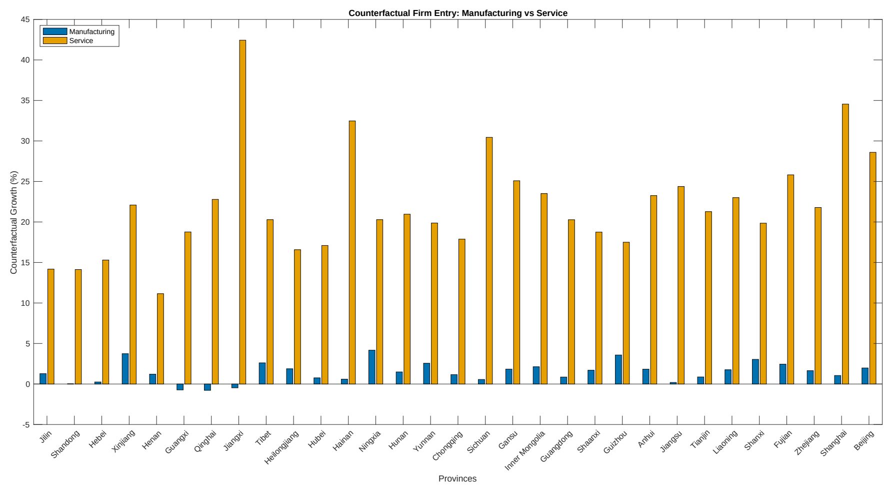


Figure 6. Firm Entry: Service vs. Manufacturing

Notes: This figure displays the model implied counterfactual increase in number of firms for the service and manufacturing sectors across space. The corresponding change in the number of firms for each region-sector pair is $\hat{M}_{ok} = \frac{\hat{r}_{ok}}{\hat{c}_{ok}}$. I calculate the province-level firm growth by multiplying the change in the number of firms in each region-sector by the number of firms in the base year (2012) to get the counterfactual number of firms in level, then calculate the regional growth. All regions experience an increase in service growth, while the number of manufacturing firms decreases in some regions with a small magnitude.

Recall that $\text{Welfare}_d = \frac{wL_d+D_d+T_d}{P_d} = \frac{wL_d}{P_d} + \frac{D_d}{P_d} + \frac{T_d}{P_d}$, comprising real wages, real trade balances, and real tax revenue. Figure 5 decomposes the welfare changes across Chinese provinces into these key components, focusing on the two most important: changes in real wages and real tax revenue. The trade balance component is held exogenous therefore not shown here in the analysis, while price index changes are already incorporated into the real wages and real revenue effects through their impact on purchasing power.

The decomposition reveals distinct patterns across space. Real wage effects are consistently positive throughout China, reflecting increased labor income as the reform enhances economic expansion. In contrast, all provinces experience a decrease in real tax revenue. The overall welfare change for each province is the sum of these three components, revealing substantial regional disparities in the reform's impact. These variations are due to differences in economic structure, particularly the relative size of the service sector.

The results presented here assume that local tax revenues are fully rebated to local residents — a simplifying assumption that isolates the direct economic effects of the reform without interregional fiscal redistribution. Under this assumption, the reform increases regional inequality, as service-intensive provinces, which tend to be wealthier, experience larger welfare gains.

However, this distributional outcome is not inevitable. The central government maintains substantial capacity to modify these effects through interregional fiscal transfers. If tax revenue were pooled nationally rather than rebated locally, the welfare distribution would differ. Therefore, Figure 4 illustrates the potential welfare gains under local revenue retention, the actual distributional consequences depend critically on the fiscal arrangements chosen by policymakers. These results should be interpreted as showing the economic efficiency gains rather than the final welfare outcomes after any redistributive interventions.

7.2 Lower Tax Burden or Lower Frictions in Production Networks?

To disentangle the welfare effects arising from tax burden reduction versus efficiency improvements, I conduct a revenue-neutral counterfactual analysis. This exercise eliminates cross-sector tax frictions (as in the main analysis) but proportionally increases tax rates across all sectors and regions to maintain aggregate tax revenue at baseline levels.

The revenue-neutral reform generates substantial welfare gains of 4.06% and real wage increases of 4.25%—representing 94% and 81% of the total gains from the actual reform, respectively. For comparison, the main counterfactual (which includes a 7% reduction in aggregate tax revenue) produces welfare gains of 4.3% and real wage increases of 5.3%. This decomposition reveals that the vast majority of welfare improvements stem from eliminating across-region-sector trade frictions rather than simply reducing the tax burden. Even when government revenue is held constant, the reform delivers more than 90% of its efficiency benefits.

The smaller gap between revenue-neutral and actual reform effects for welfare (0.24 percentage points) compared to real wages (1.05 percentage points) reflects the fact that maintaining tax revenue requires higher tax rates, which reduces take-home wages but provides offsetting benefits through increased government transfer to residents. The efficiency gains from eliminating cross-region-sector distortions remain largely intact regardless of the overall higher tax level. This finding has important implications for tax policy design, suggesting that structural reforms to eliminate distortions can generate substantial welfare gains while maintaining the tax revenue.

7.3 Different Initial Level of In-house Production

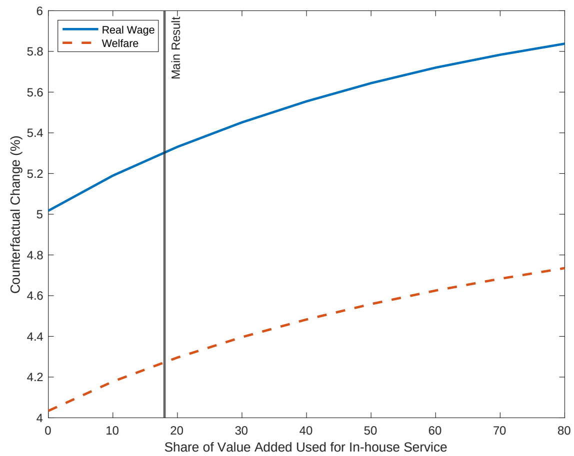
While I report baseline in-house service shares as 18% of value-added for manufacturing sectors and 0% for service sectors, Figure 7 also illustrates how average counterfactual outcomes vary with different manufacturing sector in-house shares. I focus specifically on adjusting the in-house share for manufacturing because this sector typically consists of larger firms with greater potential for vertical integration. Thus, changes in this sector's vertical integration level offer more realistic implications.

Higher initial in-house share in the manufacturing sector are associated with more substantial gains in real wages, firm entry, and overall welfare following the reform. This relationship suggests that economies with greater initial inefficiencies, indicated by higher levels of vertical integration, have more room for improvement, and therefore lead to more significant observed gains. Panel (A) in Figure 7 demonstrates that an economy with no in-house service production (0% of value-added) corresponds to a 5.02% rise in real wages and a 4.03% increase in welfare. In contrast, if we assume a 80% initial reliance on in-house services (indicating substantial vertical integration), ii results in a 5.84% increase in real wages rise 4.74% in welfare. Panel (B) shows a range of possible firm entry growth with different levels of in-house service shares. To put this comparison into perspective, if we do not allow any vertical integration, we will miss 6 percentage of the welfare improvement (4.3% compared to 4.02%).

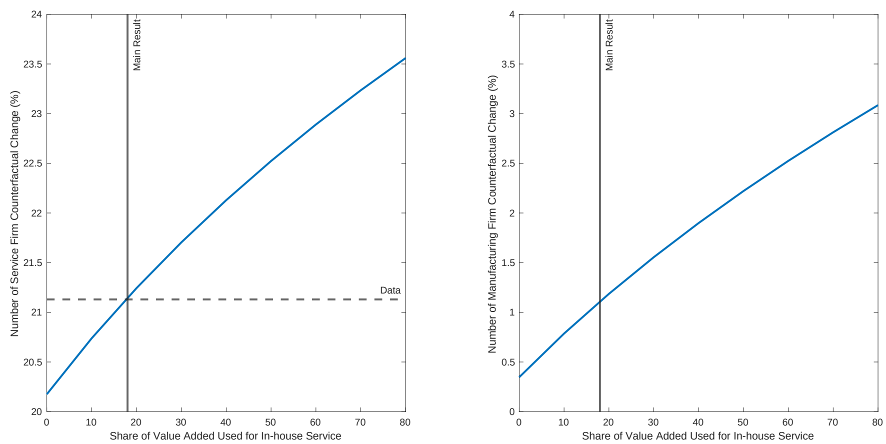
Higher in-house service shares in the manufacturing sector correlate with greater increases in real wages, growth of new firms, and welfare. This suggests that if the economy has more margin to be improved (more inefficient vertical integration to begin with), the more actual improvement we observe after the reform. Panel (A) in Figure 7 shows that a 0% of value added used by in-house production corresponds to a 4.04% increase in real wages and 3.05% increase in welfare, while a of value added used by in-house production corresponds to a 4.73% increase in real wages and 3.61% increase in welfare. Panel (B) shows how I match the result to the empirical setting.

It is crucial to recognize that a higher in-house share does not inherently indicate more favorable economic conditions simply because it leads to greater counterfactual growth. The in-house share

Figure 7. In-house Share and Counterfactual Change



(A) Welfare and Real Wage



(B) Counterfactual Firm Entry

Notes: This figure illustrates the relationship between initial in-house service share and various economic outcomes in the counterfactual scenario following the Business to Value-added Tax (B2V) reform. The x-axis represents different levels of initial in-house service share in the manufacturing sector, ranging from 0 to 0.1 (0% to 10%). Panel A shows the impacts on welfare (left y-axis) and real wages (right y-axis), while Panel B displays the effects on firm entry and real output (both on the left y-axis).

can be expressed as the proportion of intermediate cost minus the market purchases. Different initial in-house shares in the baseline economy reflect variations in underlying economic conditions and the relative costs faced by firms. A higher initial in-house share suggests that in-house production is relatively more cost-effective for firms compared to sourcing from the market. This scenario often results from a high tax burden or cascading taxes on service sectors, as firms attempt to minimize their overall tax liabilities by internalizing production. Conversely, a low initial in-house share indicates that the service sector is highly productive, allowing it to offer intermediate goods at a lower cost despite the presence of distortionary taxes. In this case, firms find it more economical to outsource rather than internalize their production.

In such an economic state, I observe a smaller increase in new firm entry and output growth in the counterfactual equilibrium. This is because the primary mechanism—labor shifting from in-house production to other firms in the market—plays a weaker or negligible role. Nevertheless, this economic state represents a more efficient equilibrium to begin with, all else being equal, as it reflects higher productivity in the service sector. Despite more economic activity and firm growth in the higher initial in-house share scenario, the overall welfare remains relatively unaffected. This indicates that while increased entry and growth are visible outcomes, the broader welfare implications are contingent on the efficiency of resource allocation and the cost structure in place.

8 Conclusion

This paper investigates how removing taxes on intermediate goods affects regional trade, production networks, and economic welfare by exploiting China's business-to-value-added tax reform as a natural experiment. The reform replaced business taxes with value-added taxes for service sectors and eliminated taxes on service inputs for manufacturing firms.

I find substantial positive effects across multiple dimensions. Firm entry increased by approximately 22% in treated service sectors, while firm revenues grew by similar magnitudes. Trade flows involving service sectors increased by 35.6% overall, with particularly strong growth of 47.9% within provinces. These results provide robust evidence that eliminating tax cascading stimulated economic activity and enhanced market integration.

To quantify welfare effects and understand the underlying mechanism, I develop a spatial general equilibrium model that incorporates firms' vertical integration decision — a key margin through which business taxes distortion production choices. The model reveals that the reform generated aggregate welfare gains of 4.3% and real wage increases of 5.3%. Crucially, a revenue-neutral counterfactual shows that 94% of these welfare gains stem from eliminating cross-region-sector trade frictions rather than reducing tax burdens.

The analysis reveals important heterogeneity in reform impacts. Regions with larger service sectors experienced greater benefits, though they also faced larger reductions in tax revenue. Manufacturing sectors with higher initial vertical integration saw larger productivity gains as they shifted from inefficient in-house production to market-based procurement of services.

These findings provide key insights for tax policy design. First, the structure of taxation matters more than level. Eliminating cascading taxes generates substantial efficiency gains event when overall tax rates are adjusted to maintain revenue. Second, tax reforms can have distributional consequences that may require complementary policies to address regional heterogeneity. Third, the benefits of efficient tax system extend beyond fiscal effects to include improved resource allocation and stronger production networks.

The analytical framework developed here could also provide a foundation for future research examining how firms adapt their organizational boundaries under different economic conditions, such as supply uncertainty, trade policies, or technological changes.

References

Acemoglu, Daron, Vasco M. Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi, "The Network Origins of Aggregate Fluctuations," *Econometrica*, 2012, *80*, 1977–2016.

Antràs, Pol, "Firms, contracts, and trade structure," *The Quarterly Journal of Economics*, 2003, *118* (4), 1375–1418.

Asker, John, Allan Collard-Wexler, and Jan De Loecker, "Dynamic Inputs and Resource (Mis)Allocation," *Journal of Political Economy*, 2014, *122*, 1013–1063.

Basu, Susanto, "Intermediate Goods and Business Cycles: Implications for Productivity and Welfare," *American Economic Review*, 1995, *85*, 512–531.

Bau, Natalie and Adrien Matray, "Misallocation and capital market integration: Evidence from India," *Econometrica*, 2023, *91* (1), 267–310.

Benguria, Felipe, Felipe Saffie, and Sergio Urzua, "The Transmission of Commodity Price Super-Cycles," *The Review of Economic Studies*, 08 2023, *91* (4), 1923–1955.

Bernard, Andrew B., Andreas Moxnes, and Yukiko U. Saito, "Production Networks, Geography, and Firm Performance," *Journal of Political Economy*, 2019, *127* (2), 639–688.

Best, Michael Carlos, Anne Brockmeyer, Henrik Jacobsen Kleven, Johannes Spinnewijn, and Mazhar Waseem, "Production versus revenue efficiency with limited tax capacity: theory and evidence from Pakistan," *Journal of Political Economy*, 2015, *123*, 1311–1355.

Bigio, Saki and Jennifer La'O, "Distortions in Production Networks*," *The Quarterly Journal of Economics*, 05 2020, *135* (4), 2187–2253.

Caliendo, Lorenzo and Fernando Parro, "Estimates of the Trade and Welfare Effects of NAFTA," *The Review of Economic Studies*, 2015, *82*, 1–44.

_ , _ , and Aleh Tsyvinski, "Distortions and the structure of the world economy," *American Economic Journal: Macroeconomics*, 2022, *14* (4), 274–308.

_ , _ , and _ , "Distortions and the Structure of the World Economy," *American Economic Journal: Macroeconomics*, 2022, *14* (4), 274–308.

Callaway, Brantly and Pedro H.C. Sant'Anna, "Difference-in-Differences with multiple time periods," *Journal of Econometrics*, December 2021, *225* (2), 200–230. .

Carvalho, Vasco and Xavier Gabaix, "The Great Diversification and Its Undoing," *American Economic Review*, 2013, *103*, 1697–1727.

Chaney, Thomas, "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *American Economic Review*, 2008, *98* (4), 1707–1721.

Ciccone, Antonio, "Input Chains and Industrialization," *Review of Economic Studies*, 2002, *69*, 565–589.

Coase, R., "The Nature of the Firm," *Economica*, 1937, *4* (16), 386–405.

di Giovanni, Julian, Andrei A. Levchenko, and Isabelle Mejean, "Firms, Destinations, and Aggregate Fluctuations," *Econometrica*, 2014, *82*, 1303–1340.

_ , _ , and _ , "The Micro Origins of International Business-Cycle Comovement," *American Economic Review*, 2018, *108*, 82–108.

Donaldson, Dave, "The Gains from Market Integration," *The Annual Review of Economics*, 2015, *7*, 619–647.

— **and Richard Hornbeck**, “Railroads and American economic growth: A "market access" approach,” *The Quarterly Journal of Economics*, 2016, *131* (2), 799–858.

Eaton, Jonathan and Samuel Kortum, “Technology, Geography, and Trade,” *Econometrica*, 2002, *70*, 1741–1779.

Faber, Benjamin, “Trade integration, market size, and industrialization: evidence from China’s National Trunk Highway System,” *The Review of Economic Studies*, 2014, *81* (3), 1046–1070.

Fajgelbaum, Pablo D, Eduardo Morales, Juan Carlos Suárez Serrato, and Owen Zidar, “State taxes and spatial misallocation,” *The Review of Economic Studies*, 2019, *86* (1), 333–376.

Fan, Jingting, “Internal Geography, Labor Mobility, and the Distributional Impacts of Trade,” *American Economic Journal: Macroeconomics*, July 2019, *11* (3), 252–288.

Foerster, Andrew T., Pierre-Daniel G. Sarte, and Mark W. Watson, “Sectoral versus Aggregate Shocks: A Structural Factor Analysis of Industrial Production,” *Journal of Political Economy*, 2011, *119*, 1–38.

Gabaix, Xavier, “The granular origins of aggregate fluctuations,” *Econometrica*, 2011, *79* (3), 733–772.

Gadenne, Lucie, Tushar K Nandi, and Roland Rathelot, “Taxation and Supplier Networks: Evidence from India,” CEPR Discussion Paper 13971, CEPR Discussion Papers 2019.

Gopinath, Gita, Şebnem Kalemli-Özcan, Loukas Karabarbounis, and Carolina Villegas-Sanchez, “Capital allocation and productivity in South Europe,” *The Quarterly Journal of Economics*, 2017, *132* (4), 1915–1967.

Hansen, Benjamin, Keaton Miller, and Caroline Weber, “Vertical integration and production inefficiency in the presence of a gross receipts tax,” *Journal of Public Economics*, 2022, *212*, 104693.

— , — , and — , “Vertical integration and production inefficiency in the presence of a gross receipts tax,” *Journal of Public Economics*, 2022, *212*, 104693.

Hsieh, Chang-Tai and Peter J. Klenow, “Misallocation and Manufacturing TFP in China and India,” *The Quarterly Journal of Economics*, November 2009, *124* (4), 1403–1448. Published: 01 November 2009.

Jones, Charles I., “Intermediate Goods and Weak Links in the Theory of Economic Development,” *American Economic Journal: Macroeconomics*, 2011, *3*, 1–28.

Krugman, Paul, “Scale economies, product differentiation, and the pattern of trade,” *American Economic Review*, 1980, *70* (5), 950–959.

Li, Shantong, Chen Pan, Jianwu He, and Jie Chen, *China’s Inter-Provincial Input-Output Tables for 2012 and 2017: Compilation and Applications*, Beijing: Economic Science Press, 2023.

Melitz, Marc J., “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity,” *Econometrica*, 2003, *71* (6), 1695–1725.

Phillips, A. and M. Ibaid, “The impact of imposing sales taxes on business inputs,” Ernst and Young LLP and the Council on State Taxation, Washington DC 2019.

Redding, Stephen J and Esteban Rossi-Hansberg, “Quantitative spatial economics,” *Annual Review of Economics*, 2017, *9*, 21–58.

Silva, JMC Santos and Silvana Tenreyro, “The log of gravity,” *The Review of Economics and statistics*, 2006, *88* (4), 641–658.

Tombe, Trevor and Xiaodong Zhu, “Trade, migration, and productivity: A quantitative analysis of China,” *American Economic Review*, 2019, *109* (5), 1843–72.

Williamson, O. E., "The Vertical Integration of Production: Market Failure Considerations," *American Economic Review*, 1971, *61* (2), 112–123.

Xing, Jing, Katarzyna Bilicka, Xipei Hou, and Sepideh Raei, "How distortive are turnover taxes? Evidence from China," *Journal of Development Economics*, October 2024, *171*, 103332.

Yi, Kei-Mu, "Can Vertical Specialization Explain the Growth of World Trade?," *Journal of Political Economy*, 2003, *111*, 52–102.

Zilibotti, Fabrizio, "Growing and Slowing Down Like China," *Journal of the European Economic Association*, 2017, *15*, 943–988.

A Derivations

A.1 Marginal Cost

The marginal cost c_{dj} for a firm with productivity $\varphi = 1$ is obtained by solving the following cost minimization problem subject to the production constraint $y_{dj} \geq 1$:

$$\min_{\ell_{dj},\, Y_{dj}^M} \; \tilde{w}_{dj}\, \ell_{dj} \; + \; \tilde{P}_{dj}^M(\omega)\, Y_{dj}^M(\omega), \tag{A1}$$

where the effective input prices are

$$\tilde{w}_{dj} \equiv t_j w_d, \qquad \tilde{P}_{dj}^M(\omega) \equiv \Big(1 + \mathbf{1}[S=1]\mathbf{1}[R=0](t_j - 1)\Big) P_{dj}^M(\omega).$$

The production technology is Cobb–Douglas:

$$y_{dj}(\omega) = \varphi\, \ell_{dj}(\omega)^{\alpha_{dj}} \left(Y_{dj}^M(\omega)\right)^{1-\alpha_{dj}}. \tag{A2}$$

First-order conditions imply the cost share rules:

$$\frac{\tilde{w}_{dj}\, \ell_{dj}}{\tilde{P}_{dj}^M Y_{dj}^M} = \frac{\alpha_{dj}}{1 - \alpha_{dj}}.$$

Thus the unit cost function is

$$c_{dj}(\omega) = \kappa(\alpha_{dj})\, \tilde{w}_{dj}^{\alpha_{dj}} \left(\tilde{P}_{dj}^M(\omega)\right)^{1-\alpha_{dj}}, \qquad \kappa(\alpha) \equiv \alpha^{-\alpha}(1-\alpha)^{-(1-\alpha)}. \tag{A3}$$

The input demands for one unit of output are:

$$\ell_{dj}^*(\omega) = \frac{\alpha_{dj}\, c_{dj}(\omega)}{\tilde{w}_{dj}} = \alpha_{dj}\, \kappa(\alpha_{dj})\, \tilde{w}_{dj}^{\alpha_{dj}-1} \left(\tilde{P}_{dj}^M(\omega)\right)^{1-\alpha_{dj}}, \tag{A4}$$

$$\left(Y_{dj}^M\right)^*(\omega) = \frac{(1-\alpha_{dj})\, c_{dj}(\omega)}{\tilde{P}_{dj}^M(\omega)} = (1-\alpha_{dj})\, \kappa(\alpha_{dj})\, \tilde{w}_{dj}^{\alpha_{dj}} \left(\tilde{P}_{dj}^M(\omega)\right)^{-\alpha_{dj}}. \tag{A5}$$

Substituting the effective prices, the marginal cost can be expressed as

$$c_{dj}(\omega) = \alpha_{dj}^{-\alpha_{dj}} (1-\alpha_{dj})^{-(1-\alpha_{dj})} (t_j w_d)^{\alpha_{dj}} \Big([1 + \mathbf{1}[S{=}1]\mathbf{1}[R{=}0](t_j{-}1)]\, P_{dj}^M(\omega)\Big)^{1-\alpha_{dj}}. \tag{A6}$$

A.2 Intermediate Price Index

The inner optimization problem associated with the choice of demanding intermediate inputs subject to the constraint $Y_{dj}^M \geq 1$ is:

$$\min_{Y_{ok,dj}^m(\omega')} \sum_{o=1}^{N+1} \sum_{k=1}^{J} \left[\int_{\omega' \in \Omega} p_{ok,dj}(\omega') Y_{ok,dj}^m(\omega') d\omega'\right] \tag{A7}$$

This gives the demand the following demand function:

$$Y_{ok,dj}^m(\omega', \omega) = \iota_{ok,dj} \left[\frac{p_{ok,dj}(\omega')}{P_{dj}^M}\right]^{-(1+\theta)} Y_{dj}^M(\omega) \tag{A8}$$

and the following price index:

$$P_{dj}^{M} = \left[\sum_{o=1}^{N+1} \sum_{k=1}^{J} \iota_{ok,dj} \int_{\Omega_{ok}} [p_{ok,dj}(\omega') t_{ok,dj}]^{-\theta} d\omega' \right]^{-\frac{1}{\theta}}, \tag{A9}$$

A.3 Profit Maximization and Price

Under the monopolistic market structure, a firm in region o, sector k with productivity φ sets its own price to other producers and consumers by solving the following profit maximization problem:

$$\begin{aligned} \max_{Y_{ok,dj}^{m}(\varphi), Y_{ok,d}^{C}(\varphi)} & \sum_{d=1}^{N} \sum_{j=1}^{J} \left(\int_{\Omega_{dj}} [p_{ok,dj}(\varphi) Y_{ok,dj}^{m}(\varphi) - \frac{c_{ok,dj}}{\varphi} Y_{ok,dj}^{m}(\varphi)] d\omega - f_{ok,dj} \right) \\ & \sum_{d=1}^{N} \left(p_{ok,d}(\varphi) Y_{ok,d}^{C}(\varphi) - \frac{c_{ok,d}}{\varphi} Y_{ok,d}^{C}(\varphi) - f_{ok,d} \right) \end{aligned} \tag{A10}$$

Substituting the demand of producers and consumers into the profit maximization problem (A10), the first order conditions imply firm in region o, sector k with productivity φ will set its price to producers *conditional on* exporting to pair d,j at:

$$p_{ok,dj}(\varphi) = \frac{1+\theta}{\theta} \frac{c_{ok} t_{ok,dj} \kappa_{ok,dj}}{\varphi} \tag{A11}$$

and *conditional on* exporting to consumers in d at:

$$p_{ok,d}(\varphi) = \frac{1+\theta}{\theta} \frac{c_{ok} t_{ok,dj} \kappa_{ok,d}}{\varphi} \tag{A12}$$

Total revenues of a firm with productivity φ from region o sector k to region d and sector j is from producers and consumers are:

$$\begin{aligned} x_{ok,dj}^{P}(\varphi) &= \int_{\Omega_{dj}} p_{ok,dj}(\omega') Y_{ok,dj}^{m}(\omega', \omega) d\omega \\ &= \int_{\Omega_{dj}} p_{ok,dj}(\omega') \left(\frac{t_{ok,dj} p_{ok,dj}(\omega')}{\iota_{ok,dj} P_{dj}^{M}} \right)^{-(1+\theta)} Y_{dj}^{M}(\omega) \\ &= p_{ok,dj}(\omega')^{-\theta} \left(\frac{t_{ok,dj}}{\iota_{ok,dj} P_{dj}^{M}} \right)^{-(1+\theta)} \int_{\Omega_{dj}} Y_{dj}^{M}(\omega) d\omega \\ &= \left(\frac{1+\theta}{\theta} \frac{c_{ok} t_{ok,dj} \kappa_{ok,dj}}{\varphi} \right)^{-\theta} \left(\iota_{ok,dj} P_{dj}^{M} \right)^{(1+\theta)} \int_{\Omega_{dj}} Y_{dj}^{M}(\omega) d\omega \\ &= \left(\frac{1+\theta}{\theta} \frac{c_{ok} t_{ok,dj} \kappa_{ok,dj}}{\varphi} \right)^{-\theta} \left(\iota_{ok,dj} P_{dj}^{M} \right)^{(1+\theta)} Y_{dj}^{M} \end{aligned} \tag{A13}$$

operating profits from all producers:

$$\pi^{P}_{ok,dj}(\varphi) = \int_{\Omega_{dj}} \left(p_{ok,dj}(\omega') - \frac{c_{ok} t_{ok,dj} \kappa_{ok,dj}}{\varphi} \right) Y^{m}_{ok,dj}(\omega', \omega) d\omega - f_{ok,dj} \tag{A14}$$
$$= \frac{1}{1+\theta} x^{P}_{ok,dj}(\varphi) - f_{ok,dj}$$

Zero cutoff productivity:

$$\varphi \geq \varphi^{*}_{ok,dj} \equiv \left[\frac{(1+\theta) c_{ok} f_{ok,dj} (\frac{1+\theta}{\theta} c_{ok} t_{ok,dj} \kappa_{ok,dj})^{\theta}}{(\iota_{ok,dj} P^{M}_{dj})^{1+\theta} Y^{M}_{dj}} \right]^{\frac{1}{\theta}} \tag{A15}$$

$$\begin{aligned} r^{C}_{ok,d}(\varphi) &= p_{ok,d}(\omega') Y^{C}_{ok,d}(\omega') \\ &= p_{ok,d}(\omega') \times s_{ok,d} p_{ok,d}(\omega')^{-(1+\theta)} P^{\theta}_{d} I_{d} \\ &= s_{ok,d} p_{ok,d}(\omega')^{\theta} P^{\theta}_{d} I_{d} \\ &= s_{ok,d} \left(\frac{1+\theta}{\theta} \frac{c_{ok} \kappa_{ok,d}}{\varphi} \right)^{-\theta} P^{\theta}_{d} I_{d} \end{aligned} \tag{A16}$$

operating profits from all consumers:

$$\pi^{C}_{ok,d}(\varphi) = \frac{1}{1+\theta} r^{C}_{ok,d}(\varphi) - f_{ok,d} \tag{A17}$$

$$\begin{aligned} \tilde{\varphi}^{\theta}_{ok,dj}(\varphi^{*}_{ok,dj}) &= \frac{1}{1 - G_{ok}(\varphi^{*}_{ok,dj})} \int_{\varphi^{*}_{ok,dj}}^{\infty} \varphi^{\theta} dG_{ok}(\varphi) \\ &= \frac{1}{1 - G_{ok}(\varphi^{*}_{ok,dj})} \frac{\rho}{\rho - \theta} (\varphi^{*}_{ok,dj})^{\theta - \rho} \\ &= \frac{\rho}{\rho - \theta} \left[\frac{(1+\theta) c_{ok} f_{ok,dj} (\frac{1+\theta}{\theta} c_{ok} t_{ok,dj} \kappa_{ok,dj})^{\theta}}{(\iota_{ok,dj} P^{M}_{dj})^{1+\theta} Y^{M}_{dj}} \right] \end{aligned} \tag{A18}$$

A.4 Gravity

We now aggregate firm-level demands into bilateral trade flows. Consider first intermediate goods. The total expenditure on varieties sourced from region–sector (o, k) and used as intermediate inputs by firms in (d, j) is

$$X^{P}_{ok,dj} = \int_{\Omega_{dj}} \left(\int_{\Omega_{ok}} p_{ok,dj}(\omega', \omega)\, Y^{m}_{ok,dj}(\omega', \omega)\, d\omega' \right) d\omega \tag{A19}$$
$$= \int_{\Omega_{dj}} \left(Y^{M}_{dj}(\omega) \big(\iota_{ok,dj} P^{M}_{dj}\big)^{1+\theta} \int_{\Omega_{ok}} p_{ok,dj}(\omega', \omega)^{-\theta}\, d\omega' \right) d\omega.$$

Here, $p_{ok,dj}(\omega')$ is the delivered price of variety ω' produced in origin (o, k) and sold to destination (d, j). It incorporates both marginal costs and sector–region specific trade frictions (introduced in the next subsection). The inner integral represents the effective contribution of all varieties from (o, k) to the CES bundle of intermediates in (d, j), while the outer integral aggregates over all firms in (d, j) that demand these inputs. The scaling term $\iota_{ok,dj}$ captures the relative weight of input (o, k) in the CES aggregator for (d, j).

Turning to final goods, aggregate expenditure by consumers in region d on varieties from origin (o, k) is

$$\begin{aligned} X^C_{ok,d} &= \int_{\Omega_{ok}} p_{ok,d}(\omega')\, Y^C_{ok,d}(\omega')\, d\omega' \\ &= s_{ok,d}\, P_d^{\theta} I_d \int_{\Omega_{ok}} p_{ok,d}(\omega')^{-\theta}\, d\omega'. \end{aligned} \tag{A20}$$

The first expression follows directly from integrating demand over all varieties produced in (o, k). The second equality uses CES demand: the expenditure share $s_{ok,d}$ reflects consumer preferences for goods from (o, k), while I_d is total income in region d and P_d is the final-good price index.

Finally, the implied aggregate factor share of intermediate input (o, k) in the production of (d, j) can be expressed as

$$\xi_{ok,dj} = (1 - \alpha_{dj})\, \gamma^P_{ok,dj}, \tag{A21}$$

where $\gamma^P_{ok,dj}$ is the bilateral trade share of (o, k) in total intermediate input expenditure of (d, j). This decomposition highlights that the effective contribution of an origin input depends both on technological input shares $(1-\alpha_{dj})$ and on trade shares that reflect equilibrium prices and frictions.

A.5 Aggregation of All Firms

To move from firm-level to aggregate trade flows, I now aggregate over the distribution of firm productivities. Let $\mu_{ok,dj}(\varphi)$ denote the (equilibrium) density of productivities for firms from region–sector (o, k) that export to (d, j), and let $M_{ok,dj}$ be the corresponding measure of active exporters. The average contribution of these firms to the price index in (d, j) is given by:

$$\begin{aligned} \int_{\Omega_{ok}} p_{ok,dj}(\omega)^{-\theta} d\omega &= \int_0^{\infty} M_{ok,dj} \left(\frac{1+\theta}{\theta} \frac{c_{ok} t_{ok,dj} \kappa_{ok,dj}}{\varphi} \right)^{-\theta} \mu_{ok,dj}(\varphi)\, d\varphi \\ &= M_{ok,dj} \left(\frac{1+\theta}{\theta} c_{ok} t_{ok,dj} \kappa_{ok,dj} \right)^{-\theta} \int_0^{\infty} \varphi^{\theta} \mu_{ok,dj}(\varphi)\, d\varphi \\ &= M^e_{ok} [1 - G_{ok}(\varphi^*_{ok,dj})] \left(\frac{1+\theta}{\theta} c_{ok} t_{ok,dj} \kappa_{ok,dj} \right)^{-\theta} \left(\frac{1}{1 - G_{ok}(\varphi^*_{ok,dj})} \int_{\varphi^*_{ok,dj}}^{\infty} \varphi^{\theta}\, dG_{ok}(\varphi) \right) \\ &= M^e_{ok} \left(\frac{1+\theta}{\theta} c_{ok} t_{ok,dj} \kappa_{ok,dj} \right)^{-\theta} \frac{\rho}{\rho - \theta} (\varphi^*_{ok,dj})^{\theta - \rho}. \end{aligned} \tag{A22}$$

For convenience, define the average productivity term:

$$\tilde{\varphi}_{ok,dj} \equiv \left[\int_0^{\infty} \varphi^{\theta} \mu_{ok,dj}(\varphi)\, d\varphi \right]^{1/\theta}.$$

Consumers. Similarly, the average contribution to the consumer price index in region d is:

$$\int_{\Omega_{ok}} p_{ok,d}(\omega)^{-\theta} d\omega = \int_0^\infty \left(\frac{1+\theta}{\theta} \frac{c_{ok}\kappa_{ok,d}}{\varphi} \right)^{-\theta} \mu_{ok,d}(\varphi)\, d\varphi, \tag{A23}$$

with $\tilde{\varphi}_{ok,d} = \left[\int_0^\infty \varphi^\theta \mu_{ok,d}(\varphi)\, d\varphi \right]^{1/\theta}$.

Intermediate-goods gravity. The aggregate expenditure on intermediate inputs sourced from (o,k) to (d,j) is

$$\begin{aligned} X^P_{ok,dj} &= \int_{\Omega_{dj}} \left(Y^M_{dj}(\omega)\, \iota_{ok,dj} \left(P^M_{dj}(\omega)\right)^{1+\theta} \int_{\Omega_{ok}} p_{ok,dj}(\omega', \omega)^{-\theta} d\omega' \right) d\omega \\ &= M_{ok,dj} \left(\frac{1+\theta}{\theta} t_{ok,dj} c_{ok} \kappa_{ok,dj} \right)^{-\theta} \tilde{\varphi}^\theta_{ok,dj} \iota^{1+\theta}_{ok,dj} (P^M_{dj})^{1+\theta} \int_{\Omega_{dj}} Y^M_{dj}(\omega)\, d\omega. \end{aligned} \tag{A24}$$

Normalizing by total intermediate expenditure yields the bilateral intermediate-goods trade share:

$$\begin{aligned} \gamma^P_{ok,dj} &= \frac{X^P_{ok,dj}}{P^M_{dj} Y^M_{dj}} \\ &= \frac{M_{ok,dj} \left(\frac{1+\theta}{\theta} t_{ok,dj} c_{ok} \kappa_{ok,dj} \right)^{-\theta} \tilde{\varphi}^\theta_{ok,dj} \iota^{1+\theta}_{ok,dj}}{(P^M_{dj})^{-\theta}}. \end{aligned} \tag{A25}$$

Final-goods gravity. For final consumption in region d, aggregate demand takes the form:

$$\begin{aligned} X^C_{ok,d} &= M_{ok,d} \left(\frac{1+\theta}{\theta} c_{ok} \kappa_{ok,d} \right)^{-\theta} \tilde{\varphi}^\theta_{ok,d} s_{ok,d} P^\theta_d I_d \\ &= C_3 (c_{ok}\kappa_{ok,d})^{-\rho} f_{ok,d}^{\frac{\theta-\rho}{\theta}} M_{ok} (I_d P^\theta_d)^{\frac{\rho}{\theta}} s_{ok,d}^{\frac{\rho}{\theta}}. \end{aligned} \tag{A26}$$

Normalizing by total expenditure yields the bilateral final-goods trade share:

$$\begin{aligned} \gamma^C_{ok,d} &= \frac{X^C_{ok,d}}{I_d} \\ &= \frac{M_{ok,d} \left(\frac{1+\theta}{\theta} c_{ok} \kappa_{ok,d} \right)^{-\theta} \tilde{\varphi}^\theta_{ok,d} s_{ok,d}}{\sum_{o=1}^N \sum_{k=1}^J M_{ok,d} \left(\frac{1+\theta}{\theta} c_{ok} \kappa_{ok,d} \right)^{-\theta} \tilde{\varphi}^\theta_{ok,d} s_{ok,d}}. \end{aligned} \tag{A27}$$

Discussion. These expressions show how firm heterogeneity aggregates into standard gravity equations. The Pareto distribution collapses the integrals into closed-form terms involving $(\rho - \theta)$ and the cutoff productivity $\varphi^*_{ok,dj}$. As a result, aggregate trade shares depend on: (i) the number of entrants M_{ok}, (ii) marginal costs c_{ok} and trade frictions $\kappa_{ok,dj}$, (iii) the selection margin governed by $\varphi^*_{ok,dj}$, and (iv) the preference and technology weights $\iota_{ok,dj}, s_{ok,d}$. This aggregation step is crucial, as it links the micro-level productivity distribution with the macro-level trade shares that drive the general equilibrium of the model.

A.6 Firm Entry and Zero Profit Condition

A.6.1 Selection into Exporting

In order to determine the equilibrium number of entrants $M_{ok,dj}$ and the average productivity of entrants $\tilde{\varphi}_{ok,dj}$, I have to consider the export decisions of firms. A firm from region o and sector k with productivity φ conditional on producing will export to producers in region d, sector j and consumers in region d if and only if:

$$\pi^P_{ok,dj}(\varphi) \geq c_{ok} f_{ok,dj}, \;\; \pi^C_{ok,dj}(\varphi) \geq c_{ok} f_{ok,d}$$

This implies the cut-off productivity for exporting from o,k to d,j is:

$$\frac{1}{1+\theta}\left(\frac{1+\theta}{\theta}\frac{c_{ok} t_{ok,dj}\kappa_{ok,dj}}{\varphi}\right)^{-\theta}\left(\iota_{ok,dj}P^M_{dj}\right)^{(1+\theta)} Y^M_{dj} \geq f_{ok,dj}c_{ok}$$

$$\varphi \geq \varphi^*_{ok,dj} \equiv \left[\frac{(1+\theta)c_{ok}f_{ok,dj}(\frac{1+\theta}{\theta}c_{ok}t_{ok,dj}\kappa_{ok,dj})^{\theta}}{(\iota_{ok,dj}P^M_{dj})^{1+\theta}Y^M_{dj}}\right]^{\frac{1}{\theta}} \tag{A28}$$

Hence, only firms that are sufficiently productive will find it profitable to incur the fixed cost of exporting to a destination region-sector pair.

The cut-off productivity for exporting from o, k to d to consumers is:

$$s_{ok,d}\frac{1}{1+\theta}\left(\frac{1+\theta}{\theta}\frac{c_{ok}\kappa_{ok,d}}{\varphi}\right)^{-\theta} P^{\theta}_d I_d \geq f_{ok,d}c_{ok}$$

$$\varphi \geq \varphi^*_{ok,d} \equiv \left[\frac{(1+\theta)c_{ok}f_{ok,d}(\frac{1+\theta}{\theta}c_{ok}\kappa_{ok,d})^{\theta}}{s_{ok,d}P^{\theta}_d I_d}\right]^{\frac{1}{\theta}} \tag{A29}$$

Together, equations (A28) and (A29) allow us to determine the "average" productivity of producers selling from o,k to producers in d,j:

$$\tilde{\varphi}_{ok,dj} = \left(\frac{1}{1-G_{ok}(\varphi^*_{ok,dj})}\int_{\varphi^*_{ok,dj}}^{\infty}\varphi^{\theta}dG_{ok}(\varphi)\right)^{\frac{1}{\theta}} \tag{A30}$$

"average" productivity of producers selling from o,k to consumers in d:

$$\tilde{\varphi}_{ok,d} = \left(\frac{1}{1-G_{ok}(\varphi^*_{ok,d})}\int_{\varphi^*_{ok,d}}^{\infty}\varphi^{\theta}dG_{ok}(\varphi)\right)^{\frac{1}{\theta}} \tag{A31}$$

The density of firms selling from o, k to d, j is:

$$M_{ok,dj} = M^e_{ok}(1 - G_{ok}(\varphi^*_{ok,dj})) \tag{A32}$$

from o,k to consumers in d is:

$$M_{ok,d} = M^e_{ok}(1 - G_{ok}(\varphi^*_{ok,d})) \tag{A33}$$

The gravity equations can again be rewritten as:

$$X^P_{ok,dj} = M^e_{ok} \left(\frac{1+\theta}{\theta} c_{ok} t_{ok,dj} \kappa_{ok,dj} \right)^{-\theta} \iota^{(1+\theta)}_{ok,dj} \int_{\varphi^*_{ok,dj}}^{\infty} \varphi^\theta dG_{ok}(\varphi) (P^M_{dj})^{1+\theta} \int_{\Omega_{dj}} Y^M_{dj}(\omega) d\omega \tag{A34}$$

$$X^C_{ok,d} = M^e_{ok} \left(\frac{1+\theta}{\theta} c_{ok} \kappa_{ok,d} \right)^{-\theta} s_{ok,d} \int_{\varphi^*_{ok,d}}^{\infty} \varphi^\theta dG_{ok}(\varphi) P^\theta_d I_d \tag{A35}$$

A.6.2 Free Entry and the Allocation of Factors Across Firms

Following Krugman (1980), the number of entrants in each region–sector (o,k) is determined endogenously by a free-entry condition. Before learning their productivity φ, firms must incur a fixed entry cost $f^e_{ok} > 0$. Free entry requires that the expected value of entry equals this cost. Formally, expected profits from serving all possible markets must satisfy:

$$c_{ok} f^e_{ok} = E_\varphi \left[\sum_{d=1}^{N} \sum_{j=1}^{J} \max \left\{ \pi^P_{ok,dj}(\varphi) - f_{ok,dj} c_{ok}, 0 \right\} + \sum_{d=1}^{N} \max \left\{ \pi^C_{ok,d}(\varphi) - f_{ok,d} c_{ok}, 0 \right\} \right].$$

Using the cutoff productivity levels $\varphi^*_{ok,dj}$ and $\varphi^*_{ok,d}$, the condition can be rewritten as integrals over the distribution of productivities:

$$\begin{aligned} c_{ok} f^e_{ok} = & \sum_{d=1}^{N} \sum_{j=1}^{J} \int_{\max\{\varphi^*_{ok}, \varphi^*_{ok,dj}\}}^{\infty} \left(\pi^P_{ok,dj}(\varphi) - f_{ok,dj} c_{ok} \right) dG_{ok}(\varphi) \\ & + \sum_{d=1}^{N} \int_{\max\{\varphi^*_{ok}, \varphi^*_{ok,d}\}}^{\infty} \left(\pi^C_{ok,d}(\varphi) - f_{ok,d} c_{ok} \right) dG_{ok}(\varphi). \end{aligned}$$

The operating firms $M_{ok} = M^e_{ok} \times (1 - G_{ok}(\varphi^*_{ok}))$:

$$\begin{aligned} M_{ok} &= \frac{r^P_{ok} + r^C_{ok}}{r_{ok}(\tilde{\varphi}_{ok})} \\ &= \frac{r_{ok}}{(1+\theta) c_{ok} \left(f^e_{ok} + \sum_{d=1}^{N} \sum_{j=1}^{J} f_{ok,dj} + \sum_{d=1}^{N} f_{ok,d} \right)}. \end{aligned}$$

Collecting terms, the final expression for the equilibrium number of incumbent firms is:

$$M_{ok} = \frac{\sum_{d=1}^{N} \left(\sum_{j=1}^{J} r^P_{ok,dj} + r^C_{ok,d} \right)}{(1+\theta)(f^e_{ok} + \sum_{d=1}^{N} (\sum_{j=1}^{J} f_{ok,dj} + f_{ok,d})) c_{ok}}. \tag{A36}$$

This condition shows that the number of firms in each region–sector (o,k) increases with the total expected revenue r_{ok} and decreases with marginal costs c_{ok} and the scale of entry and fixed costs.

A.7 Incorporating the Distribution of Firm Productivity

Following Chaney (2008), I assume that firm productivity in each region–sector (o, k) follows a Pareto distribution:

$$G_{ok}(\varphi) = G(\varphi) = 1 - \varphi^{-\rho}, \qquad \rho > 0.$$

The parameter ρ governs the dispersion of productivity draws: a smaller ρ corresponds to greater heterogeneity across firms. I require $\rho > \theta$ to ensure that trade flows are finite.

Average productivity moments. The key object entering gravity equations is the conditional expectation of productivity above the cutoff φ^*. For intermediate trade from (o, k) to (d, j), the moment is:

$$\begin{aligned}\int_{\varphi^*_{ok,dj}}^{\infty} \varphi^{\theta}\, dG(\varphi) &= \int_{\varphi^*_{ok,dj}}^{\infty} \varphi^{\theta}\, \rho\varphi^{-\rho-1} d\varphi \\ &= \rho \int_{\varphi^*_{ok,dj}}^{\infty} \varphi^{\theta-\rho-1} d\varphi \\ &= \frac{\rho}{\rho-\theta} (\varphi^*_{ok,dj})^{\theta-\rho}.\end{aligned}$$

Substituting the cutoff productivity expression, this becomes:

$$\int_{\varphi^*_{ok,dj}}^{\infty} \varphi^{\theta}\, dG(\varphi) = \frac{\rho}{\rho-\theta} \left[\frac{f_{ok,dj} c_{ok} \left(\frac{1+\theta}{\theta} c_{ok} t_{ok,dj} \kappa_{ok,dj} \right)^{\theta}}{(\iota_{ok,dj} P^M_{dj})^{1+\theta} Y^M_{dj}} \right]^{\frac{\theta-\rho}{\theta}}. \tag{A37}$$

Similarly, for final goods trade from (o, k) to d:

$$\begin{aligned}\int_{\varphi^*_{ok,d}}^{\infty} \varphi^{\theta}\, dG(\varphi) &= \rho \int_{\varphi^*_{ok,d}}^{\infty} \varphi^{\theta-\rho-1} d\varphi \\ &= \frac{\rho}{\rho-\theta} (\varphi^*_{ok,d})^{\theta-\rho},\end{aligned}$$

which, after substituting the cutoff, yields:

$$\int_{\varphi^*_{ok,d}}^{\infty} \varphi^{\theta}\, dG(\varphi) = \frac{\rho}{\rho-\theta} \left[\frac{(1+\theta) f_{ok,d} c_{ok} \left(\frac{1+\theta}{\theta} c_{ok} \kappa_{ok,d} \right)^{\theta}}{s_{ok,d} P^{\theta}_d I_d} \right]^{\frac{\theta-\rho}{\theta}}. \tag{A38}$$

Gravity equations. Substituting these expressions into the bilateral gravity equations gives:

$$X^P_{ok,dj} = C_1 \,(t_{ok,dj} c_{ok} \kappa_{ok,dj})^{-\rho} f_{ok,dj}^{\frac{\theta-\rho}{\theta}} M^e_{ok} [Y^M_{dj} (P^M_{dj})^{1+\theta}]^{\frac{\rho}{\theta}} \iota_{ok,dj}^{\frac{\rho(1+\theta)}{\theta}}, \tag{A39}$$

$$X^C_{ok,d} = C_2 \,(c_{ok} \kappa_{ok,d})^{-\rho} f_{ok,d}^{\frac{\theta-\rho}{\theta}} M^e_{ok} (I_d P^{\theta}_d)^{\frac{\rho}{\theta}} s_{ok,d}^{\frac{\rho}{\theta}}. \tag{A40}$$

Trade shares Rewritten. Normalizing by total expenditures yields bilateral trade shares. For intermediates:

$$\gamma^P_{ok,dj} = \frac{(c_{ok}t_{ok,dj}\kappa_{ok,dj})^{-\rho} f_{ok,dj}^{\frac{\theta-\rho}{\theta}} M^e_{ok}\, \iota_{ok,dj}^{\frac{\rho(1+\theta)}{\theta}}}{\sum_{o'=1}^{N+1}\sum_{k'=1}^{J}(c_{o'k'}t_{o'k',dj}\kappa_{o'k',dj})^{-\rho} f_{o'k',dj}^{\frac{\theta-\rho}{\theta}} M^e_{o'k'}\, \iota_{o'k',dj}^{\frac{\rho(1+\theta)}{\theta}}}. \tag{A41}$$

For final goods:

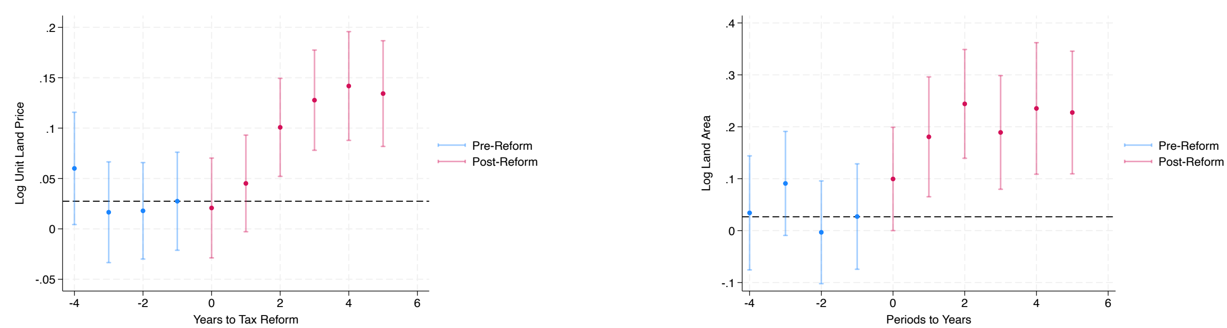
$$\gamma^C_{ok,d} = \frac{(c_{ok}\kappa_{ok,d})^{-\rho} f_{ok,d}^{\frac{\theta-\rho}{\theta}}\, M^e_{ok}\, s_{ok,d}^{\frac{\rho}{\theta}}}{\sum_{o'=1}^{N}\sum_{k'=1}^{J}(c_{o'k'}\kappa_{o'k',d})^{-\rho} f_{o'k',d}^{\frac{\theta-\rho}{\theta}} M^e_{o'k'}\, s_{o'k',d}^{\frac{\rho}{\theta}}}. \tag{A42}$$

These expressions highlight how the Pareto tail parameter ρ interacts with the cutoff productivity and fixed costs to shape trade flows. Larger dispersion (lower ρ) amplifies the role of extensive margins, while higher ρ converges the model toward a representative-firm gravity structure.

B Impact on Land Finance, Additional Source for Welfare Improvement

The results from the quantitative model presented in the paper provide insights into the effects of the Business to Value-Added Tax (B2V) reform. However, one limitation of the baseline analysis is the omission of land, a crucial factor in China's economy and a central source of fiscal revenue for local governments. This omission may bias the measured welfare effects, particularly in regions where land finance plays an outsized role.

In the model, provinces such as Shanghai and Beijing experience a large decline in tax revenue due to the reduction in business tax receipts following the reform. Yet when we examine land market outcomes, a complementary dynamic emerges that mitigates some of these fiscal pressures. Using a similar event-study specification as in the firm-level analysis, I document in Figure A.1 that the reform is associated with a differential increase in both the unit price of land and the total area of land transacted by service firms compared to manufacturing firms.



(A) Land Price **(B)** Land Area

Figure A.1. Effects of the B2V Reform on Land Finance

The estimates show that the average growth in land prices and land area transacted attributable to the B2V reform are 9.3% and 20.3% higher, respectively, in the service sectors compared to manufacturing counterparts. These findings are significant when considering that, as of 2012, land use rights transactions accounted for roughly 40% of local government revenue nationwide. Thus, while the direct removal of the business tax reduced fiscal inflows, the reform indirectly stimulated land-based revenue through higher demand from expanding service firms. Provinces such as Shanghai and Beijing, with both large service sectors and heavy reliance on land finance, likely experienced a partial offset of lost business tax revenue through this channel.

Robustness of firm-level analysis. The land finance results also provide an independent validation of the main firm-level evidence. The increase in land demand and prices is consistent with service firms expanding operations after the reform, mirroring the observed surge in entry and revenue growth in the tax survey and registry data. In particular, land acquisition is a costly and strategic investment decision, typically undertaken only when firms anticipate long-run expansion and profitability. The fact that land market outcomes exhibit similar reform-induced dynamics strengthens the interpretation that the B2V reform not only raised firm revenues but also shifted investment behavior and resource allocation. Taken together, these complementary results reinforce the conclusion that the reform stimulated service sector expansion through both output and investment margins, with important implications for local fiscal capacity and welfare.

C Counterfactual In-house Share

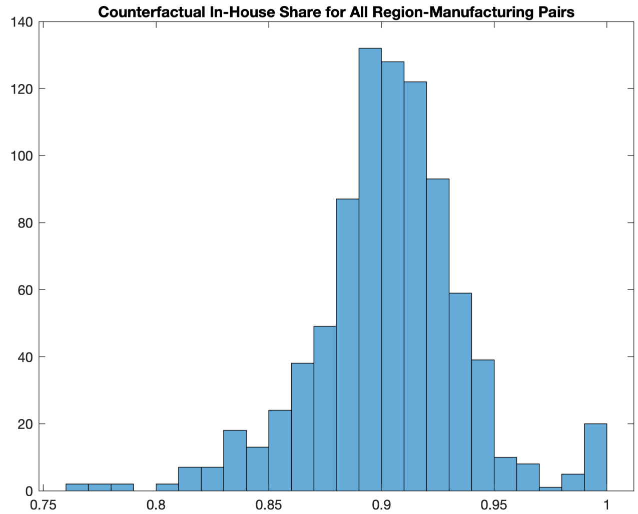


Figure A.2. Counterfactual In-house Share

Notes: This figure plots the distribution of the counterfactual in-house production share across all region–manufacturing pairs. A value of 0.9 on the x-axis indicates that the region–sector pair uses 90% of its baseline in-house share in the counterfactual equilibrium. The histogram shows that most pairs cluster between 0.88 and 0.95, implying a substantial but heterogeneous decline in in-house production following the reform, consistent with reallocation toward market-based inputs.

D How to Solve the Model Numerically

This section outlines the numerical procedure used to solve the model in equilibrium. The model is expressed in exact hat algebra, meaning all variables are computed as changes relative to the baseline equilibrium. The solution requires iterating over a set of endogenous variables until market clearing, firm entry, and free-entry conditions are simultaneously satisfied.

D.1 Variables Solved in the Model

The following endogenous variables are jointly solved in equilibrium:

- **Nominal wages in change:** $\hat{\mathbf{w}} = \{\hat{w}_1, \hat{w}_2, \dots, \hat{w}_N\}$ for N regions.
- **Intermediate input price indices in change:**

$$\hat{\mathbf{P}}^M = \{\hat{P}^M_{1,1}, \hat{P}^M_{1,2}, \dots, \hat{P}^M_{1,J}, \dots, \hat{P}^M_{N,1}, \hat{P}^M_{N,2}, \dots, \hat{P}^M_{N,J}\}_{1\times(N\times J)}.$$

- **Final good price indices in change:** $\hat{\mathbf{P}} = \{\hat{P}_1, \hat{P}_2, \dots, \hat{P}_N\}_{1\times N}$.
- **Number of incumbent firms in change:**

$$\hat{\mathbf{M}}_{ok} = \{\hat{M}_{1,1}, \hat{M}_{1,2}, \dots, \hat{M}_{1,J}, \dots, \hat{M}_{N,1}, \hat{M}_{N,2}, \dots, \hat{M}_{N,J}\}_{1\times(N\times J)}.$$

- **Nominal output in change:**

$$\hat{\mathbf{y}}_{ok} = \{\hat{y}_{1,1}, \hat{y}_{1,2}, \dots, \hat{y}_{1,J}, \dots, \hat{y}_{N,1}, \hat{y}_{N,2}, \dots, \hat{y}_{N,J}\}_{1\times(N\times J)}.$$

- **Cutoff productivities in change:** for each exporting region–sector pair (o, k) and destination (d, j),

$$\hat{\varphi}^*_{ok,dj}, \qquad \hat{\varphi}^*_{ok,d},$$

which determine the extensive margin of trade and the average productivity of active exporters.

D.2 Parameters Calibrated or Estimated Outside the Model

The following parameters are taken from estimation or calibration:

- Inter-sector–region trade elasticity ρ.
- Elasticity of substitution between labor and intermediate input η.
- Baseline intermediate trade shares $\gamma^P_{ok,dj}$.
- Baseline final trade shares $\gamma^C_{ok,d}$.
- Counterfactual tax matrix $t_{k,j}$:

$$\hat{\mathbf{t}} = \begin{bmatrix} \hat{t}^{1,1} & \hat{t}^{1,2} & \dots & \hat{t}^{1,J} \\ \hat{t}^{2,1} & \hat{t}^{2,2} & \dots & \hat{t}^{2,J} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{t}^{J,1} & \hat{t}^{J,2} & \dots & \hat{t}^{J,J} \\ t^{H,1} & t^{H,2} & \dots & t^{H,J} \end{bmatrix}.$$

D.3 Numerical Solution Procedure

The model is solved using the following iterative steps:

- **Step 1.** Guess an initial vector of wages $\hat{\mathbf{w}}$ and final good price indices $\hat{\mathbf{P}}$.
- **Step 2.** Using equilibrium conditions, compute marginal costs in change $\hat{c}_{ok,dj}$, the intermediate price index $\hat{P}^M_{dj}$, and the final good price index $\hat{P}_d$.
- **Step 3.** Compute new bilateral trade shares $\hat{\gamma}^P_{ok,dj}$ by combining baseline shares, the estimated elasticities, and the updated $\hat{P}^M_{dj}$ and $\hat{c}_{ok,dj}$.
- **Step 4.** Given $\hat{\gamma}^P_{ok,dj}$, the counterfactual tax matrix $\hat{t}$, and updated value-added shares $(\zeta_{ok,dj})'$, compute counterfactual expenditures from origin (o, k) to destination (d, j):

$$(X_{o'k',ok})' = (X_{ok})'(\gamma^P_{o'k',ok})' = (\xi_{o'k',ok})' \sum_{o'} \left[\sum_{k'} (X_{o'k'})'(\gamma^P_{ok,o'k'})' + (\gamma^C_{ok,o'})'(I_{ok,o'})' \right]. \quad \text{(A1)}$$

  After updating expenditures, iterate again on value added for each region–sector pair to ensure internal consistency.

- **Step 5.** Calculate the cutoff productivities $\hat{\varphi}^*_{ok,dj}$ and $\hat{\varphi}^*_{ok,d}$ using the zero-profit conditions. These determine the average productivity of exporters and update the extensive margin of trade.

- **Step 6.** Impose trade balance conditions. Substituting counterfactual expenditures into the trade imbalance condition yields:

$$\sum_{o'=1}^{N}\sum_{k'=1}^{J}\left[\sum_{k=1}^{J} X_{ok}\gamma^{P}_{o'k',ok} + Y^{C}_{o'k',o}\right] - D_o = \sum_{o'=1}^{N}\sum_{k'=1}^{J}\left[\sum_{k=1}^{J} X_{o'k'}\gamma^{P}_{ok,o'k'} + Y^{C}_{ok,o'}\right]. \tag{A2}$$

- **Step 7.** Check convergence. If implied wages $\hat{\mathbf{w}}$ are not consistent with market clearing, update the guess for $\hat{\mathbf{w}}$ and return to Step 1. Iterate until convergence.

At convergence, the algorithm delivers a full set of equilibrium outcomes: wages, price indices, trade shares, firm entry, and cutoff productivities. These allow us to compute welfare and output changes under the counterfactual tax regime.