Essays on Trade and Economic Geography

Ningyuan Jia

Department of Economics

London School of Economics

A thesis submitted to the Department of Economics
of the London School of Economics for the Degree of Doctor of Philosophy

London, July 2024

Declaration

I certify that the thesis I have presented for examination for the Ph.D. degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorization does not, to the best of my belief, infringe the rights of any third party.

I confirm that Chapter 2 is jointly co-authored with Swati Dhingra, Gianmarco Ottaviano, Thomas Sampson, and Catherine Thomas.

I confirm that Chapter 3 is jointly co-authored with Swati Dhingra, Emily Fry, and Sophie Hale.

I declare that my thesis consists of approximately 44735 words.

Acknowledgements

I am deeply grateful to Swati Dhingra for her unwavering support, encouragement, and guidance throughout my academic journey. Her consistent support and belief in my abilities have been invaluable, and I am truly appreciative of her mentorship. Swati has been a supervisor and a source of inspiration, providing valuable academic resources and demonstrating what it means to pursue one's dreams with determination and perseverance. I am immensely fortunate to have had Swati as a mentor, and I am deeply grateful for her influence on my academic and personal development.

I extend my heartfelt thanks to Catherine Thomas, Daniel M. Sturm, and Matthias Doepke for their invaluable support during the final stages of my PhD journey. Catherine Thomas has been a consistent source of support and assistance throughout my PhD. Her willingness to provide help whenever needed has been truly appreciated. Our collaboration on the second chapter of this thesis has been a valuable learning experience, and I am grateful for the opportunity to work alongside her. Daniel M. Sturm has played a pivotal role in shaping my academic interests, particularly in the field of economic geography. Matthias Doepke's advice on family economics and weekly PhD group meetings have been immensely beneficial during my job market preparation, and I am grateful for his guidance and support during this crucial time.

Finally, I am fortunate to have had consistent support from my family as I pursued academia. My close friend, Leo, has been invaluable in exchanging ideas and providing encouragement. I am also grateful to my girlfriend, Shiying, for her steadfast company and support. Many thanks to all of them for their unwavering support and companionship over the years.

Abstract

This thesis consists of three chapters on international trade and economic geography. The first chapter examines Family Planning Programs (FPPs) as a driver of structural transformation. Many nations have implemented FPPs as a strategic response to demographic shifts, aiming to optimize economic growth. This chapter introduces a comprehensive dynamic economic geography model, incorporating dynamic fertility decisions, multi-regional and multi-sectoral frameworks, non-homothetic preferences, technological advancements, and region-specific constraints. The core mechanism suggests that changes in the total labor force predominantly favor labor-intensive non-agricultural sectors, thereby facilitating structural transformation. Counterfactual analyses indicate that China's FPPs initially boost the share of non-agricultural employment by 0.7 percentage points but lead to a 2.7 percentage point reduction in the long term.

The second chapter investigates the global liquefied natural gas (LNG) market. Around three-quarters of global trade in LNG is transacted within long-term contracts between buyers and sellers. Reduced-form evidence from transactions between 2009 and 2019 shows contracts operate as call options. We model a contractual relationship where non-contractible effort from the seller contributes to the relative value of a contract shipment. Estimates of our model suggest the seller's effort is important, and that sellers extract a large share of the contract value. The results suggest there are significant frictions in the LNG spot market due to technology and geography constraints and also its relatively small size.

The third chapter examines the implications of the EU-UK Trade and Cooperation Agreement (TCA) using a dynamic spatial general equilibrium model. Unlike previous studies that primarily projected potential impacts of Brexit based on speculative scenarios, this analysis introduces new estimates of tariffs and non-tariff measures (NTMs) specified by the TCA. The findings show that while the TCA will necessitate substantial adjustments for some sectors, Brexit will not fundamentally alter the overall structure of the UK economy. However, the main impacts will manifest as significant hits to real wages and productivity, exacerbating the long-standing economic challenges faced by the UK.

Table of Contents

List of Figures

List of Tables

Chapter 1

Demographic Transition and Structural Transformation in China

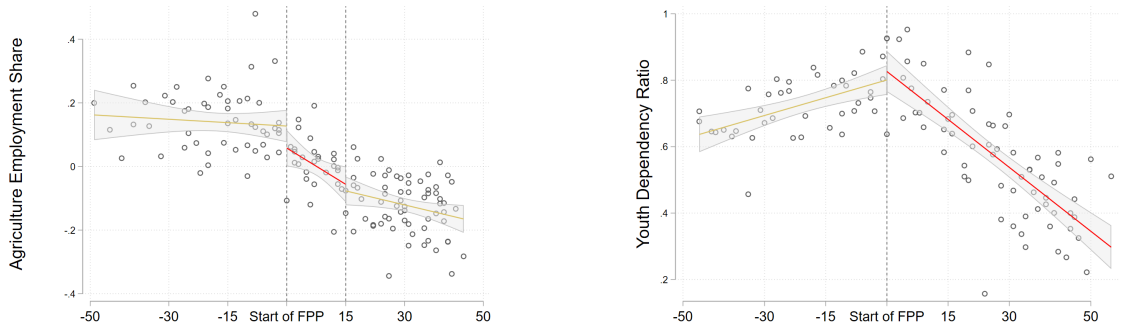
1.1 Introduction

Demographic transition and structural transformation are two crucial processes that influence a country's economic development and societal change. Demographic transition involves the shift from high to low rates of fertility and mortality, thereby creating a window of opportunity for industrialization and urbanization. Concurrently, structural transformation entails the transition from an agriculture-centric economy to one focused on industry and services, which is vital for improving living standards. In the early stages of their development, many countries initiated Family Planning Programs (FPPs) to capitalize on this demographic opportunity. By the mid-1990s, such programs were operational in 115 countries[1]. Decades after the implementation of FPPs, the total fertility rate in developing countries was halved. Despite extensive research on family-level outcomes, there remains a significant gap in understanding the relationship between demographic transition and structural transformation. Additionally, there is limited insight into the trade-offs between short-term dependency ratios and the long-term labor force composition, particularly as age structures evolve. This study aims to address these gaps.

Figure 1.1 offers an overview of demographic transition and structural transformation trends in countries that have implemented FPPs. Within this overarching context, two specific trends are examined in detail. First, Figure 1.1a explores the relationship between the initiation of FPPs and the proportion of agricultural employment across 19 countries, with data ranging from 1900 to 2008.[2] This figure reveals a significant decline in the agri-

[1]For a comprehensive review of global family planning programs, see Robinson and Ross (2007).

[2]The countries included in this study—Bangladesh, Chile, Colombia, Egypt, India, Indonesia, Iran, Malaysia, Morocco, Nepal, Pakistan, Philippines, Singapore, South Korea, Sri Lanka, Thailand, Tunisia, and



(a) Industrialization and FPPs

(b) Dependency Ratio and FPPs

Figure 1.1: Demographic Transition and Structural Transformation for Countries with FPPs

Source: International Historical Statistics

cultural employment share in the 15 years following the implementation of FPPs. Importantly, the rate of this decline appears to decelerate after the initial 15-year period. Second, Figure 1.1b analyzes the dependency ratio, defined as the ratio of children to working-age individuals, in the same set of countries. A consistent decrease in this ratio is observed in the aftermath of FPP implementation. Collectively, these figures unveil compelling patterns that underpin the research questions this study aims to investigate. Specifically, the initial sharp reduction in agricultural employment share, followed by a slower rate of decline, suggests a potential trade-off between the short-term and long-term effects of FPPs on structural transformation. Similarly, the sustained decrease in the dependency ratio raises questions about its long-term implications for labor force composition. These patterns underscore the need for a clear understanding of how FPPs may interact with broader economic and demographic shifts.

In this chapter, I examine how demographic transition caused by FPPs has influenced the process of structural transformation. The chapter consists of three parts. First, I develop a dynamic economic geography model that incorporates the key mechanism underlying demographic transition: differences in sectoral factor intensity coupled with shifts in the labor force size. Second, I use exogenous variation in the specific FPP—China's One Child Policy (OCP)—across prefectures to provide suggestive evidence on the relationship between demographic transition and structural transformation and empirically evaluate the channels depicted in the model. Third, I take the model to China's data and perform counterfactual analysis by removing China's FPP fines. This model enables

Turkey—are selected based on the availability of historical data. The start years of FPPs are sourced from Robinson and Ross (2007).

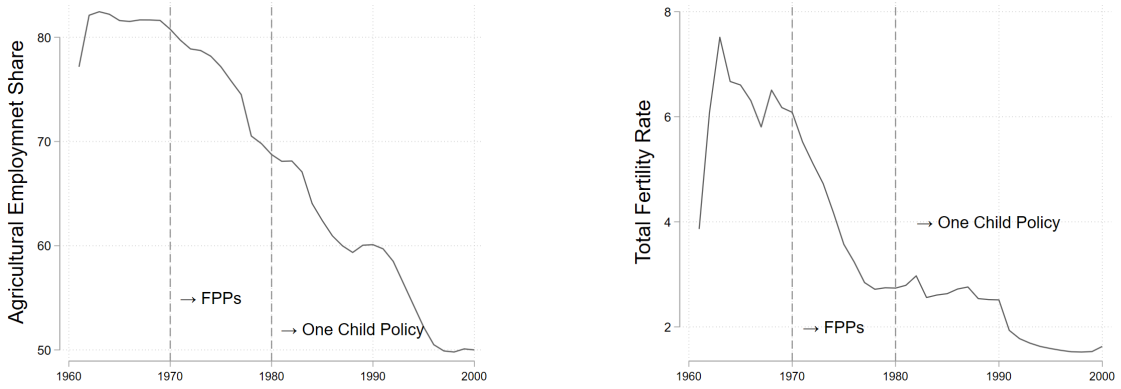
me to examine the driving forces behind structural transformation while considering the demographic transition's impact and assess aggregate welfare effects across prefectures. By adding China-specific frictions and other known drivers of structural transformation in the model, I can distinguish the contributions of FPPs from other traditional forces including technological change, entry barriers to the non-agricultural sector, migration costs, trade costs, and Hukou restrictions.[3]

My research focuses on China as an ideal case study to explore the relationship between demographic transition and structural transformation. Two critical facts make China unique for this investigation. Firstly, China has undergone a profound economic evolution, transitioning from an agriculture-focused economy to a manufacturing-driven model at the end of 20th century. Over the decades, the share of agricultural sector employment in China fell from 81% in 1970 to 50% in 2000 (as depicted in Figure 1.2a). By 2010, this figure had further reduced to 39%. While numerous influential factors contributed to this transition, my specific focus is on one pivotal factor: the demographic transition.

Secondly, the implementation of a series of FPPs during the 1970s and the introduction of the OCP in the early 1980s exerted substantial impacts on China's population dynamics.[4] Figure 1.2b illustrates the fertility trend, which represents the number of children per mother from 1960 to 2000. The significant decrease in fertility is closely associated with the extensive implementation of these FPPs. Moreover, this decline in fertility rate has induced profound changes in the age structure of the population and the dependency ratio. Figure 1.3a compares the population pyramids in 1982 and 2000. In 1982, the population pyramid displayed a higher proportion of individuals under 15, indicative of a youthful population. By 2000, the age structure shifted, with a greater share of individuals between 15 and 65, reflecting a transition towards a higher share of working-age population. This change has significant implications for dependency ratios, as measured by the number of young dependents (under 15) over total working age population who aged between 15 and 64 in Figure 1.3b, which shows a consistent decline. This chapter investigates how changes in age structure resulting from these FPPs influence short-term and long-term

[3] In China, "Hukou" refers to the household registration system. It is a crucial administrative system that identifies and records individuals' official residential status in the country. Every Chinese citizen is assigned a Hukou, which can be either agricultural (rural) or non-agricultural (urban) in a specific administrative unit. Chan (2010) and Bian (2002) discuss the Hukou system in detail.

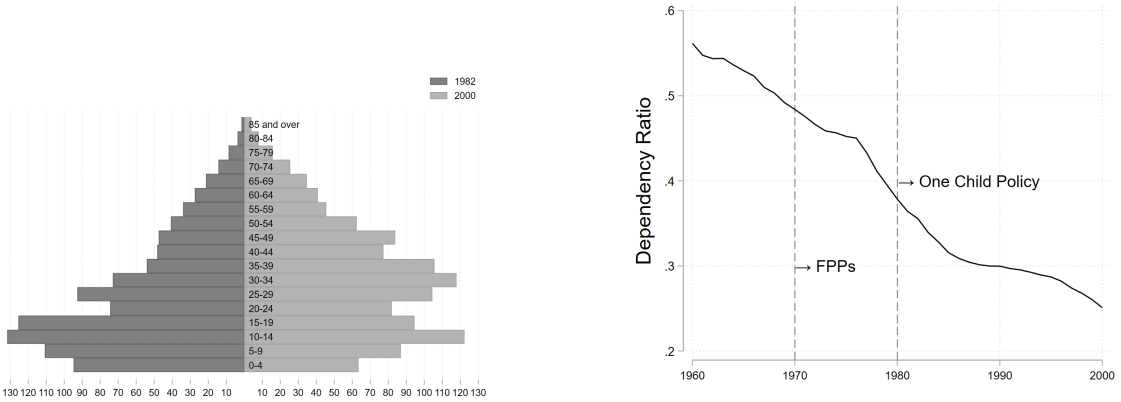
[4] The Chinese FPP has undergone multiple transitions over the years. It began in the 1970s with a moderate approach, encapsulated by the slogan "wan, xi, shao", which advocated for late marriage, late childbearing, and limited fertility. This was succeeded by the strict implementation of the OCP in early 1980s. Further details on the evolution of these policies will be elaborated upon in the Background section.



(a) Industrialization in China (b) Fertility Trend in China

Figure 1.2: Demographic Transition and Structural Transformation in China

Source: World Bank and National Bureau of Statistics of China



(a) Population Pyramid (in Millions) (b) Dependency Ratio

Figure 1.3: Age Structure and Dependency Ratio in China

Source: China's 1982 and 2000 1% population census

structural transformation. By delving into China's developmental trajectory, my chapter offers inspiration for policies on demographic and economic transition.

This chapter conducts a region-level analysis by building an economic geography model, motivated by three salient factors. First, the Hukou system in China imposes stringent restrictions on migration, rendering each prefecture a de facto self-contained economic unit. To put this into perspective, the proportion of inter-provincial migrants in China constituted a mere 2.4% of the total population in the year 2000. This is in stark contrast to the United States, where inter-state migrants accounted for over 30% of the population in the same year.[5] This low level of migration suggests that demographic transitions in

[5]Data for China's migration statistics in 2000 are sourced from the 2000 1% Population Census. U.S. inter-state migration data are derived from Molloy et al. (2011).

China, influenced by FPPs, are largely localized phenomena. Second, despite constrained labor mobility, intra-province trade is notably robust, comprising 22% of the total GDP. This indicates that while labor may be immobile, economic interconnectivity is maintained through trade networks. Third, there is a pronounced disparity in regional development across China. The fluctuation in the share of agricultural employment is considerable, ranging from a decrease of 61% to an increase of 42%, with a standard deviation of 14% during 1982-2000. Collectively, these three factors—restricted migration, significant intra-province trade, and uneven regional development— motivate region-level model to investigate China's demographic transition and its impact on structural transformation.

The dynamic economic geography model in this chapter analyzes the effect of fertility related shocks, describes its determinants, quantifies its aggregate impact, and studies its interaction with other traditional forces. The model features the following key components: (1) dynamic fertility decision-making, (2) standard multi-region and multi-sector urban model component, (3) non-homothetic preferences and technology progress in different sectors, and (4) China-specific frictions: Hukou costs and non-agricultural entry barriers. To this end, I incorporate endogenous fertility in line with canonical theory in family economics. The model also employs an overlapping generations setting to capture the intertemporal aspects of demographic and economic change. Additionally, a fine for each extra child is incorporated to reflect China's unique policy environment, particularly the fines associated with non-compliance with FPPs. By integrating these components, the model is well-equipped to match China's specific demographic and economic characteristics. It enables a thorough investigation of the key channels through which demographic transition interacts with structural transformation, thereby contributing to a more comprehensive understanding of China's economic development trajectory.

In my model, individuals are assumed to live for two periods. During the first period, they are children and do not participate in any decision-making. In the second period, they enter the labor market and make work and fertility choices. When individuals decide to raise children, they not only need to allocate some time out of their working hours to take care of their children, but they also face FPP-specific punishment if they have more children than allowed by the policy. Additionally, individuals derive utility from having children, and this utility is based on the expected future utility of their offspring, as expressed in Barro and Becker (1989). On the production side, each location is characterized by two sectors: the agricultural sector and the non-agricultural sector. The agricultural sector is modeled as land-intensive and operates under perfect competition, while the non-agricultural sector

is modeled as labor-intensive and operates under monopolistic competition with region-specific entry costs.

The model's intuition posits that FPPs, by altering the cost of child-rearing, influence fertility rates and subsequently the labor force.[6] In the short term, FPPs create a "demographic window of opportunity", where reduced fertility and a lower dependency ratio lead to an increased effective labor supply, thereby fostering economic growth and industrialization. However, in the long term, the reduced fertility rates result in a smaller cohort of new labor force entrants, leading to a contraction in the total labor supply. This dynamic interplay between short-term and long-term labor force changes has complex implications for structural transformation. Specifically, the labor-intensive (non-agricultural) sector expands more in the short term due to increased labor supply but contracts in the long term as the labor supply diminishes. Thus, the model captures the nuanced effects of FPPs on structural transformation by considering sectoral differences in factor input intensity and the opposing trends in total labor supply over different time horizons.

The geographic setting of the model plays a crucial role in ensuring its basic intuition. (1) The presence of migration frictions implies that a significant share of newborns are absorbed locally. If there were no migration costs, individuals would be able to migrate frictionlessly to other prefectures, and local shocks would have a minimal impact on the labor force dynamics in each region. In such a scenario, a shock affecting one region would be equivalent to a shock affecting all prefectures uniformly, and there would be no discernible difference between prefectures with different shocks. The consideration of migration frictions in the model is essential for capturing the localized impacts of the FPPs on demographic changes and structural transformation. In China's setting, the geographic restrictions on migration imply that an decrease in fertility in a certain region would primarily affect the local labor supply, and this effect may not be immediately offset by inward migration from other prefectures. As a result, prefectures with different OCP policy exposures may experience varying changes in fertility rates and labor force sizes, leading to differences in their economic and industrial development trajectories over time. (2) The presence of inter-regional trade is also a crucial aspect of the model, impacting the interpretation of the FPPs effect. In the absence of trade with other prefectures, an increase in local labor supply would lead to a greater boost in output in the relatively labor-intensive

[6]The main channel through which FPPs affect structural transformation is changing labor forces in the chapter. However, FPPs may also have other implications which could affect structural transformation. For example, FPPs increase saving rates as in Choukhmane et al. (2023) and facilitate capital deepening. This will lead to faster growth in sectors that rely more on capital (non-agricultural sector). Since the contribution of capital in my model is captured by the sectoral TFP, I include this channel by allowing non-agricultural sector to have faster growth in TFP in regions with lower fertility rates.

sector (non-agricultural sector) compared to the land-intensive sector (agricultural sector), assuming constant prices. As a result, the relative price of agricultural products and wages would rise, attracting more labor into the agricultural sector. However, when trade is considered, local sectoral prices become influenced by costs in other prefectures and are less affected by local output changes. This dampens the equilibrium price response to regional shocks, thereby allowing the factor input intensity effect to play a dominant role in shaping the structural transformation pattern. In an extreme case, such as a small open economy where prices are globally determined, this would converge to the Rybczynski effect (Rybczynski, 1955), which states that an increase in the endowment of one factor leads to a more-than-proportional expansion in the output of the sector that intensively uses that factor, while causing an absolute decline in the output of the other sector.

Along with the traditional drivers of structural transformation such as non-homothetic preferences and technological progress, my model comprehensively captures the majority of the important factors that have influenced China's industrialization since the reform and opening up in 1978. It accounts for rapid technological advances in both sectors, lower entry barriers to the non-agricultural sector, lowering Hukou and trade costs, and FPPs. Through these features, the model enables me to disentangle the specific effect of the FPPs from other factors, study their interactions, and accurately assess the contribution of each factor to the overall process of structural transformation in China's economy.

After exploring the key channels and qualitative insights in the model, I use the OCP as a natural experiment to provide suggestive evidence for the mechanism and predictions behind the model. There are several difficulties in identifying the effect of demographic transition using the OCP. One of the main difficulties is that the OCP was not randomly assigned, which means that the groups of people who were affected by the policy may differ in other ways that affect the outcomes of interest. Additionally, the implementation of OCP at the regional level was based on local socioeconomic considerations, which means that the variation in OCP enforcement partly comes from variation in regional fundamentals.

To address the challenges of endogeneity, I exploit the post-1982 relaxation of the OCP in most provinces, which allowed families to have a second child if the first was a girl.[7] In this context, prefectures with a higher proportion of single-daughter families experienced a more significant policy shock. Given the absence of evidence for gender selection at the first birth in 1982, which I confirm in my empirical study, the gender of the first-born serves

[7]This approach aligns with the methodologies used in Huang et al. (2019) and Qian (2018).

as a plausibly exogenous family-level variable. In my primary identification strategy, I use the 1982 proportion of single-daughter families as a regional policy exposure indicator.

The identification of this measure is based on the assumption that, in the absence of the policy, the trends in outcome variables should be parallel across prefectures with different policy exposure. In other words, prefectures with higher and lower policy exposure should have experienced similar trajectories in employment outcomes, had they been subject to the same economic and policy conditions. To provide evidence supporting the parallel trend assumption, I conduct a comprehensive analysis of pre-treatment trends in employment outcomes in both sets of prefectures. I examine the employment data for agricultural and non-agricultural sectors from the years leading up to the policy, before any differential policy effects could have taken place.

My difference-in-difference (DID) estimation compares changes in agricultural and non-agricultural employment between prefectures with different policy exposures before and after the 1982 using data from the 1982, 1990, 2000 1% population censuses, and the 2005 1‰ mini population census. The baseline results indicate that prefectures with higher policy exposure experienced a greater increase in non-agricultural employment share in the long run (15 years after the policy), suggesting a faster pace of industrialization during that time. However, in the short run (within 15 years after the policy), these same prefectures witnessed a decrease in non-agricultural employment share that is slower industrialization. These contrasting outcomes shed light on the complexities of the connection between demographic changes and structural transformation.

Why would the OCP relaxation contribute to structural transformation, and why does the trend reverse in the short and long run? For any factor to account for the reversing trend in the main findings, it needs to satisfy three criteria. (1) It should be affected by the policy exposure. (2) It should also have different trends in the short and long run. (3) It should be associated with structural changes. The main explanation in this chapter, which satisfies all three criteria, relates to changes in the total labor supply combined with differences in factor intensity in the agricultural and non-agricultural sectors.

By replacing the outcome variable with the total labor force in the baseline regression, I find that the total labor supply shows different trends in the short run and long run due to changes in fertility and the dependency ratio. The most direct effect of higher exposure is that it led to higher fertility rates after policy. It takes at least 15 years for these newborns to enter the labor market; therefore, the total labor supply increased in these prefectures in the long run but not in the short run. Another aspect of the increase in fertility is the

rise in the dependency ratio, which indeed led to a decrease in effective labor supply in the short run. This is because more families needed to reduce their working hours to take care of their children, reducing the available labor force in the short run.

These changes in fertility and the dependency ratio contributed to the observed trend reversal in the short run and long run, impacting the overall labor supply and influencing the process of structural transformation in the economy. Specifically, in China during the period I study, the agricultural sector was more land-intensive, while the non-agricultural sector was more labor-intensive.[8] As a result, an increase in the local labor supply would have led to a greater proportion of labor working in the non-agricultural sector compared to the agricultural sector when prefectures are interconnected through trade.

Apart from the total labor force channel, structural transformation may also be affected through other factors, such as the child quantity-quality trade-off and the sex ratio. Prefectures with higher shares of single-daughter families might have experienced changes in parental investment and educational opportunities. This improvement in human capital could have contributed to the long-term process of structural transformation, as a better-educated and skilled workforce is essential for the growth of industries and the economy. Additionally, a sex ratio imbalance resulting from the OCP might have led to a surplus of males in certain prefectures. This could also affect structural transformation if males and females have different propensities toward different sectors. In this chapter, I empirically assess these channels and demonstrate that they are not the main drivers behind the findings in China's context.

After providing empirical evidence on the key mechanism, I take the model to data to quantify the aggregate effect of China's FPPs on structural transformation. The quantitative results from the model show that FPPs had divergent effects on China's industrialization. In the short run during 1982-1990, industrialization at the national level experienced a 0.7% faster growth rate. However, the long-run consequences of reduced labor supply during 1982-2000 due to the FPPs led to 2.7% slower industrialization. Overall, these effects were more pronounced in prefectures with high relative non-agricultural Total Factor Productivity (TFP), low Hukou costs, low trade costs, and high migration costs. Furthermore, Hukou cost reductions and TFP increases are identified as major drivers of China's overall structural transformation. Though real wage increased in the long run, welfare decreased largely due to the lower utility from children arising from high fines.

[8] This is shown in the following section where I plot the labor share in agriculture and non-agriculture sectors during 1980-2005 in China using IO tables and estimates of land share from Cao and Birchenall (2013). The labor share in the non-agricultural sector is consistently higher than in the agricultural sector during this period.

Related literature. This chapter mainly contributes to three strands of literature. The first strand is the family economics literature in understanding the causes and consequences of fertility (Doepke et al., 2023; Greenwood et al., 2021; Greenwood et al.,2017; Doepke and Tertilt, 2016; Jones et al., 2008; Hotz et al., 1997). This literature identifies two major drivers of fertility to account for the observed negative income-fertility relationship in most countries including China. The first one is the quantity-quality trade off (Delventhal et al., 2021; Croix and Licandro, 2013; Doepke and Tertilt, 2009; Manuelli and Seshadri, 2009; Moav, 2005; Doepke, 2005; Becker, 1960). However, in the context of China, the evidence on the quantity quality trade off is mixed. Qian (2009) finds that having an additional child increased the likelihood the first child enrolled in school of by about 16 percentage points, suggesting a complementary relationship between quantity and quality in education. Rosenzweig and Zhang (2009) found a positive but a modest effect of the quantity on quality. My empirical result adds to the literature and shows that on the regional level, there is no evidence for the quantity-quality trade-off. Another driver for fertility is the opportunity cost of women's time (Coskun and Dalgic, 2022; Currie and Schwandt, 2014; Hotz et al., 1997; Schultz, 1997; Galor and Weil, 1993). One effect of this channel is the negative relationship between fertility and female labor force participation (Aaronson et al., 2021). In this chapter, I show that the time cost of raising children is key to understanding the changes in fertility in response to the family planning program shocks and the main driver behind the reversing trend in total labor supply.

The second strand is the macroeconomic literature on understanding drivers of structural transformation. This literature explores the contribution of various factors including technological progress, frictions, non-homothetic tastes, capital intensities, substitutability, and trade, on structural transformation (Bar and Leukhina, 2010; Brandt et al., 2008; Buera and Kaboski, 2012; Erten and Leight, 2021; Fajgelbaum and Redding, 2022; Herrendorf et al., 2014; Ngai and Petrongolo, 2017; Ngai and Pissarides, 2007; Święcki, 2017). However, the literature has not focused on the role of population growth. Leukhina and Turnovsky (2016a) and Leukhina and Turnovsky (2016b) provide a general equilibrium model analysing the effect of population growth on structural change and show that the population effects depend on the preference-side and production-side characteristics of the economy, and trade. Barham et al. (2023) find that family planning programs in Bangladesh slow down structural transformation. Models in this literature miss the spatial dimension and focus on the aggregate economy. So they cannot explain regional variations in family planning programs. This is important because regional variations allow me to hold national level policies constant when assessing the impact of family planning

programs. This could be valuable in many countries where several national policies are adopted at the same time to promote growth. Ngai et al. (2019) highlight the spatial dimension using China's Hukou system, studying its effect on industrialization. To this end, the detailed mechanism underlining the population effect on a regional level is still unclear and the short-term and long-term trade-offs remain unexplored.

The third strand is the urban and trade literature with the quantitative spatial models (Eckert and Peters, 2022; Fan et al., 2021; Desmet et al., 2018; Caliendo and Parro, 2021; Caliendo et al., 2019; Redding and Rossi- Hansberg, 2017; Allen and Arkolakis, 2014). Within this fast growing literature, some papers focus on the spatial dimension of structural transformation. For example, Eckert and Peters (2022) propose a spatial model of structural transformation with the key driver being catch-up growth. As in most spatial models, population growth is assumed to be exogenous. My model incorporates endogenous fertility decision making that allows me to study the effect of fertility-related shocks on structural change.

This chapter proceeds as follows: Section 1.2 describes the institutional background and model intuition in a basic model. Section 1.3 presents the structural model. Section 1.4 contains the reduced-form analysis supporting the key model mechanisms and robustness tests. Section 1.5 presents the calibration and counterfactual analysis. Section 1.6 concludes the chapter.

1.2 Background and Basic Model

1.2.1 Family Planning in China

The genesis of China's family planning policies can be traced back to the early years of the People's Republic of China, established in 1949 under the leadership of Mao Zedong. Initially, the government encouraged population growth, subscribing to the belief that a larger population equated to greater national strength. This stance was reflective of Mao's famous dictum, "More people, more power". The early policies were influenced by Soviet models, such as the Mother Heroine, and were characterized by a condemnation of birth control and a ban on the import of contraceptives.

However, by the early 1950s, the implications of rapid population growth began to manifest. The First National Population Census of China in 1953 revealed a population exceeding 600 million, with an annual growth rate of over 2.2%. This prompted a shift in government policy towards family planning, leading to the repeal of the ban on contraceptives in 1954

and the commencement of official birth planning campaigns in 1956. The language around contraception also evolved, moving from terms associated with control to those suggesting planned and proactive measures.

Despite these early efforts, Mao's Great Leap Forward, initiated in 1958, marked a return to pro-natalist policies. The subsequent famine from 1959 to 1961, one of the deadliest in human history, led to a temporary decline in population growth. However, fertility rates rebounded sharply in the early 1960s, prompting the government to advocate birth planning in urban and densely populated rural areas. Family planning commissions were established at national and provincial levels to oversee these initiatives (Ebenstein, 2010).

The Cultural Revolution, beginning in 1966, disrupted many state functions, including family planning efforts. Yet, by the early 1970s, China's leadership, recognizing the severe strain that overpopulation was placing on the country's resources and economic development, began to advocate for more stringent family planning measures. In 1971, a significant campaign was launched with the slogan "One isn't too few, two are just fine, and three are too many", reflecting a shift towards a policy of having fewer children. The State Council's Leading Group for Family Planning in 1973 and the national birth planning conference in December of the same year solidified the "Later, Longer, and Fewer" strategy, which encouraged later marriages, longer intervals between births, and smaller family sizes.

While these policies were officially voluntary, they were enforced with varying degrees of coercion (Whyte et al., 2015). Local officials were tasked with monitoring compliance, leading to an increase in contraceptive use, sterilizations, and abortions. The campaign was effective in reducing fertility rates, which halved from 1971 to 1978.

By the end of the 1970s, China's family planning policies had become more institutionalized and coercive. The one-child policy, introduced in 1979 and formally implemented in 1980, marked the culmination of China's efforts to control its population growth. This policy was far more stringent than its predecessors, with significant implications for Chinese society and the global perception of China's approach to population management. The one-child policy was a radical departure from earlier measures, reflecting the government's determination to address the challenges posed by resource constraints and the need to sustain economic growth without the burden of an excessive population (Fong, 2016).

Under the OCP, almost all families in China were restricted to having only one child, and second or higher-parity births were penalized. The policy was rigorously enforced, and measures such as forced abortions, fines, and other penalties were used to ensure

compliance. In the initial years of the policy, there were significant efforts to control population growth, and the fertility rate in China declined as a result.

A crucial turning point occurred in 1982 when the Chinese government included the One Child Policy in the national constitution. By doing so, the OCP became a compulsory and legally binding policy for all Chinese citizens. Its inclusion in the constitution strengthened its authority and provided a clear legal basis for its enforcement, solidifying its position as a long-term population management measure.

However, as time passed, the OCP faced growing opposition and criticism from the public. Many Chinese citizens expressed dissatisfaction with the strict one-child policy, especially in rural areas, where larger family sizes were often considered necessary for agricultural labor and for caring for elderly family members. Additionally, the social preference for male children to carry on the family lineage contributed to discontent with the policy.

In response to public concerns, the Chinese government began to gradually relax the OCP in the 1980s. Local governments took the initiative to issue permits for a second child as early as 1982. However, the most significant relaxation was the issuance of "Document 7" by the Central Party Committee on April 13, 1984. This document served two main purposes: first, to curb female infanticide, forced abortion, and forced sterilization; and second, to devolve responsibility from the central government to the local and provincial governments, allowing for regional variation in family planning policies.

The key relaxation introduced after "Document 7" was the "1-son-2-children" rule, which allowed rural couples to have a second child if their first child was a girl. Initially, only a small percentage of rural families were granted permits for second children in 1982. However, with the implementation of "Document 7", the number of permits increased significantly. By 1986, approximately 50% of rural families were allowed to have a second child if the first child was a girl. Following this relaxation, the fertility policy remained relatively stable and was rigorously implemented across the country until recent years.

After the relaxation of the OCP, the implementation of the "1-son-2-children" rule had a considerable scope across China. As documented by Huang et al. (2019), among the 31 provinces in mainland China, 25 provinces offered this exemption of fine to at least some rural population in their respective prefectures. 19 provinces extended this exemption to cover all rural residents, accounting for 71.3% of the total GDP and 74.9% of the total population in mainland China.

1.2.2 Hukou System

The Hukou system in China is a household registration system that has played a significant role in shaping the country's socio-economic landscape. It was introduced in the 1950s as a means of controlling population movement and resource allocation during China's planned economy era. Under the Hukou system, every individual is assigned a Hukou status, either rural or urban, based on their birthplace and family background. This Hukou status is linked to various entitlements and benefits, including access to education, healthcare, and social services (Chan, 2010).

This led to significant barriers and costs for rural Hukou holders. (i) Barriers to working in different sectors: the Hukou system created obstacles for rural Hukou holders when accessing non-agricultural sectors. Rural migrants face occupational segregation with urban Hukou holders. They usually face hiring discrimination, wage discrimination, and have a much higher probability of facing wage arrears in non-agricultural sectors (Song and Smith, 2019). (ii) Barriers to moving across prefectures: the Hukou system restricts geographic mobility. Relocating to different prefectures involves complex administrative procedures, approvals, and documentation, making it difficult for individuals, especially rural Hukou holders, to seek better economic opportunities or job prospects (Solinger, 1999). Rural Hukou holders receive land for cultivation but lose land rents when migrating due to land market frictions. Access to social services, like education and healthcare, also depends on local Hukou status (Ngai et al., 2019).

1.2.3 Basic Model and Intuition

In this section, I present a simplified model that captures the core channels of the full model to convey the underlying intuition. I focus on how FPPs drive structural transformation by shaping the labor force. Due to the model's simplicity, I can offer an intuitive explanation of these dynamics. The model highlights how important the labor endowment is in changing the economy. FPPs change the number of births, which then changes the labor supply both in the short run and in the long run. This change in the labor supply affects different parts of the economy differently. Industries that rely more on the labor input will grow faster if there is more labor available and vice versa.

The basic model is as follows: consider a country endowed with labor L and land H. These two inputs can be used to produce two consumer goods - agricultural goods (A) and non-agricultural goods (M). Assume both factors are freely mobile within the country across sectors but are immobile across countries.

Production of both goods occurs under conditions of perfect competition and constant returns to scale. I assume that the production technology takes the Cobb–Douglas form, so that output of the good in sector j is

$$q^j = A^j (L^j)^{\gamma_L^j} (H^j)^{\gamma_H^j}, j \in \{A, M\}$$

where, A^j denotes the total factor productivity in sector j and γ_L^j, γ_H^j are the shares of labor and land used in production respectively and $\gamma_L^j + \gamma_H^j = 1$. The agricultural sector is more land intensive and non-agricultural sector is more labor intensive such that $\gamma_L^M > \gamma_L^A$.

In the absence of frictions, profit maximization in each sector necessitates that the marginal profit equals the marginal cost for each factor:

$$P_j \gamma_L^j A^j (L^j)^{\gamma_L^j - 1} (H^j)^{\gamma_H^j} = w$$

$$P_j \gamma_H^j A^j (L^j)^{\gamma_L^j} (H^j)^{\gamma_H^j - 1} = r$$

These optimization conditions, when combined with the market-clearing conditions for labor and land, yield the share of employment in agriculture $s^A \equiv \frac{L^A}{L}$:

$$s^A = C_1 \frac{\gamma_H^A \gamma_L^M}{\gamma_L^M - \gamma_L^A} \frac{1}{L} \left(\frac{P_A}{P_M}\right)^{\frac{1}{\gamma_L^M - \gamma_L^A}} - C_2 \tag{1.1}$$

where C_1 and C_2 are two constants.[9] From this equation, the agricultural employment share is determined by relative factor intensities, changes in labor endowment L, and relative prices.

Small open economy. To highlight the basic channel through which changes in labor endowment affect agricultural employment share, I firstly consider a small open economy scenario. In such a context, relative prices are given by the world prices $\frac{p^A}{p^M} = \frac{p_w^A}{p_w^M}$ and do not respond to any changes in domestic endowments. Equation (1.1) becomes:

$$s^A = C_1 \underbrace{\frac{\gamma_H^A \gamma_L^M}{\gamma_L^M - \gamma_L^A}}_{>0} \frac{1}{L} \Big(\underbrace{\frac{p_w^A}{p_w^M}}_{constant} \Big)^{\frac{1}{\gamma_L^M - \gamma_L^A}} - C_2$$

[9] Specifically, $C_1 = H \left(\frac{A^A (\gamma_L^A)^{\gamma_L^M} (\gamma_H^A)^{\gamma_H^M}}{A^M (\gamma_L^M)^{\gamma_L^M} (\gamma_H^M)^{\gamma_H^M}}\right)^{\frac{1}{\gamma_L^M - \gamma_L^A}}$ and $C_2 = \frac{\gamma_H^M \gamma_L^A}{\gamma_L^M - \gamma_L^A}$.

Therefore, given the factor intensity differences $\gamma_L^M > \gamma_L^A$, an increase in labor endowment L decreases the agricultural employment share. This is a direct result of the Rybczynski Effect in the trade literature (Rybczynski, 1955), which states that an increase in the endowment of one factor will lead to a more than proportional expansion of the output of the good using that factor intensively, and an absolute decline in the output of the other good, holding factor and good prices constant.

This establishes the fundamental intuition that persists throughout my complete model. Notably, this result does not depend on functional form assumptions about preferences, which traditionally underpin the forces of structural transformation. Such forces include income and price effects that rely on non-homothetic preferences (Kongsamut et al., 2001) and the complementarity between consumption goods (Ngai and Pissarides, 2007).

Extending this intuition to a context where prices are endogenously determined, the effect of an increase in L on the agricultural employment share would be mitigated. If the share of agricultural employment decreases, the relative output of agriculture also decreases. The goods market-clearing condition ensures that total output aligns with total demand. Consequently, a reduced demand for agricultural goods would signify an increase in their relative price at equilibrium. This elevation in relative price could, in turn, lead to an offsetting increase in the share of agricultural employment. However, as I will now show in a two-country model that incorporates trade costs, this effect is unlikely to completely neutralize the primary factor intensity effect.

Two-country with trade costs. Suppose there are two countries in the world, differing in terms of endowments and productivity. Assume that transaction costs in international trade take the iceberg form. Labor is mobile across sectors but not across countries. Without loss of generality, I consider the case where Country 1 exports agricultural goods in exchange for non-agricultural goods from Country 2.[10] Taking non-agricultural goods in Country 1 as the numeraire, we have $p_1^M = 1$ and denote $p \equiv p_1^A$. Iceberg trade costs imply that the goods prices in the two countries satisfy $p_2^M = k_1 \cdot p_1^M$ and $p = k_1 \cdot p_2^A$.

In this model, the objective is to investigate the dynamics of the agricultural employment share under the condition of endogenous pricing. World prices are determined by the

[10]Trade patterns within this model are determined by relative autarky prices and the trade cost. These autarky prices are, in turn, a function of fundamentals, including the endowments and productivities across the two sectors. Consequently, the trade pattern varies based on the specific values of these fundamentals and trade costs. It is assumed that these values fall within a feasible domain that allows for the emergence of the trade pattern under consideration. While alternative trade patterns are possible, the implications are similar. For a comprehensive discussion of all potential trade patterns, see Cheng et al. (2004).

global goods market clearing condition. To isolate the impact of changes in endowment from traditional forces, the model assumes identical Cobb-Douglas preference functions for both countries, represented by the aggregate consumption $C = (C^A)^{\phi}(C^M)^{1-\phi}$, where consumers in both countries allocate a fraction ϕ of their total income to agricultural goods and the rest to non-agricultural goods.[11] The equilibrium relative price is given by:

$$p \propto (L_1 + C_3 L_2)^{\gamma_L^M - \gamma_L^A}$$

The relative price of agricultural goods is shown to be positively correlated with the labor endowment in both countries, where $C_3 > 0$ is a constant collecting the model's fundamentals, with the exception of L. The underlying rationale for this correlation is due to the labor intensity disparities between the agricultural and non-agricultural sectors. Specifically, an increase in global labor endowment disproportionately enhances output in the more labor-intensive non-agricultural sector compared to the agricultural sector. To ensure markets clear, the relative price of agricultural goods must increase to preserve the equilibrium share of agricultural consumption.

Given the relative prices, the share of agricultural employment in country 1 is given by

$$s_1^A = C_4 \frac{\gamma_H^A \gamma_L^M}{\gamma_L^M - \gamma_L^A} \frac{L_2}{L_1} + C_5 \tag{1.2}$$

where $C_4 > 0$ and $C_5 > 0$ are constants. Equation (1.2) is similar to equation (1.1). This similarity indicates that even when prices are endogenous, the presence of trade between two countries results in a decline in the share of agricultural employment in country 1 as its own labor endowment increases. This suggests that although the relative price of agricultural goods rises with the labor endowment, this does not completely offset the first order effect in the presence of trade. Consequently, the conclusion reached in the context of a small open economy continues to hold: the share of agricultural employment decreases with an increase in labor endowment.

Demographic transition. I have established the relationship between labor endowment and the share of agricultural employment. Then what is the effect of family planning policies on structural change? The primary effect of FPPs is a reduction in the fertility

[11] With Cobb-Douglas preferences, the price effect and income effect are exactly offset. These forces will be incorporated comprehensively in the full model at a later stage.

rate. To model this, assume households derive utility from both consumption and the number of children, as represented by the following utility function:

$$U = (1-k)C + kn^{\eta}\Pi$$

where k denotes the relative preference for children, C represents aggregate consumption, n is the number of children, and Π is the utility derived from each child by the parents, governed by the parameter η.

Child-rearing requires a time cost; specifically, for each child, parents must allocate a share q of their working hours to childcare. Thus, the budget constraint is given by:

$$P^A C^A + P^M C^M + \chi(n-1)w \leq (1-qn)w$$

where P^A and P^M are the prices of agricultural and non-agricultural goods, respectively, C^A and C^M denote the consumption of these goods, χ represents the additional cost or subsidy per child beyond the first due to the FPPs, and w is the wage rate.

The optimal number of children is determined by equating the marginal utility derived from an additional child to the marginal cost of child-rearing. This yields the following condition for optimal fertility:

$$n = \left(\frac{qw + \chi w}{k\eta\Pi}\right)^{\frac{1}{\eta-1}}$$

Given the initial population $\bar{L}$, the short-run labor supply, which reflects the period when this population makes fertility decisions, is given by $(1-qn)\bar{L}$. This represents the labor force available after accounting for the time parents allocate to child-rearing. In contrast, the long-run labor supply, which pertains to the period when the population is replaced by the offspring, is given by $n\bar{L}(1-qn')$, where n' denotes the fertility decisions made by these offspring when they reach maturity. This captures the generational change in the labor force as a function of fertility rates and the associated time costs of child-rearing.

FPPs induce a decrease in n when $\eta < 1$, which in turn affects labor supply and structural transformation across different time frames. In the short run, the decline in fertility rates leads to a reduction in the time parents dedicate to child-rearing (qn), resulting in an immediate increase in the labor supply. This boost in labor availability can accelerate structural transformation through the channel discussed above. In contrast, the lower fertility rate leads to a contraction in the labor force in the long run as the current population ages and fewer individuals enter the working-age group. This declining labor force can impede the progress of structural transformation. Hence, while FPPs may facilitate a more rapid

structural transformation in the short run due to an expanded labor supply, they have the potential to hamper it in the long run as the labor force diminishes.

In this basic model, I integrate the concept of factor intensity differences alongside trade and endogenous fertility choices to demonstrate that FPPs precipitate a trade-off in structural transformation between the short run and the long run. The underlying intuition is driven by a Rybczynski type effect, which elucidates the relationship between factor endowments and input in different sectors in the context of international trade, and by the demographic transition, which is influenced by the time costs associated with child-rearing. However, the basic model does not yet encompass traditional forces such as non-homothetic preferences, Hukou system frictions, TFP progress, and migration patterns. These elements are critical for a comprehensive analysis and will be integrated into the full model. Their separate effects will be quantified to delineate their distinct impacts on structural transformation, providing a more complete picture of the economic transformation process.

1.3 The Model

In this section, I develop the full dynamic geographic model to study the effect of the FPPs while maintaining the underlying mechanisms: the changes in total labor force combined with sector factor intensity differences. While my model retains the traditional driving forces of structural transformation, such as technological change and non-homothetic preferences, it goes beyond existing spatial structural change models, for example Eckert and Peters (2022), by integrating the endogenous dynamic fertility decision-making framework from family economics. This feature allows for an examination of how fertility changes interact with labor force dynamics and economic structure.

The basic environment is as follows. Time is discrete. The economy consists of N regions in China indexed by i, n and one rest of the world. Each region stands for one prefecture in China. There are two sectors in each region indexed by j, k. The agricultural sector is perfectly competitive indexed by A, with a continuum of firms in each region. The non-agricultural sector is monopolistically competitive, indexed by M, and characterised by firm entry cost. There are competitive labor markets in each region sector subject to migration costs, meaning that labor can move across prefectures and sectors subject to a cost. Individuals have sector-specific Hukou indexed by $h \in \{A, M\}$, where A means rural Hukou (agricultural Hukou) and M means urban Hukou (non-agricultural Hukou). The Hukou affects individual's cost of shifting between sectors as well as their migration cost across prefectures. Prefectures are different in terms of sectoral technology as well

as fundamental amenities. There are both inter-regional trade and international trade in final goods in two sectors.

1.3.1 Timing

I consider an overlapping generation (OLG) framework. In this framework, the population is composed of individuals belonging to different generations, each living through two distinct life periods: childhood and adulthood. These individuals are characterized by their place of birth (region i) and Hukou status (h), which are predetermined attributes.

At the beginning of time t, individuals of generation t are born in region i with Hukou h and spend their childhood in that region. This Hukou status is inherited from their parents and remains fixed throughout their life.[12] During their childhood, individuals do not make any decisions.

At the end of each period t, individuals observe an idiosyncratic preference shock for each location. This shock influences their future decisions regarding migration and sectoral employment. At this point, individuals make migration decisions, selecting a new region n and a sector j based on the highest expected utility, considering preference shocks, migration costs, and Hukou restrictions as well as the expected utility they get from future fertility decision making in each region.

When the next period $(t+1)$ begins, a new generation $(t+1)$ is born in the region n to which their parents migrated. The newborn generation starts their childhood in this new location. During period $t+1$, individuals belonging to the previous generation (t) make fertility choices based on the expected utility associated with each child and the costs of child-rearing. These fertility decisions directly affect the size and composition of the next generation and, in turn, have implications for the future labor force and population distribution across prefectures.

At the end of period $t+1$, individuals of generation t pass away, and individuals of generation $t+1$ migrate to region n' and work in sector j'. This process of migration, sectoral employment, and fertility choices continues over time, generating dynamics in labor supply and economic outcomes in the model. I summarize the timeline below:

- At time t, generation t is born in region i with Hukou h and enjoy their childhood in region i.

[12] This is largely true in reality but can also change in some cases. For example, there are 0.2% permits in most provinces during 1990s that allowed people to change from rural Hukou to urban Hukou.

- At the end of t, generation t migrates to region n and works in sector j with highest expected utility accounting for preference shock, migration cost and Hukou restrictions.
- At the beginning of $t+1$, they make fertility choices based on the expected utility of each child and the child raising cost.
  - generation $t+1$ are born and start their childhood in region n.
- At the end of $t+1$, generation t die.
  - generation $t+1$ migrate to region n' and sector j'.

1.3.2 Preferences

Preferences of each individual are determined by two components: consumption and offspring. (1) Individuals derive utility from consuming goods produced in the agricultural and non-agricultural sectors. This part of the preference structure is assumed to follow a non-homothetic preference of the "PIGL" (Price-Independent Generalized Linear) class. While it does not have an explicit utility representation, the indirect utility functional form makes it convenient for investigating the trade-offs between consumption of goods and fertility decisions when there is a shock in the cost of raising children. (2) Individuals derive utility from the expected utility that their children will have in the next generation. This is determined by the future migration decision, work decision, and fertility decisions of the next generation.

Raising children comes with significant costs for parents. It requires time and effort to care for and nurture children during their childhood. This involves taking time out of the labor force to provide parental care, which can lead to a reduction in the labor supply and total income. The opportunity cost of the time spent on child-rearing is an important consideration for parents when making fertility decisions and is a key driver of the labor force dynamics in the short run.

Additionally, FPPs impose penalties on parents who choose to have more than the permitted number of children. One common form is a financial penalty, where parents are required to pay a fine or a share of their income as a punishment for exceeding the allowed family size.

Given these components, individuals face a trade-off between spending money on consumption and spending time on raising children. The structure of preferences is as follows:

$$U_{i,h,t} = (1-k_i)(\frac{1}{\eta}(\frac{y_{i,h,t}}{p_{i,A,t}^{\phi} p_{i,M,t}^{1-\phi}})^{\eta} - \nu \ln(\frac{p_{i,A,t}}{p_{i,M,t}})) + k_i n_{i,h,t}^{\eta_f} \Pi_{i,h,t+1}$$

where $U_{i,h,t}$ represents the utility of an individual living in region i with Hukou h at time t. $\frac{1}{\eta}(\frac{y_{i,h,t}}{p_{i,A,t}^{\phi}p_{i,M,t}^{1-\phi}})^{\eta} - \nu \ln(\frac{p_{i,A,t}}{p_{i,M,t}})$ is the PIGL consumption component where $y_{i,h,t}$ is the expenditure spent on total consumption and $n_{i,h,t}^{\eta_f}\Pi_{i,h,t+1}$ is utility from offspring. k_i is a regional specific parameter that determines the weight of the offspring component in the utility function. It captures the regional-specific variation in individuals' preferences for offspring and allows me to account for the fact that different prefectures may have different cultural, social, and economic factors that influence individuals' fertility decisions other than the effect of FPPs. $\eta \in (0,1)$ and $\nu > 0$ are structural parameters of the PIGL function which determine the shape of Engel curve. $\phi \in (0,1)$ is the asymptotic expenditure share on agricultural goods. $\eta_f < 1$ is a parameter that determines the individual's preference for having more children. $p_{i,A}$ and $p_{i,M}$ represent the prices of agricultural and non-agricultural goods, respectively, in region i. $\Pi'_{i,h}$ represents the expected utility that the individual's children can obtain in the future.

The incorporation of the PIGL utility function represents a significant departure from the basic model in two distinct ways.

Firstly, it introduces non-homothetic preferences, which imply that the proportion of income spent on agricultural consumption diminishes as income increases. This characteristic generates an income effect that plays an important role in driving structural transformation. An increase in the labor force exerts downward pressure on real wages, which, in turn, amplifies the share of agricultural consumption relative to the model with homothetic preferences, where the share remains constant. Consequently, the presence of non-homothetic preferences moderates the primary channel of structural transformation delineated in the basic model.

Secondly, the PIGL utility function yields income-contingent consumption shares. This adds a layer of complexity when considering the Hukou system. Given that individuals with different Hukou statuses typically have different income levels, particularly when employed in distinct sectors, the aggregate consumption effect in each region becomes a function of the labor composition by Hukou status and sector choices. The Hukou costs affect not only individual income and consumption patterns but also regional aggregate structural changes. Specifically, the share of expenditure on agriculture as a share of total consumption is given by:

$$\theta_{A,h,t}(y_{i,h,t}, P_t) = \phi + \nu(\frac{y_{i,h,t}}{p_{i,A,t}^{\phi}p_{i,M,t}^{1-\phi}})^{-\eta_c} \tag{1.3}$$

As GDP per capita increases, the limit share of expenditure on agriculture converges to ϕ, which I assume to be the share of expenditure on agriculture for local authorities.

Given the preference, individuals in each location also have the following budget constraint to make fertility and consumption choices simultaneously:

$$y_{i,h,t} + \chi_{i,h,t}(n_{i,h,t} - 1)I_{i,h,t} = (1 - qn_{i,h,t}^{\eta_q})I_{i,h,t} \tag{1.4}$$

where $qn_{i,h,t}^{\eta_q}$ is the total time cost associated with child-rearing, expressed as a proportion of the total working time devoted to each child. The parameter q represents the baseline time cost of raising a single child, while the exponent η_q captures the degree of diminishing marginal time cost per child. This formulation reflects the hypothesis that the time required for childcare may decrease at the margin as the number of children increases, due to potential economies of scale in parental time investment. $\chi_{i,h,t}$ is a penalty parameter that reflects the cost or penalty imposed on parents who choose to have more children than the permitted number allowed by the FPPs. The FPPs penalties are collected by local authorities and spent on consumption. This penalty is proportional to the number of excess children beyond the permitted number.[13] $I_{i,h,t}$ denotes the total income of the individual in region i with Hukou status h.

With this structure at hand, optimal fertility is given by the following equation:

$$n_{i,h,t}^{\eta_f - 1}(1 - qn_{i,h,t} - \chi_{i,h,t}(n_{i,h,t} - 1))^{1-\eta} = \left[\frac{(1 - k_i)(q\eta_q n_{i,h,t}^{\eta_q - 1} + \chi_{i,h,t}) \left(\frac{I_{i,h,t}}{p_{i,A,t}^{\phi} p_{i,M,t}^{1-\phi}} \right)^{\eta}}{\eta_f k_i \beta \Pi_{i,h,t+1}} \right] \tag{1.5}$$

Equation (1.5) shows the determinants of fertility rates within the model, given parameter assumptions. It demonstrates that fertility rates are decreasing in the time cost of child-rearing and the financial penalties imposed by FPPs, as well as to real wage levels. Conversely, fertility rates are increasing in the relative preference for children and the expected utility derived from each child. This relationship suggests that as the costs associated with raising children — in terms of financial resources — increase, the incentive for having additional children diminishes, thereby capturing the effect of FPPs on fertility rates.

[13] In most cases in China, the fines take the form of a share of total income for each child exceeding the permit. However, there are also other non-monetary punishments such as losing jobs. I aim to calibrate the model to back out the effective monetary equivalent punishment for the actual monetary and non-monetary fines.

It is also important to acknowledge that in the absence of FPPs, the model inherently captures the natural evolution of fertility rates. These rates are endogenously determined by the time costs associated with child-rearing, shifts in preferences towards childbearing, and the prevailing real wage levels. This mechanism allows for the examination of demographic changes as a function of economic and preference variables, distinguishing the effect of policy interventions.

When $\eta = 1, \eta_q = 1$, the PIGL preference reduces to a Cobb-Douglas preference and the time cost for childcare is constant. Then it is convenient to back out the optimal number of offspring as a function of costs as well as real wages: $n_{i,h,t} = [\frac{(1-k)(q+\chi_{i,h,t})\left(\frac{I_{i,h,t}}{p_{i,A,t}^{\phi} p_{i,M,t}^{1-\phi}}\right)^{\eta}}{\eta_f k \beta \Pi_{i,h,t+1}}]^{\frac{1}{\eta_f - 1}}$. An increase in the FPPs fines reduces the optimal number of children.

The unique aspect of this dynamic spatial model is that the fertility decisions $n_{i,h,t}$ made by individuals of generation $t-1$ at time t are influenced by future utilities which affect labor distribution at time $t+1$. This introduces a forward-looking element to the fertility decision-making process, making it more dynamic compared to other spatial models.

1.3.3 Income and Migration

Individuals make decisions about their location and sector choices based on expected utility, migration costs and idiosyncratic preference shock. The migration process consists of two steps: location choice and sector choice upon arrival.

In the first step, individuals belonging to generation $t-1$ decide whether to migrate from their current region r to another region n that offers the highest utility $\mathscr{U}^i_{rn,h,t}$. This is determined by (i) the utility in region n, which is determined by the expected utility $E(U_{n,h,t})$, which captures the consumption and offspring utilities an individual with Hukou h can derive from living in region n; (ii) amenity level $A_{n,t}$ in region n, reflecting the attractiveness of the location; (iii) idiosyncratic preference shocks $\epsilon^i_{n,h,t}$ which follow an i.i.d. Fréchet distribution $\mathscr{F}(\epsilon^i_{n,h,t}) = e^{-(\epsilon^i_{n,h,t})^{-\epsilon}}$; and (iv) the migration cost from region r to region n for Hukou h denoted by $\tau_{rn,h,t}$:

$$\mathscr{U}^i_{n,h,t} = A_{n,t} E(U_{n,h,t}) \frac{\epsilon^i_{n,h,t}}{\tau_{rn,h,t}}$$

Given this structure and the distribution of idiosyncratic preference shocks, the share of migrants from region r to region n for Hukou h is represented by:

$$m_{rn,h,t} = \frac{(\bar{u}_{n,h,t}/\tau_{rn,t})^{\epsilon}}{\Sigma_m(\bar{u}_{m,h,t}/\tau_{rm,t})^{\epsilon}} \tag{1.6}$$

where $\bar{u}_{n,h,t} \equiv A_{n,t}E(U_{n,h,t})$ and ϵ is the parameter of the Fréchet distribution that governs the elasticity of migration across prefectures.

In the second step, individuals who migrate to region n make sector choices upon arrival. They select the sector j that offers the highest income, taking into account the match productivity $\phi^j_{n,t}$ drawn from a sector-specific Fréchet distribution $F(\phi^j_{i,t}) = e^{-T_j(\phi^j_{i,t})^{-\zeta}}$ and the wage per efficient labor $w^j_{n,t}$ in region n sector j.

Additionally, individuals with Hukou h face costs $\delta^n_{j,h,t}$ for working in sector j due to Hukou restrictions. These costs represent any constraints imposed on individuals with Hukou h working in specific sectors, particularly for rural Hukou working in non-agricultural sectors. In the current framework, the multifaceted costs associated with Hukou restrictions are not disaggregated. Following Ngai et al., (2019), the primary costs are identified as the loss of land income for rural Hukou holders working in non-agricultural sectors and the ineligibility for local social benefits when migrating to another region. My model abstracts from these specificities to only consider the aggregate cost of Hukou restrictions on sectoral labor choices.

Sectoral choice is made based on the highest income after taking the Hukou-specific cost into account. The total wage income individuals earn in each sector is given by the match productivity and wage per effective labor, $\phi^j_{n,t}w^j_{n,t}$. However, they need to pay the Hukou cost to the local authority to receive the respective income in each sector. The net wage income in each sector is therefore $\frac{\phi^j_{n,t}w^j_{n,t}}{\delta^n_{j,h,t}}$, and the remaining amount, $\frac{(\delta^n_{j,h,t}-1)\phi^j_{n,t}w^j_{n,t}}{\delta^n_{j,h,t}}$, is collected by local authorities and spent on consumption. Individuals then choose the sector that maximizes the following income:

$$I_{n,h,t} = \max_j \left\{ \frac{\phi^j_{n,t}w^j_{n,t}}{\delta^n_{j,h,t}} \right\}$$

Given this maximization structure and the distribution of productivity preference shocks, the share of individuals with Hukou h working in sector j upon arriving in region n is given by:

$$s^j_{n,h,t} \equiv \frac{L^j_{n,h,t}}{L_{n,h,t}} = \frac{T_j(w^j_{n,t}/\delta_{j,h,t})^\zeta}{\Sigma_i T_i(w^i_{n,t}/\delta_{i,h,t})^\zeta} \tag{1.7}$$

where $L^j_{n,h,t}$ is the number of people in region n with Hukou h who choose to work in sector j and $L_{n,h,t}$ is the total number of people in region n with Hukou h. T_j is the structural parameter that determines the average match probability in sector j, and ζ is the parameter of the Fréchet distribution that governs the elasticity of migration across sectors.

Income and expenditure: The distribution of income is also given by the Fréchet distribution:

$$G_{n,h}(I) = e^{-\Sigma_j T_j \left(\frac{\delta_{j,h}}{w^j_n} I\right)^{-\zeta}} \tag{1.8}$$

The expenditure spent on each sector in each region comes from the consumption by labor and local authorities. The share of consumption from labor on agriculture is given by equation (1.3), and the share of consumption on agriculture from local authorities is the limit share ϕ. The total expenditure on agriculture in each region is given by:

$$E^A_{i,t} = \int \theta_{A,h,t} y_{h,t} L^A_{n,t}\, dG_{n,h,t}(I) + \phi \left(\Sigma_j X^{i,j}_t - \int y_{A,h,t} L^A_{n,t}\, dG_{n,h,t}(I)\right) \tag{1.9}$$

where $\Sigma_j X^{i,j}_t$ is the total output in each region. Under the trade balance condition, the total output in each region equals the total income in each region. The difference between the total income and wage income, and the total expenditure on non-agriculture in each region is given by:

$$E^M_{i,t} = \int (1-\theta_{A,h,t}) y_{h,t} L^A_{n,t}\, dG_{n,h,t}(I) + (1-\phi) \left(\Sigma_j X^{i,j}_t - \int y_{A,h,t} L^A_{n,t}\, dG_{n,h,t}(I)\right) \tag{1.10}$$

Demographic transition: With the migration decision $m_{rn,h,t}$ at hand, given the equilibrium labor supply $L_{r,h,t-1}$ at time $t-1$ in each region r for each Hukou h and the fertility decision $n_{i,h,t-1}$ at time $t-1$, the equilibrium labor supply (generation t) at time t in each region n is given by:

$$L_{n,h,t} = \Sigma_r m_{rn,h,t} \bar{L}_{r,h,t} \int (1 - q n^{\eta_q}_{n,h,t}) dG_{n,h,t}(I) \tag{1.11}$$

The integral is taken over the distribution of income since consumers with different realizations of income will have different fertility decisions. The aggregate labor supply is the sum over all Hukou under the realization of income following the $G_{n,h,t}(I)$ distribution, net of the time share spent on childcare. The $\bar{L}_{r,h,t-1}$ represents the population making migration choices at the end of period $t-1$. This is the total population of generation $t-1$, determined by the fertility decisions and labor allocation at time $t-1$:

$$\bar{L}_{r,h,t} = L_{r,h,t-1} \int \frac{n_{r,h,t-1}}{1 - q n_{r,h,t}^{\eta_q}} dG_{r,h,t-1}(I) \tag{1.12}$$

where $L_{r,h,t-1}$ denotes the total labor force in region r with Hukou status h at time $t-1$, post-fertility decision-making. The term $L_{r,h,t-1}/(1-qn_{n,t-1})$ represents the adult population at the onset of period $t-1$, who are then engaged in fertility decisions. This adult cohort, when multiplied by the fertility rate $n_{n,t-1}$, yields the count of the subsequent generation still in childhood at the end of period $t-1$ and residing with their parents. As they transition into period t, this younger generation, having made their migration decisions denoted by $m_{rn,h,t}$, enters the labor force in their respective regions after accounting for the time allocated to childrearing, as indicated by their fertility choices $n_{n,t}$. Aggregating these figures across all destination prefectures r gives the total labor supply in region r at time t for each Hukou h.

Equation (1.11) is the law of motion for labor supply, which shows the demographic transition mechanism within the model. This equation indicates that the labor supply in each region is contingent upon both antecedent and contemporaneous fertility choices. The immediate, or short-run, effect on labor supply is embodied by the term $n_{n,t}$, reflecting the reduction in labor supply following the current period's fertility decisions. Conversely, the long-run effect is captured by $n_{n-1,t}$, which increases the labor supply in the next period. These two effect combine gives the demographic transition in the model following FPPs.

Since labor has different productivities $\phi_{n,t}^{j}$ following the Fréchet distribution, denote variables in effective unit as $\tilde{x}$, the effective unit of labor supply $\tilde{L}_{n,h,t}^{j}$ is given by:

$$\tilde{L}_{n,h,t}^{j} = L_{n,h,t} T_j^{\frac{1}{\zeta}} \Gamma\left(1 - \frac{1}{\zeta}\right) \left(s_{n,h,t}^{j}\right)^{1-\frac{1}{\zeta}}$$

1.3.4 Production

In each region, firms operate in both the agricultural and non-agricultural sectors, utilizing a Constant Returns to Scale (CRS) production technology that employs land and labor as

inputs. These regions exhibit heterogeneity in their sectoral TFP. Firms can make entry or exit decisions but are unable to change their locations.

Non-agricultural sector: Each firm produces a unique variety of non-agricultural goods denoted as $\omega_{n,t}^M$. The production technology follows a Cobb-Douglas production function with labor ($L_{n,t}^M$) and land ($H_{n,t}^M$) as inputs:

$$q_{n,t}^M(\omega_{n,t}^M) = A_{n,t}^M(\tilde{L}_{n,t}^M(\omega_{n,t}^M))^{\gamma_L^M}(H_{n,t}^M(\omega_{n,t}^M))^{\gamma_H^M}$$

where $A_{n,t}^M$ denotes the total factor productivity in the non-agricultural sector in region n and γ_L^M, γ_H^M are the shares of labor and land used in production, respectively. I assume a constant returns to scale technology such that they add up to 1.

In the current framework, capital is not explicitly modeled as a separate input; instead, its effects are subsumed within the measure of TFP. This approach is justified if, during the period in my study, capital deepening has not been substantial and has not exhibited a significant correlation with the FPPs. Though accounting for a large share of GDP, the capital's share in GDP is relatively constant during the period under study. Nonetheless, it is acknowledged in the literature that FPPs can influence savings rates, which in turn may contribute to capital accumulation. For instance, Choukhmane et al. (2023) suggests that the One-Child Policy could explain at least 30% of the increase in aggregate savings. To incorporate this dynamic, the model posits that non-agricultural TFP is inversely related to fertility rates. This assumption allows for the possibility that households with fewer children may save more, thereby potentially accelerating capital deepening and structural transformation. This indirect linkage between fertility decisions and economic growth through savings and investment is an important consideration in the broader analysis of FPPs' impact on economic development. The non-agricultural TFP takes the following form:

$$A_{n,t}^M = \bar{A}_{n,t}^M(1 - qn_{n,t}^{\eta_q})^g \tag{1.13}$$

where $\bar{A}_{n,t}^M$ represents the baseline level of non-agricultural TFP, and g captures the contribution of the capital deepening effect to TFP. The term $(1 - qn_{n,t}^{\eta_q})$ models the influence of fertility rates on savings and, consequently, on capital accumulation. This formulation aligns with the approach taken by Choukhmane et al. (2023) to examine the savings rate

within an overlapping generations model, providing insights into the economic ramifications of China's FPPs.[14]

Final non-agricultural goods are produced using varieties from all locations. The aggregate production of non-agricultural goods in region n is denoted as $Q_{n,t}^M$ and is given by the following expression:

$$Q_{n,t}^M = \left(\Sigma_i \int (q_{i,n,t}^M(\omega))^{1-1/\eta_t^M} \, d\omega \right)^{\eta_t^M/(\eta_t^M-1)}$$

where the integral sums up the quantities of all varieties produced in different prefectures, and η_t^M represents the elasticity of substitution between different varieties. The final goods produced in the non-agricultural sector are sold only in region n.

Each variety producer in the non-agricultural sector faces monopolistic competition, which means they are able to set its price. Given the demand from the final non-agricultural goods producer, the price for each variety $\omega_{n,t}^M$ produced in region i and sold in region n is given a constant markup over the unit cost $c_{i,t}^M$:

$$p_{i,n,t}^M(\omega) = \frac{\eta_t^M \kappa_{in,t}^M c_{i,t}^M(\omega)}{\eta_t^M - 1}$$

where $\kappa_{in,t}^M \geq 1$ is the iceberg transportation cost between region i and n.

The monopolistic competition is employed to reflect the non-agricultural market conditions in China throughout the period, which is characterized by the pronounced market power of State-Owned Enterprises (SOEs). This market structure allows for the examination of the economic implications of SOEs' dominance and the consequences of SOE reforms during the period.

Firms need to pay a region-specific fixed entry cost $f_{e,t}^n$ measured in effective unit labor. The free entry condition implies that profits in the non-agricultural sector are equal to the entry cost, which determines the number of firms in each region:

$$N_{n,t}^M \left(f_{e,t}^n + \gamma_L^M (\eta^M - 1) f_{e,t}^n \right) = \tilde{L}_{n,t}^M \tag{1.14}$$

[14] In their OLG model, the asset holding of a middle-aged individual is given by fertility rates, cost on children, and inter-generational transfers. Plugging in the setting in my model gives the functional form of asset holding, which is a linear function of the effective labor supply $1 - q n_{n,t}^{\eta_q}$.

In the monopolistic competition market, the share of total expenditure in region n on non-agricultural varieties from region i is represented by the following equation:

$$\lambda_t^{in,M} = \frac{N_{i,t}^M (\kappa_{in,t}^M c_{i,t}^M)^{1-\eta^M}}{\Sigma_m N_{m,t}^M (\kappa_{mn,t}^M c_{m,t}^M(\omega))^{1-\eta^M}} \tag{1.15}$$

Given the expenditure share from the consumer side, the non-agricultural goods market clearing condition gives the wage as follows:

$$w_{n,t}^M = \frac{1}{\eta^M f_e} \left(\frac{\eta^M}{\eta^M - 1} \right)^{1-\eta^M} c_{n,t}^M(\omega)^{1-\eta^M} \Sigma_i \left[\left(\frac{\kappa_{in,t}^M}{P_{n,t}^M} \right)^{1-\eta^M} E_{n,t}^M \right] \tag{1.16}$$

where $E_{n,t}^M$ is the total expenditure in region n on sector M.

The agricultural sector: This follows the Eaton-Kortum structure. Each location produces a continuum of agricultural varieties $\omega \in [0, 1]$ using Cobb-Douglas technology:

$$q_{n,t}^A(\omega) = z_{n,t}^A(\omega) A_{n,t}^A (\tilde{L}_{n,t}^A)^{\gamma_L^A} (H_{n,t}^A)^{\gamma_H^A}$$

where $\tilde{L}_{n,t}^A$ represents the labor input in agriculture, $H_{n,t}^A$ denotes the land input in agriculture, and $z_{n,t}^A(\omega)$ is a productivity factor specific to each variety ω in region n. The productivity factor $z_{n,t}^A(\omega)$ is drawn from an i.i.d. Fréchet distribution with a cumulative distribution function given by $F(z_{n,t}^A) = e^{-(z_{n,t}^A)^{-\theta^A}}$. The parameters γ_L^A and γ_H^A represent the shares of labor and land, respectively, in the production of agricultural goods, while $A_{n,t}^A$ denotes the average total factor productivity in the agricultural sector in region n.

The final agricultural goods are produced by combining all varieties $\omega \in [0, 1]$ available in the market. The quantity of final goods produced in region n, denoted by $Q_{n,t}^A$, is determined by aggregating the production of each variety using a CES aggregator:

$$Q_{n,t}^A = \left(\int (\tilde{q}_{n,t}^A(\omega))^{1-1/\eta_t^A} \, d\omega \right)^{\eta_t^A/(\eta_t^A - 1)}$$

The CES aggregator combines the individual varieties $\tilde{q}_{n,t}^A(\omega)$ from different prefectures. The final agricultural goods are sold exclusively in region n.

The final agricultural goods producers source each variety ω from prefectures with the lowest price. Given the Fréchet distribution of $z_{n,t}^A(\omega)$, the share of total expenditure in region n on agricultural varieties from region i is given by:

$$\lambda_t^{in,A} = \frac{(\kappa_{in,t}^A c_t^{iA})^{-\theta^A}}{\Sigma_{m=1}^N (\kappa_{nm,t}^A c_t^{mA})^{-\theta^A}} \tag{1.17}$$

where $\kappa_{in,t}^A$ represents the iceberg transportation cost from region i to region n for agricultural varieties, and c_t^{iA} is the unit cost of agricultural varieties in region i. The price of final agricultural goods in region n, denoted by $P_{n,t}^A$, is determined by aggregating the prices of individual varieties using a CES aggregator: $P_{n,t}^A = \Gamma^{nA} \left(\Sigma_{m=1}^N (\kappa_{mn,t}^A c_t^{mA})^{-\theta^A}\right)^{-1/\theta^A}$.[15]

Given the expenditure share from the consumer side, the agricultural goods market clearing condition gives the wage as follows:

$$w_{n,t}^A = \frac{\gamma_L^A}{\tilde{L}_{n,t}^A} (c_t^{nA})^{-\theta^A} \Sigma_i \left(\frac{\kappa_{in,t}^A}{P_{n,t}^A}\right)^{-\theta^A} E_{i,t}^A \tag{1.18}$$

where $E_{n,t}^A$ is the total expenditure in region n on sector A.

1.3.5 Equilibrium and Determinants of Industrialization

In this section, I define the equilibrium in the economy and characterize the determinants of industrialization.

Definition: Given the initial distribution of labor endowments $\{L_{r,h,0}\}_{r\in N,h\in\{A,M\}}$, land endowments $\{H_{r,0}\}_{r\in N}$, fertility $\{n_{r,h,0}\}_{r\in N,h\in\{A,M\}}$, a path of fundamentals including FPPs fines $\{\chi_{r,h,t}\}_{r\in N,h\in\{A,M\},t\geq 0}$, productivity $\{A_{r,t}^A, A_{r,t}^M\}_{r\in N,t\geq 0}$, Hukou friction $\{\delta_{r,t,h}^A, \delta_{r,t,h}^M\}_{r\in N,h\in\{A,M\},t\geq 0}$, migration cost $\{\tau_{o,r,t,h}\}_{o,r\in N,h\in\{A,M\},t\geq 0}$, and trade cost $\{\kappa_{o,r,t,s}\}_{o,r,s\in N,t\geq 0}$, an equilibrium is a sequence of prices $\{p_{r,t}^A, p_{r,t}^M\}_{r\in N,t\geq 0}$ and wages $\{w_{r,t}^j, r_{r,t}^j\}_{r\in N,t\geq 0,j\in\{A,M\}}$ such that the following conditions are satisfied:

- Goods market clearing ensures that the total output $X_t^{i,j}$ in both agricultural and non-agricultural sectors in region i at time t must equal the total demand for that output:

  $$X_t^{i,j} = \Sigma_n \lambda_t^{in,j} E_{n,t}^j$$

  where the trade share $\lambda_t^{in,j}$ is given by equations (1.17) and (1.15), and the local sectoral expenditure $E_{n,t}^j$ is given by equations (1.9) and (1.10).

[15] $\Gamma^{nA} \equiv \Gamma\left(1 + \frac{1-\eta^A}{\theta^A}\right)$ and $\Gamma(.)$ is the Gamma function.

- Labor market clearing conditions satisfy equations (1.18) and (1.16), where the unit costs are:

$$c_{i,t}^{M}(\omega) = \frac{1}{A_{i,t}^{M}}(w_{i,t}^{M})^{\gamma_L^M}(r_{i,t}^{M})^{\gamma_H^M}k_M,$$

$$c_{i,t}^{A}(\omega) = \frac{1}{A_{i,t}^{A}}k_A(w_{i,t}^{A})^{\gamma_L^A}(r_{n,t}^{A})^{\gamma_H^A}$$

  and the price indices are given by:

$$P_t^{nM} = \left(\Sigma_i N_{i,t}^{M}\left(\frac{\eta_t^M \kappa_{in,t}^{M} c_{i,t}^{M}(\omega)}{\eta_t^M - 1}\right)^{1-\eta_t^M}\right)^{\frac{1}{1-\eta_t^M}}$$

$$P_t^{nA} = \Gamma^{nA}\left(\Sigma_{m=1}^{N}(\kappa_{mn,t}^{A}c_t^{mA})^{-\theta^A}\right)^{-1/\theta^A}$$

  and the number of firms is given by equation (1.14).
- Land market clearing condition and the first-order condition from the firm's problem give land prices:

$$\frac{w_{n,t}^{j}\tilde{L}_{n,t}^{j}}{r_{n,t}^{j}H_{n,t}^{j}} = \frac{\gamma_L^j}{\gamma_H^j}$$

- Law of motion for labor satisfies equations (1.11) and (1.12) where the migration share is given by equation (1.6).
- Labor supply in each sector is:

$$L_{n,h,t}^{j} = s_{n,h,t}^{j}L_{n,h,t}$$

  where the share in each sector $s_{n,h,t}^{j}$ is given by equation (1.7).
- The expected utility in each region is given by:

$$E(U_{i,h,t}) = \int_0^\infty \left[(1-k_i)\left(\frac{1}{\eta}\left(\frac{y_{i,h,t}^{*}}{p_{i,A,t}^{\phi}p_{i,M,t}^{1-\phi}}\right)^{\eta} - \nu\ln\left(\frac{p_{i,A,t}}{p_{i,M,t}}\right)\right) + k_i n_{i,h,t}^{*\eta_f}\Pi_{i,h,t+1}\right]dG_{n,h}(I) \tag{1.19}$$

  with $\Pi_{r,h,t} = E\left(\max_n\left\{\bar{u}_{n,h,t}\frac{\epsilon_{n,h,t}^{i}}{\tau_{rn,h,t}}\right\}\right) = \Gamma\left(1-\frac{1}{\epsilon}\right)\left(\Sigma_m\left(\frac{\bar{u}_{m,h,t}}{\tau_{rm,t}}\right)^{\epsilon}\right)^{\frac{1}{\epsilon}}$, and the optimal fertility $n_{i,h,t}^{*}$ and expenditure on consumption $y_{i,h,t}^{*}$ are determined by the budget constraint (1.4) in each period and the consumer utility maximization problem (1.5). The distribution of income follows equation (1.8).

The dynamic equilibrium describes the full transition of economic activities over time and space. The structural transformation in each region can be characterized by the following proposition:

Proposition 1. *The changes in agricultural employment share in region n is determined by the following forces:*

$$
\begin{aligned}
d\ln \tilde{s}^A_{n,t} = & \underbrace{[\frac{\theta^a}{1+\theta^a} d\ln(A^A_{n,t}) - \frac{(\eta^M-1)}{\eta^M} d\ln(A^M_{n,t})]}_{\text{comparative advantage}} + \underbrace{[\frac{1}{1+\theta^a} d\ln \mathscr{M}^A_{n,t} - \frac{1}{\eta^M} d\ln \mathscr{M}^M_{n,t}]}_{\text{difference in market access}} \\
& + \Sigma d\ln \tilde{L}_{n,t,h} \left[\left(\underbrace{\frac{\tilde{L}^M_{n,t,h}}{\tilde{L}^M_{n,t}} \frac{\gamma^M_H(\eta^M-1)}{\eta^M} - \frac{\tilde{L}^A_{n,t,h}}{\tilde{L}^A_{n,t}} \frac{(\theta^A\gamma^A_H+1)}{1+\theta^A}}_{\text{factor intensity effect}} \right) + \left(\underbrace{\frac{\tilde{L}^A_{n,t,h}}{\tilde{L}^A_{n,t}} - \frac{\tilde{L}_{n,t,h}}{\tilde{L}_{n,t}}}_{\text{hukou composition effect}} \right) \right] \\
& + \underbrace{\frac{1}{\eta^M} d\ln f^e_{n,t}}_{\text{entry cost}} + \frac{\zeta}{|C|} \frac{1}{\tilde{s}^M_{n,t,M}} \frac{\tilde{L}^A_{n,t,A}}{\tilde{L}^A_{n,t}} \underbrace{d\ln \delta^n_{M,t,A}}_{\text{hukou cost}}
\end{aligned}
$$

where $\tilde{x}$ represents variables in effective units adjusted by individuals' match productivity.

The pace and nature of regional structural transformation are shaped by several key economic variables. These include productivity levels across sectors, the ease of market access, the dynamics of labor supply, the costs associated with firm entry, and the institutional constraints imposed by Hukou systems. Each of these factors contributes to the shifting landscape of employment and sectoral growth within a region.

Productivity: The productivity in agricultural and non-agricultural sectors fundamentally shapes the comparative advantage of a region. Higher productivity in agriculture can retain labor within the sector, while higher productivity in manufacturing or services can draw labor away, fostering a shift towards these sectors. This observation may seem at odds with the prevailing discourse in structural transformation literature, which typically argues that a faster increase in agricultural TFP is conducive to structural transformation. However, it is crucial to recognize that the first term in the equation primarily reflects the immediate impact of comparative advantage driven by productivity differentials. In a trade-integrated economy, these first-order effects can dominate the income and price effects, dictating the direction of labor movement within and across regions.

Market access: This reflects the dynamics of demand shifts across regions, shaped by variables such as real wages, non-homothetic preferences, and trade costs. The facility with which producers and manufacturers engage with markets is crucial for the expansion and evolution of economic sectors. An increase in real wages results in a decreased proportion of agricultural consumption, thereby reducing market access for agricultural goods and accelerating industrialization. This trend can be observed when there is a reduction in

effective labor supply. Furthermore, reductions in trade costs can differentially impact market access. For instance, if trade tariffs are reduced more significantly for manufactured goods than for agricultural products upon opening to trade, there would be a relatively larger enhancement in market access for the non-agricultural sector. Such a disparity can further expedite the transition towards industrialization.

Labor supply: The dynamics of labor availability and distribution are pivotal to the process of structural transformation. Variations in labor supply, influenced by demographic changes, migration policies, or urbanization patterns, have the potential to either hasten or slow down the shift from the agricultural to the non-agricultural sectors. The Hukou system introduces additional complexity, allowing us to dissect the labor supply effect into two distinct components: the factor intensity effect and the Hukou composition effect. The former is primarily a consequence of differences in factor intensities, whereas the latter emerges due to the disparate employment distributions of residents with different Hukou statuses. To illustrate the impact of labor force changes without the influence of the Hukou system, consider a uniform increase in the effective labor supply across all Hukou types in region n, denoted by $d \ln \tilde{L}_{r,h} = l > 0$ for all h. Under this scenario, the latter effects sum to zero so the net effect of labor supply can be expressed as:

$$[(\gamma_L^A - \gamma_L^M) - (\frac{\gamma_H^M}{\eta^M} + \frac{\gamma_L^A}{1+\theta^a})]l < 0$$

Under the assumption that the agricultural sector is less labor-intensive than the non-agricultural sector, the expression consistently yields a negative value. This indicates that the first order effect of labor supply on structural transformation is governed by the relative factor intensities between sectors, aligning with the predictions of the basic model. This effect, combined with the demographic transition, gives the short-run and long-run implications of FPPs on structural transformation in my model.

Entry cost: The cost of entry for new firms affects the structural transformation by either encouraging or deterring the establishment of new firms in the non-agricultural sector. Lower entry costs can lead to a more competitive market, with new firms and more employment entering the non-agricultural sector. Conversely, high entry costs can stifle competition and slow down the process of structural transformation.

Hukou cost: Institutional factors, such as the Hukou system in China, impose costs that can significantly influence labor mobility and, by extension, ST. By restricting access to non-agricultural sector for agricultural Hukou, the Hukou system can deter rural-urban migration, affecting the labor allocation in both sectors and shaping the overall pattern

of structural transformation. Reduction in the Hukou cost contributes to the decrease in agricultural employment share.

1.4 Reduced Form Analysis

1.4.1 Data

The primary dataset employed in this empirical study consists of the 1% sample China population census data for the years 1982, 1990, and 2000, which has been sourced from IPUMS. With its extensive coverage, encompassing approximately 10-11 million observations for each year, this dataset provides a rich source of valuable information on various demographic characteristics of individuals in China. At the individual level, the census data includes a wide array of essential demographic attributes, such as age, gender, region of residence, region of registration, Hukou status, education level, and employment status.

Crucially, the dataset also contains valuable household-level information, allowing for the identification and linkage of children within their respective families. The relationship to the household head variable serves as a key instrument for discerning family structures within each household. This variable categorizes parents as "household head" and "spouse of the household head" and labels children as "child". Using this variable combined with household ID, I match all children with their parents at the household level and calculate the share of single-daughter families.

While some cases may exist where children have left home and are not accounted for in the sample, it is essential to note that such occurrences predominantly pertain to grown-up children who have independently departed from their original families. However, for the purpose of my analysis, this aspect will not introduce bias into the share calculation. This is because my analysis exclusively focuses on newborns in 1982, and none of these newborns would have left their parents' households at the time of calculation.

For the outcome variables related to sectoral employment, the census provides sector information that allows for a classification of individuals into primary and secondary sectors. This also enables the exploration of heterogeneity effects within the non-agricultural sector.

The mechanism analysis and model calibration utilize additional datasets, including panel data on 285 China prefectures from the China City Statistics Yearbook (1980 onwards), the 2002 National Input-Output (IO) table from the National Bureau of Statistics of China, province-level bilateral railway trade flow data (1985 onwards), and child-caring data from the China Health and Nutrition Survey (1991 onward).

1.4.2 Empirical Strategy

To investigate the effect of the OCP relaxation on structural transformation, I define the policy exposure using the share of single-daughter families as follows:

$$S_r = \frac{\text{\# of Households with Single-Daughter in 1982}}{\text{\# of Households with Single-Child in 1982}}$$

The numerator represents the total number of families with only one child, and that child is a girl. The denominator represents the total number of single-child families. This metric, S_r, allows me to quantify the prevalence of single-daughter families in each region, serving as a proxy for the policy exposure to the OCP relaxation. A higher value of S_r suggests a greater prevalence of single-daughter families and indicates a larger impact of the OCP relaxation in that region. In the following analysis, I restrict my sample to include only provinces that offer this relaxation and compare the variations only among prefectures within these provinces.

An important assumption for this measure being exogenous is that there is no gender selection at the first birth. This is likely to be true in practice as prenatal sex determination technologies, such as ultrasound, were not widely available in China at that time. However, the measurement is limited by the fact that only surviving children are observed in the census data. This means that the share of single-daughter families is calculated based on the gender of the child surviving at the time of the survey, which might not fully represent the true prevalence of single-daughter families if certain families engaged in gender selection or abandonment practices before the survey.

The possibility of families actively abandoning children, particularly girls, to avoid penalties under the OCP could introduce bias when calculating the share of single-daughter families based on surviving children. This may lead to underestimation of the actual prevalence of single-daughter families. More importantly, the act of sex selection based on surviving children may introduce endogeneity concerns in the analysis, arising from the correlation between the calculated share of single-daughter families and the level of OCP enforcement. Prefectures with stricter OCP enforcement may be more likely to engage in active abandonment of girls, resulting in a lower observed share of single-daughter families.

Figure 1.4 shows the proportion of females across varying birth orders. It is evident that the sex ratio for first births remains relatively stable at approximately 0.49 both prior to and subsequent to the implementation of the OCP in 1980, aligning with the normal range of sex ratio. Notably, a significant decline in the ratio is observed only for second and

subsequent births, suggesting that sex-selective practices or child abandonment predominantly occur at higher birth orders rather than the first. This pattern substantiates the methodology employed in this study, wherein the share calculation is confined solely to first births, thereby mitigating concerns regarding the distortion of data due to sex selection. Furthermore, as these fractions are calculated from data on surviving children as recorded in census data, the potential confounding factor of child abandonment does not pose a significant concern in this analysis.

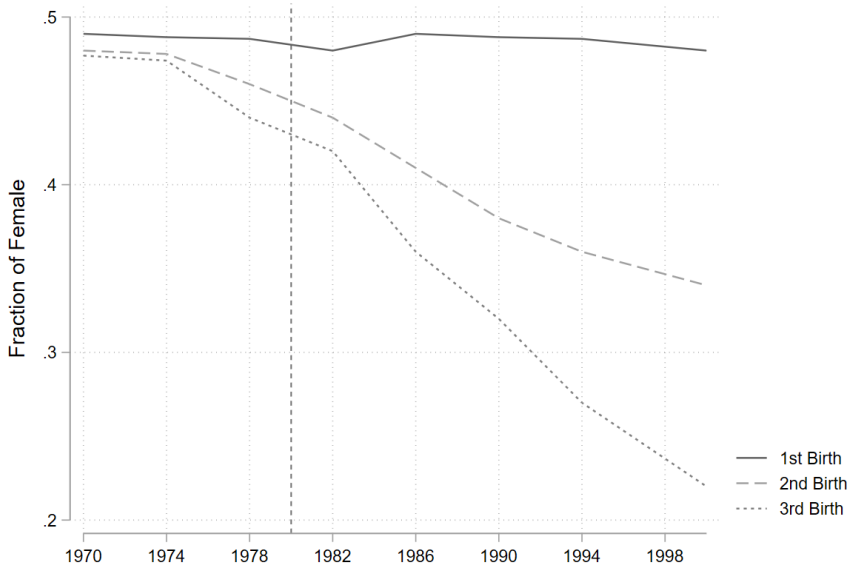


Figure 1.4: Fraction of Female for Different Birth Orders

Concerns may arise regarding the variations of the share of single-daughter families within each prefecture in China. Due to the law of large numbers, these shares should converge closely to the mean, resulting in a small variation. If substantial variation is observed, it could suggest the influence of regional unobservables rather than biological randomness. Figure 1.5 presents the distribution of single-daughter family shares across all prefectures, with values ranging from 0.44 to 0.57 and a standard deviation of 1.67%. To ascertain whether this variation stems from random biological factors, I conducted a simulation by randomly assigning the sex of children for the same sample, maintaining the same mean as observed in the actual data. The resulting distribution of single-daughter family shares from this simulated data closely mirrors the empirical distribution, with a mean standard deviation of 1.71% for 100 times draws. This comparison suggests that the observed variation in the data can indeed be attributed to random biological variation rather than systematic regional differences.

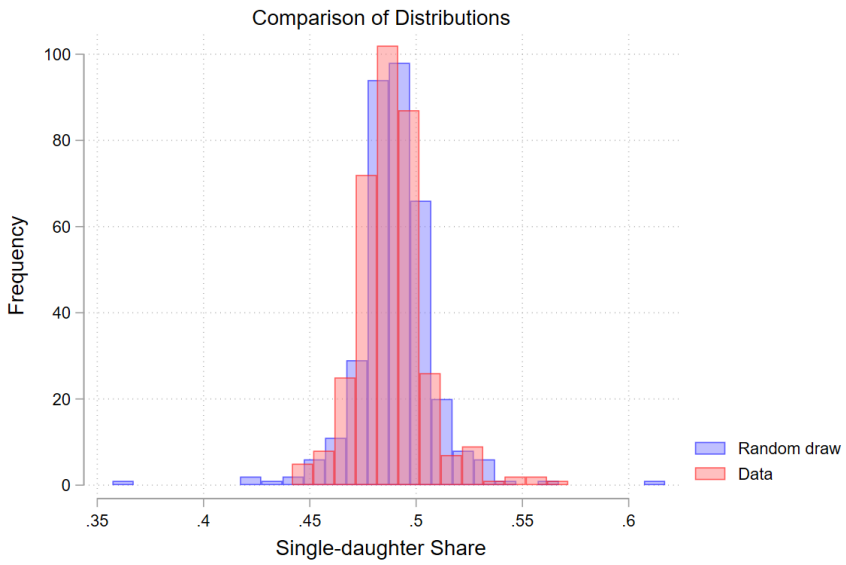


Figure 1.5: Random Draw vs Data Distribution

Figure 1.6 presents the trend in total fertility rate (TFR) before and after the 1982 OCP relaxation. The black line represents the average TFR for prefectures with a share of single-daughter families above the mean, while the blue line represents prefectures with a share below the mean. Prior to 1982, both prefectures exhibited similar TFR trends. However, after 1982, prefectures with a higher share of single-daughter families consistently demonstrated higher TFR than those with a lower share for 15 years, indicating that the former experienced a more substantial policy shock due to the OCP relaxation. This divergence in TFR trends provides suggestive support of using this share as a measure of policy exposure.

After defining the policy exposure, the main DID estimation equation in the chapter is as follows:

$$Y_{rt} = \beta \cdot post_t \cdot S_r + \delta_r + \gamma_t + \epsilon_{rt} \tag{1.20}$$

where Y_{rt} is the outcome variable, such as agricultural employment share and total employment in region r at time t. $post_t$ is a dummy variable taking the value of 1 for years after 1982 and 0 otherwise. S_r represents the share of single-daughter families in 1982. β is the key parameter of interest, capturing the effect of the One Child Policy relaxation on

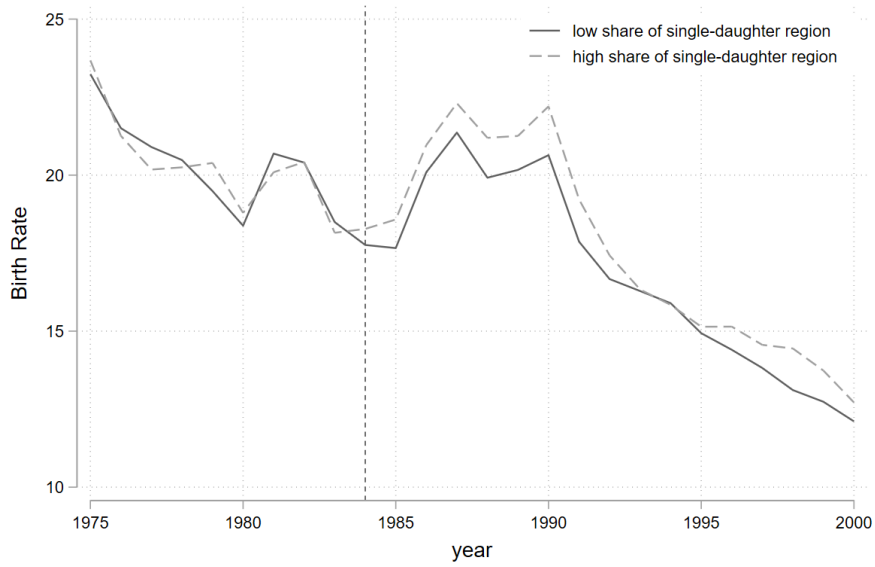


Figure 1.6: Total Fertility Rate Trend in Prefectures with Different Policy Exposure

the outcome variable. δ_r represents region fixed effects, and γ_t denotes year fixed effects. The term ϵ_{rt} is the error term accounting for unobserved factors and random variations.

1.4.3 Identification Assumptions

The key identification assumption for the DID estimation is the parallel-trend assumption: in the absence of the OCP relaxation, prefectures with different policy exposures (i.e., varying shares of single-daughter families) would have experienced similar trends in employment outcomes under the same economic and policy conditions. This assumption is vital because it allows me to attribute any divergent changes in employment outcomes post-OCP relaxation to the policy change itself, rather than pre-existing disparities among prefectures.

Since prefecture-level employment data is not available before 1980, I utilize province-level employment data and calculate the corresponding share of single-daughter families at the province level. I then classify provinces with a share of single-daughter families higher than the mean share as "high share" prefectures and provinces with a share below the mean as "low share" prefectures. Figure 1.7 compares the trend in the average share of primary employment between these two groups of prefectures. The figure shows that before the OCP relaxation in 1984, both high and low share prefectures experienced similar trends in primary employment share, with the share of primary employment declining from around

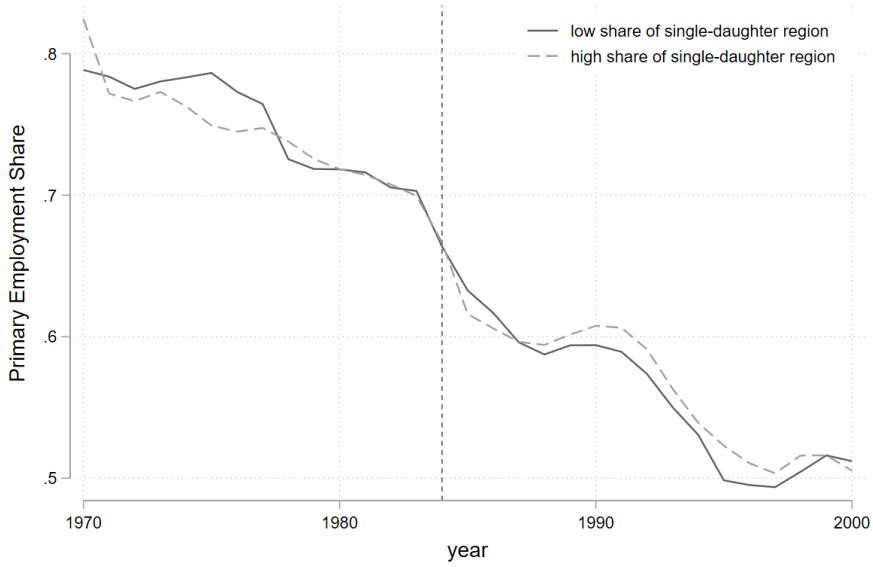


Figure 1.7: Primary Employment Share Trend in prefectures with Different Policy Exposure

82% to 67% during the period 1970-1984. This evidence supports the validity of the parallel trend assumption.

After the 1984 OCP relaxation, the trends in primary employment share diverge between prefectures with high and low shares of single-daughter families. In the first few years from 1984 to 1990, high share prefectures experience a slightly higher share of primary employment compared to low share prefectures. However, until 2000, high share prefectures undergo a larger decrease in primary employment share compared to low share prefectures, which offsets the initial slightly higher share. This indicates that the OCP relaxation may have had a varying impact on the structural transformation of the labor market across prefectures with different shares of single-daughter families.

1.4.4 Baseline Results

Table 1.1 presents the results of estimating equation (1.20) for the periods 1982-2000 and 1982-1990. The dependent variables are fertility rates, log of total employment and agricultural employment share. The first three columns show the long-run effects using data from 1982 to 1990 census. The last two columns show the long-run effect using data from 1982 to 2000 census.

I firstly examine the effect of the shock on fertility changes during the period 1982-1990

by regressing fertility on the share of single-daughter families. Column 1 present the regression results on fertility changes. The result reveals that the OCP relaxation had a significant impact on fertility, leading to an increase in fertility during the period 1982-1990 for prefectures with higher share of single-daughter families. Column 2 display the results for the total labor force. The findings show that in the long run, there was a significant increase in the total labor force in 2000, as a consequence of the increase in fertility following the OCP relaxation. However, in the short run, there was a decrease in the total labor force in 1990.

How does the policy affect structural transformation? Columns 3 and 5 show the short-run and long-run effects on the agricultural employment share. The results indicate that prefectures with a higher share of single-daughter families experienced a larger increase in agricultural employment share after the OCP relaxation in the short run until 1990. However, when examining the long-run results using data from 1982 to 2000 in column 5, the pattern reverses. The higher share of single-daughter families in prefectures is associated with lower agricultural employment share during this period. The results are robust after controlling for year and prefecture fixed effects. Specifically, one standard deviation increase in the share of single-daughter families leads to a 2% (-1.22*0.0167) decrease in the agricultural employment share in the long run but a 3.7% (2.24*0.0167) increase in the agricultural employment share in the short run.

While the magnitude of the short-run effect (2.24) appears larger than the long-run effect (-1.22), it is crucial to consider the net effect over time. The baseline results suggest that the long-run effect of the OCP relaxation dominates the short-run effect, resulting in an overall net positive effect on industrialization.

The key takeaway from the baseline results is that prefectures with a larger OCP relaxation shock experienced deindustrialization in the short run, but they eventually bounced back and became more industrialized in the long run. This indicates that there is a trade-off between short-run and long-run effects when implementing similar family planning programs or other shocks to fertility. I will delve into the fundamental drivers that underlie this trade-off in Subsection 1.4.6.

1.4.5 Robustness Tests

In this section, I conduct a series of robustness tests to validate the findings from the baseline results.

Alternative years for share of single-daughter family. To ensure that the share of

Table 1.1: Baseline Regression Results for 1982-1990 & 1982-2000

	1982-1990			1982-2000	
	Fertility	log Emp	Agri Share	log Emp	Agri Share
$\text{post}_t \cdot S_r$	1.23^a	-0.56^b	2.24^c	0.37^c	-1.22^a
	(0.01)	(0.12)	(1.34)	(0.19)	(0.26)
year FE	Y	Y	Y	Y	Y
prefecture FE	Y	Y	Y	Y	Y

Note: Significance levels are denoted as a for 99%, b for 95%, and c for 90%. Standard errors clustered at prefecture level. Standard errors are shown in the bracket. Dependent variables are fertility rates, log total employment and agricultural employment share during 1982-1990 and 1982-2000 respectively. All regressions use population in 1982 as weights.

single-daughter families identified in the baseline regression is not systematically influenced by regional-specific unobservables, I calculate this share using data from the pre-FPP period of 1970. Figure 1.8 compares this share with the baseline share. There is no correlation between these two shares. Table 1.2 provides detailed regression results with this newly defined proportion. The regression results indicate that none of the coefficients are statistically significant, thereby substantiating the robustness of the baseline results and affirming that they are not compromised by regional discrepancies in unobserved variables.

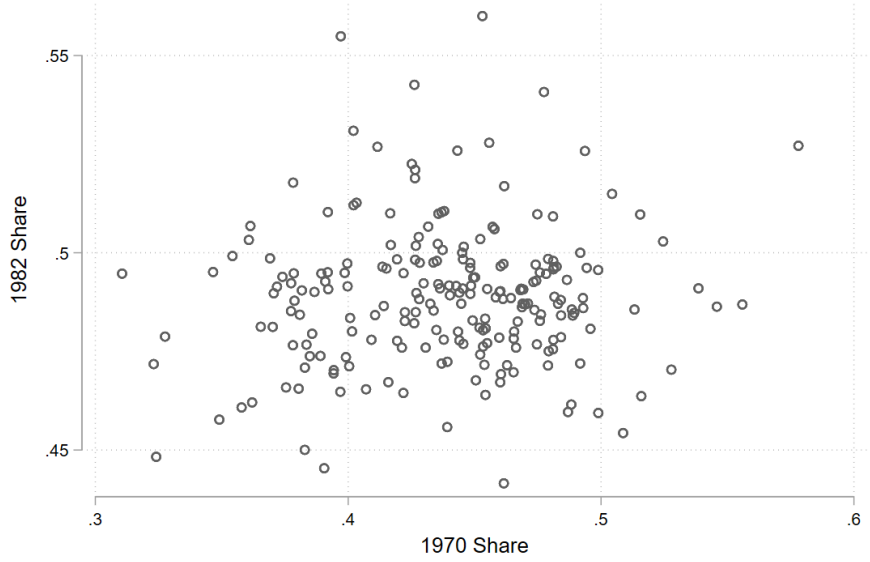


Figure 1.8: 1970 Share and 1982 Share

Non-treated provinces. I focus on provinces that experienced the OCP relaxation, as discussed in Section 1.2, and exclude provinces with no relaxation or only partial relaxations. This approach ensures that the share of single-daughter families, which serves as

Table 1.2: Robustness Test: Alternative Years for Share of Single-daughter Family

	1982-1990			1982-2000	
	Fertility	log Emp	Agri Share	log Emp	Agri Share
$\text{post}_t \cdot S_r^{1970}$	0.31	0.98	1.23	−1.19	0.33
	(0.98)	(2.36)	(2.44)	(0.65)	(1.66)
year FE	Y	Y	Y	Y	Y
prefecture FE	Y	Y	Y	Y	Y

Note: Significance levels are denoted as a for 99%, b for 95%, and c for 90%. Standard errors clustered at prefecture level. Standard errors are shown in the bracket. Dependent variables are fertility rates, log total employment and agricultural employment share during 1982-1990 and 1982-2000 respectively. All regressions use population in 1970 as weights.

the policy exposure measure of OCP relaxation, is only relevant in provinces where these relaxation policies were implemented. By examining provinces without the OCP relaxation, I can assess the validity of using the share of single-daughter families as a measure of policy exposure. In provinces where the OCP relaxation did not occur, the outcome variable should not respond to changes in this share.

Table 1.E.1 in Appendix 1.4.5 presents the regression results using the same baseline regression as equation (1.20) for provinces without the OCP relaxation shock. None of the coefficients is significant at the 5% level, indicating that neither the short-run nor the long-run effect responds to the exposure measurement in prefectures without the OCP relaxation policy.

Migrants. Another concern is related to the treatment of migrants in the baseline regression. Migrants in China were subject to different policies in some provinces and could have experienced different policy shocks compared to local residents. To ensure the accuracy of the policy exposure measurement, I recompute the share of single-daughter families by excluding all migrants, and then re-run the baseline regression.

Table 1.E.2 in Appendix 1.4.5 presents the regression results after excluding migrants from the calculation of the share of single-daughter families. The coefficients in this robustness test are very close to the baseline results in terms of both magnitude and significance level. This indicates that the exclusion of migrants from the calculation of the share of single-daughter families has little impact on the estimated policy exposure. It is worth noting that migrants only accounted for around 1% of the total working-age population in 1982, so it is not surprising that their impact on the regional OCP exposure during the early stage of the OCP is likely limited.

Rural and urban. In addition to the different treatment towards migrants, the differential treatment of individuals with agricultural Hukou (rural Hukou) and non-agricultural Hukou (urban Hukou) under the family planning policies can also introduce potential biases in the estimation of policy effects. To address this concern, I calculate the policy exposure for the baseline regression using individuals with rural Hukou, as the OCP relaxation only applies to rural Hukou individuals. This allows me to use non-treated urban individuals for a placebo test. If the baseline effect is driven by region-specific factors that affect all population in the same region other than fertility, using non-treated individuals in the same region should generate a similar pattern.

Table 1.E.3 in Appendix 1.4.5 displays the regression results using the policy exposure calculated from the share of single-daughter families for urban Hukou individuals. The coefficients are not significant at the 5% level. This finding further supports the validity of the baseline results and indicates that the observed policy effects are primarily driven by the relaxation of the One Child Policy for the rural population rather than being influenced by other region-specific factors.

Ethnic minorities. Another factor to consider is the treatment of ethnic minorities. The one-child policy primarily applied to the Han ethnic majority, and there may have been different policy provisions for ethnic minorities. This disparity could potentially introduce bias when estimating the policy effects. To address this concern, I conducted robustness tests in which I restricted the sample to only include the Han ethnic group, excluding ethnic minorities' impact on the policy effects. The results of these tests are presented in Table 1.E.4 in Appendix 1.4.5, and the coefficients are close to the baseline regression results. This suggests that the policy effects are consistent and robust when focusing solely on the Han ethnic group.

1.4.6 Mechanism Discussion

In this section, I explore potential mechanisms through which the family planning program affects structural transformation. To be considered as a potential channel underlying the baseline findings, any factor should meet the following criteria:

(i) It should have a direct effect on structural transformation, meaning that changes in this factor can influence the shift from agricultural sector to non-agricultural sector in the economy.

(ii) It should be correlated with the share of single-daughter families, so that the OCP

relaxation could lead to changes in this factor, which in turn affects structural transformation.

(iii) The pattern of this factor should change in both the short run and the long run to account for the observed reverse pattern in the reduced form.

Total labor force: Changes in fertility resulting from family planning policies can directly impact the total labor force in the long run. Moreover, changes in the dependency ratio due to changes in fertility can also affect the labor force in the short run, influencing the structural transformation process. The changes in fertility and labor supply have been confirmed in the baseline results.

Next, I turn to examine if this explains the reversing trend in structural transformation. To do that, I include the interaction term of the total labor force in the baseline regression. Table 1.3 presents the regression results with the inclusion of the interaction term of the total labor force in the baseline regression. Notably, the coefficient on the labor force interaction term is significant in both periods. Moreover, the coefficient on the policy exposure term becomes insignificant when the labor force interaction term is included, indicating that the total labor force is a significant channel through which the policy affects structural transformation. The coefficients on the total labor force are significantly negative for both periods, aligning with the basic model's intuition. An increase in labor endowment leads to a decrease in the agricultural employment share. Given that the total labor force experiences a decrease in the short run and an increase in the long run for regions with a high share of single-daughter families, this dynamic of the labor force channel is capable of explaining the reversing trend in structural transformation observed in the baseline findings.

To provide more evidence on the factor intensity channel, I explore the subsectors within the non-agricultural sector with different levels of labor intensity and run the baseline regression separately based on the labor intensity.[16] The results are shown in Table 1.4, which provide further evidence supporting the factor intensity channel. In both periods (1982-2000 and 1982-1990), the coefficients on the policy exposure term for high-intensity

[16]High-intensity manufacturing includes textile and apparel manufacturing, footwear manufacturing, leather goods manufacturing, furniture manufacturing, paper and paper products manufacturing, printing and related support activities, and food processing industries (e.g., meat and dairy products). Low-intensity manufacturing includes electronic manufacturing, automobile production, machinery manufacturing, chemical manufacturing (excluding certain chemical industries), pharmaceutical manufacturing, aerospace and defense equipment manufacturing, electrical equipment manufacturing, and computer and electronic products manufacturing.

Table 1.3: Regression Results with Labor Force

	1982-2000	1982-1990
$post_t \cdot S_r$	1.31	-1.124
	(5.16)	(3.62)
$post_t \cdot labor_r$	-0.53[a]	-0.51[a]
	(0.12)	(0.14)
year FE	Y	Y
prefecture FE	Y	Y
Observations	408	390
R-squared	0.943	0.912
Adj R-squared	0.914	0.892

Note: Significance levels are denoted as *a* for 99%, *b* for 95%, and *c* for 90%. Standard errors clustered at prefecture level. Standard errors are shown in the bracket. Dependent variables are share of agricultural employment. All regressions use population in 1982 as weights.

Table 1.4: Regression Results for Different Non-agricultural Sectors

	1982-2000		1982-1990	
	High intensity	low intensity	High intensity	low intensity
$post_t \cdot S_r$	7.422[a]	1.243[c]	-7.826[c]	-2.854[c]
	(1.77)	(1.01)	(6.11)	(2.01)
year FE	Y	Y	Y	Y
prefecture FE	Y	Y	Y	Y
Observations	408	408	390	390
R-squared	0.977	0.894	0.965	0.896
Adj R2	0.943	0.874	0.901	0.879

Note: Significance levels are denoted as *a* for 99%, *b* for 95%, and *c* for 90%. Standard errors clustered at prefecture level. Standard errors are shown in the bracket. Dependent variables are share of log of employment in each subsector.

manufacturing subsectors are significantly larger in magnitude compared to those for low-intensity manufacturing subsectors.

Based on the above analysis, it is evident that the labor force channel is a crucial factor explaining the reversing trend in the baseline findings regarding structural transformation.

Quantity-quality trade-off: Family planning policies may influence investment in human capital and the composition of the labor force in terms of skills and education levels, which in turn affect the skill distribution of the workforce and shape the structural transformation.

To examine this channel, I first analyze the data on education levels before and after the relaxation of the OCP. Specifically, I investigate whether there were any changes in the

Table 1.5: Regression Results for Skilled Share and Sex Ratio

	1982-2000		1982-1990		
	Total skilled	Sex ratio	Total skilled	New skilled	Sex ratio
$post_t \cdot S_r$	0.01	-0.75[a]	0.03	0.06	-0.65[a]
	(0.12)	(0.24)	(0.25)	(0.34)	(0.23)
year FE	Y	Y	Y	Y	Y
prefecture FE	Y	Y	Y	Y	Y
Observations	408	408	390	390	390
R-squared	0.989	0.912	0.942	0.942	0.879
Adj R2	0.978	0.864	0.912	0.912	0.861

Note: Significance levels are denoted as a for 99%, b for 95%, and c for 90%. Standard errors clustered at prefecture level. Standard errors are shown in the bracket.

proportion of skilled workers (e.g., those with higher education levels) in response to the policy relaxation.

Table 1.5 presents the regression results for the quantity-quality trade-off channel, specifically focusing on the share of skilled labor in the labor force. The dependent variable in columns 1 and 3 is the share of skilled labor as a proportion of the total labor force, while the dependent variable in column 4 is the share of skilled labor among those under 19 years old within the labor force. None of the coefficients for the quantity-quality trade-off channel are statistically significant, indicating that the policy relaxation did not result in a trade-off between quantity (i.e., total labor force) and quality (i.e., skill levels) of the labor force at least at the regional level in my study. Therefore, this is less likely to be a channel explaining my main findings.

Next, I explore the potential role of **sex ratio distortions** as a potential channel through which family planning policies might have influenced structural transformation. This may occur through the following channel, as the policy relaxation only happens to families with single-daughter, the sex ratio would be distorted for newborns. If newborns of different genders show varying propensities towards working in the non-agricultural sector or agricultural sector, this could significantly influence the overall structural transformation of the economy.

The results in Table 1.5, columns 2 and 5, show the relationship between the sex ratio of newborns and the share of single-daughter families. The dependent variable in column 2 is the sex ratio (male to female) within the age group of 0-9, representing the sex ratio of newborns during the policy relaxation period. On the other hand, the dependent variable in column 5 is the sex ratio within the labor force aged under 19, capturing the sex ratio

of individuals who entered the labor force during that period. Both are significant and negative, suggesting that the sex ratio is distorted after the OCP relaxation.

To assess the impact of sex ratio distortions on the process of structural change, I approach the question from two angles. First, I include the sex ratio interaction term in the baseline regression to investigate its aggregate effect at the regional level. Second, I turn to individual-level data from the year 2000 to delve into the propensity of male and female individuals toward different sectors to explore how sex ratio distortions might have impacted labor choices at the micro-level.

Table 1.6 presents the regression results for the baseline model with the inclusion of sex ratio interactions (columns 1 and 2). While the coefficients in these regressions show some decrease compared to the baseline regression, they are not statistically different. Specifically, the coefficient on the sex ratio interaction term is also not significant, indicating that sex ratio distortions do not appear to be a key channel explaining the observed trend in structural change. In columns 3 and 4 of Table 1.6, I investigate the relationship between gender and sectoral employment patterns specifically for the labor force aged below 20. The dependent variables in these regressions are binary variables indicating whether individuals in this age group are employed in the non-agricultural or agricultural sectors, respectively. Coefficients are not significant, which means that among the newborns after the OCP relaxation, there are no significant differences in the propensity to choose different sectors between males and females within the labor force aged below 20. The findings from both the aggregate and individual-level analyses suggest that the sex ratio distortions may not be a significant factor influencing the overall process of structural change.

1.5 Quantitative Analysis

In the quantitative analysis section, I calibrate the model using various data sources and model structures to study the aggregate effects of the FPP and its interactions with other forces in the economy. The calibration process involves estimating the values of various parameters in the model based on available data and model equations. Once the model is calibrated, I perform counterfactual analyses to explore the implications of different scenarios including changes to the FPPs, as well as other fundamentals in the economy.

1.5.1 Calibration

In this section, I present the methodology employed to calibrate the model for each parameter. The region in the model is defined to encompass 285 prefectures and one rest

Table 1.6: Regression Results for Sex Ratio

	1982-2000	1982-1990	2000	
	Agri share	Agri share	Non-agri	Agri
$post_t \cdot S_r$	-1.21^{a}	2.13^{c}		
	(0.29)	(1.76)		
$post_t \cdot Sex_r$	5.49	4.98		
	(8.59)	(6.53)		
$Female_i$			0.13	-0.21
			(1.24)	(0.86)
year FE	Y	Y	N	N
prefecture FE	Y	Y	Y	Y
Observations	408	390	3, 321, 125	3, 321, 125
R-squared	0.957	0.909	0.103	0.153
Adj R2	0.898	0.794	0.098	0.143

Note: Significance levels are denoted as a for 99%, b for 95%, and c for 90%. Standard errors clustered at prefecture level. Standard errors are shown in the bracket.

of the world, and panel data from these prefectures is collected for calibration purposes. However, as data availability may vary, not all variables might be directly accessible at the prefecture level. In such cases, I use data from higher administrative levels, such as province-level data, and make reasonable assumptions to ensure the model accurately matches the observed data.

I estimate a set of structural parameters and fundamentals using model implies equations, reduced form regressions from structural equations, model inversion and simulated method of moments (SMM). Firstly, production shares (γ_s^A, γ_s^M), Hukou costs $(\delta_{j,h,t}^n)$ and entry fixed cost $(f_{e,t})$ are backed out directly from the model equations to match data moments. The capital deepening parameter g is set to match the capital share in production. Then cross-region and cross-sector migration elasticities (ϵ and ζ), the share of time cost for each child (q, η_q) are estimated through reduced-form regression, while trade elasticities (θ^A, η) are obtained from existing literature. With these elasticities, region-sector productivity A_n^j is calibrated to match sectoral GDP in each region in the data. Since detailed trade and migration data are not available at prefecture level, I use province level data and make some assumptions to estimate migration costs $\tau_{rn,h,t}$ and trade costs $\kappa_{rn,t}^j$. Regional preference for offspring k_i is calibrated to match the model with the fertility rate in each region before1982.[17] FPPs fines $\chi_{n,h}$ are calibrated to match the fertility after 1982. Fertility elasticity η_f is estimated using SMM to match the model predicted elasticity with data.

[17]Here I match fertility rate in 1970 in the sense that fertility may already affected by the OCP during 1970-1982 (Chen and Fang, 2021).

Consumption parameters η_c, ν are backed out using SMM to match aggregate GDP in two sectors and relative price ratios in three years. Table 1.7 summarizes the moments and data used in the estimation.

Production shares (γ_s^A, γ_s^M), capital deepening g and fixed entry cost $f_{e,t}^n$. Due to the structure of the Cobb-Douglas production function, production shares are expressed as the share of input costs in proportion to total output. These shares are derived from China's national Input-Output (IO) table. In China's IO table for the period 1980-2000, returns to land are attributed to labor in the agricultural sector but to operating surplus in non-agricultural sector. To disaggregate the labor share in the agricultural sector, I utilize the estimation from Cao and Birchenall (2013), which estimates the average land and labor shares in the agricultural sector for the years 1952-2003 in China. This yields an average factor share of land at 37% and labor at 38%. I assume that the share of returns to land in total labor income in the agricultural sector remains constant during 1980-2000. This assumption, combined with the total labor share from the IO table, allows for the construction of the time trend of labor shares in both sectors, as depicted in Figure 1.9. The labor share in the non-agricultural sector is consistently higher than in the agricultural sector, which substantiates the basic factor intensity assumption in my model.

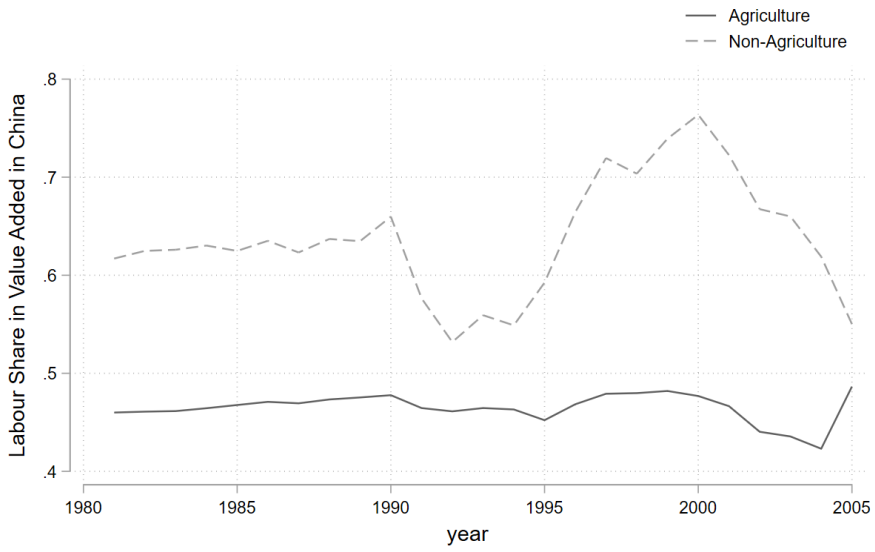


Figure 1.9: Labor Share in Production in China

In the absence of explicit capital and sectoral linkages within my model, I proportionally allocate the share of capital to land and labor, ensuring their combined contributions sum to unity. I maintain the constancy of land and labor shares for the purposes of quantification,

adopting their values from the year 2000. This yields a labor share in agriculture denoted by $\gamma_L^A = 0.51$ and a labor share in non-agriculture represented by $\gamma_L^M = 0.88$. The parameter that dictates capital deepening, denoted by g, is set at the capital share in non-agricultural production, which is 0.32 in 2000.

From the free entry condition for the non-agricultural sector, the fixed entry cost (f_e^n) can be expressed as a function of only observable and other model parameters, which can be estimated separately:

$$f_{e,t}^n = \frac{L_{n,t}^M}{(1 + \gamma_L^M(\eta^j - 1))N_{n,t}^M}$$

This equation shows that the entry cost is inversely related to the number of firms in a region. As the number of firms increases, the fixed entry cost decreases, given the same amount of labor supply. On average, the entry cost decrease by 20% during 1982-2000 (from 2.24 to 1.79), in terms of efficient labor requirement.

Hukou costs $\delta_{j,h,t}^n$. The Hukou costs in the model represent the frictions associated with changing sectors, which are distinct from an individual's Hukou type. By manipulating the sectoral choice share $s_{n,h}^j$, I derive an expression for the relative Hukou restriction as $\frac{s_{n,A,t}^A}{s_{n,A,t}^M} / \frac{s_{n,M,t}^A}{s_{n,M,t}^M} = (\frac{\delta_{M,A,t}^n}{\delta_{A,M,t}^n})^\zeta$[18] where $\delta_{M,A,t}^n$ is the Hukou friction for agricultural Hukou to work in non-agricultural sector and $\delta_{A,M,t}^n$ is the Hukou friction for non-agricultural people to work in agricultural sector[19]. If this ratio is large, it indicates that the Hukou system imposes greater restrictions on rural Hukou people to work in the non-agricultural sector. The intuition for this equation is as follows: the share of employment in each sector is determined by two relative forces—sectoral wages and Hukou restrictions. Individuals with different Hukou types face the same wage efficiency rates for a given sector within a region, dividing the relative shares of employment in agricultural and non-agricultural sectors effectively cancels out the influence of relative sectoral wages. As a result, the remaining impact is solely attributed to Hukou restrictions. Figure 1.10 displays the distribution of the log relative Hukou cost during the period 1990-2005, revealing a consistent decrease in Hukou costs over time. This decline can be attributed to the several Hukou reforms implemented after the 1990s. On average, relative Hukou restrictions decrease by 68%.

Cross-sector migration elasticities ζ. I calculate the relative Hukou frictions using

[18] I do not estimate the Hukou cost in levels but only in relative terms because the Hukou restriction in the same region can only be estimated up to a scale when there also exists regional level migration cost. A simultaneous increase in Hukou cost in the same region is equivalent to an increase in the migration cost to that region.

[19] Generally, this no friction for non-agricultural people to work in either sector arising from Hukou system. I keep this symmetric form to maintain the generality of the model.

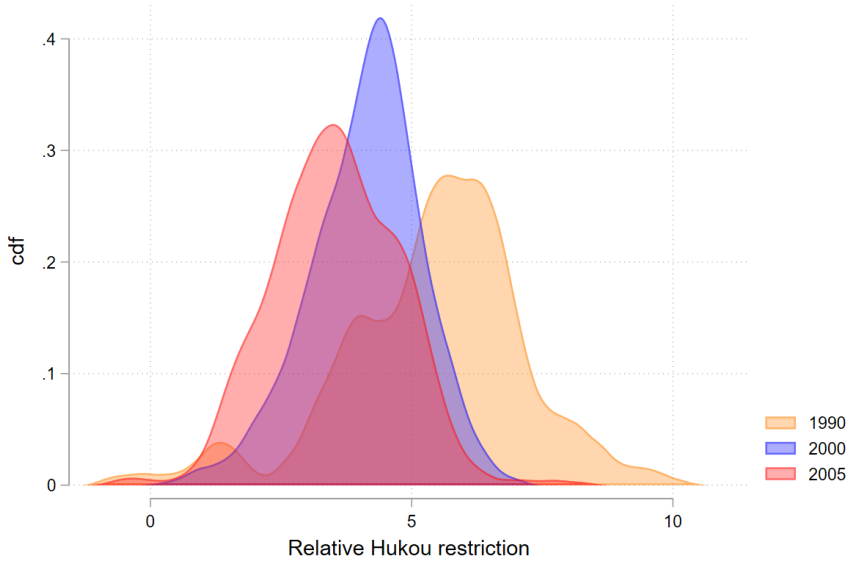


Figure 1.10: Relative Hukou Restrictions

observed data on the share of employment in each sector for each Hukou after getting the elasticity from the following regression. By taking log of the relative share expression, the model implies the reduced form type regression:

$$\ln \frac{s^A_{n,h}}{s^M_{n,h}} = \beta_0 + \zeta \ln \frac{\delta^n_{M,A}}{\delta^n_{A,M}} + \zeta \ln \frac{w^A_n}{w^M_n} + \epsilon_{n,h}$$

where I express the relative share of employment in each region as a function of relative Hukou frictions and relative wages. Using relative wage data and employment data, I empirically run this regression, where I use the wage data for the previous year as an instrument. This gives me an estimates of 1.55, which is close to 1.5 in Tombe and Zhu (2019).

Migration cost $\tau_{rn,h}$ and cross-region migration elasticity ϵ. I assume that the migration cost takes the form of quasi-symmetric structure $\tau_{in,h} = \tau^A_{i,h}\tau^B_{n,h}d^{\eta_\tau}_{in}$, where $\tau_{nn,h} = 1$ (Allen et al., 2020). The origin component $\tau^A_{i,h}$ captures the friction faced by individuals when migrating out of region i, reflecting the cost of forgoing the benefits of their local Hukou. The destination component $\tau^B_{n,h}$ measures the friction experienced by individuals when migrating to region i, representing the cost of working in a region with no local Hukou. d_{in} is the distance between region i and region n. η_τ measures the elasticity of

migration with respect to distance. Plugging the functional form of migration cost into the share of migration yields the following reduced-form regression equation:

$$\ln m_{rn,h} = -\epsilon\eta_\tau \ln d_{rn} - \epsilon \ln \frac{\bar{I}_{n,h}}{P_n} + \alpha_r + \epsilon$$

Using migration data, distance and instrumenting the expected real wage with past years' wage, I get the estimation for these elasticities $\epsilon = 1.5$, $\epsilon\eta_\tau = 0.8$. The functional form of migration cost reduces the degree of freedom of the migration cost from $N * N * 2$ to $2N+2N$, which means that the bilateral migration data is not necessary for the estimation. I calibrate the origin component $\tau^A_{i,h}$ and destination component $\tau^B_{n,h}$ by matching the share of migrants who migrate to the destination and the share of migrants who migrate out of the origin. The migration costs component are only identified up to a scale: any value satisfy $\kappa^A_i = t\tilde{\kappa}^A_i$, $\kappa^A_i = \tilde{\kappa}^A_i/t$ will match the moment in the data. My estimation shows that the migration cost is 2 times higher for agricultural Hukou than non-agricultural Hukou on average.

Trade cost κ^j_{rn}. The trade cost κ^j_{in} is assumed to follow a power function of distance, given by $\kappa^j_{in} = \kappa^j d^{\eta_\kappa}_{in}$, where $\kappa_{nn} = 1$. This formulation implies that trade costs increase with distance between prefectures, and it allows for different trade costs for trade between prefectures and sectors. η_κ is the trade elasticity to distance and κ^j is the overall level of trade cost in sector j. I assume that η_κ is the same across sectors. The difference in the trade cost is driven by the parameter κ^j. The distance elasticity η_κ can be estimated from the trade gravity equation derived from the model:

$$\ln \pi_{jrn} = -\eta_\kappa \theta_a \ln d_{rn} + \alpha_r + \alpha_n + \epsilon$$

Bilateral trade data is not available at the prefecture level. I use province level bilateral railway trade data to estimate the elasticity of trade to distance. To calibrate the overall level of trade cost, I match the model with the total trade share for each sector[20]. The overall level of trade cost decrease by 13% in my estimation.

FPP fines $\chi_{n,h}$**, fertility elasticity** η_f **and regional preference for offspring** k_i. I select values for the regional preference for offspring k_i, to match the regional fertility rates prior to the implementation of the FPPs. By doing so, the model can account for the influence of all other factors that affect fertility rates and are consistent over time without

[20] In 1982, I do not have total trade share but only railway trade share. Therefore I use the relative change in the share of trade by railway from 1985-2000 as the relative change in total trade value to back out the trade value in 1982.

being affected by FPPs. The benchmark year preceding the FPPs is established as 1970 to account for any non-monetary penalties arising from the initial series of FPPs during the 1970s. Consequently, the decline in fertility from 1970 to 1980 can be attributed to both the natural decrease in birth rates and the enforcement of FPPs. Thus, my analysis captures not only the impact of the OCP post-1980 but also the enduring effects of the FPPs from the preceding decade. To capture the effect of the FPPs, I choose the fines $\chi_{n,h}$ in such a way that it exactly matches the regional fertility rate after the implementation of the policy, taking the value of k_i into account. I match the fertility rates after FPPs in 1982 and 1990 to back out the fines.

Figure 1.11 presents a comparison between the fine rates stipulated in policy documents in 1990 for all provinces in mainland China and the fine rates estimated by the model for urban and rural Hukou holders. As anticipated, the model-estimated fine rate for urban Hukou holders is significantly higher than the rates specified in policy documents. This discrepancy arises because the model's back-calculation of fine rates, achieved by matching them with fertility rates, inherently incorporates the influence of all forms of punishment that could affect an individual's fertility decisions. This includes non-monetary penalties, such as the potential loss of jobs, which elevates the monetary equivalent of the fine rate for urban Hukou holders to approximately twice that of the documented rates. Conversely, the FPPs are relatively less stringent for rural Hukou holders, and non-monetary punishments are less prevalent. Consequently, the estimated fine rates for rural residents align more closely with those recorded in policy documents.

For the fertility elasticity η_f, I simulate the model with different values of η_f and get the coefficient from regressing fertility rate on regional real wage. Then I choose the value of η_f that best aligns with the observed relationship between fertility and regional real wages. The estimated fertility elasticity parameter η is 0.75, which demonstrate a decreasing return to fertility. The coefficient generated by the model is -0.234, while the corresponding coefficient obtained from the real data is -0.221. This coefficient shows that my model generate a negative relationship between fertility and income without incorporating the quantity-quality trade-off. which is widely documented in the literature and often explained by quantity-quality trade-off (Doepke et al., 2023).

Share of time cost for each child q, η_q. Using data from the China Health and Nutrition Survey, which includes questions regarding the daily hours spent on childcare, the number of children, and detailed employment information such as average work hours each day, I establish the relationship between time dedicated to childcare and the number of children

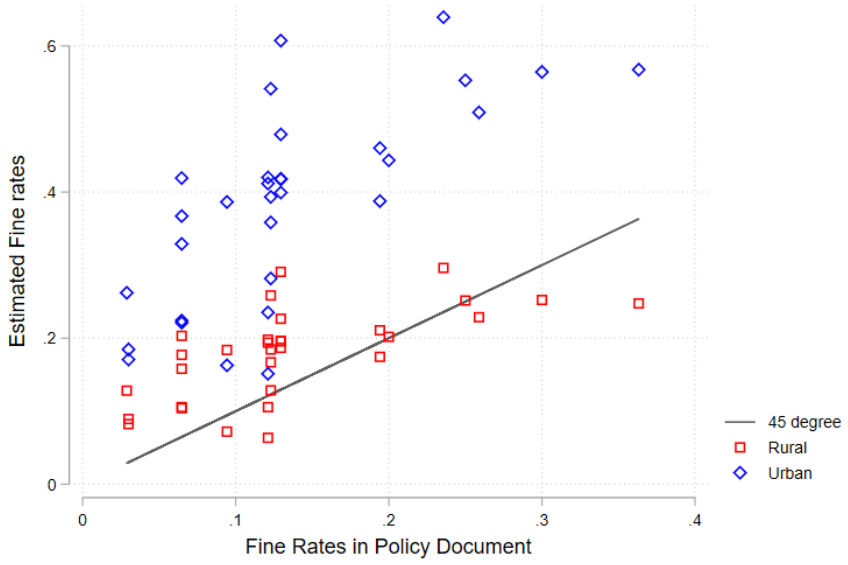


Figure 1.11: Estimated FPPs Fines vs Fines in Policy Documents

Source: Data for fines in policy documents are from Ebenstein (2010)

in a household. Figure 1.12 illustrates the inverse relationship between the average hours spent on each child each day and the total number of children in the household. A regression of the mean childcare hours on the number of children yields an estimate of $\eta_q = 0.67$. The regression's constant term reveals the baseline number of hours required to care for one child. By dividing this constant term by the mean working hours, the estimate for q is identified as amounting to 15% of the average working hours.

Productivity A_n^j and consumption parameters η_c, ν. Region-sector specific productivity is calibrated by selecting values that precisely align the model with sectoral GDP in each region. To compare productivity across time, I also ensure that the model generated aggregate price index over the years matches the Consumer Price Index (CPI) data. To obtain the consumption parameters η_c and ν, I match the model with the aggregate share of GDP for both sectors across different years and also with the relative price index between the two sectors during these periods. On average, productivity increases by 8% and 5% each year for agricultural and non-agricultural sectors.

1.5.2 Model Validation

In this section, the model's validity is assessed by matching untargeted moments observed in the data. In the calibration, FPPs fines were inferred by aligning with the fertility rates documented in 1990 and 1982. I assume that fines are held constant after 1990 and simulate

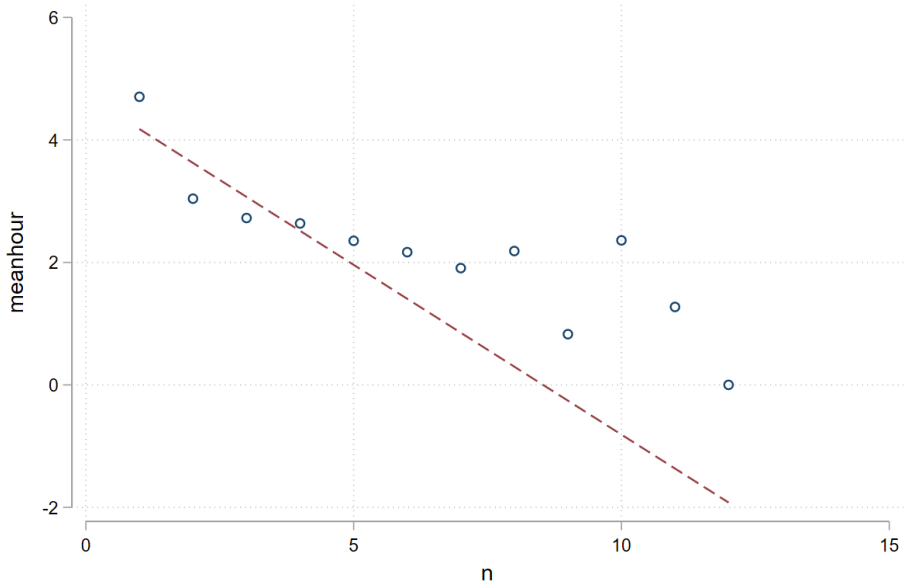
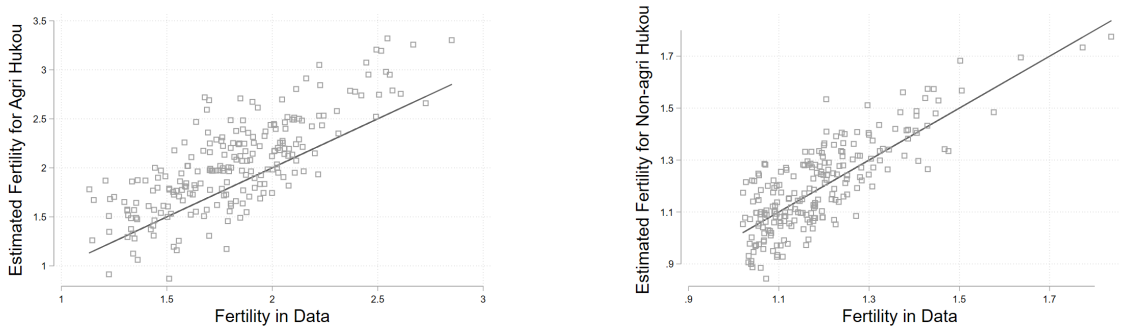


Figure 1.12: Parental Time on Each Child and Number of Children

Source: China Health and Nutrition Survey (CHNS)

Table 1.7: Parameters and Moments

Parameter	Notation	Approach/Targets	Dimension
Production shares	γ_L^A, γ_L^M	labor share in production	2
Migration cost	$\tau_{or,h,t}$	Migration shares	$N \times N \times 2 \times 2$
Trade cost	$\kappa_{or,t}^j$	Trade shares	$N \times N \times 2 \times 2$
Migration elasticity	ζ	Migration gravity equation	1
Productivity	$A_{r,t}^j$	Sectoral GDP and price	$N \times 2 \times 2$
Capital deepening	g	Capital share	1
Hukou cost	$\delta_{j,h,t}^r$	Employment share by Hukou	$N \times 2 \times 2$
Entry fixed cost	$f_{r,t}^e$	# of firms	$N \times 2$
FPP fines	$\chi_{r,h}$	Mean fertility	$N \times 2$
Consumption	η_c, ν	Aggregate share of exp.	2
Fertility	η_f, q, η_q	Fertility elasticity, mean exp.	3
Preference shifter	k_r	Fertility before FPPs	N



(a) Agri Hukou Fertility

(b) Non-Agri Hukou Fertility

Figure 1.13: Model Estimated Fertility vs Data in 2000

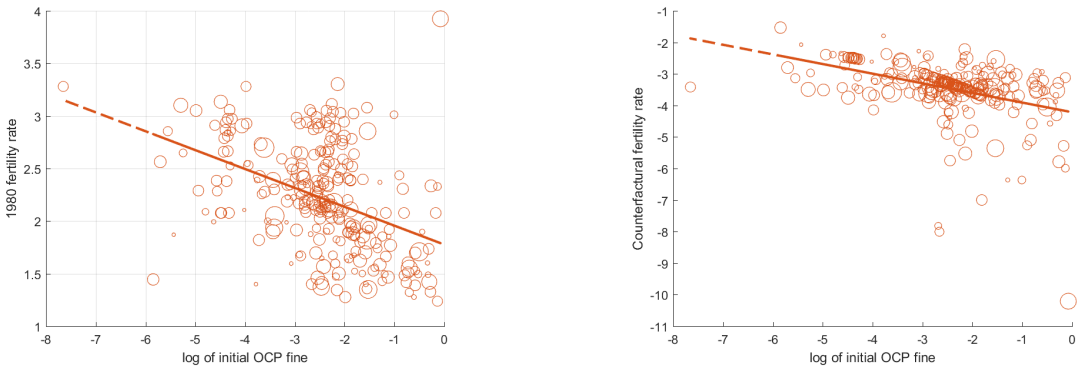
the model to project fertility in 2000. Figures 1.13a and 1.13b depict the model-estimated fertility alongside the actual fertility data for agricultural and non-agricultural Hukou in 2000, respectively, with the inclusion of the 45-degree line to facilitate comparison.

The estimated fertility rates for both agricultural and non-agricultural Hukou exhibit a strong correlation with the empirical data, indicating the model's robust capacity to generate endogenous fertility outcomes. As anticipated, agricultural Hukou demonstrate higher fertility rates than their non-agricultural counterparts, due to the lesser FPP fines imposed on them. Despite the high degree of correlation, the mean estimated fertility for agricultural Hukou is marginally elevated compared to the data. This discrepancy can be attributed to potential shifts in preference factors or slight alterations in FPP fines that occurred in 2000.

1.5.3 FPPs Effect in Regions

After the calibration and validation of the model, I use the model to investigate the short-term and long-term impacts of the FPPs in China. Firstly, I explore regions heterogeneities in FPPs to validate the main predictions of the model and to illustrate the impact of the FPPs at a regional level. Secondly, I study the implication on regional inequalities, particularly on regional convergence in structural transformation.

Demographic transition. To begin with, I present the trend between the estimated FPPs fines and the fertility rate before and after removing the FPPs fines. Figure 1.14a illustrates the relationship between the estimated FPPs fines (in log) and the observed fertility rate (in percentage point) in 1980. As expected, prefectures with high FPPs fines experience lower fertility rates, indicating the influence of the policy on fertility choices. In Figure 1.14b, I plot the changes in fertility if FPPs fines were removed. This is calculated as



(a) FPPs Fines vs Observed Fertility. (b) FPPs Fines vs Fertility Changes

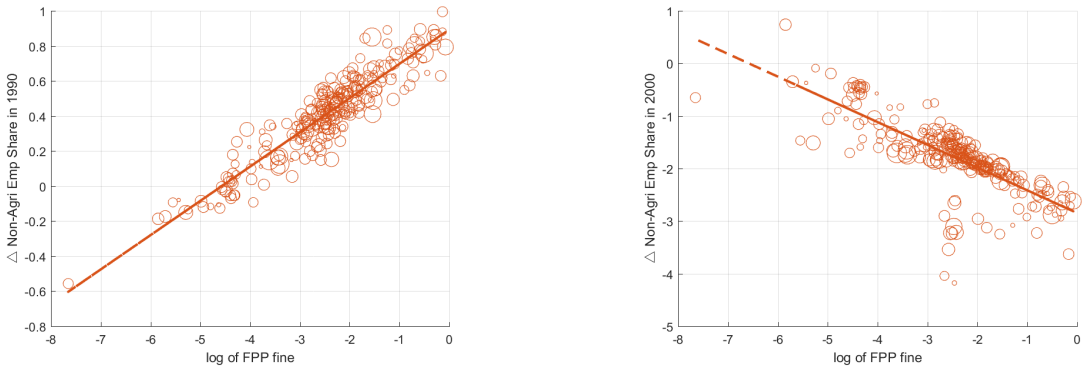
Figure 1.14: FPPs Fines vs Fertility

the difference between the observed fertility rate and the fertility rate in a world without FPPs fines. This difference represents the effect of FPPs on fertility rate. Prefectures with higher FPPs fines have larger differences, indicating that they would have had higher fertility rates in the absence of the FPPs fines.

Because of the structure of the model, higher fertility leads to a greater need for child care inputs in the first period, and these children only become adults in the second period. As a result, in the short run, the labor supply decreases mechanically due to the higher fertility. The long-run labor supply increases when these newborns enter the labor market in the presence of high migration cost, which means a large share of the new generation would stay in the same region as opposed to migrating to other prefectures. The model structure and high migration costs give the reversing trend of total labor supply following the shock on fertility, in this case, the FPPs.

Structural transformation. Next, I turn to check the response of non-agricultural employment share in 1990 and 2000 following the removal of FPPs. Figure 1.15a presents the short-term changes in the share of non-agricultural employment by comparing the actual data with the counterfactual non-agricultural employment share without the FPPs. The plot reveals a positive relationship between the two, indicating that the decrease in fertility following the FPPs led to an increase in non-agricultural employment share. However, Figure 1.15b shows that in the long run, there is an inverse correlation between them. This suggests that the non-agricultural employment share actually declines, falling below initial levels in regions that initially had higher fertility rates.

To this end, my model generates the same pattern as shown in the empirical findings. The immediate decrease in fertility leads to an increase in labor supply in the short run due to



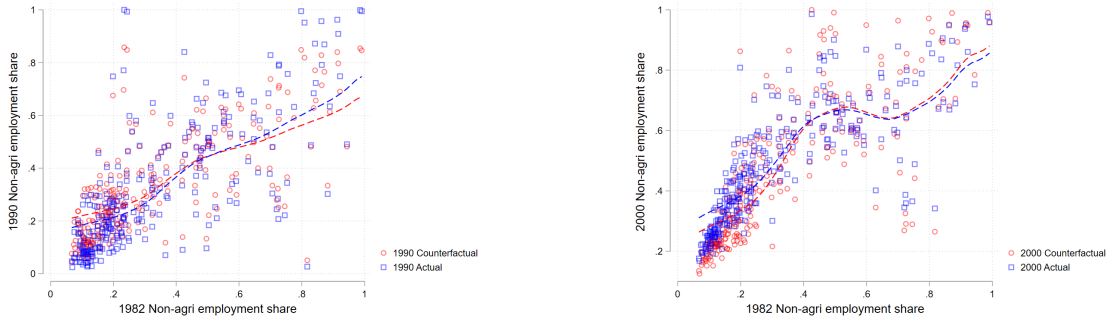
(a) Short Run Effect (b) Long Run Effect

Figure 1.15: FPPs and Structural Transformation

the lower demand for child care inputs. However, in the long run, the negative relationship between fertility and non-agricultural employment share emerges.

Regional convergence. The FPPs in China were more strictly enforced in urban areas than in rural ones, leading to greater decline in fertility rates in the cities. This difference in demographic change has influenced structural transformation differently across regions as is shown above. One implication of these heterogeneities is their effects on regional convergence. Rural areas, with their relatively higher fertility rates, have seen a slower structural transformation initially compared to urban areas. However, over time, these rural areas undergo a more rapid structural transformation as the labor force increases. This pattern suggests that FPPs have played a role in regional convergence, moderating ST in more developed areas while accelerating it in less developed ones through shifts in fertility in the long run.

To illustrate this further, Figure 1.16a presents the non-agricultural employment share, with the x-axis representing the share in 1982 and the y-axis representing the share in 1990. The blue points represent the actual shares observed in the data, while the red points depict a counterfactual scenario without the FPPs. The greater dispersion of the blue points suggests that the FPPs have resulted in increased short-term regional inequality in terms of structural transformation. Conversely, Figure 1.16b illustrates the same relationship but in a long-term context. Here, the blue points are less scattered, indicating that the FPPs have contributed to long-term regional convergence in structural transformation. Specifically, the standard deviation of the non-agricultural share in the long run (represented by the blue points) with FPPs is 0.198 compared to 0.238 without FPPs. A lower standard deviation suggests a stronger trend towards regional convergence.



(a) Short Run Convergence

(b) Long Run Convergence

Figure 1.16: FPPs and Regional Convergence

1.5.4 Aggregate Implications

In this section, I evaluate the overall impact of the FPPs at the national level in China. Additionally, I undertake a decomposition of the drivers of structural transformation to study the contribution of each factor. These factors include the expansion of international trade, increases in sector-specific total factor productivity, reductions in Hukou costs and decreases in the costs of entering non-agricultural markets.

Figure 1.17 illustrates the aggregate effects of different factors on structural transformation over the short and long term. In the short term, the FPPs lead to a 0.7% increase in the non-agricultural employment share. This is due to the immediate reduction in fertility rates, which increases labor supply at the national level and facilitates the more labor-intensive non-agricultural sector. Conversely, in the long term, the FPPs result in a 2.7% reduction in the share, due to the long-run decrease in labor supply.

Other factors also contribute to the structural change. The percentages indicate the magnitude of changes in the non-agricultural employment share for each factor compared to a counterfactual scenario where there is no change in that factor during 1982-1990. (1) The expansion of international trade had a modest impact on China's structural change from 1982 to 1990, increasing non-agricultural employment by only 0.2%. However, by 2000, as China's trade share of GDP grew from under 8% to over 20%, the contribution of international trade to structural change rose to 2.3%. (2) The rise in TFP in both agricultural and non-agricultural sectors contributed to an increase in the share of non-agricultural employment. However, the impact of agricultural TFP growth on structural transformation is considerably less pronounced than that of non-agricultural TFP growth. This discrepancy can be attributed to the interplay of income effects, which are beneficial for both types of TFP increases, and the comparative advantage effect present when

international trade is considered. With the expansion of trade, a country tends to specialize in sectors where it has a comparative advantage. As a result, when agricultural TFP increases, the comparative advantage effect may actually draw labor towards the agricultural sector, counteracting the traditional income and price effects that would typically encourage a shift away from agriculture. In contrast, an increase in non-agricultural TFP reinforces the country's comparative advantage in non-agricultural sectors, leading to a more pronounced shift of labor out of agriculture and a greater boost to structural transformation. (3) The reduction of Hukou system costs, a significant institutional friction in China, has played a substantial role in the country's structural transformation, second only to the increase in non-agricultural TFP. The Hukou system, which historically has regulated population movement within China, effectively restricted labor mobility between regions and sectors. As these costs have been reduced, it has facilitated a more fluid labor market, allowing workers to move more freely in response to economic opportunities. This reduction in mobility costs has been a key driver in the reallocation of labor from agriculture to non-agriculture, thereby accelerating the process of structural transformation. (4) A reduction in non-agricultural market entry costs also facilitates structural transformation by lowering the barriers for new firms, which boosts competition and job creation in the non-agricultural sector. However, this effect is smaller compared to changes in TFP and Hukou costs.

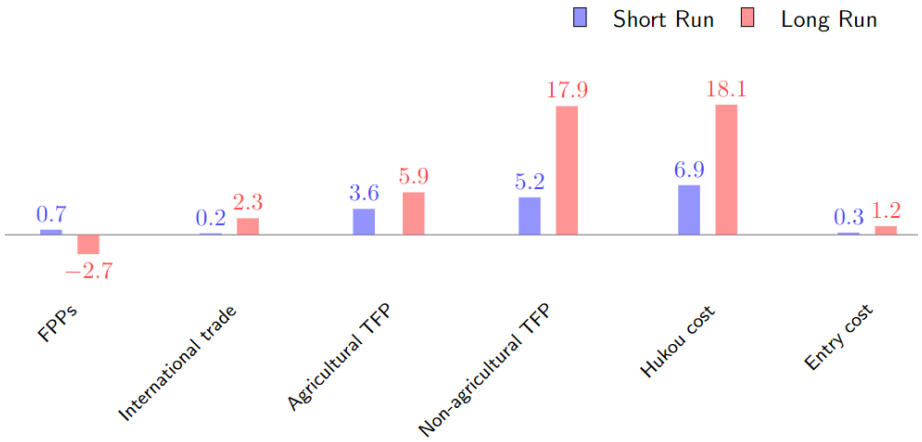


Figure 1.17: Aggregate Effect of Different Factors on Structural Transformation

Upon examining the implications for structural transformation, I turn my attention to the effects on real wages and welfare as depicted in Figure 1.18, which outlines the long-term consequences. Although the FPPs result in a 2.2% increase in real wages by reducing the labor supply, they also lead to a significant 10.1% reduction in welfare. Welfare is derived

from the utility of consumption and the utility from having children. Hence, the marked decrease in welfare is primarily due to a substantial reduction in the utility from children, which the increase in real wages does not fully compensate for. The increase in TFP within both the agricultural and non-agricultural sectors is the principal driver of the increase in welfare and real wages, contributing to an approximate 15% rise in each. Furthermore, the reduction in Hukou system costs boosts welfare and real wages by about 6%, highlighting the impact of reduced internal migration barriers on economic well-being. Finally, the decrease in entry costs into the non-agricultural market and the expansion of international trade each contribute only about 1% to the increase in welfare and real wages which are relatively modest compared to the substantial gains from increases in TFP and the easing of Hukou restrictions.

Figure 1.18: Aggregate Effect of Different Factors on Welfare and Real Wage

1.6 Conclusion

In this chapter, I investigate the relationship between demographic transition and structural transformation in China, focusing on the implications of family planning programs for economic development. By examining the effects of the FPPs on fertility rates, total labor supply, and industrialization, I aim to shed light on the complex interplay between demographic changes and economic outcomes. The empirical findings, along with insights from the dynamic economic geography model, contribute to understanding how demographic policies influence structural transformation and offer valuable guidance for policymakers facing similar challenges worldwide.

To understand the channel and quantify the effects of FPPs, I build a dynamic economic

geography model that allows me to quantify the aggregate impact of the FPPs and other mechanisms on China's industrialization. The model incorporates a dynamic fertility decision-making process, multiple regions and sectors, non-homothetic preferences, technological progress, and China-specific frictions like Hukou costs and non-agricultural entry barriers. With these features, the model captures China's unique demographic and economic characteristics, providing a comprehensive analysis of the FPPs' effects. The model demonstrates a Rybczynski-type effect, where an increase in labor supply disproportionately boosts the output of labor-intensive sectors. With FPPs altering labor availability, this effect allows for quantifying the policy's implication on China's sectoral outputs.

My empirical findings reveal the short-run and long-run effects of the OCP relaxation on China's industrialization process. The analysis highlights the total labor force as the primary driver behind these effects. In the short run, relaxing the policy increases fertility rates, leading to a higher dependency ratio, reducing the effective total labor force and resulting in a slower structural transformation. However, this trend reverses in the long run as the children born during the period of increased fertility rates eventually enter the labor market, leading to an increase in the total labor supply. Due to the differences in factor intensity between the two sectors, more labor supply enters the labor-intensive non-agricultural sector than the land-intensive agricultural sector in the long run.

Using the model and the calibrated FPPs fines, I conduct counterfactual analyses to investigate the aggregate effect of the FPPs on structural change in China. The findings reveal that the FPPs had divergent effects on China's industrialization. In the short run, the FPPs led to a 0.7% faster growth rate of industrialization, particularly in prefectures with stricter enforcement of the policy. This faster pace of industrialization during the early years of the FPPs was a consequence of the decreased fertility rates, leading to a lower dependency ratio and an increased effective total labor force. However, in the long run, the FPPs had an overall negative impact on industrialization. The reduced labor supply due to the FPPs eventually led to 2.7% less industrialization by 2000.

1.A Derivation of the Consumer Decision Problem

In this section, I provide a detailed derivation of the consumer decision problem. Consumers sequentially make decisions about their location, sector of employment, consumption, and fertility. Upon making location choices and observing their income, consumers decide on their consumption and fertility based on their income and the expected utility of their offspring in the next period.

The process is as follows:

- **Location Choices**: Consumers choose the location that provides the highest expected utility.
- **Sector Choices**: Upon choosing a location, consumers select the sector that offers the highest income at that location.
- **Consumption and Fertility Decisions**: After determining their income, consumers make decisions regarding consumption and fertility, considering the expected utility of their offspring in the subsequent period.

To solve this problem, I proceed by working backwards from the final decisions (consumption and fertility) to the initial decisions (location and sector choices).

1.A.1 Consumption and Fertility Decisions

Consumers choose fertility $n_{i,h,t}$ and expenditure on total consumption $y_{i,h,t}$ to maximize the following utility function:

$$U_{i,h,t} = (1 - k_i) \left(\frac{1}{\eta} \left(\frac{y_{i,h,t}}{p_{i,A,t}^{\phi} p_{i,M,t}^{1-\phi}} \right)^{\eta} - \nu \ln \left(\frac{p_{i,A,t}}{p_{i,M,t}} \right) \right) + k_i n_{i,h,t}^{\eta_f} \Pi_{i,h,t+1}$$

subject to the budget constraint:

$$y_{i,h,t} + \chi_{i,h,t}(n_{i,h,t} - 1) I_{i,h,t} = (1 - q n_{i,h,t}^{\eta_q}) I_{i,h,t}$$

First, express $y_{i,h,t}$ in terms of $n_{i,h,t}$ and substitute $y_{i,h,t}$ back into the utility function:

$$U_{i,h,t} = (1-k_i) \left(\frac{1}{\eta} \left(\frac{I_{i,h,t} \left(1 - q n_{i,h,t}^{\eta_q} - \chi_{i,h,t}(n_{i,h,t} - 1) \right)}{p_{i,A,t}^{\phi} p_{i,M,t}^{1-\phi}} \right)^{\eta} - \nu \ln \left(\frac{p_{i,A,t}}{p_{i,M,t}} \right) \right) + k_i n_{i,h,t}^{\eta_f} \Pi_{i,h,t+1}$$

Taking the first-order condition (FOC) with respect to $n_{i,h,t}$ yields:

$$-(1-k_i)\left(\frac{1-qn_{i,h,t}-\chi_{i,h,t}(n_{i,h,t}-1)}{p_{i,A,t}^{\phi}p_{i,M,t}^{1-\phi}}\right)^{\eta-1}\left(\frac{q\eta_q n_{i,h,t}^{\eta_q-1}+\chi_{i,h,t}}{p_{i,A,t}^{\phi}p_{i,M,t}^{1-\phi}}\right)I_{i,h,t}^{\eta}+\eta_f k_i n_{i,h,t}^{\eta_f-1}\Pi_{i,h,t+1}=0$$

Rearranging, this gives the equation (1.5) in the chapter:

$$n_{i,h,t}^{\eta_f-1}(1-qn_{i,h,t}-\chi_{i,h,t}(n_{i,h,t}-1))^{1-\eta}=\left[\frac{(1-k_i)(q\eta_q n_{i,h,t}^{\eta_q-1}+\chi_{i,h,t})\left(\frac{I_{i,h,t}}{p_{i,A,t}^{\phi}p_{i,M,t}^{1-\phi}}\right)^{\eta}}{\eta_f k_i\beta\Pi_{i,h,t+1}}\right] \quad (1.21)$$

Equation (1.21) determines optimal fertility given income $I_{i,h,t}$. Plugging the optimal fertility back into the budget constraint gives the expenditure on consumption:

$$y_{i,h,t}=I_{i,h,t}\left(1-qn_{i,h,t}^{\eta_q}-\chi_{i,h,t}(n_{i,h,t}-1)\right)$$

From the envelope theorem, the optimal utility increases with the income.

Next, I derive the expenditure share as a share of total consumption for individuals in each sector. The expenditure share from PIGL non-homothetic preferences, by Roy's identity, is given by:

$$\begin{aligned}
\theta_{A,h,t}(y_{i,h,t},P_t) &= -\frac{\partial V/\partial P}{\partial V/\partial y_{i,h,t}}\frac{p_{i,A,t}}{y_{i,h,t}} \\
&= -\frac{\left(\left(\frac{y_{i,h,t}}{p_{i,A,t}^{\phi}p_{i,M,t}^{1-\phi}}\right)^{\eta-1}\frac{-\phi y_{i,h,t}p_{i,A,t}^{-\phi-1}}{p_{i,A,t}^{1-\phi}}-\nu\frac{1}{P_{i,A,t}}\right)}{\left(\left(\frac{y_{i,h,t}}{p_{i,A,t}^{\phi}p_{i,M,t}^{1-\phi}}\right)^{\eta-1}\frac{1}{p_{i,A,t}^{\phi}p_{i,M,t}^{1-\phi}}\right)} \\
&= \phi y_{i,h,t}p_{i,A,t}^{-1}\frac{p_{i,A,t}}{y_{i,h,t}}+\frac{\nu\frac{1}{P_{i,A,t}}}{(y_{i,h,t})^{\eta-1}(p_{i,A,t}^{\phi}p_{i,M,t}^{1-\phi})^{-\eta}}\frac{p_{i,A,t}}{y_{i,h,t}} \\
&= \phi+\nu\left(\frac{y_{i,h,t}}{p_{i,A,t}^{\phi}p_{i,M,t}^{1-\phi}}\right)^{-\eta}
\end{aligned}$$

This gives the equation (1.3) in the chapter. The limit share of expenditure as GDP per capita increases on agriculture is ϕ. I assume that both rental income and any cost due to Hukou restrictions are income for local authorities who spend these incomes on both final

goods based on the limit share of expenditure. Then the total expenditure on agricultural final goods in each region is given by:

$$E_{i,t}^{A} = \int \theta_{A,h,t} y_{h,t} L_{n,t}^{A} \, dG_{n,h,t}(I) + \phi \left(\Sigma_j X_t^{i,j} - \int y_{A,h,t} L_{n,t}^{A} \, dG_{n,h,t}(I) \right)$$

where the first part is the total expenditure of all workers with both Hukou on agricultural final goods, and the integration is taken over the distribution of income, which is determined in the following section. The second part is the difference between all income (which is the total output) and the total expenditure of all workers with both Hukou on consumption. This includes rental income, income from FPP costs, and Hukou costs.

Similarly, for the non-agricultural sector:

$$E_{i,t}^{M} = \int (1 - \theta_{A,h,t}) y_{h,t} L_{n,t}^{A} \, dG_{n,h,t}(I) + (1 - \phi) \left(\Sigma_j X_t^{i,j} - \int y_{A,h,t} L_{n,t}^{A} \, dG_{n,h,t}(I) \right)$$

These give equation (1.9) (1.10) in the chapter.

1.A.2 Sector Choice

Since the optimal utility increases with income, when making sector choices, labor chooses the sector that gives them the highest income, which is given by:

$$I_{n,h,t} = \max_j \left\{ \frac{\phi_{n,t}^{j} w_{n,t}^{j}}{\delta_{j,h,t}^{n}} \right\}$$

where the match productivity $\phi_{n,t}^{j}$ is drawn from a sector-specific Fréchet distribution $F(\phi_{i,t}^{j}) = e^{-T_j (\phi_{i,t}^{j})^{-\zeta}}$.

Denote $\tilde{w}_{n,h,t}^{j} \equiv \frac{\phi_{n,t}^{j} w_{n,t}^{j}}{\delta_{j,h,t}^{n}}$. Then $\tilde{w}_{n,h,t}^{j}$ also follows a Fréchet distribution:

$$
\begin{aligned}
G_{n,h,t,j}(w) &\equiv \Pr(\tilde{w}^j_{n,h,t} < w) \\
&= \Pr\left(\frac{\phi^j_{n,t} w^j_{n,t}}{\delta^n_{j,h,t}} < w\right) \\
&= \Pr\left(\phi^j_{n,t} < \frac{\delta^n_{j,h,t}}{w^j_{n,t}} w\right) \\
&= F\left(\frac{\delta^n_{j,h,t}}{w^j_{n,t}} w\right) \\
&= e^{-T_j\left(\frac{\delta^n_{j,h,t}}{w^j_{n,t}} w\right)^{-\zeta}}
\end{aligned}
$$

Then the share of labor employed in each sector is given by $s^j_{n,h,t}$. This is the probability that $\tilde{w}^j_{n,h,t}$ is greater in sector j than in any other sectors:

$$
\begin{aligned}
s^j_{n,h,t} &\equiv \Pr[\tilde{w}^j_{n,h,t} \geq \tilde{w}^i_{n,h,t}, \forall i \neq j] \\
&= \int_0^\infty \Pr(w \geq \tilde{w}^i_{n,h,t}, \forall i \neq j)\, dG_{n,h,t,j}(w) \\
&= \int_0^\infty \Pi_{i \neq j} G_{n,h,t,i}(w)\, dG_{n,h,t,j}(w) \\
&= T_j \left(\frac{w^j_{n,t}}{\delta^n_{j,h,t}}\right)^\zeta \int_0^\infty \zeta w^{-\zeta-1} \Pi_i e^{-T_i w^{-\zeta}\left(\frac{w^i_{n,t}}{\delta^n_{i,h,t}}\right)^\zeta} dw \\
&= \frac{T_j \left(\frac{w^j_{n,t}}{\delta^n_{j,h,t}}\right)^\zeta}{\Sigma_i T_i \left(\frac{w^i_{n,t}}{\delta^n_{i,h,t}}\right)^\zeta} \int_0^\infty \zeta \Sigma_i T_i \left(\frac{w^i_{n,t}}{\delta^n_{i,h,t}}\right)^\zeta w^{-\zeta-1} e^{-w^{-\zeta}\Sigma_i T_i \left(\frac{w^i_{n,t}}{\delta^n_{i,h,t}}\right)^\zeta} dw \\
&= \frac{T_j \left(\frac{w^j_{n,t}}{\delta^n_{j,h,t}}\right)^\zeta}{\Sigma_i T_i \left(\frac{w^i_{n,t}}{\delta^n_{i,h,t}}\right)^\zeta} \left[-e^{-w^{-\zeta}\Sigma_i T_i \left(\frac{w^i_{n,t}}{\delta^n_{i,h,t}}\right)^\zeta}\right]_0^\infty \\
&= \frac{T_j \left(\frac{w^j_{n,t}}{\delta^n_{j,h,t}}\right)^\zeta}{\Sigma_i T_i \left(\frac{w^i_{n,t}}{\delta^n_{i,h,t}}\right)^\zeta}
\end{aligned}
$$

This gives the equation (1.7) in the chapter:

$$
s^j_{n,h,t} \equiv \frac{L^j_{n,h,t}}{L_{n,h,t}} = \frac{T_j \left(\frac{w^j_{n,t}}{\delta^n_{j,h,t}}\right)^\zeta}{\Sigma_i T_i \left(\frac{w^i_{n,t}}{\delta^n_{i,h,t}}\right)^\zeta}
$$

It is also useful to derive the relative share of employment in each sector, which is the basis for identifying relative Hukou cost parameters:

$$\frac{s^A_{n,A,t}}{s^M_{n,A,t}} \Big/ \frac{s^A_{n,M,t}}{s^M_{n,M,t}} = \left(\frac{\delta^n_{M,A,t}}{\delta^n_{A,M,t}}\right)^{\zeta}$$

We can also write down the distribution of $I_{n,h,t}$, which follows a Fréchet distribution:

$$\begin{aligned}
G_{n,h,t}(I) &\equiv \Pr\left(\max_j \left\{\frac{\phi^j_{n,t} w^j_{n,t}}{\delta^n_{j,h,t}}\right\} < I\right) \\
&= \Pr\left(\frac{\phi^j_{n,t} w^j_{n,t}}{\delta^n_{j,h,t}} < I \ \forall j\right) \\
&= \Pi_j \Pr(\tilde{w}^j_{n,h,t} < I) \\
&= \Pi_j G_{n,h,t,j}(I) \\
&= e^{-\Sigma_j T_j \left(\frac{\delta^n_{j,h,t}}{w^j_{n,t}} I\right)^{-\zeta}}
\end{aligned}$$

This gives the equation (1.8) in the chapter.

The expected value for the maximum income is:

$$\begin{aligned}
\bar{I}_{n,h,t} &\equiv E(I_{n,h,t}) \\
&= \int_0^\infty I \, dG_{n,h,t}(I) \\
&= \zeta \Sigma_j T_j \left(\frac{w^j_{n,t}}{\delta^n_{j,h,t}}\right)^{\zeta} \int_0^\infty I^{-\zeta} e^{-\Sigma_j T_j \left(\frac{\delta^n_{j,h,t}}{w^j_{n,t}} I\right)^{-\zeta}} dI
\end{aligned}$$

With a change of variable, denote $y \equiv \Sigma_j T_j \left(\frac{w^j_{n,t}}{\delta^n_{j,h,t}}\right)^{\zeta} I^{-\zeta}$, we have $dy = \zeta I^{-\zeta-1} \Sigma_j T_j \left(\frac{w^j_{n,t}}{\delta^n_{j,h,t}}\right)^{\zeta} dI$. Plugging this into the expected value for the maximum income and using the definition of the Gamma function $\Gamma(t) = \int_0^\infty y^{t-1} e^{-y} \, dy$:

$$\begin{aligned}
E(I_{n,h,t}) &= \left[\Sigma_j T_j \left(\frac{w^j_{n,t}}{\delta^n_{j,h,t}}\right)^{\zeta}\right]^{1/\zeta} \int_0^\infty y^{-1/\zeta} e^{-y} \, dy \\
&= \gamma \left[\Sigma_j T_j \left(\frac{w^j_{n,t}}{\delta^n_{j,h,t}}\right)^{\zeta}\right]^{1/\zeta}
\end{aligned}$$

where $\gamma \equiv \Gamma(1 - 1/\zeta)$.

Thus, the expected wage is given by:

$$\bar{I}_{n,h,t} = E(I_{n,h,t}) = \gamma \left(\Sigma_j T_j \left(\frac{w^j_{n,t}}{\delta^n_{j,h,t}} \right)^{\zeta} \right)^{\frac{1}{\zeta}}$$

Given the distribution of income in equation (1.8), expected utility is given by:

$$E(U_{i,h}) = \int_0^{\infty} \left[(1-k_i) \left(\frac{1}{\eta} \left(\frac{y_{i,h}}{p^{\phi}_{i,A} p^{1-\phi}_{i,M}} \right)^{\eta} - \nu \ln \left(\frac{p_{i,A}}{p_{i,M}} \right) \right) + k_i n^{\eta_f}_{i,h} \Pi'_{i,h} \right] dG_{n,h}(I)$$

This gives the equation (1.19).

To determine the effective unit labor in each market, we first calculate the joint distribution of ϕ^A_n and ϕ^M_n:

$$F_{\phi^A_n, \phi^M_n}(\phi^A_{n,i}, \phi^M_{n,i}) = e^{-T_A(\phi^A_{n,i})^{-\zeta} - T_M(\phi^M_{n,i})^{-\zeta}}$$

The expected value of ϕ^j_n given that $\frac{\phi^j_n w^j_n}{\delta_{j,h}} > \frac{\phi^{-j}_n w^{-j}_n}{\delta_{-j,h}}$ is:

$$
\begin{aligned}
E(\phi_n^j \mid \frac{\phi_n^j w_n^j}{\delta_{j,h,t}^n} > \frac{\phi_n^{-j} w_n^{-j}}{\delta_{-j,h,t}^n}) &= \int_0^\infty \int_0^{\frac{\delta_{-j,h,t}^n \phi_n^j w_n^j}{w_n^{-j} \delta_{j,h,t}^n}} \phi_n^j \\
&\times f_{\phi_n^A, \phi_n^M}(\phi_{n,i}^A, \phi_{n,i}^M)\, d\phi_n^{-j}\, d\phi_n^j \\
&= \int_0^\infty \phi_n^j e^{-T_{-j}(\frac{\delta_{-j,h,t}^n \phi_n^j w_n^j}{w_n^{-j} \delta_{j,h,t}^n})^{-\zeta}} \\
&\times (\zeta T_j (\phi_{n,i}^j)^{-\zeta-1}) e^{-T_j(\phi_{n,i}^j)^{-\zeta}}\, d\phi_n^j \\
&= \int_0^\infty (\zeta T_j (\phi_{n,i}^j)^{-\zeta}) e^{-((\frac{\delta_{-j,h,t}^n w_n^j}{w_n^{-j} \delta_{j,h,t}^n})^{-\zeta} T_{-j} + T_j)(\phi_{n,i}^j)^{-\zeta}} d\phi_{n,i}^j \\
&= \frac{T_j}{((\frac{\delta_{-j,h,t}^n w_n^j}{w_n^{-j} \delta_{j,h,t}^n})^{-\zeta} T_{-j} + T_j)} \\
&\times \int_0^\infty \phi_{n,i}^j (\zeta (\phi_{n,i}^j)^{-\zeta-1}) ((\frac{\delta_{-j,h,t}^n w_n^j}{w_n^{-j} \delta_{j,h,t}^n})^{-\zeta} T_{-j} + T_j) \\
&\times e^{-((\frac{\delta_{-j,h,t}^n w_n^j}{w_n^{-j} \delta_{j,h,t}^n})^{-\zeta} T_{-j} + T_j)(\phi_{n,i}^j)^{-\zeta}} d\phi_{n,i}^j \\
&= T_j \Gamma(1 - \frac{1}{\zeta}) \left(((\frac{\delta_{-j,h,t}^n w_n^j}{w_n^{-j} \delta_{j,h,t}^n})^{-\zeta} T_{-j} + T_j)^{\frac{1}{\zeta}-1} \right) \\
&= T_j^{\frac{1}{\zeta}} \Gamma(1 - \frac{1}{\zeta}) \left(\frac{T_j (w_n^j / \delta_{j,h,t}^n)^\zeta}{\Sigma_j T_j (w_n^j / \delta_{j,h,t}^n)^\zeta} \right)^{1-\frac{1}{\zeta}} \\
&= T_j^{\frac{1}{\zeta}} \Gamma(1 - \frac{1}{\zeta}) (s_{n,h,t}^j)^{1-\frac{1}{\zeta}}
\end{aligned}
$$

The total effective labor supply is:

$$
\tilde{L}_{n,h}^j = L_{n,h} T_j^{\frac{1}{\zeta}} \Gamma(1 - \frac{1}{\zeta}) (s_{n,h}^j)^{1-\frac{1}{\zeta}}
$$

1.A.3 Location Choice

Individuals belonging to generation $t-1$ decide whether to migrate from their current region r to another region n that offers the highest utility $\mathscr{U}_{rn,h,t}^i$. This is determined by (i) the utility in region n, which is determined by the expected utility $E(U_{n,h,t})$, capturing the consumption and offspring utilities an individual with Hukou h can derive from living in region n; (ii) the amenity level $A_{n,t}$ in region n, reflecting the attractiveness of the location; (iii) idiosyncratic preference shocks $\epsilon_{n,h,t}^i$, which follow an i.i.d. Fréchet distribution

$\mathscr{F}(\epsilon^i_{n,h,t}) = e^{-(\epsilon^i_{n,h,t})^{-\epsilon}}$; and (iv) the migration cost from region r to region n for Hukou h, denoted by $\tau_{rn,h,t}$:

$$\mathscr{U}^i_{n,h,t} = A_{n,t}E(U_{n,h,t})\frac{\epsilon^i_{n,h,t}}{\tau_{rn,h,t}}$$

Denote $\tilde{\mathscr{U}}^i_{rn,h,t} \equiv \bar{u}_{n,h,t}\frac{\epsilon^i_{n,h,t}}{\tau_{rn,h,t}}$, where $\bar{u}_{n,h,t} \equiv A_{n,t}E(U_{n,h,t})$. Then $\tilde{\mathscr{U}}^i_{rn,h,t}$ also follows a Fréchet distribution:

$$\begin{aligned}
\mathscr{G}_{rn,h,t}(u) &\equiv \Pr(\tilde{\mathscr{U}}^i_{rn,h,t} < u) \\
&= \Pr\left(\bar{u}_{n,h,t}\frac{\epsilon^i_{n,h,t}}{\tau_{rn,h,t}} < u\right) \\
&= \Pr\left(\epsilon^i_{n,h,t} < \frac{\tau_{rn,h,t}}{\bar{u}_{n,h,t}}u\right) \\
&= \mathscr{F}\left(\frac{\tau_{rn,h,t}}{\bar{u}_{n,h,t}}u\right) \\
&= e^{-\left(\frac{\tau_{rn,h,t}}{\bar{u}_{n,h,t}}u\right)^{-\epsilon}}
\end{aligned}$$

Then the share of migrants from region r to region n is given by:

$$\begin{aligned}
m_{rn,h,t} &\equiv \Pr[\tilde{\mathscr{U}}^i_{rn,h,t} \geq \tilde{\mathscr{U}}^i_{rm,h,t}, \forall m \neq n] \\
&= \int_0^\infty \Pr(u \geq \tilde{\mathscr{U}}^i_{rm,h,t}, \forall m \neq n)\, d\mathscr{G}_{rn,h,t}(u) \\
&= \int_0^\infty \Pi_{m\neq n}\mathscr{G}_{rm,h,t}(u)\, d\mathscr{G}_{rn,h,t}(u) \\
&= \left(\frac{\bar{u}_{n,h,t}}{\tau_{rn,h,t}}\right)^\epsilon \int_0^\infty \epsilon u^{-\epsilon-1}\Pi_m e^{-\left(\frac{\tau_{rm,h,t}}{\bar{u}_{m,h,t}}u\right)^{-\epsilon}} du \\
&= \frac{\left(\frac{\bar{u}_{n,h,t}}{\tau_{rn,h,t}}\right)^\epsilon}{\Sigma_m\left(\frac{\bar{u}_{m,h,t}}{\tau_{rm,h,t}}\right)^\epsilon} \int_0^\infty \epsilon\Sigma_m\left(\frac{\bar{u}_{m,h,t}}{\tau_{rm,h,t}}\right)^\epsilon u^{-\epsilon-1} e^{-u^{-\epsilon}\Sigma_m\left(\frac{\bar{u}_{m,h,t}}{\tau_{rm,h,t}}\right)^\epsilon} du \\
&= \frac{\left(\frac{\bar{u}_{n,h,t}}{\tau_{rn,h,t}}\right)^\epsilon}{\Sigma_m\left(\frac{\bar{u}_{m,h,t}}{\tau_{rm,h,t}}\right)^\epsilon}\left[-e^{-u^{-\epsilon}\Sigma_m\left(\frac{\bar{u}_{m,h,t}}{\tau_{rm,h,t}}\right)^\epsilon}\right]_0^\infty \\
&= \frac{\left(\frac{\bar{u}_{n,h,t}}{\tau_{rn,h,t}}\right)^\epsilon}{\Sigma_m\left(\frac{\bar{u}_{m,h,t}}{\tau_{rm,h,t}}\right)^\epsilon}
\end{aligned}$$

This gives equation (1.6) in the chapter.

Then the expected value for the maximum utility is given by:

$$\Pi_{r,h,t} \equiv E(\max_n\{\tilde{\mathscr{U}}^i_{rn,h,t}\})$$

Calculating this expected value by plugging in distribution $\mathscr{G}_{r,h,t}(u)$:

$$\begin{aligned}
\Pi_{r,h,t} &= \int_0^\infty u\, d\mathscr{G}_{r,h,t}(u) \\
&= \epsilon\Sigma_m \left(\frac{\bar{u}_{m,h,t}}{\tau_{rm,h,t}}\right)^\epsilon \int_0^\infty u^{-\epsilon} e^{-\Sigma_n \left(\frac{\tau_{rn,h,t}}{\bar{u}_{n,h,t}} u\right)^{-\epsilon}} du \\
&= \left[\Sigma_m \left(\frac{\bar{u}_{m,h,t}}{\tau_{rm,h,t}}\right)^\epsilon\right]^{1/\epsilon} \int_0^\infty y^{-1/\epsilon} e^{-y}\, dy \\
&= \Gamma(1-1/\epsilon)\left[\Sigma_m \left(\frac{\bar{u}_{m,h,t}}{\tau_{rm,h,t}}\right)^\epsilon\right]^{1/\epsilon}
\end{aligned}$$

Thus, the expected value for the maximum utility is:

$$\Pi_{r,h,t} = E(\max_n\{\tilde{\mathscr{U}}^i_{rn,h,t}\}) = \Gamma(1-1/\epsilon)\left[\Sigma_m \left(\frac{\bar{u}_{m,h,t}}{\tau_{rm,h,t}}\right)^\epsilon\right]^{1/\epsilon}$$

This gives the expected future utility for offspring in consumers problem.

Finally, the law of motion for labor is given by:

$$L_{n,h,t} = \Sigma_r m_{rn,h,t} \bar{L}_{r,h,t} \int (1 - q n^{\eta_q}_{n,h,t})\, dG_{n,h,t}(I)$$

The term $\bar{L}_{r,h,t}$ represents the population at the end of period $t-1$ that is making migration choices. This population is determined by the fertility decisions and labor allocation at time $t-1$:

$$\bar{L}_{r,h,t} = L_{r,h,t-1} \int \frac{n_{r,h,t-1}}{1 - q n^{\eta_q}_{r,h,t-1}}\, dG_{r,h,t-1}(I)$$

The integral is taken over the distribution of income since consumers with different realizations of income will make different fertility decisions. The aggregate labor supply, $L_{n,h,t}$, is the sum over all Hukou, adjusted for the time share spent on childcare, under the realization of income following the $G_{n,h,t}(I)$ distribution.

This gives (1.11) (1.12) in the chapter.

1.B Derivation of the Production Decision Problem

In this section, I provide a detailed derivation of the production equations presented in the chapter. Each region has two sectors: agriculture and non-agriculture. Both sectors use labor and land as inputs, and they choose these inputs to maximize their profit.

The agricultural sector is modeled as perfectly competitive, following the Eaton-Kortum structure, while the non-agricultural sector is modeled as monopolistic competition. Both sectors are tradable and subject to sector-specific iceberg trade costs.

1.B.1 Agricultural Production

The agricultural production function follows a Cobb-Douglas production function, using effective labor $\tilde{L}^A_{n,t}$ and land $H^A_{n,t}$ as inputs. Each location produces a continuum of products $\omega \in [0,1]$:

$$q^A_{n,t}(\omega) = z^{nA}(\omega) A^A_{n,t} (\tilde{L}^A_{n,t})^{\gamma^A_L} (H^A_{n,t})^{\gamma^A_H}$$

where $z^{nA}(\omega)$ is a variety-specific productivity factor drawn from a Fréchet distribution with cumulative distribution function $F(z^{nA}) = e^{-(z^{nA})^{-\theta^A}}$. The parameters γ^A_L and γ^A_H represent the shares of labor and land in agricultural production, and $A^A_{n,t}$ denotes the average TFP in the agricultural sector in region n at time t.

Since the agricultural sector is under perfect competition, the first-order condition gives the unit cost:

$$c^A_{i,t}(\omega) = \frac{1}{A^A_{i,t}} k_A (w^A_{i,t})^{\gamma^A_L} (r^A_{n,t})^{\gamma^A_H}$$

where $k_A = \left(\left(\frac{\gamma^A_L}{\gamma^A_H} \right)^{\gamma^A_H} + \left(\frac{\gamma^A_H}{\gamma^A_L} \right)^{\gamma^A_L} \right)$ is a sector-specific constant.

Goods are tradable subject to iceberg trade costs. The price of ω sold from location i to location n is given by:

$$p^A_{in,t}(\omega) = \frac{\kappa^A_{in,t}}{z^{nA}(\omega)} c^A_{i,t}(\omega)$$

Final goods in the agricultural sector are produced using $\omega \in [0,1]$ from locations that have the lowest price:

$$Q_{n,t}^A = \left(\int (\tilde{q}_{n,t}^A(\omega))^{1-1/\eta^A} \, d\omega \right)^{\eta^A/(\eta^A-1)}$$

The Fréchet distribution of prices implies that the share of total expenditure in region n on agriculture from region i is given by:

$$\lambda_t^{in,A} = \frac{(\kappa_{in,t}^A c_t^{iA})^{-\theta^A}}{\Sigma_{m=1}^N (\kappa_{nm,t}^A c_t^{mA})^{-\theta^A}} \tag{1.22}$$

Given the trade share and the total expenditure in each region, the trade value from region i to region n is given by:

$$X_{in,t}^A = \frac{(\kappa_{in,t}^A c_{i,t}^A)^{-\theta^A}}{\Sigma_{m=1}^N (\kappa_{nm,t}^A c_{m,t}^A)^{-\theta^A}} E_{n,t}^A$$

This gives the equation (1.17) in the chapter.

The price of agricultural goods in region n is given by:

$$P_{n,t}^A = \Gamma_{n,t}^A \left(\Sigma_{m=1}^N (\kappa_{mn,t}^A c_{m,t}^A)^{-\theta^A} \right)^{-1/\theta^A}$$

where $\Gamma_{n,t}^A \equiv \Gamma\left(1 + \frac{1-\eta^A}{\theta^A}\right)$.

The first-order conditions for labor and land give:

$$w_{n,t}^A = \frac{\gamma_L^A}{\tilde{L}_{n,t}^A} \Sigma_i \lambda_t^{in,A} E_{i,t}^A \tag{1.23}$$

$$r_{n,t}^A = \frac{\gamma_H^A}{H_{n,t}^A} \Sigma_i \lambda_t^{in,A} E_{i,t}^A$$

Plugging the trade share (1.22) into the FOC for labor (1.23) gives the equation (1.18) in the chapter:

$$w_{n,t}^A = \frac{\gamma_L^A}{\tilde{L}_{n,t}^A} (c_t^{nA})^{-\theta^A} \Sigma_i \left(\frac{\kappa_{in,t}^A}{P_{n,t}^A} \right)^{-\theta^A} E_{i,t}^A$$

1.B.2 Non-Agricultural Production

Non-agricultural production is under monopolistic competition. In each region, the non-agricultural variety $\omega_{n,t}^M$ is produced using a Cobb-Douglas production technology, utilizing land and effective labor as inputs:

$$q_{n,t}^M(\omega_{n,t}^M) = A_{n,t}^M(\tilde{L}_{n,t}^M(\omega_{n,t}^M))^{\gamma_L^M}(H_{n,t}^M(\omega_{n,t}^M))^{\gamma_H^M}$$

Final non-agricultural goods producers produce final goods using all varieties from all locations:

$$Q_{n,t}^M = \left(\Sigma_i \int (q_{in,t}^M(\omega))^{1-1/\eta^M}\, d\omega\right)^{\eta^M/(\eta^M-1)}$$

For the final goods producer problem, they choose the optimal demand by solving the following problem:

$$\begin{aligned}&\max\ P_{n,t}^M Q_{n,t}^M - E_{n,t}^M \\ &= P_{n,t}^M \left(\Sigma_i \int (q_{in,t}^M(\omega))^{1-1/\eta^M}\, d\omega\right)^{\eta^M/(\eta^M-1)} - \Sigma_i \int p_{in,t}^M(\omega) q_{in,t}^M(\omega)\, d\omega\end{aligned}$$

The first-order condition requires:

$$P_{n,t}^M \left(\Sigma_i \int (q_{in,t}^M(\omega))^{1-1/\eta^M}\, d\omega\right)^{\frac{1}{\eta^M-1}} q_{in,t}^M(\omega)^{-\frac{1}{\eta^M}} = p_{in,t}^M(\omega)$$

$$\frac{q_{in,t}^M(\omega)^{-\frac{1}{\eta^M}}}{q_{mn,t}^M(\omega)^{-\frac{1}{\eta^M}}} = \frac{p_{in,t}^M(\omega)}{p_{mn,t}^M(\omega)} \tag{1.24}$$

Plugging (1.24) back into the total expenditure function gives the demand:

$$E_{n,t}^M = \Sigma_i \int p_{in,t}^M(\omega) \left(\frac{p_{in,t}^M(\omega)}{p_{mn,t}^M(\omega)}\right)^{-\eta^M} q_{mn,t}^M(\omega)\, d\omega$$

$$q_{mn,t}^M(\omega) = \frac{p_{mn,t}^M(\omega)^{-\eta^M} E_{n,t}^M}{\Sigma_i \int p_{in,t}^M(\omega)^{1-\eta^M}\, d\omega} \tag{1.25}$$

Final goods producers make zero profit, thus:

$$P_{n,t}^M Q_{n,t}^M - E_{n,t}^M = 0 \tag{1.26}$$

Plugging the demand (1.25) into the production function of final goods and combining with (1.26):

$$E_{n,t}^M = \Sigma_i \int p_{in,t}^M(\omega) \left(\frac{p_{in,t}^M(\omega)}{P_{n,t}^M \left(\Sigma_i \int (q_{in,t}^M(\omega))^{1-1/\eta^M} \, d\omega \right)^{\frac{1}{\eta^M - 1}}} \right)^{\eta^M} d\omega$$

Rearranging gives:

$$P_{n,t}^M Q_{n,t}^M = (P_{n,t}^M)^{\eta^M} Q_{n,t}^M \Sigma_i \int (p_{in,t}^M(\omega))^{1-\eta^M} \, d\omega$$

This gives the price index:

$$P_{n,t}^M = \left(\Sigma_i \int (p_{in,t}^M(\omega))^{1-\eta^M} \, d\omega \right)^{\frac{1}{1-\eta^M}} \tag{1.27}$$

Plugging the price index (1.27) back into the demand function (1.25) gives the final demand function for each variety:

$$q_{in,t}^M(\omega) = (p_{in,t}^M(\omega))^{-\eta^M} (P_{n,t}^M)^{\eta^M - 1} E_{n,t}^M \tag{1.28}$$

Given the demand function, a non-agricultural firm in location i supplies $q_{in,t}^M$ to all locations n to maximize its total profit. Total profit is given by:

$$\max_{\{q_{in,t}^M\}_{n \in N}} \Sigma_n \left(p_{in,t}^M(\omega) q_{in,t}^M(\omega) - \kappa_{in,t}^M q_{in,t}^M(\omega) c_{i,t}^M(\omega) \right)$$

subject to

$$q_{in,t}^M(\omega) = (p_{in,t}^M(\omega))^{-\eta^M} (P_{n,t}^M)^{\eta^M - 1} E_{n,t}^M$$

where the unit cost function is given by the firm's cost minimization problem:

$$c_{i,t}^M(\omega) = \frac{1}{A_{i,t}^M}(w_{i,t}^M)^{\gamma_L^M}(r_{i,t}^M)^{\gamma_H^M} k_M$$

with $k_M \equiv \left(\frac{\gamma_L^M}{\gamma_H^M}\right)^{\gamma_L^M} + \left(\frac{\gamma_H^M}{\gamma_L^M}\right)^{\gamma_H^M}$ being a constant.

The first-order condition gives the price for firms in location i selling to location n as a constant markup over unit cost:

$$p_{in,t}^M(\omega) = \frac{\eta^M \kappa_{in,t}^M c_{i,t}^M(\omega)}{\eta^M - 1}$$

Plugging this into the price index (1.27), the aggregate price is given by:

$$P_{n,t}^M = \left(\Sigma_i N_{i,t}^M \left(\frac{\eta^M \kappa_{in,t}^M c_{i,t}^M(\omega)}{\eta^M - 1}\right)^{1-\eta^M}\right)^{\frac{1}{1-\eta^M}}$$

Given the demand and price, bilateral trade from region i to region n is given by:

$$X_{in,t}^M = N_{i,t}^M \left(\frac{\eta^M}{\eta^M - 1}\frac{\kappa_{in,t}^M c_{i,t}^M}{P_{n,t}^M}\right)^{1-\eta^M} E_{n,t}^M = \frac{N_{i,t}^M(\kappa_{in,t}^M c_{i,t}^M)^{1-\eta^M}}{\Sigma_i N_{i,t}^M(\kappa_{in,t}^M c_{i,t}^M(\omega))^{1-\eta^M}} E_{n,t}^M$$

Therefore, the share of expenditure in equation (1.15) in the chapter is given by:

$$\lambda_t^{in,M} = \frac{N_{i,t}^M(\kappa_{in,t}^M c_{i,t}^M)^{1-\eta^M}}{\Sigma_m N_{m,t}^M(\kappa_{mn,t}^M c_{m,t}^M(\omega))^{1-\eta^M}}$$

Total revenue for each firm is given by:

$$\begin{aligned} R_{i,t}^M(\omega) &\equiv \Sigma_n p_{in,t}^M(\omega) q_{in,t}^M(\omega) \\ &= \Sigma_n \frac{\eta^M \kappa_{in,t}^M c_{i,t}^M(\omega)}{\eta^M - 1}\left(\frac{\eta^M \kappa_{in,t}^M c_{i,t}^M(\omega)}{\eta^M - 1}\right)^{-\eta^M}(P_{n,t}^M)^{\eta^M - 1} E_{n,t}^M \\ &= \left(\frac{\eta^M}{\eta^M - 1}\right)^{1-\eta^M}(c_{i,t}^M(\omega))^{1-\eta^M}\Sigma_n\left(\frac{\kappa_{in,t}^M}{P_{n,t}^M}\right)^{1-\eta^M} E_{n,t}^M \end{aligned}$$

Total costs are a fraction of total revenue. Plugging in the price function, we get:

$$C_{i,t}^M(\omega) \equiv \Sigma_n \kappa_{in,t}^M q_{in,t}^M(\omega) c_{i,t}^M(\omega) = \frac{\eta^M - 1}{\eta^M} R_{i,t}^M(\omega)$$

The first-order condition for the firm's cost minimization problem gives the total payment to production workers as a fraction of the total cost for each firm:

$$w_{n,t}^M \tilde{L}_{n,t}^M(\omega) = \gamma_L^M \frac{\eta^M - 1}{\eta^M} R_{i,t}^M(\omega) \tag{1.29}$$

Profit for a firm in country i from selling to country n is:

$$\pi_{in,t}^M(\omega) = \frac{x_{in,t}^M(\omega)}{\eta^M}$$

where

$$x_{in,t}^M(\omega) \equiv p_{in,t}^M(\omega) q_{in,t}^M(\omega) = \left(\frac{\eta^M}{\eta^M - 1} \frac{\kappa_{in,t}^M c_{i,t}^M}{P_{n,t}^M} \right)^{1-\eta^M} E_{n,t}^M$$

Total profit for a firm in country i selling to all countries is given by:

$$\Pi_{i,t}^M(\omega) = \Sigma_n \pi_{in,t}^M(\omega) = \frac{R_{i,t}^M(\omega)}{\eta^M}$$

The number of firms is determined by the free entry condition, assuming firms only operate for one period:

$$\Pi_{i,t}^M(\omega) = w_{n,t}^M f_e$$

Combining the total profit function and payment to production workers gives:

$$\frac{w_{n,t}^M \tilde{l}_{n,t}^M(\omega)}{\gamma_L^M(\eta^M - 1)} = w_{n,t}^M f_e$$

where we derive the total effective labor input for each firm:

$$\tilde{l}_{n,t}^M(\omega) = \gamma_L^M(\eta^M - 1) f_e$$

Given total labor, the labor market clearing condition gives the number of firms in each region, which is the equation (1.14) in the chapter:

$$N_{n,t}^{M}(f_e + \gamma_L^M(\eta^M - 1)f_e) = \tilde{L}_{n,t}^{M}$$

Plugging in the revenue function into equation (1.29) and combining with equation (1.14) gives the equation (1.16) in the chapter:

$$w_{n,t}^{M} = \frac{1}{\eta^M f_e}\left(\frac{\eta^M}{\eta^M - 1}\right)^{1-\eta^M}(c_{n,t}^{M}(\omega))^{1-\eta^M}\Sigma_i\left[\left(\frac{\kappa_{in,t}^{M}}{P_{n,t}^{M}}\right)^{1-\eta^M}E_{n,t}^{M}\right]$$

1.C Solving Algorithm

In this section, I provide the solving algorithm for the dynamic spatial model presented in the chapter. Firstly, the equilibrium conditions can be summarized by the following equations:

$$X_t^{i,j} = \Sigma_n \lambda_t^{in,j} E_{n,t}^{j} \tag{1.30}$$

$$\lambda_t^{in,M} = \frac{N_{i,t}^{M}(\kappa_{in,t}^{M}c_{i,t}^{M})^{1-\eta^M}}{\Sigma_m N_{m,t}^{M}(\kappa_{mn,t}^{M}c_{m,t}^{M}(\omega))^{1-\eta^M}} \tag{1.31}$$

$$\lambda_t^{in,A} = \frac{(\kappa_{in,t}^{A}c_t^{iA})^{-\theta^A}}{\Sigma_{m=1}^{N}(\kappa_{nm,t}^{A}c_t^{mA})^{-\theta^A}} \tag{1.32}$$

$$N_{n,t}^{M}(f_e + \gamma_L^M(\eta^M - 1)f_e) = \tilde{L}_{n,t}^{M} \tag{1.33}$$

$$w_{n,t}^{M} = \frac{1}{\eta^M f_{e,t}}\left(\frac{\eta^M}{\eta^M - 1}\right)^{1-\eta^M}(c_{n,t}^{M}(\omega))^{1-\eta^M}\Sigma_n\left(\frac{\kappa_{in,t}^{M}}{P_{n,t}^{M}}\right)^{1-\eta^M}E_{n,t}^{M} \tag{1.34}$$

$$w_{n,t}^{A} = \frac{\gamma_L^A}{\tilde{L}_{n,t}^{A}}(c_t^{nA})^{-\theta^A}\Sigma_i\left(\frac{\kappa_{in,t}^{A}}{P_{n,t}^{A}}\right)^{-\theta^A}E_{i,t}^{A} \tag{1.35}$$

$$c_{i,t}^{M}(\omega) = \frac{1}{A_{i,t}^{M}}(w_{i,t}^{M})^{\gamma_L^M}(r_{i,t}^{M})^{\gamma_H^M}k_M \tag{1.36}$$

$$c_{i,t}^{A}(\omega) = \frac{1}{A_{i,t}^{A}} k_A (w_{i,t}^{A})^{\gamma_L^A} (r_{n,t}^{A})^{\gamma_H^A} \tag{1.37}$$

$$P_t^{nM} = \left(\Sigma_i N_{i,t}^{M} \left(\frac{\eta^M \kappa_{in,t}^{M} c_{i,t}^{M}(\omega)}{\eta^M - 1} \right)^{1-\eta^M} \right)^{\frac{1}{1-\eta^M}} \tag{1.38}$$

$$P_t^{nA} = \Gamma^{nA} \left(\Sigma_{m=1}^{N} (\kappa_{mn,t}^{A} c_t^{mA})^{-\theta^A} \right)^{-1/\theta^A} \tag{1.39}$$

$$E_{i,t}^{A} = \int \theta_{A,h,t} y_{h,t} L_{n,t}^{A} \, dG_{n,h,t}(I) + \phi \left(\Sigma_j X_t^{i,j} - \int y_{A,h,t} L_{n,t}^{A} \, dG_{n,h,t}(I) \right) \tag{1.40}$$

$$E_{i,t}^{M} = \int (1 - \theta_{A,h,t}) y_{h,t} L_{n,t}^{A} \, dG_{n,h,t}(I) + (1-\phi) \left(\Sigma_j X_t^{i,j} - \int y_{A,h,t} L_{n,t}^{A} \, dG_{n,h,t}(I) \right) \tag{1.41}$$

$$\frac{w_{n,t}^{j} \tilde{L}_{n,t}^{j}}{r_{n,t}^{j} H_{n,t}^{j}} = \frac{\gamma_L^j}{\gamma_H^j} \tag{1.42}$$

$$L_{n,h,t}^{j} = s_{n,h}^{j} L_{n,h,t} \tag{1.43}$$

$$\tilde{L}_{n,h}^{j} = L_{n,h} T_j^{\frac{1}{\zeta}} \Gamma \left(1 - \frac{1}{\zeta} \right) (s_{n,h}^{j})^{1-\frac{1}{\zeta}} \tag{1.44}$$

$$L_{n,h,t} = \Sigma_r m_{rn,h,t} \bar{L}_{r,h,t} \int (1 - q n_{n,h,t}^{\eta_q}) \, dG_{n,h,t}(I) \tag{1.45}$$

$$\bar{L}_{r,h,t} = L_{r,h,t-1} \int \frac{n_{r,h,t-1}}{1 - q n_{r,h,t}^{\eta_q}} \, dG_{r,h,t-1}(I) \tag{1.46}$$

$$s_{n,h,t}^{j} = \frac{T_j \left(w_{n,t}^{j} / \delta_{j,h,t}^{n} \right)^{\zeta}}{\Sigma_i T_i \left(w_{n,t}^{i} / \delta_{i,h,t}^{n} \right)^{\zeta}} \tag{1.47}$$

$$m_{rn,h,t} = \frac{(\bar{u}_{n,h,t} / \tau_{rn,h,t})^{\epsilon}}{\Sigma_m \left(\bar{u}_{m,h,t} / \tau_{rm,h,t} \right)^{\epsilon}} \tag{1.48}$$

$$E(U_{i,h,t}) = \int_0^\infty \left[(1-k_i) \left(\frac{1}{\eta} \left(\frac{y^*_{i,h,t}}{p^\phi_{i,A,t} p^{1-\phi}_{i,M,t}} \right)^\eta - \nu \ln \left(\frac{p_{i,A,t}}{p_{i,M,t}} \right) \right) + k_i n^{*\eta_f}_{i,h,t} \Pi_{i,h,t+1} \right] dG_{n,h}(I) \tag{1.49}$$

$$\Pi_{r,h,t} = \Gamma \left(1 - \frac{1}{\epsilon} \right) \left(\Sigma_m \left(\bar{u}_{m,h,t} / \tau_{rm,t} \right)^\epsilon \right)^{\frac{1}{\epsilon}} \tag{1.50}$$

$$y^*_{i,h,t} + \chi_{i,h,t}(n^*_{i,h,t} - 1) I_{i,h,t} = (1 - q n^{*\eta_q}_{i,h,t}) I_{i,h,t} \tag{1.51}$$

$$-(1-k_i) \left(\frac{1 - q n^*_{i,h,t} - \chi_{i,h}(n^*_{i,h,t} - 1)}{p^\phi_A p^{1-\phi}_M} \right)^{\eta-1} \left(\frac{q \eta_q n^{*\eta_q - 1}_{i,h,t} + \chi_{i,h}}{p^\phi_A p^{1-\phi}_M} \right) I^\eta_{i,h,t} + \eta_f k_i n^{*\eta_f - 1}_{i,h,t} \Pi_{i,h,t+1} = 0 \tag{1.52}$$

$$G_{n,h}(I) = e^{-\Sigma_j T_j \left(\frac{\delta_{j,h}}{w^j_n} I \right)^{-\zeta}} \tag{1.53}$$

Given the initial distribution of labor endowments $\{L_{r,h,0}\}_{r \in N, h \in \{A,M\}}$, land endowments $\{H_{r,0}\}_{r \in N}$, fertility $\{n_{r,h,0}\}_{r \in N, h \in \{A,M\}}$, and a path of fundamentals including FPPs fines $\{\chi_{r,h,t}\}_{r \in N, h \in \{A,M\}, t \geq 0}$, productivity $\{A^A_{r,t}, A^M_{r,t}\}_{r \in N, t \geq 0}$, Hukou friction $\{\delta^A_{r,t,h}, \delta^M_{r,t,h}\}_{r \in N, h \in \{A,M\}, t \geq 0}$, migration cost $\{\tau_{or,t,h}\}_{o,r \in N, h \in \{A,M\}, t \geq 0}$, and trade cost $\{\kappa_{or,t,s}\}_{o,r,s \in N, t \geq 0}$, I solve the model in the following steps:

1. Define $V_{n,h,t} \equiv E(U_{n,h,t})$. Guess a convergent sequence of $\{V^0_{n,h,t}\}^{T+1}_{t=0}$ with $V^0_{n,h,T+1} = V^0_{n,h,T}$.
2. For $0 \leq t \leq T$, given the allocation of labor at time t, $L_{r,h,t}$, calculate the equilibrium at time t in the following ways:
   (a) Guess a vector of wages $w^j_{i,t}$.
   (b) Given labor supply at time t, calculate the effective unit labor supply using (1.44).
   (c) Given labor and land, which we assume are at fixed supply, calculate land prices using equation (1.42).
   (d) Calculate the price index using equations (1.37), (1.36), (1.39), (1.38), and (1.33).

(e) Given wages and land prices, solve for the expected fertility, income, and expenditure share:
   i. Generate a realization of income $I_{i,h,t}$ following the $G_{n,h}(I)$ Fréchet distribution as in equation (1.53) with the number of draws equal to the population for $L_{r,h,t}$.
   ii. Given the price index and $V_{n,h,t}$, solve the optimal expenditure on consumption $y^*_{i,h,t}$ and optimal fertility $n^*_{i,h,t}$ for each realization of income $I_{i,h,t}$ using equations (1.51) and (1.52).
   iii. Solve consumers' expenditure share on each sector $\int \theta_{A,h,t} y_{h,t} L^A_{n,t}\, dG_{n,h,t}(I)$.

(f) Plug the expenditure share into the market clearing condition (1.30) and combine with (1.32), (1.31), (1.40), and (1.41) to solve the total expenditure $E^{i,j}_t$.

(g) Calculate a new vector of wages $w^j_{i,t}$ from the labor market clearing condition (1.35) and (1.34). Iterate until the wages converge.

(h) Given the optimal fertility $n^*_{i,h,t}$ for each realization of income, plug into the law of motion for labor (1.43) and (1.46) to get the next period's allocation of labor $L_{r,h,t+1}$.

3. Given the new sequence of prices and optimal consumption $y^*_{i,h,t}$ and optimal fertility $n^*_{i,h,t}$, calculate the new path for $\{V^1_{n,h,t}\}^T_{t=0}$ using equations (1.50) and (1.50).
4. Check the distance between the new path $\{V^1_{n,h,t}\}^T_{t=0}$ and $\{V^0_{n,h,t}\}^T_{t=0}$, and iterate until the path converges.

1.D Proof for Proposition 1

I start with equations (1.18) and (1.16) in the chapter:

$$w^A_{n,t} = \frac{\gamma^A_L}{\tilde{L}^A_{n,t}} (c^{nA}_t)^{-\theta^A} \Sigma_i \left(\frac{\kappa^A_{in,t}}{P^A_{n,t}}\right)^{-\theta^A} E^A_{i,t}$$

$$w^M_{n,t} = \frac{1}{\eta^M f_{e,t}} \left(\frac{\eta^M}{\eta^M - 1}\right)^{1-\eta^M} (c^M_{n,t}(\omega))^{1-\eta^M} \Sigma_n \left(\frac{\kappa^M_{in,t}}{P^M_{n,t}}\right)^{1-\eta^M} E^M_{n,t}$$

Taking the log and then the total differential, we get:

$$(1+\theta^A) d\ln w^A_{n,t} = [-(\theta^A \gamma^A_H + 1) d\ln L^A_{n,t} + \theta^A d\ln A^A_{n,t} + d\ln \mathscr{M}^A_{n,t}] \tag{1.54}$$

$$\eta^M d\ln w^M_{n,t} = -d\ln f^M_{e,t} + \gamma^M_H(1-\eta^M)d\ln(\tilde{L}^M_{n,t}) + (\eta^M - 1)d\ln(A^M_{n,t}) + d\ln \mathscr{M}^M_{n,t} \quad (1.55)$$

where $\mathscr{M}^A_{n,t}$ and $\mathscr{M}^M_{n,t}$ are the respective market conditions for agriculture and non-agriculture sectors.

We can also define the share of effective labor as:

$$\tilde{s}^j_{n,h,t} = \frac{\tilde{L}^j_{n,h,t}}{\Sigma_j \tilde{L}^j_{n,h,t}} = \frac{L_{n,h,t}T_j^{\frac{1}{\zeta}}(s^j_{n,h,t})^{1-\frac{1}{\zeta}}}{\Sigma_j L_{n,h,t}T_j^{\frac{1}{\zeta}}(s^j_{n,h,t})^{1-\frac{1}{\zeta}}} = \frac{L^j_{n,h,t}T_j^{\frac{1}{\zeta}}(s^j_{n,h,t})^{-\frac{1}{\zeta}}}{\Sigma_j L^j_{n,h,t}T_j^{\frac{1}{\zeta}}(s^j_{n,h,t})^{-\frac{1}{\zeta}}}$$

$$= \frac{L^j_{n,h,t}\bar{w}_{n,h,t}\delta^n_{j,h,t}/w^j_{n,h,t}}{\Sigma_j L^j_{n,h,t}\bar{w}_{n,h,t}\delta^n_{j,h,t}/w^j_{n,h,t}} = \frac{s^j_{n,h,t}\delta^n_{j,h,t}/w^j_{n,h,t}}{\Sigma_j s^j_{n,h,t}\delta^n_{j,h,t}/w^j_{n,h,t}}$$

Then, the ratio of the shares of effective labor is given by:

$$\frac{\tilde{s}^A_{n,h,t}}{\tilde{s}^M_{n,h,t}} = \frac{s^A_{n,h,t}\delta^n_{A,h,t}/w^A_{n,h,t}}{s^M_{n,h,t}\delta^n_{M,h,t}/w^M_{n,h,t}}$$

Taking the logarithm and then the total differential, we get:

$$d\ln\frac{\tilde{s}^A_{n,h,t}}{\tilde{s}^M_{n,h,t}} = d\ln\frac{s^A_{n,h,t}}{s^M_{n,h,t}} + d\ln\frac{\delta^n_{A,h,t}}{\delta^n_{M,h,t}} + d\ln\frac{w^M_{n,h,t}}{w^A_{n,h,t}} \quad (1.56)$$

Additionally, we have the shares summing up to 1:

$$\tilde{s}^A_{n,h,t} + \tilde{s}^M_{n,h,t} = 1$$

Taking the differential, we get:

$$\tilde{s}^A_{n,h,t}d\ln\tilde{s}^A_{n,h,t} + \tilde{s}^M_{n,h,t}d\ln\tilde{s}^M_{n,h,t} = 0$$

The changes in relative labor share can also be expressed as:

$$d\ln\frac{s^A_{n,h,t}}{s^M_{n,h,t}} = \zeta\left(d\ln w^A_{n,t} - d\ln w^M_{n,t}\right) + \zeta\left(d\ln\delta^n_{M,h,t} - d\ln\delta^n_{A,h,t}\right) \quad (1.57)$$

We have changes in effective unit labor as:

$$\tilde{L}^j_{n,t} = \Sigma_h \tilde{s}^j_{n,h,t} \tilde{L}_{n,h,t}$$

$$d\tilde{L}^j_{n,t} = \Sigma_h d\tilde{s}^j_{n,h,t} \tilde{L}_{n,h,t} + \tilde{s}^j_{n,h,t} d\tilde{L}_{n,h,t}$$

$$d \ln \tilde{L}^j_{n,t} = \Sigma_h \left(d \ln \tilde{s}^j_{n,h,t} + d \ln \tilde{L}_{n,h,t} \right) \frac{\tilde{L}^j_{n,h,t}}{\tilde{L}^j_{n,t}} \tag{1.58}$$

Plugging equations (1.54), and (1.55) into equation (1.56) gives:

$$\begin{aligned}
d \ln \frac{\tilde{s}^A_{n,h,t}}{\tilde{s}^M_{n,h,t}} &= \zeta \left(d \ln w^A_{n,t} - d \ln w^M_{n,t} \right) + \zeta \left(d \ln \delta^n_{M,h,t} - d \ln \delta^n_{A,h,t} \right) \\
&\quad + d \ln \frac{\delta^n_{A,h,t}}{\delta^n_{M,h,t}} + d \ln \frac{w^M_{n,h,t}}{w^A_{n,h,t}} \\
&= \frac{\zeta - 1}{1 + \theta^A} \left[-(\theta^A \gamma^A_H + 1) d \ln \tilde{L}^A_{n,t} + \theta^A d \ln A^A_{n,t} + d \ln \mathscr{M}^A_{n,t} \right] \\
&\quad - \frac{\zeta - 1}{\eta^M} \left[-d \ln f^M_{e,t} + \gamma^M_H (1 - \eta^M) d \ln \tilde{L}^M_{n,t} + (\eta^M - 1) d \ln A^M_{n,t} + d \ln \mathscr{M}^M_{n,t} \right] \\
&\quad + (\zeta - 1) \left(d \ln \delta^n_{M,h,t} - d \ln \delta^n_{A,h,t} \right) \\
&= \frac{\zeta - 1}{1 + \theta^A} \left[-(\theta^A \gamma^A_H + 1) \sum_h (d \ln \tilde{s}^A_{n,h,t} + d \ln \tilde{L}_{n,h,t}) \frac{\tilde{L}^A_{n,h,t}}{\tilde{L}^A_{n,t}} \right] \\
&\quad + \frac{\zeta - 1}{1 + \theta^A} \left[\theta^A d \ln A^A_{n,t} + d \ln \mathscr{M}^A_{n,t} \right] \\
&\quad + (\zeta - 1) \left(d \ln \delta^n_{M,h,t} - d \ln \delta^n_{A,h,t} \right) \\
&\quad - \frac{\zeta - 1}{\eta^M} \left[-d \ln f^M_{e,t} + \gamma^M_H (1 - \eta^M) \sum_h (d \ln \tilde{s}^M_{n,h,t} + d \ln \tilde{L}_{n,h,t}) \frac{\tilde{L}^M_{n,h,t}}{\tilde{L}^M_{n,t}} \right] \\
&\quad - \frac{\zeta - 1}{\eta^M} \left[(\eta^M - 1) d \ln A^M_{n,t} + d \ln \mathscr{M}^M_{n,t} \right]
\end{aligned}$$

Write the equation for both Hukou types and combine with (1.57), then write in matrix form to solve for the share of agriculture employment for both Hukou types:

$$\begin{pmatrix} d \ln \tilde{s}^A_{n,A,t} \\ d \ln \tilde{s}^A_{n,M,t} \end{pmatrix} = C^{-1} B + (\zeta - 1) C^{-1} \begin{pmatrix} \Delta^n_{A,t} \\ \Delta^n_{M,t} \end{pmatrix} \tag{1.59}$$

where matrix C is

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \tag{1.60}$$

$$C_{11} = (\zeta - 1)\left(\frac{(\theta^A\gamma_H^A + 1)}{1+\theta^A}\frac{\tilde{L}_{n,A}^A}{\tilde{L}_n^A} - \frac{(\eta^M - 1)\gamma_H^M}{\eta^M}\frac{\tilde{L}_{n,A}^M}{\tilde{L}_n^M}\frac{\tilde{s}_{n,A}^A}{\tilde{s}_{n,A}^M}\right) + \frac{1}{\tilde{s}_{n,A}^M}$$

$$C_{12} = (\zeta - 1)\left(\frac{(\theta^A\gamma_H^A + 1)}{1+\theta^A}\frac{\tilde{L}_{n,M}^A}{\tilde{L}_n^A} - \frac{(\eta^M - 1)\gamma_H^M}{\eta^M}\frac{\tilde{L}_{n,M}^M}{\tilde{L}_n^M}\frac{\tilde{s}_{n,M}^A}{\tilde{s}_{n,M}^M}\right)$$

$$C_{21} = (\zeta - 1)\left(\frac{(\theta^A\gamma_H^A + 1)}{1+\theta^A}\frac{\tilde{L}_{n,A}^A}{\tilde{L}_n^A} - \frac{(\eta^M - 1)\gamma_H^M}{\eta^M}\frac{\tilde{L}_{n,A}^M}{\tilde{L}_n^M}\frac{\tilde{s}_{n,A}^A}{\tilde{s}_{n,A}^M}\right)$$

$$C_{22} = (\zeta - 1)\left(\frac{(\theta^A\gamma_H^A + 1)}{1+\theta^A}\frac{\tilde{L}_{n,M}^A}{\tilde{L}_n^A} - \frac{(\eta^M - 1)\gamma_H^M}{\eta^M}\frac{\tilde{L}_{n,M}^M}{\tilde{L}_n^M}\frac{\tilde{s}_{n,M}^A}{\tilde{s}_{n,M}^M}\right) + \frac{1}{\tilde{s}_{n,M}^M}$$

Where B is defined as:

$$B = A + (\zeta - 1)\sum_h d\ln\tilde{L}_{n,h}\left(\frac{\tilde{L}_{n,h}^M}{\tilde{L}_n^M}\frac{\gamma_H^M(\eta^M - 1)}{\eta^M} - \frac{\tilde{L}_{n,h}^A}{\tilde{L}_n^A}\frac{(\theta^A\gamma_H^A + 1)}{1+\theta^A}\right)$$

and A is:

$$A = (\zeta - 1)\left[\frac{\theta^A}{1+\theta^A}d\ln A_{n,t}^A - \frac{(\eta^M - 1)}{\eta^M}d\ln A_{n,t}^M + \frac{1}{\eta^M}d\ln f_{e,t}^M + \frac{1}{1+\theta^A}d\ln\mathscr{M}_{n,t}^A - \frac{1}{\eta^M}d\ln\mathscr{M}_{n,t}^M\right]$$

$\Delta_{h,t}^n$ is defined as:

$$\Delta_{h,t}^n = (d\ln\delta_{M,h,t}^n - d\ln\delta_{A,h,t}^n)$$

Then the aggregate share of effective labor in each region is the weighted sum of shares for both Hukou types:

$$\tilde{s}_{n,t}^j = \frac{\tilde{L}_{n,t}^j}{\tilde{L}_{n,t}} = \Sigma_h\tilde{s}_{n,h,t}^j\frac{\tilde{L}_{n,h,t}}{\tilde{L}_{n,t}}$$

$$d\tilde{s}^j_{n,t} = \Sigma_h d\tilde{s}^j_{n,h,t}\frac{\tilde{L}_{n,h,t}}{\tilde{L}_{n,t}} + \tilde{s}^j_{n,h,t} d\frac{\tilde{L}_{n,h,t}}{\tilde{L}_{n,t}}$$

$$d\ln \tilde{s}^j_{n,t} = \Sigma_h \left(d\ln \tilde{s}^j_{n,h,t} + d\ln \frac{\tilde{L}_{n,h,t}}{\tilde{L}_{n,t}} \right) \frac{\tilde{s}^j_{n,h,t}\tilde{L}_{n,h,t}}{\tilde{L}_{n,t}\tilde{s}^j_{n,t}}$$

$$d\ln \tilde{s}^j_{n,t} = \Sigma_h \left(d\ln \tilde{s}^j_{n,h,t} + d\ln \frac{\tilde{L}_{n,h,t}}{\tilde{L}_{n,t}} \right) \frac{\tilde{L}^j_{n,h,t}}{\tilde{L}^j_{n,t}} \tag{1.61}$$

By plugging equation (1.59) into equation (1.61), we get:

$$\begin{aligned}
d\ln \tilde{s}^A_{n,t} &= \left(d\ln \tilde{s}^A_{n,A,t} + d\ln \frac{\tilde{L}_{n,A,t}}{\tilde{L}_{n,t}} \right) \frac{\tilde{L}^A_{n,A,t}}{\tilde{L}^A_{n,t}} + \left(d\ln \tilde{s}^A_{n,M,t} + d\ln \frac{\tilde{L}_{n,M,t}}{\tilde{L}_{n,t}} \right) \frac{\tilde{L}^A_{n,M,t}}{\tilde{L}^A_{n,t}} \\
&= d\ln \tilde{s}^A_{n,A,t}\frac{\tilde{L}^A_{n,A,t}}{\tilde{L}^A_{n,t}} + d\ln \tilde{s}^A_{n,M,t}\frac{\tilde{L}^A_{n,M,t}}{\tilde{L}^A_{n,t}} + d\ln \frac{\tilde{L}_{n,A,t}}{\tilde{L}_{n,t}}\frac{\tilde{L}^A_{n,A,t}}{\tilde{L}^A_{n,t}} + d\ln \frac{\tilde{L}_{n,M,t}}{\tilde{L}_{n,t}}\frac{\tilde{L}^A_{n,M,t}}{\tilde{L}^A_{n,t}} \\
&= \frac{1}{|C|}\left(\frac{B}{\tilde{s}^M_{n,M,t}} + (\zeta - 1)\Delta^n_{A,t} \right) \frac{\tilde{L}^A_{n,A,t}}{\tilde{L}^A_{n,t}} + \frac{1}{|C|}\left(\frac{B}{\tilde{s}^M_{n,A,t}} + (\zeta - 1)\Delta^n_{M,t} \right) \frac{\tilde{L}^A_{n,M,t}}{\tilde{L}^A_{n,t}} \\
&\quad + d\ln \frac{\tilde{L}_{n,A,t}}{\tilde{L}_{n,t}}\frac{\tilde{L}^A_{n,A,t}}{\tilde{L}^A_{n,t}} + d\ln \frac{\tilde{L}_{n,M,t}}{\tilde{L}_{n,t}}\frac{\tilde{L}^A_{n,M,t}}{\tilde{L}^A_{n,t}} \\
&= \frac{1}{|C|}\left(\frac{1}{\tilde{s}^M_{n,M,t}}\frac{\tilde{L}^A_{n,A,t}}{\tilde{L}^A_{n,t}} + \frac{1}{\tilde{s}^M_{n,A,t}}\frac{\tilde{L}^A_{n,M,t}}{\tilde{L}^A_{n,t}} \right) B \\
&\quad + \frac{\zeta - 1}{|C|}\left(\Delta^n_{A,t}\frac{\tilde{L}^A_{n,A,t}}{\tilde{L}^A_{n,t}} + \Delta^n_{M,t}\frac{\tilde{L}^A_{n,M,t}}{\tilde{L}^A_{n,t}} \right) \\
&\quad + d\ln \frac{\tilde{L}_{n,A,t}}{\tilde{L}_{n,t}}\frac{\tilde{L}^A_{n,A,t}}{\tilde{L}^A_{n,t}} + d\ln \frac{\tilde{L}_{n,M,t}}{\tilde{L}_{n,t}}\frac{\tilde{L}^A_{n,M,t}}{\tilde{L}^A_{n,t}}
\end{aligned}$$

$\Delta^n_{A,t}\frac{\tilde{L}^A_{n,A,t}}{\tilde{L}^A_{n,t}} + \Delta^n_{M,t}\frac{\tilde{L}^A_{n,M,t}}{\tilde{L}^A_{n,t}}$ can be further simplified as follows:

$$
\begin{aligned}
&\Delta^n_{A,t}\frac{\tilde{L}^A_{n,A,t}}{\tilde{L}^A_{n,t}}+\Delta^n_{M,t}\frac{\tilde{L}^A_{n,M,t}}{\tilde{L}^A_{n,t}}\\
&=[((\zeta-1)(\frac{(\theta^a\gamma^a_H+1)}{1+\theta^a}\frac{\tilde{L}^A_{n,M,t}}{\tilde{L}^A_{n,t}}-\frac{(\eta^M-1)\gamma^M_H}{\eta^M}\frac{\tilde{L}^M_{n,M,t}}{\tilde{L}^M_{n,t}}\frac{\tilde{s}^A_{n,M,t}}{s^M_{n,M,t}})+\frac{1}{\tilde{s}^M_{n,M,t}})d\ln\frac{\delta^n_{M,A,t}}{\delta^n_{A,A,t}}\\
&\quad-((\zeta-1)(\frac{(\theta^a\gamma^a_H+1)}{1+\theta^a}\frac{\tilde{L}^A_{n,M,t}}{\tilde{L}^A_{n,t}}-\frac{(\eta^M-1)\gamma^M_H}{\eta^M}\frac{\tilde{L}^M_{n,M,t}}{\tilde{L}^M_{n,t}}\frac{\tilde{s}^A_{n,M,t}}{s^M_{n,M,t}}))d\ln\frac{\delta^n_{M,M,t}}{\delta^n_{A,M,t}}]\times\frac{\tilde{L}^A_{n,A,t}}{\tilde{L}^A_{n,t}}\\
&\quad-[((\zeta-1)(\frac{(\theta^a\gamma^a_H+1)}{1+\theta^a}\frac{\tilde{L}^A_{n,A,t}}{\tilde{L}^A_{n,t}}-\frac{(\eta^M-1)\gamma^M_H}{\eta^M}\frac{\tilde{L}^M_{n,A,t}}{\tilde{L}^M_{n,t}}\frac{\tilde{s}^A_{n,A,t}}{\tilde{s}^M_{n,A,t}})+\frac{1}{\tilde{s}^M_{n,A,t}})d\ln\frac{\delta^n_{M,M,t}}{\delta^n_{A,M,t}}\\
&\quad-(\zeta-1)(\frac{(\theta^a\gamma^a_H+1)}{1+\theta^a}\frac{\tilde{L}^A_{n,A,t}}{\tilde{L}^A_{n,t}}-\frac{(\eta^M-1)\gamma^M_H}{\eta^M}\frac{\tilde{L}^M_{n,A,t}}{\tilde{L}^M_{n,t}}\frac{\tilde{s}^A_{n,A,t}}{\tilde{s}^M_{n,A,t}})d\ln\frac{\delta^n_{M,A,t}}{\delta^n_{A,A,t}}]\times\frac{\tilde{L}^A_{n,M,t}}{\tilde{L}^A_{n,t}}
\end{aligned}
$$

$$
\begin{aligned}
&=[(\zeta-1)\left(\frac{(\theta^a\gamma^a_H+1)}{1+\theta^a}\frac{\tilde{L}^A_{n,M,t}}{\tilde{L}^A_{n,t}}-\frac{(\eta^M-1)\gamma^M_H}{\eta^M}\frac{\tilde{L}^M_{n,M,t}}{\tilde{L}^M_{n,t}}\frac{\tilde{s}^A_{n,M,t}}{s^M_{n,M,t}}\right)d\left(\ln\delta^n_{A,M,t}+\ln\delta^n_{M,A,t}\right)\\
&\quad+\frac{1}{\tilde{s}^M_{n,M,t}}d\ln\delta^n_{M,A,t}]\frac{\tilde{L}^A_{n,A,t}}{\tilde{L}^A_{n,t}}\\
&\quad-[(\zeta-1)\left(\frac{(\theta^a\gamma^a_H+1)}{1+\theta^a}\frac{\tilde{L}^A_{n,A,t}}{\tilde{L}^A_{n,t}}-\frac{(\eta^M-1)\gamma^M_H}{\eta^M}\frac{\tilde{L}^M_{n,A,t}}{\tilde{L}^M_{n,t}}\frac{\tilde{s}^A_{n,A,t}}{\tilde{s}^M_{n,A,t}}\right)\\
&\quad d\left(\ln\delta^n_{M,A,t}+\ln\delta^n_{A,M,t}\right)+\frac{1}{\tilde{s}^M_{n,A,t}}d\ln\delta^n_{A,M,t}]\frac{\tilde{L}^A_{n,M,t}}{\tilde{L}^A_{n,t}}
\end{aligned}
$$

$$
\begin{aligned}
&=(\zeta-1)\left(\frac{(\theta^a\gamma^a_H+1)}{1+\theta^a}\frac{\tilde{L}^A_{n,M}\tilde{L}^A_{n,A}-\tilde{L}^A_{n,A}\tilde{L}^A_{n,M}}{\tilde{L}^A_n\tilde{L}^A_n}\right.\\
&\quad\left.-\frac{(\eta^M-1)\gamma^M_H}{\eta^M}\left(\frac{\tilde{L}^M_{n,M}}{\tilde{L}^M_n}\frac{\tilde{L}^A_{n,M}}{\tilde{L}^M_{n,M}}\frac{\tilde{L}^A_{n,A}}{\tilde{L}^A_n}-\frac{\tilde{L}^M_{n,A}}{\tilde{L}^M_n}\frac{\tilde{L}^A_{n,A}}{\tilde{L}^M_{n,A}}\frac{\tilde{L}^A_{n,M}}{\tilde{L}^A_n}\right)\right)d(\ln\delta_{A,M}+\ln\delta_{M,A})\\
&\quad+\frac{1}{\tilde{s}^M_{n,M}}\frac{\tilde{L}^A_{n,A}}{\tilde{L}^A_n}d\ln\delta_{M,A}-\frac{1}{\tilde{s}^M_{n,A}}\frac{\tilde{L}^A_{n,M}}{\tilde{L}^A_n}d\ln\delta_{A,M}\\
&\quad=\frac{1}{\tilde{s}^M_{n,M}}\frac{\tilde{L}^A_{n,A}}{\tilde{L}^A_n}d\ln\delta_{M,A}-\frac{1}{\tilde{s}^M_{n,A}}\frac{\tilde{L}^A_{n,M}}{\tilde{L}^A_n}d\ln\delta_{A,M}
\end{aligned}
$$

Plugging B, A, and Δ into the expression of $d\ln\tilde{s}^A_{n,t}$ gives the final equation in proposition 1, where I assume that only agriculture Hukou face a hukou cost:

Table 1.E.1: Robustness Test: Non-treated Provinces

	1982-1990			1982-2000	
	Fertility	log Emp	Agri Share	log Emp	Agri Share
$\text{post}_t \cdot S_r^{1970}$	−0.21	0.48	−0.23	0.15	−0.12
	(0.78)	(1.36)	(1.44)	(0.75)	(1.36)
year FE	Y	Y	Y	Y	Y
prefecture FE	Y	Y	Y	Y	Y

Note: Significance levels are denoted as a for 99%, b for 95%, and c for 90%. Standard errors clustered at prefecture level. Standard errors are shown in the bracket. Dependent variables are fertility rates, log total employment and agricultural employment share during 1982-1990 and 1982-2000 respectively. All regressions use population in 1970 as weights.

$$
\begin{aligned}
d\ln \tilde{s}^A_{n,t} = & \underbrace{[\frac{\theta^a}{1+\theta^a} d\ln(A^A_{n,t}) - \frac{(\eta^M-1)}{\eta^M} d\ln(A^M_{n,t})]}_{\text{comparative advantage}} + \underbrace{[\frac{1}{1+\theta^a} d\ln \mathscr{M}^A_{n,t} - \frac{1}{\eta^M} d\ln \mathscr{M}^M_{n,t}]}_{\text{difference in market access}} \\
& + \Sigma d\ln \tilde{L}_{n,t,h} \left[\left(\underbrace{\frac{\tilde{L}^M_{n,t,h}}{\tilde{L}^M_{n,t}} \frac{\gamma^M_H(\eta^M-1)}{\eta^M} - \frac{\tilde{L}^A_{n,t,h}}{\tilde{L}^A_{n,t}} \frac{(\theta^A\gamma^A_H+1)}{1+\theta^A}}_{\text{factor intensity effect}} \right) + \left(\underbrace{\frac{\tilde{L}^A_{n,t,h}}{\tilde{L}^A_{n,t}} - \frac{\tilde{L}_{n,t,h}}{\tilde{L}_{n,t}}}_{\text{hukou composition effect}} \right) \right] \\
& + \underbrace{\frac{1}{\eta^M} d\ln f^e_{n,t}}_{\text{entry cost}} + \frac{\zeta}{|C|} \frac{1}{\tilde{s}^M_{n,t,M}} \frac{\tilde{L}^A_{n,t,A}}{\tilde{L}^A_{n,t}} \underbrace{d\ln \delta_{M,t,A}}_{\text{hukou cost}}
\end{aligned}
$$

1.E Robustness Test

In this section, I present the detail regress results for the robustness tests.

Non-treated provinces: To validate the measurement of policy exposure, I examine provinces that did not experience the OCP relaxation. These provinces are excluded from the main analysis to ensure relevance. The results in Table 1.E.1 show that none of the coefficients are significant at the 5% level.

Migrants: I recompute the share of single-daughter families by excluding migrants and re-run the baseline regression. Table 1.E.2 shows the results. The coefficients are close to the baseline, indicating minimal impact from excluding migrants. Migrants were about 1% of the working-age population in 1982, so their effect on OCP exposure is limited.

Rural and urban: The different treatment of rural and urban Hukou under family planning policies could bias the results. To address this, I calculate policy exposure using rural

Hukou individuals for the baseline regression, as the OCP relaxation only applies to them. Urban Hukou individuals serve as a placebo test. Table 1.E.3 shows regression results using policy exposure from the share of single-daughter families for urban Hukou individuals. The coefficients are not significant at the 5% level.

Ethnic minorities: The one-child policy primarily applied to the Han ethnic majority, with different provisions for ethnic minorities. This could introduce bias in estimating policy effects. To address this, I conducted robustness tests by restricting the sample to the Han ethnic group, excluding ethnic minorities. Table 1.E.4 shows the results, which are close to the baseline regression results.

Table 1.E.2: Robustness Test: Excluding Migrants

	1982-1990			1982-2000	
	Fertility	log Emp	Agri Share	log Emp	Agri Share
$\text{post}_t \cdot S_r$	1.20^a	-0.55^b	2.20^c	0.35^c	-1.20^a
	(0.02)	(0.11)	(1.30)	(0.18)	(0.25)
year FE	Y	Y	Y	Y	Y
prefecture FE	Y	Y	Y	Y	Y

Note: Significance levels are denoted as a for 99%, b for 95%, and c for 90%. Standard errors clustered at the prefecture level. Standard errors are shown in brackets. Dependent variables are fertility rates, log total employment, and agricultural employment share during 1982-1990 and 1982-2000 respectively. All regressions use population in 1982 as weights.

Table 1.E.3: Robustness Test: Urban Hukou Individuals

	1982-1990			1982-2000	
	Fertility	log Emp	Agri Share	log Emp	Agri Share
$\text{post}_t \cdot S_r$	0.15	-0.10	0.20	0.05	-0.08
	(0.25)	(0.30)	(0.40)	(0.22)	(0.35)
year FE	Y	Y	Y	Y	Y
prefecture FE	Y	Y	Y	Y	Y

Note: Significance levels are denoted as a for 99%, b for 95%, and c for 90%. Standard errors clustered at the prefecture level. Standard errors are shown in brackets. Dependent variables are fertility rates, log total employment, and agricultural employment share during 1982-1990 and 1982-2000 respectively. All regressions use population in 1982 as weights.

Table 1.E.4: Robustness Test: Han Ethnic Group Only

	1982-1990			1982-2000	
	Fertility	log Emp	Agri Share	log Emp	Agri Share
$\text{post}_t \cdot S_r$	1.15^a	-0.50^b	2.10^c	0.32^c	-1.15^a
	(0.03)	(0.10)	(1.25)	(0.18)	(0.25)
year FE	Y	Y	Y	Y	Y
prefecture FE	Y	Y	Y	Y	Y

Note: Significance levels are denoted as a for 99%, b for 95%, and c for 90%. Standard errors clustered at the prefecture level. Standard errors are shown in brackets. Dependent variables are fertility rates, log total employment, and agricultural employment share during 1982-1990 and 1982-2000 respectively. All regressions use population in 1982 as weights.

Chapter 2

Incomplete Contracts in Commodities Trade: Evidence from LNG

2.1 Introduction

Long-term contracts are common in situations where buyers and sellers are locked into bilateral relationships having made transaction-specific durable investments (Klein et al., 1978; Masten and Crocker, 1985; Williamson, 1979). Contracts increase relationship surplus by reducing costly future contention about how to divide the quasi-rents. However, ex ante investments with long time horizons require contractual arrangements that are flexible enough to adapt to unpredictable changes to supply and demand without being impractical to design and implement.

The chapter examines the efficiency of long-term contracts in the global liquefied natural gas (LNG) industry that last up to twenty years. There are two main reasons to study this setting: First, it is a growing source of global energy, supplying around 10% of world demand, with bilateral contracts governing around three quarters of this supply.[1] Second, it provides a market where the value of customized intermediate inputs in a global supply chain can be directly estimated because weather conditions provide exogenous shocks to demand and data reveal the times that individual shipments arrive at their destinations.

Our research question is how well contracts achieve potential gains from trade by mitigating supply chain frictions, enabling supply to be timed to better match buyer demand. We also estimate the distribution of the gains between buyers and sellers, within contracts and in local spot markets.

We first describe the salient industry features that motivate our research question and

[1]Natural gas provides around 23% of the world's energy and nearly 40% of natural gas is supplied in its liquefied form (IEA, 2021).

guide our modeling choices. On the supply side, sellers of LNG have made large capital investments in liquefaction plants, called trains, and face very low marginal production costs. As a result, plants tend to produce at close to capacity and total output is unresponsive to spot prices. LNG is transported from sellers to buyers in large, specialized vessels that are often owned (or long-term leased) by one of the two parties. Buyers, who have also made large capacity investments in regasification terminals, and who typically operate with some excess capacity, are subject to demand shocks due to unexpected weather conditions. Contracts specify an annual quantity to be transacted between the buyer and seller at a price that is typically indexed to oil prices rather than destination-specific LNG spot prices. Zahur (2022) analyzes the timing of contract signing relative to capital investments and documents that contracts duration can be as long as the typical productive life of the initial investments.

This chapter focuses on how contracts operate from week to week, as buyers and sellers bilaterally negotiate the arrival times of individual contracted shipments throughout the year. Data on shipment times and weather conditions between 2009 and 2019 in the largest LNG destinations show that the quantity of contracted LNG arriving at a destination responds to local weather shocks. On the other hand, spot volumes are not related to weather shocks. In weeks when it is unexpectedly cold, relative to seasonal norms, buyers receive more LNG under contract. That is, sellers deliver LNG at times when it is particularly valuable to their contract partners, treating the buyers as "priority customers". It is notable that these times coincide with when the seller's outside option, selling on spot markets, is also more valuable because spot prices are also positively correlated with weather shocks.

Our analysis shows that delivered contract shipments are valued more than the shipments available to buyers on their local spot markets at the same time. Although LNG is itself a commodity, it is transported in large vessels that take time to travel the often long distances between sellers and buyers. The relatively small number of shipments in each spot market may not coincide with buyer demand at any point in time. We find that sellers exert relationship-specific effort within contractual transactions to deliver shipments in volumes and at times that match buyer demand. Because seller effort is costly, and sellers forgo favorable spot market outside options when they fulfil contract requests, sellers must anticipate extracting around one third of the overall estimated contract relationship value via the contract price in order for them to find contracts incentive compatible. We show that contracts contribute to total industry gains from trade. A counterfactual industry-level analysis that converts all contract shipments to spot is welfare-reducing.[2]

[2]Consistent with this conclusion, contractual relationships have proved to be very resilient during the energy

The shipment-level data used in this chapter were made available to us by the company Kpler, a provider of data and analytic solutions for commodity markets. The data include information on the origin and destination of each LNG shipment, the volumes at origin and destination, the nature of the vessel, and the voyage duration. The data on LNG spot prices was downloaded from EIKON and destination-specific weather data came from the World Bank Climate Knowledge Portal.

We present three sets of empirical analyses. The first is a reduced-form investigation of variation in volumes at the route-week level. Larger quantities flow to destinations when it is unseasonably cold[3]. We show that it is contracted volumes that respond to weather shocks and that spot volumes do not adjust. These findings are evidence that sellers try to send contract shipments to their relationship partners at times when buyers have the highest value for those shipments.

The second analysis explores the hypothesis that buyers value the contract shipments they receive over available spot shipments, given observed spot prices and quantities. The logic underlying this hypothesis is that if there were no such premium, there would be no contract relationship surplus that could be shared with the seller, and no economic rationale to explain why sellers are consistently willing to send contract shipments rather than sell on the spot market exactly when the spot prices are high. Using buyers' revealed preference for contract shipments, we estimate a lower bound on the typical contract shipment value premium that is shared between the buyer and the seller. Intuitively, this premium must be sufficient to compensate sellers for not deviating from the contract when spot prices are high and compensate buyers for not deviating from the contract when spot prices are low. The estimated mean of the relative value for the contract shipments studied is \$27.07 per cubic metre of LNG, which is around 15% of the mean spot price in the sample used here.

This second analysis is limited in that it produces lower bounds of relative contract value compensating sellers only for foregoing spot markets when contract shipments are requested. It does not allow for sellers to be exerting any relationship-specific effort to increase buyer value. To ask the more fundamental question of how value is created and shared along the value chain, and how much value contracts create compared to the counterfactual

crisis in 2022 arising because of the war in Ukraine. Sellers continued to deliver contract shipments despite all-time high spot prices. Two notable exceptions were Eni and Gunvor in early 2022, who did not deliver anticipated contract shipments to Pakistan S&P (2022). While the total share of spot volumes had been increasing up to 2020, this trend has recently reversed, "LNG term contract volumes have leapt this year, as energy security becomes paramount worldwide" S&P (2023).

[3]By "unseasonably cold", we mean temperatures that are lower than the normal temperature for that specific month when compared across multiple years.

of no contracts at all, we turn to the third analysis in the chapter, which constructs and estimates a structural model of the LNG industry.

In the model, we allow buyer value for LNG to be a function of local weather. For contract shipments, buyer value is also a function of relationship-specific seller effort. The amount of effort the seller exerts depends on how much the effort contributes to seller value, effort costs, and the share of contract surplus the seller extracts. In each period, buyers observe weather conditions and choose whether to request a contract shipment or to buy on the spot market. The spot value depends on an idiosyncratic shock drawn from a distribution with an expected value that is an increasing function of local spot market thickness in that period.[4] The model delivers a set of theoretical moments, derived from incentive compatibility constraints, that give the probability a contract buyer requests a contract shipment in a given week as a function of the weather and model parameters.

Because the model focuses on route-week-level buyer decisions, it does not explicitly take into account seller capacity constraints or any trade-offs the seller has to make between demand from multiple contract partners in any one period. We do two things to accommodate these seller-level constraints. First, we include a fixed cost of fulfilling contract shipments that is a function of how many contract requests a seller fulfills that month relative to their typical number. Second, we estimate the model parameters on a subset of the data: route-weeks where the seller sells at least some positive quantities on the spot markets. These are quantities that sellers could have otherwise allocated to contract partners without being capacity constrained.

We estimate the model using Nonlinear Least Squares. Parameter identification is made possible by independent variation in the relationships between contract shipment timing and spot prices, holding weather constant, and between contract shipment timing and weather shocks, holding spot prices constant. Using the model estimates to predict the untargeted moments of the elasticity of contract shipments with respect to weather shocks and spot prices gives values that are close to those in the data.

The results include distributions of estimates of the willingness to pay for the LNG in contract and spot shipments on each route-week in the sample. A cubic metre via a contract shipment is more than twice as valuable, on average, given the current size of the spot markets.[5] The lower bound on the share of contract shipment surplus that sellers

[4] Spot market thickness is measured as the relative size of the spot market in the region and how well the distribution of vessel sizes in the spot market fits the distribution of local contract demand.

[5] We note that a part of this difference arises because buyers tend to request contract shipments during weeks when they experience positive demand shocks.

must extract via the contract price in order for contractual relationships to be incentive compatible is estimated to be 29%. We also find that, for around one half of contract shipments, sellers exert relationship-specific effort to deliver contract shipments that buyers value more highly, for example to arrive at the ideal time during a week or in desirable quantities.

We then turn to a counterfactual analysis that asks how much long-term contracts contribute to industry value compared to a world where all LNG is traded on local spot markets. We are able to do this by holding fixed total global supply and sellers' allocations of shipments to destinations in each week. We convert all arriving contract shipments to spot and add them to the spot shipments already arriving at the destination in a given week, giving us a counterfactual spot market in each destination. Because we have estimated the relationship between local spot market thickness and expected spot value, we can simulate the expected values buyers would have obtained in these enlarged counterfactual spot markets.[6]

We find that the efficiency of the LNG industry would have been significantly reduced in the absence of contracts between 2009 and 2019. The gains from contracts outweighed the fixed and variable costs of maintaining them, contributing around 32% of the total gains from trade in the industry over this period.

In all the analyses in this chapter, we take the existing contract network as fixed. There is some reduced form evidence that sellers tend to have contracts with buyers in destinations with more unpredictable weather, but we do not model why some routes have contracts and others do not. We have also not modeled transport costs explicitly.[7] One other important issue our analysis does not consider is seller market power. In theory, it could be that sellers choose to fulfil contract shipment requests in order to withhold quantities from spot markets and gain from the resulting higher spot prices. We do not allow sellers' supply decisions to include these strategic considerations. The main reason for this choice is that the routes between sellers and buyers tend to either be contract or spot. That is, if a seller has a contract with a buyer, it is unlikely that they are also selling on the spot market at that destination, and vice versa.[8]

[6] This approach allows us to sidestep the challenge of estimating counterfactual spot prices because we focus on the total surplus created by each shipment as a function of weather conditions, rather than how the unobserved counterfactual spot price divides this surplus. Setting up the counterfactual in this way does not take into account any allocative efficiency gains possible in the spot market if sellers were to divert spot shipments to locations where they anticipate higher prices.

[7] The counterfactual estimation holds the distances traveled by each vessel fixed and, as such, transport costs play no role in the welfare calculations.

[8] While there are several large firms in this industry that have minority ownership stakes in multiple sellers, the majority ownership stake is often a state energy company with no cross-seller links.

This chapter relates to several literature. First, industry-specific studies of the pros and cons of long-term quantity contracts. These include Joskow (1988), Pirrong (1993), and the discussion in Lafontaine and Slade (2012). Zahur (2022) has studied the LNG industry in depth and has focused on the long-term contracts that we investigate here. He shows that contracts allow greater capital investments at the cost of allocative efficiency but does not examine whether contractual relationships create value through adaptation to shocks. Our analysis is motivated by Zahur's observation that contracts are often signed around the time of the capital investments, which, due to the long life of capital, raises the issue we explore here of the efficiency implications of contract flexibility in delivery timing.[9]

In our setting, unpredictable weather causes exogenous demand shocks that we exploit to identify how supply chains respond. We establish that long-term contracts are sufficiently flexible to allow the supply side to adapt and create overall industry value. This finding complements work showing that vertical integration can permit adaptation to unforeseen circumstances. Forbes and Lederman (2009) and (2010) study vertical integration in the airline industry. They show that major airlines tend to own regional airline partners when adverse weather is more likely and where a route is more integrated into the major network. Integrated routes also respond to weather shocks and perform better than non-integrated routes at the same airport on the same day. Costinot et al. (2011) show that the boundaries of U.S. multinational firms reflect the extent to which sectors involve non-routine tasks. Vertical integration along the supply chain is more likely in sectors with high problem-solving intensity, consistent with ownership facilitating adaptation to changing circumstances.

We contribute to the literature and differ from previous papers in two ways. First, we focus on global integration, where transaction frictions play important roles. Second, we examine interactions through long-term contracts, rather than through ownership or vertical integration as emphasized in the literature. We find that long-term contracts enable adaptation in global LNG supply chains in terms of delivery volumes and timing. We highlight that this finding implies contracts increase the resilience of the industry supply chain to demand shocks.[10]

A recent literature on relational contracts establishes how the prospect of future rents effectively deters short term opportunism in a range of empirical settings (Macchiavello, 2022).

[9] Because our counterfactual does not vary each shipment's destination, we do not explore allocative efficiency.

[10] Recent global events, perhaps most notably the Covid-19 pandemic, have prompted much political debate and study of the efficiency implications of potential supply-side disruptions within global value chains (see Grossman et al., (2023); Grossman et al., (2023), and the papers cited therein).

Macchiavello and Morjaria (2015) use the dynamic incentive compatibility constraints implied by repeated seller-buyer interactions to estimate lower bounds on the value of relationships in the Kenyan flower industry, a setting of imperfect contract enforcement. Strikingly, we show that the future value of formal, and yet incomplete, contracts in the global LNG industry governs the short term actions of the sellers in the industry. Buyers, however, compare contemporaneous costs and benefits, interpreting contracts as call options to request shipments at times when they are particularly valuable.

One recent industry-specific study that examines the externalities that contracts exert on spot markets is Harris and Nguyen (2022).[11] They estimate the incentives that govern both relationship formation and the actions of buyers and sellers in U.S. trucking. Relationships form between trading partners with a high match quality, but the large number of contract relationships leads to thin spot markets, increasing the search costs for trading partners. Tolvanen et al. (2021) also find that frictions in spot markets provide an insurance role for buyers, in analysis of transactions data from a large producer in the U.S. pulp industry. In our setting, we show that relationship surplus is an endogenous outcome reflecting demand shocks and seller incentives to exert costly relationship-specific effort to increase buyer value. Even though these costs are saved in the counterfactual that eliminates contracts, the increased thickness of the spot market is insufficient to compensate for the lost returns of seller effort.

Our interpretation of the relatively low buyer values for spot market shipments is that spot markets do not supply the quantities of LNG that buyers need at the ideal delivery times, even in the counterfactual thicker spot markets, due to frictions related to bulky shipping technology and the long distances between buyers and sellers. Brancaccio et al. (2020) document how frictions arise in global shipping when country-level trade imbalances in dry bulk commodities lead to shipping price differences across markets. We note that LNG vessels are specialized and there is no two-way trade. In addition, vessels are very large relative to total weekly demand, an indivisibility that amplifies the distance-related time lag between requesting and receiving a shipment. We infer that contracts were efficient in this industry because the total market size was too small, over the time period studied, for spot markets to overcome these geographical and technological frictions.

This chapter proceeds as follows: Section 2.2 describes the industry and summarizes the data. Section 2.3 is the reduced-form analysis relating contract volumes to weather conditions. Section 2.4 estimates lower bounds on buyers' values for contract shipments given

[11] Kranton (1996) sets out theoretically how spot market externalities can be large enough or small enough to determine whether relationships are efficient.

prevailing spot prices. Section 2.5 presents the structural model, Section 2.6 discusses estimation, and Section 2.7 presents the results. Section 2.8 is the counterfactual analysis. Section 2.9 concludes.

2.2 Industry and Data

2.2.1 Industry Overview

The global LNG industry allows the gas produced in dispersed, oil-rich, locations to be liquefied by cooling it to -160° centigrade and transported via ships to locations with regasification capacity. The industry is characterised by large, long-lived, location-specific capital investments at both origins and destinations typically located far apart from each other, and connected via a costly shipping technology. Demand is concentrated in East Asian countries, although regas terminals continue to be built at ports worldwide.[12]

Long-term sales and purchase agreements (SPAs) are contracts between sellers and buyers that govern the bulk of delivered volumes and that are signed before or after capital investments are made (Zahur, 2022). Contracts specify annual volumes and a pricing formula. Much of the detail in the contracts relates to their "take or pay" nature, a feature shared with pipeline gas contracts (Masten and Crocker, 1985). Prices are often indexed to local oil prices, which means that the seller bears risk related to the forgone opportunity of selling at high LNG spot prices when they occur. The buyer typically bears the quantity risk, in that there is an obligation to take the total quantity over the course of the year, whether they need it or not, or incur demurrage charges to compensate the seller. Contracts also contain various clauses related to unforeseen contingencies and the length of acceptable delivery windows.

The SPAs, however, do not specify the exact timing of individual contract shipments during the year. While parties typically operate under a three-month forward-looking schedule, shipments are governed by separate Confirmation Notices that describe timing, quantity, and other shipment-specific details. Confirmation Notices are agreed bilaterally prior to shipping. Therefore, the governing contracts are incomplete in that sellers and buyers must arrange exactly when and in what quantities each shipment is delivered (or collected, when the buyer is responsible for the shipping). Because buyers are subject to varying weather conditions that impact shipment value, and shipment delivery dates are observable, this

[12] Recent disruptions to pipeline gas in the EU due to the Russia-Ukraine war have led to many new regasification investments. The import capacity in the EU and the UK is set to increase by one third by 2024, after expanding only modestly in the 10 years to 2022. https://www.eia.gov/todayinenergy/detail.php?id=54780.

industry offers a setting where we can observe if contractual partners are able to arrange transactions at valuable times.[13]

Residual LNG volumes are traded on global spot markets. There is significant variation in the destination-specific local spot price over time, and also across countries at any one point in time. Figure 2.1 shows the price data we use in this chapter.

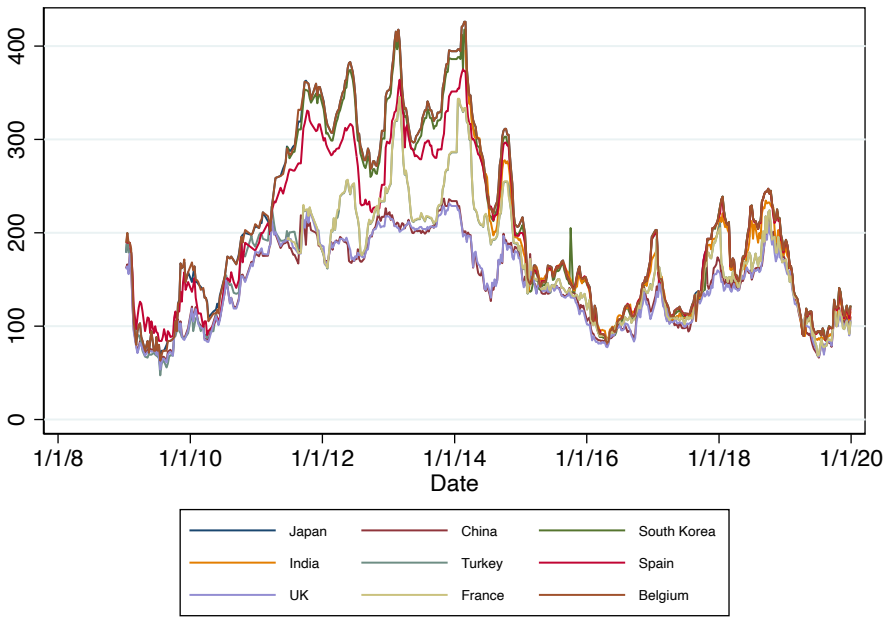


Figure 2.1: LNG Spot Prices by Destination Country over Time, USD

2.2.2 Global Supply in 2019

The shipment-level data we use, which runs from the start of 2009 to the end of 2019, contains shipments from 19 countries. Qatar has been the largest exporter historically, with over 500 shipments in each of the 11 years. Australia has increased exports over the time period, and exported more than 500 shipments in the last four years in the data. After Qatar and Australia, the largest exporter countries are Malaysia, Indonesia, Algeria, and Nigeria. Figure 2.2 focuses on 2019, and shows total volumes exported by source country

[13] Although contractual agreements are proprietary, template SPAs are publicly available. For example, the company BP publishes its master sales and purchase agreement template on its website, https://www.bp.com/content/dam/bp/business-sites/en/global/bp-trading-and-shipping/documents/bp-master-ex-ship-lng-sale-and-purchase-agreement-2019-edition%20(3)%202.pdf. A Confirmation Notice template for a specific shipment is given in the Appendix, Schedule 1, on pages 46 to 49.

in cubic metres. The USA, Russia, and Trinidad and Tobago are more prominent in 2019 than in the period as a whole.

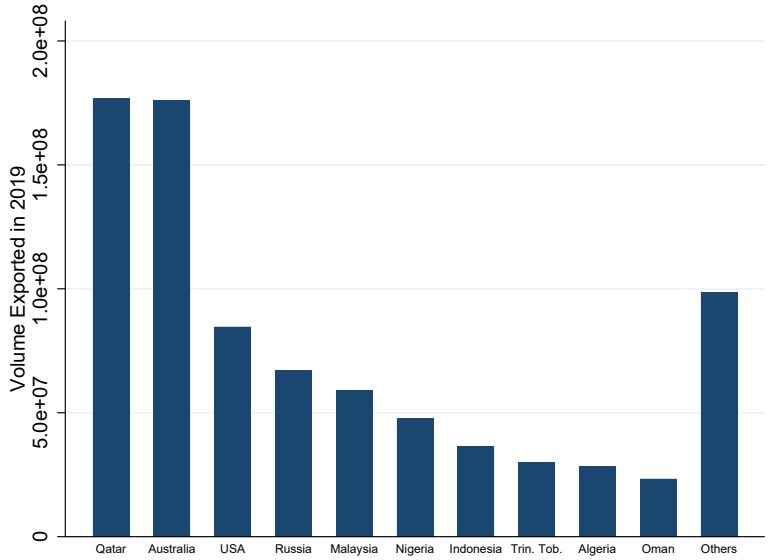


Figure 2.2: Largest Exporter Countries in 2019

Because the LNG industry involves costly shipping technology for a commodity product, both sellers and buyers have incentives to minimize the distance travelled. The overall flows confirm this, in general. For example, Figure 2.3 shows that the two largest destinations for Australia are the relatively close Japan and China, where Australia supplies 39% and 47% of these countries' LNG, respectively. Qatar supplies less of these countries' demand, at 11% and 13%. The figures below the country name show the average mileage travelled on routes between the source and destination countries. The LNG travelling to Japan and China from Australia travels much shorter distance than the LNG from Qatar.

However, there are exceptions. For example, Qatar supplies 27% of South Korea's LNG and 29% of Taiwan's despite LNG having to travel over 5,000 miles to get there. Australia is much closer to both destinations and supplies only 19% and 27%, respectively.[14] We suspect deviations from gravity relate to the staggered timing at which liquefaction and regasification capacity was installed in different countries and the agreed long term contracts between buyers and sellers.

Figure 2.3 also offers some insight about how shipments are organized. The bulk of LNG

[14]We have some data on production costs. Qatar has a much lower operating production cost than the installations in Australia (see supply functions in earlier update), so the delivered cost from Qatar may still be much lower than from Australia.

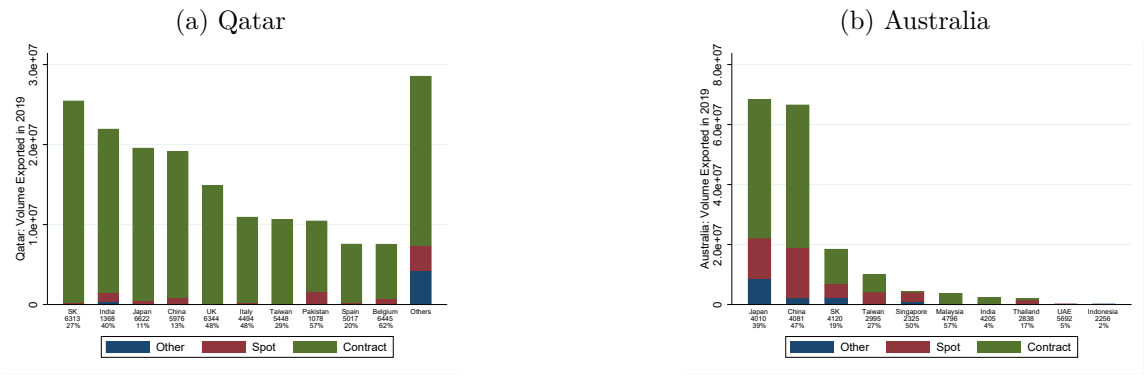


Figure 2.3: Volume by Destination, 2019

exports from Qatar in 2019 were contracted quantities, meaning very little LNG from Qatar ended up in the spot market. For Australia, spot volumes were relatively more common.

Figure 2.4 focuses on 2019, and shows total volumes imported by destination country.

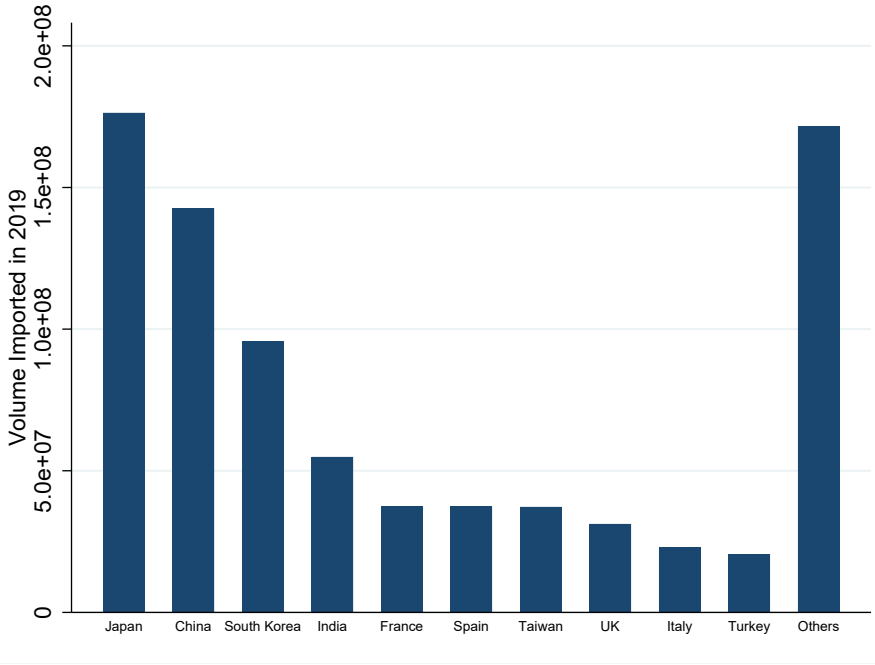


Figure 2.4: Largest Importer Countries in 2019

Figure 2.5 shows the volumes arriving in Japan in 2019 by source country. The figures below the country names are the average distance travelled and the share of that country's LNG exports arriving in Japan. There are some interesting comparisons. Australia sends

Japan 39% of its output but is much further away than Malaysia, which sends a similar share of its total output to Japan.

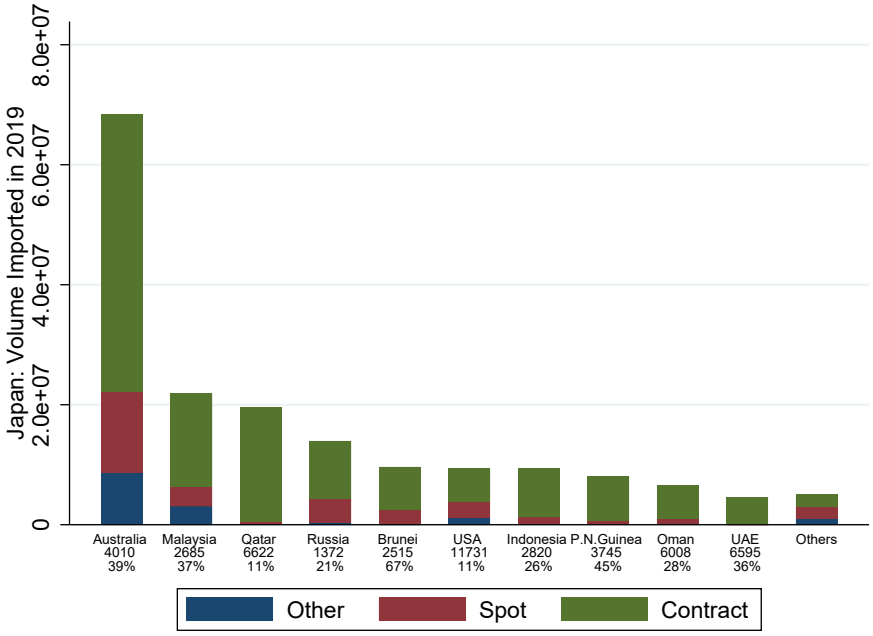


Figure 2.5: Japan Imported Volumes by Origin, 2019

2.2.3 Data

We have shipment-level data on 44,848 shipments from January 2009 to December 2019 that were either contract or spot market transactions. 13 countries imported more than 500 shipments in total over the 11 years in the data. Japan has been, by far, the largest importer historically, receiving over 15,000 shipments in total during the 11 years. South Korea had the second highest number of shipments, at nearly 6,000. China has steadily increased shipment imports, from around 100 in 2009 to nearly 1,000 in 2019. For each shipment, our data contain distance traveled, time taken, vessel capacity and type, as well as volume at origin and volume at destination, which permits a boil off calculation.[15]

We have 36 origins in the data that we currently use. Of these, 17 are countries where liquefaction is under control of one (often state-owned) company, and integrated oil companies often have minority stakes in these state-owned oil companies. The countries are: Algeria, Angola, Brunei, Cameroon, Egypt, Equatorial Guinea, Indonesia, Malaysia, Ni-

[15]The volume arriving at the destination is always lower than the volume that left the origin. A note on conversion factors. To convert mmBtu into cubic metres of LNG: $1mmBtu = 0.048$ cubic metres of LNG. Equivalently, $1m^3 = 21.04mmBtu$. Source: http://www.lngplants.com/conversiontables.html

geria, Norway, Oman, Papua New Guinea, Peru, Qatar, Trinidad and Tobago, United Arab Emirates, and Yemen. The other 19 origins are separate liquefaction facilities in Australia (10), Russia (3), and the USA (6). We treat these plants separately because they have different owners and/or geographies: The Russian plant Sakhalin is in the far East, very near Japan, and Yamal and Vysotsk are in the West of Russia. The Australian plants are also located on different sides of the country, as well as being separate companies, albeit with some overlapping ownership. Similarly, the US plants are distributed across various states, each with distinct ownership and geographical considerations.

We have 39 destinations, which are the regasification terminals aggregated to the country level. The largest individual buyers are often state or city energy providers. For example, the largest single buyer in the data is the only South Korean buyer, KOGAS, a public natural gas company. The next largest is CPC Corporation, the sole LNG importer in Taiwan, which is a state-owned company. CNOOC, the state owned China National Offshore Oil Corporation, is the largest importer in China. In Japan, the three largest importers are the Tokyo Electric Power Company (TEPCO), which is majority-owned by the government of Japan, and its joint venture JERA, as well as the Tokyo Gas Company.

We hence have $1,404$ possible origin-destination pair routes in each of 11 years, but, since some facilities came online during the sample, we have $8,570$ route-years pairs where the plant and terminal are both active during the year. Of all feasible route-years, only $2,178$ contain any positive flows.

In the empirical analysis, we restrict attention to the 10 destinations with large LNG spot markets, where weekly spot price data is available. These are the destinations shown in Figure 2.1.[16] These 10 destinations account for $31,267$, or 70% of all the contract and spot shipments over the time period. The shipment-level data allow us to see the total volume at origin and at destination traded on a given route in a given week by summing up the shipment volumes. Because we can see whether any given shipment is contract or spot, we can also compute the total contract and spot volumes.

We use data from several other data sources. The location-specific LNG spot prices shown in Figure 2.1 are from EIKON. These are weekly data for the 10 different spot markets we

[16]The countries are Belgium, China, France, India, Japan, South Korea, Spain, Taiwan, Turkey, and the United Kingdom. There is a spot price also in the US, but we have excluded the US as a destination as it is predominantly an exporter of LNG. We also exclude the LNG spot markets in Argentina, Brazil, and Mexico. This is because they are relatively small markets and also because demand appears not to vary with weather. Also note that we assume buyers in Taiwan can purchase on the China spot market due to its proximity.

study. Country-specific temperatures are from the World Bank climate knowledge portal.[17] We select the mean monthly temperature in each country for each year.[18]

In this chapter, we are interested in how well contracts allow parties to adapt to unforeseen circumstances, in our case, unexpected weather shocks. We follow the climatology literature (see, for example, Nicholson (1993)) and compute weather anomalies as the deviations from the destination's long-term mean, divided by its long-run standard deviation. To adjust for seasonality, we calculate this measure of the shock for each destination-month. That is, the temperature shock is the actual mean temperature in a destination that month minus the mean temperature for that destination-month divided by its standard deviation. Since each observation is standardized, the average value of this variable is zero and the standard deviation is one. The variation in weather anomalies is shown in Figure 2.6.

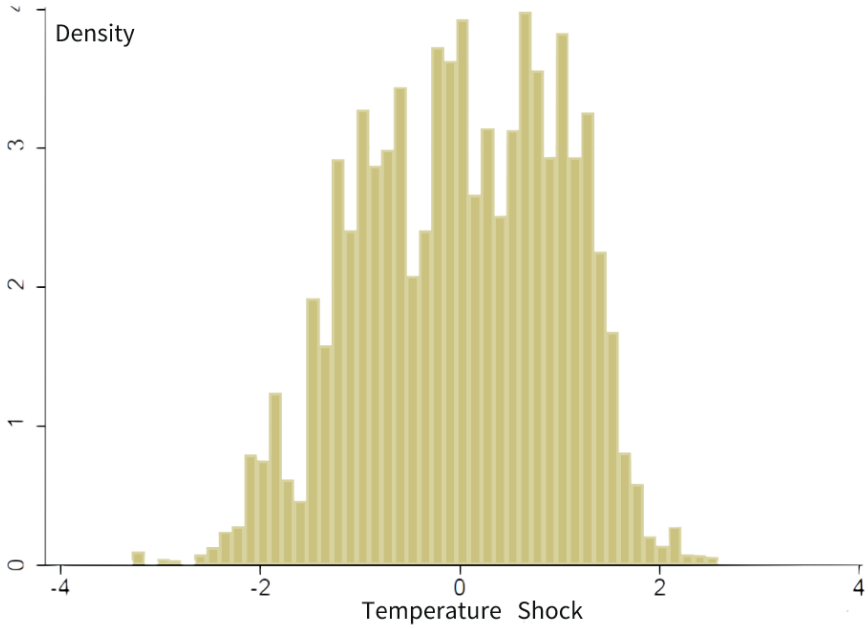


Figure 2.6: Destination-month Temperature Shocks, Standardized

2.2.4 Descriptive Statistics

First, we examine which 2,178 of the 8,570 potentially active route-years actually have some positive flows. Table 2.1 examines whether flows follow gravity. It contains three regressions of the indicator for a route-year being active on the log of the direct distance (in ′000 km) between the origin and destination and variables relating to the total volume

[17] https://climateknowledgeportal.worldbank.org/download-data

[18] All the results are robust to using the minimum monthly temperature in each year to measure weather shocks.

produced (received) by the origin (destination). Column 1 shows that an increase in route distance of $1,000$ km lowers the probability of the route-year being active by 20% (the coefficient is 0.204).[19] Column 2 includes the total volume of the origin and destination as controls, and column 3 includes origin and destination fixed effects. In this final column, an increase in distance of 1000 km reduces the probability of being active by 16%. In other words, distance matters in that having any positive flows follows gravity.

Table 2.1: Probability Route Active in a Year

	(1)	(2)	(3)
Log of Airline Distance (km)	-0.204***	-0.187***	-0.164***
	(0.007)	(0.006)	(0.006)
Total Volume by Origin (Year)		0.003***	
		(0.000)	
Total Volume by Destination (Year)		0.005***	
		(0.000)	
Observations	8,570	8,570	8,570
R^2	0.103	0.305	0.525

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Turning to the set of route-years with positive flows, the volumes on a given route in the year tend to be either contract or spot volumes. Figure 2.7 shows the bi-modal share of volumes on a route-year that is under contract.

Based on this figure, we construct an indicator variable for whether an active route-year is mostly under contract, defined as having at least 50% of total flows in the year being contract flows and investigate selection into being a contract route.

In Table 2.2, the binary dependent variable is equal to one if the volumes on a route-year are more than 50% contract and all columns include year fixed effects. Column 1 shows there is no evidence that distance matters, in that shorter routes are no more or less likely to be contract routes in any year. Instead, the total volumes of the buyer and seller are positively associated with being a contract route. Column 2 shows that route-years originating at larger sellers and ending at larger buyers are significantly more likely to be contract route-years. Column 3 includes buyer fixed effects and shows that the typical

[19] For instance, Engel and Rogers (1996) estimates that the US-Canada border plays an important role in terms of explaining price dispersion by adding approximately 11.9×10^{-3} to the standard deviation of prices, equivalent to a distance of about 75,000 miles. Even after accounting for nominal price stickiness, the border explains around 18.9% to 33.3% of the price dispersion.

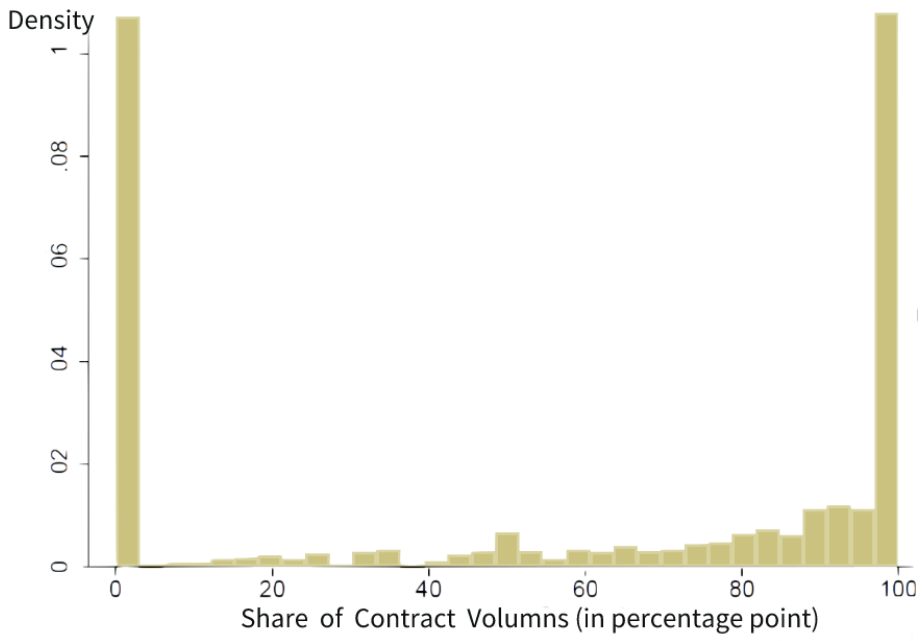


Figure 2.7: Share of Route-year Volumes that is under Contract

buyer is more likely to have contract routes with its larger sellers. Column 4 includes seller fixed effects and shows that the typical seller is more likely to have contracts with its larger buyers.

Column 5 of Table 2.2 presents some evidence that routes select into having contracts when the weather at the destination is more unpredictable. It controls for year and seller fixed effects and the dependent variable is the mean absolute value of the monthly weather abnormality at the destination during that year, where weather abnormality is the standardized temperature shock shown in Figure 2.6. The mean of this measure is 0.663 and the standard deviation is 0.598. The coefficient of 0.299 suggests that a one standard deviation decrease in route-year weather predictability (an increase in unpredictability) is associated with an 18% increase in the probability that a given route-year is mostly contract. While our structural model will take the existence of a contract on a route as given, the evidence in column 5 is consistent with there being a contract in place when the potential gains from adaptation to weather shocks are largest.

The empirical analysis in this chapter focuses on the volumes and number of shipments at the route level in a given week. In the $2,178$ route-years with some positive flows, there are $110,745$ route-week observations. We match $52,201$ of these route-week observations to a destination-level LNG spot price because they are in one of the 10 locations we consider.

Table 2.2: Probability Active Route-year is Contract.

	(1)	(2)	(3)	(4)	(5)
Distance, ln(km)	0.017	0.006	0.005	-0.014	
	(0.015)	(0.015)	(0.016)	(0.015)	
Seller volume, year		0.001***	0.001**		
		(0.000)	(0.000)		
Buyer volume, year		0.001***		0.002***	
		(0.000)		(0.000)	
Buyer-year weather shock					0.299***
					(0.044)
Constant	0.566***	0.528***	0.566***	0.553***	0.432***
	(0.027)	(0.029)	(0.031)	(0.026)	(0.031)
Observations	2,171	2,171	2,169	2,171	1,449
R^2	0.011	0.021	0.130	0.273	0.289
Fixed effects	Year	Year	Year, buyer	Year, seller	Year, seller

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Of these, only 17,516 route-weeks have positive volumes. That is, the route-weeks with zero shipments are important in the data.

Table 2.3 summarizes the data used in Section 2.3, which are volumes of LNG at the route-week level between 2009 and 2019. The first row of column 2 shows that the mean volume per route-week is 82.75 thousand cubic meters. Conditional on there being some positive volumes, the mean volume is 246.61 thousand cubic meters, and the mean number of shipments on a route-week is 1.79. The mean vessel capacity in the shipment-level data is 130 thousand cubic meters.

Table 2.3: Active Route-weeks, 2009 to 2019

		Volumes, '000 cubic metres			
	Obs	Mean	Std.dev	Min	Max
All	52,201	82.75	152.25	0	1904.41
Contract	34,144	110.98	173.81	0	1904.41
Non-contract	18,057	29.37	74.95	0	722.51
Local spot price, USD per m^3	52,201	182.40	82.19	47.34	426.06
Local mean temp, °C	52,201	12.77	9.24	-8.99	31.25
Local temp shock, °C	52,201	0.00	1.00	-3.29	2.59

The second row of Table 2.3 shows that 65% of route-week observations are on routes where at least 50% of volumes that year are under contract. Hence, 35% of routes are not contract routes, and the majority of volumes on these routes is spot.

The final three rows of Table 2.3 show that the mean spot price for the route-weeks in the data is \$182.40 per cubic metre, which corresponds to \$8.67 per mmbtu. The mean monthly temperature across the 10 destinations is 12.77 degrees centigrade, and the standard deviation is large. The final row summarizes the weather shock used in the analysis (see also Figure 2.6).

2.3 Reduced Form Analysis

This section asks if route volumes respond to weather shocks at the destination, and if the responsiveness varies for contract and spot volumes. We consider only active routes, defined as those where plants at both the origin and destination are operational and there is at least one shipment on that origin-destination route during the calendar year. The weather shock variable is as defined in Section 2.2.4, as the mean monthly temperature at the destination standardized by subtracting the mean of that destination-month's mean temperature and dividing by its standard error all years in the data, and shown in Figure 2.6.

In the first set of specifications, we regress total volumes on a given route-week (gt), in $'000$ cubic meters, on the weather shock that month, as follows:

$$Q_{lgt} = \alpha w_{gt} * O_{lg} + \Gamma_{gm} + \epsilon_{lgt}, \tag{2.1}$$

where Q_{lgt} is the total volume on active route lg in week t, w_{gt} the weather shock that month, and Γ_{gm} are destination-month fixed effects. The variable O_{lg}, included only in the second specification, is an indicator variables for how the route is organized, equal to 1 for contract routes.

The results of estimating equation (2.1) are given in Table 2.4. Column 1 shows that the total volume is higher by 2.820 thousand cubic meters when the weather shock is lower by one standard deviation. This corresponds to a 3.4% increase in total volume relative to the mean. Column 2 includes the interaction of temperature with the indicator for whether the route is a contract route. The coefficient on non-contract routes becomes a small positive and significant number; that is, volumes actually decrease on non-contract routes when there is an adverse weather shock (a positive demand shock). For contract routes, the sum

of the relevant coefficients shows that a one standard deviation negative weather shock is associated with a 4.11 thousand cubic metre increase in volumes on the typical contract route. This is a 3.7% increase in total volume relative to mean contract route volumes.

Table 2.4: Route-week Volume Responsiveness to Local Temperature Shock.

	(1) Total Volumes	(2) Total Volumes
Weather shock	-2.820***	1.555*
	(0.725)	(0.805)
Contract route * Weather shock		-5.666***
		(1.338)
Contract route		78.826***
		(4.648)
Constant	82.750***	31.144***
	(0.000)	(3.044)
Observations	52,201	52,201
R^2	0.094	0.152
Fixed effects	destination-month	destination-month

Standard errors in parentheses, clustered at buyer-month level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The second set of reduced form specifications changes the dependent variable to total volumes, contract volumes, or spot volumes on a route-week, controlling for the average quantity with route-month fixed effects. The estimated equation is:

$$Q_{lgt} = f(w_{gt})'\alpha + \Gamma_{lgm} + \epsilon_{lgt}, \tag{2.2}$$

where Q_{lgt} are the volumes in question, and Γ_{lgm} are the route-month fixed effects, rather than the destination-month fixed effects included in Table 2.4. The results are given in Table 2.5. In columns 1, 3, and 5, $f(w_{gt}) = w_{gt}$ and α is a scalar, investigating whether there is a linear relationship between the weather shock and volumes. In columns 2, 4, and 6, $f(w_{gt}) = w_{gt} + w_{gt}^2$ and α is a vector of two coefficients, allowing for a quadratic relationship.

Because there are route-month fixed effects in these specifications, the coefficients reflect variation in volumes delivered in the same month across years on a given route. In columns 1 and 2, the dependent variable is the total volumes on a route-week. The negative significant coefficient of -1.319 tells us that volumes are around 1.6% lower when the temperature is one degree higher. Column 2 shows that the relationship is non-linear, in that respons-

iveness is increasing in the weather shock. Columns 3 and 4 use the contract volumes as the dependent variable, and the coefficients suggest a negative linear relationship between the weather shock and volumes. Columns 5 and 6 find no significant relationship between spot volumes and the weather shock.

Table 2.5: Route-week Volumes Responsiveness to Local Temperature Shock

	(1) Total	(2) Total	(3) Contract	(4) Contract	(5) Spot	(6) Spot
Weather shock	-1.319**	-1.518***	-1.522***	-1.648***	-0.093	-0.116
	(0.565)	(0.569)	(0.505)	(0.509)	(0.237)	(0.243)
Weather shock squared		-0.892*		-0.566		-0.101
		(0.496)		(0.426)		(0.214)
Constant	82.766***	83.656***	65.849***	66.413***	11.997***	12.098***
	(0.000)	(0.495)	(0.000)	(0.425)	(0.000)	(0.213)
Observations	52,187	52,187	52,187	52,187	52,187	52,187
R^2	0.624	0.624	0.643	0.643	0.177	0.177
Fixed effects	lg-month	lg-month	lg-month	lg-month	lg-month	lg-month

Standard errors in parentheses, clustered at route-month level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

As additional evidence that weather shocks are indeed shocks to local demand at the destination, we show that spot prices vary with the temperature shock. We regress the weekly local spot price in a destination-week on the weather shock. We have $4,423$ destination-week observations and include destination-month fixed effects to control for predictable seasonality at each destination. The results are in Table 2.6. The estimates suggest a negative linear relationship: spot prices are high when the destination is unexpectedly cold.

2.4 Lower Bounds on Buyer Values for Contract Shipments

This second empirical analysis estimates lower bounds on the value of receiving LNG in a contract shipment rather than a spot shipment, given the thickness of the spot market and the status quo spot prices. Our approach builds on the intuition developed in Section 2.3 that the buyer requests contract shipments in periods when they have positive demand, which is also when spot prices tend to be high. We add to this the constraint that a contract relationship is also incentive compatible for the seller over the duration of a contracting period, here assumed to be a calendar year.

We first introduce some notation, which will also be used in the model in Section 2.5.

Table 2.6: Destination Spot Price Responsiveness to Local Temperature Shock

	(1) Local spot price	(2) Local spot price
Weather shock	-12.419***	-12.367***
	(2.175)	(2.244)
Weather shock squared		0.176
		(1.388)
Constant	180.682***	180.499***
	(0.124)	(1.441)
Observations	4423	4423
R^2	0.161	0.161
Fixed effects	g-month	g-month

Standard errors in parentheses, clustered at g-month level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

When there is a contract in place on a given seller-buyer route, the downstream buyer, g, has a value V_{lgt} for one cubic metre of LNG[20] arriving from seller l in a contract shipment at time t. The value g has for the same LNG quantity in a spot shipment in t is V^s_{gt}, which is not l-specific.[21] The per cubic metre prices for LNG arriving via contract and spot are denoted p^c_{lgt} and p^s_{gt}, respectively. A seller can deliver the shipments they have at time t to any of their contract partners or sell on any spot market, so the relevant spot price for the seller is the highest available spot price in t. We denote this p^s_{lt}. The object of interest in this section is $V^c_{lgt} - V^s_{gt}$ and the challenge is that only p^s_{gt} are observed in the data.

We start by considering the buyer's surplus from a contract shipment, $V^c_{lgt} - p^c_{lgt}$, and appeal to revealed preference: whenever the buyer receives a contract shipment, it must create more value for them than would a spot shipment at the local prevailing spot price. That is, buyer incentive compatibility requires:

$$V^c_{lgt} - p^c_{lgt} \geq V^s_{gt} - p^s_{gt},$$

or, equivalently,

$$V^c_{lgt} - V^s_{gt} \geq p^c_{lgt} - p^s_{gt}. \tag{2.3}$$

Turning next to the seller's incentive compatibility constraint, we argue that sellers must, on average, receive at least as much as what they could earn selling the same quantities

[20] The commodity in our model, which is LNG product, is homogeneous in either the contract market or the spot market.

[21] Note that s is used to denote a spot shipment and not a seller.

on spot markets over the contract relationship. For simplicity, in this section, we let the seller be forward-looking only over the month of a relevant contract shipment. The seller's dynamic incentive compatibility constraint implies that the contract prices they receive in expectation:

$$E(p^c_{lgt}) \geq E(max_G(p^s_{gt})), \tag{2.4}$$

where expectations are taken over all spot markets, G, over the contracted time period. We note that this is certainly a lower bound on the surplus required to align sellers' incentives in the relationship since it does not account for any variable or differential fixed costs incurred on contract shipments.

If both parties' expectations are consistent, we can combine the incentive compatibility constraints (2.3) and (2.4) to give:

$$E\left(V^c_{lgt} - V^s_{gt}\right) \geq E(p^c_{lgt}) - p^s_{gt} \geq E(max(p^s_{gt})) - p^s_{gt}. \tag{2.5}$$

The right had side of inequality (2.5) can be calculated using data on spot prices at all destinations at times when contracts are received by buyer g.

The data yield an estimate of a lower bound on $E\left(V^c_{lgt} - V^s_{gt}\right)$ per cubic metre of LNG received under contract by destination g in month t, which is the contract period under consideration in this section. We find estimates for all g-months where buyer g receives at least one contract shipment.

While we have made some strong assumptions to derive these lower bounds on the value contract shipments create for buyers under prevailing spot prices, the logic follows the simple intuition that the relative expected value from a contract is sufficient to compensate the seller for not switching to the spot market when the spot price is high together with the amount required to compensate the buyer for not switching to the spot market when the price is low at times of similar demand.

Figure 2.8 is a histogram of the estimated lower bounds of the relative value per cubic metre of LNG arriving at the destination in contract shipments compared to available spot shipments for the $1,158$ buyer-months in this analysis.

The mean value is \$27.07, the standard deviation is \$40.77 and the median is \$10.02. A regression of destination-month value on 11 year fixed effects has an adjusted R-squared of 0.24. Including only the 10 buyer fixed effects gives an adjusted R-squared of 0.36. A regression with fixed effects for the interaction of buyer and year fixed effects gives an adjusted R-squared of 0.83. This high value comes about because the approach suggests

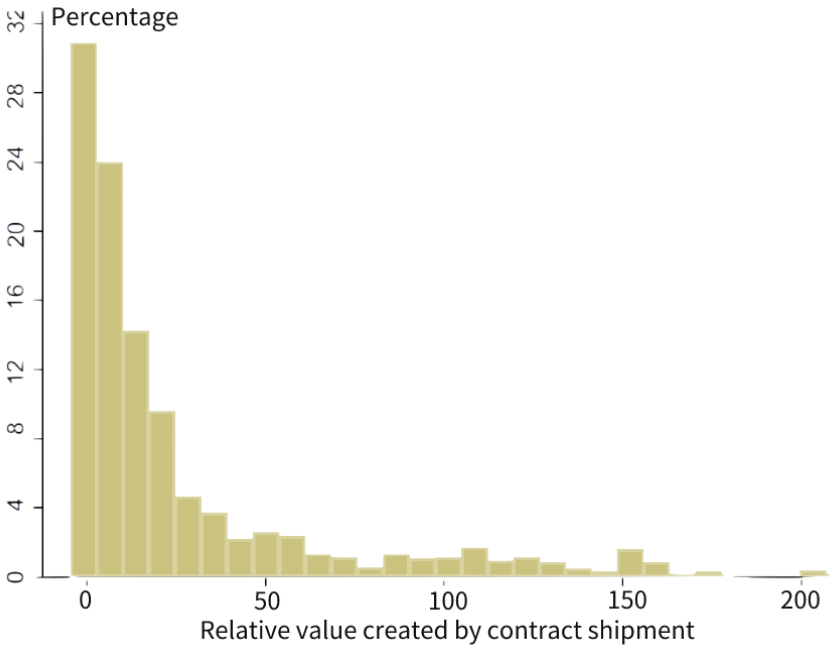


Figure 2.8: Relative Value of Contract Shipments

the contract shipments dispatched when there are high spot prices somewhere in the world must be particularly valuable.[22]

The estimated lower bounds can be interpreted as the value of a contract shipment over a unilateral deviation to a spot shipment in the same week, conditional on the vector of global spot prices at that time. In order to derive estimates of the relative value, rather than lower bounds, we need a structural model that delivers estimates of LNG demand and the relationship between spot market size and buyer spot value, as well as the nature of seller effort in contractual relationships and the division of contract surplus. This is the purpose of the Section 2.5 where we exploit exogenous weather shocks as well as local spot price variation.

[22] A similar exercise grouping time periods by destination-specific demand shock and assuming the difference between contract shipment value and contract price is at least as large as the difference between the highest and lowest prices paid for spot shipments under similar demand conditions yields similar estimates of relative contract value lower bounds.

2.5 Model

2.5.1 Environment

Time is discrete and each period t denotes a week. There are two types of agents, L sellers, $l \in 1, 2, ..., L$, and G buyers, $g \in 1, 2..., G$. All agents are risk neutral and have a discount factor δ.[23] The buyers and sellers are globally dispersed, and vessels containing LNG travel from sellers to buyers in each period. All possible lg pairs are feasible trading routes. We assume shipping capacity is not a constraint.

Each seller l has a fixed installed liquefaction capacity and produces an exogenous quantity of LNG in each t. Each buyer g has fixed installed regasification capacity. Buyer demand for LNG varies across time, and per period demand is stochastic.

In $t = 0$, l and g decide whether to write a long-term contract specifying an annual quantity of LNG to be shipped from l to g and a formula that determines the contract price P^c_{lgt}. We assume that whether or not there is a contract on a given route is exogenous to factors in the model.[24,25] Hence, at the start of each time period $t > 0$, each l has a pre-determined set of contract buyers and each g has a predetermined set of contract sellers. Each contract is incomplete in that the exact timing and quantity of any one delivery are not contracted and shipments are agreed bilaterally throughout each year of the contract.

Transactions on any lgt route-week can take one of two organizational forms: a contract shipment or a spot shipment. For each period $t > 0$, the timing is as follows:

- Each buyer g:
  - experiences a demand shock and learns the nature of its demand for LNG in that period.[26] All parties form expectations of the vector of all gt-specific spot prices p^s_{gt} based on observed shocks across all g.
  - makes a draw from a distribution that determines how well a shipment bought in the local spot market in time t will fit its demand conditions. The shape of the distribution reflects the expected t-specific thickness in g's local spot market.

[23] We do not explicitly model the shipping stage of the value chain in the current version. Specialized LNG vessels are usually owned by or on long-leases to sellers or buyers.

[24] Selection into having a contract on a route appears consistent with reduced form evidence about the adaptation benefits of contracts. As shown in Table 2.2, a seller is more likely to have a contract route with a buyer experiencing less predictable weather.

[25] Each contract's price is indexed to the destination region's oil price, P^o_{gt}, but the exact indexing rule is unobserved. Actual contract prices are also unobserved in our data.

[26] We have in mind that the buyer learns the ideal day of delivery and ideal volume it needs, if any, in week t.

  - decides whether to request a contract shipment from each of their contract partners.
- Then, each seller l:
  - chooses whether to fulfil each contract request received, and, conditional on fulfilling a request, decides how much costly relationship-specific effort to exert to ensure vessel capacity and arrival time meets buyer needs.
  - dispatches contract shipments and sends any remaining output to the spot market with the highest expected price.
- Destination spot prices p^s_{gt} are realized, equilibrating the residual spot demand and supply in each destination, and payments for contract and spot shipments are made.

2.5.2 Contract Shipments

When l fulfils g's request for a contract shipment in t, g's value for the shipment, V^c_{gt}, depends on their local demand shock, w_{gt}, and on the amount of relationship-specific effort made by the seller to fit the shipment to g's needs, $h_{lgt} \geq 0$. We specify a functional form for V^c_{gt}:

$$V^c_{gt}(w_{gt}, h_{lgc}) = v(w_{gt})(1 + h_{lgt})^{\mu} \tag{2.6}$$

In this "contract value production function", $\mu \leq 1$ is the intensity with which l's effort creates value. This parameter also captures any returns to scale in l's effort.

There is a constant marginal cost of seller effort, c. Each contract shipment also incurs a fixed cost F_{lt}. We allow F_{lt} to be an increasing function of the number of contract shipments fulfilled by l in t relative to l's typical number of contracts. This flexibility is intended to capture any increasing cost of contract fulfillment that is not g-specific. For example, perhaps it is harder to find the ideal vessel and send it at the ideal time for each contract shipment when l has more to manage.

An endogenous fraction β of contract value, net of fixed cost, goes to l and is built into the pre-agreed contract price index formula. Hence, for every contract shipment, the first order condition determining l's level of relationship-specific effort is:

$$\beta v(w_{gt})\mu(1 + h_{lgt})^{\mu-1} = c \tag{2.7}$$

Rearranging equation (2.7) and simplifying gives the seller's choice of effort (plus one) as:

$$(1 + h_{lgt}) = v(w_{gt})^{\frac{1}{1-\mu}} \left(\frac{\beta\mu}{c}\right)^{\frac{1}{1-\mu}}. \tag{2.8}$$

Seller effort is increasing in effort intensity, μ, and decreasing in marginal effort costs, c. A larger share of surplus going to l, β, also leads to higher levels of effort. Note that the production function implies potential complementarities between the responsiveness of buyer value to the gt demand shock, l's effort choice, and l's share of surplus.

This gives variable effort costs, borne by l, as:

$$c(1+h_{lgt}) = \beta\mu v(w_{gt}) \left(v(w_{gt})^{\frac{\mu}{1-\mu}} \left(\frac{\beta\mu}{c}\right)^{\frac{\mu}{1-\mu}} \right). \tag{2.9}$$

The contract shipment surplus, defined as the value to the buyer less the fixed cost, in terms of the model parameters, the endogenous variable β, and the function $v(w_{gt})$, is:

$$S^c_{lgt} = v(w_{gt}) \left(v(w_{gt})^{\frac{\mu}{1-\mu}} \left(\frac{\beta\mu}{c}\right)^{\frac{\mu}{1-\mu}} \right) - F_{lt} = v(w_{gt})^{\frac{1}{1-\mu}} \left(\frac{\beta\mu}{c}\right)^{\frac{\mu}{1-\mu}} - F_{lt} \tag{2.10}$$

Note that S^c_{lgt} is deterministic conditional on demand shocks w_{gt}, the number of contract requests that l receives in t, which determines F_{lt}, and model parameters. It is increasing in g's demand shock. Contract shipment surplus is also increasing in l's effort intensity and in the share going to the seller, and decreasing in the marginal cost of seller effort and in fixed cost.

2.5.3 Spot Shipments

When g buys from the spot market in t, its value for the spot shipment is V^s_{gt}, and it has to pay the local spot market price, p^s_{gt}. We assume there is no relationship-specific investment made by the seller for these shipments. For simplicity, we also assume spot shipments do not incur any fixed cost.

Shipments in the spot market are differentiated by vessel type and arrival time, which affect spot market value. The gt-specific shock to the spot value is ϕ^s_{gt}:

$$V^s_{gt} = \frac{v(w_{gt})}{\phi^s_{gt}}. \tag{2.11}$$

We let $\phi^s_{gt} \geq 1$ follow a Pareto distribution $\mathcal{F}_\phi$ with pdf $f_\phi = \frac{\lambda_{gt}}{\phi^{\lambda_{gt}+1}}$. We allow the λ_{gt} parameter to be an increasing function of the thickness of the local spot market in gt. This captures the idea that g is likely to find a better fit for its needs when there are relatively more spot shipments available and when those shipments' capacities are similar to the capacities of the contract shipments that are typically received.[27] The expected value of

[27] Appendix 2.A describes how we measure spot market thickness.

$\left(\frac{\lambda_{gt}}{\phi^{\lambda_{gt}+1}}\right)$ is an increasing function of λ_{gt}. Larger realized draws of ϕ^s_{gt} mean that g can find only a poor match for its needs in the spot market in t.

The buyer, g, observes its time-specific ϕ^s_{gt} as well as all w_{gt} demand shocks before deciding whether to request a contract shipment in t, buy from the local spot market, or not receive any shipments.

Seller l's payoff from selling a given shipment in the spot market is the highest spot price available at time t, $max_G(p^s_{gt})$.

2.5.4 Equilibrium

Whenever a contract shipment is observed, g has found it incentive compatible to request that shipment in t, and l has found it incentive compatible to fulfil the request. Following the evidence in Section 2.3, showing that contract volumes increase at times when there is an adverse weather shock, we assume that g's requests are consistent with per-period incentive compatibility.

To explain the fact that sellers choose to fulfil contract shipment requests even in times when their spot market outside alternative is particularly valuable, we assume the contractual relationship is dynamically incentive compatible for sellers because they extract value from the continuation of the relationship. Failing to deliver a contract shipment that fits buyer needs on request would end the relationship. That is, buyers play a trigger strategy in equilibrium.

Buyers' Incentive Compatibility Constraints

The state variables for buyer g in t are its pre-determined set of contract partners, the demand shock, w_{gt}, and the realized spot market value shock ϕ_{gt}. Having also observed all other demand shocks, g forms an expectation of the local spot price, $E(p^s_{gt})$.

Requesting a contract shipment is incentive compatible for g in t iff:

$$(1-\beta)S^c_{lgt} \geq V^s_{gt} - E(p^s_{gt}). \tag{2.12}$$

From equations (2.10) and (2.11), this inequality is satisfied when ϕ_{gt} exceeds ϕ^{s*}_{lgt}, which is given by:

$$\phi^{s*}_{lgt} = \frac{v(w_{gt})}{(1-\beta)\left[v(w_{gt})^{\frac{1}{1-\mu}}\left(\frac{\beta\mu}{c}\right)^{\frac{\mu}{1-\mu}} - F_{lt}\right] + E(p^s_{gt})}. \tag{2.13}$$

Intuitively, g calls in a contract shipment from contract partner l when the demand shock is

sufficiently high that g's share of the contract surplus, given l's relationship-specific effort, exceeds g's surplus from the spot market, given the expected spot price and the realized ϕ^s_{gt} draw. The threshold value ϕ^{s*}_{lgt} varies with the demand shock because it is a function of contract shipment surplus.[28]

If gt faces a positive demand shock that is sufficiently large such that $S^c_{lgt} \geq 0$, the probability that g requests a contract shipment from l is the probability $\phi_{gt} \geq \phi^{s*}_{lgt}$, which is:

$$\Pr(\phi^s_{gt} \geq \phi^{s*}_{lgt}) = 1 - F_\phi(\phi^{s*}_{lgt}) = (\phi^{s*}_{lgt})^{-\lambda},$$

and the probability that g goes to the spot market is the probability that ϕ^s_{gt} is less than ϕ^{s*}_{lgt}. This probability is given by:

$$\Pr(\phi^s_{gt} \leq \phi^{s*}_{lgt}) = F_\phi(\phi^{s*}_{lgt}) = 1 - (\phi^{s*}_{lgt})^{-\lambda},$$

Since whenever $S^c_{lgt} \geq 0$, the buyer always has the option of requesting a contract shipment to retrieve the positive surplus, the probability that there are no shipments is zero.

Sellers' Incentive Compatibility Constraint

We note that g is more likely to request a shipment when they receive a more positive demand shock, which is exactly when l's outside option of selling on the spot market is likely to be of high value (since demand shocks are correlated within each region containing multiple spot markets). We impose that the seller's equilibrium strategy satisfies their dynamic incentive compatibility constraint. That is, l fulfils a contract request if, over the course of all future periods ($t' > t$), the share of contract surplus it will receive, β, less the variable effort costs it will incur, outweighs the revenue it would earn from selling the same quantities in the spot market with the highest spot price in each period $t' \geq t$ net of seller-specific transport costs. That is, we compare the seller's value from a contract shipment to the value that seller would obtain on the spot market, given global spot prices and the transport costs from the seller to each market.

[28] Although we do not currently use it in the estimation, there is a second threshold value of ϕ^s_{gt}, above which the buyer prefers no shipment to a spot shipment, that is $\frac{v(w_{gt})}{\phi^s_{gt}} - E(p^s_{gt})) \leq 0$. This threshold value is:

$$\bar{\phi}^s_{gt} = \frac{v(w_{gt})}{E(p^s_{gt})}. \tag{2.14}$$

Seller l's incentive compatibility constraint for all future periods $t' \geq t$ can be written as follows:

$$\sum_{t' \geq t} \delta_{lt'}(\beta S^c_{lgt'} - c(1 + h_{lgt'})) \geq \sum_{t' \geq t} \delta_{lt'} E(max_{g \in G}(p^s_{gt'} - \tau_{lgt})), \tag{2.15}$$

where $g \in G$ indexes all destinations with spot markets and τ_{lgt} is the transport cost to each spot market destination from the seller in question in that time period. The discount factor is $\delta_{lt'} = \frac{1}{1+r_{lt'}}$, where $r_{lt'}$ is the seller-time-specific real interest rate.[29]

2.6 Estimation

We now describe how we estimate the model set out in Section 2.5. We derive theoretical moments that govern whether g requests a contract shipment in time t that is fulfilled by l. We then discuss the subsample of data used for estimation, which is chosen to reflect some of the abstractions made in the model. Finally, we describe the empirical moments and the estimation procedure.

2.6.1 Theoretical Moments

The parameters of interest are those that relate the demand shock to shipment value (in the function $v(w_{lgt})$), and to endogenous effort via effort intensity, μ, the share of seller surplus, β, and the marginal costs of contract effort, c. The other parameters relevant to contract shipments are those that affect the fixed costs, F_{lt}. We also focus on how the shape parameter of the Pareto distribution of spot value shocks, λ, relates to spot market thickness. We now describe three ways that we extend the model so as to, later, bring additional data to bear on estimating these relationships.

First, focusing on the value function $v(w_{lgt})$, we want this function to reflect how destinations in the data vary in their reliance on contemporaneous LNG supply as an energy source.[30] We hence allow the responsiveness of demand to weather shocks to depend on storage capacity relative to annual demand. We specify the shipment value function to be a power function governed by three parameters, k, α_i, for $i = 0, 1$. We set $v(w_{lgt}) = k w_{gt}^{\alpha_0 R_{gt}^{\alpha_1}}$, where w_{gt} is the weather shock at g at time t and R_{gt} is the ratio of total annual demand relative to storage capacity. Our hypothesis is that a higher ratio of demand to storage capacity means destinations are less able to use inventories to smooth out demand shocks

[29] These country-year interest rates are taken from the World Development Indicators.

[30] In some countries, it is a core energy source and in others it serves as a "peak shaver". Related countries vary in their storage capacity relative to total demand. Investments in storage capacity are an endogenous response to the contribution LNG makes to mean energy supply and to its variation.

and so this ratio will be associated with a greater responsiveness of willingness to pay to a weather shock, implying $\alpha_1 > 1$.

Second, because the relationship between spot market thickness and its value will be of key importance in the counterfactual analysis, we estimate λ flexibly to be $\lambda = \lambda_0(Th_{gt})^{\lambda_1}$,[31] where Th_{gt} is the measure of spot market thickness.

Third, in order to allow the fixed costs of fulfilling a contract shipment to vary with deviations from the typical number of l's contract shipments, we specify $F = F_0(N_{lt})^{F_1}$, where N_{lt} is the deviation from the mean number of contract shipments a seller fulfils in each period.

Considering also effort intensity, μ, the variable cost of seller effort, c, as well as the share of contract surplus going to the seller, β, we have a total of ten parameters to estimate, $\theta = \{k, \alpha_0, \alpha_1, \lambda_0, \lambda_1, \mu, c, F_0, F_1, \beta\}$.

We first look at the implications of g's incentive compatibility constraint in inequality (2.12). The threshold value of the shock above which a contract shipment is preferred to a spot shipment, ϕ^{s*}_{lgt}, from equation (2.13), can now be written:

$$\phi^{s*}_{lgt} \equiv \frac{kw_{gt}^{\alpha_0 R_{gt}^{\alpha_1}}}{(1-\beta)S^c_{lgt} + E(p^s_{gt})}.$$

Hence, the theoretical moments giving the probability a contract shipment is observed in lgt, conditional on $S^c_{lgt} \geq 0$, are:

$$\Pr(\phi^s_{gt} \geq \phi^{s*}_{lgt}) = \begin{cases} \left(\frac{kw_{gt}^{\alpha_0 R_{gt}^{\alpha_1}}}{(1-\beta)S^c_{lgt} + E(p^s_{gt})} \right)^{-\lambda} & S^c_{lgt} \geq 0 \\ 0 & S^c_{lgt} < 0 \end{cases} \tag{2.16}$$

where the term S^c_{lgt} is the surplus from a contract shipment and, under the functional form assumptions, equation (2.10) becomes:

$$S^c_{lgt} = \left(kw_{gt}^{\alpha_0 R_{gt}^{\alpha_1}} \right)^{\frac{1}{1-\mu}} \left(\frac{\beta\mu}{c} \right)^{\frac{\mu}{1-\mu}} - F_{lt},$$

and the theoretical moments in expression (2.16) are functions of all the model parameters of interest.

Some intuition for parameter identification can be found by observing that if w_{gt} is held

[31] A similar formulation is used for firms' cost draw by Boehm et al. (2022).

constant, then the numerator and first term in the denominator of the theoretical moments are fixed, and variation in the probability of observing a contract shipment is determined only by variation in expected spot prices, allowing us to identify λ. Similarly, given the functional form assumptions for V^o_{gt}, $o \in \{c, s\}$, and holding $E(p^s_{gt})$ fixed, variation in the probability of observing a contract shipment comes only from variation in w_{gt}, allowing us to estimate $k, \alpha_0, \alpha_1, \mu$ and a term that is a function of both β and c.

We then use l's dynamic incentive compatibility constraint, inequality (2.15), to derive lower bounds for β and permit an estimate of c. Although the theoretical moments given in equation (2.16) are a non-linear function of all the parameters, we show in Appendix 2.B that each parameter is identified.

2.6.2 Data Selection for Model Estimation

Because the model in Section 2.5 is at the lgt level, it abstracts away from several key aspects of the buyer and seller's broader contexts. For the purposes of estimation, we select a subset of the available data to limit the impact of this simplification.

First, we restrict attention to lg routes where there is a contract in place.[32] For the seller, the route-week model we use does not account for the fact that the total quantity a seller can dispatch in any one week is more or less fixed. To ensure these seller constraints do not introduce biases into our estimation, we focus only on lgt observations where the seller sells positive quantities in at least one spot market, meaning they could have sold more via contract if they had received further requests that they chose to fulfil.

We also only include lgt observations where the buyer received at least one contract shipment and/or spot shipment from some seller $g \in G$. This ensures that $S^c_{lgt} \geq 0$. Hence, the buyer's decision is only whether or not to request a contract shipment, conditional on having positive demand.[33]

There are two further lgt-specific variables that we construct to use as model inputs. The first is the number of contract shipments l fulfils in t relative to the typical number, N_{lt}, which helps us understand how sellers' fixed costs of contract shipments vary with demand. N_{lt} is defined as the de-meaned and normalized number of contract shipment requests that the seller delivers that month, where the number is de-meaned by the mean

[32] These are the route-years where we see at least one contract shipment on the route that year.

[33] Our model suggests weeks where buyers receive both contract and spot shipments are times when the buyer would have preferred to receive more contract shipments, but their additional requests were unfulfilled by the seller(s). We implicitly assume that, when they are asked, sellers $l \in L$ send sufficient contract shipments on a route to avoid relationship breakdown.

number of contract requests l fulfils and the normalization ensures $N_{lt} \in (0, 1)$. Scheduling vessel capacity and loading slots to fulfil specific contract requests is an ongoing seller-level management challenge and the more requests they receive in the weeks around the week in question, the more costly it could be to allocate resources to any one individual contract request. We use deviations from the typical share of contracts to allow for the possibility that sellers who routinely handle a large number of contract shipments have likely made sunk investments in this fixed managerial capacity.

The second is the measure of spot market thickness, Th_{gt}, that allows us to estimate how the idiosyncractic shock to spot shipment value varies with the nature of the local spot market. Our preferred measure has two components. We consider the relative size of the spot market in the region in that period compared to the destination's contract demand in that period. If buyers have total demand for a large number of shipments, and much of it is fulfilled via contract, then their spot market options are more limited. We make an adjustment to this measure by multiplying it by the extent of overlap between the distribution of vessel capacities requested by the buyer under contract and the distribution of vessel capacities available on the region-specific spot market. This adjustment accounts for the fact that even if the regional spot market is relatively large, a buyer might not be able to find a vessel that fits their demand specifications at any point in time.[34]

2.6.3 Empirical Moments and Estimation

There are $23,541$ lgt observations with positive trade flows that satisfy the restrictions in Subsection 2.6.2. We construct an indicator, C_{lgt} equal to 1 if there is at least one contract shipment in lgt. This indicator is equal to 0 if there are no contract shipments, which, given sample restrictions, implies g buys on the local spot market in t. $C_{lgt} = 1$ for $13,689$ of the $23,541$ observations, which is 58%.

Because l is defined at the country level, other than for Australia, Russia, and the United States, and g is also at the country level, there are many lgt observations where there is more than one contract shipment and/or g buys more than one spot shipment. In the estimation, we weight observations by the sum of the number of spot shipments g buys in t plus the number of contract shipments in lgt. There are $60,287$ weighted observations, of which $C_{lgt} = 1$ for $50,327$, or 83%.

As in Section 2.3, we use weather shocks as the demand shocks w_{gt}. In the structural estimation, the weather shock is defined as the deviation from the average mean temperature

[34] Appendix 2.A gives a detailed description of how Th_{gt} is constructed.

in a destination in a given month so, for example, it tells us when January is particularly cold relative to a typical January in g.[35] Although local spot price variation is associated with weather shocks, there is sufficient independent variation in the data that both spot price and weather shocks are significantly correlated with the probability of receiving a contract shipment when both are included in a regression. We will use the coefficients in this regression as untargeted moments in our estimation as a robustness check for the parameter estimates. The weighted regression generating these two coefficients from the data sample is:

$$C_{lgt} = \alpha_1 w_{gt} + \alpha_2 p^s_{gt} + \gamma_{gm} + \varepsilon_{lgt} \tag{2.17}$$

The estimated coefficients are $\hat{\alpha}_1 = 0.069$ and $\hat{\alpha}_2 = 0.074$, and both are significant at the 5% level.

We employ Nonlinear Least Squares estimation to identify the model parameters, denoted as $\boldsymbol{\theta}$. Specifically, we seek parameter values that minimize the following expression:

$$\min_{\boldsymbol{\theta}} \sum_{t=1}^{T} \sum_{g=1}^{G} \sum_{l=1}^{L} \left((\phi^{s*}_{lgt})^{-\lambda_{gt}} - D_{lgt} \right)^2 \tag{2.18}$$

Here, ϕ^{s*}_{lgt} represents the model-implied probability of having a contract shipment, and D_{lgt} denotes the observed share of contract shipments in our selected sample. Given that our model is highly nonlinear and includes numerous parameters, it is critical to ensure that our model is identifiable. In Appendix 2.B, we demonstrate this aspect of the model.

2.7 Results

2.7.1 Parameter Estimates

Estimating the model by the NLLS (2.18) gives the following parameter estimates:

All estimates are significantly different from zero. The positive value function parameters α_0 and α_1 show that the value g receives from a shipment is higher when there is a weather shock, that is, when it is unexpectedly cold. The positive estimates of λ_0 and λ_1 show that the value g receives from spot market shipments is positively associated with Th_{gt}, our measure of local spot market thickness at time t. The coefficient μ of 0.45 shows that relationship-specific seller effort contributes positively to contract shipment value. The

[35] Work in progress estimates the model using the standardized weather shock described in Section 2.2.4 rather than simply the deviation from the destination-month mean temperature. Comparing the reduced form results suggests the findings will be qualitatively similar.

Parameter	Description	Estimate	Confidence Interval
α_0	Value function normalization	9.95	[9.83, 10.07]
α_1	Value function elasticity	0.21	[0.17, 0.25]
μ	Effort intensity	0.45	[0.44, 0.45]
c	Marginal effort cost	1.11	[1.10, 1.13]
F_0	Contract fixed cost normalization	4.98	[2.53, 7.42]
F_1	Contract fixed cost elasticity	0.61	[0.31, 0.84]
λ_0	Spot market thickness normalization	0.79	[0.69, 0.97]
λ_1	Spot market thickness elasticity	0.33	[0.29, 0.37]

positive coefficients F_0 and F_1 suggest there are fixed costs associated with each contract shipment that are increasing in the deviation from the typical share of contract requests that a seller fulfills that month.

To find the β parameter estimate, we return to inequality (2.15):

$$\sum_{t' \geq t} \delta_{lt'}(\beta S^c_{lgt'} - c(1 + h_{lgt'})) \geq \sum_{t' \geq t} \delta_{lt'} E(max_{g \in G}(p^s_{gt'})),$$

which delivers an estimate of β for each lgt. The distribution of these estimates is given in Figure 2.9. Since we need our estimate of β to satisfy all of these constraints, we take it as

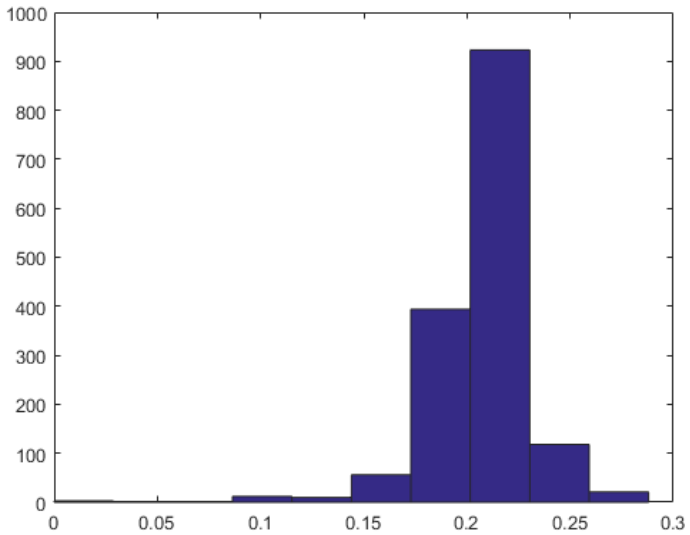


Figure 2.9: Estimated Lower Bounds on β

the maximum of the distribution in Figure 2.9, which is 0.2880, with a confidence interval $[0.2846, 0.2915]$.

The results give that the seller exerts relationship-specific effort, that is, $h > 0$, for 49% of

contract shipments. The seller incurs additional variable costs in these cases. This suggests that the seller's share of the contractual relationship surplus is sufficient to compensate them for foregoing spot markets and also exerting effort for these shipments.[36] In the model, h is an increasing function of the weather shock, and there is a strong positive correlation between estimated h_{lgt} values and w_{gt}. That is, sellers exert effort within contractual relationships when their contract partners face unexpectedly cold weather.

The parameter predictions allow us to predict the responsiveness of C_{lgt}, whether or not a contract shipment is observed in a route-week as determined by the estimated model, to observed variation in weather and spot prices using equation (2.17). These untargeted moments are close to the actual data, with the model predicting slightly more variation with spot prices than we observe in the data.

Coefficient	Data	Model predictions
w_{gt}	0.069**	0.070***
	(0.028)	(0.001)
p^s_{gt}	0.074**	0.100***
	(0.012)	(0.001)

2.7.2 Shipment Value Estimates

The estimated values of the observed contract shipments are shown in Figure 2.10. The far right histogram presents calculated shipment values from equation (2.6):

$$V^c_{gt}(w_{gt}, h_{lgt}) = \alpha_0 w^{\alpha_1}_{gt}(1 + h_{lgt})^{\mu}$$

The middle histogram subtracts estimated fixed costs, F_{lt}, giving the distribution of S^c_{lgt}. The left (purple) histogram is the surplus of contract shipments net of estimated variable seller effort costs, $c(1 + h_{lgt})$. The mean of the left (purple) distribution is 5.58 and the standard deviation is 0.67. These values can be compared to the normalized distribution of the local spot price per mmbtu, the mean of which is 0.31. That is, we estimate that the mean value of a contract shipment is eighteen times higher than the mean spot price over the period studied.

The expected spot market shipment value can be calculated from local weather and the

[36] We note that the seller receives only a β share of the additional surplus generated through its effort, and so is suboptimal.

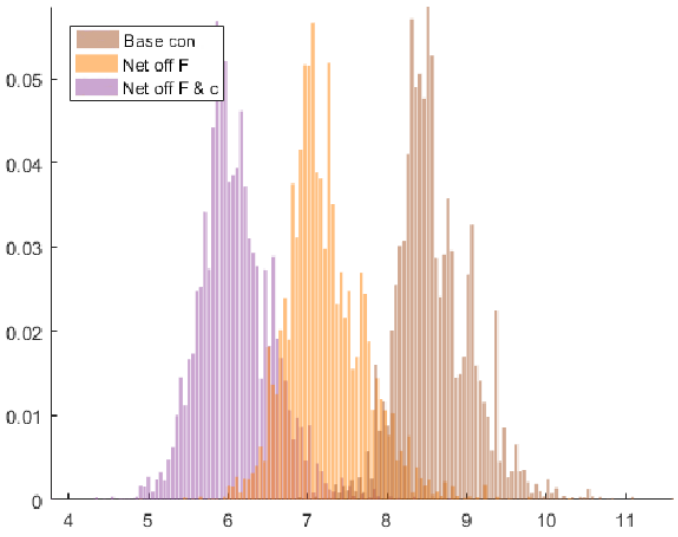


Figure 2.10: Estimated Value of Contract Shipments

estimate of $\lambda_{gt} = \lambda_0 Th_{gt}^{\lambda_1}$, where Th_{gt} is the observed thickness of the local spot market.[37] It is:

$$E(\mathscr{V}_{gt}^s) = E\left(\frac{\alpha_0 w_{gt}^{\alpha_1}}{\phi_{gt}^s}\right) = \alpha_0 w_{gt}^{\alpha_1} \frac{\lambda}{\lambda+1} \tag{2.19}$$

Figure 2.11 shows the calculated expected spot values together with the net contract surplus net of variable effort costs (purple histogram) from Figure 2.10. In the figure, the blue histogram gives the expected values of spot shipments when at least one spot shipment is observed in lgt, weighted by the actual number of spot shipments observed in the data. The mean of the distribution is 2.19 and the standard deviation is 0.41. The estimated mean value of a spot shipment is hence only 40% of the mean value of a contract shipment, but is still around 7 times larger than the normalized mean spot price. In other words, we find that around 15% of the gains from trade in the current spot market go to the seller and around 85% goes to buyers. Note that we estimate $\beta = 0.29$, which is the share of surplus going to sellers in the contract relationships, suggesting that the sellers extract a smaller share of the smaller gains from trade in spot shipments.

[37] The derivation is: $E(\mathscr{V}_{gt}^s) = E\left(\frac{\alpha_0 w_{gt}^{\alpha_1}}{\phi_{gt}^s}\right) = \alpha_0 w_{gt}^{\alpha_1} \frac{\lambda_{gt}}{\lambda_{gt}+1} = \alpha_0 w_{gt}^{\alpha_1} E(\frac{1}{\phi_{gt}^s}) = \alpha_0 w_{gt}^{\alpha_1} \int \frac{1}{\phi_{gt}^s} \frac{\lambda_{gt}}{(\phi_{gt}^s)^{\lambda_{gt}+1}} d\phi_{gt}^s = \alpha_0 w_{gt}^{\alpha_1} \lambda_{gt} \int \frac{1}{(\phi_{gt}^s)^{\lambda_{gt}+2}} d\phi_{gt}^s = -\alpha_0 w_{gt}^{\alpha_1} \frac{\lambda_{gt}}{\lambda_{gt}+1} (\phi_{gt}^s)^{-(\lambda_{gt}+1)} |_1^\infty = \alpha_0 w_{gt}^{\alpha_1} \frac{\lambda_{gt}}{\lambda_{gt}+1}$.

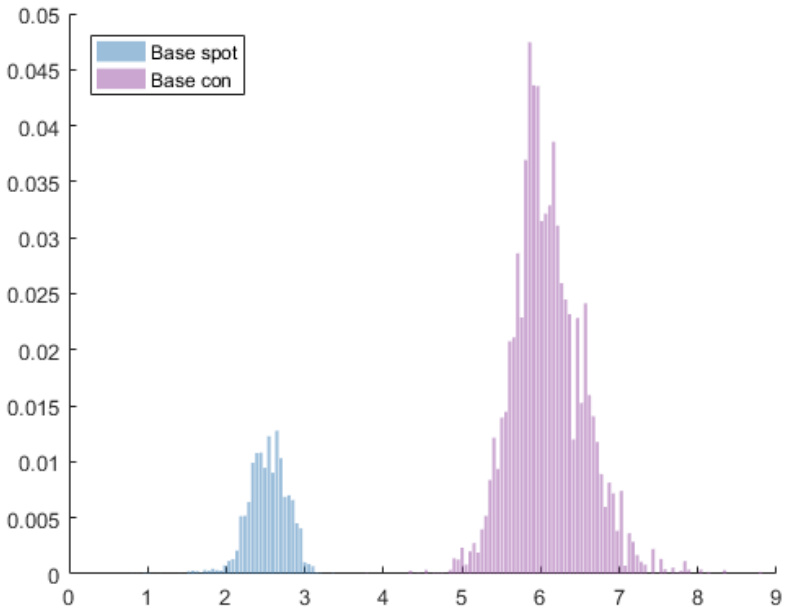


Figure 2.11: Expected Value of Spot and Contract Shipments

2.8 Counterfactual Analysis

We now ask how much long-term contractual relationships contribute to industry value relative to a situation where the same vessels arrived in each destination in each period and were transacted in the spot market. That is, the counterfactual is larger spot markets in each lgt where contract shipments are currently observed. The trade-off is that all shipments become "anonymous" arm's length transactions, and sellers exert no relationship-specific effort and incur zero contract fixed and variable costs.

We are able to estimate this counterfactual because the λ parameter estimates from Section 2.7 allow us to calculate the expected spot shipment value as a function of the observed weather shock and the counterfactual spot market thickness, Th_{gt}, using equation (2.19). The measure of Th_{gt} becomes equal to 1 because the share of regional supply that is spot is now 100%. In addition, because the vessel sizes that were demanded under contract by each buyer are now a subset of the distribution of vessel sizes available via spot, the distributions of vessel sizes become equivalent (see Appendix 2.A).

One major appeal of this approach is that we do not need to simulate counterfactual spot prices. We also avoid needing to address the seller's allocation problem, that is, where they would send vessels in the absence of contracts. A related benefit is that our counterfactual

estimates are invariant to transport costs, which are currently excluded from the analysis, but which remain constant in the counterfactual.

For this analysis, we estimate the counterfactual values of all contract and spot shipments, including those that did not satisfy the estimation sample restriction in Subsection 2.6.2. This gives us $31,428$ contract shipments and $6,201$ spot shipments, so $37,449$ shipments in total.

The predicted values of these $37,449$ shipments sold in the enlarged spot markets is given in the red histogram in Figure 2.12. For comparison, the blue histogram is the distribution of actual spot shipment values given the observed much smaller number of spot shipments, as shown in Figure 2.11. Figure 2.12 also includes the observed shipment surplus for contracts net of variable costs (the purple histogram). The red histogram reflects higher counterfactual spot values than in the blue histogram because buyers expect to receive a spot shipment that is a closer fit to their needs in larger spot markets. This counterfactual distribution has a mean value of 3.41 and a standard deviation of at 0.35, which is smaller than the standard deviation in the observed spot value distribution. This is because there is no variation in Th_{gt} in the counterfactual.

Because the counterfactual holds all transport costs fixed, and varies only the value created for buyers, it is possible to compare total welfare in the observed and counterfactual cases.

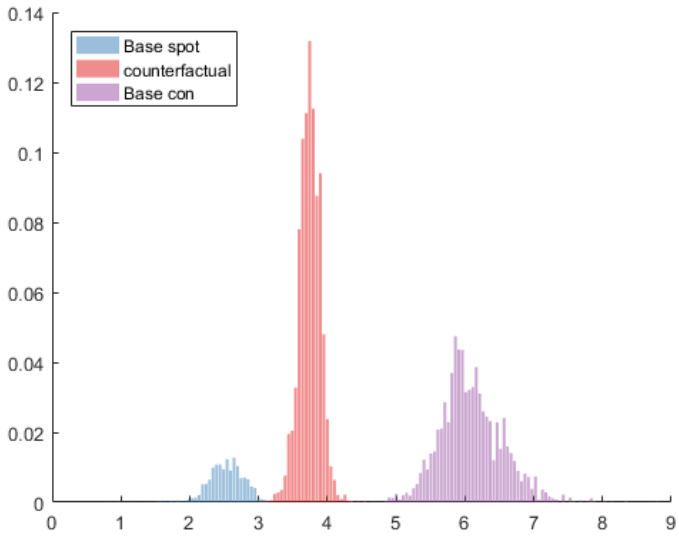


Figure 2.12: Observed and Counterfactual Shipment Values

The total change in welfare can be decomposed into three parts, where the variables with asterisks denote the number of counterfactual shipments, Q_s^*, and the expected value of the distribution of the expected value of those counterfactual shipments, $E(V_s^*)$:

$$\begin{aligned}\Delta W &= \Delta W_s + \Delta W_c \\ &= Q_s^* E(V_s^*) - (Q_s E(V_s) + Q_c(V(h) - F - c(1+h)) \\ &= (Q_s + Q_c)E(V_s^*) - (Q_s E(V_s) + Q_c(V(h) - F - c(1+h))) \\ &= Q_s(E(V_s^*) - E(V_s)) + Q_c(E(V_s^*) - V(h)) + Q_c(F + c(1+h)) \end{aligned} \tag{2.20}$$

The first term in equation (2.20) is the percentage increase in the value of spot markets in the counterfactual where they are much larger. This change is 3.43%. The second term is the percentage reduction in the value of the shipments that are currently transacted via contracts, and is estimated to be 73.65%. This loss can be attributed to the loss in shipment value from the fact that sellers no longer match these shipments to buyers' exact needs and exert no relationship-specific costly effort to ensure the quality of the match. The third term in equation (2.20) is the welfare gain coming from the savings in the seller's variable effort costs and fixed costs associated with contract shipments. This saving is equivalent to 38.06% of industry value over the time period studied.

The net impact of the removal of contracts is a reduction in industry value of 32.12%. We infer that total industry size remains too small for spot markets to deliver shipments to buyers that fit their demand even if all industry supply were transacted on global spot markets rather than via contracts.

2.9 Conclusion

Long-term contracts dominate the LNG industry and the nature of these relationships shapes industry value. This chapter brings insights from recent advances in the theory of incomplete contracts to bear on detailed transaction-level data. Our main contribution is the finding that these contracts allow transactions to flexibly adjust to unexpected demand circumstances. We also quantify the significant role played by relationship-specific seller effort in creating value for the buyer. Adaptation, often enabled by non-contractible effort, increases industry efficiency relative to a counterfactual of LNG spot markets characterized by frictions from technology and geography. The analysis reveals that the value of the global LNG industry would only have been two thirds of its actual size between 2009 and 2019 in the absence of adaptation within long-term contracts.

2.A Measuring Spot Market Thickness

For each destination-week, the measure Th_{gt} is the product of two terms.

First, we find the share of spot shipments in the region (Europe or Asia) in that month, which is the total number of spot shipments delivered to all destinations in the region divided by the total number of spot and contract shipments delivered in the region. The higher is this share, the more likely any buyer in the region could purchase spot shipments at the time they demand it.

Second, we turn to the match of spot shipment capacities with those demanded via contracts. We find the distribution of spot vessel sizes that have been delivered in the region over the last three months. We then add to these shipments the contract shipments delivered to the destination that month to give the capacity distribution of typical region spot shipments together with the destination's contract capacity demand. We use the Kolmogorov-Smirnov statistic to tell us how similar these two distributions are. If they are very similar, we infer that the destination's demand contract capacities could be catered to by the region's spot market. The more similar the distributions the thicker the local spot market, in terms of shipment capacity availability.

In the counterfactual, by construction, both of these terms equal one. Hence, the spot market in each destination-week achieves the maximum spot market thickness that is ever observed in the actual data. The buyer values V^s_{gt} for this enlarged spot market are still discounted by the gt-specific shock ϕ^s_{gt}, whose distribution is determined by the estimated λ_0 and λ_1 parameters presented in Section 2.7.

2.B Model Identification

We employ Nonlinear Least Squares estimation to identify the model parameters, denoted $\boldsymbol{\theta}$. Specifically, we choose parameter values that solve the following minimization problem:

$$\min_{\boldsymbol{\theta}} \sum_{t=1}^{T} \sum_{g=1}^{G} \sum_{l=1}^{L} \left((\phi^{s*}_{lgt})^{-\lambda_{gt}} - D_{lgt} \right)^2 .$$

Given that our model is highly nonlinear and includes numerous parameters, it is critical to ensure that our model is identifiable. According to Silvey (1970), a model is identifiable if no two different sets of parameter values result in the same model. An obvious cause of

non-identifiability is parameter redundancy. This phenomenon occurs when a model can be respecified in terms of a smaller set of parameters (Catchpole and Morgan (1997)).

Cole et al. (2010) provided a comprehensive framework for detecting parameter redundancy in our case by defining the vector $\boldsymbol{\kappa} = [\kappa_{111}, \kappa_{121}, ...\kappa_{lgt}.., \kappa_{LGT}]^T$ with $\kappa_{lgt} = (\phi_{lgt}^{s*})^{-\lambda_{gt}}$. They showed that if the rank of $\boldsymbol{D} \equiv \frac{\partial \boldsymbol{\kappa}}{\partial \boldsymbol{\theta}}$, denoted by k, is smaller than the number of parameters q, then the model is parameter redundant and will be non-identifiable. This scenario aligns with the concept of rank-deficient nonlinear least squares, which typically does not yield a unique solution Eriksson and Gulliksson (2004). Conversely, if $k = q$, the model is of full rank, not parameter redundant, and will be at least locally identifiable (Theorem 2a of Cole et al. (2010)).

Given our functional forms of interest, $v(w_{gt}) = k{w_{gt}}^{\alpha_0 R_{gt}^{\alpha_1}}$, $F_{lt} = F_0 CS_{lt}^{F_1}$, and $\lambda_{gt} = \lambda_0 TH_{gt}^{\lambda_1}$, with $k > 0, \alpha_0 > 0, F_0 > 0, \lambda_0 > 0, \mu > 0$, where R_{gt} is the ratio of annual demand to storage capacity, w_{gt} is the normalized weather shock, CS_{lt} is the share of contract shipments, and TH_{gt} is spot market thickness, we can express κ_{lgt} as follows:

$$
\begin{aligned}
\kappa_{lgt} &= \left(\frac{v(w_{gt})}{(1-\beta)v(w_{gt})^{\frac{1}{1-\mu}}(\frac{\beta\mu}{c})^{\frac{\mu}{1-\mu}} - (1-\beta)F_{lt} + E(P_{gt})}\right)^{-\lambda_{gt}} \\
&= \left(\frac{E(P_{gt}) - (1-\beta)F_{lt} + Cv(w_{gt})^{\frac{1}{1-\mu}}}{v(w_{gt})}\right)^{\lambda_{gt}} \\
&= \left(E(P_{gt})\frac{1}{k}{w_{gt}}^{-\alpha_0 R_{gt}^{\alpha_1}} - (1-\beta)F_0 CS_{lt}^{F_1}\frac{1}{k}{w_{gt}}^{-\alpha_0 R_{gt}^{\alpha_1}} + (1-\beta)(\frac{\beta\mu}{c})^{\frac{\mu}{1-\mu}}(k{w_{gt}}^{\alpha_0 R_{gt}^{\alpha_1}})^{\frac{\mu}{1-\mu}}\right)^{\lambda_0 TH_{gt}^{\lambda_1}}
\end{aligned}
$$

We can redefine the parameters as

$$
\begin{aligned}
\tilde{k} &= \frac{1}{k}, \\
\tilde{F_0} &= (1-\beta)F_0, \\
C &\equiv (1-\beta)\left(\frac{\beta\mu}{c}\right)^{\frac{\mu}{1-\mu}}
\end{aligned}
$$

It is easy to see that c cannot be separately identified as an increase in c can be offset by changes in F_0 and β, resulting in the exactly the same model results. Therefore, the parameters that need to be checked for identifiability are $\theta = [\tilde{k}, \alpha_0, \alpha_1, \tilde{F_0}, F_1, C, \mu, \lambda_0, \lambda_1]$.

This gives the following expression for κ_{lgt}:

$$
\kappa_{lgt} = \left(E(P_{gt})\tilde{k}{w_{gt}}^{-\alpha_0 R_{gt}^{\alpha_1}} - \tilde{F_0} CS_{lt}^{F_1}\tilde{k}{w_{gt}}^{-\alpha_0 R_{gt}^{\alpha_1}} + C(\tilde{k}{w_{gt}}^{-\alpha_0 R_{gt}^{\alpha_1}})^{\frac{\mu}{\mu-1}}\right)^{\lambda_0 TH_{gt}^{\lambda_1}}
$$

The first derivative matrix $\boldsymbol{D} \equiv \frac{\partial \boldsymbol{\kappa}}{\partial \boldsymbol{\theta}}$ is given by

$$\boldsymbol{D} = \frac{\partial \boldsymbol{\kappa}}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial \kappa_{111}}{\partial \theta_1} & \frac{\partial \kappa_{121}}{\partial \theta_1} & \dots & \frac{\partial \kappa_{LGT}}{\partial \theta_1} \\ \frac{\partial \kappa_{111}}{\partial \theta_2} & \frac{\partial \kappa_{121}}{\partial \theta_2} & \dots & \frac{\partial \kappa_{LGT}}{\partial \theta_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \kappa_{111}}{\partial \theta_9} & \frac{\partial \kappa_2}{\partial \theta_9} & \dots & \frac{\partial \kappa_{LGT}}{\partial \theta_9} \end{bmatrix}$$

If the model is not parameter redundant, then the rank of the matrix $\boldsymbol{D}$ should be equal to the number of parameters, which is 9, implying it has full row rank. To demonstrate this, we derive the first derivatives contained within the matrix:

$$\frac{\partial \kappa_{lgt}}{\partial \tilde{k}} = \lambda_{gt} \kappa_{lgt}^{\frac{\lambda_{gt}-1}{\lambda_{gt}}} (E(P_{gt}) - \tilde{F}_0 C S_{lt}^{F_1} \tilde{k} w_{gt}^{-\alpha_0 R_{gt}^{\alpha_1}} + C \frac{\mu}{\mu - 1} (\tilde{k} w_{gt}^{-\alpha_0 R_{gt}^{\alpha_1}})^{\frac{1}{\mu-1}}) w_{gt}^{-\alpha_0 R_{gt}^{\alpha_1}}$$

$$\frac{\partial \kappa_{lgt}}{\partial \alpha_0} = -\lambda_{gt} \kappa_{lgt}^{\frac{\lambda_{gt}-1}{\lambda_{gt}}} (E(P_{gt}) - \tilde{F}_0 C S_{lt}^{F_1} \tilde{k} w_{gt}^{-\alpha_0 R_{gt}^{\alpha_1}} + C \frac{\mu}{\mu - 1} (\tilde{k} w_{gt}^{-\alpha_0 R_{gt}^{\alpha_1}})) \tilde{k} w_{gt}^{-\alpha_0 R_{gt}^{\alpha_1}} R_{gt}^{\alpha_1} \ln w_{gt}$$

$$\frac{\partial \kappa_{lgt}}{\partial \alpha_1} = -\lambda_{gt} \kappa_{lgt}^{\frac{\lambda_{gt}-1}{\lambda_{gt}}} (E(P_{gt}) - \tilde{F}_0 C S_{lt}^{F_1} \tilde{k} w_{gt}^{-\alpha_0 R_{gt}^{\alpha_1}} + C \frac{\mu}{\mu - 1} (\tilde{k} w_{gt}^{-\alpha_0 R_{gt}^{\alpha_1}})) \tilde{k} w_{gt}^{-\alpha_0 R_{gt}^{\alpha_1}} \alpha_0 R_{gt}^{\alpha_1} \ln w_{gt} \ln R_{gt}$$

$$\frac{\partial \kappa_{lgt}}{\partial F_0} = \lambda_{gt} \kappa_{lgt}^{\frac{\lambda_{gt}-1}{\lambda_{gt}}} C S_{lt}^{F_1} \tilde{k} w_{gt}^{-\alpha_0 R_{gt}^{\alpha_1}}$$

$$\frac{\partial \kappa_{lgt}}{\partial F_1} = \lambda_{gt} \kappa_{lgt}^{\frac{\lambda_{gt}-1}{\lambda_{gt}}} \tilde{F}_0 C S_{lt}^{F_1} \tilde{k} w_{gt}^{-\alpha_0 R_{gt}^{\alpha_1}} \ln C S_{lt}$$

$$\frac{\partial \kappa_{lgt}}{\partial C} = \lambda_{gt} \kappa_{lgt}^{\frac{\lambda_{gt}-1}{\lambda_{gt}}} (\tilde{k} w_{gt}^{-\alpha_0 R_{gt}^{\alpha_1}})^{\frac{\mu}{\mu-1}}$$

$$\frac{\partial \kappa_{lgt}}{\partial \mu} = -\lambda_{gt} \kappa_{lgt}^{\frac{\lambda_{gt}-1}{\lambda_{gt}}} (C (\tilde{k} w_{gt}^{-\alpha_0 R_{gt}^{\alpha_1}})^{\frac{\mu}{\mu-1}} \frac{\ln(\tilde{k} w_{gt}^{-\alpha_0 R_{gt}^{\alpha_1}})}{(\mu - 1)^2})$$

$$\frac{\partial \kappa_{lgt}}{\partial \lambda_0} = \kappa_{lgt} \ln \kappa_{lgt}^{\frac{1}{\lambda_{gt}}} T H_{gt}^{\lambda_1}$$

$$\frac{\partial \kappa_{lgt}}{\partial \lambda_1} = \kappa_{lgt} \ln \kappa_{lgt}^{\frac{1}{\lambda_{gt}}} \lambda_0 T H_{gt}^{\lambda_1} \ln T H_{gt}$$

To prove that the matrix $\boldsymbol{D}$ is of full row rank, we require that the only solution to the following equation satisfies simultaneously for all lgt-level data is $\beta^i = 0, \forall i \in \boldsymbol{\theta}$ (Horn and Johnson, 2012).

$$\beta^{\tilde{k}}\frac{\partial \kappa_{lgt}}{\partial \tilde{k}}+\beta^{\alpha_0}\frac{\partial \kappa_{lgt}}{\partial \alpha_0}+\beta^{\alpha_1}\frac{\partial \kappa_{lgt}}{\partial \alpha_1}+\beta^{F_0}\frac{\partial \kappa_{lgt}}{\partial F_0}+\beta^{F_1}\frac{\partial \kappa_{lgt}}{\partial F_1}+\beta^{C}\frac{\partial \kappa_{lgt}}{\partial C}+\beta^{\mu}\frac{\partial \kappa_{lgt}}{\partial \mu}+\beta^{\lambda_0}\frac{\partial \kappa_{lgt}}{\partial \lambda_0}+\beta^{\lambda_1}\frac{\partial \kappa_{lgt}}{\partial \lambda_1} = 0 \tag{2.21}$$

where β^i represents the coefficients associated with each parameter i in the parameter set $\boldsymbol{\theta}$. This condition indicates that each parameter i in $\boldsymbol{\theta}$ does not contribute linearly to the model outcomes, ensuring that $\boldsymbol{D}$ is of full row rank by demonstrating the linear independence of its rows.

Upon substituting the first derivatives into the matrix $\boldsymbol{D}$ and collecting terms, we obtain the following expression:

$$\begin{aligned}
0 =&(\beta^{\tilde{k}} - \beta^{\alpha_0}\tilde{k}R_{gt}^{\alpha_1}\ln w_{gt} - \beta^{\alpha_1}\tilde{k}\alpha_0 R_{gt}^{\alpha_1}\ln w_{gt}\ln R_{gt})H_{lgt}^{v}\\
&+(\beta^{F_0} + \beta^{F_1}\tilde{F}_0\ln CS_{lt})H_{lgt}^{F}\\
&+(\beta^{C} - \beta^{\mu}(C\frac{\ln(\tilde{k}w_{gt}^{-\alpha_0 R_{gt}^{\alpha_1}})}{(\mu-1)^2}))H_{lgt}^{C}\\
&+(\beta^{\lambda_0}\frac{1}{\lambda_0} + \beta^{\lambda_1}\ln TH_{gt})H_{lgt}^{\lambda}
\end{aligned} \tag{2.22}$$

where $H_{lgt}^{v} \equiv (E(P_{gt}) - \tilde{F}_0 CS_{lt}^{F_1}\tilde{k}w_{gt}^{-\alpha_0 R_{gt}^{\alpha_1}} + C\frac{\mu}{\mu-1}(\tilde{k}w_{gt}^{-\alpha_0 R_{gt}^{\alpha_1}})^{\frac{1}{\mu-1}})w_{gt}^{-\alpha_0 R_{gt}^{\alpha_1}}$, $H_{lgt}^{F} \equiv CS_{lt}^{F_1}\tilde{k}w_{gt}^{-\alpha_0 R_{gt}^{\alpha_1}}$, $H_{lgt}^{C} \equiv (\tilde{k}w_{gt}^{\alpha_0 R_{gt}^{\alpha_1}})^{\frac{\mu}{\mu-1}}$, and $H_{lgt}^{\lambda} \equiv \kappa_{lgt}^{\frac{1}{\lambda_{gt}}}\ln \kappa_{lgt}^{\frac{1}{\lambda_{gt}}}$.

Parameter redundancy in a model can be categorized into two types: intrinsic parameter redundancy, which is inherent to the model structure regardless of the data, and extrinsic parameter redundancy, which arises due to the characteristics of the dataset used. For instance, consider a scenario where $w_{gt} = 1\,\forall gt$. In such a case, the parameters β^{α_0} and β^{α_1} can assume any values, while all other parameters $\beta^i = 0\,\forall i \neq \alpha_0, \alpha_1$, thus satisfying equation (2.21) and leading to parameter redundancy induced by this specific dataset. This prevents the identification of β^{α_0} and β^{α_1}.

Given a sufficiently large and varied dataset, our analysis is concentrated on intrinsic parameter redundancy, under the assumption of an ideal dataset. This entails that the functions $H_{lgt}^{v}, H_{lgt}^{F}, H_{lgt}^{C}, H_{lgt}^{\lambda}$ are linearly independent, given our parameter assumptions

that $k > 0, \alpha_0 > 0, F_0 > 0, \lambda_0 > 0, \mu > 0$. Consequently, it must follow that each coefficient in front of $H^i, i \in \{v, F, C, \lambda\}$, is equal to 0:

$$\begin{aligned}
\beta^{\tilde{k}} - \beta^{\alpha_0} \tilde{k} R_{gt}^{\alpha_1} \ln w_{gt} - \beta^{\alpha_1} \tilde{k} \alpha_0 R_{gt}^{\alpha_1} \ln w_{gt} \ln R_{gt} &= 0 \\
\beta^{F_0} + \beta^{F_1} \tilde{F}_0 \ln CS_{lt} &= 0 \\
\beta^{C} - \beta^{\mu} (C \frac{\ln(\tilde{k} w_{gt}^{-\alpha_0 R_{gt}^{\alpha_1}})}{(\mu - 1)^2}) &= 0 \\
\beta^{\lambda_0} \frac{1}{\lambda_0} + \beta^{\lambda_1} \ln TH_{gt} &= 0
\end{aligned} \tag{2.23}$$

These four conditions must hold for any lgt, within a perfect dataset. Consequently, the solution to these equations is $\beta^i = 0 \, \forall i \in \boldsymbol{\theta}$. This establishes the proof for the identifiability of our nonlinear least squares model.

Equations (2.22) and (2.23) also explain the variations utilized to separately identify each parameter. From equation (2.22), variations in $H_{lgt}^v, H_{lgt}^F, H_{lgt}^C$, and H_{lgt}^λ enable the separate identification of parameters associated with the value function, fixed costs, returns to scale, and market thickness, respectively. Similarly, equation (2.23) indicates that variation in $R_{gt}^{\alpha_1} \ln w_{gt}$ allows for the separate identification of α_0 from $\tilde{k}$, and variation in $\ln R_{gt}$ allow for the identification of α_1 from α_0.

The same intuition also applies to the other parameters.

Chapter 3

Counting the Costs: A Quantitative Assessment of Brexit's Effect on UK Regional Economies

3.1 Introduction

On 23 June 2016, the United Kingdom voted to leave the European Union, leading to the formation of the EU-UK Trade and Cooperation Agreement (TCA), signed on 30 December 2020. This comprehensive agreement, effective from 1 January 2021, governs free trade in goods with zero tariffs and quotas under specific rules, allows limited mutual market access for services, and establishes cooperation in areas like law enforcement and energy. It also ends the UK's participation in the European Single Market, Customs Union, and most EU programs, significantly altering UK-EU relations post-Brexit.

In this chapter, we examine the implications of the TCA using a dynamic spatial general equilibrium model. Previous studies primarily projected potential impacts of Brexit based on a potential range of scenarios, and very few have assessed the finalized deal or modeled the granular impacts across various regions and sectors of the economy. This study examines the actual changes in tariffs and non-tariff measures (NTMs) stipulated by the TCA and provides an assessment of regional outcomes and the adjustment paths in response to the Brexit shock. Based on a detailed reading of the TCA and the associated literature, the new relationship with the EU is shown to increase trade costs by 10.8 percent for UK exports to the EU and 11.0 percent for imports from the EU. These costs are projected to rise to 16.2 percent and 16.6 percent, respectively, should the EU undergo further integration in the coming years. It is also crucial to note that the increase in trade costs within the UK is not equally distributed across different regions and sectors.

In this analysis, we recognize that the full adaptation to the new trading arrangements outlined in the TCA will span several years. This extended period is attributed not only to the phased implementation of the new trading protocols but also to the time required for labor markets to adjust to these changes. Our study aims to capture the dynamic adjustment of the labor market and the gradual implementation of the TCA to quantify its effects on the UK's trade and welfare. Therefore, we employ a dynamic spatial model that incorporates labor market adjustments to understand the evolving economic landscape.

Our model incorporates forward-looking migration decisions, connects multiple sectors through an input-output structure and enables sector-level trade in line with the gravity equation. On the household side, forward-looking households select their sector and region of employment for the upcoming period and supply one unit of labor inelastically each period. Operating on a hand-to-mouth basis, households spend their entire income within the period, receiving total income from factor payoffs, tariff revenue, and trade imbalances, which are redistributed to them as lump sum transfers. On the production side, we utilize the Eaton and Kortum (2002) framework, in which a continuum of heterogeneous firms operates under competitive conditions with constant-returns-to-scale technology. These firms demand labor, local factors, and intermediate inputs from all other markets in the economy. The model encompasses both goods and services sectors, which are tradable both interregionally and internationally, and are subject to iceberg trade costs, NTMs, and tariffs.

To assess the impact of the TCA through our model, we first provide new estimates of tariffs and NTMs. Despite the TCA stipulating zero tariffs on all traded goods, the requirement for exporters to comply with Rules of Origin (ROO) can restrict some firms from accessing these tariff-free benefits. This restriction arises either because their products include significant non-EU sourced components, or because the compliance costs outweigh the benefits, particularly when Most Favoured Nation (MFN) tariffs are low. As preference utilisation rates are below 100 percent, we expect that some firms will continue paying tariffs, even in the long-run. We infer the tariff inputs by analyzing recent EU trade data, focusing on preference utilisation and tariff-free rates.

Most trade barriers introduced by the TCA are NTMs, which present challenges in quantification due to their varied and complex nature. To estimate these NTMs, the approach aligns with the UK Trade Policy Observatory's (UKTPO) assessment of the TCA. Following the methodology described by Cadot and Gourdon (2016), we estimate the ad valorem equivalents of the non-tariff barriers, differentiating between those barriers impacted by

the presence or absence of a deep regional trade agreement (RTA). We also incorporate the UKTPO's evaluation of sector-specific provisions within the TCA.

We compile a new dataset for regionalized World Input-Output Database (WIOD) to feed our model.[1] This dataset encompasses 66 regions (12 UK regions, 53 additional countries, and one rest of the world) with 44 sectors in each region for 2014. Utilizing this data, we construct sector-region level input-output linkages, trade flows, and shares of value added and consumption, which are integral for our model's quantifications. The dataset is constructed based on detailed UK trade flow data, the WIOD, and a series of proportionality assumptions that effectively distribute economic activities across regions and sectors.

We construct a sector-region level transition matrix for labor migration within the UK using data from the 2016-2021 labor force survey. This matrix captures the dynamic movement of households across different regions and sectors. The migration matrix is integrated into our model to estimate migration elasticity and migration costs, assuming these costs remain consistent with their 2021 levels throughout our analysis.

To quantify the effects of economic shocks in our model, we use the dynamic "exact-hat algebra" approach, following the methodology outlined by Caliendo et al. (2019). This approach ensures that our model exactly matches sector-region level production, trade, and reallocation patterns observed in the base year. Subsequently, we introduce the gradual evolution of tariffs and NTMs as stipulated by the TCA over specific years. In setting the evolution of these shocks, we also take into account the forgone benefits of EU integration that the UK would have enjoyed had it remained a member. This consideration is crucial as it captures the potential additional impacts arising from further integration within the EU, from which the UK is now excluded (Dhingra et al., 2017).

Our measurements of trade shocks show that trade barriers post-Brexit are set to disproportionately affect the UK's service sectors, especially those like finance and professional services that are heavily regulated. The TCA offers limited access compared to the full benefits previously enjoyed within the EU's single market. This is particularly detrimental for sectors where the UK is a major player in EU markets.

Our model quantification starts from the year 2021, using the 2014 regionalized WIOD to represent 2021 trade flow and consumption patterns. We use data from the 2021 Labor Force Survey (LFS) for the initial labor distribution and incorporate the 2021 migration transition matrix. In simulating the model, we assume that all fundamentals remain as

[1] We also use 2018 regionalized OECD Input Output table in our sensitivity check.

they were in 2021, with the only changes over time being trade shocks resulting from the implementation of the TCA. However, this quantification, while providing insights into the scale and expected path of adjustment, does not account for the disruptive impacts of other policy changes and global shocks, such as Covid-19.

Our model indicates that the TCA has led to increased trade barriers that affect various sectors of the UK economy differently. In services, especially regulated ones like finance and professional services, barriers have increased significantly, threatening to reduce exports to the EU, which form a large part of the UK's market engagement. This is due to the TCA securing far less market access than what was available under EU membership. In contrast, agriculture might see some benefits from reduced competition from EU imports, potentially offsetting losses from reduced export opportunities. However, sectors such as fishing are likely to suffer greatly due to reliance on EU markets and new trade barriers, expecting a 30% reduction in output. Manufacturing shows a mixed impact; food manufacturing might grow by 5% due to less competition and a focus on domestic markets, whereas sectors like basic metals manufacturing face declines.

Despite the significant impacts on certain sectors, the new trading relationship with the EU is not expected to prompt a large or rapid labor market adjustment across the UK economy as a whole due to limited within-country migration. Our modeling indicates that less than 0.5% of the workforce, equivalent to approximately 132,000 people, will relocate across region-sectors due to Brexit. This suggests that the UK's comparative advantage and the overall structure of the economy will not undergo the fundamental transformations some anticipated or feared. In particular, tradeable professional services, such as finance and business services, are predicted to experience the most substantial losses, with their share of gross output decreasing slightly by 0.5 percentage points to 24.6%. It is important to emphasize that this projection is about long-term trends and not immediate changes. Additionally, this projection is affected by other concurrent events like COVID-19; therefore, in the short run, the realized effect of Brexit could be different in reality. While some envisioned Brexit as a catalyst for revitalizing manufacturing, our findings suggest otherwise; the sector's share of the economy is expected to decline marginally by just 0.1 percentage point. Consequently, the broad industrial structure of the UK will remain relatively stable, continuing as a service-dominated economy with a smaller manufacturing sector compared to countries like France.

The new barriers introduced under the TCA are expected to reduce trade openness—measured as total trade as a share of GDP. While these barriers will substantially reduce the volume

of trade, they are not expected to fundamentally alter the UK's export specialization. The specialization in goods trade will likely continue to align with the UK's existing comparative advantages. Although exports from highly specialized services sectors like financial services and other business services are projected to decline more than average, these shifts in export specialization will be relatively minor, especially when compared to the gradual evolution of the UK's comparative advantage over the past decade.

Brexit's impact varies significantly across the UK, with considerable attention on its potential to help economically weaker regions 'level up.' Our model reveals that the North East, one of the UK's poorest regions, will suffer significantly, exacerbating its already substantial productivity and income disparities. In contrast, the impact on London, a major driver of regional inequality due to its economic size, remains uncertain. Although London is highly exposed to sectors heavily affected by Brexit, its ability to adapt appears more robust.

Brexit is unlikely to significantly transform the nature of the UK economy. Its effects on the industrial structure, export specialization, and regional inequalities are expected to be minimal. Instead, the broader impact of Brexit can be viewed as a general decrease in workers' pay and productivity. A less open UK is projected to be poorer and less productive by the end of the decade, with real wages expected to decrease by 2.1%, amounting to a loss of £550 per worker per year, and labor productivity anticipated to drop by 1.5%. This decline represents more than a quarter of the productivity growth achieved in the last decade.

This chapter proceeds as follows: Section 3.2 develops the dynamic spatial model. Section 3.3 describes the data and calibration. Section 3.4 provides quantification results of the model to study the effect of TCA. Section 3.5 concludes.

3.2 Model Environment

The model is based on Caliendo et al. (2019) and Caliendo and Parro (2022). To investigate the effects of the TCA, we define the bilateral trade costs in our model to include traditional iceberg transport costs, tariffs, and NTMs. The economy consists of $n \in N$ regions and $j \in J$ sectors in each region. Each region-sector combination houses a market with a continuum of perfectly competitive firms that produce intermediate goods. There is also a final goods producer in each market that exclusively uses these intermediate goods to manufacture non-tradable final goods. These intermediate goods, while freely tradable, are used solely for the production of final goods. The final goods produced are either consumed

locally or utilized as intermediate inputs by other producers of intermediate goods. Time is discrete.

3.2.1 Households

Consumption

In each region n at time t, households working in sector j maximize their instantaneous utility by choosing from a variety of consumption goods. The instantaneous utility function for a household in sector j and region n is represented by a Cobb–Douglas utility function as follows:

$$U_t^{nj} = \prod_{k=1}^{J} (c_t^{nj,k})^{\alpha^k} \tag{3.1}$$

where $c_t^{nj,k}$ denotes the consumption of good k by a household in sector j and region n at time t, and α^k are the constants representing the share of consumption for each good, such that $\sum_{k=1}^{J} \alpha^k = 1$.

Households in our model do not make investment or saving decisions; instead, they are characterized as living hand-to-mouth[2]. Consequently, at each time t, they expend all of their wage income. The budget constraint for each household is thus defined as $\sum_{k=1}^{J} p_t^k c_t^{nj,k} \leq w_t^{nj}$

where p_t^k represents the price of consumption good k at time t, and w_t^{nj} is the wage rate of households in sector j and region n at time t. Additionally, households may be non-employed, receiving a fixed consumption from home production, denoted as b^n, where $b^n > 0$. We define sector zero as non-employment such that $U_t^{n0} = b^n$.

Given the property of Cobb-Douglas (CD) preferences, the indirect utility function is given by:

$$u_t^{nj} = \begin{cases} b^n & \text{if } j = 0, \\ \frac{w_t^{nj}}{P_t^n} & \text{otherwise,} \end{cases} \tag{3.2}$$

where the price index is defined as

$$P_t^n = \Pi_{k=1}^{J} (P_t^{nk}/\alpha^k)^{\alpha^k} \tag{3.3}$$

[2]The welfare results should therefore be interpreted as consumption that has been smoothed over the longer horizon by households before and after Brexit.

Migration

After supplying labor and spending their income on consumption in each period t, workers observe idiosyncratic mobility shocks ϵ_t^{ik}, and make decisions regarding their movement. The value function for a worker in location i and sector j at time t (v_t^{nj}) is comprised of the current flow of utility from that location, alongside the expected continuation value from making an optimal location choice in the future:

$$v_t^{nj} = \ln(u_t^{nj}) + \max_{i \in N, k \in J} \left\{ \beta \mathbb{E}[v_{t+1}^{ik}] - m^{nj,ik} + \nu \epsilon_t^{ik} \right\} \tag{3.4}$$

where $\ln(u_t^{nj})$ denotes the logarithm of the utility derived from consumption for a worker in location i and sector j at time t, β is the discount factor, $\mathbb{E}$ represents the expectations over future idiosyncratic shocks, $m^{nj,ik}$ is the cost of moving from location-sector pair nj to ik, and $\nu \epsilon_t^{ik}$ captures the idiosyncratic preference shock for moving to location i in sector k at time t.

Assuming that the idiosyncratic shock ϵ_t^{ik} is i.i.d. over time and follows a Type-I Extreme Value distribution with a zero mean, the cumulative distribution function (CDF) is given by $F(\epsilon) = e^{-e^{(-\epsilon - \bar{\gamma})}}$, where $\bar{\gamma}$ is the Euler–Mascheroni constant.

Define $V_{t+1}^{ik} \equiv \mathbb{E}[v_{t+1}^{ik}]$. The share of labor that transitions from market nj to ik is given by the following expression:

$$\mu_t^{nj,ik} = \frac{\exp\left((\beta V_{t+1}^{ik} - m^{nj,ik})^{1/\nu} \right)}{\sum_{i'=1}^{N} \sum_{k'=0}^{J} \exp\left((\beta V_{t+1}^{i'k'} - m^{nj,i'k'})^{1/\nu} \right)} \tag{3.5}$$

Given the properties of the Type-I Extreme Value distribution, the expected utility is given by:

$$V_t^{nj} = \ln(u_t^{nj}) + \nu \log \left(\sum_{i=1}^{N} \sum_{k=0}^{J} \exp\left((\beta V_{t+1}^{ik} - m^{nj,ik})^{1/\nu} \right) \right) \tag{3.6}$$

3.2.2 Production

In each region-sector market (n, j), there exists a continuum of perfectly competitive goods producers. The production of each good utilizes labor, structures, and materials from all sectors as inputs. Labor is imperfectly mobile across regions and sectors, while structures are in fixed supply, denoted as H^{nj}, in each market (n, j). The production technology is characterized by a Cobb–Douglas function:

$$q_t^{nj} = z^{nj} \left(A_t^{nj} \left(h_t^{nj} \right)^{\xi^n} \left(l_t^{nj} \right)^{1-\xi^n} \right)^{\gamma^{nj}} \prod_{k=1}^{J} \left(M_t^{nj,nk} \right)^{\gamma^{nj,nk}} \tag{3.7}$$

where q_t^{nj} is the quantity of intermediate goods produced in market (n, j) at time t, z^{nj} is the idiosyncratic productivity parameter, A_t^{nj} is the average productivity specific to the region-sector market, h_t^{nj} represents the structured inputs, l_t^{nj} is the labor input, and $M_t^{nj,nk}$ denotes the quantity of materials used from sector k in region n at time t. Parameters γ^{nj} and $\gamma^{nj,nk}$ represent the share of value-added and the share of material input cost in the total output, respectively. The parameter ξ^n denotes the share of structure in value-added. We assume the production function has constant returns to scale, which is represented by $\sum_{k=1}^{J} \gamma^{nj,nk} + \gamma^{nj} = 1$. In a perfectly competitive market, this condition implies that firms earn zero profit. The optimal decision-making process leads to the following expression for the unit cost:

$$x_t^{nj} = B^{nj} \left(\left(r_t^{nj} \right)^{\xi^n} \left(w_t^{nj} \right)^{1-\xi^n} \right)^{\gamma^{nj}} \prod_{k=1}^{J} \left(P_t^{nj,nk} \right)^{\gamma^{nj,nk}} \tag{3.8}$$

To accommodate the large observed regional trade imbalances, we assume a fixed population (mass 1) of rentiers in each region, who are immobile and cannot relocate across regions. These rentiers own the structures within their respective regions, rent these structures to local firms, and then redirect all their rental income to a global investment portfolio. Each rentier receives a constant share ι^n from this global portfolio, where $\sum_{n=1}^{N} \iota^n = 1$, ensuring that the sum of all regional shares equals one. The total income χ collected from rents across all regions and sectors is given by $\chi = \sum_{i=1}^{N} \sum_{k=1}^{J} r_t^{ik} H^{ik}$, Each rentier spends their income share from the global portfolio exclusively on goods produced within her own region, hence the expenditure of rentiers in region n is $\iota^n \chi$. We calibrate ι to match the initial trade imbalance and assume this share is constant across all years.

Trade

Trade within the model incurs costs that take the form of iceberg costs. These include tariffs τ, transportation costs, and NTMs denoted by κ. For each unit of intermediate goods from sector j shipped from region i to region n, $\kappa_t^{nj,ij} \tau_t^{nj,ij}$ units of the good are required to produce one unit in the destination region. Under the assumption of perfect

competition, the price paid for a particular variety of goods is determined by the minimum unit cost across all regions. This can be mathematically expressed as:

$$p_t^{nj}(z^j) = \min_i \left\{ \frac{\kappa_t^{nj,ij} \tau_t^{nj,ij} x_t^{ij}}{z^{ij} (A_t^{ij})^{\gamma^{ij}}} \right\}$$

Following Eaton and Kortum (2002), we assume that the productivity term z^{nj} follows a Fréchet distribution with the cumulative distribution function (CDF) given by $F(z^{nj}) = \exp\left\{ - \left(z^{nj}\right)^{-\theta^j} \right\}$.

The bilateral trade share $\pi_t^{nj,ij}$ can then be expressed as follows:

$$\pi_t^{nj,ij} = \frac{\left(\kappa_t^{nj,ij} \tau_t^{nj,ij} x_t^{ij}\right)^{-\theta^j} \left(A_t^{ij}\right)^{\theta^j \gamma^{ij}}}{\sum_{m=1}^{N} \left(k_t^{nj,mj} \tau_t^{nj,mj} x_t^{mj}\right)^{-\theta^j} \left(A_t^{mj}\right)^{\theta^j \gamma^{mj}}} \tag{3.9}$$

Given the properties of the Fréchet distribution, the price index is determined as:

$$P_t^{nj} = \Gamma^{nj} \left(\sum_{i=1}^{N} \left(\tau_t^{nj,ij} \kappa_t^{nj,ij} x_t^{ij}\right)^{-\theta^j} \left(A_t^{ij}\right)^{\theta^j \gamma^{ij}} \right)^{-1/\theta^j} \tag{3.10}$$

3.2.3 Market Clearing

Goods market clearing in the model stipulates that the total output X_t^{nj} in each sector j in region n at time t must equal the total demand for that output. This total demand includes its use as an intermediate input across all sectors in all regions and its use in final consumption:

$$X_t^{nj} = \sum_{k=1}^{J} \gamma^{nj,nk} \sum_{i=1}^{N} \pi_t^{ik,nk} X_t^{ik} + \alpha^j I_t^{nj} \tag{3.11}$$

The total income I_t^{nj} in region n for sector j is defined as the sum of wages, tariff transfers, and rentier income from the global portfolio:

$$I_t^{nj} = \sum_{k=1}^{J} w_t^{nk} L_t^{nk} + TR_t^{nj} + \iota^n \chi_t \tag{3.12}$$

where w_t^{nk} represents the wage rate in sector k of region n, L_t^{nk} is the corresponding labor

employed, TR_t^{nj} denotes the tariff revenue, and $\iota^n \chi_t$ is the rentier income from the global portfolio. The tariff revenue TR_t^{nj} is given by $TR_t^{nj} = \sum_{j=1}^{N} \sum_{k=1}^{J} \frac{(\tau_t^{jk,nj}-1)(\pi_t^{jk,nj} X_t^{jk})}{\tau_t^{jk,nj}}$

From the first order conditions of the firm's problem, a fixed fraction of expenditure is spent on labor costs and structures. Therefore, the labor market clearing condition can be expressed as:

$$L_t^{nj} = \frac{\gamma^{nj}(1-\xi^n)}{w_t^{nj}} \sum_{i=1}^{N} \pi_t^{ik,nk} \frac{X_t^{ij}}{\tau_t^{ik,nk}} \tag{3.13}$$

Similarly, the structure market clearing condition is given by:

$$H^{nj} = \frac{\gamma^{nj}\xi^n}{r_t^{nj}} \sum_{i=1}^{N} \pi_t^{ik,nk} \frac{X_t^{ij}}{\tau_t^{ik,nk}} \tag{3.14}$$

Finally, we account for endogenous trade imbalances where the trade deficit for region i at time t, denoted D_t^i, is defined as the difference between the income share from the global portfolio $\iota^i \chi_t$ and the total rents paid for the global portfolio $D_t^i = \iota^i \chi_t - \sum_{k=1}^{J} r_t^{ik} H^{ik}$

Accordingly, the trade balance condition, which ensures that the total imports equal total exports plus the trade deficit, is given by:

$$\sum_{k=1}^{J} \sum_{n=1}^{N} \frac{\pi_t^{ik,nk} X_t^{ik}}{\tau_t^{ik,nk}} = \sum_{k=1}^{J} \sum_{j=1}^{N} \frac{\pi_t^{jk,ik} X_t^{jk}}{\tau_t^{jk,ik}} + D_t^i \tag{3.15}$$

This condition essentially mirrors the labor market clearing condition. To see this, sum over sectors for the goods market clearing condition (3.11) and subtract it from the labor market clearing condition (3.13), which yields the trade balance condition as formulated above.

3.2.4 General Equilibrium

Definition 1: *Given the initial allocation of labor L_0^{nj}, a fixed supply of structures H^{nj}, and an exogenous evolution of fundamentals over time represented by $\{A_t^{nj}, \kappa_t^{nj,ij}, \tau_t^{nj,ij}, m_t^{nj,ik}\}_{t=0}^{\infty}$, along with a set of parameter values $\{\alpha^j, \gamma^{nj}, \gamma^{nj,nk}, \xi^n, \iota^n\}$, the general equilibrium of the economy is defined by a vector of allocations and prices $\{l_t^{nj}, w_t^{nj}, r_t^{nj}, v_t^{nj}\}_{t=0}^{\infty}$ such that the following conditions are satisfied:*

- The **indirect utility function** is given by consumer utility maximization, as specified in equations (3.2) and (3.3).

- The **law of motion for labor** describes how labor supply in market (n, j) at time $t+1$ evolves according to the migration shares and expected utility, captured by:

$$L_{t+1}^{nj} = \sum_{i=1}^{N} \sum_{k=0}^{J} \mu_t^{ik,nj} L_t^{ik} \tag{3.16}$$

  where migration share and expected utility calculations are detailed in equations (3.5) and (3.6).
- **Trade share** is given by equation (3.9), where price index and unit cost is given by (3.10) and (3.8).
- The **goods market clearing condition** is satisfied in each period, as outlined in equation (3.11). **Total income** for each sector and region is computed as defined in equation (3.12).
- The **labor market clearing condition** is satisfied in each period, as described in equation (3.13).
- The **structure market clearing condition** is satisfied in each period, as described in equation (3.14).

To characterise the existence and uniqueness of the general equilibrium, firstly, given the evolution of labor, the general equilibrium in each period is determined as in a standard static international trade model as in Caliendo and Parro (2015). There exists a unique vector of price (up to scale) in each period that is determined by equilibrium condition (3.8) (3.9) (3.10) (3.11) (3.12) (3.13) (3.14). Then Between periods, the dynamics of the population distribution are determined by the law of motion for labor (3.16) (3.5) and (3.6).[3]

To solve the model and apply it to data for our policy evaluation, we utilize the dynamic exact-hat algebra method as discussed in Caliendo, Dvorkin and Parro (2019) and Kleinman et al. (2023). This method involves representing the change in any scalar or vector between two consecutive periods, t and $t+1$, using the notation $\dot{y}_{t+1} \equiv \frac{y_{t+1}}{y_t}$. We then restate our general equilibrium as follows in terms of changes:

Definition 2: The dynamic exact-hat algebra: *Given an initial observed allocation of the economy, including trade shares* $\pi_0^{nj,ik}$*, initial outputs* X_0^{ij}*, labor allocations* L_0^{nk}*, migration matrix* $\mu_{-1}^{nj,ik}$*, initial trade deficits* D_0^i*, and a convergent sequence of changes in economic fundamentals over time* $\{\dot{A}_t^{nj}, \dot{\kappa}_t^{nj,ij}, \dot{\tau}_t^{nj,ij}, \dot{m}_t^{nj,ik}\}_{t=1}^{\infty}$*, the solution for the sequence*

[3] Allen, Arkolakis and Li (2023) provide a sufficient condition for the existence of a unique general equilibrium of the forward-looking migration model.

of changes in the model's endogenous variables $\{\dot{l}_t^{nj}, \dot{w}_t^{nj}, \dot{r}_t^{nj}, \dot{v}_t^{nj}\}_{t=1}^{\infty}$ *does not require information on the level of fundamentals, and is determined by the following equations:*

$$\dot{x}_{t+1}^{nj} = (\dot{L}_{t+1}^{nj})^{\gamma^{nj}\xi^n}(\dot{w}_{t+1}^{nj})^{\gamma^{nj}}\Pi_{k=1}^{J}(\dot{P}_{t+1}^{nk})^{\gamma^{nj,nk}} \tag{3.17}$$

$$\dot{P}_t^{nj} = (\Sigma_{i=1}^{N}\pi_t^{nj,nk}(\dot{x}_{t+1}^{ij}\dot{\tau}_{t+1}^{nj,ij}\dot{k}_{t+1}^{nj,ij})^{-\theta^j}(\dot{A}_{t+1}^{ij})^{\theta^j\gamma^{ij}})^{-1/\theta^j} \tag{3.18}$$

$$\pi_{t+1}^{nj,nk} = \pi_t^{nj,ik}(\frac{\dot{x}_{t+1}^{ij}\dot{\tau}_{t+1}^{nj,ij}\dot{k}_{t+1}^{nj,ij}}{\dot{P}_{t+1}^{nk}})^{-\theta^j}(\dot{A}_{t+1}^{ij})^{\theta^j\gamma^{ij}} \tag{3.19}$$

$$X_{t+1}^{nj} = \Sigma_{k=1}^{J}\gamma^{nj,nk}\Sigma_{i=1}^{N}\pi_t^{ik,nk}\frac{X_t^{ik}}{\tau_t^{ik,nk}} + \alpha^j(\Sigma_{k=1}^{J}\dot{w}_{t+1}^{nk}\dot{L}_{t+1}^{nk}w_t^{nk}L_t^{nk} + TR_{t+1}^{nj} + \iota^n\chi_{t+1}) \tag{3.20}$$

$$\dot{w}_{t+1}^{nk}\dot{L}_{t+1}^{nk}w_t^{nk}L_t^{nk} = \gamma^{nj}(1-\xi^n)\Sigma_{i=1}^{N}\pi_{t+1}^{ij,nk}\frac{X_{t+1}^{ij}}{\tau_t^{ik,nk}} \tag{3.21}$$

$$\Sigma_{k=1}^{J}\Sigma_{n=1}^{N}\frac{\pi_{t+1}^{ik,nk}X_{t+1}^{ik}}{\tau_{t+1}^{ik,nk}} = \Sigma_{k=1}^{J}\Sigma_{j=1}^{N}\frac{\pi_{t+1}^{jk,ik}X_{t+1}^{jk}}{\tau_{t+1}^{jk,ik}} + D_{t+1}^{i} \tag{3.22}$$

where $\chi_{t+1} = \Sigma_{i=1}^{N}\Sigma_{k=1}^{J}\frac{\xi^i}{1-\xi^i}\dot{w}_{t+1}^{ik}\dot{L}_{t+1}^{ik}w_t^{ik}L_t^{ik}, TR_{t+1}^{nj} = \Sigma_{j=1}^{N}\Sigma_{k=1}^{J}(\tau_t^{jk,nj}-1)\frac{\pi_t^{jk,nj}X_t^{jk}}{\tau_t^{jk,nj}}$

$$\mu_t^{nj,ik} = \frac{\mu_t^{nj,jk}(\dot{v}_{t+2}^{ik})^{\beta/\nu}}{\Sigma_{m=1}^{N}\Sigma_{h=0}^{J}\mu_t^{nj,mh}(\dot{v}_{t+2}^{mh})^{\beta/\nu}} \tag{3.23}$$

$$\dot{v}_{t+1}^{nj} = \frac{\dot{w}_{t+1}^{nj}}{\dot{P}_{t+1}^{n}}(\Sigma_{i=1}^{N}\Sigma_{k=0}^{J}\mu_t^{nj,ik}(\dot{v}_{t+2}^{ik})^{\beta/\nu})^{\nu} \tag{3.24}$$

$$L_{t+1}^{nj} = \Sigma_{i=1}^{N}\Sigma_{k=0}^{J}\mu_t^{ik,nj}L_t^{ik} \tag{3.25}$$

We employ the numerical algorithm as proposed by Caliendo et al. (2019) to solve the general equilibrium model. The computational process begins with an initial guess of the value function path for all sectors, regions, and periods, denoted by $\dot{v}_{nj,t}^0$ for $t \in [1, \ldots, T+1]$. We impose $\dot{v}_{nj,T+1} = 1$ at the final period, meaning that the economy reaches a steady state after a sufficiently long time period. Starting with this guess, and given the initial labor allocations L_0^{nk} along with the migration matrix $\mu_{-1}^{nj,ik}$, we proceed to solve for the path of all future labor allocations $(L_t^{nk})_{t=1}^{\infty}$ from equation (3.23) and (3.25). Given the path of labor supply, we proceed to determine the equilibrium prices and wages for each period t by solving equations (3.17) through (3.22). This set of equations systematically computes the prices and wages based on the prevailing economic conditions and labor distribution. We calculate the next iteration of the value function path, $\dot{v}_{nj,t}^1$, using equation (3.24). This calculation provides updated estimates for the value function

that reflect the newly determined economic variables. Ultimately, if the newly computed path for changes in values differs from the initial estimates, the model employs an updating rule to compute the system until the model converges. Appendix 3.C provides the detailed algorithm.

3.3 Data and Calibration

In this section, we outline the data required for solving the model and conducting counterfactual analyses as described in Definition 2. The dataset includes the initial observed allocation of the economy, encompassing trade shares $\pi_t^{nj,ik}$, initial outputs X_0^{ij}, labor allocations L_0^{nk}, migration matrices $\mu_{-1}^{nj,ik}$, and initial trade deficits D_0^i. To address counterfactual questions, we also require a convergent sequence of changes in economic fundamentals over time, denoted by $\{\dot{A}_t^{nj}, \dot{\kappa}_t^{nj,ij}, \dot{\tau}_t^{nj,ij}, \dot{m}_t^{nj,ik}\}_{t=1}^{\infty}$. We assume that all other fundamentals remain the same and only the trade shock changes.[4] Additionally, calibration of model parameters is essential, including the share of value added in gross output γ^{nj}, the share of intermediate inputs $\gamma^{nj,nk}$, labor compensation share in value added ξ^n, sectoral trade elasticities θ^j, consumption shares α^j, global portfolio shares ι^n, the discount factor β, and the migration elasticity $\frac{1}{\nu}$.

3.3.1 Data

Our model incorporates 12 UK regions,[5] 53 additional countries, and one "rest of the world" category. It features 19 industry sectors along with one additional sector designated for unemployment. Specifically, these include 1 agriculture sector, 1 manufacturing sector, 17 service sectors, and mining. This sectoral configuration is essential for capturing the major changes in the NTM for service sectors due to the TCA.[6] Each of the 12 UK regions[7] in our model features a distinct labor market for every sector, facilitating labor mobility within these regions across different sectors. However, labor mobility across countries is

[4] After 2019, the outbreak of COVID-19 also represented a significant shock that could affect productivity and migration. However, in the current model, we are not considering the COVID-19 shock due to a lack of data. Therefore, the results should be interpreted solely as the effect of Brexit in the absence of all other shocks.

[5] Specifically, North East, North West, Yorkshire and the Humber, East Midlands, West Midlands, East, London, South East, South West, Wales, Scotland, and Northern Ireland

[6] Sectors are aggregated at different levels in various data sources. For example, the Labour Flow Survey comprises 22 sectors, while the WIOT encompasses 44 sectors. We merge them, resulting in the final count of 19 sectors.

[7] We are not using travel-to-work areas as most data are available at the 12 UK administrative areas level, especially bilateral trade data. For understanding regional development, such as structural change, the UK administrative areas provide a good starting point.

not assumed. For the other countries in our model, we assume the existence of a single labor market for each sector in each country.

Bilateral flows: The model requires bilateral trade flows across all sectors between each pair of regions in the sample. We compile these flows by regionalizing the 2014 WIOT. Firstly, we use the Multiregional Input-Output Tables for Europe (EUREGIO), which provide detailed input-output tables for all EU regions and major countries worldwide as of 2010.[8] We then convert the EUREGIO tables to a format where European regions are aggregated to the country level, retaining only the UK at the regional level. Next, we use EUREGIO values to establish regional shares; for example, calculating the North East's share of the UK's total for each cell. Finally, we create a WIOT table that includes UK regions in place of UK totals by multiplying the UK shares from the EUREGIO dataset with WIOT 2014 UK values, assuming that these shares remained constant from 2010 to 2014.

Migration matrix: Labor force data is available at the sector-region level for 20 sectors and one unemployment sector across the years 2009-2015 and 2016-2021 from the Labour Force Survey (LFS). Let $M^{(nk,ij)}$ represent the annual transition matrix, indicating the share of labor transitioning from nk to ij. From the data, we have the transition matrices for the periods 2009-2015 ($M^{(nk,ij)}_{2009-2015}$) and 2016-2021 ($M^{(nk,ij)}_{2016-2021}$). We assume the transition matrix remains constant each year and calculate the annual migration matrices to match the total migration ratio from 2009-2015.

TCA tariffs: While many assessments have assumed that tariffs would remain at zero in the event of avoiding a 'no deal' Brexit, firms now encounter higher regulatory barriers to access preferential tariffs, particularly due to rules of origin requirements. Consequently, some analyses have increased their estimates of non-tariff barriers. Trade data shows that many UK firms are actually paying tariffs to export to the EU, either because they cannot meet the requirements or because the cost of compliance outweighs the benefits of tariff exemption. This chapter incorporates the preference utilisation rate to account for these paid tariffs, resulting in an additional 4.5 percentage points for the estimated non-tariff barriers in sectors such as textile manufacturing. It assumes that the costs associated with complying with rules of origin requirements for those using preferential tariffs are included in the non-tariff estimates. The details of calculating the tariff are discussed in the appendix 3.A.

[8]The EUREGIO data is publicly available and can be accessed online. European Commission, Joint Research Centre (JRC) (2020).

Services trade barriers: Although the provisions in the TCA are comparable with past EU trade agreements considered to be deep on services, such agreements have not been successful at securing substantial services market access. Reflecting this, the TCA only secures a very small share of the services market access the UK enjoyed in the single market (between 80 percent and 100 percent of their levels in the absence of a deal). As with other trade agreements covering services, provisions in the TCA typically lock in the liberalisation already applied to non-EU trading partners rather than securing additional liberalisation, for example, equivalence agreements that would replicate the existing market access such as replacing EU passporting rights for financial services.

In this chapter, these NTMs are estimated in line with Fusacchia et al. (2022)'s assessment of the Trade and Cooperation Agreement. This approach is based on Cadot and Gourdon (2016), which estimates ad valorem equivalents of the non-tariff barriers, differentiating between barriers with and without a deep regional trade agreement (RTA). To estimate these NTMs, we use data from the OECD Services Trade Restrictiveness Index (STRI), Department for International Trade assessments of UK-Canada CETA and UK-Japan EPA, and Developing Trade Consultants' ad valorem equivalent estimates of MFN and EEA trade barriers. We provide further details on estimating the NTM for all sectors in the appendix 3.A.

EU integration: Over time, the EU has deepened integration by introducing further measures to align regulatory standards and practices across member states. This is expected to continue over the next decade, but as the UK will be outside the EU, it will not benefit from these lower barriers. Our assumption on further integration is aligned with past analysis by Dhingra et al. (2017), which assumes further EU integration will be equivalent to a 5.6 percent tariff on all trade in their optimistic 'soft' Brexit scenario of how UK-EU trade will evolve. However, instead of a flat 5.6 percent ad valorem equivalent barrier applied across sectors, this chapter assumes that EU integration varies across sectors to account for higher expected future integration in sectors the EU has successfully liberalised in the past. For example, it is assumed to be higher for professional services and lower in sectors where it has been less successful, such as utilities.

TCA shock timeline: The dynamic nature of the model allows for delayed implementation of certain barriers. To reflect this, 25 percent of the goods barriers introduced on EU imports to the UK are delayed to account for the implementation period for customs checks, and 50 percent of the financial services barriers on imports from the EU are delayed until 2024 to account for the temporary permissions regime, which allows financial services

firms already operating in the UK to continue doing so. No changes have been modeled to non-EU trade barriers. This means we assume the rolled-over trade agreements were successful in maintaining market access to these markets, while new agreements signed with Australia and New Zealand, and those being negotiated, for example with India, are not assessed.

3.3.2 Calibration

The dynamic exact hat algebra in the chapter requires only the initial allocation of economic distribution and a path for changes in fundamentals. We assume that the only changes in fundamentals come from the TCA shock, while all other fundamentals remain constant over time, and there is no need to calibrate these fundamentals in level.

To fully solve the model, we need to calibrate the following parameters: the share of value added in gross output (γ^{nj}), the share of intermediate inputs ($\gamma^{nj,nk}$), labor compensation share in value added (ξ^n), sectoral trade elasticities (θ^j), consumption shares (α^j), global portfolio shares (ι^n), the discount factor (β), and the migration elasticity ($\frac{1}{\nu}$).

For the share of value added in gross output (γ^{nj}), the share of intermediate inputs ($\gamma^{nj,nk}$), labor compensation share in value added (ξ^n), consumption shares (α^j), and global portfolio shares (ι^n), we can directly compute these shares given observed gross output in each sector, intermediate input at the sector-region level from regional WIOT, and labor compensation data in each region-sector in 2014. We assume all these parameters remain the same over the period of our simulation and take values as they were in 2014. The discount factor (β) is set at 0.99 as in Caliendo et al. (2019).

To estimate the migration elasticity $\frac{1}{\nu}$, we follow the procedure in Caliendo et al. (2019) which is based on Artuç et al. (2010). The model implies the following migration equation between periods:

$$\log\left(\frac{\mu_t^{nj,nk}}{\mu_t^{nj,nj}}\right) = \tilde{C} + \frac{\beta}{\nu}\log\left(\frac{w_{t+1}^{nk}}{w_{t+1}^{nj}}\right) + \beta\log\left(\frac{\mu_{t+1}^{nj,nk}}{\mu_{t+1}^{nk,nk}}\right) + \epsilon_{t+1},$$

where ϵ_{t+1} is a random error term and $\tilde{C}$ is a constant. We estimated this equation using GMM, instrumenting with past values of flows and wages. This yields an estimate of $\nu = 5.84$, representing the inverse migration elasticity. This value is higher than the estimate by Caliendo et al. (2019), who found an annual elasticity of $\nu = 2.02$ for the US.

We estimate the sectoral trade elasticities (θ^j) using the same method as in Caliendo and

Parro (2015). The model equation (3.9) implies that for any three countries and any sector j, the following relationship holds:

$$\frac{X_{ni}^{j}X_{ih}^{j}X_{hn}^{j}}{X_{nh}^{j}X_{hi}^{j}X_{in}^{j}} = \left(\frac{\tilde{\kappa}_{ni}^{j}\tilde{\kappa}_{ih}^{j}\tilde{\kappa}_{hn}^{j}}{\tilde{\kappa}_{in}^{j}\tilde{\kappa}_{hi}^{j}\tilde{\kappa}_{nh}^{j}}\right)^{-\theta^{j}}$$

where $\tilde{\kappa}_{ni}^{j} \equiv \kappa_{t}^{nj,ij}\tau_{t}^{nj,ij}$. This gives the following reduced form relationship:

$$\ln\left(\frac{X_{ni}^{j}X_{ih}^{j}X_{hn}^{j}}{X_{nh}^{j}X_{hi}^{j}X_{in}^{j}}\right) = -\theta^{j}\ln\left(\frac{\tilde{\kappa}_{ni}^{j}\tilde{\kappa}_{ih}^{j}\tilde{\kappa}_{hn}^{j}}{\tilde{\kappa}_{in}^{j}\tilde{\kappa}_{hi}^{j}\tilde{\kappa}_{nh}^{j}}\right) + \varepsilon_{nih}^{j} \tag{3.26}$$

where ε_{nih}^{j} is the random error term and is assumed to be orthogonal to tariffs and NTMs.

We estimate equation (3.26) by OLS. We use 2021 bilateral trade data on goods and services, which is the first year after the implementation of TCA, from the Bilateral Trade in Goods by Industry and End-use Category (BTDIxE) from OECD (2021b) and the Balanced Trade in Services dataset (BaTIS) from OECD (2021a). In estimating equation (3.26), we use our measured NTMs on both goods and services as $\kappa_{t}^{nj,ij}$ and estimated goods tariffs as $\tau_{t}^{nj,ij}$. We estimate θ^{j} sector by sector and Table 3.1 presents the estimation results for sectors where data are available and compares them with Caliendo and Parro (2015).

Since Brexit affects trade between the UK and all EU countries, this provides more observations with non-zero relative tariff pairs in our estimation compared to Caliendo and Parro (2015). Additionally, since we have non-tariff measures (NTMs) for not only goods but also services, using data on service trade allows our estimates to include NTMs for the services sector, providing elasticity estimates for each service sector. Our estimates are broadly in line with the literature[9] but still show significant differences compared to Caliendo and Parro (2015) at the sector level. For any sectors where data are not available, we replace them with the mean elasticity.

3.4 Quantification

In this section, we quantify the dynamic effects of Brexit on the UK's regional economy by comparing two distinct scenarios. First, we simulate a scenario where the UK retains its EU membership, maintaining all economic fundamentals as they were in the base year of

[9]Hillberry et al. (2005) document an average elasticity of 17. Broda and Weinstein (2006) find an average elasticity of 17 (seven-digit TSUSA), 12 (ten-digit HTS), and 7 (three-digit TSUSA). Yi (2003) suggests a value of 15 for Armington-type models. Romalis (2007) finds elasticity between 4 and 13.

Sector	θ^j	s.e.	N	adj R^2	CP2015
Crop production	2.98	(2.13)	8430	0.52	8.11
Fishing aquaculture	11.27	(4.70)	2416	0.45	8.11
Mining quarrying	19.46	(5.82)	5714	0.45	15.72
Food products	0.86	(1.85)	11802	0.65	2.55
Textiles apparel	3.71	(2.68)	12442	0.76	5.56
Wood products	17.37	(5.52)	7130	0.67	10.83
Paper products	70.66	(16.19)	7654	0.66	9.07
Printing media	48.70	(18.42)	4958	0.48	9.07
Coke petroleum	11.62	(7.94)	4484	0.34	51.08
Chemical products	6.75	(3.64)	10554	0.67	4.75
Pharmaceuticals	18.27	(7.04)	6562	0.61	4.75
Rubber plastic	13.75	(4.52)	10158	0.75	1.66
Non-metal minerals	9.56	(3.62)	8442	0.72	2.76
Basic metals	9.22	(6.58)	7254	0.50	7.99
Fabricated metals	2.10	(3.13)	10142	0.74	7.99
Electronic products	17.31	(3.23)	12722	0.82	10.6
Electrical equipment	1.68	(2.00)	11234	0.77	10.6
Machinery equipment	21.50	(3.35)	11290	0.78	1.52
Other transport	5.71	(1.84)	7234	0.65	0.37
Furniture	29.16	(3.47)	11694	0.75	5
Maintenance repair	8.67	(1.22)	15316	0.85	5
Transport	12.79	(1.09)	20185	0.86	5
Travel	8.70	(1.12)	19929	0.82	5
Construction	7.47	(2.93)	14724	0.77	5
Insurance pension	10.48	(2.07)	17786	0.86	5
Financial services	5.71	(0.71)	18956	0.90	5
Intellectual property	17.57	(1.45)	15901	0.92	5
Telecommunications	17.64	(3.94)	19831	0.91	5
Business services	10.54	(1.08)	19486	0.89	5
Recreational services	11.24	(1.09)	16840	0.87	5
Government services	29.91	(1.96)	17991	0.80	5

Table 3.1: Estimation Results for Each Sector

2021. This hypothetical scenario serves to illustrate the economic conditions that would have persisted had the UK decided against leaving the EU.

In contrast, our second scenario incorporates the changes in tariffs and NTMs that have arisen as a consequence of the TCA implementation, reflecting the new economic realities post-Brexit. By comparing these scenarios, our analysis aims to pinpoint the specific impacts on UK regional markets resulting from the TCA implementation.

Additionally, over the past decades, the EU has progressively deepened integration by harmonizing regulatory standards and practices across member states. This trend is expected

to continue over the next decade. However, as the UK has exited the EU, it will not partake in the benefits derived from these reduced barriers. In our analysis, assumptions about further EU integration align with projections made by Dhingra et al. (2017), which suggests that further integration within the EU could equate to effectively lowering trade barriers by an equivalent of 5.6 percent tariffs on all trade.

Unlike Dhingra et al. (2017), which assumes a uniform 5.6 percent ad valorem equivalent barrier reduction across all sectors, our analysis adopts a different approach. We contend that the degree of integration will vary significantly across different sectors, mirroring the EU's historical patterns of liberalization success. For example, we anticipate greater integration in sectors such as professional services, where the EU has already achieved notable liberalization successes. In contrast, in sectors like utilities, which have seen less effective barrier reduction, we assume a correspondingly lower level of future integration.

The core scenario used throughout this section includes EU integration, but where relevant, the results without the forgone EU integration are shown to demonstrate the direct impact of the new barriers introduced by the implementation of the TCA relative to the longer-term losses from forgone integration.

3.4.1 Sectoral Trade and Output

The overall significant shock to trade resulting from Brexit is not uniformly distributed across sectors, as detailed in Figure 3.1 and Figure 3.2. These figures illustrate the difference in aggregate trade value between the scenario where the UK leaves the EU and the baseline scenario in which the UK remains within the EU.

Certain sectors, such as agriculture, are expected to experience substantial shocks to both imports and exports. In contrast, other sectors exhibit more asymmetric shock profiles. For instance, in the financial services sector, imports are projected to decline by 64 percent, marking the second-largest drop across all sectors. However, exports in this sector are expected to decrease by only 26 percent, which is slightly less than the average reduction in exports.

Moreover, the manufacturing sector, which represents approximately 40 percent of UK exports, faces below-average shocks. Despite these relatively moderate reductions, UK exports in manufacturing are anticipated to suffer more significantly compared to imports, underscoring the sector-specific vulnerabilities and the complex nature of trade adjustments post-Brexit.

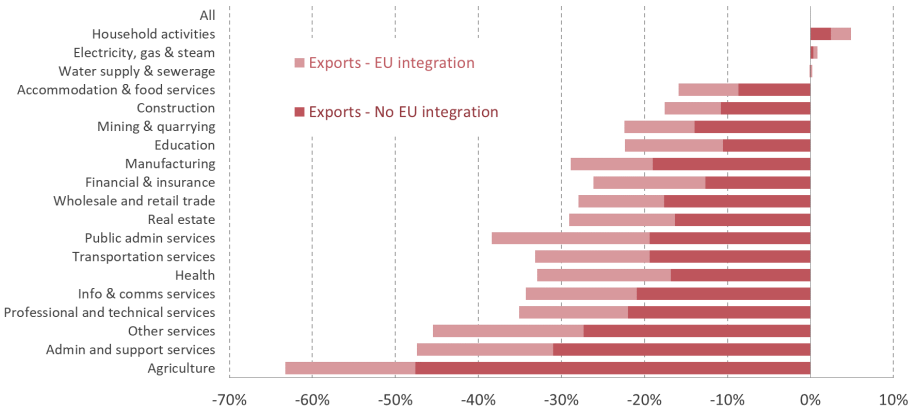


Figure 3.1: Percentage Change in UK Exports With and Without Further EU Integration Relative to Remaining in the EU

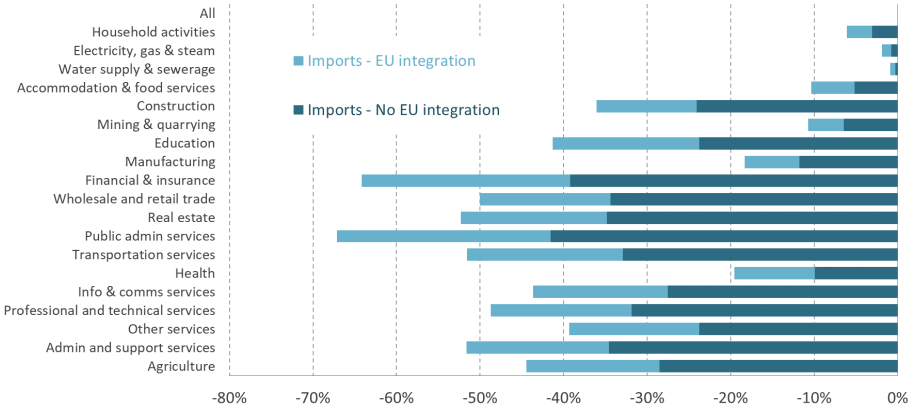


Figure 3.2: Percentage Change in UK Imports With and Without Further EU Integration Relative to Remaining in the EU

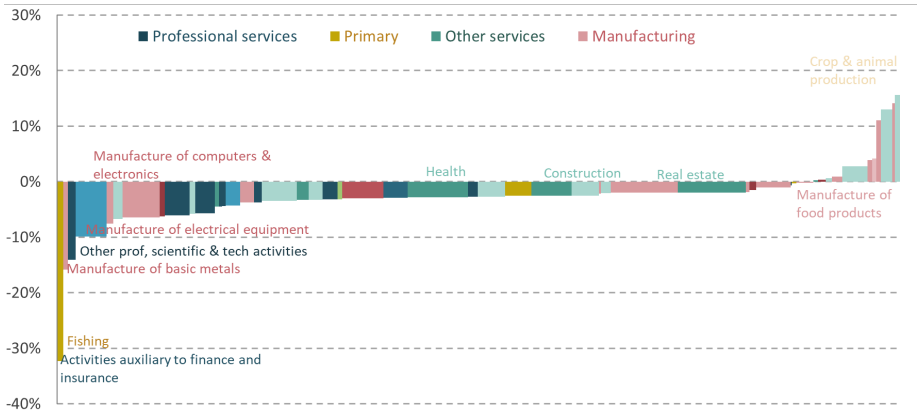


Figure 3.3: Estimated Change in Gross Output due to Brexit by Sector in 2030: UK

Trade shocks resulting from the TCA have diverse effects on sectoral output—decreased exports may hinder a sector's economic performance, while reduced imports might decrease import competition. Nonetheless, diminished imports also influence the pricing of inputs, thus carrying wider economic repercussions. Given that the regionalized WIOD offers more detailed sectoral data than is available for migration in the dynamic model, we also implemented a static version of our model. This model utilizes solely the regionalized WIOD data and assumes no migration, enabling a more granular analysis of sectoral impacts. Due to the low elasticity of migration, this static approach yields results similar to the dynamic model in the short term. The outcomes of this detailed sector-specific analysis are presented in Figure 3.3, which illustrates the variation in output across different subsectors.

Professional services, including finance, professional and technical services, and information and communication services, are expected to face the most significant trade barriers as the UK transitions to trading under the TCA with the EU. These sectors are likely to experience larger than average output shocks, with no sectors within this category anticipated to show growth in the near term.

In contrast, the UK's primary industries, such as agriculture and mining, exhibit more variability in terms of Brexit impact. For instance, while some changes in the trading arrangements with the EU are projected to yield benefits for British agriculture, the fishing sector is poised to encounter severe challenges. Specifically, fishing is expected to sustain a 30 percent negative output shock, positioning it as one of the sectors most adversely affected by the transition.

Reconciling the concerns of British farmers with the expected outcomes post-Brexit presents

a multifaceted challenge. While the agriculture sector is anticipated to experience the most substantial decline in exports to the EU—more than a 60 percent fall as illustrated in Figure 3.9—this sector also stands to benefit from reduced import competition. The expected decrease in imports, over 40 percent relative to a no Brexit scenario, suggests that domestic consumers may substitute EU goods with locally produced agricultural products, potentially benefiting a broad base of agricultural producers.

However, these theoretical gains contrast sharply with the challenges highlighted by British farmers. The adjustment to the new trade environment is expected to be particularly onerous for existing exporters due to the drastic reduction in export opportunities. Furthermore, Brexit's impact extends beyond trade policy to migration policy, significantly affecting the availability of seasonal labor critical for agriculture. These changes in migration policy, which lead to increased labor costs, have not been fully accounted for in the modelling. As a result, the projected gains for sectors where labor migration is an important factor may not materialize as anticipated.

Additionally, British fishers, who rely heavily on exports to the EU, now face formidable new barriers that inhibit their access to EU markets. The costs associated with these new barriers are likely to outweigh any benefits derived from reduced competition from EU fish exporters, compounding the sector's challenges.

The impact of Brexit on manufacturing subsectors exhibits significant variation, as illustrated in Figure 3.3. The effects are distinctly polarized: certain subsectors face substantial negative output shocks, whereas others are poised for growth. Notably, the manufacturing of food and beverages, one of the largest manufacturing sectors, is expected to expand by approximately 5 percent as a direct result of Brexit. This growth contrasts sharply with sectors like the manufacture of electrical equipment and basic metals, which are anticipated to suffer the most considerable declines, with expected output reductions of 7 percent and 14 percent respectively compared to scenarios within the EU.

This divergence in sectoral outcomes can largely be attributed to their varying degrees of exposure to EU exports and imports. For instance, the manufacturing of food benefits significantly from reduced import competition, given its high exposure to EU imports. This advantage sufficiently offsets the negative impacts stemming from lost market access, leading to an overall positive growth in output. Conversely, sectors like electrical equipment and basic metals, with potentially less benefit from import competition and greater reliance on EU markets, face more severe challenges.

Despite the diverse trade and output shocks across sectors, our analysis suggests that the

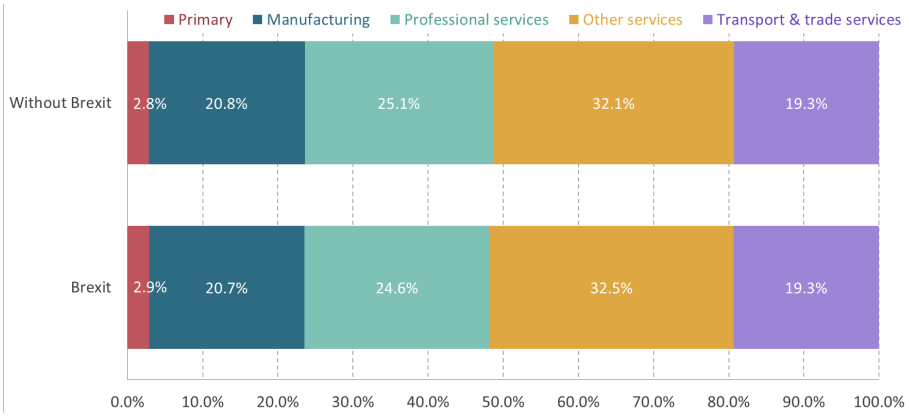


Figure 3.4: The Share of Output from Professional Services is Expected to Fall due to Brexit

long-term changes to the broad sectoral structure of the UK economy, trade specialization, and regional inequality are expected to be relatively minor. Notably, tradeable and professional services, including finance and business services, are anticipated to experience the most significant output reductions due to Brexit. We estimate that the contribution of these sectors to gross output will decrease slightly, by 0.5 percentage points, from 25.1 percent to 24.6 percent of gross output, as illustrated in Figure 3.4.[10]

The manufacturing sector's contribution to gross output is anticipated to decline marginally by 0.1 percentage points in the core EU integration scenario, decreasing from 20.8 percent to 20.7 percent. This change, while seemingly negligible, represents less than half the average annual decline experienced by the manufacturing sector's share of the economy throughout this century.

These findings indicate that the broader economic and labor market structures are expected to endure only limited impacts from Brexit. The labor market disruptions are projected to be minor and gradual, with less than 0.5 percent of the workforce, or approximately 132,000 individuals, anticipated to relocate from their current region-sector as a direct consequence of Brexit. These transitions are expected to occur slowly over several years. Thus, while Brexit may exert additional downward pressure on real wage growth amidst an ongoing cost of living crisis, it is unlikely to trigger significant workforce movements to new regions or sectors in search of employment.

The assessment of the new trading arrangements with the EU reveals substantial variation

[10]Primary industry includes agriculture and mining sectors. Professional services encompass finance and insurance, information and communication services, and professional, scientific, and technical services. Transport and trade services include transportation services and wholesale and retail trade. Other services comprise all other service sectors.

Figure 3.5: Long-term Estimated Change in Exports due to Brexit by Sector and Revealed Comparative Advantage in 2019: UK

in the trade shocks experienced across different sectors. This variation raises important questions regarding the implications for the trade specialization of the UK economy. Figure 3.5 illustrates the expected shifts in economic specialization. Specifically, the UK is projected to become less specialized in services overall, while goods trade is likely to further concentrate on sectors where the UK already has established strengths as measured by the revealed comparative advantage,[11] such as vehicles, aircraft, and pharmaceuticals.

This shift in specialization is particularly pronounced in services. The sectors of financial services and other business services, where the UK has traditionally been highly specialized and are significant exporters, are expected to experience larger-than-average declines in exports due to Brexit. Conversely, the insurance sector, another area of high specialization for the UK, is expected to see an increase in specialization post-Brexit.

3.4.2 Regional Variation

The differential impacts of Brexit across various sectors imply that its overall effects will not be uniformly distributed across the UK's regions. In public debates, considerable attention has been focused on whether Brexit will exacerbate or mitigate existing regional inequalities. Our analysis indicates that the North East, one of the UK's economically weaker regions, is poised to be disproportionately affected due to its high exposure to the EU market. By 2030, we project a decline in manufacturing output by 2.7 percent

[11] Note: Revealed Comparative Advantage (RCA), a measure of the UK's relative export specialization compared to global exports, in 2019 is along the vertical axis. Change in the share of UK exports is on the horizontal axis. Sectors from the model are matched to the HS2 and EBOPS sectors (the classification typically used for sectors and products in international trade data) used to estimate the RCA (De Lyon et al., 2022 and Dhingra et al., 2017).

relative to the baseline scenario, accompanied by significant reductions in regulated services such as professional, scientific, and technical services, and finance and insurance, which are expected to decrease by 4.1 and 4.0 percent, respectively. These declines are likely to exacerbate the already substantial productivity and income disparities in the region. Conversely, regions like the East of England and Scotland are anticipated to perform slightly better than the national average, suggesting a complex mosaic of regional impacts stemming from Brexit.

Under the Northern Ireland Protocol (NIP), Northern Ireland continues to adhere to the EU's rules for the free movement of goods and customs union regulations, distinguishing it from the rest of the UK. This special status shields Northern Ireland from the standard trade barriers that now exist between the UK and the EU. As a result, instead of imposing checks on goods moving between Northern Ireland and the EU, inspections are carried out on transfers from Great Britain to Northern Ireland to align with the NIP's provisions.

This arrangement has mitigated the economic impact of Brexit on Northern Ireland compared to other regions. Specifically, Northern Ireland has experienced a relatively minor total output decline of just 0.7% relative to the baseline. Should the NIP be discontinued, Northern Ireland would still be the fourth least affected region, facing an output shock of 1.1%—still more favorable than the UK average of 1.3%. Conversely, for all other UK regions, the NIP has had a marginal positive impact, lessening the reduction in output by about 0.01 percentage points.

Public attention has frequently focused on how London will fare post-Brexit, given its crucial role in regional inequality across the UK. Studies on Brexit's impact have varied widely, with some identifying London as either the most or least affected region.[12] Where findings have suggested a below-average impact on London, this is often attributed to the city's relatively low exposure to EU trade. Specifically, trade with the EU accounts for only about 7 percent of London's total output, a proportion that is lower than that of any other UK region. This lower dependency on EU markets may shield London from more severe economic repercussions experienced by regions with higher EU exposure.

The impact of Brexit on London—and consequently on overall regional inequality—remains uncertain. Unlike previous studies, which mainly considered general trade patterns, our model uses sectoral trade shares with various countries to model London's specific exposure to Brexit. Additionally, product and labor markets within the UK are assumed to be integrated with the rest of the country, though integration is not perfect.

[12]For example, see Los et al. (2017), Dhingra et al., (2017), Chen et al., (2018).

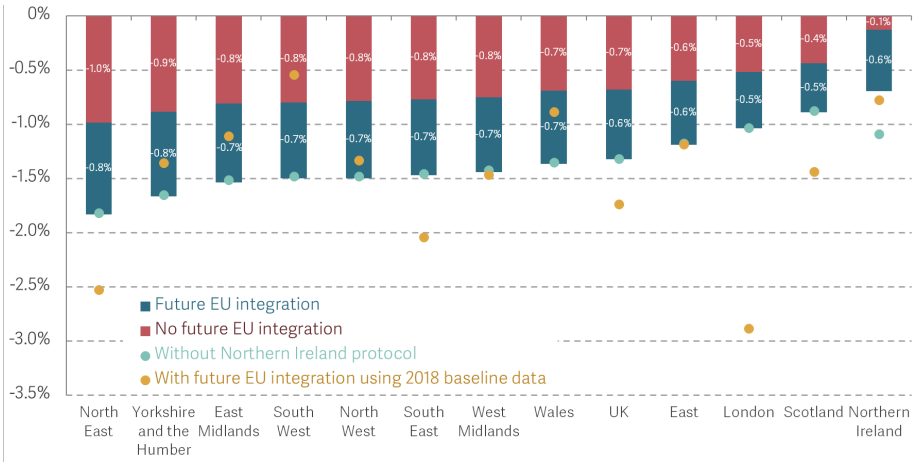


Figure 3.6: Estimated Falls in Gross Output by UK Region in 2030

However, London's exposure to Brexit varies annually, leading to uncertainty about how changes in EU market access might affect London's output, although the sectors most at risk are clear. The core EU integration scenario, depicted in Figure 3.6, indicates that London might experience a smaller than average decline in output. This scenario suggests a marginal widening of existing regional inequalities. However, updating the base year in our models to 2018 data leads to a starkly different conclusion, with London emerging as the region most adversely affected. This shift is driven by changes in the underlying data concerning London's EU exposure across sectors, which has increased from below average in 2014 to above average in 2018. This variability in data underscores the complexity of predicting Brexit's regional impacts, especially for a globally connected metropolis like London.

While there are some differences in how regions are affected by Brexit, these differences are modest and do not support the idea that Brexit will significantly boost productivity in poorer areas, as some supporters have suggested. These supporters have argued that Brexit would help balance out economic growth across different regions, a concept often referred to as "leveling up". However, the evidence indicates that the effects of Brexit on regional disparities are expected to be minimal.

3.4.3 Trade Openness, Productivity and Welfare

While Brexit is not anticipated to fundamentally alter the structure of the UK economy, it is expected to lead to a broad-based reduction in worker productivity and real wage. The introduction of new trade barriers with the EU, a major and proximate trading partner, is likely to significantly diminish the UK's trade openness. This reduction is challenging

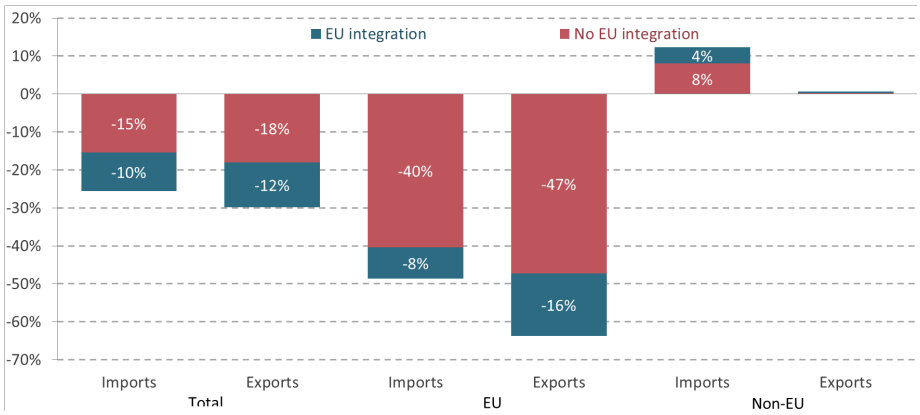


Figure 3.7: Estimated Falls in Exports and Imports with the EU, Outside the EU, and with the World in 2030: UK

to counterbalance with other trade policies, given the limitations on the UK's capacity to liberalize trade with major partners or to forge agreements as comprehensive as those provided by the Single Market. Consequently, a less open UK is projected to be poorer and less productive by the end of the decade. In the sections that follow, we will discuss the trade openness, welfare and productivity through our model results.

By 2030, exports to the EU are projected to be 47 percent lower than they would have been had the UK remained within the EU, compounded by an additional 16 percent decrease due to the missed opportunities for further integration with the EU. As illustrated in 3.7, by the end of the decade, the proportion of UK exports to—and imports from—the EU is anticipated to decline by 63 percentage points and 48 percentage points respectively. Current trade data shows that the share of imports from the EU has decreased by less than 5 percentage points, while the export share has remained stable. This suggests that significant adjustments in trade are yet to occur as global trade normalizes. These adjustments are expected to lead to further declines in real incomes and economic output.

A less open UK is expected to be both poorer and less productive. The decrease in trade openness will coincide with a 1.5 percent decline in labor productivity and a 2.1 percent reduction in real wages by the end of 2030, amounting to a loss of £551 per worker per year. This represents a significant setback, equivalent to losing a quarter of the productivity growth achieved over the past decade. Almost half of these negative impacts are attributable to the forgone opportunities for further EU integration. This missed integration intensifies the relative size of the barriers faced by UK firms, as EU firms now encounter even fewer frictions within the EU, exacerbating the disparities in market conditions.

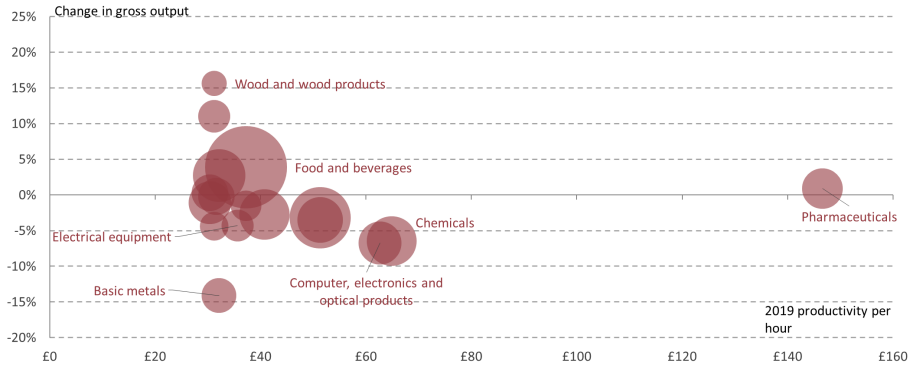


Figure 3.8: Expected Long-run Percentage Change in Gross Output Compared to Productivity per Hour across Manufacturing Sectors: UK

In response to the new trading arrangement, UK firms are expected to focus more on domestic markets. While Brexit might revive British manufacturing, our analysis indicates that the overall contribution of manufacturing to the economy will remain largely unchanged. The shift to domestic markets will primarily benefit lower productivity sectors that struggled against EU imports. Figure 3.8 shows that growing sectors had below-average productivity in 2019, with shrinking sectors averaging £47 per hour compared to £37 per hour for growing ones. This suggests that any manufacturing revival will likely be in lower productivity, lower-paid sectors, limiting improvements in overall economic prosperity or productivity.

While the overall scale of the productivity impact from Brexit, estimated at a 1.5 percent decrease in the long run, appears small relative to other studies, our focus is primarily on the impacts mediated through trade and labor market adjustment channels. For instance, the Office for Budget Responsibility (OBR) bases its productivity impact estimates on a range of studies, using an average value of a 4 percent total impact on productivity. Our analysis, however, specifically assesses the direct impacts stemming from trade and local labor market adjustment. It's important to note that our analysis does not account for impacts on investment or adjustments in other policy areas, such as changes in migration cost.

Estimates from comparable models regarding the trade impacts of Brexit show consistent results. For example, Bevington et al. (2019) which examined the trade impact of an average Free Trade Agreement (FTA) similar to the final Trade and Cooperation Agreement (TCA) but assuming higher future EU integration, projected a 2.5 percent decline in per capita incomes by 2030. Similarly, general equilibrium modeling conducted by the

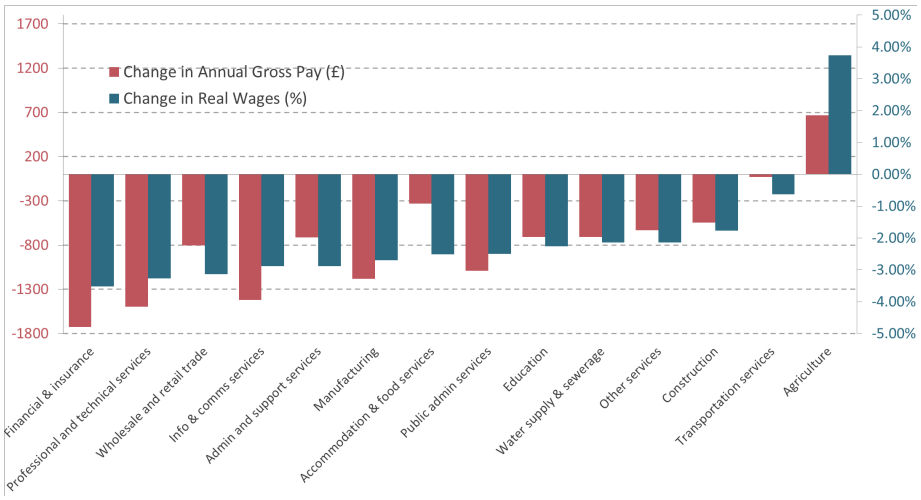


Figure 3.9: Annual Wage Falls by Selected Sectors: UK, 2030

IMF (2018) indicates that UK GDP would decrease by 2.5 percent in an FTA scenario akin to the final TCA.

Overall, our estimated impact on productivity, should be viewed as a conservative lower bound—it captures only the direct impacts resulting from changes in trade and labor market adjustment. Brexit will also introduce additional shocks to the UK economy through related policy reforms and impacts on investment, many of which have already materialized. Thus, while our estimates provide valuable insights, they represent only a partial view of the broader economic repercussions of Brexit.

Figure 3.9 shows that the finance and insurance sectors suffer the most from Brexit, with real wages decreasing by an average of 3.5 percent. This decrease translates into an annual loss of £1726 per employee compared to a scenario without Brexit. Other service industries also experience significant wage impacts. For example, wages in the professional, scientific, and technical services, as well as in the information and communication sectors, will be 3.2 percent and 2.9 percent lower, respectively, than if Brexit had not occurred. These reductions equate to annual wage losses of £1497 and £1418 per worker, respectively.

Regional variations in real wage declines, compared to a scenario without Brexit, reflect differences in sectoral composition and wage levels across UK regions. London sees the most significant drop in nominal wages, with an annual decrease of £880 per person, marking a 2 percent reduction in real wage as highlighted in Figure 3.10. Meanwhile, Wales and the North East experience the highest percentage declines in real wages, at 2.6 and 2.3 percent respectively. However, in terms of gross annual pay, these figures translate to decreases of only £590 and £580 per person per year, slightly above the UK average.

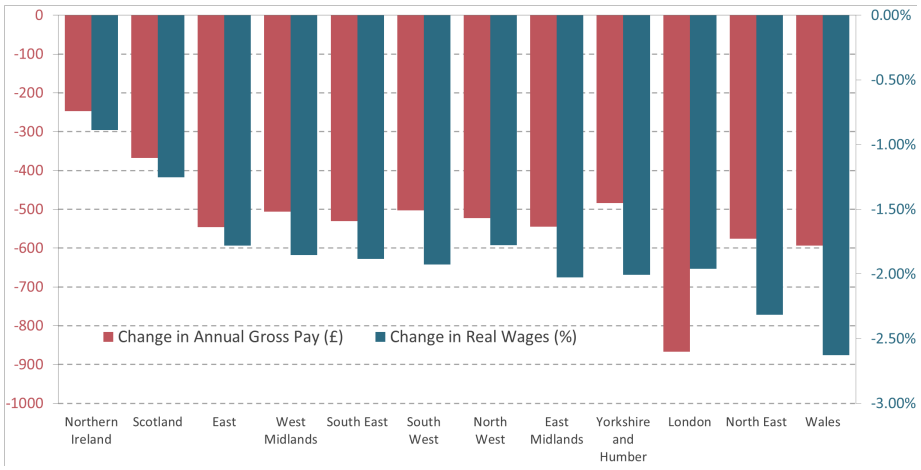


Figure 3.10: Annual Real Wage Falls and Effect on Annual Gross Pay by Region: UK, 2030

3.5 Conclusion

This chapter has explored the economic implications of the EU-UK Trade and Cooperation Agreement using a dynamic spatial general equilibrium model, offering a detailed assessment that moves beyond the speculative scenarios commonly found in earlier research. Our analysis has incorporated newly computed estimates of tariffs and NTMs defined by the TCA, regionalized WIOD and labor flow data to provide a comprehensive analysis.

The findings indicate that while the TCA necessitates significant sector-specific adjustments, it does not precipitate a fundamental restructuring of the UK economy. Instead, the principal effects are manifested in substantial declines in real wages and productivity. These impacts are not uniformly distributed across the economy but are particularly pronounced in sectors directly affected by the new trading barriers established by the TCA.

Furthermore, our findings suggest that the economic repercussions of Brexit, as facilitated by the TCA, will likely lead to long-term economic strain rather than the short-term disruptions initially anticipated. This underscores the necessity for policy interventions aimed at mitigating these impacts, particularly in the most vulnerable sectors and regions. Future policy frameworks should consider the detailed sectoral and regional analyses provided here to tailor strategies that can more effectively support the UK's economic adjustment in the post-Brexit landscape.

3.A Trade and Cooperation Agreement Model Inputs

In this section, we provide detailed methods for estimating tariff and non-tariff barriers (NTB).

3.A.1 Tariff Inputs

Although the Trade and Cooperation Agreement maintained zero tariffs on all goods trade, exporters are now required to meet rules of origin (ROO) requirements. These requirements prevent some firms from utilising the preferential access set out in the agreement, either because the goods they export do not meet the requirements (e.g., they contain a high share of foreign value added from inputs sourced from outside the EU) or because the regulatory burden to meet them is perceived to exceed the cost of paying the tariffs (especially where the most favoured nation, or MFN, tariffs are relatively small). Past work has estimated that the costs for businesses associated with rules of origin requirements are equivalent to a 3 to 5 percent tariff.

The EU collects data on the utilisation of preferential tariff rates on its imports across products and trading partners. Since the implementation of the TCA, this includes the UK. This data can be used to calculate two related but distinct measures:

- Preference utilisation rates, which are the share of trade using preferential tariffs over the share of trade where the MFN tariff is not zero.
- Tariff-free trade, which is the share of trade that paid no tariff over the total value of trade.

Issues with Using Preference Utilisation Rates to Inform Tariff Inputs

Although preference utilisation rates provide evidence on the value of trade that is paying tariffs to enter the EU despite the zero preferential tariffs available, there are some concerns when using them to inform modelling inputs addressed below.

1. Preference Utilisation Rates Increase Over Time Over time, preference utilisation tends to rise as firms learn how to utilise the preferences available, adapt supply chains to meet ROO requirements where beneficial, and stop trading where tariffs have put them at a disadvantage. The UK is no exception; tariff utilisation has increased in many products over time, rising overall from 59 percent in January 2021 to 71 percent in August 2021, as shown in Figure 3.A.1. However, since August, the rate has not increased but has been maintained around 70 percent. As preference utilisation rates are generally not

increasing month to month, using the latest preference utilisation rates gives a reasonable estimate of the share of trade that fails to utilise preferences and instead pays tariffs.

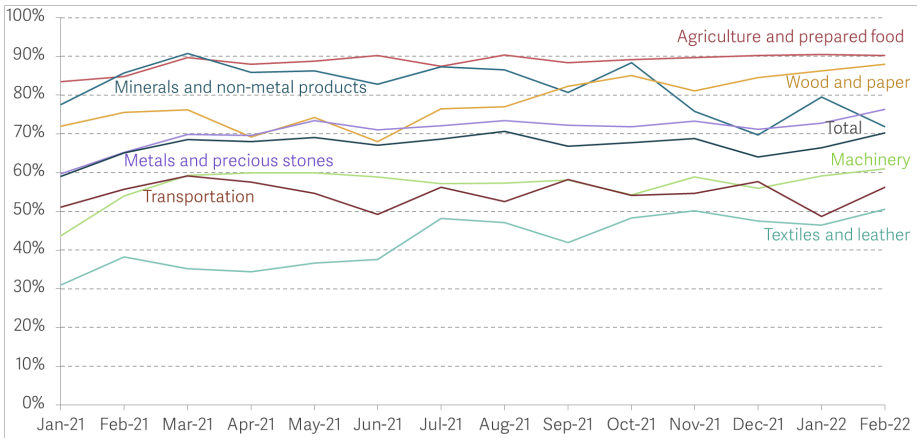


Figure 3.A.1: Utilisation of Tariffs Increased in the First Half of 2021 but have Remained Flat since August 2021

Notes: Preference utilisation rates are calculated as the share of trade imported using preferences over total trade imported using preferences or MFN non-zero tariffs.
Source: Eurostat adjusted extra-EU imports by tariff regime.

2. Preference Utilisation Rates Are Endogenous Preference utilisation rates are not independent of the barriers introduced in the agreement. The rates have been established as firms optimised their behaviour following the implementation of the UK-EU agreement; for example, some firms unable to meet rules of origin may have already dropped out. However, our judgement is that they still give a useful sense of the potential tariff costs across sectors to inform the estimated relative barriers facing firms.

Given these issues, UK utilisation rates are compared against those for exporters based in countries with similar EU agreements, specifically Canada, South Korea, and Japan. For sectors where UK utilisation rates are below the minimum of the other three countries, the median preference utilisation rate is used as a proxy.

Method for Estimating Model Tariff Inputs

As preference utilisation rates are below 100 percent, it is expected that some firms will continue paying tariffs, even in the long run. The method below uses tariff utilisation rates to estimate tariff barriers between the EU and UK across sectors after the UK left the EU:

1. Calculate the most recent 3-month preference utilisation rate and tariff-free rate using the EU trade data (December 2021 – February 2022).
2. Adjust preference utilisation rates for products where the UK preference utilisation

rate is below the expected level based on comparators. This adjustment applies to musical instruments, beverages, articles of apparel, man-made filaments, and special woven fabrics.

3. Adjust the average HS2-digit MFN tariffs to exclude zero tariff lines from the averages. To estimate the average tariff on goods that have non-zero tariffs for each HS2 product, divide the average tariff rate by the share of trade with a non-zero tariff.
4. Estimate the average tariff paid by sector under the TCA by multiplying the adjusted non-zero tariff rate by the share of trade that pays a tariff (equal to total trade minus zero MFN trade and zero preferential trade).
5. Convert the HS2 tariffs to WIOD sectors, weighting lines using UK-EU trade in 2021.

In this methodology, it is implicitly assumed that the costs of complying with rules of origin for firms utilising preferences are captured in the NTM estimates applied separately. Remaining tariffs are generally very low as preference utilisation and MFN tariffs are generally low, as shown in Figure 3.A.2.

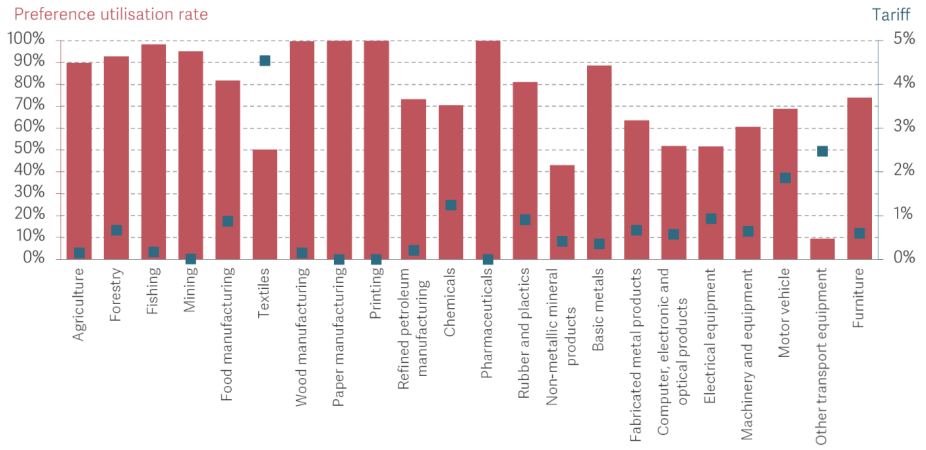


Figure 3.A.2: Remaining Tariffs are Mostly Small, Except for Textiles and Other Transport, Where Preference Utilisation Rates are Low

Notes: Preference utilisation rates are calculated as the share of trade imported using preferences over total trade imported using preferences or MFN non-zero tariffs.
*Source:*Eurostat adjusted extra-EU imports by tariff regime and WITS weighted average applied tariff data.

3.A.2 Goods Non-Tariff Inputs

In this chapter, these NTMs are estimated in line with Fusacchia et al. (2022)'s assessment of the Trade and Cooperation Agreement. This approach is based on Cadot and Gourdon (2016), which estimates ad valorem equivalents of the non-tariff barriers, differentiating between barriers with and without a deep regional trade agreement (RTA). The

only sectors considered to have sufficiently broad and deep mutual recognition of conformity assessment are the vehicles and pharmaceuticals sectors.

A difference between the Cadot and Gourdon (2016) approach and the approach taken in this chapter is that this chapter assumes border and rules of origin costs are captured in the Cadot and Gourdon non-tariff estimates. The estimations are shown in Table 3.A.1.

Table 3.A.1: Goods NTM Estimates

Sector	NTM (%)
Crop and animal production, hunting and related service activities	20.7
Forestry and logging	11.8
Fishing and aquaculture	20.5
Mining and quarrying	9.4
Manufacture of food products, beverages and tobacco products	18.7
Manufacture of textiles, wearing apparel and leather products	5.6
Manufacture of wood and of products of wood and cork	6.5
Manufacture of paper and paper products	3.3
Printing and reproduction of recorded media	3.3
Manufacture of coke and refined petroleum products	9.4
Manufacture of chemicals and chemical products	7.9
Manufacture of basic pharmaceutical products and preparations	5.4
Manufacture of rubber and plastic products	7.0
Manufacture of other non-metallic mineral products	9.4
Manufacture of basic metals	5.8
Manufacture of fabricated metal products	5.8
Manufacture of computer, electronic and optical products	6.7
Manufacture of electrical equipment	6.7
Manufacture of machinery and equipment n.e.c.	6.7
Manufacture of motor vehicles, trailers and semi-trailers	7.9
Manufacture of other transport equipment	7.9
Manufacture of furniture; other manufacturing	5.5

3.A.3 Services Inputs

The following data is used:

1. **OECD Services Trade Restrictiveness Index (STRI):** The STRI identifies, catalogues, and quantifies cross-cutting barriers to trade across services sectors, representing the applied MFN services barriers. Additionally, the OECD calculated a GATS-STRI, which estimates the maximum level of restrictiveness a country could have while still meeting commitments made in the WTO General Agreement on Trade in Services. The Intra-EEA-STRI estimates the maximum level of restrictive-

ness that any European Economic Area (EEA) member (EU members plus Iceland, Liechtenstein, and Norway) can apply towards another EEA member.

2. **Department for International Trade Assessments:** These assessments cover services restrictions in the UK-Canada Comprehensive Economic and Trade Agreement (CETA) and UK-Japan Economic Partnership Agreement (EPA). The Department for International Trade commissioned trade experts to produce STRI scores for these UK trade agreements.
3. **Developing Trade Consultants' Estimates:** These ad valorem equivalent estimates of MFN and EEA trade barriers are used in the UKTPO TCA assessment. Other studies provide estimates of UK services barriers, both within the EEA and facing countries outside the EEA (facing MFN barriers). These estimates are based on government-commissioned studies by Developing Trade Consultants (DTC) combined with estimates from other studies. For each of the MFN and EEA barriers, a GDP-weighted average across the four UK nations is calculated, taking the difference as a Single Market effect. DTC's results map closely to GTAP sectors in seven cases; in the remaining four, the Fontagne-based results are scaled up by the (UK-trade-weighted) average ratio of DTC to Fontagne results.

Methodology for calculating services NTMs

1. **Create Combined Bound and Applied STRI Indices:** Calculate indices which combine the applied and bound restrictions, with bound restrictions weighted so the impact of removing water (i.e., locking in elements of existing services liberalisation) is valued at 40 percent of the changes to the applied restrictions, in line with recent academic work and the Department for International Trade methodology for estimating NTM changes.

   $$\text{Total restrictiveness index} = \text{Applied trade restrictiveness index} + 0.4 \times (\text{Bound restrictiveness index} - \text{Applied restrictiveness index})$$

   For the within-EU index, the applied restrictiveness index is the lower of the STRI and the intra-EEA STRI for each sector. The bound restrictiveness index is the Intra-EEA STRI. For the MFN index, the applied restrictiveness index is the STRI, and the bound restrictiveness index is the GATS-STRI.
2. **Estimate the Share of Water Removed in Existing EU Agreements and Adjust for Differences in the TCA:** Calculate the average of the Department for International Trade published estimates of water removed by EU agreements with

Canada and Japan to represent a typical deep EU FTA. Adjust the average deep EU FTA in sectors where the depth of provisions in the TCA goes above or below the Canada and Japan agreements. Based on an assessment of the coverage and provisions of the TCA, three sectors are adjusted: audio-visual services, legal, and telecoms. Audio-visual liberalisation is removed as it is not covered by the TCA; legal and telecoms both have novel and deeper provisions, so liberalisation is boosted by 50 percent.

3. **Estimate the TCA Restrictiveness Composite Index and Use This to Estimate the Change in the Services Barrier:**

$$\begin{aligned}\text{TCA restrictiveness index} = {} & \text{MFN restrictiveness index} \\ & - 0.4 \times (\text{GATS STRI} - \text{STRI}) \\ & \times \text{estimated share of water removed under the TCA}\end{aligned}$$

Use the TCA index to estimate the share of the EU liberalisation that is captured under an FTA by calculating the difference between the TCA and EU index over the difference between the MFN and EU index. This percentage is then applied to the difference between the estimate of the EEA and MFN ad valorem equivalent services barriers which is shown in Table 3.A.2.

3.B Sensitivity Analysis

3.B.1 Future EU Integration and the Northern Ireland Protocol

The core scenario includes both future EU integration and the Northern Ireland Protocol, creating a total output shock in the dynamic model of 1.3 percent and in the static model of 2.2 percent. To identify the magnitude of each of these effects, we model the dynamic model over time and the static model without future EU integration and the Northern Ireland Protocol. Over time, the contribution of future EU integration to the total shock increases. In 2021, future EU integration contributes just 0.06 percentage points, which rises to 0.64 percentage points by 2030, explaining 49 percent of the total output shock (Figure 3.B.1). EU integration contributes 1 percentage point to the static scenario, which explains 46 percent of the total shock. The NIP contributes less than 0.005 percentage points in output reduction in 2030 to the dynamic model, rising from 0.001 in 2021. The NIP marginally improves the static reduction by 0.03 percentage points.

The effects of EU integration and the Northern Ireland Protocol on real wages have a

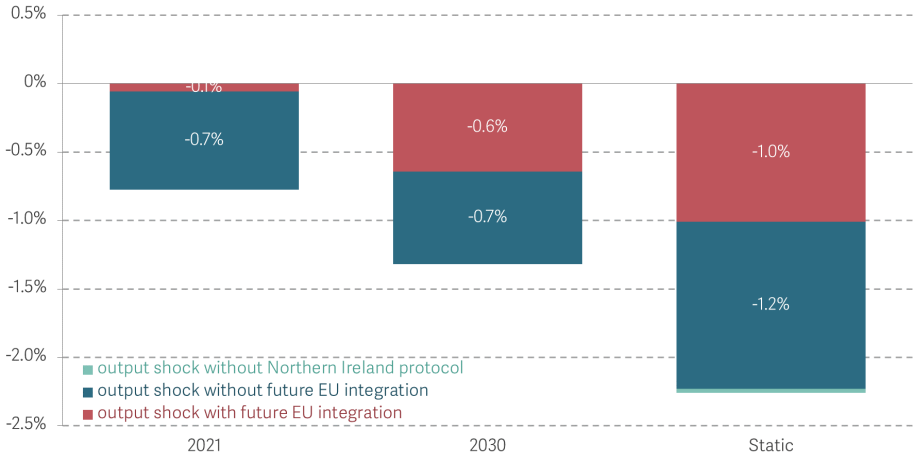


Figure 3.B.1: Contribution of Future EU Integration to Total Output Shock

similar effect to output. The total real wage shock in 2021 is just a 0.9 percent fall relative to the baseline, rising to 1.8 percent in 2030. EU integration contributes 44 percent of the real wage shock by 2030. The Northern Ireland protocol reduces the wage shock by just 0.01 percentage point in 2021, increasing to just 0.02 percentage points by 2030.

The core scenario uses 2014 data inputs to ensure that we model a pre-Brexit UK, without contamination either from anticipating the referendum or the outcomes of the referendum. We assume that firms in each region are able to trade independently with Europe and the rest of the world as well as with other UK regions. we consider the following two cases:

- Using 2014 data inputs with an assumption that the UK trades centrally. The core scenario assumes that each region-sector in the UK is able to independently trade with international markets. While this assumption will hold for many organizations, the extreme alternative assumption would be that the firm buys from the national warehouse. Then the relevant shares for the region are the sector-country shares and not the region-sector-country shares. At a high level, using a central trading assumption leads to an output shock of 1.3 percent in 2030, the same as the core scenario used in this chapter.
- Using 2018 data inputs. To compare changes over time, we used 2018 data as an input in the dynamic model, regionalizing using ONS regional trade data, combined with OECD IO tables. While 2018 data is more recent than 2014 data, there had already been effects from the Brexit referendum by this date. For example, the 12 percent depreciation caused imports to become more expensive, causing inflation. When using this data, the modeled output and real wages shocks are higher than to the scenario using 2014 data. For example, the combined output shock was 1.7

percent in 2030 compared with 1.3 percent in the core scenario. The NIP reduces the total output shock by 0.02 percentage points, and including future EU integration contributes 0.8 percentage points to the output shock in 2030 (47 percent of the total shock) in line with the 2014 scenario. However, the 2018 data provides a different assessment of the effects on regions and sectors.

3.B.2 Regional Sensitivities

Across scenarios using 2014 base data, the ordering of worst to least hit regions remains relatively constant. The North East sees the largest falls in output of 1.8 percent, while London sees a relatively low fall of 1 percent. However, varying the assumption of the UK trading centrally versus regionally, and changing the base data to 2018 does affect the regional ordering, as shown in Figure 3.B.2.

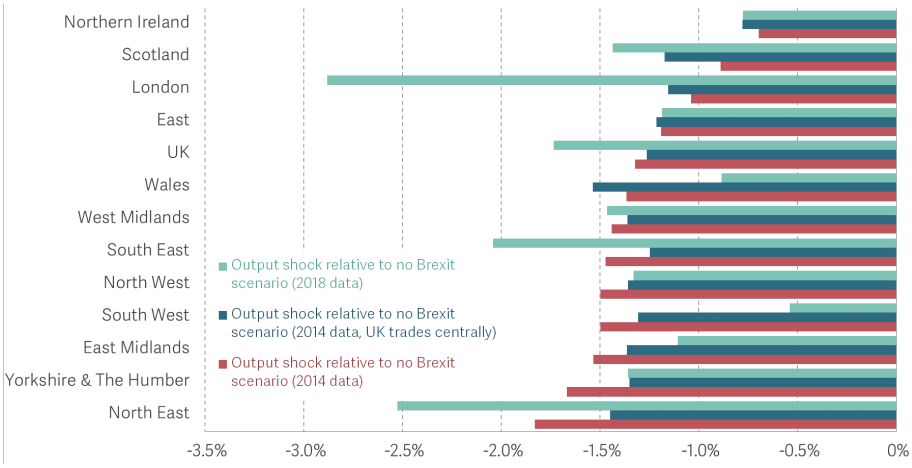


Figure 3.B.2: Variation in Regional Output Shocks due to Base Data and Assumptions

Comparing the 2014 core scenario with the assumption that the UK trades centrally reveals limited divergence between the output shocks for these two assumptions. The total output shock for the UK remains at 1.3 percent in both scenarios. The North East and East Midlands continue to have the worst output shocks in both scenarios. Additionally, London, Scotland, and Northern Ireland continue to perform the best in both scenarios. Wales experiences the largest movement, falling from a 1.4 percent output shock relative to the baseline to a 1.5 percent output shock, which places the region from mid-table to the worst affected. Otherwise, all other moves are relatively small, and regions are not sensitive to this scenario. In reality, the actual outcome is likely to lie somewhere between these two assumptions.

Comparing the use of 2014 base data with 2018 base data reveals shifts in the regions most

and least affected. The North East continues to experience a high output shock relative to the UK average, while Northern Ireland continues to see a relatively low output shock compared to the UK average. However, London's output shock shifts from being one of the lowest at 1 percent to the highest at 2.9 percent, affecting the UK average due to the size of the London economy and increasing the UK output shock from 1.3 percent to 1.7 percent. A comparison of real wage shocks is shown in Figure 3.B.3, with Wales and the South West experiencing lower real wage shocks using 2018 data compared to when using 2014 base data.

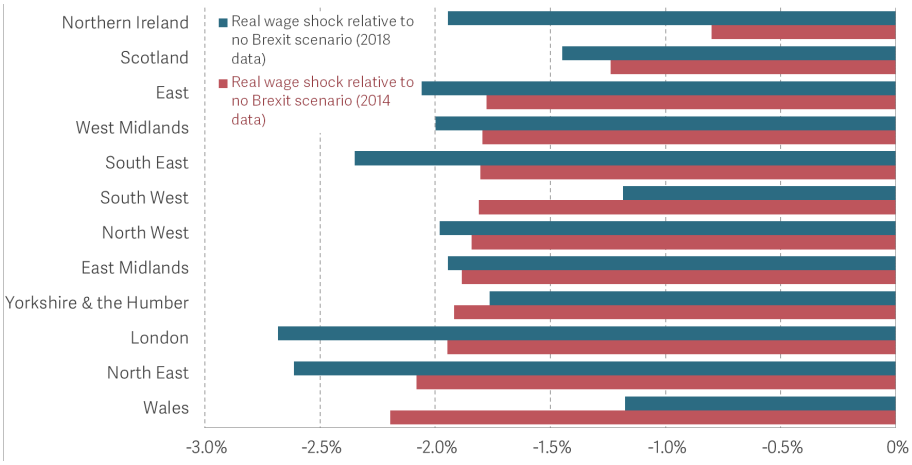


Figure 3.B.3: Variation in Regional Real Wage Shocks by Base Data

3.B.3 Sectoral Sensitivities

Sectors are less affected by the changes in base data for both output and real wages. However, the size of the shocks in 2018 base data is significantly higher than 2014 base data for certain regions' output shocks, and for almost all regions' real wage shocks. For both real wages and for output, agriculture and forestry benefit in both 2018 and 2014 base data. Sectors seeing the largest output and real wage shocks include regulated services such as finance and insurance, professional and scientific, and information and communications, as well as mining and quarrying. For both output and real wage changes, accommodation and food and manufacturing are the biggest movers between 2014 and 2018 base data. For example, manufacturing sees a 0.3 percent output fall relative to a no-Brexit scenario with 2018 base data compared with a 1.7 percent fall relative to a no-Brexit scenario using 2014 data (see Figures 3.B.4 and 3.B.5).

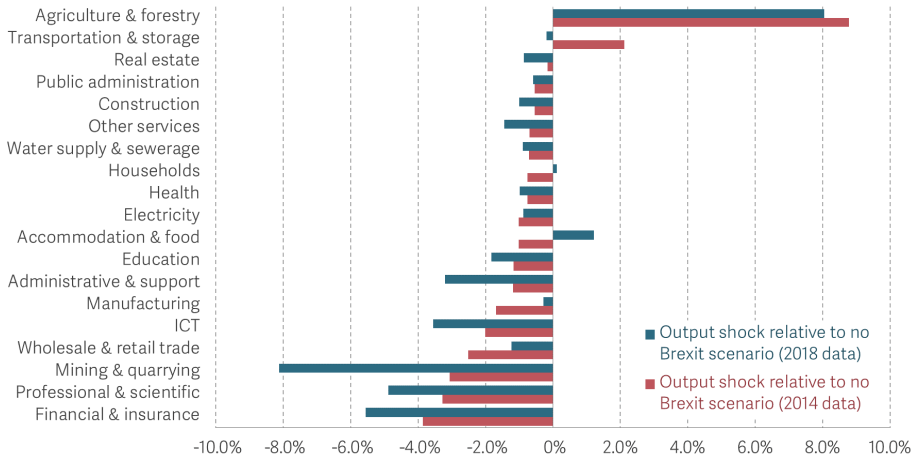


Figure 3.B.4: Sectoral Output Shock by Base Data

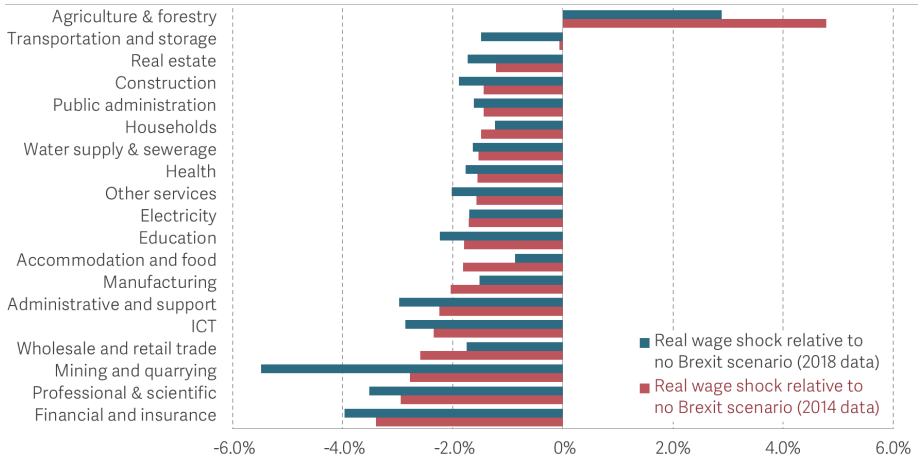


Figure 3.B.5: Sectoral Real Wage Shock by Base Data

3.C Solving Algorithm

In this section, we provide detailed steps for solving the dynamic exact-hat algebra.

The equilibrium conditions are summarized in (3.17) - (3.25). Given an initial observed allocation of the economy, including trade shares $\pi_0^{nj,ik}$, initial outputs X_0^{ij}, labor allocations L_0^{nk}, migration matrix $\mu_{-1}^{nj,ik}$, initial trade deficits D_0^i, and a convergent sequence of changes in economic fundamentals over time $\{\dot{A}_t^{nj}, \dot{\kappa}_t^{nj,ij}, \dot{\tau}_t^{nj,ij}, \dot{m}_t^{nj,ik}\}_{t=1}^{\infty}$, the solution for the sequence of changes in the model's endogenous variables does not require information on the level of fundamentals and can be solved in the following steps:

1. Guess a sequence of $\{\dot{v}_t^{nj,0}\}_1^{T+1}$ with $\dot{v}_{T+1}^{nj,0} = 1$.
2. Given the initial labor migration matrix and the sequence of $\dot{v}_t^{nj,0}$, plug in Equation (3.23) to obtain the sequence of migration matrix μ_t.

3. Given the migration matrix μ_t and initial labor allocation, solve the path of labor distribution using Equation (3.25).
4. For $0 \leq t \leq T$, solve price changes and wage changes in each period given the trade share and total output at time t:
   (a) Guess a sequence of wage changes in period $t+1$.
   (b) Solve price index changes given wages and the path of labor using Equations (3.17) and (3.18).
   (c) Solve trade shares at time $t+1$ using Equation (3.19).
   (d) Solve total output X_{t+1}^{ij} using Equation (3.20).
   (e) Plug the total output from step (d) and trade share from step (c) into Equation (3.22) to check if the trade balance condition Equation (3.22) is satisfied or not; iterate wage changes until Equation (3.22) holds.
5. Given a sequence of equilibrium wage changes and price changes, solve the new sequence of $\{\dot{v}_t^{nj,1}\}_1^T$ using Equation (3.24).
6. Check if the new $\{\dot{v}_t^{nj,1}\}_1^T$ is close to $\{\dot{v}_t^{nj,0}\}_1^T$, plug the new $\{\dot{v}_t^{nj,1}\}_1^T$ with $\dot{v}_{T+1}^{nj,1} = 1$ into step 1 until the new value equals the initial value.

Table 3.A.2: Difference Between MFN and EEA Treatment Estimates

Services Sectors	MFN - EEA (%)
Repair and installation of machinery and equipment	15
Electricity, gas, steam, and air conditioning supply	0
Water collection, treatment, and supply	0
Sewerage; waste collection, treatment, and disposal	0
Construction	8
Wholesale and retail trade and repair of motor vehicles	8
Wholesale trade, except motor vehicles	8
Retail trade, except motor vehicles	8
Land transport and pipelines	15
Water transport	13
Air transport	20
Warehousing and support for transportation	14
Postal and courier activities	16
Accommodation and food services	15
Publishing activities	16
Motion picture, video, TV production, sound recording, and broadcasting	16
Telecommunications	4
Computer programming, consultancy, and information services	21
Financial services, except insurance and pension funding	26
Insurance, reinsurance, and pension funding	11
Activities auxiliary to financial services and insurance	22
Real estate activities	12
Legal, accounting, head offices, and management consultancy	12
Architectural and engineering activities	9
Scientific research and development	12
Advertising and market research	12
Other professional, scientific, and technical activities	12
Administrative and support services	12
Public administration and defence	19
Education	19
Human health and social work	19
Other service activities	15

Bibliography

Aaronson, D., Dehejia, R., Jordan, A., Pop-Eleches, C., Samii, C., & Schulze, K. (2021). The effect of fertility on mothers' labor supply over the last two centuries. *The Economic Journal*, *131*(633), 1–32.

Allen, T., & Arkolakis, C. (2014). Trade and the topography of the spatial economy. *Nber Working Papers*, *129*(3), 1085–1140.

Allen, T., Arkolakis, C., & Li, X. (2023). On the equilibrium properties of spatial models. *Forthcoming In American Economic Review: Insights*.

Allen, T., Arkolakis, C., & Takahashi, Y. (2020). Universal gravity. *Journal of Political Economy*, *128*(2), 393–433.

Artuç, E., Chaudhuri, S., & McLaren, J. (2010). Trade shocks and labor adjustment: A structural empirical approach. *American economic review*, *100*(3), 1008–1045.

Bar, M., & Leukhina, O. (2010). Demographic transition and industrial revolution: A macroeconomic investigation. *Review of Economic Dynamics*, *13*(2), 424–451.

Barham, T., Kuhn, R., McCully, B., & Turner, P. (2023). The demographic transition and structural transformation: Evidence from bangladesh.

Barro, R. J., & Becker, G. S. (1989). Fertility choice in a model of economic growth. *Econometrica: journal of the Econometric Society*, 481–501.

Becker, G. S. (1960). An economic analysis of fertility. In *Demographic and economic change in developed countries* (pp. 209–240). Columbia University Press.

Bevington, M., Huang, H., Menon, A., Portes, J., Rutter, J., & Sampson, T. (2019). *The economic impact of boris johnson's brexit proposals* (tech. rep.). Centre for Economic Performance, LSE.

Bian, Y. (2002). Chinese social stratification and social mobility. *Annual review of sociology*, *28*(1), 91–116.

Boehm, J., Dhingra, S., & Morrow, J. (2022). The comparative advantage of firms. *Journal of Political Economy*, *130*(12), 3025–3100.

Brancaccio, G., Kalouptsidi, M., & Papageorgiou, T. (2020). Geography, transportation, and endogenous trade costs. *Econometrica*, *88*(2), 657–691.

Brandt, L., Hsieh, C.-T., & Zhu, X. (2008). Growth and structural transformation in china. *China's great economic transformation*, 683–728.

Broda, C., & Weinstein, D. E. (2006). Globalization and the gains from variety. *The Quarterly journal of economics*, *121*(2), 541–585.

Buera, F. J., & Kaboski, J. P. (2012). The rise of the service economy. *American Economic Review*, *102*(6), 2540–2569.

Cadot, O., & Gourdon, J. (2016). Non-tariff measures, preferential trade agreements, and prices: New evidence. *Review of World Economics*, *152*, 227–249.

Caliendo, L., Dvorkin, M., & Parro, F. (2019). Trade and labor market dynamics: General equilibrium analysis of the china trade shock. *Econometrica*, *87*(3), 741–835.

Caliendo, L., & Parro, F. (2015). Estimates of the trade and welfare effects of nafta. *The Review of Economic Studies*, *82*(1), 1–44.

Caliendo, L., & Parro, F. (2021). Trade policy. https://api.semanticscholar.org/CorpusID:240779656

Caliendo, L., & Parro, F. (2022). Trade policy. *Handbook of international economics*, *5*, 219–295.

Cao, K. H., & Birchenall, J. A. (2013). Agricultural productivity, structural change, and economic growth in post-reform china. *Journal of Development Economics*, *104*, 165–180.

Catchpole, E. A., & Morgan, B. J. (1997). Detecting parameter redundancy. *Biometrika*, *84*(1), 187–196.

Chan, K. W. (2010). The household registration system and migrant labor in china: Notes on a debate. *Population and development review*, *36*(2), 357–364.

Chen, W., Los, B., McCann, P., Ortega-Argilés, R., Thissen, M., & Van Oort, F. (2018). The continental divide? economic exposure to brexit in regions and countries on both sides of the channel. *Papers in regional science*, *97*(1), 25–55.

Chen, Y., & Fang, H. (2021). The long-term consequences of china's "later, longer, fewer" campaign in old age. *Journal of Development Economics*, *151*, 102664.

Cheng, W., Sachs, J., & Yang, X. (2004). An extended heckscher-ohlin model with transaction costs and technological comparative advantage. *Economic Theory*, *23*, 671–688.

Choukhmane, T., Coeurdacier, N., & Jin, K. (2023). The one-child policy and household saving. *Journal of the European Economic Association*, *21*(3), 987–1032.

Cole, D. J., Morgan, B. J., & Titterington, D. (2010). Determining the parametric structure of models. *Mathematical biosciences*, *228*(1), 16–30.

Coskun, S., & Dalgic, H. (2022). The emergence of procyclical fertility: The role of gender differences in employment risk.

Costinot, A., Oldenski, L., & Rauch, J. (2011). Adaptation and the boundary of multinational firms. *The Review of Economics and Statistics*, *93*(1), 298–308.

Croix, D. d. l., & Licandro, O. (2013). The child is father of the man: Implications for the demographic transition. *The Economic Journal*, *123*(567), 236–261.

Currie, J., & Schwandt, H. (2014). Short-and long-term effects of unemployment on fertility. *Proceedings of the National Academy of Sciences*, *111*(41), 14734–14739.

De Lyon, J., Martin, R., Oliveira Cunha, J., Shah, A., Shah, K., Thwaites, G., & Valero, A. (2022). Enduring strengths: Analysing the uk's current and potential economic strengths, and what they mean for its economic strategy, at the start of the decisive decade.

Delventhal, M. J., Fernández-Villaverde, J., & Guner, N. (2021). *Demographic transitions across time and space* (tech. rep.). National Bureau of Economic Research.

Desmet, K., Nagy, D. K., & Rossi-Hansberg, E. (2018). The geography of development. *Journal of Political Economy*, *126*, 903–983. https://api.semanticscholar.org/CorpusID:477870

Dhingra, S., Huang, H., Ottaviano, G., Paulo Pessoa, J., Sampson, T., & Van Reenen, J. (2017). The costs and benefits of leaving the eu: Trade effects. *Economic Policy*, *32*(92), 651–705.

Dhingra, S., Machin, S., & Overman, H. (2017). Local economic effects of brexit. *National Institute Economic Review*, *242*, R24–R36.

Doepke, M. (2005). Child mortality and fertility decline: Does the barro-becker model fit the facts? *Journal of population Economics*, *18*(2), 337–366.

Doepke, M., Hannusch, A., Kindermann, F., & Tertilt, M. (2023). The economics of fertility: A new era. *Handbook of the Economics of the Family*, *1*(1), 151–254.

Doepke, M., & Tertilt, M. (2016). Families in macroeconomics. In *Handbook of macroeconomics* (pp. 1789–1891, Vol. 2). Elsevier.

Doepke, M., & Tertilt, M. (2009). Women's liberation: What's in it for men? *The Quarterly Journal of Economics*, *124*(4), 1541–1591.

Eaton, J., & Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, *70*(5), 1741–1779.

Ebenstein, A. (2010). The "missing girls" of china and the unintended consequences of the one child policy. *Journal of Human resources*, *45*(1), 87–115.

Eckert, F., & Peters, M. (2022). Spatial structural change. *SSRN Electronic Journal*. https://api.semanticscholar.org/CorpusID:46908249

Engel, C., & Rogers, J. H. (1996). How wide is the border? *The American Economic Review*, 1112–1125.

Eriksson, J., & Gulliksson, M. (2004). Local results for the gauss-newton method on constrained rank-deficient nonlinear least squares. *Mathematics of Computation*, *73*(248), 1865–1883.

Erten, B., & Leight, J. (2021). Exporting out of agriculture: The impact of wto accession on structural transformation in china. *Review of Economics and Statistics*, *103*(2), 364–380.

European Commission, Joint Research Centre (JRC). (2020). Regional trade data for europe [Persistent Identifier: http://data.europa.eu/89h/432cf8a7-fd5e-4816-a70c-633a7380c77c].

Fajgelbaum, P., & Redding, S. J. (2022). Trade, structural transformation, and development: Evidence from argentina 1869–1914. *Journal of political economy*, *130*(5), 1249–1318.

Fan, T., Peters, M., & Zilibotti, F. (2021). Service-led or service-biased growth? equilibrium development accounting across indian districts. *NBER Working Paper Series*. https://api.semanticscholar.org/CorpusID:233650147

Fong, M. (2016). *One child: The story of china's most radical experiment*. Simon; Schuster.

Forbes, S. J., & Lederman, M. (2010). Does vertical integration affect firm performance? evidence from the airline industry. *The RAND Journal of Economics*, *41*(4), 765–790.

Forbes, S. J., & Lederman, M. (2009). Adaptation and vertical integration in the airline industry. *American Economic Review*, *99*(5), 1831–1849.

Fusacchia, I., Salvatici, L., & Winters, L. A. (2022). The consequences of the trade and cooperation agreement for the uk's international trade. *Oxford Review of Economic Policy*, *38*(1), 27–49.

Galor, O., & Weil, D. N. (1993). The gender gap, fertility, and growth.

Greenwood, J., Guner, N., & Marto, R. (2021). *The great transition: Kuznets facts for family-economists* (tech. rep.). National Bureau of Economic Research.

Greenwood, J., Guner, N., & Vandenbroucke, G. (2017). Family economics writ large. *Journal of Economic Literature*, *55*(4), 1346–1434.

Grossman, G. M., Helpman, E., & Lhuillier, H. (2023). Supply chain resilience: Should policy promote international diversification or reshoring? *Journal of Political Economy, forthcoming*.

Grossman, G. M., Helpman, E., & Sabal, A. (2023). *Resilience in vertical supply chains* (tech. rep.). National Bureau of Economic Research.

Harris, A., & Nguyen, T. M. A. (2022). Long-term relationships and the spot market: Evidence from us trucking.

Herrendorf, B., Rogerson, R., & Valentinyi, A. (2014). Growth and structural transformation. *Handbook of economic growth*, *2*, 855–941.

Hillberry, R. H., Anderson, M. A., Balistreri, E. J., & Fox, A. K. (2005). Taste parameters as model residuals: Assessing the "fit" of an armington trade model. *Review of International Economics*, *13*(5), 973–984.

Horn, R. A., & Johnson, C. R. (2012). *Matrix analysis*. Cambridge university press.

Hotz, V. J., Klerman, J. A., & Willis, R. J. (1997). The economics of fertility in developed countries. *Handbook of population and family economics*, *1*, 275–347.

Huang, Z., Lin, L., & Zhang, J. (2019). Fertility, child gender, and parental migration decision: Evidence from one child policy in china. *Available at SSRN 3375122*.

IEA. (2021). *World energy outlook 2021*. International Energy Agency. https://www.iea.org/reports/world-energy-outlook-2021

IMF, I. M. F. (2018). *United kingdom: Selected issues*. International Monetary Fund.

Jones, L. E., Schoonbroodt, A., & Tertilt, M. (2008). *Fertility theories: Can they explain the negative fertility-income relationship?* (Tech. rep.). National Bureau of Economic Research.

Joskow, P. L. (1988). Asset specificity and the structure of vertical relationships: Empirical evidence. *The Journal of Law, Economics, and Organization*, *4*(1), 95–117.

Klein, B., Crawford, R. G., & Alchian, A. A. (1978). Vertical integration, appropriable rents, and the competitive contracting process. *The journal of Law and Economics*, *21*(2), 297–326.

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

Kongsamut, P., Rebelo, S., & Xie, D. (2001). Beyond balanced growth. *The Review of Economic Studies*, *68*(4), 869–882.

Kranton, R. E. (1996). Reciprocal exchange: A self-sustaining system. *The American Economic Review*, 830–851.

Lafontaine, F., & Slade, M. (2012). Inter-firm contracts: Evidence. *Handbook of organizational economics*, 958–1013.

Leukhina, O. M., & Turnovsky, S. J. (2016a). Population size effects in the structural development of england. *American Economic Journal: Macroeconomics*, *8*(3), 195–229.

Leukhina, O. M., & Turnovsky, S. J. (2016b). Push, pull, and population size effects in structural development: Long-run trade-offs. *Journal of Demographic Economics*, *82*(4), 423–457.

Los, B., McCann, P., Springford, J., & Thissen, M. (2017). The mismatch between local voting and the local economic consequences of brexit. *Regional studies*, *51*(5), 786–799.

Macchiavello, R. (2022). Relational contracts and development. *Annual Review of Economics*, *14*, 337–362.

Macchiavello, R., & Morjaria, A. (2015). The value of relationships: Evidence from a supply shock to kenyan rose exports. *American Economic Review*, *105*(9), 2911–2945.

Manuelli, R. E., & Seshadri, A. (2009). Explaining international fertility differences. *The Quarterly Journal of Economics*, *124*(2), 771–807.

Masten, S. E., & Crocker, K. J. (1985). Efficient adaptation in long-term contracts: Take-or-pay provisions for natural gas. *The American Economic Review*, *75*(5), 1083–1093.

Moav, O. (2005). Cheap children and the persistence of poverty. *The economic journal*, *115*(500), 88–110.

Molloy, R., Smith, C. L., & Wozniak, A. (2011). Internal migration in the united states. *Journal of Economic perspectives*, *25*(3), 173–196.

Ngai, L. R., & Petrongolo, B. (2017). Gender gaps and the rise of the service economy. *American Economic Journal: Macroeconomics*, *9*(4), 1–44.

Ngai, L. R., & Pissarides, C. A. (2007). Structural change in a multisector model of growth. *American economic review*, *97*(1), 429–443.

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

Nicholson, S. E. (1993). An overview of african rainfall fluctuations of the last decade. *Journal of climate*, 1463–1466.

OECD. (2021a). Balanced trade in services (batis) dataset [Accessed: 20240612]. https://www.oecd.org/trade/balanced-trade-in-services-batis.htm

OECD. (2021b). Bilateral trade database by industry and end-use (btdixe) [Accessed: 20240612]. http://oe.cd/btd/

Pirrong, S. C. (1993). Contracting practices in bulk shipping markets: A transactions cost explanation. *The Journal of Law and Economics*, *36*(2), 937–976.

Qian, N. (2009). *Quantity-quality and the one child policy: The only-child disadvantage in school enrollment in rural china* (tech. rep.). National Bureau of Economic Research.

Qian, N. (2018). The effect of china's one child policy on sex selection, family size, and the school enrolment of daughters. *Towards gender equity in development*, *296*.

Redding, S. J., & Rossi-Hansberg, E. (2017). Quantitative spatial economics. *Annual Review of Economics*, *9*, 21–58.

Robinson, W. C., & Ross, J. A. (2007). *The global family planning revolution: Three decades of population policies and programs*. World Bank Publications.

Romalis, J. (2007). Nafta's and cusfta's impact on international trade. *The review of Economics and Statistics*, *89*(3), 416–435.

Rosenzweig, M. R., & Zhang, J. (2009). Do population control policies induce more human capital investment? twins, birth weight and china's "one-child" policy. *The Review of Economic Studies*, *76*(3), 1149–1174.

Rybczynski, T. M. (1955). Factor endowment and relative commodity prices. *Economica*, *22*(88), 336–341.

Schultz, T. (1997). Demand for children in low income countries. chapter 8 in handbook in population and family economics, vol. 1a, edited by mr rosenzweig and o stark.

Silvey, S. (1970). Statistical inference penguin books. *Middlesex, England*.

Solinger, D. J. (1999). *Contesting citizenship in urban china: Peasant migrants, the state, and the logic of the market*. Univ of California Press.

Song, Q., & Smith, J. P. (2019). Hukou system, mechanisms, and health stratification across the life course in rural and urban china. *Health & place*, *58*, 102150.

S&P. (2022). Gunvor backs out of lng deliveries to pakistan for april-june deliveries [Accessed: 2023-09-12].

S&P. (2023). Demand for term lng contracts firms amid supply security concerns [Accessed: 2023-09-12].

Święcki, T. (2017). Determinants of structural change. *Review of Economic Dynamics*, *24*, 95–131.

Tolvanen, J., Darmouni, O., & Essig Aberg, S. (2021). Pulp friction: The value of quantity contracts in decentralized markets. *Available at SSRN 4381156*.

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

Whyte, M. K., Feng, W., & Cai, Y. (2015). Challenging myths about china's one-child policy. *The China Journal*, (74), 144–159.

Williamson, O. E. (1979). Transaction-cost economics: The governance of contractual relations. *The journal of Law and Economics*, *22*(2), 233–261.

Yi, K.-M. (2003). Can vertical specialization explain the growth of world trade? *Journal of political Economy*, *111*(1), 52–102.

Zahur, N. B. (2022). Long-term contracts and efficiency in the liquefied natural gas industry. *Available at SSRN 4222408*.