ABSTRACT

Title of dissertation: ESSAYS ON THE WELFARE IMPLICATIONS OF INTERNATIONAL ECONOMIC INTEGRATION

Jingting Fan, Doctor of Philosophy, 2017

Dissertation directed by: Professor Nuno Limão
Department of Economics

The unprecedented international economic integration in the past few decades, in the form of global trade and activities of multinational corporations, has spurred heated discussions among policy makers and academics on the costs and benefits of globalization. Despite active research in this area, however, many aspects of globalization and its consequences are still not well understood. This dissertation examines the welfare implications of globalization, focusing on two specific aspects.

The first part of this dissertation studies the determinants and welfare implications of multinational corporations' decisions to perform R&D outside their home countries, or offshore R&D. In the first chapter, I develop a quantitative model that incorporates two motives for offshore R&D: the talent-acquisition motive, and the market-access motive. I calibrate the model and perform counterfactual experiments to understand the welfare implications of offshore R&D. I find that offshore R&D increases countries' gains from global integration by a factor of 1.2 on average, with much larger increases for developing than for developed countries. I also find

that incorporating offshore R&D has important implications for understanding the welfare impact of traditional forms of global integration, namely trade and offshore production. In the second chapter, I test the key implications of the two offshore R&D motives using firm-level data from the United States Patent and Trademark Office. The evidence supports the theory.

The second part of this dissertation, the third chapter, studies the effect of international trade on income and inequality of a country characterized by large domestic trade costs and migration restrictions. I develop a multi-region general equilibrium model featuring domestic trade and migration, both of which are subject to spatial frictions. Quantifying the model using data from China, I find that the trade between China and the rest of the world increases China's real income, but at the same time exacerbates the inequality in China. More than half of the rise in inequality comes from between-region inequality, while the rest comes from the skill premium. Moreover, there is an interaction between the spatial and the skill dimension of the effect of trade on inequality. Both results underscore the importance of incorporating domestic geographic frictions in understanding the welfare impacts of trade. As an additional contribution, I construct a city-level panel of the Hukou policies in China, and use it to quantify the interaction between Hukou reforms with China's international trade integration.

ESSAYS ON THE WELFARE IMPLICATIONS OF INTERNATIONAL ECONOMIC INTEGRATION

by

Jingting Fan

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2017

Advisory Committee:
Professor Nuno Limão, Chair/Advisor
Professor Serguey Braguinsky
Professor Şebnem Kalemli-Özcan
Professor Eunhee Lee
Professor John Shea

Dedication

To my parents, and other people who have helped me

致我的父母，和其他所有帮助过我的人

Acknowledgments

I am extremely grateful to my advisor, Professor Nuno Limão, who provided outstanding guidance and support through the entire process of this dissertation, without which this dissertation would not have been possible. I am also greatly indebted to the rest of my dissertation committee. Professor Şebnem Kalemli-Özcan always pushed me to confront ideas with more and better data. She also taught me a lot by bringing me on a project at an early stage of my graduate study. Professor John Shea gave sharp and the most detailed comments on my paper drafts, round after round, which helped improve my research and writing tremendously. I also thank Eunhee Lee for her insightful comments, and Serguey Braguinsky for serving on my dissertation committee.

I benefit a lot from interacting with Pablo D'Erasmo and Rafael Dix-Carneiro. Their second-year field classes shaped my research interests, and taught me the necessary skills to complete this dissertation. Their support continued even after their departed from the University of Maryland. The RA job from Pablo D'Erasmo exempted me from fixed-schedule teaching, giving me great flexibility in terms of when and where to work. Rafael Dix-Carneiro continued to offer advice on a paper I started in my second year, which became a chapter of this dissertation. My gratitude extends to Felipe Saffie, who, despite not a member of my dissertation committee, was always available, and provided valuable and timely feedback to my ideas.

I thank my friends and colleagues in graduate school: Wenlan Luo, Jongho Park, Marisol Rodríguez Chatruc, Lixin Tang, Ozlem Tonguc, and Ben Zou. They not only offered helpful suggestions on research, but also gave me encouragement and support. As my coauthors, Wenlan Luo, Lixin Tang, and Ben Zou bore with me with patience for the slow progress when I had to focus on my dissertation, and were always willing to help when I was stuck.

I am enormously indebted to my parents. The past six years have been a humbling experience for me, and it was their unconditional support that kept me motivated to finish this dissertation.

Finally, I acknowledge the financial support from the University of Maryland Graduate School.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

CDF	Cumulative distribution function
CHNS	Chinese Health and Nutrition Survey
EPO	European Patent Office
FDI	Foreign direct investment
IPR	Intellectual property right
PCT	Patent Cooperation Treaty
PDF	Probability density function
PWT	Penn World Table
R&D	Research and development
RoW	The rest of the world
USPTO	The United Patent and Trademark Office

Chapter 1: Talent, Geography, and Offshore R&D

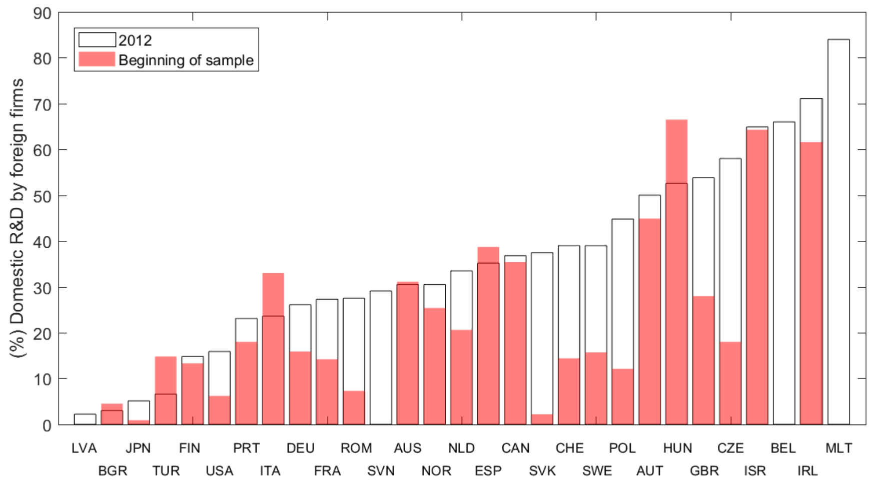
1.1 Introduction

Global integration in the form of international trade and multinational activities is one of the most significant economic phenomena of the past decades. Its impact has become an important topic for policy discussion and academic research. Existing studies on globalization focus on trade and multinationals' *offshore production* activities, but abstract from their *offshore R&D* activities, which also occur at significant levels. Figure 1.1 plots the share of R&D expenditures in a country incurred by the affiliates of foreign multinationals located in that country as a measure of offshore R&D. Uncolored bars are for 2012, and colored bars are for the first year with available data for each country, dating back to as early as 1985. By this measure, offshore R&D increased in most countries in the past two decades. In 2012, foreign affiliates accounted for more than 30% of R&D expenditures in the median country in the sample.[1]

The offshore R&D decisions of multinationals could have important aggregate implications. By determining the location and efficiency of R&D activities, offshore

[1] In appendix I show that the importance of offshore R&D can also be established using international patent statistics.

Figure 1.1: The Level and Growth of Offshore R&D, 1985-2012



Notes: The measure for offshore R&D in country i is $\frac{\text{R\&D expenditures in country } i \text{ by foreign firms}}{\text{Total private R\&D expenditures in country } i}$. Uncolored bars indicate the value of this variable in 2012; colored bars indicate the value at the beginning of the sample, which differs by country and dates back to as early as 1985. Data source: OECD.

R&D directly affects the income of countries. Moreover, in a world interconnected through trade and offshore production, offshore R&D can affect income indirectly, by shaping countries' specialization in innovation or production.

In this chapter I model and quantify the impact of offshore R&D. I address three questions. First, what are the determinants of offshore R&D? Second, how large are the welfare gains of opening up to offshore R&D? Third, how do these gains depend on and interact with the traditional forms of economic integration, namely trade and offshore production?[2]

I develop a unified framework for firms' global R&D and production decisions. In the model, firms differ along two dimensions: innovation efficiency, which governs how effective a firm is in converting researcher input into new product blu-

[2]Throughout this dissertation, I use the term offshore production to refer to cases in which a product is produced in a location different from where it is developed. This is related to the term "multinational production" used in recent studies (Ramondo, 2014; Ramondo and Rodríguez-Clare, 2013; Irarrazabal et al., 2013; Arkolakis et al., 2014; and Tintelnot, 2016).

eprints, and production efficiency, which governs a firm's productivity in converting production labor into output. Researchers differ in their talent. Firms can enter foreign countries (hosts) to perform offshore R&D. In each host, the firm matches with local researchers to develop new varieties. I model R&D as an assignment problem between firms and researchers, in which researcher talent and firm efficiency are complements. This setup deviates from the efficiency units assumption, and implies that quality and quantity of researchers are not perfect substitutes, an important feature of R&D in reality.[3]

I embed this offshore R&D decision into a multi-country general equilibrium model of global production and trade (Arkolakis et al., 2014). Specifically, after a product is developed by an R&D center, whether onshore or offshore, the firm first chooses which countries to sell it to, and then decides where to produce it. A firm from the U.S. therefore can develop a new product in the U.K., produce it in China, and export from there to India. These flexible decisions capture the complex strategies employed by modern multinationals.[4]

The model allows for two motives for offshore R&D commonly cited by firms: "market-access" and "talent-acquisition".[5] The former is straightforward: firms want to produce near their markets to save on trade costs. If separating the locations of innovation from production is costly, firms have incentives to offshore

[3]The output distribution of researchers is highly skewed. Akcigit et al. (2016) shows that the average top 1% inventor has 1019 lifetime citations, while the median inventor has only 11.

[4]DuPont offers a good example. Headquartered in Delaware, U.S., it has major R&D centers located in the U.S., Brazil, China, Switzerland, Korea, Germany, and Japan. Moreover, it has production facilities in 19 countries, from which it serves around 90 countries.

[5]According to firm-level surveys (see, for example, Thursby and Thursby, 2006), the quality of research personnel and host country market potential are the two most important factors firms consider, when choosing where to build their offshore R&D centers.

their R&D to large markets. The latter motive depends on both firm and host country characteristics. First, it reflects the host country's inventor wage, which depends on the abundance of talented inventors—an input supply effect, and the abundance of efficient firms competing for talent—an input demand effect. Second, because of the complementarity in innovation, host inventor wage interacts with firm efficiency to reinforce the talent-acquisition motive for high-efficiency firms.

Despite being rich, the model is tractable. I derive an analytic expression for the model-implied gains from openness, which augments the expression in Arkolakis et al. (2014) with an additional term that captures the importance of foreign companies in domestic R&D.[6] The expression makes it clear that offshore R&D represents a new channel for countries to benefit from global integration.

I study the quantitative importance of the two offshore R&D motives, and the magnitudes of the welfare gains. Specifically, I calibrate the model to 25 countries and a composite of 22 other countries. I parameterize each country's distribution of firm efficiency using the World Management Survey developed by Bloom et al. (2012), and its talent distribution using the international cognitive test score database developed by Hanushek and Woessmann (2012). I determine other parameters by matching various statistics of the firm size distribution in the U.S. and the intensities of bilateral international activities, including trade, offshore production, and offshore R&D. The model matches several non-targeted patterns in the data, better than an otherwise similar model without complementarity between firm efficiency

[6] Gains from openness is defined as the change in real income as a country moves from complete isolation to the observed equilibrium.

and researcher talent.

I quantify the importance of international differences in the distributions of firm efficiency and researcher talent in explaining the observed level of offshore R&D. I eliminate the incentives of offshore R&D arising from these distribution differences by first giving each individual country the management distribution of the U.S. (the highest in the world), and then the talent distribution of Brazil (the lowest in the world). The former reduces the average level of offshore R&D by around three quarters, whereas the latter reduces this average by around one third. So differences in the distributions of talent and management efficiency are an important driving force for offshore R&D.

I further examine how a country's access to foreign markets through exporting, and to foreign producers through offshore production, affect its attractiveness as a destination for R&D. While both consumer and producer access increase the return to innovation in partial equilibrium, I find that they have opposite general equilibrium effects: consumer access reduces inward offshore R&D, while producer access increases it. Therefore, increasing access to foreign markets through reducing exporting costs would not necessarily help a country in attracting R&D-intensive FDI. Country specialization in innovation or production is the key to understanding this result. When a country loses access to foreign consumers through exporting, its competitiveness in production weakens, which lowers wages and makes it more attractive as a host for offshore R&D centers. As a result, it specializes more in innovation, and firms do R&D there and offshore their production to other countries. Such specialization is not possible without offshore production, so when both consumer and producer access are shut down, the average offshore R&D across countries decreases to less than half of the benchmark level.

Together, these two sets of experiments suggest that the talent-acquisition and market-access motives in the model are strong enough to account for the observed level of offshore R&D on average.

I further examine the normative implications of offshore R&D. The median welfare gains from offshore R&D, defined analogously to the gains from trade, are around 2.2% of real income. Compared to a restricted version of the model with only trade and offshore production, the welfare gains from openness in the full model with offshore R&D are larger by a factor of 1.2. Importantly, this amplification is substantially larger for emerging countries than for developed countries, mainly because a larger share of R&D in emerging countries is carried out by foreign affiliates. Overlooking this channel therefore will not only result in underestimating the gains from globalization, but also bias the assessment of the relative size of the welfare gains across countries.

Existing quantitative studies on multinational activities do not separately model offshore R&D and offshore production, even though they are very different activities that can be targeted by specific policies.[7] Is this an innocuous assumption for policy simulations? To answer this question, I compare the effects of policies designed to promote these two multinational activities, focusing on China and India as an example.[8] First, I reduce the inward offshore R&D costs in these two

[7]For example, countries can grant tax credits or open their borders specifically to R&D intensive FDI. An example is the U.K. "patent box," which reduced the corporate tax rate on revenues from R&D by 10 p.p.

[8]This policy evaluation is interesting in its own right because these two emerging giants are

countries; second, I reduce inward offshore production costs. I find that, in the first experiment, China and India reap most of the benefits, whereas in the second experiment, developed countries also benefit significantly. The gains are small for developed countries in the first experiment because offshore R&D liberalization weakens the comparative advantage of China and India in production, which reduces the welfare gains from global specialization for everyone. In the second experiment, in contrast, the changes are more aligned with countries' comparative advantage. This comparison highlights the different implications for other countries of liberalization in offshore R&D and production. Such differences are especially relevant for studying multilateral investment agreements.

Offshore R&D also has implications for the welfare gains from other types of economic openness. To make this point, I perform an experiment with the same unilateral reductions in inward offshore production costs as in the previous experiment, but in a restricted version of the model without offshore R&D. Compared to the previous experiment, this experiment leads to substantially higher welfare gains for developed countries, and lower welfare gains for India and China. The distribution of profit from innovation is the key to the difference. More offshore production in China and India increases wages and reduces the profits from performing R&D there. At the equilibrium level of offshore R&D, the profit decreases are shared among domestic and foreign firms in these two countries; without offshore R&D, all the losses would be borne by domestic firms. This experiment shows that it is

becoming popular destinations for offshore activities. Related to this trend, their governments are attempting to attract more foreign companies, especially R&D intensive ones, by cutting red tape and speeding up the approval process.

important to model offshore R&D, even if one's goal is to evaluate the effects of offshore production.

1.2 Related Literature

This chapter is related to the recent literature that quantifies the gains from globalization, especially studies on the aggregate implications of technological transfer through multinational activities (see, among others, McGrattan and Prescott, 2009; Ramondo and Rodríguez-Clare, 2013; Irarrazabal et al., 2013; Arkolakis et al., 2014; Tintelnot, 2016; Burstein and Monge-Naranjo, 2009; Alviarez, 2016; and Holmes et al., 2015).[9] Within this literature, the most closely related paper is Arkolakis et al. (2014), which studies the welfare gains from trade and offshore production. The present chapter differs in two aspects. First, rather than treating innovation efficiency of a country as a single exogenous parameter, I decompose it into two measurable components, firm innovation efficiency and researcher talent, and examine the role of each in shaping a country's comparative advantage in innovation. Second, I allow firms to perform offshore R&D by mobilizing their managerial capacity abroad, so a country's comparative advantage in innovation is endogenous. I show that this channel has quantitatively important implications for both the gains from openness, and the effect of specific policy changes.

This chapter is also related to the literature explaining the pattern of FDI,

[9] See Antràs and Yeaple (2014) for a recent review of the literature on multinational corporations. Also see Costinot and Rodriguez-Clare (2014) for a review of quantitative studies on the aggregate implications of international trade, which encompasses the bulk of the research on the gains from globalization.

dating at least as far back as the theoretical work by Helpman (1984) and Markusen (1984) (for vertical and horizonal FDI, respectively). More recently, researchers have examined the determinants of M&A FDI (Nocke and Yeaple, 2007; Nocke and Yeaple, 2008; and Head and Ries, 2008), and have incorporated firm heterogeneity into the model (Helpman et al., 2004).[10] This chapter contributes to this literature in two ways. Theoretically, I outline a rich model of R&D, which can be viewed as a model of FDI with two-tiered vertical linkage: one between headquarters and R&D centers, and one between R&D centers and production sites. This structure allows the model to capture the complex strategies frequently seen in modern multinationals, in a way that existing two-country models of offshore R&D cannot (Gersbach and Schmutzler, 2011). Quantitatively, I calibrate a general equilibrium model to assess the strength of each factor.

In terms of modeling, this chapter is related to a number of studies that use an assignment framework to understand international trade and offshoring.[11] I apply a matching framework to innovation decisions in a model of multinational production and trade, and quantify the effects of complementarity between firms and researchers. In doing so, I develop a computational algorithm that can solve the matching function efficiently in the presence of multiple countries and when endogenous offshore R&D decisions lead to discontinuities in innovation efficiency distributions. This setup and computational algorithm could have applications in

[10]Studies have also examined empirically the impact on FDI flows of various factors, including skill endowments (Yeaple, 2003), institutions (Alfaro et al., 2008), and taxes and corruption (Wei, 2000).

[11]See, among others, Grossman and Maggi (2000), Yeaple (2005), Costinot and Vogel (2010), and Antras et al. (2006).

other contexts.[12]

This chapter's focus on international cooperation in R&D is shared by several recent papers (Kerr and Kerr, 2014; Kerr et al., 2016; and Branstetter et al., 2013). These papers discuss international cooperation either among inventors from different countries, or between inventors and firms from different countries, made possible by international migration or multinational activities. This chapter contributes to this literature by developing a model of offshore R&D, testing its specific predictions, and quantifying the aggregate implications of offshore R&D.

I organize the remainder of this chapter as follows. I set up a general equilibrium model of offshore R&D in Section 1.3. I parameterize this model to match the data in Section 1.4, and perform counterfactual experiments using the parameterized model in Section 1.5. I conclude and discuss directions for future research in Section 1.6.

1.3 A Model of Offshore R&D and Production

This section sets up the model and describes firms' global innovation and operation decisions.

[12] Roys and Seshadri (2014) quantifies a general equilibrium model of team production based on Antras et al. (2006) in a closed-economy setting. Their model fixes the team size each of manager exogenously, so wages do not play an allocative role. In the present chapter, wages determine team size and firm size distribution.

1.3.1 Environment

There are N countries in the model, indexed by $i = 1, 2, ...N$. Country i is endowed with L_i^R measure of researchers, who differ in their talent, $\theta \in \Theta$, distributed according to $H_i(\theta)$, and L_i^P measure of homogenous production workers. There are no immigration or education choices in the model, so L_i^R, L_i^P, and $H_i(\theta)$ are both exogenous.[13]

Researchers work with R&D centers to develop new differentiated varieties. Production workers manufacture these varieties and perform operational tasks for R&D centers (in the form of fixed costs). Country i is also endowed with E_i measure of heterogeneous firms with different innovation efficiencies, $\tilde{z}^R \in \tilde{\mathbb{Z}}^R$, distributed according to $G_i^E(\tilde{z}^R)$. Firms build R&D centers in different countries, which then recruit local researchers to develop new varieties. I use R_i to denote the measure of R&D centers in country i. In equilibrium R_i is an endogenous outcome determined by firms' offshore R&D decisions.

The representative consumer in country i decides how much to spend on each

[13] The talent distribution in a country reflects the quality of the education system, education choice, as well as cultural traits such as openness to innovation. By taking the talent distribution as given, this chapter abstracts from the effect of international integration on these factors. To better focus on offshore R&D, this chapter also abstracts from immigration, which is especially relevant when it comes to highly skilled workers such as the inventors in this model. In an extended version of the model that incorporate immigration, high skill immigration and offshore R&D likely work as substitutes for firms and inventors from different countries to work together, so the gains from offshore R&D might be smaller compared to the benchmark model presented in this chapter. The magnitude of the difference, however, depends crucially on the leniency of the immigration policy. I leave a quantitative evaluation of the interaction between immigration and offshore R&D policies to future work.

variety, according to the following preference:

$$U_i = (\int_{\Omega_i} q_i(\omega)^{\frac{\sigma-1}{\sigma}} d\omega)^{\frac{\sigma}{\sigma-1}},$$

where Ω_i denotes the set of product varieties available in country i, $q_i(\omega)$ is the consumption of variety ω, and $\sigma > 1$ is the elasticity of substitution. Let the aggregate consumption expenditure in country i be X_i. The demand for variety ω is:

$$q_i(\omega) = p_i(\omega)^{-\sigma} \frac{X_i}{P_i^{1-\sigma}},$$

where $P_i^{1-\sigma} = \int_{\Omega_i} p_i(\omega)^{1-\sigma} d\omega$ is the ideal demand price index aggregated over $p_i(\omega)$, the price of variety ω in country i.

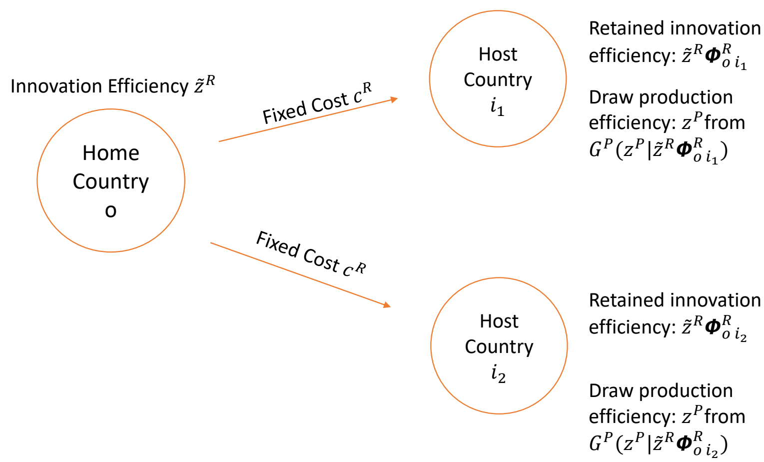
1.3.2 Firm Decisions: Overview

This subsection overviews firms' decisions. In the model, firms operate in multiple countries, and make sequential decisions on R&D, production, and exporting. I will use the following indexing conventions throughout this dissertation: o denotes a firm's headquarters, that is, the country where a firm originates; i denotes the country where a product is developed—the location of the R&D center; l denotes the country where the product is manufactured; and d denotes the destination country where it is consumed.

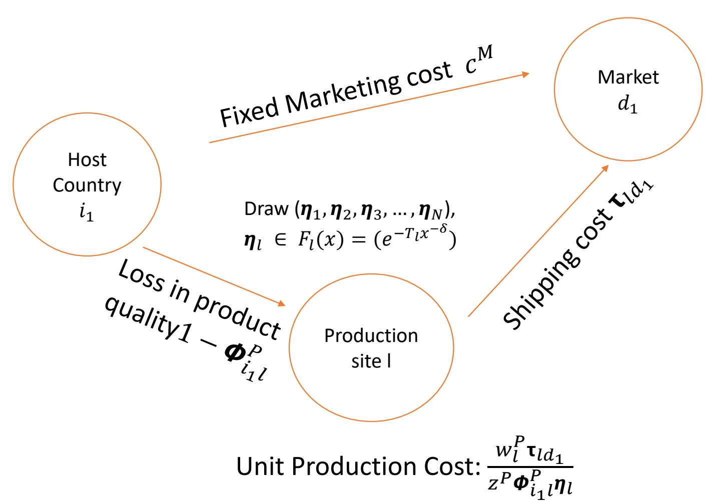
Consider a firm from country o. Knowing its innovation efficiency in the home country, $\tilde{z}^R$, the firm decides how many R&D centers to open and in which

Figure 1.2: Firm's Two-tiered Decisions

(a) Offshore R&D Decisions



(b) Offshore Production and Export



countries. To open an R&D center in country i, it pays a fixed cost of c_i^R in country i production labor. An R&D center's innovation efficiency depends on that of its parent.[14] Motivated by evidence on spatial frictions in knowledge transfers within firms (see, for example, Irarrazabal et al., 2013; Keller and Yeaple, 2013), I assume that firms can only transfer part of their innovation management efficiency to offshore R&D centers. Letting $\phi_{oi}^R \leq 1$ be the proportion of innovation efficiency that can be transferred, the innovation management efficiency for an R&D center in country i operated by a country o firm is $z^R = \tilde{z}^R \phi_{oi}^R$. This efficiency governs how many varieties can be developed by a given number of researchers.

Innovative firms are not always the most efficient in carrying out manufacturing. To allow for this heterogeneity, each R&D center upon entry also obtains a random draw of production management efficiency, denoted $z^P \in \mathbb{Z}^P$, which is common to all products developed by the R&D center. To capture positive correlation between innovation efficiency and production efficiency, the distribution from which z^P is drawn increases in z^R in the sense of first-order stochastic dominance. I use $G^P(z^P|z^R)$ to denote the CDF for production efficiency draws, with $g^P(z^P|z^R)$ being the corresponding probability density function (PDF).[15] This offshore R&D module is illustrated in Figure 1.2a. As the figure indicates, firms can open multiple R&D centers in different countries, but at most one R&D center in each country.

[14] This assumption follows a long tradition in the theory of multinationals, see, for example, Helpman, 1984; Helpman et al. (2004); and Nocke and Yeaple (2008). Empirically, Guadalupe et al. (2012) documents an increase in innovation and adoption of foreign technology upon acquisition by foreign companies.

[15] Under this assumption, the production management efficiency is specific to each R&D center. R&D centers with different innovation management efficiencies affiliated with the same parent will draw from different distributions. An alternative interpretation of this production management efficiency is the quality of products developed by an R&D center.

Given the production and innovation efficiency of affiliated R&D centers, (z^P, z^R), firms recruit researchers in each center to develop new differentiated varieties, and decide which countries to sell their products to. To sell products to destination country d, a per-variety fixed marketing cost of c_d^M in terms of country d production labor needs to be paid.

As Figure 1.2b indicates, firms can potentially manufacture products developed by their R&D centers in a third country l , where they do not necessarily perform R&D, and then export to destination countries. By separating production from R&D (offshore production), firms can take advantage of cheaper production labor and save on shipping fees. However, geographic separation makes it difficult for R&D centers to communicate with production plants, reducing production efficiency. I use $\phi_{il}^P \leq 1$ to denote the fraction of productivity that a firm can transfer from its R&D center in country i to production site in country l. For an R&D center with production efficiency z^P, the preserved *plant-level* offshore productivity in country l is $z^P \phi_{il}^P$. I further assume that there is a stochastic element, η_l, idiosyncratic to a production site and a variety, which enters productivity multiplicatively, so the *variety-level* productivity in l is $z^P \phi_{il}^P \eta_l$. The cost of producing and delivering one unit of product is $\frac{w_l^P \tau_{ld}}{z^P \phi_{il}^P \eta_l}$, which takes into account the cost of production labor, w_l^P, and shipping fee, τ_{ld}.

In the model, firms perform offshore R&D for several reasons. First, if a country is relatively abundant in talented inventors, foreign firms might want to enter to make full use of their skills. Anecdotes abound about MNCs establishing offshore R&D centers in order to tap into the local talent pool. Google, for instance, recently announced a plan to train two million Android developers in India within the next three years. According to a survey of 200 R&D executives (Thursby and Thursby, 2006), MNCs rank being close to highly qualified R&D personnel as the most important factor for the location choice of R&D centers in their home countries and other developed countries, and as the second most important factor, right after growth potential, for their new R&D centers in emerging economies.

The aforementioned production and trade decisions also imply that firms might choose to perform R&D in places close to major destination markets, or places with good access to countries with cheap production labor, in order to produce and distribute their products more efficiently.

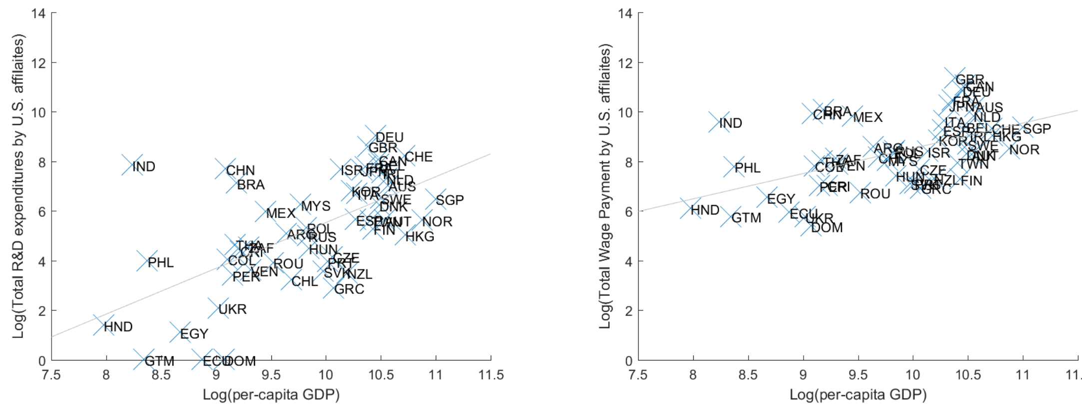
By allowing for both offshore R&D and production decisions, the model captures the complexity of multinationals' global strategies. This stands in contrast to existing quantitative studies of multinationals that do not allow for offshore R&D. Such restriction might not be important if R&D activities performed by foreign affiliates are simply product adaption to local markets, a "by-product" of offshore production. Figure 1.3 demonstrates that, while U.S. multinationals' R&D expenditures and total wage payment both increase with host income, the former increases much faster (note that the two panels are on the same scale, so the slope of the fitted line in the first panel is twice of that in the second panel). This figure highlights that product adaption unlikely to be the whole story, and offshore R&D is not simply a by-product of offshore production.

Importantly, I assume that different varieties developed by a firm, either in the same or in different R&D centers, are differentiated from each other and from vari-

Figure 1.3: Overseas R&D and Employment by U.S. Multinationals

(a) U.S. Multinational Affiliate R&D

(b) U.S. Multinational Affiliate Wage Payment



Notes: The left panel plots the log of total R&D expenditures by U.S. multinationals in each host country against host income. The right panel plots the total wage payment of U.S. multinationals against host income. Data source: Bureau of Economic Analysis.

eties developed by all other firms. Such an assumption is consistent with how R&D is organized in many multinational firms. General Electric, for example, organizes its ten research labs by scientific disciplines in five countries (the U.S., Germany, India, China, and Brazil).[16], [17] This assumption implies that firms make offshore R&D decisions for each country independently and that R&D centers affiliated with the same firm operate as if they are independent from each other.

Given this independence, in the remainder of this section, I first consider the

[16]Alternatively, this assumption can be interpreted as capturing M&A FDI. More than 70% of FDI flows in the data are in the form of mergers and acquisitions (Nocke and Yeaple, 2008). One explanation for this observation is that, by transferring know-how and managerial capacity to targets, acquiring firms can improve the operating efficiency of the targets. The differentiated-variety assumption adopted in the present chapter is consistent with this perspective of FDI—multinationals transfer their managerial technology to newly acquired foreign R&D centers, and increase the efficiency of these R&D centers in carrying out their independent product development.

[17]This assumption treats R&D at headquarters and R&D in offshore centers symmetrically. Recently, Bilir and Morales (2016) estimates the effects of R&D on productivity for multinational firms. They find that R&D at headquarters have stronger spillover effects to foreign affiliates than R&D at affiliates to other affiliates. The current model cannot account for this finding. But an extension of the model that allows firms to first invest in R&D to build up ”core management capacity” before performing product innovation at home and abroad would be consistent with this finding.

production and trade decision of a firm, after a variety has been developed. I then describe the innovation decision of each R&D center, and firms' decisions to build offshore R&D centers. Finally, I characterize the market for researchers and analyze the welfare gains from openness under a special case.

1.3.3 Production and Trade

Consider a variety developed by an R&D center (z^P, z^R) in country i, which can potentially be produced in any country by production labor using a linear production technology. For each variety, an R&D center obtains a vector of N idiosyncratic productivity draws, one for each potential production site, denoted $\boldsymbol{\eta} = (\eta_1, \eta_2, .., \eta_N)$. I assume that η_l is independent across countries, and follows a Frechet distribution: $F(x) \equiv \text{Prob}(\eta \leq x) = \exp(-\Lambda_l x^{-\delta})$, where Λ_l governs the mean of the draws for country l, and δ governs the dispersion of the draws across varieties and countries. The productivity for a variety produced in country l is: $z^P \phi_{il} \eta_l$.

Letting w_l^p denote the wage rate for each unit of production labor in country l, the cost of serving country d by producing in country l is $c_{ild} = \frac{w_l^p \tau_{ld}}{z^P \phi_{il}^p \eta_l}$, where τ_{ld} is the iceberg shipping cost from l to d. Given the monopolistic competition market structure, the price for a variety sold in country d, if produced in country l, is

$$p_{ild} = \frac{\sigma}{\sigma - 1} \frac{w_l^p \tau_{ld}}{z^P \phi_{il}^p \eta_l}.$$

Conditional on serving destination market d, a firm chooses the lowest cost production location for each of its varieties. Because there are no fixed costs in offshore production, all countries are potentially production sites. The price of this variety in country d is simply the lowest one among all possible choices:

$$p_{id}(\boldsymbol{\eta}) = \min_l \{ \frac{\sigma}{\sigma-1} \frac{w_l^p \tau_{ld}}{z^P \phi_{il}^p \eta_l} \}.$$

For each variety and each destination market, production will take place in one country. However, since each R&D center develops a continuum of varieties, in equilibrium, a firm will serve each destination through all countries in the world.[18] For tractability, I assume that each R&D center needs to decide first which destination markets to enter and pays the fixed marketing cost before knowing the idiosyncratic country-specific productivity draws, so firms make destination market entry decisions based on expected profits. The expected per-variety profit from market d for the R&D center from country i , defined as $\pi_i^d(z^P)$, is

$$\pi_i^d(z^P) = \frac{1}{\sigma}(\frac{\sigma}{\sigma-1})^{1-\sigma}\Gamma(\frac{\delta+1-\sigma}{\delta})P_d^{\sigma-1}X_d(\frac{1}{z^P})^{1-\sigma}\boldsymbol{\Psi}_{id}^{\frac{\sigma-1}{\delta}} - c_d^M w_d^p,$$

where Γ is the Gamma function, and $\boldsymbol{\Psi}_{id} = \sum_l \Lambda_l(\frac{w_l^p \tau_{ln}}{\phi_{il}^p})^{-\delta}$. The first term in this expression is calculated from $\frac{1}{\sigma}P_d^{\sigma-1}X_d \int \min_l(p_{ild}(\boldsymbol{\eta})^{1-\sigma})d\tilde{F}(\boldsymbol{\eta})$, with $\tilde{F}(\boldsymbol{\eta})$ being the distribution of $\boldsymbol{\eta} = (\eta_1, \eta_2, ..., \eta_N)$.

This expected profit increases in the production efficiency of an R&D center,

[18] This result implies that the model cannot capture the extensive-margin of firms' offshore *production* decisions. This is not necessarily an important drawback, as the focus of this chapter is on offshore R&D and its interaction with offshore production in the aggregate. In the next chapter I show the model predictions on firms' offshore R&D decisions are supported empirically.

z^P, so there exists a threshold $\hat{z}^p_{id}$ such that R&D centers from i will expend marketing costs and enter country d if and only if their production efficiency is above this threshold. This cutoff is given by:

$$\pi_i^d(\hat{z}^P_{id}) = 0. \tag{1.1}$$

A firm makes an independent entry decision for each destination market. The per-variety expected profit for a firm with production efficiency draw z^P, taking into account its potential entry into all destination markets, is

$$\pi_i(z^P) = \sum_d \mathbb{I}_{z^P \geq \hat{z}^p_{id}} \pi_i^d(z^P). \tag{1.2}$$

1.3.4 Innovation and the Market for Researchers

R&D centers choose the talent of researchers, θ, and their quantity, $l(\theta)$, to develop new differentiated varieties. Let y be the measure of differentiated varieties developed:

$$y = f(z^R, \theta) l(\theta)^\gamma,$$

where γ measures the return to the number of researchers, and $f(z^R, \theta)$ captures how firm innovation efficiency and researcher talent affect innovation output. I assume that $0 < \gamma < 1$, implying decreasing returns to scale in the number of researchers. This assumption has several interpretations. First, it can be thought of as a reduced-form approximation to a model in which R&D requires supervision from top management, but managerial time is limited in a company. In such a context, hiring more researchers results in less supervision time for each of them, reducing researcher productivity.[19] An alternative is to think of innovation output as a function of both accumulated knowledge capital and researcher input. In a static model in which the distribution of knowhow and accumulated knowledge is given, the research output features decreasing returns to researcher input. Finally, decreasing returns to scale might stem from increases in coordination costs, free-riding, and disagreement among researchers as teams expand.[20]

Given $\pi_i(z^P)$, the per-variety expected profit, the optimization problem for the R&D center is

$$\pi_i^R(z^P, z^R) = \max_{\theta\in\Theta, l(\theta)}[\pi_i(z^P)f(z^R,\theta)l(\theta)^\gamma - w_i(\theta)l(\theta)],$$

where $w_i(\theta)$ is the wage for a researcher with talent θ. As is clear from the equation, the production efficiency of a firm affects innovation incentives because it determines the profit for each variety. I make the following assumption about f:

Assumption 1 *f is twice continuously differentiable and increasing in its arguments, i.e., $f_1, f_2 > 0$. Further, f is log-supermodular, i.e., $\frac{\partial^2 log f(z^R,\theta)}{\partial z^R \partial\theta} > 0$.*

The assumption that $f_1, f_2 > 0$ simply means that more efficient firms and more able researchers are more productive in innovation. The log-supermodularity

[19]See Antras et al. (2006) for an analysis of the effects of offshoring in a model in which managers can only supervise a fixed number of workers.

[20]Such coordination costs have been documented empirically. For example, Haas and Choudhury (2015) finds that, while total patenting increases with the number of members in a team, the increase is smaller than the increase in the team size—there is decreasing returns in the number of researchers in a team.

assumption implies strong complementarity between researcher ability and firm efficiency. Under this assumption, more productive firms have a comparative advantage in working with more able researchers.[21] R&D activities require cooperation between researchers, and a large amount of managerial and monetary resources. Moreover, after a product prototype is developed, testing and marketing costs are big hurdles to clear before the product can reach consumers. A well-managed firm can do all of these tasks better, so it is especially profitable for them to work with talented researchers. The model captures this idea with the log-supermodularity of f.

The setup here deviates from the efficiency units assumption. A researcher with high talent is more valuable than multiple researchers with lower talent. Similarly, a firm with high innovation efficiency is more productive in R&D than multiple firms with lower efficiencies. These implications are in line with a few observations in the literature. First, as mentioned earlier, the *quality* of research talent is one of the top considerations when firms choose where to build their offshore R&D centers, along with the *cost* of research labor.[22] Second, it is well documented that there are a large number of small and less productive firms in developing countries, the prevalence of which can account for an important fraction of cross-country income differences (Hsieh and Klenow, 2009). Management efficiency might be a source

[21] The log-supermodularity assumption has been adopted in a growing literature in international trade which uses assignment models to study questions such as the determinants of specialization and the impacts of trade integration, as reviewed recently by Costinot et al. (2015). The framework here is similar to the one in Grossman et al. (forthcoming).

[22] Branstetter et al. (2013) conducts interviews with foreign-affiliated R&D centers in China. The interview responses stress the scale and quality of the research talent in China, rather than its cost.

of performance differences between firms (Bloom et al., 2013). To the extent that many developing countries have a large number of very small firms, they might not necessarily lack a sufficient *stock* of management efficiency. The model here is consistent with view that it is not necessarily a lack of management efficiency stock, but rather the lack of exceptional firms like Apple and Google, that explains the low incomes in developing countries.[23] Finally, complementarity also implies that the same inventor will be paid more to work in a more efficient firm. This is consistent with the finding that larger and more productive firms pay a wage premium (see, for example, Schank et al., 2007), and the evidence on positive assortative matching between firms and inventors I provide in the appendix.

I now characterize the market for researchers. Let $T_i(z^P, z^R) : (\mathbb{Z}_i^P, \mathbb{Z}_i^R) \to \Theta$ be the optimal choice of talent for an R&D center characterized by (z^P, z^R). We have the following lemma:

Lemma 1 (Assortative matching between firms and inventors) *T_i is continuous and strictly increasing in z^R. Moreover, T_i is independent of z^P.*

Proof See appendix.

The proof of Lemma 1 is an extension of assortative matching results in the literature (see, for example, Grossman and Helpman, 2014; Grossman et al., forthcoming; Sampson, 2014) to the case with an additional source of heterogeneity, namely production efficiency. Because high z^R R&D centers enjoy a higher margi-

[23] Roys and Seshadri (2014) builds a model of matching between heterogeneous entrepreneurs and workers, enriched with human capital accumulation, to show that the model can account for the differences in life cycle dynamics between firms in rich and poor countries, and can explain a substantial share of income differences between countries.

nal productivity increase from hiring better researchers, they have a comparative advantage in working with high-ability researchers, leading to assortative matching. Since z^P enters firms' innovation output multiplicatively in the form of $\pi_i(z^P)$, higher z^P does not affect the type of researchers hired by an R&D center, but only their quantity. In the following I will write the matching function simply as $T_i(z^R)$, omitting the argument z^P.

Given the equilibrium $w_i(\theta)$, the demand of an R&D center for researchers, if it chooses researchers with talent θ, is

$$l_i(z^P, z^R) = (\frac{\gamma \pi_i(z^P) f(z^R, \theta)}{w_i(\theta)})^{\frac{1}{1-\gamma}}. \tag{1.3}$$

The corresponding measures of invention and profit are therefore:

$$y_i(z^P, z^R) = (\gamma \frac{\pi_i(z^P)}{w_i(\theta)})^{\frac{\gamma}{1-\gamma}} f(z^R, \theta)^{\frac{1}{1-\gamma}}, \tag{1.4}$$

$$\pi_i^R(z^P, z^R) = (\gamma^{\frac{\gamma}{1-\gamma}} - \gamma^{\frac{1}{1-\gamma}}) w_i(\theta)^{-\frac{\gamma}{1-\gamma}} [\pi_i(z^P) f(z^R, \theta)]^{\frac{1}{1-\gamma}}. \tag{1.5}$$

In equilibrium, firms choose the type of researchers to maximize profit. This requires the improvement in marginal output from higher-quality researchers to be exactly offset by their higher wages. We can obtain this equation by differentiating Equation (1.5) with respect to θ:

Lemma 2 (Optimal talent choice) $w_i(\theta)$ *satisfies the following relationship:*

$$\frac{w_i'(\theta)}{w_i(\theta)} = \frac{f_2(z^R,\theta)}{\gamma f(z^R,\theta)}|_{\theta=T_i(z^R)}. \tag{1.6}$$

Proof See appendix.

The formal proof of Lemma 2 establishes the differentiability of $w_i(\theta)$. The proof is similar to that in Sampson (2014) and is relegated to the appendix.

Since researchers are heterogeneous, labor market clearing requires that the total demand equals total supply for each type. Let $\underline{\theta}_i$ and $\overline{\theta}_i$ be the lower and upper limits of the support for the researcher talent distribution, and let $\underline{z}_i^R$ and $\overline{z}_i^R$ denote the lower and upper limit of the support for the innovation efficiency distribution, respectively. To derive the researcher market clearing conditions for each type, I start with an aggregate version: for all $\underline{\theta}_i < \theta < \overline{\theta}_i$, the number of researchers with talent lower than θ is equal to the total demand for researchers with talent below θ. Formally,

$$\begin{aligned} L_i^R \int_{\underline{\theta}_i}^{T_i(z^R)} dH_i(\theta) &= R_i \int_{\underline{z}_i^R}^{z^R} [\int_{\mathbf{Z}^P} l_i(z^P,z) g_i^P(z^P|z) dz^P] g_i^R(z) dz \\ &= R_i \int_{\underline{z}_i^R}^{z^R} (\frac{\gamma f(z,T_i(z))}{w_i(T_i(z))})^{\frac{1}{1-\gamma}} [\int_{\mathbf{Z}^p} \pi_i(z^p)^{\frac{1}{1-\gamma}} g_i^P(z^P|z) dz^p] g_i^R(z) dz, \end{aligned}$$

where R_i is the measure of R&D centers in country i and $g_i^R(z)$ is their PDF, both of which are determined in equilibrium by firms' offshore R&D decisions. On the left side of this equation is the total number of researchers with talent below $T_i(z^R)$, and on the right side is the corresponding total demand.

Differentiating this equation with respect to z^R, we have the following equation that characterizes $T'(z^R)$:[24]

$$L_i^R T'(z^R) h_i(T_i(z^R)) = R_i(\frac{\gamma f(z^R, T_i(z^R))}{w_i(T_i(z^R))})^{\frac{1}{1-\gamma}} \int_{\mathbf{Z}^P} g_i^R(z^R) \pi_i(z^P)^{\frac{1}{1-\gamma}} g^P(z^P|z^R) dz^P \tag{1.7}$$

Equation (1.7) then characterizes the market clearing condition for each researcher type. Equations 1.6 and 1.7, together with two boundary values,

$$T_i(\underline{z}_i^R) = \underline{\theta}_i, \quad T_i(\overline{z}_i^R) = \overline{\theta}_i, \tag{1.8}$$

determine the matching function $T_i(z^R)$ and the wage schedule $w_i(\theta)$. In summary, we have the following results:

Proposition 1 *Under Assumption 1,*

1) Firms with higher innovation efficiency hire strictly better researchers. Firms with the same innovation efficiency but different production efficiencies hire the same type of researchers in different quantities.

2) The researcher labor market is characterized by Equations 1.6, 1.7, and 1.8.

How does the output of R&D centers with different innovation efficiencies depend on the talent distribution of a country? Since a change in the talent distribution affects the entire matching function, characterizing the effect of a general

[24] Because of offshore R&D decisions, g_i^R is not necessarily continuous. At the finite discontinuous points of g_i^R, the matching function might not be differentiable. In this case, Equation 1.7 is not defined on the discontinuous points of g_i^R. While T_i is still well defined and continuous, the kinks in T_i' make it challenging to solve the matching function numerically. In the quantitative section, I describe a computational algorithm suited for this context.

change is difficult. I consider changes to the talent distribution that can be ranked by the following criterion:

Definition 1 *Consider $h(\theta)$ and $\tilde{h}(\theta)$, probability density functions for the talent distribution. $\tilde{h}(\theta)$ is more talent abundant relative to $h(\theta)$, if $\frac{\tilde{h}(\theta_2)}{\tilde{h}(\theta_1)} > \frac{h(\theta_2)}{h(\theta_1)}, \forall \theta_1 < \theta_2$.*

This notion of factor abundance, which is stronger than first-order stochastic dominance, is introduced by Costinot and Vogel (2010) to characterize how relative factor supply and factor demand determine allocation and prices. According to this definition, a more talent abundant distribution has a higher relative share of the higher-skill type than a less talent abundant distribution. Letting $y(z^P, z^R; h(\theta))$ denote the measure of R&D output by an R&D center (z^P, z^R) when the talent distribution is $h(\theta)$. We have the following proposition:

Proposition 2 *Consider two R&D centers in country i, with innovation efficiencies $z_2^R > z_1^R$ and a common production efficiency z^P. Further assume that $\hat{h}(\theta)$ is more talent abundant than $h(\theta)$. Then $\frac{y(z^P, z_2^R; \hat{h}(\theta))}{y(z^P, z_1^R; \hat{h}(\theta))} > \frac{y(z^P, z_2^R; h(\theta))}{y(z^P, z_1^R; h(\theta))}$, if either 1) z^P and z^R are independent; or 2) There are no fixed marketing costs, that is, $c_d^M = 0$.*

Proof See appendix.

Proposition 2 states that, the R&D output of firms with high z^R relative to that of firms with low z^R is larger, when the researcher distribution is more talent abundant. The intuition for this result is that, under the additional conditions stated above, increases in talent abundance improve the quality of researchers for all firms. This benefits efficient firms disproportionately more, because of the complementarity between talent and efficiency.[25]

Importantly, since this proposition works through improving of match quality for firms, it also applies to a change in the firm innovation efficiency distribution that results in improvements in match quality for all firms between (z_1^R, z_2^R). An example of such a change is a decrease in the "efficiency abundance" of the firm distribution in the spirit of Definition 1. The talent abundance in the proposition should thus be broadly interpreted as a *relative* measure—the log difference of R&D output between the two centers increases as the talent abundance of the inventor distribution increases, relative to the efficiency abundance of the firm distribution.

Although Proposition 2 is stated in the context of domestic firms, it applies to all active R&D centers in a host country. We can test the model by comparing the innovation output of R&D centers affiliated with companies with different innovation efficiencies. If Proposition 2 is correct, then this difference will be larger in host countries with higher relative talent abundance. In the next section I show that the complementarity channel underlying this prediction is quantitatively relevant in determining the pattern of offshore R&D across host countries with different talent distributions. In the next chapter, I test this implication directly and show that the data supports this prediction.

[25] Grossman et al. (forthcoming), Sampson (2014), and Costinot and Vogel (2010) obtain similar results on the effects of trade on income inequality under the log-supermodularity assumption. Compared to these papers, additional technical assumption is needed to ensure that the general equilibrium changes in return to R&D, $\pi_i(z^P)$, due to the distribution change do not decrease the quality of match for any firms.

1.3.5 Offshore R&D

Now we can characterize firms' decisions to open offshore R&D centers. I make the following assumption about $g^P(z^P|z^R)$.

Assumption 2 *The distribution from which an R&D center draws its production efficiency z^P increases in its innovation efficiency in the sense of first-order stochastic dominance.*

Define $\pi_i^R(z^R)$ as the expected profit (over the possible z^P draws) for an R&D center in country i, with innovation efficiency z^R:

$$\pi_i^R(z^R) = \int_{\mathbf{Z}^P} \pi_i^R(z^P, z^R) g^P(z^P|z^R) dz^P$$

Firms compare the expected profit from building an offshore R&D center to the fixed cost of setting up the center, $c_i^R w_i^P$. By definition (Equations 1.5 and 1.2) , $\pi_i^R(z^P, z^R)$ increases in z^P. We can also show that $\pi_i^R(z^P, z^R)$ increases with z^R.[26] Assumption 2 then implies that $\pi_i^R(z^R)$ increases strictly in z^R, so the decision to offshore R&D follows a threshold rule: there exists a cutoff $\hat{z}_{oi}^R$, so that a firm from country o will perform offshore R&D in country i if and only if its innovation efficiency is above $\hat{z}_{oi}^R$. This cutoff is given by the following zero profit condition:

$$\pi_i^R(\hat{z}_{oi}^R \phi_{oi}^R) = c_i^R w_i^P. \tag{1.9}$$

[26] From Equation 1.5, $\frac{\partial \log(\pi_i^R(z^P, z^R))}{\partial z^R}|_{\theta = T_i(z^R)} = \frac{1}{1-\gamma} \frac{\partial \log(f(z^R, \theta))}{\partial z^R} > 0$.

1.3.6 R&D Center Efficiency Distribution

Firms' offshore R&D decisions determine g_i^R, the distribution of innovation management efficiency, and hence the distribution of production management efficiency, in each country. Given $\hat{z}_{oi}^R$, we can now derive R&D centers' production and innovation efficiency distributions. Let $G_i^R(z^R)$ be the CDF for innovation management efficiency of the R&D centers active in country i, and let $G_o^E(\tilde{z}^R)$ be the CDF of the distribution of innovation efficiency for firms from country o. Then we have the following equation:

$$R_i G_i^R(z^R) = \sum_{o=1}^{N} \mathbb{I}_{\frac{z^R}{\phi_{oi}^R} > \hat{z}_{oi}^R} E_o G_o^E(\frac{z^R}{\phi_{oi}^R})$$

Differentiating this equation with respect to z^R, we obtain the density function:

$$g_i^R(z^R) = \frac{1}{R_i} \sum_{o=1}^{N} \mathbb{I}_{\frac{z^R}{\phi_{oi}^R} > \hat{z}_{oi}^R} E_o g_o^E(\frac{z^R}{\phi_{oi}^R}) \frac{1}{\phi_{oi}^R}. \tag{1.10}$$

The PDF for R&D centers with (z^P, z^R) is $g_i(z^P, z^R) = g^P(z^P|z^R) g_i^R(z^R)$.

1.3.7 Aggregation

Knowing $g_i(z^P, z^R)$, I derive the total measure of varieties that are invented in a country, denoted M_i, and the distribution of these varieties over different production efficiencies. Letting $m_i(z^P)$ be the measure of varieties innovated in country

i by R&D centers with a production efficiency of z^P, then we have:

$$\begin{aligned} m_i(z^P) &= R_i \int_{\mathbb{Z}^R} y_i(z^P, z^R) g_i(z^P, z^R) dz^R \\ M_i &= \int_{\mathbb{Z}^P} m_i(z^P) dz^P, \end{aligned} \tag{1.11}$$

where $y_i(z^P, z^R)$ is given by Equation 1.4. The price index in country d is then given by the following equation:

$$\begin{aligned} P_d^{1-\sigma} &= \sum_i \int_{z^P > \hat{z}_{id}^p} m_i(z^P) [\int \min_l \{p_{ild}(\boldsymbol{\eta})^{1-\sigma}\} d\tilde{F}(\boldsymbol{\eta})] dz^P \\ &= \Gamma(\frac{\delta + 1 - \sigma}{\delta})(\frac{\sigma}{\sigma - 1})^{1-\sigma} \sum_i \Psi_{id}^{\frac{\sigma-1}{\delta}} \int_{z^P > \hat{z}_{id}^p} m_i(z^P) z^{P^{\sigma-1}} dz^P \end{aligned} \tag{1.12}$$

To express the aggregate objects in the model, let X_{id} be the total sales in country d of the products developed in country i. We have the following:

$$\begin{aligned} X_{id} &= P_d^{\sigma-1} X_d \int_{\hat{z}_{id}^P}^{\infty} m_i(z^P) [\int \min_l \{p_{ild}(\boldsymbol{\eta})^{1-\sigma}\} d\tilde{F}(\boldsymbol{\eta})] dz^P \\ &= (\frac{\sigma}{\sigma - 1})^{1-\sigma} \Gamma(\frac{\delta + 1 - \sigma}{\delta}) P_d^{\sigma-1} X_d \Psi_{id}^{\frac{\sigma-1}{\delta}} \int_{\hat{z}_{id}^P}^{\infty} m_i(z^P)(z^P)^{\sigma-1} dz^P. \end{aligned} \tag{1.13}$$

These sales can be fulfilled through production in any country. Letting X_{ild} denote the value of production in country l, then we have $\sum_l X_{ild} = X_{id}$. I further define Y_l to be the total production of the varieties in country l, so $\sum_{i,d} X_{ild} = Y_l$. The Frechet assumption on idiosyncratic productivity draws also implies that, for each R&D center located in country i, the share of products it sells in country d

that are fulfilled through production in country l is:

$$\psi_{ild} = \frac{\Lambda_l(\frac{w_l^P \tau_{ld}}{\phi_{il}^P})^{-\delta}}{\boldsymbol{\Psi}_{id}},$$

with $\boldsymbol{\Psi}_{id} = \sum_l \Lambda_l(\frac{w_l^p \tau_{ld}}{\phi_{il}^p})^{-\delta}$. Because this probability is the same for all R&D centers from country i, it also applies to the aggregate sales:

$$X_{ild} = \psi_{ild} X_{id}. \tag{1.14}$$

Production workers are used to produce output, and to pay fixed R&D and marketing costs. The production labor market clearing condition is:

$$w_d^P L_d^P = \underbrace{\frac{\sigma-1}{\sigma} Y_d}_{\text{Production}} + \underbrace{\sum_o E_o c_d^R w_d^P (1 - G_o^R(\hat{z}_{od}^R))}_{\text{Fixed R\&D center setup costs}} + \underbrace{c_d^M w_d^P \sum_i \int_{\hat{z}_{id}^P}^{\infty} m_i(z^P) dz^P}_{\text{Fixed marketing costs}}. \tag{1.15}$$

Recall that the density of firms from country o with innovation efficiency $\tilde{z}^R$ is $g_o^E(\tilde{z}^R)$. We can integrate π_i^R over $g_o^E(\tilde{z}^R)$ to compute the total profits made by country i R&D centers affiliated to firms from country o, denoted Π_{oi}:

$$\Pi_{oi} = E_o \int_{\hat{z}_{oi}^R}^{\bar{\tilde{z}}_i^R} \pi_i^R(\tilde{z}^R \phi_{oi}^R) g_o^E(\tilde{z}^R) d\tilde{z}^R,$$

This profit is after deducting R&D, marketing, and production costs, but before deducting fixed costs for building R&D centers.

Let I_i be the total R&D expenditures in country i, defined as total compensation to researchers in country i. Let I_{oi} be the expenditures in I_i that are incurred by affiliates of firms from country o. Equations 1.3 and 1.4 imply that:

$$I_i = \sum_o I_{oi} = \frac{\gamma}{1-\gamma} \sum_o \Pi_{oi}$$

The income of country d comes from three sources: wages of production labor, compensation to researchers, and the net profit made by domestic firms from the country. Current account balance requires that total consumption of each country equals total income:

$$X_d = \underbrace{w_d^P L_d^P}_{\text{Production Labor}} + \sum_i \underbrace{[\Pi_{di} - E_d c_i^R w_i^P (1 - G_d^R(\hat{z}_{di}^R))]}_{\text{Net Profit}} + \underbrace{I_d}_{\text{Researcher Compensation}} . \tag{1.16}$$

Definition 2 *The competitive equilibrium is defined as a set of allocations and prices, such that:*

1. *Firms' market entry decisions satisfy Equation 1.1.*

2. *The matching function, T_i, and wage schedule for researchers, w_i satisfy Equations 1.6, 1.7, and 1.8.*

3. *Firms' offshore R&D decisions satisfy Equation 1.9.*

4. *The distribution of R&D center innovation efficiency in each country satisfies Equation 1.10.*

5. *The distribution of productivity efficiency for varieties satisfies Equation 1.11.*

6. *The price index in each country satisfies Equation 1.12.*

7. *The wage for production labor satisfies Equation 1.15.*

8. *The total expenditure in each country satisfies Equation 1.16.*

1.3.8 The Gains from Openness

In this subsection I focus on a special case to derive an expression for the welfare gains from openness, defined as the percentage change in real income ($\frac{X_d}{P_d}$), as a country moves from complete isolation to the degree of openness observed in the data. This expression makes it clear that offshore R&D is a new channel for countries to benefit from globalization. It also relates the size of this benefit to observable information and model parameters. Specifically, I make the following assumption:

Assumption 3 *1)* $f(z^R, \theta) = z^R \theta^\beta$*;*

2) Production efficiency, z^P*, is independent of* z^R*, and follows a Pareto distribution:* $G_d^P(x) = 1 - (\frac{x}{\underline{z}_d^P})^{-\kappa_P}$;

3) There is no fixed marketing cost: $\forall d, c_d^M = 0$;

4) Firm innovation efficiency, $\tilde{z}^R$*, follows a Pareto distribution:* $G_d^E(x) = 1 - (\frac{x}{\underline{\tilde{z}}_d^R})^{-\kappa_R}$.

The first part of the assumption maintains that $f(z^R, \theta)$ takes a multiplicative form.[27] Under this assumption, $\frac{\partial^2 \log f(z^R,\theta)}{\partial z \partial \theta} = 0$, so $f(z^R, \theta)$ no longer satisfies the strict log-supermodularity requirement in Assumption 1. Since a CES function with elasticity of substitution smaller than 1 satisfies strict log-supermodularity, the multiplicative case represents the limiting case as the elasticity approaches 1. This simplification will allow us to solve for the equilibrium wage schedule and firm-level decisions analytically.

In the general model, because firms endogenously choose how many varieties to develop, aggregation is difficult. The first three components of Assumption 3, however, imply a Pareto distribution of production efficiency for varieties, which admits analytical aggregation. The fourth component in turn allows us to derive the total fixed costs of R&D in each country. With these simplifications, we have the following:

Proposition 3 *Under Assumption 3, the gains from openness for country d, defined as the percentage change in $\frac{X_d}{P_d}$ as a country moves from complete isolation to the observed equilibrium, is*

$$GO_d = \underbrace{\left(\frac{X_{ddd}}{\sum_l X_{dld}}\right)^{-\frac{1}{\delta}} \left(\frac{\sum_l X_{dld}}{X_d}\right)^{-\frac{1}{\sigma-1}} \left(\frac{I_{dd}}{I_d}\right)^{-\frac{1-\gamma}{\sigma-1}}}_{\textit{Direct Effect}} \overbrace{\left(\frac{\frac{\sigma-1}{\sigma}}{\frac{\sigma-1}{\sigma}\frac{Y_d}{X_d} + \frac{(1-\gamma)\kappa_R - 1}{\gamma\kappa_R}\frac{I_d}{X_d}\left(1-\frac{I_{dd}}{I_d}\right)}\right)}^{\textit{Indirect Effect}} - 1. \tag{1.17}$$

Proof See appendix.

This expression highlights various forces through which a country benefits from

[27] The assumption that the power of z^R is 1 is without loss of generality, because the units of z^R can always be scaled so that it enters $f(z^R, \theta)$ with a power of 1.

economic integration. The first term, $\frac{X_{ddd}}{\sum_l X_{dld}}$, captures the benefits from offshore production for consumption. The second term, $\frac{\sum_l X_{dld}}{X_d}$, captures the benefits from foreign innovation for consumption. These two terms are direct effects of offshore production and trade in the model. The third term, $\frac{I_{dd}}{I_d}$, captures the importance of foreign firms in domestic R&D. Intuitively, the smaller is this ratio, the more a country relies on foreign affiliates for R&D, and the more significant are the welfare gains from offshore R&D. The last term in the equation captures the effects of profit flows on welfare through their impacts on total expenditures. This indirect effect tends to bring positive welfare impacts, for countries that specialize in R&D (smaller $\frac{Y_d}{X_d}$), and countries that rely more on domestic firms in R&D (smaller $\frac{I_d - I_{dd}}{X_d}$).[28]

I compare the gains-from-openness expression in this model to the expression in Ramondo et al. (2015) and Arkolakis et al. (2014), both of which feature trade and offshore production, but not offshore R&D. Their formulas are as follows:

$$GO_d = \overbrace{\underbrace{[(\frac{X_{ddd}}{X_d})^{-\frac{1}{\delta}}(\frac{\sum_l X_{dld}}{X_d})^{\frac{1}{\delta}-\frac{1}{\kappa_P}}]}_{\text{Direct Effect (Ramondo et al., 2015)}}\,[(\frac{X_d}{Y_d})^{\frac{\kappa_P+1-\sigma}{\sigma-1}\frac{1}{\kappa_P}+1}]}^{\text{Arkolakis et al. (2014)}} - 1 \qquad (1.18)$$

Like Equation 1.17, this equation consists of a direct and an indirect effect. There are three main differences between the two equations. First and most important, Equation 1.17 features an extra term, $\frac{I_{dd}}{I_d}$, the gains from having foreign affiliates doing R&D domestically. The second difference is that, the power on the direct effect is different across these two equations. Specifically, κ_P, the dispersion

[28]Finite aggregate fixed R&D costs require $(1-\gamma)\kappa_R - 1 > 0$.

parameter for production efficiency distribution, does not appear in Equation 1.17. This is because by assuming away the fixed marketing costs, the extensive margin of exporting vanishes, and the elasticity of substitution between varieties alone determines trade elasticity. Third, while in Equation 1.18, the strength of the indirect effect only depends on $\frac{Y_d}{X_d}$, in Equation 1.17, it depends on $\frac{I_d}{X_d}$ and $\frac{I_{dd}}{I_d}$, too.

The comparison across the two expressions highlight the novel role of offshore R&D for countries to benefit from globalization. To see this, consider two calibrations, based on my model and the model in Arkolakis et al. (2014), respectively. With flexible international frictions, both model are able to match the observed bilateral linkages perfectly, so $\frac{X_{ddd}}{X_d}$ and $\frac{\sum_l X_{dld}}{X_d}$ are both equal to the data. As K_R approaches $\frac{1}{1-\gamma}$ from above,[29] the indirect effect in Equation 1.17 converges to $\frac{X_d}{Y_d}$, the indirect effect in Equation 1.18. Under suitable choice of κ_P and σ, the only remaining difference between the two expressions is that in Equation 1.17, there is an extra term $\frac{I_{dd}}{I_d}$, which captures the importance of foreign firms in domestic R&D activities.

To have an idea of how large this term is, consider the median country in the quantitative section, with about 30% of its R&D done by foreign affiliates. The value of $(\frac{I_{dd}}{I_d})^{-\frac{1-\gamma}{\sigma-1}}$ is around 1.055, when $\gamma = 0.4$ and $\sigma = 5$. All else equal, this term generates a 5% real income change. So offshore R&D indeed represents a quantitatively important channel through which countries benefit from global integration.

[29] κ_R must be greater than $\frac{1}{1-\gamma}$ for the equilibrium to be well defined.

1.4 Parameterization

I perform a quantitative analysis of the determinants and the welfare implications of offshore R&D, using the model developed in Section 1.3. I focus on a sample with 25 countries and a statistical aggregation of another 22 countries.[30] I parameterize the model to be consistent with the data in its predictions on the interactions between countries and the size distribution of firms within the U.S. This section describes the parameterization procedures, starting with the functional form assumptions.

1.4.1 Additional Assumptions

In the quantification, I embed an occupational choice into the model. Throughout the rest of the chapter, I assume that each country is endowed with L_i number of workers, with talent distribution $H_i(\theta)$. Workers sort into production labor or research. Each production worker has one unit of production labor, and each researcher has θ units of talent in research. Adding occupational choice generates endogenous responses in the supply of inventors in the counterfactual experiments.

The function $f(z^R, \theta)$ determines the complementarity between the innovation management efficiency of firms and the talent of researchers. I assume that f is a CES function with elasticity of substitution $\alpha < 1$:

$$f(z^R, \theta) = ({z^R}^{\frac{\alpha-1}{\alpha}} + \theta^{\frac{\alpha-1}{\alpha}})^{\frac{\alpha}{\alpha-1}}.$$

[30] The list of countries in this statistical aggregation is provided in the appendix.

This specification satisfies the log-supermodularity assumption. As α approaches 1, the complementarity between researcher talent and firm efficiency weakens.

The capture the long right tail in inventor and firm R&D output size distribution, the distributions of worker talent and firm innovation efficiency are parameterized to be truncated Pareto distributions. A more commonly used parameterization in the literature is the Pareto distribution. By truncating the distribution at a potentially arbitrarily large value that is determined by the data, I reduce the computational burden of solving the matching problem. The specifications are given below:

$$H_i(\theta) = \frac{(\underline{\theta}_i^{-\kappa_i^\theta} - \theta^{-\kappa_i^\theta})}{(\underline{\theta}_i{}^{-\kappa_i^\theta} - \overline{\theta}_i{}^{-\kappa_i^\theta})},\ G_i^E(z^R) = \frac{(\underline{z}_i^{R^{-\kappa_i^R}} - z_i^{R^{-\kappa_i^R}})}{(\underline{z}_i^{R^{-\kappa_i^R}} - \overline{z}_i^{R^{-\kappa_i^R}})}.$$

In these expressions, the letters with upper and lower bars indicate the upper and lower bounds for their respective distributions. κ_i^R and κ_i^θ are the truncated-Pareto counterparts to the shape parameter in the Pareto distribution.

To capture the correlation between innovation and production efficiency at the firm level, I assume that there are two distributions, indicated by H and L (for high and low, respectively), from which firms draw their productivity z^P. The probability of drawing from the high distribution depends on a firm's innovation efficiency in the following fashion:

$$\text{Prob}(z^P \in H|z^R) = \frac{\exp(A + B \times z^R)}{1 + \exp(A + B \times z^R)}, \tag{1.19}$$

where A and B are parameters to be estimated. A positive value for B means that more innovative firms tend to be more productive as well. H and L are both Pareto distributed with the same shape parameter κ^P:

$$G_H(z^P) = 1 - (\frac{\underline{z}_H^P}{z^P})^{\kappa_P},\ G_L(z^P) = 1 - (\frac{\underline{z}_L^P}{z^P})^{\kappa_P}.$$

I assume that $\underline{z}_L^P < \underline{z}_H^P$, so the H distribution first-order stochastically dominates the L distribution.

1.4.2 Parameters Assigned Directly

I set the number of workers in a country, L, to total employment from the Penn World Tables. To focus on differences in the firm efficiency distributions and to abstract from differences in the number of firms, I set E, the measure of firms, to be proportional to L. This proportion is chosen so that the average employment per firm in the model equals the average employment per firm in the U.S.

I directly assign values to a few parameters in the model. Parameter σ, the elasticity of substitution between varieties, determines the markup charged by firms. I set this parameter to be 5, following recent studies in international trade (see Simonovska and Waugh, 2014, for example). This value also implies that 20% of sales are variable profits. In the U.S., R&D expenditures account for about 8% of manufacturing sales. The model counterpart of R&D expenditures is researcher compensation. I set γ, the share of researcher compensation in variable profit, to 0.4, so that researcher compensation accounts for about 8% of sales in the model.

Equation 1.14 implies $\log(X_{ild}) = \alpha_{id} + \beta_{il} - \delta \log(\tau_{ld})$, where α_{id} and β_{ld} are pair fixed effects. δ therefore determines the elasticity of X_{ilds} with respect to the cost of shipping from l to d. Based on Arkolakis et al. (2014), which estimates this specification using the affiliate production and sales data of U.S. multinationals ($i = U.S.$), I set δ to 10.9.

Calibrating firm efficiency and worker talent distributions for each country requires comparable data across countries. I use the World Management Survey by Bloom et al. (2012) and the cognitive test score data by Hanushek and Woessmann (2012) to calibrate these distributions, as explained below.

The World Management Survey provides firm-level management scores for each country in the sample. In the survey, interviewers rate each firm based on its talent management policy and production efficiency along various dimensions. The overall management score for a firm is then averaged over these subscores. The talent management score intends to capture whether firms follow good managerial practice for retaining and incentivizing its talent, so it is closely related to whether a firm is able to make full use of its research talent. I use it to calibrate innovation management distributions. I obtain three distribution statistics of $\tilde{z}^R$ for each country: mean, standard deviation, and skewness.

I use firm-level talent and production management scores to estimate A and B, the coefficients linking a firm's innovation and production efficiencies. Specifically, I classify a firm as being from a high productivity distribution if its production management score falls into the top 1% of the distribution in the world (the top 4% in the U.S.). I then estimate the relationship between a firm's innovation management score and the probability that it is from a high productivity distribution using the Logit model given by Equation 1.19. This procedure determines $A = -6.3$, $B = 0.167$.[31]

In using the World Management Survey to calibrate firm efficiency distribution, I take the stand that survey scores capture fundamental differences about management practises across countries which, in turn, lead to different innovation and production performances. Consistent with this assumption, Bloom et al. (2012) shows that the average management score of a country correlates strongly with per-capita income. An alternative calibration strategy to this approach would be to use a measure for the output of R&D—for example, the patent data—to capture firms' innovation efficiency distribution in a country. This alternative approach has two shortcomings compared to the current approach. First, the interpretation of R&D in this model is broader than activities that generate patents. In the model, firms perform only R&D and manufacturing. This R&D should therefore be interpreted as non-production activities that add values to products, including product invention, development, marketing, etc. A firm's efficiency in all these activities likely depend on its management practise, captured by the World Management Survey. Using patent data would miss an important part of this difference among countries. Second, and perhaps more importantly, firms that apply for patents at the USPTO are a selected subset of all firms. Such selection depends on firms' costs and benefits

[31] The choice of the top 1% cutoff is motivated by the importance of the most productive firms in international business and in production in general. A high cutoff allows me to better capture the distribution of the very top firms. The implicit assumption underlying this calibration strategy is that firms drawing their production efficiency from the L distribution constitute the bottom 99% in the production efficiency distribution, whereas firms drawing from the H distribution constitute the top 1% of production efficiency. This assumption does not hold exactly because under the Pareto assumption, $G_L(z^P)$ will always overlap with $G_H(z^P)$. Given the choice of the cutoff (1%), however, the calibrated $\underline{Z}_H^P$ will be large enough so the overlap is negligible.

from patenting in the U.S. For example, firms from countries that export intensively to the U.S., or countries that enforce a strong IPR protection policy, are more likely to apply for patents from the USPTO. Without explicitly modelling patenting decision, this differential selections across countries might affect the measured firm efficiency distribution.

For the talent distribution, I obtain average cognitive score and the share of students reaching "basic" and "top" performance from the test score database. These measures are defined based on a common absolute level across countries. To pin down the relative scale of management efficiency and talent, I take the U.S. as the benchmark. Specifically, I set $H_{US}(\theta)$ and $G^R_{US}(\tilde{z}^R)$ to be the same, and use the three statistics on the talent management score to pin down all three parameters in $G^R_{US}(\tilde{z}^R)$ (and hence $H_{US}(\theta)$).[32] I then determine the distributions for other countries, by relating their distribution statistics to those of the U.S.[33]

Table 1.1 summarizes the information on the parameters determined directly. I choose additional parameters jointly in equilibrium, a process I describe below.

[32] The talent management score is approximately normal in the data. Since it is well known that the firm size distribution has a fat tail, I take the exponential of original scores and use that to match firms' innovation efficiency distribution. The statistics I use to pin down each country's distribution are based on these exponents of scores. A few countries in the quantitative analysis are not covered by the World Management Survey. I impute their statistics based on country characteristics. The calibration appendix reports the procedures used in the imputation process.

[33] For firm innovation distributions, the three moments can be perfectly matched by the three parameters in the truncated Pareto distribution. For the talent distribution, however, the truncated Pareto distribution cannot perfectly match all three moments. I therefore use only the average score and the top student share to pin down the upper bound and the shape parameter, while setting the lower bound to be the same across countries. This simplification, however, does not leave out important information, as the correlation between the share of students reaching basic performance and the average score is 0.92.

Table 1.1: Parameters Calibrated Externally

Symbol	Descriptions	Value	Source
σ	Elasticity of substitution between varieties	5	Simonovska and Waugh (2014)
γ	Return to research team size in R&D	0.4	Manufacturing R&D share
δ	Dispersion in offshore production efficiency draws	10.9	Arkolakis et al. (2014)
A	Probability of having a high production efficiency	-6.3	Estimated
B	Dependence of z^P on z^R	0.17	Estimated

1.4.3 Parameters Determined in Equilibrium

Overview The remaining parameters to be determined include international frictions, $\{\tau_{ld}\}$, $\{\phi_{oi}^R\}$, $\{\phi_{il}^P\}$, $\{c_d^M\}$, and $\{c_i^R\}$; country-specific productivity, $\{\Lambda_l\}$; production efficiency distribution parameters, $\underline{z_L^P}$, $\underline{z_H^P}$, and κ_P; and complementarity between management and talent, α. Although in equilibrium these parameters are jointly identified, for certain parameters some moments are more informative than others. I describe below how each parameter is determined.

The iceberg components of international frictions, $\{\tau_{ld}\}$, $\{\phi_{oi}^R\}$, and $\{\phi_{il}^P\}$, determine the aggregate flows of international integration. I use them to match bilateral trade shares, offshore R&D shares, and offshore production shares. To reduce measurement errors, I average bilateral patenting and trade data over the period 1998-2007.[34] The data sources for these bilateral relationships include: the multinational production data sets introduced in Ramondo et al. (2015); bilateral trade including domestic absorption from the World Input-Output Database; and bilateral offshore R&D information based on patenting statistics at USPTO from the OECD patent database.

The offshore R&D measure warrants some explanation. When filing for a

[34] The multinational production database is averaged over 1996-2001 in the original source.

patent at the USPTO, the applicant needs to write the address of the inventors, and the address of the assignee, or the owner of the patent. The OECD patent database documents the number of patents invented in country A (defined as the location of the inventor) but assigned to owners located in another country B, which I use to measure the extent of offshore R&D by firms from country B to country A.[35] In Appendix A, I show that this measure of offshore R&D correlates strongly not only with similarly defined measures based on patents at the European Patent Office and the Patent Cooperation Treaty, but also with an expenditure-based offshore R&D measure.

The fixed components of international costs, $\{c_d^M\}$ (for marketing) and $\{c_i^R\}$ (for R&D), determine the extensive margins of firms' global operations. Due to the lack of this information for a large sample of countries, I assume that these fixed costs are the same for all country pairs, and choose them to match the share of exporters (0.35) and the share of foreign affiliates among research active firms (0.037) in the U.S. manufacturing sector, respectively. The assumption that the fixed costs are the same for firms from all countries performing offshore R&D in, or exporting to, all destinations, is obviously a violation of the reality. This assumption likely affects the model's predictions on the extensive margins of offshore R&D and exporting, however, it might not be very important for the aggregate outcomes we are interested in. In the context of exporting, the literature has shown that, when firms'

[35]In the next chapter, I use the firm-level data based on the same underlying database for empirical analysis. The notion of offshore R&D in this OECD harmonized data and the firm-level data I construct is the same, but the OECD took extra efforts to ensure, to the extent possible, that patents filed under the name of affiliates in host countries are rightly classified as invented in by foreign affiliates, rather than domestic firms. For example, a patent filed by Apple China should be classified as invented by a foreign affiliate, rather than a domestic Chinese company.

productivity distribution is Pareto, the import share predicts the welfare gains from trade (Arkolakis et al., 2012). In the current context, I show analytically (Section 1.3.8) and quantitatively (Section 1.8.2) that the gains from offshore R&D are also strongly tied to the share of foreign firms in domestic R&D. In the calibration, while the fixed costs are the same across countries, the iceberg exporting and offshore R&D costs differ across country pairs, which allows me to match offshore R&D and trade shares perfectly for all country pairs through the intensive margin. As a result, even without variation in fixed costs, the model is still able to capture the strength of trade and offshore R&D for countries.

I calibrate $\{\Lambda_l\}$, the labor productivity in production, by matching the real per-capita income of each country. I normalize $\underline{z_L^P}$ to 1, and determine α, $\underline{z_H^P}$, and κ_P jointly. α affects both the pattern of matching between firms and researchers, and the firm size distribution. Strong complementarity (small α) puts efficient firms at an advantage in working with talent, which affects the shape of the matching function and the concentration of researchers. Figure 1.4a plots the model matching function under various α. The matching functions corresponding to smaller α tend to be more convex, with a larger share of researchers working for the top firms.[36] I measure the overall convexity of the matching function using the ratio between the average slope of the matching function for the top 50% z^R firms, and the average slope for the bottom 50% firms. This convexity conveys information about the value of α, and will be used as a calibration target. I discuss below how I construct the

[36]From Equation 1.7, other things equal, the slope of the matching function reflects the size of the research teams. The steeper the curve, the larger is the research team. A more convex matching function thus means a more unequal distribution of research team size, similar to the Lorenz curve.

model and empirical measures for this convexity.

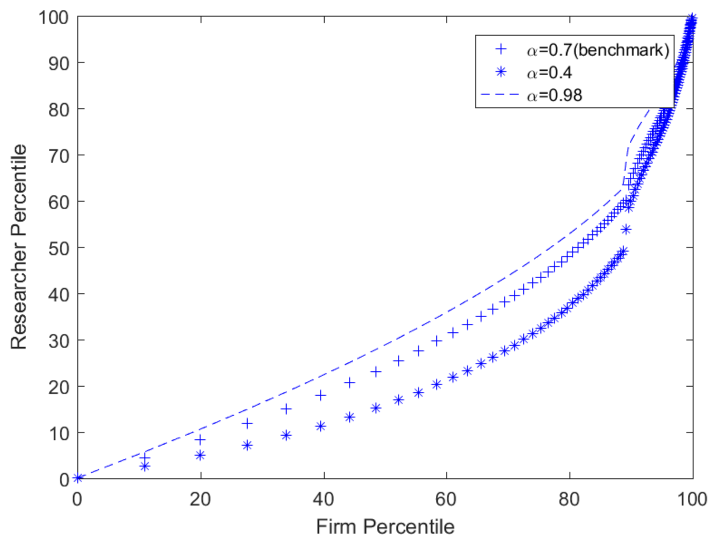
By determining the distribution of talent across firms, α also affects the number of products a firm develops, and hence the firm size distribution. Numerically, it is mostly informative about the size of firms in the top 1%. In addition to α, κ_P and $\underline{z}_H^P$ are also important for the firm size distribution: κ_P directly affects the Pareto shape of the firm size distribution at the very top, while $\underline{z}_H^P$ effectively determines the scale of the top 4% firms relative to the bottom 96%, as about 4% of U.S. firms draw from the H distribution.

Specifics about the matching function I estimate the parameters of matching function using evidence of positive assortative matching between inventors and firms in the next chapter. Specifically, using inventor-firm linked patenting data from the USPTO, I measure firm innovation efficiency using the per-inventor innovation output, and inventor talent using past innovation. Focusing on a sample of job switchers, I then estimate nonparametrically how the talent of an inventor is related to the innovation efficiency of the new firm, controlling for inventor and firm characteristics as well as time and patent category fixed effects. A positive correlation indicates positive assortative matching. The solid line in Figure 1.4b presents the estimates, along with a 2 s.e. band. The overall convexity measure of this empirical matching function is 1.71.

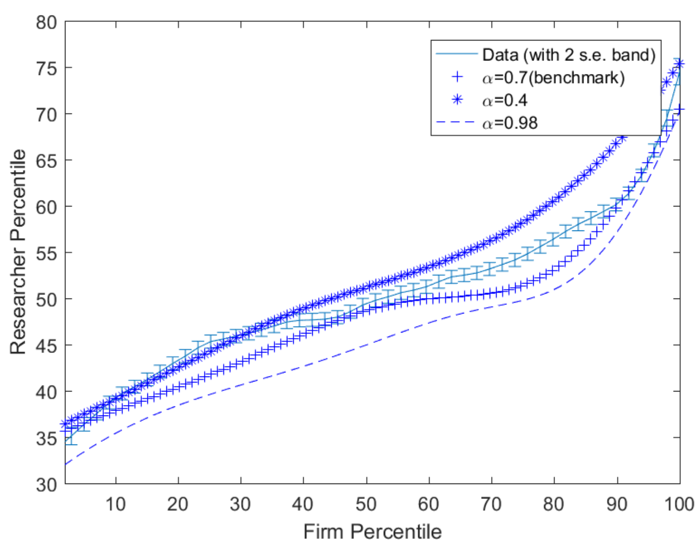
We cannot directly compare the model convexity measure to its data counterpart. In the data, matches are noisy, so the range of the estimated matching function is not $[1, 100]$, whereas in the model, this is always the case. To make the two comparable, I take the stand that in choosing the optimal types of researchers,

Figure 1.4: The Model and Empirical Matching Function

(a) Model matching function



(b) Empirical and “noisy” model matching function



Notes: In both panels, the horizontal axis plots percentiles of firm innovation efficiency, and the vertical axis plots percentiles of researcher talent. The upper panel is the model matching function under different α. The lower panel shows the empirical matching function estimated by the author using the USPTO data (solid line), and simulated “noisy” model matching functions under different α.

firms make mistakes. They cannot differentiate among workers whose talent satisfies firms' first order condition (Equation 1.6) within a certain "error margin". I fix the wage schedule at the benchmark equilibrium, and then choose the size of this margin so that the estimated matching function using simulated data has the same range as the empirical matching function. I then compute the convexity measure based on this simulated matching function.[37]

Figure 1.4b plots the simulated noisy matching function when α is 0.7, which will be the benchmark calibration, and two different values. The benchmark value offers the best fit for the overall concavity, determined by the value of the matching function at the 50th firm percentile. A smaller α could fit the overall shape reasonably well, but misses the top range. A larger α, on the other hand, is a poor fit overall.

1.4.4 Computational Algorithm

A detailed account of the computational algorithm is provided in the appendix. This section briefly describes the nested procedure I use. In the outer loop, I choose $\underline{z}_H^P$, κ^P, c^M, c^R, and α to match the targets described above. In the middle loop, I iterate over $\{\tau\}$, $\{\phi^R\}$, $\{\phi^P\}$, and $\{T\}$ to match all bilateral shares and per-capita real income of countries. The inner loop solves the model given exogenous parameters.

[37] For different α, the size of the "error margin" needed to match the range of the empirical matching function varies. But as long as $\alpha < 1$, the simulated matching function can always match the range of its empirical counterpart. When α approaches 1, firms become increasingly indifferent between different researchers. A small amount of mistakes in recruiting would then result in a flat matching function.

This computation algorithm requires solving the researcher market equilibrium for all countries at different parameter values. With offshore R&D decisions in the model, the distribution of R&D center innovation efficiency, $g_i^R(z^R)$, is an endogenous outcome. The cutoff rule in offshore R&D decisions implies that $g_i^R(z^R)$ could have multiple discontinuities. As a result, the matching function, $T_i(z^R)$, is not necessarily differentiable. In this case, general boundary value problem solvers routinely fail or takes a long time to find the solution. In the appendix, I develop a computational algorithm that is well-suited for this exercise. In essence, I show the solution to the boundary value problem can be found by solving a sequence of initial value problems.

1.4.5 Model Fit

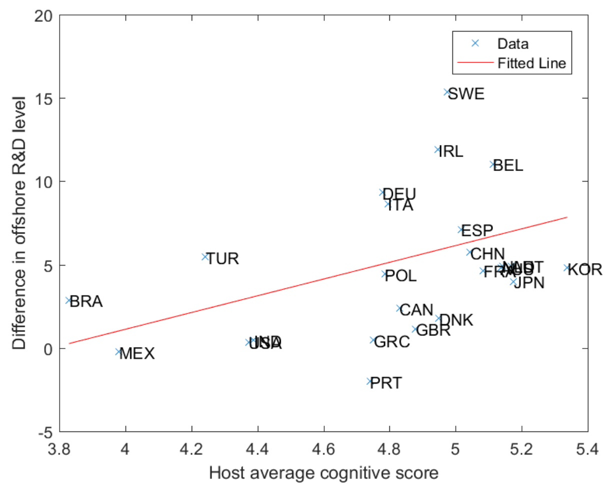
Table 1.2: Fit of the Targeted Moments

Parameter	Value	Moments	Model	Data
c^M	0.0693	Share of exporters	0.35	0.35
c^R	2.6	Share of foreign affiliates	0.042	0.037
$\underline{z}_H^P$, α, κ^P	$\underline{z}_H^P = 1.2$	Fraction of firms with emp.<100	0.99	0.99
	$\alpha = 0.7$	Fraction of firms with emp.<20	0.95	0.95
	$\kappa^P = 8.16$	Matching function slope between 0%-50%/ slope between 50-100%	1.58	1.71
		Share of emp. in firms with >500 emp.	0.41	0.47
		Power law coefficient of firm size dist.	1.04	1.05

The calibration process determines that $c^M = 0.069$, $c^R = 2.6$, $\underline{z}_H^P = 1.2$, $\kappa^P = 8.16$, and $\alpha = 0.7$. Table 1.2 reports the value of parameters and the model moments that help pin down these parameters. Overall, the model is able to fit data along these dimensions well.

The value of α suggests strong complementarity between innovation efficiency and researcher talent. Since the complementarity and the resulting talent-acquisition motive are an important channel in the model, in the following, I first discuss the role of α in determining the model predictions and explaining the patterns in the data. I then present additional non-targeted implications of the model under the benchmark calibration and compare them to the data whenever possible.

Figure 1.5: Complementarity and Offshore R&D



Notes: The vertical axis plots the percentage point difference between the benchmark parameterization and an alternative parameterization with $\alpha = 0.98$ in the share of R&D expenditures by foreign affiliates. The horizontal axis is host average talent. Host average innovation efficiency is netted out from both axis.

The importance of complementarity To understand the role of complementarity in shaping offshore R&D between countries, I solve a counterfactual experiment with $\alpha = 0.98$, keeping other parameters at the benchmark. This parameter value implies much weaker complementarity than the benchmark calibration. The vertical axis in Figure 1.5 shows the percentage point difference between the benchmark and the counterfactual equilibrium in the share of domestic R&D done by foreign affiliates. The horizontal axis is host average talent quality. The figure indicates that higher complementarity increases offshore R&D, particularly in host countries with high talent, so complementarity is an important force for the pattern of offshore R&D.

The calibration suggests that a relatively strong complementarity ($\alpha = 0.7$) fits the pattern of matches and moments of the firm size distribution well. Does it also explain the pattern of offshore R&D better than under weak complementarity (as α approaches 1)? The next chapter shows using firm-level patent data that offshore R&D increases with firm innovation efficiency, host talent, and their interaction. Here I evaluate the model's ability to generate these features under $\alpha = 0.7$ and $\alpha = 0.98$.

Because the calibration exactly matches bilateral offshore R&D and the distribution of talent and innovation efficiency in the cross section, I evaluate the model in changes. I simulate a counterfactual equilibrium in which countries receive random shocks to their distributions of talent or efficiency.[38] I then use the simulated data to perform a difference-in-difference regression of changes in offshore R&D on changes in the distributions of host talent and home innovation efficiency, in which each pair of country is an observation.

Columns 1-3 of Table 1.3 report the results under the benchmark specification. Bilateral pair fixed effects are included in all three columns, so the model is identified from changes. The first two columns show that the host country talent and home country innovation efficiency both have significant positive impacts. The third column in addition adds an interaction term. The interaction is positive and

[38]Specifically, I reduce the upper bound of the talent distribution by a random fraction for one third of the countries, reduce the upper bound of innovation efficiency by a random fraction for one third of the countries, and then keep the remaining one third of countries intact.

Table 1.3: Complementarity and the Patterns of Offshore R&D: Simulated Data

	Benchmark Calibration			Weak Complementarity		
	(1)	(2)	(3)	(4)	(5)	(6)
Log(average home mgt. efficiency)	0.157***		-0.650*	0.010***		-0.018
	(0.044)		(0.375)	(0.003)		(0.025)
Log(average host talent)		0.071*	-0.625**		0.006**	-0.020
		(0.038)	(0.298)		(0.003)	(0.019)
Interaction			0.353**			0.012
			(0.164)			(0.011)
Bilateral FE	X	X	X	X	X	X
Observations	1352	1352	1352	1352	1352	1352
Within R^2	0.019	0.005	0.025	0.019	0.007	0.021

Standard errors in parentheses

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

significant, while the non-interactive terms turn negative. So the effects are concentrated in the pairs of countries that experience improvements in both host talent and home efficiency, consistent with the empirical findings. Columns 4-6 of Table 1.3 report the same specifications under the case where $\alpha = 0.98$. In this case, host talent and home efficiency both have significant marginal impacts on offshore R&D. The within R square terms in the first two columns are also similar to those under the benchmark specification. However, the interaction term is not significant, in contrast to the data presented in the next chapter.

Together, these results suggest that under the benchmark calibration, the model is able to generate a relationship between offshore R&D and the distributions of endowments that is similar qualitatively to that observed in the data, while a model with weak complementarity cannot.

The management score difference between large and small firms The calibration procedure for κ_P and z_H^P takes the stand that management efficiency differences are the fundamental cause of performance differences among firms.[39] To validate this assumption, I examine the model's performance in matching the mapping from management score to firm size. This is a valid test because the calibration uses the information on management scores and the firm size distribution separately, but puts no restrictions on how a one-point increase in the management score at different percentiles of the firm size distribution translates into increases in firm size. For this comparison, I use the total management score, defined as the sum of the innovation and production score, for a consistent comparison with the empirical evidence.

Table 1.4: Additional Untargeted Moments

Management Score and Firm Size	Model	Data
Management score difference between large and small firms	1.18	1.32
The Management Efficiency of Foreign Affiliates		
Foreign affiliate advantage	1.33	1.16
Coefficient of variation across countries	0.094	0.075
Correlation with domestic average score	-0.67	-0.84

The first panel of Table 1.4 reports the difference in average total management score between firms with 10000 employees and firms with 10 employees for the model and data.[40] In the model, the difference in total management score between an average firm with 10000 employees and an average firm with 10 employees is 1.18 times the standard deviation of the management score, which is close to the empirical counterpart of 1.32.

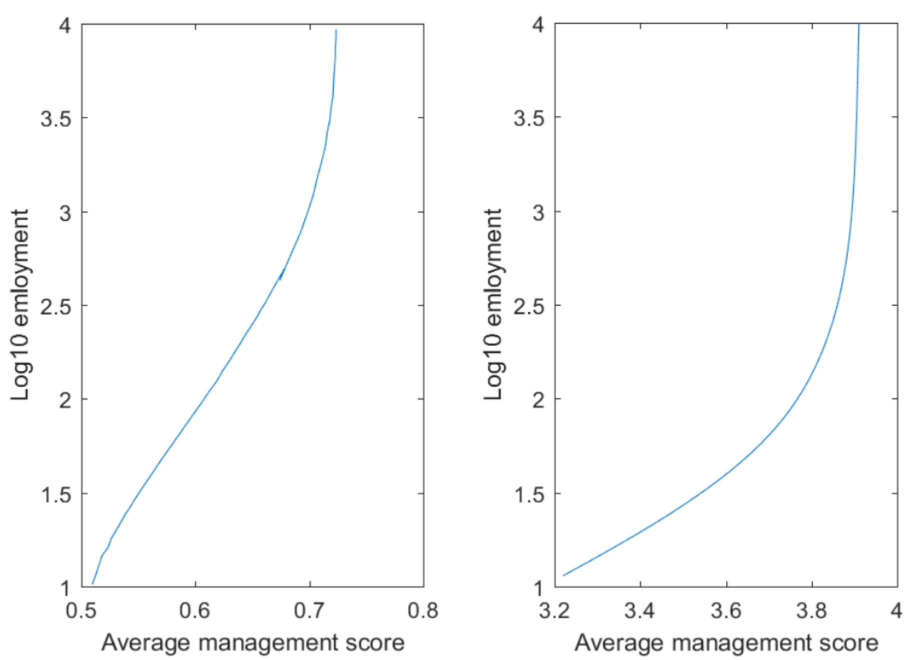
[39]The calibration essentially takes the management score distribution from data, and chooses κ_P and z_H^P so that the variation in firm size is close to that in the data.

[40]The empirical counterpart of this number is from Bloom et al. (2014), which estimates this relationship nonparametrically, focusing on medium-sized U.S. manufacturing firms with 10-10000 employees. Because the two surveys have different scales for scoring, I normalize the increase by the standard deviation of total management score.

Figure 1.6 plots nonparametrically the relationship between management score and firm size from the model and the data. The estimated curve from the data, in the left panel, displays some convexity: initially, firm size increases relatively slowly with management score; at the top range, however, a small increase in management score results in a larger percentage increase in firm size. Such a relationship can always be captured by the model by choosing how management score scales into productivity, which is partially determined by z_H^P. The question is whether the scale chosen to match other moments is able to generate this relationship. The right panel is the model relationship between management score and employment. Consistent with the data, the model also generates some convexity.

Figure 1.6: Management Score and Firm Size

(a) Data (b) Model



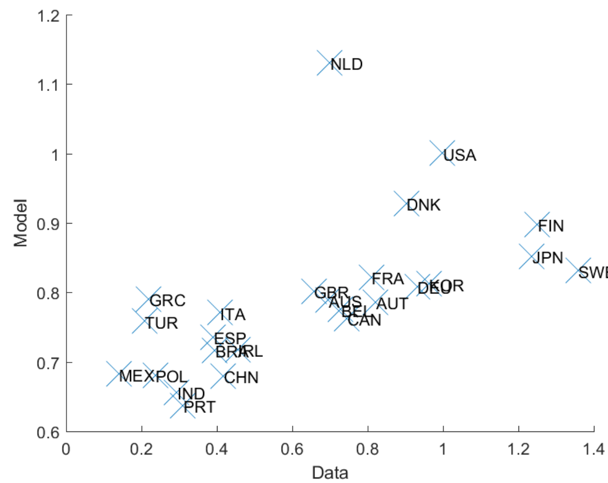
Notes: The left panel shows the model relationship between management score and firm employment in the data estimated in Bloom et al. (2014); the right panel shows the model counterpart. Both are based on the sub-sample of firms with employment between 10 and 10000. The range of variation in x-axis is 1.18 times the standard deviation of the management score in the data, and 1.32 in the model.

The multinational managerial advantage One important assumption of the model is that affiliates' innovation efficiency depends on that of their parents, rather than that of host country domestic firms. This assumption, together with the self-selection mechanism, implies that foreign affiliates tend to be more management efficient than domestic firms, and that the managerial advantage of foreign affiliates is larger in countries with worse domestic innovation management efficiency.

I validate these implications quantitatively by calculating the foreign affiliate managerial advantage for each country. The measure I use is the ratio between average foreign affiliate innovation efficiency and average domestic firm innovation efficiency. I then compare the statistics of this measure among the model countries to their data counterpart, constructed using the database introduced in Bloom et al. (2012).

The bottom panel of Table 1.4 reports the statistics of the foreign affiliate managerial advantage for the sample countries. Both the model and the data indicates a larger innovation management score for foreign affiliates compared to domestic firms, although the difference is larger in the model (33%) than in the data (16%). The variability of the foreign affiliate advantage measure across countries, captured by its coefficient of variation, is 0.094 in the model, and 0.075 in the data. The correlation between this measure and the host country average domestic innovation score is -0.67 in the model, and -0.84 in the data. So quantitatively, the model fits the cross-country pattern of foreign affiliate innovation advantage well. In the appendix, I also plot the model foreign affiliate managerial advantage against its data counterpart for individual countries that are common to both samples. Overall the

Figure 1.7: Share of R&D in Income: Model versus Data



Notes: This figure plots the share of income from non-production labor in the model against the share of R&D in GDP in the data across countries. The measure for the U.S. is normalized to have 1 in both the model and the data.

model is a reasonable fit.

Share of non-production income in GDP The model predicts countries will specialize differently in R&D or production. Figure 1.7 plots the share of income from non-production labor in the model against its counterpart in the data, the share of R&D in GDP. There is a strong correlation between the model and the data across countries, even though the model best captures the manufacturing industry while the data is from the aggregate economy.[41]

International Frictions Finally, I check if the calibrated bilateral frictions are reasonable by comparing their correlations with geographic distance. The correlations between the logs of τ, ϕ^R, and ϕ^P and the log of distance are 0.2, -0.22, and -0.42, respectively. The signs of these correlations are consistent with larger international frictions for longer distances (ϕ^P and ϕ^R are the inverse of costs). The

[41]The model prediction better matches the ratio between R&D expenditures and manufacturing value added.

difference between offshore production and offshore R&D in distance elasticity also supports that these two activities are different in nature.

1.5 Counterfactual Experiments

In this section I perform counterfactual experiments using the parameterized model to shed light on the determinants and impacts of offshore R&D.

1.5.1 What Determines Offshore R&D

I first examine the quantitative importance of the talent-acquisition and market-access motives for offshore R&D. This is a relevant exercise, because policy makers around the world are looking to attract R&D intensive FDI. Domestic research talent and access to foreign countries through trade and offshore production are cited as important determinants of the attractiveness of a country as a host for offshore R&D centers (Guimón, 2009). I perform a set of experiments in which I either change the distribution of talent or management endowments, or the market access of a country. To isolate the effects from changes in other countries, when computing these counterfactual equilibria, I change parameters for one country at a time, keeping model parameters at the benchmark for all other countries.

The role of endowment distributions The first set of experiments aim to quantify the importance of cross-country differences in the distributions of firm efficiency and researcher talent in determining offshore R&D. Specifically, I increase innovation efficiency of each host country, and decrease the talent of their workforce,

Table 1.5: The Determinants of Offshore R&D

		"Talent Acquisition"			**"Market Access"**			**All**
Country	**Benchmark**	**Efficiency**	**Talent**	**Both**	**Consumers**	**Producers**	**Both**	
	1	2	3	4	5	6	7	8
Developed								
AUS	26.83	0.05	7.61	0.00	30.30	17.83	0.27	0.00
AUT	50.21	5.87	25.81	0.00	51.46	39.88	0.00	0.00
BEL	57.12	12.25	14.98	2.48	68.15	31.62	0.00	0.00
CAN	33.52	13.38	12.19	0.13	38.88	22.73	0.00	0.00
DEU	23.85	4.22	7.53	1.03	36.32	11.51	1.65	0.00
DNK	33.55	0.93	15.98	0.13	41.22	18.88	0.00	0.00
ESP	42.92	0.28	30.96	0.00	45.64	41.03	29.53	0.00
FIN	17.93	0.00	0.46	0.00	26.86	1.62	0.00	0.00
FRA	33.74	1.00	17.66	0.31	40.32	23.61	2.89	0.00
GBR	45.65	17.83	29.04	7.13	54.25	33.23	5.63	0.08
GRC	58.00	10.46	53.50	5.72	56.90	57.02	49.54	1.37
IRL	55.20	30.23	29.72	0.01	55.21	51.22	0.00	0.00
ITA	29.20	0.33	22.11	0.21	32.05	26.31	19.19	0.10
JPN	5.05	0.00	2.03	0.00	9.00	2.39	2.35	0.00
KOR	4.81	0.00	0.00	0.00	8.80	3.34	1.36	0.00
NLD	34.66	1.13	2.83	0.07	54.67	0.81	0.00	0.00
POL	60.79	31.85	49.68	21.07	60.72	60.52	43.06	0.25
PRT	50.21	0.14	41.17	0.00	60.27	46.95	0.00	0.00
SWE	26.97	0.65	3.22	0.00	31.92	7.07	0.00	0.00
USA	7.93	7.93	2.97	2.99	13.89	1.92	1.18	0.45
Median	33.64	1.07	15.48	0.10	40.77	23.17	0.72	0.00
Emerging								
BRA	37.14	5.37	37.14	5.37	40.15	36.49	34.23	4.68
CHN	52.13	22.66	46.33	19.40	53.46	52.12	51.28	18.73
IND	57.88	35.06	57.08	34.66	58.67	57.74	56.96	33.99
MEX	49.26	24.32	49.49	24.40	54.50	48.07	36.78	14.80
TUR	51.79	12.04	46.24	9.38	52.43	51.60	49.40	8.37
Median	51.79	22.66	46.33	19.40	53.46	51.60	49.40	14.80
Median	37.14	5.37	22.11	0.21	45.64	31.62	1.65	0.00

Notes: The numbers reported in this table are the share of domestic R&D expenditures incurred by affiliates of foreign companies in each country. All numbers are in percentage points. The first column shows the results from the benchmark calibration. The second column changes the firm innovation efficiency distribution for each country to that of the U.S. The third column changes the worker talent distribution for each country to that of Brazil. The fourth column combines the changes in the second and third columns. The fifth column increases exporting costs to infinity. The sixth column increases countries' outward offshore production costs to infinity. The seventh column combines changes in the fifth and sixth columns. The last column combines changes in the fourth and seventh columns.

to see how these two factors affect the equilibrium offshore R&D. I choose the U.S. innovation efficiency distribution and the Brazilian talent distribution as the benchmarks, because these two countries respectively have the highest management efficiency, and the lowest average talent.

The second column in Table 1.5 reports the share of R&D done by foreign affiliates for each host country when it is given the U.S. management efficiency distribution. With an improvement in domestic management efficiency, domestic firms are more competitive in both labor and product markets. Domestic wages increase and prices decrease, reducing foreign firms' incentive to enter. Indeed, compared to the share of foreign R&D in the benchmark equilibrium in the first column, the shares in these counterfactual equilibria are much lower. The median share of R&D done by foreign affiliates across all countries is 5.37% , or one-eighth of the benchmark value. Perhaps surprisingly, for many developed countries, the offshore R&D share decreased significantly. For example, Germany sees its offshore R&D decreasing by 18 percentage points. This is not due to Germany having a particularly bad calibrated firm efficiency distribution—as shown in Table A.1, the average management score of Germany is 8.21, the third highest among all countries (right after the U.S. and Canada). Rather, the large change is due to the U.S. having a significantly better efficiency distribution than all other countries. Developing countries also experience significant decreases in offshore R&D—it decreases by more than 20 percentage points in all developing countries in the sample. While these countries have worse calibrated firm efficiency distributions compared to the developed countries (see Table A.1), their decrease in offshore R&D are not larger in percent, so that significant offshore R&D remains. This likely arises for two reasons. First, these developing countries in general have large domestic markets. Second, offshore production into these countries tend to be costly due to their geographic locations. These two together imply a strong product market entry motive in offshore R&D.

I then change each country's talent distribution to that of Brazil, while keeping its firm efficiency distribution at the benchmark. Intuitively, when domestic talent distribution improves, R&D outputs of both domestic and foreign-affiliated R&D centers increase. The increase in the latter is larger for two reasons. First, foreign affiliates are on average more productive, so they benefit more from the improvement in researcher quality. Second, the increase in R&D output allows more foreign firms to overcome the fixed costs and enter. Column 3 of Table 1.5 reports the share of R&D by foreign affiliates in each country. In the median country, foreign affiliates now account for about 22% of domestic R&D, which is a decrease of around one-third from the benchmark value. The size of the decrease, again, varies considerably across countries. Perhaps because developed countries had better talent distributions to begin with, they experience larger drops in inward offshore R&D.

Finally, I combine the two experiments by changing the distributions of both management efficiency and talent. As can be seen from the fourth column of Table 1.5, the global median share of R&D done by foreign affiliates is around 0.21%. Overall, cross-country differences in the distributions of talent and firm efficiency can account for most of the observed offshore R&D for developed countries, and a smaller but still significant share for large developing countries.

The role of foreign access I now examine the impact of the host country's access to foreign countries on offshore R&D. In the model, foreign access consists of two channels: access to foreign consumers through exporting, and access to foreign producers through offshore production. I consider the separate and joint impacts of these two channels.

In the first experiment, I increase each host country l's iceberg export cost, $\tau_{ld}, l \neq d$ to infinity. This shuts down host countries' direct access to foreign consumers, but R&D centers there can still indirectly access foreign consumers through offshore production. The shares of R&D by foreign affiliates in these counterfactual equilibria are reported in column 5 of Table 1.5, which shows small but universal increases in offshore R&D shares across countries.

This result might seem surprising at first glance, given the partial equilibrium intuition that eliminating access to foreign consumers through direct exporting reduces the return to doing R&D in a host country. This effect seemingly should be especially strong for more productive firms, because they export more. So fewer foreign firms should enter, and their share in total R&D should decrease. In a model with both trade and offshore production, however, this direct channel is muted. Even if they cannot export directly, firms can still serve foreign consumers indirectly by offshoring their production to other countries. Moreover, due to the lower demand for labor from production, wages for both inventors and production workers decrease, which makes the country more attractive as a host for R&D centers. An increase in export costs thus has a similar effect to a decrease in a host country's production efficiency, which strengthens its comparative advantage in innovation, driving it to specialize in R&D activities.

In the second experiment, I increase the costs of offshore production in each country to infinity (by setting $\phi_{il}^{R}, i \neq l$ to zero), so it is impossible for R&D centers to perform offshore production in other countries. The 6th column of Table 1.5 shows that, compared to the benchmark equilibrium, most countries experience a decrease in offshore R&D. The median share of R&D by foreign affiliates decreases by about 5 percentage points from the benchmark economy, to around 32%. Because firms located in emerging economies in this sample do not perform outward offshoring activities to begin with, the decrease in offshore R&D resulting from this change tends to be more significant for developed economies than for emerging economies.

The general equilibrium effect works in the same direction as the partial equilibrium effect in this case. When the option of offshore production is eliminated, R&D centers in the host countries have to produce locally to serve both foreign and domestic customers, which increases wages for production workers and inventors, making the country less attractive as a host for R&D centers. An increase in offshore production costs is therefore similar to a reduction in R&D innovation efficiency of a country, which strengthens its comparative advantage in production.

Column 7 of Table 1.5 reports the experiment when both exporting and offshore production are shut down. Compared to column 5, the share of offshore R&D is much smaller for developed countries, because when offshore production is not an option, countries can no longer specialize in innovation. For developing countries, the differences between columns 5 and 7 are small, mainly because they do not perform much outward offshoring production in the benchmark equilibrium.

Finally, I combine the two sets of experiments reported in this section, by changing the two distributions and also eliminating host access to foreign consumers and producers. The median of foreign R&D shares, reported in the last column of Table 1.5, is 0. The only countries that attract a significant share of offshore R&D are large emerging economies, such as Brazil, China, and India. The large markets of these countries, and their relative remoteness for exporting and offshore production from other countries, are the reasons for foreign firms to perform R&D in those countries.

In summary, the experiments in this section show that the two main forces incorporated in the model have significant impacts on firms' offshore R&D decisions. On average, differences in the management and worker quality distributions together explain about 86% of the equilibrium offshore R&D (the global average decrease from 37.14% in the benchmark economy to 5.38% in the fourth column). Host access to foreign customers reduces offshore R&D in the country, while its access to foreign producers increases it. Combined, international differences in the distributions of talent and firm efficiency, and access to foreign markets and producers, explain more than 92% of the average level of offshore R&D in the benchmark equilibrium (the average value in the 8th Column is 3.31%). The small remaining offshore R&D activities are concentrated in emerging economies with large domestic markets.

1.5.2 The Gains from Offshore R&D

I now turn to the normative aspect of offshore R&D. As a starting point, I examine the welfare gains from various forms of economic integrations by eliminating each channel from the model separately. I define the gains from offshore R&D as the increase in real income as a country moves from an equilibrium where offshore R&D is not allowed to the baseline equilibrium. I define the gains from trade and gains from offshore production analogously. Finally, I define the gains from openness as the combined effects of the three channels. Since these three channels interact with each other, the sum of gains from trade, offshore production, and offshore R&D, does not necessarily equal the gains from openness.

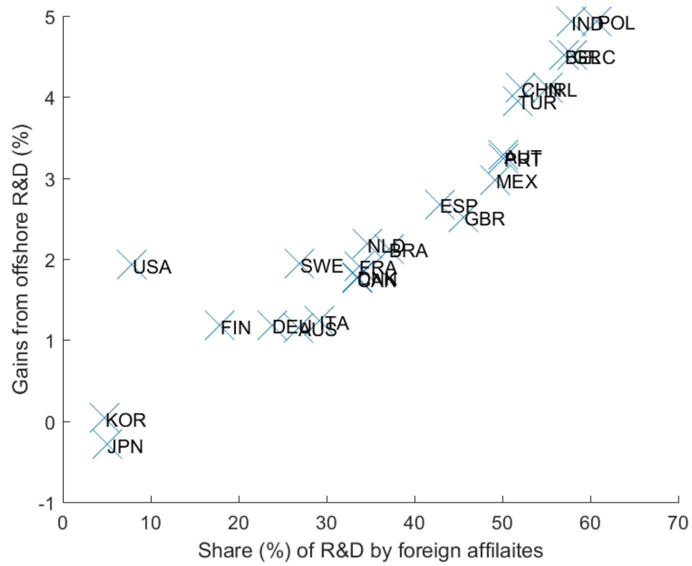
It is worth emphasizing that, defined this way, these welfare measures are about the level of welfare gains a country has *currently* achieved through international economic integration. Smaller gains from openness for a country do not mean that this country has little to benefit from *further* economic integration. On the contrary, if a country currently benefits very little from international economic integration because of its high distortions, it means there is a larger scope for future gains through eliminating these distortions. In this chapter I mainly focus on the level of achieved welfare gains so that the results are more comparable to the existing literature. I perform some experiment on further liberalizations in section 1.5.3.

The first column in Table 1.6 presents the welfare gains from offshore R&D. The median welfare gain is 2.18%. The median, however, masks a great deal of country heterogeneity. Some countries, such as China, India, and Greece, benefit by 4% or higher. Meanwhile countries like Japan and Korea barely receive any benefits or even lose, due to the general equilibrium effects from international competition. Figure 1.8 plots the gains from offshore R&D against the model share of foreign affiliates in domestic R&D for each country. Countries with a higher share of R&D

Table 1.6: The Welfare Gains from International Economic Integration

ISO	Offshore R&D	Trade	Offshore Prod.	Openness	Trade & Offshore Prod.
	1	2	3	4	5
Developed					
AUS	1.14	5.64	2.09	19.47	17.24
AUT	3.28	10.67	3.12	52.43	43.98
BEL	4.51	15.94	4.52	74.15	61.53
CAN	1.76	9.43	3.53	43.56	38.74
DEU	1.18	7.08	3.17	24.87	22.88
DNK	1.77	17.23	2.32	43.02	40.40
ESP	2.67	3.95	0.32	17.61	12.69
FIN	1.17	8.12	3.00	21.70	20.86
FRA	1.91	6.90	2.34	22.51	19.62
GBR	2.51	8.47	4.59	34.68	28.82
GRC	4.51	7.67	0.03	20.54	13.32
IRL	4.11	12.55	2.17	60.98	48.39
ITA	1.23	3.33	0.42	11.35	9.22
JPN	-0.29	3.16	1.31	4.75	4.69
KOR	0.03	3.83	0.15	6.60	6.45
NLD	2.18	19.47	9.43	64.29	62.48
POL	4.93	5.01	0.61	28.20	17.45
PRT	3.24	5.23	2.48	43.77	33.17
SWE	1.93	8.79	3.52	34.33	31.72
USA	1.93	8.92	3.97	16.08	15.90
Median	1.93	7.89	2.41	26.53	21.78
Emerging					
BRA	2.12	-0.46	0.09	6.91	3.27
CHN	4.11	-0.92	-0.05	7.96	2.40
IND	4.93	-0.53	0.04	9.48	3.70
MEX	2.97	3.61	0.73	20.73	13.60
TUR	3.95	3.27	-0.18	12.06	6.14
Median	3.95	-0.46	0.04	9.48	4.15
Median	2.18	6.90	2.17	21.70	18.28

Figure 1.8: The Welfare Gains from Offshore R&D



Notes: The vertical axis shows the welfare gains from offshore R&D (%), and the horizontal axis shows the share of R&D in a host country performed by foreign affiliates in the model economy.

done by foreigners tend to benefit more from offshore R&D.

The second and third columns report the gains from trade and the gains from offshore production, respectively. The median gain is around 7.0% for trade, and 2% for offshore production. Again, the welfare gains take a wide range of values. As expected, smaller economies and countries that are closer to major markets, such as Belgium, Netherlands, and Ireland, gain more from both trade and offshore production. Larger and more remote economies, such as India, gain less. Some countries even receive modest losses from trade and offshore production.

In the fourth column are the overall gains from openness. They range from 16% for the U.S. to 74% for Belgium, with a median of 21.7%. The gains from openness are almost always larger than the sum of the gains from the three forms of economic integration, which means these three forms of integration are substitutes—the benefit from additional openness is smaller once a country is already open in other dimensions. The substitution between trade and offshore production is intuitive—since these two are alternative ways of serving goods from where they are invented to where they are consumed, when one channel is present, the marginal benefits from the other channel are lower. The last column of Table 1.6 reports the combined gains from trade and offshore production, computed from a counterfactual scenario where both trade and offshore production are eliminated. Indeed, the values in column 5 are universally larger than the sum of columns 2 and 3. This result is consistent with the finding in Arkolakis et al. (2014), in a setting without offshore R&D.

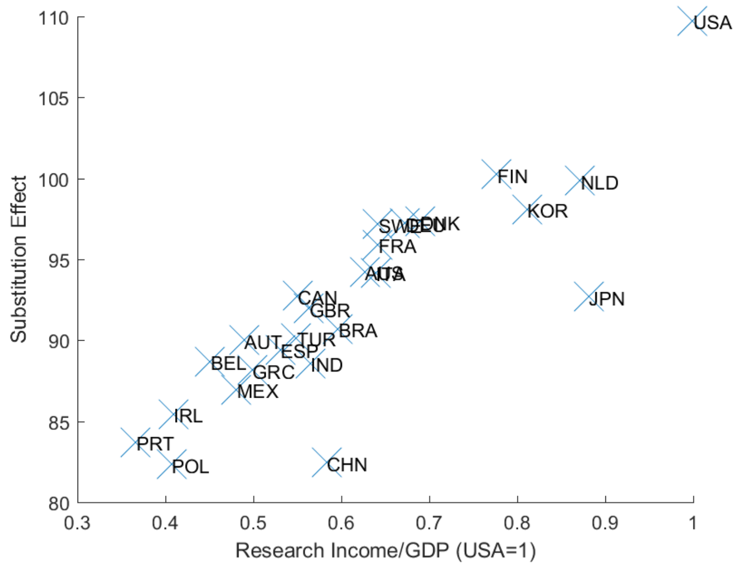
The interaction pattern between offshore R&D and the combined effect of trade and offshore production is more nuanced—while in most countries, the sum of columns 1 and 5 is still smaller than column 4, the difference is small. In countries like the U.S., the sum of gains from offshore R&D and the gains from trade and offshore production is actually larger than the overall gains from openness.

This difference is again related to the interaction among various forces through country specialization. First, there is a demand-for-R&D channel. The option to export and to produce offshore raises the return to innovation. Because of the fixed marketing cost, this benefits more efficient firms particularly, who are also the ones most likely to perform offshore R&D. This demand side channel therefore tends to increase the gains from offshore R&D when trade and offshore production are present. However, there is also a labor-supply channel. Since innovation and production compete for workers, the general equilibrium effect discussed in the previous section sets in. When one sector expands in a country, wages increase, making the country benefit less from new opportunities in the other sector.

Offshore R&D tends to increase the R&D efficiency in countries that are relatively scarce in high efficiency firms, which weakens the comparative advantage of these countries in production. Because those are also the countries that tend to specialize in production, offshore R&D reduces their gains from trade and offshore production by weakening their comparative advantage. The substitution between offshore R&D, trade and offshore production is stronger for countries with strong comparative advantage in production. We can use the ratio between the sum of gains from offshore R&D and the combined gains from trade and offshore production over the overall gains from openness as a measure of the strength of this substitution. A lower ratio means smaller marginal gains from further integrating the economy once it is already integrated through other ways, and therefore represents stronger substitution. I use the share of income generated by R&D labor in the calibrated equilibrium as a measure for comparative advantage in innovation. Figure 1.9 displays the relationship between these two measures. As conjectured, the substitution effect is more important for countries with comparative advantage in production.

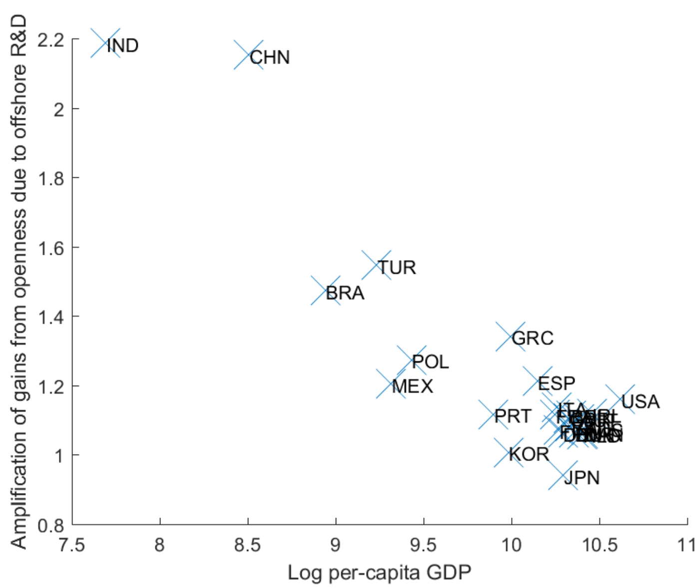
How important is accounting for offshore R&D in understanding the gains from openness? Figure 1.10 plots the relationship between host income and the ratio between the gains from openness in the benchmark model, shown in Table 1.6, and the gains from openness in a restricted-version of the model without offshore R&D. With offshore R&D as an additional channel for gains from openness, the ratio is generally larger than 1, indicating higher gains from openness in the benchmark model. The average of this ratio among the model countries is 1.2. This amplification, however differs significantly across countries. For emerging countries

Figure 1.9: The Substitution Between International Activities



Notes: The vertical axis shows the sum of gains from offshore R&D and the combined gains from trade and offshore production, divided by the overall gains from openness. The horizontal axis shows the share researcher compensation in total income.

Figure 1.10: Relative Importance of Offshore R&D



Notes: The vertical axis shows the ratio between gains from openness in the benchmark model, and the gains from openness in a model without offshore R&D. The horizontal axis shows host income.

in the sample, such as China, India, Brazil, and Turkey, the gains from openness are more than 100% higher in the benchmark model with offshore R&D. This amplification is much lower for developed countries. For example, for the U.S., the inclusion of offshore R&D only increases the gains from openness by 15%. The wide range of this ratio also underscores the importance of incorporating offshore R&D—overlooking this channel will not only understate the gains from openness, but also bias the comparison of the gains from openness across countries.

Why do developing countries benefit more from offshore R&D? Further examination of countries' participation in various forms of integration suggests that during the sample period, developing countries participated more intensively in offshore R&D than in trade and offshore production. By fitting this pattern, the model implies that the frictions impeding offshore R&D increase more slowly with distance than the frictions impeding trade and offshore production. As a result, developing countries which are far away from major home countries of innovating firms—U.S., West Europe, and Japan—participate more intensively in offshore R&D, and less intensively in offshore production.

To sum up, the counterfactual experiments in this section demonstrate that offshore R&D represents a quantitatively important new channel through which countries benefit from globalization. It is a weak substitute for trade and offshore production in general, although the substitution patterns depend on a country's specialization in innovation or production in the world economy. Further, by showing that offshore R&D and other forms of globalization have very different impacts across countries, the results also highlight the importance of modelling offshore

R&D separately, rather than treating it as part of the offshore production process.

1.5.3 Further Liberalization of China and India

Existing quantitative research on multinational activities usually does not allow firms to make independent decisions on offshore R&D and production. I evaluate whether this is an important restriction by comparing the welfare implications of liberalizing offshore R&D and offshore production. Doing so is potentially important because policy makers usually have at their disposal policies that specifically target production or innovation activities.

As an example, I focus on the case of China and India and evaluate two types of openness policies. This exercise is interesting in its own right, because both countries are becoming popular destinations for offshore production and R&D. Related to this trend, their governments are attempting to attract more foreign companies, especially R&D intensive ones, by cutting red tape and speeding up the entry approval process.

I first consider an inward offshore R&D liberalization that makes it easier for foreign firms to open R&D centers in India and China. More specifically, I reduce the fixed costs of R&D in these countries by 20%. This reduction in cost can be interpreted as a tax credit for the upfront investment in R&D, subsidized land, or speedy approval of entry. The magnitude of the reduction is well within the range of policies commonly used.[42] The first column of Table 1.7 reports the results. As

[42]For example, in 2012, to attract a $30 billion investment in a chip factory from Samsung Electronics Co. Ltd., the Chinese city Xi'An offered a package of favorable policies, including free land, infrastructure, and tax credits. The land alone was valued at $4 billion, more than 10% of

Table 1.7: Further Inward FDI Liberalization in China and India

	Policy 1	Policy 2	Policy 3
CHN	0.65	0.20	0.13
IND	0.89	0.19	0.12
DEU	0.00	0.17	0.21
JPN	0.00	0.15	0.24
USA	-0.01	0.09	0.26

Notes: All numbers are in percentage point terms. Policy 1 is a unilateral reduction of 20% in *fixed* inward offshore R&D costs for China and India from the benchmark equilibrium. Policy 2 is a unilateral reduction of 10% in inward offshore production costs for China and India. Policy 3 simulates the same shock as in Policy 2 in a restricted version of the model without offshore R&D.

we can see, China and India benefit by 0.6 and 0.9 percent in welfare from such a policy, while other countries are not significantly affected.

The second experiment is a liberalization in inward offshore production, which increases ϕ_{oi}^{P} by 10% for $i =$ India, China, $o \neq i$. Because these two types of liberalizations do not necessarily share the same fiscal costs or administrative burdens, I do not compare the levels of the welfare gains, but instead focus on the distributions of the welfare gains across countries. The second column of Table 1.7 shows that India and China still benefit from this liberalization. But differently from the first experiment, major developed countries also benefit significantly. The difference between these two experiments is due to the interaction between offshore R&D and countries' existing specialization in the world economy. Because offshore R&D into China and India reduces these two countries' existing comparative advantage in production, it pushes developed countries to be less specialized in innovation. As a result, they do not benefit much in the first experiment. Inward offshore production liberalization,

the initial investment cost.

on the other hand, allows China and India to be more specialized in production, and developed countries to be more specialized in innovation, thus benefiting everyone. These two experiments demonstrate that openness to offshore R&D and offshore production could have different welfare implications for other countries. It is thus very important to separate offshore R&D and offshore production in the model, to better evaluate specific policies.

Because of the interactions among the three forms of global integration, incorporating offshore R&D also affects our understanding of the effects of other types of policies. I focus on China and India as an example to illustrate this point. Specifically, I consider the same liberalization for China and India in inward offshore production as in the second experiment, but in a restricted version of the model without offshore R&D. The welfare impacts of this experiment are reported in the third column of Table 1.7. Compared to the second column, the welfare gains are significantly smaller for China and India, but larger for developed countries. The reason for the difference is that, when there is no offshore R&D, openness to offshore production only crowds out R&D by domestic firms, so Chinese and Indian firm owners bear all the reduction in profit from increased inward offshore production. This reduces the aggregate welfare gains in these two countries, but increases the welfare gains to developed countries. The difference in welfare impacts suggests that even if one's goal is solely to understand the effect of liberalizing offshore production, it is important to incorporate offshore R&D into the model.

1.6 Conclusion

Talented researchers and efficient firms are both necessary inputs to invention of new products, but they are distributed unevenly across countries. By carrying out their R&D activities offshore, firms mobilize their management technology across borders, which might generate important aggregate gains. This chapter develops a unified model of firms' global R&D and production decisions, featuring talent-acquisition and market-access motives for offshore R&D. I use the model to perform quantitative analysis on the determinants and welfare implications of offshore R&D.

The welfare gains from offshore R&D are on average 2.5% of real income. Incorporating this channel amplifies the welfare gains from openness by a factor of 1.2 on average, with more amplification for developing countries than for developed countries. Further experiments show that a country's openness to offshore R&D and offshore production have very different spillover effects to other countries. Moreover, because of the interaction among various forms of international integration, whether offshore R&D decisions are allowed makes a difference when evaluating the effects of liberalizing offshore production. All these results point to the importance of incorporating offshore R&D into existing trade models for a better understanding of globalization.

As a first step towards quantitatively evaluating offshore R&D, this chapter abstracts from three important features of the reality that might affect the results. First, I do not allow spillover effects between foreign and domestic R&D centers. Technological spillover from foreign firms is one of main benefit cited by policy makers across the world when making a case for FDI. Indeed, if domestic firms can learn from innovative and efficient foreign firms, then the gains from offshore R&D could be larger than predicted in the current model. More generally, not only offshore R&D, but also offshore production by multinationals might generate knowledge spillover to domestic firms. In the current model, I assume that there is no such spillover. The main reason for this assumption is that the literature has not yet reached a consensus on the size of such spillovers. Understanding the magnitude of the spillover from FDI, and incorporating this channel into the analysis is thus an important task for future research.

Second, I do not consider the migration of high skill workers across countries. In the data, a substantial fraction of world patents are invented by workers residing outside their home country (Miguelez and Fink, 2013). As an alternative for firms and inventors from different countries to work together, the migration of inventors will likely have an important effect on firms' offshore R&D decisions, which in turn will affect the impact of government policy on offshore R&D. This might lead to interactions among policies on offshore R&D and high skill immigration. For example, taxing firms' profits from offshore R&D will incentivize firms to perform more R&D domestically if high-skill immigration is allowed, compared to if it is banned. In the future, I intend to incorporate migration into the analysis and shed light on the interaction of high skill immigration and offshore R&D policies.

Third, this chapter assumes perfect intellectual property right (IPR) protection in all countries, whereas in reality, countries differ in this regard and firms likely take this difference into account when deciding where to perform R&D. Understanding the effect of IPR protection is important for policy, not least because requirements on IPR protection have become an important clause of many regional trade and investment agreements. Given the prevalence of offshore R&D documented in this chapter, it is crucial that we incorporate multinational firms' decisions when trying to understand the effect of such IPR clauses. The framework presented in this chapter provides a natural starting point to model and quantify the effect of IPR protection on the income of nations, while changes in IPR protection driven by the past trade and investment agreements provide a source of data for empirical analysis. Combining these two approaches to understand the effect of IPR protection is a direction I intend to pursue in the future.

Chapter 2: Offshore R&D: Evidence

2.1 Introduction

The model presented in the first chapter generates several predictions that relate offshore innovation to firm innovation efficiency and host country characteristics. The first goal of this chapter is to test these predictions. Specifically, firm heterogeneity implies that more efficient firms will offshore their R&D to a larger number of host countries, and perform more R&D in each of them. At the host country level, the market-access motive implies that larger countries are more attractive as a host for offshore R&D centers. The talent-acquisition motive has two implications. First, host countries with higher relative inventor talent abundance attract more offshore R&D. Second, as Proposition 2 indicates, this effect is especially strong for more efficient home country firms, because of the complementarity between researcher talent and firm innovation efficiency in R&D.

The primary source of data I use in testing these predictions is patent-level information from the USPTO. While all patents in the USPTO are filed in the U.S. for the protection in the U.S. market, they could be invented anywhere in the world. A patent can be invented in one country (based on the location of its inventor(s)), but assigned to an owner in another country (based on the location of the owner).

As an example, I classify a patent by an inventor living in India, but assigned to a company registered in the U.S., as the output of offshore R&D conducted by the American firm in India. I perform such classification patent by patent, and then aggregate the data to the firm and country level, so that I know, for example, how many patents General Electric invented in Japan, Germany, Britain, etc. I construct measures of firm efficiency and researcher talent in a country based on the shares of highly innovative firms and highly prolific inventors located in the country, using the USPTO database. These different pieces of information from the USPTO are then combined with other country characteristics, such as market size, income, and country-pair characteristics, such as geographic and cultural distances.

Evidence from the combined data set supports the model's predictions. At the firm level, along the extensive margin, a firm's R&D efficiency in its country of registry (home country), as proxied by the number of patents invented by the firm in that country, correlates strongly with the number of foreign countries it enters to perform R&D. A typical firm registered in the U.S. conducts R&D only in the U.S., while the most innovative firms, such as IBM, Microsoft, and P&G, conduct R&D in more than ten countries. Along the intensive margin, a host subsidiary whose parent has an above-median innovation efficiency in the home country performs 62% more R&D in the host country than a subsidiary with a below-median parent from the same home country.

At the host country level, a 1% increase in host country GDP increases offshore R&D output in the country by 0.08%. A 1% increase in host country relative talent abundance increases offshore R&D output by 0.09%. These results are supportive of the talent-acquisition and market-access motives. Importantly, the positive effect of host country relative talent abundance is concentrated in the most efficient firms, consistent with complementarity between firm efficiency and researcher talent.

In addition to its direct predictions on offshore R&D, the complementarity assumption in the model also implies that more innovative firms work with more talented researchers (Lemma 1). The second goal of this chapter is to test this implication using firm- and inventor- level data from the USPTO. Using past innovation as a proxy for inventor talent, and various measures of firms' R&D efficiency, I show that, among a sample of job-switching inventors, the more talented ones tend to switch to more efficient firms, consistent with assortative matching.

This chapter contributes to three strands of literature. First, I add the management literature that investigates firms' offshore R&D incentives. Most studies in this literature are either based on firms' self-reported motives, or focus on firms in or from a single region.[1] This chapter complements existing studies by testing the talent-acquisition and market-access motives of offshore R&D using patenting information for firms from a large number of countries.

Second, I contribute to the empirical economic research on the patterns of FDI, as reviewed in Chapter 1. Most of existing studies in this literature focus on examining capital investment or output foreign firms. This chapter instead focuses on R&D, which is also an important decision made by multinational corporations. The use of patent data is not new. Indeed, a growing literature on globalization of

[1]See, for example, Ambos (2005), Shimizutani and Todo (2008), and Ito and Wakasugi (2007). See also Hall (2011) for a review.

R&D has used patent data (Kerr and Kerr, 2014; Kerr et al., 2016; and Branstetter et al., 2013). This chapter differs in that it uses the patent data to construct firm-level measures for offshore R&D, as opposed to other forms of R&D globalization, such as international co-invention or high-skill immigration.

Finally, I establish positive assortative matching between inventors and firms, a prediction of the model assumption of complementarity in innovation. While several existing studies have documented positive assortative matching in general labor markets and the market for managers, this chapter is, to my knowledge, the first to document positive assortative matching between inventors and firms.[2]

2.2 Specification and Hypothesis

2.2.1 Direct Predictions on Offshore R&D

I use mainly the following specification to test the model predictions on offshore R&D:

$$\log(y_{foi}) = \beta_0\delta_f + \beta_1\gamma_i + \beta_2\gamma_i q_f + \beta_3 x_{oi} + \epsilon_{foi}, \tag{2.1}$$

where f, o and i index for parent company, home country, and host country, respectively. The dependent variable, y_{foi}, is a measure of innovation output by company f's affiliated R&D centers in country i (multiple affiliates in the same countries are aggregated into one). The first independent variable, δ_f, is the firm fixed effect,

[2]Existing research mostly focuses on the match between workers in general and firms (see for example, Abowd et al., 1999 and the references thereto). More recently, research has focused on the match between firms/projects and CEO (Terviö, 2008, among others).

which controls for characteristics that are common to all R&D centers affiliated with the same firm. I exclude the firm fixed effect in some specifications to examine the effect of firm innovation efficiency. γ_i is a vector of host country characteristics that might affect offshore R&D and patenting, including size, relative talent abundance, per-capita income, intellectual property right protection (IPR), and general human capital. When these characteristics are not of primary interest, I use host country fixed effects instead. q_f is firm innovation efficiency. The interaction term $\gamma_i q_f$ captures how host country characteristics affect firms with different efficiencies. Of prime interest among these is the interaction between host country relative talent abundance and firm efficiency. x_{oi} is a vector of variables that vary across host-home pairs, including various measures of distance. When the interest is not in host country or bilateral characteristics, I use country-pair fixed effects to capture this term. ϵ_{foi} is the error term.

The talent-acquisition and market-access motives imply that market size and relative talent abundance increase offshore R&D into a host country, with the latter having a stronger effect for more efficient firms. Moreover, as most models with firm heterogeneity would predict, more efficient firms enter more countries for offshore R&D, and innovate more in each of them. In the context of econometric specification 2.1, these model predictions imply positive coefficients for measures of host market size, host relative talent abundance, firm innovation efficiency, and the interaction between firm innovation efficiency and host market size.

2.2.2 Positive Assortative Matching

The log-supermodularity assumption implies strict positive assortative matching between firms and researchers. To test this implication, my empirical strategy is to measure the qualities of firms and inventors, and then assess whether there high-quality firms are matched with high-quality inventors. The idea is that, if there are greater values for high-talent inventors to work with high-efficiency firms, such matches should show up more in the data than other kinds of matches.

2.3 Data Description

2.3.1 Direct Predictions on Offshore R&D

I use patent data from the USPTO to construct three key measures used in the specification: offshore R&D center innovation output, firm innovation efficiency, and host country relative talent abundance.

Firm and inventor classification To construct these measures, I need to be able to identify individual inventors and firms. This is challenging because patent data is self-reported, so there are no individual or firm identifiers. Moreover, typos and misspellings are frequent, and the same company might have different abbreviations. I follow the patent literature in addressing these issues. For the firm side, I use the 2006 update of the disambiguated data set introduced by Hall et al. (2001), which covers patents granted from 1976 to 2006. By combining automatic cleaning procedures—which take care of common abbreviations in company names—

with manual checks, Hall et al. (2001) generates a unique identifier for each patent owner. For inventors, I use the unique inventor identifiers provided in Li et al. (2014), which uses a supervised learning approach to automatically generate inventor identifiers.

Offshore R&D output measure When applying for a patent at the US-PTO, the applicant, usually the owner, reports address information for both the inventor and the owner of the proposed patent. I classify a patent as invented in a country-i offshore R&D center, affiliated to firm f from country o, if its inventor is located in country i and its owner in country o.[3] Counting the total number of such patents by each firm in each host country, I obtain the benchmark measure for y_{foi}. In essence, this is a firm-level counterpart of the offshore R&D measure used in the quantitative section of Chapter 1.

Firm innovation efficiency measure I use the total number of patents invented by firm f in its home country o as a proxy for its innovation efficiency. I focus on home country innovation for this measure, and drop observations from the home country of each firm when estimation Equation 2.1, so that the results are not driven by the mechanical correlation between home innovation and the measure of innovation efficiency. To reduce measurement error, in benchmark regressions, I classify a firm according to whether its innovation efficiency is above the median in its home country. Later I will show results with different cutoffs.

Host country relative talent abundance All research firms in a country

[3] Importantly, this information is reported at the time of patent application, so transfers of patents between owners from different countries are unlikely to be important.

compete for talent. It is the abundance of talented inventors *relative to* the abundance of efficient firms that matters for the type of inventor a foreign offshore R&D center is able to recruit. Following Definition 1 from the theoretical model presented in the first chapter, I construct the measure for absolute inventor talent abundance as the share of inventors in a country that are in the top 1% most productive inventors in the world; I construct the measure for absolute firm efficiency abundance analogously. I then use the log of the ratio between the two as the benchmark measure for relative abundance. Taking the ratio also nets out some of the differential selection across countries into patenting in the U.S.[4]

I use a relative quality measure, not relative quantity measures (e.g., the number of inventors relative to the number of firms), because the model predicts that a change in the relative quantity will have no impact on the matching function or the wage schedule.[5] In robustness checks, I also include this relative quantity. The choice to use the top 1% of inventors and firms in constructing this measure is motivated by the importance of exceptional inventors and firms in aggregate innovation. In robustness exercises, I use different cutoffs for computing the top shares, and other measures of quality in constructing the ratio.

Discussion on the use of patent data to measure R&D The advantages of using patent data for these measures are obvious: in addition to having

[4]For example, patenting might be easier in some countries, so marginal firms and inventors self select into patenting, resulting lower measured average inventor and firm quality.

[5]By inspecting Equation 1.7, we can see that $\frac{L_i^R}{R_i}$ only enters as a ratio, and by taking derivatives of both sides with respect to z^R, we can see that this ratio does not affect $T_i^{''}$; hence it will not affect the matching function or the shape of the wage schedule. The ratio $\frac{L_i^R}{R_i}$ will thus only affect the research team size and output of all R&D centers in country i proportionally.

a wide country coverage, patents are also highly correlated with firm-level R&D. Figure A.1 and Table A.2 in the appendix show that a patent-based offshore R&D measure correlates reasonably well across countries with a measure based on R&D expenditures. However, the drawbacks of patent data are also well known (Pavitt, 1988). First, the benefits of patents differ across countries, so that firms might have different incentives to apply for patents in the U.S. These differences might stem from market size, intellectual property right protection, or other country characteristics, such as connections to the U.S. Second, different industries have different reliance on patents for the protection of their intellectual property. Third, patents have heterogeneous values, so patent counts are a noisy measure of firms' innovation output.

I add additional controls to address these concerns. Specifically, for the first concern, I either control directly for host country size, IPR protection, and other country characteristics, or simply include host country fixed effects. For the second concern, I use firm fixed effects to absorb firms' characteristics, including their industry. Moreover, I construct measures at the patent-category level so that host country specialization does not drive the results. Finally, to address the third concern, I also use patent citations as an alternative measure of innovation output.

Sample period The patent data spans 1976-2006. Since both the dependent variable, offshore R&D output, and the key independent variable, the relative talent abundance, are constructed based on patenting data, measurement error will lead to correlation between the two measures. To avoid this problem, I split the sample into two periods, 1976-1996 and 1997-2006. I use only information from the first period to measure the number and quality of innovating firms and individuals. I then use the 1997-2006 data to measure R&D output for each parent company and its foreign subsidiaries. Further, my regressions include only observations from *new* offshore R&D centers—those that enter in the second period—in order to prevent any R&D centers used as regression observations from affecting host talent quality measures. This sample split also prevents reverse causality, i.e., the entry of innovative and efficient foreign firms attracting more talented individuals to become inventors.

Additional data Additional variables used for the regressions are from the following sources: GDP, population, per-capita income, and a human capital index come from the Penn World Table 8.0; bilateral distance information is from the CEPIT outdistance database (Mayer and Zignago, 2011); and intellectual property protection information is taken from Park (2008). All these variables are averaged over 1997-2006 for consistency.

2.3.2 Positive Assortative Matching

I rely also on the USPTO patent-level data for this test. As in the previous subsection, I construct a panel of inventors and firms using the inventor identifier from Li et al. (2014) and the firm identifier from the NBER patent database project (Hall et al., 2001). This data set has a structure that resembles that of a matched employer-employee data set, except that here a match only shows up in a given year, when a patent is filed.

Using lagged innovation as a proxy for inventor ability, and various lagged measures of firm innovation efficiency, I investigate whether more talented inventors are more likely to switch to high-efficiency firms. Since patents are invented jointly by inventors and firms, correlating these two measures would pick up their mechanical correlation. To avoid this problem, I focus on a sample of inventors that switch firms and examine, among them, whether the more innovative ones are more likely to move to more productive firms.

2.4 Results on Offshore R&D

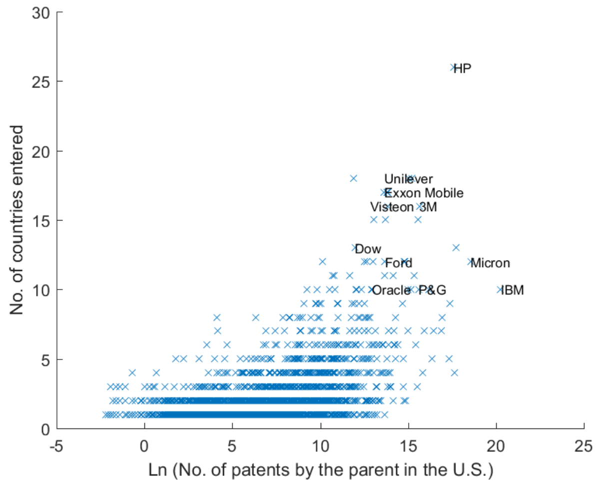
I first discuss the results on Offshore R&D.

2.4.1 Baseline Results

Figure 2.1 provides evidence on the effect of firm innovation efficiency on offshore R&D through the extensive margin over the period 1997-2006, focusing on firms headquartered in the U.S. Each dot represents a firm. The horizontal axis is the number of patents granted to the firm and invented in the U.S. The vertical axis is the number of countries in which the firm performs R&D. The figure indicates that firms with higher innovation efficiency tend to perform offshore R&D in more countries. Among the firms that enter the largest number of countries, IT and chemical companies are the most common.

I now estimate Equation 2.1 to test additional model predictions. Table 2.1 presents the baseline results. The first column includes the indicator for firm innovation efficiency, a vector of host country characteristics, bilateral distance mea-

Figure 2.1: Firm Efficiency and Offshore R&D Entry



Notes: Each dot represents a firm headquartered in the U.S. The horizontal axis is the log of the number of patents the firm invented in the U.S. The vertical axis is the number of host countries it entered for offshore R&D, defined as $\sum_i \mathbb{I}_{y_{foi}>0}$.

sures, and home country fixed effects. Consistent with the first implication of the talent-acquisition motive, host country relative talent abundance has a positive and statistically significant impact on innovation output. The estimate has an elasticity interpretation: a 1% increase in host relative talent abundance increases firm-level offshore R&D by around 0.1%. Consistent with the market-access motive, host GDP also has a positive effect with a similar point estimate. Firms with above median innovation efficiency generate 63% higher R&D output, on average, so innovation efficiency increases offshore R&D through not only the extensive margin, but also the intensive margin. The estimate for host country intellectual property right protection is small and statistically insignificant, reassuring us that differential selection into patenting due to intellectual property protection differences are not driving the results. Host per-capita income does not have a significant effect. Distance measures are mostly insignificant, except for the common language indicator.

The second column adds the interaction term between the host country relative talent abundance and firm innovation efficiency. This interaction term is positive and statistically significant, with a point estimate of 0.15. While most other coefficients do not change, the coefficient for host relative talent abundance is no longer significant: consistent with the prediction from Proposition 2, the impacts of host talent quality are mainly concentrated in the top half of firms as ranked by innovation efficiency.

Results so far are supportive of a market-access motive and a talent-acquisition motive. Since the market-access motive is closely related to the extensive existing literature on the effects of market size on innovation (Acemoglu and Linn, 2004) and the location choice of multinational firms (Head and Mayer, 2004), I now focus on the talent-acquisition motive by further examining the interaction term. In the third column, I add host country and parent firm fixed effects to further absorb unobserved heterogeneity. The point estimate of the interaction term rises to 0.177, meaning that a 1% increase in relative talent abundance in the host country increases the R&D output by 0.17% more for R&D centers with above-median efficiency. After adding these better controls for host country and firm heterogeneity, bilateral geographic distance becomes significant, with an elasticity of -0.119. Common language, on the other hand, is no longer significant. The fourth column adds country pair fixed effects to capture differential economic connections between countries. The point estimate of the interaction term barely changes.

Table 2.1: Determinants of Offshore R&D: Baseline Results

Dependent Variable:	Log (Offshore patents invented in a host country)			
	(1)	(2)	(3)	(4)
Host relative talent abundance	0.091**	0.020		
	(0.039)	(0.041)		
I (Parent R&D >median)	0.625***	0.994***		
	(0.076)	(0.077)		
Host inventor relative abundance * I (Parent R&D >median)		0.147***	0.177***	0.171***
		(0.026)	(0.040)	(0.040)
Host GDP	0.083**	0.088**		
	(0.035)	(0.035)		
Host per-capita income	-0.004	-0.001		
	(0.076)	(0.075)		
Host IPR protection	0.032	0.026		
	(0.117)	(0.118)		
Distance	-0.007	0.003	-0.119**	
	(0.026)	(0.028)	(0.055)	
Common border	-0.009	0.003	0.203	
	(0.051)	(0.052)	(0.131)	
Common language	0.152***	0.151**	0.095	
	(0.056)	(0.056)	(0.094)	
Colonial tie	-0.075	-0.074	0.027	
	(0.049)	(0.051)	(0.092)	
Home country FE	X	X	X	
Host country FE			X	
Home-Host FE				X
Firm FE			X	X
Observations	14803	14803	7914	7716
R^2	0.050	0.053	0.454	0.490

Notes: The level of observation is host country-parent company. The dependent variable is the log of the total number of patents invented by an affiliate of a parent company in a host country over 1997-2006. For the I(Parent R&D ¿median) indicator, Parent R&D is measured by the total number of patents invented by the parent in its home country during the same period, and median is computed for all patenting firms in the home country of the parent company. *Host relative talent abundance* is defined as the log difference between the share of inventors in a host country that fall into the global top 1%, and the share of firms in that country that falls into the global top 1%. This measure is constructed using only patenting information for 1976-1996 to avoid mechanical correlation. *Host IPR* is the intellectual property right protection index from Park (2008), averaged over 1997-2006. Other host country characteristics are from the Penn World Table 8.0, averaged over 1997-2006.
Standard errors (two way clustered at the host-country and parent-company levels) are in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.4.2 Robustness and Heterogeneous Effects

As discussed earlier, firms' incentives to perform R&D and to patent their R&D output are potentially affected by country characteristics. There may be plausible alternative theories that generate heterogeneous effects of these other characteristics for higher productivity firms. I now examine whether such alternatives can explain the baseline findings and generally find they cannot.

First, relative talent abundance might pick up an income effect. High-income consumers prefer high-quality products, which might be more R&D intensive than low-quality products. If efficient firms have comparative advantage in doing R&D, they might perform more R&D in high income host countries. I capture this by including the interaction between host country per-capita income and firm efficiency.

Second, the returns to both R&D and patenting are higher in large countries. Firms with higher efficiency might benefit disproportionately more because they tend to be more efficient in production. This concern motivates me to include the interaction between host country GDP and firm efficiency. Following the same reasoning, the effect of stronger patent protection enforcement might also benefit efficient firms more, encouraging them to patent more. Therefore I further include the interaction between firm efficiency and the IPR protection index.

Finally, the complementarity between talent and firm efficiency might happen in the production stage. Firms with better management can make better use of skilled workers in production, which reduces production costs and increases the return to R&D. I incorporate the interaction between the host country human capital index and firm efficiency to address this concern.

The first column of Table 2.2 reports the regression with these additional terms. The interaction between host talent and home efficiency is still significant, although it shrinks by about 40%. The variable that explains this drop is the interaction between host GDP and firm efficiency. Other interaction terms do not have strong effects.

I use a quality-based relative talent abundance measure in my baseline regressions because according to the model, the relative *quantity* of inventors and research firms will affect firms with different efficiencies proportionally. To make sure the empirical finding is not the result of improperly measuring the relative talent abundance, in the second column, I add the interaction between firm efficiency and the ratio of the *number* of inventors and the number of R&D active firms. Reassuringly, this quantity ratio does not have a statistically significant impact itself and does not substantially change the interaction term between host relative talent quality and firm efficiency.

For a fuller picture of how host talent affects firms with different efficiencies, in the third column, I add indicators for firms with R&D efficiency above the 25th, 75th, and 90th percentiles of the R&D efficiency distribution in their home countries, as well as the interaction of these indicators with the full set of controls in the second column. The effect of a better host talent distribution is substantially larger for firms in the upper tail of the distribution. A 1% increase in host relative talent abundance leads to a 0.45% larger increase in R&D output for firms in the top 10% of the firm efficiency distribution than for firms in the bottom 25% of the distribution.

So far, all the regressions have pooled patents over all categories to construct measures for both the dependent and independent variables. Aggregation reduces measurement errors, but given that industries do not equally rely on patents for IPR, using aggregate patenting data might confound sectoral composition with country-level relative talent quality. For robustness, I also construct all variables within each individual patent category, classified by Hall et al. (2001).[6] For each firm, I keep only the category in which it patents most. Columns four and five perform regressions using category-level data. The fourth column controls for host-category fixed effects, and the last column controls for bilateral pair-category fixed effects. Both columns confirm that the effect of host country talent is significantly larger for more efficient firms. Although the sample size is substantially smaller as more fixed effects are added, the coefficients are quantitatively similar to those in the third column, so the differential sectoral specializations of countries are unlikely to be the explanation for the benchmark results.

To ensure that the results are not driven by the specific ways in which I measure efficiency and talent, I perform additional robustness checks, using alternative measures of host relative talent abundance, R&D center innovation output, and firm innovation efficiency. The results are reported in Table 2.3. In columns 1-4, I aggregate data across all patent categories to construct measure for firm R&D output and host relative talent abundance. In columns 5-8, I use category-level data to construct R&D output as well as talent abundance (in regression, only the main

[6]There are in total six categories: chemical (excluding drugs), computers and communications, drugs and medical, electrical and electronics, mechanical, and others.

Table 2.2: Determinants of Offshore R&D: Alternative Explanations and Heterogeneous Effects

Dependent Variable:	Log (Offshore patents invented in a host country)				
	Aggregated Across All Categories			Only Firms' Main Category	
	(1)	(2)	(3)	(4)	(5)
Host relative talent abundance * I (Parent R&D >25%)			0.131	0.107	0.187
			(0.084)	(0.142)	(0.158)
Host relative talent abundance * I (Parent R&D >median)	0.105*	0.093*	0.117	0.134	0.217
	(0.054)	(0.051)	(0.091)	(0.159)	(0.175)
Host relative talent abundance * I (Parent R&D >75%)			0.208**	0.126	0.217
			(0.095)	(0.148)	(0.185)
Host relative talent abundance * I (Parent R&D >90%)			0.465***	0.354*	0.579**
			(0.146)	(0.173)	(0.217)
Host per-capita income * I (Parent R&D >median)	-0.093	-0.063			
	(0.148)	(0.146)			
Host GDP * I (Parent R&D >median)	0.119***	0.101**			
	(0.039)	(0.043)			
Host human capital * I (Parent R&D >median)	-0.102	-0.161			
	(0.214)	(0.224)			
Host IPR * I (Parent R&D >median)	0.167	0.155			
	(0.196)	(0.197)			
Host inventor relative *quantity* * I (Parent R&D >median)		0.125			
		(0.146)			
Home-Host FE	X	X	X	X	X
Firm FE	X	X	X	X	X
Host-Category FE				X	
Host-Home-Category FE					X
Host Characteristics *Full set of firm efficiency indicators			X	X	X
Observations	7687	7687	7687	4381	3907
R^2	0.490	0.490	0.496	0.562	0.605

Notes: See Table 2.1 for descriptions of variables and sample period. The first three columns use information aggregated over all patent categories. The last two columns use variables similarly constructed at the patent category level. For each firm, only the category in which it generates the most patents is used in regressions. Although not reported, columns 3-5 also include the interaction of host country characteristics in column 2 with the full set of firm R&D efficiency indicators. Standard error (two way clustered at the host-country and parent-company levels) are in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

category of each firm is kept). In all these regressions, I include full control variables from the second column in Table 2.2.

Within each of these two sets, I vary how I measure key variable to see if the results are sensitive. In columns 1-3 and 5-7, I use the same measure for the dependent variable, but vary how I construct independent variables. Specifically, I first vary the cutoff in defining "top" inventors and "top" firms from the top 1% in the baseline analysis, to top 10% in columns 1 and 5. In columns 2 and 6, I use the ratio between the average number of patents by inventors and the average number of patents by firms as the measure for the relative abundance in talented researchers. Since the dispersion in inventors' output is primarily driven by the output of the most talented inventors in a country, in columns 3 and 7, I use the ratio between the standard deviation of inventor output and the standard deviation of firm innovation output as a proxy for the relatively abundance of talented researchers. Finally, in columns 4 and 8, I use citation counts, rather than patent counts, to measure both the outcome variables and firms' R&D efficiency. All these alternative measures yield similar results.

Table 2.3: Determinants of Offshore R&D: Alternative Measures

Regression Level:	Aggregated Across Categories				Only Firms' Main Category			
Dependent Variable:	Benchmark (patent)			Citation	Benchmark (patent)			Citation-based
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Host relative talent abundance 2* I (Parent R&D >median)	0.146*				0.259*			
	(0.083)				(0.145)			
Host relative talent abundance 3* I (Parent R&D >median)		0.110*				0.254*		
		(0.056)				(0.130)		
Host relative talent abundance 4* I (Parent R&D >median)			0.140**				0.225*	
			(0.066)				(0.127)	
Host relative talent abundance* I (Parent R&D2 >median)				0.106**				0.173*
				(0.048)				(0.093)
Home-Host FE	X	X	X	X				
Host-Home-Category FE					X	X	X	X
Firm FE	X	X	X	X	X	X	X	X
Full Controls	X	X	X	X	X	X	X	X
Observations	8083	8149	8149	7687	4882	5265	5238	3907
R^2	0.486	0.486	0.486	0.501	0.591	0.588	0.587	0.600

Notes: All regressions include the full set of controls in the second column of Table 2.2. The definition of common variables are the same as Tables 2.2. *Host relative talent abundance 2* is defined as the ratio between the share of top inventors and the share of top firms, in which "top" is defined as among global top 10%. *Host relative talent abundance 3* is defined as the ratio between the average number of patents by inventors, and the average number of patents by firms. *Host relative talent abundance 4* is defined as the ratio between the standard deviation of the number of patents by inventors, and the standard deviation of the number of patents by firms. *Parent R&D 2* is the total citation of the patents invented by a parent company in its home country.
standard errors (two way clustered at the host-country and parent-company levels) are in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.5 Results on Positive Assortative Matching

Now I turn to the results on positive assortative matching between inventors and firms. The main findings are reported in Table 2.4. There are two panels in Table 2.4, each corresponding to a set of regressions with the same outcome variable. Each specification in the table regresses a measure of firm innovation efficiency on a measure of inventor talent, on a sample of inventors that have just moved to a new firm. The independent variable, same across all panels, is my preferred measure of inventor quality, which is the lagged value of log total forward citations to the patents filed by the inventor to date, adjusted by the number of inventors on each of these patents. The lagged value refers to the previous observation of the inventor in the database when he/she does not work for the present employer. This might be a few years back, however, if an inventor's last patent is from the distant past. The dependent variable in panel A is the lagged value of the log of total number of forward citations to the patents a firm has been granted. I use lag value here to ensure that the inventor under investigation is not also included in the outcome variable, leading to a mechanical correlation.

The first column adds no control variables. In the second column, I add the years since first patenting for firms and inventors to capture the life cycle effects, as it is plausible that inventors with different ages prefers firms at different stages of growth, for reasons not necessarily related to firms' innovation efficiency. In the third column, I add year fixed effects as well as category fixed effects. After controlling for these fixed effects, the point estimate shrinks somewhat, but is still

Table 2.4: The Match Between Firms and Inventors

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A	**Outcome: ln (Total citation to patents of the firm)** $_{t-1}$					
Inventor Quality: Measure 1$_{t-1}$	0.192***	0.290***	0.228***	0.025***	0.095***	0.025***
	(0.006)	(0.006)	(0.006)	(0.004)	(0.007)	(0.005)
Observations	54733	54733	54733	54733	54733	54733
Firm/Inventor Controls		X	X	X	X	X
Year/Category FE			X	X	X	X
New Employer FE				X		X
Previous Employer FE					X	X
R^2	0.017	0.240	0.315	0.898	0.664	0.927
Panel B	**Outcome: Firm Productivity (total citation per inventor)** $_{t-1}$					
Inventor Quality: Measure 1$_{t-1}$	2.460***	3.598***	3.244***	0.301***	1.480***	0.283***
	(0.080)	(0.092)	(0.091)	(0.061)	(0.098)	(0.066)
Observations	56096	56096	56096	56096	56096	56096
Firm/Inventor Controls		X	X	X	X	X
Year/Category FE			X	X	X	X
New Employer FE				X		X
Previous Employer FE					X	X
R^2	0.037	0.059	0.101	0.673	0.458	0.801

Notes: The regressions reported in this table use a sample of inventors that have switched firms. The independent variable is the lagged value of the log of total forward citations to the patents filed by the inventor to date, adjusted for the number of inventors for each of these patents. *Firm/Inventor Controls* refers to years since first patenting for the firm and for the inventor. *Year/Category FE* refers to year fixed effects and category fixed effects. Categories here are defined by Hall et al. (2001). There are in total six categories: chemical (excluding drugs), Computers and Communications, drugs and medical, electrical and electronics, mechanical, and others.

Robust standard errors in parentheses.

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

statistically significant. Column 3 is my preferred specification. The point estimate indicates that, an inventor with talent that is 1% higher will be matched to a firm with 0.2% higher innovation efficiency.

Columns 4 through 6 push further by adding fixed effects for previous employers, current employers, and both, respectively. Identifying the effect from *overtime* changes for a given origin or destination employer help overcome biases that might arise from unobserved firm heterogeneity. The cost is that the attenuation effect might be stronger. As indicated by the R^2, when both current and former employer fixed effects are added in Column 6, they absorb most of the variation. The coefficient decreases by more than 80%. However, it is still statistically significant.

Total forward citations might be a noisy measure of invention efficiency for the firm. For example, firms with larger researcher teams tend to have more inventions, hence higher citations, to their patents. Although researcher team size is a theory-consistent measure for innovation efficiency, some firms might have more inventors for reasons outside the model. Panel B uses a measure similar in spirit to firms' "labor productivity", defined as the per-inventor total forward citations in a given year, to address this concern. The results are all statistically and economically significant. The preferred specification in column 3 suggests that one percent increase in inventor productivity would increase the per-inventor citation of his/her employer by 3.

2.6 Conclusion

This chapter presents two sets of empirical results to test the model developed in the first chapter. The first set of results focuses on the model predictions on the impact of firm efficiency and host country characteristics on offshore R&D. The baseline specification finds that more efficient firms perform offshore R&D in a larger number of countries, and more R&D in each of them. Larger and more talent abundant countries are more attractive as a destination of offshore R&D. Moreover, the effect of talent abundance is stronger for more efficient firms. All of these results are consistent with the model. Additional exercises that consider alternative explanations and measurements do not substantially weaken the evidence.

The second set of results focuses on testing a direct implication of the complementarity assumption. I find that more efficient firms work with more talented inventors, consistent with positive assortative matching implied by the assumption.

Taken together, these results provide support for the key ingredients of the model of offshore R&D presented in the first chapter: talent-acquisition and market-access motives of offshore R&D, and the complementarity between researcher talent and firm efficiency.

Chapter 3: Internal Geography, Labor Mobility, and the Distributional Impacts of Trade

3.1 Introduction

In recent decades we have witnessed increasing integration of large developing countries, such as Brazil, China, India, and Mexico, into global trade. This trend has renewed the interest of policymakers and academics in understanding the impacts of globalization on income and inequality.

Existing research on this topic focuses on the impacts of international trade on workers in different industries, or with different skills, but abstracts from the geographic dimension of the distributional impacts.[1] Consider workers living far away from a nation's ports. Because of the high *intra-national* trade costs, they might not benefit much from cheaper imported products, and international trade can exacerbate the intra-national inequality in living standards. Moreover, in a world with both skilled and unskilled workers, if one type of worker is more mobile than the other, and responds to trade liberalization by migrating to the coast, then the workers left behind might even lose from trade. These losses can be independent of regional sectoral specializations. This geographic margin in the distributional

[1] More discussion of existing literature may be found in the next section.

impacts of trade is not only plausible, but also empirically relevant.[2]

The scenario discussed above naturally leads to the following questions: First, in the presence of *intra-national* trade and migration costs, how does *international* trade liberalization affect the aggregate income of a country, and its within-country inequality—including both the between-region inequality among workers with similar level of skills, and the within-region inequality between different types of workers? Second, most countries mentioned in the opening paragraph are investing in infrastructure and launching structural reforms, with the aim of reducing the within-country spatial frictions. To what extent would these changes affect our answer to the first question?

With a focus on China, this chapter answers these two questions through the lens of a quantitative model. The coexistence of rapid trade growth, large spatial inequality, and recent regional reforms that significantly reduced the internal migration costs makes China a useful setting for this study. As is well known, China has experienced rapid integration into world trade since its economic reform in 1978, and the process accelerated after China joined the WTO in 2001. At the same time, China has historically had high intra-national trade costs and strict controls on worker migration. Perhaps partially due to these intra-national frictions, China's economic growth over the past decades has been uneven. Indeed, as shown in Kanbur

[2] Limão and Venables (2001) documents that poor infrastructures dampen a country's participation in international trade; Atkin and Donaldson (2012) estimates the intra-national trade costs to be 4-6 times larger in their sample of African countries than in the United States. Topalova (2010) shows that in India, trade liberalization hurt the poorest workers because of their limited inter-regional and inter-sectoral mobility. See also Kanbur and Venables (2005) for an excellent overview of the UNU-Wider project on "Spatial disparities in development," which analyzes evidence in over 50 developing countries, and concludes that international trade and the lack of infrastructure are two important factors in the increasing spatial disparities in many of these countries.

and Zhang (2005), inter-regional inequality grew rapidly during the period of fast trade expansion in China. In terms of migration restriction, China has gone through several waves of labor market reforms since the late 90's. Gradually, Chinese cities have relaxed the once-strict restriction Hukou restriction—a household registration certificate that ties individuals to their home region—so that people migrating from other cities have an easier access to employment and local public goods. Importantly, these reforms differ in timing and strength, providing variation to estimate the impact of such structural reforms on worker mobility. This estimate can then be fed into structural models to quantify the interaction between domestic reforms and international trade integration.

This chapter proceeds in three steps. In the first step, I document large spatial inequalities among Chinese cities in terms of participation in international trade, income, and urbanization, which underscore important domestic trade and migration costs in China. I also construct the first city-level panel of Hukou reform policies, spanning 1997-2010, using information from databases on law and government regulations. Combining this database with micro-data from the Chinese population censuses, I show that relaxing the Hukou restriction increases migrants into a city. From this estimate, I back out the underlying change in migration costs due to the Hukou restriction, which is used in counterfactual experiments.

In the second step, I develop a spatial equilibrium model with multiple regions representing Chinese regions and a statistical aggregation of the rest of the world. Regions are connected to each other through costly migration and trade. The trade block of the model builds on the quantitative trade theory (Eaton and Kortum, 2002). To capture the effect of international trade on income and the skill premium, the model incorporates several channels emphasized in the literature: the factor content of trade dating back to Stolper and Samuleson, trade in capital goods and capital-skill complementarity (Burstein et al., 2013 and Parro, 2013), and input-output linkages (Caliendo and Parro, Forthcoming). While these ingredients are not new, in this chapter, these channels will have differential impacts on Chinese regions due to these regions' differences in skill composition, sectoral specialization, and participation in international trade, all of which are endogenous. On the worker side, workers decide where to work, based on the utility they would obtain from all potential destinations, which in turn depends on region-specific amenities, prices, and wages, as well as city-specific labor productivity draws. The differential impacts of trade across regions and the migration of workers will prove important in shaping how trade affects skilled and unskilled workers from different parts of China.

The third step parameterizes the model and performs counterfactual experiments. I estimate migration costs using individual level data from the 2000 population census, separately for skilled and unskilled workers. This estimate includes both workers' home biases and restrictions arising from policies such as the Hukou system. I take this estimate as the benchmark measure. In some counterfactuals, I reduce the estimated migration costs by an amount implied by the empirically estimated effect of actual Hukou reforms in the first step, to investigate how migration reforms affect international trade integration. I estimate domestic trade costs using regional trade flows. I calibrate additional parameters to match the distribution of income and employment, and trade openness in 2005, using various micro and macro data sources,

I examine the distribution of the gains from international trade within China, by shutting down trade between the model economy and the rest of the world. In line with results from papers without internal geography, the average gains from trade (across regions and worker groups, weighted by population) are around 7.5%. However, the welfare gains are distributed unevenly. Along the skill dimension, skilled workers gain 11% on average, while unskilled workers gain only 5%, so the average skill premium increases by 6%. Impacts also differ among workers with similar skills. For example, among urban unskilled workers, some gain as much as 20%, while others experience 2% welfare losses. The geographic location of a region is important: regions on the coast reap most of the welfare gains, and regions in the interior benefit little. Aggregate inequality, as measured by the Theil index, increases by 8% after the international trade liberalization in China. The geographic dimension—the increases in inequality between geographic regions—accounts for 64% of the overall increase in inequality, while the skill dimension—the increase in within-region inequality—accounts for the rest. Despite the rise in inequality, however, only the unskilled workers from a dozen of regions experience welfare losses. Therefore, there is scope for government redistribution to ensure that trade liberalization is a Pareto improvement for China.[3]

Consistent with existing reduced-form evidence (Han et al., 2012), there is an active interaction between the geographic dimension and the skill dimension: the

[3]Since the workers that experience welfare losses are geographically concentrated, place-based transfers might be a way to target these people. In the text I also discuss the impacts on the welfare of these people of structural reforms that reduce domestic trade or migration.

rises in regional skill premia are larger in coastal regions, that is, there is a negative gradient in changes in skill premia with respect to regions' distances to the coast. In addition to capital-skill complementarity, which intuitively increases skill premia more in regions that import more capital goods, I uncover two forces behind the gradient of changes in skill premia with respect to location, both of which are inherently related to geography. First, because capital and other manufacturing industries use intermediate varieties more intensively, they tend to locate in regions with better access to suppliers. After trade liberalization, the coastal regions experience a larger increase in access to foreign suppliers and therefore have stronger comparative advantages in these industries. As a result, coastal regions specialize in capital and manufacturing industries, which hire skilled workers more intensively, while interior regions specialize in the unskilled-intensive agricultural industry. This change in the *domestic* specialization pattern following trade liberalization increases skill premia on the coast and decreases skill premia in the interior. Because this channel works through the factor content of intra-national trade, I call it the "Domestic Stolper-Samuelson Effect."[4] To the best of my knowledge, this is a novel channel in the literature. An implication of this channel for empirical studies is that in measuring a region's exposure to international trade, it is important to take into account the responses of both international trade and domestic trade to the reductions in international trade barriers.

The second force is related to the differential mobility of skilled and unskilled

[4]The standard Stolper-Samuelson effect, which operates through international specialization, is less important here, as the trade between China and the ROW is largely within sector.

workers. Because the estimated migration costs are lower for skilled workers, more skilled workers respond to trade liberalization by migrating out of their hometowns in the interior towards the coast, resulting in higher skill shares on the coast, and lower skill shares in the interior. This channel decreases skill premia on the coast and increases skill premia in the interior, offsetting the "Domestic Stolper Samuelson Effect".

I show that both forces are quantitatively important for changes in skill premia after trade. Incorporating the internal geography of a country is thus relevant for our understanding of the distributional impacts of trade, along both the geographic dimension and the skill dimension.

To shed light on how domestic frictions affect the welfare impacts of trade, I conduct four additional hypothetical international trade liberalization experiments. The only difference among these experiments is that the model "China" in these economies has either inter-provincial trade costs, or migration costs, or both, reduced through structural reforms. To make sure that these reforms are realistic—in the sense that the decrease in frictions in these reforms are attainable—I use the U.S. as the benchmark for the scale of the domestic trade reform, and the estimate of the mobility effect of Hukou policy changes for scale of the migration reform.

I find that reductions in both domestic trade and migration costs can help distribute the gains from international trade more evenly across the country. However, while Hukou reforms increase the gains from international trade, domestic trade reforms decrease gains from trade. The intuition is as follows: when domestic trade costs are smaller, coastal regions trade more with the interior region. Moreover, as the interior region becomes more attractive due to the decrease in the internal trade cost, more workers are willing to move there, increasing the size of the region that does not trade much with the rest of the world. The country as a whole therefore participates less actively in and benefits less from international trade. Overall, these results highlight the potential impact of domestic infrastructure investment and structural reforms on international trade. The different outcomes across these experiments also call for a quantitative approach for a concrete understanding of how domestic reforms interact with international trade liberalization.

3.2 Related Literature

This chapter contributes to the literature on the impact of trade on inequality. Inequality manifests itself in many dimensions. While existing studies have analyzed this issue from different angles, [5] the inequality between skill and unskilled workers is an important dimension emphasized by the literature. Different from most existing work in this literature, this chapter also studies the geographic dimension, that is, the inequality between geographic regions. It makes two contribution to this literature. First, I find that the geographic dimension accounts for a larger share of the increase in inequality from trade than the skill dimension, highlighting the

[5] Goldberg and Pavcnik (2007) offers a review on the effects of globalization on various measures of inequality, including the skill premium; Ma (2013) and Tang (2014) study inequality between top income earners and the rest of the population; Helpman et al. (2016) studies trade and the exporter premium both theoretically and empirically; Dix-Carneiro (2014) and Cosar (n.d.) study the impacts of trade on workers with different sector-specific experience; Levchenko and Zhang (2013) uses a framework similar to this chapter to assess the impacts of trade under different assumptions on the inter-sectoral mobility of labor and capital; Autor et al. (2013), Kovak (2013) and Dix-Carneiro and Kovak (2015) focus on the impacts of trade on different local labor markets, but in settings in which the differential impacts are only driven by sectoral specializations of regions, not geography. Rodriguez Chatruc (2016) and Kumar (2007) emphasize the role of geography.

importance internal geography for the relationship between trade and inequality. Second, I find that international trade has differential impacts on regional skill premia due to domestic trade and migration costs. Because of this interaction, modelling the geographic dimension is important even when one's main interest is in understanding how international trade affects the skill premium.

The spatial equilibrium model in this chapter builds on Redding (2012), which studies the gains from international trade, taking into account the mobility of labor within a country.[6] Contemporaneously, Tombe and Zhu (2015) also extends Redding (2012) to examine issues related to trade and labor mobility frictions in China. Relative to these two papers, the first contribution of the present chapter is that, in addition to internal geography, it also incorporate skill and unskilled workers. Doing so allows me to decompose the aggregate inequality effect of trade into a geography and a skill dimension, and demonstrate the relative importance of the two. Moreover, I show these two dimensions interact with each other to shape the distribution of the welfare gains from trade.[7] There exist other work that looks at *either* domestic trade costs, *or* migration costs in China, but not both. See, for example, Poncet (2005), Tombe and Winter (2014), and Au and Henderson (2006). This chapter models both trade and migration costs in a unified framework for

[6] Also closely related to this literature is an economic geography literature that examines the interaction between international trade and the distribution of economic activities within a country. See for example, Krugman and Elizondo (1996), Venables and Limão (2002), and Hanson (2001,0) for earlier contributions. More recent studies, such as Allen and Arkolakis (2014) and Cosar and Fajgelbaum (2016), develop quantitative models to take to the data. This strand of literature, however, typically treats workers as perfectly mobile, ruling out the analysis of the distributional impacts.

[7] To analyze how international trade affects skill premia across regions, the model here is richer than Redding (2012) and Tombe and Zhu (2015) in that it incorporates three channels: trade in capital good and capital-skill complementarity, regional specialization and the skill content of trade, and differential mobility of skilled and unskilled workers.

quantification. Doing so is important as the welfare impacts of international trade integration will only have a spatial dimension when both channels are present.

The second contribution is the empirical application. In the application to China, to understand the scope for domestic migration reforms to interact with international trade, it is necessary to isolate the part of migration cost that could potentially be changed by policies. Building on recent work Kinnan et al. (2015) and Sun et al. (2011), I construct a new city-level panel database on Hukou reforms. The database spans the period 1997-2010, and documents more than 600 changes to Hukou policies at city-year level. I use this data to estimate the effect of Hukou policy changes on migration. This allows me to predict what would happen to trade if there was a comprehensive across-the-board Hukou reform.

This chapter thus also contributes to empirical research on internal migration in China. The Hukou reform database constructed in this chapter might be of use to other people interested in this topic. The effect of Hukou reform on migration, estimated in this chapter, is also an important question yet to be settled.[8]

3.3 Background

This section documents some important stylized facts about the economic geography of China and the background of the Hukou system. This information will be incorporated into the quantitative framework.

[8] Using province-level Hukou reform information, Sun et al. (2011) finds little evidence for impact of Hukou reform on migration. On the other hand, using province-level Hukou reform tally interacted with network effect between province pairs, Kinnan et al. (2015) finds strong effect of Hukou reform on migration.

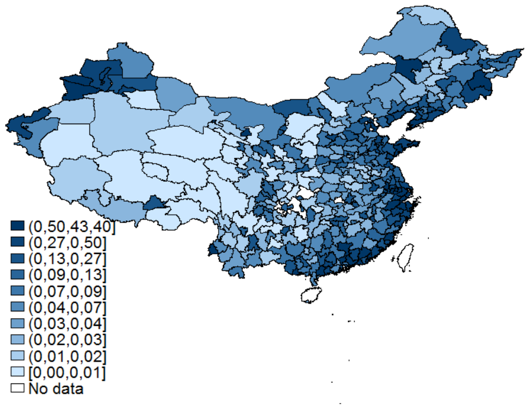
3.3.1 The Economic Geography of China

Panels (a) and (b) of Figure (3.1) plot trade openness and urbanization rates for cities in China.[9] The border regions in China, especially those on the east coast, trade with the rest of the world much more intensively than the interior cities. At the same time, as Panel (b) shows, the east coast also tends to have a much larger urban sector. These spatial differences can be due to the large intra-national trade costs in China, or to differential regional comparative advantages. The quantitative exercise below incorporates both domestic trade costs and regional differences in sectoral productivity. I will estimate the former using domestic trade data, and back out the latter using the distribution of sectoral production.

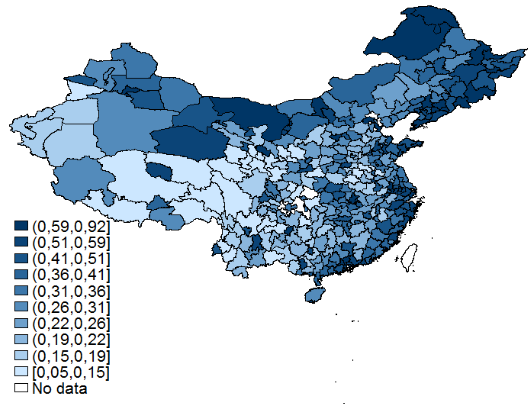
Panel (c) plots the log average wage relative to Beijing, taking into account worker force differences in worker characteristics across cities.[10] The southeastern coast tends to offer higher wages than the interior. (The exceptions are a few cities in the northeast, which are mostly natural resource cities with low population density.) The wage differences across cities are on a magnitude of 30-40%. Panel (d) plots the size of cities. Despite the higher wages in the coast, a large number of people are concentrated in central China, which is consistent with significant migration costs. This initial distribution of population also implies that when a reform reduces restrictions on migration, not all people will flock to the coast cities. Instead, many people in the central will migrate to productive cities nearby.

[9] There are, in total, 34 provinces and 340 cities in China.

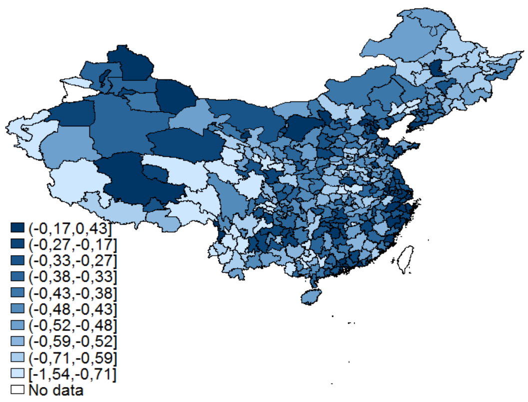
[10] The regional average wages are measured as the regional fixed effects in an individual-level Mincer regression, so worker composition differences are controlled for.



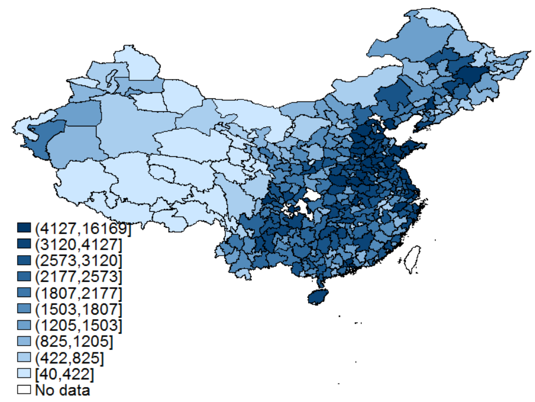
(a) Trade Openness



(b) Share of Urban Employment



(c) Average Wage



(d) City-level Employment

Figure 3.1: City-level Statistics
Source: Author's calculation based on 2005 City Statistics Yearbook (Panel a), 2000 population censuses (b,d), and 2005 population survey (c)

In all figures, there is considerable heterogeneity, even among cities that are geographically close to each other. This motivates a city-level analysis in the quantitative section.

3.3.2 The Hukou System and Reforms

The Hukou system is one of the reasons why migration has been limited in China. Hukou is a household certificate system that ties individuals to locations. Introduced in the 1950's, its original goal was to manage individual mobility and occupation. In the era of a command economy, since most jobs were controlled by the state, foods rationed according to Hukou, the Hukou system could be strictly enforced. The boom in the private economy in the 80's and 90's made enforcement difficult. People started to move to cities for job opportunities. However, without official Hukou, migrants were ineligible for many local public goods, such as health care, schooling and social security. As a result, even though it was possible to find a job in the private sector, the Hukou system still imposes restrictions on migration.

Beginning the mid 1990's, Chinese cities gradually started to reform the Hukou system, allowing qualified people from the rural area and other cities to obtain local Hukou. Experimental in nature, these reforms were initially carried out in a very small number of cities. In 1997, the State Council and the Ministry of Public Security launched a large-scale experiment that relaxed the strict constraint imposed by the central government on provincial and local authorities in terms of what types of Hukou policies were allowed. Under this experiment, each province was *allowed to* select up to 10-20 counties (there are 34 provinces and about 2800 counties in China) to experiment with a relaxed Hukou policy for 2 years. At the end of this "trial period', in 2001, the reform was then scaled up to potentially all counties. Importantly, since the original Hukou policy was a top-down constraint, the reforms were in the form of the central government allowing local government to relax Hukou policies. Substantial freedom was given to provincial and local governments to decide how far they wanted to go in the reform, but they were not allowed to go beyond the framework given by the central government. Indeed, in the official statement, the central government specified that the reforms should be determined based on local economic conditions.

To understand the potential impact of a comprehensive reform on how China responds to international trade, we need to isolate the component of migration costs that are due to the Hukou system. For this purpose, I construct a database with city-level information on Hukou policies to estimate the effect of Hukou reforms on migration. The construction of the database follows recent work by Sun et al. (2011) and Kinnan et al. (2015). I manually searched a list of key words related to Hukou reforms in the most comprehensive online law library in China, and the webpage of the official news agency of the China's Communist Party.[11] Based on the content of the Hukou policies from these two sources, I rate the Hukou openness of each city on a scale of 0-6, with 0 being completely closed, and 6 being the most open. I focus on the period 1997-2010, as 1997 is when the large-scale experimentation first started, and 2010 is the year of the latest census, the best source of reliable information on migration. A detailed account of how I constructed this database is provided in the appendix.

Table 3.1 summarizes the database by time. As Column 1 in the table shows, the average reform index across cities is virtually 0 in 1997. It increases gradually over time, and reaches 3.31 in 2010. The cross-sectional standard deviation of the index starts small—as most cities controlled Hukou strictly in the beginning—and eventually converged to around 1.3. So the average increase in Hukou openness over the period is more than 200% of the standard deviation of the openness across cities.

[11]The address of these two sources are http://www.pkulaw.cn/ and http://www.xinhuanet.com/, respectively.

The third column reports the number of cities that actually experience a change in the reform index in a given year. In most years, 15–50 cities experience a change in the index. In 2001, however, two thirds of cities relaxed their Hukou restriction. This is consistent with the switch from experimentation in selected towns within a province to a comprehensive reform in 2001.

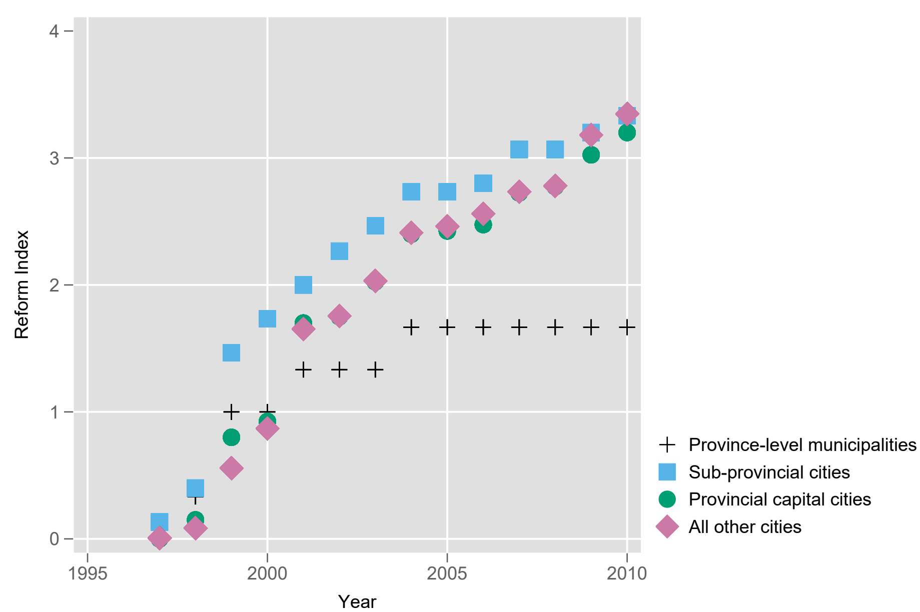
Table 3.1: Summary Statistics of Hukou Reforms

	Reform Index		
Year	Mean	Standard deviation	No. of reforms
1997	0.01	0.15	
1998	0.11	0.52	14
1999	0.63	1.14	73
2000	0.91	1.19	60
2001	1.67	0.9	236
2002	1.77	0.93	33
2003	2.04	1.12	41
2004	2.42	1.24	65
2005	2.46	1.23	12
2006	2.55	1.23	18
2007	2.74	1.27	44
2008	2.78	1.3	6
2009	3.15	1.34	51
2010	3.31	1.45	25

Table 3.1 shows large variation in changes to Hukou policies across cities. Indeed, official statements frequently state the requirement that Hukou policy should depend on the development stage of a city. In the early period of the reform, the concern seemed to be that cities might not be able to provide enough job to all migrants, so more developed cities were encouraged to be more open. In the late 2000's, however, the concern became congestion and pollution in the largest cities. Policy statements from this period emphasized controlling the size of large cities and encouraging more people to move to small cities.

This change in attitude is visible from the database. In Figure 3.2, I plot the average scores for four different types of Chinese cities. The province-level municipalities, which include Beijing, Shanghai, and Chongqing, have the same status as a province in China's political hierarchy. They are also among the largest cities in China. One tier below are the sub-provincial cities, many of which are provincial capitals. They also tend to be the most economically vibrant cities in China. Further below are the remaining provincial capitals, and finally all other cities. Figure 3.2 shows that the provincial municipalities and sub-provincial cities were more open than other cities at the beginning of the sample period. But other cities started to catch up. After the 2000's the province-level municipalities became more restricted than other cities. By the end of the sample period, the sub-provincial cities were also less open than other cities.

Figure 3.2: Hukou Reform in Different Types of Cities



Source: author's calculation.

I merge the Hukou reform panel data with information from the 2000 and 2010 population censuses, the 2005 population survey, and other city characteristics from city statistics yearbooks. Because the population census and survey are only available for these three years, I take the average of the Hukou openness measure over three intervals: 1997-2000, 2001-2005, and 2006-2010.

I first analyze the determinants of reforms using a regression framework. Table 3.2 reports the results in which the Hukou reform index is the dependent variable. In the first column, I include only provincial fixed effects. These fixed effects alone explain about 24% of the variation in the reform index. In the second column, I include time dummies for the three time periods. Consistent with the gradual opening up, the time fixed effects tend to increase over time. Together, provincial and time fixed effects account for around 80% of the variation in the reform index. This reflects the fact that policies changes tend to be correlated across cities within a province. The third column adds city fixed effects. The R squared increases further to 85%. In the fourth column, I add the interaction between the administrative level of a city and time trend. The administrative level of a city ranges from 0 to 3, with the highest value representing province-level cities, and the lowest corresponding to cities that are not provincial capitals. Consistent with the previous narrative, the regression indicates that more important cities became relatively more closed to migrants over time.

The discussion so far supports the claim that the Hukou reforms depend on a city's economic and political status. This gives rise to an endogeneity concern when we estimate the effect of Hukou reforms on migration. The last column of

Table 3.2: Determinants of Hukou Reforms

	(1)	(2)	(3)	(4)	(5)
	Hukou Reform Index				
time=2		1.654***	1.659***	1.684***	0.805***
		(0.042)	(0.045)	(0.047)	(0.185)
time=3		2.489***	2.493***	2.543***	0.800**
		(0.054)	(0.057)	(0.064)	(0.368)
city level=1 × trend				-0.086	-0.129
				(0.087)	(0.087)
city level=2 × trend				-0.192*	-0.400***
				(0.105)	(0.126)
city level=3 × trend				-0.730***	-1.064***
				(0.185)	(0.222)
log per capita GDP					-0.069
					(0.131)
log per capita GDP × trend					0.002
					(0.046)
log population					-0.174
					(0.225)
log population × trend					0.166***
					(0.030)
Constant	1.667***	0.286	-0.467	-0.742*	-0.523
	(0.276)	(0.359)	(0.658)	(0.450)	(0.947)
City Fixed Effects			yes	yes	yes
Provincial Fixed Effects	yes	yes			
Observations	1019	1019	1019	1019	1009
R^2	0.239	0.792	0.850	0.852	0.859

Notes: *city level* indicates the administrative level of a city, with 3 being the highest, and 0 being the lowest. *time* indicates the time period, which takes a value between 1 and 3. Reform index is averaged over the following interval: 1997-2000 (period 1), 2001-2005 (period 2), and 2006-2010 (period 3).

Robust standard errors in parentheses

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

Table 3.3: Migration in the 20000 Census

	Variable	Mean	Median	Std	N
Urban Sector	Share of inter-province migrants	0.11	0.07	0.11	340
	Share of intra-province migrants	0.19	0.17	0.10	340
Rural Sector	Share of inter-province migrants	0.05	0.02	0.10	339
	Share of intra-province migrants	0.05	0.03	0.06	339

Notes: Source: authors' calculation based on the 2000 census. Sample includes all prefecture jurisdictions. Migration is defined based on the difference between the place of residence and the place of birth.

Table 3.2 reassures us that the endogeneity problem might not be too severe, conditional on the set of fixed effects we include. Specifically, I add per-capita GDP and population—two most important characteristics of a city—into the regression. Although some coefficients are different, because of the correlation among variables, the explanatory power of the specification, as measured by the R squared, barely changes. This result suggests that after controlling for city fixed effects, time fixed effects, and differential trend between cities with different administrative levels, Hukou reforms are unlikely to be correlated with other time-varying city characteristics.

3.3.3 Mobility and the Effect of Hukou Reforms

This subsection describes the evidence on migration in China, and assesses the effect of Hukou reforms on mobility.

While the full-fledged Hukou reforms did not start until the late 1990's, migration to many cities began growing in the 1980s as the market economy developed. Indeed, using the 2000 population census, Table 3.3 shows that about a third of the people in the urban sector, and 10% of the people in the rural sector, are migrants.

I estimate the effect of Hukou reforms carried out in the sample period on labor mobility. I focus mainly on two outcome variables: the share of residents with local Hukou, and the number of migrants in a city. To isolate the policy change from time-invariant unobserved regional heterogeneity that can affect migration decisions, I use the following first-difference specification:

$$\triangle y_{i,t} = \beta_1 p_t + \beta_2 c_i + \beta_3 \triangle x_{i,t} + \epsilon_{i,t} \tag{3.1}$$

In the specification, $\triangle y_{i,t}$, is the change in an outcome variable of city i between two consecutive periods, t-1 and t. p_t are the period fixed effects. To capture differential trends by city types, I include city administrative level fixed effects, c_i. $\triangle x_{i,t}$ captures contemporaneous changes in the economic environment in city i.

Table 3.4 reports the effect of Hukou reforms on the share of residents with local Hukou. There are three snapshots for this outcome variable, so we have a two-period panel for specification 3.1. In the first column, only the reform index and time fixed effects are included. The coefficient for the change in the Hukou reform index is statistically significant. The point estimate is 1.1 percentage point. To put this number into perspective, at the time of 2000, the average share of residents in a city without a local Hukou is 6%. A one-point increase in the Hukou reform index therefore decreases the share of residents without local Hukou by 16%. The second column adds city administrative level indicators to allow for differential trends by city types. The coefficients for these indicators suggest that larger and more important cities are becoming stricter over time in granting residents local

Hukou. The inclusion of these variables, however, does not affect the magnitude and significance of the coefficient for Hukou reforms. To further capture the differential trends among cities, the third column adds changes in per-capita GDP, population, and a proxy for local public good, the teacher-to-student ratio in local public primary schools. Reassuringly, none of these variables have a significant impact. Moreover, the coefficient for the Hukou reform index barely changes.

Table 3.4: The Effect of Hukou Reform on Access to Local Hukou

	(1)	(2)	(3)
	$\triangle$ Share of residents with local Hukou		
$\triangle$ Hukou Reform Index	0.011***	0.008**	0.009**
	(0.004)	(0.004)	(0.004)
time=3	0.070***	0.068***	0.065***
	(0.007)	(0.007)	(0.007)
city level=1		-0.021**	-0.021**
		(0.010)	(0.009)
city level=2		-0.072***	-0.071***
		(0.012)	(0.012)
city level=3		-0.122***	-0.120***
		(0.029)	(0.028)
$\triangle$ log per capita GDP			0.012
			(0.011)
$\triangle$ log population			-0.005
			(0.025)
$\triangle$ TeacherStudentRatio			0.000
			(0.001)
Constant	-0.009	0.002	-0.006
	(0.007)	(0.007)	(0.010)
Observations	679	679	672
R^2	0.152	0.207	0.209

Notes: see Table 3.2 for the definition of variables. *TeacherStudentRatio* is the teacher-to-student ratio in local public primary schools.

Robust standard errors in parentheses

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

Table 3.4 provides a direct validation of the reform database by showing that the reforms documented indeed made it easier for migrants to obtain local Hukou. Did the better prospect of obtaining Hukou makes a city more attractive to migrants? Table 3.5 reports the effect of Hukou reforms on inward migration. Columns 1-3 focus on log changes in the number of migrants in a city that moved in during the past year, and Columns 4-6 focus on log changes in the number of migrants that moved in during the past five years. Such information is only available for 2000 and 2005, so after taking the log fist difference, we have only one snapshot.[12] Columns 1 and 4 include only the change in reform index as the independent variable. Coefficients in both columns are positive and statistically significant. According to the estimates, a one-point increase in the reform index increases the one-year migration by 16%, and increases the five-year migration by 21%. To rule out differential trends, I gradually add city administrative level fixed effects and additional control variables, including changes in per-capita GDP, population, and the teacher-to-student ratio in local primary schools. After the inclusion of these variables, the coefficient of Hukou reform loses its significance in predicting one year migration, although the point estimate remains positive and economically meaningful. On the other hand, the coefficient for five year migration remains robust. According to the preferred specification in Column 6, a one-point increase in the Hukou reform index increases the five year migration by 18%. I will use this estimate to back out the implied change in migration costs from a Hukou reform in the quantitative section. Now let me describe the quantitative model.

[12]The 2010 census only publicizes county-level population, which does not include one year and five year migration.

Table 3.5: The Effect of Hukou Reforms on Migration

	(1)	(2)	(3)	(4)	(5)	(6)
	△Log(No. of migrants in the past year)			△Log(No. of migrants in the past 5 years)		
△Hukou Reform Index	0.161*	0.139	0.093	0.209***	0.206***	0.184***
	(0.097)	(0.098)	(0.098)	(0.072)	(0.072)	(0.069)
city level=1		-0.567***	-0.560***		-0.338**	-0.334**
		(0.205)	(0.209)		(0.146)	(0.144)
city level=2		-0.594***	-0.542*		-0.137	-0.076
		(0.152)	(0.314)		(0.162)	(0.162)
city level=3		-0.330**	-0.259		0.333***	0.386***
		(0.149)	(0.689)		(0.095)	(0.102)
△ log per capita GDP			-0.631***			-0.328
			(0.223)			(0.229)
△ log population			0.017			0.366
			(0.331)			(0.282)
△ TeacherStudentRatio			-0.032			-0.040
			(0.038)			(0.030)
Constant	-0.088	0.051	0.519**	-0.347***	-0.299**	-0.091
	(0.180)	(0.190)	(0.242)	(0.133)	(0.141)	(0.208)
Observations	295	295	292	331	331	327
R^2	0.009	0.041	0.085	0.025	0.041	0.088

Notes: Notes: see Table 3.2 for the definition of variables. *TeacherStudentRatio* is the per-student number of teachers in local primary schools.

Robust standard errors in parentheses

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

3.4 Theoretical Framework

3.4.1 The Environment

There are $2N + 1$ regions in the economy. These regions consist of rural and urban sectors of the N Chinese cities, in total $2N$ regions, and one last region that represents the rest of the world (ROW). Denote the set of regions $\mathbf{G}$. I will use $o \in \mathbf{G}$ and $d \in \mathbf{G}$ to refer to the origin and destination of trade and migration flows. I also introduce $\mathbf{R}$ and $\mathbf{U}$ to denote the rural and urban subsets of $\mathbf{G}$: $\mathbf{G} = \mathbf{R} \bigcup \mathbf{U}$. There are four production industries in the economy: agricultural (A), capital and equipment (K), other manufacturing (M), and services (S). The agricultural industry is located in rural regions, and the three other industries are located in urban regions. A, K, and M are tradable; S is non-tradable. In the following, I describe the decisions of workers and firms, and define the equilibrium of the economy.

3.4.2 Workers

There are two types of workers, with different levels of skill. I use e, $e \in \{h, l\}$ to denote the skill level of a worker, where h and l stand for high-skill and low-skill, respectively. A worker's sole source of income is his or her wage, which depends on the wage rate for each labor unit, and the number of labor units a worker possesses—or a worker's productivity—in the local labor market. I assume a worker's productivity in any region is a random draw from a given distribution, to be specified below. The random draw assumption captures in a reduced-form way the match quality between a worker and a region. Workers value both amenities and consumption goods. They choose where to live *within* the country based on regional outcomes—amenities, wages and prices—and individual-level outcomes—their productivity draws across regions.[13] The idiosyncracy of workers' productivity draws allows the model to generate bilateral migration flows, a prominent feature of the data.

3.4.2.1 Preference

Based on a migrant survey, Akay et al. (2012) documents that migrants in China remit on average 10% of their earnings to their hometowns. Remittances

[13] I model migration as driven by idiosyncratic productivity draws and use wage data to discipline the distribution that governs the productivity draws. An alternative is to model migration as driven by idiosyncratic preference shocks.

could potentially lead to trade imbalances. To account for this phenomenon, I assume that, for worker i, born in region o (origin), working in region d (destination), the consumption optimization problem is:

$$\begin{aligned} \max_{C_o,\, C_d} C_{o,d} = (B_o C_o)^{\lambda} (B_d C_d)^{1-\lambda} \\ \text{s.t. } P_o C_o + P_d C_d \leq W_d^e z_d(i), \end{aligned} \tag{3.2}$$

where P_o and P_d are prices for C_o and C_d— the final consumption goods in regions o and d, respectively; B_o and B_d are the amenity values of regions o and d to workers; W_d^e is the nominal wage for each unit of type-e effective labor; and $z_d(i)$ is worker i's productivity in region d. According to this utility function, workers value amenities and consumption in both their hometowns and destinations, with a weight of λ placed on home consumption. Therefore, a λ share of income will be remitted.

Final consumption goods C_o is a bundle of industry final outputs from all four industries:

$$C_o = (C_o^A)^{s_A} (C_o^M)^{s_M} (C_o^S)^{s_S} (C_o^K)^{s_K}, \tag{3.3}$$

where $s_A + s_M + s_S + s_K = 1$. Let P_o^s be the price of $C_o^s, s \in \{A, M, K, S\}$. Then the price of C_o is

$$P_o = \kappa_p (P_o^A)^{s_A} (P_o^K)^{s_K} (P_o^M)^{s_M} (P_o^S)^{s_S}, \tag{3.4}$$

where κ_p is a constant. Worker i's indirect utility from the consumption of goods and amenities, in both origin and destination is

$$C_{o,d}(W_d^e, z_d, P_o, P_d) = \kappa_c B_o^\lambda B_d^{1-\lambda} \frac{W_d^e z_d(i)}{P_o^\lambda P_d^{1-\lambda}}, \tag{3.5}$$

where κ_c is a constant.

3.4.2.2 Migration Decision and Labor Supply Across Regions

Migration is a once-for-life choice. Upon birth, workers learn their draws of productivity in all regions within the country and decide where to work, taking into account their utility from consumption, $C_{o,d}$, and the migration costs they will have to incur. Migration costs, denoted as $d_{o,d}^e$, are both skill-specific and source-destination specific.

Formally, given productivity draws, $\{z_d(i) : d \in \mathbf{G}\}$, worker i chooses the destination d to maximize welfare:

$$U_o(\{W_d^e\}, \{z_d(i)\}, \{P_d\}) = \max_{d \in \mathbf{G}} \{ \frac{C_{o,d}(W_d^e, z_d(i), P_o, P_d)}{d_{o,d}^e} \} = \frac{\kappa_c B_o^\lambda}{P_o^\lambda} \max_{d \in \mathbf{G}} \{ \frac{W_d^e B_d^{1-\lambda} z_d(i)}{P_d^{1-\lambda} d_{o,d}^e} \} \tag{3.6}$$

Notice that migration costs enter the utility function only through $\frac{W_d^e}{d_{od}^e}$. Therefore, it can also be interpreted as how workers discount income from the destination. This cost is similar to the iceberg cost assumption used in international trade literature.[14]

[14] The underlying assumption is that the migration cost for any given origin-destination pair is a fixed share of income, regardless of worker i's income or productivity. In reality, of course, migration costs have both fixed and variable components, and I abstract from the fixed costs for tractability.

For ease of notation, denote $v_d^e = \frac{W_d^e B_d^{1-\lambda}}{P_d^{1-\lambda}}$. Then v_d is the amenity-adjusted real wage *rate* in region d. Worker i will move to region d if and only if this move gives the highest utility:

$$\begin{aligned} \frac{\kappa_c B_o^\lambda}{P_o^\lambda} \frac{v_d^e z_d(i)}{d_{o,d}^e} &\geq \frac{\kappa_c B_o^\lambda}{P_o^\lambda} \frac{v_g^e z_g(i)}{d_{o,g}^e}, \forall g \in \mathbf{G} \\ \Leftrightarrow \frac{v_d^e z_d(i)}{d_{o,d}^e} &\geq \frac{v_g^e z_g(i)}{d_{o,g}^e}, \forall g \in \mathbf{G} \end{aligned} \tag{3.7}$$

Following Ahlfeldt et al. (2015), I assume $\{z_d(i) : d \in \mathbf{G}\}$ are generated from the Frechet distribution. To capture the individual-specific component in workers' productivity, I allow each worker's draws to be correlated across regions. Specifically, the vector of productivity draws for any given worker is generated from the following CDF:

$$F(z_1(i), z_2(i), ...z_d(i)...) = \exp(-(\sum_{d \in \mathbf{G}} z_d(i)^{-\epsilon_e})^{1-\rho}), \tag{3.8}$$

where ρ controls the inter-regional correlation of productivity draws and ϵ_e controls their cross-sectional dispersion.[15] Under this assumption, the probability that a

[15] Hsieh et al. (2013) also uses this parametric assumption to model individuals' comparative advantage in different occupations. I normalize the mean of the productivity distributions to be the same across regions. Differences in regional productivity enter the economy from the production side.

worker from origin o moves to destination d (derived in the appendix), is:

$$\begin{aligned}
\pi^e_{o,d} :&= \mathbf{Pr}(\frac{v^e_d z_d}{d_{o,d}} \geq \frac{v^e_g z_g}{d_{o,g}}, \forall g \in \mathbf{G}) \\
&= \mathbf{Pr}(z_d \geq (\frac{v^e_g d_{o,d}}{v^e_d d_{o,g}}) z_g, \forall g \in \mathbf{G}) \\
&= \frac{(\frac{v^e_d}{d_{o,d}})^{\epsilon_e}}{\sum_{g \in \mathbf{G}} (\frac{v^e_g}{d_{o,g}})^{\epsilon_e}} \\
&= \frac{(\frac{v^e_d}{d_{o,d}})^{\epsilon_e}}{\sum_{g \in \mathbf{G}} (\frac{v^e_g}{d_{o,g}})^{\epsilon_e}}
\end{aligned} \tag{3.9}$$

Let L^e_d denote the *number of workers* with skill level e who are working in d, and let l^e_o denote the number of workers who are born in o. Then $L^e_{o,d} := l^e_o \pi^e_{o,d}$ is the number of workers moving from o to d, and we have the following:

$$L^e_d := \sum_{o \in \mathbf{G}} L^e_{o,d} = \sum_{o \in \mathbf{G}} l^e_o \pi^e_{o,d}. \tag{3.10}$$

L^e_d is different from the supply of *effective labor units* in region d, due to the self-selection on productivity in migration. To derive the supply of effective labor units, I first derive the expected productivity of migrants from region o to region d, denoted $E(z^e_d | L^e_{o,d})$, in two steps.

In the first step, I derive the expected value of the destination-specific component in workers' indirect utility (3.6), u^e_o, for workers moving from o to d, denoted $E(u^e_o | L^e_{o,d})$, where $u^e_o := \max_{d \in \mathbf{G}} \{\frac{W^e_d B^{1-\lambda}_d z_d}{P^{1-\lambda}_d d^e_{o,d}}\} = \max_{d \in \mathbf{G}} \{\frac{v^e_d z_d}{d^e_{o,d}}\}$. I show in the appendix that $E(u^e_o | L_{o,d})$ is given by the following expression:

$$E(u_o^e|L_{o,d}^e) = \Phi_o^{e\frac{1}{\epsilon_e}}\Gamma(1 - \frac{1}{\epsilon_e(1-\rho)}), \tag{3.11}$$

where $\Phi_o^e := \sum_{g\in\mathbf{G}}(\frac{v_g^e}{d_{o,g}^e})^{\epsilon^e}$ measures the welfare of being born in region o. The more connected region o is to other labor markets (smaller $d_{o,g}$), and the more attractive the nearby regions are (higher v_g^e), the higher the utility workers born in region o enjoy.

Notice this expression is independent of d—for workers from the same region, their average utility will be the same regardless of their destination. The intuition is as follows: a destination with higher wages attracts more marginal workers, who obtain lower welfare from the move, pushing down the average utility for the group of workers making the move. This selection along the extensive margin exactly offsets the higher welfare received by the infra-marginal migrants with the same destination, under the parametric assumption of productivity draws. This selection channel is present under more general distributional assumptions, although it might not exactly cancel the effect from infra-marginal migrants.[16]

In the second step, we use $E(u_o^e|L_{o,d}^e)$ to derive $E(z_d^e|L_{o,d}^e)$, the expected pro-

[16] This channel is similar to the selection in trading partner in Eaton and Kortum (2002), in which a country with lower production costs export more marginal goods, and the higher costs of these marginal products offsets the cost advantage of the country. As a result, varieties from different countries have the same average price. See also Hsieh et al. (2013) for the discussion of a similar channel in an occupation-choice context.

ductivity of the workers who move from o to d,

$$
\begin{aligned}
E(z_d^e | L_{o,d}^e) &= E(\frac{u_o^e d_{o,d}^e}{v_d^e} | L_{o,d}^e)(\text{for } L_{o,d}^e,\ u_o^e = \max_{g\in\mathbf{G}}\{\frac{v_g^e z_g}{d_{o,g}^e}\} = \frac{v_d^e z_d}{d_{o,d}}) \\
&= \frac{d_{o,d}^e}{v_d^e} E(u_o^e | L_{o,d}) \\
&= \Phi_o^{e\frac{1}{\epsilon_e}} \Gamma(1 - \frac{1}{\epsilon_e(1-\rho)}) \frac{d_{o,d}^e}{v_d^e}
\end{aligned}
\tag{3.12}
$$

Let $E_{o,d}^e$ denote the total number of effective labor units, brought to d by workers from o. Then $E_{o,d}^e = E(z_d^e|L_{o,d}^e)L_{o,d}^e = E(z_d^e|L_{o,d}^e)l_o^e\pi_{o,d}^e$. Aggregating over migrants from all origins, the total supply of effective labor units in d, E_d^e, is given by

$$
\begin{aligned}
E_d^e &= \sum_{o\in\mathbf{G}} E_{o,d}^e \\
&= \sum_{o\in\mathbf{G}} E(z_d^e|L_{o,d}^e) l_o^e \pi_{o,d}^e.
\end{aligned}
\tag{3.13}
$$

3.4.2.3 The Distribution of Consumption Expenditures

Workers remit a share λ of income to their hometowns for the purchase of consumption goods. The total remittances sent to location o by its out-immigrants in location d is then

$$
\begin{aligned}
R_{o,d}^o &= \sum_{e\in\{h,l\}} L_{o,d}^e E(z_d^e|L_{o,d}^e) W_d^e \lambda \\
&= \sum_{e\in\{h,l\}} E_{o,d}^e W_d^e \lambda
\end{aligned}
\tag{3.14}
$$

The remaining expenditures, $R^d_{o,d}$, are spent in d:

$$R^d_{o,d} = \sum_{e \in \{h,l\}} E^e_{o,d} W^e_d (1 - \lambda). \tag{3.15}$$

The total expenditure on consumption goods in region d is given by

$$R_d = \underbrace{\sum_{o \in \mathbf{G}} R^d_{o,d}}_{\text{Spending from migrants and stayers}} + \underbrace{\sum_{o \in \mathbf{G}} R^d_{d,o}}_{\text{Remittances and spending from stayers}} \tag{3.16}$$

3.4.3 Production and Trade

The production side of the economy is a multi-sector version of Eaton and Kortum (2002), extended to incorporate input-output linkages and capital-skill complementarity.

3.4.3.1 Intermediate Variety Production

Within industry $s, s \in \{A, M, K, S\}$, there is a continuum of intermediate varieties, denoted $\{\omega : \omega \in \mathbf{\Omega}_s\}$. Intermediate varieties are produced using industry final outputs and equipped composite labor, both of which are introduced below. In many developing countries, there is segmentation between rural and urban labor markets (Swiecki, 2017). To capture this, I assume intermediate variety producers in urban industries (industries M, K, and S) are located only in urban regions and hire equipped composite labor from urban labor markets; intermediate variety producers in the agricultural industry are located only in rural regions and hire equipped composite labor from rural labor markets.

The production function for intermediate variety ω, in region d, industry s, is

$$\begin{aligned} & y_d^s(\omega) = t_d^s(\omega) a_d^{s\gamma_s^A}(\omega) m_d^{s\gamma_s^M}(\omega) s_d^{s\gamma_s^S}(\omega) l_d^{s\gamma_s^L}(\omega), \\ & s \in \{A\} \text{ if } d \in \mathbf{R}; \quad s \in \{M, K, S\} \text{ if } d \in \mathbf{U}, \end{aligned} \tag{3.17}$$

where $a_d^s(\omega)$, $m_d^s(\omega)$, and $s_d^s(\omega)$ are the amounts of industry final outputs in agricultural, manufacturing (non-capital), and service industries that are used in production. $l_d^s(\omega)$ is the employment of equipped composite labor. $\gamma_s^{s'}, s, s' \in \{A, M, S, L\}$, are the shares of different inputs in production. $t_d^s(\omega)$ is region d's efficiency in producing variety ω.

Recall that P_d^s is the price of the final outputs of industry s in region d; let W_d be the price for one unit of equipped composite labor in region d. The marginal cost of production is

$$mc_d^s(\omega) = \frac{K_\gamma P_d^{A\gamma_s^A} P_d^{M\gamma_s^M} P_d^{S\gamma_s^S} W_d^{\gamma_s^L}}{t_d^s(\omega)} := \frac{c_d^s}{t_d^s(\omega)}, \tag{3.18}$$

where K_γ is a constant. c_d^s, introduced for ease of notation, is the marginal cost of ω for a producer with unit productivity. The optimal input choice of intermediate variety producers requires

$$\frac{P_o^A a_d^s(\omega)}{y_d^s(\omega) mc_d^s(\omega)} = \gamma_s^A, \frac{P_o^M m_d^s(\omega)}{y_d^s(\omega) mc_d^s(\omega)} = \gamma_s^M, \frac{P_o^S s_d^s(\omega)}{y_d^s(\omega) mc_d^s(\omega)} = \gamma_s^S, \frac{W_d l_d^s(\omega)}{y_d^s(\omega) mc_d^s(\omega)} = \gamma_s^L. \tag{3.19}$$

3.4.3.2 Industry Final Goods Production

In each city, there is a representative final goods producer in each industry. Industry final goods producers combine intermediate varieties of the same industry into final outputs, to be used for final consumption and the production of intermediate varieties. I assume industry final outputs are non-tradable across cities, but freely tradable between the rural and urban regions *within* each city. Therefore, residents and intermediate variety producers in rural and urban regions of the same city have the same access to industry final goods of all sectors, despite their different specializations in intermediate variety production. The production technology for industry s, region d, is the following:

$$Q_d^s = [\int_{\omega \in \Omega_s} q_d^s(\omega)^{\frac{\sigma_s - 1}{\sigma_s}} d\omega]^{\frac{\sigma_s}{\sigma_s - 1}}, s \in \{A\} \text{ if } d \in \mathbf{R}; s \in \{M, K, S\} \text{ if } d \in \mathbf{U}, \tag{3.20}$$

where is $q_d^s(\omega)$ is the quantity of variety ω used.

3.4.3.3 Trade in Intermediate Varieties

Intermediate varieties in A, M, and K industries are tradable, both domestically and internationally; intermediate varieties in the service industry are non-tradable.[17] Final goods producers source the intermediate varieties they use from the cheapest source, taking into account trade costs. I further assume markets for intermediate varieties are competitive, so the producers of intermediate varieties sell their products at marginal costs. Region d's price of intermediate variety ω, produced in another region d', is

$$p^s_{d,d'}(\omega) = mc^s_{d'}(\omega)\tau_{d,d'}, \tag{3.21}$$

where $\tau_{d,d'}$, the iceberg trade cost, is the amount of goods needed to be shipped by producers in d' for one unit to arrive at d. The price for variety ω that a producer in region d actually pays is the cheapest price among all sources:

$$p^s_d(\omega) = \min_{d'}\{p^s_{d,d'}(\omega)\} = \min_{d'}\{\frac{c^s_{d'}}{t^s_{d'}(\omega)}\tau_{d,d'}\}. \tag{3.22}$$

As in Eaton and Kortum (2002), I assume $\{t^s_d(\omega) : \omega \in \Omega_s\}$ are generated from the Frechet distribution with location parameter T^s_d and dispersion parameter θ, with the following CDF:

$$F^s_d(t) = \exp(-\mathrm{T}^s_d t^{-\theta}). \tag{3.23}$$

Under this distribution, among the expenditures spent on intermediate varieties in

[17]In the following, I assume trade costs are infinite for intermediate varieties in the service industry, and proceed as if services were tradable.

region d, industry s, the share allocated to varieties produced in region d' is

$$\delta^s_{d,d'} = \frac{T^s_{d'}(c^s_{d'}\tau_{d,d'})^{-\theta}}{\sum_{d''} T^s_{d''}(c^s_{d''}\tau_{d,d''})^{-\theta}}, \tag{3.24}$$

where the denominator sums over all urban regions if $s \in \{M, K, S\}$, that is, if s indexes an urban industry, and over all rural regions if $s \in \{A\}$.

The parametric assumption on productivity also implies that region d's distribution of prices for intermediate varieties in industry s is

$$F^s_d(p) = 1 - \exp(-\Psi^s_d p^{-\theta}), \tag{3.25}$$

where $\Psi^s_d = \sum_{d'} T^s_{d'}(c^s_{d'}\tau_{d,d'})^{-\theta}$. Again, the summation is taken over urban regions for urban industries, and over rural regions for the agricultural industry. The unit price for industry final goods corresponding to production function (3.20) is

$$\begin{aligned} P^s_d &= [\int_0^\infty {p^s_d}^{1-\sigma_s} dF^s_d(p)]^{\frac{1}{1-\sigma_s}} \\ &= [\Gamma(\frac{\theta + 1 - \sigma_s}{\theta})]^{\frac{1}{1-\sigma_s}} (\Psi^s_d)^{-\frac{1}{\theta}}. \end{aligned} \tag{3.26}$$

3.4.3.4 Equipped Composite Labor Production

Equipped composite labor is produced by a representative producer in each region, from capital and two types of labor units. I incorporate capital-skill complementarity by specifying the production function of equipped composite labor in a nested CES form, with capital being complementary to high-skill labor, and substitutable to low-skill labor.[18]

Formally, effective high-skill labor units, E_d^h, low-skill labor units, E_d^l, and capital and equipment, K_d, are combined into *equipped composite labor*, E_d, through the following technology:

$$E_d^{eh} = [(1-\eta_d^h)^{\frac{1}{\rho_{kh}}}(K_d)^{\frac{\rho_{kh}-1}{\rho_{kh}}} + (\eta_d^h)^{\frac{1}{\rho_{kh}}}(E_d^h)^{\frac{\rho_{kh}-1}{\rho_{kh}}}]^{\frac{\rho_{kh}}{\rho_{kh}-1}} \tag{3.27}$$

$$E_d = [(1-\eta_d^l)^{\frac{1}{\rho_{lkh}}}(E_d^l)^{\frac{\rho_{lkh}-1}{\rho_{lkh}}} + (\eta_d^l)^{\frac{1}{\rho_{lkh}}}(E_d^{eh})^{\frac{\rho_{lkh}-1}{\rho_{lkh}}}]^{\frac{\rho_{lkh}}{\rho_{lkh}-1}}, \tag{3.28}$$

where E_d^{eh} is *equipped high-skill labor*, the output from the inner nest. ρ_{kh} $(\rho_{kh} < 1)$ is the elasticity of substitution between high-skill labor and capital, and ρ_{lkh} $(\rho_{lkh} > 1)$ is the elasticity of substitution between equipped high-skill labor and low-skill labor. η_d^h and η_d^l determine the region-specific shares of different factors in equipped composite labor.

Let W_d^h/W_d^l be the wage rate for high-/low-skill labor, W_d^{eh} the unit price for equipped high-skill labor, and W_d the unit price for equipped composite labor. The optimization decision and the zero-profit conditions of equipped composite labor production imply the following:

[18] This formulation has a tradition in macroeconomics (see, for example Krusell et al., 2000), and has recently been adapted to the international setting to examine impacts of globalization on wage inequality by Burstein et al. (2013) and Parro (2013).

$$
\begin{aligned}
W_d^{eh} &= [(1-\eta_d^h)(P_d^K)^{1-\rho_{kh}} + (\eta_d^h)(W_d^h)^{1-\rho_{kh}}]^{\frac{1}{1-\rho_{kh}}} \\
W_d &= [(1-\eta_d^l)(W_d^l)^{1-\rho_{lkh}} + (\eta_d^l)(W_d^{eh})^{1-\rho_{lkh}}]^{\frac{1}{1-\rho_{lkh}}}
\end{aligned}
\tag{3.29}
$$

$$
\begin{aligned}
\frac{P_d^K K_d}{W_d^h E_d^h} &= (\frac{P_d^K}{W_d^h})^{1-\rho_{kh}} \frac{1-\eta_d^h}{\eta_d^h} \\
\frac{W_d^{eh} E_d^{eh}}{W_d^l E_d^l} &= (\frac{W_d^{eh}}{W_d^l})^{1-\rho_{lkh}} \frac{1-\eta_d^l}{\eta_d^l}
\end{aligned}
\tag{3.30}
$$

Equation (3.30) expresses the ratios between the shares of different factors in equipped composite labor as functions of relative prices and technological parameters, η_d^h and η_d^l. Factors' shares vary by regions and prices. Nonetheless, to simplify notation, I use β_d^K, β_d^h, and β_d^l to denote the shares of capital, high-, and low-skill labor in equipped composite labor in region d:

$$
\beta_d^K + \beta_d^h + \beta_d^l = 1. \tag{3.31}
$$

3.4.3.5 Goods and Labor Markets Clearing Conditions

Let X_d^s be region d's production of final output in industry s. Since final output producers add no value in converting intermediate varieties into industry final outputs, X_d^s is also the value of their input demand for intermediate varieties. The total demand for the intermediate varieties in industry s , produced in region

d, is:

$$D_d^s = \sum_{d' \in \{\mathbf{U}\}} X_{d'}^s \delta_{d',d}^s, s \in \{M, K, S\}$$
$$D_d^s = \sum_{d' \in \{\mathbf{R}\}} X_{d'}^s \delta_{d',d}^s, s \in \{A\}. \tag{3.32}$$

To produce D_d^s amount of intermediate varieties in industry s, the producers in region d use, respectively, $D_d^s \gamma_s^A, D_d^s \gamma_s^M,$ and $D_d^s \gamma_s^S$ amounts of the industry final outputs of agricultural, manufacturing, and service industries. The producers also employ $D_d^s \gamma_s^L$ worth of equipped composite labor, whose income will be distributed to capital and workers. The labor market clearing conditions, which are different for rural and urban labor markets in each city, are

$$\text{Rural (d}\in \mathbf{R}\text{): } E_d^h W_d^h = D_d^A \gamma_A^L \beta_d^h; \quad E_d^l W_d^l = D_d^A \gamma_A^L \beta_d^l,$$
$$\text{Urban (d}\in \mathbf{U}\text{): } E_d^h W_d^h = \beta_d^h \sum_{s \in \{M,K,S\}} D_d^s \gamma_s^L; \quad E_d^l W_d^l = \beta_d^l \sum_{s \in \{M,K,S\}} D_d^s \gamma_s^L. \tag{3.33}$$

The demand for industry final outputs in each region comprises demand from residents and intermediate variety producers. Since residents and producers in both the rural region and the urban region of a city purchase industry final outputs from the same representative producer in that city, to express market clearing conditions for industry final goods, I use d to denote an urban region, and d' to denote the rural region of the same city. The market clearing conditions for industry final outputs are:

$$
\begin{aligned}
X_{d'}^{A} &= (C_d^A + C_{d'}^A) + D_{d'}^A \gamma_A^A + \sum_{s\in\{M,K,S\}} D_d^s \gamma_s^A, \quad d' \in \mathbf{R} \\
X_{d}^{M} &= (C_d^M + C_{d'}^M) + D_{d'}^A \gamma_A^M + \sum_{s\in\{M,K,S\}} D_d^s \gamma_s^M, \quad d \in \mathbf{U} \\
X_{d}^{S} &= (C_d^S + C_{d'}^S) + D_{d'}^A \gamma_A^S + \sum_{s\in\{M,K,S\}} D_d^s \gamma_s^S, \quad d \in \mathbf{U} \\
X_{d}^{K} &= (C_d^K + C_{d'}^K) + D_{d'}^A \gamma_A^L \beta_{d'}^K + \sum_{s\in\{M,K,S\}} D_d^s \gamma_s^L \beta_d^K. \quad d \in \mathbf{U}
\end{aligned}
\tag{3.34}
$$

In Equation (3.34), the left side is the total supply of industry final outputs in the city; on the right side, $D_{d'}^A \gamma_A^{s'}$ and $\sum_{s\in\{M,K,S\}} D_d^s \gamma_s^{s'}$ are the demands from intermediate variety producers in the agricultural industry and the three urban industries, respectively; $C_d^s + C_{d'}^s$ is the sum of consumption demands in rural and urban regions of the city. The consumption demand term is calculated as $s_s[R_d + R_{d'} - (S_d + S_{d'})]$, where R_d is region d's aggregate income, remittances included; $S_d + S_{d'}$ is the city's international trade surplus taken as exogenous from the data, scaled to the model economy;[19] and s_s is the share of industry s in the final consumption bundle. Adjusting for trade surpluses ensures that the calibration of regional productivity takes into account the international trade imbalances, about 5% of the GDP of China in 2005. After calibration, however, in all counterfactual experiments, I focus on the competitive equilibrium defined below, without international trade imbalances (but allowing for intra-national imbalances arising from remittances).

[19] I provide details on the construction of city-level surpluses in the appendix.

3.4.4 Definition of Equilibrium

The parameters in the economy are the following: preference parameters, including $\{\sigma_A, \sigma_M, \sigma_K\}$ $\{s_A, s_M, s_K, s_S\}$, and λ; spatial frictions, including migration costs $\{d_{o,d}\}$ and trade costs $\{\tau_{o,d}\}$; production technology, including $\{\gamma_s^{s'}\}$, $\{\eta_d\}$, $\{\rho_{ks}, \rho_{lks}\}$, and θ; local productivity and amenities, $\{T_d^s\}$ and $\{B_d^e\}$; and initial labor endowments in each region, $\{l_o^e\}$.

Definition 3 *A competitive equilibrium of the economy is defined as a set of prices and allocations that satisfy the following conditions:*

1. *Workers' migration decisions are optimal, that is, Equation (3.7) is satisfied. In aggregate, this implies Equation (3.9).*
2. *The distribution of effective labor units E_d^e, and final consumption expenditures, R_d, are consistent with workers' migration choices—Equations (3.13) and (3.16).*
3. *The decisions of intermediate variety producers are optimal—Equations (3.18) and (3.19).*
4. *The decisions of composite labor producers are optimal—Equations (3.29) and (3.30).*
5. *Industry final goods producers' production and sourcing decisions are optimal—Equations (3.24) and (3.26).*
6. *Workers' consumption decisions are optimal.*
7. *Labor markets and goods markets clear—Equations (3.32)-(3.34).*

The definition of the equilibrium also highlights the key departure from the existing applications of similar quantitative trade models in cross-country settings: exogenous labor supply is replaced with migration decisions, summarized by Equation (3.9), so the distribution of labor across regions is endogenous.

3.5 Parameterization

Before conducting counterfactual experiments, I calibrate the model to data from the Chinese economy in 2005.[20] This section explains how I determine the parameters in the model, starting with data sources.

3.5.1 Data Descriptions

Quantifying the model primarily requires the following information: to calibrate regional productivity, we need, by skill level, the average wage in each region, and the employment for each city-industry pair; to calibrate region-specific parameters in equipped composite labor production function, we need the shares of different factors in equipped composite labor; to estimate domestic migration costs we need migration flows; to estimate trade costs we need information on domestic trade flows; finally we need the measures of geographic and cultural distances between regions. This section describes briefly the sources of data; the appendix provides more details.

[20] An alternative is to solve the model in changes, as in Dekle et al. (2008). This alternative is infeasible because the level of aggregation in migration and trade are different, between the model and the data. An additional advantage of calibrating the benchmark equilibrium rather than solving the model in changes is that we can assess the fit of the model by looking at moments that are not direct targets of calibration.

I use the 2005 mini population census to estimate the wage rates for Chinese regions. I estimate the average wage for unskilled workers and the skill premium in each region as the regional fixed effects and the region-specific skill dummies, in an individual wage regression that controls for a rich set of individual demographic and occupation variables. This regression approach nets out the differences in demographics and detailed industry structures across regions, which are not explicitly modeled. The specification and the results of this regression are reported in the appendix.

I also use the 2005 mini census to construct the *number of workers* employed in each city-industry. Once we have the estimates for migration costs and regional amenity-adjusted real wages, we can use Equation (3.13) to convert the number of workers into the employment of *effective labor units*. Combining this with the regional wages estimated above, I obtain the total wage bill for high- and low-skill workers at the city-industry level.[21]

Using the data described above, we can readily compute the relative shares of wage payments to high- and low-skill workers. Determining η_d^h and η_d^l, the region-specific parameters in the equipped composite labor production function, further requires the relative shares between capital and equipment (K) and labor. For the urban sector, I use the 2004 Annual Survey of Industrial Production to construct

[21] We run into a small sample problem and end up with zeros for the employment capital and equipment industry in some cities, as my sample is only a 1% sub-sample of the mini-census. To overcome this problem, I tabulate employments, differentiating only between agricultural and urban industries, even though the data contains employment by two-digit industry. (As a result, we do not know the distribution of employments across the three urban industries in each city.) I supplement this information with the ratio of employment in industry K over industry M, constructed from the manufacturing sub-sample of the 2004 economic census, to obtain the employment information at the city-industry level. I provide more details in the appendix.

wage bill and capital expenditures for each city, which I combine with the relative shares of skilled over unskilled workers, to obtain the shares for all three inputs into equipped composite labor;[22] for the rural sector, due to the lack of regional data, I assume all cities have the same capital/labor share, and determine this share using the national input-output table.

To construct a database of inter-regional and inter-sectoral migration, I use the 2000 population census. The 2000 census serves the purpose best because it reports birthplace information, which is essential for the definition of migration, a lifetime decision in the theoretical framework.[23] For each worker, I identify his or her skill level, current city, birth province, type of Hukou, and whether he or she is currently working in a rural or urban industry, and then determine his or her migration status based on this information.[24]

I construct proxies for geographic distance and cultural distance between Chinese cities. For any two cities, their geographic distance is calculated as the greater-circle distance between the coordinates of their city centers, proxied by the locations of their local governments, extracted from Google Maps. The cultural distance is

[22] This is an annual firm-level survey, containing detailed financial information for all state-owned enterprises, as well as private firms with sales over 5 million RMB yuan, in the industrial sector. I aggregate firm-level expenditures on capital and equipment and labor to compute the city-specific labor share in equipped composite labor.

[23] The 2005 mini population census, on the other hand, reports only migration information during the past 5 years and, therefore, is inconsistent with the notation of long-term migration adopted here. We cannot combine 5-year migration with long-term migration constructed from the 2000 census, because of the possibility of repeat migrants or return migrants.

[24] Hukou is the household registration system in China, which records the place of legal residence and the sector of origin for Chinese residents; the information on birth place is only up to the provincial level in the census, so I tabulate only the source province for migrants. The census does record the source city of the most recent migration move for each individual. That city, however, is not necessarily the same as an individual's birth city, as he or she may be a repeat migrant. In the appendix, I provide additional background information on the Hukou system and a discussion of the drawbacks of alternative ways of constructing migration flows.

constructed as $1 - \text{corr}(V_o, V_d)$, where V_o is a vector, the elements of which are the shares of various ethnic groups in the total ethnic minority population in o in the 1990 census. The cultural distance between two cities is small if two cities had similar compositions of ethnic minorities in the 1990s.[25]

Finally, I use the 2002 inter-regional input-output table of China to construct trade flows between Chinese provinces, which are then used for the estimation of domestic trade costs.

3.5.2 Parameters Calibrated Independently

I calibrate the following parameters independently. The dispersion parameter ϵ_e governs the variance of the idiosyncratic component of workers' productivity draws. The parametric assumption in Equation (3.8) implies that, the wage distribution of workers *staying* in their hometown follows a Frechet distribution with dispersion parameter $(1-\rho)\epsilon_e$ (proved in the appendix). A property of the Frechet distribution is that its coefficient of variation satisfies the following relationship:

$$\frac{\text{Variance}}{\text{Mean}^2} = \frac{\Gamma(1 - \frac{2}{\epsilon_e(1-\rho)})}{(\Gamma(1 - \frac{1}{\epsilon_e(1-\rho)}))^2} - 1. \tag{3.35}$$

Guided by this relationship, I use the wage distribution of *stayers* to recover $\epsilon_e(1-\rho)$.[26] Specifically, I regress the log wage of stayers on regional fixed effects,

[25] Migrations were less common prior to 1990; therefore correlation constructed this way captures the historical cultural distance between regions, and is unlikely to be driven by current migration. I provide background information on ethnicity in China and the summary statistics of cultural distance in the appendix.

[26] Hsieh et al. (2013) also uses this relationship to recover the ability dispersion and follows a similar strategy, described below, to calibrate ρ in an occupation-choice model.

individual demographics, and industry fixed effects, for high- and low-skill worker samples separately. I then take the exponents of the residuals, compute their coefficients of variations, and choose $\epsilon_e(1-\rho)$ so that Equation (3.35) gives the same value. This procedure determines $\epsilon^{\mathrm{h}}(1-\rho) = 2.72$ and $\epsilon^{\mathrm{l}}(1-\rho) = 2.88$. By deriving statistics for only stayers' wage distribution, and matching them to their data counterparts, this procedure takes into account the self-selection on productivity in migration.[27]

The parameter ρ controls the correlation of individuals' productivity draws across regions. My strategy for calibrating it is first to compute the explanatory power of individual fixed effects in an individual-panel wage regression using real data. Then I choose ρ so that, in the simulated data, individual fixed effects have the same level of explanatory power. This procedure determines ρ to be 0.4. In the appendix I provide more details on this procedure.

Productivity dispersion in intermediate varieties, θ, is not separately identifiable from trade costs using the data I have. I assign a value of 4, the preferred estimate of Simonovska and Waugh (2014), to the productivity dispersion for A, M, and K industries.[28] The elasticities of substitution between high-skill labor and capital, and between low-skill labor and equipped high-skill labor, are set to the

[27] We can also use migrants for this calibration. In that case, the model predicts that, only for migrants sharing the same origin *and* destination, will the wage distribution follow a Frechet distribution. This approach is infeasible for two reasons: first, in the data, we identify the source region only up to the provincial level, and second, for many origin-destination pairs, there are only a few workers.

[28] Simonovska and Waugh (2014) focuses on aggregate trade flows. Papers focusing on agricultural trade alone, for example, Donaldson (2017) and Sotelo (2014), report similar estimates for the elasticity of trade. In this model, trade is driven by Ricardian comparative advantage, so the love-for-variety parameters do not have impacts on the levels and elasticities of trade; they are used solely for computing the aggregate price indices, and the only requirement is that $\theta - \sigma > 1$, so that the price indices are integrable.

estimates in Krusell et al. (2000)—0.67 and 1.67, respectively.[29] These values imply that capital and high-skill labor are complements, and both are substitutes to low-skill labor.

The share of remittances in migrants' income is calibrated to 10%, following Akay et al. (2012). The shares of different industries in the final consumption bundle, $\{s_A, s_M, s_K, s_S\}$, are calibrated to the shares of these industries in final consumption. The calibration determines $s_A = 0.23, s_M = 0.24, s_K = 0.01, s_S = 0.52$. The shares of different inputs in intermediate variety production, $\{\gamma_s^{s'}\}$, are calibrated to the 2002 national input-output table.

The upper panel of Table (3.6) summarizes the sources and values of these parameters. The lower panel provides information on other parameters, which I discuss in the rest of this section.

[29] Parro (2013) and Burstein et al. (2013) use the same parameter values when examining the roles of skill-biased technological change and globalization in explaining the rise in the skill premium.

Table 3.6: Model Parameterization

A: Parameters Calibrated Independently			
Parameter	Description	Target/Source	Value
ρ	Correlation in worker productivity draws	Idiosyncratic component of individual wage	0.4
ϵ^h,ϵ^l	Dispersion in worker productivity draws	Equation (3.35)	$\epsilon^h = \frac{2.73}{1-\rho}$, $\epsilon^l = \frac{2.88}{1-\rho}$
θ	Elasticity of trade	Simonovska and Waugh (2014)	4
λ	Share of remittances	Akay et al. (2012)	0.1
ρ_{kh}, ρ_{lkh}	Elasticities in equipped composite labor	Krusell et al. (2000)	$\rho_{kh} = 0.67$, $\rho_{lkh} = 1.67$
s_A, s_M, s_S, s_K	Sectoral shares in final consumption	Aggregate consumption in the economy	$s_A = 0.22, s_M = 0.24$ $s_S = 0.52, s_K = 0.01$
$\gamma_s^{s'}$	Input-output linkages	National input-output table	See the appendix
B: Parameters Estimates/Calibrated in Equilibrium			
Parameter	Description	Target/Source	Value
$\{d_{o,d}\}$	Migration Costs	Migration Flow	See Table(3.7)
$\{\tau_{o,d}\}$	Domestic Trade Costs	Domestic Trade Flow	See Table(3.8)
t_a, t_m, t_k	International Trade Costs	International Openness	See Table(3.8)
$\{\eta_d^h\}$, $\{\eta_d^e\}$	Factor weights in equipped composite labor	Corresponding factor shares in the data	-

3.5.3 Migration Cost Estimation

3.5.3.1 The Specification of Migration Cost

I specify the cost of a migration from o to d as

$$\ln(d^e_{o,d}) = \beta^e_1 I_1 + \beta^e_2 * I_1 * \text{dist}_{o,d} + \beta^e_3 I_2 + \beta^e_4 * I_2 * \text{dist}_{o,d} + \beta^e_5 I_3 + \beta^e_6 * I_3 * \text{dist}_{o,d} + \beta^e_7 * I_4 + \beta^e_8 * \text{Cdist}_{o,d}, \tag{3.36}$$

where I_1-I_4 are mutually exclusive dummy variables: I_1 indicates if o and d belong to different cities within the same province; I_2 indicates if o and d belong to different provinces within the same large region (of which there are seven in China, each containing five provinces on average); I_3 indicates if o and d belong to different large regions; and I_4 is the indicator for rural-urban migration. These dummy variables capture different kinds of institutional barriers to the free mobility of labor. $\text{dist}_{o,d}$ is the geographic (great-circle) distance between o and d, while $Cdist_{o,d}$ is the cultural distance: these two variables capture the geographic and cultural barriers to migration.

3.5.3.2 Estimation Strategy

If migration flows are recorded at the city-to-city level in the data, both migration cost parameters, $\{\beta\}$, and amenity-adjusted real wages, $\{v^e_d\}$, can be estimated in linear regression. In this section, I use this simpler case to illustrate the source of identification; as discussed in the data section, however, the migration data is recorded at province-to-city level, so in actual implementation, I use non-linear least squares and estimate the parameters by minimizing the distance between the data and the model-predicted province-to-city flows. I provide the details on estimation in the appendix.

From Equation (3.9), if we divide $\pi_{o,d}$ by $\pi_{o,o}$, the resulting equation is:

$$\ln(\frac{\pi_{o,d}}{\pi_{o,o}}) = \underbrace{\epsilon^e}_{\text{Ability Dispersion}} [\underbrace{\ln(v_d^e) - \ln(v_o^e)}_{\text{Fixed Effects}} - \underbrace{\ln d_{o,o}}_{0} + \underbrace{\ln d_{o,d}}_{\text{Migration Cost}}] \tag{3.37}$$

We can then substitute Equation (3.36) into the migration costs component in the expression, and estimate this specification using linear regression. The specification demonstrates clearly that the dispersion of workers' productivity, ϵ^e, is not separately identifiable from migration costs, and therefore I calibrate it using information on wage dispersion. Parameters governing migration costs, $\{\beta\}$, are identified from within variation; if, within a region, the majority of workers are from regions that are far away, the estimated migration costs will be small. The logs of the amenity-adjusted real wages, v_d^e and v_o^e, are identified as destination and origin fixed effects; intuitively, if a region employees a larger number of workers (relative to the number of workers born in the region), it either pays a good wage, or offers attractive amenities. Once we calibrate the remaining parameters in the benchmark economy and solve the model, we can back out amenities, B_d^e, by subtracting wages and prices from v_d^e.[30]

[30]Note that I use the 2000 data to estimate the long-run migration costs; the wage and employment data, on the other hand, represents the 2005 economy. To ensure that the recovered $\{v_d^e\}$ are consistent with the 2005 employment distribution, after estimating $\{\beta\}$, I use workers' birthplace and employment distribution in 2005, and solve the migration model again. Specifically, I solve

3.5.4 Calibrating the Rest of the World

With the estimated migration costs and $\{v_d^e\}$, Equation (3.13) predicts the supply of effective labor in each place. Together with the regional wage estimated before, I compute the regional and, in turn, the national labor value added in China. I then use the share of Chinese value added in the world, calculated from Penn World Table 6.1, to determine the GDP of the ROW:

$$\text{GDP}_{\text{ROW}} = \text{GDP}_{\text{China}} * \frac{\text{Data GDP}_{\text{ROW}}}{\text{Data GDP}_{\text{China}}} \tag{3.38}$$

To calibrate the total number of effective labor units available in the ROW, I assume that $E_{\text{ROW}} = E_{\text{China}} \frac{\text{Population}_{\text{ROW}}}{\text{Population}_{\text{China}}}$, where E stands for number of effective labor units.[31] The wage for each effective labor unit in the ROW is then $\frac{\text{GDP}_{\text{ROW}}}{E_{\text{ROW}}}$.

for $\{v_d^e\}$, so that the model-predicted total number of workers in each region is the same as that in data for the year of 2005, i.e., $\sum_{o\in\mathbf{G}} l_o^e \pi_{o,d}^e = \sum_{p\in P} L_{p,d}^e$, where $L_{p,d}^e$ (data) is numbers of workers working in region d; l_o^e (data) is the number of workers born in region o; and $\pi_{o,d}^e$ is the model-predicted probability of migration, as functions of $\{v_d^e\}$. The workers employment distribution in 2005 comes from the 2005 mini census directly. Because the 2005 mini census does not provide birthplace information, I construct the birthplace distribution using the 2000 census, focusing on appropriately adjusted age groups. The appendix provides more details. One additional benefit of updating $\{v_d^e\}$ is that, even if there are changes in migration costs during the period 2000–2005, our benchmark calibration still ensures that the number of workers in each place is the same as that in the data.

[31] I assume that in the ROW, intermediate varieties are produced using industry final outputs and effective labor units directly, without equipped composite labor, so there is neither the distinction between skilled and unskilled workers nor capital-skill complementarity.

3.5.5 Joint Estimation of Trade Costs and Regional Productivity

3.5.5.1 The Specification of Trade Cost

Following the gravity literature in international trade, I specify the trade costs between any two regions *within* China as a log linear function of the geographic, institutional and cultural distance between them:

$$\log(\tau_{o,d}) = \gamma_1 I_1' + \gamma_2 * (1 - I_2' - I_3') * \text{dist}_{o,d} + \gamma_3 I_2' + \gamma_4 * I_2' * \text{dist}_{o,d} + \gamma_5 I_3' + \gamma_6 * I_3' * \text{dist}_{o,d} + \gamma_7 * I_4' + \gamma_8 * \text{Cdist}_{o,d} \tag{3.39}$$

Dummy variables I_1'-I_3' in this specification are the same as I_1-I_3 in the migration cost specification.[32] I_4' is an indicator for common provincial border. $\text{Cdist}_{o,d}$ and $\text{dist}_{o,d}$ are also defined in the same way as in the migration cost specification, except that here I allow for positive trade costs within the same city—$\text{dist}_{o,o} > 0$. I proxy within-city distance, $\text{dist}_{o,o}$, using city o's radius, constructed as half of city o's average distance to its five closest neighboring cities.

I further specify the trade cost between a given Chinese city and the ROW as the trade cost between that Chinese city and its nearest port city, plus a parameter for international trade cost that captures tariffs, non-tariff barriers, and information flow costs, among other barriers to trade. These costs likely differ across industries, so I allow international trade costs to be industry-specific, too. The international

[32] Under this specification, the marginal impact of within-city distance on trade cost is γ_2, the same as that of between-city, within-province distance. I also try treating these two variables separately, and it seems the data does not allow for simultaneous identification of these two variables.

trade costs will be calibrated to match industry trade openness in China in 2005. The targets for industry openness are reported in Panel B of Table (3.8).

3.5.5.2 Estimation Strategy

Similar to the migration cost estimation, if we have city-pair trade flow data, we can estimate the specification using a linear regression. For the clarity of exposition, in this section, I focus on this simple setup to illustrate the sources of identification in estimation. I provide more details on the actual computational algorithm in the online appendix, in which, to accommodate the aggregate nature of the trade data, I jointly determine trade costs and region-specific productivity, by solving the full model and choosing the parameters, so that the distance between the model-predicted trade flows and the data is minimized.

From Equation (3.24) we can derive the following equation:

$$\ln(\frac{\delta^j_{d,o}}{\delta^j_{d,d}}) = \underbrace{-\theta}_{\text{Trade Elasticity}} [\underbrace{(\frac{1}{\theta}\ln(T^j_d) - \ln(c^j_d)) - (\frac{1}{\theta}\ln(T^j_o) - \ln(c^j_o))}_{\text{Fixed Effects}} - \underbrace{\ln\tau_{o,o}}_{\text{Trade Cost (within city)}} + \underbrace{\ln\tau_{o,d}}_{\text{Trade Cost}}]. \tag{3.40}$$

We can then obtain the estimation specification by substituting Equation (3.39) into trade cost in this equation. Applying this specification to domestic trade flow data, we can estimate $\ln(\frac{T_d^{j\,\frac{1}{\theta}}}{c^j_d})$, the cost-adjusted productivity, with origin and destination fixed effects. Intuitively, if region d purchases a lot from local producers (large $\delta^j_{d,d}$), region d is either very productive in converting input bundle into output, or it has access to cheap input bundles, both of which imply large $\ln(\frac{T_d^{j\,\frac{1}{\theta}}}{c_d^j})$. We can also recover $\{\gamma\}$, the coefficients determining domestic trade costs, where the source of identification comes from the extent to which region d sources from regions that are at different distances.[33]

Once we have the estimates, $\widehat{\ln(\frac{T_d^{j\,\frac{1}{\theta}}}{c_d^j})}$, we can use the model structure to separate T_d^j and c_d^j and to calibrate the industry-level international trade costs. Note that c_d^j depends on local wages and the prices of industry final outputs, which in turn depend on productivity and wages in all regions in the economy. Therefore, fixing wages at the observed value, for any given level of industry trade barriers, we can choose a distribution of $\{T_d^j\}$ such that, given wages and productivity across regions, the equilibrium distribution of costs $\{c_d^j\}$ from the trade model satisfy that $\ln(\frac{T_d^{j\,\frac{1}{\theta}}}{c_d^j}) = \widehat{\ln(\frac{T_d^{j\,\frac{1}{\theta}}}{c_d^j})}$. We can then determine the international trade costs so that the model exhibits the same level of international trade openness as in the data.

3.5.6 Estimation Results

3.5.6.1 Migration Costs

Table (3.7) reports the estimates for the migration costs. The model fits the data well, as indicated by the high R^2s. The signs of coefficients are as expected:

[33] Since I actually implement the estimation using provincial-level trade data, the identification of the two variables that vary only within a province, the inter-city dummy and the within-provincial geographic distance, relies on differences in trade patterns for provinces with different internal structures. Intuitively, the larger these coefficients are, the higher the within-province trade costs are and the more intensive the inter-provincial trade is. Suppose a province trades intensively with itself. All else equal, if this province has many cities, then its low trade-penetration rate will be reflected in a large estimate for the intercity dummy; if, on the other hand, this province has only a small number of (geographically) large cities, then the low trade-penetration rate will be reflected in a large estimate for the coefficient for within-province distance.

all measures of distance increase migration costs. In terms of magnitude, the cost of migrating to other cities within the same province is around 117 log points for both types of workers. As a migration move covers more distance, it incurs a larger cost: for skilled workers, the additional cost of crossing a provincial border is about 30 log points, and the additional cost of crossing a regional border is another 20 log points; these costs are slightly higher for unskilled workers.

Table 3.7: Estimates of Migration Costs

	Skilled Workers	Unskilled Workers
I(Different Cities, Same Province)	1.167	1.192
	(0.0492)	(0.043)
I(Different Provinces, Same Region)	1.502	1.555
	(0.0318)	(0.03)
I(Different Regions)	1.719	1.812
	(0.0275)	(0.0305)
I(Rural to Urban)	0.586	0.606
	(0.0191)	(0.0172)
I(Different Cities, Same Province)*Distance	0.378	0.332
	(0.2761)	(0.1984)
I(Different Provinces, Same Area)*Distance	0.367	0.738
	(0.0578)	(0.048)
I(Different Regions)* Distance	0.215	0.539
	(0.0225)	(0.0292)
Cultural Distance	0.141	0.148
	(0.0312)	(0.0359)
Observations	42160	42160
R^2	0.92	0.67

Notes: This table reports the estimates of domestic migration costs. Robust standard errors are in parentheses. *Distance* is measured as the great circle distance between cities (in 1000 km); *Cultural Distance* is measured as one minus the correlation in lagged ethnic minority shares between cities.

The continuous components of geographic distance have nonlinear effects on migration costs: when the origin and the destination are in the same province, the marginal cost of moving an extra 1000 kilometers is not significant; when the origin and destination are in different provinces within the same large region, the marginal cost is sizable and statistically significant; when the origin and destination are in different large regions, the marginal cost becomes somewhat smaller, but is still significant. This pattern holds for both types of workers, but the coefficients are much larger for unskilled workers.

The estimation also reveals substantial costs, about 60 log points, associated with rural-urban migration. This magnitude, however, is only about one-third of the calibrated "labor wedge" for China in Swiecki (2017). The difference underscores the importance of accounting for the geographic dimension: a large proportion of the measured rural-urban wedge could be a joint product of regional inequality and spatial frictions.

Finally, for both types of workers, the coefficients for cultural distance are positive and significant. The standard deviation of cultural distance is 0.3, so increasing cultural distance by one standard deviation leads to an increase of around 5 log points in migration costs.

It is instructive to compare my estimates to those based on the U.S. data. While no existing papers use the exact same specification to this paper, some recent studies estimate a spatial equilibrium model, in which workers choose where to work, taking into account real wages, amenities, and migration costs. For example, Diamond (2016) specifies a discrete choice model for workers' migration decision, and estimates the structural parameters governing migration costs in workers' utility function. Since these are "deep" parameters in the sense that they capture something fundamental about American workers and their mobility, they can be compared to my estimates. Diamond (2016) estimates the model separately for four worker groups with different races (black and non-black) and education (college and non-college). Since most worker in my sample period are not college graduates, I compare my results to the non-black and non-college worker group.

The results from the full model of Diamond (2016) suggest that living in a city outside the state of birth is equivalent to a 55 log point decrease in the real wage, and living in a city outside the census division of birth is equivalent to a 82 log point decrease in the real wage.[34] In terms of size, a state in the U.S. is similar to a province in China, and the American census division is similar to the big geographic region used in my estimation. Therefore these estimates are most comparable to the migration cost associated with crossing provincial and regional borders in my specification. For both skilled and unskilled workers, my estimates suggest that crossing a provincial border is equivalent to a 150 log point decrease in the real wage, and crossing a regional border is equivalent to a 180 log point decrease in the real wage. Comparison based on these two sets of coefficients suggests that migration in China is two to three times as costly as it is in the U.S., consistent with strong restriction on worker mobility in China.

[34] According to Tables 4 and 5 of Diamond (2016), in workers' utility function, the coefficient associated with living in the same state of birth is 3.44, and that associated living in the same region of birth is 1.219 (the leave-out category is living outside the census division of birth), whereas the coefficient for wage is 4.026. Therefore living outside the state of birth, but within the same census division is equivalent to $\frac{3.433-1.219}{4.026} * 100$ log points of the real wage, and living outside the census division of birth is equivalent to $\frac{3.433}{4.206} * 100$ log points in the real wage.

3.5.6.2 Trade Costs

Panel A in Table (3.8) presents the estimates for domestic trade costs. The model fits the data well, with an R^2 of 0.7. According to the estimates, crossing a provincial border increases trade costs by about 100 log points; crossing a regional border adds another 20 log points; sharing a common provincial border, on the other hand, could reduce the costs by 6.5 log points. If the dummies variables indeed capture the institutional barriers to domestic trade, the estimates indicate that these barriers are large.

Geographic distance significantly increases trade costs: for trading partners from different provinces within the same large region, distance has a large impact—each additional 1000 kilometers increases trade costs by 18 log points; for trading partners from two different regions, the impacts of distance are smaller: each additional 1000 kilometers increases trade costs by 8 log points. Cultural distance does not appear to affect trade costs. Perhaps due to the lack of variation in the number of cities *within* a province, and the radiuses of these cities, the estimation does not identify any trade costs associated with crossing city borders, or with additional kilometers between cities within the same province.

Overall, the estimates suggest that the trade costs between cities within China increase with both institutional and geographic distances. The former, captured by dummy variables in the regression, play a more important role, especially for close trade partners. The size of the inter-provincial dummy is smaller than in studies examining market fragmentation in China using earlier data (Poncet, 2005; Poncet, 2003). On the other hand, relative to comparable estimates for the U.S. (Wolf, 2000; Crafts and Klein, 2014), my estimate of the provincial-border effect is about twice as large, reflecting larger barriers to trade flows at provincial border in China. Since my estimates use variation in province-level trade costs, one valid concern is whether, due to the aggregate nature of the data, I might misattribute the cost of trading within a province to provincial borders. If that is the case, a further concern is whether the results from counterfactual experiments would be affected. In the appendix, I discuss related issues arising in the literature focusing on U.S. and perform a robustness exercise, in which I decrease the provincial border dummy in the economy to the U.S. level while increasing the coefficients for continuous distance components, keeping the overall level of domestic trade costs the same. All of the counterfactual results are robust to this alternative domestic trade cost structure.

Panel B of Table (3.8) presents the level of sectoral international openness by sector in China, defined as trade over production. Capital and manufacturing industries are more open compared to the agricultural industry, and this is reflected in the higher calibrated sectoral trade costs for the agricultural industry. Consistent with anecdotal evidence, international trade costs are smaller than the estimated inter-provincial costs, capturing the feature in the data that coastal provinces trade much more intensively with the ROW than with interior provinces.[35]

[35] In a state council meeting in 2014, Prime Minister Keqiang Li mentioned the complaint of producers in Shanghai that shipping costs within China were so high that it was cheaper to ship goods to California than to Beijing.

Table 3.8: Domestic and International Trade Costs

A. Domestic Trade Cost Estimates		
	Coefficient	Standard Error
I(Different Cities, Same Province)	0.0001	(0.0572)
I(Different Provinces, Same Region)	1.0897	(0.0719)
I(Different Regions)	1.2276	(0.0507)
I(Sharing Provincial Border)	-0.0648	(0.0393)
I(Same Province)*Distance	-0.0003	(0.1829)
I(Different Provinces, Same Region)*Distance	0.1836	(0.0683)
I(Different Regions)* Distance	0.0833	(0.0206)
Cultural Distance	0	(0.0554)
Observations	900	
R^2	0.70	

B. International Trade Cost Calibration: Targets and Parameter Values		
	Trade/Production	International Trade Costs
Agricultural Industry	0.12	0.93
Manufacturing Industry	0.36	0.75
Capital and Equipment Industry	0.46	0.67

Notes: Panel A of this table reports the estimates of domestic trade costs. Robust standard errors are in parentheses; *Distance* is measured as the great circle distance between cities (in 1000 km); *Cultural Distance* is measured as one minus the correlation in lagged ethnic minority shares between each city pair. Panel B of this table reports the level of industry openness in the data, and the calibrated international trade costs. The data on sectoral-level trade is aggregated from the 2005 UN Comtrade database. Production data is from the 2005 statistics yearbook.

Table 3.9: Non-targeted Moments

	Data	Model
Trade/Labor VA: mean	0.45	0.41
Trade/Labor VA: std	0.86	0.58
Corr (Trade/VA, Wage)		
For Worker Group:		
Urban unskilled	0.29	0.26
Rural unskilled	0.37	0.33
Urban skilled	0.25	0.14
Rural skilled	0.41	0.23

Notes: The data sample excludes the top 1% most open cites, with trade/GDP greater than 3, which is also the highest level of openness predicted by the model. *Trade/Labor VA* refers to the ratio between trade (imports+exports) and total payments to labor.

3.5.6.3 Additional Validations of the Model

The good fit of the migration and trade regressions suggests that the model is a reasonable approximation of the Chinese economy. Since I fully parameterize the model, I can look at additional moments that are not targets of calibration or estimation, to further assess the fit of the model.

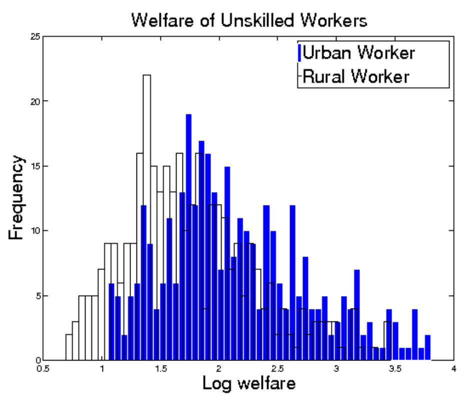
Table (3.9) presents summary statistics for city-level openness, and the correlation between city openness and average wage.[36] Overall, the model performs reasonably well in this test. It reproduces the mean city openness and the correlation between openness and wages for unskilled workers. It also captures the higher correlations between wages and trade in the rural sector compared to the urban sector for both types of workers. However, the model under-predicts the dispersion of city level openness, and the correlation between openness and wage, for high skilled workers.

3.5.6.4 Welfare and Productivity Distribution in the 2005 Equilibrium

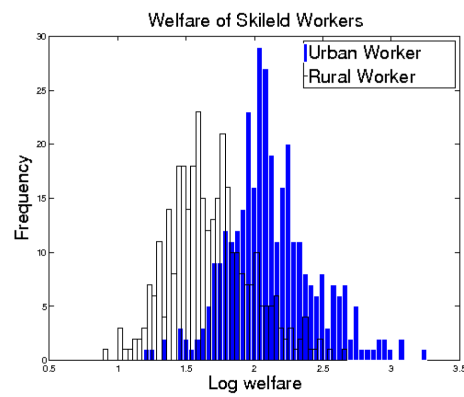
I compute the expected value of the welfare of workers, defined by Equation (3.6), by their places of birth. Figure (3.3) plots the density distribution of the log welfare for different worker groups. There is considerable dispersion in welfare among all worker groups. Among skilled workers, those born in cities with desirable

[36] City openness is computed as the sum of imports and exports, divided by local labor value-added. Since the model does not incorporate all primary inputs to GDP (e.g., land), for consistent comparison, I normalize city-level trade by wage payment instead of GDP.

amenities or high wages can be 150-250 log points better off than those born in other cities; among unskilled workers, the dispersion is even larger. Figure (3.4) plots the wage for each region (y axis) against the region's calibrated productivity (x axis). In both rural and urban sectors, wages clearly increase with local productivity; through the lens of the trade model, high wages imply high productivity in equilibrium. But because of the differences in market access across regions, the relationship is not perfect; if the trade costs were identical for each trading partners, the competition between producers in different regions would impose a perfect relationship between wages and productivity.



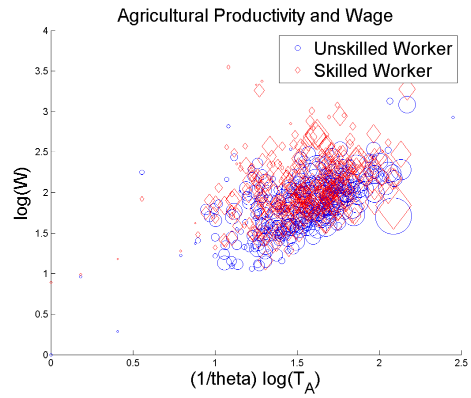
(a) Welfare of Unskilled Workers



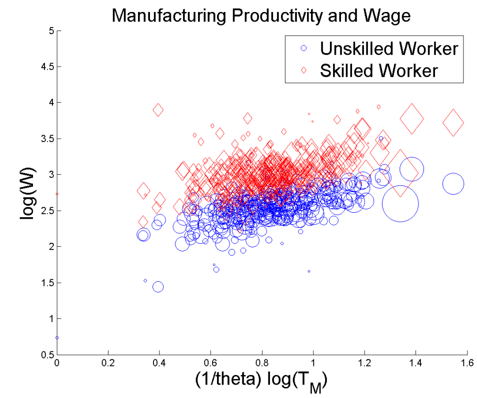
(b) Welfare of Skilled Workers

Figure 3.3: Welfare Distribution

These two sets of figures underscore the importance of limited worker mobility and internal geography in determining both trade and welfare. It is therefore critical to take these two elements into account when studying the welfare implications of trade.



(a) Rural Wage and Agricultural Productivity



(b) Urban Wage and Manufacturing Productivity

Figure 3.4: Wage and Productivity

3.6 Counterfactual Experiments

3.6.1 Benchmark Experiment

I use the model as a laboratory to conduct a sequence of policy experiments in order to examine the impacts of trade on welfare and inequality and the roles of within-country frictions in determining these impacts. In the first experiment, I keep all parameters of the model at the calibrated values and shut down international trade between China and the ROW by increasing the international trade costs to infinity.

3.6.1.1 Impacts of Trade on Welfare and Inequality

I compute the welfare gains from trade for each type of worker by calculating the relative changes in their welfare as China moves from the autarky to the open economy equilibrium.[37] Panel A of Table (3.10) reports the mean, standard deviation, and 5% and 95% percentiles of the distribution of welfare gains from trade, by worker skill groups. I compute the national average gains from trade by averaging over all worker groups in all regions, weighted by population. The national average gains from trade are 7.61%, similar in magnitude to the predictions of models without within-country heterogeneity and internal frictions.[38] However, the welfare gains do not accrue to everyone in the economy equally. First, different types of workers benefit differently from trade. The average gains from trade are about 11% for skilled workers, and 5% for unskilled workers. Within skill group, the impacts of trade also differ dramatically; among all worker groups, the standard deviations of the distributions of the welfare gains are similar to, or larger than, the respective means. The most-benefited group receives a welfare improvement of 20-30%, while some workers, likely unskilled ones, could experience welfare losses.

These patterns suggests that international trade might have important impacts on inequality, between workers with different skills, and among similar workers from different regions. I use the Theil index to measure the overall inequality in real wages in China, decomposing it into between-region and within-region components, and examine the impacts of international trade on each component.

Panel B of Table (3.10) presents the results. The first row is the decomposition for the benchmark economy. The between-region component constitutes about 90% of the overall inequality in China, while the within-region inequality between

[37] If we use per-capita real wage as the proxy for income and welfare, both qualitative and quantitative results hold.

[38] See, for example, Parro (2013) and Burstein et al. (2013)

Table 3.10: Aggregate and Distributional Impacts of Trade

A. Gains from Trade for Different Worker Groups				
	Mean	std	5%	95%
Urban Skilled	11.48	11.29	1.50	30.86
Urban Unskilled	5.19	8.43	-1.86	19.61
Rural Skilled	11.02	10.78	1.74	29.98
Rural Unskilled	4.98	8.31	-1.62	19.71
National Average	7.612			

B. Impacts of Trade on Inter- and Intra-Regional Inequality			
	Between Region	Within Region	Theil Index
Open Economy	0.182	0.031	0.213
Autarky	0.172	0.025	0.197
Increase (%)	5.8%	24.0%	8.15%
Relative Contribution	63.5%	36.5%	100%

Notes: Panel A of this table reports the summary statistics of the city-level welfare gains from trade for different worker skill groups. All numbers are in percentage points. *National Average* is the population-weighted average (across regions and worker skill groups) gains from trade. Panel B reports the decomposition of inequality, measured by the Theil index, into within- and between-region components in both the autarky and the open economy. The last row reports the relative contributions of the two components to the increase in the aggregate inequality after trade liberalization.

skilled and unskilled workers contributes only 10%. The second row of the table is the decomposition for the autarky economy. Again, the between-region component contributes more than 80% to the overall inequality.

As reported in the third row of Panel B, moving from the autarky economy to the open economy, the overall inequality in the country increases by 8%; both between- and within-region inequality increase. Although the within-region component accounts for only about 10% of inequality, its contribution to the increase is 36%. The between-region component accounts for the remaining 64% of the increase in aggregate inequality.

3.6.1.2 Trade and Inequality: the Role of Internal Geography

The decomposition in the previous section suggests that both within- and between-region components matter in the context of the impacts of trade on inequality. Since one important difference in the model between regions is their geographic environments, in particular, their accesses to foreign markets, in this section, I examine to what extent geography can explain the impacts of trade on different regions.

Each panel in Figure (3.5) plots the relationship between access to foreign markets and the city-level average welfare gains from trade for one worker group. The vertical axis is the ratio between the average welfare in the open economy and the average welfare in the autarky; the horizontal axis is each city's distance to its nearest port; the size of bubbles indicates city size. In all panels, regions form two groups in terms of their gains from trade: a coastal group that reaps most of the benefits and an interior group that benefits very little. For unskilled workers, some interior regions lie below 1, indicating that residents there bear welfare losses. The segregation of gains from trade is reminiscent of the segregation in terms of international trade integration in panel (a) of Figure (3.1): cities in the coastal provinces trade much more intensively with the ROW than with most interior cities. By limiting free mobility of goods within the country, intra-national trade costs indeed prevent interior regions from benefiting from trade.[39]

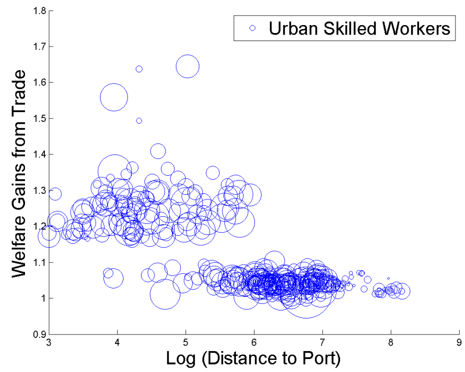
To illustrate the impacts of international trade on within-region inequality and how the impacts differ along the geographic dimension, Figure (3.6) plots *changes* in skill premia in rural and urban regions against regions' distances to their nearest ports. In the urban sector, except for a couple of regions in the hinterland, almost all regions experience increases in their skill premia after trade. The increase is around 10% in the coastal areas and about 5% in the interior. In the rural sector, on average, skill premia increase by 5% in coastal regions but decrease by 1-2% in the interior. The negative correlation between changes in skill premia and distances to the coast is consistent with empirical findings from a diff-in-diff approach (Han et al., 2012).

These figures illustrate clearly that within-country geography is relevant for

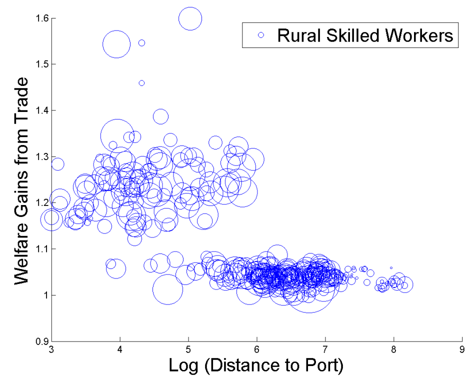
[39]The discontinuities in the gains from trade in Figure (3.5) as we move along the horizontal axis from the interior to the coast are largely driven by the large estimated value of the inter-provincial dummy in the domestic trade specification. As discussed in Section 5.5, and more in the appendix, I might potentially misattribute the costs of shipping over geographic distance to the provincial dummy. To address this concern I perform a robustness check, reported in the appendix, with a more continuous domestic trade cost structure (while keeping the overall level of domestic trade costs unchanged) and show all results are robust to this alternative domestic trade cost structure.

Figure 3.5: Geographic Distribution of Gains from Trade

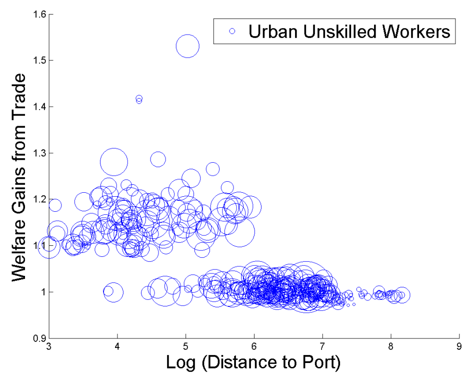
(a) Welfare Effect: Urban Skilled Workers



(b) Welfare Effect: Rural Skilled Workers



(c) Welfare Effect: Urban Unskilled Workers



(d) Welfare Effect: Rural Unskilled Workers

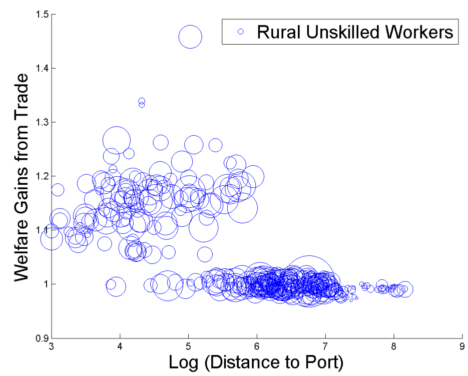
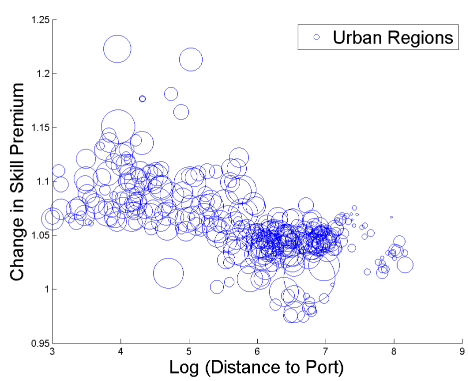
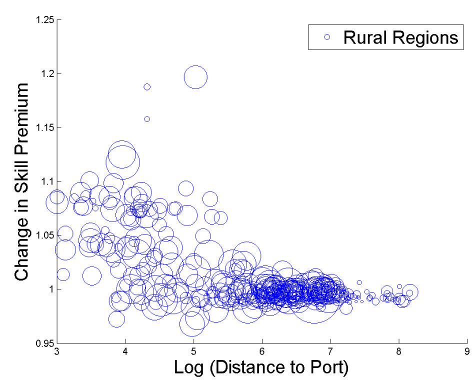


Figure 3.6: Trade and the Skill Premium

(a) Change in Urban Skill Premium



(b) Change in Rural Skill Premium



workers' gains from trade. The prediction that, within each skill group, workers from the coastal regions benefit more, is intuitive: international trade is, on average, welfare-improving, and since coastal regions trade more, workers there also benefit more. On the other hand, the forces behind the differential impacts of trade on skilled and unskilled workers within the same region (i.e., the changes in skill premia) and how the differential impacts vary across locations (i.e., the negative gradient of the changes in skill premia with respect to regions' distances to the coast) are less obvious. The next section will explain these patterns.

3.6.1.3 Explaining the Gradient of Changes in Skill Premia

The impacts of international trade on skill premia rest on its impacts on the relative demand and the relative supply of skilled versus unskilled workers. I discuss forces affecting these two factors separately.

The factor content theory of trade predicts that in an open economy, a developing country with abundant unskilled labor will specialize in producing and exporting unskilled-intensive products. Opening to trade causes a change in the pattern of specialization, and the relative demand for unskilled labor increases. The "Stolper-Samuelson Theorem" then predicts a decrease in the skill premium in developing countries following trade liberalization. This channel, however, is *not* an important channel in the current context: trade between China and ROW is largely within sector; therefore the change in relative demand for workers induced by the factor content of trade is unlikely to be large.[40]

[40]In the model, all industries in the urban sector uses the same composite equipped labor,

However, there is an mportant channel related to the factor content theory of trade, but operating through within country specialization. Because urban tradable industries (K and M industries) employ intermediate goods more heavily than the agricultural industry, they are more "transportation intensive." When a country opens up to trade, the coastal regions, due to their proximity to foreign suppliers, have stronger comparative advantage in these industries, and increase their specialization in producing capital and manufacturing products. The interior regions, on the other hand, increases their specialization in the agricultural industry. This shift in specialization patterns increases the relative demand for skilled workers in the coastal regions and decrease it in the hinterland, resulting in a negative relationship between the increases in regional skill premia, and regions' distance to the nearest port, as shown in Figure (3.6).

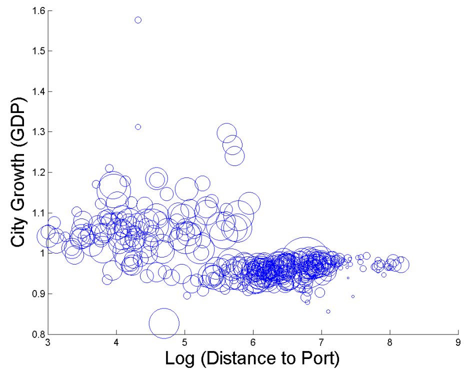
A second channel that affects relative demand for skilled workers is the capital-skill complementarity in production. China is a net importer of capital goods, which are complements to skilled workers. As a result, after international trade liberalization, skill premia increase across the board. Because coastal regions experience larger drops in the prices of capital goods, skill premia increase more on the coast.

Now consider changes in skill composition across regions after trade liberalization. A region's change in skill composition is determined by the *net* numbers of skilled/unskilled labor units migrating into that region, which in turn depend

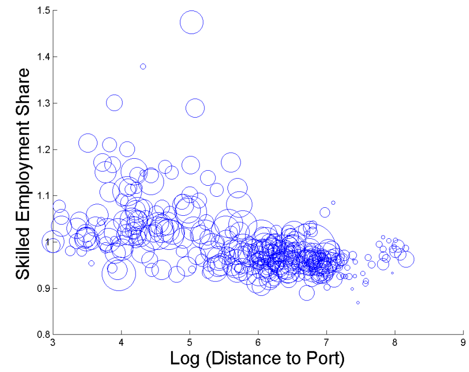
with the same skill intensity, so the factor content of trade theory can only operate through the reallocation of workers between urban and agricultural sectors; and this force is less important, because surplus is only 1% of production in the agricultural industry. Even if we allow capital and equipment industry to have a different skill intensity from the manufacturing industry, the factor content of trade is unlikely to change much, because the deficit in capital and equipment industry is less than 2% of production.

Figure 3.7: Reallocation After Trade

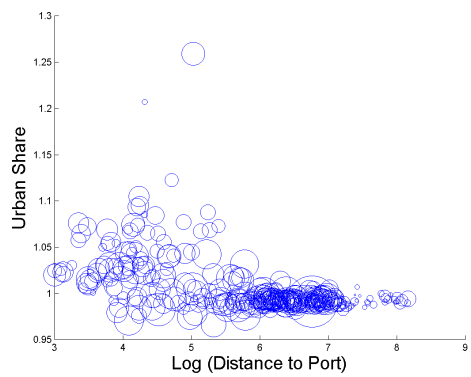
(a) City Growth



(b) Skill Share in Employment



(c) Urban Share in Value Added



on regions' access to labor pools and workers' costs of migration. Since the coastal regions gain more from trade, they will experience a net gain in population. As the estimated migration costs are lower for skilled workers, there will be more skilled workers moving from the interior into the coastal regions, pushing down skill premia on the coast and driving them up in the interior. The differential mobility between skilled and unskilled workers constitutes a third channel that tends to flatten the gradient of changes in skill premia and offset the channels described above.

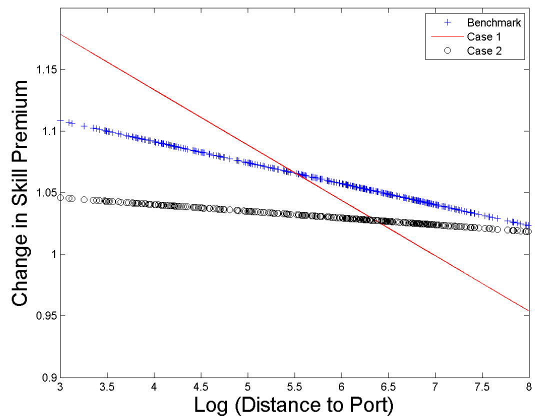
Apart from capital-skill complementarity, the two other channels work through reallocation of workers across regions or sectors. I provide evidence of the reallocation pattern predicted by these channels in Figure (3.7). Panel (a) plots the changes in a city's GDP against its distance to port and shows that coastal cities expand after trade at the expense of interior cities, due to the movements of workers from the interior to the coast. Panel (b) plots each city's skill share in local employment. Consistent with the prediction from the differential mobility channel, skill shares increase in the coast, and decrease in the hinterland. Finally, Panel (c) plots the share of urban value added in each city. As predicted by the "Domestic Stolper-Samuelson Effect", the share of urban value added in local economies increases on the coast and decreases in the hinterland.

To illustrate the quantitative importance of these channels, I conduct a sequence of counterfactual exercises and plot the changes in skill premia in these experiments in Figure (3.8).[41] "Benchmark" refers to the previous experiment. "Case 1" increases skilled workers' migration costs to the level of unskilled workers; experiment "Case 2" further shuts down capital skill complementarity by setting both the elasticity between capital and skilled worker and the elasticity between equipped and unskilled worker to 1.1, the estimates of Dix-Carneiro (2014) using a symmetric CES specification. In both cases, I compute the open economy and autarky equilibria and calculate the changes in wages and welfare as the country opens up to trade.

In Figure (3.8), when migration costs are the same for skilled and unskilled workers, the gradient for the changes in skill premia with respect to distance to

[41]For ease of comparison, I plot only the fitted value from a weighted least squares regression of changes in skill premia on regions' distances to port.

Figure 3.8: Changes in the Skill Premium



Notes: Each line is the fitted value from the weighted (by city size) least squares regression of changes in skill premia on the log of cities' distances to coast. Case 1: equal migration costs for both types of workers. Case 2: equal migration costs *and* no capital-skill complementarity.

the nearest port becomes steeper. The coastal regions now experience around 15% increases in skill premia, 5% points higher than in the benchmark experiment, and the interior regions experience roughly 5% decreases in skill premia. When I further shut down capital-skill complementarity, while there is still a mild gradient, the fitted line shifts downward, and becomes flatter, as expected. Globalization now increases within-region inequality more evenly across regions.

These experiments suggest that the various channels related to internal geography are all quantitatively important for both the geographic dimension, and the skill dimension of the distributional impacts of trade. In particular, the "Domestic Stolper-Samuelson Effect" is unexplored previously. Operating through changes in domestic specialization patterns, this channel has important implications for measuring regional trade exposures: since the interior regions in the country trade little with the ROW, most conventional measures of trade exposure will overlook these regions' exposures. However, because of international trade liberalization, the economic environment of these regions change dramatically. It is therefore important to take into account not only regions' international trade participation but also their trade with domestic partners, in measuring the regional impacts of trade.

3.6.2 Trade Liberalization Under Alternative Internal Geographies

In this section, I perform four additional experiments to examine how the distributional impacts of international trade differ in economies with different internal frictions. This question bears policy significance, as many countries that have recently experienced trade reforms are also liberalizing their domestic labor markets, or constructing transportation infrastructures aimed at lowering domestic trade costs.

In all four experiments, I solve the model for its open-economy equilibrium, in which international trade costs are set at the calibrated values, and the autarky equilibrium, in which international trade is shut down. Across these experiments the model economies differ in its intra-national frictions, which capture hypothetical reforms. In the first experiment, the provincial "border effect" in China's domestic trade—the trade costs associated with crossing a provincial border per se—is set to the U.S. level, estimated in the literature. In addition to domestic protectionism that might be at play in China, the domestic "border effect" might also capture geographic frictions that are not easily measured. Therefore rather than completely eliminating this effect, setting it to the U.S. level appears more reasonable.

In the second and third experiments, I eliminate the migration frictions arising from the Hukou policy for skilled and unskilled workers, respectively. In the fourth experiment, I combine all changes in the first three experiments.

The Hukou system affects many aspects of migration. In particular, its importance differ by city. Measuring the heterogenous effect of Hukou across cities is beyond the scope of this paper. As a starting point, I assume that the empirical estimate in the third section reflects the effect of Hukou on migration costs. Specifically, the empirical result suggests that each additional point of the Hukou reform index increases gross migration into a city by around 20%. Liberalizing all cities from the average degree of Hukou restriction in 2000 to complete openness, a score of 6 in my database, implies a 5-point change. According to this estimate, this means doubling the gross migrant inflow into a city.[42] Based on the model structure, I back out the corresponding average migration cost change corresponding to a 100% increase in inward migration, and use this in the experiments. In the appendix, I report the sources and values of the geographic parameters for the U.S. economy, and the way I back out the change in migration costs from the estimated effect of Hukou reform.

Table (3.11) reports the results of these experiments. The first column is the benchmark experiment. The second column is the experiment with lower intranational trade costs. Panel (A) reports summary statistics for welfare impacts by worker group, while Panel (B) reports aggregate outcomes. Compared with the ben-

[42] Obviously, when all cities implement the reform simultaneously, not all of them will experience a net increase in inward migration. But the gross migration could still increase.

chmark experiment, there are two major differences. First, with smaller domestic trade costs, the effect of international trade on inequality is smaller. The standard deviations of the welfare impact of trade for all worker groups decrease by about 50%. Measured the by Theil index, the trade-driven inequality increase shrinks by more than half. The shrinkage is mostly due to a smaller between-region inequality component. Intuitively, smaller domestic trade costs allow the impact of international trade to be spread more evenly across geographic regions. The second finding is somewhat surprising: with lower domestic trade costs, the overall gains from trade are smaller for all worker groups, and the country as a whole trades less intensively with the rest of the world. This result stands in contrast to the empirical findings (see, for example, Coşar and Demir, 2016) that better domestic infrastructures increase regional exports. Two reasons explains why the aggregate effects are different from the empirically identified effects: first, with lower domestic frictions, the coastal regions now trade more with the interior and less with the ROW; second, as the interior regions become better connected, they are more attractive as destinations, so more workers stay or migrate there. The size of the coastal regions—the regions that originally trade more intensively with the ROW—shrink. Therefore, the country's aggregate international trade decreases.

In the third and fourth columns, I report the experiments in which the Hukou system is abolished, for skilled and unskilled workers, separately. In both cases, the increase in the Theil index due to trade is smaller than in the benchmark economy. When the Hukou constraint is abolished, more workers are able to respond to the international trade liberalization by migrating to the coast. If more skilled workers

Table 3.11: Trade and Inequality: Different Domestic Frictions

A. Statistics by Worker Category										
	Benchmark		(2) TC		(3) SMC		(4) UMC		(2)+(3)+(4)	
	Mean	std	Mean	std	Mean	std	Mean	std	Mean	std
Urban Skilled	11.48	11.29	9.01	5.78	11.57	8.06	12.03	10.71	7.22	3.97
Urban Unskilled	5.19	8.43	4.49	3.86	7.67	8.77	7.15	10.58	4.32	3.24
Rural Skilled	11.02	10.78	8.81	5.22	11.53	8.42	12.09	10.65	7.40	3.08
Rural Unskilled	4.98	8.31	3.79	3.39	6.64	9.06	6.39	7.58	4.04	3.93

B: Aggregate Statistics					
	Benchmark	(2) TC	(3) SMC	(4) UMC	(2)+(3)+(4)
National Average	7.61	4.95	7.78	7.66	4.76
Trade Openness	60.13	41.18	57.89	55.63	37.35
Increase in Inequality	8.15%	3.50%	5.29%	6.15%	2.00%
Contribution-Between	63.54%	36.32%	73.00%	54.18%	23.45%
Contribution-Within	36.46%	63.68%	27.16%	45.82%	76.55%

Notes: This table reports the effects of trade on welfare and inequality in economies with different internal geographies. All numbers are in percentage points. The first column, *Benchmark*, is the same as the experiment reported in Table (3.10); *TC* refers to the case with lower domestic trade costs; *SMC* refers to the case with lower skilled migration costs; *UMC* refers to the case with lower unskilled migration costs; the final column combines the reductions to trade costs, skilled migration costs, and unskilled migration costs. Panel A reports the means and standard deviations of city-level welfare gains from trade for different worker groups. Panel B reports national average welfare gains, changes in inequality after trade, and the compositions of the changes in inequality. The measures for average gains from trade and inequality are the same as in Table (3.10). *Trade openness* is defined as the sum of imports and exports over GDP.

migrate, the increase in the skill premium in the coast will be smaller. Conversely, if more unskilled workers migrate, the skill premium in the coast will be even higher. Since the increase in the skill premium in the coast is the main source of the increase in the with-in region inequality component, the above discussion implies that if Hukou is abolished only for skilled (unskilled) workers, the within-region component will be less (more) important than in the benchmark case. This implication is consistent with the decomposition of inequality changes in Columns three and four of Table 3.11. In both cases, the gains from trade are slightly higher than the benchmark case, despite the fact that the economy trade less. So liberalizing domestic labor market through Hukou reforms can not only help with inequality, but also amplify the gains from trade.

The last column in Table 3.11 reports the result from the experiment in which reductions in trade and migration costs are combined. In this scenario, the Theil index increases by 2%, only a quarter of the benchmark case, and lower than all previous experiments. Reforms in goods and labor markets are complements in distributing the gains from trade more evenly. On the flip side, the aggregate gains from trade also decrease, and are even lower than the case with reduced domestic trade costs (the second column). With lower migration costs, more workers are attracted to the interior China as a response to the reduction in domestic trade costs. This decreases the country's participation in international trade, and reduces the welfare gains from trade.

In summary, the experiments in this section show that, reducing the domestic "border effect" and abolishing the Hukou system can both help the gains from trade to be spread more evenly across the country. Importantly, these reforms ensures almost everyone benefits through international trade liberalization in China (unskilled workers from one region still lose in the case of domestic trade reform, and all workers benefit in all other cases). Therefore these reforms offer an alternative way to place-based transfers to make international trade liberalization a Pareto improvement for China. On the other hand, these reforms also can have different effects on aggregate trade and the gains from trade. This result calls for a quantitative approach to evaluate the effect of domestic reform on international trade.

3.6.3 Sensitivity Analysis

This section reports the sensitivity analysis of the results to the parameters that are calibrated outside the model. Since the choice of these parameters affect the equilibrium distribution of wage, for each new parameters, I recalibrate regional productivity to match the 2005 equilibrium, and then solve the corresponding autarky equilibrium for the welfare effects of international trade.[43]

Table 3.12: Sensitivity Analysis

Parameters	Openness	Average Gain	Inequality Increase	Contribution (%)	
				Between	Within
Benchmark	60.13	7.61	8.15	63.54	36.46
$\rho_{kh} = 0.67, \rho_{lkh} = 1.1$	60.14	7.83	7.48	67.70	32.30
$\rho_{kh} = 1.1, \rho_{lkh} = 1.67$	60.29	7.69	6.56	69.94	30.06
$\rho_{kh} = 1.1, \rho_{lkh} = 1.1$	60.28	8.01	5.12	78.11	21.89
$\rho = 0.2 : R^2 = 0.6$	58.34	7.09	9.52	66.39	33.61
$\rho = 0.55 : R^2 = 0.8$	60.45	7.78	7.17	61.49	38.51
$\theta = 4.5$	46.57	4.89	6.72	64.03	35.97
$\theta = 5$	35.81	3.16	5.43	63.36	36.64

Notes: This table reports the effects of trade on welfare and inequality under alternative parameterizations. All numbers are in percentage points. Measures for average gains from trade and inequality are the same as in Table (3.10). *Trade openness* is defined as the sum of imports and exports over GDP.

The elasticity of substitution between capital and skilled workers, ρ_{kh}, and the elasticity between equipped skilled workers and unskilled workers, ρ_{lkh}, are important parameters in the model. For robustness, I first reduce ρ_{lkh}, to 1.1, implying that the upper nest is close to the Cobb-Douglas production function. I then increase ρ_{kh} to 1.1, keeping ρ_{lkh} at the benchmark level. Finally, I treat capital, skilled workers and unskilled workers as symmetric input into composite labor production by setting both ρ_{lkh} and ρ_{kh}, to 1.1. Rows (2)–(4) in Table (3.12) report

[43] This is different from the previously reported experiments in which we change only the exogenous parameters, without calibrating the economy to the 2005 equilibrium again. There, the goal was to understand the impacts of trade, in an otherwise similar economy with different structural parameters, so I kept regional productivity at the calibrated level.

the findings. As we can see, the aggregate gains from trade remain similar, while the changes in aggregate inequality and the contributions from the within-region component become smaller, as expected.

In the previous analysis, I calibrate the correlation between an individual's productivity draws across regions, ρ, to 0.4, to match the explanatory power of individual fixed effects in panel wage regression: individual fixed effects explain 70% of the remaining variation in wages, after controlling for regional fixed effects and individual demographics. I perform the policy experiment again, for $\rho = 0.2$ and $\rho = 0.55$, corresponding to an explanation power of 60% and 80%. Rows (5) and (6) in Table (3.12) report the findings. The results do not change much.

Finally, I increase the elasticity of trade, θ, from 4 to 4.4 and 5, and conduct the same exercise. The last two rows in Table (3.12) reports the results. When trade is more elastic, the country, in particular the coastal regions, benefits less from trade liberalization. Because of this, fewer people will migrate to the coast, further reducing the country's trade with the ROW. The increases in overall inequality are also smaller in these two cases.

Overall, the experiments suggests that the conclusions of the chapter are robust to alternative parameter values.

3.6.4 Discussion of Modeling Assumptions

In modeling the economy, I make several assumptions. In this section, I discuss how the violations of these assumptions would affect the main results.

In terms of the timing of migration, I assume that workers learn their idiosyncratic productivity draws in all regions prior to their move. Admittedly, in reality, there is substantial uncertainty about the payoffs to migration, which can be inferred from the fact that many migrants return to their birthplaces shortly after their migration (Kennan and Walker, 2011). In the empirical analysis, I classify workers as migrants if they are currently not in their birthplaces. Some of them might be temporary migrants who will shortly return to their hometowns. However, these migrants are unlikely to constitute an important part of the total migrants: even if 50% of migrants are temporary workers who return to their hometowns within two months, over a period of twenty years, the stock of migrants in each place will mostly be the permanent ones. My estimates of the migration costs, then, correspond to the long-run migration costs, which could be interpreted as reduced-form approximations of the real migration costs when there is uncertainty.

I use the Frechet distribution to model individuals' productivity draws. This distribution is a reasonable approximation of the wage distribution. In particular, it has a fat right tail. Most existing work in the migration literature makes similar parametric assumptions, using Logit or Pareto distributions. Instead of treating idiosyncratic migration decisions as outcomes of idiosyncratic individual preference shocks, an approach commonly adopted in migration literature, I assume they are driven by idiosyncratic productivity shocks, as in Ahlfeldt et al. (2015). The advantage of this approach is that, while individual preference is unobservable, parameters governing productivity shocks can be inferred directly from the wage distribution. The Frechet distribution is particularly attractive because under this assumption, we have tractable expressions to aggregate supply of efficiency units in each region. However, all the channels discussed in the chapter would apply under other distributional assumptions.

Since China's economy is growing quickly, and my model is static, one might worry that this discrepancy will make my results less useful. In analyzing the potential problems, it is important to be clear what dynamics one has in mind. First of all, the demographic structures are changing over time. My framework is general enough to incorporate multiple age groups, but I abstract from this mainly because of the limited sample size. Hence, my estimates could be interpreted as average migration costs across different age groups. If we want to simulate how the economy would evolve in the long run for a future policy change, it would be problematic because the future demography is different. However, the counterfactual experiments are backward looking; the counterfactuals aim to analyze the implications of China's *past* trade integration on welfare, when there are different magnitudes of internal frictions. Hence, the changing demography will not invalidate the results.

Another potential threat is that in 2005, the domestic labor markets are not yet in equilibrium; that is, there is potential migration that has not been realized. The existence of those workers will result in overestimating regional fixed effects for the regions experiencing migration outflows, and given the observed wages, this will in turn be reflected in overestimated amenities in these regions; similarly, I will under-estimate the amenities in popular migration destinations. In quantification, I find large dispersion in amenities, and if this argument is true, the real dispersion will be larger. In counterfactual experiments, however, since I keep the amenities fixed, the biases in the measured amenities will not affect the relative changes in the variables of interest, between trade and autarky equilibrium.

3.7 Conclusion

This chapter studies the aggregate and distributional impacts of trade on an economy with internal trade costs and migration costs. Focusing on China, I find that relative to the aggregate welfare gains, the distributional impacts of trade are large: the average welfare gain is about 7%, and the increase in overall inequality, as measured by the Theil index, is around 8%. Both the between-region inequality among workers with similar skill levels, and the within-region inequality between workers with different skill levels, contribute significantly to the increase in overall inequality. Reforms in domestic markets for factors or goods increase the internal integration of the country and reduce the effects of international trade on inter-regional inequality, but also reduce the welfare gains from international trade integration.

The impacts of trade on skill premia are not even across regions. Counterfactual experiments show that differential mobility between skilled and unskilled workers and changes in the specialization pattern of regions *within* the economy are important for the change in the gradient of skill premia. As neither of these forces arises in a model without frictional domestic trade or migration, it is important to take into account the role of internal geography, even when the interest is in the impacts of trade on the skill premium.

This chapter abstracts from some interesting and important aspects of the real world that could affect the impacts of international trade liberalization. For example, regional agglomeration effects might amplify both the distributional and the aggregate impact. Both agglomeration and dynamic effects are potentially important features to incorporate into future research.

As an independent contribution, this chapter constructs the first panel data of city-level Hukou reforms in China. Using this database, I show that Hukou reforms have a large impact on workers' mobility, and the potential to shape China's responses to international trade. The database could be of use to other researchers interested in the Chinese economy.

Appendix A: Chapter 1 Appendix

A.1 Theory

A.1.1 Lemma 1

Proof Country index i is omitted in this proof. Consider two R&D centers characterized by management scores (z_1^P, z_1^R), and (z_2^P, z_2^R), with $z_2^R > z_1^R$. Let $T : \mathbb{Z}^{\mathbb{P}} \times \mathbb{Z}^{\mathbb{R}} \to \Theta$ be the mapping from the type of an R&D center to the type(s) of researchers it recruits. Let $\theta_1 = T(z_1^P, z_1^R)$, $\theta_2 = T(z_2^R, z_2^P)$, so the first R&D center recruits θ_1 and the second θ_2.

I prove by contradiction that $\theta_1 <= \theta_2$. Suppose $\theta_1 > \theta_2$, given that (z_1^P, z_1^R) hires θ_1, it must be the case that it at least weakly prefer researchers with talent θ_1 to researchers with ability θ_2. From Equation 1.5, this implies

$$\pi(z_1^P)^{\frac{1}{1-\gamma}} w(\theta_2)^{-\frac{\gamma}{1-\gamma}} f(z_1^R, \theta_2)^{\frac{1}{1-\gamma}} \leq \pi_i(z_1^P)^{\frac{1}{1-\gamma}} w(\theta_1)^{-\frac{\gamma}{1-\gamma}} f(z_1^R, \theta_1)^{\frac{1}{1-\gamma}}$$

$$[\frac{w(\theta_1)}{w(\theta_2)}]^{\frac{\gamma}{1-\gamma}} \leq [\frac{f(z_1^R, \theta_1)}{f(z_1^R, \theta_2)}]^{\frac{1}{1-\gamma}}.$$

Similarly, for (z_2^P, z_2^R):

$$[\frac{w(\theta_2)}{w(\theta_1)}]^{\frac{\gamma}{1-\gamma}} \leq [\frac{f(z_2^R, \theta_2)}{f(z_2^R, \theta_1)}]^{\frac{1}{1-\gamma}}. \tag{A.1}$$

We therefore have:

$$[\frac{w(\theta_1)}{w(\theta_2)}]^{\frac{\gamma}{1-\gamma}} \geq [\frac{f(z_2^R, \theta_1)}{f(z_2^R, \theta_2)}]^{\frac{1}{1-\gamma}} > [\frac{f(z_1^R, \theta_1)}{f(z_1^R, \theta_2)}]^{\frac{1}{1-\gamma}} \geq [\frac{w(\theta_1)}{w(\theta_2)}]^{\frac{\gamma}{1-\gamma}},$$

where the first inequality is from Equation A.1, and the second from Assumption 1.

The above contradiction suggests that $\theta_1 \leq \theta_2$, and that $T(z^P, z^R)$ is weakly increasing in z^R. Now suppose $\theta_1 = \theta_2$, given the weak monotonicity of T, for R&D centers with $z^R \in (z_1^R, z_2^R)$, regardless of their production efficiency, will also recruit θ_1. Therefore in equilibrium, the demand for researchers with ability θ_1 will have a mass point, which contradicts with the assumption that talent distribution in each country has no mass point. Therefore the equilibrium matching function, $T(z^P, z^R)$, will be strictly increasing in z^R.

Now consider (z_1^P, z_1^R) and (z_2^P, z_1^R). If these two R&D centers hire different types of researchers, θ_1, and θ_2, then from monotonicity, all researchers with ability between θ_1 and θ_2 would be recruited by R&D centers with innovation efficiency z_1^R. The demand for researchers by R&D centers with this efficiency will be a positive mass, which contradicts that the distribution of efficiency for R&D centers have no mass point. Therefore $T_i(z^P, z^R)$ is independent of z^P.

A.1.2 Lemma 2

Proof Country index i is omitted. To show that $w(\theta)$ is differentiable, we consider an R&D center with innovation efficiency z_1^R, which is matched to $T(z_1^R)$. Consider z^R and $\theta = T(z_1^R)$. By the definition of $T(z_1^R)$, R&D centers with innovation efficiency z_1^R prefers researchers with ability θ instead of those with $\theta + d\theta$. Following Equation A.1, this implies:

$$[\frac{w(\theta + d\theta)}{w(\theta)}]^{\gamma} \geq [\frac{f(z_1^R, \theta + d\theta)}{f(z_1^R, \theta)}]$$

Consider $z_2^R = T^{-1}(\theta + d\theta)$, then similarly, we also have:

$$[\frac{w(\theta + d\theta)}{w(\theta)}]^{\gamma} \leq [\frac{f(z_2^R, \theta + d\theta)}{f(z_2^R, \theta)}]$$

From these two equations, we have:

$$\frac{f(z_1^R, \theta + d\theta) - f(z_1^R, \theta)}{f(z_1^R, \theta)} \leq \frac{w(\theta + d\theta)^{\gamma} - w(\theta)^{\gamma}}{w(\theta)^{\gamma}} \leq \frac{f(z_2^R, \theta + d\theta) - f(z_2^R, \theta)}{f(z_2^R, \theta)}$$

Dividing all three terms in the above inequality by $d\theta$, and letting $d\theta \to 0$, the first term approaches $\frac{f_2(z_1^R, \theta)}{f(z_1^R, \theta)}$, and the third term approaches $\frac{f_2(z_2^R, \theta)}{f(z_2^R, \theta)}$, which in term equals $\frac{f_2(z_1^R, \theta)}{f(z_1^R, \theta)}$, given the continuity of $f(z^R, \theta)$ and $T(z^R)$. Therefore we have

$$\lim_{d\theta \to 0} \frac{w(\theta + d\theta)^{\gamma} - w(\theta)^{\gamma}}{w(\theta)^{\gamma} d\theta} = \frac{f_2(z_1^R, \theta)}{f(z_1^R, \theta)}$$

Therefore $w(\theta)^\gamma$ is differentiable, with derivative being $w(\theta)^\gamma \frac{f_2(z_1^R,\theta)}{f(z_1^R,\theta)}$. This implies $w(\theta)$ is also differentiable, and its derive satisfies the following equation:

$$\frac{w'(\theta)}{w(\theta)} = \frac{f_2(z^R,\theta)}{\gamma f(z^R,\theta)}.$$

A.1.3 Proposition 2

Proof Country index i is omitted. Consider two talent distributions $H(\theta)$ and $\tilde{H}(\theta)$, with $\tilde{H}(\theta)$ more talent abundant than $H(\theta)$ according to Definition 1, and $h\tilde{(\theta)}$ and $h(\theta)$ being the corresponding PDFs. I use tilde to denote variable under $\tilde{H}(\theta)$. I first show that $\tilde{T}(z^R) \geq T(z^R)$, i.e., firms are matched with more talented researchers under $\tilde{H}(\theta)$ than under $H(\theta)$.

I prove by contradiction. From the definition of talent abundance, $\tilde{T}(\underline{z}^R) = \underline{\tilde{\theta}} > T(\underline{z}^R) = \underline{\theta}$, and $\tilde{T}(\overline{z}^R) = \tilde{\overline{\theta}} > T(\overline{z}^R) = \overline{\theta}$. Suppose for $z^R \in (\underline{z}^R, \overline{z}^R)$, $\tilde{T}(z^R) < T(z^R)$, then there must be $z_1^R < z^R$ and $z_2^R > z^R$, so that $\tilde{T}$ crosses T from above at z_1^R, and crosses it again from below at z_2^R. In $z^R \in (z_1^R, z_2^R)$, $\tilde{T}(z^R) < T(z^R)$.

For this to be possible, it must be the case that $\frac{\tilde{T}'(z_2^R)}{\tilde{T}'(z_1^R)} > \frac{T'(z_2^R)}{T'(z_1^R)}$. Using

Equation 1.7, and that $T(z_1^R) = \tilde{T}(z_1^R)$, $T(z_2^R) = \tilde{T}(z_2^R)$, this implies:

$$[\frac{\tilde{w}(\tilde{T}(z_1^R))}{\tilde{w}(\tilde{T}_i(z_2^R))}]^{\frac{1}{1-\gamma}} \frac{\int_{\mathbf{Z}^P} \tilde{\pi}(z^P)^{\frac{1}{1-\gamma}} g^P(z^P|z_2^R)dz^P}{\int_{\mathbf{Z}^P} \tilde{\pi}(z^P)^{\frac{1}{1-\gamma}} g^P(z^P|z_1^R)dz^P} \frac{\tilde{h}(\tilde{T}(z_1^R))}{\tilde{h}(\tilde{T}(z_2^R))} >$$
$$[\frac{w(T(z_1^R))}{w(T(z_2^R))}]^{\frac{1}{1-\gamma}} \frac{\int_{\mathbf{Z}^P} \pi(z^P)^{\frac{1}{1-\gamma}} g^P(z^P|z_2^R)dz^P}{\int_{\mathbf{Z}^P} \pi(z^P)^{\frac{1}{1-\gamma}} g^P(z^P|z_1^R)dz^P} \frac{h(T(z_1^R))}{h(T(z_2^R))}.$$

Note that $\pi(z^P)$ depend on the talent distribution because the latter determines general equilibrium outcomes, such as X and P. However, if one of the two additional conditions stated in Proposition 2 is satisfied,

$$\frac{\int_{\mathbf{Z}^P} \tilde{\pi}(z^P)^{\frac{1}{1-\gamma}} g^P(z^P|z_2^R)dz^P}{\int_{\mathbf{Z}^P} \tilde{\pi}(z^P)^{\frac{1}{1-\gamma}} g^P(z^P|z_1^R)dz^P} = \frac{\int_{\mathbf{Z}^P} \pi(z^P)^{\frac{1}{1-\gamma}} g^P(z^P|z_2^R)dz^P}{\int_{\mathbf{Z}^P} \pi(z^P)^{\frac{1}{1-\gamma}} g^P(z^P|z_1^R)dz^P}.$$

Then the above inequality further simplifies to:

$$[\frac{\tilde{w}(\tilde{T}(z_1^R))}{\tilde{w}(\tilde{T}(z_2^R))}]^{\frac{1}{1-\gamma}} \frac{\tilde{h}(\tilde{T}(z_1^R))}{\tilde{h}(\tilde{T}(z_2^R))} > [\frac{w(T(z_1^R))}{w(T(z_2^R))}]^{\frac{1}{1-\gamma}} \frac{h(T(z_1^R))}{h(T(z_2^R))}.$$

From the definition of talent abundance, $\frac{\tilde{h}(\tilde{T}(z_1^R))}{\tilde{h}(\tilde{T}(z_2^R))} < \frac{h(T(z_1^R))}{h(T(z_2^R))}$. From Equation 1.6 and log-supermodularity, $[\frac{\tilde{w}(\tilde{T}(z_1^R))}{\tilde{w}(\tilde{T}(z_2^R))}]^{\frac{1}{1-\gamma}} < [\frac{w(T(z_1^R))}{w(T(z_2^R))}]^{\frac{1}{1-\gamma}}$, so the above inequality cannot hold. Thus we have proved that $\tilde{T} \geq T, \forall z^R \in (\underline{z}^R, \overline{z}^R)$.

Let $y(z^P, z^R)$ and $\tilde{y}(z^P, z^R)$ denote the number of varieties an R&D center with efficiency (z^P, z^R) develops when the talent distribution is $H(\theta)$ and $\tilde{H}(\theta)$, respectively. Now consider the output difference between R&D centers with $z_1^R <$

z_2^R. From Equation 1.4, we have:

$$\begin{aligned}\log(\frac{y(z_P, z_1^R)}{y(z^P, z_2^R)}) &= \log(y(z^P, z_1^R)) - \log(y(z^P, z_2^R)) \\ &= \int_{z_1^R}^{z_2^R} \frac{\partial \log(y(z^P, z^R))}{\partial z^R} dz^R \\ &= \frac{1}{1-\gamma} \int_{z_1^R}^{z_2^R} \frac{f_1(z^R, T(z^R))}{f(z^R, T(z^R))} dz^R \\ &\leq \frac{1}{1-\gamma} \int_{z_1^R}^{z_2^R} \frac{f_1(z^R, \tilde{T}(z^R))}{f(z^R, \tilde{T}(z^R))} dz^R \\ &= \log(\frac{\tilde{y}(z_P, z_1^R)}{\tilde{y}(z^P, z_2^R)}),\end{aligned}$$

where the inequality uses the definition of log-supermodularity and the above conclusion that $\tilde{T}(z^R) \geq T(z^R)$.

A.1.4 Proposition 3

Proof To derive the gains from openness under Assumption 3, I proceed in three steps. The first step is to derive expression for production workers' real wage, $\frac{w_i^P}{P_i}$, in terms of measurable flows and total number of domestically invented varieties. The second step is to derive the relationship between production wage and total expenditure, $\frac{X_i}{w_i^P}$ in order to obtain $\frac{X_i}{P_i}$, the real income of a country. In the final step, I use $\frac{X_i}{P_i}$ to derive the gains from openness.

Step1: real wage for production worker I first derive real wage for production workers, $\frac{w_i^P}{P_i}$. The key step is to derive the total measure of varieties in each country. Under the assumption that $f(z^R, \theta) = z^R \theta^\beta$, Equation 1.6 becomes:

$\frac{w_i'(\theta)}{w_i(\theta)} = \frac{\beta}{\gamma}\frac{1}{\theta}$. Therefore the researcher wage schedule can be solved directly:

$$w_i(\theta) = \underline{w}_i \theta^{\frac{\beta}{\gamma}}$$

Under the assumption of zero fixed marketing cost, the per-variety variable profit $\pi_i(z^P)$ given by Equation 1.2 becomes ${z^P}^{\sigma-1} \sum_d \frac{1}{\sigma}(\frac{\sigma}{\sigma-1})^{1-\sigma}\Gamma(\frac{\delta+1-\sigma}{\delta})P_d^{\sigma-1}X_d\Psi_{id}^{\frac{\sigma-1}{\delta}}$. The total innovation output by an R&D center with (z^P, z^R) is therefore:

$$\begin{aligned} y_i(z^P, z^R) &= (\frac{\gamma}{\underline{w}_i})^{\frac{\gamma}{1-\gamma}} \pi_i(z^P)^{\frac{\gamma}{1-\gamma}} {z^R}^{\frac{1}{1-\gamma}} \\ &= (\frac{\gamma}{\underline{w}_i})^{\frac{\gamma}{1-\gamma}} [\sum_d \frac{1}{\sigma}(\frac{\sigma}{\sigma-1})^{1-\sigma}\Gamma(\frac{\delta+1-\sigma}{\delta})P_d^{\sigma-1}X_d\Psi_{id}^{\frac{\sigma-1}{\delta}}]^{\frac{\gamma}{1-\gamma}} {z^P}^{\frac{\gamma(\sigma-1)}{1-\gamma}} {z^R}^{\frac{1}{1-\gamma}} \end{aligned}$$

The measure of varieties invented in country i that can be produced with z^P is

$$\begin{aligned} m_i(z^P) &= R_i \int_{\mathbb{Z}_i^R} y_i(z^P, z^R) g(z^P|z^R) g_i(z^R) dz^R \\ &= R_i g(z^P) \int_{\mathbb{Z}_i^R} y_i(z^P, z^R) g_i(z^R) dz^R \\ &= K_i {z^P}^{\frac{\gamma(\sigma-1)}{1-\gamma} - \kappa_P - 1}, \end{aligned}$$

$$\text{Where } K_i = {\underline{z}_i^P}^{\kappa_P} \kappa_P (\frac{\gamma}{\underline{w}_i})^{\frac{\gamma}{1-\gamma}} [\sum_d \frac{1}{\sigma}(\frac{\sigma}{\sigma-1})^{1-\sigma}\Gamma(\frac{\delta+1-\sigma}{\delta})P_d^{\sigma-1}X_d\Psi_{id}^{\frac{\sigma-1}{\delta}}]^{\frac{\gamma}{1-\gamma}} R_i \int_{\mathbb{Z}_i^R} {z^R}^{\frac{1}{1-\gamma}} g_i(z^R) dz^R \tag{A.2}$$

The total measure of varieties developed in country i, M_i is

$$M_i = \int_{\underline{z}_i^P}^{\infty} m_i(z^P) dz^P = \frac{K_i}{\kappa^P - \frac{\gamma(\sigma-1)}{1-\gamma}} {\underline{z}_i^P}^{\frac{\gamma(\sigma-1)}{1-\gamma} - \kappa_P} \tag{A.3}$$

From Equation A.2, the productivity distribution for varieties developed in country i follows a Pareto distribution, with minimum $\underline{z_i}^P$ and dispersion parameter

$\frac{\gamma(\sigma-1)}{1-\gamma} - \kappa_P$. Given the measure of new varieties, M_i, the offshore production and trade block of the model corresponds to the model in Arkolakis et al. (2014) with exogenous entry. Following their notation, I define $\lambda^E_{id} = \frac{\sum_l X_{ild}}{X_d}$, $\lambda^T_{ld} = \frac{\sum_i X_{ild}}{X_d}$. Then λ^E_{id} denotes the share of consumption expenditure in country d that are spent on goods invented in country i, and λ^T_{ld} denotes the share of consumption in country d that are imported from country l.

Given that the total measure of varieties developed in country i is M_i, and that their productivity distribution is Pareto, the ideal price index in country d is given by:

$$\begin{aligned}
P_d^{1-\sigma} &= \Gamma(\frac{\delta+1-\sigma}{\delta})(\frac{\sigma}{\sigma-1})^{1-\sigma}\sum_i \Psi_{id}^{\frac{\sigma-1}{\delta}} M_i \frac{\kappa_P - \frac{\gamma(\sigma-1)}{1-\gamma}}{\underline{z}_i^{P\,\frac{\gamma(\sigma-1)}{1-\gamma}-\kappa_P}} \int_{\underline{z}_i^P}^{\infty} z^{P(\sigma-1)+\frac{\gamma(\sigma-1)}{1-\gamma}-\kappa_P-1} dz^P \\
&= \Gamma(\frac{\delta+1-\sigma}{\delta})(\frac{\sigma}{\sigma-1})^{1-\sigma}\sum_i \Psi_{id}^{\frac{\sigma-1}{\delta}} M_i \frac{\kappa_P - \frac{\gamma(\sigma-1)}{1-\gamma}}{\kappa_P - \frac{\sigma-1}{1-\gamma}} \underline{z}_i^{P(\sigma-1)}.
\end{aligned}$$

By definition of X_{id} from Equation 1.13, using the above expression for price, we have

$$\begin{aligned}
\lambda^E_{id} &= \frac{X_{id}}{X_d} \\
&= \frac{\Psi_{id}^{\frac{\sigma-1}{\delta}} M_i \underline{z}_i^{P(\sigma-1)}}{\sum_i \Psi_{id}^{\frac{\sigma-1}{\delta}} M_i \underline{z}_i^{P(\sigma-1)}} \\
&= \Gamma(\frac{\delta+1-\sigma}{\delta})(\frac{\sigma}{\sigma-1})^{1-\sigma} \frac{\kappa_P - \frac{\gamma(\sigma-1)}{1-\gamma}}{\kappa_P - \frac{\sigma-1}{1-\gamma}} \frac{\Psi_{id}^{\frac{\sigma-1}{\delta}} M_i \underline{z}_i^{P(\sigma-1)}}{P_d^{1-\sigma}}.
\end{aligned}$$

Therefore,

$$P_d^{1-\sigma} = \frac{\Psi_{dd}^{\frac{\sigma-1}{\delta}} M_d \underline{z}_d^{P(\sigma-1)}}{\lambda_{dd}^E} \Gamma(\frac{\delta+1-\sigma}{\delta})(\frac{\sigma}{\sigma-1})^{1-\sigma} \frac{\kappa_P - \frac{\gamma(\sigma-1)}{1-\gamma}}{\kappa_P - \frac{\sigma-1}{1-\gamma}}. \tag{A.4}$$

Note that $\lambda_{dd}^T = \sum_i \psi_{idd} \lambda_{id}^E$, where $\psi_{idd} = \frac{T_d(\frac{w_d^P \tau_{dd}}{\phi_{id}^P})^{-\delta}}{\Psi_{id}}$, so production wage satisfies:

$$w_d^{P\delta} = \frac{1}{\lambda_{dd}^T}(\sum_i \frac{T_d \phi_{id}^{P\,\delta} \lambda_{id}^E}{\Psi_{id}}) \tag{A.5}$$

To express $\sum_i \frac{T_d \phi_{id}^{P\,\delta} \lambda_{id}^E}{\Psi_{id}}$ in flow units, consider:

$$\begin{aligned}
\frac{X_{idd}}{X_d} &= \lambda_{id}^E \frac{T_d(\frac{w_d^P \tau_{dd}}{\phi_{id}^P})^{-\delta}}{\Psi_{id}} \\
\Leftrightarrow \frac{X_{idd}}{X_d} \Psi_{dd} w_d^{P\delta} &= \lambda_{id}^E T_d (\frac{\tau_{dd}}{\phi_{id}^P})^{-\delta} \frac{\Psi_{dd}}{\Psi_{id}} \\
\Leftrightarrow \frac{\sum_i X_{idd}}{X_d} \Psi_{dd} w_d^{P\delta} &= \sum_i \lambda_{id}^E T_d \phi_{id}^{P\,\delta} \frac{\Psi_{dd}}{\Psi_{id}} \\
\Leftrightarrow \frac{\sum_i X_{idd}}{X_d} \frac{T_d}{\psi_{ddd}} &= \sum_i \lambda_{id}^E T_d \phi_{id}^{P\,\delta} \frac{\Psi_{dd}}{\Psi_{id}} \\
\Leftrightarrow T_d \frac{\sum_i X_{idd}}{X_d} \frac{\sum_l X_{dld}}{X_{ddd}} &= \sum_i \lambda_{id}^E T_d \phi_{id}^{P\,\delta} \frac{\Psi_{dd}}{\Psi_{id}}
\end{aligned}$$

The real wage for production workers, $\frac{w_d^P}{P_d}$, is

$$
\begin{aligned}
\frac{w_d^P}{P_d} &= \underline{z}_d^P \Gamma(\frac{\delta+1-\sigma}{\delta})^{\frac{1}{\sigma-1}} (\frac{\sigma}{\sigma-1}) (\frac{\kappa_P - \frac{\gamma(\sigma-1)}{1-\gamma}}{\kappa_P - \frac{\sigma-1}{1-\gamma}})^{\frac{1}{\sigma-1}} M_d^{\frac{1}{\sigma-1}} {\lambda_{dd}^T}^{-\frac{1}{\delta}} {\lambda_{dd}^E}^{-\frac{1}{\sigma-1}} (\sum_i \lambda_{id}^E T_d {\phi_{id}^P}^{\delta} \frac{\Psi_{dd}}{\Psi_{id}})^{\frac{1}{\delta}} \\
&= \underline{z}_d^P \Gamma(\frac{\delta+1-\sigma}{\delta})^{\frac{1}{\sigma-1}} (\frac{\sigma}{\sigma-1}) (\frac{\kappa_P - \frac{\gamma(\sigma-1)}{1-\gamma}}{\kappa_P - \frac{\sigma-1}{1-\gamma}})^{\frac{1}{\sigma-1}} M_d^{\frac{1}{\sigma-1}} {\lambda_{dd}^T}^{-\frac{1}{\delta}} {\lambda_{dd}^E}^{-\frac{1}{\sigma-1}} (T_d \frac{\sum_i X_{idd}}{X_d} \frac{\sum_l X_{dld}}{X_{ddd}})^{\frac{1}{\delta}} \\
&= \underline{z}_d^P \Gamma(\frac{\delta+1-\sigma}{\delta})^{\frac{1}{\sigma-1}} (\frac{\sigma}{\sigma-1}) (\frac{\kappa_P - \frac{\gamma(\sigma-1)}{1-\gamma}}{\kappa_P - \frac{\sigma-1}{1-\gamma}})^{\frac{1}{\sigma-1}} M_d^{\frac{1}{\sigma-1}} {\lambda_{dd}^T}^{-\frac{1}{\delta}} {\lambda_{dd}^E}^{-\frac{1}{\sigma-1}} (T_d \lambda_{dd}^E \frac{\sum_i X_{idd}}{X_{ddd}})^{\frac{1}{\delta}} \\
&= \underline{z}_d^P \Gamma(\frac{\delta+1-\sigma}{\delta})^{\frac{1}{\sigma-1}} (\frac{\sigma}{\sigma-1}) (\frac{\kappa_P - \frac{\gamma(\sigma-1)}{1-\gamma}}{\kappa_P - \frac{\sigma-1}{1-\gamma}})^{\frac{1}{\sigma-1}} T_d^{\frac{1}{\delta}} M_d^{\frac{1}{\sigma-1}} {\lambda_{dd}^T}^{-\frac{1}{\delta}} {\lambda_{dd}^E}^{\frac{1}{\delta}-\frac{1}{\sigma-1}} (\frac{X_{ddd}}{\sum_i X_{idd}})^{-\frac{1}{\delta}}
\end{aligned}
\tag{A.6}
$$

Step 2: Relating consumption to wage: In the second step, I derive the ratio between production wage and expenditures, $\frac{X_d}{P_d}$. I start with the market clearing condition for production workers:

$$
\begin{aligned}
w_d^P L_d^P &= \frac{\sigma-1}{\sigma} Y_d + \sum_o E_o c_d^R w_d^P (1 - G_o(\hat{z}_{od}^R)) \\
&= \frac{\sigma-1}{\sigma} Y_d + \sum_o E_o c_d^R w_d^P (\frac{\hat{z}_{od}^R}{\underline{z}_o^R})^{-\kappa_R}
\end{aligned}
\tag{A.7}
$$

The first term on the right hand side is total demand for production workers from production, while the second term on the right hand side is demand from the overhead of R&D centers. The second line uses the fact that $G_o^E(\tilde{z}^R)$ follows Pareto distribution. The goal here is to express the second term in the right hand side in the form of aggregate flow variables.

Consider the expected profit for a firm with z^R to perform offshore R&D in country i:

$$\begin{aligned}\pi_i^R(z^R) &= \int_{\mathbb{Z}_i^P} \pi_i^R(z^P, z^R) g(z^P|z^R) dz^P \\ &= (1-\gamma)\int_{\mathbb{Z}_i^P} y_i(z^P, z^R)\pi_i^P(z^P) g(z^P) dz^P \\ &= (1-\gamma)\kappa_P(\frac{\gamma}{\underline{w}_i})^{\frac{\gamma}{1-\gamma}}[\sum_d \frac{1}{\sigma}(\frac{\sigma}{\sigma-1})^{1-\sigma}\Gamma(\frac{\delta+1-\sigma}{\delta})P_d^{\sigma-1}X_d\Psi_{id}^{\frac{\sigma-1}{\delta}}]^{\frac{1}{1-\gamma}} z^{R\frac{1}{1-\gamma}} \underline{z}_i^{P\kappa_P}\int_{\underline{z}_i^P}^{\infty} z^{P\frac{\sigma-1}{1-\gamma}-\kappa_P-1} dz^P \\ &= K_i' z^{R\frac{1}{1-\gamma}},\end{aligned}$$

where $K_i' = \frac{\kappa_P}{\kappa_P - \frac{\sigma-1}{1-\gamma}}(1-\gamma)(\frac{\gamma}{\underline{w}_i})^{\frac{\gamma}{1-\gamma}}[\sum_d \frac{1}{\sigma}(\frac{\sigma}{\sigma-1})^{1-\sigma}\Gamma(\frac{\delta+1-\sigma}{\delta})P_d^{\sigma-1}X_d\Psi_{id}^{\frac{\sigma-1}{\delta}}]^{\frac{1}{1-\gamma}}\underline{z}_i^{P\frac{\sigma-1}{1-\gamma}}$. Therefore the cutoff innovation efficiency level for a firm from o to open R&D center in country i is given by:

$$(\hat{z}_{oi}^R)^{\frac{1}{1-\gamma}} = \frac{c_i^R w_i^P}{K_i' \phi_{oi}^{R\frac{1}{1-\gamma}}}$$

The total fixed costs paid by firms from country o, doing R&D in country i, is

$$E_o c_d^R w_d^P (\frac{\hat{z}_{od}^R}{\underline{z}_o^R})^{-\kappa_R} = E_o(\underline{z}_o^R)^{\kappa_R}(\hat{z}_{oi}^R)^{\frac{1}{1-\gamma}-\kappa_R}(\phi_{oi}^R)^{\frac{1}{1-\gamma}} K_i'$$

Now consider total R&D expenditures incurred by offshore R&D centers of country o firms:

$$\begin{aligned}I_{oi} &= E_o\frac{\gamma}{1-\gamma}\int_{\hat{z}_{oi}^R}^{\infty}\int_{\underline{z}_i^P}^{\infty} \pi_i^R(z^P, z^R\phi_{oi}^R) g_i(z^P) dz^P g_o(z^R) dz^R \\ &= E_o\frac{\gamma}{1-\gamma}K_i'\phi_{oi}^{R\frac{1}{1-\gamma}}\frac{\kappa_R}{\kappa_R - \frac{1}{1-\gamma}}(\underline{z}_o^R)^{\kappa_R}(\hat{z}_{oi}^R)^{\frac{1}{1-\gamma}-\kappa_R},\end{aligned}$$

so the overhead cost is a fixed share, $\frac{(1-\gamma)\kappa_R - 1}{\gamma\kappa_R}$, of total R&D expenditure by foreign firms. Noting that this ratio holds true for offshore R&D center from all other countries, except for the home country because they do not incur additional fixed costs, the labor market clearing condition, A.7, becomes:

$$w_d^P L_d^P = \frac{\sigma-1}{\sigma}Y_d + \frac{(1-\gamma)\kappa_R - 1}{\gamma\kappa_R}(1-\frac{I_{dd}}{I_d})I_d,$$

and we have:

$$\frac{X_d}{w_d^P} = \frac{L_d^P}{\frac{\sigma-1}{\sigma}\frac{Y_d}{X_d} + \frac{(1-\gamma)\kappa_R-1}{\gamma\kappa_R}\frac{I_d}{X_d}(1-\frac{I_{dd}}{I_d})}.$$

We can combine this equation with Equation A.6 to obtain the expression for gains from openness:

$$GO_d = (\frac{M_d}{M_d'})^{\frac{1}{\sigma-1}}(\frac{X_{ddd}}{\sum_i X_{idd}})^{-\frac{1}{\delta}}{\lambda_{dd}^T}^{-\frac{1}{\delta}}{\lambda_{dd}^E}^{\frac{1}{\delta}-\frac{1}{\sigma-1}}\frac{\frac{\sigma-1}{\sigma}}{\frac{\sigma-1}{\sigma}\frac{Y_d}{X_d} + \frac{(1-\gamma)\kappa_R-1}{\gamma\kappa_R}\frac{I_d}{X_d}(1-\frac{I_{dd}}{I_d})} - 1, \tag{A.8}$$

where M_d is the measure of varieties innovated in country d in the benchmark equilibrium, while M_d' is the measure of output invented under the counterfactual autarky equilibrium.

Step 3: deriving relative change in measure of varieties: The final step is to express $\frac{M_d}{M_d'}$ in terms of observable flows. To do this, we first derive $\underline{w}_d$, the wage for the bottom researcher in country d. Notice that under the multiplicative assumption, wage schedule is $w_i(\theta) = \underline{w}_i\theta^{\frac{\beta}{\gamma}}$, and the optimal demand for researcher satisfies $l_i(z^P, z^R) = [\frac{\gamma}{\underline{w}_i}\pi_i^P(z^P)z^R]^{\frac{1}{1-\gamma}}\theta^{-\frac{\beta}{\gamma}}$.

In this case, the wage schedule can be interpreted as each unit of researcher efficiency, defined as $\theta^{\frac{\beta}{\gamma}}$, is paid a unit of wage, $\underline{w}_i$. Therefore the payment to a researcher whose ability is $2\theta^{\frac{\beta}{\theta}}$ is simply twice the payment to a researcher with ability $\theta^{\frac{\beta}{\theta}}$. The labor demand equation can be manipulated into:

$$\frac{l_i(z^P, z^R)\theta^{\frac{\beta}{\gamma}}}{z^{R\frac{1}{1-\gamma}}} = (\frac{\gamma}{\underline{w}_i})^{\frac{1}{1-\gamma}}\pi_i(z^P)^{\frac{1}{1-\gamma}},$$

which states that for an R&D center (z^P, z^R), each unit of innovation management efficiency, $z^{R\frac{1}{1-\gamma}}$, is matched with $(\frac{\gamma}{\underline{w}_i})^{\frac{1}{1-\gamma}}\pi_i(z^P)^{\frac{1}{1-\gamma}}$ unit of researcher efficiency talent. The labor market clearing condition for researcher efficiency unit is then:

$$L_i^R \int_{\Theta_i} \theta^{\frac{\beta}{\gamma}} h_i(\theta) d\theta = R_i(\frac{\gamma}{\underline{w}_i})^{\frac{1}{1-\gamma}}[\int_{\mathbb{Z}_i^P} \pi_i(z^P)^{\frac{1}{1-\gamma}} g_i(z^P)dz^P][\int_{\mathbb{Z}_i^R} z^{R\frac{1}{1-\gamma}} g_i(z^R)dz^R]$$

Therefore

$$(\frac{\gamma}{\underline{w}_i})^{\frac{1}{1-\gamma}} = \frac{L_i^R \int_{\Theta_i} \theta^{\frac{\beta}{\gamma}} h_i(\theta) d\theta}{R_i \frac{\kappa_P}{\kappa_P - \frac{\sigma-1}{1-\gamma}} \underline{z}_i^{P\frac{\sigma-1}{1-\gamma}} [\sum_d \frac{1}{\sigma}(\frac{\sigma}{\sigma-1})^{1-\sigma}\Gamma(\frac{\delta+1-\sigma}{\delta})P_d^{\sigma-1} X_d \Psi_{id}^{\frac{\sigma-1}{\delta}}]^{\frac{1}{1-\gamma}} \int_{\mathbb{Z}_i^R} z^{R\frac{1}{1-\gamma}} g_i(z^R)dz^R}.$$

Substituting this into Equation A.3, we obtain the expression for the measure of varieties developed:

$$M_i = \frac{\kappa_P - \frac{\sigma-1}{1-\gamma}}{\kappa_P - \frac{\gamma(\sigma-1)}{1-\gamma}}(L_i^R \int_{\Theta_i} \theta^{\frac{\beta}{\gamma}} h_i(\theta) d\theta)^{\gamma}(R_i \int_{\mathbb{Z}_i^R} z^{R\frac{1}{1-\gamma}} g_i(z^R)dz^R)^{1-\gamma}$$

As this expression makes clear, under the multiplicative assumption of $f(z^R, \theta)$, the aggregate innovation output is a Cobb-Douglas function of total stock of innovation efficiency stock, and researcher talent stock in an economy. This expression also implies that in the absence of immigration, the ratio between the measure of varieties in the benchmark equilibrium and in the autarky equilibrium is

$\frac{M_i}{M_i'} = (\frac{R_i \int_{\mathbb{Z}_i^R} z^{R\frac{1}{1-\gamma}} g_i(z^R) dz^R}{R_i' \int_{(\mathbb{Z}_i')^R} z^{R\frac{1}{1-\gamma}} g_i'(z^R) dz^R})^{1-\gamma}$, where the denominator is the stock of innovation efficiency units in autarky equilibrium when offshore R&D is not possible.

Recall that each unit of $z^{R\frac{1}{1-\gamma}}$ is matched with $\frac{\gamma}{\underline{w}_i}^{\frac{1}{1-\gamma}}$ units of researcher efficiency units. Since the draw of z^P is independent of z^R, the share of researcher efficiency units recruited by foreign R&D centers in the open economy is proportional to the share of innovation efficiency units of these R&D centers in the country, that is, $\frac{R_i \int_{\mathbb{Z}_i^R} z^{R\frac{1}{1-\gamma}} g_i(z^R) dz^R}{R_i' \int_{(\mathbb{Z}_i')^R} z^{R\frac{1}{1-\gamma}} g_i'(z^R) dz^R} = \frac{I_i}{I_{ii}}$.

The expression for the welfare gains from openness therefore is:

$$
\begin{aligned}
GO_d &= (\frac{X_{ddd}}{\sum_i X_{idd}})^{-\frac{1}{\delta}} \lambda_{dd}^{T\,-\frac{1}{\delta}} \lambda_{dd}^{E\,\frac{1}{\delta}-\frac{1}{\sigma-1}} (\frac{I_{dd}}{I_d})^{-\frac{1-\gamma}{\sigma-1}} \frac{\frac{\sigma-1}{\sigma}}{\frac{\sigma-1}{\sigma}\frac{Y_d}{X_d} + \frac{(1-\gamma)\kappa_R-1}{\gamma\kappa_R}\frac{I_d}{X_d}(1-\frac{I_{dd}}{I_d})} - 1 \\
&= (\frac{X_{ddd}}{\sum_i X_{idd}})^{-\frac{1}{\delta}} (\frac{\sum_i X_{idd}}{X_d})^{-\frac{1}{\delta}} (\frac{\sum_l X_{dld}}{X_d})^{\frac{1}{\delta}-\frac{1}{\sigma-1}} (\frac{I_{dd}}{I_d})^{-\frac{1-\gamma}{\sigma-1}} \frac{\frac{\sigma-1}{\sigma}}{\frac{\sigma-1}{\sigma}\frac{Y_d}{X_d} + \frac{(1-\gamma)\kappa_R-1}{\gamma\kappa_R}\frac{I_d}{X_d}(1-\frac{I_{dd}}{I_d})} - 1,
\end{aligned}
\tag{A.9}
$$

which is equivalent to Equation 1.17 in Section 2.

A.2 Quantification

This section provides additional information on the quantification section of the first chapter.

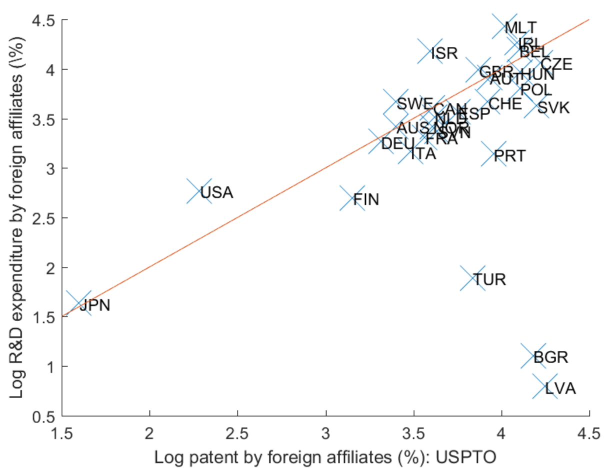
A.2.1 Data

I use the OECD harmonized USPTO data to construct the bilateral offshore R&D measure. Specifically, for each country, the OECD counts the number of

USPTO patents invented in that country that are assigned to firms from other countries. The data is up to country pair level, so we know, for example, the number of patents invented by individuals residing in Germany, but are assigned to firms located in the U.S. I use this data set to measure the share of R&D performed within the border of a host country by firms from each foreign country. Based on this measure, the first column in Table A.1 reports combined share of offshore R&D by foreign firms in total R&D in each host country.

One important drawback of using patent data is that, it might be biased due to differential selection into patenting across countries. For example, if only firms selling to the U.S. patent at USPTO, then the measure of offshore R&D will be biased towards these firms. I present two pieces of evidence to show this selection is not important in the context of this paper.

Figure A.1: The Comparison of Two Measure



Notes: The figure plots measure of offshore R&D based on R&D expenditure against the measure based on patenting from the OECD harmonized USPTO data, averaged over 1998-2007. Redline indicates perfect correlation.

First, in Figure A.1, I compare the patent-based measure of offshore R&D to the expenditure-based measure, to show that such selection is unlikely to be

Table A.1: Country Characteristics

		Innovation Mgt. Dist.				Talent Dist.		
ISO	Offshore R&D	Mean	Std.	Skewness	Imputed	Mean	Top Share	Basic Share
AUS	27.09	6.43	3.64	1.88		5.09	11.241	93.84
AUT	50.41	6.14	3.80	2.49	Yes	5.09	9.7359	93.106
BEL	57.25	6.20	3.85	2.48	Yes	5.04	9.3775	93.129
BRA	37.29	5.26	3.33	2.35		3.64	1.087	33.846
CAN	33.73	8.40	6.09	2.01		5.04	8.3302	94.843
CHN	52.24	5.94	2.74	1.85		4.94	8.3393	93.478
DEU	23.99	8.21	5.25	2.20		4.96	8.1654	92.694
DNK	33.69	6.87	5.00	2.67	Yes	4.96	8.7538	88.78
ESP	43.23	5.29	3.46	2.21		4.83	7.9337	85.88
FIN	18.22	7.00	5.10	2.67	Yes	5.13	12.386	95.775
FRA	33.97	6.43	4.25	2.52		5.04	8.4914	92.621
GBR	45.88	7.33	4.84	2.06		4.95	8.7857	92.885
GRC	58.26	5.63	3.70	1.89		4.61	4.2429	79.772
IND	58.03	5.93	5.00	3.02		4.28	1.2742	92.188
IRL	55.46	7.14	6.73	3.86		4.99	9.4048	91.371
ITA	29.70	6.47	4.15	2.17		4.76	5.4455	87.543
JPN	5.33	7.83	5.57	1.82		5.31	16.758	96.669
KOR	5.05	6.76	4.05	1.88	Yes	5.34	17.835	96.159
MEX	49.42	6.90	4.43	1.66		4.00	0.88366	48.933
NLD	34.82	6.56	4.14	2.46	Yes	5.12	9.1594	96.536
POL	60.79	7.25	4.60	1.73		4.85	9.8585	83.764
PRT	50.24	5.38	2.99	1.94		4.56	3.1628	80.269
SWE	27.25	7.06	4.17	1.99		5.01	8.7552	93.943
TUR	51.82	5.86	2.58	2.09		4.13	3.9199	58.233
USA	8.08	10.94	8.15	2.15		4.90	7.3299	91.824

Notes:"*Offshore R&D*" refers to the share (%) of patents invented in a country but owned by firms from foreign countries, based on the USPTO data. "*Innovation Mgt. Dist.*" refers to the sample distribution statistics constructed from the World Management Survey as described in text. "*Imputed*" indicates whether the innovation management distribution statistics are imputed. "*Talent*" refers talent distribution statistics from Hanushek and Woessmann (2012), in which "*Mean*" is the mean score for a country, and "*Top Share*" and "Basic Share" are share of students achieving "top" and "basic" performance, respectively. The performance standards are common across countries.

Table A.2: Correlation Between Various Measures of Offshore R&D

Full Sample

	R&D	USPTO	EPO	PCT
R&D	1.00			
USPTO	0.37	1.00		
EPO	0.42	0.89	1.00	
PCT	0.57	0.88	0.93	1.00

Excluding Three Outliers

	R&D	USPTO	EPO	PCT
R&D	1.00			
USPTO	0.72	1.00		
EPO	0.67	0.90	1.00	
PCT	0.76	0.93	0.93	1.00

Notes: The upper panel presents correlations between four measures of the share of R&D activities done by foreign firms in a host country. The four measures are based on R&D expenditures, and patenting data from three patent offices: USPTO, EPO, and PCT. The lower panel present the correlation excluding three outlier countries: Latvia, Bulgaria, and Turkey.

important in the context of offshore R&D. As the figure indicates, other than three outliers, Latvia, Turkey, and Bulgaria, the two measures line up closely.

Second, if differential selection in patenting in the U.S. due to product market consideration is important, one should expect the same measure based on the European Patent Office (EPO) and the Patent Cooperation Treaty (PCT) patents to give different results. The upper panel in Table A.2 presents the correlation matrix of the four measures. As the table indicates, the three patent-based measures are close to each other, and they are all different from the expenditure-based measure. However, the discrepancies are mainly driven by the three outliers, Latvia, Turkey, and Bulgaria. Once the three outliers are excluded, as the lower panel in Table A.2 shows, all patent-based measures are strongly correlated with the expenditure-based measure.

The calibration uses the World Management Survey (Bloom et al., 2012) and an internationally comparable cognitive ability score database (Hanushek and Woessmann, 2012). I compute the mean, standard deviation, and skewness of innovation management efficiency distribution in each country, by computing the corresponding statistics of the exponent of firm-level talent management scores for each country. I take exponent so that the distribution of scores has a right tail that resembles the firm size distribution. The distribution statistics for cognitive test scores are directly from Hanushek and Woessmann (2012). These statistics include the average cognitive score for high school students in a country, the share of students that achieve "top" performance, and the share of student that achieve "basic" performance. Thresholds for "top" and "basic" performance are defined in absolute level so the shares are comparable internationally. These statistics are reported in Table A.1.

A few countries in the sample are not included in the world management survey. I impute their management distribution statistics by regressing each statistics on income, R&D share, and geographic-region fixed effects, where geographic regions are at sub-continent level. The R^2 of these regressions are all above 0.85. In general, geographic-region dummies have biggest explanatory power. Table A.1 indicates which countries have imputed management scores.

The model economy consists of the 25 countries reported in the table, and a statistical aggregation of another 23 countries: Argentina, Belarus, Switzerland, Chile, Colombia, Costa Rica, Guatemala, Croatia, Iran, Islamic Rep, Israel, Lebanon, Malaysia, Norway, New Zealand, Saudi Arabia, Singapore, El Salvador, Thailand, Tunisia, Uruguay, Venezuela, South Africa. The main constraint in modelling these countries explicitly is the availability of World Management Survey and the World Input-Output Database. In calibrating the distributions for this "country", I use the same imputation method as described above when World Management Survey is not available, and then use country population as weights to compute the average distribution statistics.

A.2.2 Calibration

A.2.2.1 Relating Production Efficiency to Innovation Efficiency

To discipline the relationship between firms' innovation and production management efficiencies, I use micro data from the World Management Survey to estimate the following equation:

$$\text{Prob}(z^P \in H | z^R) = \frac{\exp(A + B \times z^R)}{1 + \exp(A + B \times z^R)}. \tag{A.10}$$

This data base covers around 11000 firms from 34 countries. I classify a firm as being a H type, if its production management scores falls in the top 1% in the sample. Because in calibration, I assume the management score in the model, z^R, is exponent of the management score in the data, in this estimation, I transform the innovation score accordingly. Table A.3 presents summary statistics on innovation management score, defined this way, and the indicator for H type.

Table A.3: Firm Management Score Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
z^R	11338	6.68	4.92	1	54.6
$\mathbb{I}_{z^P \in G_H^P}$	11340	0.0109	0.104	0	1

Notes: This table presents the summary statistics for firm-level innovation management score and the indicator for whether a firm is in the top 1% production efficiency.

Table A.4: Estimates for A and B

Dependent Variable:	$\mathbb{I}_{z^P \in G_H^P}$
	Coeff
z^R	0.167***
	(0.009)
Constant	-6.30***
	(0.17)
Pseudo R^2	0.2545
N	11338

Notes: This table presents results from a Logit regression of the high production efficiency indicator $\mathbb{I}_{z^P \in G_H^P}$, on firms' innovation efficiency, z^R. The high production efficiency indicator takes a value of 1 if the production management score of a firm is in the top 1% in the world.
Standard errors are in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.4 presents result from Logit estimation of Equation A.10, using the full sample. Consistent with positive correlation between innovation and management efficiency, the estimate for A is positive and statistically significant.

A.2.2.2 Estimating the Matching Function

In calibration, I use a nonparametric patching function to determine the value for the complementarity parameter. I estimate this matching function based on Panel B of Table 2.4. The measure for firm innovation efficiency and inventor talent is the same as in the regression. I estimate a local linear regression of inventor talent on firm innovation efficiency, focusing on the job-switching inventors. I control for the fixed effects for year and patent category, as well as firm and inventor age, defined as years since firs time the firm/inventor appears in the USPTO database.

A.2.2.3 Model Fit: Additional Figures

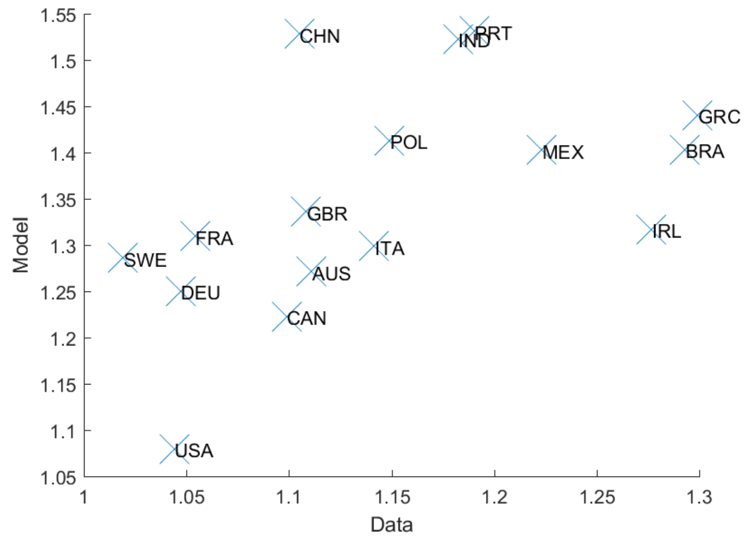
In assessing the model fit, I construct a measure of foreign affiliates' managerial advantage. The measure I use is the ratio between the average innovation management score of foreign affiliates in each country, and the average score of domestic firms in that country. Figure A.2 plots this ratio to its data counterpart for each country. The figure shows that, consistent with the summary statistics in Table 1.2, there is a positive relationship between the model and the data, although the model over predicts the premium.

A.2.2.4 Computational Algorithm

To solve the model and calibrate it to match the data, I use the following calibration algorithm:

1. Choose c^M, c^R, z_H^R, α, κ^P.

Figure A.2: Foreign Affiliate Managerial Advantage



Notes: The vertical axis shows the model-based foreign affiliate innovation managerial advantage, defined as the ratio between the average innovation management score of foreign affiliates in each country and the average score of domestic firms. The horizontal axis shows the empirical counterpart of this ratio, based on the data from Bloom et al. (2012). The correlation between the model and the data is 0.53.

2. Choose $\{\Lambda_l\}$, country-specific production efficiency.

3. Choose $\{\phi^R_{oi}\}$, $\{\phi^P_{il}\}$, $\{\tau_{ld}\}$, bilateral frictions.

4. Solve the model, compare the model-predicted bilateral shares in offshore R&D, offshore production, and trade, to the data. If they are not the same, go back to Step 3, and update international frictions accordingly. Otherwise proceed to the next step.

5. Compare the model-generated GDP with the data. If they are not the same, go back to step 2, and update country-level productivity accordingly.

6. Compute the moments on the firm size distribution, and the extensive margin of offshore R&D and trade, reported in Table 1.2. Compare these moments to their data counterparts. If these moments are not the same with the data,

go back to Step 2. If they are the same, then the calibration procedure is finished.

I solve the model in step 4 of this calibration procedure, using the following algorithm. I start with a guess of aggregate variables $\{X_i\}$, $\{P_i\}$, $\{w_i^P\}$, cutoffs for workers to become researchers, $\{\hat{\theta}_i\}$, and cutoffs for offshore R&D, $\{\hat{z}_{oi}^R\}$. Given the aggregate variables, I solve for the cutoffs to export for each producer-consumer country pairs, $\{\hat{z}_{ld}^P\}$, and the corresponding per-variety profit, $\{\pi_i^P(z^P)\}$. I then use $\{\hat{z}_{oi}^R\}$ to solve for the measure and efficiency distribution of R&D centers in each country. With this information, I solve the researcher labor market equilibrium in each country, finding the researcher wage schedule, $w_i(\theta)$ and the matching function $T_i(z^R)$, which further allows me to solve for the number of varieties in each country, and the productivity for these varieties. The offshore and trade block of the model then determines $\{X_{ild}\}$. Based on $\{X_{ild}\}$, I distribute all revenues from sales to the production workers, researchers, and firm owners from different countries.

I then update the guess $\{X_d\}$ using the current account balance conditions, $\{w_d\}$ using the production labor market clearing condition, and $\{P_d\}$ using the model-implied price indices (Equation 1.12). I also update the guess for occupation choice and offshore R&D based using their respective indifference conditions. I continue this process until the updated aggregate objects and the cutoffs are the same as the input.

A crucial step in solving the model is to solve for the research market equilibrium for each country, characterized by Equations 1.6, 1.7, and 1.8. This is computationally difficult for two reasons. First, due to firms' offshore R&D decisions, the density of R&D center efficiency distribution is discontinuous. This discontinuity leads to kinks in the matching function. Commonly used boundary value problem solver take a long time or, under many parameter values, fail to find the solution. Second, the density of R&D center efficiency distribution depends on the offshore R&D decision of firms from all over the world. Evaluating the R&D density function therefore requires summing over all home countries (Equation 1.10). Similarly, we also need to evaluate $\pi_i^P(z^P)$, which depends on aggregate variables of all countries. The computational burden increases quadratically in the number of countries in the economy.

I solve the first problem by using the "shooting" method, that is, to recast the boundary value problem as a sequence of initial value problems. Specifically, given a wage for the bottom researcher, $w_i(\hat{\theta}) = \underline{w}_i$, Equations 1.6, 1.7 constitute an initial value problem. This problem can be solved by simply forward integrating the two Equations starting from the initial value of wage. I use the Runge-Kutta Cash–Karp method in solving the initial value problems.

Let the solution to the initial problem be $T_i(z^R|\underline{w}_i)$. If $T_i(\overline{z}^R|\underline{w}_i) = \overline{\theta}_i$, then the solution to the initial problem is also the solution to the original boundary value problem. We can therefore search over the initial wage, $w_i(\hat{\theta}_i)$ and solve a sequence of initial value problems until we find the solution to the original problem. Further, as shown in Proposition 4 at the end of this section, $T_i(\overline{z}^R|\underline{w}_i)$ decreases monotonically in $\underline{w}_i$. This feature of the model makes this search efficient and robust.

To further speed up the process, I use the model feature that given all the aggregate variables, the research market equilibrium in each country is independent. I use the OpenMP protocol to parallelize the computation. In solving for each researcher market equilibrium, evaluating $g_i^R(z^R)$ and $\pi_i^P(z^P)$ requires summing over all home countries. I further parallelize this process by using SIMD.

In the following proposition, I prove the monotonicity of $T_i(\overline{z}^R|\underline{w}_i)$ in w_i:

Proposition 4 *Define $T_i(z^R|\underline{w})$ and $w_i(\theta|\underline{w})$ as the solution to the initial problem given by Equations 1.7, 1.6, and initial conditions $w_i(\hat{\theta}_i) = \underline{w}$, $T_i(\underline{z}^R) = \underline{\theta}_i$. Then the end value of the solution to this initial problem, $T_i(\overline{z}_i^R|\underline{w})$, decreases in $\underline{w}$.*

Proof Consider two wages for the bottom researcher in country i, $w^1 < w^2$. This proposition claims that $T_i(\overline{z}_i^R|w^1) \geq T_i(\overline{z}_i^R|w^2)$. I prove by contradiction.

Suppose $T_i(\overline{z}_i^R|w_1) < T_i(\overline{z}_i^R|w_2)$. Given that $w^1 < w^2$, Equation 1.7 implies that at $T_i'(\underline{z}^R|w_1) > T_i'(\underline{z}_i^R|w_2)$, that is, at least initially at $\underline{z}_i^R$, when facing a lower wage w_i, R&D centers will hire a larger number of researchers. This means that at an ϵ interval to the right of $\underline{z}_i^R$, $T_i(z^R|w_1) > T_i(z^R|w_2)$. Since $T_i(\overline{z}_i^R|w_1)$ and $T_i(\overline{z}_i^R|w_2)$ are both continuous function of z^R, for $T_i(\overline{z}_i^R|w_1) \leq T_i(\overline{z}_i^R|w_2)$ to hold, there must be at a point $\tilde{z}^R$, such that $T_i(z^R|w_2)$ crosses $T_i(z^R|w_1)$ at $\tilde{z}^R$ for the first time from below. Suppose $\tilde{\theta} = T_i(\tilde{z}^R|w_1) = T_i(\tilde{z}^R|w_2)$. From Equation 1.7, $w_i(\tilde{\theta}|w_1) > w_i(\tilde{\theta}|w_2)$.

From Equation 1.6, $w_i(\theta|\underline{w}_i) = \underline{w}_i \exp \int_{\hat{\theta}_i}^{\theta} \frac{f_2(T_i^{-1}(x),x)}{\gamma f(T_i^{-1}(x),x)} dx$. Under the log supermodularity assumption of function f, the integrant on the right hand side increases with z^R. Because $\tilde{z}^R$ is the first point where the two matching functions intercept, for all $\theta \in (\hat{\theta}_i, \tilde{\theta})$, $T_i^{-1}(\theta|w_1) < T_i^{-1}(\theta|w_2)$. Therefore $w_i(\tilde{\theta}|w_1) < w_i(\tilde{\theta}|w_2)$, which contradicts the above result.

Appendix B: Chapter 3 Appendix

This section provides additional theoretical and empirical materials on the third chapter.

B.1 Theory Appendix

B.1.1 Deriving Equation (3.9)

$$
\begin{aligned}
\pi^e_{o,d} &= \Pr(\frac{v^e_d z_d}{d_{o,d}} \geq \frac{v^d_g z_g}{d_{o,g}}, \forall g \in G) \\
&= \Pr(z_g \leq \frac{\frac{v^e_d}{d_{o,d}}}{\frac{v^e_g}{d_{o,g}}} z_d, \forall g \in G,) \\
&= \int_0^\infty \Pr(z_g \leq \frac{\frac{v^e_d}{d_{o,d}}}{\frac{v^e_g}{d_{o,g}}} z_d, \forall g \in G | z_d) f(z_d) dz_d \\
&= \int_0^\infty F_d(\frac{\frac{v^e_d}{d_{o,d}}}{\frac{v^e_1}{d_{o,1}}} z_d, \frac{\frac{v^e_d}{d_{o,d}}}{\frac{v^e_2}{d_{o,2}}} z_d, \ldots \frac{\frac{v^e_d}{d_{o,d}}}{\frac{v^e_g}{d_{o,g}}} z_d, \ldots) dz_d,
\end{aligned}
$$

Where $F_d(\frac{\frac{v^e_d}{d_{o,d}}}{\frac{v^e_1}{d_{o,1}}} z_d, \frac{\frac{v^e_d}{d_{o,d}}}{\frac{v^e_2}{d_{o,2}}} z_d, \ldots \frac{\frac{v^e_d}{d_{o,d}}}{\frac{v^e_g}{d_{o,g}}} z_d, \ldots) := \frac{dF}{dz_d}|_{z_g = \frac{\frac{v^e_d}{d_{o,d}}}{\frac{v^e_g}{d_{o,g}}} z_d, \forall g \in G}$ is the probability that the draw from region d is z_d *and* this draw dominates all other draws.

Use the functional form of F, it follows that

$$\pi^e_{o,d} = \int_0^\infty \exp(-(\sum_{g\in G}(\frac{\frac{v^e_d}{d_{o,d}}}{\frac{v^e_g}{d_{o,g}}}z_d)^{-\epsilon_e})^{1-\rho}) * (1-\rho)\epsilon_e(\sum_{g\in G}(\frac{\frac{v^e_d}{d_{o,d}}}{\frac{v^e_g}{d_{o,g}}}z_d)^{-\epsilon_e})^{\rho} z_d^{-\epsilon_e-1} dz_d$$
$$= \frac{\int_0^\infty d\exp(-(\sum_{g\in G}(\frac{\frac{v^e_d}{d_{o,d}}}{\frac{v^e_g}{d_{o,g}}}z_d)^{-\epsilon_e})^{1-\rho})}{\sum_{g\in G}(\frac{\frac{v^e_d}{d_{o,d}}}{\frac{v^e_g}{d_{o,g}}})^{-\epsilon_e}}$$
$$= \frac{(\frac{v^e_d}{d_{o,d}})^{\epsilon_e}}{\sum_{g\in G}(\frac{v^e_g}{d_{o,g}})^{\epsilon_e}}$$

B.1.2 Deriving Equation (3.11)

We first derive the distribution of u^e_o, $u^e_o = \max_{d\in G}\{\frac{v^e_d z_d}{d_{o,d}}\}$,

$$\begin{aligned}
F_{u^e_o}(u) :&= \text{Prob}(u^e_o \le u) \\
&= \text{Prob}(\frac{v^e_d z_d}{d_{o,d}} \le u, \quad \forall d \in G) \\
&= \text{Prob}(z_d \le \frac{u d_{o,d}}{v^e_d}, \quad \forall d \in G) \\
&= F(\frac{u d_{o,1}}{v^e_1}, \frac{u d_{o,2}}{v^e_2}, ..., \frac{u d_{o,d}}{v^e_d}, ...) \\
&= \exp(-[\sum_{d\in G}(\frac{u d_{o,d}}{v^e_d})^{-\epsilon_e}]^{1-\rho}) \\
&= \exp(-[\sum_{d\in G}(\frac{d_{o,d}}{v^e_d})^{-\epsilon_e}]^{1-\rho} u^{-(1-\rho)\epsilon_e}) \\
&= \exp(-\Phi^{e\,1-\rho}_o u^{-(1-\rho)\epsilon_e})
\end{aligned}$$

It can be shown that, $\forall d \in G$, the cumulative distribution function of u for workers moving from o, to d, is

$$F_{u^e_{o,d}}(u) = F_{u^e_o}(u) = \exp(-\Phi^{e\,1-\rho}_o u^{-(1-\rho)\epsilon_e}),$$

which is a Frechet distribution with position parameter $\Phi^{e\,1-\rho}_o$ and dispersion parameter $(1-\rho)\epsilon_e$.[1]

$$\begin{aligned}
E(u^e_o|L_{o,d}) &= \int u dF_{u^e_o}(u) \\
&= \int u d(\exp(-[\sum_{d\in G}(\frac{d_{o,d}}{v^e_d})^{-\epsilon_e}]^{1-\rho}u^{-(1-\rho)\epsilon_e})) \\
&= \int u\epsilon_e(1-\rho)\exp(-[\sum_{d\in G}(\frac{d_{o,d}}{v^e_d})^{-\epsilon_e}]^{1-\rho}u^{-(1-\rho)\epsilon_e})[\sum_{d\in G}(\frac{d_{o,d}}{v^e_d})^{-\epsilon_e}]^{1-\rho}u^{-(1-\rho)\epsilon_e}du \\
&= \int u\epsilon_e(1-\rho)\exp(-\Phi^{e\,1-\rho}_o u^{-(1-\rho)\epsilon_e})\Phi^{e\,1-\rho}_o u^{-(1-\rho)\epsilon_e}du \\
&= -\int \epsilon_e(1-\rho)\exp(-y)y du \quad (\text{change of variable} : y = \Phi^{e\,1-\rho}_o u^{-(1-\rho)\epsilon_e}) \\
&= -\int \epsilon_e(1-\rho)\exp(-y)y d(\frac{y}{\Phi^{e\,1-\rho}_o})^{\frac{1}{-(1-\rho)\epsilon_e}} \\
&= \int \exp(-y)y^{-\frac{1}{(1-\rho)\epsilon_e}}\Phi^{e\,\frac{1}{\epsilon_e}}_o dy \\
&= \Phi^{e\,\frac{1}{\epsilon_e}}_o\Gamma(1-\frac{1}{\epsilon_e(1-\rho)}) \quad (\text{Definition of Gamma function})
\end{aligned}$$

B.1.3 Deriving Equation (3.35)

For workers staying in their hometown, $u^e_{o,o} = \frac{v^e_o z_o}{d_{o,o}} = v^e_o z_o$, hence the distribution of productivity draws for workers choosing to stay in o is:

[1]This is obtained by showing $F_{u^e_{o,d}}(u) := \text{Prob}(u^e_{o,d} \leq u | u^e_{o,d} \text{is the highest}) = \frac{\text{Prob}(u^e_{o,d}\leq u, u^e_{o,d}\text{is highest})}{\pi^e_{o,d}} = \frac{\int_0^u F_d(z_d)dz_d}{\pi^e_{o,d}} = F_{u^e_o}(u)$.

$$
\begin{aligned}
F_{z^e_{o,o}}(z) :&= \Pr(z^e_{o,o} < z) \\
&= \Pr(\frac{u^e_{o,o}}{v^e_o} < z) \quad \text{(using } d_{o,o} = 1) \\
&= F_{u^e_{o,o}}(zv^e_o) \\
&= \exp(-[v_o^{-(1-\rho)\epsilon_e} \mathbf{\Phi}_o^{e\,1-\rho}] z^{-(1-\rho)\epsilon_e}),
\end{aligned}
$$

which is also a Frechet distribution. For different regions, the productivity distribution of stayers there have different means, but their dispersions will be the same. Therefore, I regress stayers' log wages on regional fixed effects to net out the different average regional productivity draws and interpret the exponents of the residuals as random draws from a Frechet distribution with dispersion parameter $\epsilon_e(1-\rho)$. The coefficient of variations for this distribution is given by Equation (3.35).

B.1.4 Proposition 1

Proposition 1 is used in Section 3.2 of this appendix, in estimating migration costs.

Proposition 5 *Given migration costs $\{d_{o,d}\}$, there exists a unique set of $\{v_d\}$ (up to normalization), such that the model-predicted number of workers employed in each region equals that in the data, i.e., $L^e_d = \sum_{o \in \boldsymbol{G}} \pi^e_{o,d} l^e_o$ is satisfied, where L^e_d is the number of workers working in d (data), l^e_o is the number of workers born in o (data), and $\pi^e_{o,d}$ is the model-predicted probability of workers born in o to move to d.*

Proof The proof follows Michaels et al. (2011) and Lemma 1, Lemma 2 in Ahlfeldt et al. (2015), so I only sketch the key steps here.

Consider Equation (3.10) in the text

$$L_d^e = \sum_{o \in \mathbf{G}} \pi_{o,d}^e l_o^e,$$

Where L_d^e and l_o^e are data, and $\pi_{o,d}^e = \frac{(\frac{v_d^e}{d_{o,d}})^{\epsilon^e}}{\sum_{g \in \mathbf{G}} (\frac{v_g^e}{d_{o,g}})^{\epsilon^e}}$. Given $\{d_{o,d}\}$, l_o^e, and L_d^e, the only unknowns in this equation is $\{v_d^e\}$. Let $\mathbf{v}^e$ be the vector $(v_1^e, v_2^e, ..., v_d^e, ...)$. Define $\mathrm{WD}(\mathbf{v}^e)$ (worker deficits) as

$$\mathrm{WD}(\mathbf{v}^e) = L_d^e - \sum_{o \in \mathbf{G}} \pi_{o,d}^e l_o^e.$$

WD is simply the gap between the number of workers working in region d in the data, and the number predicted by the model. $\mathrm{WD}(\mathbf{v}^e)$ is a function of $\mathbf{v}^e$. To prove Proposition 1 we show the following:

1. $\mathrm{WD}(\mathbf{v}^e)$ is continuous;
2. $\mathrm{WD}(\mathbf{v}^e)$ is homogeneous of degree zero;
3. $\sum_{d \in \mathbf{G}} \mathrm{WD}_d(\mathbf{v}^e) = 0, \forall \mathbf{v}^e \in \mathbf{R}_+^G$
4. $\mathrm{WD}(\mathbf{v}^e)$ exhibits gross substitute property.

It is easy to verify that requirement (1) and (2) are satisfied. Requirement (3) can be shown to be satisfied by noting that $\sum_{d \in G} \pi_{o,d}^e = 1$; requirement (4) can be shown to be satisfied by computing the derivatives directly.

Requirements (1)–(2) guarantee the existence of a solution. The proof is a constructive one: by homogeneous of degree zero, we can normalize $\mathbf{v}^e$ to the simplex $\{\mathbf{v}^e \in R_+ : \sum \mathbf{v}^e = 1\}$. Define $\mathbf{WD}^+ = \max\{0, \mathbf{WD}\}$, and $\mathbf{f}(\mathbf{v}) = \frac{\mathbf{v}+\mathbf{WD}^+}{\sum_{d\in\mathbf{G}} v_d + \sum_{d\in textbfG} \mathbf{WD}(v)_d}$, then $\mathbf{f}$ is a continuous function mapping the unit simplex onto itself. The existence of a solution to $\mathbf{v} = \mathbf{f}(\mathbf{v})$ then follows from the Brouwer's fixed point theorem.

Requirement (3)-(4) then guarantee the uniqueness of the solution, see Ahlfeldt et al. (2015) for a more detailed explanation.

The implication of proposition 1 is that, given migration costs, we can solve Equation (3.10) for the unique set of amenity-adjusted real wages for all locations.

B.2 Background Information and Data Appendix

B.2.1 Background Information on the Chinese Hukou System

The Hukou system, dating back to the 1950, is the household registration system in China. It was originally established to control the rural-urban migration in China. Back then, residents in cities were subsidized with downward-distorted prices for agricultural products, so there was a strong incentive for people to live in cities. There are two types of Hukous, one for rural residents, the other for urban residents, in each city.[2] Before the 1980's, people were tied to where their Hukous were and were not allowed to move to any other places without permission from the authority. As a result, there were only minimum rural-urban or urban-urban

[2] Therefore an urban Hukou in Beijing is different from either an urban Hukou in Shanghai or a rural Hukou in Beijing

migrant workers.

Although people are free to travel now, the Hukou system is still important for many aspects of life, as it determines the eligibility for health care, social insurance, housing, and education, etc. In many aspects, it acts like within country visa system, distorting the free mobility of labor.[3] Therefore, a policy change that makes it easier for migrants to obtain local Hukou still makes a material difference for their quality of life.

I construct the database for Hukou reforms following the approach by Sun et al. (2011) and Kinnan et al. (2015). The primary source of information I rely on is one of the most comprehensive online law libraries, Peking University Law Information Database (`http://www.pkulaw.cn/`), and the webpage of the official news agency for the communist party (`http://www.xinhuanet.com/`). On these two databases, I search for the following keywords: first, any combinations of "hukou" or "huji"(also means Hukou) with "gaige" (reform) or "guanli" (management), which are the keywords used in Kinnan et al. (2015). I supplement these keywords with the following words: "chengzhenhua"/"chengshihua" (both mean urbanization) and "luohu"/"ruhu" (both mean granting Hukou to someone). These additional keywords expand the number of policy changes recorded by around 40%.

I read the news articles and law documents about the Hukou policy from the keyword search, and rate them on a scale of 0–6. A 0 means a strict control on Hukou, with virtually no room for mechanical growth (new Hukou due to migra-

[3]See the May 6th, 2010 issue of the magazine The Economist, available at http://www.economist.com/node/16058750, for more information about the Hukou system in China.

tion), while a 6 means an open-door policy that grants Hukou to anyone with legal residence and a job in a city. The rating is done in the following way. First, I separate a policy in terms of its geographic coverage within a city. Some policies apply only to outside the central district of a city, while others apply to the central district as well, so I rate policies on two sub scores based on their influence on the central districts and other parts of a city separately. Each sub score takes a score of 0–3. The reform index will be the sum of two sub scores, so it takes a value of 0–6.

To evaluate each sub score, I focus on the policies that are relevant to a substantial part of population, so those that only apply to advanced degree holders with overseas experience or high-tech entrepreneurs are excluded. Within the remaining policies, the details of the eligibility requirements differ, but most of the requirements are based on the following criteria: 1) a migrant's job prospect and job stability, 2) his/her residential condition, and 3) his/her history of contribution to local social security. I give a value of 0 to cities that grant virtually no Hukou to migrants other than those initiated by public-sector employers. I give a value of 1, if a migrant can obtain Hukou by purchasing an apartment above certain size or value. I give a value of 2, if a migrant can obtain Hukou by purchasing an apartment (or renting a subsidized apartment from his/her employer), with no specific requirement on its values, or by working and contributing to the social security by more than 5 years. I give a value of 3, if a migrant can obtain Hukou by working and residing in a city, and contributing to the local social security for a relatively short period of time.

Importantly, in Chinese cities, the migration into the central districts are controlled much more strictly than into the outskirts. When I see a policy mentioning only relaxing Hukou restriction in its central district, with no explicit statement on the policy for the other parts of the city, I take the stand that such policy implicitly apply to the entire city.

B.2.2 Sources of Additional Data and Sample Construction

The primary individual- and firm-level data I use are the following: the 2005 Mini Population Census, the 2000 Population Census, the 2004 Economic Census, and the 2004 Annual Survey of Industrial Production. In addition to these micro data sources, I also use the 2002 inter-regional and inter-sectoral input-output table, and the data from national accounts and provincial statistical yearbooks.

The 2005 Mini Population Census covers 1% of Chinese population. It records individual demographic and employment information. To my knowledge, this is the only data set that provides individual-level income information for the entire country, so I use it to estimate the average income in each region. I also choose 2005 as the benchmark year, as the calibration procedure requires wage information. The sample I use in this paper is a 1% sub-sample of this data set.

The 2000 Population Census covers the entire Chinese population. My sample is its 0.095% sub-sample. Respondents in this sub-sample fill a longer form than others, which asks for information on migration, education, occupation, industry, and housing conditions, but unfortunately, not for information on income.

The 2004 Economic Census covers the universe of registered firms. The sample I have access to is its manufacturing sub-sample, with firm-level revenue and employment information.

The 2004 Annual Survey of Industrial Production covers all state-owned enterprises, as well as private enterprises with annual sales over 5 million RMB yuan. Different from the 2004 Economic Census, this data set contains detailed firm-level financial information, rather than only employment and revenue information.[4]

The rest of this section covers details in sample construction.

B.2.2.1 Wage

There are two types of workers, two types of local labor markets (rural and urban), and N cities in the economy, so in total there are 4N wages (mean wages for skilled and unskilled workers in all regions in the economy) to estimate. The data I use for this purpose is the 2005 mini census.

I estimate the following specification:

$$\begin{aligned} log(\mathrm{Wage}_{e,i}) = & \beta_0 + \beta_1 \mathrm{age} + \beta_2 \mathrm{age}^2 + \beta_3 \mathrm{sex} + \beta_4 I_{\mathrm{Skilled}} * I_{\mathrm{Agriculture}} \\ & + \mathrm{F}_i + \mathrm{S}_i * \mathrm{F}_i I_{\mathrm{Skilled}} + \mathrm{A}_i * \mathrm{F}_i I_{\mathrm{Agriculture}} + \epsilon_{e,i}, \end{aligned}$$

where F_i is the regional fixed effect, $\mathrm{F}_i * I_{\mathrm{Skilled}}$ is the interaction between regional fixed effect and high-skill dummy, and $\mathrm{F}_i * I_{\mathrm{Agriculture}}$ is the interaction between regional fixed effect and a dummy for agricultural sector. In this specification, I restrict the relative skill premium in the agricultural sector (relative to the skill

[4] The 2004 Economic Census also covers detailed financial information, but I do not have access to other variables.

premium in urban sector of the same city) to be the same across cities (β_4 is not city-specific). This choice is constrained by the power of the regression, as in the sample, in many cities, the rural sector only employee a small number of high-skill workers. The omitted group in the regression is the unskilled worker in the urban sector in Beijing, whose average wage is β_0. Average wages for other groups of workers can be calculated as follows:

Table B.1: Average wage for different groups

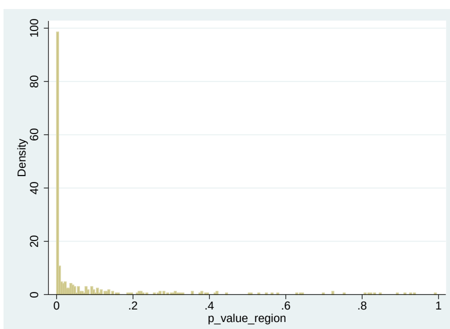
Education	Sector	Region	Wage
Unskilled	Urban	i	$\beta_0 + F_i$
Unskilled	Rural	i	$\beta_0 + F_i + A_i$
Skilled	Urban	i	$\beta_0 + F_i + S_i$
Skilled	Rural	i	$\beta_0 + \beta_4 + F_i + S_i + A_i$

Table B.2: Wage Regressions

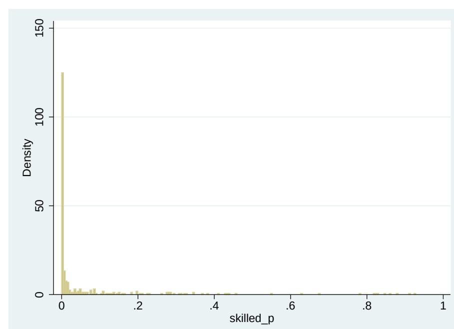
	(1) log_wage
Age	0.0327***
	(22.32)
Age_square	-0.000413***
	(-22.70)
Sex	-0.206***
	(-42.36)
Skilled_agri	-0.296***
	(-16.57)
Observations	62138
R^2	0.576

t statistics in parentheses
* p<0.10, ** p<0.05, *** p<0.01

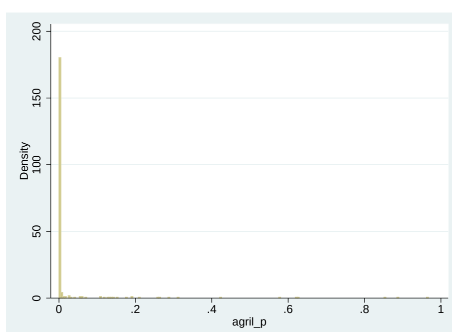
The output of the regressions are presented in Table (B.2). The signs and magnitudes of coefficients are reasonable. The R^2 of the regression is 0.58, indicating that the regression has a strong explanatory power. Figure (B.1) presents the distribution of the p-values for the fixed effects in the wage regression. The distribution is heavily concentrated around zero (the spike in the figures corresponds to p-value<0.0005), suggesting that the fixed effects are very precisely estimated. Figure (B.2) shows the distribution of average wage for different worker groups across regions. Two patterns emerge: first, there is considerable heterogeneity across regions; second, overall, wages are higher for high-skill workers and urban workers. Figures 3.1c in the text cast the estimates for average wages of workers on the map of China. The dispersions in wages shows up on the map as the difference both across and within geographic areas.



(a) Regional Fixed Effects



(b) Skilled*Regional Fixed Effects

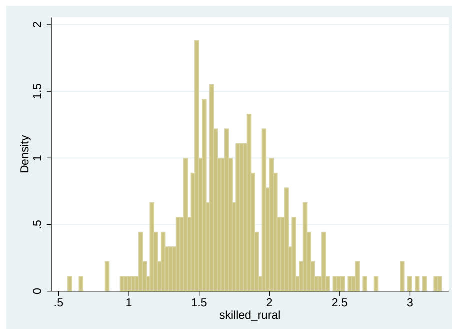


(c) Agriculture*Regional Fixed Effects

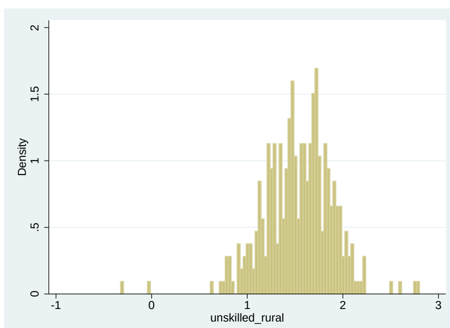
Figure B.1: Distribution of the P-value for Fixed Effects

Figure B.2: Average Wages for Different Worker Groups

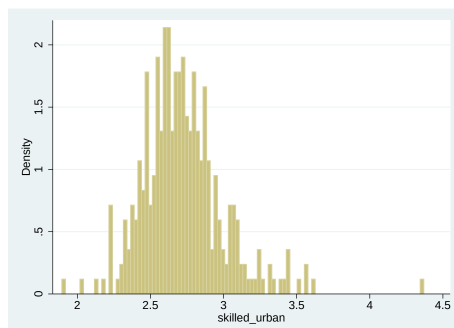
(a) Rural high-skill workers



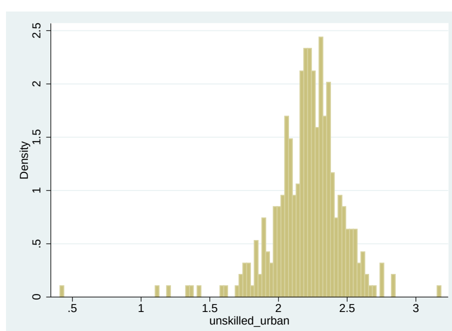
(b) Rural low-skill workers



(c) Urban high-skill workers



(d) Urban low-skill workers



B.2.2.2 Migration

Since I use the 2005 mini census to estimate regional wage and calibrate the model to the 2005 economy, ideally I would like to use this data set to estimate migration costs, too. Since the model neglects dynamic choice of individuals, the migration decision in the model should be best interpreted as a life-time choice. So the model-consistent definition of migration is one that is based on birthplace. However, the 2005 data does not cover birthplace information, so I use the 2000 census to estimate the long-run migration costs.[5] The underlying assumption is that the long-run migration costs do not change much over the period of 2000-2005. It is of course possible that some migration restrictions have been lifted during the period; in that case, the counterfactual experiments in the paper should be interpreted as: what are the welfare implications of international trade for China in 2005, had the migration costs stayed at the 2000 level.

The following are the procedures I use to construct migration flow: first of all, I restrict the sample to those who already finished their schooling, aged between 20 and 60 (60 is the official retirement age for urban male non-physical-labor workers in China), I also drop those who are currently not working, unless the reason for not working is either "on vacation" or "on sick leave". I classify a worker as a

[5] The mini census does report the place of residence in 2000. Therefore one alternative is to combine the migration over the period of 2000-2005 with the long-term migration in 2000, to construct the long-term migration in 2005. This is problematic, as a large fraction of the workers that migrated during 2000–2005 might had been already living outside their birthplace in 2000, i.e., they are repeat migrants. Empirical studies focusing in the U.S. have documented the phenomena of repeat migrant or return migrant (Kennan and Walker, 2011), and the fact that migrants are more likely to respond to economic shocks by migrating, than native workers (Cadena and Kovak, 2013). In light of the evidence, this approach will double count return migrants and repeat migrants, overestimating the long-term migration in 2005.

migrant, if he or she is not working in her or his birthplace. I identify the source sector (rural or urban) of a worker with the type of Hukou (rural or urban) the worker currently holds, and the destination sector of a worker by the locality the survey respondent.[6] Given the small proportion of workers with college degrees in China in 2000, I classify a worker to be high-skill, if he or she has received more than nine years' formal education, equivalent to finishing junior high school.[7] From these procedures, for all workers in the economy, I identify their education level, source province, source sector, destination city, and destination sector. I use this to estimate inter-regional and inter-sectoral migration costs.

B.2.2.3 Worker Employment and Birthplace Distributions in 2005

Recovering $\{v_d^e\}$ After estimating the parameters governing migration costs, I solve the labor market clearing conditions (Equation 3.10 in the text) for one more time, to obtain $\{v_d^e\}$ for 2005, the regional fixed effects that are consistent with employment distribution in 2005. For this purpose, I need workers' birthplace and employment distributions in 2005, by workers' level of skills.

I construct the employment distribution from the 2005 mini census. For some cities, due to the small sample size and the small share of skilled workers, there are few skilled workers sampled. For these cities, I supplement the employment

[6] To the extent that some rural Hukou holders have switched an urban Hukou in 2000, this classification underestimate rural-urban migration. However, until recently, switching a rural Hukou for an urban one was highly restricted.

[7] The higher education reform started in 1999 in China, which expanded the scale of the higher education sector dramatically. Before the reform, the college admission rate in China was below 5%; in 1999, the college admission increased by 40%. The following years saw additional increase. But until 2005, college graduates constitute only a small proportion of the Chinese labor market.

distribution aggregated up from the micro data with the published aggregate city-level statistics on employment from the same survey.

I construct workers' birthplace distribution from the 2000 census. I restrict the sample to workers aged 15–55 in 2000. The distribution of this sample will be the distribution for workers aged 20–60 in 2005. To determine the skill level of workers for this sample, if a worker has finished schooling in 2000, I classify his or her skill level based on the education attainment directly; for workers that are above 15, but have not yet finished schooling, I assume they are skilled—by this age, a typical Chinese kid has received 8-9 years of education, so the possibility of (wrongly) classifying a student receiving less than 9 years education as skilled is minimized.

Recovering $\{T_d^s\}$ The employment distribution constructed above gives us the *number of workers* employed in each region. Once we have the estimates for migration costs and regional amenity-adjusted real wages, we can use Equation (3.13) in the text to convert these into the employment of *effective labor units*. Since there are three industries in urban regions, we still do not know the distribution of employment across industries in each urban region, which is needed for the calibration of productivity at city-industry level.[8] I supplement the regional employment information with the share of employment in industry K over industry M, constructed from the manufacturing sub-sample of the 2004 economic census, and use the service mar-

[8] In the main text, I analyze the intuition behind the quantification strategy in the context of a linear-regression setup, where we need the trade flows between cities for estimation. Such data is not available, so I use a joint quantification strategy, discussed in section 3.2 of this appendix, for which I need employment distribution in each city-industry to determine the corresponding productivity.

ket clearing conditions to obtain the employment information at the city-industry level.[9]

Specifically, let $E^h_{d,s}$ and $E^l_{d,s}$, $s \in \{A, M, K, S\}, d \in \{\mathbf{U},\ \mathbf{R}\}$ be the sectoral effective labor unit employment, then regional labor market clearing conditions are:

$$\begin{aligned} &E^h_{d,A} = E^h_d,\ E^l_{d,A} = E^l_d,\ d \in \mathbf{R} \\ &E^h_{d,M} + E^h_{d,K} + E^h_{d,s} = E^h_d,\ E^l_{d,M} + E^l_{d,K} + E^l_{d,S} = E^l_d,\ d \in \mathbf{U} \end{aligned} \tag{B.1}$$

The right sides of these equations are already constructed from the data. Since only agricultural industry is located in rural regions, from the above equation we know labor effective unit employment in the agricultural industry.

From the optimality conditions of intermediate variety producers, given by Equation (3.19), the production of intermediate varieties in each place can be calculated, and this should equal to the total demand, D^s_d:

$$\begin{aligned} D^A_d &= \frac{E^h_{d,A} W^h_d}{\beta^h_d \gamma^L_A} = \frac{E^l_{d,A} W^l_d}{\beta^l_d \gamma^L_A},\ d \in \mathbf{R} \\ D^s_d &= \frac{E^h_{d,s} W^h_d}{\beta^h_d \gamma^L_s} = \frac{E^l_{d,s} W^l_d}{\beta^l_d \gamma^L_s},\ s \in \{M, K, S\},\ d \in \mathbf{U}, \end{aligned} \tag{B.2}$$

With $\{D^s_d : s \in \{A, M, K, S\}\}$ we can compute the city-level demand for industry final output in the service sector, which must equal D^S_d,

$$D^S_d = C^S_d + C^S_{d'} + + D^A_{d'} \gamma^S_A + \sum_{s \in \{M,K,S\}} D^s_d \gamma^A_s, \quad d \in \mathbf{U}, \tag{B.3}$$

[9]I do not directly use the 2005 mini census to construct industry-level employment because due to the limited sample size, in some cities, there are no or only a small number of high-skill employment in the capital and equipment industry.

where C_d^S is the urban service consumption in region d; d' indicates the rural region in the same city as urban region d and $C_{d'}^S$ is the service consumption of this rural region. $C_d^S + C_{d'}^S$ is determined directly by workers' wage and employment distribution. Combine Equations (B.1), (B.2) and (B.3),[10] we have a linear equation system, with 4N unknowns: $E_{d,A}^h$, $E_{d,M}^h$, $E_{d,K}^h$, $E_{d,s}^h$, and 3N equations—(B.3) and the subset of (B.1) for high skilled workers. We combine these three equations with one more data moment—regional employment share in capital and equipment (K) versus other manufacturing industries (M), $\frac{E_{d,K}^h}{E_{d,M}^h}$ to solve for employments of effective labor units in all city-industry.

Once we obtain these employments, we can also use Equation (B.2) to compute the production of intermediate varieties in each industry in all cities.

B.2.2.4 Factor Shares in Equipped Composite Labor

We need the shares of payments to capital, high-skill workers, and low-skill workers in each region, to calibrate the region-specific equipped composite labor production functions. I compute the ratios between payments to high-skill workers over low-skill workers directly from the estimated wages and the distribution of effective labor units, both of which have been constructed previously. I further need the ratio between the payment to capital, and the payment to labor, in each region.

For the urban regions, I use the 2004 Survey of Industrial Production. I aggregate firm-level data to obtain the city-level ratio between wage bill and expenditures on capital and equipment. The firm-level wage bill is the "total salary payments"

[10]We use Equation (B.2) to eliminate $E_{d,s}^l$

entry in the data set; the firm-level expenditures on capital and equipment is the "total capital depreciations" entry in the data set. The total depreciations entry includes, in addition to depreciations to capital and equipment, depreciations to properties and buildings. Therefore I adjust for this by subtracting the share of buildings among aggregate tangible fixed capital stock in China in 2004, calculated from the national statistical yearbook. The mean ratio across cities, constructed this way, is similar to the corresponding ratio from the national input-output table for the urban sector.

For the rural regions, since I am not aware of any data sources that contain information on capital share at the regional level, I assume the capital shares are the same for all rural regions and use the national input-output table to determine it.

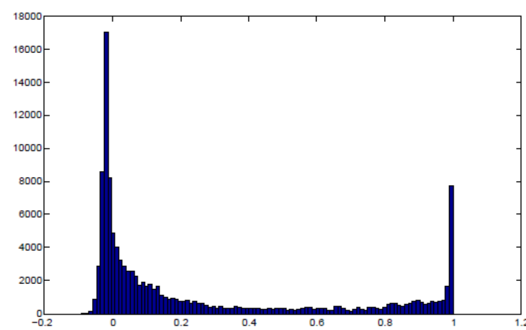
B.2.2.5 Cultural Distance

To proxy for the cultural distance between cities, I construct a cultural similarity index based on the compositions of ethnic minority groups. I extract the prefecture-level information on the compositions of ethnic minorities from the 1990 census. Migrations was not as pervasive in 1990 as it was in 2000, and therefore the ethnic compositions largely reflect the cultural root of a city. Using the 1990 census data helps us avoid the endogeneity problem that would arise, if we used the 2000 census to construct cultural distance.

There are 56 ethnic groups in China, with Han ethnic being the dominating one. I exclude it, because the share of Han population is so large that including it eliminates most of the variation in the similarity index. For each city, I am left with a 55 by 1 vector, each element of which is the share of one ethnic group in the total local ethnic minority population. I then compute the correlations between the vectors of all city pairs, and use these as the values of my cultural similarity index; the cultural distance is then defined as one minus this similarity index.

Figure (B.3) is the density distribution of the index. The mean, median and standard deviation of the similarity index are 0.2569, 0.0608, and 0.3645, respectively.

Figure B.3: Density Distribution of the Similarity Index



Source: Author's calculation based on the 1990 census

B.2.2.6 City-level International Trade Surplus

To incorporate international trade imbalances into the calibration, I construct a data set of city-level international trade surplus.

Each city's trade surplus in 2005 is extracted directly from the provincial statistical yearbook. I make two more adjustments. First, Beijing trades a lot with the ROW, but the majority of the trade is done by big companies (especially those

SOEs) with headquarters in Beijing. It is plausible that the trade is actually carried in the subsidiaries of these companies, spread out over the country. Fortunately, Beijing statistical yearbook reports “local trade” and “total trade” separately, the later including trade done by SOEs. I assign “local trade” to Beijing, and the remaining component of “total trade” to all Chinese cities, based on their relative size. The implicit assumption is that the operation of those SOEs headquartered in Beijing are distributed across all cities, proportionally to their size.

Second, sometimes the data is not well-behaved. For example, for Shaoshan, a city in Guangdong Province, one of the coastal provinces, the trade surplus is 13 times of its GDP. My conjecture is that there are many trade intermediaries. I make the following adjustments: I aggregate city-level trade surplus to the province level, and then allocate the trade surplus of a province to the cities in the province, according to the GDP of these cities. The underlying assumption is that those trade intermediaries mostly work with other companies in the same province, and trade surplus is proportional to size of economy within a province.

To determine the city-level trade surplus in the scale of the model economy, I first calculate the aggregate trade surplus from the data. I convert the aggregate surplus into the scale of the model and distribute it to all cities, proportionally to each city’s contribution to the aggregate trade surplus in the data, constructed above. These are the surplus terms, S_d, in Equation (3.34).

B.2.2.7 Input-output Linkages for China and the ROW

In the model, the input-output parameters for China are constructed from the 2002 national input-output table, which records, at the 2-digit industry level, the usages of inputs in the economy. I aggregate the data to four industries—agricultural, capital and equipment, other manufacturing, and service, and four inputs—industry final outputs in the agricultural, other manufacturing, and service industries, as well as equipped composite labor.

The input shares of the ROW are assume to be the same as the median country in Parro (2013). Since the industry classification is finer in this paper, for values not directly available in Parro (2013), I use the corresponding value from China, scaled appropriately. The underlying assumption behind this imputation that, input-output linkages are similar across different countries, are strongly supported by Iones (2013). All results in the paper are robust to changes in the input shares.

Table (B.3) report the shares of inputs in each industry.

B.3 Estimation and Calibration Appendix

B.3.1 Calibrating ρ

I obtain an individual panel data from China (China Nutrition and Health Survey), and estimate a Mincer regression with regional fixed effects, along with gender, education, age, and age square as control variables. I then add individual

Table B.3: Input Shares in China and the ROW

$\gamma_s^{s'}$	Output Industry: China			
Input	A	M	K	S
L	0.57	0.30	0.59	0.48
A	0.19	0.07	0.00	0.03
M	0.15	0.44	0.26	0.21
S	0.09	0.20	0.16	0.28

$\gamma_s^{s'}$	Output Industry: ROW			
Input	A	M	K	S
L	0.58	0.42	0.56	0.63
A	0.19	0.00	0.00	0.00
M	0.16	0.41	0.26	0.11
S	0.07	0.17	0.18	0.26

Notes: This table reports the input shares for different industries in China and the ROW. The source of the values for China is the national input-output table for 2002; the values for the ROW are calculated based on Parro (2013). *L* stands for the equipped composite labor.

fixed effects to the specification. I compare the R^2 of these two regressions and see how much of the variation unexplained in the first Mincer regression is explained by the individual fixed effects.

As it turns out, about 70% of the unexplained variations can be explained by individual fixed effects. Note that the correlation parameter, ρ, maps one-to-one into the explanatory power of individual fixed effects in the wage regression. For each given value of ρ, I simulate workers' productivity draws from different locations, then estimate a regression specification with *only* individual fixed effects, and calculate the R^2. I chose the correlation parameter so that this R^2 is 70%. This procedure determines a value of 0.4 for ρ.

B.3.2 Estimating Migration Cost

I use nonlinear least squares to estimate the migration cost, in which $\{\beta\}$ is determined by minimizing the difference between the model-predicted migration flows and their data counterparts. Since the data is at the province-to-city level, I aggregate the predicted city-to-city flows to province-to-city level and take as the objective function the sum of square of the differences between the model's predictions and the data.

Formally, let $p \in \mathbf{P}$ indexes a province in the set of all provinces, $\mathbf{P}$, and $o \in p$ indexes a region o belonging to province p. Recall that l_o^e is the number of workers born in o, and $\pi_{o,d}^e$ is the model-predicted probability for workers to move from o to d, then $l_o^e \pi_{o,d}^e$ is the model-predicted flow from o to d and $\sum_{o \in p} l_o^e \pi_{o,d}^e$ is the aggregate flow from *province* p to *region* d. Let $L^e_{p,d}$ be the flow from p to d in the data, the estimation problem can be formulated in the following way:

$$\min_{\{\beta\}} \sum_{p\in P, d\in G} (\log(\sum_{o\in p} l^e_o \pi^e_{o,d}) - \log(L^e_{p,d}))^2 \tag{B.4}$$

To predict the migration flows using the model, we need to know the regional amenity-adjusted real wages, v^e_d. Because there are more than six hundred regions (rural and urban sectors in 340 cities), it is infeasible to estimate all $\{v^e_d\}$ and $\{\beta\}$ simultaneously. I adopt a nested procedure, similar in spirit to Berry et al. (1995), as follows: in the inner loop, for each given $\{\beta\}$, I solve the migration model for the amenity-adjusted real wages, $\{v^e_d\}$, so that the model-predicted total number of workers in each region is the same as that in data, that is, $\sum_{o\in \mathbf{G}} l^e_o \pi^e_{o,d} = \sum_{p\in P} L^e_{p,d}$, $\forall d \in \mathbf{G}$. Once we have $\{v^e_d\}$, we can compute the model-predicted migration flows, and evaluate the objective function for the given $\{\beta\}$. In the outer loop, I then search over the space of $\{\beta\}$ to minimize the objective function.[11] Proposition 1 in Section 1 of this appendix ensures the feasibility of this approach by establishing the existence and uniqueness of the solution to the problem in the inner loop.

We use the 2000 migration data, constructed in section 1 of this appendix, to estimate $\{\beta\}$. After obtaining the estimates, to ensure the recovered $\{v^e_d\}$ are

[11]This nested approach is equivalent to imposing a constraint that the (model-predicted) total numbers of workers migrating to each place equals the total number of workers in that place in the data, and therefore is similar in spirit to what is referred to as "structural gravity estimation" in trade literature. See Fally (2015) for a discussion of the relationship between this and alternative approaches of gravity estimation.

consistent with the 2005 employment distribution, we solve Equation $\sum_{o\in\mathbf{G}} l_o^e \pi_{o,d}^e = \sum_{p\in P} L_{p,d}^e$, $\forall d \in \mathbf{G}$ again, using L_d^e and l_o^e from 2005, to obtain the new $\{v_d^e\}$.

B.3.3 Jointly Estimating Trade Cost and Productivity

I determine international trade costs, domestic trade costs, and regional productivity jointly.

As discussed in the text, due to the aggregate nature of the data, I use nonlinear least square in estimation, which requires solving the model for the predictions of trade flows. In solving the model, to ensure the size and specialization of the cities in the model are consistent with the data, I compute the production of intermediate varieties in each industries in all cities (details in Section 2.2 of this appendix), and force the joint estimation algorithm to respect this distribution of intermediate variety production.

Figure (B.4) explains the joint estimation algorithm. I start with an initial guess for international trade costs, and the parameters governing domestic trade costs, $\{\gamma\}$, with which I compute the trade cost between any trade partners, $\{\tau_{o,d}\}$. I then guess a distribution for regional productivity, solve the trade model for prices and trade shares, and check if the demand for intermediate varieties produced by each region equals the supply.[12] If not, I update the guess for the distribution by

[12]In the step where we solve the trade model, if we know η_d^h and η_d^s, Equations (3.18), (3.26), and (3.29) in the text can be viewed as a system of equations with prices being the only unknowns. Once we solve these equations for the prices, we can obtain trade shares. Although η_d^h and η_d^s are unknown before the model is parameterized, in section 3.4 of this appendix I show that, conditional on information on the shares of different factors in the equipped composite labor, η_d^h and η_d^s are unnecessary in solving the model. Once the model is solved, however, we can use Equation (3.30) to back out η_d^h and η_d^s, to be used in policy experiments.

increasing productivity in regions with excess supply, and decrease productivity in regions with excess demand. The intuition behind this is that, if a region faces excess demand, it means the intermediate varieties produced there is competitive in the international market. To restore the market clearing condition for this region, I make the intermediate varieties produced in that regions more expensive by decreasing the productivity.[13]

Once the distribution of regional productivity that clear all intermediate variety markets are found, I compute the bilateral trade flows, and evaluate the objective function (B.5).

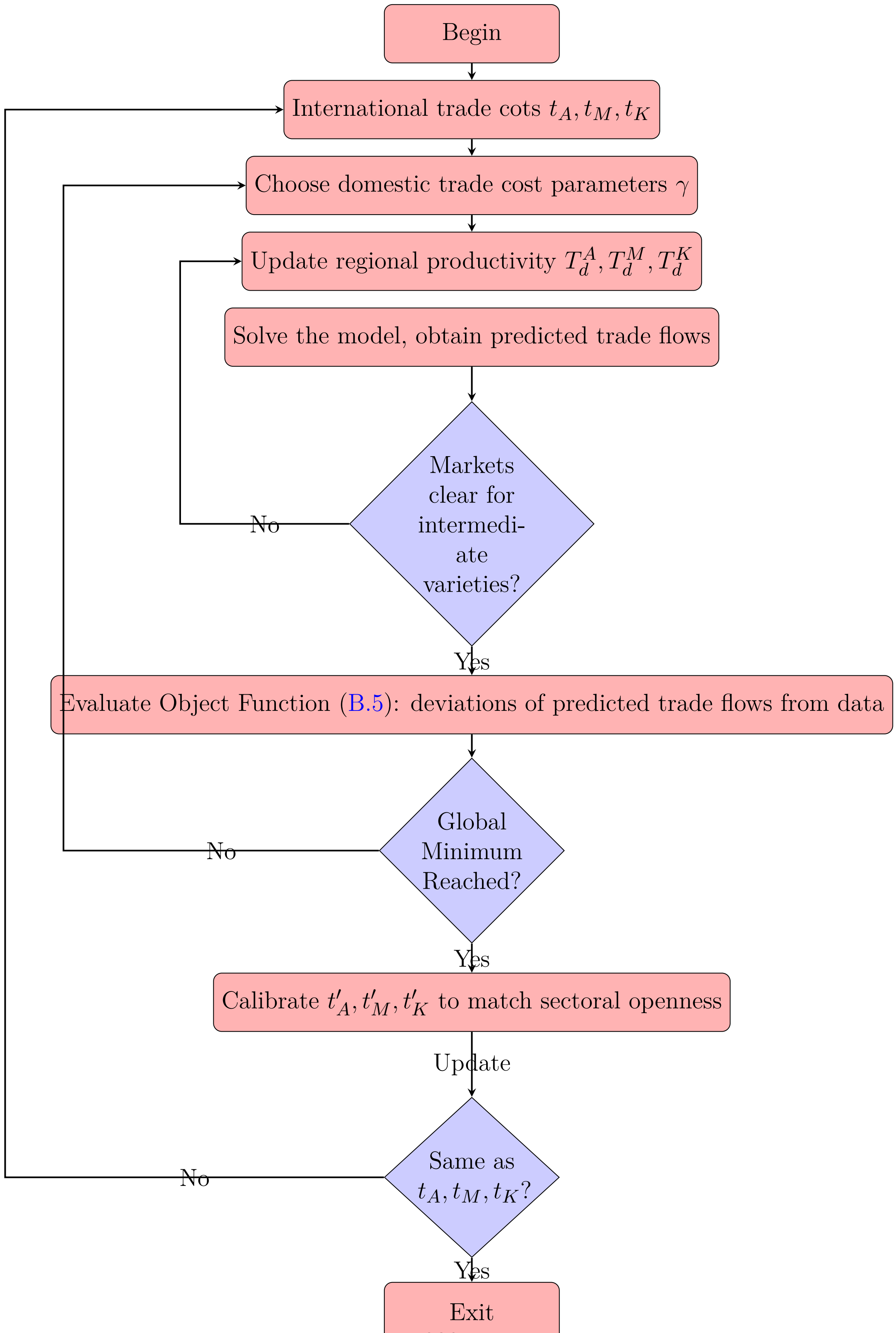
$$\sum_{\text{All P1, P2}} [\log(\frac{X_{P1,P2}}{\text{Domestic Sales}_{P1}}) - \text{the model counterpart}]^2, \tag{B.5}$$

where $X_{P1,P2}$ is the export of goods from province P1 to province P2 in the data. In specifying the objective function, since the domestic trade data is at provincial level, to bring the model and the data together, I aggregate the model-predicted trade flows to provincial level. I normalize the trade flows by aggregate *domestic* sales of the *source* provinces, so that the estimates are not affected by the change in international trade openness between 2002 and 2005.[14]

[13]The feasibility of this approach requires that, for any given level of trade costs, we can find a set of unique T_d^s that clear all intermediate variety markets in all locations. Redding (2012) proves this is true in a single-sector model. An earlier version of this paper extends the proof to a multi-sector model with input-output linkages within the same broad sector. In the general model here with flexible input-output linkages and capital-skill complementarity, the uniqueness cannot be established. But in implementation, I find the update rule always converge uniformly to one unique object.

[14]The domestic trade data is from 2002, whereas the employment data used to determine production and consumption is from 2005. By normalizing the flows using domestic sales of source provinces, I effectively use only the domestic trade patterns in 2002 for estimation.

Figure B.4: Estimation Algorithm



I search over the space of $\{\gamma\}$ until the global minimum is reached, after which I calibrate international trade costs to match the sectoral openness, keeping both domestic trade costs and regional productivity fixed. I repeat the process until convergence.

B.3.4 Additional Information on the Joint Estimation

In solving the trade model, we need to compute the prices of tradable goods, for the estimated regional wages and given distribution of technology $\{T_d^s\}$. Computing the prices, however, requires η_d^h and η_d^l (see footnote (12)).

To proceed with the estimation algorithm, not knowing η_d^h, η_d^l, I substitute the relative factor shares, $\frac{\text{Capital Share}}{\text{Skilled Share}}$ and $\frac{\text{Equipped Skilled Share}}{\text{Unskilled Share}}$, at the regional level, to the left hand side of Equation(3.29) in the text, and express η_d^h, η_d^l as

$$\eta_d^h = \frac{(\frac{P_d^K}{W_d^h})^{1-\rho_{kh}}}{\frac{\text{Capital Share}}{\text{Skilled Share}} + (\frac{P_d^K}{W_d^h})^{1-\rho_{kh}}}, \quad \eta_d^l = \frac{(\frac{W_d^{eh}}{W_d^l})^{1-\rho_{lkh}}}{\frac{\text{Equipped Skilled Share}}{\text{Unskilled Share}} + (\frac{W_d^{eh}}{W_d^l})^{1-\rho_{lkh}}} \tag{B.6}$$

I then substitute Equation (B.6) into (3.29), and solve the model without actually knowing η_d^h or η_d^l. The idea is that, η_d^h and η_d^l must be consistent with the optimal choices of equipped composite labor producers, and therefore when we vary the prices, we also adjust η_d^h and η_d^l so that the optimal factor shares are consistent with data. Once the whole procedure is over and the model is solved, we can then back out η_d^h and η_d^l from (B.6). These are interpreted as the true parameter values, which I keep fixed for all counterfactual experiments.

B.3.5 Parameters for the Counterfactual Experiments with Different Internal Geographies

In the counterfactual experiments with alternative internal geographies, reported in Section 7.2, I reduce the values of inter-provincial dummies in the trade cost specifications in China to the U.S. level. I also liberalize migration in China through hypothetical Hukou reforms that decrease the destination-city specific migration costs. In this section I describe the sources and values of these parameters.

The value of inter-state trade costs are from Crafts and Klein (2014), which estimates U.S. inter-state trade using the latest data. Under different specifications, their estimates for the inter-state dummy range between 2 to 2.55. To be conservative, I use the upper bound of their estimates, 2.55. This estimate of the inter-state dummy bundles together trade elasticity and trade costs, so I recover the inter-state trade cost by dividing 2.55 by 4, the trade elasticity, arriving at 0.65. Therefore in relevant experiments I reduce the provincial border effect from the benchmark level of 1.1 to 0.65.

The change in migration cost from the hypothetical Hukou reforms are calculated as follows. Conceptually, I model the impact of Hukou in the following way. Consider the migration cost from the origin o to destination d, $d^e_{o,d}$. This can be decomposed into a destination-specific component, d_d, and a pair and skill type specific component, $\underline{d}^e_{od}$, so that $d^e_{o,d} = d_d \underline{d}^e_{od}$.

From Equation 3.9, the total number of workers that migrate to city d from else where is the following:

$$
\begin{aligned}
&\sum_{o\neq d}(l_o^h\pi_{od}^h + l_o^l\pi_{od}^l)\\
=&\sum_{o\neq d}\Big(\frac{(\frac{v_d^h}{d_{o,d}^h})^{\epsilon_h}}{\sum_{g\in\mathbf{G}}(\frac{v_g^h}{d_{o,g}^h})^{\epsilon_h}}l_o^h + \frac{(\frac{v_d^l}{d_{o,d}^l})^{\epsilon_l}}{\sum_{g\in\mathbf{G}}(\frac{v_g^l}{d_{o,g}^l})^{\epsilon_l}}l_o^l\Big)\\
=&\sum_{o\neq d}\Big(d_d^{-\epsilon_h}\frac{(\frac{v_d^h}{\underline{d}_{o,d}^h})^{\epsilon_h}}{\sum_{g\in\mathbf{G}}(\frac{v_g^h}{d_{o,g}^h})^{\epsilon_h}}l_o^h + d_d^{-\epsilon_l}\frac{(\frac{v_d^l}{\underline{d}_{o,d}^l})^{\epsilon_l}}{\sum_{g\in\mathbf{G}}(\frac{v_g^l}{d_{o,g}^l})^{\epsilon_l}}l_o^l\Big)\\
\approx& d_d^{-\epsilon_h}\sum_{o\neq d}\Big(\frac{(\frac{v_d^h}{\underline{d}_{o,d}^h})^{\epsilon_h}}{\sum_{g\in\mathbf{G}}(\frac{v_g^h}{d_{o,g}^h})^{\epsilon_h}}l_o^h + \frac{(\frac{v_d^l}{\underline{d}_{o,d}^l})^{\epsilon_l}}{\sum_{g\in\mathbf{G}}(\frac{v_g^l}{d_{o,g}^l})^{\epsilon_l}}l_o^l\Big),
\end{aligned}
\tag{B.7}
$$

where $\approx$ follows from the fact that in calibration, $\epsilon_h \approx \epsilon_l \approx 4.67$. For a local change in d_d, the term in bracket does not change much. After taking log and first difference, we have:

$$
\begin{aligned}
\triangle\ln\Big(\sum_{o\neq d}(l_o^h\pi_{od}^h + l_o^l\pi_{od}^l)\Big) &\approx \triangle - \epsilon_h\ln(d_d)\\
&\approx -4.67\triangle\ln(d_d).
\end{aligned}
\tag{B.8}
$$

Under the assumption that the bracketed term in the last line of B.7 does not change,[15] Equation B.8 can be interpreted as the model-based specification for the results in Table 3.5. The estimate in Table 3.5 suggests that an one-point increase in the reform index increases migration by 20%. Plug this into Equation B.8, we have the following: each point of Hukou reform index implies a $0.2/4.67 = 4.3\%$

[15]This is certainly a strong assumption, but the empirical specification incorporates contemporaneous city characteristics and time fixed effects, which captures the change the city-specific real wage, v_d^e, and other terms in the bracket.

decrease in the city-specific migration cost, d_d. Liberalizing the entire economy from the average Hukou restriction in 2000 to complete openness implies a 5-point increase in the Hukou reform index for all cities. Therefore for this counterfactual experiment, I assume that the Hukou reform would decrease the migration cost into each city (from elsewhere) by 22%.

B.3.6 Discussion on the Estimated Inter-Provincial Effect and Additional Robustness

In Section 6.5.2, I report my estimates of the domestic trade costs. It is useful to compare my estimates to those obtained using the U.S. Commodity Flow Survey data. In the literature, the comparable coefficient for state border, after scaled appropriately by the elasticity of trade, is on the range of 0.38 (Wolf, 2000) to 0.65 (Crafts and Klein, 2014, using 2007 data). So my estimate of the state-border effect is about twice as large as the comparable estimates for the U.S., reflecting larger barriers to trade flows at provincial borders in China. One lesson from the U.S. state border literature is that, the estimates might be driven up by the wholesale industry (Hillberry and Hummels, 2003), and might suffer from the aggregation bias—a lot of trade costs are actually due to geographic distance, but might be captured by the state-border dummy when state-level aggregate data is used. When these two factors are taken into account, the estimates shrink (Hillberry and Hummels, 2008).

Therefore, as discussed in Section 6.5.2, one natural concern is whether in China, due to the quality, or the level of aggregation, of the data, the estimates might also misattribute the impacts of geographic distance to the provincial borders; and if that is the case, whether the results from the counterfactual experiments are still valid.

Without detailed micro-level trade flow data available for China, I cannot examine the bias of the estimates. Instead, I use an additional experiment to show that even if there is bias in the estimation, it will not affect main conclusions of the counterfactual experiment. Specifically, I perform a robustness exercise, in which I reduce inter-provincial and inter-regional trade costs to 0.65, the level of the U.S. economy, while at the same time increase the coefficients for the continuous geographic components, so that the overall domestic trade costs and international trade participation are similar to those of the benchmark economy. Effectively, I change the composition of the domestic trade costs, keeping its overall level same as before. I shut down international trade in this economy, and compute the welfare gains from trade, as well as other outcome variables discussed in the text. The results, reported in Table (B.4), are very similar to those of the benchmark experiment, reported in the first column of Table (3.11).

Table B.4: Counterfactual Experiment with an Alternative Domestic Trade Cost Structure

Panel A: Statistics by Worker Group		
	Mean	std
Urban Skilled	11.52	9.89
Urban Unskilled	5.43	7.41
Rural Skilled	11.10	9.12
Rural Unskilled	5.29	6.74
Panel B: Aggregate Statistics		
National Average	7.47	
Trade Openness	60.60	
Increase in Inequality	6.7	
Contribution-Between(%)	56.88	
Contribution-Within (%)	43.12	

References

Abowd, John M, Francis Kramarz, and David N Margolis, "High Wage Workers and High Wage Firms," *Econometrica*, 1999, *67* (2), 251–333.

Acemoglu, Daron and Joshua Linn, "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry," *The Quarterly journal of economics*, 2004, *119* (3), 1049–1090.

Ahlfeldt, Gabriel M, Stephen J Redding, Daniel M Sturm, and Nikolaus Wolf, "The Economics of Density: Evidence from the Berlin Wall," *Econometrica*, 2015, *83* (6), 2127–2189.

Akay, Alpaslan, Corrado Giulietti, Juan D Robalino, and Klaus F Zimmermann, "Remittances and Well-being among Rural-to-urban Migrants in China," *Review of Economics of the Household*, 2012, pp. 1–30.

Akcigit, Ufuk, Salomé Baslandze, and Stefanie Stantcheva, "Taxation and the International Mobility of Inventors," *The American Economic Review*, 2016, *106* (10), 2930–2981.

Alfaro, Laura, Sebnem Kalemli-Ozcan, and Vadym Volosovych, "Why Doesn't Capital Flow from Rich to Poor Countries? An Empirical Investigation," *The Review of Economics and Statistics*, 2008, *90* (2), 347–368.

Allen, Treb and Costas Arkolakis, "Trade and the Topography of the Spatial Economy," *The Quarterly Journal of Economics*, 2014, *129* (3), 1085–1140.

Alviarez, Vanessa, "Multinational Production and Comparative Advantage," *Working Paper, University of British Columbia*, 2016.

Ambos, Björn, "Foreign Direct Investment in Industrial Research and Development: A Study of German MNCs," *Research Policy*, 2005, *34* (4), 395–410.

Antràs, Pol and Stephen Yeaple, "Multinational Firms and the Structure of International Trade," *Handbook of International Economics*, 2014, *4*, 55.

Antras, Pol, Luis Garicano, and Esteban Rossi-Hansberg, "Offshoring in a Knowledge Economy," *The Quarterly Journal of Economics*, 2006, *121* (1), 31–77.

Arkolakis, Costas, Arnaud Costinot, and Andres Rodriguez-Clare, "New Trade Models, Same Old Gains?," *The American Economic Review*, 2012, *102* (1), 94–130.

_ , Natalia Ramondo, Andrés Rodríguez-Clare, and Stephen Yeaple, "Innovation and Production in the Global Economy," *National Bureau of Economic Research Working Paper*, 2014.

Atkin, David and Dave Donaldson, "Who's Getting Globalized? The Size and Nature of Intranational Trade Costs," *Mimeo*, 2012.

Au, Chun-Chung and J Vernon Henderson, "How Migration Restrictions Limit Agglomeration and Productivity in China," *Journal of Development Economics*, 2006, *80* (2), 350–388.

Autor, David H., David Dorn, and Gordon H. Hanson, "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, 2013, *103* (6), 2121–68.

Berry, Steven, James Levinsohn, and Ariel Pakes, "Automobile Prices in Market Equilibrium," *Econometrica*, 1995, pp. 841–890.

Bilir, L Kamran and Eduardo Morales, "The Impact of Innovation in the Multinational Firm," *National Bureau of Economic Research Working Paper*, 2016.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts, "Does Management Matter? Evidence from India," *The Quarterly Journal of Economics*, 2013, *1* (51), 51.

—, Christos Genakos, Raffaella Sadun, and John Van Reenen, "Management Practices Across Firms and Countries," *The Academy of Management Perspectives*, 2012, *26* (1), 12–33.

—, Erik Brynjolfsson, Lucia Foster, Ron S Jarmin, Megha Patnaik, Itay Saporta Eksten, and John Van Reenen, "IT and Management in America," *CEPR Discussion Paper No. DP9886*, 2014.

Branstetter, Li, Guangwei Li, and Francisco Veloso, "The Globalization of R&D: China, India and the Rise of International Co-invention," *Unpublished Manuscript*, 2013.

Burstein, Ariel, Javier Cravino, and Jonathan Vogel, "Importing Skill-Biased Technology," *American Economic Journal: Macroeconomics*, 2013, *5* (2), 32–71.

Burstein, Ariel T and Alexander Monge-Naranjo, "Foreign Know-how, Firm Control, and the Income of Developing Countries," *The Quarterly Journal of Economics*, 2009, p. 149.

Cadena, Brian C and Brian K Kovak, "Immigrants Equilibrate Local Labor Markets: Evidence from the Great Recession," *National Bureau of Economic Research Working Paper*, 2013.

Caliendo, Lorenzo and Fernando Parro, "Estimates of the Trade and Welfare Effects of NAFTA," *The Review of Economic Studies*, Forthcoming.

Chatruc, Marisol Rodriguez, *Trade Shocks and Local Labor Markets' Linkages: Theory and Evidence*, University of Maryland, College Park, 2016.

Coşar, A Kerem and Banu Demir, "Domestic Road Infrastructure and International Trade: Evidence from Turkey," *Journal of Development Economics*, 2016, *118*, 232–244.

Cosar, A Kerem and Pablo D Fajgelbaum, "Internal Geography, International Trade, and Regional Specialization," *American Economic Journal: Microeconomics*, 2016, *8* (1), 24–56.

Cosar, Kerem, "Adjusting to Trade Liberalization: Reallocation and Labor Market Policies," *University of Chicago, Unpublished Manuscript.*

Costinot, Arnaud and Andres Rodriguez-Clare, "Trade Theory with Numbers: Quantifying the Consequences of Globalization," *Handbook of International Economics*, 2014, *4*, 197.

_ and Jonathan Vogel, "Matching and Inequality in the World Economy," *Journal of Political Economy*, 2010, *118* (4).

_ , _ et al., "Beyond Ricardo: Assignment Models in International Trade," *Annual Review of Economics*, 2015, *7* (1), 31–62.

Crafts, Nicholas and Alexander Klein, "Geography and Intra-national Home Bias: U.S. Domestic Trade in 1949 and 2007," *Journal of Economic Geography*, 2014.

Dekle, Robert, Jonathan Eaton, and Samuel Kortum, "Global Rebalancing with Gravity: Measuring the Burden of Adjustment," *IMF Staff Papers*, 2008, *55* (3), 511–540.

Diamond, Rebecca, "The determinants and welfare implications of US workers' diverging location choices by skill: 1980–2000," *The American Economic Review*, 2016, *106* (3), 479–524.

Dix-Carneiro, Rafael, "Trade Liberalization and Labor Market Dynamics," *Econometrica*, 2014, *82* (3), 825–885.

— and Brian K Kovak, "Trade Reform and Regional Dynamics: Evidence from 25 years of Brazilian Matched Employer-Employee data," *National Bureau of Economic Research Working Paper*, 2015.

Donaldson, Dave, "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *The American Economic Review*, 2017.

Eaton, Jonathan and Samuel Kortum, "Technology, Geography, and Trade," *Econometrica*, 2002, *70* (5), 1741–1779.

Fally, Thibault, "Structural Gravity and Fixed Effects," *Journal of International Economics*, 2015, *97* (1), 76–85.

Gersbach, Hans and Armin Schmutzler, "Foreign Direct Investment and R&D-Offshoring," *Oxford Economic Papers*, 2011, *63* (1), 134–157.

Goldberg, Pinelopi Koujianou and Nina Pavcnik, "Distributional Effects of Globalization in Developing Countries," *Journal of Economic Literature*, 2007, *45* (1), 39–82.

Grossman, Gene M and Elhanan Helpman, "Growth, Trade, and Inequality," *National Bureau of Economic Research Working Paper*, 2014.

— and Giovanni Maggi, "Diversity and Trade," *American Economic Review*, 2000, pp. 1255–1275.

—, Elhanan Helpman, Philipp Kircher et al., "Matching, Sorting, and the Distributional Effects of International Trade," *Journal of Political Economy*, forthcoming.

Guadalupe, Maria, Olga Kuzmina, and Catherine Thomas, "Innovation and Foreign Ownership," *The American Economic Review*, 2012, *102* (7), 3594–3627.

Guimón, José, "Government Strategies to Attract R&D-Intensive FDI," *The Journal of Technology Transfer*, 2009, *34* (4), 364–379.

Haas, Martine R and Prithwiraj Raj Choudhury, "Collaborate to Innovate? Teams and Patent Generation in a Global R&D Center," 2015.

Hall, Bronwyn H, "The Internationalization of R&D," *Available at SSRN 2179941*, 2011.

_, Adam B Jaffe, and Manuel Trajtenberg, "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools," *National Bureau of Economic Research Working Paper*, 2001.

Han, Jun, Runjuan Liu, and Junsen Zhang, "Globalization and Wage Inequality: Evidence from Urban China," *Journal of International Economics*, 2012, *87* (2), 288–297.

Hanson, Gordon, "Market potential, increasing returns and geographic concentration," *Journal of International Economics*, 2005, *67* (1), 1–24.

Hanson, Gordon H, "US–Mexico Integration and Regional Economies: Evidence from Border-city Pairs," *Journal of Urban Economics*, 2001, *50* (2), 259–287.

Hanushek, Eric A and Ludger Woessmann, "Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation," *Journal of Economic Growth*, 2012, *17* (4), 267–321.

Head, Keith and John Ries, "FDI as an Outcome of the Market for Corporate Control: Theory and Evidence," *Journal of International Economics*, 2008, *74* (1), 2–20.

_ and Thierry Mayer, "Market Potential and the Location of Japanese investment in the European Union," *Review of Economics and Statistics*, 2004, *86* (4), 959–972.

Helpman, Elhanan, "A Simple Theory of Ttrade with Multinational Corporations," *Journal of Political Economy*, 1984, *92* (3).

_ , Marc J Melitz, and Stephen R Yeaple, "Export Versus FDI with Heterogeneous Firms," *American Economic Review*, 2004, *94* (1), 300–316.

_ , Oleg Itskhoki, Marc-Andreas Muendler, and Stephen J Redding, "Trade and Inequality: From Theory to Estimation," *The Review of Economic Studies*, 2016, p. rdw025.

Hillberry, Russell and David Hummels, "Intranational Home Bias: Some Explanations," *Review of Economics and Statistics*, 2003, *85* (4), 1089–1092.

_ and _ , "Trade Responses to Geographic Frictions: A Decomposition Using Micro-data," *European Economic Review*, 2008, *52* (3), 527–550.

Holmes, Thomas J, Ellen R McGrattan, and Edward C Prescott, “Quid Pro Quo: Technology Capital Transfers for Market Access in China,” *The Review of Economic Studies*, 2015, p. rdv008.

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

—, Erik Hurst, Charles I. Jones, and Peter J. Klenow, “The Allocation of Talent and U.S. Economic Growth,” *National Bureau of Economic Research Working Paper*, 2013.

Iones, Charles I, “Misallocation, Economic Growth, and Input-Output Economics,” in “Advances in Economics and Econometrics: Tenth World Congress,” Vol. 2 Cambridge University Press 2013, p. 419.

Irarrazabal, Alfonso, Andreas Moxnes, and Luca David Opromolla, “The Margins of Multinational Production and the Role of Intrafirm Trade,” *Journal of Political Economy*, 2013, *121* (1), 74–126.

Ito, Banri and Ryuhei Wakasugi, “What Factors Determine the Mode of Overseas R&D by Multinationals? Empirical Evidence,” *Research Policy*, 2007, *36* (8), 1275–1287.

Kanbur, Ravi and Anthony J Venables, *Spatial Inequality and Development*, Oxford University Press, 2005.

— and Xiaobo Zhang, "Fifty Years of Regional Inequality in China: a Journey Through Central Planning, Reform, and Openness," *Review of development Economics*, 2005, *9* (1), 87–106.

Keller, Wolfgang and Stephen Ross Yeaple, "The Gravity of Knowledge," *American Economic Review*, 2013, *103* (4), 1414–1444.

Kennan, John and James R Walker, "The Effect of Expected Income on Individual Migration Decisions," *Econometrica*, 2011, *79* (1), 211–251.

Kerr, Sari Pekkala and William R Kerr, "Global Collaborative Patents," *National Bureau of Economic Research Working Paper*, 2014.

— , — , Caglar Ozden, and Christopher Robert Parsons, "Global Talent Flows," *World Bank Policy Research Working Paper*, 2016, (7852).

Kinnan, Cynthia, Shing-Yi Wang, and Yongxiang Wang, "Relaxing Migration Constraints for Rural Households," *National Bureau of Economic Research Working Paper*, 2015.

Kovak, Brian K, "Regional Effects of Trade Reform: What is the Correct Measure of Liberalization?," *The American Economic Review*, 2013, *103* (5), 1960–1976.

Krugman, Paul and Raul Livas Elizondo, "Trade Policy and the Third World Metropolis," *Journal of Development Economics*, 1996, *49* (1), 137–150.

Krusell, Per, Lee E Ohanian, José-Víctor Ríos-Rull, and Giovanni L Violante, "Capital-skill Complementarity and Inequality: A Macroeconomic Analysis," *Econometrica*, 2000, *68* (5), 1029–1053.

Kumar, Utsav, *Essays on inequality and reforms: Evidence from India*, University of Maryland, College Park, 2007.

Levchenko, Andrei A and Jing Zhang, "The Global Labor Market Impact of Emerging Giants: a Quantitative Assessment," *IMF Economic Review*, 2013, *61* (3), 479–519.

Li, Guan-Cheng, Ronald Lai, Alexander DiAmour, David M Doolin, Ye Sun, Vetle I Torvik, Z Yu Amy, and Lee Fleming, "Disambiguation and Coauthorship Networks of the US Patent Inventor Database (1975–2010)," *Research Policy*, 2014, *43* (6), 941–955.

Limão, Nuno and Anthony J Venables, "Infrastructure, Geographical Disadvantage, Transport Costs, and Trade," *The World Bank Economic Review*, 2001, *15* (3), 451–479.

Ma, Lin, "Globalization and Top Income Shares," *Mimeo*, 2013.

Markusen, James R, "Multinationals, Multi-Plant Economies, and the Gains from Trade," *Journal of International Economics*, 1984, *16* (3), 205–226.

Mayer, Thierry and Soledad Zignago, "Notes on CEPII's Distances Measures: The GeoDist Database," *CEPII Working Paper*, 2011.

McGrattan, Ellen R and Edward C Prescott, "Openness, Technology Capital, and Development," *Journal of Economic Theory*, 2009, *144* (6), 2454–2476.

Michaels, Guy, Stephen J Redding, and Ferdinand Rauch, "Technical Note: An Eaton and Kortum (2002) Model of Urbanization and Structural Transformation," *Mimeo*, 2011.

Miguelez, Ernest and Carsten Fink, "Measuring the International Mobility of Inventors: A New Database," *WIPO Economics & Statistics Series No.8*, 2013.

Nocke, Volker and Stephen Yeaple, "Cross-Border Mergers and Acquisitions vs. Greenfield Foreign Direct Investment: The Role of Firm Heterogeneity," *Journal of International Economics*, 2007, *72* (2), 336–365.

_ and _ , "An Assignment Theory of Foreign Direct Investment," *The Review of Economic Studies*, 2008, *75* (2), 529–557.

Park, Walter G, "International Patent Protection: 1960–2005," *Research Policy*, 2008, *37* (4), 761–766.

Parro, Fernando, "Capital-skill Complementarity and the Skill Premium in a Quantitative Model of Trade," *American Economic Journal: Macroeconomics*, 2013, *5* (2), 72–117.

Pavitt, Keith, "Uses and Abuses of Patent Statistics," in A. F. J. van Raan, ed., *Handbook of Quantitative Studies of Science and Technology*, Sole Distributors for the U.S.A. And Canada, Elsevier Science Pub. Co., 1988, pp. 509–536.

Poncet, Sandra, “Measuring Chinese Domestic and International Integration,” *China Economic Review*, 2003, *14* (1), 1–21.

_ , “A Fragmented China: Measure and Determinants of Chinese Domestic Market Disintegration,” *Review of international Economics*, 2005, *13* (3), 409–430.

Ramondo, Natalia, “A Quantitative Approach to Multinational Production,” *Journal of International Economics*, 2014, *93* (1), 108–122.

_ and Andrés Rodríguez-Clare, “Trade, Multinational Production, and the Gains from Openness,” *Journal of Political Economy*, 2013, *121* (2), 273–322.

_ , _ , and Felix Tintelnot, “Multinational Production: Data and Stylized Facts,” *The American Economic Review*, 2015, *105* (5), 530–536.

Redding, Stephen J., “Goods Trade, Factor Mobility and Welfare,” *National Bureau of Economic Research Working Paper*, 2012.

Roys, Nicolas and Ananth Seshadri, “Economic Development and the Organization of Production,” *Mimeo*, 2014.

Sampson, Thomas, “Selection into Trade and Wage Inequality,” *American Economic Journal: Microeconomics*, 2014, *6* (3), 157–202.

Schank, Thorsten, Claus Schnabel, and Joachim Wagner, “Do Exporters Really Pay Higher Wages? First Evidence from German Linked Employer–Employee Data,” *Journal of International Economics*, 2007, *72* (1), 52–74.

Shimizutani, Satoshi and Yasuyuki Todo, “What Determines Overseas R&D Activities? The Case of Japanese Multinational Firms,” *Research Policy*, 2008, *37* (3), 530–544.

Simonovska, Ina and Michael E Waugh, “The Elasticity of Trade: Estimates and Evidence,” *Journal of International Economics*, 2014, *92* (1), 34–50.

Sotelo, Sebastian, “Trade Frictions and Agricultural Productivity: Theory and Evidence from Peru,” *Mimeo*, 2014.

Sun, W, C Bai, and P Xie, “The Impact of Hukou Reforms on Rural Labor Migration in China,” *Economic Research Journal*, 2011, pp. 28–41.

Swiecki, Tomasz, “Intersectoral Distortions and the Welfare Gains from Trade,” *Journal of International Economics*, 2017, *104*, 138–156.

Tang, Lixin, “Top Income Inequality, Aggregate Saving and the Gains from Trade,” *Mimeo*, 2014.

Terviö, Marko, “The Difference that CEOs Make: An Assignment Model Approach,” *The American Economic Review*, 2008, pp. 642–668.

Thursby, Marie and Jerry Thursby, *Here or There?: A Survey of Factors in Multinational R&D Location–Report to the Government-University-Industry Research Roundtable*, National Academies Press, 2006.

Tintelnot, Felix, “Global Production with Export Platforms,” *The Quarterly Journal of Economics*, 2016, p. qjw037.

Tombe, Trevor and Jennifer Winter, "What's Inside Counts: Migration, Taxes, and the Internal Gains from Trade," *Mimeo*, 2014.

— and Xiaodong Zhu, "Trade, Migration and Regional Income Differences: Evidence from China," *Mimeo*, 2015.

Topalova, Petia, "Factor Immobility and Regional Impacts of Trade Liberalization: Evidence on Poverty from India," *American Economic Journal: Applied Economics*, 2010, pp. 1–41.

Venables, Anthony J and Nuno Limão, "Geographical Disadvantage: A Heckscher–Ohlin–von Thünen Model of International Specialisation," *Journal of International Economics*, 2002, *58* (2), 239–263.

Wei, Shang-Jin, "How Taxing is Corruption on International Investors?," *Review of Economics and Statistics*, 2000, *82* (1), 1–11.

Wolf, Holger C, "Intranational Home Bias in Trade," *Review of Economics and Statistics*, 2000, *82* (4), 555–563.

Yeaple, Stephen Ross, "The Role of Skill Endowments in the Structure of US Outward Foreign Direct Investment," *Review of Economics and Statistics*, 2003, *85* (3), 726–734.

— , "A Simple Model of Firm Heterogeneity, International Trade, and Wages," *Journal of International Economics*, 2005, *65* (1), 1–20.