Title
Essays on Economic Growth and International Trade

Author
Yang, Yang

Publication Date
2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Essays on Economic Growth and International Trade

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy in Management

by

Yang Yang

2017

ABSTRACT OF THE DISSERTATION

Essays on Economic Growth and International Trade by

Yang Yang
Doctor of Philosophy in Management
University of California, Los Angeles, 2017
Professor Romain T. Wacziarg, Chair

The first chapter of this dissertation examines the impact of highway expansion on aggregate productivity growth and sectoral reallocation between cities in China. To do so, I construct a unique dataset of bilateral transportation costs between Chinese cities, digitized highway network maps, and firm-level census. I first derive and estimate a market access measure for cities in China from 1995 to 2005. I then examine the channels through which the highway infrastructure affected economic outcomes. The results suggest that highways promoted aggregate productivity growth by facilitating the entry of new firms and reallocation among existing firms. I estimate the aggregate economic impact of China's national highway system and find that eliminating all highways in China would decrease aggregate productivity by 3.2%. There is also evidence that the national highway system led to a sectoral reallocation between cities in China.

In the second chapter, I investigate to what extent firms in developing economies improve their productivity by importing foreign technology. I examine the effects of machinery importing on firm productivity for Chinese manufacturing firms, or "upgrading by importing". To do so, I develop an algorithm to merge the Chinese firm-level census data with the Chinese Customs data. To address endogeneity concerns on importing decisions, I use a propensity score matching approach to identify the impact of machinery imports on firm productivity. Finally, I estimate a simple empirical model to examine the heterogeneous effects and to quantify the aggregate impact of machinery importing. I find that machinery

and equipment imports improved firm productivity in China and could potentially generate large gains in aggregate productivity. The results in this paper suggest that importing foreign machinery goods is an important channel for technology diffusion.

The third chapter analyzes the impact of fiscal and structural policies on gender inequality for countries at different stages of development. Using Bayesian Model Averaging, in addition to frequentist methods, we address model uncertainty due to the large number of possible determinants previously highlighted in the microeconomics literature. We find that better sanitation facilities, low adolescent fertility, narrower marriage age gaps and higher public spending in education contribute to closing the gender gap in education. Better infrastructure, more equal legal rights, low adolescent fertility rates, and a stronger institutional environment boost female labor force participation. At lower levels of labor market protection, stronger protection is associated with narrower labor force participation gaps, but excessive labor market rigidity weighing on female labor force participation.

The dissertation of Yang Yang is approved.

W. Walker Hanlon

Edward E. Leamer

Nico Voigtländer

Romain T. Wacziarg, Committee Chair

University of California, Los Angeles

2017

Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank my advisor, Romain Wacziarg, for his invaluable guidance and encouragement. I am also indebted to Nico Voigtländer, Edward Leamer, Walker Hanlon and Thierry Tressel for their generous support along the way.

I also thank my supervisors/coauthors at the IMF, Sonali Jain-Chandra, Kalpana Kochhar, Monique Newiak, and Edda Zoli, for creating a free and collegial environment for me when I was an intern at the Fund. Chapter 3 is a revised version of joint work with them.

I am very grateful to the Center for Global Management and the Price Center for Entrepreneurship and Innovation at the UCLA Anderson School of Management for their generous financial support for my research.

I acknowledge my colleagues Youngjin Song, Andrea Di Miceli, Alvaro Garcia, Ting Ji, Yang Cao, Menghan Xu, Cong Xie, Ruoyao Shi, Cheng Zhou, Vasily Korovkin, Stefano Fiorin, Mikhail Poyker, Shekhar Mittal, Mikhail Galashin, Xiaochen Feng, Rui Xu for their helpful comments and suggestions.

Finally, special thanks go to my parents, Sil Park, Richard Xie and Viktor Zhong. Thank you all for believing in me and making me believe in myself.

YANG YANG

Education

University of California, Los Angeles Los Angeles, CA	Ph.D. Candidate, Aug 2014 Global Economics and Management
Hong Kong University of Science and Technology Hong Kong	BSc, Jun 2011 Economics and Finance

Employment

Jun 2016 – Sep 2016	The International Monetary Fund Asia Pacific Department	Fund Intern Washington, DC
Jan 2016 – May 2016	Pepperdine University Business Administration Division	Adjunct Professor Malibu, CA
Jun 2014 – Apr 2015	The World Bank Group Development Research Group	Short-Term Consultant Washington, DC
Dec 2012 – Sep 2013	University of California, Los Angeles Anderson School of Management	Research Assistant Los Angeles, CA

Fellowships and Awards

Center for Global Management, UCLA Anderson (2014-17)
Price Center for Entrepreneurship & Innovation, UCLA Anderson (2015)
Anderson Fellowship, UCLA Anderson (2011-2015)

Chapter 1

Transport Infrastructure, City Productivity Growth and Sectoral Reallocation: Evidence from China

1.1 Introduction

While transport infrastructure projects are among the most expensive investments in the world, many questions remain as to whether and how transport infrastructure influences economic outcomes. Transport infrastructure facilitates interactions between cities by reducing the costs of transportation between them.[1] We learn from trade theories that resource reallocation at both the firm and industry levels generates gains from reductions in trade barriers.[2] A large body of research in international trade has examined how the removal of trade barriers affects industries and firms, as well as the welfare implications of trade

[1]Transportation costs determine inter-city trade and inter-city travel/migration. In the model I present, city aggregate productivity is determined through the trade channel. However, the market access term I derive is fairly general and is able to incorporate other channels as long as the effect can be diciplined by a gravity equation.

[2]The former channel is usually called industry specialization suggested by classical trade theories such as the Ricardian model and the Heckscher-Ohlin model or the New Economic Geography models pioneered by Krugman (1991). The latter channel refers to resrouce reallocation between firms within industries suggested by more recent trade theories (Melitz 2003, Bernard et al. 2003, and Chaney 2008).

liberalization.[3] While international trade barriers have been reduced drastically over the past few decades, domestic transportation costs remain high even in developed countries.[4] An interesting and important question for researchers is the following: Are the channels suggested in the literature operational in the case of reductions in domestic transportation costs? If so, to what extent?

This paper aims to make two contributions to the literature. To the best of my knowledge, this paper is the first to examine the channels through which highways affect firm and aggregate productivity growth.[5] Previous research focuses on whether a road or railroad connection affects GDP or population growth. We still know very little about the channels through which transport infrastructure affects aggregate TFP growth. In this paper, I decompose the change in aggregate productivity resulting from highway connection into four channels: within-firm productivity growth, entry of new firms, reallocation between existing firms and exit of inefficient firms.

Second, this paper adds to the identification and estimation of the effects of transport infrastructure. This paper seeks to obtain reliable estimates of transportation costs between cities using price quote data from logistics companies. I construct a time-varying instrumental variable to address endogeneity concerns on both the *location* and *timing* of highway placement.[6] This newly developed instrumental variable approach allows me to shed light on the causal impact of highways on economic outcomes.

Evaluating the impact of a transport infrastructure project is not a straightforward exercise. One approach is to estimate or calibrate all model parameters and simulate the effects of changes in policies on economic outcomes. This structural approach has been successfully

[3] There are many good references on this topic, including Melitz and Trefler (2012), Arkolakis et al. (2012), Costinot and Rodriguez-Clare (2014).

[4] See *World Development Report 2009: Reshaping Economic Geography* (2009) for a detailed account of domestic trade barriers across the world.

[5] Ghani et al. (2016) examine the effects of the Golden Quadrilateral project in India on population, GDP and labor productivity. However, they do not estimate TFP in their paper, nor do they explore the channels of productivity gains.

[6] Whereas most studies use historical routes or planned routes as instruments for actual road construction, my instrument is time-varying and is constructed from engineering and network theory.

implemented in a few recent papers (Redding 2016; Nagy 2015).[7] Another approach is to use (plausibly exogenous) variation in treatment intensity within a country to examine the *relative* effects across regions (Autor et al. 2013; Donaldson 2016). With such cross-regional variation, the standard reduced-form analysis can be implemented to evaluate the effects of the project. For this paper, I first adopt the latter approach to estimate all the direct and indirect impact of highways on productivity. I then consider the counterfactual scenario of eliminating all highways to evaluate the aggregate impact of China's national highway system.[8]

I first estimate road transportation costs between any two prefecture-level cities in China. Most early studies in the literature use distance as a proxy for transportation costs (Hanson 2005; Redding and Venables 2004). A few recent papers combine digitized maps of transport routes with selected parameters to estimate transportation costs (Baum-Snow et al. 2015; Donaldson and Hornbeck 2016). I estimate transportation costs between any two prefecture-level cities in China with price quote data and digitized maps of China's highway network. This approach is novel in the literature and improves on the methods mentioned above because I use the actual prices that firms have to pay to transport goods from the origin to the destination to measure proximity to suppliers and consumers for 339 prefecture-level cities in China.

A market access measure is derived from a multi-city trade model with heterogeneous firms. A city's market access summarizes all the direct and indirect impact of transportation costs on city aggregate productivity. I estimate city market access from 1995 to 2005 using estimated bilateral transportation costs.[9] Donaldson and Hornbeck (2016) point out that the

[7]The estimated impact, however, hinges heavily on choices of functional forms and parameters.

[8]It is interesting to consider and compare other counterfacdtual scenarios. For example, would an alternative highway system generate larger economic gains? What would happen to aggregate productivity if the railway system expanded instead of highways?

[9]Before 2000, long-distance freight was shipped primarily by rail, whereas short-distance freight was carried on local roads. However, the railway system changed very little over the period that I study. Thus, I do not expect the rail network to be a strong driver for growth after 2000. Banerjee et al. (2012) examine the effects of railways on local economic outcomes and they find railway had weak positive effects on the level of GDP, but not on GDP growth. In my empirical specification that I will discuss in detail below, I

"market access" approach addresses the methodological challenge of estimating aggregate treatment effects with considerable treatment spillover effects. In this paper, treatment intensity differs across cities even though the objective of the Chinese government was to build an integrated national highway network system. I exploit such variation across cities to identify the aggregate effects of highway infrastructure.

I use firm-level data to examine how a city's production efficiency responds to an expansion of highway network. I use the standard procedures proposed in the industrial organization literature (Olley and Pakes 1996; Levinsohn and Petrin 2003; Ackerberg et al. 2015 and Wooldridge 2009) to estimate firm TFP. Although there is a large and growing literature in international trade on firm adjustments after trade liberalization, there are very few papers that have looked at firm-level adjustments and its aggregate implications after a large-scale transportation infrastructure project. I find that average firm TFP increased as a city gained market access, which is consistent with models such as Melitz (2003) and Chaney (2008). I do not find strong evidence for within-firm productivity growth.

I then conduct a decomposition exercise to examine the channels through which the highway infrastructure promoted productivity growth in China. I follow Haltiwanger (1997) and decompose changes in aggregate TFP at the industry and city level into four components: within-firm productivity growth of continuing firms, the reallocation of market shares from less-productive continuing firms to more-productive continuing firms, the entry of productive firms and the exit of inefficient firms. The decomposition exercise suggests that the entry of new and productive firms contributed most to TFP gains. The reallocation among large incumbents and exit of inefficient firms also contributed to TFP growth.[10] I find similar effects of access to the international market through domestic transportation cost reductions.[11] These findings are broadly in line with the findings in the international

include the railway network as a control variable.

[10]The exit of inefficient firms made the smallest contribution to TFP gains, which suggests that exit decisions may be subject to other economic and political factors.

[11]Similar to the results from the firm-level regressions, I do not find strong evidence for within-firm productivity gains from an increase in domestic market access. I do find that a reduction in transportation costs to ports promoted within-firm TFP growth.

trade literature (Brandt et al. 2016).

To evaluate the aggregate economic impact of China's national highway system, I choose the scenario of removing all highways as the baseline counterfactual.[12] I calculate the counterfactual trade costs between all city-pairs and cities' counterfactual market access in absence of highways. Based on the estimated impact of market access on productivity and the calculated decline in cities' market access in the counterfactual scenario, I estimate that eliminating all highways in China would decrease aggregate productivity by 3.2%. If we allow highways to influence the distribution of population across cities, the estimated impact of highways on productivity is 3.8%. The counterfactual analysis suggests a sizable effect of highway infrastructure on aggregate TFP.

To evaluate the effects of a transport infrastructure project, we need to deal with the threat to identification posed by endogenous highway assignment. I employ an instrumental variable approach to address the potential endogeneity of both the *location* and *timing* of highway assignment. I construct a *time-varying* least-cost path spanning tree network to instrument for the expansion of China's highway network over the period 1995–2005. I borrow from engineering and network theory to construct the hypothetical network. I then construct the market access measure with the least-cost path network and use it as an instrument for the actual city market access. The results from two-stage least squares (2SLS) regressions confirm my findings from the baseline OLS regressions. I perform various robustness checks to test the validity of my results.

I also examine the heterogeneous effects of the national highway system on the specialization pattern of industries. I examine different dimensions of industry characteristics that have been emphasized by the classical trade theories such as Heckscher–Ohlin model and the New Economic Geography models. In particular, I examine the differential effects of market access on industries with different transportation costs, capital intensities, and product differentiation. I find that industries with larger transportation costs and higher capital intensity tend to concentrate in locations with better highway access. These findings improve our understanding of are consistent with New Economic Geography models and comple-

[12] Donaldson and Hornbeck (2016) develop a general methodology for evaluating the aggregate economic impact of transport network projects.

ment empirical studies in urban economics that investigate the geographic concentration of industries (Frye 2014).

This paper is related to several strands of literature in urban and spatial economics. First, this research is related to a recent and rapidly-growing empirical literature on the economic impacts of transport infrastructure. Donaldson (2016) is a seminal paper that investigates the effects of India's railroad construction in the colonial era on trade, prices and income. Ghani et al. (2016), Alder (2015) and Asturias et al. (2015) focus on the effects of the Golden Quadrilateral highway project in India. For China, Faber (2014)[13] and Baum-Snow et al. (2015) look at the effects of China's highway construction on local GDP and population growth. Banerjee et al. (2012) examines the effect of railway network on per capita GDP levels and growth in China. Second, this paper is also related to classic theories of international trade and new economic geography theory pioneered by Krugman (1991). Hanson (2005), Redding and Venables (2004), Hanson and X. (2004) and Donaldson and Hornbeck (2016) are influential empirical studies that connect the models to the data tightly.

My paper also expands the literature on trade barriers and firm-level adjustments. Recent literature has been increasingly focusing on how firms respond to trade liberalization. This trend originates from theoretical work by Melitz (2003) and Bernard et al. (2003), which introduce a new margin of gain from trade–reallocation of resources from the less productive firms to the more productive firms. Recent papers in urban economics emphasize the sorting of firms across space (Behrens et al. 2014; Gaubert 2014. Lileeva and Trefler 2010 and Bustos 2011a examine the effects of trade liberalization on firm productivity and exporting behavior. It is little known, however, how domestic market integration through expansions of transportation infrastructure affects production efficiency of firms. This study departs from the literature and aims to shed light on how productivity responds to a reduction in trade costs due to highway connection.

The remainder of the article is organized as follows. Section 2 briefly discusses the background of the construction of China's national highway system. Section 3 presents a multi-city trade model with heterogeneous firms to illustrate how transportation infrastruc-

[13]Contrary to Faber (2014), in this paper I treat a city center along with its periphery as one metropolitan area. I do not investigate how the spatial organization of production within each metropolitan area.

ture affects firms and industries. Section 4 describes the data I use and how I construct transportation costs and market access at the city level and industry level from the data. Section 5 discusses the instrumental variable and the identification strategy I use. Section 6 presents and discusses my findings on aggregate TFP. Section 7 discusses findings on the spatial reallocation of industries. The last section concludes.

1.2 Background

China has shown greater interest in investing in its domestic infrastructure than have many other large developing countries over the past two decades. Some argue that large investments in transport infrastructure contributed to China's "growth miracle". China experienced a period of rapid highway expansion in the late 90s to mid-2000s, from virtually no highway to an integrated national highway network. In 1992, the Chinese State Council approved the construction of the "7-5" network. The objectives stated by the government were to connect all provincial capitals and cities with an urban population above 500,000 through a National Trunk Highway System (NTHS) by 2020. The plan outlined the construction of 12 trunk highway roads, including five longitudinal roads and seven latitudinal roads. Most of the projects, however, were completed during a 6-year period from 1998 to 2003. For example, Faber (2014) estimates that 81% of NTHS opened to traffic between mid-1997 and end of 2003. The completion of that project marked the beginning of an integrated domestic market supported by a national highway system.[14] Figure 1.3 shows the original plan for the NTHS.[15]

The drastic expansion of Chinese cities' access to the international market, especially in coastal areas, is equally pronounced. This expansion stirred the movement of goods, services

[14]The majority (76%) of the NTHS is expressways, the rest are highways.

[15]there were many local constructions of highways which did not belong the the National Trunk Highway System. When we construct the measure of market access, I use both the national trunk roads and local highways. When I construct my instrumental variable, I only use the trunk roads, as the construction of those trunk roads were much less subject to local economic conditions.

and people across cities. As shown in Figure 1.4, the total value of inter-provincial trade as a percentage of GDP increased from 82% to 108% for the period from 2002 to 2007. The rapid expansion of the highway network is also evident in the change in average travel distance for different transport modes. Road transportation experienced the largest increase in average travel distance among all transport modes. In addition to the reduction in transportation costs from highway expansion, China also experienced a massive trade liberalization during the same period of time. The country's WTO accession at the end of 2001 represented a large milestones for China's effort to liberalize trade and domestic market. The importance of international trade increased thereafter until the outbreak of the global financial crisis. Although WTO accession and associated policy changes were largely a national experiment, the impact of the trade reform differs across cities. This national shock allows us to use the city-level data to assess the effects of international market access on industrial clustering and production efficiency from a reduction in domestic transportation costs[16].

1.3 A Multi-City Trade Model with Heterogeneous Firms

1.3.1 Preferences

Consumer preferences in city k are defined over the consumption of goods produced in sectors $s \in 0, 1, ...S$:

$$U_k = Q_0^{\beta_0} \prod_{s=0}^{S} Q_{i,s}^{[\sigma_s/(\sigma_s-1)]\beta_s}, \qquad \sum_{s=0}^{S} \beta_s = 1, \beta_s \geq 0. \tag{1.1}$$

$$Q_{i,s} = \left[\int_{\omega \in \Omega_s} q_s(\omega)^{(\sigma_s-1)/\sigma_s} d\omega \right]^{\sigma_s/(\sigma_s-1)}, \quad \sigma_s > 1, s \geq 1. \tag{1.2}$$

A homogeneous good 0 is produced in every city, is freely traded and is used as a numeraire. It is produced under constant returns to scale and one unit of labor in city k can

[16] In the regressions below, I control for industry-year fixed effects to account for effects stemming from tariff and other industry-specific policy changes

produce w_k units of good 0. Its price is set to 1 so the wage of city k is w_k.[17] In this case, the wage in each city is exogenously determined by the city's efficiency in producing the numeraire good.

1.3.2 Production

Each firm uses labor to produce a variety. I assume firms pay a fixed cost of f_{kj}^s units of labor and an iceberg variable trade cost τ_{kj}^s to serve market n, and $\tau_{kk}^s = 0$ without loss of generality. Each firm uses labor to produce a variety. Costs are specified in terms of origin's labor. For a firm with productivity φ, the total amount of labor required to produce q units of a variety and sell them to market n is:

$$c_k^s = f_{kj}^s + \frac{w_k^s \tau_{kj}^s q^s}{\varphi}. \tag{1.3}$$

A representative consumer in city k solves a constrained maximization problem, and the relative demand between any two varieties can be expressed as follows:[18]

$$\frac{q^s(\omega_1)}{q^s(\omega_2)} = \left(\frac{p^s(\omega_1)}{p^s(\omega_2)}\right)^{-\sigma_s}. \tag{1.4}$$

Multiply both sides by $p^s(\omega_1)$ and rearrange:

$$\beta_s Y_k = \int_0^N p^s(\omega_1) q(\omega_1) d\omega_1 = p(\omega_2)^{\sigma} q(\omega_2) \int_0^N p(\omega_1)^{1-\sigma_s} d\omega_1 \tag{1.5}$$

If we define a price index of sector s in city j to be $P_j^s = \left(\int_0^N p(\omega_1)^{1-\sigma_s} d\omega_1\right)^{\frac{1}{1-\sigma_s}}$, then city j's Marshallian demand for any variety ω produced in city k is

$$q_{kj}^s(\omega) = \beta_s Y_j \left(P_j^s\right)^{\sigma_s - 1} p_{kj}^s(\omega)^{-\sigma_s}, \tag{1.6}$$

[17] I assume labor is perfectly mobile across sectors so w_k^s, the wage in sector s in city k, should be equal to w_k for all sectors in city k.

[18] Note that we are using CES utility function, and from (5) it is immediate that $\frac{d(q(\omega_1)/q(\omega_2))/(q(\omega_1)/q(\omega_2))}{d(U_{\omega_1}/U_{\omega_2})/(U_{\omega_1}/U_{\omega_2})} = \frac{d(q(\omega_1)/q(\omega_2))/(q(\omega_1)/q(\omega_2))}{d(p(\omega_1)/p(\omega_2))/(p(\omega_1)/p(\omega_2))} = \frac{-dln(q(\omega_1)/q(\omega_2))}{-dln(p(\omega_1)/p(\omega_2))} = \sigma_s$

Solving a firm's profit maximization problem, we arrive at the standard expression for the firm's optimal price as its marginal cost multiplied by a constant markup:

$$p^s_{kj}(\varphi) = \frac{\sigma_s}{\sigma_s - 1}\frac{w^s_k \tau^s_{kj}}{\varphi}. \tag{1.7}$$

Equilibrium revenue of the firm is given by

$$r^s_{kj}(\varphi) = \beta_s Y_k P_s^{\sigma_s - 1} p^s_{kj}(\varphi)^{1-\sigma_s}, \tag{1.8}$$

Profit is given by

$$\pi^s_{kj}(\varphi) = \frac{r^s_{kj}(\varphi)}{\sigma_s} - w^s_k f^s_{kj} = \frac{(\sigma_s - 1)^{\sigma_s - 1}}{\sigma_s^{\sigma_s}} w^{1-\sigma_s} \beta_s Y_j P_s^{\sigma_s - 1} \varphi^{\sigma_s - 1} - w^s_k f^s_{kj}, \tag{1.9}$$

Productivity Cutoff.—After observing its productivity, a firm decides on whether to exit or to sell in that market. We can pin down the export cutoff productivity $\overline{\varphi}$ by the zero-profit condition:

$$\overline{\varphi}^s_{kj} = \lambda_s \left(\frac{f^s_{kj}}{\beta_s Y_j}\right)^{1/(\sigma_s - 1)} \frac{w_k \tau^s_{kj}}{P^s_j}, \tag{1.10}$$

where λ_s is a constant[19] and f^s_{kj} is fixed cost[20]. A special case is a firm selling to its domestic market. In that case, $\tau_{kk} = 1$ and the firm will exit if its productivity is lower than the survival productivity.

$$\overline{\varphi}^s_k = \lambda_s \left(\frac{f^s}{\beta_s Y_k}\right)^{1/(\sigma_s - 1)} \frac{w_k}{P^s_k}. \tag{1.11}$$

1.3.3 Quantitative Predictions

In order to generate predictions from the model, I make a few simplifying assumptions. I first assume firm productivity is drawn from a common Pareto distribution in each city.

[19] $\lambda_s = (\sigma_s/\beta_s)^{1/(\sigma_s - 1)}(\sigma_s/(\sigma_s - 1))$

[20] For simplicity, below I assume fixed cost of exporting is independent of origins and destinations, or $f^s_{kj} = f^s$.

$$g_s(\varphi) = \gamma_s \varphi_{min}^{\gamma_s} \varphi^{-(\gamma_s+1)}, \quad G_s(\varphi) = 1 - \left(\frac{\varphi_{min}}{\varphi}\right)^{\gamma_s} \tag{1.12}$$

where $\varphi_{min} > 0$ is the lower bound of the support of the productivity distribution and γ_s is the shape parameter (lower values of the shape parameter correspond to greater dispersion in productivity). Without loss of generality, I assume productivity is distributed over $[1, +\infty]$.[21]

We also assume that the total mass of potential entrants in city j in each differentiated sector is proportional to $w_j L_j$, similar to Eaton and Kortum (2002) and same as Chaney (2008). We abandon the free entry condition in the Melitz model to simplify the analysis for now, although imposing the free entry condition will not change predictions from the model. Since we do not impose free entry, firms produce positive profits that need to be redistributed to the workers. Following Chaney (2008), I assume a global fund collects profits from all firms and redistribute them to workers in units of the numeraire good. Each worker owns w_i shares of the fund.

The price index P_j^s takes the following form:

$$P_j^s = \left(\sum_{k=1}^{N} w_k L_k \int_{\overline{\varphi}_{kj}^s}^{\infty} \left(\frac{\sigma_s}{\sigma_s - 1} \frac{w_k \tau_{kj}^s}{\varphi}\right)^{1-\sigma_s} dG_s(\varphi)\right)^{1/(1-\sigma_s)} \tag{1.13}$$

dividends per share, d, is defined as

$$d = \frac{\sum_{s=1}^{S} \sum_{k=1}^{N} w_k L_k \left(\int_{\overline{\varphi}_{kj}^s}^{\infty} \pi_{kl}^s(\varphi) dG_s(\varphi)\right)}{\sum_{n=1}^{N} w_n L_n}, \tag{1.14}$$

$$Y_j = (1 + d) \times w_j L_j \tag{1.15}$$

To unburden us from the heavy notation, I ignore the industry subscript s from now on. The other industries are analogous. Given the productivity cutoff $\overline{\varphi}$,

$$P_j = \kappa_1 Y_j^{1-\sigma} \Theta_j; \tag{1.16}$$

[21] If we assume the Pareto distribution is bounded from above, we will need to assume that $\gamma_s > \sigma_s - 1$ for firm size to be finite.

where $\Theta_j^{-\gamma} = \sum_{k=1}^{K}(Y_k/Y) \times (w_k\tau_{kj})^{-\gamma} \times f^{-[\gamma/(\sigma-1)-1]}$. This term is very similar to the "multilateral resistance variable" coined by Anderson and Van Wincoop (2003).

If a firm's productivity is larger than the cutoff $\overline{\varphi}_{kj}$, then export value is given by

$$x_{kj} = \kappa_2 \times \left(\frac{Y_j}{Y}\right)^{(\sigma-1)/} \times \left(\frac{\Theta_j}{w_k\tau_{kj}}\right)^{\sigma-1} \times \varphi^{\sigma-1}. \tag{1.17}$$

plugging this into (10), we have

$$\overline{\varphi}_{kj} = \kappa_3 \times \left(\frac{Y}{Y_j}\right)^{1/\gamma} \times \left(\frac{w_k\tau_{kj}}{\Theta_j}\right) \times f^{1/(\sigma-1)} \tag{1.18}$$

Productivity cutoff in city k is given by

$$\overline{\varphi}_k = \kappa_4 \times \frac{w_k}{Y_k^{1/\gamma}} \times \left(\sum_{j=1}^{N} \frac{w_j^{-\gamma}Y_j}{\tau_{kj}^{\gamma}}\right)^{1/\gamma} \tag{1.19}$$

The last term on the right resembles the market potential term coined by Harris (1954) and market access term in Redding and Venables (2004) and Donaldson and Hornbeck (2016). More specifically, a city that can reach big cities (large Y_k) in a cheap way (small τ_{kj}) tends to have higher productivity cutoff, meaning that the average firm productivity in the city tends to be higher. Later we introduce the data and methods to estimate this market access term.[22]

The average firm productivity in a city is defined as the following

$$\overline{TFP}_k = \int_{\overline{\varphi}_k}^{\infty} \varphi dG_k(\varphi) = \frac{\gamma}{\gamma-1}\overline{\varphi}_k = \frac{\gamma}{\gamma-1}\kappa_4 \times \frac{w_k}{Y_k^{1/\gamma}} \times \left(\sum_{j=1}^{N} \frac{w_j^{-\gamma}Y_j}{\tau_{kj}^{\gamma}}\right)^{1/\gamma} \tag{1.20}$$

The aggregate productivity in a city is defined as a revenue-weighted TFP:

$$\widehat{TFP}_k = \int_{\overline{\varphi}_k}^{\infty} \varphi s(\varphi) dG_k(\varphi) = \frac{\gamma_s}{\gamma-\sigma}\overline{\varphi}_k^{\sigma} = \frac{\gamma}{\gamma-\sigma}\kappa_4^{\sigma} \times \left(\frac{w_k}{Y_k^{1/\gamma}}\right)^{\sigma} \times \left(\sum_{j=1}^{N} \frac{w_j^{-\gamma}Y_j}{\tau_{kj}^{\gamma}}\right)^{\sigma/\gamma} \tag{1.21}$$

[22] In the Appendix, we introduce a closely related concept of market access that is adopted in Redding and Venables (2004) and Donaldson and Hornbeck (2016). The results remain largely unchanged with their definitions.

where κ_1, κ_2, κ_3, κ_4 are all constants, and $s(\varphi)$ is the share of a firm's revenue in total industry revenue in city k.

As mentioned above, the term $\sum_{j=1}^{N} \frac{w_j^{-\gamma} Y_j}{\tau_{kj}^{\gamma}}$ is closely related to the "market access" term that I will construct in Section 4. The derivations from the model show that firms located in cities with larger market access tend to have higher productivity on average, and cities with larger market access tend to have higher aggregate productivity. If a transport infrastructure project reduces the transportation costs between a city and its neighbors hence increases the market access of a city *relative to* other cities, we should expect to see a *relative* increase in the city's production efficiency.[23] In the model, the efficiency gain is a combination of the entry of new and productive firms, the exit of inefficient firms and the reallocation across firms. In Section 6, I will examine the relative importance of each channel.

1.4 Data and Measurement

There are mainly four data sources I use for this project. First, I use digitized maps of China's highway network system for the period from 1992 to 2015 in combination with local non-highway, railway and waterway maps.[24] The vector data representing China's expressways, highways and local paved roads was built up over many years from a wide variety of published road atlases. That compilation of information on the current expressway routes has then been coded to indicate the type and status of expressways that were depicted in road atlases. The early road atlases from the last century have maps that are quite small scale, whereas

[23] In the model, I assume labor is immobile. If labor is perfectly mobile, then we should not expect to see a positive relationship between city market access and production efficiency in the data because the relative increase in demand for labor would be completed offset by new immigrants. The reality in China was that labor were somewhat mobile but with strong restrictions. In that sense, the estimated effect of highways below should serve as a lower bound for the actual effect.

[24] The data mainly come from the Australian Consortium for the Asian Spatial Information and Analysis Network (ACASIAN). For data inquiries, please contact lcrissman@optusnet.com.au. I would like to thank Prof. Dai at the Geology Department of Beijing Normal University for helping me get the newest transport route map.

the better atlases from more recent years have much more detail. As a result, information in older atlases were coded onto the routes shown in more recent ones to avoid any discrepancy over time. I use these digitized highway maps to construct least cost path between any two cities for 1995, 2001, and 2005. The construction of least cost paths and choice of years are explained below. Figure 1.5 present highways construction in China for the period of this study.

Second, firm-level data and industry-level data are constructed from the Annual Survey of Industrial Enterprises (ASIE) from 1998 to 2007. The survey conducted by the National Bureau of Statistics of China span the period from 1998 to 2009. The survey contains all State-Owned Enterprises (SOEs) and all private enterprises with annual sales of $650,000 and above. The data set contains very detailed information on firm's balance sheet and income statement, as well as information on ownership, export status, employment among others. For the year 2004, we have information on average education and skill level of labor. There are over 100 variables in the data. The most useful variables for this project are firms identifier, industry identifier, gross output, total sales, wage bill, employment, stock of fixed capital, value of intermediate inputs, export status, year of establishment, ownership, skill/education level and location identifier. I also use Economic Census for 1995, 2004, 2008 and Basic Units Census for 1996 and 2001 to construct aggregate output and employment for each industry at the prefecture city level.

Finally, to estimate bilateral transportation costs between any two cities, I collect prices quoted by logistics companies for freight transportation between any two cities in China. More specifically, I gather all price quotes from two of the largest logistics companies in China from their websites[25] to estimate transportation costs between any two cities. These two companies cover freight transportation over most of the cities in China. I will argue below that price quotes they publish are good estimates of the actual costs that firms or their clients have to pay to deliver the goods. By using actual transportation costs data, I

[25] I use quotes from Arima World Group (website: www.hoau.com) and Deppon (website: www.deppon.com). Please see Appendix A.2. for details of the price information on their websites. They are two of the largest logistics companies in China. They both have a very extensive network in the whole nation.

contribute to the literature by estimating the bilateral transportation costs between Chinese cities.

1.4.1 Estimation of Bilateral Transportation Costs

To construct a measure of market access for each city, it is necessary to first gauge the cost of transporting goods between cities. A challenge in the literature is the lack of reliable estimates of transportation costs. Some of the earlier papers use the Euclidean distance between locations to proxy for transportation costs. There are two problems with this approach. First, Euclidean distance is a poor proxy for transportation costs. Euclidean distance is not able to take into account road availability and conditions, topography and many other factors that will affect transportation costs. Second, Euclidean distance is time-invariant and thus prevents researchers from using changes in transportation costs over time to evaluate the impact of a transport infrastructure project.

Some of the recent work in this literature uses digitized transport route data to estimate transportation costs. With digitized transport route data, the most widely used approach is to estimate travel distance or time first and then to estimate transportation costs between locations with parameter assumptions, such as dollars per ton-mile (Donaldson and Hornbeck 2016) or dollars per hour (Baum-Snow et al. 2015). This approach is arguably a substantial advance since transport infrastructure is a key determinant of trade costs between locations and has attracted considerable interest in the literature. However, there are still a number of pitfalls associated this approach. First, the assumed per unit cost parameters are usually taken from previous studies that investigate a different question in a different country for a different time period. There are many potential issues in using, for example, overland shipping costs in the US in the 1980s to investigate the current Chinese market, which could undermine the relevance of the estimated transportation costs. Second, this simplistic approach ignores many other factors that we are known to affect transportation costs, such as congestion, economies of scale and local labor costs. It is the acknowledgment of these pitfalls that lead me to adopt a different strategy for estimating transportation costs in this paper. One contribution of this paper is to estimate the transportation costs between any two cities in China with data on actual transportation costs.

As mentioned at the beginning of this section, I collect all price quotes from two of the largest logistics companies in China. The data indicate how much the seller or the buyer has to pay to have the product shipped from the factory to the user. The price quotes incorporate all information, observable or unobservable, such as travel distance, road availability, congestion, costs of labor, gasoline and other inputs. Thus, in contrast to previous papers, I measure the actual transportation costs between any two cities in China. A potential drawback of using the price quotes is that companies may not strictly follow the quotes when they finalize the contracts with their clients. I checked with representatives from these two companies and with some of their clients to determine whether the price quotes they publish on their websites truly reflect the actual transaction prices. To the best of my knowledge, the two companies very closely follow the price quotes in practice. An established client will be able to secure a 10%-20% discount, while the logistics companies occasionally award discounts as high as 50% to very large orders. Although they do exercise discretion on an order-by-order basis, the price quotes are reasonably good estimates of the actual shipping costs that a firm needs to pay if it chooses to use one of the two logistics companies. There are many other smaller logistics companies in China, and many of the smaller companies only maintain a local presence. It is impossible to collect price quotes from all logistics companies in China, but to the extent that there are many companies and the market is competitive, I will assume that the price quotes from these two largest companies reflect the "market price".[26] Since I need to estimate iceberg trade costs, I also need to obtain value-to-weight ratio of manufacturing goods. I use revenue and quantity data of Chinese manufacturing firms to estimate the average value-to-weight ratio for each manufacturing industry.

Table 1.1 shows some of the estimated shipping costs between cities. The first lesson we learn from the table is that shipping costs vary widely within China. We see that for some cities that are close to each other and/or have easy highway access, shipping costs between them are only a very small fraction of total value of manufacturing goods. However, for many cities that are far away from each other or do not have highway access, then shipping

[26]For any city pair, I use the lower price of those offered by the two companies as the actual shipping costs that a firm has to pay to transport goods between the two cities, conditional on firms using the same class of delivery service

costs can be as high as 30% of total goods value. An example would be that the estimated shipping costs between the two largest cities in China, Beijing and Shanghai, is 5.5% of total goods value. Second, these estimated costs are very different from the parameters assumed in the literature. For example, Limao and Venables (2001) find that the cost of shipping 1 ton of freight overland for 1000 miles is about $2,100, or about 2% of its value. Some of the papers that use these estimates will end up estimating a much smaller transportation cost[27].

While I can collect the current price quotes from the two companies' websites, they do not publish any information on historical prices. To estimate transportation costs for previous years, I employ a simple linear regression model incorporating both macro data and origin- and destination-specific data. To obtain transportation costs, I run a simple regression model to estimate transportation costs for previous years.

The first step is to estimate the following regression specification:

$$\tau_{o,d,15} = \alpha + \beta_1 hwy_{o,d,15} + \beta_2 local_{o,d,15} + fe_o + fe_d + \epsilon_{o,d,15}$$

where $\tau_{o,d,15}$ is the iceberg transportation cost between city o and d in 2015, expressed as a percentage of goods value. $hwy_{o,d,15}$ is the length of highway on the optimal route, whereas $local_{o,d,15}$ is the length of local roads on the optimal route in 2015. By estimating this regression specification, I obtain $\hat{\alpha}$, $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{fe}_o$, $\hat{fe}_d$, which are estimates for α, β_1, β_2, fe_o, fe_d,respectively. The second step is to estimate transportation costs between origin and destination.

$$\hat{\tau}_{o,d,t} = \hat{\alpha} + \hat{\beta}_1 hwy_{o,d,t} + \hat{\beta}_2 local_{o,d,t} + \hat{fe}_o + \hat{fe}_d$$

One way to assess the quality of estimated historical transportation costs is to examine the goodness-of-fit of the regression. In Column (2) of Table 1.5, I find that the model can explain 85% of the variation in bilateral transportation costs. If I exclude all of the origin and destination fixed effects and only use the road data, the model with road data

[27]For example, an earlier version of Baum-Snow et al. (2015) estimates that travel along the 1990 road network from Beijing to Shanghai would cost 2% of value

alone explains 69% of the variation in transportation costs (see Column (1) of Table 1.5). I also add nonlinear terms of highway and non-highway length, as well as total travel time and its nonlinear terms to control for factors that are not collinear with highway and non-highway distance. In Columns (3) and (4) of Table 1.5, I find that the R-squared does increase, but only marginally. These results suggest that road data can effectively predict transportation costs, and city-specific factors such as labor costs and congestion also matter for transportation costs. The results offer some confidence that this simple linear regression model with road data can estimate historical transportation costs reasonably well. Please also note that it is the *relative*, not the *absolute* transportation costs that matter for my purpose. Any common factors such as oil price and wage growth that affected the entire country symmetrically will not invalidate my analysis. One may be temped to estimate the above regression at the industry level if one believes that the cost structure is heterogeneous across industries. However, the price quotes from the logistics companies do not distinguish between product groups. Thus, there is no need to separately estimate transportation costs for each industry.

1.4.2 Construction of Market Access

The idea of "market access" dates back to Harris (1954). Harris argued that the potential demand for goods produced in a location depends on the sum of distance-weighted GDP from all locations. Mathematically, Harris's "market potential" term equals $\sum_d (d_{od})^{-1} N_d$, where d_{od} is the distance and N_d is location's population or GDP. Redding and Venables (2004) and Donaldson and Hornbeck (2016) derive a similar term–"market access"–that measures each location's proximity to markets. To construct market access for all 340 prefectures in China, I assemble a prefecture-level dataset of employment and the number of firms in all manufacturing industries. To assess the effects of highway access on productivity, I use data for all manufacturing firms in China. Over the period that I study, the administrative boundaries of many prefectures changed because counties were occasionally reassigned to a different prefecture. I establish a county-level correspondence from 1995 to 2008 and construct time-consistent prefecture boundaries.

Table 1.1: Transportation cost (% of Good Value)

Panel A: Low Iceberg Transportation cost		
origin	destination	Iceberg Cost
Yangquan	Taiyuan	1.04%
Liangshan	Ya'an	1.30%
Chengdu	Zigong	1.30%
Kaifeng	Zhengzhou	1.51%
Panel B: High Iceberg Transportation cost		
origin	destination	Iceberg Cost
Huainan	Lasa	20.32%
Hetian	Wuzhou	20.97%
Lasa	Kelamayi	23.70%
Shannan	Bayingguole	28.11%
Panel C: Between Beijing and Shanghai		
Beijing	Shanghai	5.54%

Notes: This table shows iceberg trade costs between city pairs. Panel A shows the top-5 cities pairs that have the lowest transportation costs. Panel B shows the bottom-5 city pairs that have the highest transportation costs. Panel C shows the transportation costs between Beijing and Shanghai, the two largest cities in China.

After the massive and rapid expansion of highway network during the period 1998-2005, China established the National Trunk Highway System (NTHS). Thanks to this massive infrastructure project, highway transportation gradually became the main transportation method for inter-city trade after 2000, whereas railroads were little changed and have been used primarily for transporting commodities such as coal and metal since then. In fact, estimates from different sources suggest that road transportation accounted for nearly 70% of freight value and logistics costs in 2007. Since I only consider manufacturing industries from 1998 to 2007, it makes sense for me to focus on the highway "shock" to Chinese cities. Therefore, I use digitized highway network data and local road data to construct each city's market access after China's highway network expansion. As mentioned above, I control for access to railways and waterways in my empirical specification.

I use firm census data for the years 1995, 1996, 2001 and 2008 to construct employment, the number of firms and sales for 42 2-digit manufacturing industries at the prefecture level. The reason for only using manufacturing firms is that it is generally accepted that manufactured goods are much more tradable than services. For example, a KPMG report estimates that the movement of industrial products accounted for 87% of the value of goods moved in 2006. I use the Annual Survey for Industrial Enterprises (ASIE) from 1998 to 2007 to estimate firm-level productivity. I use the standard Olley-Pakes estimation method to estimate firm TFP.[28] I also use firm-level census data to construct output and employee data at the industry-prefecture level from 1995 to 2008. The reason that I use the average growth rates from the period 2001-2008 is twofold. First, China did not have a national highway network until the early 2000s. Thus, by using the period from 2001 to 2008, I ensure that my measures of market access using highway data truly capture proximity to markets and supplies. Second, the Chinese government announced a 4-trillion-yuan stimulus package after the global financial crisis began in 2008. In the following years, they committed considerable resources to building roads as part of the stimulus package in an effort to boost domestic demand. Thus, the period after 2008 might be affected by the crisis and the stimulus package, which might need additional attention.

[28] There is now a large literature on the estimation of firm productivity. Other popular methods include Levinsohn and Petrin (2003), Ackerberg et al. (2015) and Wooldridge (2009).

As pointed out by Baum-Snow et al. (2015), there were very few expressways before 2000 and almost all long-distance shipping was through railways or waterways. With the rapid and massive expansion of national highway network from 1998-2003, road transportation emerged as the prominent transportation method except for a few commodities. In fact, coal alone consists of roughly 50% of inter-provincial trade by railway in 2007. Coal and a few other commodities such as iron ore, lumber and other metals account for over 90% of freight transportation by railway [29]. Moreover, China's railway network has been little changed since 1990. In this paper, I only consider manufacturing industries, which should further alleviate the concern that constructing market access solely based on roads data may distort results. In addition, I include access to railways and waterways in the regressions to address any potential bias. The effects of the highway network expansion on regional trade can be seen in Figure 1.4 inter-provincial trade exploded since early 2000, reaching 106% of GDP from 85%. Even though there is no data on inter-prefecture trade, it is very likely that the increase in trade was even more drastic. Since then, China continued to expand the reach of its highway network. The average transport distance of highways nearly tripled from 60 kilometers in 2000 to 172 kilometers in 2008, whereas the average transport distance of railways and waterways did not change at all and stayed at 760 kilometers and 1800 kilometers respectively during the same period. Figure 1.5 shows how China's highway network evolved over the years.

The market access term derived in Section 3 resembles the "market potential" approach of Harris (1954) and the "market access" approach developed by Redding and Venables (2004) and Donaldson and Hornbeck (2016). The Harris approach is simple, whereas the Donaldson and Hornbeck approach requires numerically solving 318 nonlinear equations simultaneously. Donaldson and Hornbeck (2016) verify that the numerically solved market access term is highly correlated with the simpler market potential term, and the results do not depend on what term they use. In the model, I show that productivity is determined by a "market potential" term, whereas industry specialization patterns are more closely related to the "market access" term. Henceforth, I use the term "market access" to refer to a location's proximity to suppliers and consumers, and use the market potential term as

[29]Data used for calculating these statistics are from the National Bureau of Statistics

my baseline. I also numerically solve for the market access term and verify that these two approaches generate similar results. I use the following equation to estimate each location's market access:

$$MA_o \approx \sum_d \tau_{od}^{-\theta} Y_d,$$

where $\tau_{od,t}$ represents the iceberg trade costs and Y_d is the size of the economy of destination d. Redding and Turner (2015) note that most shipments cover over very short distances and that the time cost of freight seems to be important. This is why I choose the least-time-cost paths as the "optimal" route for ground transportation. The choice of routes is also confirmed by representatives from logistics companies in China.

What I have constructed is *domestic* market access. Cities also trade with the rest of the world, and this is especially true for the cities in coastal areas. Two forces increased Chinese cities' access to the international market during the period of study. First, tariffs declined rapidly following China's accession to the WTO at the end of 2001. Second, highway connections reduced the domestic transportation costs from the origin cities to the ports. Since I am interested in transportation infrastructure in this paper, I control for industry-year fixed effects in the empirical specification to abstract from tariff effects.[30] I use the transportation cost from the origin city to the nearest port[31] to measure the city's access to the international market.

A concern associated with using only highway data is that railways also played an important role in freight transportation in China, especially before the highway era. Even at present, a very large share of coal, coke and metals are transported via rail, not highways. While short-distance travel is dominated by highway transportation, rail remains important for long-distance travel. However, rail only constitutes a small share of transportation except for those commodities mentioned above. A first argument against this concern is that

[30]The magnitude of the tariff reduction was industry specific. A caveat is that the HS classification used for exports and imports differs from the Chinese Industry Classification.

[31]I calculate the transportation costs from every city to 9 of the largest ports in China. Then, I select the lowest cost from the 9 for each city.

my empirical strategy identifies the effect of market access over the time dimension. The railway system for freight transportation only changed very slightly over the period of study (Baum-Snow et al. 2015). Therefore, any time-invariant effect of railways on productivity and industry specialization should be absorbed by the city fixed effects. However, one may still be worried any time-varying effect of the almost static railway system. To address this issue, I multiply cities' access to railways by a year dummy and include this interaction term in my empirical specifications to allow for potentially time-varying effects of railways.

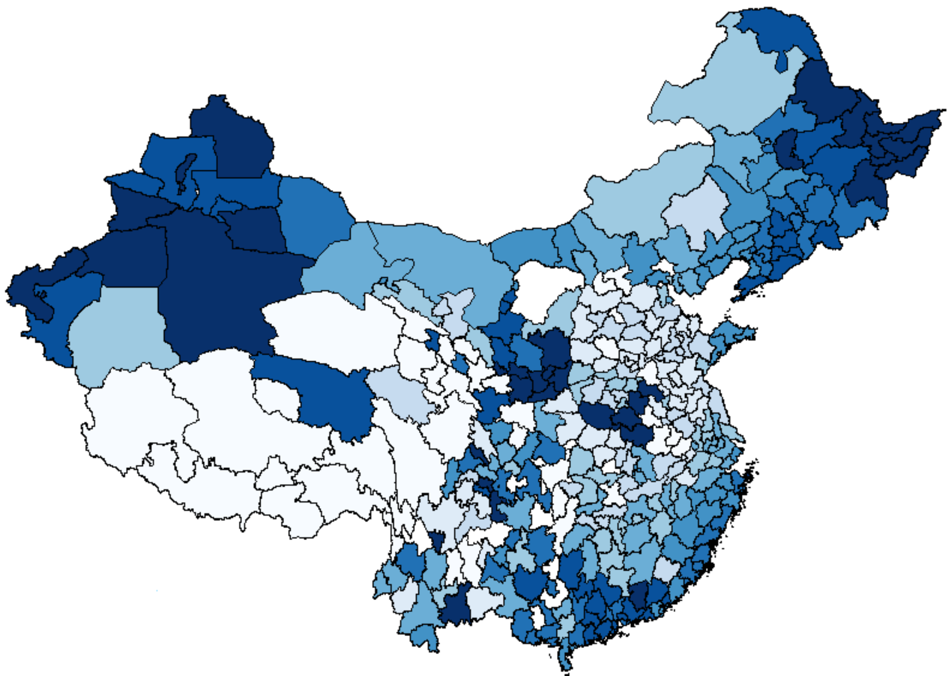
One may also be concerned about industry heterogeneity in transportation costs. Indeed, the iceberg transportation costs depend on the weight-to-value ratio of the industry and can differ markedly across different industries. I use the firm production quantity data set along with ASIE to estimate the weight-to-value ratio for each industry. Then, I construct city- and industry-specific market access. I verify that the estimated effects using industry- and city-specific market access are very similar to the city-level market access measure I construct. Figure 1.1 shows the change in market access for all Chinese cities from 1995 to 2005. Figure 1.6 presents the market access measure I construct for each city in 1995, 2001 and 2005.

1.4.3 Firm Productivity and Markups

Similar to Feenstra et al. (2014a), I use the augmented Olley and Pakes (1996) approach to estimate and calculate the TFP.[32] There are two approaches to estimating firm TFP. One is the Olley-Pakes approach that uses value added, and the other is the Levinsohn and Petrins approach that uses total output. Feenstra et al. (2014a) argues that the Olley-Pakes approach is more appropriate in the Chinese context because processing trade in China accounts for more than a half of the country's total trade since 1995. The prices of imported intermediate inputs are different from those of domestic intermediate inputs. Using the domestic deflator to deflate imported intermediate input would create another unnecessary source of estimation bias. A potential issue with the Olley-Pakes approach is that a large number of firms that have zero investment will be dropped from the estimation exercise.

[32]All nominal variables are deflated by input and output deflators. Deflators are taken from Brandt et al. (2016). I use the perpetual inventory method to construct real capital stock, similar to Brandt et al. (2016).

Figure 1.1: Estimated Changes in Market Access



Notes: this figure shows the changes in market access from 1995 to 2005 for 339 Chinese cities. Darker color means larger increase in market access. We can see that areas in the west and the northeast gained large increases in market access. A few cities in the central area also experienced large increases in market access.

However, As shown in Brandt et al. (2012), in the Chinese data there is only negative real investment for 1% of continuing firms. Moreover, I do not observe investment decisions directly, but estimate investment from the capital stock series, which will smooth out most of the zero investment decisions. I follow De Loecker and Warzynski (2012) to estimate firm markups. Since I use the value-added approach to estimate TFP, I choose labor as the "flexibly adjustable" input.[33] I also back out prices, marginal costs and physical productivity for a subset of firms that report quantity data. Prices and marginal costs will be used in Section 6 to disentangle gains retained by firms from gains passed on to consumers. Table

[33]Although it is more appropriate to use the value-added approach to estimate firm TFP, it is somewhat difficult to justify the assumption that labor input is fully flexible in China. Therefore, I also employ the output approach to estimate firm TFP and use material inputs as the "adjustable" input. These results are consistent with my baseline results.

1.6 presents the estimated industry-level TFP by aggregating estimated firm TFP.

1.5 Instrumental Variable Approach

One of the greatest challenges in the literature is the endogenous placement of transport infrastructure, which may bias the estimated effects of transportation infrastructure. The official documents from the Chinese government state that the objective of the National Trunk Highway System was to connect all cities with populations over 500,000. It is highly probable that when choosing a route to connect two large cities, policymakers and urban planners would choose to build highways near the cities that they expected to have higher economic growth for reasons unobserved by the researcher. Our OLS estimates would have an upward bias in the existence of such correlation. Another possibility is that the location and funding decisions for highways were the product of a bargaining process between the central and local governments. Cities that had better political connection could get preferential policies from the central government. It is possible that highway was just one of the placed-based preferential policies during my period of study that shaped the subsequent development of the preferred cities. If that was the case, then the effect of market access on TFP may also be biased upward.

Researchers have proposed historical roads (Donaldson 2016; Baum-Snow et al. 2015), planned networks (Baum-Snow 2007; Michaels 2008; Duranton and Turner 2011; Duranton et al. 2014), and algorithm-generated networks (Faber 2014) as instruments for actual highway or railway networks. The validity of these instruments hinges on whether these networks only influence productivity growth and industry specialization patterns through their predictive power for the actual transportation network conditional on control variables. In other words, these instruments may fail the exclusion restriction if they are correlated with unobserved economic fundamental or policy variables which also affect productivity growth and spatial reallocation of industries.

To address the identification issues discussed above, I construct a time-varying least-cost path spanning-tree network as an instrument for actual highway connections to evaluate

China's highway expansion.[34] Note that I need to construct a time-varying instrument since the goal is to evaluate the effects of the NTHS on productivity and industry specialization. Thus, I also instrument for the *timing* of highway placement, which could also be endogenous as planners might choose to build roads near their most preferred cities first. The construction of the time-varying instrument is executed in two steps. In the first step, I use the Kruskal minimum spanning tree algorithm to calculate the least-cost path connections between any two city centers.[35] I use remote sensing data on terrain ruggedness collected by satellite images to estimate the construction cost of each small piece of land in China. Please refer to Appendix A.4. for all the estimated construction costs. I then construct least cost path connections for any two cities in China. Given all the least cost path connections, a least-cost path spanning-tree network is constructed that connects all the node cities and minimizes construction costs at the same time.

In the second step, I use the Girvan-Newman algorithm from network theory to predict the optimal timing of the construction of each connection.[36] The Girvan-Newman Algorithm ranks each edge by counting the number of shortest paths that move along that edge. The top-ranked edges are the most "important" ones to the network and should be built first. After the sequence of construction has been solved by the algorithm, I determine highway construction on the least-cost path network for each year. To do so, I first calculate the *actual* length of highway construction in the entire country for a given year and take that as an exogenous variable–one can regard this as a pre-approved government budget devoted to highway construction. Given the constraint that the planner faces, he or she constructs the least-cost path network by performing constrained optimization for each year.

[34] In this paper, I am interested in the effects of market access on TFP growth and industry reallocation. A city's market access is affected by the expansion of the highway network even far away from that city. Thus, I am not testing a simple highway connection effect, which is likely to be correlated with city TFP growth. Instead, I use market access to summarize all the direct and indrect impact of transport costs on city TFP.

[35] Faber (2014) constructs a similar least-cost path network. However, my instrument varies over time and predicts the timing of highway construction.

[36] Frye (2014) uses the Girvan-Norman algorithm to construct an instrument for the US interstate highway system.

The final step is to estimate city market access with the algorithm-generated highway networks. The way to estimate the algorithm-based market access is the same as described in 4.2. The only difference is that now the least-cost path spanning-tree network is used for estimation instead of the actual highway network. We use it to instrument for the market access measure based on actual transport routes.

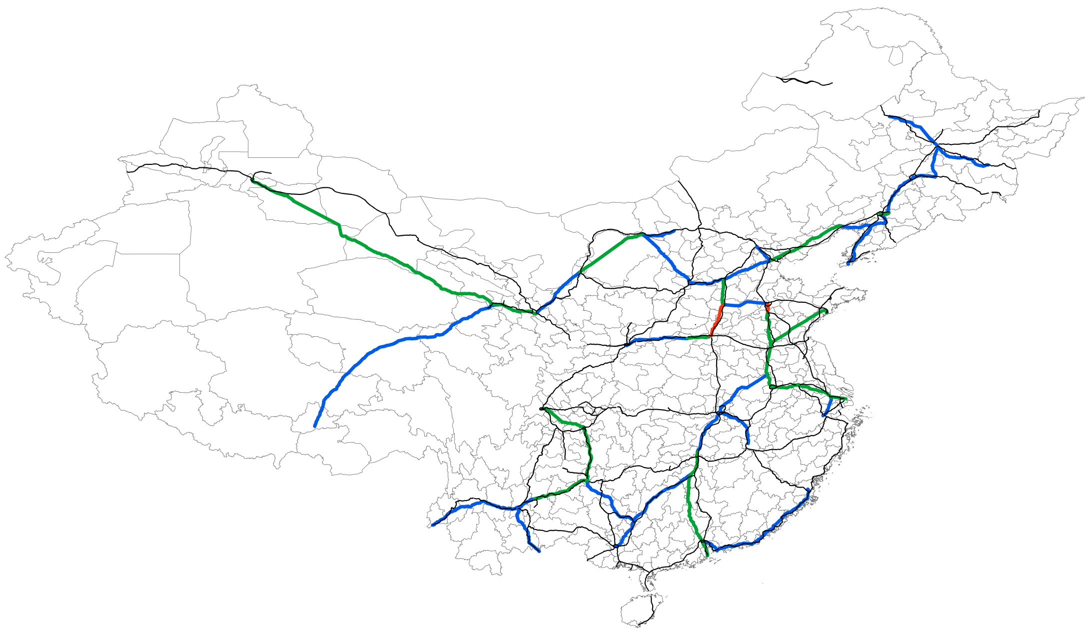
My time-varying instrument addresses the identification concerns on both the *location* and *timing* of highway connection. The location dimension of the instrument is constructed based on cost minimization. The timing dimension of the instrument is constructed by on a "centrality" measure from network theory. Figure 1.2 plots the hypothetical construction of the least-cost path network. The highway routes in red represent hypothetical construction before 1995, the routes in green represent hypothetical construction during the period 1995-2000, and the routes in blue represent hypothetical construction during the period 2000-2005. The highway routes in black are the actual National Trunk Highway Network. It is evident from the figure that the least-cost-path spanning tree network resembles the actual highway network, but the two also differ in both the location and timing of highway assignment.[37]

The identifying assumption is that the hypothetical highway network should affect city productivity and the spatial allocation of industries only through the actual highway network, conditional on time-invariant city characteristics, city population and pre-existing railway and waterway access. I will discuss two threats to my exclusion restriction. First, local construction costs may be correlated with the potential economic returns of highway connection. For instance, it is costly to build highways in a hilly region, and the economic benefits of highway connection for the region is very high because the region is not well connected by any other transport mode. The inclusion of pre-existing access to railway and waterway in the empirical specifications should mitigate this issue. I nevertheless construct an alternative Euclidean distance spanning tree network to replace the cost-based network. The Euclidean distance spanning tree network minimizes the total distance, not total construction costs of the network. None of the results are drastically different with this alternative instrument. We may also be concerned that placed-policies is correlated with the "centrality" of an edge,

[37]For the timing of actual highway assignment, please refer to Figure 1.5. Details of construction for each year is available upon request.

which is related to the "centrality" of its connecting cities. It is reassuring that I do not find evidence that city centrality is strongly correlated with pre-trends in the data.

Figure 1.2: Instrument: A Least-Cost Path-Spanning Tree Network



Notes: this figure shows the time-varying instrument. The black lines are the actual National Trunk Highway System (NTHS). The red lines are hypothetical highway construction in 1995; the green lines are hypothetical highway construction between 1996 and 2001; the blue lines are hypotheticali highway construction between 2002 and 2005. The complete network is the least-cost path spanning-tree network.

1.6 Market Access and City Productivity

In this section, I first examine whether the large increase in Chinese cities' market access raised the production efficiency with firm-level data. In Section 6.2, I aggregate TFP growth at the industry level and decompose the aggregate TFP growth into four channels and quantify the importance of each channel. In Section 6.3, I then estimate the extent of the revenue-based TFP that are attributable to changes in markups and changes in physical productivity. Section 6.4 presents robustness checks to ensure the validity of my findings.

1.6.1 Firm-Level Regressions

First, I use the following baseline specification to examine the effects of market access on firm TFP:

$$logTFP^s_{i,k,t} = \alpha + \delta logMA^s_{k,t} + \left(X^s_{i,k,t}\right)' \beta + \lambda_s + \mu_{s,t} + \epsilon^s_{i,k,t}$$

where $TFP^s_{i,j,k,t}$ is measured productivity for firm i in industry s and city k in year t. $MA^s_{k,t}$ is the market access of city k in industry s in year t.

I am interested in δ, which measures the effect of market access on firm productivity. I include industry × year fixed effects, $\mu_{s,t}$, and city fixed effects, λ_k. I also include additional control variables, $X^s_{i,k,t}$. The industry × year fixed effects account for any time-varying industry characteristics that may be correlated with the location of highway construction. As the tariff reductions were industry-specific after China's accession to the WTO in 2001, and the government's positive attitude towards active industrial policies, it is crucial to control for these potential confounding factors. By including city fixed effects, I control for any city characteristics that are not time-varying and that may be correlated with highway construction. Thus, the treatment effect of highway construction is only identified from variation within a city over time. I include city population, the interaction between distance to a railway and a year dummy, and the interaction of distance to a waterway and a year dummy as additional controls. The reason for including the interaction is because railways and waterways were little changed over the period I study. Thus, I interact them with year dummies to account for the potentially time-varying effects of these variables. I also include firm-level control variables, i.e., firm markup, firm size, ownership structure, and an exporter status dummy.

I choose market access in three periods, 1995, 2001 and 2005, to perform my empirical analysis.[38] The reason for choosing these three years is twofold. The first reason is mainly

[38]For firm-level data, I use 1998,2003 and 2007 to account for any lagged effects of highway construction. It is plausible that the effect is not immediate after the completion of the construction project for several

data driven. It is not easy to identify what exactly was on the ground, even with the officially published highway network maps. 2001 and 2005 are the years in which there are more sources, and thus, I can cross check to ensure that the truly functioning highways are in my dataset. The second reason is that China had a "structural break" within each of the two time intervals. The project accelerated after the 1997 Asian Financial Crisis, as the central government decided to invest in the NTHS as part of its fiscal stimulus plan, whereas China joined the WTO at the end of 2001. To allow for the possibility that the effect of highway connection was not instantaneous, I use firm TFP data in 1998, 2003 and 2007 to allow for a possible lagged effect of highway connection.

OLS Regressions

Table 1.2 shows that firms in a city became more productive on average as the city's market access expanded. To check the stability of coefficients, I include a slew of controls in the regression. The estimated coefficient remains stable as I include aggregate and firm-level covariates.[39] More specifically, a 1% increase in market access increases firm productivity by 0.05%. Please note that market access increased due to either GDP growth or transportation cost reduction. Combining these two sources implies that there are large variations in market access across cities and cities in general experienced large changes in market access over the years. A one-standard-deviation increase in market access would boost firm TFP by 5%. In one of the robustness checks below, I fix GDP and only look at the effect coming from reductions in transportation costs. In Column (2), I look at how firm TFP responds to changes in a city's access to international market. I find that a reduction in transportation costs from origin to the nearest port–an effective increase in access to the international market–raise average firm productivity. A 1% reduction in domestic transportation costs to ports boosted firm TFP in the city by 0.6 %. When I include both domestic market access

reasons. First, firms need to learn about the faster routes and build connections in other cities first. Second, it also takes time for logistics companies to update their routes and allocate labor and capital to the new routes.

[39]This is not because these control variables are not relevant. In fact, the R-squared increases as more covariates are added.

and international market access, I find that both promoted firm TFP growth.

I then include firm fixed effects into the regression specification and test if an increase in market access raised TFP within firms. This channel has been emphasized in a few recent papers (Lileeva and Trefler 2010; Bustos 2011a; Garcia and Voigtländer 2013). I do not find strong evidence for that channel. As we can see in Table 1.8, the coefficient on market access is insignificant once I include firm fixed effects. I do find that a reduction in transportation costs to the nearest port led to within-firm productivity growth. However, we need to construct measures of output market access and measures of input market to investigate whether the effect comes from easier access to cheaper and higher quality foreign inputs or from larger access to demand.[40]

Table 1.9 presents results on markups. I do not find strong evidence that firm markups respond to market access. Since higher productivity is another way of saying a firm has a cost advantage over other firms, the fact that markups do not respond to market access implies that the higher production efficiency is translated into lower product prices. This finding potentially has important welfare implications–both the more productive firms and consumers seem to reap the benefits from highway infrastructure. I will decompose the revenue-based TFP into markups, marginal costs and physical productivity.

2SLS Regressions

As explained in Section 5, the instrument I construct is the hypothetical city market access constructed from a least-cost path network. I present results from the first stage and followed by the second stage. In the first stage, I regress the actual market access on the hypothetical market access, along with all the control variables. Results from the first stage of the two-stage-least-squares regressions are presented in Table 1.7. The hypothetically constructed market access is highly correlated with the market access estimated with the actual highway

[40]Currently I am using China's input-output tables to construct output market access and input market access at the city- and the industry-level. This exercise will allow us to disentangle the effects of proximity to suppliers from the effects of proximity to consumers.

networks.[41] In Table 1.2, we see that the estimated effect is slightly smaller than the results from OLS regressions, which suggests that there were indeed unobserved forces that were correlated with cities' expansion of market access and affected TFP growth.

Table 1.2: City Market Access and Firm TFP

Dependent variable:			OLS					IV		
Firm TFP	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Market access	0.134***	0.065***		0.064***	0.069***	0.064***	0.042**		0.040**	0.046**
	(0.023)	(0.019)		(0.018)	(0.019)	(0.022)	(0.020)		(0.020)	(0.020)
Access to port			0.087**	0.087**	0.088**			0.087**	0.087**	0.088**
			(0.034)	(0.035)	(0.034)			(0.034)	(0.035)	(0.034)
Exporter dummy		-0.008	-0.008	-0.008	-0.008		-0.008	-0.008	-0.008	-0.008
		(0.009)	(0.009)	(0.009)	(0.009)		(0.009)	(0.009)	(0.009)	(0.009)
City population		-0.031	0.027	0.027	0.027		-0.031	0.027	0.027	0.027
		(0.080)	(0.070)	(0.073)	(0.074)		(0.080)	(0.070)	(0.071)	(0.073)
Node city × year					Yes					Yes
Railway Dist × year					Yes					Yes
Waterway Dist × year					Yes					Yes
Ownership structure		Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes
Industry-year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. of observations	612484	598381	597085	597085	594476	612484	598381	597085	597085	594476
Adjusted. R squared	0.380	0.655	0.656	0.656	0.656	-	-	-	-	-
No. of clusters	339	339	335	335	335	339	339	335	335	335

Notes: This table includes results from OLS and IV regressions. The dependent variable is firm TFP and the main explanatory variable is market access. All regressions include industry-year and city fixed effects. Standard errors are clustered at the city level. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

1.6.2 Counterfactual Impact of Removing Highways

Based on the estimated effect of market access on city productivity,[42] I evaluate the economic impact of China's national highway system. I consider a baseline counterfactual scenario

[41] The F statistic in the first stage is very large, which suggests a reassuring strong first stage. However, it also rings some alarm bells about the validity of the instrument. Spatial correlation in the error term could potentially lead to the very high F-statistic for weak instrument test. A solution to spatial correlation in panel IV regressions has been proposed in König et al. (2015). As a next step, I will follow their procedure and correct for spatial correlation in the error term.

[42] We need to assume that the effect of market access on aggregate productivity is log linear.

of removing all highways in China.[43] Other counterfactual scenarios, such as replacing the national highway network with a more extensive railway system or a different highway network, can be analyzed in a similar fashion. I first calculate the decline in market access for each city if we remove all highways in China. The median loss in market access by removing all highways is 62%. I then estimate the decline in productivity in each city under the counterfactual scenario based on the estimated effect of market access on productivity. Finally, the loss in productivity for each city is weighed by city size to estimate the total national decline in productivity in absence of the national highway system.

The counterfactual analysis suggests that eliminating all highways in China would lower aggregate TFP by 3.2%.[44] In the baseline counterfactural, population is held fixed but it is likely that removing highways would change the distribution of population and production across cities. In fact, the effect of transport infrastructure on the distribution of population is the focus for a few studies mentioned above (Faber 2014; Baum-Snow et al. 2015).[45] I relax this assumption to allow highways to influence the distribution of population across cities. I use the observed population distribution in 1995 as a proxy for the counterfactual population distribution without highways.[46] I find that removing all highways would lower aggregate TFP by 3.8%. It is interesting that we get a larger increase in aggregate TFP by allowing population redistribution. The larger economic impact of highways suggests that workers moved to cities which experienced larger gain in productivity. Table 1.3 presents the estimated impact of removing all highways.

[43] Donaldson and Hornbeck (2016) conduct a similar counterfactural analysis to evaluate the loss of total land value in the US if all the railways were eliminated.

[44] Alternatively, I fix population to 1995 level, and GDP to 1995 and 2005 level. Results are not sensitive to different weights.

[45] With perfect labor mobility, city productivity is not determined by its market access according to the model I present in this paper. In that sense, the estimated impact from the empirical part should serve as a lower bound of the actual effect of highways on aggregate TFP.

[46] It is possible that the counterfactual population distribution differs from the population distribution in 1995 in important ways. The model I present does not predict migration. Donaldson and Hornbeck (2016) and Baum-Snow et al. (2015) do model population change due to transport infrastructure projects.

Table 1.3: Counterfactual Impacts on Aggregate TFP

	Estimated Decrease in Aggregate TFP without Highways
Baseline counterfactural scenario in absence of highways (holding city population constant from 2005)	3.16%
1. Holding the population distribution from 2000	3.12%
2. Holding the population distribution from 1995	3.09%
Allowing for changes in the distribution of population (holding total population constant)	3.81%

Notes: This table shows the estimated impact of removing all highways. Row 1-3 reports the counterfactual impact on aggregate TFP from eliminating all highways in China. In the baseline scenario, population distribution is held fixed from 2005. In Row 2, population distribution is assumed to be from 2000. In Row 3, population distribution is held fixed from 1995. In Row 4, I allow for changes in population distribution over time. The population distribution from 1995 is used as a proxy for the counterfactual population distribution without highways. All results assume that total population is held unchanged.

1.6.3 Decomposition of Productivity Gains from Highway Access

From the firm-level regressions, I have established that highways increased Chinese cities' productivity. The firm-level regressions only indicate that highway connections increased the production efficiency of a city. However, the channels through which highway connections enhanced aggregate production efficiency are unclear. In this section, I explore the mechanisms underlying the aggregate TFP gains resulting from highway connections. There are potentially four channels through which market access could affect a city's production efficiency: a within-firm productivity enhancement among continuing firms, the reallocation of market shares across continuing firms, the entry of productive new firms[47] and the exit of inefficient firms. All channels except for the first are present in the model and the central argument in the recent international trade literature (Pavcnik 2002; Bernard et al. 2003; Melitz 2003), but the first is not. However, many more recent papers (Lileeva and Trefler 2010; Bustos 2011a; Garcia and Voigtländer 2013) document sizable within-plant or within-firm productivity growth after trade liberalization. To quantify the effects of each channel, I follow Haltiwanger (1997) to decompose changes in city productivity at the industry level into four terms. First, a productivity index for industry s is defined as follows:

$$lnTFP^s_{k,t} = \sum_i \theta^s_{i,k,t} lnTFP^s_{i,k,t}, \tag{1.22}$$

where $\theta^s_{i,t}$ is the share of output for firm i in industry s in city k at time t.

[47] For entry, we can further decompose it into the extensive margin (number of new firms) and the intensive margin (productivity of new firms).

$$
\begin{aligned}
\Delta lnTFP^s_{k,t} = &\sum_{continuers} \theta^s_{i,k,t-1}\Delta lnTFP^s_{i,k,t}+\\
&\sum_{continuers} \Delta\theta^s_{i,k,t}(lnTFP^s_{i,k,t-1} - lnTFP^s_{k,t-1})+\\
&\sum_{enteringfirms} \theta^s_{i,k,t}(lnTFP^s_{i,k,t} - lnTFP^s_{i,k,t-1})-\\
&\sum_{exitingfirms} \theta^s_{i,k,t-1}(lnTFP^s_{i,k,t-1} - lnTFP^s_{k,t-1})
\end{aligned}
$$

I regress each of the four components on changes in city market access, transportation cost to the nearest port, and other city-level characteristics and industry fixed effects. Table 1.4 reports the results from the decomposition exercise. Overall, the effect is very similar to that observed in the firm-level regressions – a 1% increase in market access leads to a 0.04% increase in the productivity index. I find that the entry of productive firms contributed the most to the aggregate gains in TFP. Reallocation among large incumbent firms and exit of inefficient firms also contributed to the productivity gain. These findings are consistent with evidence from other strands of the literature.[48] Consistent with the firm-level regressions, I do not find strong evidence of within-firm productivity improvements from an increase in domestic market access. Although the OLS result is significant and large, the IV result is much smaller and insignificant.

The decomposition exercise reveals the sources of TFP gains from highway infrastructure investments in China: transportation infrastructure affects city productivity mainly though the entry and expansion of relatively new and small firms and the contraction and exit of inefficient firms. Table 1.4 also presents the sources of TFP gains from easier access to the international market. Overall, a 1 % reduction in transportation costs to the nearest port increases aggregate TFP by 0.14%. Similar to domestic market access, the entry and expansion of productive young firms contributed to the TFP gains. In contrast to domestic

[48] However, the exit effect is small. This may be because large firms, especially state-owned enterprises, were given many preferential policies to keep them alive, to the extent that the shut-down decision itself was subject to local politics.

market access, I find that within-firm productivity growth was an important source of aggregate TFP growth. I find little evidence on reallocation across large incumbents. This is consistent with findings from Lileeva and Trefler (2010) on the liberalization of trade between the US and Canada.

Table 1.4: Decomposition of TFP gains

(a) OLS	Total effect (1)	Within (2)	Between (3)	Entry (4)	Exit (5)
Change in domestic market access	0.061**	0.039***	0.014***	0.045**	0.009***
	(0.031)	(0.012)	(0.005)	(0.020)	(0.002)
Change in access to ports	0.139***	0.065***	0.006	0.097***	-0.000
	(0.027)	(0.018)	(0.008)	(0.020)	(0.004)
No. of Observations	6413	6413	6413	6413	6413

(b) IV	Total effect (1)	Within (2)	Between (3)	Entry (4)	Exit (5)
Change in domestic market access	0.080**	0.007	0.021***	0.055**	0.011***
	(0.032)	(0.012)	(0.006)	(0.023)	(0.003)
Change in access to ports	0.134***	0.062***	0.006	0.095***	-0.001
	(0.027)	(0.017)	(0.008)	(0.020)	(0.004)
No. of Observations	6413	6413	6413	6413	6413

Notes: This table includes results from OLS and IV regressions. I decompose aggregate TFP into four components (the within-firm productivity growth, the reallocation between firms, the entry of new firms and the exit of incumbents). I then regress the change in each component on market access. Control variables include city population, provincial capital dummy, average years of schooling, access to railways, access to waterways and industry fixed effects. All regressions include industry fixed effects. Robust standard errors in parentheses. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

1.6.4 Robustness Checks

Revenue Productivity and Prices

The distinction between revenue productivity and physical productivity has become increasingly important in the literature. Up to now, we have measured revenue productivity, not physical productivity (the connection and difference between the two will become clear below). As a result, we can not rule out a somewhat counter-intuitive case: the average firm's physical productivity decreases as a city's market access expands, which implies an increase in marginal cost. Since both terms enter revenue productivity and offset each other, it is still possible that revenue productivity goes up. Next, I investigate how firms' marginal costs and prices respond to changes in market access.[49]

Results on marginal cost and price are presented in Table 1.10. Similar to earlier results, I do not find a strong effect of market access on markups.[50] However, we find that a 1% increase in market access leads to a 0.12% reduction in firms' marginal costs, which suggests that the effect of market access on firm productivity is more than twice as large as suggested by the regressions on TFPR if we consider the price effect. An interesting and important take-away from these regressions is that firms pass most of their cost advantages to consumers, implying economic integration through transportation costs reductions could potentially sizable welfare gains for consumers.

Isolating transportation costs reduction from GDP growth

In my market access measures, two factors affect a city's market access: 1) transportation costs 2) the size of its neighbors. One might be concerned that changes in a city's market access due to changes in its neighbors' size might be correlated with some unobserved factors that also affect firm productivity in the city. Moreover, since I am mainly interested in the effects of transportation costs in this paper, I would like to isolate the changes in market access due to a reduction in transportation costs from changes in sizes of neighbors. In Table 1.11, I show results when I use population of all cities in 1995 and the only time variation in a city's market access comes from changes in bilateral transportation costs. It is clear that shutting down the size channel does not substantially alter my results. If anything, the effects from the transportation costs channel are slightly larger than the corresponding baseline results.

[49] Please refer to Appendix A.3 for an illustration of the idea in Garcia and Voigtländer (2013).

[50] This result at first seems to be in contrast with one of the main predictions in Melitz and Ottaviano (2008). However, De Loecker et al. (2016) also find markups did not decrease after trade liberalization India. In fact, they find markups went up because firms' marginal costs decreased and firms could cut prices and raise markup simultaneously.

Excluding the "Node" Cities

Another threat to my identification strategy is the potential correlation between highway placement and some unobserved placed-based policies. In the baseline regressions, I remove city GDP to alleviate the this concern, but one may still be concerned that those "node" cities targeted by the government may have other characteristics that affected both highway placement and TFP growth. I re-run the same regressions as in the baseline but exclude the 52 cities that were targeted as "nodes" by the government when it planned the national highway network. The results are presented in Table 1.12. Note that the exclusion of node cities does not change the results. Since my results are not driven by those target cities, I can be confident that unobserved city characteristics do not threaten the validity of my results.

An Alternative Way for Estimating Market Access

Donaldson and Hornbeck (2016) show that a location's market access can be expressed as the sum over the transportation costs with each other county, that other location's population, and that other location's access to other markets. The derivation is relegated to the the Appendix. They numerically solve for market access for all US counties. Donaldson and Hornbeck (2016) also verify that the the results they get from using their market access measure is consistent with their baseline results from using the "market potential" term by Harris (1954). I follow their strategy of estimating city market access by solving a system of nonlinear equations. I then run regressions with the numerically solved market access measure, and confirm that the two measures generate very similar results.[51]

1.7 Spatial Reallocation of Industries

Now I turn to the effects of highways on the spatial reallocation of industries. As explained in the introduction, classic trade theories predict industry specialization patterns after trade

[51]The estimation of market access is implemented in MATLAB. MATLAB code is available upon request.

liberalization based on either technological differences, endowment differences or returns to scale. Following the literature, I explore three important dimensions of industry characteristics that will lead to heterogeneous responses to highway connection. I use weight to value ratio to measure the iceberg transportation cost of an industry. I use the median firm's capital to labor ratio to measure industry's capital intensity. I also take the trade elasticities estimated in the trade literature as a measure of the degree of product differentiation of an industry.[52] I use the following baseline specification[53] for estimating the effects of market access on employment growth and firm growth:

$$\Delta logY_{j,k} = \alpha + \delta \Delta logMA_k \times Ind_j + \beta X_{k,j} + \mu_k + \nu_j + \epsilon_{j,k}$$

where $logMA_k$ is city k's change in market access, Ind_j is a ranking of industries depending on what aspects we are interested in. I am most interested in three sources of industry heterogeneity: industry's degree of product differentiation, as measured by the elasticity of substitution; industry's capital intensity, as measured by the median capital-labor ratio of firms; industry's transportation cost, as measured by the median weight-to-value ratio. I include initial share of the industry in total output as a control variable to account for any convergence effect.

In Table 12-14, I present the effects of highway network on employment growth across industries and across cities. In Table 1.13, I find that the coefficient on the interaction term between market access and capital to labor ratio is positive, meaning that industries with higher capital intensity grew disproportionately faster in cities with large market access. Results do not change very much if I include the interaction between market access and city population, as well as the inclusion of interaction between market access and driving time to nearest port. The instrument variable approach generates similar results to OLS regressions.

[52]Please note that here the market access term is very similar to Donaldson and Hornbeck (2016) The model in Section 3 does not generate precise predictions for industry reallocation. So, I follow the literature and use the market access term used in recent papers.

[53]This type of specification has been use extensively in the literature to examine the effects of various frictions on growth and trade, such as financial constraints (Rajan and L. 1998; Manova 2013), labor market protection (Tang 2012).

One way to interpret the coefficients of the interaction term is the following: a one-standard deviation of a city's market access would increase annual employment growth in the sector at the 75th percentile of the distribution by capital intensity by 1.6 percentage points more than annual employment growth in the sector at the 25th percentile for the period from 2001 to 2008. Cities with large market access will find themselves have a comparative advantage in capital-intensive industries.[54] In Table 1.14, the coefficient on the interaction term between market access and weight to value ratio is again positive, suggesting that industries with weight-to-value ratio (high transportation costs) grew disproportionately faster in cities with large market access. Similarly, Table 2.1 presents results on product differentiation. The results suggest that industries with a low degree of product differentiation grew disproportionately faster in cities with large market access. This result is in contrast to Hanson and X. (2004), in which they find more industries with differentiated products tend to locate in large *countries*.[55] Overall, my results show that transport infrastructure redistributes industries spatially, and the findings are in general consistent with theories of trade and economic geography.

1.8 Discussion and Conclusion

Transport infrastructure is one of the most expensive public goods in the world. Governments across the world spend billions of dollars to build highways and railways to facilitate the movement of goods and people. Existing empirical studies focus on the effects of railways or highways on GDP or population growth and have found mixed results. In this paper, I examine the channels through which transport infrastructure affects economic outcomes. I find that the national highway network in China promoted production efficiency, delivered siz-

[54] More analysis will need to be done to identify the underlying mechanisms. Gaubert (2014) argues that larger cities tend to specialize in capital intensive industries because wage is higher in big cities. Hanson (2005) also finds that wages tend to be higher in locations with larger market access.

[55] As mentioned in Hanson and X. (2004), the robustness of the home-market effect in Krugman (1980) is still an open question. The prediction hinges critically on model assumptions.

able welfare gains to consumers, and led to a sectoral reallocation between cities. Aggregate TFP growth resulting from reduced transportation costs is attributable primarily to firm entry and resource reallocation. I also find that cities with large market access specialized in industries that have low unit cost, higher capital intensity, and low product differentiation.

The findings presented in this paper have important policy implications. Facing the threat of secular stagnation, policymakers are searching for tools to boost aggregate demand in the short run and promote economic growth in the long run. After the global recession, there has been a growing interest among policymakers worldwide in using infrastructure investments both as a short-term fiscal policy instrument and as a long-term growth generator.[56] For example, the World Bank has consistently dedicated itself to investing in infrastructure in low-income countries to fight poverty. The International Monetary Fund is also actively advocating for more infrastructure investments in Latin America and Africa to meet the infrastructure needs and boost economic growth in these regions. The two recently-founded development banks–the Asian Infrastructure Investment Bank and the New Development Bank, were established under the leadership of China to address the increasing infrastructure needs in Asia. The US president-elect Donald Trump envisions a trillion-dollar infrastructure plan. The increasing use of infrastructure projects by policymakers begs the question of whether the huge amount of tax dollars spent on infrastructure is well justified by their potential benefits.

This paper provides novel evidence that transport infrastructure promotes economic growth and sheds light on the mechanisms underlying the gains from infrastructure investment. I examine the channels of productivity gains resulting from infrastructure investments and quantify the relative importance of each channel. I highlight the role of highway infrastructure in raising aggregate TFP by facilitating resource reallocation between heterogeneous firms. This paper also evaluates the aggregate economic impact of China's national highway system. Findings in this paper suggest a sizable economic impact of infrastructure investments and market integration when domestic transportation costs are large and misal-

[56]During the crisis, China, India and Korea led the way to spend on infrastructure as part of their stimulus packages.

location is pervasive.[57] To make investment decisions, however, we will also need to evaluate potential dynamic gains from such as a large-scale transport infrastructure project and its economic costs, which is beyond the scope of this paper.

There are a few avenues for future research. As mentioned in the introduction, a natural byproduct of this paper is a quantitative exercise that compares aggregate productivity growth under the actual highway construction with counterfactual scenarios. Given the enormous cost of infrastructure projects, it is important to compare the potential economic growth resulting from infrastructure to both its direct and indirect cost to aid policy recommendations. Another interesting area for future research is inter-sectoral linkages. So far, I have ignored the inter-connectedness of sectors when estimating market access. One interesting extension would be to incorporate inter-sectoral linkages and construct output market access and input market access measures using input-output tables. Such output and input market access measures would not only allow for the use of cross-industry and within-city variation to identify the impact of transportation infrastructure but also make it possible to differentiate the impact of output access from input access. Finally, it would also be interesting to conduct an in-depth focus study on some of the most remote and poorest areas. To do so, one would need to examine the county or even village level data. Such studies at the more micro-level would help to shed light on the role of infrastructure in reducing poverty and affecting income inequality.

[57]This paper focuses on productivity gains from highways in the manufacturing sector but neglects potential gains in other sectors.

Figures

Figure 1.3: The National Trunk Highway System

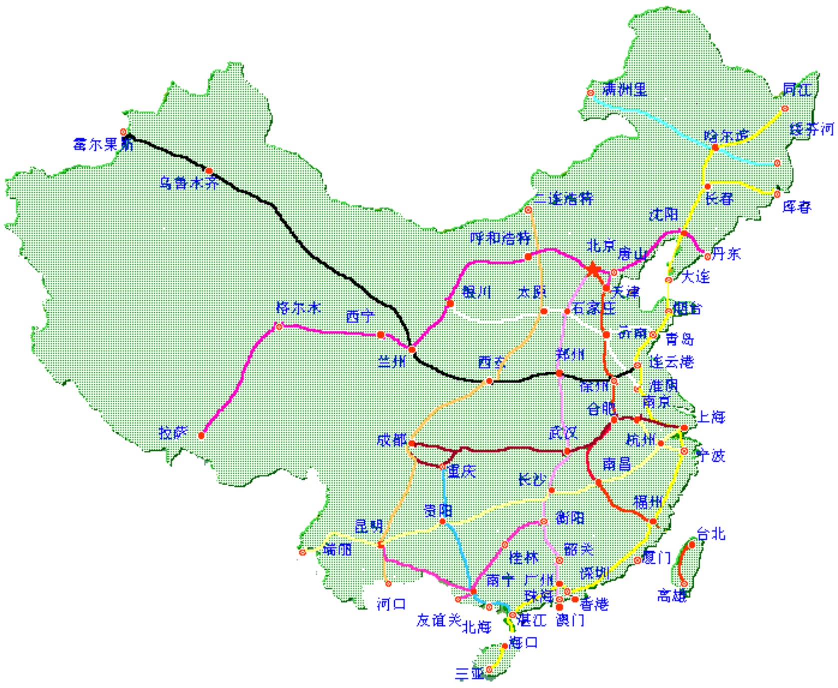


Figure 1.4: Inter-Provincial Trade

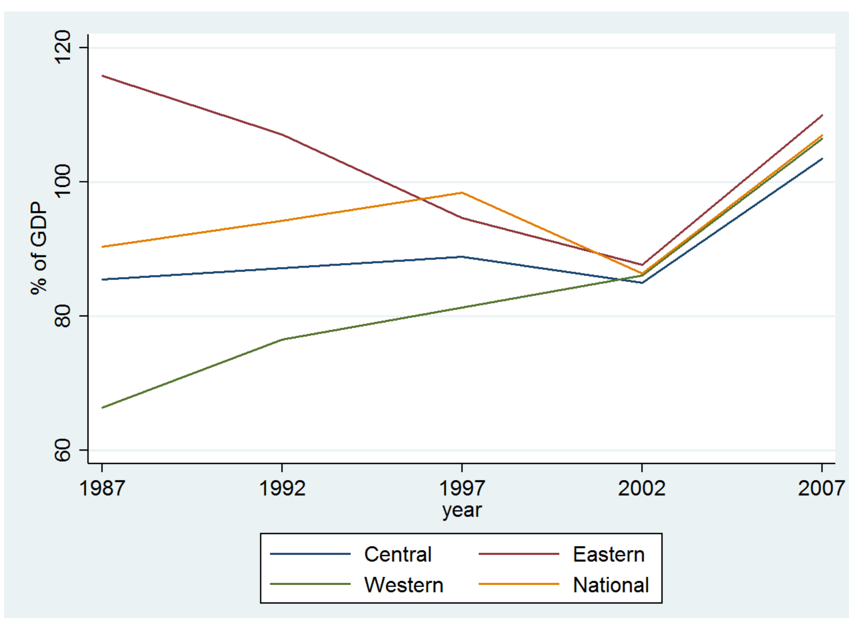
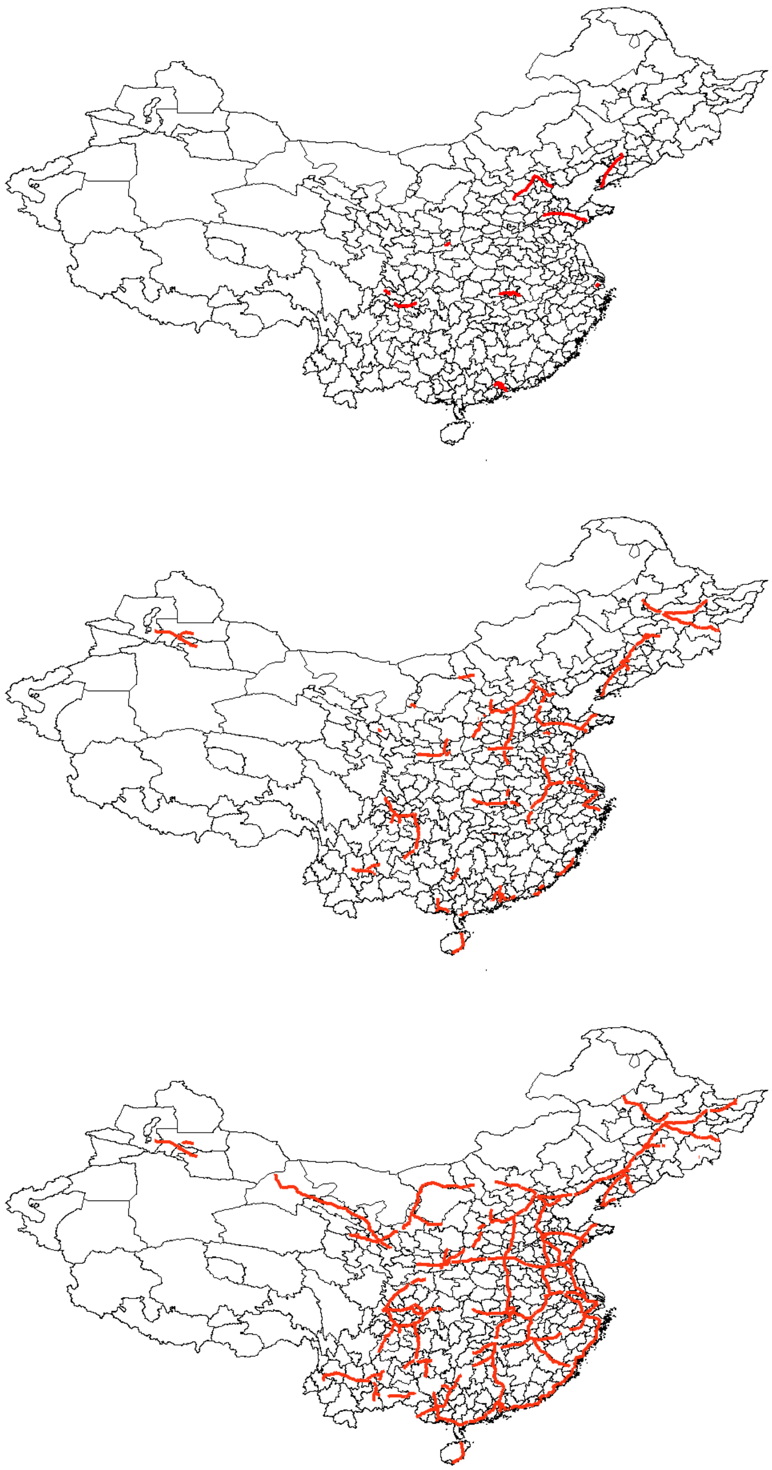
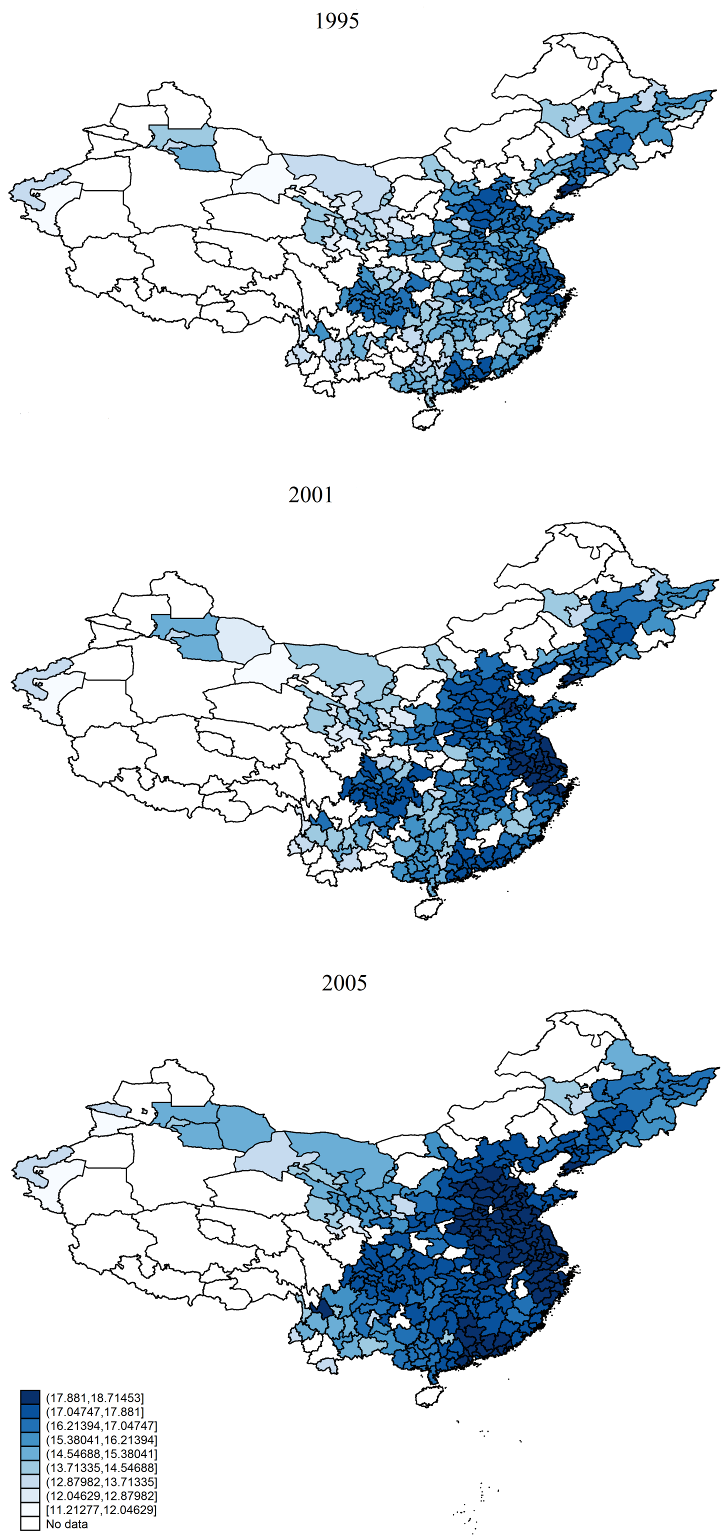


Figure 1.5: Highway network in 1995, 2001, 2005



Notes: this figure shows how China's National Trunk Highway System evolved over time. Please note that there were many local construction that were not part of the NTHS. These highways were built for intra-city connections or connecions to nearby cities.

Figure 1.6: Estimated changes in market access 1995-2005



Notes: this figure shows how market access evolved over time for 339 Chinese cities. Darker color means larger market access.

Tables

Table 1.5: Estimation of Transportation Costs

Dependent Variable: Iceberg Transport cost	(1)	(2)	(3)	(4)
Length of highway	0.454***	0.539***	1.331***	1.105***
	(0.001)	(0.001)	(0.007)	(0.008)
Length of local road	1.092***	0.712***	1.215***	1.199***
	(0.005)	(0.026)	(0.045)	(0.045)
Length of highway (squared)			-0.381***	-0.315***
			(0.004)	(0.004)
Length of local road (squared)			-0.938***	-0.780***
			(0.056)	(0.055)
Length of highway (3 squared)			0.048***	0.041***
			(0.001)	(0.001)
Length of local road (3 squared)			0.327***	0.258***
			(0.017)	(0.016)
Travel time				0.264***
				(0.005)
Travel time (squared)				-0.022***
				(0.001)
Travel time (3 squared)				0.001***
				(0.000)
Origin fixed effects	No	Yes	Yes	Yes
Destination fixed effects	No	Yes	Yes	Yes
No. of observations	97253	97253	97253	97253
Adjusted. R squared	0.689	0.850	0.869	0.875

Notes: This table shows estimation results from a simple linear regression model designed to estimate the relationship between transport costs and the road network structure. All regressions include origin and destination fixed effects. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

Table 1.6: Productivity of Chinese Manufacturing Firms

	Industry Aggregate TFP (log)									
Chinese Industrial Classfication (2-digit)	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Processing of Foods (13)	5.05	4.96	5.09	5.18	5.35	5.46	5.57	5.62	5.67	5.80
Manufacturing of Foods (14)	5.49	5.16	5.53	5.54	5.62	5.79	5.89	6.00	6.09	6.14
Manufacture of Beverages (15)	3.88	3.77	3.89	3.99	4.05	4.18	4.30	4.55	4.56	4.60
Manufacture of Tobacco (16)	2.49	2.47	2.49	2.64	3.05	3.21	3.13	3.15	3.34	3.42
Manufacture of Textile (17)	5.97	6.01	6.07	6.21	6.33	6.45	6.46	6.70	6.78	6.96
Manufacture of Apparel, Footwear & Caps (18)	5.46	5.16	5.18	5.33	5.32	5.45	5.54	5.74	5.95	5.89
Manufacture of Leather, Fur, & Feather (19)	5.04	5.04	5.00	5.08	5.13	5.18	5.29	5.28	5.44	5.54
Processing of Timber, Manufacture of Wood and Bamboo Products (20)	4.68	3.85	3.85	3.93	4.07	4.09	4.33	4.27	4.35	4.57
Manufacture of Furniture (21)	5.05	4.66	4.77	4.77	4.96	4.92	5.14	5.07	5.26	5.41
Manufacture of Paper & Paper Products (22)	5.08	5.08	5.18	5.31	5.48	5.64	6.00	5.88	6.00	6.17
Printing, Reproduction of Recording Media (23)	5.52	5.45	5.57	5.80	5.87	5.97	5.99	6.25	6.32	6.43
Manufacture of Articles For Culture, Education & Sport Activities (24)	6.27	6.26	6.31	6.31	6.41	6.56	6.54	6.69	6.85	6.97
Processing of Petroleum, Coking, &Fuel (25)	4.95	4.86	4.93	5.05	5.15	5.29	5.36	5.39	5.46	5.48
Manufacture of Raw Chemical Materials (26)	5.64	5.47	5.69	5.81	5.94	6.31	6.69	6.52	6.56	6.69
Manufacture of Medicines (27)	6.16	6.18	6.23	6.30	6.50	6.55	6.58	6.66	6.77	6.83
Manufacture of Chemical Fibers (28)	4.35	4.51	4.79	4.68	4.80	5.12	5.21	5.23	5.38	5.48
Manufacture of Rubber (29)	5.77	5.73	5.82	5.98	6.24	6.42	6.47	6.75	6.70	6.75
Manufacture of Plastics (30)	4.72	4.58	4.65	4.74	4.87	4.96	4.91	4.96	5.14	5.26
Manufacture of Non-metallic Mineral goods (31)	4.81	4.73	4.82	4.97	5.08	5.24	5.49	5.80	5.68	5.87
Smelting & Pressing of Ferrous Metals (32)	4.49	4.32	4.43	4.62	4.83	5.22	5.52	5.52	5.51	5.50
Smelting & Pressing of Non-ferrous Metals (33)	6.30	6.49	6.60	6.69	7.02	7.01	7.24	7.47	7.87	7.96
Manufacture of Metal Products (34)	5.58	5.49	5.57	5.66	5.92	6.00	5.94	5.96	6.07	6.32
Manufacture of General Purpose Machinery (35)	5.34	5.39	5.44	5.65	5.86	6.10	6.40	6.55	6.74	6.90
Manufacture of Special Purpose Machinery (36)	5.01	5.13	5.30	5.39	5.75	6.01	6.06	6.05	6.27	6.42
Manufacture of Transport Equipment (37)	4.85	4.85	5.00	5.33	5.60	6.01	6.06	6.11	6.14	6.22
Electrical Machinery & Equipment (39)	5.41	5.62	5.69	5.74	5.81	5.97	6.05	6.11	6.28	6.34
Computers & Other Electronic Equipment (40)	7.78	7.97	8.00	8.20	8.52	8.40	8.77	8.70	8.61	8.35
Measuring Instruments & Machinery for Cultural Activity & Office Work (41)	5.85	5.41	5.38	5.69	5.91	6.18	6.01	6.23	6.34	6.42
Manufacture of Artwork (42)	5.52	5.32	5.44	5.60	5.68	5.81	5.68	6.08	5.93	6.46

Notes: This table shows the industry-level productivity index in China from 1998 to 2007. The construction of the productivity index follows (22) in 6.2. The index is expressed in log terms. The 2-digit industry classification changed in 2003, and the years after 2003 are converted to the classification before the change to ensure consistency over time.

Table 1.7: First Stage Regressions

	log MA	log MA (excluding nodes)	log MA (constant GDP)
log MA (IV)	0.967***	0.922***	0.960***
	(0.035)	(0.041)	(0.036)
Access to Port	0.022	0.060	0.028
	(0.033)	(0.041)	(0.034)
Exporter dummy	0.002**	0.004***	0.002**
	(0.001)	(0.001)	(0.001)
City population	-0.001	-0.001	-0.000
	(0.001)	(0.001)	(0.001)
Railway Dist × year	Yes	Yes	Yes
Waterway Dist × year	Yes	Yes	Yes
Ownership structure	Yes	Yes	Yes
Industry-year FE	Yes	Yes	Yes
City FE	Yes	Yes	Yes
No. of observations	608208	370615	608208
Adjusted. R squared	0.991	0.992	0.982
First-stage F statistic	723.433	489.353	656.805

Notes: This table presents results from the first-stage regressions. The dependent variable is the market access estimated with actual highway network data. The independent variable is the hypothetical market access constructed from the least-cost path spanning tree network. Standard errors are clustered at the city level. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

Table 1.8: Market Access and Within-Firm TFP Growth

Dependent variable:			OLS					IV		
Firm TFP	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Market access	0.038	0.019		0.016	0.022	0.023	0.017		0.013	0.021
	(0.024)	(0.014)		(0.014)	(0.015)	(0.026)	(0.015)		(0.016)	(0.016)
Access to port			0.081**	0.080**	0.072**			0.081**	0.080**	0.072**
			(0.033)	(0.033)	(0.033)			(0.033)	(0.033)	(0.034)
Exporter dummy		-0.006	-0.006	-0.006	-0.002		-0.006	-0.006	-0.006	-0.002
		(0.009)	(0.009)	(0.009)	(0.009)		(0.009)	(0.009)	(0.009)	(0.009)
City population					Yes					Yes
					Yes					Yes
Node city × year					Yes					Yes
Railway Dist × year					Yes					Yes
Waterway Dist × year					Yes					Yes
Ownership structure		Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes
Industry-year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. of observations	322546	313209	312381	312381	312381	322546	313209	312381	312381	312381
Adjusted. R squared	0.691	0.822	0.823	0.823	0.823	-	-	-	-	-
No. of clusters	339	339	335	335	335	339	339	335	335	335

Notes: This table includes results from OLS and IV regressions. The dependent variable is firm TFP and the main explanatory variable is market access. All regressions include industry-year and city fixed effects. Standard errors are clustered at the city level. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

Table 1.9: Market Access and Markups

Dependent variable:		OLS			IV	
Firm Markup	(1)	(2)	(3)	(4)	(5)	(6)
Market Access	-0.007		-0.002	-0.003		-0.001
	(0.020)		(0.018)	(0.022)		(0.019)
Access to port		0.072	0.072		0.073	0.075
		(0.046)	(0.046)		(0.045)	(0.045)
Exporter dummy	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
City population	-0.173*	-0.140	-0.140	-0.171	-0.140	-0.140
	(0.097)	(0.086)	(0.086)	(0.098)	(0.086)	(0.086)
Node city × year	Yes	Yes	Yes	Yes	Yes	Yes
Railway Dist × year	Yes	Yes	Yes	Yes	Yes	Yes
Waterway Dist × year	Yes	Yes	Yes	Yes	Yes	Yes
Ownership structure	Yes	Yes	Yes	Yes	Yes	Yes
Industry-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
No. of observations	595635	594476	594476	595635	594476	594476
Adjusted. R squared	0.84	0.841	0.841	-	-	-
No. of clusters	333	331	331	333	331	331

Notes: This table includes results from OLS and IV regressions. The dependent variable is firm markups and the main explanatory variable is market access. All regressions include industry-year and city fixed effects. Standard errors are clustered at the city level. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

Table 1.10: Market Access, Marginal Cost and Price

Dependent variable:	Marginal Cost			Price		
	(1)	(2)	(3)	(4)	(5)	(6)
Market Access	-0.155***	-0.170***	-0.120**	-0.084*	-0.086*	-0.095**
	(0.048)	(0.049)	(0.046)	(0.046)	(0.047)	(0.047)
Access to port		-0.036	-0.022		-0.036	-0.022
		(0.100)	(0.094)		(0.100)	(0.094)
Exporter dummy			0.323***			0.323***
			(0.046)			(0.043)
City population		0.366*	0.210		0.363*	0.242
		(0.209)	(0.190)		(0.208)	(0.192)
Node city × year		Yes	Yes		Yes	Yes
Railway Dist × year		Yes	Yes		Yes	Yes
Waterway Dist × year		Yes	Yes		Yes	Yes
Ownership structure	Yes	Yes	Yes	Yes	Yes	Yes
Industry-year FE	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes
No. of observations	197656	196145	196143	197656	196145	196143
Adjusted. R squared	0.389	0.389	0.406	-	-	-
No. of clusters	338	331	331	338	331	331

Notes: The table presents results from a regression of marginal cost and price on market access. All regressions include industry-year and city fixed effects. All standard errors are clustered at the city level. Standard errors in parentheses. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

Table 1.11: The NTHS and Firm TFP: Fixing City GDP

Dependent variable:		OLS			IV	
Firm TFP	(1)	(2)	(3)	(1)	(2)	(3)
Market Access	0.052***		0.053***	0.051***		0.053***
	(0.015)		(0.014)	(0.015)		(0.014)
Access to port		0.075**	0.076**		0.075**	0.076**
		(0.031)	(0.031)		(0.031)	(0.033)
City population	-0.051	-0.004	-0.007	-0.051	-0.004	-0.004
	(0.070)	(0.066)	(0.066)	(0.070)	(0.066)	(0.066)
Exporter dummy	-0.007	-0.008	-0.008	-0.007	-0.008	-0.002
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Ownership structure	Yes	Yes	Yes	Yes	Yes	Yes
Node city × year	Yes	Yes	Yes	Yes	Yes	Yes
Railway Dist × year	Yes	Yes	Yes	Yes	Yes	Yes
Waterway Dist × year	Yes	Yes	Yes	Yes	Yes	Yes
Industry-year FE	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes
No. of observations	595635	594476	594476	595635	594476	594476
Adjusted. R squared	0.654	0.654	0.654	-	-	-
No. of clusters	333	331	331	333	331	331

Notes: This table includes results from OLS and IV regressions. The dependent variable is firm TFP and the main explanatory variable is market access. All regressions include industry-year and city fixed effects. Standard errors are clustered at the city level. Standard errors in parentheses. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

Table 1.12: The NTHS and Firm TFP: Excluding the "Node" Cities

Dependent variable:			OLS					IV		
Firm TFP	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Market Access	0.152***	0.053***		0.050***	0.057***	0.096***	0.048***		0.047***	0.053***
	(0.023)	(0.014)		(0.014)	(0.015)	(0.020)	(0.016)		(0.016)	(0.016)
Access to port			0.066**	0.066**	0.072**			0.068**	0.072**	0.072**
			(0.029)	(0.029)	(0.029)			(0.029)	(0.030)	(0.031)
Exporter dummy		-0.008	-0.008	-0.008	-0.008		-0.008	-0.008	-0.008	-0.008
		(0.009)	(0.009)	(0.009)	(0.009)		(0.009)	(0.009)	(0.009)	(0.009)
City population					0.043					0.0064
					(0.069)					(0.069)
Railway Dist × year					Yes					Yes
Waterway Dist × year					Yes					Yes
Ownership structure		Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes
Industry-year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. of observations	428739	418867	417960	417960	416133	428739	418867	417960	417960	416133
Adjusted. R squared	0.421	0.682	0.693	0.694	0.694	-	-	-	-	-
No. of clusters	285	285	280	280	280	285	285	280	280	280

Notes: This table includes results from OLS and IV regressions. The dependent variable is firm TFP and the main explanatory variable is market access. All regressions include industry-year and city fixed effects. Standard errors are clustered at the city level. Standard errors in parentheses. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

Table 1.13: Employment and Firm Growth: Capital Intensity

Dependent Variable:	Growth of Employment		Growth of No. of firms	
	OLS	IV	OLS	IV
K/L*Δ Market Access	0.642***	0.644***	0.462***	0.463***
	(0.136)	(0.136)	(0.076)	(0.078)
K/L*Δ Population	0.006	0.004	-0.005	-0.006
	(0.010)	(0.009)	(0.006)	(0.006)
K/L*Δ Access to port	0.022*	0.026*	0.012	0.013
	(0.013)	(0.014)	(0.009)	(0.010)
K/L*Rail Access	0.010**	0.010**	0.005**	0.005**
	(0.000)	(0.000)	(0.002)	(0.002)
Initial industry Share	-0.001***	-0.001***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Industry FE	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes
No. of observations	8250	8250	8251	8251
No. of clusters	319	319	319	319

Note: the dependent variable in the two columns on the left is the change in employment. The dependent variable on the right is the change in number of firms. The coefficient of K/L*Δ Market Access is positive and significant, which implies that industries with high capital intensity grew disproportionately faster in cities that gained large market access. All regressions include sector and city fixed effects. Standard errors are clustered at the city level and province-industry level. Standard errors in parentheses. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

Table 1.14: Employment and Firm Growth: Weight-to-Value Ratio

Dependent Variable:	Growth of Employment		Growth of No. of firms	
	OLS	IV	OLS	IV
Weight/Value*Δ Market Access	0.004***	0.004***	0.002***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
Weight/Value*Δ Population	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Weight/Value*Δ Access to port	0.000	0.000*	0.001**	0.001**
	(0.000)	(0.000)	(0.000)	(0.000)
Weight/Value*Rail Access	0.001	0.000	0.002	0.001
	(0.000)	(0.000)	(0.003)	(0.000)
Initial industry Share	-0.001***	-0.001***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Industry FE	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes
No. of observations	8250	8250	8251	8251
No. of clusters	319	319	319	319

Note: the dependent variable in the two columns on the left is the change in employment. The dependent variable on the right is the change in number of firms. The coefficient of Weight/Value*Δ Market Access is positive and significant, which implies that industries with large weight-to-value ratio grew disproportionately faster in cities that gained large market access. All regressions include sector and city fixed effects. Standard errors are clustered at the city level and province-industry level. Standard errors in parentheses. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

Table 1.15: Employment and Firm Growth: Product Differentiation

Dependent Variable:	Growth of Employment		Growth of No. of firms	
	OLS	IV	OLS	IV
ProdDiff*Δ Market Access	-0.008**	-0.009**	-0.007***	-0.009***
	(0.004)	(0.004)	(0.002)	(0.002)
ProdDiff*Δ Population	0.000	0.002	0.000	0.001
	(0.001)	(0.001)	(0.000)	(0.001)
ProdDiff*Δ Access to port	-0.000	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
ProdDiff*Rail Access	-0.000	-0.000	-0.000*	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)
Initial industry Share	-0.001***	-0.002***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Industry FE	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes
No. of observations	8250	8250	8251	8251
No. of clusters	319	319	319	319

Note: the dependent variable in the two columns on the left is the change in employment. The dependent variable on the right is the change in number of firms. The coefficient of ProdDiff*Δ Market Access is negative and significant, which implies that industries with low product differentiation grew disproportionately faster in cities that gained large market access. All regressions include sector and city fixed effects. Standard errors are clustered at the city level and province-industry level. Standard errors in parentheses. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

Appendix

A.1. China's National Trunk Highway System

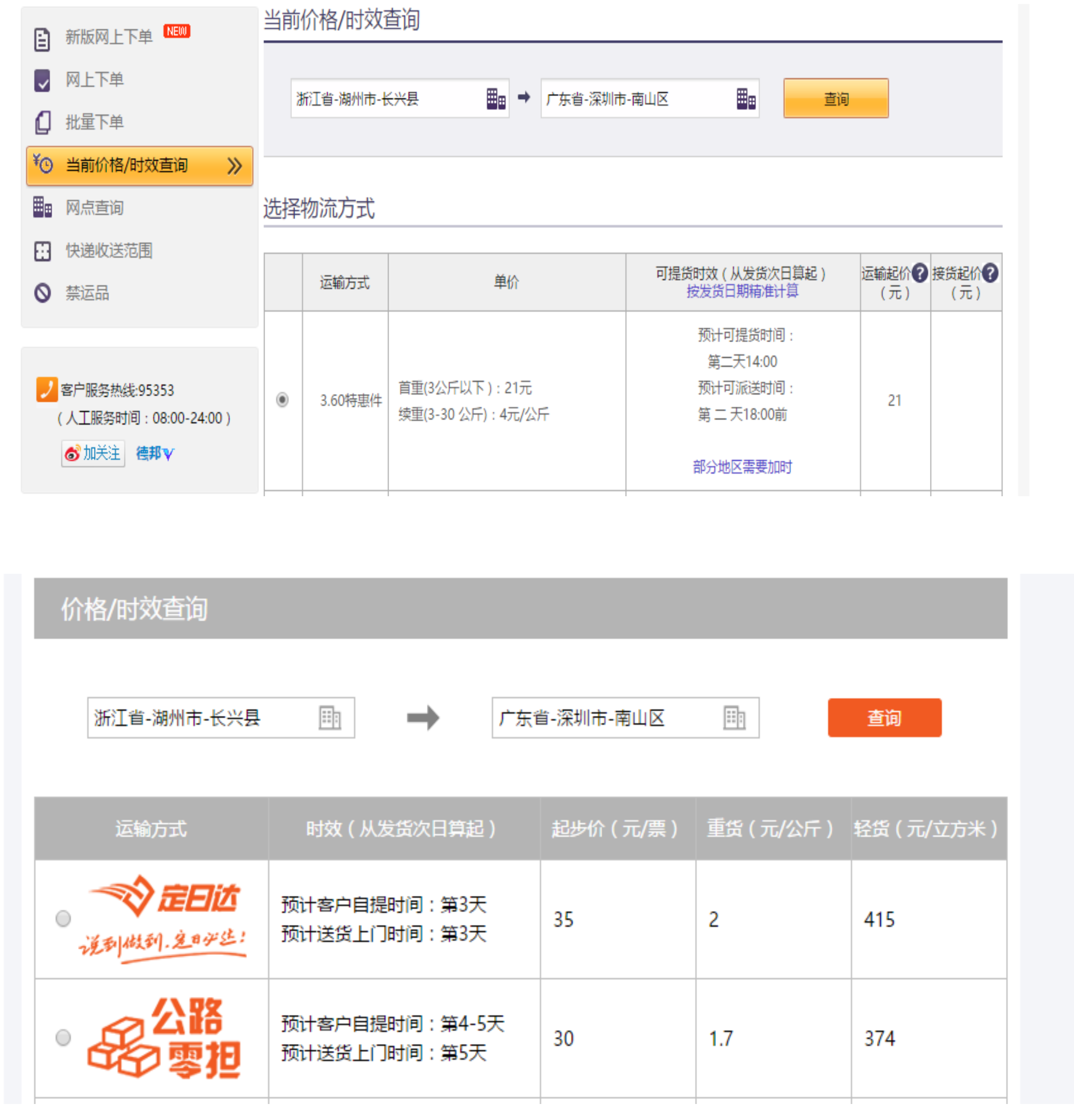
The Chinese government has long held the belief that a highway system is important for developing regional economies, boosting efficiency of the logistics system and raising living standard. The highway construction plan aimed to promote trade, increase free flow of commodities, and subsequently raise competition in the domestic market. The government also wanted to ease regional inequalities and increase employment through connecting different parts of the country through highway. The rapid economic growth in China during the early years of the economic reform has resulted in an immense demand for an effective transportation system. The Chinese government wanted to solve the problems of heavy traffic and delay due to the backward infrastructure. To fund the construction project, the government established new regulations in 1984, imposing vehicle purchase tax and raising roadway tolls. In addition, it decided to make loans to build highways and repay the loans by charging highway tools.

According to the original plan, the new transportation system is 34422 kilometers long, of which 25765 kilometers are highway, 1479 kilometers are Rate I freeway, and 7178 kilometers are Rate II freeway. The three kinds of express way takes up 74.85%, 4.3% and 20.85% of the total distance respectively. The government planned to finish the construction around 2020. The plan also stated that the new highway system would link the capital Beijing to other provincial cities, connecting 93% of the major cities with other 1 million population and big cities with over 0.5 million population. The number of cities that are linked together will exceed 200, and 0.6 billion people will be affected by the highways. The government planned five vertical expressways and seven horizontal expressways to connect the country into a single network.

The plan was approved by the National People's Congress in 1992 and established by the Department of Transportation in June, 1993. The construction of the highway system can be divided into four time periods: the mid-1980s to 1991, 1992 to 1997, 1998 to 2003, 2003 to 2007. It started in the major cities in the mid-1980s. Since the approval of the construction plan in 1992, the project entered the regular phase. Until the end of 1997, the total distance of highways in China was 4771 kilometers, of which 70% was the major national highway. In 1998, the government decided to build infrastructure and speed up the highway construction project in response to the financial crisis in Asia. From 1998, the highway construction project entered the rapid development phase, and in 2001, the total distance of highways in China exceeded 19,000 kilometers, the second longest in the world after the United States. From 2003 to 2007, the construction project was further accelerated and was finished in 2007.

A.2. Price Quote Data

Figure 1.7: Price Quotes from Deppon



Notes: this figure shows the websites of the two logistics companies. I collect all the price quotes from these two companies for any pair of cities in China. There are three transport modes that the two companies provide. The fastest mode comes with a higher price. I always choose the fastest mode to be consistent. I then compare the quotes for every city pair and choose the lower quote as the transportation cost between the city pair.

A.3. Revenue Productivity, Prices and Marginal Cost

If firms with high productivity tend to pass some of the efficiency advantage to consumers in the form of lower prices, then regressing revenue productivity will produce downward bias.[58]. The reallocation effect may also be dampened by changes in prices. I use production quantity data at the firm level from 2000 to 2006 to estimate firm marginal costs and prices for a subset of firms in the Annual Survey of Industrial Enterprises.[59] I follow Garcia and Voigtländer (2013) to decompose revenue productivity into markup, marginal costs and physical productivity:

$$TFPR_{it} = p_{it}A_{it} = \mu_{it} \cdot MC_{it}(A_{it}, w_{it}) \cdot A_{it} \tag{1.23}$$

where p_{it} is the output price that firm i charges, A_{it} is firm i's physical productivity,μ_{it} is markup, MC_{it} is firm i's marginal cost and w_{it} is the input price of firm i.

As we can see from the above expression, firms with higher physical productivity will have lower marginal cost, and the two effects offset each other. Assuming markups do not change, then the change in revenue productivity is ambiguous. The fact that I find positive significant effects on revenue productivity suggests that physical productivity might have responded even more to highway access. Also note that the input costs of firms are also very likely to respond to changes in market access. If inputs shipped from other cities becomes cheaper in respond to increase in market access, then revenue TFP will cause even larger downward bias. If labor inputs become more expensive after highway expansion and the effect dominates the cheaper prices of other inputs, then revenue TFP will give us smaller biases.[60]

[58]Smeets and Warzynski (2013) construct a firm level price index to deflate revenue productivity and show that this correction yields larger international trade premia in a panel of Danish manufacturers. Garcia and Voigtländer (2013) show within-plant physical productivity growth after exporting, in contrast to most previous studies that use revenue productivity

[59]The firm-level quantity data is used by the National Bureau of Statistics in China to estimate the total quantity of production of major manufacturing products. Though the data does not cover all firms in the ASIE, it is meant to capture all median and large firms.

[60]Theoretically, if the effect of labor input is sufficiently large, it is possible that the effect will exactly

A.4. Bilateral Trade and Market Access

I derive an alternative measure of market access that is very similar to Redding and Venables (2005) and Donaldson and Hornbeck (2015) with the model in Section 3. For now, I will consider a one-sector version of the model to simplify derivations. An extension to the multi-sector version would be an interesting exercise but requires more theoretical and data work.[61]

Bilateral trade:

$$X_{ij}^s = \beta_s \times \frac{Y_i \times Y_j}{Y} \times \left(\frac{w_i \tau_{ij}^s}{\Theta_j^s}\right)^{-\gamma_s} \times \left(f_{ij}^s\right)^{-[\gamma/(\sigma_s-1)-1]}. \tag{1.24}$$

where $\Theta_j^{-\gamma} \equiv \sum_{k=1}^N \left(\frac{Y_k}{Y}\right) \times (w_k \tau_{kj})^{-\gamma} \times f_{kj}^{-[\gamma/(\sigma-1)-1]}$.

Similar to Redding and Venables (2005) and Donaldson and Hornbeck (2015), I define Consumer Market Access (CMA) of city j to be

$$\Theta_j^{-\gamma} \equiv CMA_j. \tag{1.25}$$

Rewrite (9)

$$X_{ij} = \frac{Y_i}{Y}(w_i \tau_{ij})^{-\gamma} CMA_j^{-1} Y_j. \tag{1.26}$$

Similarly, Firm Market Access (FMA) of city j is defined as:

$$FMA_i \equiv \sum_j \tau_{ij}^{-\gamma} CMA_j^{-1} Y_j. \tag{1.27}$$

Note we can also write CMA_j as:

$$CMA_j = \sum_i \tau_{ij}^{-\gamma} FMA_i^{-1} Y_i. \tag{1.28}$$

Donaldson and Hornbeck (2015) shows that $FMA_i = \rho CMA_i$ for some scalar $\rho > 0$. If

offset or even dominate physical productivity.

[61] Using China's Input-Output tables for 1997, 2002 and 2007, I am constructing industry-specific input supply access and output demand access.

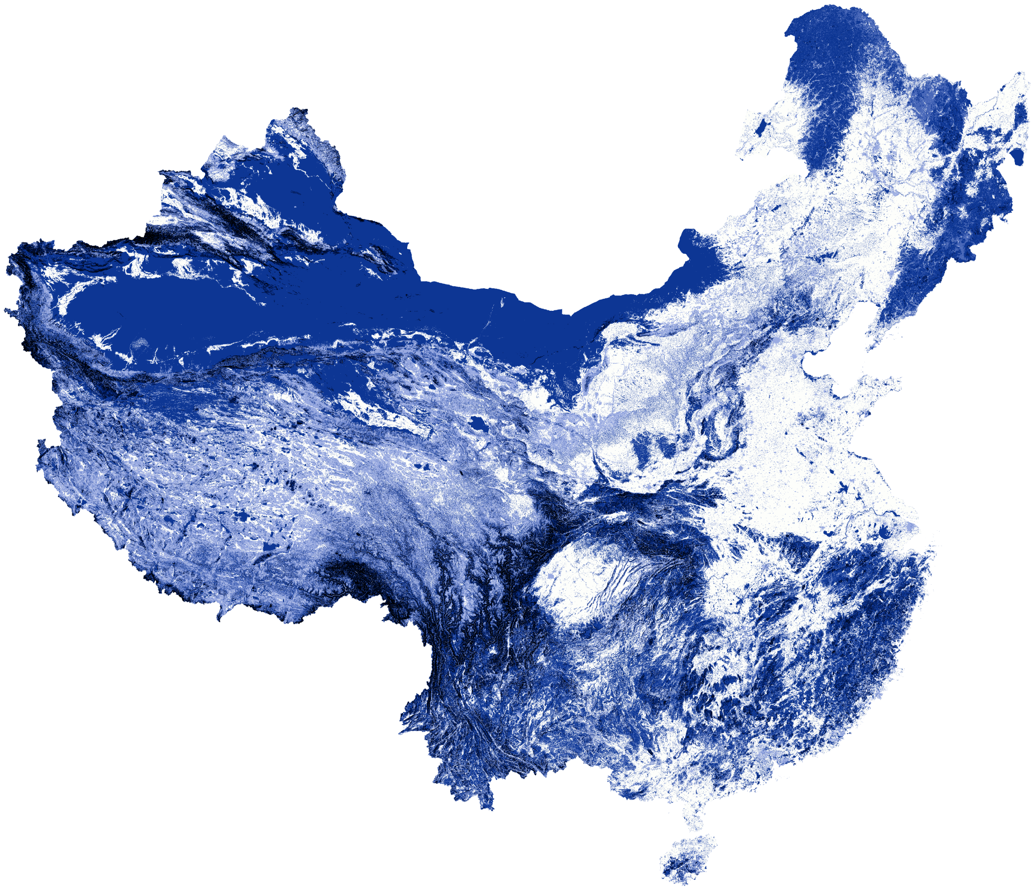
we let $MA_i \equiv FMA_i = \rho CMA_i$ equation (12) implies that:

$$MA_i = \kappa \sum_j \tau_{ij}^{-\gamma} MA_j^{\frac{-(1+\gamma)}{\gamma}} Y_j. \tag{1.29}$$

From the expression above, we see that a city's market access is the weighted sum of all its neighboring cities' market access. To solve the market access measure, I solve a system of 339 equations with MatLab.

A.5. Highway Construction Costs in China

Figure 1.8: Estimated Construction Costs



Notes: this figure presents the estimated construction costs for each pixel. I follow Faber (2014) to construct the construction costs using remote sensing data on terrain ruggedness. The darker the pixel is, the higher the construction costs.

Chapter 2

Upgrading by Importing: Machinery Imports and Productivity Growth

2.1 Introduction

It is commonly agreed among economists that productivity growth is the most fundamental driver of economic growth. The crucial role of technology in production and economic growth has long been recognized by various theoretical and empirical work on macroeconomics.[1] Theories of technology diffusion argue for the "trickle-down" effect: new technologies are invented in advanced countries and then diffuse to poor countries. A few channels have been proposed regarding how technologies trickle down from inventors to imitators. Given the large body of literature on gains from trade, a natural question is whether trade affects growth through the technology adoption channel. We still know little about the answer to this question.[2]

In this paper, I provide direct empirical evidence for a largely unexplored channel of technology diffusion through trade. I find that technology-scarce countries upgrade their production technology and increase production efficiency by importing machines and equipment from foreign countries. I examine the relationship between machinery importing and productivity growth using data for Chinese manufacturing firms. I argue that firms in technology-scarce countries can upgrade their production technology by importing machines

[1]See for example, Comin and Mestieri (2014)

[2]There is some indirect evidence for the effect of trade on technology diffusion. See, e.g., Coe and Helpman (1995)

and equipment from foreign countries.

The most crucial finding of this paper is that Chinese manufacturing firms improve their productivity by importing foreign machines and equipment, which I call the "Upgrading by Importing" (UBI) channel. A firm's productivity will gain by 6.7% after importing foreign machines. Machines from technologically savvy countries tend to improve firm productivity more than those from imitators. These results are robust to various regression specifications.

I find a fair amount of heterogeneity across industries. 12 of the 20 industries exhibit "upgrading by importing" (UBI), while I do not find strong productivity response to machinery importing in the rest. Table 2.9 in the Appendix presents estimation results by industry. We can see that material manufacturing industries and equipment manufacturing industries are the ones that exhibit strong "Upgrading by Importing" effect, whereas I do not find strong effects for wood, paper and furniture manufacturing, as well as chemical and petroleum products.

The baseline empirical analysis presented in this paper is silent on the underlying channels through which foreign machines improve firm productivity. Firms may be able to improve the quality of their products after acquiring new machines. They may also produce the same good at a lower cost. The outcome is likely to be a combination of the two channels. I disentangle the quality effect from the cost effect in the Section 3. I find that the quality channel is present in the data: firms import foreign machines tend to charge a higher price for its product. The effect is particularly strong when firms import machines from R&D intensive countries. I do not find strong evidence for efficiency gains, as firms import machines do not seem to improve their physical productivity.

The importance of the effect of imported capital on firm performance and aggregate productivity growth is twofold. First, the crucial role of capital in production and economic growth has long been recognized by theoretical work and empirical work. For example, an empirical literature pioneered by De Long and Summers (1991) has been trying to investigate the importance of capital goods investment on economic growth. De Long and Summers find that machinery and equipment investment has a strong association with growth, more than any other factors in their regressions. Second, recently economists have come to realize importing as a way of raising firms' productivity because imported intermediate inputs could be imperfect substitutes to domestic inputs. For example, Caselli and Coleman (2001) use imports of computers as a case study of technology diffusion and find a strong association between computer adoption and human capital across countries. Imported capital goods are a component of capital in a simple Cobb-Douglas production function, but more importantly, they also embody foreign technology that will enhance production efficiency. Following this argument, importing capital goods can be thought of as a way for firms to upgrade technology, especially in developing countries.

Very few studies in the literature have focused on the impact of machinery imports on productivity at the firm level. The challenge of identifying the effect of importing machines at the micro level comes from two empirical issues. The first challenge is due to limited availability of data. Very few firm-level datasets have detailed information on firms' importing behavior.[3] I develop a matching algorithm to merge a large Chinese firm panel dataset with the Chinese Customs transaction-level dataset. By merging the two datasets, I am able to document what products each firm imports at a very disaggregated product level (HS 8 digit). The second challenge stems from identification concerns. A firm's decision on whether and how much to import machines is likely to depend on the firm's productivity and other firm-specific characteristics. For example, if a firm decides to import machines and equipment in response to a positive productivity shock, then we will have a reverse causality issue.[4] Moreover, if a firm's importing decision is correlated with other endogenous decisions unknown to the econometrician, then the positive correlation between importing and productivity we observe may be spurious.

Given the difficulty of finding a valid instrumental variable in the context of this paper, I first use matching techniques and a structural approach to address endogeneity. I develop a propensity score matching method to select firms that are similar in many observable characteristics that may influence importing decisions. I only compare firms within groups that share similar characteristics before importing machinery goods. This method has been

[3]The firm-level datasets for Chile and Colombia do contain information on firm importing.

[4]An extensive literature in Industrial Organization has been trying to deal with the relationship between unobserved firm-level TFP shock and observed firm decisions (see Olley and Pakes 1996).

successfully implemented in many studies, especially in labor economics.[5]

To further address endogeneity concerns, I construct an empirical structural model to estimate firm TFP, explicitly incorporating firms' importing decision into an AR(1) productivity process. By imposing more structure on how a firm's productivity evolves over time, I will be able to examine firms' heterogeneous responses to machinery importing and conduct counterfactual analysis. With the structural approach, I aim to answer what the aggregate TFP gain would be if tariffs on machinery imports had been reduced or eliminated.

This study builds on a large literature on technology upgrading, investment and growth. The seminar work by De Long and Summers finds that machinery and equipment investment has a strong association with growth, more than any other factors in their regressions. Caselli and Wilson (2004) document large differences investment composition across countries and show that the composition of capital has the potential to account for some of the large observed differences in TFP across countries. Second, economists have started recently to emphasize the role of import liberalization in international technology diffusion. For example, Caselli and Coleman (2001) use imports of computers as a case study of technology diffusion and find strong association between computer adoption and human capital across countries. Imported capital can be thought of as a way for firms to upgrade technology, especially in developing countries. The existence of capital-skill complementarity, as emphasized by many papers such as Krusell et al. (2000), introduces reallocation gains of skilled labors during trade liberalization. This study departs from this literature by specifically looking at capital imports at the firm level. Imported machines and equipment embody foreign technology. Importing capital goods requires a large fixed cost and can be seen as a firm's choice to transfer advanced production technology. China is now the world's third largest economy that produces a wide range of machinery goods, which makes it particularly interesting and suitable for studying the impact of importing.[6]

[5] I also plan to construct an instrumental variable to address endogeneity concerns. I measure the (plausibly exogenous) *change* in each firm's travel distance to the nearest port and use the variable to instrument for the firm's importing decision in each year. The source of change in access to ports comes from a period of rapid expansion of China's National Trunk Highway System.

[6] Unlike many small developing countries in which firms must import certain machines because no or

As opposed to earlier papers which emphasize the exporting side of trade liberalization on technology upgrading,[7] there are a few recent papers starting to look at the effect of importing on technology upgrading and productivity. Studies have shown that improved access to foreign intermediate inputs increase firm productivity, for Indonesia (Amiti and Konings 2007), Chile (Kasahara and Rodrigue 2008), and India (Topalova and Khandelwal 2011). Amiti and Konings (2007) explore the effects of import tariff (on final output and intermediate input) changes on firms' TFP for Indonesia. The results show that a 10 percentage point fall in input tariffs leads to a productivity gain of 12 percent for firms that import their input, twice as high as any gains from reducing output tariff. In a related study, Goldberg et al. (2010) documents lower input tariffs account on average for 31% of the new products introduced by domestic firms after India's trade liberalization. I depart from this literature by specifically looking at capital imports. Halpern et al. (2015) estimate a structural model with Hungarian data and conduct counterfactual policy analysis to investigate the effect of imports on productivity.

Imported machines and equipment embody foreign technology. Importing capital goods requires a large fixed cost and can be seen as a firm's choice to transfer advanced production technology. I find two papers very closely related to this paper. Bas and Berthou (2013) investigate the complementarity between imported capital goods and imported intermediate inputs. They provide theoretical and empirical evidence on that intermediate input tariff reductions increase the possibility of importing capital goods for those firms in the middle range of the productivity distribution, due to capital-input complementarity. Halpern et al. (2015) empirically test the effect of capital imports on firm productivity with Hungarian

few domestic producers make such machines, Chinese firms choose to purchase domestic or foreign machines and that decision reflects technological differences or product quality differences between China and foreign countries. For example, Halpern et al. (2015) focus on Hungry, in which domestic producers can only produce a small subset of machinery goods.

[7] Yeaple (2005) is a seminal paper in which he builds a tractable model where ex ante homogeneous firms choose different level of technology and work skill. In his model, a reduction in exporting cost induces more firms adopt new technology and become more skill intensive. Bustos (2011b) builds on Yeaple (2005) to develop a Melitz-type of heterogeneous firm model with endogenous technology choice. He finds that the regional free trade agreement between Brazil and Argentina induced exporters to upgrade technology.

firm-level data. They find that the "R&D content" of capital is strongly positively correlated to total factor productivity. They also show that the share of imported capital is strongly positively related to productivity both within and across firms. This paper also focus on the effect of imports from trade liberalization.

There are also a few recent papers on importing and the demand for skill. Burstein and Vogel (2012) separate the Heckscher-Ohlin Mechanism at the industry level and the Skill-biased Technical Change at the firm level. Their model predicts trade liberalizations increase skill premium rises in all countries, even skill-scarce countries. Burstein and Cravino (2015) argue that if we take capital-skill complementarity as an important feature of technology, a reduction in world's trade costs may have important impact on countries' skill premium through sectoral reallocation of skilled and unskilled labor. At the firm level, Verhoogen (2008) first proposes the quality-upgrading mechanism linking trade and wage inequality in developing countries empirically examines its implications with a panel data on Mexican manufacturing plants. Koren and Csillag (2011) use Hungarian linked employer-employee data to estimate the effect of imported machines on the wages of machine operators. They match specialized machines with the type of workers that will operate these machines and find that workers exposed to imported machines earn 8 percent higher wages at the same firm. Voigtlaender et al. (2015) build an O-Ring type model with quality complementarity across input tasks to examine the impact of imported inputs on firms' skill demand. Fieler et al. (2014) estimate a model with heterogeneous firms and endogenous quality choices with Colombian data. They find skill premium and skill intensity in manufacturing increased, and the size of the firms decreased in Colombia.

The remainder of the article is organized as follows. Section 2 discusses the data and the algorithm I develop to merge the firm-level census with the Customs transaction-level dataset. Section 3 presents the baseline empirical specification and propensity score matching (PSM) method. Section 4 describes the instrumental variable strategy. Section 5 estimates an empirical structural model and conducts counterfactural analysis. The last section concludes.

2.2 Data

I use two large panel datasets for this study. The first one is the Annual Surveys of Manufacturing Firms, which is the only large scale manufacturing firm-level dataset available for China. The second dataset I use is the Chinese Customs dataset, which documents details of each international transaction conducted by Chinese firms. I develop an algorithm to merge these two datasets to combine detailed production and financial information for each firm with its importing and exporting activities for the period 2000-2007. The merge of the two dataset is essential for this project because I need firms' output, labor, capital and intermediate inputs to estimate firm-level TFP. With information on each firm's machinery imports, I can examine the effects of machinery importing on firm TFP.

The Annual Survey of Industrial Enterprises (ASIE) conducted by the National Bureau of Statistics of China span the period from 1998 to 2007.[8] The survey contains all State-Owned Enterprises (SOEs) and all private enterprises with an annual sales of 5 million RMB (roughly \$606,000 USD at the exchange rate of 8.25) and above. According to Brandt et al. (2012), firms in ASIE represent 90% of gross output in the manufacturing sector. The dataset contains very detailed information on firm's balance sheet and income statement, as well as information on ownership, export status, employment, among others.[9] For the year 2004, I have information on the education and skill level of firms' labor force. There are over 100 variables in the data, and the variables that are particularly useful for this project are firms identifier, industry identifier, gross output, total sales, wage bill, employment, stock of fixed capital, value of intermediate inputs, export status, year of establishment, ownership, skill/education level and location. For details, Please refer to Table 2.5 in the Appendix.

The China Customs data collect information of every export or import transaction on a monthly basis from 2000 to 2006. The data set contains price, quantity and value at the HS 8-digit level. For each trade, it also includes each Chinese firm's name, contact person,

[8]later years are available for both the ASIE and the Customs dataset. For this study, however, we only use data for the period 2000-2006.

[9]Please note that for many other countries, the unit of observation is a plant. For the Chinese data, however, the unit of observation is a firm.

telephone number, physical address, origin/destination country, and nature (processing or normal trade). One issue with the Customs data is that if a trade is done by a trading intermediary, the database only reports information of the trade intermediary. So, I am not able to identify firms which buys imported machines trough an intermediary. This data issue will potentially bias my results against finding any significant effect even if there is one. Fortunately, even with this potential bias, I still find a significant effect of machinery importing on firm TFP. To that end, the estimated coefficient should be viewed as the lower bound of the actual effect. All the 2-digit HS product codes that fall between 82 and 91 are defined as machinery and equipment products.[10]

There are two supplementary datasets I use. The first one is firm production quantity dataset. This dataset reports the quantity of production for each 6-digit product for a subset of firms in the ASIE. I use the data to estimate firm physical productivity. Digitized micro-geographic data is also sued to measure each firm's travel time to the nearest port. The change in travel time came from the expansion of China's National Trunk Highway system from 1998 to 2005.

2.2.1 Merging Datasets

In order to identify those manufacturing firms that directly import capital goods from abroad, I need to merge the two data sets. To do so, I develop a fuzzy-matching algorithm to merge the ASIE dataset with the Customs dataset.[11] The algorithm matches information on firm name, location, contact information from the two data sets using techniques of entity resolution. The merged sample shows that more than 90% of the matched firms are exactly matched by firm name, and the others are matched by information of firm name, telephone, contact person, and location. To compare the matching rate of my algorithm to other papers out there which focus on exporters, I also match exporters from the two data sets to check the performance of the matching algorithm. My algorithm can match 60% of the exporters

[10]There is a possibility that some of the machinery and equipment products are in fact used as intermediate inputs by some industries.

[11]The code for the matching algorithm is available upon request.

in the manufacturing firms' survey. This matching rate is higher than what is reported in a number of studies that also does the matching.[12]

I first follow Brandt et al. (2012) to create a firm-level panel for the period 1999 - 2007. I use id, official name, address, contact, and telephone numbers to merge firm survey data from different years.[13] After creating the firm panel dataset, I proceed in two steps to merge the ASIE with the Customs data.

The first step is to match firms with exactly the same name from the two databases. Since firm names often contains information such as location, ownership, and typos, I first clean both database to leave out those unnecessary but sometimes erroneous information. Also, the text encoding for Chinese is different from English, resulting in errors when the names are written with different encoding. Then I match the two datasets based on firm names. Over 90% of the firms are matched in this step.

The second step involves a fussy matching process that matches firms based on various information, such as firm name, contact person, location, and phone number. It is not required that firms have exactly the same information to be matched. Rather, similarity is scored based on the information and we match two firms if their similarity score is above a threshold.

2.2.2 Firm Productivity Estimation

In order to assess the impact of machinery importing on firm productivity, first I need to measure firm total factor productivity (TFP). There is a large literature in industrial organization that deals with various issues when I estimate firm productivity. A few influential papers propose methods to tackle the endogeneity issue between a firm's productivity shock and input usage (For details, see Olley and Pakes 1996; Levinsohn and Petrin 2003; Ackerberg et al. 2015).

I use the augmented Olley-Pakes (Olley and Pakes 1996) approach to estimate and calcu-

[12]For example, Fan et al. (2015) also merge the two datasets and obtain a matching rate of 50%.

[13]The official ASIE data is a repeated cross section. We are not able to use only firm id or name to merge all firms because firm id and name changed sometime over the sample period. For more details of the merging, please refer to Brandt et al. (2012).

late the TFP.[14] I use the perpetual inventory method to construct real capital stock, similar to Brandt et al. (2012). Other studies have used the same method to estimate firm TFP for China (Brandt et al. 2012; Feenstra et al. 2014b). There are mainly two approaches to estimating firm TFP. One is the Olley-Pakes approach that uses value added, and the other is the Levinsohn and Petrin (2003) approach that uses total output. The main difference between the two approaches is the proxy variable. Olley and Pakes 1996 use firm investment as the proxy for productivity shock whereas Levinsohn and Petrin (2003) use intermediate inputs.

Feenstra et al. (2014a) argues that the Olley-Pakes approach is more appropriate in the Chinese context because processing trade in China accounts for more than a half of the country's total trade since 1995. The prices of imported intermediate inputs are different from those of domestic intermediate inputs. Using the domestic deflator to deflate imported intermediate input would create another unnecessary source of estimation bias. A potential issue with the Olley-Pakes approach is that a large number of firms that have zero investment will be dropped from the estimation exercise. However, As shown in Brandt et al. (2012), in the Chinese data there is only negative real investment for 1% of continuing firms. Moreover, I do not observe investment decisions directly, but estimate investment from the capital stock series, which will smooth out most of the zero investment decisions. For robustness checks, I also estimate firm productivity using the ACF approach (Ackerberg et al. 2015). As we will see in Section 3, results are largely independent of the estimation approaches I use. I follow De Loecker and Warzynski (2012) to estimate firm markups. Since I use the value-added approach to estimate TFP, I choose labor as the "flexibly adjustable" input.[15]

[14] All nominal variables are deflated by input and output deflators. Deflators are taken from Brandt et al. (2012).

[15] Although it is more appropriate to use the value-added approach to estimate firm TFP, it is somewhat difficult to justify the assumption that labor input is fully flexible in China. Therefore, I also employ the output approach to estimate firm TFP and use material inputs as the "adjustable" input.

2.2.3 Revenue vs. Physical Productivity

The standard methods of TFP estimation mentioned above measures revenue-based total factor productivity (TFPR). TFPR has been widely used in the literature as a measure of efficiency. TFPR as an efficiency measure will be bias if prices respond to efficiency. Garcia and Voigtländer (2013) provide a simply illustration of the idea by decomposing TFPR into prices, P, and physical productivity, A: $ln(TFPR) = ln(P) + ln(A)$. Using TFPR as a proxy for A introduces bias when prices respond to efficiency. For example, when facing downward-sloping demand, firms typically respond to efficiency gains by expanding production and reducing prices. This generates a negative correlation between P and A, so that TFPR will underestimate physical productivity. Despite these shortcomings of TFPR, the majority of studies have used this measure to analyze productivity gains from exporting. One practical reason is the lack of information on physical quantities. While some corrections to the estimation of production functions have been proposed, only a few studies have derived A directly.

I use a subset of the firms[16] in the ASIE dataset who report quantities of production to estimate physical TFP. The existence of multi-product firms complicates the problem, and a solution to the estimation of TFPQ in the existence of multi-product firms is beyond the scope of this paper.[17] I only include single-product firms to circumvent this issue.

2.2.4 Descriptive Statistics

Table 2.5 shows the number of firms at the 2-digit industry level documented in the ASIE for the period 2000-2006. Table 2.6 shows summary statistics for a few variables that we are particularly interested in. We see that firm output, export, machinery import and productivity all grew substantially during the sample period. While roughly 25% of the firms are exporters, only 8% are machinery importers. I also compare the mean difference of characteristics between machinery importers and non machinery importers. We use the 2004

[16] These firms are usually the largest firms in each industry as the quantity dataset is used for estimating total quantity of production for each good by the Chinese Statistics Bureau.

[17] Garcia and Voigtländer (2013) provide a solution to the problem by using marginal cost as a proxy.

data because it has information on the skill levels of workers and firm R&D expenditures. Results are presented in Table 2.1. I find that firms importing machinery goods are larger and more productive than those do not. Firms that import machines also spend more on R&D and have more skilled labor force. It can be concluded that machinery importers differ from other firms in many important dimensions.

Table 2.1: Mean difference between machinery importers and non-importers

	Non importer		Machinery importer		
Variable	No. of obs	Mean	No. of obs	Mean	Difference of mean
Output (log)	248051	9.76	24704	10.87	1.11***
Value added (log)	243371	8.37	23740	9.44	1.07***
Employment (log)	251453	4.55	24728	5.46	0.92***
Exporter (dummy)	251741	0.23	24732	0.80	0.57***
Material importer (dummy)	251741	0.05	24732	0.83	0.79***
University graduates (% of total)	251453	3.68	24728	7.51	3.83***
No. of computers	251741	18	24732	109	91.09***
Productivity (log)	223421	3.77	23605	3.90	0.13***

Notes: This table summarizes statistics for machinery importers and non-importers in 2004. Difference of mean for each variable is also presented. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

2.3 Empirical Strategy

2.3.1 Baseline OLS

I use two baseline empirical specifications to estimate the effect of machinery importing on firm TFP. I first use a dummy variable indicating whether a firm imported machines and equipment. I also use the value of imported machines as a percent of firm's total capital stock. This normalization serves two purposes.

$$\omega_{i,t} = c + \beta_1 im_{i,t}^k + X_{i,t}\beta_X + \omega_{i,t-1} + \sigma_i + \epsilon_t \tag{2.1}$$

where $\omega_{i,t}$ is firm productivity this period; im_{it}^k is either a dummy variable indicating whether a firm imported machines or the value of machinery imports as percentage of total capital stock in period t; ω_{it-1} is firm productivity the previous period; $X_{i,t}$ are firm characteristics that I want to control for, such as exporter status and ownership structure; σ_i is firm-fixed effect that I also control for. This empirical specification can be rationalized by a simple firm investment model as in Das et al. (2007) or Halpern et al. (2015).

A firm's capital investment decisions should be made prior to the productivity shock realizes in this period. By including $\omega_{i,t-1}$, the previous period's productivity, we control for potential impact of productivity innovation last time on capital importing in the current period.

I present our baseline results in Table 2.2. I find a significant and positive effect of capital importing on firm productivity. On average, a firm would enhance its productivity by 7% by adding imported machines to its capital stock. Consistent with some of the previous studies, exporters are more productive than non-exporters, which may be a results of either selection or "learning by exporting". I also find that importing intermediate inputs improves firm productivity, which confirms findings in a few recent papers.

I also break capital imports by origin country. I rank country's R&D capability using OECD data on countries' R&D expenditures.[18] I find that the effect of imports from R&D intensive countries (im^r) are larger than hat of imports from countries that are poor in R&D activities (im^p). I do find larger capital imports from R&D intensive countries, which is consistent with trade theory emphasizing trade originated from comparative advantages.

I also present results for value of imported machines as a percentage of total value of capital stock. This exercise recognizes the possibility that the intensive margin may play an important role, i.e, "upgrading by importing" depends not only on whether a firm imports machines, but also on the value of machinery imports. Table 2.3. In Column (4)-(6), I only include firms that import machines in period t. The effect is larger and stronger compared to the case in which I include those firms with zero machinery imports. This result suggest either the matching process conducted induced some bias to the results, or that intensive margin is more important for "upgrading by importing".

As a robustness check, I estimate firm productivity using ACF approach (Ackerberg et

[18]Countries that are more R&D intensive than China are regarded as "R&D intensive". Please refer to Table 2.10 in the Appendix for a R&D intensity ranking.

al., 2015). As we can see from Table 2.7 and Table 2.8 in the Appendix, the results are not sensitive to which estimation approach we employ.

I find substantial heterogeneity across industries. Table 2.9 Some industries such as exhibit strong "upgrading by importing", but I do not find effects for industries such as.

I discuss potential threats to the baseline strategy. First, the matching process only matches 50% of the exporting firms. Second, the existence of trading intermediaries complicates the issue. Second, whether to import machines is an endogenous decision made by the firm at each period.

Table 2.2: Upgrading by importing: machinery importing dummy

Dep var: product price (p)		All firms			Only machinery importers	
	(1)	(2)	(3)	(4)	(5)	(6)
ω_{t-1}	0.120***	0.119***	0.119***	0.120***	0.119***	0.119***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
im^k	0.071***	0.068***	0.067***			
	(0.014)	(0.014)	(0.014)			
exporter		0.063***	0.060***		0.064***	0.060***
		(0.012)	(0.011)		(0.012)	(0.011)
im^{mat}			0.000			0.000
			(0.000)			(0.000)
im^k_r				0.057***	0.055***	0.054***
				(0.011)	(0.011)	(0.011)
im^k_p				0.037***	0.035**	0.035**
				(0.013)	(0.013)	(0.013)
Ownership control	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
No. of obs	1308606	1308606	1306367	1308606	1308606	1306367
Adjusted R^2	0.738	0.739	0.740	0.738	0.739	0.740

Notes: This table includes results from the baseline OLS regressions. The dependent variable is firm-level TFP at time t and the main explanatory variable im^k is a dummy variable indicating whether a firm imported machinery goods at time t. *exporter* is a dummy variable indicating whether a firm is an exporter or not. im^{mat} is the value of a firm's material imports. Numbers in the parenthesis are standard errors. Standard errors are clustered at the province level. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

Table 2.3: Upgrading by importing: value of imported machines (% of K)

Dep var: product price (p)		All firms			Only machinery importers	
	(1)	(2)	(3)	(4)	(5)	(6)
ω_{t-1}	0.120***	0.119***	0.120***	0.048***	0.048***	0.050***
	(0.013)	(0.013)	(0.013)	(0.008)	(0.008)	(0.009)
sim^k/K	0.002**	0.002**	0.002**	0.004***	0.004***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
exporter		0.064***	0.060***		0.019	0.015
		(0.012)	(0.011)		(0.016)	(0.016)
im^{mat}			0.000			0.000*
			(0.000)			(0.000)
Ownership control	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
No. of obs	1308606	1308606	1306367	107456	107456	107326
Adjusted R^2	0.738	0.739	0.739	0.721	0.721	0.723

Notes: This table includes results from the baseline OLS regressions. The dependent variable is firm-level TFP at time t and the main explanatory variable sim^k is the value of imported machinery goods as a share of total capital stock at time t. *exporter* is a dummy variable indicating whether a firm is an exporter or not. im^{mat} is the value of a firm's material imports. Numbers in the parenthesis are standard errors. Standard errors are clustered at the province level. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

2.3.2 Propensity Score Matching

I apply the propensity score matching (PSM) method as proposed by Rosenbaum and Rubin (1983). Propensity score matching makes the same identifying assumptions as OLS but avoids two additional restrictive assumptions. First, unlike OLS, a matching algorithm compares only comparable firms. This reduces the risk of comparing oranges to apples. Second, a matching algorithm identifies parameters in a non-parametric way. This allows for heterogeneous upgrading by importing effect based on firm characteristics. Despite all the differences between OLS and PSM, the consensus in the literature is that these two methods should not produce drastically difference results.[19] Otherwise, we may need to worry about serious model mis-specification issues, which would make it difficult to argue the validity of my baseline OLS results. It is reassuring that results with PSM method are very much in line with the baseline OLS results. Table 2.4 reports the effect of machinery importing on firm TFP for the propensity score matching method.

Table 2.4: Machinery Importing and Firm TFP: Propensity Score Matching

	All	20th	40th	60th	80th	100th
Avg treatment effect on the treated (ATT)	0.074***	0.017	0.134***	0.144***	0.172***	0.057***
	(0.006)	(0.117)	(0.052)	(0.028)	(0.011)	(0.007)
No. of obs	1059167	208786	212152	212796	212735	212698

Notes: This table includes results from propensity score matching. The dependent variable is firm TFP at time t and the main explanatory variable is machinery imports im_t^k at time t. I include all observations in Column 1. I then divide the sample into 5 subgroups based on firms' propensity scores. Results for the 5 subgroups are presented in Column 2-6, respectively. Numbers in the parenthesis are standard errors. Standard errors are clustered at the province level. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

[19] See Wooldridge (2009) for a discussion of the matching methods.

2.3.3 Quality vs. Costs

As mentioned in the introduction, revenue TFP measures can not disentangle quality upgrading from cost reduction. In this subsection, I investigate whether firms upgrade their product quality or reduce their production costs upon importing foreign machines. I decompose TFPR into prices, P, and physical productivity, A: $ln(TFPR) = ln(P) + ln(A)$. I use production quantity data for a subset of firms in the ASIE to estimate both price and physical productivity.[20] I divide total value of production by quantity to estimate product price. I then use estimated revenue TFP and price to estimate physical productivity. I regress both price and physical productivity on firms' machinery importing activity (both the dummy and the share). We find that most of the action came from the quality channel. Table 2.11 and Table 2.13 both show that importing machines raised firm product price, which is often used as a proxy for product quality. On the other hand, I do not find strong evidence for a cost reduction effect (see Table 2.12 and Table 2.14).

2.4 A Structural Approach

The goal of this paper is to argue for an "Upgrading by Importing" (UBI) channel. However, when we estimate productivity using either ACF or OP, we do not explicitly model how machinery importing affects the endogenous evolution of firm productivity. Instead, we only estimate a reduced-form relationship between machinery importing and estimated productivity. This reduced form relationship we see in the previous section is useful but suffers from two drawbacks. First, it is difficult to establish a causal relationship between machinery importing and productivity growth. Reverse causality, simultaneity and other types of endogeneity could all potentially bias my baseline OLS estimates. For example, a reserse causality statment can also explain the positive association between machinery importing and productivity growth: firms that experienced or expect a positive productivity shock are more likely to import machines and equipment. Second, without a structural model, it would be infeasible to consider counterfactual scenarios and estimate the aggregate impact

[20] I only use single product firms to avoid issues when dealing with multi-product firms

of machinery importing. While it is important to understand how much individual firms gain from foreign technology, we would also like to get a sense of the full extent of technology diffusion through the UBI channel.

In this section, I use an empirical structural model to estimate the impact of capital importing on productivity. The advantage of structural estimation is twofold. First, it tackles the endogeneity concerns mentioned above. Although instrumental variable approach has generated fruitful research and remained a powerful tool to address endogeneity, it is very difficult to find a valid instrument in the context of this paper. Structural approach on the other hand, explicitly model the productivity process, and allows us to estimate the heterogeneous effects across firms. Second, the model allows me to back out model parameters and conduct counterfactual analysis. I outline our structural approach and present our estimation results.

A firm's production function is assumed to take the standard log linear form (all the variables are log values).

$$y_{it} = \beta_l l_{it} + \beta_k k_{it} + \omega_{it} + \epsilon_{it} \tag{2.2}$$

Similar to what I do before, I use material inputs as the proxy variable, as Levinsohn and Petrin (2003) to control for unobserved productivity shock. The underlying rationale for using material as a proxy for productivity is that material inputs respond to productivity - if a firm receives a positive productivity shock this period, then it will use more material inputs to produce more output.

$$m_{it} = h(\omega_{it}, k_{it}, l_{it}) \tag{2.3}$$

Therefore, I can express productivity as a function of material inputs, capital and labor, assuming the function is invertible.

$$\omega_{it} = f(m_{it}, k_{it}, l_{it}) \tag{2.4}$$

I generalize the productivity process by explicitly including machinery imports (im_{it}^k) into the commonly use AR(1) productivity process. I are going to structually estimate the parameters that govern this process. In principle, I should estimate the parameters nonparametrically. However, as shown by Levinsohn and Petrin (2003), third order polynomials actually work well. I will follow their parametric approach and use third order polynomials.

$$\omega_{it+1} = g(\omega_{it}, im_{it}^{k}) + \xi_{it+1} \tag{2.5}$$

I use generalized method of moments to estimate the parameters. Identification comes from the widely used assumptions that capital takes more adjustment time than labor. The two moment conditions are as follows:

$$E[\xi_{it+1}(l_{it}, k_{it+1})'] = 0 \tag{2.6}$$

The estimation procedure has five steps:

- Regress y_{it} on l_{it}, k_{it} and the proxy variable for demand, intermediate inputs, to get predicted output $\hat{\phi}_{it}$.
- $\omega_{it+1}(\beta_l, \beta_k) = \hat{\phi}_{it} - \beta_l l_{it+1} - \beta_k k_{it+1}$
- Regress ω_{it+1} on $(\omega_{it}, im_{it}^{k})$ nonparametrically to obtain ξ_{it+1}
- Estimate a GMM estimator of (β_l, β_k) by minimizing the moment conditions
- Regress nonparametrically ω_{it+1} on $(\omega_{it}, im_{it+1}^{k})$ to get the effect of capital imports on productivity

In Figure 2.1, I show that there is large heterogeneity in the effect of machinery imports on productivity. However, over 90% of the firms would upgrade by importing. On average, foreign machines raised firm productivity by 5%. 80% of firms will gain 0% to 10% by importing foreign capital. This result is consistent with our baseline OLS regression and results from our propensity score matching. We also highlight that about 10% of the firms would have negative productivity growth if they imported foreign machinery. Not surprisingly, 86% of these firms did not report capital imports in the data. For the 90% of firms that would have positive productivity growth if they imported foreign machinery, only 8.5% of them did actually report capital imports. The TFP gains I estimate from machinery imports and the small percentage of importers I observe in the data have important policy implications. If policies could reduce the costs for firms to import foreign machines and equipment, there could be sizable aggregate TFP gains from machinery importing. For example, policies that reduces tariffs and other trade restrictions might have aggregate implications apart from the existing channels from trade theories.

2.5 Conclusion

In this paper, I study the effect of machinery imports on firm productivity. My baseline results suggest that a firm would raise its productivity by 7% by importing machines. This finding is robust to adding various control variables. Using a propensity score method and only comparing firms with similar observables, I find very similar results. The gain in revenue productivity mainly comes from the quality upgrading channel. I do not find evidence that firms reduced their production costs upon importing machines. Finally, I structurally estimate parameters that govern the productivity process, explicitly allowing machinery imports to affect the endogenous change in productivity. The estimation results confirm our baseline results. In addition, I find that a majority of firms do not import foreign capital, even though our estimation shows they could benefit from machinery imports.

The findings in this paper have important policy implications. I argue policies that are geared towards reducing tariffs or other trade barriers might introduce a new gain from trade liberalization through the technology-upgrading channel proposed by this study. This channel is still very much under-exploited . This study on Chinese manufacturing firms provide evidence for a sizable gain from machinery importing, and various trade frictions prevent firms from benefiting from this channel. It would be interesting to extend the analysis to other developing countries.

2.6 Appendix

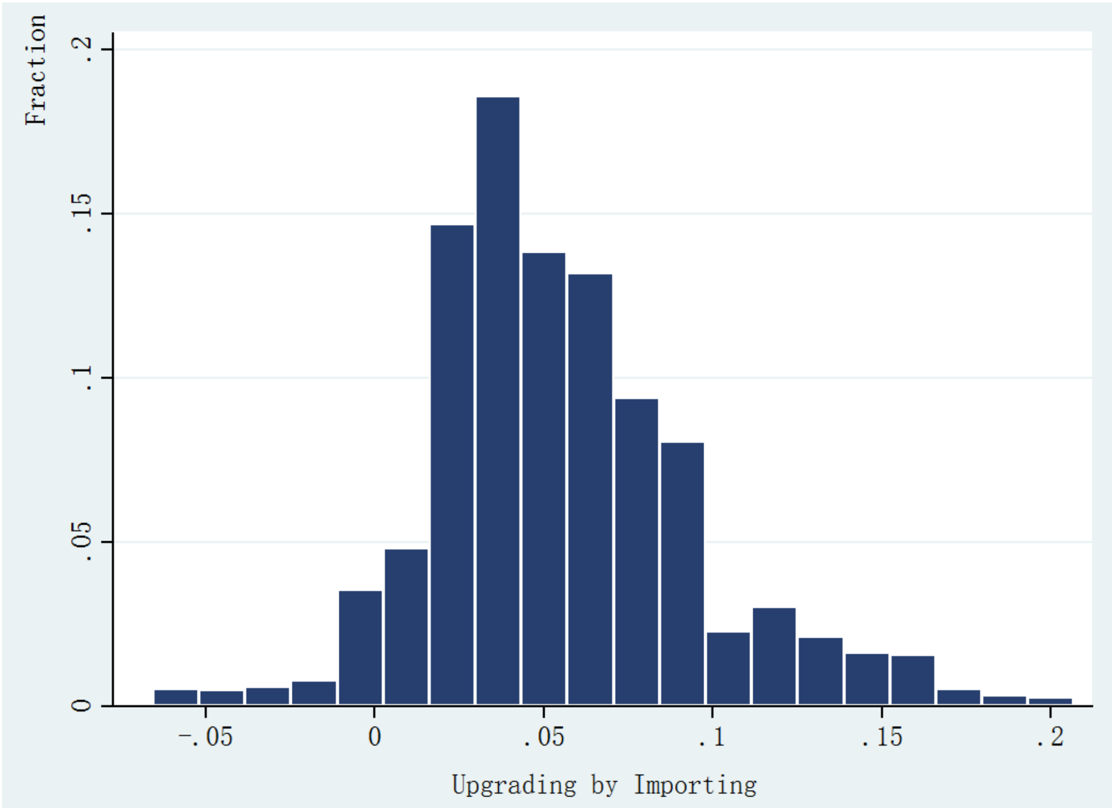


Figure 2.1: Heterogeneous Productivity Responses

Notes: This table includes results from the structural estimation. The y-axis shows percentage of firms in each bin and the x-axis shows the “UBI” impact.

Table 2.5: Annual Survey of Industrial Enterprises in China

CIC industry classification code	2000	2001	2002	2003	2004	2005	2006
Processing of Foods (13)	10676	10381	10413	11192	14097	14575	16356
Manufacturing of Foods (14)	4691	4563	4615	4636	5528	5553	6056
Manufacture of Beverages (15)	3409	3307	3287	3194	3469	3519	3914
Manufacture of Textile (17)	10968	12065	13248	14862	24192	22569	25345
Manufacture of Apparel, Footwear & Caps (18)	7064	8037	9061	9716	12029	11865	13072
Manufacture of Leather, Fur & Feather (19)	3164	3539	3932	4518	6393	6227	6859
Manufacture of Wood, Bamboo, Rattan, Palm & Straw Products (20)	2552	2808	3033	3501	5017	5397	6374
Manufacture of Furniture (21)	1498	1625	1767	2046	3025	3074	3603
Manufacture of Paper & Paper Products (22)	4672	5027	5285	5570	7473	7461	7892
Printing, Reproduction of Recording Media (23)	3701	3691	3806	4084	5139	4826	5029
Manufacture of Articles For Culture, Education & Sport Activities (24)	1879	2024	2327	2516	3382	3378	3633
Processing of Petroleum, Coking, &Fuel (25)	993	1027	1144	1323	2019	1990	2160
Manufacture of Raw Chemical Materials (26)	11430	12031	12637	13803	18759	18716	20715
Manufacture of Medicines (27)	3301	3488	3681	4063	4709	4971	5367
Manufacture of Chemical Fibers (28)	834	885	909	937	1536	1306	1402
Manufacture of Rubber (29)	1783	1777	1822	2016	3168	3034	3353
Manufacture of Plastics (30)	6230	6884	7665	8382	12269	12041	13504
Manufacture of Non-metallic Mineral goods (31)	14540	14707	15305	16245	19960	20111	21936
Smelting & Pressing of Ferrous Metals (32)	2997	3176	3333	4119	7141	6649	6999
Smelting & Pressing of Non-ferrous Metals (33)	2538	2823	2942	3367	5300	5163	5863
Manufacture of Metal Products (34)	8376	9274	10039	9746	14131	13802	15573
Manufacture of General Purpose Machinery (35)	9338	10027	10767	12546	20568	19981	22905
Manufacture of Special Purpose Machinery (36)	6406	6391	6546	7129	10925	10260	11615
Manufacture of Transport Equipment (37)	6850	7114	7470	8281	11823	11315	12586
Electrical Machinery & Equipment (39)	127	123	114	10400	16145	15366	16905
Computers & Other Electronic Equipment (40)	7845	8675	9385	5856	9161	8868	9709
Manufacture of Measuring Instruments & Machinery (41)	4459	4824	5320	2515	3916	3723	4084
Manufacture of Artwork (42)	1860	2018	2146	4259	5128	5131	5764
Total	144181	152311	161999	180822	256402	250871	278573

Notes: This table includes the number of firms in each 2-digit industry for the period 2000-2006.

Table 2.6: Descriptive Statistics

Year	No. of firms	Avg output	Avg employment	No. of exporters	No. of machinery importers	log TFP
2000	147253	51115	311	36818	12260	3.38
2001	155731	55282	289	40416	13914	3.54
2002	165861	62282	277	44953	15804	3.71
2003	181077	72923	270	50589	17393	3.86
2004	256612	66528	223	76596	25666	3.86
2005	251061	81351	236	74373	23877	3.98
2006	278752	90562	227	78185	26521	4.11

Notes: This table presents summary statistics for a few key variables for this paper. Roughly 25% of firms are exporters. only 8% are machinery importers.

Table 2.7: Upgrading by importing: machinery importing dummy (gross output approach)

Dep var: product price (p)		All firms			Only machinery importers	
	(1)	(2)	(3)	(4)	(5)	(6)
im^k	0.001***	0.001***	0.001***	0.002***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
exporter		-0.072**	-0.072**		-0.038	-0.037
		(0.034)	(0.034)		(0.042)	(0.042)
im^{mat}			0.000			0.001
			(0.000)			(0.000)
Ownership control	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
No. of obs	400418	400418	399855	34715	34715	34698
Adjusted R^2	0.827	0.828	0.828	0.907	0.907	0.907

Notes: This table includes results from the baseline OLS regressions. The dependent variable is firm-level TFP at time t and the main explanatory variable im^k is a dummy variable indicating whether a firm imported machinery goods at time t. *exporter* is a dummy variable indicating whether a firm is an exporter or not. im^{mat} is the value of a firm's material imports. Numbers in the parenthesis are standard errors. Standard errors are clustered at the province level. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

Table 2.8: Upgrading by importing: value of imported machines, % of K (gross output approach)

Dep var: physical TFP (ω)	All firms			Only machinery importers		
	(1)	(2)	(3)	(4)	(5)	(6)
sim^k	0.002	0.002	0.002	0.002	0.002	0.002
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
exporter		0.276***	0.275***		0.002	0.002
		(0.056)	(0.056)		(0.057)	(0.057)
im^{mat}			0.000			-0.001
			(0.000)			(0.003)
Ownership control	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
No. of obs	245539	245539	245179	21498	21498	21486
Adjusted R^2	0.871	0.872	0.872	0.904	0.904	0.904

Notes: This table includes results from the baseline OLS regressions. The dependent variable is firm-level TFP at time t and the main explanatory variable sim^k is the value of imported machinery goods as a share of total capital stock at time t. *exporter* is a dummy variable indicating whether a firm is an exporter or not. im^{mat} is the value of a firm's material imports. Numbers in the parenthesis are standard errors. Standard errors are clustered at the province level. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

Table 2.9: Heterogeneous effects by industry

Chinese Industrial Classification (2-digit)	Machinery importing dummy	Value of imported machines, as % of total capital stock
Upgrading by importing		
Processing of Foods (13)	0.102***	0.109***
Manufacture of Beverages (15)	0.072*	0.239*
Manufacture of Textile (17)	0.046**	0.058***
Manufacture of Apparel, Footwear & Caps (18)	0.046**	0.050***
Manufacture of Raw Chemical Materials (26)	0.052*	0.066***
Manufacture of Non-metallic Mineral goods (31)	0.092***	-0.002*
Smelting & Pressing of Ferrous Metals (32)	0.079**	0.140***
Manufacture of General Purpose Machinery (35)	0.045***	0.003**
Manufacture of Special Purpose Machinery (36)	0.048*	0.006**
Manufacture of Transport Equipment (37)	0.059**	0.009***
Electrical Machinery & Equipment (39)	0.074***	0.007***
Computers & Other Electronic Equipment (40)	0.053***	0.002***
No upgrading by importing		
Manufacturing of Foods (14)	0.062	0.035
Manufacture of Leather, Fur & Feather (19)	0.036	0.040
Manufacture of Wood, Bamboo, Rattan, Palm & Straw Products (20)	0.129***	0.023
Manufacture of Furniture (21)	0.067**	0.122
Manufacture of Paper & Paper Products (22)	0.042	-0.010
Printing, Reproduction of Recording Media (23)	0.003	0.078**
Manufacture of Articles For Culture, Education & Sport Activities (24)	0.024	0.020**
Processing of Petroleum, Coking, &Fuel (25)	0.040	0.047***
Manufacture of Medicines (27)	0.055**	-0.056
Manufacture of Chemical Fibers (28)	0.008	0.244**
Manufacture of Rubber (29)	0.030	-0.001
Manufacture of Plastics (30)	0.071**	0.000
Smelting & Pressing of Non-ferrous Metals (33)	0.039	-0.034
Manufacture of Metal Products (34)	0.037	0.009
Manufacture of Measuring Instruments & Machinery (41)	0.001	0.001*
Manufacture of Artwork (42)	0.062	0.009

Notes: This table includes results from the OLS regression above. The dependent variable is firm TFP at time t and the main explanatory variable is machinery imports m_t^k at time t. Numbers in the parenthesis are standard errors. Standard errors are clustered at the province level. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

Table 2.10: R&D intensive countries

R&D intensive countries	R&D intensity (relative to China)
Sweden	4.71
Israel	4.63
Finland	4.17
Japan	3.97
United States	3.47
Switzerland	3.33
Germany	3.16
Iceland	3.03
Denmark	2.87
Korea	2.86
France	2.84
Netherlands	2.61
Belgium	2.55
Austria	2.50
United Kingdom	2.39
Canada	2.37
Luxembourg	2.17
Norway	2.16
Australia	1.93
The Republic of Slovenia	1.80
Ireland	1.55
Czech Republic	1.50
Italy	1.34
The Russian Federation	1.32
new Zealand	1.29
Spain	1.13

Notes: This table list all countries that are classified as R&D intensive. Sweden, the most innovative country in the world for example, sees the technology content of its products are 4.7 times that of its Chinese counterparts.

Table 2.11: Quality Effect: machinery importing dummy

Dep var: product price (p)	(1)	(2)	(3)	(4)	(5)	(6)
im^k	0.037*	0.039**	0.039**			
	(0.019)	(0.019)	(0.019)			
exporter		-0.072**	-0.072**		-0.072**	-0.072**
		(0.034)	(0.034)		(0.034)	(0.034)
im^{mat}			0.000			0.000
			(0.000)			(0.000)
im^k_r				0.044**	0.045**	0.045**
				(0.020)	(0.020)	(0.020)
im^k_p				-0.001	0.000	0.002
				(0.023)	(0.023)	(0.023)
Ownership control	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
No. of obs	401124	401124	400483	401124	401124	400483
Adjusted R^2	0.827	0.828	0.827	0.827	0.828	0.827

Notes: This table includes results from regressing product price on machinery imports. The dependent variable is product price at time t and the main explanatory variable im^k is a dummy variable indicating whether a firm imported machinery goods at time t. *exporter* is a dummy variable indicating whether a firm is an exporter or not. im^{mat} is the value of a firm's material imports. Numbers in the parenthesis are standard errors. Standard errors are clustered at the province level. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

Table 2.12: Cost Reduction: machinery importing dummy

Dep var: physical TFP (ω)	(1)	(2)	(3)	(4)	(5)	(6)
im^k	0.038*	0.026	0.026			
	(0.022)	(0.022)	(0.022)			
exporter		0.275***	0.274***		0.275***	0.274***
		(0.056)	(0.056)		(0.056)	(0.056)
im^{mat}			0.000			0.000
			(0.000)			(0.000)
im^k_r				0.028	0.019	0.019
				(0.024)	(0.024)	(0.024)
im^k_p				0.026	0.021	0.020
				(0.028)	(0.028)	(0.028)
Ownership control	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
No. of obs	245539	245539	245179	245539	245539	245179
Adjusted R^2	0.871	0.872	0.872	0.871	0.872	0.872

Notes: This table includes results from regressing physical TFP on machinery imports. The dependent variable is firm-level physical productivity at time t and the main explanatory variable im^k is a dummy variable indicating whether a firm imported machinery goods at time t. *exporter* is a dummy variable indicating whether a firm is an exporter or not. im^{mat} is the value of a firm's material imports. Numbers in the parenthesis are standard errors. Standard errors are clustered at the province level. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

Table 2.13: Quality Effect: value of imported machines (% of K)

Dep var: product price (p)	(1)	(2)	(3)	(4)	(5)	(6)
im^k	0.001***	0.001***	0.001***	0.002***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
$exporter$		-0.072**	-0.072**		-0.038	-0.037
		(0.034)	(0.034)		(0.042)	(0.042)
im^{mat}			0.000			0.001
			(0.000)			(0.000)
Ownership control	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
No. of obs	400418	400418	399855	34715	34715	34698
Adjusted R^2	0.827	0.828	0.828	0.907	0.907	0.907

Notes: This table includes results from regressing product price on machinery imports. The dependent variable is product price at time t and the main explanatory variable sim^k is the value of imported machinery goods as a share of total capital stock at time t. $exporter$ is a dummy variable indicating whether a firm is an exporter or not. im^{mat} is the value of a firm's material imports. Numbers in the parenthesis are standard errors. Standard errors are clustered at the province level. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

Table 2.14: Cost reduction: value of imported machines (% of K)

Dep var: physical TFP (ω)	(1)	(2)	(3)	(4)	(5)	(6)
im^k	0.002	0.002	0.002	0.002	0.002	0.002
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
exporter		0.276***	0.275***		0.002	0.002
		(0.056)	(0.056)		(0.057)	(0.057)
im^{mat}			0.000			-0.001
			(0.000)			(0.003)
Ownership control	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
No. of obs	245539	245539	245179	21498	21498	21486
Adjusted R^2	0.871	0.872	0.872	0.904	0.904	0.904

Notes: This table includes results from regressing physical TFP on machinery imports. The dependent variable is firm-level physical TFP at time t and the main explanatory variable sim^k is the value of imported machinery goods as a share of total capital stock at time t. *exporter* is a dummy variable indicating whether a firm is an exporter or not. im^{mat} is the value of a firm's material imports. Numbers in the parenthesis are standard errors. Standard errors are clustered at the province level. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

Chapter 3

Gender Equality: What Policies Have the Biggest Bang for the Buck?

3.1 Introduction

In spite of the progress of the recent decades, gender gaps in various areas of economic opportunities and outcomes remain. Female labor force participation is lower than male participation in most countries, women's access to education is more limited than that of men and gender gaps in accessing social and financial services and in legal rights persist, especially in emerging markets and low-income countries. These disparities have adverse implications for women's economic productivity, income equality, and, ultimately, growth and economic development (Kochhar et al. (2017)). Indeed, vast empirical evidence (see Table 3.7)exists to show that significant macroeconomic gains can be realized when women are able to develop their full labor market potential (Elborgh-Woytek et al. 2013, Cuberes and Teignier 2016; Kochhar et al. 2017).

What can policymakers do, then, to promote gender equality in economic empowerment? What policies are more effective in fostering female labor force participation and women's opportunity to access education? Building on evidence from numerous micro studies, this paper conducts an empirical investigation based on macroeconomic data to estimate the impact of fiscal and structural policies on gender inequality in a sample of 100 countries from 1980 to 2014. Our study in particular focuses on policies that could address gender inequality in emerging and developing countries. With inequality of opportunity being a binding constraint in particular in these countries, we also assess their impact on educational opportunities for women, as measured by years of schooling and years in tertiary education. Realizing that there is likely no silver-bullet for policies to take effect across all income levels, we also examine whether certain policies are more critical in emerging markets and low-income countries (LICs) than in advanced economies.

By systematically assessing drivers of gender inequality discussed in the literature, this study provides new insights into which policies offer most effective solutions. One novelty of our paper is that it applies Bayesian Model Averaging (BMA) to identify the most fundamental and robust determinants of gender inequality. Against the background of the large number of potential determinants of gender inequality suggested in the literature, this methodology helps address model uncertainty as part of the statistical methodology (see, e.g., Fernndez et al. 2001; Sala-i Martin et al. 2004; Masanjala and Papageorgiou 2008).

We find that the scope for policies to narrow gender gaps is substantial and we identify policy actions that are likely to provide the largest "bang for the buck". Adequate infrastructure, more specifically, good sanitation facilities, contributes to closing the education gap between males and females in emerging markets and low income economies. This finding–already emphasized in case studies and other micro level analysis (e.g, Hannan and Andersson 2002, and Adukia 2016)–is for the first time confirmed with macro data. Low adolescent fertility, a narrower marriage age gap and public spending on education help lowering gender gaps in education. Good infrastructure, stronger legal rights for females, low adolescent fertility, a narrower marriage age gap and a stronger institutional environment (as measured by the corruption index) also support female labor force participation. A very important finding of our work is that labor market protection appears to have a non-linear effect on labor market participation gaps. At lower levels of labor market protection, stronger protection is associated with narrower labor force participation gaps, but at higher levels of protection, excessive labor market rigidity weighs on female labor force participation. This result is consistent with World Bank (2013) that finds that "employment protection legislation and minimum wage can shift employment away from young people, women, the less skilled". But there exist a sort of "plateu" for labor market protection at which changes in employment protection have minimal effect on employment (and productivity). "However, when the edge of the plateau is reached (either on the too-strict or the too-loose side) impacts are more negative".

The paper is organized as follows. Section II reviews the existing literature; Section III presents stylized facts on the relationship between macroeconomic and structural policies, and gender gaps in advanced economies, emerging markets and LICs. The data and empirical methodology are discussed in section IV, while Section V presents the results. Section VI discusses the conclusions and policy implications.

3.2 Literature Review

Gender equality encompasses a variety of aspects. It includes dimensions, such as equality in the access to education, health and financial access for women and men, equality in labor force participation, and political representation. When analyzing the impact of policies on gender equality, the cross-country literature has typically focused on labor market outcomes, namely, female labor force participation or employment, while studies at the country level have also more deeply analyzed how policies impact inequality of opportunity, such as school enrollment rates. This overview provides a short summary of the main areas of work. Since the theoretical and empirical literature on female labor force participation is vast, this overview does not aim to be exhaustive.

Female labor force participation is positively correlated with educational attainment for women. Calibrating a dynamic model of labor supply, Eckstein and Lifshitz (2011) find that one-third of the increase in female employment during the last century in the United States can be attributed to education. In an empirical exercise, Steinberg and Nakane (2012), show that a one standard deviation increase in the education level in Organization for Economic Cooperation and Development (OECD) countries is associated with a 3 percentage point increase in female labor force participation.

A number of studies have pointed to the theoretical underpinning of female labor supply. Female labor supply is often modeled using the framework of the time allocation model (Becker 1965), which posits that women make their labor supply decisions not only considering leisure and labor, but also home-based production of goods and services (including caring for children). Working for a wage is chosen only if earnings at least make up for the lost home production (and the associated costs), implying a higher elasticity of female labor supply to wages. Most studies have emphasized the importance of education in models of female labor supply. A number of studies have also included wages as a key in modeling female labor supply models (Heckman and Macurdy 1980). Fernández and Wong (2014) develop a dynamic life-cycle model with incomplete markets and risk-averse agents who differ in their educational endowments and make work, consumption, and savings decisions. They find that, in addition to the above factors, divorce risk has a large impact on married women's participation rates. Eckstein and Lifshitz (2011) estimate a dynamic stochastic female labor supply model with discrete choice (contained in Eckstein and Wolpin 1989) and find that changes in education (accounting for a third of the increase in female employment) and wages (explaining about 20 percent) play a large role in explaining female employment. They also formulate a new framework that models intra-family dynamics (using dynamic stochastic games) and relate it to the household's labor supply decision.

Fertility and higher marriage rates have been shown to significantly affect female labor force participation. For individual countries, there is evidence of a negative relationship between fertility and women's participation in the labor force. For instance, Bloom et al. (2009) find that the number of births is significantly negatively related to women's labor supply, with each birth on average decreasing women's labor supply by almost two years during a women's reproductive life. Mishra and Smyth (2010) estimate that a 1 percent increase in the fertility rate results in a 0.4 percent decrease in female labor force participation rates in G7 countries. While there is a negative relationship between the variables at the individual country level, there is a positive relationship between fertility and female labor force participation at the cross-country level. Using data from OECD economies, De Laat and Sevilla-Sanz (2011) explain the puzzle of a negative relationship at the individual country level, but a positive one across countries, by taking into account mens contribution to home production. They find that women living in countries where men participate more in home production are better able to combine motherhood with work outside the house, leading to greater participation in the labor force at relatively high fertility levels. The trade-off between family and work is also reflected in a negative correlation between female labor force participation and marriage rates.

Fiscal policies that are tailored to country-circumstances can significantly increase female labor force participation (Aguirre et al. 2012; Duflo 2012; Revenga and Shetty 2012; Sen 2001; Thévenon 2013; Kalb 2009). On the revenue side, tax credits or benefits for low-wage earners can stimulate labor force participation, including among women. By reducing the net tax liability or even turning it negative, tax credits increase the net income gain from accepting a job. Such credits are usually phased out as income rises. Policies can also build on the fact that female labor supply is more responsive to taxes than male labor supply (IMF 2012). For example, a switch from family income taxation to individual income taxation that reduces the tax burden for (predominantly female) secondary earners can support female labor force participation, while it would affect the less-tax-elastic male labor supply to a smaller extent.

As for expenditure policy, better access to comprehensive, affordable, and high-quality child care frees up women's time for formal employment (Gong et al. 2010). The elasticity of female labor supply with respect to the price of child care has been estimated to range from -0.13 to -0.2. Thus, reducing the price of childcare by 50 percent could be associated with an increase of 6.5 to 10 percent in the labor supply of young mothers. Other studies document the importance of public infrastructure to boost the participation of women in the labor force. Cubas Norando (2010) finds that a large part of the difference in female labor force participation rates in 1990 between the United States, on the one hand, and Brazil and Mexico, on the other, can be explained by the availability of basic infrastructure (electricity and running water). Ghani et al. (2013) note that inadequate infrastructure affects women's participation more than that of men because women are more often responsible for household activities. Das et al. (2015) estimate that female labor force participation in India would rise by 2 percentage points if Indian states increased education spending by 1 percent of GDP.

Gender-based legal restrictions impede women's empowerment and thus their economic participation. Gonzales et al. (2015) examine the effect of gender-based legal restrictions and other policy choices and demographic characteristics on female labor force participation. Drawing on a large and novel panel data set of gender-related legal restrictions, they find that restrictions on women's rights to inheritance and property, as well as legal impediments to undertaking economic activities such as opening a bank account or freely pursuing a profession, are strongly associated with larger gender gaps in labor force participation.

A lower female age at marriage, and large age gaps between men and women at marriage have been associated with high gender inequality. Marrying at a younger age is associated with becoming a parent at younger age and thus impacts an individuals' educational investment decision. For instance, using data from rural Bangladesh, Field and Ambrus (2008) show that each additional year that marriage is delayed is associated with about one fifth additional year of schooling and 5.6 percent higher literacy for women. Arguing that a lower age of marriage for women may simply reflect lower marriage ages for both men and women in a society, Stimpfle and Stadelmann (2016) test the relationship between the gap between men and women at marriage and estimate that an additional age difference between husband and wife of one year reduces female secondary schooling completion rates by 14 percentage points.This, in turn, adversely impacts female education more than male education, therefore increasing the education gap.

3.3 Stylized Facts

Before turning to the empirical investigation, we take a preliminary look at the data on our sample of 100 countries from 1980 to 2014, to uncover possible correlations among policy variables and gender inequality in education and labor force participation, as well as to highlight any differences between developing countries, emerging markets and advanced countries.

Our data suggest that higher public spending on education is associated with a narrower gender gap in years of schooling in both developed and developing countries, although the relationship appears to be stronger in the latter (Figure 3.4), possibly reflecting larger marginal returns where gaps are higher. Better infrastructure, as measured by improved access to sanitation facilities, is negatively correlated with the gender gap in years of schooling in emerging markets (Figure 3.5). This may be because in countries with better infrastructure, females can reduce the time spent on household activities, and increase the time in school. In most societies, females are responsible for household water supply and sanitation. This activity can be very time consuming in areas lacking of adequate access to water and infrastructure. For example, collecting water is estimated taking as much as 26 percent of women's time in rural Africa (Lamb 2015). This effect is not evident in advanced economies, where access to sanitation facilities is similar across countries. Finally, the data suggest that improved access to finance is associated with a narrower gender gap in tertiary enrollment, but not in secondary or primary enrollment, possibly reflecting higher cost of attending tertiary school (Figure 3.6).

Our data confirm the finding by Gonzales et al. (2015) that more equal legal rights are associated with narrower labor force participation gaps (Figure 3.2). Higher public spending on education is also associated with lower labor force participation gaps: the former is associated with lower education gaps, and more highly educated women have larger incentives to join the labor market (Figure 3.3). However, the effect seems to be mainly driven by advanced economies. This may reflect different efficiencies in public spending in providing the skill sets that would make it easier for women to access the labor market. The impact of labor market regulation on labor force participation gaps seems to be equally ambiguous, with some relationship in advanced economies, but less so in emerging and developing economies. (Figure 3.1)

3.4 Empirical Specification

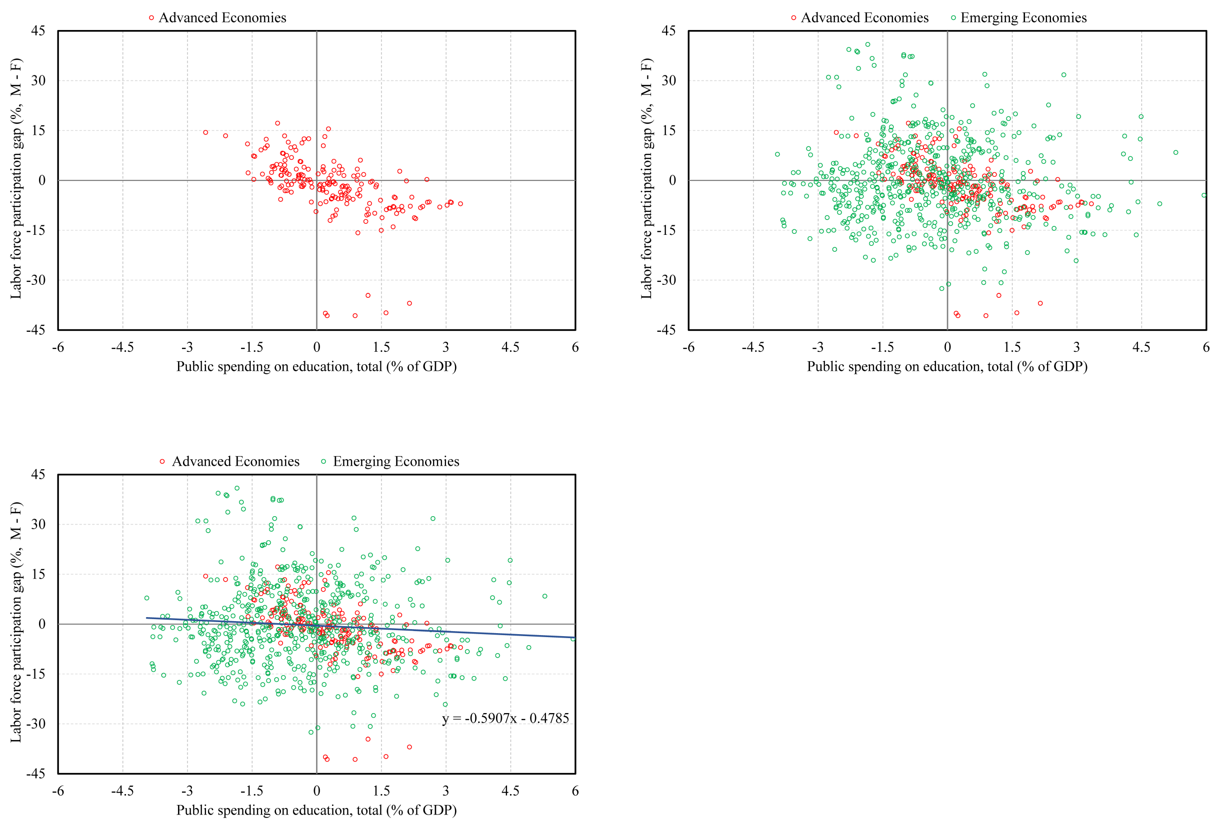
A fixed-effect panel regression is estimated to assess the impact of policies on gender gaps, while controlling for country structural characteristics. The model allows accounting for potential unobserved non-time variant factors at the country level, and controlling for global factors which may have influenced the gaps similarly in certain points of time. Specifically, we estimate the following relationship:

$$Gap_{k,i,t} = \alpha + \delta Policy_{k,i,t} + \beta X_{k,i,t} + \mu_k + \nu_t + \epsilon_{k,i,t}$$

where $Gap_{k,i,t}$ is gender gap in either labor force participation or education in county i in region k and in year t. $Policy_{k,i,t}$ captures various types of policies, specifically:

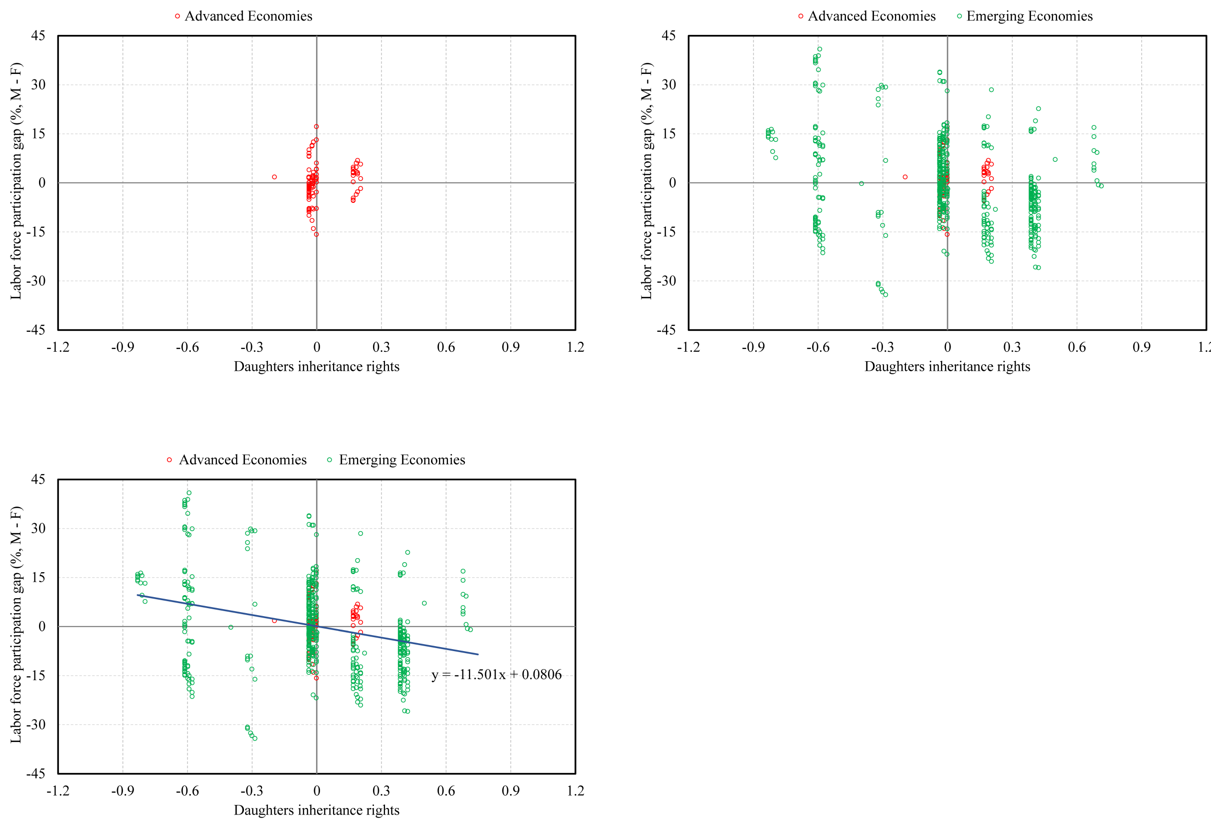
Fiscal policies: public spending on education and on health, that are expected to

Figure 3.1: Gender gap in labor force participation and public spending on education



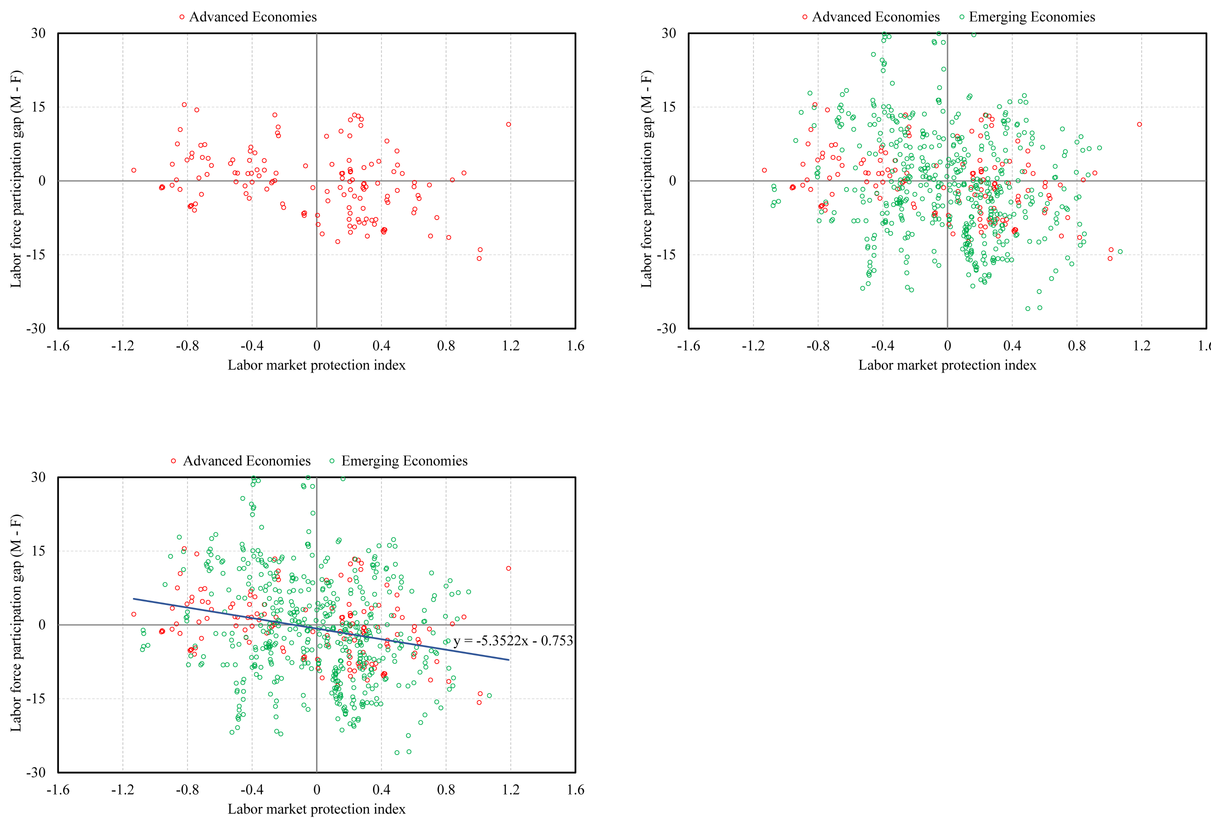
Notes: The figures show the relationship between gender gap in LFP and public spending on education. I first plot advanced economies and emerging markets separately, then plot all countries together. I find that education spending reduces gender gap in LFP only in advanced economies.

Figure 3.2: Gender gap in labor force participation and daughter inheritance rights



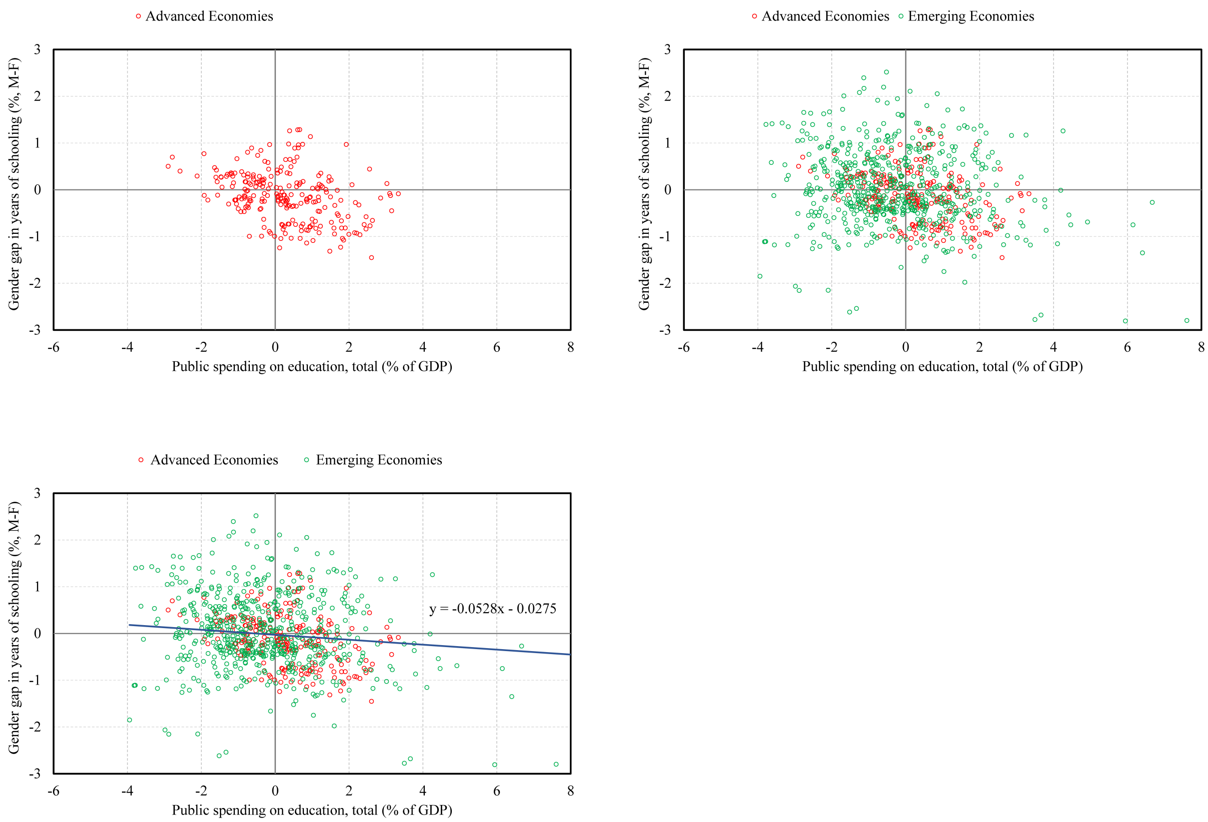
Notes: The figures show the relationship between gender gap in LFP and public spending on education. I first plot advanced economies and emerging markets separately, then plot all countries together. I show that more equal legal rights boosts female labor force participation.

Figure 3.3: Gender gap in labor force participation and employment protection



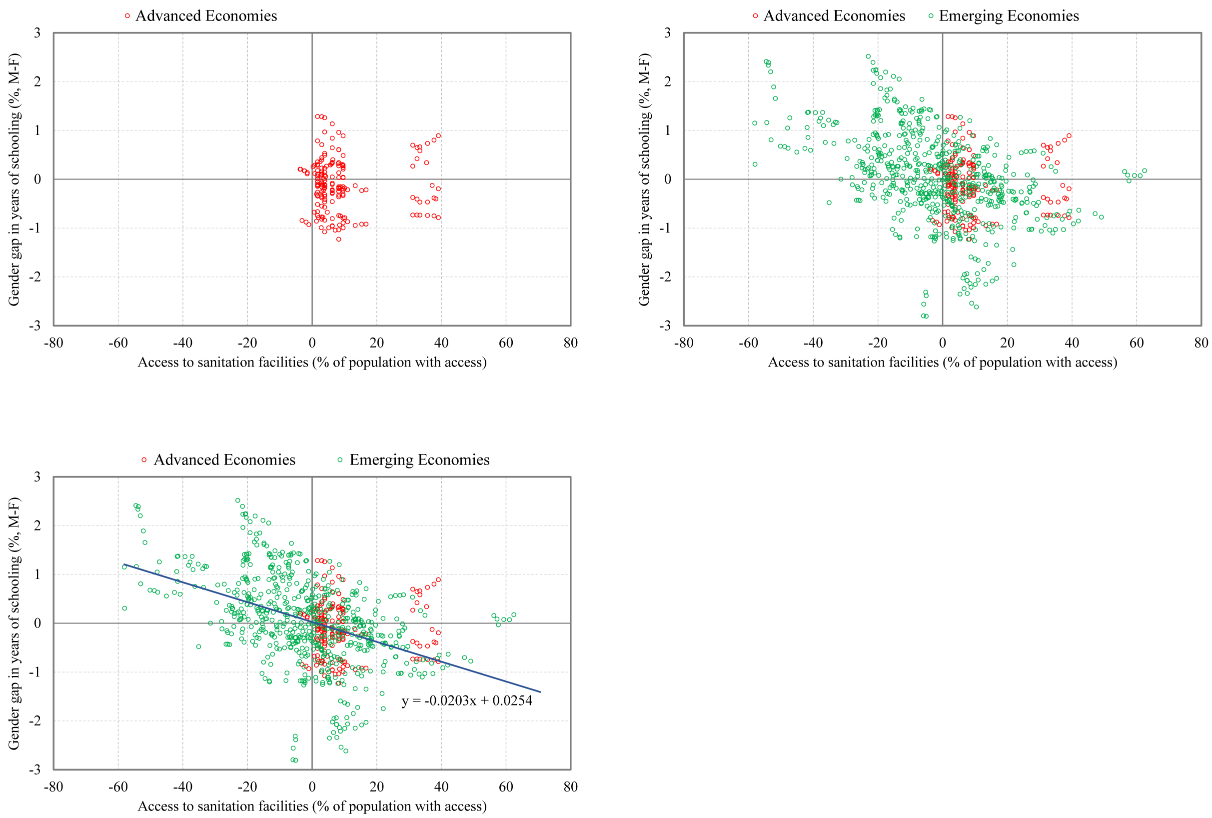
Notes: The figures show the relationship between gender gap in LFP and public spending on education. I first plot advanced economies and emerging markets separately, then plot all countries together. Employment protection laws narrow gender gap in LFP.

Figure 3.4: Gender gap in education and public education spending



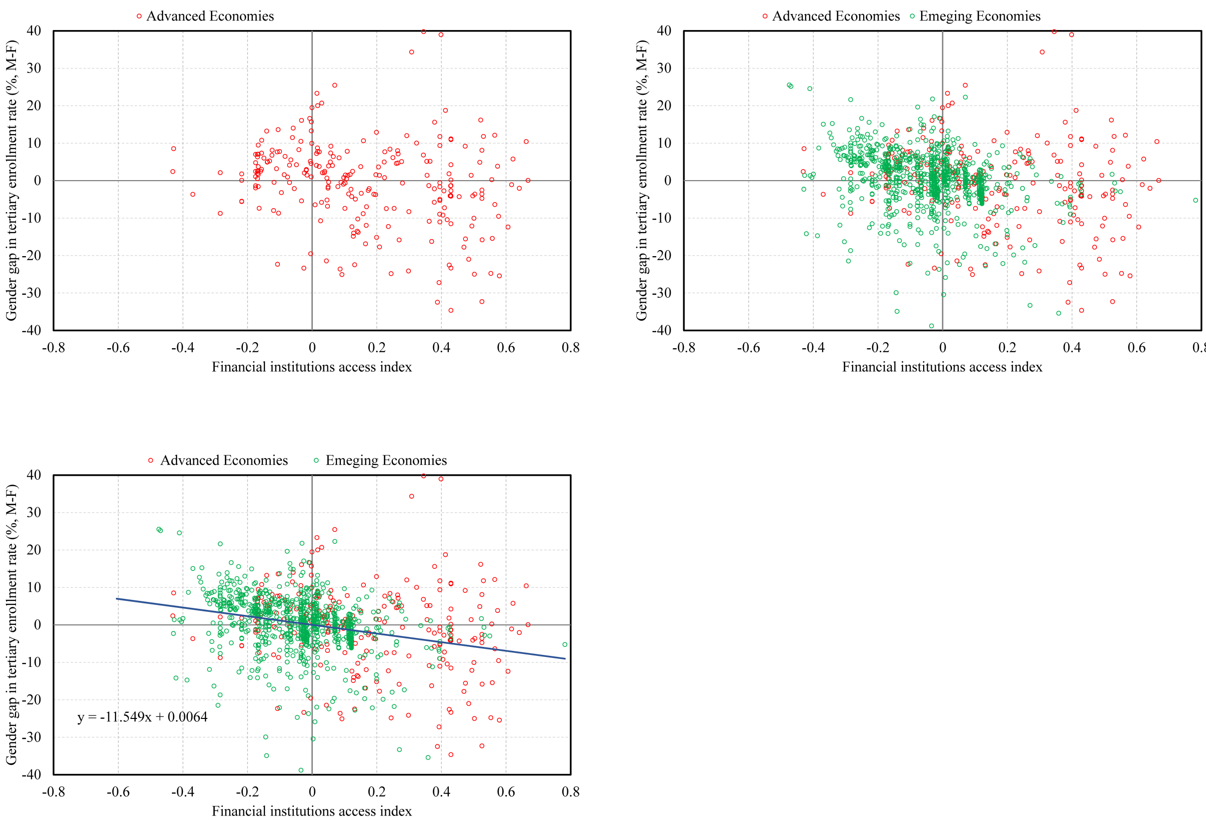
Notes: The figures show the relationship between gender gap in LFP and public spending on education. I first plot advanced economies and emerging markets separately, then plot all countries together. We see that fiscal expenditures matter for reducing education gap.

Figure 3.5: Gender gap in education and access to sanitation facilities



Notes: The figures show the relationship between gender gap in LFP and public spending on education. I first plot advanced economies and emerging markets separately, then plot all countries together. I find evidence that better infrastructure reduces education gap.

Figure 3.6: Gender gap in tertiary enrollment and access to Finance



Notes: The figures show the relationship between gender gap in LFP and public spending on education. I first plot advanced economies and emerging markets separately, then plot all countries together. Better access to finance promotes female tertiary education.

provide both opportunities to attend school and better skills to join the labor market; and the income tax rate as a proxy for incentives to join the labor market.

Several indicators of *infrastructure*: access to sanitation facilities, access to electricity, access to improved water source, and telephone subscription rates. Good infrastructure would decrease the time needed to spent on household activities-often disproportionately allocated to women and girls-thus freeing up their time to attend school or join the labor market. Moreover, adequate infrastructure provides a safer environment to travel to and attend school.

Structural policies related to the labor market, the product market, public safety as well as access to finance. These policies are proxied by a labor market protection index, trade openness, the political risk rating, and the control of corruption. Annex II gives an overview of the labor market protection indices examined in our analysis.

A range of *legal variables* to capture the equality between men and women under the law, including: equal inheritance rights for daughters and sons, women's right to head a household, and guaranteed equality of women and men before the law, from the World Bank's Women, Business and the Law database.

$X_{k,i,t}$ is a vector of control variables. They include (i) GDP per capita and its square to capture economic development, (ii) fertility and neonatal mortality rates, and (iii) other social factors, such as the gap in the marriage age between men and women and absolute age at marriage which determine women's educational investment decisions and time to spend in the labor market relative to men, and serve as a proxy for attitude towards women. μ_k are region fixed effects, and ν_t are year fixed effects.

3.5 Results

3.5.1 Gender Gaps in Education

Countries' level of development and demographics help explain the variation of gender gaps in education across countries, and the scope for policies to narrow these gaps is large (Table

3.1).

The relationship between gender gaps and GDP per capita is nonlinear. At early stages of development, higher development outcomes lower the gender education gap, but seems to increase the gap in economies at a higher development level, reflecting gaps in tertiary education. However, at higher income levels, increases in GDP per capita are associated with narrower gender gaps in tertiary education (Table 3.1).

Fertility and marriage gaps are closely related to gender gaps in education. Higher fertility rates at the adolescent stage are strongly associated with wider gender gaps in education, reflecting girls dropping out from school to take care of their children. A larger marriage age gap–with the average men being older than the average women or girl at the time of marriage in most countries–is associated with larger gender gaps in education, as men on average have more time to pursue education before starting a family.

A stable and safe environment, as measured by indices of political stability and public safety, narrows the gender gap in education. Improved access to sanitation facilities, is associated with narrower gaps in years of schooling between girls and boys, as it frees time from household activity and provides a safer environment at schools. Higher financial development seems to help lowering gender gaps in tertiary education, likely reflecting the higher cost of these services in many countries. Higher public spending on education and health is associated with lower gender gaps in tertiary education.

3.5.2 Gender Gaps in Labor Force Participation

Several factors help explain the gender gap in labor force participation, and policies can play an important role (Table 3.2).

The results confirm the well-documented relationship between the gender gap in labor force participation and the level of development. At low stages of development, female labor force participation is high, since both women and men need to work for subsistence. At higher levels of GDP per capita, female labor force participation rates decrease (and the gender gap widens), reflecting trade-offs between family care and joining the labor markets. Finally, female labor force participation rises again (gender gap narrows) when income levels move beyond a certain threshold, reflecting greater opportunities in the labor market.

Table 3.1: Gender Gap in Years of Schooling: OLS Regressions

Dependent Var: gender gap in yrs of schooling (%, M-F)	(1)	(2)	(3)	(4)*	(5)	(6)*
Stages of development						
Log(GDP per capita)	-0.327	-1.635***	-1.173***	7.956**	-0.635*	13.477***
	(0.421)	(0.313)	(0.295)	(3.524)	(0.328)	(3.687)
Log(GDP per capita) squared	0.010	0.091***	0.070***	-0.601***	0.047**	-0.971***
	(0.025)	(0.018)	(0.017)	(0.223)	(0.019)	(0.229)
Health and Demographics						
Maternal mortality ratio	0.001***	0.000**	0.000	-0.010***	0.000	-0.001
	(0.000)	(0.000)	(0.000)	(0.002)	(0.000)	(0.001)
Adolescent fertility rate	0.006***	0.006***	0.000	0.056***	0.003**	0.057***
	(0.001)	(0.001)	(0.001)	(0.013)	(0.001)	(0.015)
Marriage age gap	0.045**	0.157***	0.158***	0.496***	0.084***	0.210
	(0.018)	(0.020)	(0.020)	(0.127)	(0.018)	(0.131)
Institutions						
Political Stability and Public Safety		-0.123***	-0.125***	-3.440***	-0.058**	-2.426***
		(0.028)	(0.026)	(0.345)	(0.027)	(0.387)
Infrastructure						
Access to improved sanitation facilities			-0.016***	-0.014	-0.017***	0.010
			(0.001)	(0.014)	(0.002)	(0.015)
Financial Access						
Financial institutions access				-8.489***	0.092	-5.430**
				(2.141)	(0.096)	(2.317)
Fiscal Policies						
Public education expenditure					-0.130***	-0.202*
					(0.019)	(0.116)
Public health expenditure					0.003	-1.519***
					(0.017)	(0.276)
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	1773	1688	1656	1076	1448	928
Adjusted R-squared	0.432	0.377	0.429	0.516	0.469	0.561

Robust standard errors in parentheses. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

* The dependent variable is gender gap in tertiary school enrollment rate (%, M-F).

Notes: This table includes results from the OLS regression. The dependent variable is gender gap in education. Numbers in the parenthesis are standard errors. Standard errors are clustered at the province level. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

Fertility and social factors are significantly related to labor force participation gaps. Mirroring the results on education gaps, higher fertility rates are strongly associated with wider gender gaps in labor force participation. Higher age gaps at marriage are also related to larger labor force participation gaps, as women's time to participate in the labor market becomes more limited relative to men. Conversely, the absolute age at marriage for women is not significantly related to the labor force participation gap, possibly reflecting the fact that factors affect both male and female young marriage age jointly in some countries.

Legal rights boost female labor force participation. In particular, equal inheritance rights are strongly related to lower labor force participation gaps, as is a better institutional environment, as measured by the control of corruption, and the economic risk rating of the country.

Adequate infrastructure matters, and improved sanitation facilities play the most important role in narrowing the gender gap in labor force participation. The impact of improved sanitation facilities likely arises through their effect on the time needed for household work and on security.. The marginal impact of these measures, however, becomes lower at higher levels of development. Improved access to telephone lines can help provide women better access to the labor market. However, the effect is nonlinear, with men benefiting first from these improvements, while the gender gap in labor force participation only shrinks once access to telephone lines has reached a critical level.

Higher public expenditure on education is associated with narrower labor force participation gaps in advanced economies. Conversely, the impact of public expenditures on labor force participation gaps is smaller for emerging markets. This may reflect better targeting of education expenditures to the needs of the labor market in more advanced economies.

The impact of labor market protection is non-linear. At lower levels of labor market protection, an increase in the strength of protection is associated with narrower labor force participation gaps, as women seem to disproportionately benefit from stricter regulation. However, the marginal impact of protection decreases at higher levels of protection, with excessive labor market rigidity weighing on the benefits.

The results on labor market protection raises the question on whether the effect arises through higher female labor force participation or possibly lower male labor force participation. To examine these possible asymmetries for men and women, we report also the results of separate regression of male and female labor force participation rates on all the determinants. We find that stronger labor protection laws significantly increase female labor force participation rate, whereas there is some evidence, albeit weak, that stronger labor protection laws lower male labor force participation rate (see column 1 and 2 of Table 3.2). These findings suggest that the stronger labor protection narrows the gender gap in the labor market because better protection encourages females to participate in the labor market, whereas male workers do not benefit much from better protection. We find similar, but even stronger effects if we regress employment to population ratio on our explanatory variable set (see column 3 and 4 of Table 3.2). We use different labor protection indices and different time periods for robustness checks (see Table 3.4 and Table 3.5). All results are in line with the baseline results.

3.6 Robust Determinants of Gender Inequality: A Bayesian Model Averaging Approach

Our results confirm the impact of the wide range of variables that has been found to be determinants of gender inequality. However, it is still unclear to policymakers what are the most fundamental and robust determinants of gender inequality. Mis-specified econometric models lead to biased estimates, and classical statistical approaches offers little help with model uncertainty especially when the sample is small. Large panels, like the one we are using–covering a vast number of countries over the past three decades–alleviate the small sample issue. For policy recommendations, however, it is important to test the robustness of the determinants of gender inequality.

Hence, we use Bayesian Model Averaging (BMA) to address model uncertainty and examine the robustness of each potential determinant. BMA is a statistical technique which offers a way to think about model uncertainty (Leamer 1978; Raftery (1995); Sala-i Martin et al. 2004). The intuition behind it is that since it is difficult to know what the "true" model is, we can attach probabilities to different models. This requires departing from the

Table 3.2: Gender Gap in Labor Force Participation: OLS Regressions

Dependent var: gender gap in labor force participation (%, M-F)	(1)	(2)	(3)	(4)	(5)	(6)
Stages of development						
Log(GDP per capita)	57.474***	45.773***	45.123***	50.814***	46.610***	59.483***
	(5.070)	(4.013)	(5.513)	(6.888)	(5.614)	(5.872)
Log(GDP per capita) squared	-2.957***	-2.338***	-2.182***	-2.865***	-2.382***	-3.050***
	(0.282)	(0.215)	(0.304)	(0.457)	(0.317)	(0.344)
Health and Demographics						
Neonatal mortality rate	0.139*	0.055	-0.054	0.174**	0.131*	0.072
	(0.078)	(0.065)	(0.065)	(0.071)	(0.073)	(0.078)
Fertility rate	3.923***	4.149***	2.279***	3.231***	3.142***	4.620***
	(0.477)	(0.475)	(0.589)	(0.569)	(0.558)	(0.608)
Marriage age gap	2.306***	2.498***	1.036***	0.572*	0.526	1.165***
	(0.334)	(0.337)	(0.354)	(0.341)	(0.328)	(0.345)
Female avg. years of schooling	-0.758***	-0.541***	-1.658***	-0.779***	-2.438***	-3.926***
	(0.243)	(0.200)	(0.220)	(0.295)	(0.515)	(0.579)
Institutions						
Daughter inheritance rights		-5.734***	-9.528***	-8.446***	-8.848***	-4.702***
		(1.003)	(1.123)	(1.071)	(1.049)	(0.946)
Female as head of household		-8.926***	-9.846***	-10.338***	-9.681***	-7.238***
		(0.739)	(0.828)	(0.851)	(0.801)	(0.861)
Control of Corruption		-0.923**	-1.767***	-1.463***	-1.333***	-0.071
		(0.451)	(0.418)	(0.457)	(0.449)	(0.539)
Structural Policies						
Economic stability rating			-0.214***	-0.152***	-0.127**	-0.101
			(0.062)	(0.056)	(0.058)	(0.062)
Labor market protection Index			-8.107***	-9.173***	-9.336***	-15.908***
			(2.697)	(2.863)	(2.919)	(2.771)
Labor market protection index squared			1.471*	1.800**	1.770**	3.554***
			(0.818)	(0.868)	(0.887)	(0.831)
Infrastructure						
Telephone subscription rate				0.409***	0.693***	0.709***
				(0.089)	(0.054)	(0.147)
Telephone subscription rate x female education				-0.046***	-0.078***	-0.077***
				(0.009)	(0.009)	(0.014)
Access to improved sanitation facilities (%)				-0.375*	-0.043	-0.286***
				(0.203)	(0.070)	(0.069)
Access to improved sanitation facilities (%) x log GDP per capita				0.064***	0.028***	0.044***
				(0.023)	(0.008)	(0.008)
Financial Access						
Financial institutions access					1.268	-5.850
					(6.667)	(6.439)
Financial institutions access x education					0.196	0.851
					(0.676)	(0.657)
Fiscal Policies						
Public education expenditure						-1.257***
						(0.299)
Public education expenditure x Emerging Economy						0.935***
						(0.222)
Public health expenditure						0.292
						(0.190)
Observations	1842	1769	1387	1360	1360	1172
Adjusted R-squared	0.679	0.664	0.746	0.762	0.762	0.781

Robust standard errors in parentheses. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

Notes: This table includes results from the OLS regression above. The dependent variable is gender gap in labor force participation. Numbers in the parenthesis are standard errors. Standard errors are clustered at the province level. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

classical statistics framework and adopt Bayesian updating.

More specifically, BMA averages across a large set of models for a given set of priors. Each model receives a weight and the final estimates are constructed as a weighted average of the parameter estimates from each of the models. Bayesian information criterion (BIC) or Schwarz criterion (SBIC) is used to assign a weight to each model. The weights also depend on the choice of priors specified. The general rules for interpreting posterior probabilities is the following: posterior probabilities of <50%, 50% – 75% and >95% are usually interpreted as no evidence, weak evidence and strong evidence of an effect, respectively. Following this classification, we use BMA to select the most robust determinants of gender inequality, while the level of development and its square is set to be included in all regressions as a basic control.

The results (see Table 3.6) highlight a small set of variables that is robustly related with lower education gaps. In particular, improved sanitation facilities, higher public expenditure on education, and a lower age gap between men and women at marriage, are all robustly associated with lower educational gaps, with a posterior inclusion probability of 100 percent. A substantial number of country characteristics and polices are robustly related to gender gaps in labor force participation.

Demographics

There is strong evidence of an impact of higher fertility rates and wider age gaps at marriage between men and women and wider gender gaps in the labor market, with both variables showing a posterior inclusion probability of more than 95 percent. Equal minimum ages by law for men and women could therefore help decrease the gap.

Legal rights

There is strong evidence that legal rights are associated with narrower gaps in labor force participation, with equal inheritance rights for daughters and sons, women's right to be head of a household entering with a posterior inclusion probability of 100 percent, and in an economically significant way.

Infrastructure

Infrastructure is also strongly and robustly related to lower gaps in labor force participation. In particular, improved sanitation facilities decrease the gap in particular at early stages of development, while the impact moderates as GDP per capita increases. Telephone subscription rates, on the other hand, are related to narrower gaps in countries where female education is on average higher, likely because access to telephones helps carrying out jobs that require a certain level of training or education and ability to work from home. For instance, Ivanova et al. (2017) show that access to a cellphone for all women, and access to a computer for married women is significantly related to higher female labor force participation in Costa Rica.

Labor market protection

Stronger labor market protection is associated with lower labor force participation gaps, but the marginal impact of stronger protection declines as protection strengthens. At lower levels of labor market protection, an increase in the strength of protection is associated with narrower labor force participation gaps, as women seem to disproportionately benefit from stricter regulation. However, the marginal impact of protection decreases at higher levels of protection, when more excessive labor market rigidity weighs on the benefits (see also World Bank 2013).

3.7 Conclusions and Policy Implications

The novel contribution of this paper is the ability to identify the most important and robust determinants of gender inequality, thus providing a more useful guide to policy action. We complement the existing literature by explicitly addressing model uncertainty that is inherent in all empirical estimations where the proposed set of potential determinants is large. In fact, in addition to standard fixed-effects estimations for a large panel, our paper is the first to apply Bayesian Model Averaging–a methodology that is specifically designed to highlight factors that are robustly related to a variable of interest–to the literature on gender gaps in education and labor force participation. With that, our paper is able to highlight policy areas that are likely to yield the biggest bang for the buck.

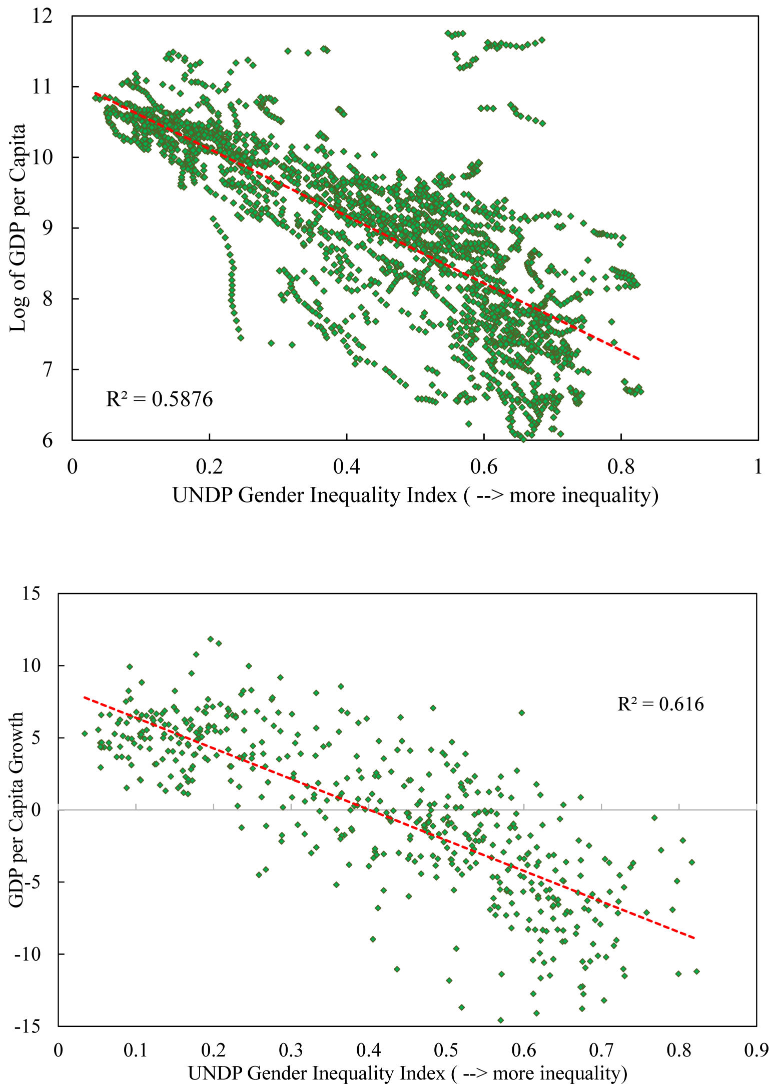
The paper finds that policies can play an important role in lowering gender inequality. Education gaps. Good sanitation facilities contribute to narrow gender education gaps by freeing up time from household activity and providing a safer environment at school. Low adolescent fertility and small marriage age gap prolong the time in which girls do not face the trade-off between family and education and are associated with lower gender gaps in education. The result that adequate sanitation facilities and narrower marriage age gaps reduce the gender gap in education is robust across many different model specifications. In addition, a stable and safe environment, as measured by indices of political stability and public safety, is also associated with a narrower education gap between males and females. Moreover, higher financial development contributes lower gender gaps in tertiary education, likely reflecting the higher cost of these services in many countries. Higher public spending on education and health is associated with narrower gender gaps in tertiary education.

Labor force participation gaps. Adequate infrastructure help reduce the gender gap in labor force participation, and improved sanitation facilities play the most important role in this respect. Stronger legal rights for females, low adolescent fertility, a small marriage age gap. Labor market protection has a non-linear impact of labor market participation gaps. At lower levels of labor market protection, stronger protection is associated with lower labor force participation gaps, but the marginal impact of protection decreases at higher levels of protection, with excessive labor market rigidity weighing on female labor force participation. All these factors are robustly related to gender gaps in labor force participation. Finally, higher public expenditure on education is associated with narrower labor force participation gaps in advanced economies, and there is some evidence that a stronger institutional environment–as measured by better control of corruption and lower economic riskboost female labor force participation

What types of economic policies can then boost gender equality? First of all, fiscal policies aiming at improving infrastructure, especially sanitation facilities in LICS. Moreover, social campaigns and policies to cut adolescent fertility and narrow the marriage age gap. Labor market policies can also support gender equality. However, our results suggest that strengthening job protection can help narrow the labor force participation gaps only in institutional settings where job protection is low. Instead, making the labor market too rigid could widen the labor force participation gap.

3.8 Appendix

Figure 3.7: Gender inequality vs. GDP per capita and GDP per capita growth



Notes: This figure shows that gender inequality is negatively associated with GDP per capita and GDP growth.

Data

We construct a large cross-country dataset of economic, political, institutional, and social variables for the period 1980-2014. We collect data from various sources, such as the World

Development Indicators, Women, Business and the Law, IMF Fiscal Affairs, IMF WEO, United Nations, International Labor Organization, Polity, International Country Risk Guide, Barro and Lee (2013), Botero et al. (2004), and Campos and Nugent (2012). Please refer to Table 3.3 below for summary statistics of all the policy variables that we focus on in this paper.

Table 3.3: Sample: A Panel Dataset for 100 Countries from 1980 to 2014

Variable Names	No. of Observations	Mean	Standard Deviation
Gender Gaps			
Labor Force Participation Gap (M - F)	4200	24.02	17.20
Managerial Positions Gap (M - F)	1366	43.54	21.50
Literacy Gap (M - F)	3106	9.73	10.09
Primary School Enrollment Gap (M - F)	5425	6.16	10.21
Secondary School Enrollment Gap (M - F)	5114	1.67	9.19
Tertiary School Enrollment Gap (M - F)	4570	-2.44	9.94
Years of Schooling Gap	4402	0.82	0.93
Insitutions			
Daughter Inheritance Rights	6369	0.75	0.43
Female Head of Household	6369	0.74	0.44
Guaranteed Equality by Law	3079	0.91	0.29
Political Risk Rating	4880	62.03	14.54
Government Effectiveness	3464	-0.05	0.98
Regulatory Quality	3466	-0.05	0.96
Rule of Law	3523	-0.07	0.98
Control of Corruption	3466	-0.05	1.00
Demographics			
Female Avg. Years of Total Schooling	4402	6.48	3.16
Contraceptive Prevalence	3580	45.16	23.45
Maternal Mortality Ratio	4248	245.17	342.23
Adolescent Fertility Rate	2633	63.88	37.32
Marriage Age Gap	4562	3.71	1.61
Infrastructure			
Telephone Subscription Rate	6347	15.09	17.44
Access to Electricity	3906	71.38	32.90
Access to Improved Water Source	4063	83.26	18.72
Access to Improved Sanitation Facilities	3966	67.52	31.50
Access to Finance			
Financial Institutions Access	5950	0.17	0.28
Fiscal Policies			
Public Education Expenditure (% of GDP)	3790	4.56	2.46
Public Health Expenditure (% of GDP)	3229	3.69	2.34
Personal Income Tax Rate	4886	38.82	15.49
Structural Policies			
Employment Protection Index	4633	1.56	0.51
Trade Openness	5793	82.20	53.09

Notes: This table shows the dataset we collect for this paper. We collect various country characteristics and policy variables for 120 countries over 4 decades.

Table 3.4: Regressions with BDLLS and Doing Business Index

Dependent variable: gender gap in labor force participation	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
WB Doing Business Index									
Can the workweek for a single worker extend to 50 hours per week	-5.558								
	(4.783)								
Are there restrictions on night work?		3.598							
		(2.784)							
Are there restrictions on "weekly holiday" work?			-2.918						
			(2.856)						
What is the maximum number of working days per week?				8.008***					
				(2.73)					
Is there a retraining or reassignment obligation before an employer can make a worker redundant?					-4.037				
					(3.121)				
Are there priority rules that apply to redundancy dismissals or lay-offs?						-3.870*			
						(2.299)			
notice period for redundancy dismissal after 20 years of continuous employment							-0.538***		
							(0.177)		
severance pay for redundancy dismissal after 20 years of continuous employment								0.045	
								(0.031)	
BDLLS									
Measures the protection of labor and employment laws									-13.918**
									(5.959)
Observations	121	121	121	120	120	120	120	120	72
Adjusted R-squared	0.437	0.435	0.433	0.458	0.439	0.441	0.464	0.441	0.55

Notes: This table includes regressions using components in the Doing Business Index as the independent variable. Numbers in the parenthesis are standard errors. Standard errors are clustered at the province level. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

Table 3.5: Labor Market Protection and Labor Force Participation by Men and Women

Dependent variable:	Female LFP rate	Male LFP rate	Female employment to pop ratio	Male employment to pop ratio
	(1)	(2)	(3)	(4)
Labor Market Protection Index	15.365***	-1.841	20.448***	4.484**
	(3.282)	(1.484)	(3.781)	(1.899)
Labor Market Protection Index (squared)	-2.961***	0.752*	-4.763***	-1.216**
	(0.976)	(0.430)	(1.143)	(0.572)
Control variables:				
Development stage	✓	✓	✓	✓
Health and demographics	✓	✓	✓	✓
Institutions	✓	✓	✓	✓
Structural policies	✓	✓	✓	✓
Infrastructure	✓	✓	✓	✓
Financial access	✓	✓	✓	✓
Fiscal policies	✓	✓	✓	✓
Region fixed effects	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓
No. of Observations	1126	1126	1126	1126
Adjusted R-squared	0.741	0.507	0.711	0.586

Notes: The dependent variable is either male LFP rate or female LFP rate, and the main explanatory variable is strenght of labor market protection. Numbers in the parenthesis are standard errors. Standard errors are clustered at the province level. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

Table 3.6: Gender Gap in LFP: A BMA Exercise

Gender gap in labor force participation (M-F)	Coefficient	Posterior Inclusion Prob
log GDP per capita	58.62	1.00
log GDP per capita squared	-3.36	1.00
Fertility Rate	4.25	1.00
Daughter Inheritance Rights	-4.37	1.00
Female as Head of Household	-9.13	1.00
Female Avg. Years of Total Schooling	-1.26	0.97
Labor Market Protection Index	-16.93	1.00
Labor Market Protection Index (squared)	3.47	0.96
Telephone Subscription Rate x Female Eduation	-0.04	1.00
Improved Sanitation Facilities	-0.68	0.99
Improved Sanitation Facilities x log GDP per capita	0.09	0.99
Control of Corruption	-1.41	0.74
Marriage Age Gap	0.89	0.96
Telephone Subscription Rate	0.31	0.88
Personal Income Tax Rate	0.03	0.48
Economic Risk Rating (higher means lower risks)	-0.10	0.56
Neonatal Mortality Rate	0.05	0.31
Public Education Expenditure	-0.07	0.17
Political Risk Rating (higher means lower risks)	0.02	0.23
Public Health Expenditure	0.00	0.03
Financial Institutions Access x female Eduation	0.04	0.05
Financial Institutions Access	-0.40	0.05
Number of countries = 96		
Number of observations = 1126		

Notes: this table shows the BMA exercise for gender gap in LFP. All models include GDP per capita, region and time fixed effects. A posterior probability ¿0.5 is considered strong evidence for an effect.

Table 3.7: Gender Gap in Years of Schooling: A BMA Exercise

Gender Gap in years of schooling (M-F)	Coefficient	Posterior Inclusion Prob
log GDP per capita	-0.504	1.00
log GDP per capita squared	0.038	1.00
Improved Sanitation Facilities	-0.017	1.00
Public Education Expenditure	-0.130	1.00
Marriage Age Gap	0.095	1.00
Adolescent Fertility Rate	0.001	0.36
Female Head of Household	0.003	0.05
Maternal Mortality Ratio	0.000	0.05
Improved Water Source	0.000	0.04
Financial Institutions Access	0.005	0.04
Political Risk Rating	0.000	0.03
Trade Openness	0.000	0.03
Daughter Inheritance Rights	0.001	0.03
Public Health Expenditure	0.000	0.03
Number of countries = 120		
Number of observations = 1468		

Notes: this table shows the BMA exercise for gender gap in education. All models include GDP per capita, region and time fixed effects. A posterior probability ¿0.5 is considered strong evidence for an effect.

Bibliography

Ackerberg, D. A., K. Caves, and G. Frazer, "Identification properties of recent production function estimators," *Econometrica*, 2015, *83(6)*, 2411–2451.

Adukia, Anjali, "Sanitation and Education," 2016.

Aguirre, D, Leila Hoteit, Christine Rupp, and Karim Sabbagh, "Empowering the third billion: Women and the world of work in 2012," *Booz and Company*, 2012.

Alder, S., "Chinese Roads in India: The Effect of Transport Infrastructure on Economic Development," in "2015 Meeting Papers" number 1447 Society for Economic Dynamics 2015.

Amiti, Mary and Jozef Konings, "Trade Liberalization, Intermediate Inputs, and Productivity: Evidence from Indonesia," *American Economic Review*, December 2007, *97* (5), 1611–1638.

Anderson, J. E. and E. Van Wincoop, "Gravity with gravitas: a solution to the border puzzle," *the american economic review*, 2003, *93* (1), 170–192.

Arkolakis, C., A. Costinot, and A. Rodríguez-Clare, "New trade models, same old gains?," *The American Economic Review*, 2012, *102* (1), 94–130.

Asturias, J., M. García-Santana, and R. Ramos Magdaleno, "Competition and the welfare gains from transportation infrastructure: Evidence from the Golden Quadrilateral of India," 2015.

Autor, D, David Dorn, and Gordon H Hanson, "The China syndrome: Local labor market effects of import competition in the United States," *The American Economic Review*, 2013, *103* (6), 2121–2168.

Banerjee, A., E. Duflo, and N. Qian, "On the road: Access to transportation infrastructure and economic growth in China," *National Bureau of Economic Research*, 2012, (No. w17897.).

Barro, Robert J. and Jong Wha Lee, "A new data set of educational attainment in the world, 19502010," *Journal of Development Economics*, 2013, *104*, 184 – 198.

Bas, Maria and Antoine Berthou, "Does input-trade liberalization affect firms foreign technology choice?," *The World Bank Economic Review*, 2013.

Baum-Snow, N., "Did highways cause suburbanization?," *The Quarterly Journal of Economics*, 2007, pp. 775–805.

_ , L. Brandt, V. Henderson, M. Turner, and Q. Zhang, "Transport Infrastructure, Urban Growth and Market Access in China," *ERSA conference papers. European Regional Science Association*, October 2015.

Becker, Gary S, "A Theory of the Allocation of Time," *The economic journal*, 1965, pp. 493–517.

Behrens, K., G. Duranton, and F. Robert-Nicoud, "Productive cities: Sorting, selection, and agglomeration," *Journal of Political Economy*, 2014, *122(3), 507-553.*

Bernard, Andrew B, Jonathan Eaton, J Bradford Jensen, and Samuel Kortum, "Plants and productivity in international trade," *The American Economic Review*, 2003, *93* (4), 1268–1290.

Bloom, David E, David Canning, Günther Fink, and Jocelyn E Finlay, "Fertility, female labor force participation, and the demographic dividend," *Journal of Economic Growth*, 2009, *14* (2), 79–101.

Botero, Juan C., Simeon Djankov, Rafael La Porta, Florencio Lopez de Silanes, and Andrei Shleifer, "The Regulation of Labor*," *The Quarterly Journal of Economics*, 2004, *119* (4), 1339.

Brandt, L., J. Van Biesebroeck, and Y. Zhang, "Creative accounting or creative destruction? Firm-level productivity growth in Chinese manufacturing," *Journal of Development Economics*, 2012, *97* (2), 339–351.

_ , _ , L. Wang, and Y. Zhang, "WTO accession and performance of Chinese manufacturing firms," *Center for Economic Policy Research*, 2016.

Burstein, Ariel and Javier Cravino, "Measured Aggregate Gains from International Trade," *American Economic Journal: Macroeconomics*, April 2015, *7* (2), 181–218.

_ and Jonathan Vogel, "International trade, technology, and the skill premium," *Manuscript, Columbia University and UCLA*, 2012.

Bustos, Paula, "Trade liberalization, exports, and technology upgrading: Evidence on the impact of MERCOSUR on Argentinian firms," *The American economic review*, 2011, *101* (1), 304–340.

_ , "Trade Liberalization, Exports, and Technology Upgrading: Evidence on the Impact of MERCOSUR on Argentinian Firms," *American Economic Review*, February 2011, *101* (1), 304–40.

Campos, Nauro F and Jeffrey B Nugent, "The dynamics of the regulation of labor in developing and developed countries since 1960," 2012.

Caselli, F and W J Coleman, "Cross-Country Technology Diffusion: The Case of Computers," *American Economic Review*, May 2001, *91* (2), 328–335.

Caselli, Francesco and Daniel J Wilson, "Importing technology," *Journal of Monetary Economics*, 2004, *51* (1), 1 – 32.

Chaney, T., "Distorted gravity: the intensive and extensive margins of international trade," *The American Economic Review*, 2008, *98.4*, 1707–1721.

Coe, D T. and E Helpman, "International r&d spillovers," *European economic review*, 1995, *39* (5), 859–887.

Comin, D and M Mestieri, "Chapter 2 - Technology Diffusion: Measurement, Causes, and Consequences," in P Aghion and S N. Durlauf, eds., *Handbook of Economic Growth*, Vol. 2 of *Handbook of Economic Growth*, Elsevier, 2014, pp. 565 – 622.

Costinot, A. and A. Rodriguez-Clare, "Trade Theory with Numbers: Quantifying the Consequences of Globalization," *The Handbook of International Economics*, 2014, pp. 197–261.

Cuberes, David and Marc Teignier, "Aggregate effects of gender gaps in the labor market: A quantitative estimate," *Journal of Human Capital*, 2016, *10* (1), 1–32.

Das, Sanghamitra, Mark J. Roberts, and James R. Tybout, "Market Entry Costs, Producer Heterogeneity, and Export Dynamics," *Econometrica*, 2007, *75* (3), 837–873.

Das, Sonali, Sonali Jain-Chandra, Kalpana Kochhar, and Naresh Kumar, "Women Workers in India: Why So Few Among So Many?," 2015, p. 55.

Donaldson, D., "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *forthcoming, American Economic Review*, 2016.

Donaldson, Dave and Richard Hornbeck, "Railroads and American Economic Growth: A "Market Access" Approach," *The Quarterly Journal of Economics*, 2016, p. qjw002.

Duflo, Esther, "Women Empowerment and Economic Development," *Journal of Economic Literature*, December 2012, *50* (4), 1051–79.

Duranton, G. and M. A. Turner, "The fundamental law of road congestion: Evidence from US cities," *The American Economic Review*, 2011, *101* (6), 2616–2652.

_ , P. M. Morrow, and A. Turner, "Roads and Trade: Evidence from the US," *The Review of Economic Studies*, 2014, *81* (2), 681–724.

Eaton, J. and S. Kortum, "Technology, geography, and trade," *Econometrica*, 2002, *70* (5), 1741–1779.

Eckstein, Zvi and Kenneth I Wolpin, "Dynamic labour force participation of married women and endogenous work experience," *The Review of Economic Studies*, 1989, *56* (3), 375–390.

— **and Osnat Lifshitz**, "Dynamic Female Labor Supply," *Econometrica*, 2011, *79* (6), 1675–1726.

Elborgh-Woytek, Ms Katrin, Ms Monique Newiak, Ms Kalpana Kochhar, Ms Stefania Fabrizio, Kangni Kpodar, Mr Philippe Wingender, Mr Benedict J Clements, and Mr Gerd Schwartz, *Women, work, and the economy: Macroeconomic gains from gender equity*, International Monetary Fund, 2013.

Faber, B., "Trade integration, market size, and industrialization: evidence from China's National Trunk Highway System," *The Review of Economic Studies*, 2014, *81(3)* (pp.1046-1070).

Fan, Haichao, Yao Amber Li, and Stephen R. Yeaple, "Trade Liberalization, Quality, and Export Prices," *The Review of Economics and Statistics*, 2015, *97* (5), 1033–1051.

Feenstra, R. C., Z. Li, and M. Yu, "Exports and credit constraints under incomplete information: Theory and evidence from China," *Review of Economics and Statistics*, 2014, *96* (4), 729–744.

Fernández, Raquel and Joyce Cheng Wong, "Divorce Risk, Wages and Working Wives: A Quantitative Life-Cycle Analysis of Female Labour Force Participation," *The Economic Journal*, 2014, *124* (576), 319–358.

Fernndez, Carmen, Eduardo Ley, and Mark F. J. Steel, "Model uncertainty in cross-country growth regressions," *Journal of Applied Econometrics*, 2001, *16* (5), 563–576.

Field, Erica and Attila Ambrus, "Early marriage, age of menarche, and female schooling attainment in Bangladesh," *Journal of political Economy*, 2008, *116* (5), 881–930.

Fieler, Ana Cecília, Marcela Eslava, and Daniel Xu, "Trade, Skills, and Quality Upgrading: A Theory with Evidence from Colombia," Technical Report, National Bureau of Economic Research 2014.

Frye, D., "Transportation Networks and the Geographic Concentration of Industry," 2014.

Garcia, A. and N. Voigtländer, "Exporting and Plant-Level Efficiency Gains: It's in the Measure," Technical Report, National Bureau of Economic Research 2013.

Gaubert, C., "Firm sorting and agglomeration," *Unpublished manuscript, UC Berkeley*, 2014.

Ghani, E., A. G. Goswami, and W. R. Kerr, "Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing," *Econ Journal*, 2016, *126*, 317–357 doi:10.1111/ecoj.12207.

Ghani, Ejaz, William R. Kerr, and Stephen D. OConnell, "Local industrial structures and female entrepreneurship in India," *Journal of Economic Geography*, 2013, *13* (6), 929.

Goldberg, Pinelopi Koujianou, Amit Kumar Khandelwal, Nina Pavcnik, and Petia Topalova, "Imported intermediate inputs and domestic product growth: Evidence from India," *The Quarterly Journal of Economics*, 2010, *125* (4), 1727–1767.

Gong, Xiaodong, Robert Breunig, and Anthony King, "How Responsive is Female Labour Supply to Child Care Costs: New Australian Estimates," Technical Report, Institute for the Study of Labor (IZA) 2010.

Gonzales, Mr Christian, Sonali Jain-Chandra, Ms Kalpana Kochhar, and Ms Monique Newiak, *Fair Play:: More Equal Laws Boost Female Labor Force Participation*, International Monetary Fund, 2015.

Halpern, Lszl, Mikls Koren, and Adam Szeidl, "Imported Inputs and Productivity," *American Economic Review*, December 2015, *105* (12), 3660–3703.

Haltiwanger, J. C., "Measuring and analyzing aggregate fluctuations: the importance of building from microeconomic evidence," *St. Louis Fed Review*, 1997, *79.*

Hannan, Carolyn and Ingvar Andersson, "Gender perspectives on ecological sanitation," 2002.

Hanson, G. H., "Market potential, increasing returns and geographic concentration," *Journal of international economics*, 2005, *67(1)*, 1–24.

_ **and Chong X.**, "The Home-Market Effect and Bilateral Trade Patterns," *American Economic Review,*, 2004, *94(4)*, 1108–1129.

Harris, C. D., "The, Market as a Factor in the Localization of Industry in the United States," *Annals of the association of American geographers*, 1954, *44* (4), 315–348.

Heckman, James and Thomas E Macurdy, "A Life Cycle Model of Female Labour Supply," *Review of Economic Studies*, 1980, *47* (1), 47–74.

i Martin, Xavier Sala, Gernot Doppelhofer, and Ronald I. Miller, "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach," *American Economic Review*, September 2004, *94* (4), 813–835.

IMF, "Fiscal Policy and Employment in Advanced and Emerging Economies," *International monetary fund, fiscal affair department*, 2012.

Ivanova, Anna, Ryo Makioka, and Joyce Wong, "Costa Rica," in Kalpana Kochhar, Monique Newiak, and Sonali Jain-Chandra, eds., *Women, Work, and Economic Growth : Leveling the Playing Field*, International Monetary Fund, 2017.

Kalb, Guyonne, "Children, Labour Supply and Child Care: Challenges for Empirical Analysis," *Australian Economic Review*, 2009, *42* (3), 276–299.

Kasahara, Hiroyuki and Joel Rodrigue, "Does the use of imported intermediates increase productivity? Plant-level evidence," *Journal of Development Economics*, 2008, *87* (1), 106 – 118.

Kochhar, Kalpana, Monique Newiak, and Sonali Jain-Chandra, "Women, Work, and Economic Growth : Leveling the Playing Field," 2017.

König, M, D Rohner, M Thoenig, and F Zilibotti, "Networks in conflict: Theory and evidence from the great war of africa," *University of Zurich, UBS International Center of Economics in Society, Working Paper*, 2015, (14).

Koren, Miklós and Márton Csillag, *Machines and machinists: Capital-skill complementarity from an international trade perspective* number MT-DP-2011/14, IEHAS Discussion Papers, 2011.

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

_ , "Increasing Returns and Economic Geography," *The Journal of Political Economy*, 1991, *99(3)*, 483–499.

Krusell, Per, Lee E. Ohanian, Jos-Vctor Ros-Rull, and Giovanni L. Violante, "Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis," *Econometrica*, 2000, *68* (5), 1029–1053.

Laat, Joost De and Almudena Sevilla-Sanz, "The fertility and women's labor force participation puzzle in OECD countries: the role of men's home production," *Feminist Economics*, 2011, *17* (2), 87–119.

Lamb, Jenny, "The Role of Water in Addressing Gender Inequality," in "Water for Women," Water Aid, 2015.

Leamer, Edward E, *Specification searches: Ad hoc inference with nonexperimental data*, Vol. 53, John Wiley & Sons Incorporated, 1978.

Levinsohn, J. and A. Petrin, "Estimating production functions using inputs to control for unobservables," *The Review of Economic Studies*, 2003, *70(2)*, 317–341.

Lileeva, A. and D. Trefler, "Improved Access to Foreign Markets Raises Plant-level Productivity…For Some Plants," *The Quarterly Journal of Economics*, 2010, *125 (3)*, 1051–1099.

Limao, N. and A. J. Venables, "Infrastructure, geographical disadvantage, transport costs, and trade," *The World Bank Economic Review*, 2001, *15* (3), 451–479.

Loecker, J. De and F. Warzynski, "Markups and firm-level export status," *The American Economic Review.*, 2012, *102(6), 2437-2471.*

_ , K. Goldberg Pinelopi, Amit K. Khandelwal, and Nina Pavcnik, "Prices, Markups, and Trade Reform," *Econometrica*, 2016, *84* (no. 2), 445–510.

Long, J. B De and L H. Summers, "Equipment Investment and Economic Growth," *Quarterly Journal of Economics*, 1991, *106* (2), 445–502.

Manova, K., "Credit constraints, heterogeneous firms, and international trade," *The Review of Economic Studies*, 2013, *80* (2), 711–744.

Masanjala, Winford H. and Chris Papageorgiou, "Rough and lonely road to prosperity: a reexamination of the sources of growth in Africa using Bayesian model averaging," *Journal of Applied Econometrics*, 2008, *23* (5), 671–682.

Melitz, M. J. and D. Trefler, "Gains from trade when firms matter," *The Journal of Economic Perspectives*, 2012, *26(2)*, 91–118.

Melitz, M.J., "The impact of trade on intraindustry reallocations and aggregate industry productivity," *Econometrica*, 2003, *71(6)*, pp.1695–1725.

_ **and G.I. Ottaviano**, "Market size, trade, and productivity," *The review of economic studies*, 2008, *75(1)*, 295–316.

Michaels, G., "The effect of trade on the demand for skill: Evidence from the interstate highway system," *The Review of Economics and Statistics*, 2008, *90* (4), 683–701.

Mishra, Vinod and Russell Smyth, "Female labor force participation and total fertility rates in the OECD: New evidence from panel cointegration and Granger causality testing," *Journal of Economics and Business*, 2010, *62* (1), 48–64.

Nagy, D., "City location and economic development," *mimeo*, 2015.

Norando, German Cubas, "Essays on infrastructure, female labor force participation and economic development," 2010.

Olley, S. and A. Pakes, "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 1996, *64 (6)*, 1263–1298.

Pavcnik, N., "Trade liberalization, exit, and productivity improvements: Evidence from Chilean plants," *The Review of Economic Studies*, 2002, *69(1)* (245-276).

R, C. Feenstra, Z. Li, and M. Yu, "Exports and credit constraints under incomplete information: Theory and evidence from China," *Review of Economics and Statistics*, 2014, *96.4*, 729–744.

Raftery, Adrian E, "Bayesian model selection in social research," *Sociological methodology*, 1995, pp. 111–163.

Rajan, R.G. and Zingales L., "Financial Dependence and Growth," *The American Economic Review 88 (3) American Economic Association: 559–86*, 1998.

Redding, S., "Goods trade, factor mobility and welfare," *Journal of International Economics*, 2016, *101*, 148–167.

_ **and A.J. Venables**, "Economic geography and international inequality," *Journal of international Economics*, 2004, *62(1)*, 53–82.

Revenga, Ana and Sudhir Shetty, "Empowering Women Is Smart Economics-Closing gender gaps benefits countries as a whole, not just women and girls," *Finance and Development-English Edition*, 2012, *49* (1), 40.

Rosenbaum, Paul R and Donald B Rubin, "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 1983, pp. 41–55.

Sen, Amartya, "The many faces of gender inequality," 2001.

Smeets, V. and F. Warzynski, "Estimating productivity with multi-product firms, pricing heterogeneity and the role of international trade," *Journal of International Economics*, 2013, *90(2)*, 237–244.

Steinberg, Chad and Mr Masato Nakane, *Can women save Japan?* number 12-248, International Monetary Fund, 2012.

Stimpfle, Alexander and David Stadelmann, "Marriage Age Affects Educational Gender Inequality: International Evidence," 2016.

Tang, H., "Labor market institutions, firm-specific skills, and trade patterns," *Journal of International Economics*, 2012, *87* (2), 337–351.

Thévenon, Olivier, "Drivers of female labour force participation in the OECD," 2013.

Topalova, Petia and Amit Khandelwal, "Trade Liberalization and Firm Productivity: The Case of India," *The Review of Economics and Statistics*, 2011, *93* (3), 995–1009.

Verhoogen, Eric A., "Trade, Quality Upgrading, and Wage Inequality in the Mexican Manufacturing Sector*," *The Quarterly Journal of Economics*, 2008, *123* (2), 489.

Voigtlaender, Nico, Diego Saravia et al., "Imported Inputs, Quality Complementarity, and Skill Demand," in "2015 Meeting Papers" number 1538 Society for Economic Dynamics 2015.

Wooldridge, J. M., "On estimating firm-level production functions using proxy variables to control for unobservables," *Economics Letters*, 2009, *104(3)*, 112–114.
World Development Report 2009: Reshaping Economic Geography

***World Development Report 2009: Reshaping Economic Geography**, The World Bank, Washington DC, 2009.*

Yeaple, Stephen Ross, *"A simple model of firm heterogeneity, international trade, and wages," Journal of International Economics, 2005, 65 (1), 1 – 20.*