5-2020

Three essays on quality of tradable products

Angdi LU
Singapore Management University, angdi.lu.2015@phdecons.smu.edu.sg

THREE ESSAYS ON QUALITY OF TRADABLE PRODUCTS

ANGDI LU

A DISSERTATION

IN

ECONOMICS

Submitted to the Singapore Management University in Partial Fulfilment

of the Requirements for the Degree of PhD in Economics

2020

Supervisor of Dissertation

PhD in Economics, Programme Director

Three Essays On Quality of Tradable Products

by Angdi Lu

(B.A., Fudan University)

Submitted to School of Economics in partial fulfillment of

the requirements for the Degree of Doctor of Philosophy in Economics

Dissertation Committee:

Pao-Li Chang (Supervisor/Chair)
Associate Professor of Economics
Lee Kong Chian Fellow
Singapore Management University

Amanda Jakobsson
Assistant Professor of Economics
Singapore Management University

Yuan Mei
Assistant Professor of Economics
Singapore Management University

Lin Tian
Assistant Professor of Economics
INSEAD

Singapore Management University

2020

Declaration

I hereby declare that this PhD dissertation is my original work and it has been written by me in its entirety.
I have duly acknowledged all the sources of information which have been used in this dissertation.

This PhD dissertation has also not been submitted for any degree in any university previously.

Angdi Lu

Angdi Lu

15 June 2020

Abstract

This dissertation includes three essays on the quality of tradable products. The first chapter studies the supply-side determinants of quality specialization across Chinese cities. Specifically, we complement the quality specialization literature in international trade and study how larger cities within a country produce goods with higher quality. In our general equilibrium model, firms in larger cities specialize in higher-quality products because agglomeration benefits (arising from the treatment effect of agglomeration and firm sorting) accrue more to skilled workers, who are also more efficient in upgrading quality, although these effects are partially mitigated by higher skill premium in larger cities. Using firm-level data from China, we structurally estimate the model and find that agglomeration and firm sorting each accounts for about 50% of the spatial variation in the quality specialization. A counterfactual policy to relax land use regulation in housing production raises product quality in big cities by 5.5% and indirect welfare of residents by 6.2%. The second chapter examines how information frictions matter in the endogenous choice of product quality made by firms. We introduce quality choice into a trade-search model with information frictions Allen (2014). In our model, producers must search to learn about the quality-augmented price index elsewhere and decide whether to enter a specific destination market. Hence, a fall in information frictions such as the building of information and communications technologies infrastructure (i.e., faster mobile networks) will induce quality upgrading. We empirically test the predictions of our model using unit value data and variation in information and communications technologies infrastructure across Chinese cities. The third chapter provides empirical evidence on the effects of falling trade costs on product quality across cities within a country. We approach this question in the context of expanding the highway system in China in the past decades, which substantially reduces the trade costs across regions within the nation. Empirically, we combine two firm-level panels that provide unit-value information of products across Chinese cities with city-level data on transportation infrastructure for 2001-2007. We find that firms choose to upgrade product quality more in cities with a greater expansion of connecting highways. In addition, this effect is more pronounced in larger cities, which speaks to changes in the spatial concentration of higher-quality products. These results

are also robust to the inclusion of an exhaustive battery of fixed effects and to changes in estimation specifications. Our findings shed important insights on the impact of falling intranational trade cost on quality specialization pattern across cities, which is difficult to model quantitatively due to the presence of agglomeration and sorting.

Contents

Acknowledgment

I would like to express my deepest gratitude to my supervisor, Prof. Pao-Li Chang, for her invaluable guidance and patience over my PhD years. This dissertation would not have been possible without her help and encouragement. In addition, I sincerely thank my dissertation committee, Prof. Amanda Jakobsson, Prof. Yuan Mei, and Prof. Lin Tian, who provided great advice to my research. I also would like to thank Prof. Davin Chor, Prof. Bingjing Li, and Prof. Lin Ma. I benefited a lot from their insightful comments and industrious scholarship in the reading group at NUS.

Special thanks to my beloved family for their selfless love and unconditional support over the past years. It is their love that makes it possible for me to complete what I started. This thesis is dedicated to them.

List of Tables

List of Figures

1 Quantifying Quality Specialization Across Space: Skills, Sorting, and Agglomeration

1.1 Introduction

Firms in big cities specialize in high-quality products (Dingel, 2017; Saito and Matuura, 2016). One explanation formalizes the insight of "Linder hypothesis" to rationalize this empirical regularity. It builds on the so-called "home-market effect" and hypothesizes that local demand in big cities is biased towards high-quality goods because demand for quality rises with income (Dingel, 2017; Picard and Okubo, 2012; Picard, 2015). Another explanation complements the demand-based conjecture and focuses on the productivity advantage of firms in big cities (Saito and Matuura, 2016). Firms become more productive in a big city, and this creates more room for costly quality upgrading. These hypotheses have provided important insights. However, none of them allows free mobility and touches on sorting behavior which are critical in the spatial context, since individuals and firms are freely mobile within a country and are free to choose their location.[1] In this sense, a supply-side explanation of the spatial pattern of quality specialization is underdeveloped, because the movement of factors and firms are what distinguish spatial models from international trade models.

Moreover, performing counterfactual experiments in a fully specified general equilibrium model is lacking in the existing literature on the spatial pattern of quality specialization which either only develops theoretical models or presents reduced-form evidence. This is important because the pattern of quality specialization provides an additional channel of gains from inter-city trade and also gains from agglomeration. Hence, it is desirable to develop quantitative models that are capable of quantifying the welfare effects of spatial policies through the channel of quality specialization. Our paper partly fills this gap.

In this paper, we provide a supply-side explanation for the quality specialization pattern across cities. The main feature of our approach is that more productive firms en-

[1] One exception is the line of work done in Picard and Okubo (2012) and Picard (2015). However, the sorting behavior in their models is related to demand-based factors instead of the productivity advantage provided by agglomeration. Furthermore, the individuals in their model are immobile across regions.

dogenously sort into larger cities because they receive more benefits from agglomeration. As a consequence, firms in big cities specialize in high-quality products because of two reasons. First, agglomeration benefits are such that their productivity is higher in larger cities. Second, firms that sorted into larger cities are also more productive firms. Quantifying the extent to which how much each channel has influenced quality specialization pattern is the main contribution that our paper aims to deliver. To our knowledge, our paper is the first to investigate such supply-side explanations in a general equilibrium quantitative model.

We develop a general equilibrium model with endogenous quality choice, endogenous spatial sorting of firms, and endogenous city formation. More productive firms upgrade the quality of their products because the marginal cost of production is lower and leaves more room for choosing high quality. This is reminiscent of the quality upgrading literature in international trade that focuses on heterogeneous firms (Feenstra and Romalis, 2014; Antoniades, 2015; Fan, Li, Xu and Yeaple, 2017). Different from these literature that assume labor as the only factor in the production, we employ a flexible production function that uses capital, unskilled labor, and skilled labor as inputs which is partly similar to the production function in Fieler, Eslava and Xu (2018). The production structure implies that skill intensity increases with quality choices. This assumption makes the identification of the quality-upgrading parameter easier and more transparent. As a consequence, there is no need to rely on any unit-price information in identifying the quality upgrading parameter which could be potentially biased. Though we do have access to both quantity and price information from the custom data, we only use this information to perform out-of-sample test to examine the empirical fit of our model.

Modeling endogenous spatial sorting in a quantitative framework is not a trivial task and can be computationally daunting. To deal with this issue, we import the spatial sorting framework developed in Gaubert (2018) to aid our investigation. We posit that firm productivity is a composite term of its innate efficiency and the size of the city it locates in. Firms are heterogeneous in their inherent efficiency. City size boosts firm productivity through two channels. The first channel is the standard agglomeration benefit, while the second is a log-supermodular term such that firms with a higher innate efficiency receive more benefit from agglomeration. The computational advantage of this framework is that city size is a sufficient statistic for the production and sorting decisions of firms. We generalize Gaubert's insight into an environment with two types of labor and quality choices. To offer a clear demonstration of how city size alone is a sufficient statistic, we first develop the benchmark model in an environment with costless trade. This also has the advantage that only supply-side factors are in play when we quantify the distribution of quality across space.

Apart from sorting, we also model the endogenous formation of cities which is a byproduct of factor demand from firms, in the sense that factor markets must be cleared locally. In our model, producing high-quality products requires hiring more skilled workers. The quantitative implication of this feature is entirely different from the existing literature such as Dingel (2017). In Dingel's paper, which quantifies the relative importance of the home-market effect and the factor abundance on the choice of quality, factor abundance is exogenously given for each core-based statistical areas (CBSA) area. In contrast, our model assumes a spatial no-arbitrage condition such that each individual must derive the same utility regardless of his location. Together with the local labor market clearing conditions, this will pin down the endogenous factor supply in each city. In this sense, our supply-side story is entirely different from that of Dingel's and is more general.

We structurally estimate our model using firm-level data from China Annual Survey of Industrial Firms. In particular, we calibrate part of the parameters using prior estimates from the literature, as these parameters are standard and have been well-studied in the past. For all other parameters that are relevant to quality upgrading and firm sorting, we structurally estimate them using an Simulated-Method-of-Moment (SMM) estimator. The intuition is to search over parameter space to minimize the weighted distance between model-generated moments that are directly governed by those parameters and the corresponding empirical moments. We find that product quality is on average 23% higher in big cities than that of small cities. There is also substantial sectoral heterogeneity in the quality specialization pattern and the quality difference could be as high as 60% in some sectors. In addition, we decompose the channels and find that quantitatively firm sorting account for half of the quality specialization pattern across cities while traditional agglomeration forces account for another half.

Finally, we quantify the general equilibrium impact of a supply-side spatial policy, which is frequently used in developing economies such as China, using the estimated model. This counterfactual examines policies that restrict land use in the production of housing. This policy directly affects the distribution of wages across space as housing is the congestion force in the model. Consequently, agglomeration is weakened due to the congested land market and firms produce goods with lower quality. We find an indirect welfare benefit of 6.2% in a counterfactual where we relax land use regulations by 20%. Furthermore, average quality across cities decreases by 5.5%. In sum, these counterfactuals are highly relevant to developing economics such as China. The policy implications and quantifying the welfare effect of these spatial policies through the lens of quality specialization are significant and non-trivial.

The rest of the paper is organized as follows. Section 1.2 provides a brief literature review and discusses the contributions of our paper. Section 1.3 qualitatively examines the empirical evidences. Section 1.4 describes the model and discuss its implications. Section 1.5 takes the model to the data and performs quantitative analysis. Section 1.6 conducts couterfactual experiment. Lastly, Section 1.7 concludes.

1.2 Related Literature

The present study is related to several strands of literature in urban economics and international trade. First, our work is related to the spatial literature on the benefits of agglomeration (Davis and Dingel, 2019; Gaubert, 2018; Tian, 2018; Handbury and Weinstein, 2015; Behrens, Duranton and Robert-Nicoud, 2014; Combes, Duranton, Gobillon, Puga and Roux, 2012; Albouy, 2012; Duranton and Puga, 2004; Rosenthal and Strange, 2004; Glaeser, Kolko and Saiz, 2001; Glaeser, 1999; Glaeser, Kallal, Scheinkman and Shleifer, 1992). Our paper complements this literature by studying an additional margin of gains from agglomeration, that is the productivity advantage of big cities also enable firms to upgrade their product quality. As metioned earlier, our work is not the first in the literature to study such effect. Under a reduced-form partial equilibirum framework, Saito and Matuura (2016) show that firms upgrade product quality in a larger city using the universe of Japanese firm-level data. In comparison to their paper, our work is the first attempt that structurally estimates a quantitative spatial equilibrium model focusing on quality. Our model is able to quantify the general equilibrium effect, perform welfare analysis, and study relevant counterfactuals. Our equilibrium model is also tractable and explicitly models firm sorting which can be a concern of endogeneity in empirical studies. In particular, we quantify the exact degree how each channel affects quality specialization pattern across space.

Our paper is also relevant to a literature in urban economics that focuses on explaining skill premium and skill compositions across cities (Davis and Dingel, 2019, 2017; Glaeser and Maré, 2001; Baum-Snow and Pavan, 2012, 2013; Baum-Snow, Freedman and Pavan, 2018; Moretti, 2013; Diamond, 2016; De La Roca and Puga, 2017; Combes, Duranton and Gobillon, 2008; Dingel, Miscio and Davis, 2019; Davis, Mengus and Michalski, 2018; Behrens and Robert-Nicoud, 2015; Farrokhi and Jinkins, 2019; Lindley and Machin, 2016; Hendricks, 2011; Bacolod, Blum and Strange, 2009; Chor, 2005; D'Costa and Overman, 2014; Florida, Mellander, Stolarick and Ross, 2012; Ma and Tang, 2018; Jiao and Tian, 2019; Ciccone and Hall, 1996). The consensus of the literature was that a spatial equilibrium model that imposes a no-arbitrage or free-mobility condition, which requires all individuals to receive same utility across cities, would only imply a constant skill premium in city size (Black, Kolesnikova and Taylor, 2009). A recent literature pioneered by Davis and Dingel (2019) provide evidences that skill premium are in fact rising in city size and they reconcile the puzzle using an inframarginal learning effect under the assumption that there is a continuum of workers heterogeneous in their ability. Our work complements this literature. In particular, we show that even with two skill types of workers, our model is able to generate rising skill premium across cities. Two elements are essential. First, we assume that there are two separate residential housing markets in each city and we microfound this assumption using a within-city sorting model with non-homothetic preference. Second, given that there are two housing markets, rising skill premium is then a consequence of increasing skill composition, which is in turn a result of skill-biased agglomeration benefits and incentive to hire more skilled workers for quality upgrading. In sum, our model argue that skill premium are higher in larger cities partly because there are more skilled workers in big cities for quality upgrading purposes. Congestion forces in the two housing markets then ensure that skill premium rises in city size.

Furthermore, our work is related to the literature in international trade that studies the quality specialization across countries which focuses on both the demand side (Piveteau and Smagghue, 2019; Dingel, 2017; Fajgelbaum, Grossman and Helpman, 2011, 2015; Hallak, 2006, 2010; Choi, Hummels and Xiang, 2009) and the supply side explanations (Fieler, Eslava and Xu, 2018; Dingel, 2017; Faber and Fally, 2017; Fan, Li, Xu and Yeaple, 2017; Antoniades, 2015; Feenstra and Romalis, 2014; Hallak and Sivadasan, 2013; Kugler and Verhoogen, 2012; Crozet, Head and Mayer, 2012; Khandelwal, 2010; Verhoogen, 2008; Schott, 2004; Hummels and Skiba, 2004). Our work is related to this literature in the sense that we complement the supply-side understanding of quality specialization pattern in a narrower definition of space, that is we narrow the definition of space from across countries to within a country and study the quality specialization pattern across cities. Similar to the international trade literature, we focus on the idea that higher productivity of heterogeneous firms enable costly quality upgrading. In addition, we also focus on the effect of firm sorting and scale effect (agglomeration) on quality specialization across space which is absent in the trade literature. We hope that our work can shed some light on how sorting and scaling effects of multinational firms and foreign direct investment affect the choice of quality across countries.

1.3 Stylized Facts

In this section, we present some stylized facts on quality specialization across Chinese cities and how it correlates with firm heterogeneity and agglomeration. We first document that firms produce higher-quality goods in big cities, after controlling for comparative advantage, product-specific time shocks, city-specific time trends, and other city-time specific characteristics. Next, we show that more productive firms would specialize in higher-quality products. Together, these two sets of facts lay down the basic elements of our model and pave the way for our structural estimation that disentangles the endogenous economic forces in equilibrium. Lastly, we show that producing high-quality goods is strongly correlated with employing more skilled labor. This fact will be useful in the design of identification strategy in our empirical structural estimation.

1.3.1 Data and Measurement

We merge two datasets that contain firm-level information on sales and output separately. The first dataset is the Annual Survey of the Industrial Firms (ASIF). This dataset contains information on various firm-level characteristics such as sales, profits, taxes, investment, intermediate input expenditure, labor expenditure, and education level of workers. The second dataset is the Industrial Firms Product Quantity Dataset.[2] This dataset contains information on the physical quantity of firm output, and it has been used in other literature to measure product quality Fan, Gao, Li and Luong (2018). These two datasets both cover the universe of Chinese manufacturing firms and use the same firm identification.[3] [4] While we use both datasets to construct the stylized facts, only the first dataset is used in the structural estimation. We only exploit the information in the second dataset to evaluate the out-of-sample performance of our structural model.

We measure product quality following two approaches in the international trade literature. The first approach exploits information on the unit price of products, which are readily available in trade data, to measure the quality of goods (Schott, 2004; Hummels and Klenow, 2005; Hallak, 2006). The intuition is that a higher-quality good commands a higher price, hence unit values are reasonable proxies for product quality, all else equal. In contrast, the second approach focuses on the market share of a product and measures product quality using a nested logit demand system (Khandelwal, 2010; Amiti and Khandelwal, 2013). The idea is that unit values may fail to reflect quality differences, as there may be other confounding factors such as production costs that are driving the price differences. Market shares in turn capture the vertical component of quality differences, in

[2]We are extremely grateful to Yao Amber Li for her suggestions of this dataset.

[3]We confirm that this is true as both datasets also report the firm names, addresses, and names of corporate representatives.

[4]Note, however, that our sample only covers the single-product firms. The reason is that the ASIF dataset only reports the sales of the entire firm while the physical quantity dataset reports the quantity information of each five-digit product that the firm produces. Since we need to construct unit price at the product level, we only include single-product firms in computing prices.

the sense that a higher-quality good would have a greater market share conditional on the same price. We follow both approaches to construct measures for quality and use them in our empirical specifications. More details on our measurement of product quality can be found in Appendix A.1.

To provide a measure of firm productivity in the second stylized fact, we implement production function estimation using the canonical methods in the empirical industrial organization literature. In particular, we first employ the semiparamteric control function approach in Olley and Pakes (1996) to obtain a baseline measure of firm productivity. Next, we check for the robustness of our estimates using Levinsohn and Petrin (2003) and Ackerberg, Caves and Frazer (2015) which address the zero investment and control function collinearity problems.[5]

Note, however, the above measure of firm productivity does not correspond to the "innate efficiency" that we define in the structural model. The reasons are as follows. There are abundant evidences suggesting that firms become more productive in larger cities and hence a positive "treatment effect" of agglomeration on firm productivity. As such, the estimates we obtain are *ex post* measures of productivity after the treatment of agglomeration, and they are different from the "innate efficiency" of firms . Including this measure in our empirical specification would have subsumed all the interactions between firm heterogeneity and agglomeration. Moreover, there is no easy way to filter out such treatment effect of agglomeration. Consider a regression of the productivity measures on city size. Ideally, this regression would have filtered out all the explanatory power city size has on productivity. However, this regression also filters out the effect of firm sorting, in the sense that firms that are more innately efficient also endogenously choose to locate in big cities (Gaubert, 2018; Tian, 2018). Hence, we will focus on a subsample of "moving firms", which choose to relocate in another city, to establish our stylized facts on how firm heterogeneity and agglomeration are related to quality specialization across Chinese cities.

[5]We implement all production function estimations using a Stata module `prodest` (Rovigatti and Mollisi, 2018).

1.3.2 Empirical Evidence

Stylized Fact 1: Firms produce higher-quality products in big cities, and this pattern is robust to adding an extensive set of controls.

Econometric Design

We now establish the first of these stylized facts, that firms in larger cities produce goods with higher quality. We largely follow Chor (2005) and Chor and Manova (2012) in adopting an extensive set of fixed effects to filter out omitted variables as much as possible. Exploiting the variation in product quality measures across cities in a given sector and year, we estimate the following specification:

$$q_{ijkt} = \beta_0 + \beta_1 \ln CitySize_{it} + \gamma' X_{it} + D_{gt} + D_i \times t + \epsilon_{ijkt} \tag{1.1}$$

where q_{ijkt} is a measure for quality of goods produced by firm j from city i in industry k, and $\ln CitySize_{it}$ is the log of employment size of city i during year t. Standard errors are clustered by city to account for any possible correlations of idiosyncratic noises within then same city. The results are qualitatively similar if the standard errors are clustered at the city-industry level. The main coefficient of interest is β_1, which captures the effect of city size on average quality of goods produced in the city. We expect $\beta_1 > 0$, so that agglomeration induces firms to produce goods with higher quality, on average.

The city-year specific vector X_{it} controls for other possible determinants of product quality besides agglomeration. First, we control for skill premium, which is defined as the ratio between wages of skilled labor and that of unskilled labor, in a given city i and year t. This variable determines the relative price of skilled labor, and hence partly determines the relative cost of producing higher-quality products (Fieler, Eslava and Xu, 2018; Dingel, 2017). We expect that it should be negatively correlated with product quality. However, given that we only have data on skill premium across cities in one year, this variable will be city-specific, and it will be subsumed by city-industry fixed effects. Next, we include a set of measures on the demand faced by the firms across cities. In particular, we focus on non-homothetic demand which is documented extensively in the international trade literature (Fajgelbaum, Grossman and Helpman, 2011, 2015; Dingel, 2017) as an important determinant of product quality. To do so, we construct measures on both domestic and foreign non-homothetic demand. We proxy the domestic non-homothetic demand measure by using the average income of markets weighted by trade cost for each city. Importantly, we use the trade cost estimates from Ma and Tang (2019) which are based on various modes of transportation network and realistic geography in China. For foreign measures, we include a firm-level control which indicates whether a firm is exporting in a given year. [6] We expect these coefficients to be positive, as firms that are closer to high-income cities and firms that are exporting should specialize in producing goods with higher quality.

To control for omitted variable bias, we include a battery of fixed effects as well as city-specific time trends in the specification. Ideally, we should include D_{ig} which are the city-product specific fixed effects. These control for comparative advantage patterns across space (Chor, 2010), which may confound the correlation between quality and city size.[7] One possible reason could be that big cities in China are mostly located in the coastal regions. These regions may be endowed with time-invariant comparative advantage in producing higher-quality products because they are natural manufacturing hubs for exports to destinations with higher income. However, given the limitation of our sample size, we would not have meaningful variation in identifying city-product specific fixed effects which are close to one-hundred thousand in number.

Instead, we choose to include product-year fixed effects, D_{gt}. These control for product-specific shocks that may affect quality choices. In particular, they subsume and control for any technological progress in the industries, for the changes in availability of high-quality inputs in each industry, and for times-series variation in export demand of different products.

Lastly, we control for linear city-specific time trends, $D_i \times t$, that may affect quality. These control for the possibility that the correlation we observe is driven by the time-

[6]To further examine the heterogeneous effects, we also add an interactive term with the log of firm export value. To avoid log of zeros, we take the log of $1 + Export$.

[7]Similar strategies have been employed in other papers to control for comparative advantage pattern. Chor and Manova (2012) include country-sector fixed effects to control for comparative advantage that may affect pattern of exports. Wang and Li (2017) use the interactions between country and industry characteristics to identify how ICT acts as a source of comparative advantage.

series variations in some other unobservable variable which affects both city size and product quality. Possible scenarios could be either about the massive urbanization due to the commercialization of housing markets or the state-owned enterprise reform in the beginning of 2000s. They also help to capture any time trend such as the trend in the availability of high-quality inputs or increasing product entry, which are usually more prevalent in big cities. The results are qualitatively similar if we include a quadratic term to control for non-linear time trends.

In sum, our specification includes an extensive set of controls and fixed effects that allow us to establish a robust *correlation* between city size and product quality. First, we control for several other factors which may affect product quality choices. Second, the set of fixed effects we include is close to exhaustive with the exception of city-time fixed effects. We could not include these as they will subsume all the effects that city size ($\ln CitySize_{it}$) has on product quality. As a result, any omitted variable that is city-time specific may confound our estimate. We partly address this concern by including city-specific time trends as well as city-time specific variation in non-homethetic demand. Together, these controls should allay concerns regarding omitted variable bias.

Note, however, the current specification does not allow us to establish causality between city size and quality choices for two reasons. First, there are potential concerns that more productive firms will sort into big cities (Gaubert, 2018). Moreover, as we will document later, more productive firms tend to produce higher-quality goods. Hence, it could be that firms in big cities produce higher-quality goods not because there is a treatment effect of agglomeration, but rather due to the fact that more productive firms choose to locate in big cities. Second, there are also concerns about reverse causality. Under a costly trade setting, the availability of high-quality goods may induce people to agglomerate. Given these concerns, our regression merely establishes a robust correlation between city size and average quality. Without a credible identification strategy, we cannot disentangle the endogenous economic forces at work. These questions are left to be answered in our structural estimation.

Results

We report the regressions in Table 1.1. As expected, the coefficients for city size are both economically and statistically significant across all three proxies for product quality except for market shares. Our estimates for Column (1) and (3) imply that on average product quality becomes 7% larger when a city grows double in size. This result also echoes our spatial equilibrium of the structurally-estimated quantitative model in Section 1.5, in which we find that product quality is about 23% higher in a big city about 4 times as large as a small city. In contrast, our reduced-form estimates here would have suggested that this number is 28% (=4×7%), which is close to that of the structural model. The results in this table reinforce our confidence in the structural model.

[Insert Table 1.1 here]

Stylized Fact 2: More productive firms tend to specialize in producing goods with higher quality.

Econometric Design

We now establish the second stylized fact and answer the question that to what extent firm heterogeneity matters in shaping the quality specialization pattern. The quality literature in international trade has supplied ample evidences that more productive firms choose to produce goods with higher quality because higher productivity provides more room for quality upgrading. However, most evidences in the trade literature focus on exporters with few papers examining the firms in the non-international universe. One exception is Saito and Matuura (2016) which estimate product quality using Japanese manufacturing census. Although our measurement of quality largely follows the trade literature and hence is similar to Saito and Matuura (2016), our empirical design is different and complements their approach. In particular, we explicitly estimate the production function and regress proxies for quality on our productivity estimates. Key to our identification is an exhaustive set of fixed effects which include city-time fixed effects D_{it}, city-industry fixed effects D_{ik}, and product-time fixed effects D_{gt}. To this end, we estimate the following econometric specification:

$$q_{ijkt} = \beta_1 z_{ijkt} + \alpha \mathbf{1}_t\{j = exporter\} + D_{gt} + D_i \times t + \varepsilon_{ijkt} \tag{1.2}$$

where q_{ijkt} is a proxy for quality of goods produced by firm j from city i in sector k. z_{ijkt} is the productivity estimate of firms. We cluster standard errors at the city level, but the results are similar under clustering by city-industry. The main coefficients of interest is β_1, which captures the extent to which heterogeneity in firm productivity shapes the choice of quality. We expect the sign to be positive, as firms that are more productive would be able to afford costly quality upgrading. To filter out the omitted variables as much as possible, we also include an exhaustive set of fixed effects which is similar to our previous specification.

Results

We report the regressions in Table 1.2. All our estimates are economically and statistically significant, although the magnitude varies across the three measures. Taken literally, the coefficients for productivity would suggest that a firm that is twice more efficient would have specialized in products that are 10.6% higher in unit price, 31.7% larger in market share, and 41.0% higher in quality. Although these are economically large coefficients, we cannot distinguish the forces at work. In our structural model, productivity estimates are an ex-post result of firm sorting and treatment effect of agglomeration. Without a clear identification strategy, it's impossible to know how much each force has contributed to the observed quality specialization pattern. We will address these questions in our structural estimation.

[Insert Table 1.2 here]

Stylized Fact 3: Firms that produce high-quality goods also employ more skilled labor.

Lastly, we document an empirical relationship between quality of goods that a firm produces and the ratio of skilled labor that it hires. Intuitively, firms that want to upgrade their product quality should employ more skilled labor, because it takes more research engineers as well as skilful technicians to design and manufacture higher-quality products. In addition, production of quality is also costly, in the sense that it takes more workers whether skilled or unskilled to upgrade quality. Therefore, we expect that more productive firms are more likely to produce higher-quality goods, and they also employ relatively more skilled workers because their productivity advantage leaves more room for costly quality upgrading.

The data are consistent with our prior. We show this in several scatter plots based on the following specification.

$$y_{ijk,2004} = \beta z_{ijk,2004} + \alpha \mathbf{1}_{2004}\{j = exporter\} + D_{ik} + D_g + \epsilon_{ijk,2004} \tag{1.3}$$

where $y_{ijk,2004}$ is an outcome variable such as skill intensity or product quality of firm j in sector k from city i in year 2004. $z_{ijk,2004}$ is the productivity estimate of the same firm in 2004, and D_{ik} is a city-sector specific fixed effects that control for comparative advantage pattern over space. Note that we only include data in 2004, because the dataset only contains information on education of employees during that year. As such, all time-specific fixed effects disappear, and we replace them with city-industry fixed effects and product fixed effects.

First, we show that firms that produce high-quality goods also employ more skilled workers. To this end, we extract the residuals from the regressions in specification (1.3) but excluding the productivity control variable (z_{ijkt}). As such, we have two sets of residuals. Each set will separately correspond to the ones extracted from the regression that uses product quality or skill intensity as the outcome variable. Then, we scatter-plot these residuals in panel A of Figure 1.1. The vertical axis corresponds to the residuals from the product quality regression while the horizontal axis corresponds to the residuals from the skill intensity regression, both excluding the productivity control variable. The results are largely consistent with our prior. Firms that specialize in higher-quality products also tend to hire more skilled workers in our sample. This partly motivates our structural estimation where we heavily use empirical moments on skill intensity to identify the quality-related parameters.

Next, we document an empirical correlation suggesting that this pattern is actually driven by heterogeneity in firm productivity. In particular, we further extract a set of residuals from regressing productivity estimates of firms on the set of control variables (except for productivity variable itself) in specification (1.3). We then plot these residuals against the set of residuals from the regressions in the previous section. The results show that both skill intensity and quality are related to firm productivity. This further motivates our structural model as we use empirical moments on firm size which is a result of higher productivity to jointly identify the parameters.

In sum, we have shown that product quality that a firm chooses is positively correlated with the skill intensity a firm employs. This pattern is also related to firm heterogeneity, in the sense that both variable are positively correlated with productivity estimates. As such, this stylized fact motivates our choice of moments in the structural estimation of the spatial-equilibrium model.

[Insert Figure 1.1 here]

1.4 The Model

1.4.1 Housing Sector

We build our model based on the framework in Gaubert (2018). There are a number of ex-ante identical "sites" which are treated as cities. Each city consists of two separate areas, downtown (D) and suburb (S). Each area is endowed with a fixed amount of land normalized to 1. To introduce congestion forces that prevent the indefinite growth of a city, we follow Gaubert (2018) in assuming that housing is constructed using land which in fixed supply and workers,

$$H = \Lambda^h \left(\frac{l_u}{1-h} \right)^{1-h}$$

where H is the amount of housing production, Λ is the amound of land input, l_u is the amount of unskilled labor input, and h is the intensity of land in building houses. This assumption of using inelastic land supply as a congestion force is well-established in the literature, see Helpman (1998), Monte, Redding and Rossi-Hansberg (2018), Rossi-Hansberg (2005), and Ahlfeldt, Redding, Sturm and Wolf (2015).

1.4.2 Demand

There are two types of workers in this economy: skilled and unskilled. We denote these types by $\zeta \in \{s, u\}$. The preferences are assumed to be homogeneous across all workers regardless of their types. In particular, we assume a three-tier utility structure. In the top tier, an individual has Stone-Geary preference for consumption C and housing H,

$$U = \left(\frac{C}{\alpha}\right)^{\alpha} \left(\frac{H - \bar{h}}{1 - \alpha}\right)^{1-\alpha}$$

where $\bar{h}$ is the minimum floor space an individual need to survive, and C is a Cobb-Douglas aggregator across traded goods from S sectors,

$$C = \prod_{j=1}^{S} C_j^{\beta_j}, \quad \text{with} \sum_{j=1}^{S} \beta_j = 1.$$

In the bottom tier, C_j is a CES aggregator over varieties φ within a sector j. Up to now, the demand structure is identical to those in Gaubert (2018) except that we used Stone-Geary preference in the top-tier utility. To introduce quality in this quantitative framework, we incorporate preference for quality such that the bottom-tier utility function is

$$C_j = \left[\int \Phi(\omega, q)^{\frac{1}{\sigma_j}} c_s(\omega)^{\frac{\sigma_j - 1}{\sigma_j}} d\omega\right]^{\frac{\sigma_j}{\sigma_j - 1}}$$

where $\Phi(\omega, q)$ is a preference shifter for variety ω with quality q, and σ_s is elasticity of substitution across varieties in sector j. We further assume that $\Phi(\cdot)$ is increasing in q so that consumers value products with higher quality. Given our assumption of Stone-Geary preference in the outer layer, the expenditure share of high-quality goods will be increasing in income.

1.4.3 Housing Sector and Wage Premium

We index cities by the sizes of skilled and unskilled labor (L_s, L_u). Conditional on living in a city (L_s, L_u), a type-ζ worker will inelastically supply a unit of labor and earn wage $w_\zeta(L_s, L_u)$. Given the city she is in and the wage she earns, a worker chooses the amount of consumption composite C and housing H to maximize her utility, subject to the budget constraint $PC + p_H(L_s, L_u)H = w_\zeta(L_s, L_u)$. Notice that consumption composite C and ideal price index P are not tied to city size, because we assume that trade cost is absent in order to abstract away from any home-market effect.

Consider the partial equilibrium in the housing sector. Landlords, who own the land in a city, will take the general equilibrium prices as given and develop houses according to the following supply equation,

$$H(L_s, L_u) = \left[\frac{p_H(L_s, L_u)}{w_u(L_s, L_u)}\right]^{\frac{1-h}{h}}$$

where $H(L_s, L_u)$ is the total amount of houses supplied by the landlords in a city with L_s skilled labor and L_u unskilled labor. For the demand side, given our assumption of Stone-Geary preference and the fact that there are two areas in a city, there will be within-city sorting pattern if we assume that the housing price in downtown is higher than that of in suburb (e.g., because amenity is higher in the city center). In the appendix, we supply a microfoundation for such sorting behavior which is built on a random utility model. For simplicity matter, we assume that there will be perfect sorting such that skilled workers sort into downtown and unskilled workers sort into the suburb. Given the general equilibrium prices, workers' utility maximization problem entails that the demand for houses and consumption composite by worker types are

$$h_s = \frac{(1-\alpha)(w_s - p_H^D \bar{h})}{p_H^D} + \bar{h}, \quad c_s = \frac{\alpha(w_s - p_H^D \bar{h})}{P}; \quad h_u = \frac{(1-\alpha)(w_u - p_H^S \bar{h})}{p_H^S} + \bar{h}, \quad c_u = \frac{\alpha(w_u - p_H^S \bar{h})}{P}$$

where (p_H^D, p_H^S) are the housing prices, and we suppress the notations of city sizes for simplicity matter. Equating the housing supply with the housing demand in each area will pin down the house price in each city,

$$L_s\left[(1-\alpha)\frac{w_s(L_s,L_u)-p_H^D(L_s,L_u)\bar{h}}{p_H^D(L_s,L_u)}+\bar{h}\right]=\left[\frac{p_H^D(L_s,L_u)}{w_u(L_s,L_u)}\right]^{\frac{1-h}{h}}, \tag{1.4}$$

$$L_u\left[(1-\alpha)\frac{w_u(L_s,L_u)-p_H^S(L_s,L_u)\bar{h}}{p_H^S(L_s,L_u)}+\bar{h}\right]=\left[\frac{p_H^S(L_s,L_u)}{w_u(L_s,L_u)}\right]^{\frac{1-h}{h}}. \tag{1.5}$$

Note that the equations above implicitly define (p_H^D, p_H^S) as a function of (L_s, L_u) conditional on wages. Substituting the housing prices (p_H^D, p_H^S) back to the utility function for both types of workers, we have

$$\bar{U}_s=\left(\frac{w_s-p_H^D\bar{h}}{P}\right)^{\alpha}\left(\frac{w_s-p_H^D\bar{h}}{p_H^D}\right)^{1-\alpha}, \tag{1.6}$$

$$\bar{U}_u=\left(\frac{w_u-p_H^S\bar{h}}{P}\right)^{\alpha}\left(\frac{w_u-p_H^S\bar{h}}{p_H^S}\right)^{1-\alpha}. \tag{1.7}$$

where $\bar{U}_s$ and $\bar{U}_u$ are constants since workers are freely mobile across space. Hence, the wages and house prices (w_s, w_u, p_H^D, p_H^S) of a particular city are jointly pinned down by equations (1.4), (1.5), (1.6), and (1.7) as a function of the city index/city size (L_s, L_u). That is, (L_s, L_u) are sufficient statistics to characterize the wages and house prices in a city, conditional on general equilibrium constants $\bar{U}_s$, $\bar{U}_u$, and P. We establish the following proposition on the behavior of our model by applying the implicit function theorem and the Cramer's rule to the system of equations.

Proposition 1.1. *House prices and wages are increasing in city size, while skill premium is proportional to the relative skill labor size across cities if necessary housing is sufficiently small in comparison to the general equilibrium price index, in the sense that,*

$$\frac{dp_H^D}{dL_s}>0,\quad \frac{dp_H^D}{dL_u}>0,\quad \frac{dp_H^S}{dL_u}>0;\quad \frac{dw_s}{dL_s}>0,\quad \frac{dw_s}{dL_u}>0,\quad \frac{dw_u}{dL_u}>0;\quad \frac{dw_s/w_u}{dL}\propto\frac{L_s}{L_u}.$$

Intuitively, house prices are higher in larger cities because of the congestion force of fixed land supply. In turn, wages must also be higher in larger cities to compensate for the higher living costs. The skill premium is positively related to the skill composition of a city and is unclear ex ante if it increases with city size. Empirically, it is increasing with respect to city size such that skill premium is higher in larger cities (Davis and Dingel, 2019; Diamond, 2016; Ma and Tang, 2018). Accommodating this empirical regularity is critical for our quantitative exercise, as quality choices of firms will be affected by the skill premium in our model.

1.4.4 Production and Quality

Similar to Gaubert (2018), we assume that a firm with innate productivity z uses capital and labor to produce a variety with quality q in sector j of a city (L_s, L_u) with total population $L = L_s + L_u$. In particular, we assume that the production function is

$$y_j(z, L.q; s_j) = k^{\gamma_j} \ell(q, \varphi)^{1-\gamma_j}$$

where $\varphi \equiv \varphi(z, L, q; s_j)$ is a labor-augmenting firm productivity that will be explained in the next section and ℓ is the effective labor composite that combines high-skill and low-skill local labor imperfectly

$$\ell = \left[\chi_u(q, \varphi)^{\frac{1}{\sigma_L}} \ell_u^{\frac{\sigma_L - 1}{\sigma_L}} + \lambda^{\frac{1}{\sigma_L}} \chi_s(q, \varphi)^{\frac{1}{\sigma_L}} \ell_s^{\frac{\sigma_L - 1}{\sigma_L}} \right]^{\frac{\sigma_L}{\sigma_L - 1}}.$$

The interpretation of our specification of the production function is as follows. $\sigma_L > 1$ measures the degree of substitution between skilled labor and unskilled labor. λ denotes the relative importance of effective skilled labor in the production. $\chi_u(q, \varphi)$ and $\chi_s(q, \varphi)$ capture the productivity of workers in a firm of productivity $\varphi(z, L, q; s_j)$ to produce outputs with quality q. In particular, we assume that $\partial \chi_\varsigma(q, \varphi)/\partial q < 0$ so that firms find it costly to upgrade product quality. We also assume $\partial \chi_\varsigma(q, \varphi)/\partial \varphi > 0$ and $\chi_s(q, \varphi)/\chi_u(q, \varphi)$ is increasing in φ, so that more productive firms face lower marignal cost and also employ more skilled labor.[8] In addition, we follow Fieler, Eslava and Xu (2018) in assuming that to produce a variety with higher quality q, a firm has to employ relatively more skilled workers. Taking the ratio over the factor demand of two labor

[8] In the empirical implementation, however, we allow that $\chi_s(q, \varphi)/\chi_u(q, \varphi)$ can be decreasing in φ

types, the expression for skill intensity can be written as

$$\frac{\ell_s^*(z, L_s, L_u)}{\ell_u^*(z, L_s, L_u)} = \lambda \frac{\chi_s(q, \varphi)}{\chi_u(q, \varphi)} \left[\frac{w_s(L_s, L_u)}{w_u(L_s, L_u)} \right]^{-\sigma_L},$$

which is increasing in the importance of skilled labor, increasing in the targeted level of quality as long as skilled workers are relatively more productive in higher quality output $\chi_s(q, \varphi)/\chi_s(q, \varphi) > 0$, increasing in the productivity of the firm φ, and decreasing in skill premium in the located city (L_s, L_u). Holding everything else constant and without considering the agglomeration effect on productivity, firms tend to choose a lower skill intensity in a larger city since skill premium is higher in big cites.

In addition, we assume that there is a fixed cost for quality upgrading $f_q q$ which is increasing in the choice of quality q. Denote the optimal choice of factors as $(k^*, \ell_s^*, \ell_u^*)$, the total profit of a firm z producing variety of quality q in sector j of a city (L_s, L_u) is then

$$\pi(k^*, \ell_s^*, \ell_u^*; L_s, L_u, q) = r_j^*(z, L_s, L_u) - [rk^* + w_s(L_s, L_u)\ell_s^* + w_u(L_s, L_u)\ell_u^*] - f_q q$$

1.4.5 Productivity and Agglomeration

Following Gaubert (2018), we assume that productivity $\varphi(z, L, q; s_j)$ of a firm z located in a city (L_s, L_u) is increasing in the innate efficiency z. There is also local agglomeration externality related to the total size of labor in the located city. The key assumption to generate sorting pattern due to agglomeration is that $\varphi(\cdot)$ presents a strong complementarity between agglomeration and innate efficiency, where s_j captures the sectoral heterogeneity of the log-supermodular forces.

Assumption 1.1. *$\varphi(z, L, q; s_j)$ is strictly log-supermodular in the size of labor $L = L_s + L_u$ and firm innate efficiency z, and is twice differentiable such that*

$$\frac{\partial^2 \log \varphi(z, L; s_j)}{\partial L \partial z} > 0.$$

Our assumption that φ is only related to the total labor size $L = L_s + L_u$ but not the skill composition is too strong and ad-hoc. However, we are only able to do this because there is no prior structural estimates on the traditional agglomeration parameters *and* the log-supermodular forces for *both* skilled and unskilled population size in the literature. In addition, we lack the city information to implement a proper structural estimation for these parameters. Nevertheless, we will also examine two extensions of our benchmark model and make sure that the quantitative implications are not too different from our benchmark model. The first extension is that we assume the benefits associated with agglomeration is solely from skilled labor. In the second extension, the agglomeration forces associated with skilled labor will be larger than that of unskilled labor. The emphasis on skilled labor is well grounded in the literature.

1.4.6 Entry and Location Choice

We assume that in order to enter into production, firms pay f_E fixed cost in terms of the final consumption composite. After entry, they draw an innate efficiency z from a distribution $F(\cdot)$. Once they draw the innate efficiency, they will choose a city (L_s, L_u) to produce goods with quality q of their choosing.

1.4.7 Firm's Problem

Formally defined, the firm's problem is to choose optimal amount of factors, level of quality, and labor sizes of a city $(k^*, \ell_s^*, \ell_u^*, q^*, L_s^*, L_u^*)$, in order to maximize its profits. To analyze the optimal behaviors of firms, we break down their decisions into three steps. In the first step, we assume that conditional on demand, quality, and the city it locates in, a firm optimally chooses the amount of factors $(k^*, \ell_s^*, \ell_u^*)$ to maximize profit. Given the assumptions and the CES preference, we can show that the consumer demand for variety z with quality q is

$$c_j^d(z;q) = \Phi_j(z,q)\left[\frac{p_j(z;q)}{P_j}\right]^{-\sigma_j}\frac{X_j}{P_j}$$

where X_j is the aggregate expenditure on sector-j good and P_j is the sectoral ideal price index

$$P_j = \left[\int \Phi_j(z';q')p_j(z';q')^{1-\sigma_j}dz'\right]^{\frac{1}{1-\sigma_j}}$$

From the cost minimization problem, the input cost function for producing one unit of output is

$$\kappa_j(z;q) = \frac{r^{\gamma_j} w^{1-\gamma_j}}{\gamma_j^{\gamma_j}(1-\gamma_j)^{1-\gamma_j}}$$

$$\text{where} \quad w(q,\varphi,L_s,L_u) = \left[\chi_u(q,\varphi) w_u(L_s,L_u)^{1-\sigma_L} + \lambda \chi_s(q,\varphi) w_s(L_s,L_u)^{1-\sigma_L}\right]^{\frac{1}{1-\sigma_L}}.$$

Given the input cost function, the firm's problem is then to set prices that maximizes its operational profit,

$$\max_{p_j} \ \pi_j(z;q) = \underbrace{[p_j(z;q) - \kappa_j(z;q)]}_{\text{per uint profit}} \underbrace{\left[\frac{p_j(z;q)}{P_j}\right]^{-\sigma_j} \frac{\Phi_j(z,q)X_j}{P_j}}_{\text{demand}}$$

Since the market structure is monopolistic and the preference is CES, firm pricing must that it charges a constant markup $\frac{\sigma_j}{\sigma_j-1}$ over the unit input cost. Substituting this into the operational profit function, we have

$$\begin{aligned} \pi_j^*(z;q,L_s,L_u) &= \frac{1}{\sigma_j}\left[\frac{\sigma_j \kappa_j(z;q)}{\sigma_j - 1}\right]^{1-\sigma_j} \Phi_j(z,q) P_j^{\sigma_j-1} X_j \\ &= \Upsilon_{1j} \frac{\Phi_j(z,q)}{w(q,\varphi,L_s,L_u)^{(1-\gamma_j)(\sigma_j-1)}} P_j^{\sigma_j-1} X_j \end{aligned}$$

where Υ_{1j} collects the sector-specific constants,

$$\Upsilon_{1j} = \sigma_j^{-\sigma_j}\left[(\sigma_j-1)\gamma_j^{\gamma_j}(1-\gamma_j)^{1-\gamma_j} r^{-\gamma_j}\right]^{\sigma_j-1}.$$

In the second step, conditional on the city size of location, firms optimally choose product quality to maximize their profits $q^* = \underset{q\geqslant 0}{\operatorname{argmax}}\, \pi_j^*(z;q,L_s,L_u) - f_q q$, where $\pi_j^*(z;q,L_s,L_u)$ is the optimal profit computed in the first step and $f_q q$ captures the fixed costs of quality upgrading. From the first-order condition, the optimal level of quality q^* chosen by firm z is characterized by the following equation,

$$\pi_j^*(z;q,L_s,L_u)\left[\underbrace{\frac{1}{\Phi_j(z,q)}\frac{\partial \Phi_j(z,q)}{\partial q}}_{\Delta \text{ in sales due to higher } q} - \underbrace{\frac{(1-\gamma_j)(\sigma_j-1)}{w(q,\varphi,L_s,L_u)}\frac{\partial w(q,\varphi,L_s,L_u)}{\partial q}}_{\Delta \text{ in cost due to quality upgrading}}\right] = f_q$$

In practice, we will only be able to solve for the optimal quality choices numerically if $f_q = 0$. To see this, note that one must know all the general equilibrium quantities in order to solve for the individual optimal choices above. However, the general equilibrium quantities can only be known after solving for the individual choices. This poses an insurmountable computational burden. To circumvent this issue, we set $f_q \approx 0$ which is supported by the empirical estimate of 4.7×10^{-5} in Fieler, Eslava and Xu (2018) using Colombian data, so that the first-order condition reduces to

$$\frac{1}{\Phi_j(z,q)}\frac{\partial \Phi_j(z,q)}{\partial q} - \frac{(1-\gamma_j)(\sigma_j-1)}{w(q,\varphi,L_s,L_u)}\frac{\partial w(q,\varphi,L_s,L_u)}{\partial q} = 0.$$

Solving the reduced first-order condition only requires information on the choice of city sizes (L_s, L_u) and is independent of the general equilibrium quantities. This is essentially the key feature in Gaubert (2018) that makes a quantitative model computationally feasible. Invoking the implicit function theorem and the second-order condition for maximizing π with respect to q, we can assess the impact of changes in firm efficiency z on the quality choice q^*. Proposition 2 summarizes our findings.

Proposition 1.2. *Conditional on the cities that the firms are located in and the parameterization of $\varphi(z, L; s_j)$, optimal choice of quality increases with firm innate efficiency z such that $\frac{\partial q^*}{\partial z} > 0$.*

Similarly, we can also show that conditional on city size, firm's choice of quality will be increasing in the size of cities. We state this result more formally in Proposition 3.

Proposition 1.3. *Conditional on its innate efficiency, a firm will choose a higher quality in a larger city if the increase in city size induces the firm to hire more skilled workers, in the sense that,*

$$\frac{\partial q^*}{\partial L} > 0, \quad \textit{if and only if} \quad \frac{\partial \chi_s/\partial L}{\chi_s/L} - \frac{\partial \chi_u/\partial L}{\chi_u/L} > (\sigma_L - 1)\left(\frac{\partial w_s/\partial L}{w_s/L} - \frac{\partial w_s/\partial L}{w_s/L}\right).$$

1.4.8 Firm Sorting to Cities

In the third step, firms choose their location to maximize operation profits.

$$(L_s, L_u) = \underset{L_s \geqslant 0; L_u \geqslant 0}{\operatorname{argmax}} \; \pi_j^*(z; q, L_s, L_u),$$

where $\pi_j^*(z; q, L_s, L_u)$ is the optimal profit that a firm z earns in a city of size (L_s, L_u). Maximizing this profit is then equivalent to maximize $w(q, \varphi, L_s, L_u)^{(1-\gamma_j)(1-\sigma_j)}$. The first-order conditions with respect to L_s and L_u are

$$\frac{\partial w(q, \varphi, L_s, L_u)}{\partial \varphi(z, L; s_j)} \frac{\partial \varphi(z, L; s_j)}{\partial L_s} \gtrless \frac{\partial w(q, \varphi, L_s, L_u)}{\partial L_s}$$

$$\frac{\partial w(q, \varphi, L_s, L_u)}{\partial \varphi(z, L; s_j)} \frac{\partial \varphi(z, L; s_j)}{\partial L_u} \gtrless \frac{\partial w(q, \varphi, L_s, L_u)}{\partial L_u}$$

which implicitly determine the optimal choice of city size in equilibrium. Note that, we do not impose any binding first-order condition because of two reasons. First, depending on the set of available cities, optimal solution may not be available for choosing. Second, by our parameterization of the productivity term, the benefits from agglomeration are the same for skilled labor zie and unskilled labor size, $\partial \varphi / \partial L_s = \partial \varphi / \partial L_u$. Optimal choices of city size by firms then require that the agglomeration benefit to be equated with marginal cost which is how a larger city size will push up the house price and hence wages. However, it often is the case that the size of skilled workers will have a different impact on wages than that of unskilled labor. It is entirely possible that one effect will dominate another and firms will want to choose a city with a larger size of one particular type of population to reap the agglomeration benefit while avoiding a city with more costly production. However, the optimal choices made by firms in the partial equilibrium will be inconsistent with the general equilibrium quantities, in particular, the local labor market clearing conditions. Regardless the firm's choice of city size, the wages for the type of labor that has a higher impact on marginal cost will not be zero in any city. Thus, in such cities, the supply will not meet the factor demand for skilled labor. General equilibrium forces will adjust to make sure that the local labor markets clear.

Nevertheless, it is clear that firms with higher innate efficiency will choose a larger city in our model. The proof of this statement relates to arriving at a contradiction if we assume otherwise. We summarize this claim in the following proposition.

Proposition 1.4. *Firms with a higher innate efficiency will choose to locate in a larger city. That is, suppose there are two firms each with innate efficiency z_H and z_L. Denote the firms' choice of city size in the general equilibrium as (L_s^{H*}, L_u^{H*}) and (L_s^{L*}, L_u^{L*}). Then $L_s^{H*} \geqslant L_s^{L*}$ and $L_u^{H*} \geqslant L_u^{L*}$ if $z_H > z_L$.*

This proposition is essentially similar to the firm sorting behavior established in Gaubert (2018) which we built our model upon, in the sense that firms that have a higher innate productivity will choose to locate in a larger city.

Given the optimal factor usage decisions, quality upgrading decisions, and city choices. The revenue and the factor demand of a firm z are such that

$$\tilde{r}_j^*(z) = \sigma_j \Upsilon_{1j} \frac{\Phi_j(z, q^*)}{w(q^*, \varphi, L_s^*, L_u^*)^{(1-\gamma_j)(\sigma_j-1)}} P_j^{\sigma_j-1} X_j,$$
$$\ell_s^*(z) = \Upsilon_{2j} \frac{\lambda \chi_s(q^*, \varphi)\Phi_j(z, q^*)}{w(q^*, \varphi, L_s^*, L_u^*)^{(1-\gamma_j)(\sigma_j-1)+1-\sigma_L} w_s^{\sigma_L}} P_j^{\sigma_j-1} X_j,$$
$$\ell_u^*(z) = \Upsilon_{2j} \frac{\chi_u(q^*, \varphi)\Phi_j(z, q^*)}{w(q^*, \varphi, L_s^*, L_u^*)^{(1-\gamma_j)(\sigma_j-1)+1-\sigma_L} w_u^{\sigma_L}} P_j^{\sigma_j-1} X_j.$$

where $\Upsilon_{2j} = (\sigma_j - 1)(1 - \gamma_j)\Upsilon_{1j}$.

Proposition 1.5. *In equilibrium, suppose $(L_s^{H*}, L_u^{H*}) > (L_s^{L*}, L_u^{L*})$, then it must be that $z_H \geqslant z_L$. In addition, $\tilde{r}_j^*(z_H) \geqslant \tilde{r}_j^*(z_L)$ and $\pi_j^*(z_H) \geqslant \pi_j^*(z_L)$.*

1.4.9 General Equilibrium

We follow Gaubert (2018) and Tian (2018) to define a spatial general equilibrium as follows. Formally, we define a spatial general equilibrium as a city size distribution $\{L_s, L_u\}$, a set of production decisions $\{p_j(z)\}$ and quality choices $\{q_j(z)\}$ made by a mass of M_j heterogeneous firms indexed by z in each sector j, a set of location choices $\{L_{s,j}(z), L_{u,j}(z)\}$ made by firms, a set of wages for skilled and unskilled workers in each city $\{w_s(L_s, L_u), w_u(L_s, L_u)\}$, a set of housing prices in each city $\{p_H(L_s, L_u)\}$, a set of price index P_j, and the utility of workers $(\bar{U}_s, \bar{U}_u)$ such that,

1. Given wages, house prices, and price indices, skilled and unskilled workers in each city maximize their utilities.

2. Given wages and house prices, landlords maximize their profits from developing houses.

3. Given the city size distributions, firms in each sector j decide their optimal choice of locations $\{L_{s,j}(z), L_{u,j}(z)\}$ and optimal production plans $\{p_j(z), q_j(z)\}$.

4. Goods markets clear. That is in each sector j, aggregate demand is equal to the aggregate sectoral outputs

$$X_j = \sigma_j \Upsilon_j P_j^{\sigma_j - 1} X_j M_j \int_z \frac{\Phi_j(z,q)}{w(q,\varphi,L_s,L_u)^{(1-\gamma_j)(\sigma_j-1)}} dF_j(z).$$

5. Local labor markets clear. That is in each city (L_s, L_u), the markets for skilled and unskilled labor clear,

$$\int_{L_{0s}}^{L_s} n f_{L_s}(n) dn = \sum_{j=1}^{S} M_j \int_0^\infty \mathbf{1}_j(L_s, L_u, z) l_s(z) dF_j(z), \quad \forall L_s > L_{0s}.$$

$$\int_{L_{0u}}^{L_u} n f_{L_u}(n) dn = \sum_{j=1}^{S} M_j \int_0^\infty \mathbf{1}_j(L_s, L_u, z) l_u(z) dF_j(z), \quad \forall L_u > L_{0u}.$$

6. National labor markets for skilled and unskilled labor clear. That is,

$$\bar{L}_s = \sum_{j=1}^{S} \Upsilon_{2j} P_j^{\sigma_j - 1} X_j M_j \int_z \frac{\lambda \chi_s(q,\varphi) \Phi_j(z,q)}{w(q,\varphi,L_s,L_u)^{(\sigma_j-1)(1-\gamma_j)+1-\sigma_L} w_s^{\sigma_L}} dF_j(z).$$

$$\bar{L}_u = \sum_{j=1}^{S} \Upsilon_{2j} P_j^{\sigma_j - 1} X_j M_j \int_z \frac{\chi_u(q,\varphi) \Phi_j(z,q)}{w(q,\varphi,L_s,L_u)^{(\sigma_j-1)(1-\gamma_j)+1-\sigma_L} w_u^{\sigma_L}} dF_j(z) + \bar{L}_u(1-h)(1-\alpha).$$

7. Capital market clears by Walras's Law.

8. The ex-ante expected profit of a firm is zero in each sector j, due to free entry,

$$f_E P = \Upsilon_{1j} P_j^{\sigma_j - 1} X_j \int_z \frac{\Phi_j(z,q)}{w(q,\varphi,L_s,L_u)^{(1-\gamma_j)(\sigma_j-1)}} dF_j(z).$$

9. Spatial no-arbitrage condition holds, such that each type of workers receive the same amount of utility regardless of the city (L_s, L_u) that they are located in.

1.4.10 Parameterization and Calibration

In order to assess the quantitative behavior of the model, we first parameterize the firm productivity term following Fieler, Eslava and Xu (2018) and Gaubert (2018),

$$\log \varphi(z, L; s_j) = a_j \log L + \log(z) (1 + \log L)^{s_j} + \epsilon_{i,L}.$$

We parameterize the term $\varphi(z, L; s_j)$ following Gaubert (2018). The terms are identical to her set up, so we will just rephrase Gaubert's interpretation of these parameters. a_j would capture the traditional agglomeration forces. The second term would capture the interaction between city size L and innate efficiency of the firm z, where sector-specific term s_j governs the quantitative magnitude of the interaction. $s_j > 0$ would ensure the log-supermodularity in our assumption. $\epsilon_{i,L}$ is a term that captures city-size and firm specific idiosyncratic shock to productivity. In particular, Gaubert assumes that firm innate efficiency z follows a truncated log-normal distribution with mean zero and variance $\nu_{z,j}$, while the idiosyncratic productivity shock follows a Gumbel distribution with mean zero and variance $\nu_{\phi,j}$.[9] [10] We also import these assumptions into our model.

Besides the agglomeration parameters, our model also features a set of parameters that characterize skill and quality choices. In particular, we parameterize $\chi_s(q, \varphi)$ and $\chi_u(q, \varphi)$ as follows,

$$\chi_s(q, \varphi) = \varphi^{\lambda_{1s}} \exp(\lambda_{2s} q); \quad \chi_u(q, \varphi) = \varphi^{\lambda_{1u}} \exp(\lambda_{2u} q)$$

which are partly similar to the set up in Fieler, Eslava and Xu (2018). The interpretations of the parameters are as follows. First, λ_{1s} and λ_{1u} capture how the productivity of firms

[9] The distribution for z is truncated so that $\log z$ will be non-negative.

[10] The assumption of Gumbel distribution can be interpreted as that each firm will draw many independent technological shocks that follow an exponential distribution. As the firm can only adopt one direction at a time, the maximum of these shocks would then follow a Gumbel distribution.

accrue to skilled and unskilled workers. If these parameters equal 1, then the $\varphi^{\lambda_{1s}}$ and $\varphi^{\lambda_{1u}}$ terms become the classical labor-augmenting productivity. In our model, we expect that $\lambda_{1s} > \lambda_{1u} > 0$ as empirical evidences suggest that skilled labor receives more benefit from agglomeration than unskilled labor does.

Next, λ_{2s} and λ_{2u} define how costly it is to produce higher-quality good using each type of labor. We expect the sign and magnitude of these parameters to be negative and that $\lambda_{2s} > \lambda_{2u}$. Given the exponential functional form, this implies that production of quality will reduce the productivity of workers, and this productivity-dampening effect is stronger for unskilled workers than for skilled workers. Intuitively, it takes longer time and more effort for workers to produce goods with higher quality, and this is more so for unskilled workers. We choose the exponential functional form because it would generate a skill intensity distribution that is close to the data, as similarly noted in Fieler, Eslava and Xu (2018).[11]

1.4.11 Solving the Model

We now present a step-by-step description on the algorithm we used to solve the model, which is also similar to the algorithm presented in Gaubert (2018).

1. For each sector j, we simulate 8,000 firms with 8,000$\times$200 random variables, where 200 is the number of cities. The simulations are then transformed to the innate efficiency of firms z_i and firm-city specific idiosyncratic shocks $\nu_{i,L}$.

2. We then simulate an initial distribution of city size (L_s, L_u) with the smallest city not smaller than the ones observed in the data.

3. Compute the local wages and house prices given the size of cities.

4. Given the wages and house prices, compute the entry decision, the optimal location choice, and the optimal quality choice made by firms over a grid of 200×200, where we discretize the choice of quality over the interval of [0,10] with a step size of 0.05.

[11] Fieler, Eslava and Xu (2018) use a slightly more complicated functional form. Still, our choices are largely similar to theirs.

5. Given firm choices, compute the sectoral quantities $\tilde{E}_{s,j}$, $\tilde{E}_{u,j}$, and $\tilde{S}_j$ as follows

$$\tilde{E}_{s,j} = \int_z \frac{\lambda \chi_s(q,\varphi)\Phi_j(z,q)}{w(q,\varphi,L_s,L_u)^{(\sigma_j-1)(1-\gamma_j)+1-\sigma_L} w_s^{\sigma_L}} dF_j(z),$$

$$\tilde{E}_{u,j} = \int_z \frac{\chi_u(q,\varphi)\Phi_j(z,q)}{w(q,\varphi,L_s,L_u)^{(\sigma_j-1)(1-\gamma_j)+1-\sigma_L} w_u^{\sigma_L}} dF_j(z),$$

$$\tilde{S}_j = \int_z \frac{\Phi_j(z,q)}{w(q,\varphi,L_s,L_u)^{(1-\gamma_j)(\sigma_j-1)}} dF_j(z).$$

6. Given the sectoral quantities from step 5, compute the general equilibrium quantities $\{X, P_j, M_j\}$ from the following system of equations that represent the goods market clearing condition, the national labor market clearing conditions, and the free-entry condition

$$\begin{aligned}
1 &= \sigma_j \Upsilon_{1j} P_j^{\sigma_j-1} M_j \tilde{S}_j, \text{ for all } j \in S, \\
\bar{N}_u &= \sum_{j=1}^{S} \Upsilon_{2j} P_j^{\sigma_j-1} \beta_j X M_j \tilde{E}_{u,j} + \bar{N}_u(1-h)(1-\alpha), \\
\bar{N}_s &= \sum_{j=1}^{S} \Upsilon_{2j} P_j^{\sigma_j-1} \beta_j X M_j \tilde{E}_{s,j}, \\
f_E P &= \Upsilon_{1j} P_j^{\sigma_j-1} \beta_j X \tilde{S}_j, \text{ for all } j \in S
\end{aligned}$$

7. Given the general equilibrium quantities $\{X, P_j, M_j\}$, compute the local labor market demand for skilled and unskilled labor.

8. If the local labor markets do not clear, then update the city size (L_s, L_u) and go to step 3. If the local labor markets clear, then stop the algorithm and extract the relevant information.

1.5 Quantifying the Model

1.5.1 Data

The dataset we use for our structural estimation is the Annual Survey of Industrial Firms collected by the National Bureau of Statistics of China (NBSC). In particular, we use the data in year 2004 in our baseline quantitative analysis. The universe of firms covered in this dataset spans over all manufacturing firms, which include both state and non-state enterprises, that generate more than 5 million RMB in revenue each year. The dataset reports information on the location, capital, output, taxation, revenue, and education level of the workers in each firm. All firms are codified in four-digit manufacturing classifications and we merge the information of subsidiaries under the same legal entity, which is the identifier that uniquely represents an enterprise in the dataset.

Following the existing literature that uses this dataset, we drop observations on firms that do not meet the following criteria: the number of employees is more than 8 people, total assets less liquid assets is positive, total assets minus total fixed assets is positive, total assets minus total net fixed assets is positive, and accumulated depreciation minus current depreciation is positive. The final sample size of manufacturing firms used in our estimation is 195,384 spanning over thirty two-digit Chinese Standard Industrial Classification (CSIC Rev. 2) sectors. We then concord the CSIC sectors to 17 sectors that are similar to those used in Gaubert (2018) and Caliendo and Parro (2015). The descriptive statistics on the value-added, employment, and proportion of skilled workers with college education are reported in Table 1.3. The concordance of sectors from CSIC to our definition of sectors is detailed in Table A.1 in the appendix.

We obtain the geographic location of firms using postal code reported in the data. A city is defined as the prefecture-level administrative unit in China. We use the prefecture-level population from 2004 China City Statistic Year Book in 2004 to proxy for city size. There are 243 cities in our firm-level dataset. In addition, our quantitative analysis requires information on the skill composition of a city, which is defined as the ratio of the skilled to unskilled workers. We compute this figure using the 1% sample of the 2005 Population Census which reports the interviewee's education level and geographic location. We define people holding bachelor degree or above as skilled workers and the rest as unskilled workers. Due to data limitation, we use the 2000 General Population Survey for Hunan, Hubei, Jilin, Yunnan, Shanxi and Tianjin province to proxy for the skill composition in 2005.

Finally, we follow Gaubert (2018) to divide cities into 4 quartiles according to their size. Different from Gaubert, we define big cities as those cities in the 4th quartile in comparison to the largest cities that account for 50% of total population. In our sample, defining big cities by the 4th quartile implies that there are 12 big cities out of 243 cities. In contrast, using Gaubert's definition of big cities that represent 50% of population translates to 45 big cities. We report the proportion of firms in each sector that is located in big cities (4th quartile) in Table 1.3. It is evident firms from sectors such as medical, machinery, transport and automotive, electrical, and computer are more likely to hire a high proportion skilled workers and also more likely to locate in big cities. To further substantiate this observation, we also plot the city size against the average skill intensity in the city in Figure 1.2. It is clear that a firm's skill employment ratio is positively associated with the city size. This effect is also robust to industry fixed effects and controlling for other firm-level characteristics.

[Insert Table 1.3 here]

[Insert Figure 1.3 here]

1.5.2 Moments and Identification

We structurally estimate the model sector by sector using the Simulated Method of Moments (SMM) estimator which minimizes the weighted distance between simulated moments generated by our model and the empirical moments in the data. The set of parameters that we wish to estimate are $\Theta = \{a_j, s_j, \nu_{R,j}, \nu_{z,j}, \lambda_{1s,j}, \lambda_{1u,j}, \lambda_{2s,j}, \lambda_{2u,j}\}$. In specific, we use the following set of 17 targeted moments to identify these parameters. In general, we want to find those moments which are sensitive to the change in parameter value in simulation, so as to provide identification. Furthermore, these parameters can be partitioned into two disjoint sets, $\Theta_1 = \{\nu_{R,j}, \nu_{z,j}\}$ and $\Theta_2 = \Theta - \Theta_1$. The first set of parameters, Θ_1, does not interact with any city-specific information given the setup of our model. In contrast, the second set of parameters will interact with city-specific labor sizes. As a consequence, the relevant simulated moments will also behave differently with different size of cities. Hence, we shall adopt simulated moments by city quartiles for the second set of parameters but not for the first set of parameters. In particular, we define city quartiles as the 25th, 50th, and 75th percentiles by city size. The moments are reported as follows and the choices are partly similar to those in Fieler, Eslava and Xu (2018) and Gaubert (2018).

Distribution of skill intensity by city size. We compute the average skill intensity (proportion of skilled workers employed by firms) in each quartile of cities and use these figures as the first set of moments $\{m_q^1\}_{q=1,2,3,4}$ to identify $\{\lambda_{1s}, \lambda_{1u}, \lambda_{2s}, \lambda_{2u}\} \in \Theta_2$.

Distribution of value-added by city size. We compute the share of total value-added and average value-added by city quartiles and use them as the second set of moments $\{m_q^2\}_{q=1,2,3,4}$ to identify $\{a_j, s_j\} \in \Theta_2$. Intuitively, both the agglomeration forces and the log-supermodularity forces affect firm's profitability in big and small cities. Therefore, value-added across cities will be a sensitive measure to changes in these parameters.

Distribution of firm size. We use normalized total revenue as a proxy for the size of firms. Then we compute the normalized value-added in the 25th, 50th, 75th, and 90th percentiles and use them as the third set of moments $\{m^3\}$ to identify $\{\nu_{R,j}, \nu_{z,j}\} \in \Theta_1$. Intuitively, firm heterogeneity will affect the distribution of firm size. Therefore our choice of moments will be sensitive to the changes of these parameters.

We then estimate the parameters $\hat{\Theta}$ by targeting the empirical moments using an SMM estimator, $\min_{\hat{\Theta}} \ [\mathbf{m}-\mathbf{m}(\hat{\Theta})]'\mathbf{W}[\mathbf{m}-\mathbf{m}(\hat{\Theta})]$, where $\mathbf{m}(\hat{\Theta})$ is the vector of simulated moments from the model under parameter values $\hat{\Theta}$, $\mathbf{m}$ is the vector of empirical moments, and $\mathbf{W}$ is the weighting matrix. For the benchmark estimation, we use the identity matrix as the weighting matrix. An alternative estimate using a generalized variance-covariance matrix $\mathbf{W}$ by bootstrapping the sample with replacement for 2,000 times following Eaton, Kortum and Kramarz (2011) is reported in the appendix for robustness check purposes. In addition, optimization involving an SMM objective is usually neither convex nor concave. Thus, we use Simulated Annealing algorithm which is a probabilistic global algorithm for our estimation. This algorithm is known for its accuracy and is widely used in the literature (Eaton, Kortum and Kramarz, 2011; Gaubert, 2018; Antràs, Fort and Tintelnot, 2017). In practice, we first search over a grid of parameters space to find an initial combination of parameter values that produces a relatively small loss. We then use these parameter values as the starting point and apply the annealing algorithm. This procedure speeds up our estimation and is robust to our choice of initial values. Starting the annealing algorithm from another random grid point converges to a set of similar estimates.

1.5.3 Structural Estimates

We will shortly update our structural estimates of the parameters with corresponding standard errors. We did not impose any restriction on the values of the parameters in the estimation. The values of the estimated parameters for $\{a_j, s_j, \nu_{R,j}, \nu_{z,j}\}$ are similar to the prior estimates in the existing literature such as Gaubert (2018) and Tian (2018). Our estimates for the traditional agglomeration parameter a_j and the parameter that governs the log-supermodular complementarity force s_j are positive for all sectors except for the manufacturing of plastic and food. The standard interpretation of the negative estimates in the literature is that these are consider mature sectors and hence are associated with different agglomeration forces Gaubert (2018).

We now discuss the estimates for $\{\lambda_{1s,j}, \lambda_{1u,j}, \lambda_{2s,j}, \lambda_{2u,j}\}$ which is the set of parameters new in our model in comparison to the literature. Our estimates suggest that the productivity advantage of big cities is skill-biased, as the estimates are positive and $\hat{\lambda}_{1s}$ is greater than $\hat{\lambda}_{1u}$ in all sectors. This echos a strand of literature which argues that agglomeration forces benefit skilled workers more, for example because high-ability individuals learn better from idea exchange (Davis and Dingel, 2019). In particular, our estimates suggest that agglomeration disproportionately benefit skilled workers in medical and computer sectors, partly due to the fact that these sectors require extensive idea exchange among engineers and professionals. Finally, our estimates of $\hat{\lambda}_{2s}$ and $\hat{\lambda}_{2u}$ suggest that production of higher-quality good is costly and requires employing more labor, since both $\hat{\lambda}_{2s}$ and $\hat{\lambda}_{2u}$ are negative. Our estimates also imply that production of higher quality good is intensive in skilled labor, since $\hat{\lambda}_{2s} > \hat{\lambda}_{2u}$.[12]

[12] Given our parameterization of the model, skill intensity of a firm z in a city (L_s, L_u) can be written as

$$\frac{\ell_s^*(z, L_s, L_u)}{\ell_u^*(z, L_s, L_u)} = \lambda \frac{\chi_s(q,\varphi)}{\chi_u(q,\varphi)} \left[\frac{w_s(L_s, L_u)}{w_u(L_s, L_u)}\right]^{-\sigma_L} = \lambda \varphi^{\lambda_{1s}-\lambda_{1u}} e^{(\lambda_{2s}-\lambda_{2u})q} \left[\frac{w_s(L_s, L_u)}{w_u(L_s, L_u)}\right]^{-\sigma_L}.$$

It implies to produce a higher-quality good, a firm will employ relatively more skilled labor if and only if $\lambda_{2s} - \lambda_{2u} > 0$.

1.5.4 Quantitative Results

We feed our parameter estimates into the model and extract average choice of product quality of the simulated firms by city quartiles. We find that our model generates significant quality differences across space. On average, product quality in the big cities (the 4th quartile) is 22.9% higher than that of the smallest cities (the 1st quartile). There is also significant sectoral heterogeneity in the quality specialization across space. For manufacturing sectors such as medical equipment, transport and automotive, food, and furniture, the average product quality difference between big and small cities can be as high as 27.4% to 59.5%. We report the entire distribution of quality choices across all firms located in different city quartiles in Figure 1.3.

[Insert Figure 1.3 here]

To further assess the contribution of firm sorting and traditional agglomeration benefit in determining the quality differences across space, we follow Gaubert (2018) and consider the following regression. We regress each simulated firm's choice of quality on the size of city that it locates in with industry fixed effect. We then repeat the exercise in a counterfactual where we shut down the sorting of firms by setting the efficiency of every firm to the average efficiency in the benchmark model and compute the reallocation of economic activities across space. The results are reported in Table 1.4. In Column (1) where we have the full model, a 10% increase in city size translate to a 1% increase in quality. In contrast, in Column (2) where sorting of firms is shut down, the effect is dampened and is only half of the effect in the full model. This suggests that sorting of firms accounts for half of the quality differences in big cities while traditional agglomeration forces account for the other half.

[Insert Figure 1.4 here]

1.5.5 Goodness of Fit: Within-Sample and Out-of-Sample

We first evaluate the fit of our model by comparing the simulated moments in the model to the empirical moments in the data. A summary of the results is reported in

Table 1.5, where we aggregated moments across sectors. We also report the goodness of fit sector by sector in Appendix A.4. In general, our model fits the data well. Our model succeeds in generating a similar average skill intensity and mean value-added (both normalized by the mean) across city quartiles in comparison to the corresponding statistics in the firm-level data. Our model also performs reasonably in generating a firm size distribution and value-added share that is close to the data, although our model implies a slightly larger value-added share in the big cities (4th quartile) and a larger revenue share among the biggest cities (90th percentile).

Our model also succeeds in fitting data moments out-of-sample. First, the city-size distribution generated by our benchmark model is able to replicate the distribution in the data. As shown in Figure 1.4, the city-size distribution implied by our model, which consists of the sum of firm's factor demand for skilled labor and unskilled labor in each city, is largely consistent with the pattern in the data. Our calibrated city-size distribution also roughly follows Zipf's Law with a slope of -1.3. (Zipf's law predicts that the slope of log rank-size regression is -1). One reason that city-size distribution in China does not perfectly follow Zipf's law is that the administrative boundary of each prefecture does not fit a commute-based definition (Dingel, Miscio and Davis, 2019). As a sensitivity check, we will also repeat our analysis using the alternative boundary of cities based on the light-based metropolitan definition in Dingel, Miscio and Davis (2019).

[Insert Table 1.5 here]

[Insert Figure 1.4 here]

1.6 Counterfactuals

We now evaluate the general equilibrium impact of a spatial policy that is frequently employed in developing economies such as China. The policies that we aim to evaluate are policies that regulate the use of land in a city such as zoning restrictions. Matching these policies to our model counterparts, land use regulation is approximated by the land use intensity in the production of housing. Whenever there are relatively few land use regulations, the land use intensity coefficient should be smaller as it is easier for developers to acquire land in their housing production. In the counterfactual, we shock the coefficient such that the coefficient for the high-end housing market becomes 20% smaller than the original value. The resulting changes are reported in Table 1.6. In overall, average quality across cities has increased by 5.5% while the aggregate welfare of all residents has increased by 20%.

[Insert Table 1.6 here]

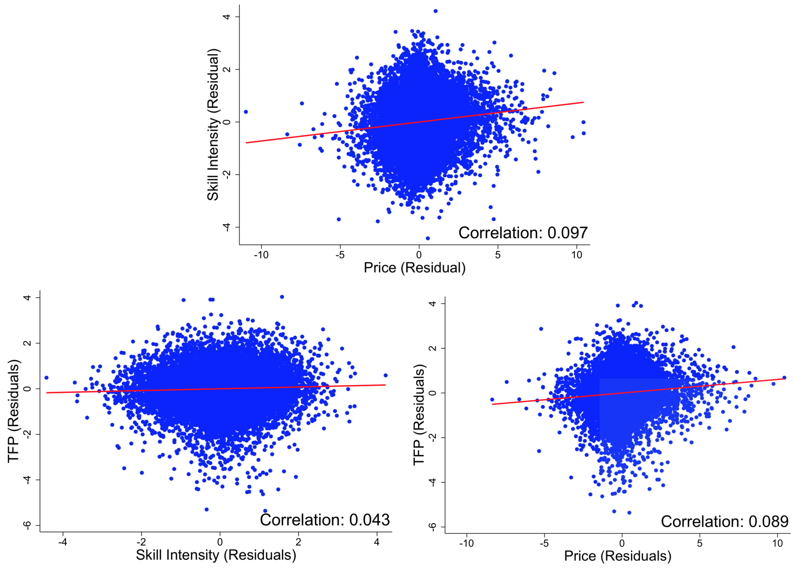
However, there are two channels that a relaxed land use regulation can affect welfare in our model. The first channel is that an increase in the supply of housing directly enters individuals' utility function. In addition, the increase in housing supply also alleviate the congestion forces and flattens the skilled wage schedule across cities. To disentangle the two effects on welfare, we follow Gaubert (2018) to first compute the reallocation of economic activities across space under the new intensity parameter. We then hold the land intensity parameter fixed at the old value and recompute the equilibrium in the housing market. The resulting indirect welfare effect is then the "pure" welfare effect resulted from changes in sorting alone. The direct welfare effect from increase in housing supply is isolated through the design of our counterfactual. We find that the indirect welfare for individuals is 6.2% higher, while the direct welfare is 13.8% higher.

1.7 Conclusion

In this paper, we study the pattern of quality specialization across Chinese cities through the lens of firm sorting and agglomeration. Extending the framework in Gaubert (2018) with two skill types and quality choices, we show theoretically both firm sorting and productivity advantage of agglomeration will induce quality upgrading. We structurally estimate and quantify the model using a plant-level dataset spanning the universe of manufacturing firms. We find that on average, product quality in big cities is 23% higher than that of small cities. A decomposition analysis shows that sorting and agglomeration each explains half of the quality pattern. Armed with the structural estimates, we then evaluate a potential policy that reduces land use regulations. We find that a 20%

relaxation of land intensity induces a 5.5% increase in quality across cities and a 6.2% indirect welfare benefit. For future work, one could further quantify the relative magnitude of both the demand and supply-side explanations, as well as incorporating input-output linkages in the present model.

Figure 1.1: Stylized Fact 3 - Skill Intensity and Quality



Notes: Observations are in 2004, the only year in which the educational levels of a firm's employees are available. "Skill Intensity" is defined as the ratio of college-graduated workers to total workers of a firm. "Price" and "Skill Intensity" are taken log in the regressions. All regressions include a constant term. Robust standard errors are clustered at city-industry level.

Figure 1.2: Correlation of Skill Intensity and City Size

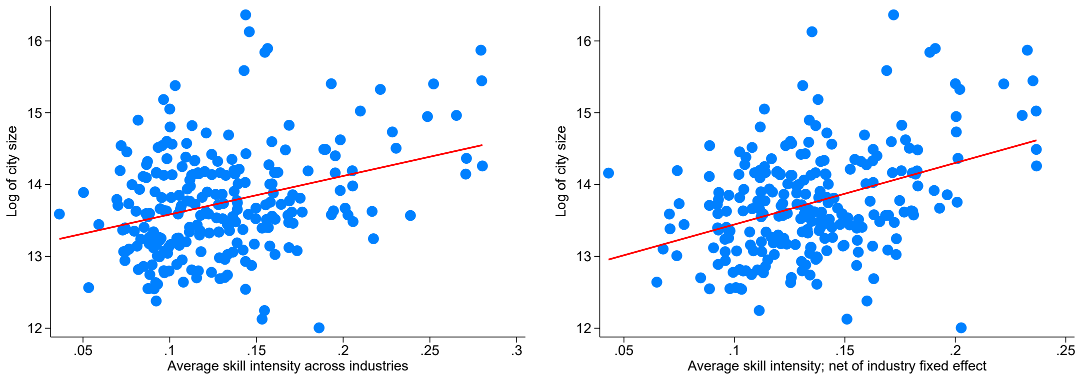


Figure 1.3: Quality Distribution in Large vs. Small Cities, by Sector

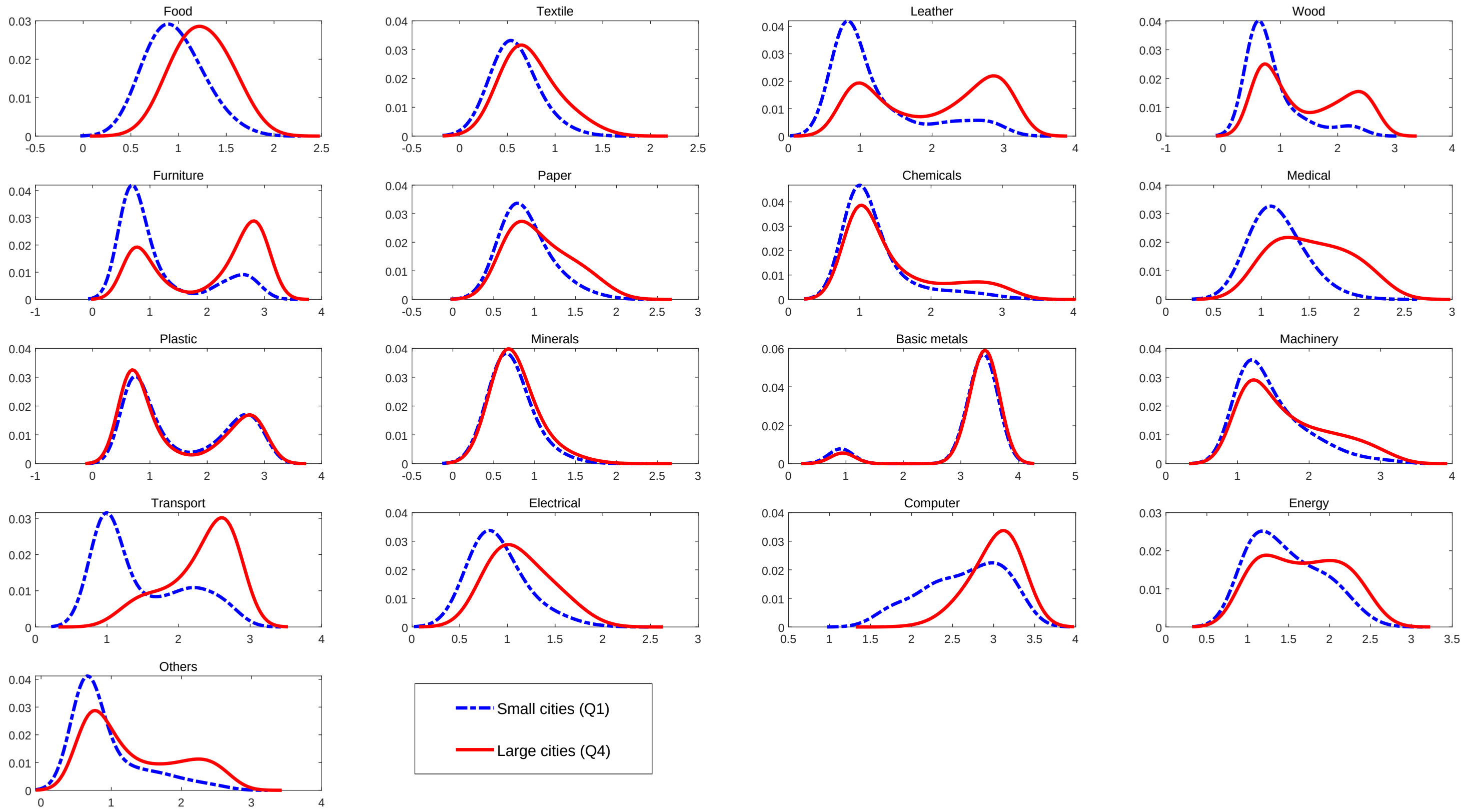


Figure 1.4: City Size Distribution, Model and Data

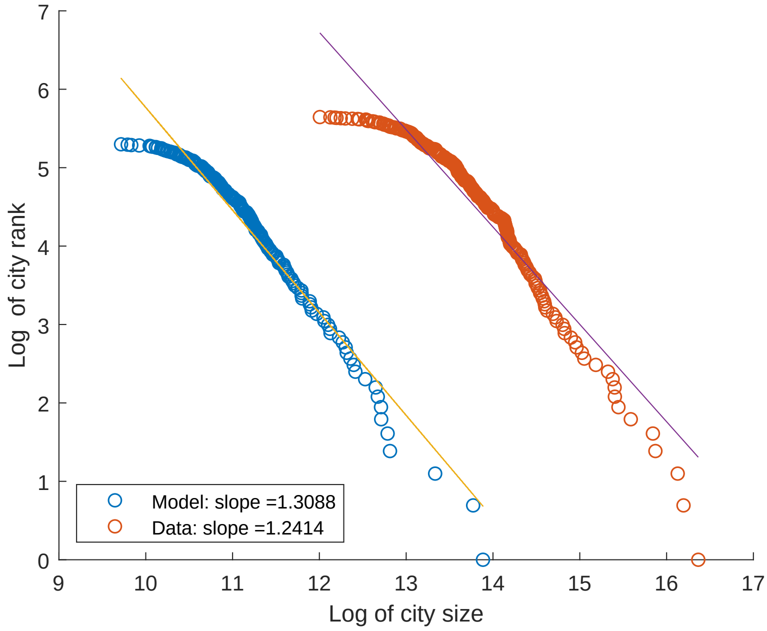


Table 1.1: Stylized Fact 1 - City Size and Quality

Dependent Variable	$\ln Prices$	$\ln\ Market\ Shares$	$Quality\ (Khdwl)$
	(1)	(2)	(3)
$\ln City\ Size$	0.070***	0.028	0.066***
	(0.017)	(0.028)	(0.025)
$\ln Market\ Access$	0.033	0.575***	0.756***
	(0.068)	(0.010)	(0.115)
$Export\ Status$	0.107***	0.671***	0.397***
	(0.019)	(0.034)	(0.022)
City FE $\times$Time trend	Yes	Yes	Yes
Product-Year FE	Yes	Yes	Yes
City Clustered SE	Yes	Yes	Yes
R-squared	0.861	0.389	0.908
Obs	313,242	313,407	313,240

Notes: $Quality\ (Khdwl)$ is the quality estimation based on Khandelwal (2010). $(\ln City\ Size)$ is the log of employment size in a prefecture. $(\ln MarketAccess)$ is the log of the sum of prefectures' GDP per capita weighted by trade costs. $ExportStatus$ is an exporter dummy. "Product" is defined as five-digit Chinese Product Classification. All regressions include a constant term. * p<0.1, ** p<0.05, *** p<0.01.

Table 1.2: Stylized Fact 2 - Productivity and Quality

Dependent Variable	$\ln Prices$	$\ln\ Market\ Shares$	$Quality\ (Khdwl)$
	(1)	(2)	(3)
TFP	0.106***	0.317***	0.410***
	(0.007)	(0.008)	(0.006)
$Export\ Status$	0.116***	0.638***	0.370***
	(0.019)	(0.032)	(0.021)
$\ln Market\ Access$	-0.056	0.295***	0.399***
	(0.064)	(0.110)	(0.104)
City FE×Time Trend	Yes	Yes	Yes
Product-Year FE	Yes	Yes	Yes
City Clustered SE	Yes	Yes	Yes
R-squared	0.863	0.426	0.934
Obs	217,749	217,750	217,748

Notes: $Quality\ (Khdwl)$ is the quality estimation based on Khandelwal (2010). $(\ln City\ Size)$ is the log of employment size in a prefecture. $(\ln MarketAccess)$ is the log of the sum of prefectures' GDP per capita weighted by trade costs. $ExportStatus$ is an exporter dummy. "Product" is defined as five-digit Chinese Product Classification. All regressions include a constant term. * p<0.1, ** p<0.05, *** p<0.01.

Table 1.3: Summary Statistics

	$\ln Value-added$			$\ln Employment$			$\%\ Skilled\ Workers$				
Sector	Mean	$Q1$	$Q4$	Mean	$Q1$	$Q4$	Mean	$Q1$	$Q4$	% in Big Cities	Obs
Food	9.89	9.12	10.92	4.74	4.04	5.57	11.30	5.40	22.10	20.90	6,712
Textile	9.80	9.18	10.57	5.01	4.36	5.67	3.80	1.80	8.00	20.20	29,948
Leather	9.90	9.24	10.69	5.20	4.58	5.89	3.30	1.60	6.90	19.90	5,055
Wood	9.52	8.92	10.28	4.60	4.08	5.15	5.50	2.50	11.10	13.50	3,880
Furniture	9.77	9.14	10.53	4.79	4.22	5.46	6.00	3.00	12.20	31.80	2,424
Paper	9.64	9.04	10.47	4.60	4.03	5.30	6.50	3.10	13.30	31.00	12,413
Chemicals	9.93	9.21	10.88	4.32	3.69	5.11	11.90	5.60	23.50	22.30	15,969
Medical	10.12	9.31	11.08	4.87	4.25	5.58	22.70	11.90	38.60	25.70	3,801
Plastic	9.66	9.06	10.45	4.50	3.91	5.19	7.00	3.40	14.30	28.00	12,902
Minerals	9.78	9.13	10.60	4.83	4.20	5.48	6.40	2.90	13.30	17.70	16,164
Basic metals	9.96	9.23	10.93	4.50	3.91	5.22	7.40	3.60	15.00	25.50	20,518
Machinery	9.68	9.08	10.51	4.55	3.99	5.22	10.5	5.00	20.60	24.70	24,953
Transport	9.97	9.24	10.95	4.79	4.19	5.58	10.80	5.20	21.20	31.10	9,365
Electrical	9.96	9.25	10.92	4.67	4.04	5.42	10.3	5.00	20.50	30.80	12,781
Computer	10.10	9.31	11.20	5.04	4.36	5.94	13.50	6.10	30.50	38.80	10,058
Energy	10.46	9.42	11.57	5.39	4.53	6.15	22.2	12.70	34.50	15.50	5,066
Others	9.67	9.08	10.42	4.98	4.30	5.69	4.50	2.10	9.70	20.80	3,825

Notes: $\ln Value-added$ is measured in thousands of RMB ("qian yuan"). The data is from the Annual Survey of Industrial Firms in 2004.

Table 1.4: Quality Choices in Simulated Equilibrium

Dep. variable	$Quality\ Choices$	
	Full Model	W/O Sorting
$\ln City\ Size$	0.094***	0.049***
	(0.001)	(0.000)
Sector FE	Yes	Yes
Obs	85,000	85,000
R-squared	0.540	0.979

Table 1.5: Goodness of Fit for Targeted Moments

Moments		Q1	Q2	Q3	Q4	P90
		Quartiles & Percentiles				
Mean skill intensity	Model	0.852	1.018	0.964	1.166	-
	Data	0.961	0.973	0.988	1.080	-
Mean value-added	Model	0.998	0.961	1.010	1.036	-
	Data	0.993	0.997	0.997	1.013	-
Value-added share	Model	0.128	0.104	0.132	0.637	-
	Data	0.209	0.207	0.292	0.293	-
Firm size (revenue)	Model	0.397	0.103	0.139	0.114	0.247
	Data	0.250	0.250	0.250	0.150	0.100

Table 1.6: Counterfactual - Relax Land Use Regulations by 20%

City quartile	Δq	Δw_s	Δw_u	Δp_H^D	Δp_H^S	ΔP	ΔW	$\Delta \tilde{W}$
	Change in variables (%)							
Q1	-	-8.3	-0.3	-47.9	2.1	-	-	-
Q2	-	-9.0	-0.1	-49.9	0.8	-	-	-
Q3	-	-9.7	-0.1	-51.3	0.8	-	-	-
Q4	-	-11.7	0.6	-56.2	-2.3	-	-	-
Overall	5.5	-9.7	0.0	-51.3	0.4	-18.7	20.0	6.2

2 Information Frictions, Pro-Competitive Effects, and the Search for Quality

2.1 Introduction

Goods are of inferior quality in remote locations. Two hypotheses may explain the persistence of low-quality products in secluded regions. First, high transportation costs or ice-berg trade costs in general prevent firms from upgrading their product quality, because they face poor market access and hence cannot afford to upgrade product quality which requires additional costs. This is often the focus of the studies on quality in the international trade literature. Second, perhaps less studied, information frictions may also prevent outside firms from competing in distant locations. Because it is costly for outside firms producing higher-quality goods to acquire information regarding the market conditions there, they may not enter and compete in such locations. As a result, lower-quality goods sold by local producers may survive and prevail in such markets.

In this paper, we study the latter hypothesis, that is how information frictions matter in determining quality of tradable products across space, through the lens of a sequential search model with trade (Allen, 2014). To do so, we introduce monopolistic competition, firm heterogeneity, and quality upgrading into (Allen, 2014). In our model, heterogeneous firms must search for market conditions elsewhere in order to decide whether to enter and compete in distant locations. As search is costly, firms hold a reservation strategy in that they will only enter a location if the market condition there is lucrative enough to recover the fixed cost of search. This is different from the setup in Melitz (2003), in which firms with constant marginal cost will always sell their goods to each destination. In particular, the reservation strategy of the firms is summarized by a reservation price index weighted by transportation cost. This index summarizes the market conditions of respective locations. A higher index would imply that the ideal price index in the region is high and hence a less competitive market either because the firms there are less productive or that the product quality there is lower. As such, firms stand to make a higher profit in such a destination.

We derive several key predictions from our model, given a positive search technology shock, such as building of information and communications technology (ICT) infrastructures, that reduces the information frictions in searching for market conditions. First, as the fixed cost of search decreases, firms would hold a higher reservation price index. The intuition is that as the cost of search decreases, firms would want to search more intensively for a lucrative destination market. As such, firms are more likely to enter the least competitive regions which are usually remote locations. Second, as the fixed cost of search decreases, firms would upgrade the quality of their products. Intuitively, information frictions would restrict the extent of the market and hence the market access that firms have. The prospective profits of firms are restricted in the presence of higher search cost. With lower information frictions and larger market access, firms would be able to earn a higher profit and be able to afford costly quality upgrading. Third, a more productive firm would choose a higher reservation price index as well as a higher production quality. This is because a more productive firm can afford to search more intensively and pay the fixed costs for quality upgrading. As a result, they are more likely to enter less competitive regions and they specialize in higher-quality products.

Together, these qualitative results deliver some novel implications. Our model implies that decreasing information frictions would lead to more competition in previously less-competitive regions, and that this pattern is more pronounced for more productive firms. This is the pro-competitive effect that we want to capture in our analysis. Building of ICT infrastructure strengthens spatial competition because more productive firms would have more information about remote locations. Therefore their products have a higher chance to penetrate in those markets. In particular, these firms are not just more competitive in terms of their cost efficiency. They are also more competitive in the sense that the quality of their products are higher. Ceteris paribus, our paper implies that building of ICT infrastructure would result in spatial penetration of cheaper and higher-quality goods. In addition, firms across space would also upgrade their product quality because their market access is larger given the positive shock to information technology.

Thus, our model features two margins of gains from building ICT infrastructure and reducing information frictions. The first margin is an “extensive margin” in which cheaper or higher-quality products are being sold to more destinations that are previously less competitive. The second margin is an "intensive margin" such that as information technology gives firms a larger market access, they are also able to afford upgrading their production quality which benefits consumer welfare.

Next, we take the model to data and examine the empirical relevance of our theoretical predictions. We first examine one specific prediction that building of ICT infrastructure induces firms to upgrade product quality, through the lens of intra-national trade and spatial variation in ICT development across Chinese cities. Key to our empirical strategy are a set of stringent fixed effects that control for omitted variable bias and an alternative measure of "information access" that address for endogeneity concerns. In general, we find that proliferation of ICT infrastructure in the origin city which supposedly would have reduced information frictions lead to substantial quality upgrading behaviors by firms. The magnitude of the effects are such that doubling the ICT penetration would have induced firm to upgrade quality by 7%. This effect is twice larger if instead the overall ICT penetration across all destination markets are doubled. We are also currently exploiting an instrumental variable that is based on least cost algorithm to provide for further identification. To examine the other predictions on pro-competitive effects, we are now using the Chinese Custom data to provide such evidences.

The rest of the paper is organized as follows. Section 2.2 provides a brief literature review and discusses the contributions of our paper. Section 2.3 describes our model and discuss its implications. Section 2.4 takes the theoretical predictions into the data and qualitatively examines the empirical evidences. Section 2.5 concludes. In future work, we also plan to fully calibrate the general equilibrium and discuss the rich quantitative implications of our model.

2.2 Literature Review

Our work is relevant to two strands of literature. First, our paper contributes to the study of quality specialization in the international trade literature. Previous works have predominantly focused on how firm heterogeneity (Feenstra and Romalis, 2014; Antoniades, 2015; Fan, Li, Xu and Yeaple, 2017) and non-homothetic demand (Hallak, 2010; Fajgelbaum, Grossman and Helpman, 2011, 2015; Dingel, 2017) interact with trade costs in shaping quality specialization across countries. Few of these literature, which are mostly based on general equilibrium frameworks, focus on how information frictions matter. Some exceptions include Chen and Wu (2016), Bai (2018), Bai, Chen, Liu and Xu (2018), and Zhao (2018) which are mostly based on partial equilibrium structural estimations of an industry model and focus more on how asymmetric information affects the quality choice of consumers. Our work, which builds on the general equilibrium framework of Allen (2014), abstracts from asymmetric information and focuses on the supply-side effect that information frictions have on quality upgrading among heterogeneous firms. In particular, our work also studies the pro-competitive effect in the sense that reducing information frictions allows higher-quality firms to penetrate more markets and fosters competition across regions along the quality channel.

In addition, our paper contributes to the study of information frictions and economic activities across space. Prior works in this literature include Allen (2014); Eaton, Eslava, Jinkins, Krizan and Tybout (2014); Arkolakis, Papageorgiou and Timoshenko (2018); Eaton, Jinkins, Tybout and Xu (2018); Steinwender (2018) and Juhsz and Steinwender (2019). Our work directly builds on Allen (2014) and extends the model with quality upgrading of heterogeneous firms. We contribute to this literature in several dimensions. First, our work is capable of studying pro-competitive effect of reducing information frictions, regardless whether there is any quality upgrading. Previous works such as Allen (2014) only study the agricultural sector in perfectly competitive regional markets. Thus, any pro-competitive effect of lowering trade barriers is absent. By introducing monopolistic competition and heterogeneous firms into the model, we will be able to study the pro-competitive effect of reducing information frictions, and this is absent in the current literature to the best of our knowledge. Second, our work complements Allen (2014) in studying an additional quality margin of gains from reducing information frictions. This is important because quality upgrading behavior may not be reflected from changes in trade flows, which is what Allen (2014) focuses on. As such, welfare effects of eliminating information frictions may be understated if the quality margin is not accounted for. Our paper partly contributes to the understanding of this question.

2.3 The Model

In this section, we will lay down the foundation of our model and derive the critical comparative statics. Though we aim to build a general equilibrium model and calibrate the equilibrium, we will only take the equilibrium prices as given in this section, so that we can meaningfully discuss the predictions of the model. The model will be closed later in general equilibrium where we introduce clearing conditions in the goods and factor markets.

2.3.1 Consumers

We follow Allen (2014) and Lucas and Prescott (1974) to consider an economy with a continuum of geographically segregated regions.[13] In each region $j \in \mathcal{J} \equiv [0, 1]$, there are L_j consumers with wage w_j. We assume that the preference of a consumer in region j is a CES function over a continuum of goods sold by firms each indexed by φ,

$$U_j = \left[\int_{\varphi \in \Omega_j} \Phi(\varphi, q)^{\frac{1}{\sigma}} c_j(\varphi)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}.$$

where Ω_j is the set of available varieties in region j, $\Phi(\varphi, q)$ is a demand-shifter for variety φ with quality q, $c_j(\varphi)$ denotes the consumption of variety φ, and σ is the elasticity of substitution.

2.3.2 Firms and Entry

A firm with productivity φ employs labor to produce goods with quality q. The marginal cost is constant but increasing in q, in the sense that it is more costly to produce goods with higher quality. For simplicity, we follow Antoniades (2015) in assuming that a firm φ uses $l = 1/\varphi$ amount of labor and αq units of cost to produce one unit of output with quality q. Furthermore, the firm must pay βq^2 units of fixed cost to upgrade

[13] This is a simplifying assumption to derive intuitive analytical solutions. However, this may resemble reality well, especially given the context of information frictions. One may think of each county in a country as a different region, and therefore there is a large number of regions which can be approximated by the continuum assumption.

their quality. In addition, we assume that trade across space is costly in the sense that firms incur an iceberg transportation cost $\tau_{ij} \geqslant 1$ to ship goods from region i to region j, with $\tau_{ii} = 1$ for all $j = i$. Suppose that a no-arbitrage condition holds, then in region j, the price of goods with quality q produced by a firm φ from region i is

$$p_{ij}(\varphi) = \tau_{ij} p_i(\varphi).$$

Under perfect information, the corresponding price index in region j then becomes

$$P_j = \left[\int_{i \in \mathcal{J}} \int_{\varphi} \Phi_i(\varphi, q) p_{ij}(\varphi, q)^{1-\sigma} dF_i(\varphi) di \right]^{\frac{1}{1-\sigma}} = \left[\int_{i \in \mathcal{J}} \int_{\varphi} \Phi_i(\varphi, q) \left[\tau_{ij} p_i(\varphi, q)\right]^{1-\sigma} dF_i(\varphi) di \right]^{\frac{1}{1-\sigma}}.$$

Thus consumption of region j on goods from region i produced by firm φ is

$$c_{ij}(\varphi, q) = \Phi(\varphi, q) \left[\frac{p_{ij}(\varphi, q)}{P_j} \right]^{-\sigma} \frac{E_j}{P_j} = \Phi(\varphi, q) \left[\frac{\tau_{ij} p_i(\varphi, q)}{P_j} \right]^{-\sigma} \frac{E_j}{P_j}. \tag{2.1}$$

Given our assumption of the geography and demand structure, we can write the profit maximization problem of a firm φ from region i as

$$\max_{p_i, q} \int_{j \in \mathcal{J}} \tau_{ij} p_i(\varphi, q) c_{ij}(\varphi, q) - (\frac{1}{\varphi} w_i + \alpha q) \tau_{ij} c_{ij}(\varphi, q) \, dj - \beta q^2.$$

We can solve this problem step by step. First, we take the quality choice as given and solve for the pricing decision of the firm. This is standard and a firm will simply charge a constant markup over its marginal cost.

$$p_i(\varphi, q) = \frac{\sigma}{\sigma - 1} (\frac{1}{\varphi} w_i + \alpha q)$$

Next, we solve for the quality choice of the firm. Substituting the pricing decision back to the optimization problem, we can first show that the profit function of the firm becomes

$$\pi(\varphi, q) = \int_{j \in \mathcal{J}} \frac{\tau_{ij}}{\sigma - 1} (\frac{1}{\varphi} w_i + \alpha q) \cdot \Phi(\varphi, q) \left[\frac{\tau_{ij} \frac{\sigma}{\sigma-1} (\frac{1}{\varphi} w_i + \alpha q)}{P_j} \right]^{-\sigma} \frac{E_j}{P_j} \, dj - \beta q^2$$

$$
\begin{aligned}
&= \int_{j\in\mathcal{J}} \frac{\Phi(\varphi,q)}{\sigma} (\tau_{ij})^{1-\sigma} \left[\frac{\frac{\sigma}{\sigma-1}(\frac{1}{\varphi}w_i+\alpha q)}{P_j}\right]^{1-\sigma} E_j\, dj - \beta q^2 \\
&= \frac{\Phi(\varphi,q)}{\sigma^\sigma(\sigma-1)^{1-\sigma}}(\frac{1}{\varphi}w_i+\alpha q)^{1-\sigma} \underbrace{\int_{j\in\mathcal{J}} (\tau_{ij})^{1-\sigma} P_j^{\sigma-1} E_j\, dj}_{\text{region i's market access}} - \beta q^2 \\
&= \frac{\Phi(\varphi,q)}{\sigma^\sigma(\sigma-1)^{1-\sigma}} \cdot \tilde{c}^{1-\sigma} \cdot MA_i - \beta q^2
\end{aligned}
$$

where we use $\tilde{c} \equiv w_i/\varphi + \alpha q$ to denote the marginal cost of production. The first-order condition of this problem then entails that the net operational benefit from quality upgrading should be equated with the increase in fixed cost of quality upgrading,

$$
\pi^*(\varphi,q^*)\left[\frac{\partial\Phi(\varphi,q^*)}{\partial q}\frac{1}{\Phi(\varphi,q^*)} + \frac{(1-\sigma)}{\tilde{c}}\frac{\partial\tilde{c}}{\partial q}\right] = 2\beta q^*
$$

where $\pi^* \equiv \sigma^{-\sigma}(\sigma-1)^{\sigma-1}\Phi(\varphi,q)\tilde{c}^{1-\sigma}MA_i$ is the operational profit of the firm excluding the fixed cost of quality upgrading. This first-order condition then pins down the optimal choice of q. Two observations are in order. First, optimal quality choice increases with firm's productivity. This is intuitive as higher productivity of firms leaves more room for costly quality upgrading. Second, a larger market access would also induce a firm to upgrade the quality of its products. Both claims are summarized in Proposition 2.1.

Proposition 2.1. *Under perfect information, the optimal choice of quality is determined by a firm's productivity φ and its location i. Furthermore, a firm will choose a higher product quality if it is more productive or has access to a larger market, i.e., $\frac{\partial q^*}{\partial z} > 0$ and $\frac{\partial q^*}{\partial MA_i} > 0$.*

Proof. Obvious. Invoking the Implicit Function Theorem and the second-order condition of profit maximization will show the positive signs of comparative statics.

Finally, we assume that firms are heterogeneous in their productivity φ, à la Melitz (2003). In particular, we assume that a firm in region i draws its productivity φ from a region-specific Pareto distribution $F_i(\varphi) = 1 - \varphi^{\theta_i}$ with shape parameter θ_i. To enter into production, a firm also has to incur a fixed cost of production f_p. Furthermore, a firm also has to incur a fixed cost of exporting f_e, in order to ship its product to another region.

2.3.3 Information Frictions and Search Process

We introduce information frictions into the model using a setup that is similar to the agriculture settings in Allen (2014). First, we assume that firms only have complete information with regard to the local market conditions of the region they locate in, such that all firms from region $i \in J$ know about the local price index P_i. However, for the other regions, firms have to engage in a sequential search process in order to learn about the market conditions.

Following Allen (2014), we assume that there are two types of information frictions: the fixed costs of each search and search probability. In each period, a firm will pay a search cost $f_i^S > 0$ to randomly draw a region $k \in J$, and learn about the price index there. After learning the price index of one destination market, the firm decides the amount of goods sold to that market and whether to stop search or not. If it decides to search again, the firm draws another region again (with replacement) and learn the price index there. This sequential search process continues until the firm stops searching. Search costs f_i^S are paid in every period.

In addition, we assume that there is a search probability $s_{ij} > 0$, which is defined as the probability that a producer from region i draws and learns information about region j in each search. For simplicity, the search process is with replacement so that the search probability for each destination remains constant during the search. In addition, the search probability is assumed to be the same for all producers from region $i \in J$.

2.3.4 Optimal Search and Production

In this section, we characterize the timing of the events and the subsequent searching behaviour. Firms in region i can sell their products home or to other regions. Specifically, a firm in region i can ship their goods to region j after incurring ice-berg transportation costs $\tau_{ij} \geqslant 1$, with $\tau_{ii} = 1$ for all $j = i$.

The timing of the event is such that firms will choose their production plans and the total mass of markets to enter before they begin to search and trade. Then, they will search for the market conditions elsewhere and decide whether to sell their manufactures in other regions.[14] We summarize the timing of events as follows. Building on Allen (2014), we

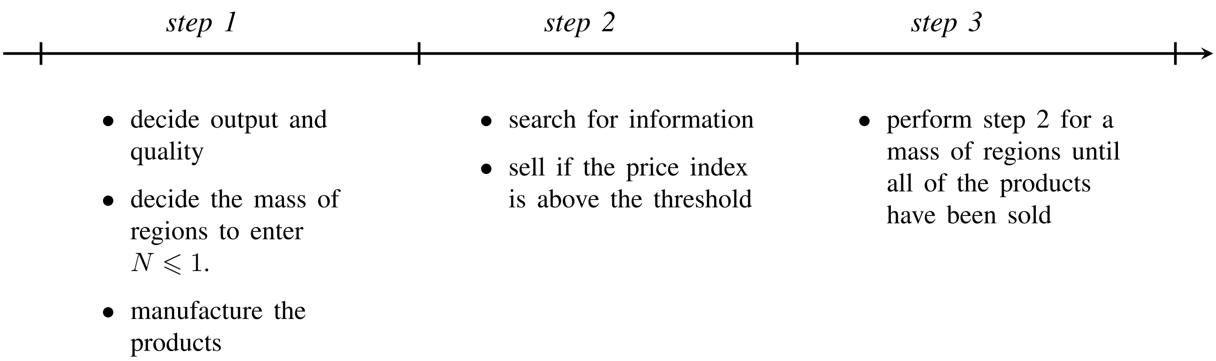


Figure 2.1: Timeline of events

assume that firms engage in a sequential search process to sell the goods that they have manufactured. If there is perfect information and the search costs are zero, then firms will keep searching for prices of all destinations until they sell all of their manufactures. With non-zero search costs, the producers cannot indefinitely search for information across all regions. They face a trade-off such that they may draw a region with larger market but in the meantime they incur a cost to search for an additional time.

Conditional on a firm φ in region i decide to enter and compete in region j, we know that its profit, excluding the fixed cost in producing quality q, from this trade is

$$\pi_{ij}(p_j, \varphi) = \frac{\Phi(\varphi, q)}{\sigma^\sigma(\sigma-1)^{1-\sigma}} \left[\frac{\tau_{ij}(\frac{1}{\varphi}w_i + \alpha q)}{P_j}\right]^{1-\sigma} E_j \tag{2.2}$$

If expenditure E_j across all regions are symmetric (we will discuss heterogeneous expenditures later), then the magnitude of π_{ij} is directly proportional to $\tilde{P}_j \equiv P_j/\tau_{ij}$ which represents the degree of competition in region j weighted by the transportation cost in shipping goods from i to j. If $\tilde{P}_j$ is small, it implies that either the competition in region j is high because most firms in this region are low-cost producers manufacturing high-quality products, or that the transportation cost in shipping to region j is high. Either way, the firm φ from region i would find it difficult to sell and compete in region j. Hence, the profit π_{ij} from this trade is small when $\tilde{P}_j$ is small.

[14]This setting is similar to the "produce then trade" setup in Angeletos and La'O (2013).

Assuming that firms have perfect foresight and know their prospective profit if they decide to enter a market, then a firm's decision can be characterized by a threshold strategy, such that they would enter a region j if and only if they have drawn a large enough $\tilde{P}_j$ which implies that either this market is less competitive or that geography makes shipping to this destination more attractive. Given our descriptions of the timeline, we can write down the value function as follows,

$$V_i(\tilde{P}_j;\varphi) = \max_{sell,search}\left\{\pi_{ij}(\tilde{P}_j,\varphi), \int_{\tilde{P}_k^{min}}^{\tilde{P}_k^{max}} V_i'(\tilde{P}_k',\varphi)dF^i(\tilde{P}_k') - f_i^S\right\} \tag{2.3}$$

If we denote the reservation transportation-weighted price index of the firm as $\tilde{P}_j^R$, then for the firm to be indifferent in selling the manufactures and search for another period, it must be the case that

$$\begin{aligned}\pi_{ij}(\tilde{P}_j^R,\varphi) &= \int_{\tilde{P}_k^{min}}^{\tilde{P}_k^{max}} V_i'(\tilde{P}_k',\varphi)dF^i(\tilde{P}_k') - f_i^S\\ &= \int_{\tilde{P}_k^{R}}^{\tilde{P}_k^{max}} \pi_{ik}(\tilde{P}_k',\varphi)dF^i(\tilde{P}_k') + \int_{\tilde{P}_k^{min}}^{\tilde{P}_k^{R}} V_i'(\tilde{P}_k',\varphi)dF^i(\tilde{P}_k') - f_i^S\end{aligned}$$

which is equivalent to

$$\begin{aligned}f_i^S &= \int_{\tilde{P}_k^{R}}^{\tilde{P}_k^{max}} \pi_{ik}(\tilde{P}_k',\varphi)dF^i(\tilde{P}_k') + \int_{\tilde{P}_k^{min}}^{\tilde{P}_k^{R}} V_i'(\tilde{P}_k',\varphi)dF^i(\tilde{P}_k') - \int_{\tilde{P}_k^{R}}^{\tilde{P}_k^{max}} \pi_{ij}(\tilde{P}_j^R,\varphi)dF^i(\tilde{P}_k') - \int_{\tilde{P}_k^{min}}^{\tilde{P}_k^{R}} \pi_{ij}(\tilde{P}_j^R,\varphi)dF^i(\tilde{P}_k')\\ &= \int_{\tilde{P}_k^{R}}^{\tilde{P}_k^{max}} \pi_{ik}(\tilde{P}_k',\varphi) - \pi_{ij}(\tilde{P}_j^R,\varphi)\, dF^i(\tilde{P}_k') + \int_{\tilde{P}_k^{min}}^{\tilde{P}_k^{R}} V'(\tilde{P}_k',\varphi)\, dF^i(\tilde{P}_k') - \int_{\tilde{P}_k^{min}}^{\tilde{P}_k^{R}} \pi_{ij}(\tilde{P}_j^R,\varphi)\, dF^i(\tilde{P}_k')\\ &= \int_{\tilde{P}_k^{R}}^{\tilde{P}_k^{max}} \pi_{ik}(\tilde{P}_k',\varphi) - \pi_{ij}(\tilde{P}_j^R,\varphi)\, dF^i(\tilde{P}_k')\end{aligned} \tag{2.4}$$

The last equality is true because if $\tilde{P}_k' \in [\tilde{P}_k^{min}, \tilde{P}_k^R]$, then the firm will not enter and compete in market k. In this case, the value of searching again is the same regardless the value of $\tilde{P}_k'$. In particular, the following is true,

$$V'(\tilde{P}_k',\varphi) = V'(\tilde{P}_k^R) = \pi_{ij}(\tilde{P}_j^R,\varphi).$$

which implies that

$$\int_{\tilde{P}_k^{min}}^{\tilde{P}_k^R} V'(\tilde{P}_k', \varphi)\, dF^i(\tilde{P}_k') = \int_{\tilde{P}_k^{min}}^{\tilde{P}_k^R} \pi_{ij}(\tilde{P}_j^R, \varphi)\, dF^i(\tilde{P}_k').$$

Given the model parameters, the reservation price index weighted by transportation costs of a firm φ is then pinned down by equation (2.3.4), which is reproduced in below

$$f_i^S = \int_{\tilde{P}_k^R}^{\tilde{P}_k^{max}} \pi_{ik}(\tilde{P}_k', \varphi) - \pi_{ij}(\tilde{P}_j^R, \varphi)\, dF^i(\tilde{P}_k').$$

The implications of this equation are as follows. First, if the fixed cost of searching decreases, the reservation price index weighted by transportation costs will increase, ceteris paribus. This is because as the fixed cost per search decreases, firms are more willing to search for one more time if they draw a lower $\tilde{P}_k$ which implies that the competition in region k is fierce or that the location is remote. As a result, firms are now more likely to enter and compete in a location with higher $\tilde{P}$, in which there is less competition or that the region is nearby the location of the firm. This is the pro-competitive effect of reducing information frictions that we want to document. Second, if the a firm is more efficient, then its reservation price index weighted by transportation costs will increase. The intuition is that as the firm is more efficient, then it can afford to search for a less competitive place more since it can make a higher profit in that destination. We summarize these discussions in Proposition 2.2.

Proposition 2.2. *Under information frictions, a firm is more likely to enter a less competitive market weighted by transportation cost if the fixed cost of search is lower or if the firm is more productive, in the sense that,* $\frac{\partial \tilde{P}_k^R}{\partial f_i^S} > 0$ *and* $\frac{\partial \tilde{P}_k^R}{\partial \varphi} > 0$.

Proof. The proof is provided jointly in Proposition 2.3.

2.3.5 Quality Upgrading Under Information Frictions

We now complete the descriptions of firm behavior in the presence of information frictions. Given that searching for information is costly, firms do not enter all destination markets. In particular, the probability that a firm φ has not drawn a transportation-cost augmented price index larger than its reservation after $m-1$ searches is $F^i(\tilde{P}_k^R)^{m-1}$, and the probability that the firm reaches market k after m searches is $F^i(\tilde{P}_k^R)^{m-1}s_{ik}$. Hence, we can write the total profit of the firm as

$$\begin{aligned}
\pi(p_i,\varphi) &= \int_{\tilde{P}_k^R}^{\tilde{P}^{max}} \sum_{m=1}^{\infty} \left[\left(\pi_{ik}(p_i,\varphi) - mf_i^S\right) F^i(\tilde{P}_k^R)^{m-1} s_{ik} \right] \, d\tilde{P}_k - \beta q^2 \\
&= \int_{\tilde{P}_k^R}^{\tilde{P}^{max}} \sum_{m=1}^{\infty} \left[\left(\frac{\Phi(\varphi,q)}{\sigma^\sigma(\sigma-1)^{1-\sigma}} \left[\frac{\tau_{ij}(\frac{1}{\varphi}w_i + \alpha q)}{P_k} \right]^{1-\sigma} E_k - mf_i^S \right) F(\tilde{P}_k^R)^{m-1} s_{ik} \right] \, d\tilde{P}_k - \beta q^2 \\
&\equiv \kappa_i(\varphi,q) \int_{\tilde{P}_k^R}^{\tilde{P}^{max}} \sum_{m=1}^{\infty} \left[\left((\tau_{ik})^{1-\sigma} P_k^{\sigma-1} E_k \right) F^i(\tilde{P}_k^R)^{m-1} s_{ik} \right] \, d\tilde{P}_k - \int_{\tilde{P}_k^R}^{\tilde{P}^{max}} \sum_{m=1}^{\infty} mf_i^S F^i(\tilde{P}_k^R)^{m-1} s_{ik} \, d\tilde{P}_k - \beta q^2 \\
&= \underbrace{\kappa_i(\varphi,q) \int_{\tilde{P}_k^R}^{\tilde{P}^{max}} \frac{s_{ik}}{1-F^i(\tilde{P}_k^R)} \cdot \tilde{P}_k^{\sigma-1} E_k \, d\tilde{P}_k}_{\text{market access of firm } \varphi \text{ in region } i} - \underbrace{\int_{\tilde{P}_k^R}^{\tilde{P}^{max}} \sum_{m=1}^{\infty} mf_i^S F^i(\tilde{P}_k^R)^{m-1} s_{ik} \, d\tilde{P}_k}_{\text{total expected search costs}} - \beta q^2 \\
&= \kappa_i(\varphi,q) \cdot \widetilde{MA}_{i,\varphi} - \underbrace{\frac{F^i(\tilde{P}_k^R)}{1-F^i(\tilde{P}_k^R)} f_i^S}_{\text{total expected search costs}} - \beta q^2
\end{aligned}$$

where if necessary we can interchange the integral and summation using Tonelli's Theorem, and we simplify the expression using the sum of a geometric series. We can apply Tonelli's Theorem because the integrand is non-negative, and that the domain is a closed interval over $\mathbb{R}$ which is Lebesgue-measurable. $\kappa_i(\varphi,q) \equiv \frac{\Phi(\varphi,q)}{\sigma^\sigma(\sigma-1)^{1-\sigma}}(\frac{1}{\varphi}w_i + \alpha q)^{1-\sigma}$ is a term that summarizes the demand shifter q and cost conditions (φ, w_i) of firm φ in region i. $\widetilde{MA}_{i,\varphi}$ is the information-augmented market access of firm φ in region i. Note that this is different from the market access expression in the absence of information frictions. First, under perfect information, firms do not face any cost in accessing information in other markets. In that case, their reservation strategies are simply that they choose to sell in the least-competitive market (weighted by transportation cost). In contrast, under information frictions, firms face a fixed cost per search. They cannot afford to search indefinitely and have a lower reservation price index. Firms are now willing to enter a more competitive market given the price index there is no smaller than its reservation strategy. In this sense, markets with less competition are protected by the presence of information frictions in that firms now are more likely to accept a draw with higher competition and hence are less likely to enter less-competitive markets.

Second, under information frictions, the information-augmented market access measure $\tilde{MA}_{i,\varphi}$ is now firm-region specific. This is because with information frictions, more efficient firms would search more intensively, in the sense that their reservation price index is higher. As a result, although more efficient firms would only consider a narrower set of markets, these destinations are markets with less competition. Hence, more efficient firms stand to earn a larger profit from these trades *on average*. Consequently, we show in Lemma 2.1 that the actual market access of a firm with higher reservation price index is larger. In contrast, without information friction, the market access measure is only region specific, and all firms from the same region would have the same access because firm heterogeneity does not affect search behavior.

Lemma 2.1. *In the presence of information frictions, a firm with a larger reservation price index weighted by transportation cost has a larger market access, i.e.,* $\frac{\partial \widetilde{MA}_{i,\varphi}}{\partial \tilde{P}_k^R} > 0$.

Proof. See Appendix B.1

Given the profit expression, the optimal choice of quality q^* is then characterized by the first-order condition,

$$\widetilde{\pi}^*(\varphi, q^*) \left[\frac{\partial \Phi(\varphi, q^*)}{\partial q} \frac{1}{\Phi(\varphi, q^*)} + \frac{1-\sigma}{\tilde{c}} \frac{\partial \tilde{c}}{\tilde{q}} \right] = 2\beta q^* \tag{2.5}$$

where $\widetilde{\pi}^*(\varphi, q^*) \equiv \sigma^{-\sigma}(\sigma-1)^{\sigma-1}\Phi(\varphi, q)\tilde{c}^{1-\sigma}\widetilde{MA}_{i,\varphi}$. Notice that equation (2.5) is similar to the first-order condition in the perfect information case, except that the operational profit expression now contains an information-augmented market access $\widetilde{MA}_{i,\varphi}$. Intuitively, similar comparative statics should hold in the presence of information frictions, in that a larger market access or a positive productivity shock would lead to quality upgrading by firms. In this sense, this is the quality upgrading behavior that we want to capture, as a smaller fixed cost of search f_i^S would lead to a higher reservation price index and hence would induce the firms to upgrade quality. However, the comparative statics are now more complicated as firm optimization are jointly determined by search behavior (2.3.4) and quality decision (2.5). We now formally state the comparative statics under information frictions in Proposition 2.3,

Proposition 2.3. *Building of ICT infrastructure, which lowers the fixed cost of search* f_i^S, *leads to a higher reservation price index and quality upgrading of firms, i.e.,* $\frac{\partial \tilde{P}_k^R}{\partial f_i^S} > 0$ *and* $\frac{\partial q^*}{\partial f_i^S} > 0$. *Similarly, a more efficient firm would also choose a higher quality and a higher reservation price index, i.e.,* $\frac{\partial \tilde{P}_k^R}{\partial \varphi} > 0$ *and* $\frac{\partial q^*}{\partial \varphi} > 0$.

Proof. See Appendix B.2.

2.3.6 General Equilibrium

So far our discussions of the model have been taking equilibrium prices as given. We will now briefly define a general equilibrium. More details can be found in the Appendix. In the future, we will also fully utilize the general equilibrium and discuss the quantitative implications of our model.

1. Consumers optimize and their demand is given by equation (2.1).

2. Firms optimize their production (2.3.4) and search (2.3.4) behavior.

3. Goods and factor markets are cleared.

4. Firms earn zero profit in expectation.

5. Firms are fully rational and have correct beliefs on the distribution of prices.

2.4 Intra-national Empirical Evidences

We now bring the theoretical predictions of our model to the data. In particular, we use several datasets to test the qualitative predictions of our model. First, we examine the prediction that building of ICT infrastructure would induce quality upgrading behaviors by firms across Chinese cities. This is done by combining two firm-level datasets in China, which are the Annual Survey of Industrial Firms (ASIF) and the Chinese Industrial Firms Product Quantity Dataset. Our empirical strategy relies on a battery of stringent fixed effects that address omitted variable bias. To address concerns that firms may choose to upgrade product quality because they become more productive given the use of ICT, we further exploit the variation of ICT infrastructure in regions other than the location of firms to identify the effect of decreasing fixed cost of search on quality upgrading. To further address endogeneity concerns that construction of ICT infrastructure may be endogenous to location characteristics, we are now constructing a separate set of regressions that utilize the opening of airports and alternative opening plans based on an least-cost algorithm as an instrumental variable.

Second, as our Chinese firm-level data do not contain any intra-national trade information, we will not be able to test the predictions concerning whether firms have sold more to less competitive markets. To provide empirical evidences on these predictions, we exploit the Chinese Custom data which contains product-level unit value information as well as destinations that the products are being sold to. For this design, we are relying on the international variation in ICT infrastructure to corroborate the empirical hypotheses. There are also less endogeneity concerns because construction of ICT infrastructures across nations is not coordinated by any single organization. We are currently constructing the variables for this regression.

2.4.1 Data

Next, we will describe the data sources used in the empirical studies. We choose to focus on the time period from 2001 to 2007, during which China experiences a drastic increase in telecommunication infrastructure construction and mobile phone penetration. Another reason that we choose to focus on this time period is that, this is the longest time span that our datasets permit, given what we have in hand.

Data on Firm-level Production. The firm-level data is from the A*nnual Survey of Industrial Firms (ASIF)*, constructed by the National Bureau of Statistics in China (NBSC). The sample period is from 2001 to 2007. The *ASIF datase*t covers manufacturing firms with gross sales more than 5 million RMB of that year, both state-owned and non-state-owned. It reports detailed firm-level financial and production information such as gross output, value added, capital stock, number of employees, wage payable and supplementary benefit payable. The approach we use to construct the panel data is largely adapted from Brandt, Van Biesebroeck and Zhang (2012). We match firms by their legal ID, firm name, legal person representative, telephone number as well as administrative area (prefecture) code sequentially.

Data on Physical Quantities and Unit Value. The product level information is obtained from the *Chinese Industrial Firms Product Quantity Dataset* collected by the NBSC from 2001 to 2007. This dataset provides information on physical output quantity at give-digit Chinese Product Code (CPC) level, which enables us to extract market share as well as price information for single product firms if the sales information for this firm in the *ASIF dataset* is available. Because the sales information is only provided at firm-level in *ASIF dataset*, we keep firms of single product in the physical quantity dataset and merge with the *ASIF dataset* according to firm identity code (ID) and firm name in each year.

Data on ICT Infrastructure. The third dataset is *China City Statistical Year Books*. "City" here refers to a prefecture, which is a geographically administrative unit in China. There were 265 prefectures in year 2000 and 283 in year 2007. The city statistical yearbooks report Chinese telecommunication development at prefecture level, including the number of broadband internet subscribers and the number of mobile phone users. On top of telecommunication information, other information such as GDP, GDP per capita, total industrial outputs are also reported in this yearbook. Figure 2.2 plots the spatial variation in the number of mobile phone owners and broadband internet subscribers. The first four Panels A, B, C, D plot the levels of telecommunication development at prefecture level in China in year 2001 and 2007. Panel E and F show the spatial distribution of telecommunication growth rate between 2001-2007 at prefecture level. For the two telecommunication measures, we find coastal areas exhibit higher level relative to those in inland areas in both year 2001 and 2007. Though, inland areas experience higher growth rates than those of coastal areas in telecommunication development.

[Insert Figure 2.2 here]

2.4.2 Empirical Strategies

The theoretical framework predicts that the reduction in information frictions would lead to larger market access and hence induce quality upgrading by firms. In the data, we proxy the decrease in information frictions by using the log of number of broadband internet and mobile phones users as proxies. The results are similar if we use the percentage of users as proxies instead. The baseline regression we run is

$$Q_{gjist} = \beta_1 \ln TeleCom_{it} + \beta_2 \ln MA_{it} + \mathbf{\Lambda}'\mathbf{F}_{jst} + D_g + D_t + D_i \times t + \varepsilon_{gjist},$$

where dependent variable Q_{gjist} represents the quality for product g produced by firm j of industry s in prefecture i at time t. We use three different measures of quality separately to proxy for product quality, which are prices, market shares and the estimated quality. We compute unit value as price of a product by dividing the sales with quantities of firm j in year t. As for market shares of a firm, we divide the output quantities of a firm's product by the sum of quantity within each five-digit product-code category. The assumption for the first two measures is that within a product group[15], products with higher prices, or products that have captured larger market shares have higher quality. We also apply the approach of Khandelwal (2010) to provide an estimate of product quality, which explores the information on both prices and market shares of a product to estimate quality based on a nested logit model.

Independent variable $\ln TeleCom_{it}$ measures the development of telecommunication in prefecture i where a firm locates at time t. We use the number of mobile phone owners and broadband internet subscribers as proxies for this variable. Given our specification, we are partly exploiting the time series variation in the number of internet or cell users in each city in identifying the effect of ICT infrastructure on product quality. The results are qualitatively similar if we use percentage of internet or mobile users instead, which may arguably be better proxies for ICT penetration ratio in a city. These results are available upon requests.

To isolate the effects of prefecture-level information frictions on product quality, we control for a transportation cost based market access measure MA_{it}. This is to distinguish between the effects of transportation cost based market access and information-based market access measures. We construct market access MA_{it} of a city i following the approach in Donaldson and Hornbeck (2016) and Allen and Atkin (2016). Following the literature, we define market access as the sum of total income in all prefectures weighted by respective trade costs between the prefectures and the corresponding location that a firm locates in. We extract the measures for trade costs from Ma and Tang (2019) which utilizes various transportation modes and realistic geography in computing the trade cost matrix. We expect the coefficient of market access to be positive, because

[15] The product group is five-digit Chinese Product Code, which is the finest disaggregated level we have.

firms closer to higher-income locations may choose a higher product quality to cater the non-homothetic demand in those location (Fajgelbaum, Grossman and Helpman, 2011, 2015; Dingel, 2017). In addition, there is also a concern with transportation cost based market access, because in our model this will also lead to quality upgrading behavior by firms.

We include an additional vector of control $\mathbf{\Lambda}'\mathbf{F}_{jt}$ to control for firm-level characteristics such as productivity, employment size, and export status.[16]. Firms with higher productivity find it more affordable to upgrade product quality whereas employment size is used to proxy for the levels of economies of scales. We expect the coefficients for these two variables to be positive, because firms of higher productivity and scale economies should find it more appealing to produce higher quality goods. Finally, export status is a dummy variable that indicates whether a firm exports in a given year. This exporter dummy controls the possible effect through non-homothetic demand from high-income foreign countries. To further control for omitted variables, we also include product fixed effects and time fixed effects which subsume all the product-specific or time-specific determinants of product quality. Finally, $D_i \times t$ are city-specific linear time trends that subsume any time-series persistence in some unobservable variables which may change both firms quality choices and ICT infrastructure construction in a prefecture.

To further alleviate the concerns on omitted variable bias, we also run a more stringent specification that includes a full battery of fixed effects which our data permit. The most stringent set of fixed effects would include product-year specific fixed effects and city-product fixed effects. Ideally, product-year fixed effects would have controlled for any product specific economic shocks over time such as technological progress, changes in availability of high-quality inputs, and for times-series variation in consumer demand and preferences for different goods. City-product fixed effects would have controlled for comparative advantage patterns across space (Chor, 2010), which may confound the empirical relationship we are interested in. One possibility is that internet or mobile penetration ratio could be high in the coastal regions which may have comparative advantage in producing higher quality products. Note, however that we do not include these fixed

[16]The productivity estimation adopts the approach of Olley and Pakes (1996)

effects in our specification because the sample size is not large enough to provide meaningful identification for tens of thousands of coefficients.

$$Q_{gjist} = \beta_1 \ln TeleCom_{it} + \beta_2 \ln MA_{it} + \mathbf{\Lambda}'\mathbf{F}_{jst} + D_{gt} + D_i \times t + \varepsilon_{gjist}.$$

Finally, we cluster standard errors at the prefecture-industry level to take into account possible correlated idiosyncratic shocks at the firm level within an industry in a prefecture. The results are similar if we cluster the standard error at the prefecture level.

2.4.3 Baseline Results

We report the baseline findings in Table 2.1 and Table 2.2. Table 2.1 reports the results with product fixed effects and year fixed effects. We run the regression for three different measures of product quality as the dependent variable. Column (1) to (4) report the results using market shares as the dependent variables. In particular, Column (1) and (3) report the first regressions without any fixed effects of city-specific time trends using log of mobile and internet users as a proxy for ICT infrastructure respectively. The results are similar if we use the percentage of internet or mobile users instead.[17] Although the estimates are both economically and statistically significant in the first regression, the coefficients for transportation based market access measures are in the wrong direction. The estimates for market access in the first regression suggest that firms in cities with larger market access would choose a lower production quality. One reason that this may be true is that there could be city-specific time persistence such that market shares of a firm's product become smaller overtime due to technological progress, the entry of new products, or increasing competition. In particular, this trend may be more pronounced in the big cities or coastal areas which usually have a larger market access. In addition, as we are running a spatial regression, there are also concerns such that there may be spatial correlation in various characteristics that may lead to over-significant estimates (Kelly, 2019). As such, it is particularly important to include fixed effects such as product fixed effects, time fixed effects, and city-specific linear time trends.

[17] These results will be included soon in the latest draft.

We run the regressions with corresponding fixed effects and city-specific time trends in Column (2) and (4) of Table 2.1. Several observations emerge from this exercise. First, the coefficient estimates for the main variables of interest are still economically significant thought it is now only statistically significant at 10% level for mobile users in Column (2). Our estimates suggest that when ICT penetration in a city doubles, firms would have on average upgraded their product quality by 10%. Second, we find that the coefficient estimates for market access become both positive and economically and statistically significant in both columns. This corroborates our previous speculation that the negative coefficient estimates in Column (1) and (3) may be due to city-specific time trends or any other omitted variables which are now subsumed by the product fixed effects or time fixed effects. Third, our coefficient estimates for other factors such as employment size or firm productivity remain economically and statistically significant throughout Column (1) to (4), and the estimates are in the intuitive directions.

Column (5) to (12) use alternative variables as proxies for quality. The results remain quantitatively and qualitatively stable except for unit value of products. In this regression, the results are not statistically significant in Column (10) and the quantitative magnitude attenuates across using either mobile or internet users. In addition, the coefficient estimates for market access are also not statistically significant, and they are intuitively in the wrong directions.

To further address for omitted variable bias, we run a more stringent specification in Table 2.2 which includes product-time specific fixed effects. Ideally, we should have included city-product fixed effects which control for comparative advantage patterns. As discussed earlier, we do not enough sample size to meaningful identify the large amount of fixed effects. We will include a set of results with city-industry fixed effects in the future. The results in Table 2.2 suggest that our findings survive a more stringent specification, and the qualitative implications remain stable. Quantitatively, our coefficient estimates for the main variables are slightly attenuated but are still economically and statistically significant. To interpret our estimates, the results in Table 2.2 suggest that doubling the ICT penetration in a city would have induced firms to upgrade product quality by 7%, in comparison to that of 10% in Table 2.1. These results suggest that our estimates in Table

2.1 are slightly biased upward due to omitted variables that are subsumed in product-time fixed effects. This could be due to any product-specific economic shocks over time such as technological progress that improves product quality over time, changes in availability of high-quality inputs which are indispensable in quality upgrading, and for time-series variation in consumer demand that maybe a result of growing income over time and non-homothetic preference.

[Insert Table 2.1 here]

[Insert Table 2.2 here]

2.4.4 Identification Challenges

One implicit threat to the identification of our baseline regression is that the proliferation of ICT technology in a city may boost firm productivity when the firms employ more ICT devices in their daily production. If we observe that ICT penetration leads to quality upgrading of firms, it could be a consequence of firms upgrading quality when their productivity is higher. To alleviate this concern, we exploit the spatial and time-series variations in ICT infrastructure in cities other than the location that the firm is in. The implicit assumption behind our logic is that, for a particular prefecture building of ICT infrastructure in other cities other than itself would not change firm productivity in that prefecture. In addition, we assume that the improvement in ICT infrastructure elsewhere would have reduced a firm's search cost to obtain information. This may be justified, in the sense that, a person who uses internet or mobile phone in an origin city would have increased her communication efficiency if there are more internet or mobile users in the destination city.

We construct the variation in ICT infrastructure in the destination cities as the sum of inverse distance weighted ICT infrastructure, and this variable is labeled as a measure of "information access", IA_{it} of local market i at time t since it follows the approach to construct a conventional measure of transportation cost based market access,

$$IA_{it} = \sum_{j \neq i} \left(\frac{1}{d_{ij}^{\epsilon}} \right) TeleCom_{jt},$$

where d_{ij} is the geographic distance between city i and city j.[18]

2.4.5 Robustness Results

We report our findings using the alternative measure of information access as the main variables of interest in Table 2.3 and Table 2.4. Our results remain qualitatively stable across all variables of interest and across each specification. Again, the results for unit value of products are not as good as the others when we use the alternative measure. And we attribute this again to the the quality of measure itself. Quantitatively, the coefficient estimates become twice larger than those using the ICT development of the origin city as proxies. Our interpretation of this finding is such that communications require inputs from both parties in the origin and destination cities. Since we are summing over ICT development across all potential destinations, it is only natural that the estimates from this regression are much larger. Intuitively, if ICT development only happens at the origin city or any single destination market, this does not reduce the search friction significantly because search is random in our model and is not directed any particular market. In contrast, a significant reduction of information frictions across all destination markets would have shrunk search frictions substantially for firms in the origin region. Hence, this alternative measure should have registered a larger effect in the robustness check regressions, as so reported in Table 2.3 and Table 2.4. In the latest draft, we will also include the results from using region-specific measures on the building of ICT infrastructures to provide an additional set of robustness checks.

[Insert Table 2.3 here]

[Insert Table 2.4 here]

As discussed earlier, though we have included an exhaustive set of fixed effects as well as an alternative measure of ICT development to address omitted variable bias and other endogeneity concerns, our specifications still suffer from endogeneity in particular from selection issues such that the central government may only choose to build the ICT

[18] In our baseline regression we simply assume $\epsilon = 1$.

infrastructures based on location specific characteristics. As we do not have the detailed geo-coded data of the ICT backbone network in China, it's impossible to construct an instrumental variable based on least-cost algorithms. In future works, we will exploit the information on the opening of airports during the same period. Opening of airports would have improved increased face-to-face contact between the origin city and the destination city, and hence it would have reduced information frictions. Opening of airports also do not affect the transportation costs as intra-national trade in China are mostly moved through railroads, high-ways, and water-ways. Since we can access the geo-coded information of the airports, we would be able to construct an alternative network of travel routes via air that could connect the destinations with least cost and use this alternative network as an instrument for our specification.

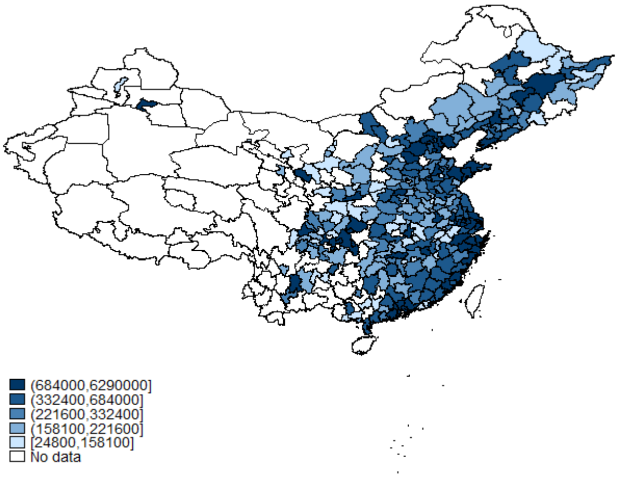
In addition, in the present scope of our paper, we only test the qualitative predictions that improving information frictions would have induced quality upgrading. Other predictions especially those concerning pro-competitive effects are less tested. We will address these issues through the lens of international trade using China Custom data. This approach is immune from endogeneity concerns such that the ICT development is associated with the location characteristics since there is no single organization that is coordinating the building of ICT infrastructure across all countries.

2.5 Conclusion and Future Works

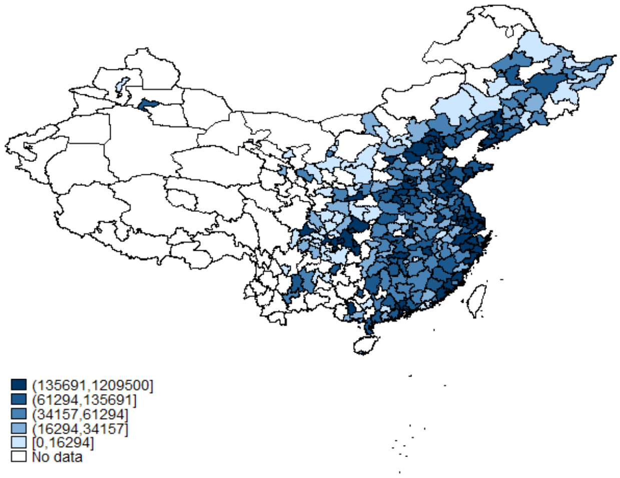
In this paper, we study the pro-competitive effects associated with reducing information frictions. We corroborate our hypothesis by introducing monopolistic competition, heterogeneous firms, and quality upgrading into a sequential search model with trade (Allen, 2014). Given the setup of our model, a firm's strategy in costly search for information is characterized by a reservation strategy, such that it will only compete in a destination if the firms are not that efficient and/or that they supply goods with inferior quality. The predictions from our model are that reducing the fixed cost of search would lead more efficient firms to compete in less competitive markets which are usually the remote locations. In addition, it would also lead to all firms to upgrade their product quality because improving information frictions provides the firms with better information-based market access.

We then take the model to firm-level data in China, together with spatial variation in ICT infrastructure across Chinese cities. The qualitative tests largely corroborates our stories using different variables as proxies for ICT development and different variables as proxies for quality. Our results are robust across specifications as well as including additional controls that address endogeneity concerns. For future works, we are currently calibrating the full-blown general equilibrium to supply richer quantitative implications of our model. Second, we are building an instrumental variable to better address other endogeneity issues in our empirical investigation. Furthermore, we are exploiting international variation in ICT infrastructure, together with the product-level information in the Chinese Custom data, to provide more credible and also richer evidence on pro-competitive effects of reducing information frictions.

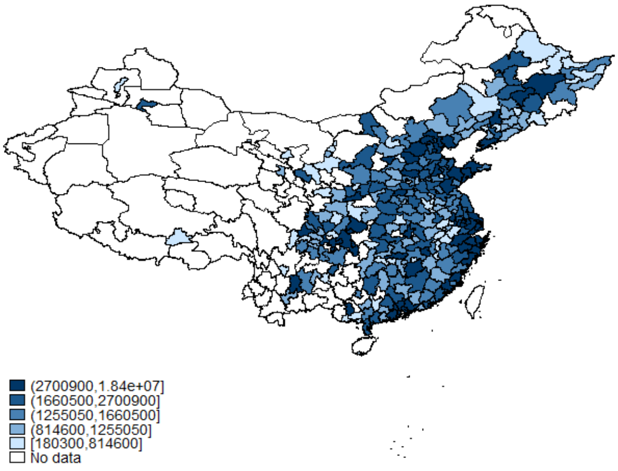
Figure 2.2: Spatial Distribution of Telecom. Development, Year 2001-2007



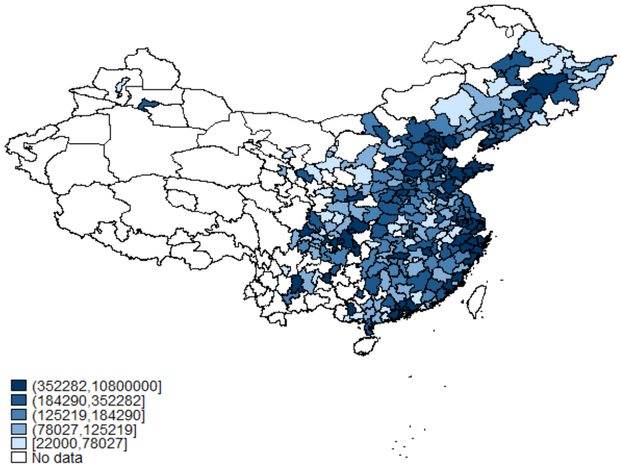
A. Mobile Phone Owners, 2001



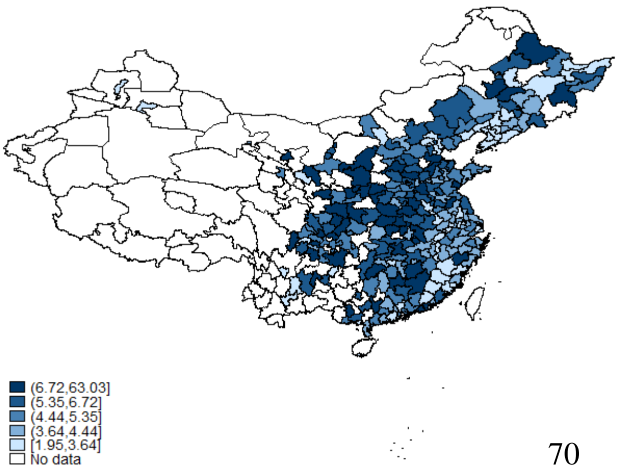
B. Broadband Internet Subscribers, 2001



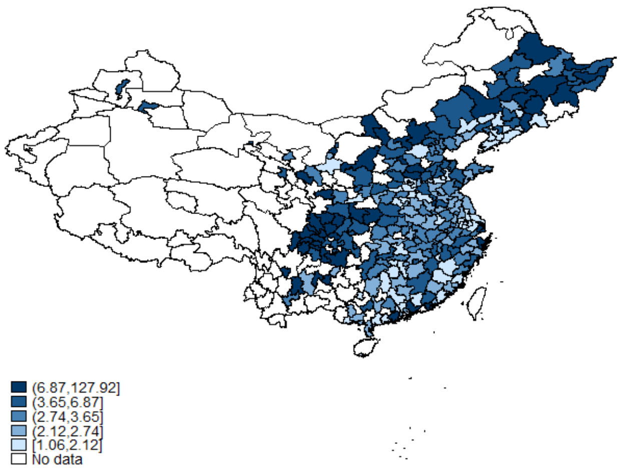
C. Mobile Phone Owners, 2007



D. Broadband Internet Subscribers, 2007



E. Growth of Mobile Phone Owners



F. Growth of Broadband Internet Subscribers

Table 2.1: Telecom. Development and Product Quality, Product FE and Year FE

Dependent Variable	$\ln Market\ Shares$				$Quality\ (Khdwl)$				$\ln Prices$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$Mobile\ Phones$	0.061***	0.082*			0.027***	0.093*			-0.074***	0.018		
	(0.004)	(0.047)			(0.006)	(0.053)			(0.007)	(0.016)		
$Internet$			0.088***	0.102***			0.072***	0.121***			-0.068***	0.032**
			(0.004)	(0.019)			(0.006)	(0.021)			(0.006)	(0.013)
$Market\ Access$	-0.394***	0.616***	-0.421***	0.480**	-0.711***	0.563***	-0.768***	0.399**	0.040	-0.070	0.021	-0.119
	(0.015)	(0.190)	(0.014)	(0.190)	(0.024)	(0.161)	(0.023)	(0.144)	(0.028)	(0.160)	(0.027)	(0.154)
$Employment$	0.629***	0.737***	0.626***	0.736***	-0.119***	0.338***	-0.123***	0.337***	-0.256***	0.115***	-0.254***	0.115***
	(0.004)	(0.017)	(0.004)	(0.017)	(0.007)	(0.015)	(0.007)	(0.015)	(0.007)	(0.018)	(0.007)	(0.018)
TFP	0.149***	0.345***	0.154***	0.345***	0.072***	0.425***	0.078***	0.426***	0.024***	0.113***	0.023***	0.113***
	(0.003)	(0.019)	(0.003)	(0.019)	(0.005)	(0.022)	(0.005)	(0.022)	(0.005)	(0.027)	(0.005)	(0.027)
City FE $\times$Time trend	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Product FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Year FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
R-Squared	0.125	0.489	0.127	0.489	0.007	0.924	0.008	0.924	0.008	0.858	0.008	0.859
Obs	185,938	185,938	185,370	185,370	185,938	185,938	185,370	185,370	185,938	185,938	185,370	185,370

Notes: " Mobile Phones" is the log of number of mobile phone owners in a city. "Internet" is the log of number of broadband Internet subscribers in a city. "Market Access" represents the log of sum of inversed distance city income. Firm-level controls are employment, TFP, and exporter dummy. Robust standard errors are clustered at city-industry level for regressions (2),(4), (6), (8), (10), (12) and are reported in parentheses. * denotes for $p < 0.1$, ** denotes for $p < 0.05$, and *** denotes for $p < 0.01$.

Table 2.2: Telecom. Development and Product Quality, Product-Year FE

Dependent Variable	$\ln Market\ Shares$				$Quality\ (Khdwl)$				$\ln Prices$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$Mobile\ Phones$	0.061***	0.058*			0.027***	0.070*			-0.074***	0.016		
	(0.004)	(0.032)			(0.006)	(0.040)			(0.007)	(0.016)		
$Internet$			0.088***	0.068***			0.072***	0.089***			-0.068***	0.030**
			(0.004)	(0.014)			(0.006)	(0.013)			(0.006)	(0.012)
$MarketAccess$	-0.394***	0.440**	-0.421***	0.353**	-0.711***	0.418***	-0.768***	0.301**	0.040	-0.037	0.021	-0.081
	(0.015)	(0.165)	(0.014)	(0.169)	(0.024)	(0.120)	(0.023)	(0.116)	(0.028)	(0.128)	(0.027)	(0.126)
$Employment$	0.629***	0.746***	0.626***	0.745***	-0.119***	0.347***	-0.123***	0.347***	-0.256***	0.116***	-0.254***	0.116***
	(0.004)	(0.017)	(0.004)	(0.017)	(0.007)	(0.015)	(0.007)	(0.015)	(0.007)	(0.018)	(0.007)	(0.018)
TFP	0.149***	0.348***	0.154***	0.348***	0.072***	0.429***	0.078***	0.429***	0.024***	0.115***	0.023***	0.115***
	(0.003)	(0.021)	(0.003)	(0.021)	(0.005)	(0.023)	(0.005)	(0.023)	(0.005)	(0.027)	(0.005)	(0.027)
City FE $\times$Time trend	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Product-Year FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
R-Squared	0.125	0.531	0.127	0.531	0.007	0.944	0.008	0.945	0.008	0.863	0.008	0.863
Obs	185,938	185,937	185,370	185,369	185,938	185,937	185,370	185,369	185,938	185,937	185,370	185,369

Notes: " Mobile Phones" is the log of number of mobile phone owners in a city. "Internet" is the log of number of broadband Internet subscribers in a city. "Market Access" represents the log of sum of inversed distance city income. Firm-level controls are employment, TFP, and exporter dummy. Robust standard errors are clustered at city-industry level for regressions (2),(4), (6), (8), (10), (12) and are reported in parentheses. * denotes for $p < 0.1$, ** denotes for $p < 0.05$, and *** denotes for $p < 0.01$.

Table 2.3: Info. Access and Product Quality, Product FE and Year FE

Dependent Variable	$\ln Market\ Shares$				$Quality\ (Khdwl)$				$\ln Prices$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$Info.\ Access\ (Mob.\ Phones)$	-0.064***	0.245**			-0.297***	0.240**			-0.427***	0.013		
	(0.009)	(0.102)			(0.014)	(0.089)			(0.016)	(0.048)		
$Info.\ Access\ (Internet)$			0.021**	0.180**			-0.308***	0.180***			-0.590***	0.016
			(0.010)	(0.076)			(0.015)	(0.062)			(0.016)	(0.043)
$MarketAccess$	-0.213***	0.411	-0.332***	0.470*	-0.215***	0.374	-0.282***	0.428*	0.589***	-0.076	0.672***	-0.081
	(0.019)	(0.258)	(0.018)	(0.250)	(0.031)	(0.252)	(0.028)	(0.230)	(0.034)	(0.147)	(0.032)	(0.153)
$Employment$	0.630***	0.734***	0.629***	0.734***	-0.119***	0.336***	-0.122***	0.336***	-0.257***	0.116***	-0.259***	0.116***
	(0.004)	(0.018)	(0.004)	(0.018)	(0.007)	(0.015)	(0.007)	(0.015)	(0.008)	(0.018)	(0.008)	(0.018)
TFP	0.145***	0.343***	0.145***	0.343***	0.067***	0.425***	0.066***	0.425***	0.025***	0.114***	0.021***	0.114***
	(0.003)	(0.019)	(0.003)	(0.019)	(0.005)	(0.023)	(0.005)	(0.023)	(0.005)	(0.027)	(0.005)	(0.028)
City FE $\times$ Time trend	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Product FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Year FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
R-Squared	0.124	0.488	0.124	0.488	0.009	0.924	0.009	0.924	0.011	0.858	0.013	0.858
Obs	178,468	178,468	178,468	178,468	178,468	178,468	178,468	178,468	178,468	178,468	178,468	178,468

Notes: " Mobile Phones" is the log of number of mobile phone owners in a city. "Internet" is the log of number of broadband Internet subscribers in a city. "Market Access" represents the log of sum of inversed distance city income. Firm-level controls are employment, TFP, and exporter dummy. Robust standard errors are clustered at city-industry level for regressions (2),(4), (6), (8), (10), (12) and are reported in parentheses. * denotes for $p < 0.1$, ** denotes for $p < 0.05$, and *** denotes for $p < 0.01$.

Table 2.4: Info. Access and Product Quality, Product-Year FE

Dependent Variable	ln $Market\ Shares$				$Quality\ (Khdwl)$				ln $Prices$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$Info.\ Access\ (Mob.\ Phones)$	-0.064***	0.258***			-0.297***	0.267***			-0.427***	0.018		
	(0.009)	(0.075)			(0.014)	(0.063)			(0.016)	(0.047)		
$Info.\ Access\ (Internet)$			0.021**	0.172***			-0.308***	0.183***			-0.590***	0.019
			(0.010)	(0.048)			(0.015)	(0.034)			(0.016)	(0.044)
$Market\ Access$	-0.213***	0.181	-0.332***	0.259	-0.215***	0.164	-0.282***	0.238	0.589***	-0.044	0.672***	-0.048
	(0.019)	(0.187)	(0.018)	(0.186)	(0.031)	(0.166)	(0.028)	(0.154)	(0.034)	(0.124)	(0.032)	(0.129)
$Employment$	0.630***	0.743***	0.629***	0.742***	-0.119***	0.345***	-0.122***	0.345***	-0.257***	0.116***	-0.259***	0.116***
	(0.004)	(0.017)	(0.004)	(0.017)	(0.007)	(0.015)	(0.007)	(0.015)	(0.008)	(0.018)	(0.008)	(0.018)
TFP	0.145***	0.347***	0.145***	0.347***	0.067***	0.430***	0.066***	0.430***	0.025***	0.116***	0.021***	0.116***
	(0.003)	(0.021)	(0.003)	(0.021)	(0.005)	(0.023)	(0.005)	(0.023)	(0.005)	(0.028)	(0.005)	(0.028)
City FE$\times$Time trend	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Product-Year FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
R-Squared	0.011	0.863	0.013	0.863	0.124	0.531	0.124	0.531	0.009	0.945	0.009	0.945
Obs	178,468	178,467	178,468	178,467	178,468	178,467	178,468	178,467	178,468	178,467	178,468	178,467

Notes: " Mobile Phones" is the log of number of mobile phone owners in a city. "Internet" is the log of number of broadband Internet subscribers in a city. "Market Access" represents the log of sum of inversed distance city income. Firm-level controls are employment, TFP, and exporter dummy. Robust standard errors are clustered at city-industry level for regressions (2),(4), (6), (8), (10), (12) and are reported in parentheses. * denotes for $p < 0.1$, ** denotes for $p < 0.05$, and *** denotes for $p < 0.01$.

3 Does Market Integration Lead to Spatial Concentration of Higher-Quality Products? Evidence from Expansion of China's Highway System

3.1 Introduction

Transportation infrastructure has been mentioned as a fundamental element for a country's economic development. The construction of the highway and railway inevitably decreases the trade costs across different regions by reducing the costs of delivering goods and people. Recent studies on transportation infrastructure mainly focus on its impact on regional economic consequences such as employment, industrialization pattern, and GDP growth (Duranton and Turner, 2012; Faber, 2014; Banerjee, Duflo and Qian, 2018). Less attention has been paid to the impact of transportation infrastructure projects on an important economic activity: quality upgrading activities of firms. In addition, the role of international trade costs in shaping quality specialization pattern across countries has been well studied while we still lack a rigorous understanding of how firms' quality decisions will response to the changes in intra-national trade costs induced by transportation infrastructure.

In this paper, we would like to study an additional margin of the impact of market integration (falling trade costs) by empirically investigating the effect of falling trade costs on quality production patterns across cities. Our analysis is based on the rapid highway system expansion in China over the past decades. China has launched its economic reform in 1978, following which the construction of transportation infrastructure has persistently grown. In particular, the nationalwide highway length has experienced a rapid increase by more than 100% over 2001-2007, from 1,702,000 km to 3,583,700 km, which indicates that there exists a substantial variation in the highway system across years. This rich time-series (cross-sectional) variation across years (regions) makes China an ideal choice to study the impact of falling trade costs on quality choices. We thus construct three measures to evaluate the extent to which a prefecture's highway is developed in our empirical analysis. The first measure is the highway length (km), which does not take the congestion issue into account. We adopt highway length per capita ($km/person$) and highway length per square urban land area size (km/km^2) to better capture the spatial and economic impact of the highway. The estimated effects are quite similar to three measures.

To construct a rigorous quality measure, we merge the *Annual Survey of Industrial Firms* and *Industrial Firms Product Quantity Dataset* and keep single-product firms to extract unit value and market share information of a product category. Our quality estimation approach is based on the nested logit model of Berry (1994) adapted by Khandelwal (2010).

We exploit the spatial variations on China's highway system expansion and manufacturing firms' output quality over 2001-2007. Our baseline analysis shows that connecting highways have a positive and significant impact on quality level across product categories, implying that firms would choose to upgrade product quality in response to a greater expansion in highways. Specifically, a 10% increase in a prefecture's highway length would result in about 1.46% increase in its product quality. We further examine the differences of this pattern between large cities such as provincial capitals, and small cities. The results suggest that the impact of highway expansion is more pronounced in large cities, which speaks to the changes in the spatial concentration of higher-quality products. Furthmore, mechanism testing shows that skill premium, the relative price for skilled labor to unskilled labor, exerts a negative and significant impact on the interaction between intra-national trade costs and firms quality choices, which is consistent with our expectation.

To alleviate the omitted bias concern, we control for a series of prefecture characteristics that could be potential determinants and affect firms' quality choices locally and a full set of fixed effects. The results are robust to the introduction of controlled variables and fixed effects.

For further robustness checks, we focus on observations in sub-samples that exclude large cities, special economic zones, coastal open cities, and Yangtze Delta River Region. The estimated results are robust to various checks.

This paper contributes to the literature by providing the empirical evidence of falling in trade costs induced by transportation infrastructure on the production pattern of quality across cities. Our study is related to the recent literature in two strands. First, it adds to the work on the economic consequences of transportation infrastructure projects in reducing delivering costs across regions within a country. Donaldson (2018) and Banerjee, Duflo and Qian (2018) focus on the aggregate and distributional impacts of transportation infrastructure. Specifically, Donaldson (2018) studied how railway network construction in India would contribute to welfare gains, while Banerjee, Duflo and Qian (2018) considered the same topic for the highway system in China. In addition, other research on this topic focus on how transportation infrastructure reshape the geographic distribution of economic activities (Faber, 2014; Duranton and Turner, 2012). Our work contributes by studying an additional margin of the impact of transportation infrastructure on the greographic economic activity distribution: the quality production pattern.

In addition, our work relates to the studies on quality specialization across countries in international trade literature, which discuss both demand-side (Piveteau and Smagghue, 2019; Dingel, 2017; Fajgelbaum, Grossman and Helpman, 2011, 2015; Hallak, 2006, 2010; Choi, Hummels and Xiang, 2009) and the supply side explanations (Fieler, Eslava and Xu, 2018; Dingel, 2017; Faber and Fally, 2017; Fan, Li, Xu and Yeaple, 2017; Antoniades, 2015; Feenstra and Romalis, 2014; Hallak and Sivadasan, 2013; Kugler and Verhoogen, 2012; Crozet, Head and Mayer, 2012; Khandelwal, 2010; Verhoogen, 2008; Schott, 2004; Hummels and Skiba, 2004). Focusing on the changes in intra-national trade costs, our reduced-form estimates provide a first look at the overall impact of market integration on the geographic distribution of quality production pattern. Nevertheless, more works need to be done to distinguish the economic forces of supply- and demand-side fundamentals in shaping quality specialization pattern.

The rest of the paper is organized as follows. In Section 3.3, we discuss the background of China's highway system and introduce the data. We also present the approach to construct various key measurements in this section. Next, We introduce the empirical strategy to explore the impact of highway expansion on product quality in Section 3.4 and discuss the baseline results in Section 3.5. In Section 3.6, we further investigate the channels through which highway expansion affects quality, and discuss how our empirical findings shed light on the theoretical prediction in spatial literature. We perform a series of robustness checks in Section 3.7 and discuss future work in Section 3.8. Finally, we conclude in Section 3.9.

3.2 Data and Descriptive Evidence

In this section, we introduce the data and variables used in the empirical investigation.

First, our data on transportation infrastructure at both the prefecture and province level comes from the *China Regional Economy Statistical Year Books* during 2001-2007. Highways in China are classified into six categories based on their road quality and congestion, which include expressway, first-class highway, second-class highway, third-class highway, fourth-class highway, and unclassified highway. In particular, expressway, first-class highway, second-class highway, third-class highway, and fourth-class highway are jointly grouped as *classified highway*, while the rest are *unclassified highway*.

The yearbooks report the length of overall highway, classified highway, expressway, first-class highway, railway, and waterway respectively at the province level. As for the prefecture level, however, the yearbooks contain only the overall highway length and classified highway length. In our baseline analysis, our geographic unit is the prefecture level. We provide a first look at the highway expansion in China over 2001-2007 across prefectures in Figure 3.1. The average prefecture highway persistently increased and experienced a substantial expansion in the year 2006, indicating a drastic decline in the transportation costs across regions over this period. Furthermore, the increased standard deviation of the highway length suggests wider spatial discrepancies in the highway system development across locations in China, which could have a profound impact on the spatial distribution of economic activities.

[Insert Figure 3.1 here]

In Figure 3.2, we scatter-plots a prefecture's population size and the corresponding highway length in the years 2001 (blue), 2004 (red), and 2007 (green). The quadratic fitted lines suggest a positive correlation between city's population size and its highway length, which becomes more pronounced over the year.

In summary, the descriptive evidence in 3.1 and 3.2 implies rich variations across regions and years in China's highway construction, and how fundamental the highways are in shaping spatial economic activities.

[Insert Figure 3.2 here]

Second, to construct the measure for quality, we use two datasets, the *Annual Survey of Industrial Firms (ASIF)* dataset and the *China Industrial Firms Product Quantity Dataset*. The ASIF dataset and the quantity dataset are both collected and constructed by the National Bureau of Statistics in China. The ASIF dataset provides detailed firm-level information on financial and production status including gross output, value-added, capital stock, export value, number of employees, ownership, and wage bill. The dataset covers all China's state-owned enterprises, as well as non-state-owned enterprises with annual sales of RMB 5 million (US$ 770,000) or above. Our approach to constructing a panel ASIF dataset is largely adapted from Brandt, Van Biesebroeck and Zhang (2012).

The *Chinese Industrial Firms Product Quantity Dataset* documents physical output quantity at five-digit Chinese Product Code (CPC) level of industrial firms. We merge the ASIF dataset with the product quantity dataset using firms' identity code (ID) and keep single-product firms to extract the information on market share and unit value (price) of a product. [19]

We restrict our study to manufacturing firms across 284 five-digit product categories, 29 two-digit industries, 247 prefectures and 28 province-level administrative units in China. The 28 province-level administrative units include 20 provinces, 4 province-level municipalities, and 4 minority autonomous regions. The 20 provinces are Hebei, Shanxi, Liaoning, Jilin, Heilongjiang, Jiangsu, Zhejiang, Anhui, Fujian, Jiangxi, Shandong, Henan, Hubei, Hunan, Guangdong, Sichuan, Guizhou, Yunnan, Shaanxi, Gansu, Qinghai. The 4 direct-administered municipalities are Beijing, Tianjin, Shanghai, and Chongqing. The 4 minority autonomous regions are Inner Mongolia, Guangxi, Xinjiang, and Ningxia.

[19] We define a product's market share as the share of sales quantity within each five-digit product category of that year following Khandelwal (2010).

Third, we collect the socio-economic variables in each prefecture from various sources. We obtain agricultural population, non-agricultural population, urban developed area size, GDP per capita, numbers of graduates of higher-education institutions, as well as inward foreign direct investment (inward FDI) from the *China Regional Economy Statistical Year Books* and *China Urban Construction Statistical Year Books*. The wage premium of a prefecture is constructed using the data from the *2005 China 1% population census* (see Section 3.6.2 for a detailed description). Lastly, we construct a series of dummy variables to indicate whether a prefecture is a provincial capital, a coastal port, a special economic zone, or a coastal open city from Wikipedia.

Table 3.1 provides the summary statistics for variables at both the prefecture and province levels.

3.3 Variables and Measurements

3.3.1 Highway Expansion

To evaluate the highway development across space and year, we construct three measures at the prefecture-year level. The first measure is the overall length of the highway in a prefecture $(Highway\ Length_{p,t})$.

Though $(Highway\ Length_{p,t})$ reflects the overall development of highway in a prefecture in a year, it is inevitably a coarse measure as it does not take into acount the population and congestion. We then use $(Highway\ per\ capita_{p,t})$ as the second measure, which unit is $(km/person)$. We compute this measure by dividing a prefecture's highway length with its population size in the same year.

$$Highway\ per\ capita_{p,t} = \frac{Highway\ Length_{p,t}}{Population\ Size_{p,t}}. \tag{3.1}$$

The third measure is $(Highway\ Geo\ Density_{p,t})$, which unit is (km/km^2). We focus on the extent to which highway expansion would benefit urban economic development and thus we divide the highway length with the corresponding urban land area size.

$$Highway\ Geo\ Density_{p,t} = \frac{Highway\ Length_{p,t}}{Urban\ Land\ Size_{p,t}}. \tag{3.2}$$

3.3.2 Quality

Quality Estimation.—— Product quality is an economic element that cannot be directly observed or measured from the data. We first follow Khandelwal (2010) 's approach to construct a measure for product quality by exploring the variations in prices and market shares of a product based on the nested logit model of Berry (1994).

Suppose $(Q^s_{f,t})$ represents product s's quality of firm f in year t. We decompose $(Q^s_{f,t})$ into three terms, product-firm fixed effect (λ^s_f), product-year fixed effect (λ^s_t), and year fixed effect (λ_t). Product-firm fixed effect (λ^s_f) captures consumer's time-invariant preferences for product quality. Year fixed effect (λ_t) measures the time trend, or macro-economic shocks common across all product categories' quality. We regard the product-year fixed effect (λ^s_t) as the product-year deviations from the average quality (fixed effect), which can only be observed by consumer but not by econometricians.

The empirical specification for estimating quality is

$$\ln MarketShare^s_{f,t} = \beta_0 + \beta_1 \ln Price^s_{f,t} + \beta_2 \ln FirmSize_{f,t} + \gamma_t + \lambda^s_f + \lambda^s_{f,t}. \quad (3.3)$$

$Price^s_{f,t}$ is the price of product s produced by firm f in year t, which is computed by dividing the sales with quantities of firm f in year t. $MarketShare^s_{f,t}$ is obtained by dividing the output quantities of a firm's product by the sum of quantity within each five-digit product category. $FirmSize_{f,t}$ is proxied by the number of employees of a firm f in year t. γ_t captures year fixed effect, which subsumes the market share of outside goods as well as time-specific variatiosn for product quality.

Identification Issues.—— There are several identification challenges in estimating quality. The first issue is ”hidden varieties”. The market share is measured at the five-digit product level. However, there could be more disaggregated product classifications within a five-digit product category. For example, a cereal producer could produce many hidden varieties (different favors) at seven-digit product classification within the five-digit product group ”fast food” and sells them at the same prices. Under this scenario, the hidden varieties would lead to an upward bias in the estimated quality at five-digit product category. We follow Khandelwal (2010) and use employment size of each firm as control of hidden varieties.

The second identification issue is from the outside product. We are unable to obtain the price and quantity information of an outside product in the dataset. Note that the market share of the outside product $MarketShare^{0}_{j,t}$ is a constant for products within the same product category of the same year, we then treat $MarketShare^{0}_{j,t}$ as a component of year fixed effects.

With parameters estimated, we can compute the estimated quality $\hat{Q^{s}_{j,t}}$ for product s of firm f in year t as

$$\widehat{Q^{s}_{j,t}} = \ln MarketShare^{s}_{j,t} - \hat{\beta}_1 \ln Price^{s}_{j,t} - \hat{\beta}_2 \ln FirmSize_{f,t} - \widehat{\gamma}_t. \tag{3.4}$$

Prefecture Quality Measure.—— To characterize the average quality choices made by firms in a prefecture, we introduce two measures as follows. We first use weighted-average of output quality proudced by firms in a prefecture. Here we choose firm sales as the weight for the measure. The measure is given by

$$WgtQuality^{h,s}_{p,t} = \frac{1}{N^{h,s}_{p,t}} \sum_{f} \frac{S^{h,s}_{f,p,t}}{S^{h,s}_{p,t}} Q^{h,s}_{f,p,t}. \tag{3.5}$$

$\left(Q^{h,s}_{f,p,t}\right)$ is the measurement for product s produced by firm f from industry h prefecture p in year t. $\left(S^{h,s}_{f,p,t}\right)$ is the corresponding sales of product s produced by firm f in industry h prefecture p year t. $\left(S^{h,s}_{f,p,t}\right)$ measures the total sales of product s in industry h prefecture p year t. $\left(N^{h,s}_{p,t}\right)$ is the total numbers of firms producing product s in industry h prefecture p year t.

In our robustness check 3.11, we present the estimated results using unweighted-average (simple average) measure $AvgQuality^{h,s}_{p,t}$ as the dependent variable. Specifically,

$$AvgQuality^{h,s}_{p,t} = \frac{\sum_f Q^{h,s}_{f,p,t}}{N^{s}_{p,t}}, \tag{3.6}$$

3.4 Empirical Strategy

In this section, we introduce our estimation strategy to examine the effects of transportation infrastructure on the weighted average quality of a product. Our baseline specification is

$$WgtQuality_{p,t}^{h,s} = \beta_0 + \beta_1 \ln Highway_{p,t} + \eta \mathbf{X}_{p,t} + \lambda_p^h + \delta_t^s + \epsilon_{p,t}^{h,s}. \tag{3.7}$$

β_0 is an intercept; dependent variable $WgtQuality_{p,t}^{h,s}$ summarizes prefecture p's firms quality choices for product s in year t, as discussed in Section 3.3. Our explanatory variable of interest is $\ln Highway_{p,t}$, which measues the highway expansion in prefecture p year t and could either be the log of highway length, highway per capita or highway geo density. This empirical specification exploits the within-product-prefecture variations to identify the impact of highway system expansion on quality production activities.

$\mathbf{X}_{p,t}$ is a set of other time-varying prefecture-level controls which could be potential determinants of the product quality of firms in a prefecture. It also alleviates the omitted variables bias concern at the prefecture level. $\mathbf{X}_{p,t}$ first include GDP per capita, which characterizes the overall level of economic development as well as the non-homothetic demand of a prefecture that could locally exert a positive impact on product quality. We also include population size to proxy for any agglomeration externalities of an urban center. Specifically, firms in larger cities would benefit more from the agglomeration externalities such as indivisible facilities sharing, buyer-supplier matching, or knowledge spillovers (Marshall (1920)), which could also positively affect quality upgrading activities. Besides, we control for the ratio of the number of students who graduated in higher education institutions to its total population as a proxy for the average human capital level of the labor force in a prefecture. The average human capital level of the labor force in a prefecture could potentially affect the average quality level of production as they would participate in production activities. The inward FDI could also play a role in attracting high-skilled workers and high-tech enterprises to a prefecture and thus could affect the overall output quality decisions of a region. We thus include the flow value of inward FDI in a prefecture as an indicator of the FDI advantage of a prefecture in that year. Finally, the urban land size could be a possible determinant of average quality level in a prefecture as it would affect the magnitude of agglomeration forces and economic density of a prefecture. We thus include the urban land size as a control variable in our prefecture characteristics.

The specification also controls for a series of fixed effects to address the omitted variable bias issue. We first introduce city-industry fixed effect λ_p^h to account for any time-invariant city-industry specific shocks that may affect the average quality across product categories and cities (Chor and Manova (2012)).

We further incorporate product-year fixed effect δ_t^s to take into account any time-varying product-specific shocks received by all firms producing the same product regardless of their locations, which could be a technological revolution, reduction in input prices, or drastic increase in market demand.

Note that we cannot control for prefecture-year fixed effects because it will subsume all the prefecture-year specific variables including our key independent variable $Highway_{p,t}$.

$\epsilon_{p,t}^{h,s}$ is an error term. We cluster the standard errors at the province level to account for any potential spatial correlations of error terms across prefectures within the same province. In our baseline regression, to deal with the heteroskedasticity problem, we use the beginning-of-period log employment of a prefecture as a weight in the estimation. Also, we cluster standard errors at the province level (28 provinces) to take into account any potential covariance between the error terms across prefectures within the same province.

3.5 Baseline Results

In this section, we examine the impact of highway system expansion on the quality choice of firms in China's various prefectures by running regression (3.7) in Section 3.4. Table 3.2-3.4 presents the main results using three different measures, highway length, highway per capita, and highway geo density for $Highway$ and $Classified\ Highway$.

[20]

Table 3.2 reports the estimated results for $Highway\ Length_{p,t}$ and $Classified\ High-way\ Length_{p,t}$. Columns (1) and (3) in Table 3.2 present the estimated coefficients of $Highway\ Length_{p,t}$ and $Classified\ Highway\ Length_{p,t}$ without controlling for any prefecture characteristics. Both coefficients are positive and statistically significant at 1% level. In Columns (2) and (4) we further include a series of time-varying prefecture characteristics that may correlate with both our outcome variable (quality) and our regressor of interest (highway) to alleviate omitted variable bias concern. The magnitude of the coefficients of both $Highway\ Length_{p,t}$ and $Classified\ Highway\ Length_{p,t}$ roughly keep unchanged. In terms of the economic magnitude of the impact, for any 10% increase in highway length in a prefecture n, the expected quality level of prefecture n would approximately increase by 1.46%. Similar results are found in Table 3.3 and 3.4 when we use highway per capita and highway geo density as the key explanatory variables.

It is interesting that in the full specification, the positive impact of $Highway\ Length_{p,t}$ on quality (0.146) is slightly larger than that of$Classified\ Highway\ Length_{p,t}$ (0.092), indicating that the increase in unclassified highway also plays a role in shaping firms' quality choices.

3.6 Channels At Work

3.6.1 Concentration of Higher-Quality Production

In the last section, the estimated results show that China's highway system expansion during 2001-2007 exerts both economically and statistically significant impact on the quality decisions made by firms. However, the channels through which the highway would affect quality choices are still unclear. In particular, we are interested in how the reduction in trade costs across regions, which is induced by the highway expansion, reshapes quality specialization pattern across space within a nation. To deal with this question, we interact highway expansion measure with a dummy variable that goes to 1 if

[20] $Highway$ covers all highway of categories in a prefecture. $Classified\ Highway$ includes expressway, first-class highway, second-class highway, third-class highway and fourth-class highway in a prefecture. The classification is based on road quality and congestion.

a prefecture is a "Large City" and goes to 0 otherwise. We then add this interaction term in the baseline specification as shown in Equation (3.8). Note that the dummy variable does not enter the specification independently as it will be subsumed by the city-industry fixed effects. We define a prefecture as a "Large City" if it is a centrally-administrated municipality or a provincial capital city.

$$WgtQuality_{p,t}^{h,s} = \beta_0 + \beta_1 \ln Highway_{p,t} + \beta_2 \ln Highway_{p,t} \times i.LargeCity_p + \eta \mathbf{X}_{p,t} + \lambda_p^h + \delta_t^s + \epsilon_{p,t}^{h,s}. \quad (3.8)$$

Columns (1)-(2), (3)-(4) and (5)-(6) in Table 3.5 report the regression results for all the three measures of highway expansion (total length, per capita, and geo. density), respectively. We find that the coefficients in Columns (1), (3), and (5) are both economically and statistically significant and positive, implying that firms have a higher probability to produce higher-quality goods when they stay in larger cities with faster highway development. Surprisingly we do not find a significant impact of city size on the effect of classified highway on product quality. Such reduced-form estimates contain both the agglomeration and firm sorting forces in the impact of highway on quality specialization pattern across space.

[Insert Table 3.5 here]

3.6.2 Skill Premium and Quality Production

Apart from the impact of city size on the effect of highway expansion on product quality, we are also interested in the role of skill premium. Higher-quality firms would use higher-quality inputs and employ relatively more skilled labor as discussed in Fieler, Eslava and Xu (2018). Skill premium, which is the ratio of skilled worker wage to unskilled worker wage, can also be interpreted as the relative price for skilled labor. The relative price for skilled labor would affect the output quality level when firms are making production decisions. Instead of taking a simple average of skilled wages over unskilled wages in an area, we apply the "Mincer approach" to estimate the skill premium.

Skill Premium Estimation.—— We explore the individual-level information in China's 2005 1% Population Census ("mini census") to construct a measure of skill premium at the prefecture level. The 2005 mini census provides information on individuals' education levels and locations when being interviewed. We follow Fan (2019) and Mincer (1958) to take into account individual characteristics such as age and gender in the estimation. The Mincer-type specification is

$$\ln Wage_{i,p} = \beta_0 + \beta_1 Age_i + \beta_2 Age_i^2 + \beta_3 Gender_i + D_p + S_p \left(I_{Skill} \times D_p \right) + \epsilon_{i,p}. \quad (3.9)$$

where $Wage_{i,p}$ is the hourly wage for individual i from prefecture p in the 2005 mini census. D_p is a prefecture level fixed effect. I_{Skill} is a dummy variable which goes to one if individual i is a skilled worker, and zero otherwise. Thus, the estimate of the interaction term of D_p and I_{Skill}, S_p, allows us to assess the skill premium across cities. We find that 90% of the estimated S_p are economically significant at 1% level. Figure 3.3 documents the geographic distribution of skill premium across China's prefectures in 2005. We also scatter-plot skill premium and employment size in Figure C.1 in the Appendix. The positive correlation shows that skill premium are higher in cities with larger population density such as Beijing, Shanghai and Shenzhen.

[Insert Figure 3.3 here]

Empirical Finding.—— We then examine the role of skill premium in firms output quality response to highway development across prefectures using the following specification,

$$WgtQuality_{p,t}^{h,s} = \beta_0 + \beta_1 \ln Highway_{p,t} + \beta_2 \ln Highway_{p,t} \times SkillPremium_{p,2005} + \eta \mathbf{X}_{p,t} + \lambda_p^h + \delta_t^s + \epsilon_{p,t}^{h,s}. \quad (3.10)$$

The coefficients of the interaction terms in Panel A, B and C in Table 3.6 are negative and statistically significant, indicating that higher skill premium (higher relative price for skilled labor) would hurt the positive impact of highway system development on higher-quality production.

[Insert Table 3.6 here]

3.7 Robustness Checks

In this section, we conduct a series of checks to test the robustness of our main results.

Alternative Geographic Unit.—— In Table 3.7, we use provincial-level variations in highway expansion to analyze the response of average product quality to transportation infrastructure. We consider two measures, total length and per capita of the routes in various transportation infrastructure including highway, classified highway, expressway, first-class highway, and railway. We find positive and significant impact of highway, classified highway, expressway, and first-class highway except for the railway on product quality. This is probably because the railway in China increased by only 13.5% between 2001-2007 (68,700 km in 2001, 77,966 km in 2007), in comparison to over 100% growth in highway system during the same period.

[Insert Table 3.7 here]

Government Promotion Policies.—— Furthermore, the results could be confounded by political determinants. The central government and the local governments in China have launched a set of promotion policies to attract FDI and incentivize the economy. Following (Fung, Iizaka and Parker, 2002; Du, Tao and Lu, 2012), we construct two dummy variables to indicate whether a prefecture is a special economic zone or open coastal city, and then exclude observations in these prefectures. The results in Table 3.8 shows that highway expansion consistently exerts a positive and significant impact on product quality.

[Insert Table 3.8 here]

Exclude Large and Developed Cities.—— Next, we wonder whether these effects are only driven by the firms in larger or developed cities. We then exclude observations in a provincial capital and centrally-administrated cities and re-run the baseline specification. In Table 3.10, we exclude the Yangtze River Delta Region. The estimated results in both Table 3.9 and Table 3.10 are quite similar to the baseline analysis.

[Insert Table 3.9 here]

[Insert Table 3.10 here]

Other Robustness Checks.—— Finally, conduct a series of to address the concerns. In Table 3.11 Column (1) presents the baseline results. In Column (2), we use the beginning-of-period prefecture population as a weight in the regression, while in Column (3) the regression is unweighted. The magnitude of the estimated coefficients is similar to that in the baseline regressions and keeps significant at 1% level. We then focus on the dependent variable and replace the weighted average quality measure $WgtQuality_{p,t}^{h,s}$ with the simple average quality measure $AvgQuality_{p,t}^{h,s}$ and re-run the regression, the estimated impact keeps quantitatively similar. We then use unit value and market shares to construct prefecture quality as alternative measures in Column (5)-(6). Surprisingly we only find the highway expansion exerts a significant impact on quality measured by $(\ln MktShare_{p,t}^{h,s})$. Finally, we cluster robust standard errors at the province and industry level to allow for any potential correlations between error terms across the same industry, and prefectures within the same province.

[Insert Table 3.11 here]

3.8 Discussion

It should be pointed out that our results suffer from endogeneity issue. Our identification strategy relies on an *ex-ante* assumption that the route development across prefectures was randomly assigned. However, as discussed in Faber (2014), some peripheral counties and prefectures are not explicitly targeted by the central and local governments during the process of policy setting. Such non-random route placements would potentially bias our estimates. In future work, we shall apply the minimum spanning tree algorithm to construct the least cost path spanning tree networks as the instrumental strategy. Also, we would like to exploit garrison information ("Yizhan" in Chinese) in the Ming Dynasty of Imperial China as an alternative instrumental variable to address the endogeneity.

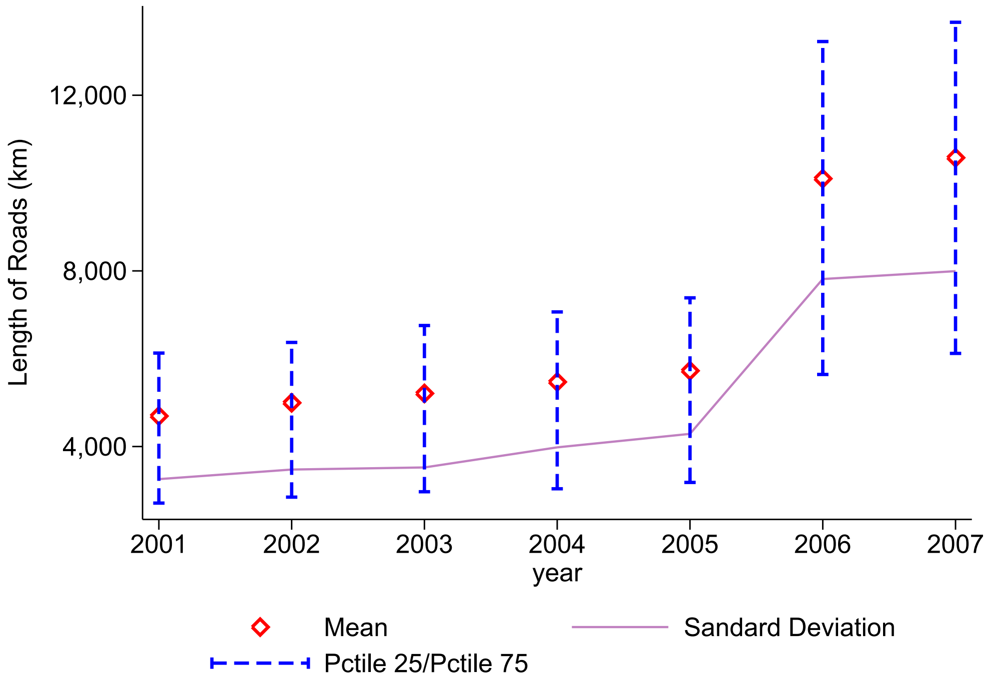
Besides, institutional quality also plays an important role in quality of tradable products as examined in Cui, Yu and Zhang (2018). We shall control for various regional institutional quality measures in our specification, such as *Contract Enforcement* and *Legal*

Institutions. We have already collected these indexes from Fan, Wang and Zhu (2003)'s *China Regional Marketization Indices*.

3.9 Conclusion

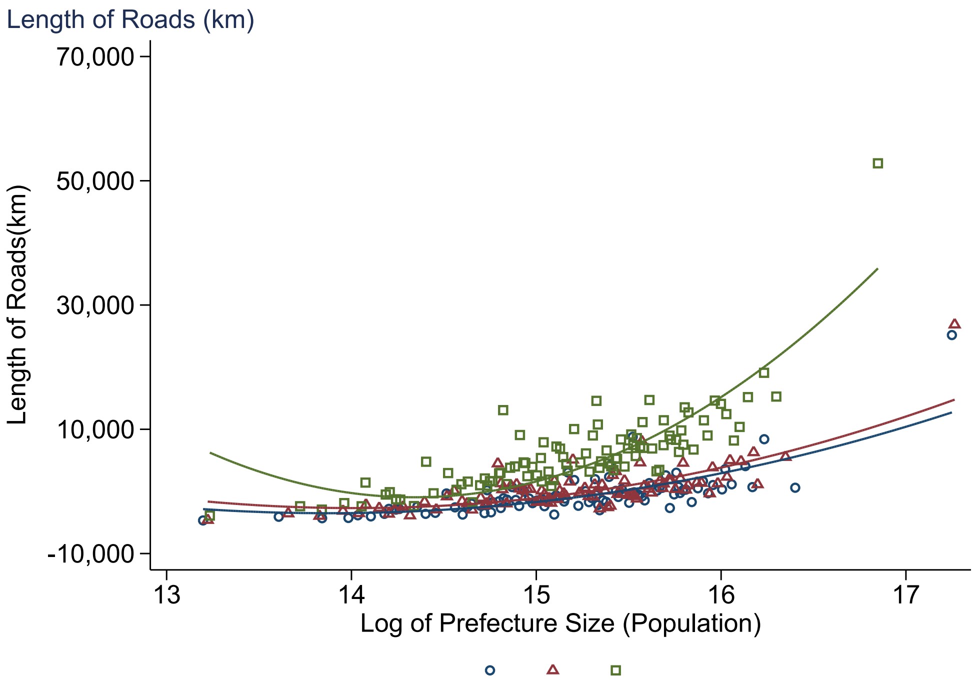
In this paper, we provide empirical evidence on the effects of falling trade costs on product quality across cities within a country. We approach this question in the context of expanding the highway system in China in the past decades, which substantially reduces the trade costs across regions within the nation. Empirically, we combine two firm-level panels that provide unit value information of products across Chinese cities with city-level data on transportation infrastructure for 2001-2007. We find that firms choose to upgrade product quality more in cities with a greater expansion of connecting highways. Besides, this effect is more pronounced in larger cities, which speaks to changes in the spatial concentration of higher-quality products. These results are also robust to the inclusion of an exhaustive battery of fixed effects and changes in estimation specifications. Our findings shed important insights on the impact of falling intranational trade cost on quality specialization pattern across cities, which is difficult to model quantitatively due to the presence of agglomeration and sorting.

Figure 3.1: Highway Expansion in China, Year 2001-2007



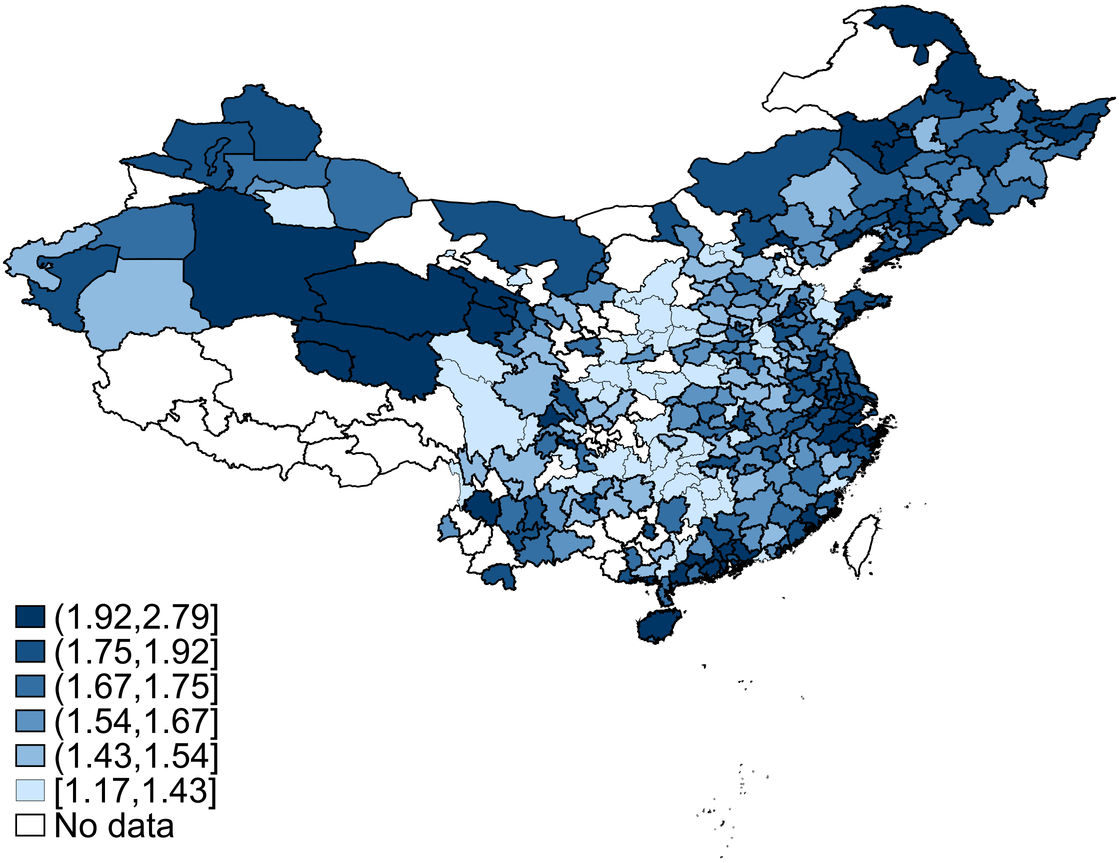
This figure plots prefecture average highway length over 2001-2007. The sample contains 243 prefecture units in each year. The data is from *Regional Economy Statistical Year Books.*

Figure 3.2: Highway Length and Population Size, Year 2001, 2004, 2007



Notes: This figure scatter-plot prefecture highway length and population size and the corresponding quadratic fitted curve during 2001, 2004 and 2007. Blue for 2001, red for 2004 and green for 2007. The sample contains 243 prefecture units in each year. The data is from *Regional Economy Statistical Year Books.*

Figure 3.3: Spatial Distribution of Skill Premium in China, 2005



Data Source: *Regional Economy Statistical Year Books, 2001-2007.*

Table 3.1: Summary Statistics

Panel A. Prefecture Level Variables

	Obs	Mean	Std. Dev.	p10	Median	p90
Average product quality	70,359	-2.931	3.116	-7.187	-2.889	0.909
Log of pref. highway length	1453	8.656	0.685	7.792	8.653	9.516
Log of pref. classified highway length	1453	8.485	0.660	7.603	8.498	9.320
Log of pref. highway per capita	1453	-6.697	0.511	-7.269	-6.794	-5.999
Log of pref. classified highway per capita	1453	-6.869	0.497	-7.446	-6.929	-6.199
Log of pref. highway geo density	1453	3.860	0.995	2.709	3.755	5.190
Log of pref. classified highway geo density	1453	3.689	0.934	2.590	3.635	4.985
Log of pref. population size	1453	15.354	0.594	14.569	15.416	16.028
Log of pref. employment size	1453	13.561	0.793	12.651	13.463	14.616
Log of pref. GDP per capita	1453	9.694	0.752	8.734	9.661	10.724
Log of pref. wage premium	1453	0.516	0.169	0.296	0.508	0.741
Log of pref. urban area	1453	4.796	0.873	3.716	4.730	6.082
Log of pref. graduate share of higher-edu inst	1453	-14.465	3.062	-17.043	-15.361	-8.130
Log of pref. inward FDI	1453	10.929	2.537	7.681	10.892	14.444
Log of pref. urban population density	1453	10.558	0.810	9.682	10.531	11.610
Coastal hub	247	0.281	0.449	0.000	0.000	1.000
Special economic zone	247	0.024	0.153	0.000	0.000	0.000
Open coastal cities	247	0.132	0.339	0.000	0.000	1.000
Economic and technological development zone	247	0.209	0.407	0.000	0.000	1.000

Panel B. Prefecture Level Variables

	Obs	Mean	Std. Dev.	p10	Median	p90
Log of prvc. highway length	172	11.281	0.623	10.729	11.268	12.092
Log of prvc. classified highway length	172	11.107	0.591	10.601	11.155	11.899
Log of prvc. expressway length	172	7.367	0.637	6.390	7.442	8.114
Log of prvc. 1st highway length	172	7.229	1.377	5.231	7.559	8.778
Log of prvc. railway length	172	7.739	0.536	7.131	7.779	8.337
Log of prvc. waterway length	172	8.343	1.391	6.184	8.637	10.082
Log of prvc. highway per capita	172	-6.545	0.489	-7.095	-6.582	-6.067
Log of prvc. classified highway per capita	172	-6.719	0.451	-7.252	-6.727	-6.227
Log of prvc. expressway per capita	172	-10.459	0.437	-10.975	-10.420	-9.990
Log of prvc. 1st highway per capita	172	-10.597	1.328	-12.219	-10.156	-9.350
Log of prvc. railway per capita	172	-10.087	0.625	-10.721	-10.231	-9.215

Notes: This table reports the main variables used in our empirical study and the corresponding summary statistics. The sample includes 28 province level administrative units, 247 prefectures, 29 two-digit CIC manufacturing industries, and 284 five-digit product categories. Panel A reports summary statistics for variables at the prefecture level. Panel B reports summary statistics for variables at the province level.

Table 3.2: Baseline Results - Highway Length

Dependent Variable	$WgtQuality_{p,t}^{h,s}$			
	(1)	(2)	(3)	(4)
$\ln Highway\ Length_{p,t}$	0.179***	0.146***		
	(0.052)	(0.042)		
$\ln Classified\ Highway\ Length_{p,t}$			0.113***	0.092***
			(0.040)	(0.029)
Prefecture Controls	No	Yes	No	Yes
City-Industry FE	Yes	Yes	Yes	Yes
Product-Year FE	Yes	Yes	Yes	Yes
R-squared	0.937	0.937	0.937	0.937
Obs	70,359	70,359	70,359	70,359

Notes: This table examines the effects of highway length on average product quality. The dependent variable $\left(WgtQuality_{p,t}^{h,s}\right)$ is the average quality of product s produced by firms in industry h prefecture p year t and weighted by firm sales. Quality is estimated by Khandelwal (2010)'s approach. The independent variables $(Highway\ Length_{p,t})$ and $(Classified\ Highway\ Length_{p,t})$ are the length of overall highway routes and classified highway routes in prefecture p year t, respectively. Prefecture controls include log of GDP per capita, population, inward FDI, shares of graduates of higher education, and urban area land size. The sample contains 247 prefecture units. All regressions are weighted by the log of beginning-of-period prefecture employment. Robust standard errors in parentheses are clustered at the province level. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.3: Baseline Results - Highway Per Capita

Dependent Variable	$WgtQuality_{p,t}^{h,s}$			
	(1)	(2)	(3)	(4)
$\ln Highway\ pc_{p,t}$	0.177***	0.146***		
	(0.049)	(0.042)		
$\ln Classified\ Highway\ pc_{p,t}$			0.116***	0.092***
			(0.041)	(0.029)
Prefecture Controls	No	Yes	No	Yes
City-Industry FE	Yes	Yes	Yes	Yes
Product-Year FE	Yes	Yes	Yes	Yes
R-squared	0.937	0.937	0.937	0.937
Obs	70,359	70,359	70,359	70,359

Notes: This table examines the effects of highway length per capita on average product quality. The dependent variable $\left(WgtQuality_{p,t}^{h,s}\right)$ is the average quality of product s produced by firms in industry h prefecture p year t and weighted by firm sales. Quality is estimated by Khandelwal (2010)'s approach. The independent variables $(Highway\ pc_{p,t})$ and $(Classified\ Highway\ pc_{p,t})$ are the ratios of overall highway length and classified highway length to population size in prefecture p year t, respectively. Prefecture controls include the log of GDP per capita, population, inward FDI, shares of graduates of higher education, and urban area land size. The sample contains 247 prefecture units. All regressions are weighted by the log of beginning-of-period prefecture employment. Robust standard errors in parentheses are clustered at the province level. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.4: Baseline Results - Highway Geo. Density

Dependent Variable	$WgtQuality^{h,s}_{p,t}$			
	(1)	(2)	(3)	(4)
$Highway\ Geo.\ Density_{p,t}$	0.088***	0.146***		
	(0.029)	(0.042)		
$Classified\ Highway\ Geo.\ Density_{p,t}$			0.073***	0.092***
			(0.024)	(0.029)
Prefecture Controls	No	Yes	No	Yes
City-Industry FE	Yes	Yes	Yes	Yes
Product-Year FE	Yes	Yes	Yes	Yes
R-squared	0.937	0.937	0.937	0.937
Obs	70,359	70,359	70,359	70,359

Notes: This table examines the effects of highway length geographic density on average product quality. The dependent variable $\left(WgtQuality^{h,s}_{p,t}\right)$ is the average quality of product s produced by firms in industry h prefecture p year t and weighted by firm sales. Quality is estimated by Khandelwal (2010)'s approach. The independent variables $(Highway\ Geo.\ Density_{p,t})$ and $(Classified\ Highway\ Geo.\ Density_{p,t})$ are the ratios of overall highway length and classified highway length to urban area land size in prefecture p year t, respectively. Prefecture controls include log of GDP per capita, population, inward FDI, shares of graduates of higher education, and urban area land size. The sample contains 247 prefecture units. All regressions are weighted by the log of beginning-of-period prefecture employment. Robust standard errors in parentheses are clustered at the province level. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.5: Concentration of Higher-Quality Production

Dependent Variable	$WgtQuality^{h,s}_{p,t}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Highway Length						
$\ln Highway\ Length_{p,t}$	0.127**					
	(0.048)					
$\ln Highway\ Length_{p,t} \times Large\ City_p$	0.071*					
	(0.036)					
$\ln Classified\ Highway\ Length_{p,t}$		0.085**				
		(0.034)				
$\ln Classified\ Highway\ Length_{p,t} \times Large\ City_p$		0.016				
		(0.031)				
Panel B. Highway Per Capita						
$\ln Highway\ pc_{p,t}$			0.125**			
			(0.049)			
$\ln Length_{p,t}\ pc \times LargeCity_p$			0.079*			
			(0.039)			
$\ln Classified\ Length\ pc_{p,t}$				0.085**		
				(0.034)		
$\ln Classified\ Length_{p,t}\ pc \times LargeCity_p$				0.016		
				(0.033)		
Panel C. Highway Geo. Density						
$\ln HighwayGeo.Density_{p,t}$					0.131***	
					(0.044)	
$\ln Highway\ Geo.\ Density_{p,t} \times Large\ City_p$					0.065**	
					(0.026)	
$\ln Classified\ Highway\ Geo.\ Density_{p,t}$						0.094***
						(0.030)
$\ln Classified\ Highway\ Geo.\ Density_{p,t} \times Large\ City_p$						-0.003
						(0.025)
Prefecture Controls	Yes	Yes	Yes	Yes	Yes	Yes
City-Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Product-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.937	0.937	0.937	0.937	0.937	0.937
Obs	70,359	70,359	70,359	70,359	70,359	70,359

Notes: This table investigates the impact of highway expansion on the concentration of higher-quality production across China's prefectures between 2001-2007. The dependent variable $\left(WgtQuality^{h,s}_{p,t}\right)$ is the average quality of product s produced by firms in industry h prefecture p year t and weighted by firm sales. Quality is estimated by Khandelwal (2010)'s approach. Panel A, B and C use highway length, highway length per capita and highway geo. density to measure highway development in a prefecture. $Large\ City_p$ is a dummy variable which goes to one if prefecture p is a direct-controlled municipality or provincial capital, and zero otherwise. Prefecture controls include the log of GDP per capita, population, inward FDI, shares of graduates of higher education, and urban area land size. The sample contains 247 prefecture units. All regressions are weighted by the log of beginning-of-period prefecture employment. Robust standard errors in parentheses are clustered at the province level. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.6: Does Skill Premium Matter?

Dependent Variable	$WgtQuality^{h,s}_{p,t}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Highway Length						
$\ln Highway\ Length_{p,t}$	0.352***					
	(0.065)					
$\ln Highway\ Length_{p,t} \times Skill\ Premium_{p,2005}$	-0.442***					
	(0.107)					
$\ln Classified\ Highway\ Length_{p,t}$		0.331***				
		(0.089)				
$\ln Classified\ Highway\ Length_{p,t} \times Skill\ Premium_{p,2005}$		-0.456***				
		(0.162)				
Panel B. Highway Per Capita						
$\ln Highway\ pc_{p,t}$			0.347***			
			(0.064)			
$\ln Length_{p,t}\ pc \times Skill\ Premium_{p,2005}$			-0.427***			
			(0.103)			
$\ln Classified\ Length\ pc_{p,t}$				0.318***		
				(0.088)		
$\ln Classified\ Length_{p,t}\ pc \times Skill\ Premium_{p,2005}$				-0.431**		
				(0.157)		
Panel C. Highway Geo. Density						
$\ln HighwayGeo.Density_{p,t}$					0.231***	
					(0.050)	
$\ln Highway\ Geo.\ Density_{p,t} \times Skill\ Premium_{p,2005}$					-0.189**	
					(0.083)	
$\ln Classified\ Highway\ Geo.\ Density_{p,t}$						0.177***
						(0.052)
$\ln Classified\ Highway\ Geo.\ Density_{p,t} \times Skill\ Premium_{p,2005}$						-0.170*
						(0.096)
Prefecture Controls	Yes	Yes	Yes	Yes	Yes	Yes
City-Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Product-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.937	0.937	0.937	0.937	0.937	0.937
Obs	70,359	70,359	70,359	70,359	70,359	70,359

Notes: This table examines the role of skill premium in firms' quality production response to highway expansion. The dependent variable $\left(WgtQuality^{h,s}_{p,t}\right)$ is the average quality of product s produced by firms in industry h prefecture p year t and weighted by firm sales. Quality is estimated by Khandelwal (2010)'s approach. Panel A, B and C use highway length, highway length per capita and highway geo. density to measure highway development in a prefecture. $Skill\ Premium_{p,2005}$ is estimated by Mincer (1958)'s approach. Prefecture controls include the log of GDP per capita, population, inward FDI, shares of graduates of higher education, and urban area land size. The sample contains 247 prefecture units. All regressions are weighted by the log of beginning-of-period prefecture employment. Robust standard errors in parentheses are clustered at the province level. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.7: Robustness Check - Province level Analysis

Dependent Variable	$WgtQuality^{h,s}_{p,t}$									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A. Route Length										
ln $Prvc.\ Highway\ Length_{k,t}$	0.276***									
	(0.081)									
ln $Prvc.\ Classified\ Highway\ Length_{k,t}$		0.210**								
		(0.079)								
ln $Prvc.\ Expressway\ Length_{k,t}$			0.096							
			(0.057)							
ln $Prvc.\ 1stClass\ Highway\ Length_{k,t}$				0.109**						
				(0.046)						
ln $Prvc.\ Railway\ Length_{k,t}$					0.073					
					(0.066)					
Panel B. Route Length Per Capita										
ln $Prvc.\ Highway\ pc_{k,t}$						0.229***				
						(0.077)				
ln $Prvc.\ Classified\ Highway\ pc_{k,t}$							0.166**			
							(0.071)			
ln $Prvc.\ Expressway\ pc_{k,t}$								0.108*		
								(0.063)		
ln $Prvc.\ 1stClass\ Highway\ pc_{k,t}$									0.109**	
									(0.047)	
ln $Prvc.\ Railway\ pc_{k,t}$										0.047
										(0.060)
Prefecture Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City-Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Product-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.937	0.937	0.937	0.937	0.937	0.937	0.937	0.937	0.937	0.937
Obs	70,503	70,503	70,503	70,503	70,503	70,503	70,503	70,503	70,503	70,503

Notes: This table examines the impact of province level highway development on product quality during 2001-2007. The dependent variable $\left(WgtQuality^{h,s}_{p,t}\right)$ is the average quality of product s produced by firms in industry h prefecture p year t and weighted by firm sales. Quality is estimated by Khandelwal (2010)'s approach. Panel A and B report results using $RouteLength_k$ and $RouteLengthpercapita_k$to measure various infrastructure in province k year t. Prefecture controls include the log of GDP per capita, population, inward FDI, shares of graduates of higher education, and urban area land size. The sample contains 247 prefecture units. All regressions are weighted by the log of beginning-of-period prefecture employment. Robust standard errors in parentheses are clustered at the province level. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.8: Robustness Check: Exclude SEZ and Open Coastal Cities

Dependent Variable	$WgtQuality^{h,s}_{p,t}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Highway Length						
$\ln Highway\ Length_{p,t}$	0.145***					
	(0.045)					
$\ln\ Classified\ Highway\ Length_{p,t}$		0.059*				
		(0.030)				
Panel B. Highway Length Per Capita						
$\ln Highway\ Length\ pc_{p,t}$			0.145***			
			(0.045)			
$\ln Classified\ Highway\ Length\ pc_{p,t}$				0.059*		
				(0.030)		
Panel C. Highway Geo. Density						
$\ln Highway\ Geo.\ Density_{p,t}$					0.145***	
					(0.045)	
$\ln Classified\ Highway\ Geo.\ Density_{p,t}$						0.058***
						(0.064)
Prefecture Controls	Yes	Yes	Yes	Yes	Yes	Yes
City-Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Product-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.935	0.935	0.935	0.935	0.935	0.935
Obs	59,360	59,360	59,360	59,360	59,360	59,360

Notes: This table examines the relationship between highway expansion and product quality in prefectures excluding special economic zones and open coastal cities during 2001-2007 . The dependent variable $\left(WgtQuality^{h,s}_{p,t}\right)$ is the average quality of product s produced by firms in industry h prefecture p year t and weighted by firm sales. Quality is estimated by Khandelwal (2010)'s approach. Panel A, B and C use highway length, highway length per capita and highway geo. density to measure highway development in a prefecture. Prefecture controls include the log of GDP per capita, population, inward FDI, shares of graduates of higher education, and urban area land size. The sample contains 247 prefecture units. All regressions are weighted by the log of beginning-of-period prefecture employment. Robust standard errors in parentheses are clustered at the province level. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.9: Robustness Check - Exclude Provincial Capital Cities and Centrally-Administrated Cities

Dependent Variable	$WgtQuality^{h,s}_{p,t}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Highway Length						
$\ln Highway\ Length_{p,t}$	0.117*					
	(0.059)					
$\ln\ Classified\ Highway\ Length_{p,t}$		0.123***				
		(0.041)				
Panel B. Highway Length Per Capita						
$\ln Highway\ Length\ pc_{p,t}$			0.117*			
			(0.059)			
$\ln Classified\ Highway\ Length\ pc_{p,t}$				0.123***		
				(0.041)		
Panel C. Highway Geo. Density						
$\ln Highway\ Geo.\ Density_{p,t}$					0.117*	
					(0.059)	
$\ln Classified\ Highway\ Geo.\ Density_{p,t}$						0.123***
						(0.041)
Prefecture Controls	Yes	Yes	Yes	Yes	Yes	Yes
City-Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Product-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.939	0.939	0.939	0.939	0.939	0.939
N	55,010	55,010	55,010	55,010	55,010	55,010

Notes: This table examines the relationship between highway expansion and product quality in prefectures excluding provincial capital cities and four centrally-administrated cities during 2001-2007. The dependent variable $\left(WgtQuality^{h,s}_{p,t}\right)$ is the average quality of product s produced by firms in industry h prefecture p year t and weighted by firm sales. Quality is estimated by Khandelwal (2010)'s approach. Panel A, B and C use highway length, highway length per capita and highway geo. density to measure highway development in a prefecture. Prefecture controls include the log of GDP per capita, population, inward FDI, shares of graduates of higher education, and urban area land size. The sample contains 247 prefecture units. All regressions are weighted by the log of beginning-of-period prefecture employment. Robust standard errors in parentheses are clustered at the province level. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.10: Robustness Check - Exclude Yangtze River Delta Region

	$WgtQuality_{p,t}^{h,s}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Highway Length						
$\ln Highway\ Length_{p,t}$	0.182***					
	(0.042)					
$\ln\ Classified\ Highway\ Length_{p,t}$		0.129***				
		(0.041)				
Panel B. Highway Length Per Capita						
$\ln Highway\ Length\ pc_{p,t}$			0.182***			
			(0.042)			
$\ln Classified\ Highway\ Length\ pc_{p,t}$				0.129***		
				(0.041)		
Panel C. Highway Geo. Density						
$\ln Highway\ Geo.\ Density_{p,t}$					0.182***	
					(0.042)	
$\ln Classified\ Highway\ Geo.\ Density_{p,t}$						0.129***
						(0.041)
Prefecture Controls	Yes	Yes	Yes	Yes	Yes	Yes
City-Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Product-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.935	0.935	0.935	0.935	0.935	0.935
Obs	53,762	53,762	53,762	53,762	53,762	53,762

Notes: This table examines the relationship between highway expansion and product quality in prefectures excluding Yangtze River Delta Region during 2001-2007. The dependent variable $\left(WgtQuality_{p,t}^{h,s}\right)$ is the average quality of product s produced by firms in industry h prefecture p year t and weighted by firm sales. Quality is estimated by Khandelwal (2010)'s approach. Panel A, B, and C use highway length, highway length per capita, and highway geo. density to measure highway development in a prefecture. Prefecture controls include the log of GDP per capita, population, inward FDI, shares of graduates of higher education, and urban area land size. The sample contains 247 prefecture units. All regressions are weighted by the log of beginning-of-period prefecture employment. Robust standard errors in parentheses are clustered at the province level. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.11: Robustness Check - Other Checks

	Baseline	Reg Weight: population	No Weight	Unweighted Est. Quality	WgtAvg Log Prices	WgtAvg Log Mkt Shares	Cluster: Prvc.-Ind.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Highway Length							
$\ln Highway\ Length_{p,t}$	0.146***	0.113**	0.108**	0.189***	-0.042	0.154*	0.146***
	(0.042)	(0.045)	(0.051)	(0.058)	(0.068)	(0.084)	(0.029)
Prefecture Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City-Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Product-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.937	0.934	0.935	0.952	0.879	0.524	0.937
Obs	70,359	70,359	70,359	70,359	70,359	70,359	70,359
Panel B. Highway Length Per Capita							
$\ln Highway\ Length\ pc_{p,t}$	0.146***	0.113**	0.108**	0.189***	-0.042	0.154*	0.146***
	(0.042)	(0.045)	(0.051)	(0.058)	(0.068)	(0.084)	(0.029)
Prefecture Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City-Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Product-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.937	0.934	0.935	0.952	0.879	0.524	0.937
Obs	70,359	70,359	70,359	70,359	70,359	70,359	70,359
Panel C. Highway Geo. Density							
$\ln Highway\ Geo\ Density_{p,t}$	0.146***	0.113**	0.108**	0.189***	-0.042	0.154*	0.146***
	(0.042)	(0.045)	(0.051)	(0.058)	(0.068)	(0.084)	(0.029)
Prefecture Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City-Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Product-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.937	0.934	0.935	0.952	0.879	0.524	0.937
Obs	70,359	70,359	70,359	70,359	70,359	70,359	70,359

Notes: This table presents several robustness check on the impact of highway expansion on product quality in China. The dependent variable in Column (1), (2), (3) and (7) $\left(WgtQuality^{h,s}_{p,t}\right)$ is the average quality of product s produced by firms in industry h prefecture p year t and weighted by firm sales. Quality is estimated by Khandelwal (2010)'s approach. The dependent variable in Column (4) is the simple average quality of product s produced by firms in industry h prefecture p year t. The dependent variable in Column (5) and (6) are the average price and market share of product s produced by firms in industry h prefecture p year t, weighted by firm sales, respectively. All regressions control Prefecture level features including the log of GDP per capita, population, inward FDI, shares of graduates of higher education, and urban area land size. The sample contains 247 prefecture units. Regressions in Column (1), (4), (5), (6) and (7) are weighted by the log of beginning-of-period prefecture employment. Regressions in Column (2) is weighted by the log of the beginning-of-period prefecture population. Regressions in Column (3) is unweighted. Robust standard errors in parentheses in Column (1)-(6) are clustered at the province level. Robust standard errors in parentheses in Column (7) are clustered at the province-industry level.

References

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer, "Identification Properties of Recent Production Function Estimators," *Econometrica*, 2015, *83* (6), 2411–2451.

Ahlfeldt, Gabriel M, Stephen J Redding, Daniel M Sturm, and Nikolaus Wolf, "The Economics of Density: Evidence from the Berlin Wall," *Econometrica*, 2015, *83* (6), 2127–2189.

Albouy, David, "Are Big Cities Bad Places to Live? Estimating Quality of Life across Metropolitan Areas," *Working Paper*, 2012.

Allen, Treb, "Information Frictions in Trade," *Econometrica*, 2014, *82* (6), 2041–2083.

_ and David Atkin, "Volatility and the Gains from Trade," *Working Paper*, 2016.

Amiti, Mary and Amit K. Khandelwal, "Import Competition and Quality Upgrading," *Review of Economics and Statistics*, 2013, *95* (2), 476–490.

Angeletos, George-Marios and Jennifer La'O, "Sentiments," *Econometrica*, 2013, *81* (2), 739–779.

Antoniades, Alexis, "Heterogeneous Firms, Quality, and Trade," *Journal of International Economics*, 2015, *95*, 263–273.

Antràs, Pol, Teresa C. Fort, and Felix Tintelnot, "The Margins of Global Sourcing: Theory and Evidence from U.S. Firms," *American Economic Review*, 2017, *107* (9), 2514–64.

Arkolakis, Costas, Theodore Papageorgiou, and Olga A. Timoshenko, "Firm Learning and Growth," *Review of Economic Dynamics*, 2018, *27*, 146–168.

Bacolod, Marigee, Bernardo S. Blum, and William C. Strange, "Skills in the City," *Journal of Urban Economics*, 2009, *65* (2), 136–153.

Bai, Jie, "Melons as Lemons: Asymmetric Information, Consumer Learning and Quality Provision," *Working Paper*, 2018.

—, Maggie Chen, Jin Liu, and Daniel Xu, "Search and Information Frictions on Global E-Commerce Platforms: Evidence from Aliexpress," *Working Paper*, 2018.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian, "On the road: Access to Transportation Infrastructure and Economic Growth in China," Technical Report, National Bureau of Economic Research Working Paper 17897 2018.

Baum-Snow, Nathaniel and Ronni Pavan, "Understanding the City Size Wage Gap," *Review of Economic Studies*, 2012, *79* (1), 88–127.

— and —, "Inequality and City Size," *Review of Economics and Statistics*, 2013, *95* (5), 1535–1548.

—, Matthew Freedman, and Ronni Pavan, "Why Has Urban Inequality Increased?," *American Economic Journal: Applied Economics*, 2018, *10* (4), 1–42.

Behrens, Kristian and Frédéric Robert-Nicoud, "Agglomeration Theory with Heterogeneous Agents," in Gilles Duranton, J. Vernon Henderson, and Wiliam C. Strange, eds., *Handbook of Regional and Urban Economics*, Vol. 5, Amsterdam: Elsevier, 2015, pp. 171–245.

—, Gilles Duranton, and Frédéric Robert-Nicoud, "Productive Cities: Sorting, Selection, and Agglomeration," *Journal of Political Economy*, 2014, *122* (3), 507–553.

Berry, Steven T, "Estimating Discrete-Choice Models of Product differentiation," *The RAND Journal of Economics*, 1994, pp. 242–262.

Black, Dan, Natalia Kolesnikova, and Lowell Taylor, "Earnings Functions When Wages and Prices Vary by Location," *Journal of Labor Economics*, 2009, *27* (1), 21–47.

Brandt, Loren, Johannes Van Biesebroeck, and Yifan Zhang, "Creative Accounting or Creative Destruction? Firm-Level Productivity Growth in Chinese Manufacturing," *Journal of Development Economics*, 2012, *97* (2), 339–351.

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

Chen, Maggie and Min Wu, "The Value of Reputation in Trade: Evidence from Alibaba," *Working Paper*, 2016.

Choi, Yo Chul, David Hummels, and CHong Xiang, "Explaining Import Quality: The Role of the Income Distribution," *Journal of International Economics*, 2009, *78* (2), 293–303.

Chor, Davin, "Institutions, Wages, and Inequality: The Case of Europe and Its Periphery (1500–1899)," *Explorations in Economic History*, 2005, *42* (4), 547–566.

_ , "Unpacking Sources of Comparative Advantage: A Quantitative Approach," *Journal of International Economics*, 2010, *82* (2), 152–167.

_ and Kalina Manova, "Off the Cliff and Back? Credit Conditions and International Trade During the Global Financial Crisis," *Journal of International Economics*, 2012, *87* (1), 117–133.

Ciccone, Antonio and Robert E. Hall, "Productivity and the Diversity of Economic Activity," *American Economic Review*, 1996, *86* (1), 54–70.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon, "Spatial Wage Disparities: Sorting Matters!," *Journal of Urban Economics*, 2008, *63* (2), 723–742.

_ , _ , _ , Diego Puga, and Sébastien Roux, "The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection," *Econometrica*, 2012, *80* (6), 2543–2594.

Couture, Victor, Cecile Gaubert, Jessie Handbury, and Erik Hurst, "Income Growth and the Distributional Effects of Urban Spatial Sorting," *Working Paper*, 2019.

Crozet, Matthieu, Keith Head, and Thierry Mayer, "Quality Sorting and Trade: Firm-Level Evidence for French Wine," *Review of Economic Studies*, 2012, *79* (2), 609–644.

Cui, Xiaomin, Miaojie Yu, and Rui Zhang, "Judicial Quality, Incomplete Contract, and Quality of Trade," 2018.

Davis, Donald R. and Jonathan I. Dingel, "The Comparative Advantage of Cities," *Working Paper*, 2017.

_ and _ , "A Spatial Knowledge Economy," *American Economic Review*, 2019, *109* (1), 153–70.

_ , Eric Mengus, and Tomasz Michalski, "Labor Market Polarization in a System of Cities: The Case of France," *Working Paper*, 2018.

D'Costa, Sabine and Henry G. Overman, "The Urban Wage Growth Premium: Sorting or Learning?," *Regional Science and Urban Economics*, 2014, *48*, 168–179.

Diamond, Rebecca, "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000," *American Economic Review*, 2016, *106* (3), 479–524.

Dingel, Jonathan I., "The Determinants of Quality Specialization," *Review of Economic Studies*, 2017, *84* (4), 1551–1582.

_ , Antonio Miscio, and Donald R. Davis, "Cities, Lights, and Skills in Developing Economies," *Journal of Urban Economics*, 2019, *Forthcoming.*

Donaldson, Dave, "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review*, 2018, *108* (4-5), 899934.

_ and Richard Hornbeck, "Railroads and American Economic Growth: A "Market Access Approach," *Quarterly Journal of Economics*, 2016, *131* (2), 799–858.

Du, Julan, Zhigang Tao, and Yi Lu, "Institutions, Culture Distance, and FDI Location Choice: Evidence from China," *Journal of Asian Economics*, 2012.

Duranton, Gilles and Diego Puga, "Micro-Foundations of Urban Agglomeration Economies," in J. Vernon Henderson and Jacques-François Thisse, eds., *Handbook of Regional and Urban Economics*, Vol. 4, Amsterdam: Elsevier, 2004, pp. 2063–2117.

_ and Matthew A Turner, "Urban Growth and Transportation," *Review of Economic Studies*, 2012, *79* (4), 1407–1440.

Eaton, Jonathan, David Jinkins, James Tybout, and Daniel Xu, "Two-Sided Matching in International Markets," *Working Paper*, 2018.

_ , Marcela Eslava, David Jinkins, C. J. Krizan, and James Tybout, "A Search and Learning Model of Export Dynamics," *Working Paper*, 2014.

_ , Samuel Kortum, and Francis Kramarz, "An Anatomy of International Trade: Evidence From French Firms," *Econometrica*, 2011, *79* (5), 1453–1498.

Faber, Benjamin, "Trade Integration, Market Size, and Industrialization: Wvidence from China's National Trunk Highway System," *Review of Economic Studies*, 2014, *81* (3), 1046–1070.

_ and Thibault Fally, "Firm Heterogeneity in Consumption Baskets: Evidence from Home and Store Scanner Data," *Working Paper*, 2017.

Fajgelbaum, Pablo, Gene M. Grossman, and Elhanan Helpman, "Income Distribution, Product Quality, and International Trade," *Journal of Political Economy*, 2011, *119* (4), 721–765.

_ , _ , and _ , "A Linder Hypothesis for Foreign Direct Investment," *Review of Economic Studies*, 2015, *82* (1), 83–121.

Fan, Gang, Xiaolu Wang, and Henpeng Zhu, "Reports on the Relative Progress of Marketization in Different Regions in China," *Economic Research Journal*, 2003, *3* (9), 18.

Fan, Haichao, Xiang Gao, Amber Yao Li, and Tuan Anh Luong, "Trade Liberalization and Markups: Micro Evidence from China," *Journal of Comparative Economics*, 2018, *46* (1), 103–130.

_ , Yao Amber Li, Sichuang Xu, and Stephen R. Yeaple, "Quality, Variable Markups, and Welfare: A Quantitative General Equilibrium Analysis of Export Prices," *Working Paper*, 2017.

Fan, Jingting, "Internal Geography, Labor Mobility, and the Distributional Impacts of Trade," *American Economic Journal: Macroeconomics*, 2019, *11* (3), 252–88.

Farrokhi, Farid and David Jinkins, "Wage Inequality and the Location of Cities," *Journal of Urban Economics*, 2019, *111*, 76–92.

Feenstra, Robert C. and John Romalis, "International Prices and Endogenous Quality," *Quarterly Journal of Economics*, 2014, *129* (2), 477–527.

Fieler, Ana Cecília, Marcela Eslava, and Daniel Yi Xu, "Trade, Quality Upgrading, and Input Linkages: Theory and Evidence from Colombia," *American Economic Review*, 2018, *108* (1), 109–46.

Florida, Richard, Charlotta Mellander, Kevin Stolarick, and Adrienne Ross, "Cities, Skills and Wages," *Journal of Economic Geography*, 2012, *12* (2), 355–377.

Fung, Kwok-Chiu, Hitomi Iizaka, and Stephen Parker, "Determinants of US and Japanese Direct Investment in China," *Journal of Comparative Economics*, 2002, *30* (3), 567–578.

Gaubert, Cecile, "Firm Sorting and Agglomeration," *American Economic Review*, 2018, *108* (11), 3117–3153.

Glaeser, Edward L., "Learning in Cities," *Journal of Urban Economics*, 1999, *46* (2), 254–277.

_ and David Maré, "Cities and Skills," *Journal of Labor Economics*, 2001, *19* (2), 316–342.

_ , Hedi D. Kallal, José A. Scheinkman, and Andrei Shleifer, "Growth in Cities," *Journal of Political Economy*, 1992, *100* (6), 1126–1152.

_ , Jed Kolko, and Albert Saiz, "Consumer City," *Journal of Economic Geography*, 2001, *1* (1), 27–50.

Hallak, Juan Carlos, "Product Quality and the Direction of Trade," *Journal of International Economics*, 2006, *68* (1), 238–265.

— , "A Product-Quality View of the Linder Hypothesis," *Review of Economics and Statistics*, 2010, *92* (3), 453–466.

— and Jagadeesh Sivadasan, "Product and Process Productivity: Implications for Quality Choice and Conditional Exporter Premia," *Journal of International Economics*, 2013, *91* (1), 53–67.

Handbury, Jessie and David E. Weinstein, "Goods Prices and Availability in Cities," *Review of Economic Studies*, 2015, *82* (1), 258–296.

Helpman, Elhanan, "The Size of Regions," *Unpublished manuscript*, 1998.

Hendricks, Lutz, "The Skill Composition of U.S. Cities," *International Economic Review*, 2011, *52* (1), 1–32.

Hummels, David and Alexandre Skiba, "Shipping the Good Apples Out? An Empirical Confirmation of the Alchian-Allen Conjecture," *Journal of Political Economy*, 2004, *112* (6), 1384–1402.

— and Peter J. Klenow, "The Variety and Quality of a Nation's Exports," *American Economic Review*, 2005, *95* (3), 704–723.

Jiao, Yang and Lin Tian, "Geographic Fragmentation in a Knowledge Economy," *Working Paper*, 2019.

Juhsz, Rka and Claudia Steinwender, "Spinning the Web: Codifiability, Information Frictions and Trade," *Working Paper*, 2019.

Kelly, Morgan, "The Standard Errors of Persistence," *Working Paper*, 2019.

Khandelwal, Amit, "The Long and Short (of) Quality Ladders," *Review of Economic Studies*, 2010, *77* (4), 1450–1476.

Kugler, Maurice and Eric A. Verhoogen, "Prices, Plant Size, and Product Quality," *Review of Economic Studies*, 2012, *79* (1), 307–339.

Levinsohn, James and Amil Petrin, "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 2003, *70* (2), 317–341.

Lindley, Joanne and Stephen Machin, "The Rising Postgraduate Wage Premium," *Economica*, 2016, *83* (330), 281–306.

Lucas, Robert E. and Edward C. Prescott, "Equilibrium Search and Unemployment," *Journal of Economic Theory*, 1974, *7* (2), 188–209.

Ma, Lin and Yang Tang, "A Tale of Two Tails: Wage Inequality and City Size," *Working Paper*, 2018.

_ and _ , "Geography, Trade, and Internal Migration in China," *Journal of Urban Economics, forthcoming*, 2019.

Marshall, Alfred, "Principles of Economics," *London: MacMillan*, 1920.

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

Mincer, Jacob, "Investment in Human Capital and Personal Income Distribution," *Journal of Political Economy*, 1958, *66* (4), 281–302.

Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg, "Commuting, Migration and Local Employment Elasticities," *American Economic Review*, 2018, *108* (12), 3855–90.

Moretti, Enrico, "Real Wage Inequality," *American Economic Journal: Applied Economics*, 2013, *5* (1), 65–103.

Olley, G. Steven and Ariel Pakes, "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 1996, *64* (6), 1263–1297.

Picard, Pierre M, "Trade, Economic Geography and the Choice of Product Quality," *Regional Science and Urban Economics*, 2015, *54*, 18–27.

Picard, Pierre M. and Toshihiro Okubo, "Firm's Locations under Demand Heterogeneity," *Regional Science and Urban Economics*, 2012, *42* (6), 961–974.

Piveteau, Paul and Gabriel Smagghue, "Foreign Competition along the Quality Ladder," *Working Paper*, 2019.

Roca, Jorge De La and Diego Puga, "Learning by Working in Big Cities," *Review of Economic Studies*, 2017, *84* (1), 106–142.

Rosenthal, Stuart S. and William C. Strange, "Evidence on the Nature and Sources of Agglomeration Economies," in J. Vernon Henderson and Jacques-François Thisse, eds., *Handbook of Regional and Urban Economics*, Vol. 4, Amsterdam: Elsevier, 2004, pp. 2119–2171.

Rossi-Hansberg, Esteban, "A Spatial Theory of Trade," *American Economic Review*, 2005, *95* (5), 1464–91.

Rovigatti, Gabriele and Vincenzo Mollisi, "PRODEST: Stata Module for Production Function Estimation Based on the Control Function Approach," 2018.

Saito, Hisamitsu and Toshiyuki Matuura, "Agglomeration Economies, Productivity, and Quality Upgrading," *RIETI Discussion Paper Series 16-E-085*, 2016.

Schott, Peter K., "Across-Product Versus Within-Product Specialization in International Trade," *Quarterly Journal of Economics*, 2004, *82* (2), 647–678.

Steinwender, Claudia, "Real Effects of Information Frictions: When the States and the Kingdom Became United," *American Economic Review*, 2018, *108* (3), 657–696.

Tian, Lin, "Division of Labor and Productivity Advantage of Cities: Theory and Evidence from Brazil," 2018.

Tsivanidis, Nick, "The Aggregate and Distributional Effects of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio," *Working Paper*, 2018.

Verhoogen, Eric A., "Trade, Quality Upgrading, and Wage Inequality in the Mexican Manufacturing Sector," *Quarterly Journal of Economics*, 2008, *123* (2), 489–530.

Wang, Yao and Jie Li, "ICT's Effect on Trade: Perspective of Comparative Advantage," *Economics Letters*, 2017, *155*, 96–99.

Yang, Rudai, "Study on the Total Factor Productivity of Chinese Manufacturing Enterprises," *Economic Research Journal*, 2015, *2*, 61–74.

Yasar, Mahmut, Rafal Raciborski, and Brian Poi, "Production Function Estimation in Stata using the Olley and Pakes Method," *The Stata Journal*, 2008, *8* (2), 221–231.

Zhao, Yingyan, "Your (Countrys) Reputation Precedes You: Information Asymmetry, Externalities and the Quality of Exports," *Working Paper*, 2018.

A Appendix to Chapter 1

A.1 Data and Measurement

A.1.1 Data

We use two datasets, both of which are collected by the National Bureau of Statistics of China. The first dataset is the Annual Survey of Industrial Firms (ASIF). This dataset provides firm-level information on sales, profits, taxes, investment, intermediate input expenditure, labor expenditure, and education level of workers. The second dataset is the Industrial Firms Product Quantity Dataset which contains information on the physical quantity of outputs produced by firms. Both datasets cover a similar universe of firms and share the same identifier. We combine the two databases in order to match the quantity data with the sales data. We use the data for three purposes. First, we construct an unbalanced panel of firms using the ASIF and use them to estimate production functions of the firms, in order to obtain an estimate of firm productivity. Second, we use both the ASIF dataset and physical quantity dataset to extract information on unit value and market share of products. These information are then used in establishing some stylized facts on firm heterogeneity and quality specialization in Section 2.4. Lastly, we use the firm-level information in the ASIF to structurally estimate a spatial-equilibrium model. We then use the unit value information extracted from both the ASIF and physical quantity dataset to externally validate our estimated model.

We will now describe the handling of each dataset first before discussing the details on how we match the two databases.

Annual Survey of Industrial Firms.—— The ASIF dataset we used covers a time period from 2000 to 2007. Although the ASIF dataset is also available for recent years, the data in this time period is well-known for its high quality and have also been the focus of many other research. This dataset covers the universe of manufacturing firms in China with an annual gross sales more than 5 million RMB. Both state-owned and private firms are included in the survey. We follow Brandt, Van Biesebroeck and Zhang (2012) to construct an unbalanced panel. Following their approach, we first match the firms by their firm ID if available, or else by firm names if available, or else by the name of legal person representative if available, or else by telephone number registered by the firm. The attrition rate of the matching each two-consecutive years ranges from 9.2% to 23.7% and exhibits a decreasing time trend.

Industrial Firms Product Quantity Database.—— The physical quantity dataset we used covers a similar universe of manufacturing firms from 2000 to 2007. This dataset provides five-digit product-level quantity information of each firm. We merge the product information and compute the average unit value of a firm across all products that it produces, since in our model a firm only produce one variety and all varieties in the same sector are essentially competing with each other. To merge the physical quantity dataset with the ASIF dataset, we match the firms by firm ID if available, or else by firm name. The attrition rate is around 60%.

A.1.2 Sector Concordance

We concord the data into a two-digit sector definition that is similar to those in Gaubert (2018) and Caliendo and Parro (2015). The reason that we do not directly use the two-digit sector definition in the Chinese classification is that there are several dozens of such sectors. As the estimation of each sector takes about 1 day, it would be computationally infeasible to estimate the model at this level of aggregation. Hence, we decided to follow the sector definition in Gaubert (2018) and Caliendo and Parro (2015) as much as possible which is primarily based on ISIC sector classifications. The details of our concordance is summarized in below as Table A.1 .

Table A.1: Concordance of Sectors

Number	Industry	Description	CSIC Rev. 2
1	Food	Food, beverages, and tobaccos	14-16
2	Textile	Textiles and apparels	17-18
3	Leather	Leather, furs, footwear, and related products	19
4	Wood	Wood and products of wood, except furniture	20
5	Furniture	Furniture	21
6	Paper	Pulp, paper, paper products, printing, and publishing	22-24
7	Chemicals	Chemical materials and chemical products	26, 28
8	Medical	Medical and pharmaceutical products	27
9	Plastic	Rubber and plastic products	29-30
10	Minerals	Nonmetallic mineral products	31
11	Basic metals	Basic metals and fabricated metals	32-34
12	Machinery	Machinery and equipment	35-36
13	Transport	Transport equipment and automotive	37
14	Electrical	Electric equipment and machinery	39
15	Computer	Computer and office machinery	40-41
16	Energy	Supplying of energy	44-46
17	Others	Manufacturing n.e.c.	42

A.1.3 Total Factor Productivity (TFP)

To measure firm's (ex-post) productivity, we first implement the approach of Olley and Pakes (1996) (hereafter O-P). The O-P approach estimates the production function semi-parametrically and addresses the simultaneity and selection issues in the OLS estimation. We assume firms within each two-digit industry share similar production pattern and estimate the coefficients for labor and capital inputs to better capture firm-level TFP in different industries. The details of constructing variables are as follows. [21]

Capital Stock and Investment. —— The ASIF database reports the book value of fixed capital stock at original purchase prices, and the book value of fixed capital stock at original purchase prices less accumulated depreciation. We can not directly use these

[21]We drop mining and petroleum industries in TFP estimation. This is because the production function assumption does not hold for industries that heavily depend on natural resources.

measures to back out firm's capital stock, because they are the sum of nominal values of the past years.

We recover the real capital investment following the approach developed by Brandt, Van Biesebroeck and Zhang (2012) and Yang (2015). To compute real capital investment after the first period a firm appearing in our database, we simply compute the difference of fixed capital stock at original purchase prices between current year and last year, and deflate this nominal capital investment using Brant-Rawski capital deflator. To proxy the real capital investment of firms appearing in the first year in the database, the ideal approach is to explore the past information on capital stocks of a firm and construct the real capital investment for each past year. Due to data limitation, we can not do this. We then make use of the information on establishment date, and assume a firm on average exhibits the same growth rate in nominal capital stocks from the established year to the first period it appears in the database. The imputed capital growth rate is obtained using the two-digit sector-province average growth rate since 1993. With the age of each firm, and the nominal capital stock of the first period a firm appears, we can back out the nominal capital stock in the year when a firm was established. The equation is

$$NK_E(1 + r_{ps})^{age} = NK_C,$$

where NK_E represents the nominal capital in the year when the firm was established, and NK_C is the nominal capital in the year when the firm first appears in the database. r_{ps} is the imputed two-digit industrial nominal capital growth rate in each province and age represents number of years since the firm established.

With nominal capital stocks of each year since the firm was established, we then apply the perpetual inventory method to back out firm's real capital stock using the depreciation rate and real deflated capital investment. The depreciation rate is assumed to be 0.09. The capital deflator we use is Brandt-Rawski deflator.

Value Added. —— We use value added as a firm's output to estimate TFP. The nominal value added can be observed directly in each year in the ASIF except 2004. We construct the nominal value added in 2004 using the nominal gross output value subtracted from value added taxes payable, and intermediary inputs.

To obtain the real value added, we construct the output deflator by exploiting the information on output price indexes for each two-digit industry in each year from China Statistical Yearbooks. The real value added can be obtained by deflating firm's nominal value added with the output deflator in each year.

Labor Payment. —— Firms report the total wage payable, employee supplementary benefit and unemployment insurance in each year. We sum across these three categories as the labor payment of each firm. Notice that this input measure is still the nominal value and should be deflated. We generate the input deflator for each sector as the weighted average of the output deflators using China's National Input-Output Table in 1998, 2002 and 2007, which also helps us to capture the changes in input prices of different industries.

With the variables constructed, we follow the literature and implement Olley and Pakes (1996) approach to estimate TFP by applying the Stata package provided by Yasar, Raciborski and Poi (2008). This approach eliminates the concerns of both simultaneity issues and selection biases, which can not be addressed in the traditional TFP estimations.

A.2 A Microfoundation for Within-City Worker Sorting

Suppose that in a city with L_s skilled workers and L_u unskilled workers, the wages of the workers are w_s and w_u respectively. We follow our assumption in the benchmark model that the city consists of two separated areas downtown (D) and suburb (S) each with 1 unit of land. Furthermore, assuming that the workers have Stone-Geary preference over consumption and housing in the sense that they must consume a minimum amount of floor space $\bar{h}$,

$$U = v\left(\frac{C}{\alpha}\right)^{\alpha}\left(\frac{H-\bar{h}}{1-\alpha}\right)^{1-\alpha}.$$

where v is a random utility component that is drawn from a Frechet distribution with a shape parameter θ and a scale parameter normalized to 1. The budget constraint of a worker of type $\zeta \in \{s, u\}$ who lives in location $n \in \{\mathrm{D}, \mathrm{S}\}$ is $PC^n + p_H^n H^n = w_\zeta$. Without loss of generality, we can label the areas such that $p_H^D \geqslant p_H^S$. The fact that house prices in the downtown area is higher than that in the suburb area could be due to a variety of reasons such as higher amenity, transportation cost, etc (Tsivanidis, 2018; Couture, Gaubert, Handbury and Hurst, 2019). We omit all these factors here for simplicity. Our model for the microfoundation can be considered as a special case of the within-city spatial sorting model in the literature (Tsivanidis, 2018; Couture, Gaubert, Handbury and Hurst, 2019) and yield similar conclusions.

Given these assumptions, we can show that the indirect utility of a ζ-type worker living in n is

$$U_\zeta^n = v\frac{w_\zeta - p_H^n\bar{h}}{(p_H^n)^{1-\alpha}P^\alpha} \equiv v\bar{U}_\zeta^n.$$

We show that in equilibrium, skilled workers sort more into downtown areas while unskilled workers choose to live more in suburb. The intuition is that, as a consequence of the Stone-Geary preference, richer workers will spend a smaller share of their income on housing and will be more likely choose to live in an area with a higher housing price. The exact argument proceeds as follows.

Given the Frechet assumption, we can write the fraction of workers with wage w_ζ that choose to live in the downtown area as,

$$\pi_D^\zeta = \text{Prob}\left\{v_D \bar{U}_\zeta^D \geqslant v_S \bar{U}_\zeta^S\right\} = \text{Prob}\left\{v_S \leqslant \frac{\bar{U}_\zeta^D}{\bar{U}_\zeta^S} v_D\right\} = \int_0^\infty \exp\left\{-\left(\frac{U_\zeta^D}{U_\zeta^S} v_D\right)^{-\theta}\right\} dF(v_D) = \frac{1}{\left(\bar{U}_\zeta^S / \bar{U}_\zeta^D\right)^\theta + 1}$$

Our goal is to show that skilled workers sort more into downtown areas than unskilled workers do, i.e., $\pi_D^s > \pi_D^u$. To show this, we first note that

$$\frac{\pi_D^s}{\pi_D^u} = \frac{\left(\bar{U}_u^S / \bar{U}_u^D\right)^\theta + 1}{\left(\bar{U}_s^S / \bar{U}_s^D\right)^\theta + 1} > 1 \quad \text{if and only if} \quad \frac{\bar{U}_u^S / \bar{U}_u^D}{\bar{U}_s^S / \bar{U}_s^D} > 1$$

Substituting the expressions for $\bar{U}_\zeta^n$, the second expression can be written as

$$\frac{\bar{U}_u^S / \bar{U}_u^D}{\bar{U}_s^S / \bar{U}_s^D} = \frac{\bar{U}_s^D / \bar{U}_u^D}{\bar{U}_s^S / \bar{U}_u^S} = \frac{(w_s - p_H^D \bar{h})/(w_u - p_H^D \bar{h})}{(w_s - p_H^S \bar{h})/(w_u - p_H^S \bar{h})} = \frac{(w_s - p_H^D \bar{h})/(w_u - p_H^D \bar{h})}{(w_s - p_H^D \bar{h} + \Delta)/(w_u - p_H^D \bar{h} + \Delta)} > 1,$$

where the inequality is true because $\Delta = p_H^D \bar{h} - p_H^S \bar{h} > 0$ and $w_s - p_H^D \bar{h} > w_u - p_H^D \bar{h}$. Therefore we conclude that $\pi_D^s > \pi_D^u$, that is, skilled workers are more likely to live in downtown area than unskilled workers.

A.3 Proofs and Derivations

A.3.1 Proof of Proposition 1

We should first notice that L_u is sufficient to compute w_u and p_H^S. Hence, we can further simplify equations and move the LHS of relevant conditions to the RHS. The transformed expressions are

$$
\begin{aligned}
F_1 &\equiv L_u \left[(1-\alpha) \frac{w_u - p_H^S \bar{h}}{p_H^S} + \bar{h} \right] - \left[\frac{p_H^S}{w_u} \right]^{\frac{1-h}{h}} = 0 \\
F_2 &\equiv \left(w_u - p_H^S \bar{h} \right) \frac{1}{P^\alpha} \frac{1}{(p_H^S)^{1-\alpha}} - \bar{U}_u = 0
\end{aligned}
$$

By implicit function theorem, we can totally differentiate these expressions by L_u and obtain

$$
\begin{aligned}
\frac{\partial F_1}{\partial w_u} \frac{\partial w_u}{\partial L_u} + \frac{\partial F_1}{\partial p_H^S} \frac{\partial p_H^S}{\partial L_u} + \frac{\partial F_1}{\partial L_u} &= 0, \\
\frac{\partial F_2}{\partial w_u} \frac{\partial w_u}{\partial L_u} + \frac{\partial F_2}{\partial p_H^S} \frac{\partial p_H^S}{\partial L_u} + \frac{\partial F_2}{\partial L_u} &= 0
\end{aligned}
$$

We can rearrange terms and write the above system of equations in matrix form as

$$
\begin{bmatrix} \frac{\partial F_1}{\partial w_u} & \frac{\partial F_1}{\partial p_H^S} \\ \frac{\partial F_2}{\partial w_u} & \frac{\partial F_2}{\partial p_H^S} \end{bmatrix} \begin{bmatrix} \frac{\partial w_u}{\partial L_u} \\ \frac{\partial p_H^S}{\partial L_u} \end{bmatrix} = \begin{bmatrix} -\frac{\partial F_1}{\partial L_u} \\ -\frac{\partial F_2}{\partial L_u} \end{bmatrix}.
$$

Solving the unknown partial derivatives requires to solve for the following,

$$
\begin{bmatrix} \frac{\partial w_u}{\partial L_u} \\ \frac{\partial p_H^S}{\partial L_u} \end{bmatrix} = \begin{bmatrix} \frac{\partial F_1}{\partial w_u} & \frac{\partial F_1}{\partial p_H^S} \\ \frac{\partial F_2}{\partial w_u} & \frac{\partial F_2}{\partial p_H^S} \end{bmatrix}^{-1} \begin{bmatrix} -\frac{\partial F_1}{\partial L_u} \\ -\frac{\partial F_2}{\partial L_u} \end{bmatrix} = \frac{1}{\frac{\partial F_1}{\partial w_u} \frac{\partial F_2}{\partial p_H^S} - \frac{\partial F_1}{\partial p_H^S} \frac{\partial F_2}{\partial w_u}} \begin{bmatrix} \frac{\partial F_2}{\partial p_H^S} & -\frac{\partial F_2}{\partial p_H^S} \\ -\frac{\partial F_2}{\partial w_u} & \frac{\partial F_1}{\partial w_u} \end{bmatrix} \begin{bmatrix} -\frac{\partial F_1}{\partial L_u} \\ -\frac{\partial F_2}{\partial L_u} \end{bmatrix}
$$

The partial derivatives can be computed as

$$
\begin{aligned}
\frac{\partial F_1}{\partial L_u} &= (1-\alpha) \frac{w_u - p_H^S \bar{h}}{p_H^S} + \bar{h} > 0 \\
\frac{\partial F_2}{\partial L_u} &= 0
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial F_1}{\partial p_H^S} &= -\frac{L_u(1-\alpha)w_u}{(p_H^S)^2} - \frac{1-h}{h}\left(\frac{p_H^S}{w_u}\right)^{\frac{1-2h}{h}} \frac{1}{w_u} < 0 \\
\frac{\partial F_2}{\partial p_H^S} &= -\bar{h}\frac{1}{P^\alpha}\frac{1}{(p_H^S)^{1-\alpha}} - (w_u - p_H^S\bar{h})\frac{1}{P^\alpha}(1-\alpha)\frac{1}{(p_H^S)^{-\alpha}}\frac{1}{(p_H^S)^2} < 0 \\
\frac{\partial F_1}{\partial w_u} &= \frac{L_u(1-\alpha)}{p_H^S} + \frac{1-h}{h}\left[\frac{p_H^S}{w_u}\right]^{\frac{1-2h}{h}} \left[\frac{p_H^S}{(w_u)^2}\right] > 0 \\
\frac{\partial F_2}{\partial w_u} &= \frac{1}{P^\alpha}\frac{1}{(p_H^S)^{1-\alpha}} > 0
\end{aligned}
$$

And we further have

$$
\begin{bmatrix} \frac{\partial w_u}{\partial L_u} \\ \frac{\partial p_H^S}{\partial L_u} \end{bmatrix} = \frac{1}{\frac{\partial F_1}{\partial w_u}\frac{\partial F_2}{\partial p_H^S} - \frac{\partial F_1}{\partial p_H^S}\frac{\partial F_2}{\partial w_u}} \begin{bmatrix} \frac{\partial F_2}{\partial p_H^S} & -\frac{\partial F_2}{\partial p_H^S} \\ -\frac{\partial F_2}{\partial w_u} & \frac{\partial F_1}{\partial w_u} \end{bmatrix} \begin{bmatrix} -\frac{\partial F_1}{\partial L_u} \\ -\frac{\partial F_2}{\partial L_u} \end{bmatrix} = \frac{1}{\frac{\partial F_1}{\partial w_u}\frac{\partial F_2}{\partial p_H^S} - \frac{\partial F_1}{\partial p_H^S}\frac{\partial F_2}{\partial w_u}} \begin{bmatrix} -\frac{\partial F_2}{\partial p_H^S}\frac{\partial F_1}{\partial L_u} + \frac{\partial F_1}{\partial p_H^S}\frac{\partial F_2}{\partial L_u} \\ \frac{\partial F_2}{\partial w_u}\frac{\partial F_1}{\partial L_u} - \frac{\partial F_1}{\partial w_u}\frac{\partial F_2}{\partial L_u} \end{bmatrix}
$$

Hence, it suffices to show that the fraction in front of the matrix is positive. Evaluating the expressions explicitly gives the following,

$$
\begin{aligned}
&\frac{\partial F_1}{\partial w_u}\frac{\partial F_2}{\partial p_H^S} - \frac{\partial F_1}{\partial p_H^S}\frac{\partial F_2}{\partial w_u} \\
&= \left[\frac{L_u(1-\alpha)}{p_H^S} + \frac{1-h}{h}\left(\frac{p_H^S}{w_u}\right)^{\frac{1-2h}{h}}\left(\frac{p_H^S}{(w_u)^2}\right)\right]\left[-\bar{h}\frac{1}{P^\alpha}\frac{1}{(p_H^S)^{1-\alpha}} - (w_u - p_H^S\bar{h})\frac{1}{P^\alpha}(1-\alpha)\frac{1}{(p_H^S)^{-\alpha}}\frac{1}{(p_H^S)^2}\right] \\
&\quad - \left[-\frac{L_u(1-\alpha)w_u}{(p_H^S)^2} - \frac{1-h}{h}\left(\frac{p_H^S}{w_u}\right)^{\frac{1-2h}{h}}\frac{1}{w_u}\right]\left[\frac{1}{P^\alpha}\frac{1}{(p_H^S)^{1-\alpha}}\right] \\
&= \left[\frac{L_u(1-\alpha)}{p_H^S} + \frac{1-h}{h}\left(\frac{p_H^S}{w_u}\right)^{\frac{1-2h}{h}}\left(\frac{p_H^S}{(w_u)^2}\right)\right]\left[-\bar{h}\frac{1}{P^\alpha}\frac{1}{(p_H^S)^{1-\alpha}} - \frac{(1-\alpha)(w_u - p_H^S\bar{h})}{p_H^S}\frac{1}{P^\alpha}\frac{1}{(p_H^S)^{1-\alpha}}\right] \\
&\quad - \left[-\frac{L_u(1-\alpha)w_u}{(p_H^S)^2} - \frac{1-h}{h}\left(\frac{p_H^S}{w_u}\right)^{\frac{1-2h}{h}}\frac{1}{w_u}\right]\left[\frac{1}{P^\alpha}\frac{1}{(p_H^S)^{1-\alpha}}\right] \\
&= \left\{\frac{L_u(1-\alpha)w_u}{(p_H^S)^2} + \frac{1-h}{h}\left(\frac{p_H^S}{w_u}\right)^{\frac{1-2h}{h}}\frac{1}{w_u} - \left[\frac{L_u(1-\alpha)}{p_H^S} + \frac{1-h}{h}\left(\frac{p_H^S}{w_u}\right)^{\frac{1-2h}{h}}\left(\frac{p_H^S}{(w_u)^2}\right)\right]\left[\bar{h} + \frac{(1-\alpha)(w_u - p_H^S\bar{h})}{p_H^S}\right]\right\} \\
&\quad \times \frac{1}{P^\alpha}\frac{1}{(p_H^S)^{1-\alpha}} \\
&= \left\{\frac{L_u(1-\alpha)w_u}{(p_H^S)^2} + \frac{1-h}{h}\left(\frac{p_H^S}{w_u}\right)^{\frac{1-2h}{h}}\frac{1}{w_u} - \left[\frac{L_u(1-\alpha)}{p_H^S} + \frac{1-h}{h}\left(\frac{p_H^S}{w_u}\right)^{\frac{1-2h}{h}}\left(\frac{p_H^S}{(w_u)^2}\right)\right]\left[\bar{h} + \frac{(1-\alpha)w_u}{p_H^S} - (1-\alpha)\bar{h}\right]\right\} \\
&\quad \times \frac{1}{P^\alpha}\frac{1}{(p_H^S)^{1-\alpha}} \\
&= \left\{\frac{L_u(1-\alpha)w_u}{(p_H^S)^2} + \frac{1-h}{h}\left(\frac{p_H^S}{w_u}\right)^{\frac{1-2h}{h}}\frac{1}{w_u} - \left[\frac{L_u(1-\alpha)}{p_H^S} + \frac{1-h}{h}\left(\frac{p_H^S}{w_u}\right)^{\frac{1-2h}{h}}\left(\frac{p_H^S}{(w_u)^2}\right)\right]\left[\frac{(1-\alpha)w_u}{p_H^S} + \alpha\bar{h}\right]\right\} \\
&\quad \times \frac{1}{P^\alpha}\frac{1}{(p_H^S)^{1-\alpha}} \\
&= \left[\frac{L_u(1-\alpha)}{p_H^S} + \frac{1-h}{h}\left(\frac{p_H^S}{w_u}\right)^{\frac{1-2h}{h}}\left(\frac{p_H^S}{(w_u)^2}\right)\right]\frac{\alpha(w_u - p_H^S\bar{h})}{p_H^S}\frac{1}{P^\alpha}\frac{1}{(p_H^S)^{1-\alpha}} \\
&\quad + \left\{\frac{L_u(1-\alpha)w_u}{(p_H^S)^2} + \frac{1-h}{h}\left(\frac{p_H^S}{w_u}\right)^{\frac{1-2h}{h}}\frac{1}{w_u} - \left[\frac{L_u(1-\alpha)}{p_H^S} + \frac{1-h}{h}\left(\frac{p_H^S}{w_u}\right)^{\frac{1-2h}{h}}\left(\frac{p_H^S}{(w_u)^2}\right)\right]\frac{w_u}{p_H^S}\right\}\frac{1}{P^\alpha}\frac{1}{(p_H^S)^{1-\alpha}} \\
&= \underbrace{\left[\frac{L_u(1-\alpha)}{p_H^S} + \frac{1-h}{h}\left(\frac{p_H^S}{w_u}\right)^{\frac{1-2h}{h}}\left(\frac{p_H^S}{(w_u)^2}\right)\right]}_{>0}\underbrace{\frac{\alpha(w_u - p_H^S\bar{h})}{p_H^S}}_{>0}\underbrace{\frac{1}{P^\alpha}\frac{1}{(p_H^S)^{1-\alpha}}}_{>0} > 0
\end{aligned}
$$

To recap, so far we have exploited part of the equilibrium conditions and proved that $\partial w_u/\partial L_u > 0$ and $\partial p_H^S/\partial L_u > 0$. We now proceed to prove the rest of the proposition related to w_s and p_H^D. Similarly, we can write the equilibrium conditions in the following form.

$$F_3 \equiv L_s\left[(1-\alpha)\frac{w_s - p_H^D\bar{h}}{p_H^D} + \bar{h}\right] - \left[\frac{p_H^D}{w_u}\right]^{\frac{1-h}{h}} = 0$$
$$F_4 \equiv (w_s - p_H^D\bar{h})\frac{1}{P^\alpha}\frac{1}{(p_H^D)^{1-\alpha}} - \bar{U}_s = 0$$

Notice that w_u, which is already solved as a function of L_u but not related to L_s, is included in F_3. We should first derive the relevant comparative statics related to L_s, as L_s does not influence the value of w_u. Performing implicit function theorem on this set of equilibrium conditions yields the following

$$\frac{\partial F_3}{\partial w_s}\frac{\partial w_s}{\partial L_s} + \frac{\partial F_3}{\partial p_H^D}\frac{\partial p_H^D}{\partial L_s} + \frac{\partial F_3}{\partial L_s} = 0,$$
$$\frac{\partial F_4}{\partial w_s}\frac{\partial w_s}{\partial L_s} + \frac{\partial F_4}{\partial p_H^D}\frac{\partial p_H^D}{\partial L_s} + \frac{\partial F_4}{\partial L_s} = 0$$

We can rearrange terms and write the above system of equations in matrix form as

$$\begin{bmatrix} \frac{\partial F_3}{\partial w_s} & \frac{\partial F_3}{\partial p_H^D} \\ \frac{\partial F_4}{\partial w_s} & \frac{\partial F_4}{\partial p_H^D} \end{bmatrix} \begin{bmatrix} \frac{\partial w_s}{\partial L_s} \\ \frac{\partial p_H^D}{\partial L_s} \end{bmatrix} = \begin{bmatrix} -\frac{\partial F_3}{\partial L_s} \\ -\frac{\partial F_4}{\partial L_s} \end{bmatrix}.$$

Solving the unknown partial derivatives requires to solve for the following,

$$\begin{bmatrix} \frac{\partial w_s}{\partial L_s} \\ \frac{\partial p_H^D}{\partial L_u} \end{bmatrix} = \begin{bmatrix} \frac{\partial F_3}{\partial w_s} & \frac{\partial F_3}{\partial p_H^D} \\ \frac{\partial F_4}{\partial w_s} & \frac{\partial F_4}{\partial p_H^D} \end{bmatrix}^{-1} \begin{bmatrix} -\frac{\partial F_3}{\partial L_s} \\ -\frac{\partial F_4}{\partial L_s} \end{bmatrix} = \frac{1}{\frac{\partial F_3}{\partial w_s}\frac{\partial F_4}{\partial p_H^D} - \frac{\partial F_3}{\partial p_H^D}\frac{\partial F_4}{\partial w_s}} \begin{bmatrix} \frac{\partial F_4}{\partial p_H^D} & -\frac{\partial F_3}{\partial p_H^D} \\ -\frac{\partial F_4}{\partial w_s} & \frac{\partial F_3}{\partial w_s} \end{bmatrix} \begin{bmatrix} -\frac{\partial F_3}{\partial L_s} \\ -\frac{\partial F_4}{\partial L_s} \end{bmatrix}$$

which can be further evaluated to

$$\begin{bmatrix} \frac{\partial w_s}{\partial L_s} \\ \frac{\partial p_H^D}{\partial L_u} \end{bmatrix} = \frac{1}{\frac{\partial F_3}{\partial w_s}\frac{\partial F_4}{\partial p_H^D} - \frac{\partial F_3}{\partial p_H^D}\frac{\partial F_4}{\partial w_s}} \begin{bmatrix} \frac{\partial F_4}{\partial p_H^D} & -\frac{\partial F_3}{\partial p_H^D} \\ -\frac{\partial F_4}{\partial w_s} & \frac{\partial F_3}{\partial w_s} \end{bmatrix} \begin{bmatrix} -\frac{\partial F_3}{\partial L_s} \\ -\frac{\partial F_4}{\partial L_s} \end{bmatrix} = \frac{1}{\frac{\partial F_3}{\partial w_s}\frac{\partial F_4}{\partial p_H^D} - \frac{\partial F_3}{\partial p_H^D}\frac{\partial F_4}{\partial w_s}} \begin{bmatrix} -\frac{\partial F_3}{\partial L_s}\frac{\partial F_4}{\partial p_H^D} + \frac{\partial F_3}{\partial p_H^D}\frac{\partial F_4}{\partial L_s} \\ \frac{\partial F_4}{\partial w_s}\frac{\partial F_3}{\partial L_s} - \frac{\partial F_3}{\partial w_s}\frac{\partial F_4}{\partial L_s} \end{bmatrix}$$

and the relevant partial derivatives are

$$
\begin{aligned}
\frac{\partial F_3}{\partial p_H^D} &= -L_s(1-\alpha)\frac{w_s}{(p_H^D)^2} - \frac{1-h}{h}\left(\frac{p_H^D}{w_u}\right)^{\frac{1=2h}{h}}\frac{1}{w_u} < 0 \\
\frac{\partial F_3}{\partial w_s} &= \frac{L_s(1-\alpha)}{p_H^D} > 0 \\
\frac{\partial F_3}{\partial L_s} &= (1-\alpha)\frac{w_s - p_H^D\bar{h}}{p_H^D} + \bar{h} \\
\frac{\partial F_4}{\partial p_H^D} &= -\bar{h}\frac{1}{P^\alpha}\frac{1}{(p_H^D)^{1-\alpha}} - (w_s - p_H^D\bar{h})\frac{1}{P^\alpha}(1-\alpha)\frac{1}{(p_H^D)^{-\alpha}}\frac{1}{(p_H^D)^2} < 0 \\
\frac{\partial F_4}{\partial w_s} &= \frac{1}{P^\alpha}\frac{1}{(p_H^D)^{1-\alpha}} > 0 \\
\frac{\partial F_4}{\partial L_s} &= 0.
\end{aligned}
$$

Two observations are in order. First, it suffices for us to prove that the fraction is positive. Second, everything is symmetric to our previous proof except that for the partial derivative $\partial F_3/\partial w_s$, there is one less term which is positive. Hence, given that $\partial F_4/\partial p_H^D$ is negative, we know that the targeted fraction is positive following a symmetry argument. We now continue the proof regarding to $\partial w_s/\partial L_u$ and $\partial p_H^D/\partial L_u$. Similarly performing implicit function theorem again we have that

$$
\begin{aligned}
&\frac{\partial F_3}{\partial w_s}\frac{\partial w_s}{\partial L_u} + \frac{\partial F_3}{\partial p_H^D}\frac{\partial p_H^D}{\partial L_u} + \frac{\partial F_3}{\partial w_u}\frac{\partial w_u}{\partial L_u} + \frac{\partial F_3}{\partial L_s} = 0 \\
&\frac{\partial F_4}{\partial w_s}\frac{\partial w_s}{\partial L_u} + \frac{\partial F_4}{\partial p_H^D}\frac{\partial p_H^D}{\partial L_u} + \frac{\partial F_4}{\partial L_u} = 0
\end{aligned}
$$

Writing it in matrix form, we have that

$$
\begin{bmatrix} \frac{\partial F_3}{\partial w_s} & \frac{\partial F_3}{\partial p_H^D} \\ \frac{\partial F_4}{\partial w_s} & \frac{\partial F_4}{\partial p_H^D} \end{bmatrix}
\begin{bmatrix} \frac{\partial w_s}{\partial L_u} \\ \frac{\partial p_H^D}{\partial L_u} \end{bmatrix}
=
\begin{bmatrix} -\frac{\partial F_3}{\partial L_u} - \frac{\partial F_3}{\partial w_u}\frac{\partial w_u}{\partial L_u} \\ -\frac{\partial F_4}{\partial L_u} \end{bmatrix}
$$

Notice that this is really similar to our previous proof. Given that we have already shown $\partial w_u/\partial L_u > 0$, we need only to show that $\partial F_3/\partial w_u > 0$ which is true.

A.3.2 Proof of Proposition 2

We can rewrite the reduced first-order condition as

$$F \equiv \frac{1}{\Phi_j(q;z)}\frac{\partial \Phi_j(q;z)}{\partial q} - \frac{(1-\gamma_j)(\sigma_j-1)}{w(q,\varphi,L_s,L_u)}\frac{\partial w(q,\varphi,L_s,L_u)}{\partial q} = 0.$$

Invoking the implicit function theorem, we can totally differentiate the LHS of the expression and show the following

$$\frac{\partial q^*}{\partial z} = -\frac{\partial F/\partial z}{\partial F/\partial q} > 0.$$

The inequality is true because of the following. First, from the SOC of the profit maximization problem with respect to q, we know that

$$\frac{\partial F}{\partial q} < 0.$$

Hence, it suffices to show that $\partial F/\partial z > 0$. Partially differentiating F with respect to z yields the following,

$$\text{Sign}\left[\frac{\partial F}{\partial z}\right] = \text{Sign}\left[w^{-2}\frac{\partial w(q,\varphi)}{\partial z}\frac{\partial w(q,\varphi)}{\partial q} - w^{-1}\frac{\partial[\partial w(q,\varphi)/\partial q]}{\partial z}\right],$$

where individual components of this expression evaluate to the following.

$$\begin{aligned}
\frac{\partial w(q,\varphi)}{\partial q} &= \frac{1}{1-\sigma_L}\left[\chi_u(q,\varphi)w_u^{1-\sigma_L} + \lambda\chi_s(q,\varphi)w_s^{1-\sigma_L}\right]^{\frac{1}{1-\sigma_L}-1}\left[\frac{\partial\chi_u(q,\varphi)}{\partial q}w_u^{1-\sigma_L} + \lambda\frac{\partial\chi_s(q,\varphi)}{\partial q}w_s^{1-\sigma_L}\right] \\
&= \frac{1}{1-\sigma_L}\left[\chi_u(q,\varphi)w_u^{1-\sigma_L} + \lambda\chi_s(q,\varphi)w_s^{1-\sigma_L}\right]^{\frac{\sigma_L}{1-\sigma_L}}\left[\frac{\partial\chi_u(q,\varphi)}{\partial q}w_u^{1-\sigma_L} + \lambda\frac{\partial\chi_s(q,\varphi)}{\partial q}w_s^{1-\sigma_L}\right] \\
&= \frac{1}{1-\sigma_L}w(q,\varphi)^{\sigma_L}\left[\frac{\partial\chi_u(q,\varphi)}{\partial q}w_u^{1-\sigma_L} + \lambda\frac{\partial\chi_s(q,\varphi)}{\partial q}w_s^{1-\sigma_L}\right] \\
\frac{\partial w(q,\varphi)}{\partial z} &= \frac{1}{1-\sigma_L}w(q,\varphi)^{\sigma_L}\left[\frac{\partial\chi_u(q,\varphi)}{\partial z}w_u^{1-\sigma_L} + \lambda\frac{\partial\chi_s(q,\varphi)}{\partial z}w_s^{1-\sigma_L}\right]
\end{aligned}$$

Further notice that given the assume functional form of $\chi_\zeta(q,\varphi)$, we know that

$$\frac{\partial\chi_\zeta(q,\varphi)}{\partial z} = \lambda_{1\zeta}\varphi^{\lambda_{1\zeta}-1}\exp(\lambda_{2\zeta}q) > 0,$$

and that

$$\frac{\partial \chi_\zeta(q,\varphi)}{\partial q} = \lambda_{2\zeta}\varphi^{\lambda_{1\zeta}}\exp(\lambda_{2\zeta}q) = \lambda_{2\zeta}\chi_\zeta(q,\varphi) < 0,$$

with $\lambda_{2\zeta} < 0$ and $\lambda_{2s} > \lambda_{2u}$. For simplicity sake, we denote $\lambda_{2\zeta}$ as λ_ζ hereafter. Hence the previous partial derivatives further evaluate to

$$\frac{\partial w(q,\varphi)}{\partial q} = \frac{1}{1-\sigma_L} w(q,\varphi)^{\sigma_L}\left[\lambda_u\chi_u(q,\varphi)w_u^{1-\sigma_L} + \lambda\lambda_s\chi_s(q,\varphi)w_s^{1-\sigma_L}\right].$$

In addition, we have that the last partial derivative evaluates to the following,

$$\begin{aligned}\frac{\partial[\partial w(q,\varphi)/\partial q]}{\partial z} =& \frac{\sigma_L}{1-\sigma_L}w(q,\varphi)^{\sigma_L-1}\left[\lambda_u\chi_u(q,\varphi)w_u^{1-\sigma_L}+\lambda\lambda_s\chi_s(q,\varphi)w_s^{1-\sigma_L}\right]\frac{\partial w(q,\varphi)}{\partial z}\\ &+\frac{1}{1-\sigma_L}w(q,\varphi)^{\sigma_L}\left[\lambda_u\frac{\partial\chi_u(q,\varphi)}{\partial z}w_u^{1-\sigma_L}+\lambda\lambda_s\frac{\partial\chi_s(q,\varphi)}{\partial z}w_s^{1-\sigma_L}\right]\\ =&\frac{\sigma_L}{(1-\sigma_L)^2}w(q,\varphi)^{2\sigma_L-1}\left[\lambda_u\chi_u(q,\varphi)w_u^{1-\sigma_L}+\lambda\lambda_s\chi_s(q,\varphi)w_s^{1-\sigma_L}\right]\left[\frac{\partial\chi_u(q,\varphi)}{\partial z}w_u^{1-\sigma_L}+\lambda\frac{\partial\chi_s(q,\varphi)}{\partial z}w_s^{1-\sigma_L}\right]\\ &+\frac{1}{1-\sigma_L}w(q,\varphi)^{\sigma_L}\left[\lambda_u\frac{\partial\chi_u(q,\varphi)}{\partial z}w_u^{1-\sigma_L}+\lambda\lambda_s\frac{\partial\chi_s(q,\varphi)}{\partial z}w_s^{1-\sigma_L}\right].\end{aligned}$$

Hence, the second component in our targeted expression evaluates to

$$\begin{aligned}w^{-1}\frac{\partial[\frac{\partial w(q,\varphi)}{\partial q}]}{\partial z} =& \frac{\sigma_L}{(1-\sigma_L)^2}w(q,\varphi)^{2\sigma_L-2}\left[\lambda_u\chi_u(q,\varphi)w_u^{1-\sigma_L}+\lambda\lambda_s\chi_s(q,\varphi)w_s^{1-\sigma_L}\right]\left[\frac{\partial\chi_u(q,\varphi)}{\partial z}w_u^{1-\sigma_L}+\lambda\frac{\partial\chi_s(q,\varphi)}{\partial z}w_s^{1-\sigma_L}\right]\\ &+\frac{1}{1-\sigma_L}w(q,\varphi)^{\sigma_L-1}\left[\lambda_u\frac{\partial\chi_u(q,\varphi)}{\partial z}w_u^{1-\sigma_L}+\lambda\lambda_s\frac{\partial\chi_s(q,\varphi)}{\partial z}w_s^{1-\sigma_L}\right],\end{aligned}$$

and the first component evaluates to

$$\begin{aligned}w^{-2}\frac{\partial w(q,\varphi)}{\partial z}\frac{\partial w(q,\varphi)}{\partial q} =& \frac{1}{(1-\sigma_L)^2}w(q,\varphi)^{2\sigma_L-2}\left[\lambda_u\chi_u(q,\varphi)w_u^{1-\sigma_L}+\lambda\lambda_s\chi_s(q,\varphi)w_s^{1-\sigma_L}\right]\\ &\times\left[\frac{\partial\chi_u(q,\varphi)}{\partial z}w_u^{1-\sigma_L}+\lambda\frac{\partial\chi_s(q,\varphi)}{\partial z}w_s^{1-\sigma_L}\right]\end{aligned}$$

Therefore, the targeted expression evaluates to the following.

$$
\begin{aligned}
& w^{-2}\frac{\partial w(q,\varphi)}{\partial z}\frac{\partial w(q,\varphi)}{\partial q} - w^{-1}\frac{\partial[\frac{\partial w(q,\varphi)}{\partial q}]}{\partial z} \\
=& \frac{1}{1-\sigma_L} w(q,\varphi)^{2\sigma_L-2}\left[\lambda_u \chi_u(q,\varphi) w_u^{1-\sigma_L} + \lambda\lambda_s \chi_s(q,\varphi) w_s^{1-\sigma_L}\right]\left[\frac{\partial \chi_u(q,\varphi)}{\partial z} w_u^{1-\sigma_L} + \lambda \frac{\partial \chi_s(q,\varphi)}{\partial z} w_s^{1-\sigma_L}\right] \\
& - \frac{1}{1-\sigma_L} w(q,\varphi)^{\sigma_L-1}\left[\lambda_u \frac{\partial \chi_u(q,\varphi)}{\partial z} w_u^{1-\sigma_L} + \lambda\lambda_s \frac{\partial \chi_s(q,\varphi)}{\partial z} w_s^{1-\sigma_L}\right] \\
=& \frac{1}{1-\sigma_L} w(q,\varphi)^{\sigma_L-1}\Bigg\{ w(q,\varphi)^{\sigma_L-1}\left[\lambda_u \chi_u(q,\varphi) w_u^{1-\sigma_L} + \lambda\lambda_s \chi_s(q,\varphi) w_s^{1-\sigma_L}\right]\left[\frac{\partial \chi_u(q,\varphi)}{\partial z} w_u^{1-\sigma_L} + \lambda \frac{\partial \chi_s(q,\varphi)}{\partial z} w_s^{1-\sigma_L}\right] \\
& - \left[\lambda_u \frac{\partial \chi_u(q,\varphi)}{\partial z} w_u^{1-\sigma_L} + \lambda\lambda_s \frac{\partial \chi_s(q,\varphi)}{\partial z} w_s^{1-\sigma_L}\right]\Bigg\}
\end{aligned}
$$

This implies that in order to show Sign $[\partial F/\partial z]$ is positive, it suffices to show that the expression in the curly bracket is negative given $1/(1-\sigma_L) < 0$. It can be shown as follows.

$$
\begin{aligned}
& w(q,\varphi)^{\sigma_L-1}\left[\lambda_u\chi_u(q,\varphi)w_u^{1-\sigma_L}+\lambda\lambda_s\chi_s(q,\varphi)w_s^{1-\sigma_L}\right]\left[\frac{\partial\chi_u(q,\varphi)}{\partial z}w_u^{1-\sigma_L}+\lambda\frac{\partial\chi_s(q,\varphi)}{\partial z}w_s^{1-\sigma_L}\right]\\
&-\left[\lambda_u\frac{\partial\chi_u(q,\varphi)}{\partial z}w_u^{1-\sigma_L}+\lambda\lambda_s\frac{\partial\chi_s(q,\varphi)}{\partial z}w_s^{1-\sigma_L}\right]\\
=&\left[\chi_u(q,\varphi)w_u^{1-\sigma_L}+\lambda\chi_s(q,\varphi)w_s^{1-\sigma_L}\right]^{-1}\left[\lambda_u\chi_u(q,\varphi)w_u^{1-\sigma_L}+\lambda\lambda_s\chi_s(q,\varphi)w_s^{1-\sigma_L}\right]\left[\frac{\partial\chi_u(q,\varphi)}{\partial z}w_u^{1-\sigma_L}+\lambda\frac{\partial\chi_s(q,\varphi)}{\partial z}w_s^{1-\sigma_L}\right]\\
&-\left[\lambda_u\frac{\partial\chi_u(q,\varphi)}{\partial z}w_u^{1-\sigma_L}+\lambda\lambda_s\frac{\partial\chi_s(q,\varphi)}{\partial z}w_s^{1-\sigma_L}\right]\\
=&\left[\chi_u(q,\varphi)w_u^{1-\sigma_L}+\lambda\chi_s(q,\varphi)w_s^{1-\sigma_L}\right]^{-1}\left[\lambda_s\chi_u(q,\varphi)w_u^{1-\sigma_L}+\lambda\lambda_s\chi_s(q,\varphi)w_s^{1-\sigma_L}\right]\left[\frac{\partial\chi_u(q,\varphi)}{\partial z}w_u^{1-\sigma_L}+\lambda\frac{\partial\chi_s(q,\varphi)}{\partial z}w_s^{1-\sigma_L}\right]\\
&-(\lambda_s-\lambda_u)\chi_u(q,\varphi)w_u^{1-\sigma_L}\left[\chi_u(q,\varphi)w_u^{1-\sigma_L}+\lambda\chi_s(q,\varphi)w_s^{1-\sigma_L}\right]^{-1}\left[\frac{\partial\chi_u(q,\varphi)}{\partial z}w_u^{1-\sigma_L}+\lambda\frac{\partial\chi_s(q,\varphi)}{\partial z}w_s^{1-\sigma_L}\right]\\
&-\left[\lambda_u\frac{\partial\chi_u(q,\varphi)}{\partial z}w_u^{1-\sigma_L}+\lambda\lambda_u\frac{\partial\chi_s(q,\varphi)}{\partial z}w_s^{1-\sigma_L}\right]\\
&-(\lambda_s-\lambda_u)\lambda\frac{\partial\chi_s(q,\varphi)}{\partial z}w_s^{1-\sigma_L}\\
=&\lambda_s\left[\chi_u(q,\varphi)w_u^{1-\sigma_L}+\lambda\chi_s(q,\varphi)w_s^{1-\sigma_L}\right]^{-1}\left[\chi_u(q,\varphi)w_u^{1-\sigma_L}+\lambda\chi_s(q,\varphi)w_s^{1-\sigma_L}\right]\left[\frac{\partial\chi_u(q,\varphi)}{\partial z}w_u^{1-\sigma_L}+\lambda\frac{\partial\chi_s(q,\varphi)}{\partial z}w_s^{1-\sigma_L}\right]\\
&-(\lambda_s-\lambda_u)\chi_u(q,\varphi)w_u^{1-\sigma_L}\left[\chi_u(q,\varphi)w_u^{1-\sigma_L}+\lambda\chi_s(q,\varphi)w_s^{1-\sigma_L}\right]^{-1}\left[\frac{\partial\chi_u(q,\varphi)}{\partial z}w_u^{1-\sigma_L}+\lambda\frac{\partial\chi_s(q,\varphi)}{\partial z}w_s^{1-\sigma_L}\right]\\
&-\lambda_u\left[\frac{\partial\chi_u(q,\varphi)}{\partial z}w_u^{1-\sigma_L}+\lambda\frac{\partial\chi_s(q,\varphi)}{\partial z}w_s^{1-\sigma_L}\right]\\
&-(\lambda_s-\lambda_u)\lambda\frac{\partial\chi_s(q,\varphi)}{\partial z}w_s^{1-\sigma_L}\\
=&(\lambda_s-\lambda_u)\left[\frac{\partial\chi_u(q,\varphi)}{\partial z}w_u^{1-\sigma_L}+\lambda\frac{\partial\chi_s(q,\varphi)}{\partial z}w_s^{1-\sigma_L}\right]-(\lambda_s-\lambda_u)\lambda\frac{\partial\chi_s(q,\varphi)}{\partial z}w_s^{1-\sigma_L}\\
&-(\lambda_s-\lambda_u)\frac{\chi_u(q,\varphi)w_u^{1-\sigma_L}}{\left[\chi_u(q,\varphi)w_u^{1-\sigma_L}+\lambda\chi_s(q,\varphi)w_s^{1-\sigma_L}\right]}\left[\frac{\partial\chi_u(q,\varphi)}{\partial z}w_u^{1-\sigma_L}+\lambda\frac{\partial\chi_s(q,\varphi)}{\partial z}w_s^{1-\sigma_L}\right]\\
=&(\lambda_s-\lambda_u)\frac{\partial\chi_u(q,\varphi)}{\partial z}w_u^{1-\sigma_L}-(\lambda_s-\lambda_u)\frac{\chi_u(q,\varphi)w_u^{1-\sigma_L}}{\left[\chi_u(q,\varphi)w_u^{1-\sigma_L}+\lambda\chi_s(q,\varphi)w_s^{1-\sigma_L}\right]}\left[\frac{\partial\chi_u(q,\varphi)}{\partial z}w_u^{1-\sigma_L}+\lambda\frac{\partial\chi_s(q,\varphi)}{\partial z}w_s^{1-\sigma_L}\right]\\
=&(\lambda_s-\lambda_u)\frac{\partial\chi_u(q,\varphi)}{\partial z}w_u^{1-\sigma_L}-(\lambda_s-\lambda_u)\frac{\chi_u(q,\varphi)w_u^{1-\sigma_L}}{\left[\chi_u(q,\varphi)w_u^{1-\sigma_L}+\lambda\chi_s(q,\varphi)w_s^{1-\sigma_L}\right]}\frac{\partial\chi_u(q,\varphi)}{\partial z}w_u^{1-\sigma_L}\\
&-(\lambda_s-\lambda_u)\frac{\chi_u(q,\varphi)w_u^{1-\sigma_L}}{\left[\chi_u(q,\varphi)w_u^{1-\sigma_L}+\lambda\chi_s(q,\varphi)w_s^{1-\sigma_L}\right]}\lambda\frac{\partial\chi_s(q,\varphi)}{\partial z}w_s^{1-\sigma_L}\\
=&(\lambda_s-\lambda_u)\frac{\partial\chi_u(q,\varphi)}{\partial z}w_u^{1-\sigma_L}-(\lambda_s-\lambda_u)\frac{\partial\chi_u(q,\varphi)}{\partial z}w_u^{1-\sigma_L}\\
&\quad+(\lambda_s-\lambda_u)\frac{\lambda\chi_s(q,\varphi)w_s^{1-\sigma_L}}{\left[\chi_u(q,\varphi)w_u^{1-\sigma_L}+\lambda\chi_s(q,\varphi)w_s^{1-\sigma_L}\right]}\frac{\partial\chi_u(q,\varphi)}{\partial z}w_u^{1-\sigma_L}\\
&\quad-(\lambda_s-\lambda_u)\frac{\chi_u(q,\varphi)w_u^{1-\sigma_L}}{\left[\chi_u(q,\varphi)w_u^{1-\sigma_L}+\lambda\chi_s(q,\varphi)w_s^{1-\sigma_L}\right]}\lambda\frac{\partial\chi_s(q,\varphi)}{\partial z}w_s^{1-\sigma_L}\\
=&(\lambda_s-\lambda_u)\frac{\lambda w_s^{1-\sigma_L}w_u^{1-\sigma_L}}{\left[\chi_u(q,\varphi)w_u^{1-\sigma_L}+\lambda\chi_s(q,\varphi)w_s^{1-\sigma_L}\right]}\left[\chi_s(q,\varphi)\frac{\partial\chi_u(q,\varphi)}{\partial z}-\chi_u(q,\varphi)\frac{\partial\chi_s(q,\varphi)}{\partial z}\right]\\
=&(\lambda_s-\lambda_u)\frac{\lambda w_s^{1-\sigma_L}w_u^{1-\sigma_L}}{\left[\chi_u(q,\varphi)w_u^{1-\sigma_L}+\lambda\chi_s(q,\varphi)w_s^{1-\sigma_L}\right]}\left[\chi_s(q,\varphi)\frac{\lambda_{1u}\varphi^{\lambda_{1u}}\exp(\lambda_{2u}q)}{\varphi}\frac{\partial\varphi}{\partial z}-\chi_u(q,\varphi)\frac{\lambda_{1s}\varphi^{\lambda_{1s}}\exp(\lambda_{2s}q)}{\varphi}\frac{\partial\varphi}{\partial z}\right]\\
=&(\lambda_s-\lambda_u)\frac{\lambda w_s^{1-\sigma_L}w_u^{1-\sigma_L}}{\left[\chi_u(q,\varphi)w_u^{1-\sigma_L}+\lambda\chi_s(q,\varphi)w_s^{1-\sigma_L}\right]}\left[\chi_s(q,\varphi)\frac{\lambda_{1u}\chi_u(q,\varphi)}{\varphi}\frac{\partial\varphi}{\partial z}-\chi_u(q,\varphi)\frac{\lambda_{1s}\chi_s(q,\varphi)}{\varphi}\frac{\partial\varphi}{\partial z}\right]\\
=&\underbrace{(\lambda_s-\lambda_u)\frac{\lambda w_s^{1-\sigma_L}w_u^{1-\sigma_L}}{\left[\chi_u(q,\varphi)w_u^{1-\sigma_L}+\lambda\chi_s(q,\varphi)w_s^{1-\sigma_L}\right]}\frac{\chi_s(q,\varphi)\chi_u(q,\varphi)}{\varphi}\frac{\partial\varphi}{\partial z}}_{>0}\underbrace{(\lambda_{1u}-\lambda_{1s})}_{<0}\\
<&0.
\end{aligned}
$$

A.3.3 Proof of Proposition 3

Similar to the proof of proposition 2, we can write the reduced first-order condition as

$$F \equiv \frac{1}{\Phi(q;z)} \frac{\partial \Phi_j(q;z)}{\partial q} - \frac{(1-\gamma_j)(\sigma_j - 1)}{w(q,\varphi,L_s,L_u)} \frac{\partial w(q,\varphi,L_s,L_u)}{\partial q} = 0.$$

Invoking the implicit function theorem, we can totally differentiate the LHS of the expression and show the following, that for any $L \in \{L_s, L_u\}$,

$$\frac{\partial q^*}{\partial L} = \frac{\partial F/\partial L}{\partial F/\partial q} > 0.$$

This is true because of the following reasoning. First, given the SOC of the profit maximization problem with respect to q, we know that $\partial F/\partial q < 0$. Hence it suffices to show that $\partial F/\partial L > 0$. This expression can be evaluated as

$$\begin{aligned} \frac{\partial F}{\partial L} &= -(1-\gamma_j)(\sigma_j - 1) \left[-w^{-2} \frac{\partial w}{\partial L} \frac{\partial w}{\partial q} + w^{-1} \frac{\partial^2 w}{\partial L \partial q} \right] \\ &= \frac{(1-\gamma_j)(\sigma_j - 1)}{w} \left[\frac{1}{w} \frac{\partial w}{\partial L} \frac{\partial w}{\partial q} - \frac{\partial^2 w}{\partial L \partial q} \right] \end{aligned}$$

which implies that

$$\text{Sign}\left(\frac{\partial F}{\partial L}\right) = \text{Sign}\left(\frac{1}{w} \frac{\partial w}{\partial L} \frac{\partial w}{\partial q} - \frac{\partial^2 w}{\partial L \partial q}\right).$$

The individual components of this expression can be evaluated as

$$\begin{aligned} \frac{\partial w}{\partial L} &= \frac{1}{1-\sigma_L} w^{\sigma_L} \left[\frac{\partial \chi_u}{\partial L} w_u^{1-\sigma_L} + \lambda \frac{\partial \chi_s}{\partial L} w_s^{1-\sigma_L} \right] + w^{\sigma_L} \left[\chi_u w_u^{-\sigma_L} \frac{\partial w_u}{\partial L} + \lambda \chi_s w_s^{-\sigma_L} \frac{\partial w_s}{\partial L} \right] \\ \frac{\partial w}{\partial q} &= \frac{1}{1-\sigma_L} w^{\sigma_L} \left[\lambda_u \chi_u w_u^{1-\sigma_L} + \lambda \lambda_s \chi_s w_s^{1-\sigma_L} \right] \\ \frac{\partial^2 w}{\partial L \partial q} &= \frac{\sigma_L}{1-\sigma_L} w^{\sigma_L - 1} \left[\lambda_u \chi_u w_u^{1-\sigma_L} + \lambda \lambda_s \chi_s w_s^{1-\sigma_L} \right] \frac{\partial w}{\partial L} + \frac{1}{1-\sigma_L} w^{\sigma_L} \left[\lambda_u \frac{\partial \chi_u}{\partial L} w_u^{1-\sigma_L} + \lambda \lambda_s \frac{\partial \chi_s}{\partial L} w_s^{1-\sigma_L} \right] \\ &\quad + w^{\sigma_L} \left[\lambda_u \chi_u w_u^{-\sigma_L} \frac{\partial w_u}{\partial L} + \lambda \lambda_s \chi_s w_s^{-\sigma_L} \frac{\partial w_s}{\partial L} \right] \end{aligned}$$

It follows that

$$\begin{aligned}
\frac{1}{w}\frac{\partial w}{\partial L}\frac{\partial w}{\partial q} =& \frac{1}{1-\sigma_L}w^{\sigma_L-1}\left[\lambda_u\chi_u w_u^{1-\sigma_L}+\lambda\lambda_s\chi_s w_s^{1-\sigma_L}\right]\frac{1}{1-\sigma_L}w^{\sigma_L}\left[\frac{\partial\chi_u}{\partial L}w_u^{1-\sigma_L}+\lambda\frac{\partial\chi_s}{\partial L}w_s^{1-\sigma_L}\right]\\
&+\frac{1}{1-\sigma_L}w^{\sigma_L-1}\left[\lambda_u\chi_u w_u^{1-\sigma_L}+\lambda\lambda_s\chi_s w_s^{1-\sigma_L}\right]w^{\sigma_L}\left[\chi_u w_u^{-\sigma_L}\frac{\partial w_u}{\partial L}+\lambda\chi_s w_s^{-\sigma_L}\frac{\partial w_s}{\partial L}\right]\\
=&\frac{1}{(1-\sigma_L)^2}w^{2\sigma_L-1}\left[\lambda_u\chi_u w_u^{1-\sigma_L}+\lambda\lambda_s\chi_s w_s^{1-\sigma_L}\right]\left[\frac{\partial\chi_u}{\partial L}w_u^{1-\sigma_L}+\lambda\frac{\partial\chi_s}{\partial L}w_s^{1-\sigma_L}\right]\\
&+\frac{1}{1-\sigma_L}w^{2\sigma_L-1}\left[\lambda_u\chi_u w^{1-\sigma_L}+\lambda\lambda_s\chi_s w_s^{1-\sigma_L}\right]\left[\chi_u w_u^{-\sigma_L}\frac{\partial w_u}{\partial L}+\lambda\chi_s w_s^{-\sigma_L}\frac{\partial w_s}{\partial L}\right]
\end{aligned}$$

and that

$$\begin{aligned}
\frac{\partial^2 w}{\partial L\partial q} =& \frac{\sigma_L}{1-\sigma_L}w^{\sigma_L-1}\left[\lambda_u\chi_u w_u^{1-\sigma_L}+\lambda\lambda_s\chi_s w_s^{1-\sigma_L}\right]\frac{\partial w}{\partial L}+\frac{1}{1-\sigma_L}w^{\sigma_L}\left[\lambda_u\frac{\partial\chi_u}{\partial L}w_u^{1-\sigma_L}+\lambda\lambda_s\frac{\partial\chi_s}{\partial L}w_s^{1-\sigma_L}\right]\\
&+w^{\sigma_L}\left[\lambda_u\chi_u w_u^{-\sigma_L}\frac{\partial w_u}{\partial L}+\lambda\lambda_s\chi_s w_s^{-\sigma_L}\frac{\partial w_s}{\partial L}\right]\\
=&\frac{\sigma_L}{1-\sigma_L}w^{\sigma_L-1}\left[\lambda_u\chi_u w_u^{1-\sigma_L}+\lambda\lambda_s\chi_s w_s^{1-\sigma_L}\right]\frac{1}{1-\sigma_L}w^{\sigma_L}\left[\frac{\partial\chi_u}{\partial L}w_u^{1-\sigma_L}+\lambda\frac{\partial\chi_s}{\partial L}w_s^{1-\sigma_L}\right]\\
&+\frac{\sigma_L}{1-\sigma_L}w^{\sigma_L-1}\left[\lambda_u\chi_u w_u^{1-\sigma_L}+\lambda\lambda_s\chi_s w_s^{1-\sigma_L}\right]w^{\sigma_L}\left[\chi_u w_u^{-\sigma_L}\frac{\partial w_u}{\partial L}+\lambda\chi_s w_s^{-\sigma_L}\frac{\partial w_s}{\partial L}\right]\\
&+\frac{1}{1-\sigma_L}w^{\sigma_L}\left[\lambda_u\frac{\partial\chi_u}{\partial L}w_u^{1-\sigma_L}+\lambda\lambda_s\frac{\partial\chi_s}{\partial L}w_s^{1-\sigma_L}\right]+w^{\sigma_L}\left[\lambda_u\chi_u w_u^{-\sigma_L}\frac{\partial w_u}{\partial L}+\lambda\lambda_s\chi_s w_s^{-\sigma_L}\frac{\partial w_s}{\partial L}\right]\\
=&\frac{\sigma_L}{(1-\sigma_L)^2}w^{2\sigma_L-1}\left[\lambda_u\chi_u w_u^{1-\sigma_L}+\lambda\lambda_s\chi_s w_s^{1-\sigma_L}\right]\left[\frac{\partial\chi_u}{\partial L}w_u^{1-\sigma_L}+\lambda\frac{\partial\chi_s}{\partial L}w_s^{1-\sigma_L}\right]\\
&+\frac{\sigma_L}{1-\sigma_L}w^{2\sigma_L-1}\left[\lambda_u\chi_u w_u^{1-\sigma_L}+\lambda\lambda_s\chi_s w_s^{1-\sigma_L}\right]\left[\chi_u w_u^{-\sigma_L}\frac{\partial w_u}{\partial L}+\lambda\chi_s w_s^{-\sigma_L}\frac{\partial w_s}{\partial L}\right]\\
&+\frac{1}{1-\sigma_L}w^{\sigma_L}\left[\lambda_u\frac{\partial\chi_u}{\partial L}w_u^{1-\sigma_L}+\lambda\lambda_s\frac{\partial\chi_s}{\partial L}w_s^{1-\sigma_L}\right]+w^{\sigma_L}\left[\lambda_u\chi_u w_u^{-\sigma_L}\frac{\partial w_u}{\partial L}+\lambda\lambda_s\chi_s w_s^{-\sigma_L}\frac{\partial w_s}{\partial L}\right]
\end{aligned}$$

Hence, we can show that

$$\begin{aligned}
\frac{1}{w}\frac{\partial w}{\partial q}\frac{\partial w}{\partial L}-\frac{\partial^2 w}{\partial L\partial q} =& \underbrace{\frac{1}{1-\sigma_L}w^{2\sigma_L-1}\left[\lambda_u\chi_u w_u^{1-\sigma_L}+\lambda\lambda_s\chi_s w_s^{1-\sigma_L}\right]\left[\frac{\partial\chi_u}{\partial L}w_u^{1-\sigma_L}+\lambda\frac{\partial\chi_s}{\partial L}w_s^{1-\sigma_L}\right]}_{A}\\
&+\underbrace{w^{2\sigma_L-1}\left[\lambda_u\chi_u w_u^{1-\sigma_L}+\lambda\lambda_s\chi_s w_s^{1-\sigma_L}\right]\left[\chi_u w_u^{-\sigma_L}\frac{\partial w_u}{\partial L}+\lambda\chi_s w_s^{-\sigma_L}\frac{\partial w_s}{\partial L}\right]}_{B}\\
&-\underbrace{\frac{1}{1-\sigma_L}w^{\sigma_L}\left[\lambda_u\frac{\partial\chi_u}{\partial L}w_u^{1-\sigma_L}+\lambda\lambda_s\frac{\partial\chi_s}{\partial L}w_w^{1-\sigma_L}\right]}_{C}-\underbrace{w^{\sigma_L}\left[\lambda_u\chi_u w_u^{-\sigma_L}\frac{\partial w_u}{\partial_L}+\lambda\lambda_s\chi_s w_s^{-\sigma_L}\frac{\partial w_s}{\partial L}\right]}_{D}\\
\equiv& A+B-C-D.
\end{aligned}$$

We shall evaluate the expression part-by-part. First, note that $w^{1-\sigma_L}=[\chi_u w_u^{1-\sigma_L}+$

$\lambda\chi_s w_s^{1-\sigma_L}]$. It follows that

$$
\begin{aligned}
A-C =& \frac{1}{1-\sigma_L} w^{2\sigma_L-1}[\lambda_u\chi_u w_u^{1-\sigma_L} + \lambda\lambda_s\chi_s w_s^{1-\sigma_L}] \left[\frac{\partial\chi_u}{\partial L} w_u^{1-\sigma_L} + \lambda\frac{\partial\chi_s}{\partial L} w_s^{1-\sigma_L}\right] \\
& - \frac{1}{1-\sigma_L} w^{\sigma_L} \left[\lambda_u \frac{\partial\chi_u}{\partial L} w_u^{1-\sigma_L} + \lambda\lambda_s \frac{\partial\chi_s}{\partial L} w_s^{1-\sigma_L}\right] \\
=& \frac{1}{1-\sigma_L} w^{2\sigma_L-1}\left[\lambda_u\chi_u w_u^{1-\sigma_L} + \lambda\lambda_u\chi_s w_s^{1-\sigma_L}\right] \left[\frac{\partial\chi_u}{\partial L} w_u^{1-\sigma_L} + \lambda\frac{\partial\chi_s}{\partial L} w_s^{1-\sigma_L}\right] \\
& + \frac{1}{1-\sigma_L} w^{2\sigma_L-1}\lambda(\lambda_s-\lambda_u)\chi_s w_s^{1-\sigma_L} \left[\frac{\partial\chi_u}{\partial L} w_u^{1-\sigma_L} + \lambda\frac{\partial\chi_s}{\partial L} w_s^{1-\sigma_L}\right] \\
& - \frac{1}{1-\sigma_L} w^{\sigma_L} \left[\lambda_s \frac{\partial\chi_u}{\partial L} w_u^{1-\sigma_L} + \lambda\lambda_s \frac{\partial\chi_s}{\partial L} w_s^{1-\sigma_L}\right] + \frac{1}{1-\sigma_L} w^{\sigma_L}(\lambda_s-\lambda_u)\frac{\partial\chi_u}{\partial L} w_u^{1-\sigma_L} \\
=& \frac{1}{1-\sigma_L} w^{\sigma_L}(\lambda_u-\lambda_s) \left[\frac{\partial\chi_u}{\partial L} w_u^{1-\sigma_L} + \lambda\frac{\partial\chi_s}{\partial L} w_s^{1-\sigma_L}\right] + \frac{1}{1-\sigma_L} w^{\sigma_L}(\lambda_s-\lambda_u)\frac{\partial\chi_u}{\partial L} w_u^{1-\sigma_L} \\
& + \frac{1}{1-\sigma_L} w^{2\sigma_L-1}\lambda(\lambda_s-\lambda_u)\chi_s w_s^{1-\sigma_L} \left[\frac{\partial\chi_u}{\partial L} w_u^{1-\sigma_L} + \lambda\frac{\partial\chi_s}{\partial L} w_s^{1-\sigma_L}\right] \\
=& \frac{1}{1-\sigma_L} w^{\sigma_L}\lambda(\lambda_u-\lambda_s)\frac{\partial\chi_s}{\partial L} w_s^{1-\sigma_L} + \frac{1}{1-\sigma_L} w^{2\sigma_L-1}\lambda(\lambda_s-\lambda_u)\chi_s w_s^{1-\sigma_L} \left[\frac{\partial\chi_u}{\partial L} w_u^{1-\sigma_L} + \lambda\frac{\partial\chi_s}{\partial L} w_s^{1-\sigma_L}\right] \\
=& \frac{1}{1-\sigma_L} w^{\sigma_L}\lambda(\lambda_u-\lambda_s)\frac{\partial\chi_s}{\partial L} w_s^{1-\sigma_L} + \frac{1}{1-\sigma_L} w^{2\sigma_L-1}\lambda(\lambda_s-\lambda_u)\chi_s w_s^{1-\sigma_L} \frac{\partial\chi_u}{\partial L} w_u^{1-\sigma_L} \\
& + \frac{1}{1-\sigma_L} w^{2\sigma_L-1}\lambda(\lambda_s-\lambda_u)\chi_s w_s^{1-\sigma_L} \lambda\frac{\partial\chi_s}{\partial L} w_s^{1-\sigma_L} \\
=& \frac{1}{1-\sigma_L} w^{\sigma_L}\lambda(\lambda_u-\lambda_s)\frac{\partial\chi_s}{\partial L} w_s^{1-\sigma_L} \left[1 - \frac{\lambda\chi_s w_s^{1-\sigma_L}}{w^{1-\sigma_L}}\right] + \frac{1}{1-\sigma_L} w^{2\sigma_L-1}\lambda(\lambda_s-\lambda_u)\chi_s w_s^{1-\sigma_L} \frac{\partial\chi_u}{\partial L} w_u^{1-\sigma_L} \\
=& \frac{1}{1-\sigma_L} w^{\sigma_L}\lambda(\lambda_u-\lambda_s)\frac{\partial\chi_s}{\partial L} w_s^{1-\sigma_L} \frac{\chi_u w_u^{1-\sigma_L}}{w^{1-\sigma_L}} + \frac{1}{1-\sigma_L} w^{2\sigma_L-1}\lambda(\lambda_s-\lambda_u)\chi_s w_s^{1-\sigma_L} \frac{\partial\chi_u}{\partial L} w_u^{1-\sigma_L} \\
=& \frac{1}{1-\sigma_L} w^{\sigma_L}\lambda(\lambda_u-\lambda_s)\frac{w_s^{1-\sigma_L} w_u^{1-\sigma_L}}{w^{1-\sigma_L}} \left[\chi_u \frac{\partial\chi_s}{\partial L} - \chi_s \frac{\partial\chi_u}{\partial L}\right].
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
B-D =& w^{2\sigma_L-1}\left[\lambda_u\chi_u w_u^{1-\sigma_L} + \lambda\lambda_s\chi_s w_s^{1-\sigma_L}\right] \left[\chi_u w_u^{-\sigma_L}\frac{\partial w_u}{\partial L} + \lambda\chi_s w_s^{-\sigma_L}\frac{\partial w_s}{\partial L}\right] \\
& - w^{\sigma_L}\left[\lambda_u\chi_u w_u^{-\sigma_L}\frac{\partial w_u}{\partial L} + \lambda\lambda_s\chi_s w_s^{-\sigma_L}\frac{\partial w_s}{\partial L}\right] \\
=& w^{2\sigma_L-1}\left[\lambda_u\chi_u w_u^{1-\sigma_L} + \lambda\lambda_u\chi_s w_s^{1-\sigma_L}\right] \left[\chi_u w_u^{-\sigma_L}\frac{\partial w_u}{\partial L} + \lambda\chi_s w_s^{-\sigma_L}\frac{\partial w_s}{\partial L}\right] \\
& + w^{2\sigma_L-1}\lambda(\lambda_s-\lambda_u)\chi_s w_s^{1-\sigma_L} \left[\chi_u w_u^{-\sigma_L}\frac{\partial w_u}{\partial L} + \lambda\chi_s w_s^{-\sigma_L}\frac{\partial w_s}{\partial L}\right] \\
& - w^{\sigma_L}\left[\lambda_s\chi_u w_u^{-\sigma_L}\frac{\partial w_u}{\partial L} + \lambda\lambda_s\chi_s w_s^{-\sigma_L}\frac{\partial w_s}{\partial L}\right] + w^{\sigma_L}(\lambda_s-\lambda_u)\chi_u w_u^{-\sigma_L}\frac{\partial w_u}{\partial L} \\
=& w^{\sigma_L}(\lambda_u-\lambda_s)\left[\chi_u w_u^{-\sigma_L}\frac{\partial w_u}{\partial L} + \lambda\chi_s w_s^{-\sigma_L}\frac{\partial w_s}{\partial L}\right] + w^{2\sigma_L-1}\lambda(\lambda_s-\lambda_u)\chi_s w_s^{1-\sigma_L}\left[\chi_u w^{-\sigma_L}\frac{\partial w_u}{\partial L} + \lambda\chi_s w_s^{-\sigma_L}\frac{\partial w_s}{\partial L}\right] \\
& + w^{\sigma_L}(\lambda_s-\lambda_u)\chi_u w_u^{-\sigma_L}\frac{\partial w_u}{\partial L} \\
=& w^{\sigma_L}(\lambda_u-\lambda_s)\frac{\chi_u w_u^{1-\sigma_L}}{w^{1-\sigma_L}}\left[\chi_u w_u^{-\sigma_L}\frac{\partial w_u}{\partial L} + \lambda\chi_s w_s^{-\sigma_L}\frac{\partial w_s}{\partial L}\right] + w^{\sigma_L}(\lambda_s-\lambda_u)\chi_u w_u^{-\sigma_L}\frac{\partial w_u}{\partial L} \\
=& w^{\sigma_L}(\lambda_s-\lambda_u)\frac{\lambda\chi_s w_s^{1-\sigma_L}}{w^{1-\sigma_L}}\chi_u w_u^{-\sigma_L}\frac{\partial w_u}{\partial L} + w^{\sigma_L}(\lambda_u-\lambda_s)\frac{\lambda\chi_u w_u^{1-\sigma_L}}{w^{1-\sigma_L}}\chi_s w_s^{-\sigma_L}\frac{\partial w_s}{\partial L} \\
=& w^{\sigma_L}(\lambda_s-\lambda_u)\frac{\lambda\chi_s\chi_u w_s^{1-\sigma_L} w_u^{1-\sigma_L}}{w^{1-\sigma_L}}\left[\frac{1}{w_u}\frac{\partial w_u}{\partial L} - \frac{1}{w_s}\frac{\partial w_s}{\partial L}\right]
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\frac{1}{w}\frac{\partial w}{\partial q}\frac{\partial w}{\partial L}-\frac{\partial^2 w}{\partial L\partial q} =&A+B-C-D\\
=&\frac{1}{1-\sigma_L}w^{\sigma_L}\lambda(\lambda_u-\lambda_s)\frac{w_s^{1-\sigma_L}w_u^{1-\sigma_L}}{w^{1-\sigma_L}}\left[\chi_u\frac{\partial\chi_s}{\partial L}-\chi_s\frac{\partial\chi_u}{\partial L}\right]\\
&+w^{\sigma_L}(\lambda_s-\lambda_u)\frac{\lambda\chi_s\chi_u w_s^{1-\sigma_L}w_u^{1-\sigma_L}}{w^{1-\sigma_L}}\left[\frac{1}{w_u}\frac{\partial w_u}{\partial L}-\frac{1}{w_s}\frac{\partial w_s}{\partial L}\right]\\
=&\frac{1}{1-\sigma_L}w^{\sigma_L}(\lambda_u-\lambda_s)\frac{\lambda\chi_s\chi_u w_s^{1-\sigma_L}w_u^{1-\sigma_L}}{w^{1-\sigma_L}}\left[\frac{1}{\chi_s}\frac{\partial\chi_s}{\partial L}-\frac{1}{\chi_u}\frac{\partial\chi_u}{\partial L}\right]\\
&+w^{\sigma_L}(\lambda_s-\lambda_u)\frac{\lambda\chi_s\chi_u w_s^{1-\sigma_L}w_u^{1-\sigma_L}}{w^{1-\sigma_L}}\left[\frac{1}{w_u}\frac{\partial w_u}{\partial L}-\frac{1}{w_s}\frac{\partial w_s}{\partial L}\right]\\
=&\frac{1}{1-\sigma_L}w^{\sigma_L}(\lambda_s-\lambda_u)\frac{\lambda\chi_s\chi_u w_s^{1-\sigma_L}w_u^{1-\sigma_L}}{w^{1-\sigma_L}}\left[\frac{1}{\chi_u}\frac{\partial\chi_u}{\partial L}-\frac{1}{\chi_s}\frac{\partial\chi_s}{\partial L}+\frac{1-\sigma_L}{w_u}\frac{\partial w_u}{\partial L}-\frac{1-\sigma_L}{w_s}\frac{\partial w_s}{\partial L}\right]\\
=&\frac{1}{1-\sigma_L}w^{\sigma_L}(\lambda_s-\lambda_u)\frac{\lambda\chi_s\chi_u w_s^{1-\sigma_L}w_u^{1-\sigma_L}}{w^{1-\sigma_L}}\left[\left(\frac{1}{\chi_u}\frac{\partial\chi_u}{\partial L}-\frac{\sigma_L-1}{w_u}\frac{\partial w_u}{\partial L}\right)-\left(\frac{1}{\chi_s}\frac{\partial\chi_s}{\partial L}-\frac{\sigma_L-1}{w_s}\frac{\partial w_s}{\partial L}\right)\right]\\
=&\underbrace{\frac{1}{1-\sigma_L}w^{\sigma_L}(\lambda_u-\lambda_s)\frac{\lambda\chi_s\chi_u w_s^{1-\sigma_L}w_u^{1-\sigma_L}}{w^{1-\sigma_L}}}_{>0}\left[\left(\frac{1}{\chi_s}\frac{\partial\chi_s}{\partial L}-\frac{\sigma_L-1}{w_s}\frac{\partial w_s}{\partial L}\right)-\left(\frac{1}{\chi_u}\frac{\partial\chi_u}{\partial L}-\frac{\sigma_L-1}{w_u}\frac{\partial w_u}{\partial L}\right)\right]
\end{aligned}
$$

Therefore,

$$
\text{Sign}\left[\frac{\partial q^*}{\partial L}\right]=\text{Sign}\left[\frac{\partial F}{\partial L}\right]=\text{Sign}\left[\left(\frac{1}{\chi_s}\frac{\partial\chi_s}{\partial L}-\frac{\sigma_L-1}{w_s}\frac{\partial w_s}{\partial L}\right)-\left(\frac{1}{\chi_u}\frac{\partial\chi_u}{\partial L}-\frac{\sigma_L-1}{w_u}\frac{\partial w_u}{\partial L}\right)\right]
$$

A.3.4 Proof of Proposition 4

The proof is essentially the same as in Gaubert (2018), with some slight modifications. It is obvious to see that the profit function is also strictly log supermodular in (L, z) due to our assumption on φ. Consider the case where $z_H > z_L$ and $L_u^H > L_u^L$. By the strict log-supermodularity of π, if the size of the skill population is fixed at $\bar{L}_s$, then $\frac{\pi(z_H,\bar{L}_s+L_u^H)}{\pi(z_H,\bar{L}_s+L_u^L)} > \frac{\pi(z_L,\bar{L}_s+L_u^H)}{\pi(z_L,\bar{L}_s+L_u^L)}$. Hence, if firm z_L has a higher profit in a city with larger skilled population $(\bar{L}_s, L_u^H)$ than in $(\bar{L}_s, L_u^H)$, then z_H must also have a higher profit in that city than the other city. Hence, $L_u^{H*} \geqslant L_u^{L*}$. The proof regarding the skilled population is similar.

A.3.5 Expression for Computing Wages

Given local wages, house prices of downtown area will be determined by the housing market clearing condition

$$L_s\left[(1-\alpha)\frac{w_s - p_H^D\bar{h}}{p_H^D} + \bar{h}\right] = \left[\frac{p_H^D}{w_u}\right]^{\frac{1-h}{h}}$$
$$L_s w_u^{\frac{1-h}{h}}\left[(1-\alpha)w_s + \alpha\bar{h}p_H^D\right] = (p_H^D)^{\frac{1}{h}}$$

$$(p_H^D)^{\frac{1}{h}} - \alpha\bar{h}L_s w_u^{\frac{1-h}{h}} p_H^D = (1-\alpha)L_s w_u^{\frac{1-h}{h}} w_s. \tag{A.1}$$

Similarly, the housing market clearing condition for suburb area can be simplified as

$$(p_H^S)^{\frac{1}{h}} - \alpha\bar{h}L_u w_u^{\frac{1-h}{h}} p_H^S = (1-\alpha)L_u w_u^{\frac{1}{h}}. \tag{A.2}$$

Recall the spatial no-arbitrage conditions for skilled and unskilled workers can be written as

$$\Gamma_s(p_H^D)^{1-\alpha} + p_H^D\bar{h} = w_s, \tag{A.3}$$

$$\Gamma_u(p_H^S)^{1-\alpha} + p_H^S\bar{h} = w_u, \tag{A.4}$$

where $\Gamma_u = \bar{U}_u P^\alpha$, and $\Gamma_s = \bar{U}_s P^\alpha$, are economic-wide constants to be pinned down in the general equilibrium. In particular, we normalize $\Gamma_u = 1$ and back out the ratio $\bar{U}_s/\bar{U}_u$ from the skill premium in the data.

The system of four equations (A.1), (A.2), (A.3) and (A.4) contain four unknowns, which can be exactly identified. Hence, given city size (L_s, L_u), the local wages w_s, w_u and house prices p_H^D, p_H^S can be computed. We can only obtain the numerical solution for these unknowns instead of the explicit analytical expressions because the system of equations is non-linear.

Plugging equation (A.4) into (A.2) to replace w_u yields the following non-linear equation to that pins down p_H^S,

$$\begin{aligned}
(1-\alpha)L_u w_u^{\frac{1}{h}} &= (p_H^S)^{\frac{1}{h}} - \alpha\bar{h}L_u w_u^{\frac{1-h}{h}} p_H^S \\
(1-\alpha)L_u &= (p_H^S)^{\frac{1}{h}} w_u^{-\frac{1}{h}} - \alpha\bar{h}L_u w_u^{-1} p_H^S \\
(1-\alpha)L_u &= (p_H^S)^{\frac{1}{h}} (\Gamma_u (p_H^S)^{1-\alpha} + p_H^S \bar{h})^{-\frac{1}{h}} - \alpha\bar{h}L_u (\Gamma_u (p_H^S)^{1-\alpha} + p_H^S \bar{h})^{-1} p_H^S \\
(1-\alpha)L_u &= (\Gamma_u (p_H^S)^{-\alpha} + \bar{h})^{-\frac{1}{h}} - \alpha\bar{h}L_u (\Gamma_u (p_H^S)^{-\alpha} + \bar{h})^{-1}
\end{aligned}$$

Given p_H^S, we can immediately compute unskilled worker wages according to labor mobility condition (A.4). Plugging w_u and equation (A.3) into (A.1) yields the equation that implicitly determines housing price for skilled labor p_H^D,

$$\begin{aligned}
(p_H^D)^{\frac{1}{h}} w_u^{\frac{h-1}{h}} - \alpha\bar{h}L_s p_H^D &= (1-\alpha)L_s(\Gamma_s (p_H^D)^{1-\alpha} + p_H^D \bar{h}) \\
(p_H^D)^{\frac{1}{h}} w_u^{\frac{h-1}{h}} &= (1-\alpha)L_s \Gamma_s (p_H^D)^{1-\alpha} + \bar{h}L_s p_H^D \\
(p_H^D)^{\frac{1}{h}-1} w_u^{\frac{h-1}{h}} &= \bar{h}L_s + (1-\alpha)L_s \Gamma_s (p_H^D)^{-\alpha}
\end{aligned}$$

The skilled labor wage w_s can thus be computed from equation (A.3).

A.3.6 Cost Function

Recall the production function of a firm is

$$y_j(z) = k^{\gamma_j} \ell(q,\varphi)^{1-\gamma_j}$$

$$\text{where} \quad \ell(q,\varphi) = \left[\chi_u(q,\varphi)^{\frac{1}{\sigma_L}} (\ell_u)^{\frac{\sigma_L-1}{\sigma_L}} + \lambda^{\frac{1}{\sigma_L}} \chi_s(q,\varphi)^{\frac{1}{\sigma_L}} (q\ell_s)^{\frac{\sigma_L-1}{\sigma_L}}\right]^{\frac{\sigma_L}{\sigma_L-1}}.$$

Since the cost function has two layers, Cobb-Douglas and CES, we solve the cost minimization problem in two steps. In the first step, we regards $\ell(q,\varphi)$ as a composite labor input with price $\tilde{w}$. The production function is Cobb-Douglas and thus the cost minimization problem is given by

$$\min_{\ell,k} \quad \tilde{r}k + \tilde{w}\ell(q,\varphi)$$

$$\text{subject to} \quad y_j(z) \leqslant k^{\gamma_j} \ell(q,\varphi)^{1-\gamma_j}$$

The Lagrangian is

$$\mathcal{L}(k,\ell,\kappa;\tilde{w},\tilde{r},q,\varphi) = \tilde{r}k + \tilde{w}\ell(q,\varphi) - \kappa\left(y_j(z) - k^{\gamma_j}\ell(q,\varphi)^{1-\gamma_j}\right).$$

Take first-order conditions of $\mathcal{L}$ w.r.t. $\ell(q,\varphi)$ and k, we can obtain the condition in which the iso-quant is tangent to the iso-cost,

$$\frac{\ell(q,\varphi)}{k} = \frac{1-\gamma_j}{\gamma_j}\left(\frac{\tilde{w}(q,\varphi)}{\tilde{r}}\right)^{-1}.$$

Solving this equation for labor yields $\ell(q,\varphi) = \frac{1-\gamma_j}{\gamma_j}\frac{\tilde{r}}{\tilde{w}(q,\varphi)}k$. Then substitute $\ell(q,\varphi)$ into the constraint,

$$y = \left(\frac{\tilde{r}}{\tilde{w}(q,\varphi)}\frac{1-\gamma_j}{\gamma_j}\right)^{1-\gamma_j} k$$

Solve for k and l in the expression of y,

$$k = \frac{y}{\left(\frac{\tilde{r}}{\tilde{w}(q,\varphi)}\frac{1-\gamma_j}{\gamma_j}\right)^{1-\gamma_j}}, \qquad l = \frac{\frac{1-\gamma_j}{\gamma_j}\frac{\tilde{r}}{\tilde{w}(q,\varphi)}y}{\left(\frac{\tilde{r}}{\tilde{w}(q,\varphi)}\frac{1-\gamma_j}{\gamma_j}\right)^{1-\gamma_j}}$$

The costs function can be expressed as

$$c(\tilde{w},\tilde{r},y) = \tilde{r}k + \tilde{w}(q,\varphi)\ell(q,\varphi) = (1-\gamma_j)^{\gamma_j-1}\gamma_j^{-\gamma_j}\tilde{r}^{\gamma_j}\tilde{w(q,\varphi)}^{1-\gamma_j}y.$$

When $y=1$, the cost function capture the unit cost of production.

In the second step, we characterize the costs function of the CES layer. The costs minimization problem of firm is such that

$$\min_{\ell_s,\ell_u} \quad w_s\ell_s + w_u\ell_u$$

$$\text{subject to} \quad \ell \leqslant \left[\chi_u(q,\varphi)^{\frac{1}{\sigma_L}}\ell_u^{\frac{\sigma_L-1}{\sigma_L}} + \lambda^{\frac{1}{\sigma_L}}\chi_s(q,\varphi)^{\frac{1}{\sigma_L}}\ell_s^{\frac{\sigma_L-1}{\sigma_L}}\right]^{\frac{\sigma_L}{\sigma_L-1}}.$$

The Lagrangian is

$$\mathcal{L}(\ell_s, \ell_u; q, \varphi, w_s, w_u) = w_s \ell_s + w_u \ell_u - \rho \left(\ell - \left[\chi_u(q,\varphi)^{\frac{1}{\sigma_L}} \ell_u^{\frac{\sigma_L - 1}{\sigma_L}} + \lambda^{\frac{1}{\sigma_L}} \chi_s(q,\varphi)^{\frac{1}{\sigma_L}} \ell_s^{\frac{\sigma_L - 1}{\sigma_L}} \right]^{\frac{\sigma_L}{\sigma_L - 1}} \right)$$

Take first-order conditions of $\mathcal{L}$ w.r.t. ℓ_s and ℓ_u and solve for ℓ_s

$$\ell_s = \lambda \frac{\chi_s(q,\varphi)}{\chi_u(q,\varphi)} \left(\frac{w_s}{w_u} \right)^{-\sigma_L} \ell_u.$$

Substituting ℓ_s into the constraint gives

$$\ell_u = \frac{\chi_u(q,\varphi)\ell}{\left[\chi_u(q,\varphi) + \lambda \chi_s(q,\varphi) \left(\frac{w_s}{w_u}\right)^{1-\sigma_L}\right]^{\frac{\sigma_L}{\sigma_L - 1}}}, \qquad \ell_s = \frac{\lambda \chi_s(q,\varphi) \left(\frac{w_s}{w_u}\right)^{-\sigma_L} \ell}{\left[\chi_u(q,\varphi) + \lambda \chi_s(q,\varphi) \left(\frac{w_s}{w_u}\right)^{1-\sigma_L}\right]^{\frac{\sigma_L}{\sigma_L - 1}}}.$$

The cost function for producing ℓ is such that

$$c(w_u, w_s, q, \varphi, \ell) = w_s \ell_s + w_u \ell_u = \frac{w_u \chi_u(q,\varphi) + w_s \lambda \chi_s(q,\varphi) \left(\frac{w_s}{w_u}\right)^{-\sigma_L}}{\left[\chi_u(q,\varphi) + \lambda \chi_s(q,\varphi) \left(\frac{w_s}{w_u}\right)^{1-\sigma_L}\right]^{\frac{\sigma_L}{\sigma_L - 1}}} \ell$$

$$= \left[\chi_u(q,\varphi) w_u^{1-\sigma_L} + \lambda \chi_s(q,\varphi) w_s^{1-\sigma_L} \right]^{\frac{1}{1-\sigma_L}} \ell.$$

The cost of producing one unit of ℓ is

$$\tilde{w}(w_u, w_s, q, \varphi) = \left[\chi_u(q,\varphi) w_u^{1-\sigma_L} + \lambda \chi_s(q,\varphi) w_s^{1-\sigma_L} \right]^{\frac{1}{1-\sigma_L}}.$$

Firms demands for skilled and unskilled labor as input are such that

$$\ell_u = \chi_u(q,\varphi) \left(\frac{w_u}{\tilde{w}(w_u, w_s, q, \varphi)} \right)^{-\sigma_L} \tilde{w}(w_u, w_s, q, \varphi) \ell,$$

$$\ell_s = \lambda \chi_s(q,\varphi) \left(\frac{w_s}{\tilde{w}(w_u, w_s, q, \varphi)} \right)^{-\sigma_L} \tilde{w}(w_u, w_s, q, \varphi) \ell.$$

The cost function for production is

$$C_j(z; q, \varphi) = \tilde{\gamma}_j \tilde{r}^{\gamma_j} \tilde{w}(q, \varphi, L_s, L_u)^{1-\gamma_j},$$

where $\quad \tilde{\gamma_j} = (1-\gamma_j)^{\gamma_j - 1}\gamma_j^{-\gamma_j}$, and $\quad \tilde{w}(q, \varphi, L_s, L_u) = [\chi_u(q, \varphi) w_u(L_s, L_u)^{1-\sigma_L} + \lambda \chi_s(q, \varphi) w_s(L_s, L_u)^{1-\sigma_L}]^{\frac{1}{1-\sigma_L}}$.

A.4 Model Fit

Figure A.1: Firm size (revenue) distribution, sector by sector

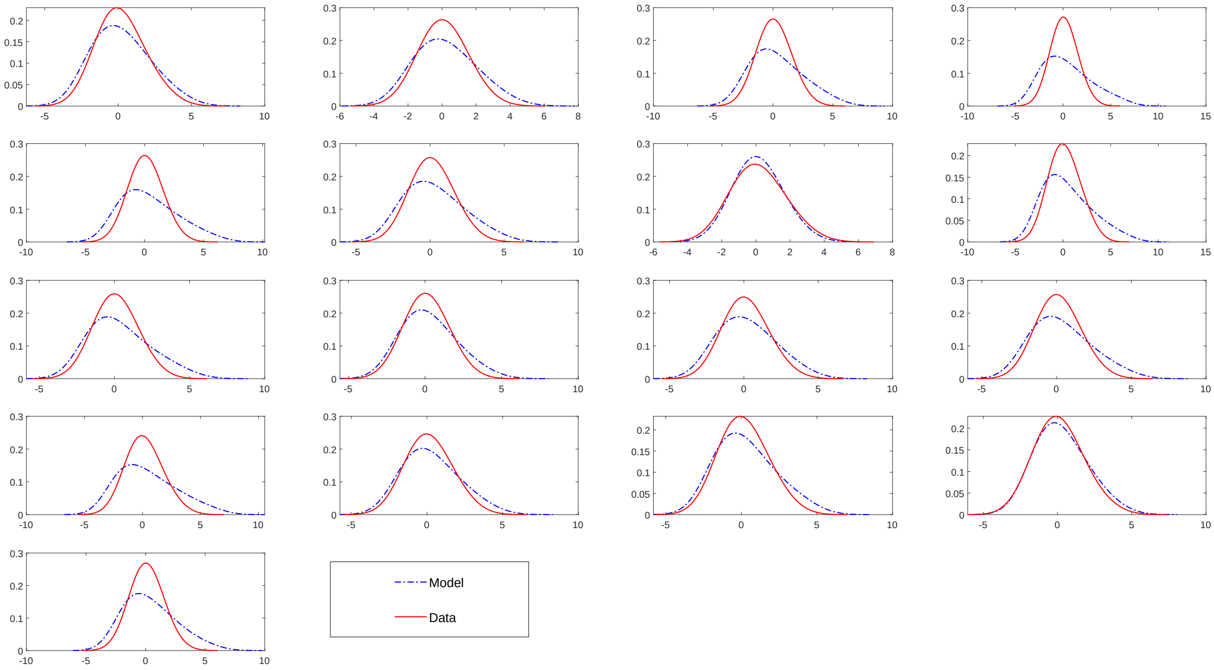


Figure A.2: Share of value added, sector by sector

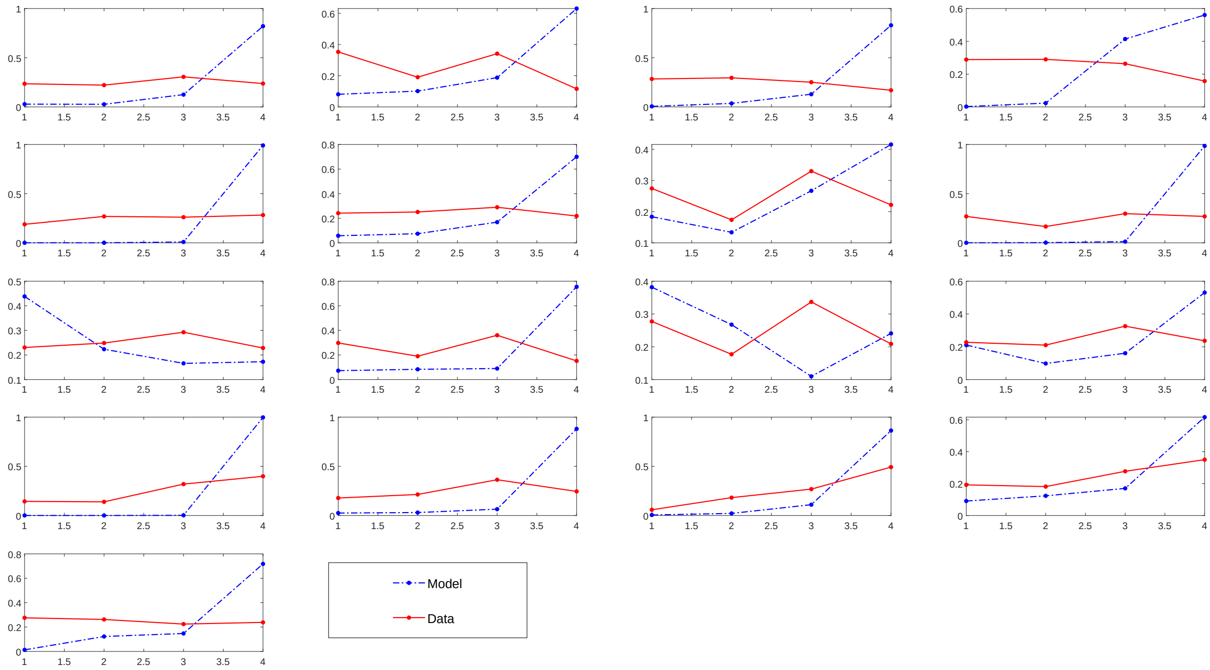


Figure A.3: Average value added, sector by sector

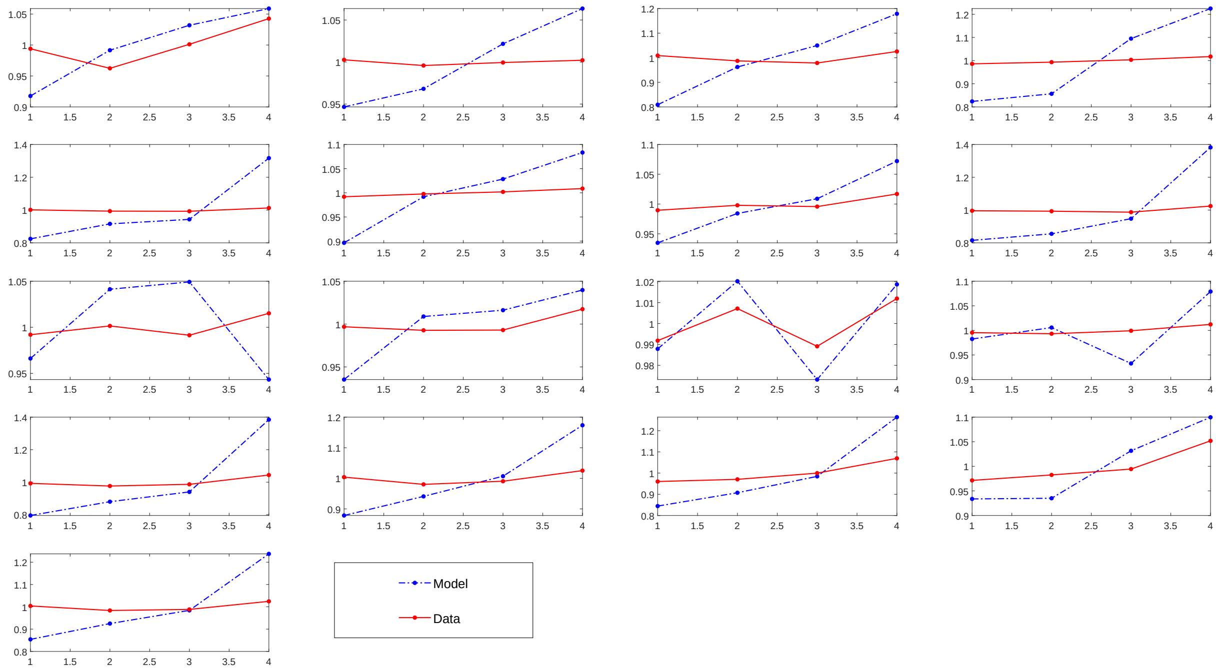
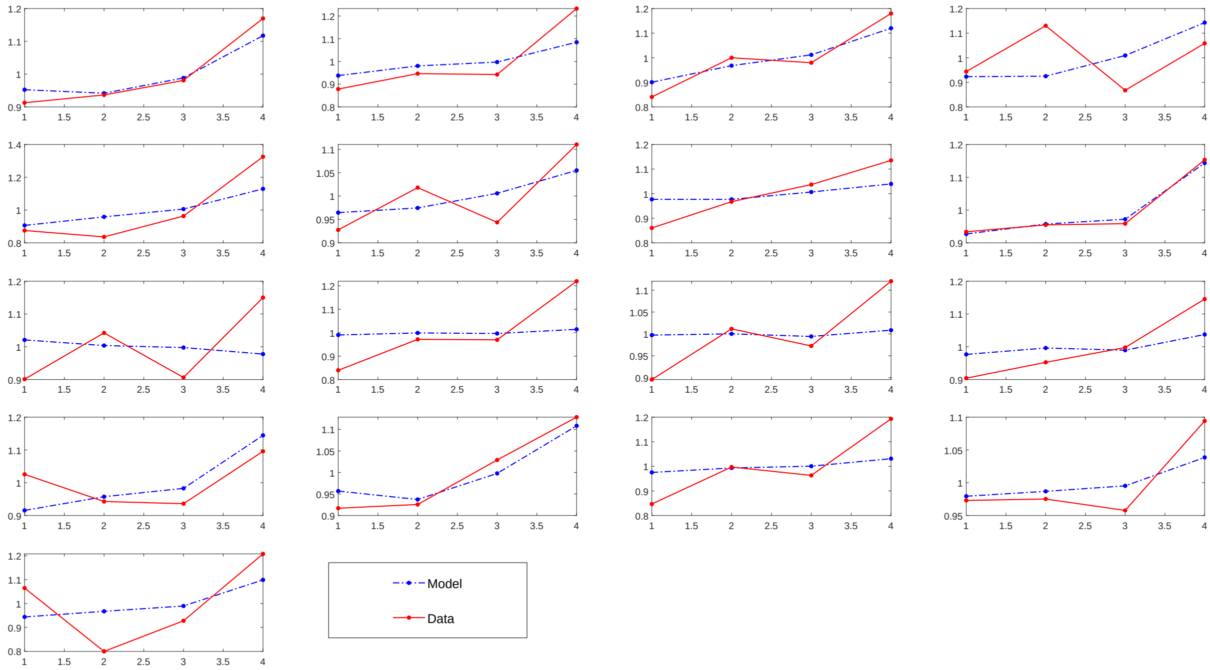
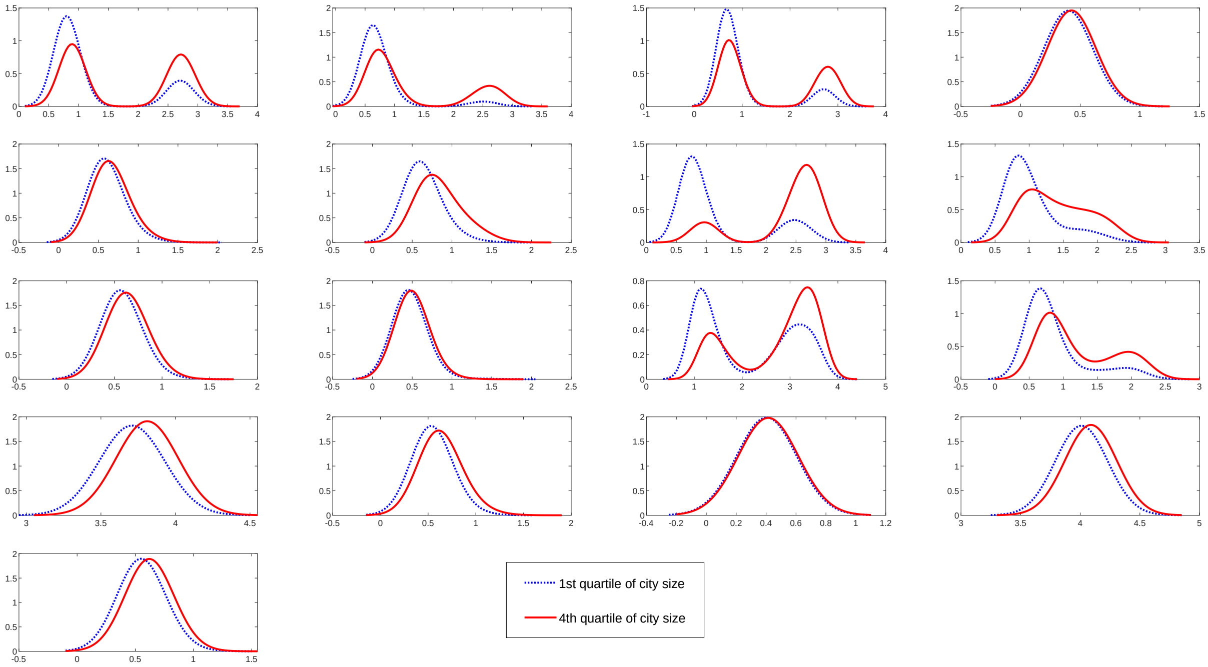


Figure A.4: Average skill intensity, sector by sector



A.5 Sensitivity Analysis

Figure A.5: Quality distribution in big vs small cities, alternative weighting matrix



B Appendix to Chapter 2

B.1 Proof of Lemma 2.1

It can be shown that we can manipulate the market access term $\widetilde{MA}_{i,\varphi}$ as follows

$$\begin{aligned}\widetilde{MA}_{i,\varphi} &= \int_{\tilde{P}_k^R}^{\tilde{P}^{max}} \frac{s_{ik}}{1-F^i\left(\tilde{P}_k^R\right)} \cdot \tilde{P}_k^{\sigma-1} E_k \, d\tilde{P}_k = \int_{\tilde{P}_k^R}^{\tilde{P}^{max}} \frac{s_{ik}}{1-F^i\left(\tilde{P}_k^R\right)} \cdot \overline{\tilde{P}_k^{\sigma-1} E_k}^R \, d\tilde{P}_k \\ &= \frac{\overline{\tilde{P}_k^{\sigma-1} E_k}^R}{1-F^i\left(\tilde{P}_k^R\right)} \int_{\tilde{P}_k^R}^{\tilde{P}^{max}} s_{ik} \, d\tilde{P}_k = \frac{\overline{\tilde{P}_k^{\sigma-1} E_k}^R}{1-F^i\left(\tilde{P}_k^R\right)} \left[1-F^i\left(\tilde{P}_k^R\right)\right] = \overline{\tilde{P}_k^{\sigma-1} E_k}^R,\end{aligned}$$

where $\overline{\tilde{P}_k^{\sigma-1} E_k}^R$ is denoted as the average value of $\tilde{P}_k^{\sigma-1} E_k$ for any $\tilde{P}_k \geqslant \tilde{P}_k^R$. Given this algebraic manipulation, it is obvious that $\frac{\partial \widetilde{MA}_{i,\varphi}}{\partial \tilde{P}_k^R} > 0$ because the average value of $\tilde{P}_k^{\sigma-1} E_k$ for $\tilde{P}_k \geqslant \tilde{P}_k^R$ increases with $\tilde{P}_k^R$.

B.2 Proof of Proposition 2.3

These comparative statics are derived from the optimal search (2.3.4) and quality decisions (2.5) which are reproduced as follows,

$$\widetilde{\pi}^*(\varphi, q^*) \left[\frac{\partial \Phi(\varphi, q^*)}{\partial q} \frac{1}{\Phi(\varphi, q^*)} + \frac{1-\sigma}{\partial \tilde{c}} \frac{\partial \tilde{c}}{\tilde{q}} \right] = 2\beta q^*$$

$$f_i^S = \int_{\tilde{P}_k^R}^{\tilde{P}_k^{max}} \pi_{ik}(\tilde{P}'_k, \varphi) - \pi_{ij}(\tilde{P}_j^R, \varphi) \, dF^i(\tilde{P}'_k)$$

We will prove the part of a lower f_i^S using Implicit Function Theorem. The proof of the second part concerning a more efficient firm is similar. We can manipulate the above equations as follows

$$F_1 \equiv \widetilde{\pi}^*(\varphi, q^*) \left[\frac{\partial \Phi(\varphi, q^*)}{\partial q} \frac{1}{\Phi(\varphi, q^*)} + \frac{1-\sigma}{\partial \tilde{c}} \frac{\partial \tilde{c}}{\tilde{q}} \right] - 2\beta q^* = 0$$

$$F_2 \equiv \int_{\tilde{P}_k^R}^{\tilde{P}_k^{max}} \pi_{ik}(\tilde{P}'_k, \varphi) - \pi_{ij}(\tilde{P}_j^R, \varphi) \, dF^i(\tilde{P}'_k) - f_i^S = 0$$

Totally differentiating these expressions with respect to f_i^S we have

$$\frac{\partial F_1}{\partial q}\frac{\partial q^*}{\partial f_i^S}+\frac{\partial F_1}{\partial \tilde{P}_k^R}\frac{\partial \tilde{P}_k^R}{\partial f_i^S}+\frac{\partial F_1}{\partial f_i^S}=0$$
$$\frac{\partial F_2}{\partial q}\frac{\partial q^*}{\partial f_i^S}+\frac{\partial F_2}{\partial \tilde{P}_k^R}\frac{\partial \tilde{P}_k^R}{\partial f_i^S}+\frac{\partial F_2}{\partial f_i^S}=0$$

Several observations are in order before we move on to derive the signs. First, $\frac{\partial F_1}{\partial q}<0$ as this is guaranteed by the second-order condition of firm's optimal quality choice. Second, $\frac{\partial F_2}{\partial q}>0$ which is very easy to check. Next, $\frac{\partial F_1}{\partial \tilde{P}_k^R}>0$ because it is proportional to $\frac{\partial \widetilde{MA}_{i,\varphi}}{\partial \tilde{P}_k^R}$ which is larger than 0 by Lemma 2.1. In addition, $\frac{\partial F_2}{\partial \tilde{P}_k^R}<0$ since the domain of the integral and the integrand would both be smaller with a larger $\tilde{P}_k^R$. Lastly, it is obvious that $\frac{\partial F_1}{\partial f_i^S}=0$ and $\frac{\partial F_2}{\partial f_i^S}=-1$.

Given these observations, we can now apply Cramer's rule as follows. First of all, the above system can be written in matrix form as

$$\begin{bmatrix}\frac{\partial F_1}{\partial q} & \frac{\partial F_1}{\partial \tilde{P}_k^R}\\ \frac{\partial F_2}{\partial q} & \frac{\partial F_2}{\partial \tilde{P}_k^R}\end{bmatrix}\begin{bmatrix}\frac{\partial q^*}{\partial f_i^S}\\ \frac{\partial \tilde{P}_k^R}{\partial f_i^S}\end{bmatrix}=\begin{bmatrix}-\frac{\partial F_1}{\partial f_i^S}\\ -\frac{\partial F_2}{\partial f_i^S}\end{bmatrix}.$$

Solving this system requires to do the following manipulation

$$\begin{bmatrix}\frac{\partial q^*}{\partial f_i^S}\\ \frac{\partial \tilde{P}_k^R}{\partial f_i^S}\end{bmatrix}=\begin{bmatrix}\frac{\partial F_1}{\partial q} & \frac{\partial F_1}{\partial \tilde{P}_k^R}\\ \frac{\partial F_2}{\partial q} & \frac{\partial F_2}{\partial \tilde{P}_k^R}\end{bmatrix}^{-1}\begin{bmatrix}-\frac{\partial F_1}{\partial f_i^S}\\ -\frac{\partial F_2}{\partial f_i^S}\end{bmatrix}=\frac{1}{\frac{\partial F_1}{\partial q}\frac{\partial F_2}{\partial \tilde{P}_k^R}-\frac{\partial F_1}{\partial \tilde{P}_k^R}\frac{\partial F_2}{\partial q}}\begin{bmatrix}\frac{\partial F_2}{\partial \tilde{P}_k^R} & -\frac{\partial F_1}{\partial \tilde{P}_k^R}\\ -\frac{\partial F_2}{\partial q} & \frac{\partial F_1}{\partial q}\end{bmatrix}\begin{bmatrix}-\frac{\partial F_1}{\partial f_i^S}\\ -\frac{\partial F_2}{\partial f_i^S}\end{bmatrix}$$

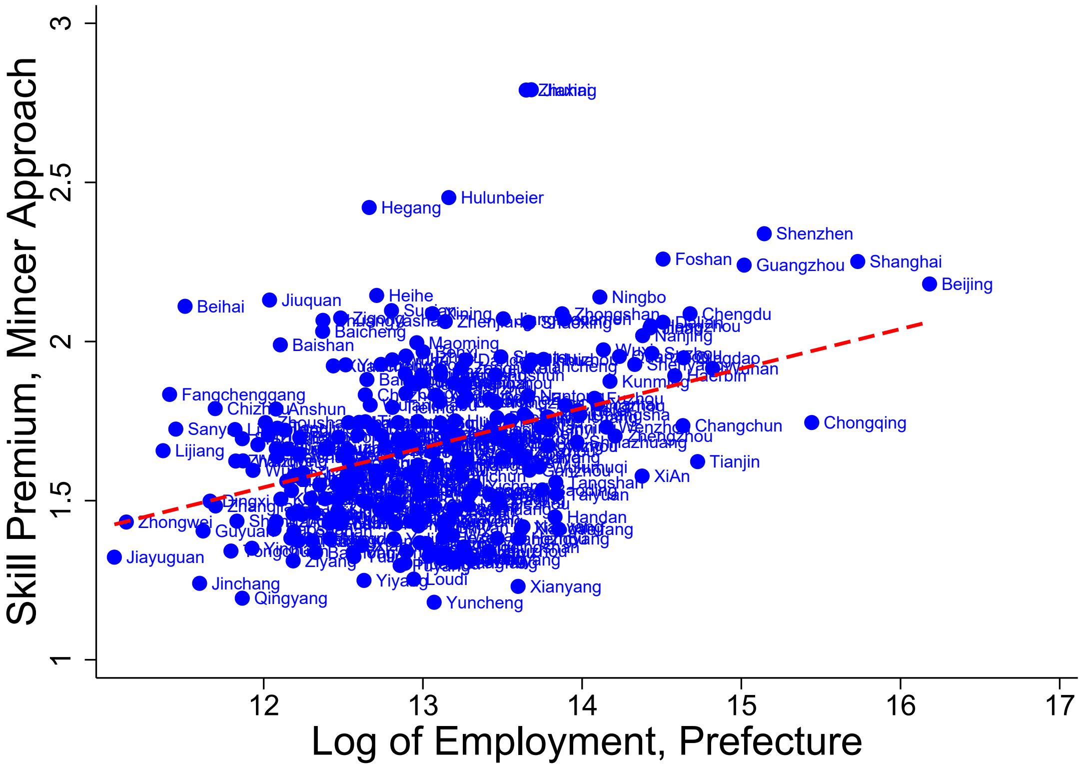
Hence, it is obvious that

$$\text{Sign}\left(\frac{\partial q^*}{\partial f_i^S}\right)=\text{Sign}\left(\frac{\partial \tilde{P}_k^R}{\partial f_i^S}\right)=-\text{Sign}\left(\frac{\partial F_1}{\partial q}\frac{\partial F_2}{\partial \tilde{P}_k^R}-\frac{\partial F_1}{\partial \tilde{P}_k^R}\frac{\partial F_2}{\partial q}\right).$$

C Appendix to Chapter 3

C.1 Skill Premium and Employment Size

Figure C.1: Skill Premium and Employment Size



Notes: The data source is 2005 1% Population Census. "Skilled Worker" is defined as college-graduated workers. "Skill Premium" is estimated with Mincer approach.