6-2020

# Three essays on international trade policies

Xin YI
*Singapore Management University*, xinyi.2015@phdecons.smu.edu.sg

SINGAPORE MANAGEMENT UNIVERSITY

PHD DISSERTATION

# Three Essays on International Trade Policies

Xin Yi

Supervised by

Associate Professor Pao-Li Chang

May 15, 2020

# Three Essays on International Trade Policies

by Xin Yi

(B.S., Nanyang Technological 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)
Associate Professor of Economics
Lee Kong Chian Fellow
Singapore Management University

Jungho Lee
Assistant Professor of Economics
Singapore Management University

Yuan Mei
Assistant Professor of Economics
Singapore Management University

Bingjing Li
Assistant Professor of Economics
National University of Singapore

Singapore Management University

2020

# Abstract

This dissertation studies the empirical and quantitative implications of trade policies. The first chapter examines the effects of trade policies on quality specialization across cities within a country. Specifically, we complement the quality specialization literature in international trade and study how larger cities within a country produce goods with higher quality. We first establish three stylized facts on how product quality is related to agglomeration, firm productivity, and worker skills. We then rationalize these facts in a spatial equilibrium model where all the elements mentioned above are present and firms are free to choose their locations. Using firm-level data from China, we structurally estimate the model and find that agglomeration and spatial sorting of firms each accounts for about 50% of the spatial variation in quality specialization. A counterfactual to relax land use regulation in housing production raises product quality in big cities by 5.5% and indirect welfare of individuals by 6.2%. The second chapter zooms into distributional issues and studies the implication of rising income inequality on product price dispersion. Using big data on a broad set of goods sold in the US (Nielsen Retail Scanner Data) from 2006 to 2017, we find that in general there is a missing middle phenomenon, where the product price distribution loses its mass in the middle price support. In addition, we find that this pattern is more pronounced in the densely populated metropolitan areas. We further link this observation to changes in income inequality, which are measured from a panel of US households from 2006 to 2017 (IPUMS ACS). The results support our conjecture that demand-side demographics has a significant influence on the missing middle phenomenon. The third chapter examines the transition dynamics of trade liberalization. In particular, we develop a multi-country, multi-sector quantitative trade model with dynamic Roy elements such as occupational choice and occupation-specific human capital accumulation. Given an abrupt trade liberalization, a country that is relatively more productive in some sectors may not have comparative advantage initially, as it takes time to accumulate occupation-specific human capital which increases occupational skill supply endogenously. We quantify this transition dynamics and its distributional consequences by calibrating the model to a North-South setup.

# Contents

# Acknowledgment

I am extremely grateful to Pao-Li Chang, Jungho Lee, and Yuan Mei for their guidance and advice. This dissertation would not have been possible without their patience, inspirations, and invitations to collaborate on research projects. I am also grateful to Davin Chor, Bingjing Li, and Lin Ma. It was my fortune to learn from their industrious scholarship in the reading group at NUS. Finally, I want to thank my parents, my aunts, and especially Yew-Kwang Ng for encouraging me to pursuit a doctorate.

# 1 Quantifying Quality Specialization Across Space

## 1.1 Introduction

Firms in big cities specialize in high-quality products (Dingel, 2017; Saito and Matuura, 2016).[1] 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.[2] 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.

[1] This chapter is a joint work with Angdi Lu who also lists it as a chapter in her dissertation. I certify to have contributed a significant amount of work which includes building the theoretical model and devising the identification strategy for the stylized facts and structural estimation.

[2] 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.

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 endogenously 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 et al., 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 et al. (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 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 plant-level data from the Chinese Manufacturing Census. 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 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 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 et al., 2014; Combes et al., 2012; Albouy, 2012; Duranton and Puga, 2004; Rosenthal and Strange, 2004; Glaeser et al., 2001; Glaeser, 1999; Glaeser et al., 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 premia and skill compositions across cities (Davis and Dingel, 2019, 2017; Glaeser and Maré, 2001; Baum-Snow and Pavan, 2012, 2013; Baum-Snow et al., 2018; Moretti, 2013; Diamond, 2016; De La Roca and Puga, 2017; Combes et al., 2008; Dingel et al., 2019; Davis et al., 2018; Behrens and Robert-Nicoud, 2015; Farrokhi and Jinkins, 2019; Lindley and Machin, 2016; Hendricks, 2011; Bacolod et al., 2009; Chor, 2005; D'Costa and Overman, 2014; Florida et al., 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 et al., 2009). A recent literature pioneered by Davis and Dingel (2019) provide evidences that skill premia 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 premia 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 premia 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 et al., 2011, 2015; Hallak, 2006, 2010; Choi et al., 2009) and the supply side explanations (Fieler et al., 2018; Dingel, 2017; Faber and Fally, 2017; Fan et al., 2017; Antoniades, 2015; Feenstra and Romalis, 2014; Hallak and Sivadasan, 2013; Kugler and Verhoogen, 2012; Crozet et al., 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.

## 1.2 Empirical 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.

### 1.2.1 Data and Measurement

We merge two databases 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 Database (IFPQD).[3] This dataset contains information on the physical quantity of firm output, and it has been used in other literature to measure product quality Fan et al. (2018). These two datasets both cover the universe of Chinese manufacturing firms and use the same firm identification.[4] [5] 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 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

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

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

[5]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 IFPQD dataset reports the quantity information of each 5-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.

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 et al. (2015) which address the zero investment and control function collinearity problems.[6]

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.

### 1.2.2 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

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

cities in a given sector and year, we estimate the following specification:

$$q_{ijkt} = \beta_1 \ln CitySize_{it} + \gamma' X_{it} + D_{gt} + D_i \times t + \epsilon_{ijkt} \quad (1)$$

where $q_{ijkt}$ is a measure for quality of goods produced by firm $j$ from city $i$ in sector $k$, and $\ln CitySize_{it}$ is the natural log of employment size of city $i$ during year $t$. Standard errors are clustered by city to account for possible correlations of idiosyncratic noises within each region. The results are qualitatively similar if the standard errors are clustered at the city-sector level. The main coefficient of interest is $\beta_1$, 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-time 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 et al., 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 et al., 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.[7] We expect these coefficients to be positive, because firms that are closer to high-income cities and firms

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

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.[8] 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 sector, 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-series variation 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

[8]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.

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. 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 1.4, 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.

Table 1: Firms in larger cities produce higher-quality goods

| Dependent Variable | Log of Prices | Log of Market Shares | Estimated Quality |
|---|---|---|---|
| | (1) | (2) | (3) |
| City Size | 0.070*** | 0.028 | 0.066*** |
| | (0.017) | (0.028) | (0.025) |
| 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 |
| $R^2$ | 0.861 | 0.389 | 0.908 |
| $N$ | 313,242 | 313,407 | 313,240 |

*Notes:* "Estimated Quality" is based on Khandelwal (2010). "City Size" is the log of employment size in a prefecture. "Market Access" is the log of the sum of prefectures' GDP per capita weighted by trade costs. "Export Status" 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.

**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:

$$q_{ijkt} = \beta_1 z_{ijkt} + \alpha \mathbf{1}_t\{j = exporter\} + D_{gt} + D_i \times t + \varepsilon_{ijkt} \tag{2}$$

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 $\beta_1$, 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 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.

Table 2: More productive firms specialize in higher-quality products

| Dependent variable | Log of Prices | Log of Market Shares | Estimated Quality |
|---|---|---|---|
| | (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) |
| 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 |
| $R^2$ | 0.863 | 0.426 | 0.934 |
| $N$ | 217,749 | 217,750 | 217,748 |

*Notes:* "Estimated Quality" is based on Khandelwal (2010). "City Size" is the log of employment size in a prefecture. "Market Access" is the log of the sum of prefectures' GDP per capita weighted by trade costs. "Export Status" 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.

**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.

$$y_{ijk,2004} = \beta z_{ijk,2004} + \alpha \mathbf{1}_{2004}\{j = exporter\} + D_{ik} + D_g + \epsilon_{ijk,2004} \tag{3}$$

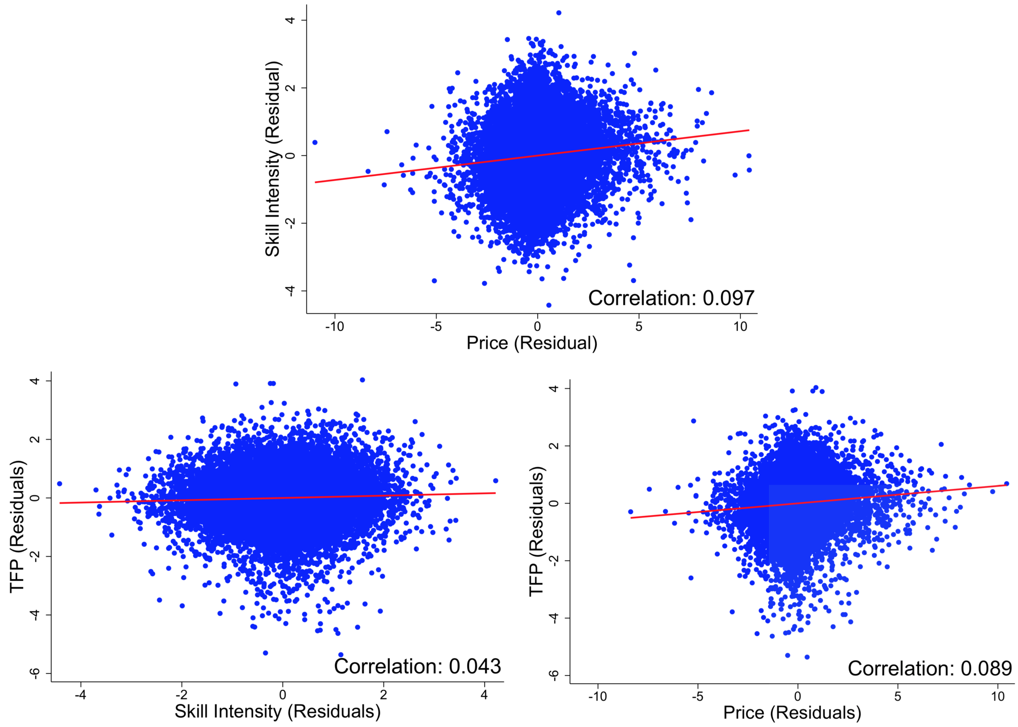
where $y_{ijk,2004}$ is an outcome variable such as skill intensity or product quality of firm $j$ in sector $k$ from city $i$ in year 2004. $z_{ijk,2004}$ is the productivity estimate of the same firm in 2004, and $D_{ik}$ is a city-sector specific fixed effects that control for comparative advantage 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 (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. 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 (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.

Figure 1: Firms specialize in higher-quality goods also employ more skilled workers




*Notes*: Observations are in 2004, the only year when the educational levels of firms' 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. Standard errors are clustered at prefecture-industry level.

## 1.3 The Model

### 1.3.1 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, $\Lambda$ 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 et al. (2018), Rossi-Hansberg (2005), and Ahlfeldt et al. (2015).

### 1.3.2 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 $\varphi$ 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 $\omega$ with quality $q$, and $\sigma_j$ 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.

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-$\zeta$ 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

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

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

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,

$$L_s \left[ (1-\alpha) \frac{w_s(L_s, L_u) - p_H^D(L_s, L_u)\bar{h}}{p_H^D(L_s, L_u)} + \bar{h} \right] = \left[ \frac{p_H^D(L_s, L_u)}{w_u(L_s, L_u)} \right]^{\frac{1-h}{h}}, \tag{4}$$

$$L_u \left[ (1-\alpha) \frac{w_u(L_s, L_u) - p_H^S(L_s, L_u)\bar{h}}{p_H^S(L_s, L_u)} + \bar{h} \right] = \left[ \frac{p_H^S(L_s, L_u)}{w_u(L_s, L_u)} \right]^{\frac{1-h}{h}}. \tag{5}$$

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

$$\bar{U}_s = \left( \frac{w_s - p_H^D \bar{h}}{P} \right)^{\alpha} \left( \frac{w_s - p_H^D \bar{h}}{p_H^D} \right)^{1-\alpha}, \tag{6}$$

$$\bar{U}_u = \left( \frac{w_u - p_H^S \bar{h}}{P} \right)^{\alpha} \left( \frac{w_u - p_H^S \bar{h}}{p_H^S} \right)^{1-\alpha}. \tag{7}$$

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 (4), (5), (6), and (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** *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.

### 1.3.3 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 $\ell$ 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. $\lambda$ 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 $\varphi$, so that more productive firms face lower marignal cost and also employ more skilled labor.[9] In addition, we follow Fieler et al. (2018) in assuming that to produce a variety with higher quality $q$, a firm has to employ relatively

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

more skilled workers. Taking the ratio over the factor demand of two labor 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 $\varphi$, 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

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

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** *$\varphi(z, L, q; s_j)$ 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 $\varphi$ 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.

### 1.3.4 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.

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

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

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 $\Upsilon_{1j}$ 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 $q^* = \operatorname{argmax}_{q\geqslant 0} \pi_j^*(z;q,L_s,L_u) - f_q q$, 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 $4.7\times10^{-5}$ in Fieler et al. (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 $\pi$ 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 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 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,*

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

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

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

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

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 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 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 $\pi_j^*(z_H) \geqslant \pi_j^*(z_L)$.*

### 1.3.5 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.

### 1.3.6 Parameterization and Calibration

In order to assess the quantitative behavior of the model, we first parameterize the firm productivity term following Fieler et al. (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 $\nu_{\phi,j}$.[10] [11] 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 et al. (2018). The interpretations of the parameters are as follows. First, $\lambda_{1s}$ and $\lambda_{1u}$ capture how the productivity of firms 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, $\lambda_{2s}$ and $\lambda_{2u}$ 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 et al. (2018).[12]

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

[11] 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.

[12] Fieler et al. (2018) use a slightly more complicated functional form. Still, our choices are largely similar to theirs.

### 1.3.7 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 8,000×200 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 $200 \times 200$, where we discretize the choice of quality over the interval of [0,10] with a step size of 0.05.

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.

## 1.4 Quantifying the Model

### 1.4.1 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 4-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 2-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 3. The concordance of sectors from CSIC to our definition of sectors is detailed in Table 20 in the appendix.

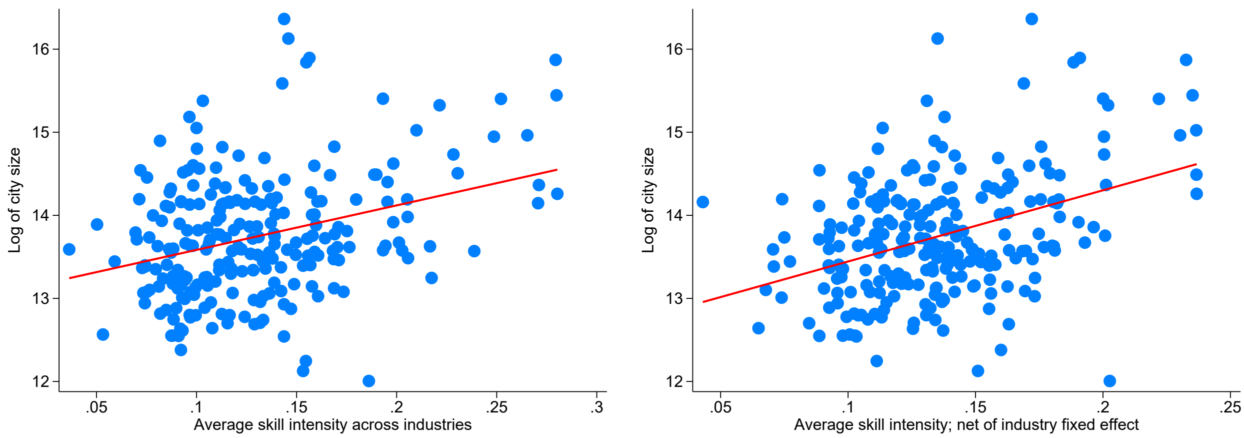
We obtain the geographic location of firms using postal code reported in the data. A "city" is defined as the prefecture unit in China. We obtain the prefecture-level population from China City Statistic Year Book in 2004 as a proxy for city size. There are 243 cities in our firm-level dataset. In addition, 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 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 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 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.

Table 3: Summary Statistics

| Sector | log Value Added | | | log Employment | | | % Skilled Workers | | | % in Big Cities | $N$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | $Q1$ | $Q4$ | Mean | $Q1$ | $Q4$ | Mean | $Q1$ | $Q4$ | | |
| Food | 9.89 | 9.12 | 10.92 | 4.74 | 4.04 | 5.57 | 11.3 | 5.4 | 22.1 | 20.9 | 6,712 |
| Textile | 9.80 | 9.18 | 10.57 | 5.01 | 4.36 | 5.67 | 3.8 | 1.8 | 8.0 | 20.2 | 29,948 |
| Leather | 9.90 | 9.24 | 10.69 | 5.20 | 4.58 | 5.89 | 3.3 | 1.6 | 6.9 | 19.9 | 5,055 |
| Wood | 9.52 | 8.92 | 10.28 | 4.60 | 4.08 | 5.15 | 5.5 | 2.5 | 11.1 | 13.5 | 3,880 |
| Furniture | 9.77 | 9.14 | 10.53 | 4.79 | 4.22 | 5.46 | 6.0 | 3.0 | 12.2 | 31.8 | 2,424 |
| Paper | 9.64 | 9.04 | 10.47 | 4.60 | 4.03 | 5.30 | 6.5 | 3.1 | 13.3 | 31.0 | 12,413 |
| Chemicals | 9.93 | 9.21 | 10.88 | 4.32 | 3.69 | 5.11 | 11.9 | 5.6 | 23.5 | 22.3 | 15,969 |
| Medical | 10.12 | 9.31 | 11.08 | 4.87 | 4.25 | 5.58 | 22.7 | 11.9 | 38.6 | 25.7 | 3,801 |
| Plastic | 9.66 | 9.06 | 10.45 | 4.50 | 3.91 | 5.19 | 7.0 | 3.4 | 14.3 | 28.0 | 12,902 |
| Minerals | 9.78 | 9.13 | 10.60 | 4.83 | 4.20 | 5.48 | 6.4 | 2.9 | 13.3 | 17.7 | 16,164 |
| Basic metals | 9.96 | 9.23 | 10.93 | 4.50 | 3.91 | 5.22 | 7.4 | 3.6 | 15.0 | 25.5 | 20,518 |
| Machinery | 9.68 | 9.08 | 10.51 | 4.55 | 3.99 | 5.22 | 10.5 | 5.0 | 20.6 | 24.7 | 24,953 |
| Transport | 9.97 | 9.24 | 10.95 | 4.79 | 4.19 | 5.58 | 10.8 | 5.2 | 21.2 | 31.1 | 9,365 |
| Electrical | 9.96 | 9.25 | 10.92 | 4.67 | 4.04 | 5.42 | 10.3 | 5.0 | 20.5 | 30.8 | 12,781 |
| Computer | 10.10 | 9.31 | 11.20 | 5.04 | 4.36 | 5.94 | 13.5 | 6.1 | 30.5 | 38.8 | 10,058 |
| Energy | 10.46 | 9.42 | 11.57 | 5.39 | 4.53 | 6.15 | 22.2 | 12.7 | 34.5 | 15.5 | 5,066 |
| Others | 9.67 | 9.08 | 10.42 | 4.98 | 4.30 | 5.69 | 4.5 | 2.1 | 9.7 | 20.8 | 3,825 |

Figure 2: Correlation of skill intensity and city size



### 1.4.2 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, $\Theta_1$, 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 et al. (2018) and Gaubert (2018).

1. **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 $\{\lambda_{1s}, \lambda_{1u}, \lambda_{2s}, \lambda_{2u}\} \in \Theta_2$.

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.

3. **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 et al. (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 et al., 2011; Gaubert, 2018; Antràs et al., 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.

### 1.4.3 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 $\hat{\lambda}_{2s} > \hat{\lambda}_{2u}$.[13]

### 1.4.4 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 3.

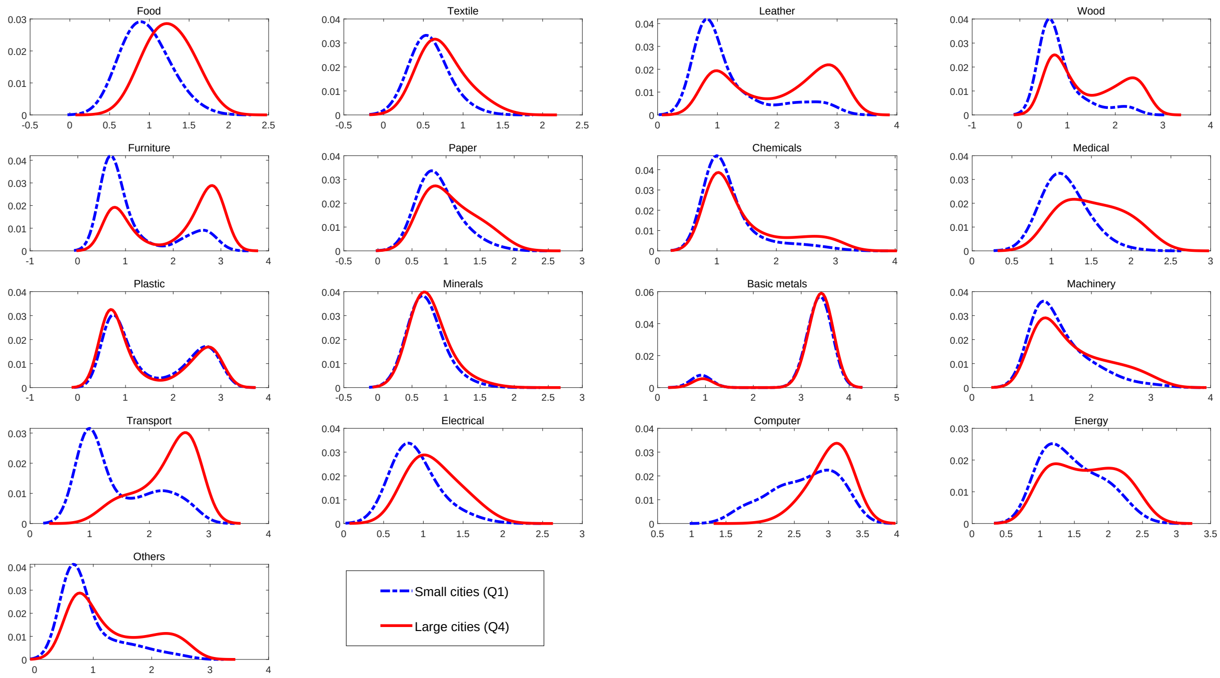
To further assess the contribution of firm sorting and traditional agglomeration benefit

[13]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 $\lambda_{2s} - \lambda_{2u} > 0$.

Figure 3: Quality distribution in big vs. small cities, sector by sector



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 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.

Table 4: Quality choices in different models

| Dep. variable: | Quality Choices | |
|---|---|---|
| | Full Model | W/O Sorting |
| log City Size | 0.094*** | 0.049*** |
| | (0.001) | (0.000) |
| Sectoral FE | Yes | Yes |
| $N$ | 85,000 | 85,000 |
| $R^2$ | 0.540 | 0.979 |

### 1.4.5 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 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 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 et al., 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 et al. (2019).

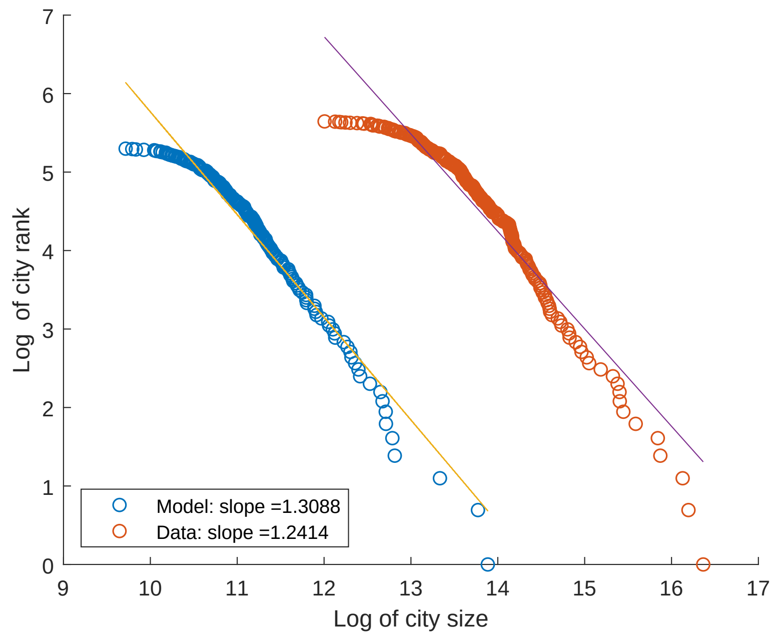
Table 5: Goodness of fit for targeted moments

| | | Quartiles & Percentiles | | | | |
|---|---|---|---|---|---|---|
| Moments | | Q1 | Q2 | Q3 | Q4 | P90 |
| 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 |

## 1.5 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

Figure 4: City size distribution, model and data




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 6. In overall, average quality across cities has increased by 5.5% while the aggregate welfare of all residents has increased by 20%.

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

Table 6: Counterfactual of relaxing land use regulations by 20%

| | Change in variables (%) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| City quartile | $\Delta q$ | $\Delta w_s$ | $\Delta w_u$ | $\Delta p_H^D$ | $\Delta p_H^S$ | $\Delta P$ | $\Delta W$ | $\Delta \tilde{W}$ |
| 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 |

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.

## 1.6 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.

# 2 The Missing Middle in Product Price Distribution

## 2.1 Introduction

The recent decades are often characterized by observers as decades of widening inequality, which includes cross-country evidences such as those in the United States (Piketty and Saez, 2003; Kopczuk et al., 2010; Saez and Zucman, 2016; Piketty et al., 2018; Autor, 2019), France (Piketty, 2003, 2011; Garbinti et al., 2018), China (Piketty et al., 2019), India (Chancel and Piketty, 2019), Japan (Moriguchi and Saez, 2008), and other countries (Piketty and Zucman, 2014; Alstadsaeter et al., 2018) as well as evidences on declining labor shares (Karabarbounis and Neiman, 2013; De Loecker et al., 2020).

In this paper, we further document another inequality phenomenon which the product price distribution loses its mass in the middle price support over time in the United States. Our work is motivated by the ad-hoc observations that the price distribution of goods and services are becoming more dispersed, in the sense that the market shares are increasing for the very cheap and very expensive products. For example, New York Times reported in 2014 that[14]

> Sears and JC Penney, retailers whose wares are aimed squarely at middle-class Americans, are both in dire straits. Last month, Sears said it would shutter its flagship store on State Street in downtown Chicago, and JC Penney announced the closings of 33 stores and 2,000 layoffs. Loehmann's, where generations of middle-class shoppers hunted for marked-down designer labels in the Back Room, is now being liquidated after three trips to bankruptcy court since 1999. ... Investors have taken notice of the shrinking middle. Shares of Sears and JC Penney have fallen more than 50 per cent since the end of 2009, even as upper-end stores like Nordstrom and bargain basement chains like Dollar Tree and Family Dollar Stores have more than doubled in value over the same period. ... But changes in the restaurant business show that the effects of rising inequality are widespread. Foot traffic at mid-tier, casual

[14]"Mid-Tier Stores close While Posh Ones Thrive", The New York Times. Feb 04, 2014

> dining properties like Red Lobster and Olive Garden has dropped in every quarter but one since 2005, according to Mr John Glass, a restaurant industry analyst at Morgan Stanley. With diners paying an average tab of US$16.50 a person at Olive Garden, he said: "The customers are middle class. They're not rich. They're not poor."

This effect is not limited in only a few sectors. Another article published on the New York Times on the same date suggests that it is rather ubiquitous[15]

> Among hotels, revenue per room in the high-end category, which includes brands like the Four Seasons and St Regis, grew 7.5 per cent last year, compared with a 4.1 per cent gain for mid-scale properties. ... And at General Electric, the increase in demand for high-end dishwashers and refrigerators dwarfs sales growth in mass-market models.

We document a more generalized version of this ad-hoc observation using big data on consumer grocery price data in the US from 2006 to 2017 (Nielsen Retail Scanner Data). This dataset contains retail information on 35,000 stores from about 90 retail chains. Weekly store-level sales data (price and volume) are recorded for approximately 3.2 million of products identified by the unique product product (UPC). These barcode-level products are then further classified into approximately 1075 product modules (PMC) and 125 product groups (PGC). We construct measures on the market share of low-, medium-, and high-price product groups for each year in each of the contiguous states (including the District of Columbia). Specifically, we first construct an indicative average price for each UPC in a given location and year by aggregating sales volumes and revenues across all weeks in a year. We regard these indicate prices of UPCs within a product group as the price distribution of the corresponding PGC. We then compute the market share measures based on these indicative prices and different thresholds for low-, medium-, and high-price products, with sales volumes of each UPC as weights. As a result, we obtain a panel of market share measures for low/middle/high-price products in the contiguous states from 2006-2017. In addition, we further construct similar market share measures at

[15] "Income Gap Changing US Consumer Landscape", The New York Times. Feb 04, 2014.

a more disaggregated geographical level, for 722 contiguous commuting zones, following similar procedures for the same time period.

Next, we link the change in market shares to the change in income distribution. In particular, we construct a corresponding set of income group share measures using the American Community Survey (ACS) dataset for each contiguous state and commuting zone from 2005 to 2017. The ACS samples include various demographic and financial information for approximately 3 million households in the US. Based on these data, we compute the household income per capita (in 1999 dollars) for each household.[16] This gives a distribution of household income in each location-year. We then infer the corresponding population shares for each income group from the distribution, using various definitions of income groups with number of persons residing in the households as weights.

In general, we find that there is indeed a missing middle phenomenon in both the population shares of income groups and the market shares of price groups across space and time. This phenomenon is robust to whether we compute the measure at a state-level or a more disaggregated commuting zone level. In particular, the time-series plots seem to suggest that the missing middle phenomenon is especially severed in the aftermath the Great Recession from 2007 to 2009. Furthermore, we introduce urban density, which is measured by the population density per square mile of land area, into both series of population/market shares. We find that the missing middle phenomenon is more pronounced in densely populated urban areas, especially for the population shares of income group measures.

We attempt to bridge the missing middle in the price distribution to that of the income distribution. To do so, we exploit the time-series variation in the data and regress the state-level market share measures for difference price groups on its lagged state population shares of income groups using OLS. We find that the changes in demand-side demographics has a significant influence on the changes in price dispersion, and this relationship is robust to the inclusion of various fixed effects as well as employing a more disaggregated specification at commuting zone level. Though the results are only correla-

[16]We excluded group definitions such as institutions and boarding houses.

tion in nature, it provides some causal evidence on how the missing middle phenomenon is driven by the change in demographics. In addition, we find that this effect is more pronounced in densely populated metropolitan areas.

Our paper is related to several literature. First, the IO literature usually studied the pricing decision of firms by taking the consumer demographics as given. We complement this literature in suggesting that changes in demand-side demographics could also be an important factor in shaping the price distribution. In particular, we argue that the rising income inequality in the last decade has driven the price distribution to be more dispersed. This changes may offer a fresh perspective on how demand-side factors should also be taken into account in government policies concerning the organization of industries.

Second, our paper is related to the literature on rising income inequality. We have already mentioned the various literature on the rising income inequality in the above. This literature is influential in the sense that studying the distributional consequences of economic trades are one of the most important research agenda of our time. However, despite all the careful analysis, none of them studied the impact of rising income inequality on price distribution to the best of our knowledge. Our work fills this gap.

Lastly, our work contributes to the growing literature that studies consumer welfare with barcode-level scanner data (Kaplan and Schulhofer-Wohl, 2017; Atkin et al., 2018; Jaravel and Sager, 2019; Faber and Fally, 2019; Jaravel, 2019; Argente and Lee, 2020; Atkin et al., 2020). Several papers from this literature is especially relevant to our work. Kaplan and Schulhofer-Wohl (2017) and Argente and Lee (2020) measure the changes in cost of living for different households using the Nielsen barcode data. Both papers find that the cost of living has risen more for the lower-income households. While their works are highly relevant to ours, they do not explicitly document the missing middle phenomenon as well as suggest any causal relationship between price and income distribution. Second, Jaravel (2019) studies how firms introduce new products in the upper price range to cater the increase in upper-income household demand. While both of our papers studies how firms react to changes in demographics, our work is significantly different from his. Our work focuses on the effect of firms shifting from middle prices to the lower and higher price products and is agnostic on how firms achieve so. In this sense, our work is more non-parametric than his. In addition, Jaravel (2019) only focuses on the price structure for the top-income groups without studying the changing structure of the entire price distribution. Finally, Jaravel (2019) studies nominal prices while we deflate the measure to study the changing distribution of real prices.

In a similar vein, Faber and Fally (2019) argues that more productive firms endogenously sort to produce higher-price products in order to cater the taste of richer households. We view the difference between our work and theirs to be similar as those comparisons with Jaravel (2019). Moreover, Faber and Fally (2019) implies a monotonic relationship between firm productivity and endogenous sorting. Our work complements theirs to cover the entire price distribution.

The rest of our paper is organized as follows. Section 2.2 documents the missing middle in product price distribution. Section 2.3 documents the missing middle in income distribution. Section 2.4 examines the correlations between changes in income and price distribution as well as suggests potential causal relationship. Finally, section 2.5 concludes and provides a blueprint for future work in adopting similar identification strategy in Jaravel (2019) and the semi-parametric welfare approach in Atkin et al. (2018).

## 2.2 Missing Middle in Product Price Distribution

This section explains how we construct market share measures for difference price groups and documents a missing middle in the price distribution over time from the Nielsen data. The Nielsen Retail Scanner data contains weekly statistics on the revenue and transaction volume for products sold in the US. This includes approximately 3.2 millions of products (UPC) in 35,000 stores under around 90 retail chains. These products are then classified into 1,075 product modules (PMC) and 125 product groups (PGC).

To construct a price distribution, our basic strategy is to impute an indicative average price from revenue and sales volume statistics. Given this goal, we focus on the price distribution of product groups (PGC) classifications as they are reasonably similar goods that are close substitutes. Hence, we compute the annual price of a product (UPC) as the total sales divided by total sales volume of all transactions under this UPC in a given location and year. The prices of all UPCs within a product group is then used to construct the price distribution of a product group (PGC). Follwing this procedure, we essentially obtain a panel of price distribution for all 119 PGCs in the contiguous states from 2006 to 2017.

### 2.2.1 Measures

In this section, we explain in details how we construct the market share measures for low-, middle-, and high-price products. The market share of products in a particular price range, $Share_{st,g,p}$ is a PGC-location-year level variable. This variable is constructed using the sum of a market share of a barcode-level product (UPC) $u$ whose price is within a certain price range characterized by a threshold $p$ in product group $g$, location $s$, and year $t$,

$$Share_{st,g,p} = \sum_{u \in \Omega(p)} (Share_{st,u})$$

where $Share_{st,u} = Sales_{st,u}/Sales_{st,g}$ is the market share of product $u$ within a product group (PGC) $g$ and $\Omega(p)$ is the set of products whose price is within the price range $Price_{s,g,p}$ characterized by threshold $p$. As mentioned above, the price of each product is imputed using sales revenue divided by sales volume (physical quantity). Take the high-price products as an example. We could define the price threshold as any of the products whose price is above the 75th percentile of the distribution. Then $Share_{st,g,75}$ would be the sum of market shares of all products in product group $g$, location $s$, and year $t$ whose price is above the 75th percentile of the price distribution based on the 2006 data.

In our baseline analysis, we define the low-price products as those products whose deflated real price is below the 25th percentile of the price distribution (of the corresponding product group $g$ in location $s$) in 2006.[17] Correspondingly, we define the high-price products as whose deflated real price is above the 75th percentile of the price distribution. The middle-price products are then defined as those between the 25th and 75th percentiles. In robustness checks, we will use alternative thresholds such as 90th/10th and 66th/33th percentiles to define the high/low-price range. Given the above definitions, we would obtain a panel of market share measures for products in each price range. Specifically, we would have a panel of variables which measure the market shares of low-, middle-, and

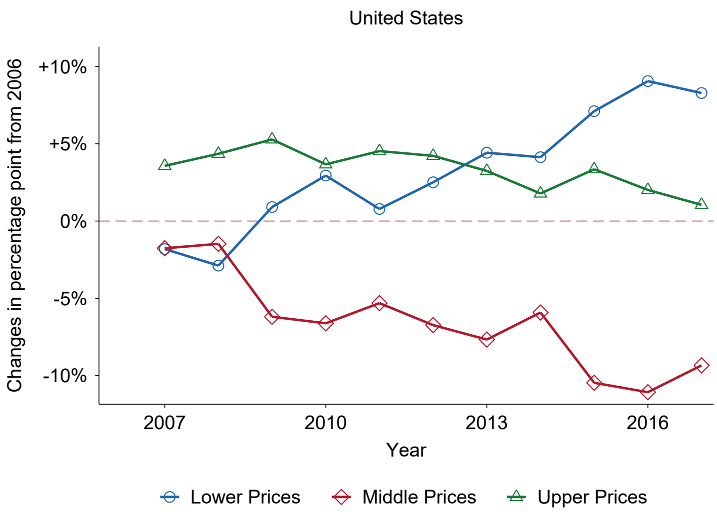
[17] We deflate the nominal price to 1999 real dollars by using CPI data from the IPUMS database.

high-price products in each location and year. Note that by definitions, the market shares should be 25%, 50%, and 25% for low-, mid-, and high-price products for any product group-location pair in the beginning year (i.e., 2006).

### 2.2.2 National Time Trend

We now document that there is indeed a missing middle in the price distribution over the years by examining the national time trend of the corresponding market share measures. We first construct an aggregate market share measure by computing the average market shares for all product groups weighted by PGC-specific market sales. This gives us three time-series of aggregate market share measures for low-, mid-, and high-price products at the national level. The numbers in each series would then represent the national weighted average of the percentage of UPC products in each PGC that belongs to the low/mid/high price range. Next, to highlight the changes, we subtract each series by the corresponding value in 2006 and the results are plotted in Figure 5.

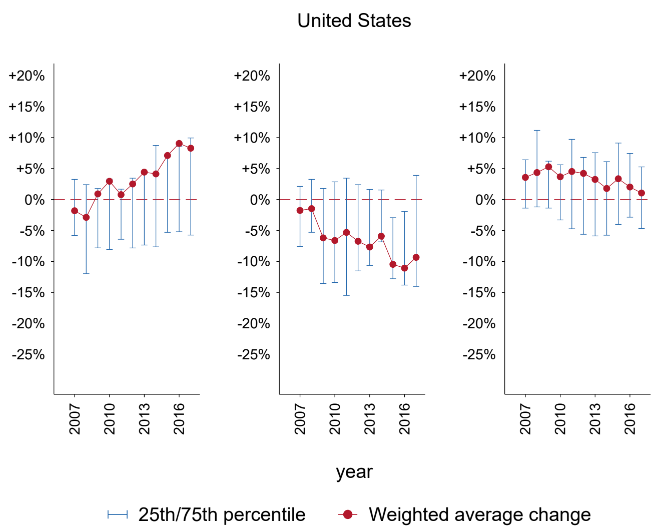
Figure 5: Cumulative Changes in Market Shares from 2006




*Notes:* Each plotted point represents the cumulative change of market share in percentage points of a particular price range in the United States from 2006 to that year. Statistics are computed using the Nielsen Retail Scanner Dataset.

Figure 5 clearly documents that the middle of the price distribution is losing its support from 2006 to 2017, and that the decline could be as large as 10 percetange points. The disappearing market shares for the middle-price products are absorbed by the rise in both low- and high-price products. The high-price products see a rise in the market share initially and declines over-time, though the cumulative change still remains positive in comparison with that of 2006. In contrast, the market share of low-price products has been increasing steadily over-time and hence absorbs the major bulk of the decline in middle-price market shares. To confirm that this hypothesis is also a universal phenomenon across the product groups, we further plotted the 75th and 25th percentile of the change in market shares for each PGC together with the weighted average in Figure 6. Although some product groups exhibits a reversed pattern, the change of market shares for the majority of the product groups is consistent with the missing middle phenomenon.

Figure 6: Dispersion of Cumulative Changes in Market Shares




*Notes:* Each plotted point represents the cumulative change of market share in percentage points of a particular price range in the United States from 2006 to that year. The market share measures are computed using 75th/25th percentile of the price distribution as thresholds

### 2.2.3 State-Level Heterogeneity

Is the missing middle phenomenon an universal phenomenon across the United States? To approach this question, we further implement the similar exercises for the states in the contiguous US. The results are plotted in Figure 7.

Figure 7: Cumulative Changes in Market Shares from 2006 Across Contiguous States


*Notes:* Each plotted point represents the cumulative change of market share in percentage points of a particular price range in the contiguous states of US from 2006 to that year. The market share measures are computed using 75th/25th percentile of the price distribution as thresholds

We find that the missing middle phenomenon is consistent across all contiguous states, in the sense that all states experience a decline in the middle share. The heterogeneity is mainly with the magnitude of the decline as well as how the shrinking middle is absorbed by the rising market shares of other price groups. In most states, the decline middle is mainly absorbed by the rise in low-price product shares over time. This is a novel and interesting finding, as most studies such as Jaravel (2019) emphasize on the effect of an increasing number of firms entering the higher-price market. We complement their findings in that, although we find the market share for higher-price products have risen over-time, the effect of rising market share in the lower-price products maybe more important in explaining the declining middle. This has profound and novel implications in thinking about the welfare effects of a more dispersed price distribution. Moreover, we find that densely populated states such as Conneticut, Delaware, Maryland, New Jersey, New York, Pennsylvania, Texas, and Washington experience a more pronounced decline in the middle price support while for less populated states such as Alabama, Nevada, and Oregon the phenonmeon is less obvious. It would be interesting to ask if this phenomenon correlates with urban density at a more disaggregated level. We examine this hypothesis in the next section.

### 2.2.4 Commuting Zones

We now introduce urban density as a possible influencing factor on product price distribution. Examining urban density would be meaningless if we collect the data at state-level. Thus, it is imperative to construct the measures at a more disaggregated level.

We choose the commuting zones as the best approximation for a disaggregated local area where people work and live in the same place. This definition is the canonical measure used in various literature such as labor economics(Autor and Dorn, 2013). We first obtain the population size of each county in 2005 as well as the land areas from NEBR and Census Bureau. We then concord the population data, land area data, and price data in Nielsen Retail Scanner to commuting zone level from county level using a concordance provided by David Dorn in his website. Next, we compute the population density (no. of people per square mile) of each commuting zone and rank the commuting zones according to its population density. We then divided the 722 contiguous commuting zones into 10 deciles with each decile represent approximately 10% of the US population. The representative areas of each decile of commuting zones is listed in Table 7, where the 10th decile represents the densest metropolitan areas. Finally, we implement similar exercises as in the previous section to compute the time-trend of market shares for each price group in each commuting zone decile. The results are reported in Figure 27.

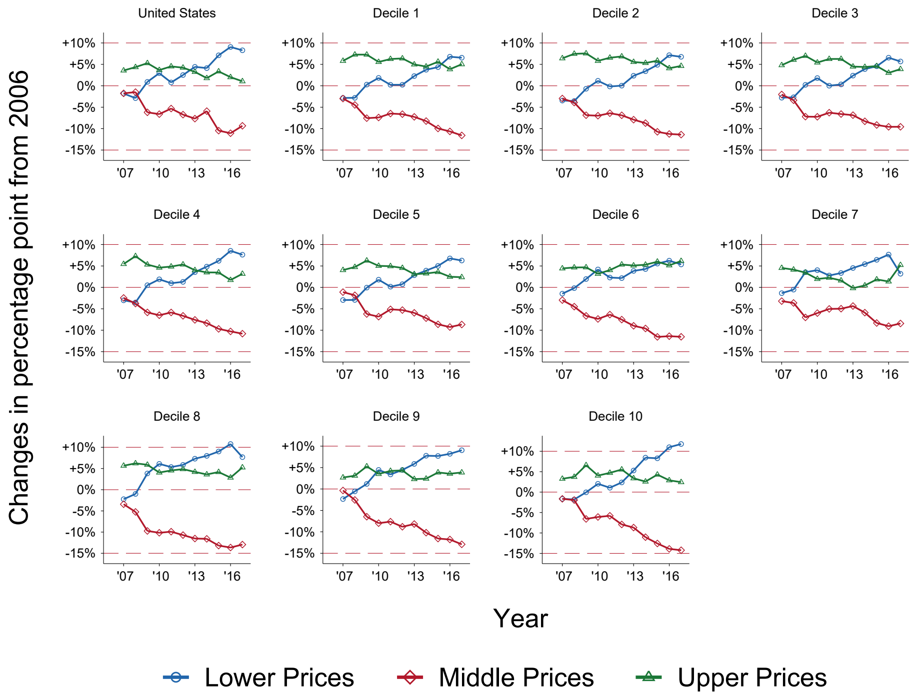
Table 7: Representative Areas for Each Decile of Commuting Zones

| Decile | Representative Areas |
|---|---|
| 1 | Freeborn-Mower, Nacogdoches-Lufton, Watertown |
| 2 | Gainesville, Decatur-Champaign-Urbana, Wheeling |
| 3 | Little Rock-North Little Rock, Rockford, Binghamton |
| 4 | Benton Harbor-South Bend, Nantucket County, Steubenville-Weirton |
| 5 | Memphis, Nashville, Youngstown-Warren-Sharon |
| 6 | Lancaster-Reading-Harrisburg-Lebanon-Carlisle, Columbus, Springfield |
| 7 | Minneapolis-St. Paul, Hamilton-Middletown-Cincinnati, Orlando |
| 8 | Atlanta, Wilmington-Newark, Detroit-Flint |
| 9 | San Francisco-Oakland, Philadelphia, Monmouth-Ocean |
| 10 | New York-Nassau-Suffolk, Chicago, Newark-Trenton |

*Notes:* The commuting zones are ranked by the population density. Population and land area data are from NEBR and the US Census Bureau. Each decile of commuting zones represents approximately 10% of the total US population

Several observations are in order with regard to the results reported in Figure 27. First, the results confirm that the missing middle in price distribution is consistent and uniform from rural to urban areas, which further speaks to the robustness and novelty of the phenomenon. Second, the results confirm our hypothesis that the missing middle phenomenon is more pronounced in densely populated metropolitan areas. In particular, the densest populated urban areas, such as those commuting zones in deciles 8, 9 and 10, experience a larger decline in the market share for middle-price products. The difference in the decline of middle shares between these densely populated areas with the rest of commuting zones could be as large as 5 percentage points. Our findings are novel and significant as previous works only examine the price structure across income groups without investigating the spatial dimension. Our paper fills this gap.

Figure 8: Cumulative Change of Market Shares in Commuting Zones




*Notes:* Each plotted point represents the cumulative change of market share in percentage points of a particular price range in the contiguous commuting zones of US from 2006 to that year. The market share measures are computed using 75th/25th percentile of the price distribution as thresholds

### 2.2.5 Robustness

The results above are reported using 75/25th percentile of the price distribution as thresholds. In this section, we further show that our results are robust to other specifications. In particular, we also computed the time-trend of at national, state, and commuting zone levels using other thresholds such as the 90th/10th percentile and the 66th/33rd percentile as thresholds. The results are reported in Appendix B.1 and are largely consistent with our baseline findings. In sum, we find that the missing middle in the price distribution, which is defined as the distribution loses its mass in the middle price support, is a pronounced, urban, and robust phenomenon in the United States for the past decade, while both the upper and lower price supports are gaining mass.

## 2.3 Missing Middle in Income Distribution

### 2.3.1 Measures

We use two inequality measures in this paper. The first inequality measure is based on a fixed set of income cutoffs in year 2005. For this measure, we compute the income cutoffs that account for the first 25% of population, the 25% to 75% of population, and the rest of 25% population with the highest income.[18] We then compute the changes in population share based on these fixed cutoffs in subsequent years. If there is a shrinking middle, then we expect population with income above the 25% and below the 75% cutoffs will have a smaller share over the years.[19]

One caveat of defining income inequality with this approach, is that the cutoffs are not time-varying. This could be biased especially if one accounts for real economic growth. In the extreme case, the population share of the middle income category could be zero, when every single person has earned a higher income than the 75th percentile in 2005 due to economic growth. Thus, we only consider this measure as a robustness check for our main results and report them in the appendix. For our baseline results, we focus on the following set of measures.

The baseline inequality measure is built on the definition of Pew Research Center. For this measure, we define the middle income class as those with income more than two-third of the median national income and no greater than twice of median.[20] This definition is time-varying and measures the relative position of an individual in the annual income distribution.

We compute the measures as defined above using the American Community Survey (ACS) data reported in the IPUMS database. To construct the population shares of income groups, first we need a plausible measure for household income as the size of household

[18] We also check for robustness for 90th/10th percentile and 66th/33rd percentile.

[19] Household income in the following years is deflated by CPI.

[20] Pew Research Center also previously used another definition, which is between 50% and 150% of median income. Another definition is by Alan Krueger when he was working for the White House. For this definition, Krueger defined middle class as those with income that is within 50% of the national median.

varies. We use two definitions of household income. The first definition is the conventional per-capita household income, in which the gross income of the household is divided by the number of persons dwelling in the same unit,

$$\text{Per-capita income} = \frac{\text{Gross household income}}{\text{Number of persons}}.$$

However, this definition does not take account into economies of scale, in the sense that the purchasing power of a family is higher with a larger household, because many items are more economical when purchased in bulks. One example cited by the Pew center is apartment renting. An one-bedroom apartment is usually more costly to rent than a two-bedroom apartment in terms of cost per bedroom due to the sharing of public space such as living room and pantry. Another example that we have in mind is the consumption of perishable fruits. A single bachelor may find it less economical to consume fruits, as he has to purchase it in smaller quantities to consume in time. In contrast, a larger family could enjoy the discount when purchasing in bulks because they could consume the fruits in a shorter period of time. To adjust for this large-household bias, we employ the following definition of adjusted per-capita income in the literature,

$$\text{Adjusted per-capita income} = \frac{\text{Gross household income}}{\text{Number of persons}^{N}}, \quad 0 \leqslant N \leqslant 1.$$

where $N$ is inversely proportional to the strength of within-household "economies of scale". If $N = 1$, then the adjusted income is equivalent to the per-capita definition, and within-household "economies of scale" is entirely absent. If $N$ is set to $0$, then the adjusted income is equivalent to the gross household income, and it is not discounted by family size. We follow the convention and set $N = 0.5$.

Table 8 reports the summary statistics for the ACS sample from 2005 to 2017. The percentile measures are computed using number of persons as weights. Observations with missing household income or negative income are dropped. Observations from Alaska and Hawaii are dropped. In addition, observations with missing census tract identifiers are dropped. Group quarters include only 1970 or 1990 household definitions, such that institutions and boarding schools are also dropped from the sample.

Table 8: Summary Statistics for ACS Sample

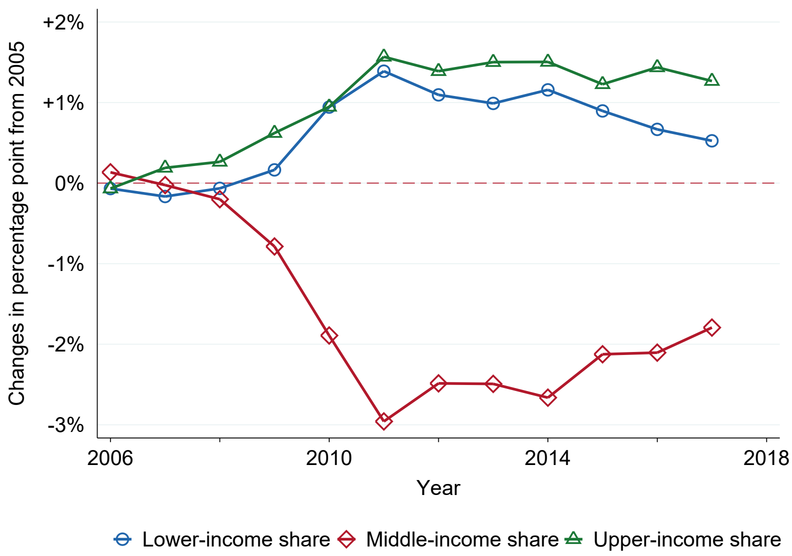
| | |
|---|---|
| The number of households (2005–2014) | 11,725,725 |
| The number of households (2015–2017) | 3,633,321 |
| The number of persons in each household (2005-2014) | 2.47 |
| The number of persons in each household (2015-2017) | 2.45 |
| 25% household income per capita in 2005 | 8,956.50 |
| 50% household income per capita in 2005 | 15,968.16 |
| 75% household income per capita in 2005 | 26,805.53 |
| 50% adjusted household income in 2005 | 27,578.87 |

### 2.3.2 Time Trend

We now report the time-series of the constructed population share of different income groups. Formally, the population share of lower-income group is defined as the share of population below 2/3 of the median income level, while that of upper income group is the share of population with more than twice the median income. We first construct this measure in a national level, as reported in Figure 9. It is evident that the middle-income population share has been declining since the Great Recession and remains low after. The decline in middle-income share could be as large as 3 percentage points on a national scale. In the meantime, both the lower- and upper-income population share have been rising though upper-income share rises slightly more than that of lower-income. The lagged-timing roughly coincides with that of the decline of middle-price product market shares. We will delay the discussion of possible correlation/causality until the next section.

Next, we examine if this trend is uniform across different contiguous states. Following a similar exercise, we construct the share of populations belonging to different income groups at the state-level, again using the national median income to define income category. We plot the time-series in each contiguous state in Figure 10. The results are rather mixed. While the majority states has seen a decline in the middle-income population share, some states experience a different pattern in contrast to the the national statistics. Second, while the declining middle share pattern for the price distribution is more uniform across the group, this is not so for the income distribution. In some areas such as District

Figure 9: Cumulative Changes of Population Shares in the United States




*Notes:* Each plotted point represents the cumulative change of population share in percentage points of a particular income group in the United States from 2006 to that year. Data are from IPUMS ACS.

of Columbia and Pennsylvania, both the middle- and lower-income share are declining. The normative message in these states seems to be that all individuals of the society are doing better, such that more and more people are lifted to upper-income class while less people are stuck at the lower-income class. Finally, it is again ex-ante interesting to observe that the pattern seems to be more pronounced in densely populated states such as Connecticut, Massachusetts, Maryland, and New Jersey. It would be interesting again to examine if urban density plays a role in determining the time pattern of population shares across income groups.

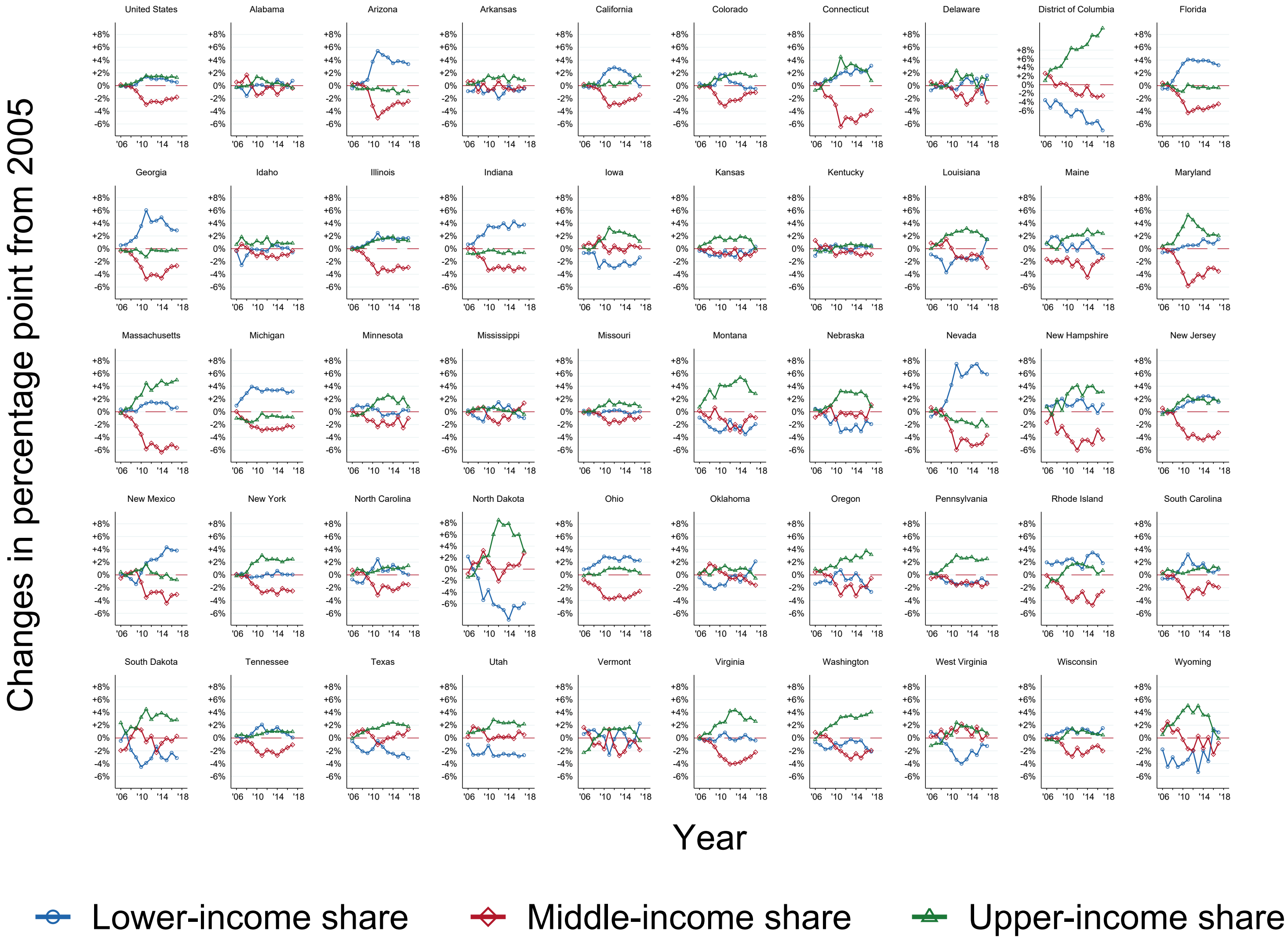
### 2.3.3 Commuting Zones

Autor (2019) finds that the population share of middle-skilled jobs has been decreasing for the past 40 years, and this is particularly evident for urban area. Given that income is positively correlated with skills, naturally we want to test if our hypothesis is also more evident in dense metropolitan areas.

We will first convert the ACS sample from Public Use Microdata Areas (PUMA) to commuting zones (CZ), which is a widely used proxy for metropolitan areas, following the approach in Autor and Dorn (2013). The problem at hand is that PUMA definition sometimes overlaps with multiple commuting zones. Autor and Dorn (2013) solves this issue by computing the probability such that an individual from a PUMA lives in a particular CZ. Given these probabilities, they then "split" the individuals into 722 parts in which each part corresponds to a different CZ in the contiguous United States. The probability attached to each of the 722 CZ would then sum to 1.

Following their approach, we first compute a per-capita household income measure based on the number of individuals reported in the ACS sample. We then split individuals into 722 pieces and concord the PUMA to CZ using a crosswalk provided on David Dorn's website. We utilizes the probabilities in the crosswalk and split each family into different CZs with reduced family size and the same per-capita income. The split sample is then combined with population density measure, which we computed based on county population in 2005 provided by the NBER, county land areas provided by the Census Bureau, and a crosswalk from county to Commuting Zones provided by David Dorn.

Figure 10: Cumulative Changes in Population Shares from 2006 Across Contiguous States




*Notes:* Each plotted point represents the cumulative change of population share in percentage points of a particular income group in the contiguous states of US from 2006 to that year. Data are from IPUMS ACS.

With a concorded ACS dataset, we are now ready to compute the relevant statistics for commuting zones with different population density. We follow the previous exercise again in dividing the sample into ten different bins ranked by urban density, with each bin (decile) accounts for approximately 10% of the total population. Then, the ACS sample is summarized in terms of household income and population shares of income groups in different deciles. The results are reported in Table 9.

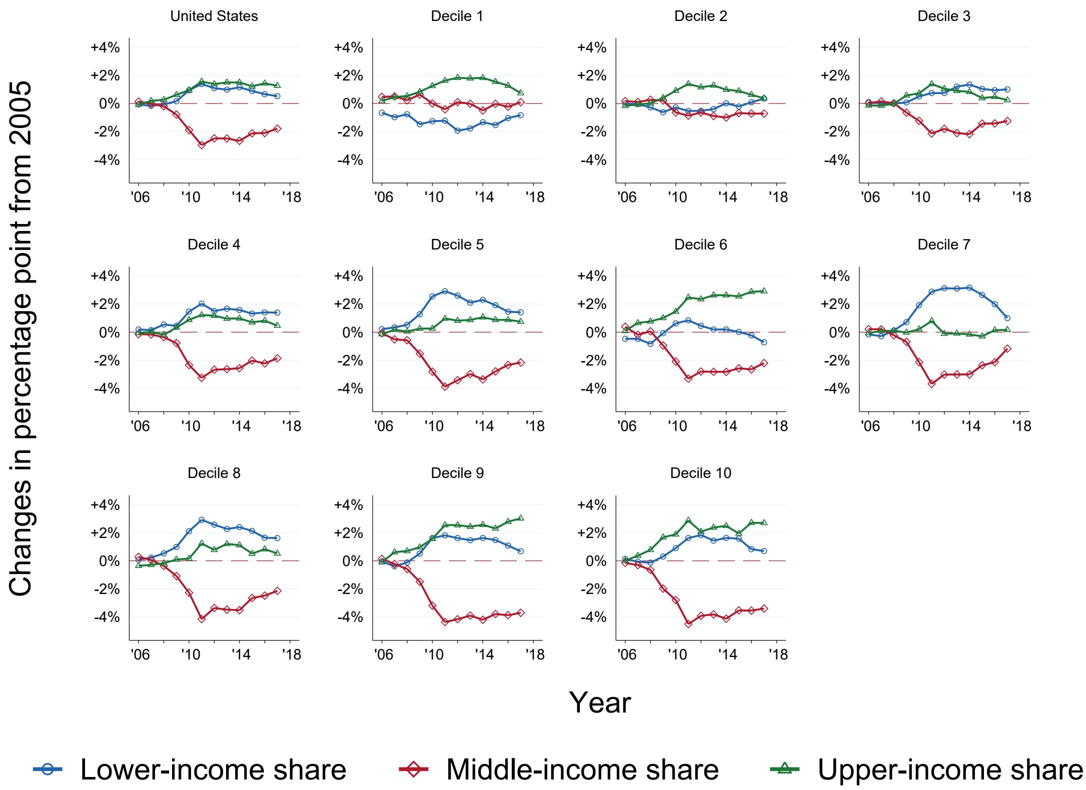
Table 9: Summary Statistics for Deciles of Commuting Zones in 2005

| Decile | Adj. income | Pct. income | % Lower | % Middle | % Upper | Density |
|---|---|---|---|---|---|---|
| 1 | 21,412 | 14,501 | 39.9 | 52.1 | 8.0 | 18 |
| 2 | 21,895 | 14,757 | 38.8 | 52.1 | 9.1 | 70 |
| 3 | 23,884 | 15,823 | 34.9 | 54.2 | 10.9 | 115 |
| 4 | 24,850 | 16,642 | 32.8 | 55.2 | 11.9 | 181 |
| 5 | 27,143 | 18,212 | 29.2 | 55.1 | 15.7 | 269 |
| 6 | 28,107 | 18,766 | 27.6 | 55.7 | 16.7 | 349 |
| 7 | 29,434 | 19,193 | 29.0 | 52.7 | 18.2 | 464 |
| 8 | 31,220 | 20,557 | 26.5 | 52.7 | 20.8 | 710 |
| 9 | 33,085 | 21,752 | 23.5 | 52.4 | 24.1 | 1,004 |
| 10 | 34,277 | 22,249 | 23.1 | 51.2 | 25.7 | 2,302 |

Table 9 shows that the income statistics are significantly different across commuting zones with different deciles of population density. First, the household income is in general higher in more densely populated metropolitan areas. This is not surprising considering the vast research on how urban life facilitates economies of scale and brings more economic opportunity. Second, in terms of income distribution, urban life is also associated with more upper-income population share and less lower-/middle-income shares. In particular, the percentage of lower-/middle-income class, as defined as those with less than twice of the national median income, is much smaller at the most densely populated commuting zones. Thus, as a level measure, urban density does not correlate with a hollowing middle in the income distribution.

Next, we examine if the same message can be found in terms of time trend. We plotted the time series of population shares of different income groups in Figure 11. First, we find that the declining population share of the middle-income group is more severe in more densely populated areas. Second, the shrinking middle over-time is associated with the rise in *both* lower- and upper-income shares, and it is more pronounced in the densely populated areas. This is in contrast with the summary statistics, in the sense that in terms of cross-sectional comparison, the metropolitan areas have more upper-income shares with less lower-/middle-income shares. In contrast, in terms of time-trend, the more densely populated areas are experiencing a more unequal income distribution over-time, where both the lower- and upper-income population shares are rising. Our findings also echo that of Autor (2019) who finds that the middle-skill employment share is declining in the more densely populated commuting zones over the past decades.

Figure 11: Cumulative Changes in Population Shares from 2006 Across Contiguous States




*Notes:* Each plotted point represents the cumulative change of population share in percentage points of a particular income group in the commuting zones of US from 2006 to that year. Data are from IPUMS ACS.

### 2.3.4 Robustness

The results above are reported using adjusted per-capita income. In this section, we further show that our results are robust to other income measures. In particular, we also computed the time-trend of at national, state, and commuting zone levels using the conventional per-capita household income to compute population share of household. Second, we computed another set of income groups using similar definition to that of price distribution. In addition, we divided the urban density into 20 bins and plotted a population share-urban density graph which is similar to that in Autor (2019). The results are reported in Appendix B.2 and are largely consistent with our baseline findings. In sum, we find that the shrinking middle in the income distribution is also a pronounced, urban, and robust phenomenon in the United States for the past decade. This is not surprising given all the reports on widening income inequality in the media. We will now proceed to link the missing middle in the price distribution with the widening income inequality in the United States.

## 2.4 Empirical Evidence for Correlations

So far we have documented a shrinking middle in both the price and income distribution, in the sense that the market shares for middle price products and the population share for middle income group are declining from 2006 to 2017. We now attempt to bridge a relationship between the two phenomena by arguing that the changes in product price dispersion are driven by the changes in income distribution. The main idea is to regress the market shares of different price groups on the population shares of different income groups, with a battery of fixed effects except for those that would absorb time variation which is main identifying variation that we exploit. Thus, our baseline specification is as follows,

$$Share_{st,g,p} = \alpha_1 Frac_{st,p} + \gamma_g + \delta_s + e_{st,g,p}$$

where $Share_{st,g,p}$ is the market share of price group $p$, of PGC $g$, at location $s$ in time $t$ and $Frac_{st,p}$ is the population share of income group $p$ at location $s$ in time $t$. $\gamma_g$ and $\delta_s$ are product and location fixed effects, while $e_{st,g,p}$ is an error term. In particular, $p$ could be any of the groups in $\{Low, Middle, High\}$. For the price groups, we construct the low-, middle, and high-price groups using the 75th/25th percentile as thresholds.

### 2.4.1 State-Level Regressions

For the baseline regression, we first investigate the relationship at the state level. The results are reported in Table 10, 11, and 12. The regressions largely confirm our hypothesis that the demand-side demographics have significant influence on the price distribution, as estimates for $\alpha_1$ are both economically and statistically significant for most regressions and specifications.

Table 10: Market Shares of Different Price Range: State-Level

| Dep. Var. | Below33 | Middle | Above66 | Below33 | Middle | Above66 |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Low Share | 0.115** | | | 0.114*** | | |
| | (0.049) | | | (0.040) | | |
| Mid Share | | 0.357*** | | | 0.350*** | |
| | | (0.042) | | | (0.038) | |
| Upp Share | | | 0.750*** | | | 0.716*** |
| | | | (0.057) | | | (0.044) |
| Robust SE | Yes | Yes | Yes | Yes | Yes | Yes |
| State FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Product FE | No | No | No | Yes | Yes | Yes |
| $R^2$ | 0.015 | 0.005 | 0.014 | 0.441 | 0.299 | 0.515 |
| $N$ | 57,346 | 57,346 | 57,346 | 57,346 | 57,346 | 57,346 |

*Notes:* Market shares measures are constructed from Nielsen Retail Scanner Data. Population share measures are constructed from IPUMS ACS Data. States include all the contiguous states in the United States including District of Columbia. Time is from 2006 to 2017. All regressions include a constant term. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

First, for the state-level regressions reported in Table 10 using 66th/33rd percentile as thresholds to define price groups, all the estimates are consistent with the intuition. In particular, our regressions suggest that on average, when the middle income population share decreases by 10 percentage point, the market share of middle price products across product groups will decrease by 3.5 percentage points, after taking into account of state and product-specific trend. In addition, the estimates for the effects of upper-income group, is more economically significant than that of middle-income and much more than that of lower-income population shares. This echoes the findings in Jaravel (2019) and Argente and Lee (2020) in that there is a more significant change in inflation inequality for the high-income group. However, our findings also add additional insights to their works as we show that the effect of the lower-income and middle-income population shares are also economically significant.

Table 11: Market Shares of Different Price Range: State-Level

| Dep. Var. | Below25 | Middle | Above75 | Below25 | Middle | Above75 |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Low Share | 0.055 | | | 0.056 | | |
| | (0.045) | | | (0.037) | | |
| Mid Share | | 0.427*** | | | 0.419*** | |
| | | (0.044) | | | (0.039) | |
| Upp Share | | | 0.733*** | | | 0.702*** |
| | | | (0.053) | | | (0.041) |
| Robust SE | Yes | Yes | Yes | Yes | Yes | Yes |
| State FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Product FE | No | No | No | Yes | Yes | Yes |
| $R^2$ | 0.014 | 0.005 | 0.015 | 0.430 | 0.350 | 0.533 |
| $N$ | 57,346 | 57,346 | 57,346 | 57,346 | 57,346 | 57,346 |

*Notes:* Market shares measures are constructed from Nielsen Retail Scanner Data. Population share measures are constructed from IPUMS ACS Data. States include all the contiguous states in the United States including District of Columbia. Time is from 2006 to 2017. All regressions include a constant term. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Next, for the state-level regressions reported in Table 11 using 66th/33rd percentile as thresholds, most the estimates are statistically significant except for the effect of lower-income population shares. Our interpretation is that as we widen the price range for the middle-price products, the significance for the lower-income shares disappears. This suggests that the statistical significance for the lower-income group reported in the previous table came from the group of products whose price are closer to the 33rd percentile threshold. In particular, whenever there is an increase in the relative poor population, i.e., those in the lower-income class, many firms who were initially in the middle-price range, will shift to selling lower-price products. However, the fact that the significance disappears in this table suggests that these firms do not go to the extreme. Rather, their response mostly falls in the range between 25th and 33rd price percentiles. This is also for case of the regressions reported in Table 12 using 90th/10th percentile as thresholds. The results reported in this table are largely similar to that of Table 11, except that the economic impacts are much smaller. These results further confirm our prior argument that the identifying variation are mainly from the percentile closer to the middle rather than in the extreme ends of the price support.

Table 12: Market Shares of Different Price Range: State-Level

| Dep. Var. | Below10 | Middle | Above90 | Below10 | Middle | Above90 |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Low Share | 0.037 | | | 0.039 | | |
| | (0.033) | | | (0.029) | | |
| Mid Share | | 0.341*** | | | 0.339*** | |
| | | (0.037) | | | (0.032) | |
| Upp Share | | | 0.525*** | | | 0.510*** |
| | | | (0.037) | | | (0.028) |
| Robust SE | Yes | Yes | Yes | Yes | Yes | Yes |
| State FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Product FE | No | No | No | Yes | Yes | Yes |
| $R^2$ | 0.008 | 0.005 | 0.015 | 0.366 | 0.375 | 0.537 |
| $N$ | 57,346 | 57,346 | 57,346 | 57,346 | 57,346 | 57,346 |

*Notes:* Market shares measures are constructed from Nielsen Retail Scanner Data. Population share measures are constructed from IPUMS ACS Data. States include all the contiguous states in the United States including District of Columbia. Time is from 2006 to 2017. All regressions include a constant term. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

### 2.4.2 Commuting Zones

We further test our hypothesis that the changes in price distribution is driven by that of income distribution in the context of commuting zones. For the baseline regression, we first investigate the relationship using an identical specification to that of state level. In the next section, we will introduce agglomeration or urban density to investigate the interactive relationship. The results for the baseline are reported in Table 13, 14, and 15.

Table 13: Market Shares of Different Price Range: Commuting Zones

| Dep. Var. | Below33 | Middle | Above66 | Below33 | Middle | Above66 |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Low Share | 0.142*** | | | 0.131*** | | |
| | (0.010) | | | (0.009) | | |
| Mid Share | | 0.075*** | | | 0.077*** | |
| | | (0.009) | | | (0.009) | |
| Upp Share | | | 0.486*** | | | 0.472*** |
| | | | (0.016) | | | (0.013) |
| Robust SE | Yes | Yes | Yes | Yes | Yes | Yes |
| CZone FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Product FE | No | No | No | Yes | Yes | Yes |
| $R^2$ | 0.261 | 0.088 | 0.115 | 0.502 | 0.279 | 0.464 |
| $N$ | 653,068 | 653,068 | 653,068 | 653,068 | 653,068 | 653,068 |

*Notes:* Market shares measures are constructed from Nielsen Retail Scanner Data. Population share measures are constructed from IPUMS ACS Data. We include all the contiguous commuting zones in the United States whenever data is available. Time is from 2006 to 2017. All regressions include a constant term. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

The results at the commuting zone level are largely consistent to that of state-level regressions. All estimates for $\alpha_1$ are economically and statistically significant. In particular, they are significant for the effects of lower-income population shares, unlike those in state-level regressions. This is because as we redefine the price thresholds in a local level, there will be more dispersion and hence more identifying variation in the extreme end of distributions. As a result, the income distribution tends to have a significant impact on the price distribution across its income support. However, the interpretation that we have for state-level regressions still holds. Specifically, the economic estimates for $\alpha_1$ attenuates when we widen the middle price range. This is consistent with our prior interpretations that the change in pricing are not too extreme.

Table 14: Market Shares of Different Price Range: Commuting Zones

| Dep. Var. | Below25 | Middle | Above75 | Below25 | Middle | Above75 |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Low Share | 0.132*** | | | 0.122*** | | |
| | (0.010) | | | (0.008) | | |
| Mid Share | | 0.111*** | | | 0.116*** | |
| | | (0.011) | | | (0.009) | |
| Upp Share | | | 0.473*** | | | 0.461*** |
| | | | (0.015) | | | (0.012) |
| Robust SE | Yes | Yes | Yes | Yes | Yes | Yes |
| CZone FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Product FE | No | No | No | Yes | Yes | Yes |
| $R^2$ | 0.321 | 0.150 | 0.091 | 0.528 | 0.373 | 0.467 |
| $N$ | 653,068 | 653,068 | 653,068 | 653,068 | 653,068 | 653,068 |

*Notes:* Market shares measures are constructed from Nielsen Retail Scanner Data. Population share measures are constructed from IPUMS ACS Data. We include all the contiguous commuting zones in the United States whenever data is available. Time is from 2006 to 2017. All regressions include a constant term. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 15: Market Shares of Different Price Range: Commuting Zones

| Dep. Var. | Below10 | Middle | Above90 | Below10 | Middle | Above90 |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Low Share | 0.079*** | | | 0.073*** | | |
| | (0.008) | | | (0.008) | | |
| Mid Share | | 0.095*** | | | 0.103*** | |
| | | (0.010) | | | (0.009) | |
| Upp Share | | | 0.328*** | | | 0.325*** |
| | | | (0.012) | | | (0.009) |
| Robust SE | Yes | Yes | Yes | Yes | Yes | Yes |
| CZone FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Product FE | No | No | No | Yes | Yes | Yes |
| $R^2$ | 0.486 | 0.326 | 0.062 | 0.598 | 0.543 | 0.506 |
| $N$ | 653,068 | 653,068 | 653,068 | 653,068 | 653,068 | 653,068 |

*Notes:* Market shares measures are constructed from Nielsen Retail Scanner Data. Population share measures are constructed from IPUMS ACS Data. We include all the contiguous commuting zones in the United States whenever data is available. Time is from 2006 to 2017. All regressions include a constant term. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

### 2.4.3 Agglomeration

We now examine if the income distribution has a larger impact on the missing middle in price distribution in the urban areas. The baseline specification is amended to incorporate an interactive effect between population density, which is a location-specific variable defined as total population in 2005 divided by total land areas, and population shares of income groups. Specifically, the specification is as follows,

$$Share_{st,g,p} = \alpha_1 Frac_{st,p} + \alpha_2 Frac_{st,p} \times log(Density_s) + \gamma_g + \delta_s + e_{st,g,p}$$

where $log(Density_s)$ is the log of population density in location $s$ and $alpha_2$ is the estimator for the interactive effect between population density and population share of income groups. Notice that we do not include a term controlling for the level of log population density, because it is a commuting zone-specific variable and has already been subsumed by the commuting zone dummies. If our hypothesis is true that the effect of demand-side demographics is larger in more densely populated metropolitan areas, then the estimates of $\alpha_2$ should be positive and significant. Otherwise, it may not be true that urban density plays a role or it could be that the effect is actually smaller in urban areas.

Table 16: Market Shares of Different Price Range: Commuting Zones

| Dep. Var. | Below33 | Middle | Above66 | Below33 | Middle | Above66 |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Low Share | 0.335*** | | | 0.316*** | | |
| | (0.027) | | | (0.0245) | | |
| Low×log(Density) | -0.057*** | | | -0.055*** | | |
| | (0.007) | | | (0.006) | | |
| Mid Share | | -0.089*** | | | -0.084*** | |
| | | (0.025) | | | (0.024) | |
| Mid×log(Density) | | 0.045*** | | | 0.045** | |
| | | (0.006) | | | (0.006) | |
| Upp Share | | | 0.380*** | | | 0.382*** |
| | | | (0.039) | | | (0.034) |
| Upp×log(Density) | | | 0.031*** | | | 0.027** |
| | | | (0.010) | | | (0.009) |
| Robust SE | Yes | Yes | Yes | Yes | Yes | Yes |
| CZone FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Product FE | No | No | No | Yes | Yes | Yes |
| $R^2$ | 0.261 | 0.088 | 0.115 | 0.502 | 0.279 | 0.464 |
| $N$ | 653,068 | 653,068 | 653,068 | 653,068 | 653,068 | 653,068 |

*Notes:* Market shares measures are constructed from Nielsen Retail Scanner Data. Population share measures are constructed from IPUMS ACS Data. We include all the contiguous commuting zones in the United States whenever data is available. Population density is computed based on population and land area data from NBER and US Census Bureau. Time is from 2006 to 2017. All regressions include a constant term. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

The results for the interactive regressions are reported in Table 16 to 18. First, the estimates of $\alpha_2$ are positive and significant (economically and statistically) for the middle income population shares. This implies that the declining population shares of middle/upper income class has a larger impact on the middle-/upper- price market shares in the most densely populated metropolitan areas. This finding echoes with the recent work in the agglomeration literature where the agglomeration benefits of larger cities are known to bias towards high-skill workers and also low-skill workers through complementarity. It is therefore natural to induce that the middle-skill workers or the middle-income workers are disappearing in the larger cities, because technological advance and globalization has shifted the US towards a knowledge economy. Our findings add novel insights to the literature, in the sense that we are the first to examine the spatial (agglomeration) dimension on how demand-side demographics influence price inequality.

Second, the interactive effect is not consistent across all income group population shares. In particular, we find that this effect is reversed for the population share of lower-income group. Although this results seems to contradict with our hypothesis, the interpretation is more subtle. The negative interaction term for the lower-income population share implies that the increase in lower-income population share has a smaller (but still positive and significant) impact on the low-price market shares. This is rather consistent with the predictions from the agglomeration literature. As what we have argued earlier, knowledge spillover and face-to-face interactions benefit high-skill workers more, especially for the case of the US economy.[21] Hence, we should have expected that rising income inequality is biased towards high-income group in densely populated areas and hence its impact on the price distribution.

The final remark concerns the interaction term when we widen the band of middle-price support. Similar to that of the baseline regressions, the interaction term loses its significance for the upper income population share when we run the regressions with 90th/10th percentiles. This is not surprising if we invoke the previous argument, in that the identifying variation mostly lies closer to the middle price support. Hence, the fact

[21] An alternative testable prediction is that, if the agglomeration benefit is mainly about labor pooling or input sharing as is the case for many developing countries, we should expect this effect to be reversed for income distribution and hence for the price distribution. The part on income distribution is the focus of one of my ongoing works.

that the interactive estimate loses its significance for a widening middle definition in Table 18 is consistent with this explanation.

Table 17: Market Shares of Different Price Range: Commuting Zones

| Dep. Var. | Below25 | Middle | Above75 | Below25 | Middle | Above75 |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Low Share | 0.291*** | | | 0.273*** | | |
| | (0.025) | | | (0.0235) | | |
| Low×log(Density) | -0.047*** | | | -0.045*** | | |
| | (0.007) | | | (0.006) | | |
| Mid Share | | -0.102*** | | | -0.094*** | |
| | | (0.028) | | | (0.026) | |
| Mid×log(Density) | | 0.059*** | | | 0.058** | |
| | | (0.007) | | | (0.006) | |
| Upp Share | | | 0.384*** | | | 0.388*** |
| | | | (0.036) | | | (0.032) |
| Upp×log(Density) | | | 0.026*** | | | 0.022** |
| | | | (0.010) | | | (0.008) |
| Robust SE | Yes | Yes | Yes | Yes | Yes | Yes |
| CZone FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Product FE | No | No | No | Yes | Yes | Yes |
| $R^2$ | 0.321 | 0.150 | 0.091 | 0.528 | 0.373 | 0.467 |
| $N$ | 653,068 | 653,068 | 653,068 | 653,068 | 653,068 | 653,068 |

*Notes:* Market shares measures are constructed from Nielsen Retail Scanner Data. Population share measures are constructed from IPUMS ACS Data. We include all the contiguous commuting zones in the United States whenever data is available. Population density is computed based on population and land area data from NBER and US Census Bureau. Time is from 2006 to 2017. All regressions include a constant term. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 18: Market Shares of Different Price Range: Commuting Zones

| Dep. Var. | Below10 | Middle | Above90 | Below10 | Middle | Above90 |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Low Share | 0.148*** | | | 0.136*** | | |
| | (0.022) | | | (0.020) | | |
| Low×log(Density) | -0.020*** | | | -0.019*** | | |
| | (0.006) | | | (0.005) | | |
| Mid Share | | -0.075*** | | | -0.064*** | |
| | | (0.027) | | | (0.025) | |
| Mid×log(Density) | | 0.047*** | | | 0.047** | |
| | | (0.007) | | | (0.006) | |
| Upp Share | | | 0.326*** | | | 0.323*** |
| | | | (0.028) | | | (0.024) |
| Upp×log(Density) | | | 0.001 | | | 0.001 |
| | | | (0.008) | | | (0.006) |
| Robust SE | Yes | Yes | Yes | Yes | Yes | Yes |
| CZone FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Product FE | No | No | No | Yes | Yes | Yes |
| $R^2$ | 0.486 | 0.327 | 0.062 | 0.598 | 0.543 | 0.506 |
| $N$ | 653,068 | 653,068 | 653,068 | 653,068 | 653,068 | 653,068 |

*Notes:* Market shares measures are constructed from Nielsen Retail Scanner Data. Population share measures are constructed from IPUMS ACS Data. We include all the contiguous commuting zones in the United States whenever data is available. Population density is computed based on population and land area data from NBER and US Census Bureau. Time is from 2006 to 2017. All regressions include a constant term. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

## 2.5 Conclusion and Future Work

In this paper, we document a novel inequality phenomenon, in which the price distribution loses its mass at the middle price support, using the Nielsen Retail Scanner Data. We find that the market shares of middle-price products are declining in the United States, while the market shares of lower- and upper-income groups are rising for the period of 2006 to 2017. The missing middle phenomenon is robust at the national, state, and commuting zone levels and is more pronounced in densely populated metropolitan areas. We further link it to the rise in income inequality and show that demand-side demographics have significant influence on shrinking middle support in the price distribution. Our results shed important insights on the effect of changing price distribution or inflation inequality such as Kaplan and Schulhofer-Wohl (2017), Jaravel (2019), and Argente and Lee (2020). Our results provide a holistic view on the effect of a non-monotonic change in the entire price distribution and emphasize on the role of shrinking middle, rather than a change of mass in a certain part of the price support.

In future works, we plan to further extend the current work in two ways. First, we wish to continue building our work to assess the welfare implications of the missing middle in product price distribution, following the semi-parametric approach in Atkin et al. (2018). Their approach also allows for welfare decomposition which we could employ to further discuss on *how* the shrinking middle in price distribution is affecting welfare of different income groups. Second, we could also use the identification strategy in Jaravel (2019) to better assess the causal effects of the demand-side demographics on the price distribution.

# 3 On the Transition Dynamics of Trade Liberalization

## 3.1 Introduction

The composition of a country's export changes over time. This is especially pronounced in the aftermath of a trade liberalization. Take China as an example. China first opened up to international trade after the Cultural Revolution. In the meantime, the United States also granted China the Most-Favored Nation (MFN) status, which extensively cut the "Column 2 Tariffs" on Chinese exports in 1980.[22] Following this (partial) trade liberalization, the composition of Chinese exports has changed dramatically. In 1980, petroleum was the biggest component of Chinese exports and accounted for 22% of total export value. After the granting of MFN status, the export of textile, clothing, and footwear took off in the 80s, and they accounted for about 40% of total exports by 1992. In contrast, exports of petroleum was less than 15%.

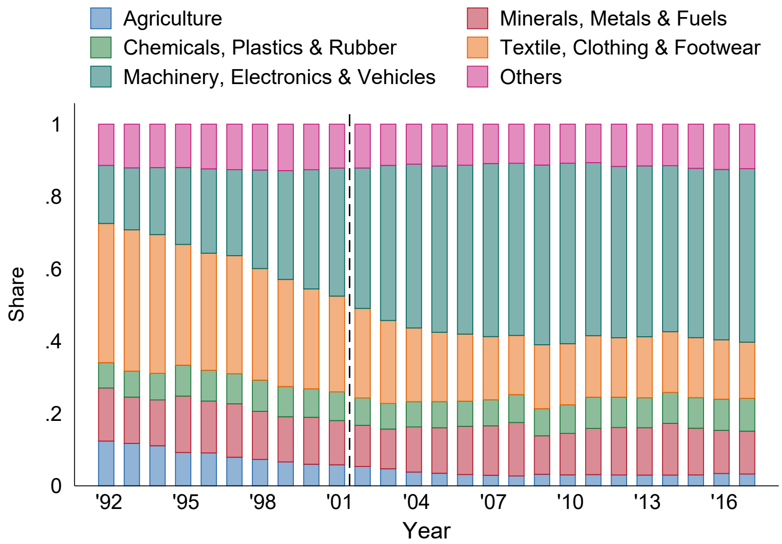
Given this change in export composition after partial trade liberalization, it seems to suggest that the comparative advantage of China is in textile products. But this pattern does not last long. Figure 31 plots the change in export composition from 1992 to 2017. It is evident that the share of textile exports has declined since 1992, and that the share of machinery & electronics has been rising. In 1992, the share of machinery, electronics, and vehicle exports was 16%. By 2004, it has increased to more than 45%, at the expense of textile export share. In particular, this seems to be a trend pre-existing before the accession to WTO in 2001.[23]

What explain this transition dynamics? What are the associated distributional consequences? In this paper, we propose a novel hypothesis that the interaction between *occupational* human capital accumulation and Heckscher-Ohlin factors matters. The key mechanism is such that as individuals accumulate more occupation-specific experiences, Heckscher-Ohlin effects would set in since occupation-specific skill supply becomes more abundant. As such, in the beginning episodes of a trade liberalization, industries with

[22] This MFN status is subject to review on an annual basis.

[23] Though the accession of WTO has certainly accelerated the growth in levels, see Figure **??** in the Appendix.

Figure 12: Composition of Chinese Exports, 1992–2017




*Notes:* HS 2-digit trade flow data are collected from Comtrade. The aggregation is done using similar classifications as those in the WITS database.

comparative advantage in the Ricardian sense would not experience immediate surge in exports. Because it takes time for workers to accumulate occupation-specific skills in industries such as electronics and vehicles , the exports of these sectors would only rise slowly. In contrast, for industries such as textiles that do not require too much occupation-specific skills, the workers can quickly pick up the job. As a result, most individuals would choose to work in those occupations, and exports will rise relatively fast in these industries for the beginning years.

The implication of this mechanism would then rationalize the transition dynamics of Chinese exports. In the initial period, exports of the textile industry rise sharply, because China is relatively more productive in this sector and the workers are relatively easy to train in textile production. In contrast, the machinery and electronics sectors hire intensively those occupations that demand more on-the-job human capital accumulation. The exports in these industries only rise slowly in the initial period, as occupation-specific skill supply is relatively scarce. Later, as more and more workers have accumulated sufficient occupation-specific human capital, the exports in these industries would take off and finally take over the textile industries in terms of export composition.

The above mechanism has profound roots in realism. Multinational corporations have long weighted the options of shifting their supply chain from China to neighboring countries with lower labor cost such as Vietnam and India. For example, there are reports from the Wall Street Journal on how Apple wanted to shift their supply chain and to reduce their dependency on China. However, one of the reasons that Apple did not do so, is that they have to "begin the multiyear process of training workers" and that "workers in India aren't ready to produce the high-end, organic light-emitting diode models".[24] These reflect that the industry is more concerned with training workers on-the-job than whether the workers become skilled through higher education.[25] In addition, the occupation margin of human capital accumulation has also been an important subject studied by labor economists. For example, an influential study (Kambourov and Manovskii, 2009) finds that occupational tenure has a substantial effect on the income that a person receives, while employer or sector tenure loses its explanatory power once occupational margin is controlled.

Following this logic, we then build a quantitative trade model with labor market dynamics. International trade in this model is pinned down by the comparative advantage pattern, with both Ricardian productivity differences and Heckscher-Ohlin factors. Labor market dynamics is characterized by Roy-type occupational choice and occupation-specific human capital accumulation. Key to our model is the interaction between Heckscher-Ohlin forces and dynamic Roy elements. In our model, skill supply is occupation-specific. This implies that the factor markets will clear by occupations. As individuals also accumulate more occupational skills on-the-job, the Heckscher-Ohlin comparative advantage is inherently dynamic. This will in turn shape the transition dynamics of exports following a trade liberalization as well as the distributional consequences.

[24]"Tim Cook and Apple Bet Everything on China. Then Coronavirus Hit." by Tripp Mickle and Yoko Kubota, March 3, 2020.

[25]In fact, this concern is even more pronounced in industries that employ workers with higher education. A fresh graduate who just becomes research engineer is vastly inexperienced in comparison to a senior engineer with decades of industrial experiences. Yet, these experiences constitute a significant proportion of skill supply in those industries.

We then quantify the relative importance of occupations in determining the transition dynamics through a simple toy calibration. Among other things, We find that the equilibrium transition dynamics are largely consistent with our prior expectations and the occupational forces significantly explain the change in export composition quantitatively. In particular, we compare the export composition over time in two simulations. The first simulation features a quantitative trade model with labor dynamics, and the second abstracts from any dynamic elements. We found that the first model is able to generate similar patterns in Figure 31, while the second simulation reports a constant export composition over-time. The change in export composition is quantitatively non-trivial as the difference between the two sectors shrinks by 70%.

Our paper contributes to three strands of literature. First, our work is related to the study of interaction between international trade and labor market dynamics. Among other recent works in this literature (Danziger, 2017; Dix-Carneiro and Kovak, 2017; Dix-Carneiro et al., 2019; Dix-Carneiro and Kovak, 2019; Erten et al., 2019) which span both empirical evidences and structurally-estimated dynamic models, our paper is especially relevant to Dix-Carneiro (2014) and Traiberman (2019). In particular, Dix-Carneiro (2014) focuses on how sector-specific human capital accumulation matters for distributional consequences of trade liberalization, while Traiberman (2019) quantifies the importance of occupation-specific human capital accumulation. They both develop a dynamic Roy model with competitive equilibrium in the domestic economy, while trade prices are calibrated as an external shock. In this sense, the class of models they construct is only applicable to small and open economies, in which changes in the domestic economy will not influence world prices.[26]

Our paper complements their works by further considering the case of multiple large countries. In our framework, Roy labor dynamics (i.e., individual occupational choice and human capital accumulation) will interact with the Heckscher-Ohlin forces in the general equilibrium and shape trade patterns during the transition periods. The intuition is that, as individuals choose different occupations and accumulate occupation-specific human

[26]Dix-Carneiro (2014) imposes the assumption that the Brazilian economy that he studies is an small and open economy, while Traiberman (2019) structurally estimates his model using Danish data.

capital, effective supply of human capital will change accordingly. This will change unit price of occupation-specific human capitals that are factors of production and influence world price through Heckscher-Ohlin forces. In contrast, this mechanism is absent in Dix-Carneiro (2014) and Traiberman (2019). For their models, changes in domestic human capital supply, be it sector-specific or occupation-specific, will have no impact on international prices because the country is perceived to be too small for the rest of the world.[27] Our paper fills this gap.

Second, our work is related to the literature of dynamic Heckscher-Ohlin models, which primarily focuses on skill acquisition through education and/or sector-specific human capital accumulation (Oniki and Uzawa, 1965; Findlay, 1970; Stiglitz, 1970; Mussa, 1978; Findlay and Kierzkowski, 1983; Borsook, 1987; Chen, 1992; Bond et al., 2003; Bajona and Kehoe, 2010; Falvey et al., 2010; Harris and Robertson, 2013; Auer, 2015; Guren et al., 2015). The basic logic behind this literature and our paper is essentially the same, that individual choices at the micro-level would have general equilibrium consequences through changes in factor price. In comparison to these papers, our work studies an additional occupational margin in which human capital accumulation may matter. This occupational margin has been shown to be quantitatively important in the case of small country models where dynamic Heckscher-Ohlin forces are absent (Traiberman, 2019). We show that this is also the case in the context of dynamic Heckscher-Ohlin model.[28]

In addition, our paper is relevant to the literature on trade and skill acquisition. Atkin (2016) finds that export expansions increase teenage school dropouts in Mexico, as the opportunity cost of attending schools (i.e., wages) becomes higher. Blanchard and Olney (2017) and Li (2018) complement this work by considering different skill types. They find that while export expansions in the unskilled sector encourage more school dropouts, ex-

[27]In this sense, closest to our work is Dix-Carneiro et al. (2019) which quantifies how labor dynamics is influenced through adjustment in the trade imbalances margin. Their model also features several large countries that have influence on world prices. While it provides a novel account of the job matching process, their model does not focus on human capital accumulation.

[28]Among other things, Traiberman (2019) finds that 57% of the variation in labor market outcomes is explained by the occupational margin in the Danish economy.

pansions in the high-skill exports incentivize more high-school and college enrollments.[29] In another paper, Chang and Huang (2014) studies how trade and education systems interact. Among other things, they show that trade may induce changes in the distribution of talents in the country, in that if a country (such as the East Asian countries) specialize in sectors that require a homogenized set of skills, then it is optimal for them to choose a centralized education system. Our work further complements this literature and consider a full Roy model in which people choose to work in different occupations including to attend schools. However, we stop short of investigating the quantitative implications of such models. The quantitative investigation of relative importance of Roy forces in comparison with that of the skill acquisition (education) is delegated to future works.

The remainder of this paper is organized as follow. Section 3.2 presents a formal model that integrates both the Roy elements and the Heckscher-Ohlin forces. Section 3.3 discusses its quantitative implications. Section 3.4 then simulates the model and discusses the dynamics of a calibrated trade liberalization. Finally, Section 3.5 concludes.

## 3.2 The Model

### 3.2.1 Consumption

There are $K$ sectors of composite goods. Consumers derive utilities from consuming these goods in a Cobb-Douglas fashion,

$$U = \prod_k \left(C^k\right)^{\gamma_k}$$

where $C^k$ is the CES composite for sector $k$ and $\sum_k \gamma_k = 1$ are the Cobb-Douglas coefficients. For each CES composite, it is combined through a continuum of sector-specific varieties

$$C^k = \left(\int q^k(\omega)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$$

[29]For other relevant papers, see Blanchard and Willmann (2016), Greenland and Lopresti (2016), and Chen et al. (2017)

where $q^k(\omega)$ is the consumption quantity of sector-$k$ variety $\omega$ and $\sigma$ is the elasticity of substitution. Given an income $E$, the budget constraint for sector $k$ composite is

$$\int q^k(\omega)p^k(\omega)d\omega = \gamma_k E$$

where $p^k(\omega)$ is the price of sector-$k$ variety $\omega$.

### 3.2.2 Production

A firm's production technology is to combine different tasks performed by different occupations, which we index by $d$, of labor. The intensity of each type is heterogeneous across sectors and captured by $\alpha_d^k$.

$$y^k(\omega) = \phi(\omega) \prod_{d \leqslant D} \left(l_d^k(\omega)\right)^{\alpha_d^k}$$

where $\phi(\omega)$ is the productivity of the firm. Given perfect competition, the firm producing variety $\omega$ in industry $k$ then charges a local price at its marginal cost

$$p_i^k(\omega) = \frac{c_i^k}{\phi(\omega)}$$

where $c_i^k$ is the cost of an unit input bundle for sector-$k$ in country $i$. This unit cost is simply a function of factor prices (occupation wages),

$$c_{i,t}^k = B^k \prod_d \left(w_{d,t}^i\right)^{\alpha_d^k}$$

and $B^k = \prod_{d \leqslant D} \left(\alpha_d^k\right)^{-\alpha_d^k}$ is a bundle of sector-$k$ specific coefficients.

### 3.2.3 Occupational Choice

Workers make two choices in each period. First, they choose which occupation they should take up in this period. Second, they choose how much of each good to consume. We model these choices as independent decisions and the latter is covered in the section on consumption. In a partial equilibrium sense, job choices are independent from consumption decisions conditional on the distribution of wages across occupations. Given our setup, the expected utility for a worker $s$ of age $a$ at calendar time $t$ is

$$V_{a,t}^s(x_{a,t}^s, \epsilon_{d,a}^s) = E_{a,t} \left\{ \max_{\{d_{a'}^s\}_{a'=t}^T} \sum_{a'=t}^T \beta^{a'-t} \left[u\left(x_{a',t'}^s\right) - \lambda(d_{a'}^s, d_{a'-1}^s) + \epsilon_{d,a'}^s\right] \right\}$$

where $x^s_{a,t}$ is a set of state variables and $\epsilon^s_{d,a}$ is the occupation-specific idiosyncratic shock. Given these notations, $u(x^s_{a,t})$ is then the current-period utility given by the real income

$$u(x^s_{a,t}) = \frac{y^s_{a,t}(x^s_{a,t})}{P_t}$$

where $y^s_{a,t}$ is the income from the chosen job and $P_t = \prod_{k=1}^{K} \left(P^k_t/\gamma^k\right)^{\gamma^k}$ is the ideal price index. We also include a non-pecuniary cost from switching occupations, $\lambda(d^s_a, d^s_{a-1})$ which measures the cost of switching from an occupation to another occupation at age $a$ for individual $s$. The purpose of this term is to discipline the model so that it's desirable to have a continuous career and hence rationalizes the empirical data. This assumption is standard in the labor and trade models that involve occupational choice.

The income of choosing occupation $d \leqslant D$ equals to prevailing wage in that occupation multiplied by the level of occupation-specific skill a worker possesses,

$$y^s_{a,t}(x^s_{a,t}) = w_{d,t} \cdot z^s_{d,a}$$

where $w_{d,t}$ is the "skill price" of human capital in occupation $d$ at calendar time $t$ and the level of occupation-specific human capital $z^s_{d,a}$ is a function of experiences.

$$z^s_{d,a} = \left[\bar{\eta}_{d,0} + \bar{\eta}_{d,1} e^s_{d,a} + \bar{\eta}_{d,2}\left(e^s_{d,a}\right)^2\right] \exp\left(\varepsilon^s_{d,a}\right)$$

such that experience or occupation-specific human capital accumulates in discrete time,

$$e^s_{d,a} = e^s_{d,a-1} + \mathbf{1}\{d^s_{a-1} = d\}$$

Finally, we assume in our model that workers have perfect foresight, in the sense that they form expectations about wages that are consistent with what the actual level will be. Rational expectation then imposes that these expected wages must coincide with the wages that clear the labor markets for each occupation. In addition, we assume that experience or occupation-specific human capital accumulates in discrete time,

### 3.2.4 International Trade

We follow the standard assumptions in canonical quantitative trade models that the productivity of producing each variety is drawn randomly from a Fréchet distribution,

$$F^k_i(\phi) = \exp\left(-\left(\phi/\phi^k_i\right)^{-\theta^k}\right)$$

where $\phi_i^k$ is the scale parameter and $\theta^k$ is the shape parameter. Perfect competition then entails that consumers would choose the lowest possible price from each origin $i$ such that

$$p_n^k(\omega) = \min_i \left\{\tau_{in} p_i^k(\omega)\right\}$$

By standard algebra, the Fréchet distribution assumption leads us to a close-form expression for trade shares

$$\pi_{in}^k = \frac{\phi_i^k \left(\tau_{in} c_i^k\right)^{-\theta^k}}{\sum_i \phi_i^k \left(\tau_{in} c_i^k\right)^{-\theta^k}}$$

To close the model, we impose that the labor market clears at occupational-level for each country $i$, which implies

$$\underbrace{w_d^i \cdot h_d^i}_{\text{skill supply}} = \sum_k \sum_n \underbrace{\alpha_d^k \pi_{in}^k \gamma^k Y_n}_{\text{skill demand}}.$$

This condition is analogous to the trade balance condition in canonical trade models such as those in Caliendo and Parro (2015)

#### 3.2.5 General Equilibrium

We define a general equilibrium in our model as an rational expectation equilibrium, such that it is a list $\{d_a^s, w_{d,t}^n, P_t^{n,k}, h_{d,t}^n, \pi_{in,t}^k\}_{s,a,n,i,d,k,t}$ that satisfies the following equilibrium conditions,

$$\begin{aligned} E_{a,t}\left[\max_{d_a^s} V_{a,t}^s(x_{a,t}^s, \epsilon_{d,a}^s)\right] = \max_{d_a^s} \quad & u(x_{a,t}^s, d_a^s) - \lambda(d_a^s, d_{a-1}^s) + \epsilon_{d,a}^s \\ & + \beta E_{a,t}\left[\max_{d_{a+1}^s} V_{a+1,t+1}^s(x_{a+1,t+1}^s, \epsilon_{d,a+1}^s)\right] \end{aligned} \tag{8}$$

$$P_t^{n,k} = A^k \left[\sum_i \phi_i^k \left(\tau_{in,t} c_{i,t}^k\right)^{-\theta^k}\right]^{-\theta^k} \tag{9}$$

$$c_{i,t}^k = B^k \prod_d \left(w_{d,t}^i\right)^{\alpha_d^k} \tag{10}$$

$$\pi_{in,t}^k = \frac{\phi_i^k \left(\tau_{in,t} c_{i,t}^k\right)^{-\theta_k}}{\sum_i \phi_i^k (\tau_{in,t} c_{i,t}^k)^{-\theta^k}} \tag{11}$$

$$w_{d,t}^i \cdot h_{d,t}^i = \sum_k \sum_n \alpha_d^k \pi_{in,t}^k \gamma^k Y_t^n, \quad \text{where} \quad h_{d,t}^i = \sum_s z_a^s \cdot \mathbf{1}\{d_a^s = d\} \tag{12}$$

and that the actual path of wages is consistent with workers' rational expectations.

### 3.2.6 Algorithm for Solving the Model

Similar to Traiberman (2019), we solve the model in the following steps given the values of the exogenous parameters.

1. Guess a path of skill price $w^i_{d,t}$ for each occupation $d$ in each country $i$ at each calendar time $t$.[30]

2. Solve for the unit cost bundle and ideal price index using conditions (9) and (10).

3. Solve for worker's dynamic problem using condition (13).

4. Solve for the trade shares using conditions (11).

5. Solve for the skill price that would clear the labor market in condition (12) using the trade shares and optimal choices of workers.

6. Update the guess in Step 1 using a weighted average of the new skill price and the guess.

7. Iterate Step 1 to 6 until convergence.[31]

## 3.3 Quantitative Implications

So far, we have built a canonical multi-sector trade model with labor dynamics. But what are its quantitative implications? Before jumping into the black box of computational simulations, we establish and discuss some quantitative priors in this section.

### 3.3.1 Roy Forces

First, we discuss the Roy elements in the partial equilibrium with the general equilibrium skill prices taken as given. Given the structure of human capital accumulations, workers in the partial equilibrium face occupational choice problem in the current period as well

[30] In practice, we set our initial guess of skill prices to be vectors of one.

[31] In practice, we iterate until the labor market imbalances at each occupation-country-year are within 0.1%.

as an inter-temporal trade-off. In particular, for an occupation with slow accumulation of human capital, the workers have to balance the trade-off between accepting a lower income in the present and receiving higher pay in the future when he accumulates more on-the-job skills. Given that the workers expect to accumulate more occupation-specific human capital over his tenure, a wage profile that features a relatively lower skill price in the early period (in comparison with the blue collar wages) would also be compatible with rational choices, since he views it as a sacrifice in the present to break into higher income in the future. But, a worker would more likely to choose a white-collar or engineer job in the early life cycle if the pay in the initial period is higher, ceteris paribus.

### 3.3.2 Heckscher-Ohlin Forces

What distinguish our work from the contemporary literature are the interactions between Roy choices in the partial equilibrium and skill prices in the general equilibrium through Heckscher-Ohlin factors. Specifically, in a trade liberalization episode, a country may experience a shortage of some occupation-specific skills because workers do not accumulate enough such human capital in the autarky. As a result, in the initial periods of trade liberalization, the skill price for these occupational human capital would be considerably high because of the scarcity of factors. This would further hamper exports of goods that employ such human capital extensively because the unit cost is driven up despite that the country may have the comparative advantage in Ricardian sense. The relatively high factor price would persist into future. However, the high skill price in these sectors would also incentivize more newborn workers to choose white-collar occupations (engineers in our model). As more individuals have accumulated enough occupation-specific human capital, the factor price will start to decline and exports finally take off. Our prior, if true, would rationalize the dynamics of Chinese export composition as shown in Figure 31, where upon the granting of MFN status by the United States in the beginning of 1980s, textile started to rise in terms of share of overall exports. But the electronics exports took over since 1990s and gradually became the largest component of Chinese exports. Our model partly attributes this dynamics to the slow human-capital accumulation in the electronics industries as these sectors employed more engineers or white-collar occupations.

Our induction is that, when China liberalized for international trade, skilled technicians were relatively scarce in the country and hence the factor price was relative expensive. As a results, electronics exports did not take off initially and only started to rise in terms of composition in later years, when the young Chinese graduates found it more lucrative to become technicians/engineers and have accumulated enough occupation-specific human capital through job tenure.

### 3.3.3 Cohort Analysis

Another important question to discuss in trade policy evaluations would be worker welfare. Given that all our workers are ex-ante identical before drawing idiosyncratic shocks, we will focus our discussions across different cohorts of workers. Focusing on the South, our prior is that newborn workers at the time of trade liberalization stand to gain most, because they only begin to enter the job market and hence can freely choose occupations without much shocks. In contrast, the old workers in the South stand to gain less. It is more costly for them to switch occupations in the later part of their life cycle, especially if they have chosen to be a blue-collar laborer for the majority of their occupational tenure.

## 3.4 Simulations

To investigate the quantitative implications of our model, we now calibrate a toy version with 2 countries (North and South), 4 industries (Aircraft, Automobile, Electronics, and Textile), and 2 occupations (Engineers and Laborers). For simplicity purposes, we do not include any education decisions in the model. This strips the ability of the model to quantify the relative importance of occupation-specific human capital accumulation and education. However, we deem this as an empirical exercise and thus it requires thorough estimation of a labor market model to justify the choice of parameters. We will delegate this task to a more rigorous work in the future.

### 3.4.1 Calibration

First, we calibrate the model in a North-South setup. The North is a technologically more advanced country such that it has absolute advantage in producing all goods in the 4 industries. Apart from the technological differences, the North and South are entirely symmetric.[32] In this sense, the autarkic equilibrium in each country is entirely driven by technological differences.

For the labor market, we assume that there are two occupations in the economy: Engineers ($e$) and Laborers ($b$). These jobs differ mainly in terms of how fast occupation-specific skills accumulate. We assume that it takes a longer time for engineers to accumulate skills on-the-job and it is relatively easier for laborers to accumulate skills. This assumption is not short of realism. Consider two occupations, laborers working on the assembly lines for shoes and process engineers working in the semiconductor industry. It takes only a few days for new laborers to be as good as experiences laborers. In contrast, it takes years for a fresh graduate to be an experienced process engineer.

We then calibrate 4 industries and label them as Aircraft ($A$), Automobiles ($M$), Electronics ($E$), and Textile ($T$) which differ in factor intensity in employing engineers and laborers. We assume that the aircraft industry employs 80% engineers and 20% laborers; the automobiles industry employs 60% engineers and 40% laborers; the electronics industry employs 50% engineers and 50% laborers; and the textile industry employs 10% engineers and 90% laborers. These coefficients are symmetric across all countries. The North and South only differ in technology in terms of the Fréchet parameters for productivity term $\phi(\omega)$.

Next, we calibrate the Ricardian elements in our model. The Ricardian elements are mainly characterized by the scale parameters $\phi_i^k$ and shape parameter $\theta^k$ of the Fréchet distribution. Following the over-identifying optimal estimates for Eaton-Kortum model in Simonovska and Waugh (2014), we set $\theta^k = 4$ for any industry $k$. For the scale parameter, we assume that in the North productivity is generally higher such that the North has absolute advantage in producing all goods. In the absence of labor market dynamics, the

[32] Note that we abstract from education decisions in the model.

comparative advantage pattern of this model is then determined by the relative ratio of the absolute productivity terms, by Corollary 1 in Costinot et al. (2012). Specifically, we assume that the South tend to have comparative advantage in products that employ less engineers,

$$\frac{\phi_N^T}{\phi_S^T} \leqslant \frac{\phi_N^E}{\phi_S^E} <<<< \frac{\phi_N^A}{\phi_S^A} \leqslant \frac{\phi_N^M}{\phi_S^M}.$$

such that in general the South has Ricardian comparative advantage in the textile and electronics industries while the North has Ricardian comparative advantage in the automobiles and aircraft industries.[33]

For the labor market parameters, we abstract from transition cost, quadratic experience term, and indiviudal heterogeneity in terms of initial human capital.[34] Given our assumptions, we calibrate the intital occupational skills of the engineers (white-collar occupations) when they first enter the labor market as 0.5, while that of the laborers (blue-collar occupations) as 4. Next, we calibrate the occupational skills accumulation parameters as 0.25 and 0.025 for the engineers and laborers respectively. The reasons that we choose these values are as follows. First, laborers and blue-collar occupations do not acquire much occupational-specific skills over the years, as their main job scope requires more physical strength rather than complicated cognitive tasks. Hence, their life-cycle skills should be such that it is initially at high-level and slowly increases over the years.[35] In contrast, the job scope of engineers or white-collar jobs in general involves more complicated cognitive tasks and relies more on on-the-job accumulation of experiences. The life-cycle skill profile of these jobs should be such that it is initially at low-level but steadily increases over the life cycle. Hence, we choose to set a low starting point for the white-collar occupations but with a relatively steeper slope.

Finally, for the consumption parameters, we assume that individuals consumes equal

[33] In our calibration, we set $\phi_N^T/\phi_S^T = 8/7$, $\phi_N^E/\phi_S^E = 6/3$, $\phi_N^M/\phi_S^M = 6/1$, and $\phi_N^A/\phi_S^A = 5/1$.

[34] Note that it is plausible to incorporate skill (education) acquisition in this model, such that individuals in the North would initially have more white collar (engineers) human capital which is related to a high-level of education.

[35] As mentioned above, we do not cater the possibility of declining skills when a worker is aging, as we abstract from the quadratic experience term.

shares of any goods such that each good accounts for 25% of their expenditure. In sum, our choice of parameters are summarized in the following table.

Table 19: Calibration of parameters

| Parameters | Value |
|---|---|
| Shape parameter of Fréchet distribution | $\theta^k = 4$ |
| Scale parameter of Fréchet distribution (North) | $(\phi_N^T, \phi_N^E, \phi_N^M, \phi_N^A) = (8, 6, 6, 5)$. |
| Scale parameter of Fréchet distribution (South) | $(\phi_S^T, \phi_S^E, \phi_S^M, \phi_S^A) = (7, 3, 1, 1)$. |
| Consumption Cobb-Douglas shares | $(\gamma^T, \gamma^E, \gamma^M, \gamma^A) = (0.25, 0.25, 0.25, 0.25)$. |
| Production Cobb-Douglas shares | $(\alpha_e^T, \alpha_b^T, \alpha_e^E, \alpha_b^E) = (0.1, 0.9, 0.7, 0.3)$. |
| Production Cobb-Douglas shares | $(\alpha_e^M, \alpha_b^M, \alpha_e^A, \alpha_b^A) = (0.8, 0.2, 0.9, 0.1)$. |
| Initial occupational skills | $(\bar{\eta}_{e,0}, \bar{\eta}_{b,0}) = (0.5, 4)$. |
| Occupational skills accumulation | $(\bar{\eta}_{e,1}, \bar{\eta}_{b,1}) = (0.25, 0.025)$. |

### 3.4.2 Transition Dynamics

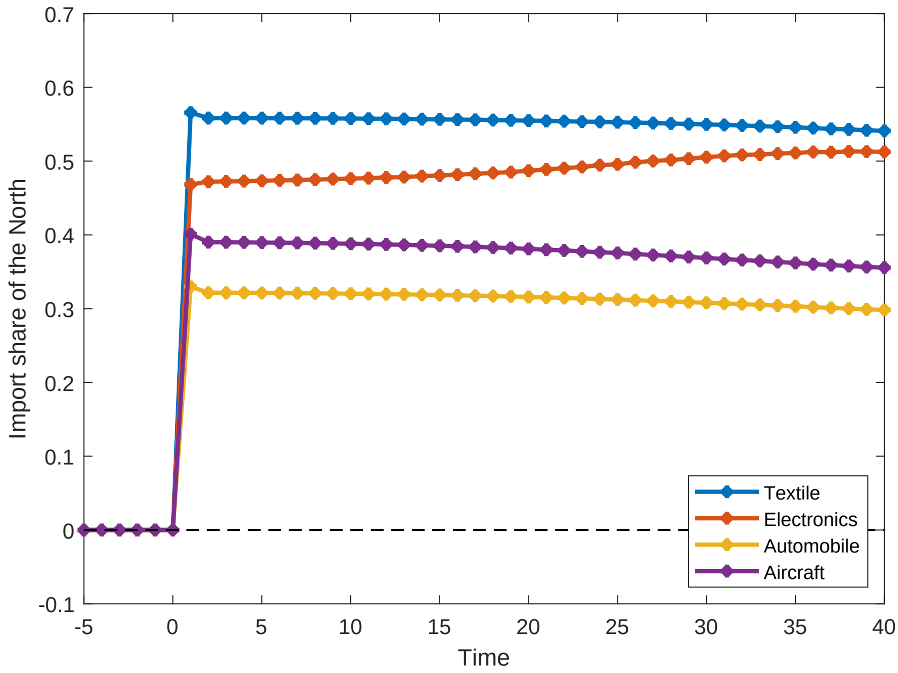
After calibrating the parameters as discussed in the above, we quantify the model with a simulation with overlapping generations. Specifically, we simulate different cohorts of individuals where each cohort contains 1,000 persons. The length of the life cycle is defined to be 40 periods, which means each cohort of individuals will live and work for 40 periods before they exit from the economy. Then, we simulate the North and South in autarky, such that the trade cost is set to an arbitrarily high value and the foreign trade shares are approximately zero. Under these conditions, we simulate overlapping generations of workers who rationally expect the general equilibrium prices under autarky and decide their occupational choice accordingly. We iterate the skill prices until the equilibrium is reached with no labor market imbalances in each occupation-country-year and define this equilibrium as the autarky equilibrium.

Next, we shock the model by the setting the trade cost to free-trade level at the end of the autarky equilibrium and continue to simulate new generations of workers. The new equilibrium is reached again by iterating over the vectors of skill prices until convergence. We define this equilibrium as the equilibrium under free trade and quantify its behavior by collecting individual choices and general equilibrium prices and quantities at each country-time. The export composition of the South is then simply defined as the import shares of the North, because it is the only destination that the South exports to.

The transition dynamics of South exports is plotted in Figure 13. We make several observations. First, the relative magnitude of export values are consistent with the Ricardian comparative advantage pattern, in the sense that the South exports textile the most and automobiles the least. Second, the transition dynamics over time are consistent with the pattern reported in Figure 31. Textiles were the largest component export given an initial trade liberalization, but its shares gradually decline overtime while that of electronics rises. This confirms our hypothesis that Roy-type occupational skill accumulation in the partial equilibrium will influence comparative advantage in the general equilibrium. Note that because the Ricardian comparative advantage is pre-determined, any changes in trade shares after the trade liberalization can only be attributed to changes in factor endowment. Third, in the meantime, the export shares of automobile and aircraft also decline overtime. The reason for this is that, trade liberalization in our model is not an unilateral one, as the reduction in trade cost is bilateral. As the North workers switch into engineers jobs or white-collar occupations, the North is also increasingly more competitive in their comparative advantage sectors. As a result, their import shares of these sectoral goods fall over time.

Next, we run another simulation in a "counterfactual" world where we eliminate all dynamic elements. Specifically, we set all dynamic occupational skill accumulation parameters to zero and reset the initial endowment to 5.5 and 4.5 respectively for the engineers and laborers. Although we maintain that the workers will choose their occupations in each period, the problem essentially reduces to a static one in every period. We report the export composition over time under this setup in Figure 14. As expected, the dynamics disappear under the new setup. In particular, the relative ratio between the textile and electronics sector exports now remain constant. This confirms our hypothesis that the transition of export composition reported in Figure 13 is driven by occupation-specific human capital accumulation. In addition, the trade shares are also in general lower in levels. This is perhaps a consequence of bad calibration. Our choice of initial human capital

Figure 13: Relative Value of South Exports

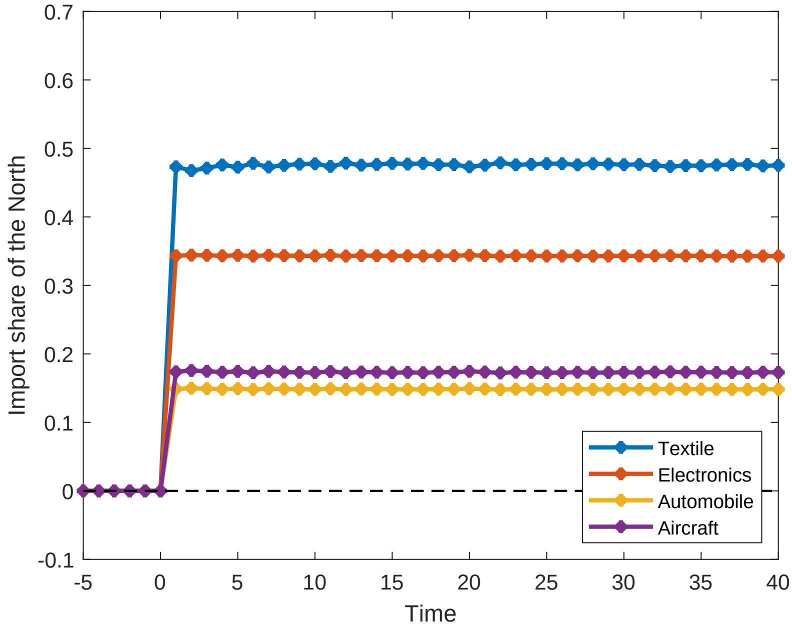



*Notes:* Each dot represents the import share of the North from the South for each sector of goods.

endowment falls short of the Roy specialization pattern under the dynamics setup. As the difference between the two countries in terms of factor endowment are smaller, the trade shares also appear to be smaller.

We also document the change in equilibrium nominal skill prices. The results are plotted in Figure 15 and 16. These figures are largely consistent with our expectations. At first, the Heckscher-Ohlin forces are at play such that the factors employed more intensively in the comparative advantage sectors sees a large jump in the nominal factor prices. Then, Roy forces are in play. As more and more workers switch to engineers occupations in both countries, the relative skill price for engineers start to decline, hence fueling the relative increase in export shares in each country. Similarly, as more and more workers switch to engineers occupations, laborers become relatively scarce and the skill price for them start to rise.

Figure 14: Relative Value of South Exports



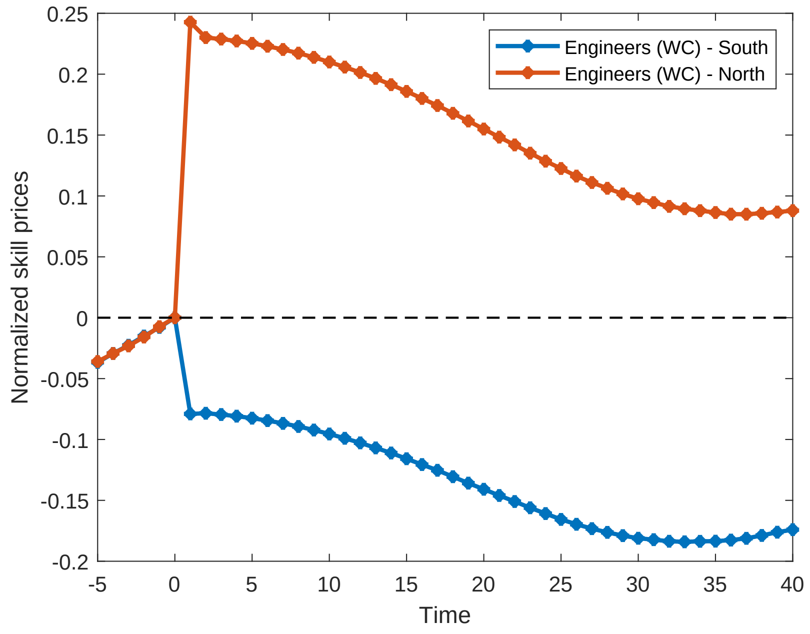

*Notes:* Each dot represents the import share of the North from the South for each sector of goods.

### 3.4.3 Welfare Implications

Next, we proceed to compare the welfare across cohorts. Notice that computing the welfare gains is meaningless here because our simulations of trade liberalization is from autarky to complete free trade. For the gains from trade literature, the welfare gains are usually computed moving from current tariffs to a no-tariff world which are substantially smaller than to a no-trade cost world. Hence, the welfare gains would be even larger for the current exercise, because our baseline is autarky. Second, as we reduce trade cost from all to nothing, the Ricardian gains from trade essentially dominates any adverse Heckscher-Ohlin effects and all occupations stand to gain from trade. For this reason, we only compare gains from trade across occupations and cohorts.

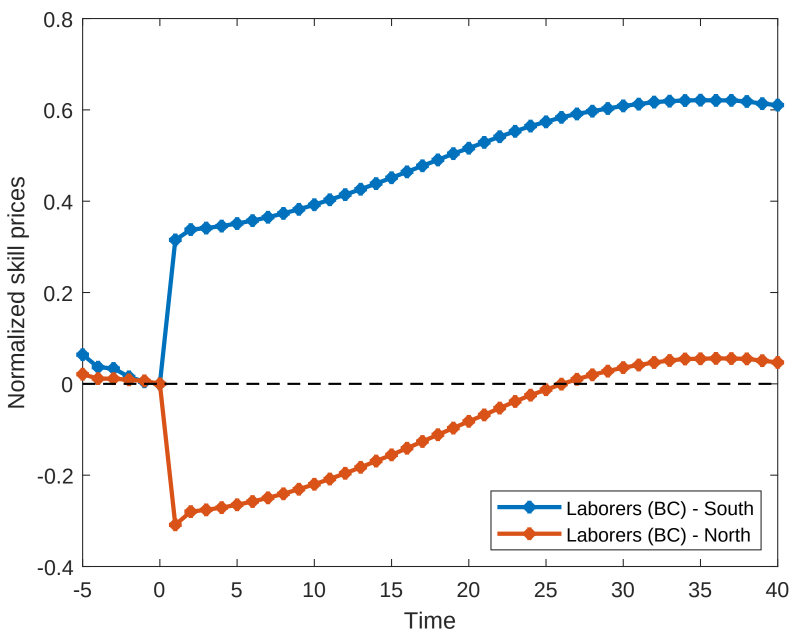
The results are plotted in Figure 17 and 18, where the normalized of log change in welfare is summarized for different age groups. Each age group is a 5-year interval for individuals at the time of trade liberalization. Group 1 stands for the youngest individuals at age 1 to 5 while group 8 stands for the oldest cohorts at age 36 to 39. The welfare changes

Figure 15: Transition Dynamics of White-Collar Skill Prices




*Notes:* Each dot represents the log of normalized skill prices.

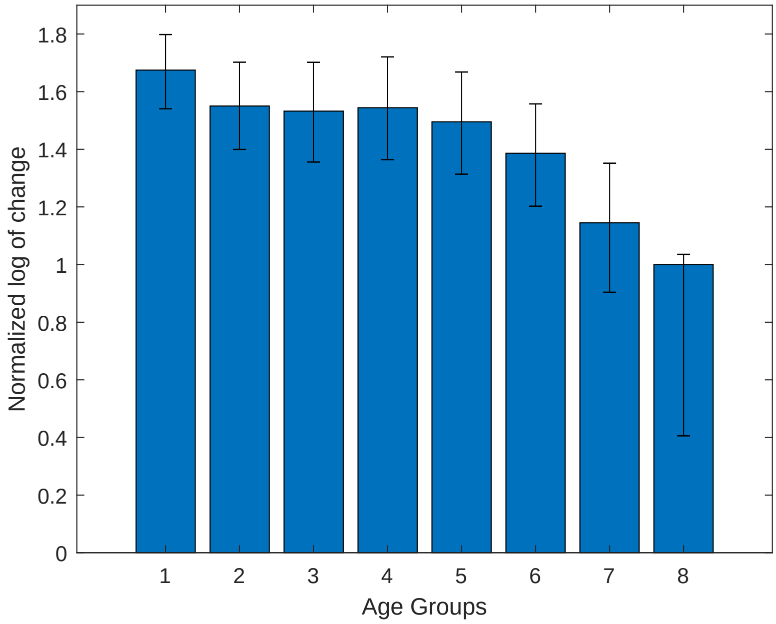
Figure 16: Transition Dynamics of Blue-Collar Skill Prices




*Notes:* Each dot represents the log of normalized skill prices.

are taken from two simulations, where the first experiment simulates trade liberalization and the second experiment simulates continuation of autarky.

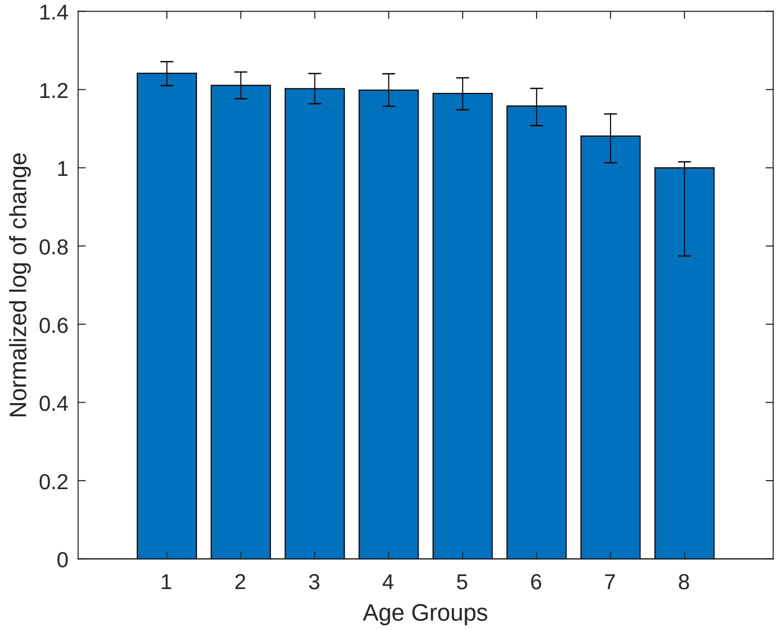
Figure 17: Distribution of Welfare Gains: North




*Notes:* Each bar represents the average of normalized log of change in welfare for each binned age group. Welfare measures are the discounted life-cycle of real consumption of workers. The change in welfare is computed based on the ratio between a simulation of trade liberalization and continuation of autarky. Age groups are 5-year intervals for individuals at the time of trade liberalization, where group 1 stands for workers at age 1 to 5 while that of group 8 is age 36 to 39. Error intervals represent the 10th and 90th percentiles of changes within the corresponding age group.

These results are consistent with our prior. First, we expected that the old workers should gain the least from the trade liberalization while young workers should gain the most. This is consistent in both the North and the South, where young workers gain significantly more than the old. Second, the distribution of welfare gains matter. In the North the distribution is more dispersed while it is less so in the South. This implies that trade liberalization contributes to the rise of inequality in the North and is consistent with many literature.

Figure 18: Distribution of Welfare Gains: South




*Notes:* Each bar represents the average of normalized log of change in welfare for each binned age group. Welfare measures are the discounted life-cycle of real consumption of workers. The change in welfare is computed based on the ratio between a simulation of trade liberalization and continuation of autarky. Age groups are 5-year intervals for individuals at the time of trade liberalization, where group 1 stands for workers at age 1 to 5 while that of group 8 is age 36 to 39. Error intervals represent the 10th and 90th percentiles of changes within the corresponding age group.

## 3.5 Conclusion

In this paper, we propose a novel mechanism that would rationalize the transition dynamics of trade liberalization that appeared in large countries such as China and Korea, which have the ability in influencing international prices. Different from the current literature on trade and labor dynamics model studying small countries, our hypothesis centers on the interaction between individual occupational choices and general equilibrium Heckscher-Ohlin forces. In particular, we argue that it takes time for the exports of certain industries to take off, because occupations in these industries usually features slow accumulation of occupation-specific human capital. As a result, the comparative advantage is endogenous as the exporting country slowly accumulates occupation-specific factors over time. We further quantify our arguments by simulating a Ricardian-Heckscher-Ohlin model with labor dynamics and overlapping generations. We find that the equilibrium transition dynamics are largely consistent with our prior expectations and the occupational forces significantly explain the change in export composition quantitatively.

In future works, we plan to further build on our current works to address our hypothesis in a more empirical sense. In particular, we wish to carefully calibrate the model to the settings of China and the United States so as to quantify the transition dynamics of Chinese export. In addition, we wish to introduce skill acquisition into the model, so that we can quantify the relative contribution of occupational forces in shaping the transition dynamics of trade liberalization in a large country, in comparison with that of education.

# 4 Bibliography


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.

Alstadsaeter, Annette, Niels Johannesen, and Gabriel Zucman, "Who owns the wealth in tax havens? Macro evidence and implications for global inequality," *Journal of Public Economics*, 2018, *162*, 89–100.

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

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.

Argente, David and Munseob Lee, "Cost of Living Inequality during the Great Recession," *Journal of the European Economic Association, forthcoming*, 2020.

Atkin, David, "Endogenous Skill Acquisition and Export Manufacturing in Mexico," *American Economic Review*, 2016, *106* (8), 2046–2085.

_ , Benjamin Faber, and Marco Gonzalez-Navarro, "Retail Globalization and Household Welfare: Evidence from Mexico," *Journal of Political Economy*, 2018, *126* (1), 1–73.

_ , _ , Thibault Fally, and Marco Gonzalez-Navarro, "A New Engel on Price Index and Welfare Estimation," *Working Paper*, 2020.

Auer, Raphael A., "Human Capital and the Dynamic Effects of Trade," *Journal of Development Economics*, 2015, *117*, 107–118.

Autor, David H., "Work of the Past, Work of the Future," *AEA Papers and Proceedings*, 2019, *109*, 1–32.

_ and David Dorn, "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market," *American Economic Review*, 2013, *103* (5), 1553–97.

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

Bajona, Claustre and Timothy J. Kehoe, "Trade, Growth, and Convergence in a Dynamic Heckscher-Ohlin Model," *Review of Economic Dynamics*, 2010, *13*, 487–513.

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.

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.

Blanchard, Emily J. and Gerald Willmann, "Trade, Education, and the Shrinking Middle Class," *Journal of International Economics*, 2016, *99*, 263–278.

— and William W. Olney, "Globalization and Human Capital Investment: Export Composition Drives Educational Attainment," *Journal of International Economics*, 2017, *106*, 165–183.

Bond, Eric W., Katheleen Trask, and Ping Wang, "Factor Accumulation and Trade: Dynamic Comparative Advantage with Endogenous Physical and Human Capital," *International Economic Review*, 2003, *44* (3), 1041–1060.

Borsook, "Earnings, Ability and International Trade," *Journal of International Economics*, 1987, *22*, 281–295.

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.

Chancel, Lucas and Thomas Piketty, "Indian Income Inequality, 1922-2015: From British Raj to Billionare Raj," *Review of Income and Wealth*, 2019, *65* (S1), 33–62.

Chang, Pao-Li and Fa-Li Huang, "Trade and Divergence in Education Systems," *International Economic Review*, 2014, *55* (4), 1251–1280.

Chen, Bo, Miaojie Yu, and Zhihao Yu, "Measured Skill Premia and Input Trade Liberalization: Evidence from Chinese Firms," *Journal of International Economics*, 2017, *109*, 31–42.

Chen, Zhiqi, "Long-Run Equilibria in a Dynamic Heckscher-Ohlin Model," *Canadian Journal of Economics*, 1992, *25* (4), 923–943.

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.

Costinot, Arnaud, Dave Donaldson, and Ivana Komunjer, "What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas," *Review of Economic Studies*, 2012, *79* (2), 581–608.

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.

Danziger, Eliav, "Skill Acquisition and the Dynamics of Trade-Induced Inequality," *Journal of International Economics*, 2017, *107*, 60–74.

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.*

Dix-Carneiro, Rafael, “Trade Liberalization and Labor Market Dynamics,” *Econometrica*, 2014, *82* (3), 825–885.

— and Biran K. Kovak, “Margins of Labor Market Adjustment to Trade,” *Journal of International Economics*, 2019, *117*, 125–142.

— and Brian K. Kovak, “Trade Liberalization and Regional Dynamics,” *American Economic Review*, 2017, *107* (10), 2908–2946.

— , Joao Pessoa, Ricardo Reyes-Heroles, and Sharon Traiberman, “Globalization, Trade Imbalances, and Labor Market Adjustment,” *Mimeo*, 2019.

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.

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

Erten, Bilge, Jessica Leight, and Fiona Tregenna, "Trade Liberalization and Local Labor Market Adjustment in South Africa," *Journal of International Economics*, 2019, *118*, 448–467.

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

— and — , "Firm Heterogeneity in Consumption Baskets: Evidence from Home and Store Scanner Data," *Working Paper*, 2019.

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.

Falvey, Rod, David Greenaway, and Joana Silva, "Trade Liberalisation and Human Capital Adjustment," *Journal of International Economics*, 2010, *81*, 230–239.

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.

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.

Findlay, Ronald, “Factor Proportions and Comparative Advantage in the Long Run,” *Journal of Political Economy*, 1970, *78* (1), 27–34.

_ and Henry Kierzkowski, “International Trade and Human Capital: A Simple General Equilibrium Model,” *Journal of Political Economy*, 1983, *91* (6), 957–978.

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

Garbinti, Bertrand, Jonathan Goupille-Lebret, and Thomas Piketty, “Income inequality in France, 1900–2014: Evidence from Distributional National Accounts (DINA),” *Journal of Public Economics*, 2018, *162*, 63–77.

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.

Greenland, Andrew and John Lopresti, “Import Exposure and Human Capital Adjustment: Evidence from the U.S.,” *Journal of International Economics*, 2016, *100*, 50–60.

Guren, Adam, David Hémous, and Morten Olsen, “Trade Dynamics with Sector-Specific Human Capital,” *Journal of International Economics*, 2015, *97*, 126–147.

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.

Harris, G. Richard and Peter E. Robertson, "Trade, Wages and Skill Accumulation in the Emerging Giants," *Journal of International Economics*, 2013, *89*, 407–421.

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.

Jaravel, Xavier, "The Unequal Gains from Product Innovations: Evidence from the U.S. Retail Sector," *Quarterly Journal of Economics*, 2019, *134* (2), 715–783.

_ and Erick Sager, "What are the Price Effects of Trade? Evidence from the U.S. and Implications for Quantitative Trade Models," *Working Paper*, 2019.

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

Kambourov, Gueorgui and Iourii Manovskii, “Occupational Specificity of Human Capital,” *International Economic Review*, 2009, *50* (1), 63–115.

Kaplan, Greg and Sam Schulhofer-Wohl, “Inflation at the household level,” *Journal of Monetary Economics*, 2017, *91*, 19–38.

Karabarbounis, Loukas and Brent Neiman, “The Global Decline of the Labor Share,” *Quarterly Journal of Economics*, 2013, *129* (1), 61–103.

Keane, Michael P. and Kenneth I. Wolpin, “The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence,” *Review of Economics and Statistics*, 1994, *76* (4), 648–672.

_ and _ , “The Career Decisions of Young Men,” *Journal of Political Economy*, 1997, *105* (3), 473–522.

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

Kopczuk, Wojciech, Emmanuel Saez, and Jae Song, “Earnings Inequality and Mobility in the United States: Evidence from Social Security Data Since 1937,” *Quarterly Journal of Economics*, 2010, *125* (1), 91–128.

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.

Li, Bingjing, “Export Expansion, Skill Acqusition and Industry Specialization: Evidence from China,” *Journal of International Economics*, 2018, *114*, 346–361.

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

Loecker, Jan De, Jan Eeckhout, and Gabriel Unger, "The Rise of Market Power and the Macroeconomic Implications," *Quarterly Journal of Economics*, 2020, *135* (2), 561–644.

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.

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.

Moriguchi, Chiaki and Emmanuel Saez, "The Evolution of Income Concentration in Japan, 1886–2005: Evidence from Income Tax Statistics," *Review of Economics and Statistics*, 2008, *90* (4), 713–734.

Mussa, Michael, "Dynamic Adjustment in the Heckscher-Ohlin-Samuelson Model," *Journal of Political Economy*, 1978, *86* (5), 775–791.

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

Oniki, H. and H. Uzawa, "Patterns of Trade and Investment in a Dynamic Model of International Trade," *Review of Economic Studies*, 1965, *32* (1), 15–38.

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.

Piketty, Thomas, "Income Inequality in France, 1901–1998," *Quarterly Journal of Economics*, 2003, *111* (5), 1004–1042.

_ , "On the Long-Run Evolution of Inheritance: France 1820–2050," *Quarterly Journal of Economics*, 2011, *126* (3), 1071–1131.

_ and Emmanuel Saez, "Income Inequality in the United States, 1913-1998," *Quarterly Journal of Economics*, 2003, *118* (1), 1–41.

_ and Gabriel Zucman, "Capital is Back: Wealth-Income Ratios in Rich Countries 1700–2010," *Quarterly Journal of Economics*, 2014, *129* (3), 1255–1310.

_ , Emmanuel Saez, and Gabriel Zucman, "Distributional National Accounts: Methods and Estimates for the United States," *Quarterly Journal of Economics*, 2018, *133* (2), 553–609.

_ , Li Yang, and Zucman Gabriel, "Capital Accumulation, Private Property, and Rising Inequality in China, 1978–2015," *American Economic Review*, 2019, *109* (7), 2469–2496.

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.

Saez, Emmanuel and Gabriel Zucman, "Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data," *Quarterly Journal of Economics*, 2016, *131* (2), 519–578.

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.

Simonovska, Ina and Michael E. Waugh, "Trade Models, Trade Elasticities, and the Gains from Trade," *Working Paper*, 2014.

Stiglitz, Joseph E., "Factor Price Equalization in a Dynamic Economy," *Journal of Political Economy*, 1970, *78* (3), 456–488.

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

Traiberman, Sharon, "Occupations and Import Competition: Evidence from Denmark," *American Economic Review*, 2019, *109* (12), 4260–4301.

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.

# 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 Database (IFPQD) 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 and the IFPQD 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 **??**. Lastly, we use the plant-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 the IFPQD database 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 et al. (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 IFPQD dataset we used covers a similar universe of manufacturing firms from 2000 to 2007. This dataset provides 5-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 IFPQD 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 2-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 20 .

## 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 $\bar{h}$,

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

Table 20: 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 |

where $v$ is a random utility component that is drawn from a Frechet distribution with a shape parameter $\theta$ 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 et al., 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 et al., 2019) and yield similar conclusions.

Given these assumptions, we can show that the indirect utility of a $\zeta$-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 $w_\zeta$ that choose to live in the downtown area as,

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

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

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

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

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

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

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

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] \\
&- \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] \\
&- \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] \\
=& \Bigg\{\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]\Bigg[\bar{h} + \frac{(1-\alpha)(\ldots}{p\ldots} \\
&\times \frac{1}{P^\alpha}\frac{1}{(p_H^S)^{1-\alpha}} \\
=& \Bigg\{\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]\Bigg[\bar{h} + \frac{(1-\alpha)w\ldots}{p_H^S} \\
&\times \frac{1}{P^\alpha}\frac{1}{(p_H^S)^{1-\alpha}} \\
=& \Bigg\{\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]\Bigg[\frac{(1-\alpha)w_u}{p_H^S} + \ldots \\
&\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}} \\
&+ \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\ldots} \\
=& \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

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

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

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

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 $\lambda_\zeta$ 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,

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

Hence, the second component in our targeted expression evaluates to

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

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.

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

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\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\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-}\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\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-}\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\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{\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] \\
=& \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

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

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}\Bigg[\frac{\partial\chi_u}{\partial L} w_u^{1-\sigma_L} + \lambda\frac{\partial\chi_s}{\partial L} w_s^{1-\sigma} \\
=& \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-} \\
=& \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 \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 $\varphi$. Consider the case where $z_H > z_L$ and $L_u^H > L_u^L$. By the strict log-supermodularity of $\pi$, 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

$$
\begin{aligned}
L_s\left[(1-\alpha)\frac{w_s - p_H^D\bar{h}}{p_H^D} + \bar{h}\right] &= \left[\frac{p_H^D}{w_u}\right]^{\frac{1-h}{h}} \\
L_s w_u^{\frac{1-h}{h}}\left[(1-\alpha)w_s + \alpha\bar{h}p_H^D\right] &= (p_H^D)^{\frac{1}{h}} \\
(p_H^D)^{\frac{1}{h}} - \alpha\bar{h}L_s w_u^{\frac{1-h}{h}} p_H^D &= (1-\alpha)L_s w_u^{\frac{1-h}{h}} w_s. \qquad (13)
\end{aligned}
$$

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

$$
(p_H^S)^{\frac{1}{h}} - \alpha\bar{h}L_u w_u^{\frac{1-h}{h}} p_H^S = (1-\alpha)L_u w_u^{\frac{1}{h}}. \qquad (14)
$$

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

$$
\Gamma_s(p_H^D)^{1-\alpha} + p_H^D\bar{h} = w_s, \qquad (15)
$$

$$
\Gamma_u(p_H^S)^{1-\alpha} + p_H^S\bar{h} = w_u, \qquad (16)
$$

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 (13), (14), (15) and (16) 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 (16) into (14) to replace $w_u$ yields the following non-linear equation to that pins down $p_H^S$,

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

Given $p_H^S$, we can immediately compute unskilled worker wages according to labor mobility condition (16). Plugging $w_u$ and equation (15) into (13) 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 (15).

### 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

$$
\begin{aligned}
\min_{\ell,k} \quad & \tilde{r}k + \tilde{w}\ell(q,\varphi) \\
\text{subject to} \quad & y_j(z) \leqslant k^{\gamma_j} \ell(q,\varphi)^{1-\gamma_j}
\end{aligned}
$$

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

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

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. $\ell_s$ and $\ell_u$ and solve for $\ell_s$

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

Substituting $\ell_s$ 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 $\ell$ 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 $\ell$ 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 $\tilde{\gamma}_j = (1-\gamma_j)^{\gamma_j-1}\gamma_j^{-\gamma_j}$, and $\tilde{w}(q,\varphi,L_s,L_u) = \left[\chi_u(q,\varphi)w_u(L_s,L_u)^{1-\sigma_L} + \lambda\chi_s(q,\varphi)w_s(L_s,L_u)^{1-\sigma}\right.$

## A.4 Model Fit

## A.5 Sensitivity Analysis

Figure 19: Firm size (revenue) distribution, sector by sector

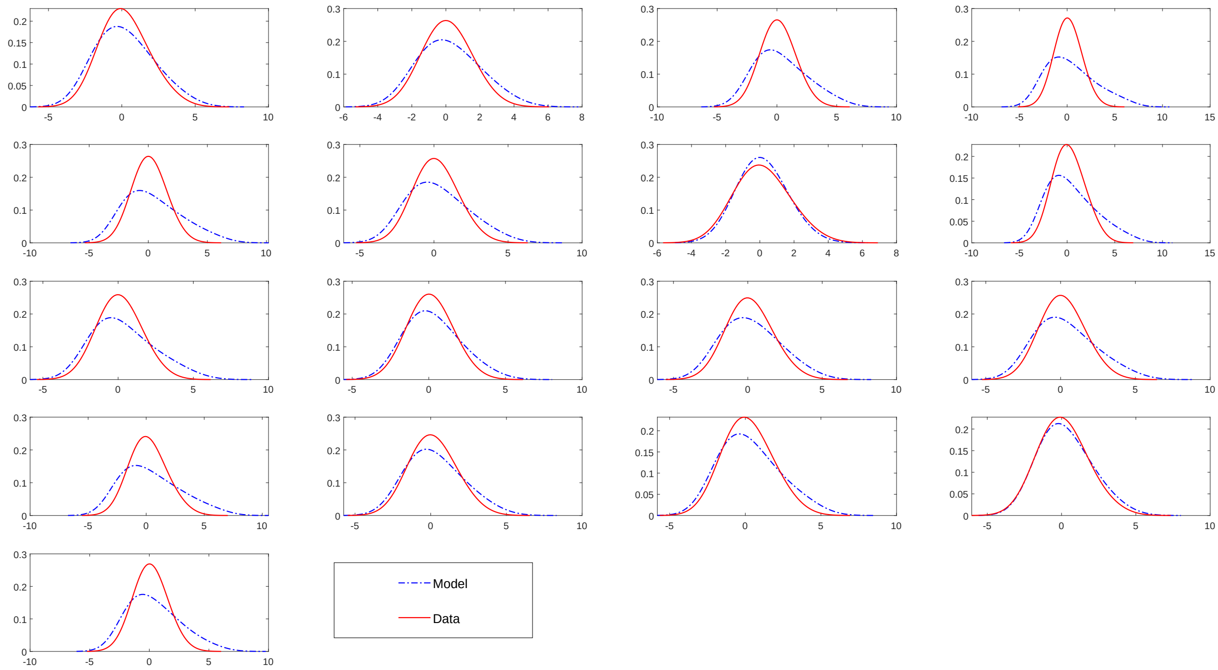


Figure 20: Share of value added, sector by sector

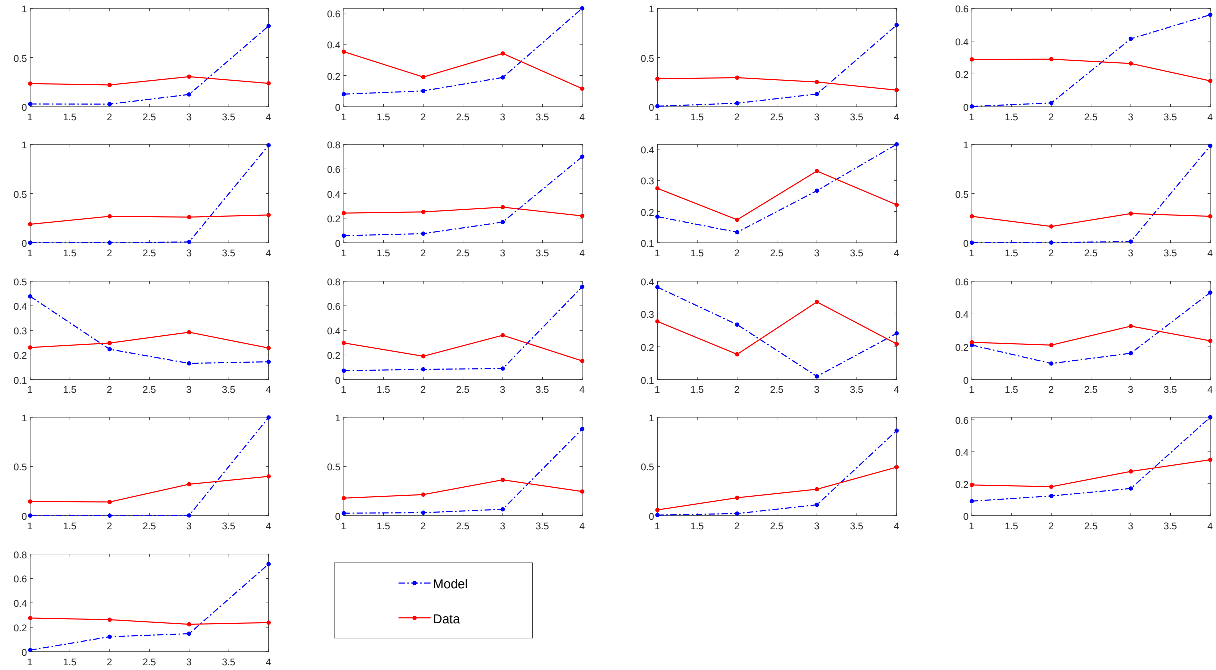


Figure 21: Average value added, sector by sector

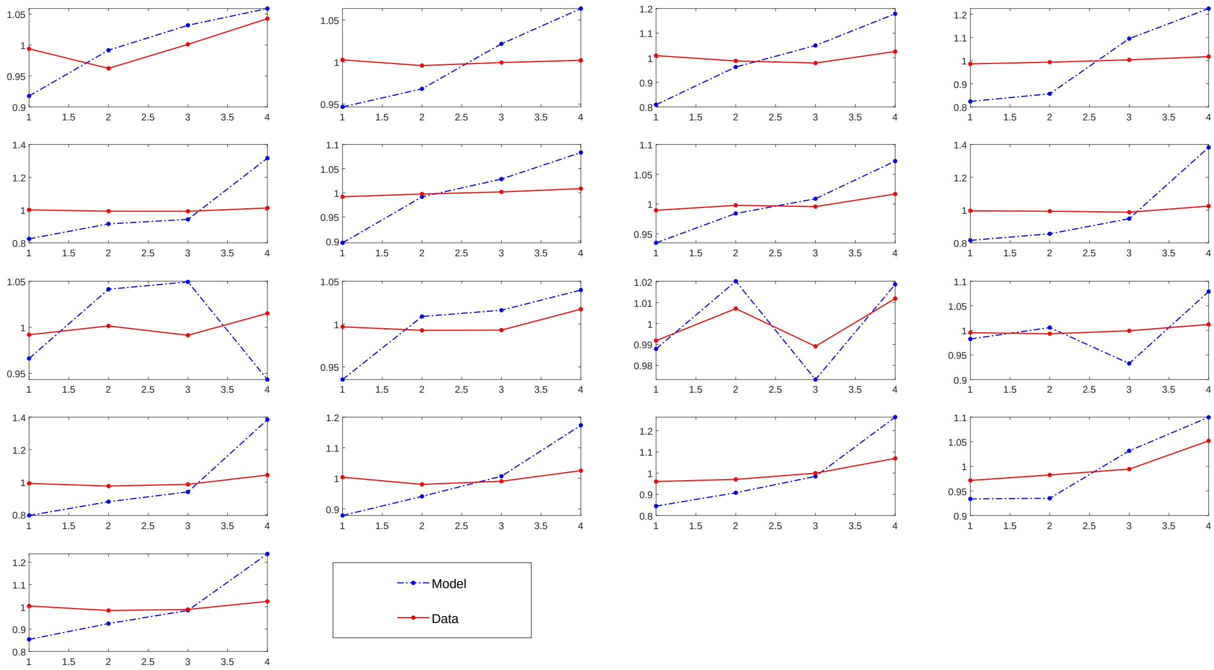


Figure 22: Average skill intensity, sector by sector

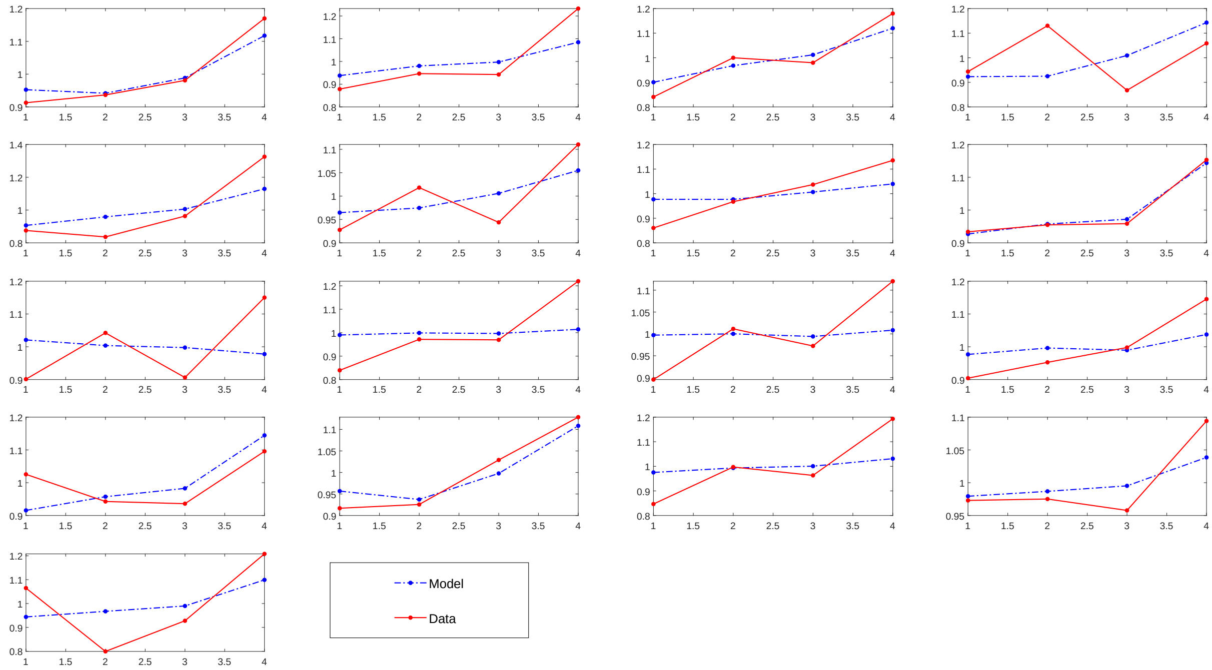
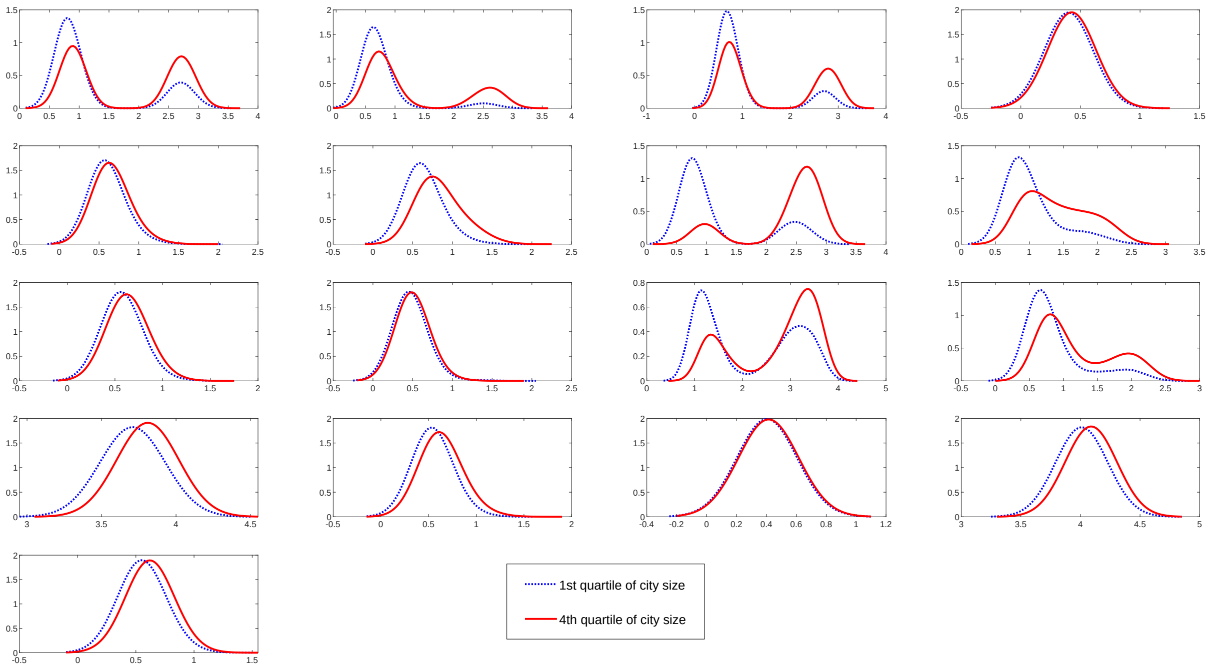


Figure 23: Quality distribution in big vs small cities, alternative weighting matrix

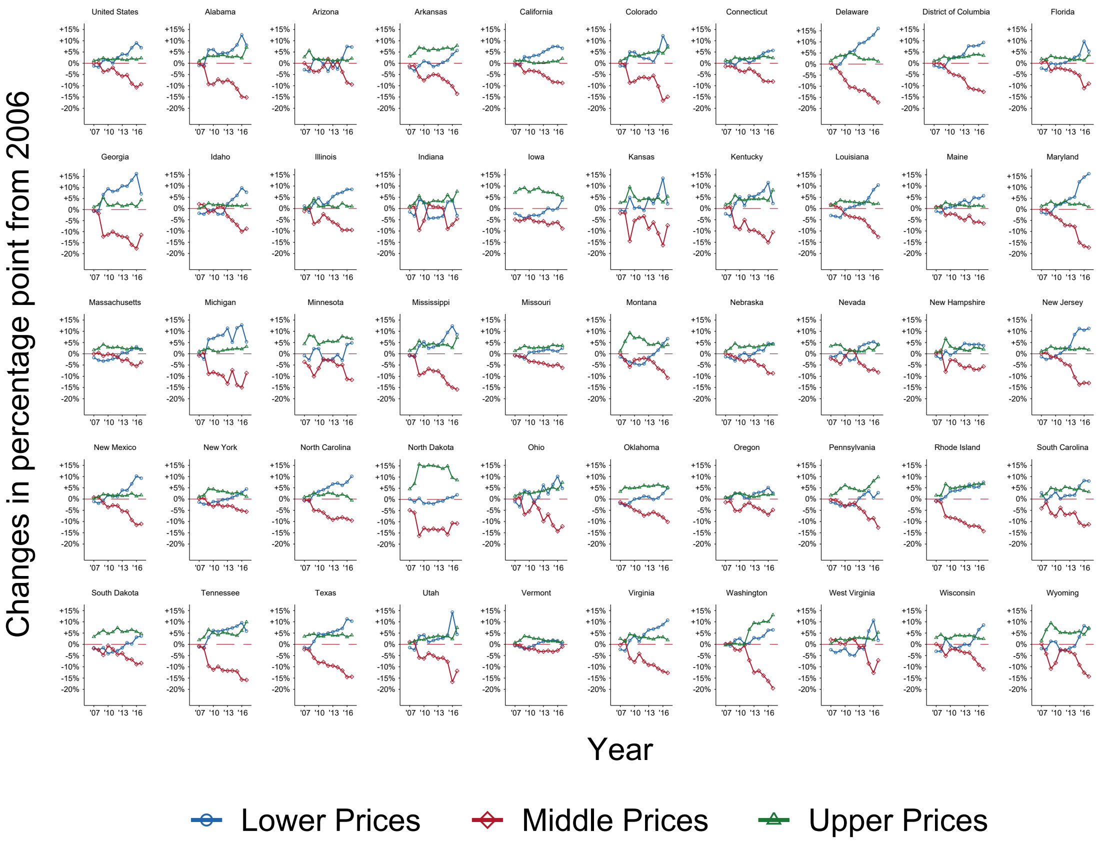


# B Appendix to Chapter 2

## B.1 Robustness Checks: Missing Middle in Price Distribution

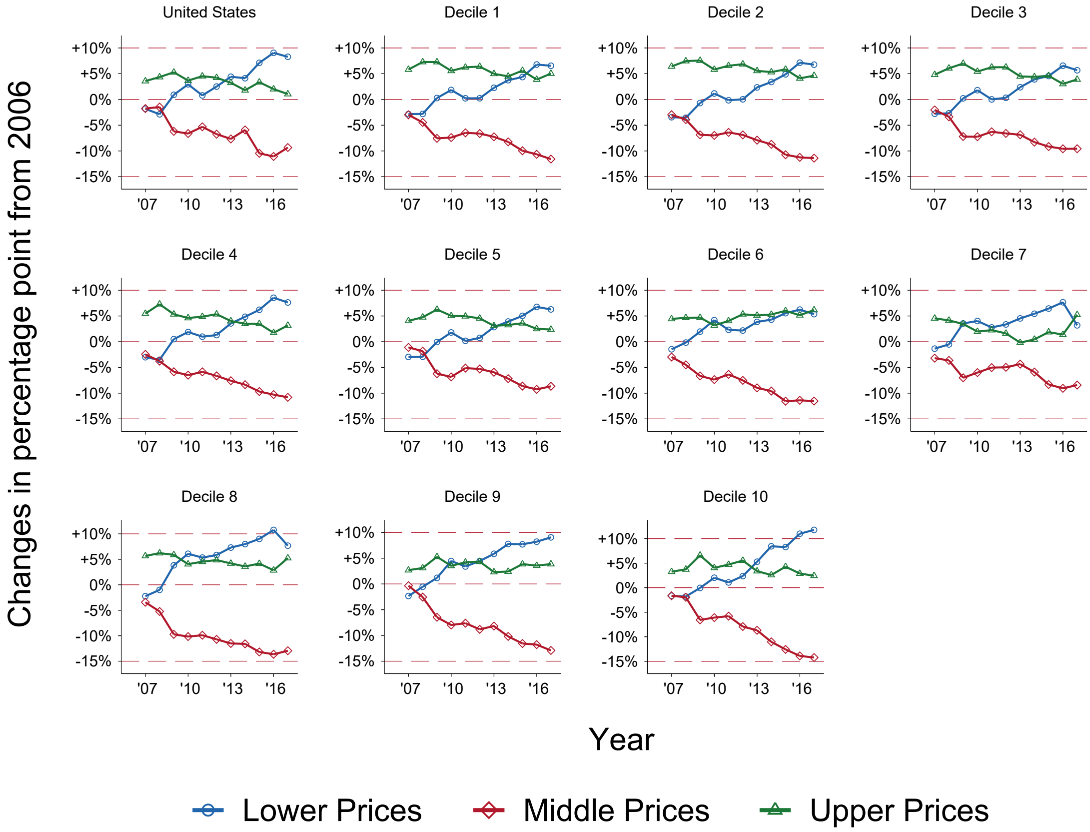
This section reports the results when we compute the market share measures using other thresholds such as 90th/10th percentile and 66th/33rd percentile.

Figure 24: Cumulative Change of Market Shares in Contiguous States



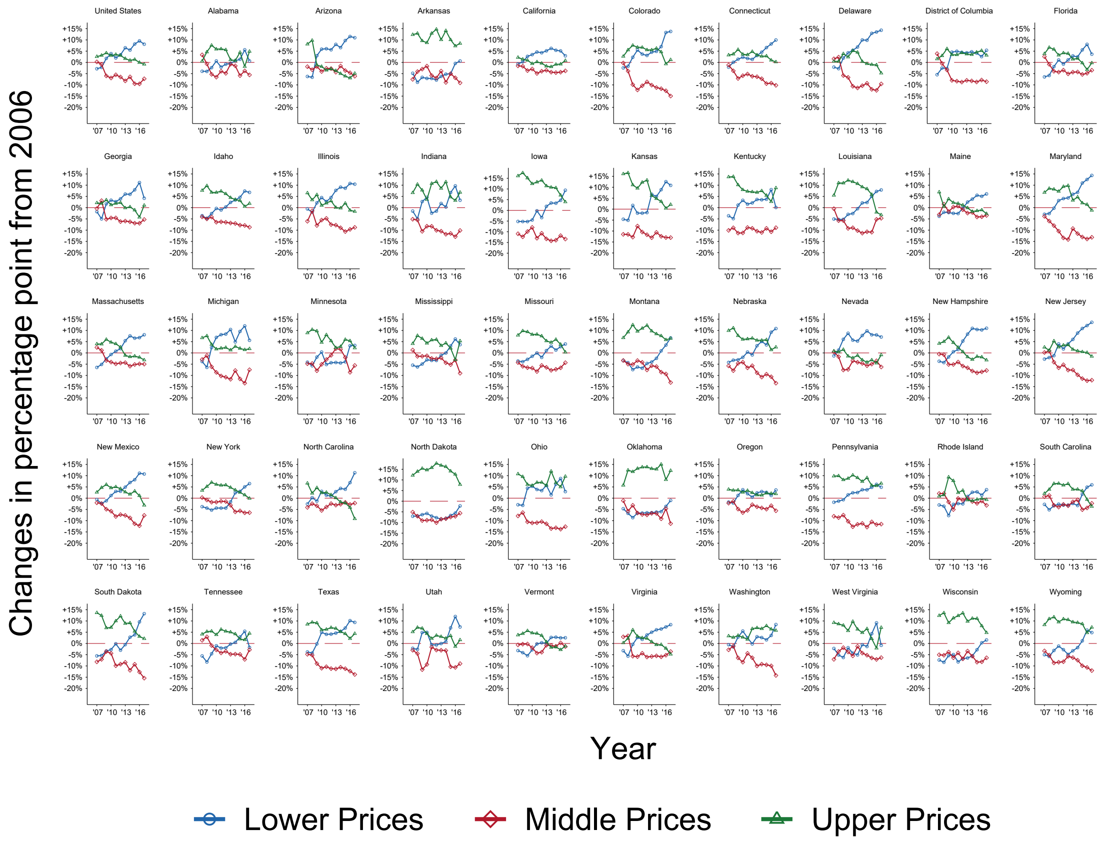

*Notes:* Each plotted point represents the cumulative change of market share in percentage points of a particular price range in the contiguous states of US from 2006 to that year. The market share measures are computed using 90th/10th percentile of the price distribution as thresholds

Figure 25: Cumulative Change of Market Shares in Commuting Zones




*Notes:* Each plotted point represents the cumulative change of market share in percentage points of a particular price range in the contiguous commuting zones of US from 2006 to that year. The market share measures are computed using 90th/10th percentile of the price distribution as thresholds

Figure 26: Cumulative Change of Market Shares in Contiguous States




*Notes:* Each plotted point represents the cumulative change of market share in percentage points of a particular price range in the contiguous commuting zones of US from 2006 to that year. The market share measures are computed using 66th/33rd percentile of the price distribution as thresholds

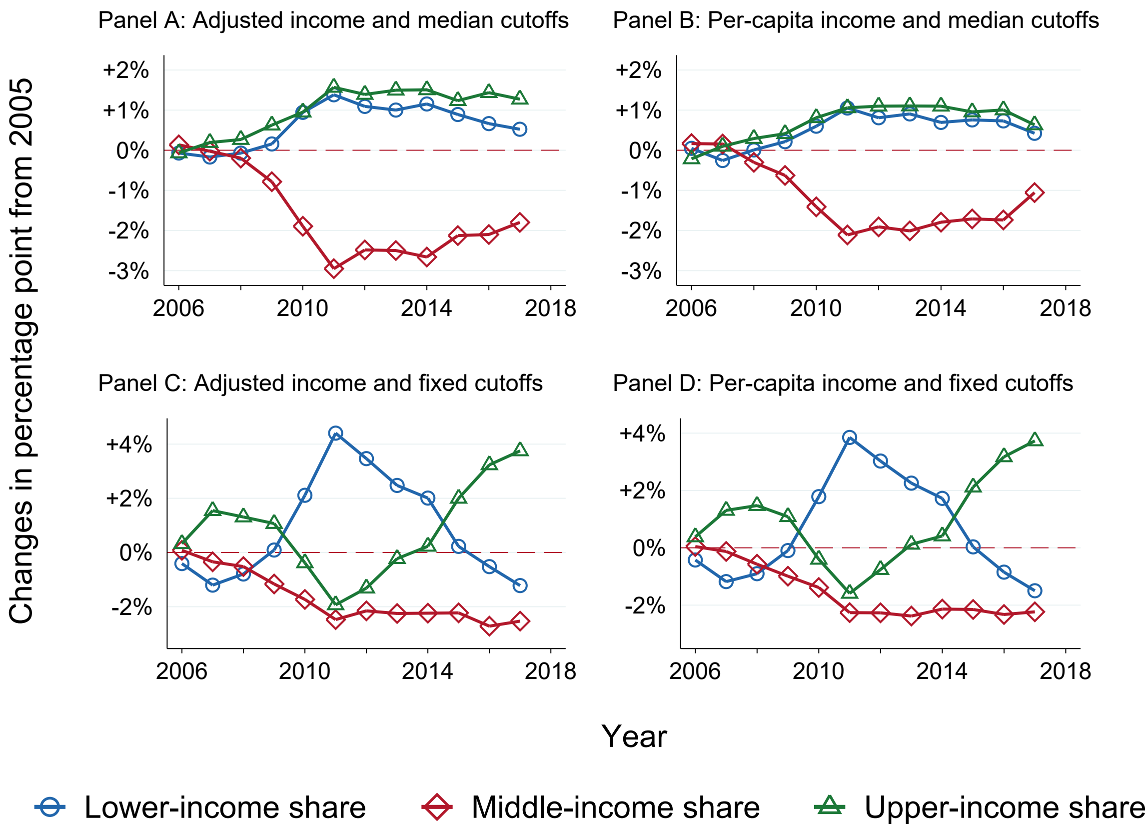
Figure 27: Cumulative Change of Market Shares in Commuting Zones


*Notes:* Each plotted point represents the cumulative change of market share in percentage points of a particular price range in the contiguous commuting zones of US from 2006 to that year. The market share measures are computed using 66th/33rd percentile of the price distribution as thresholds

## B.2 Robustness Checks: Missing Middle in Income Distribution

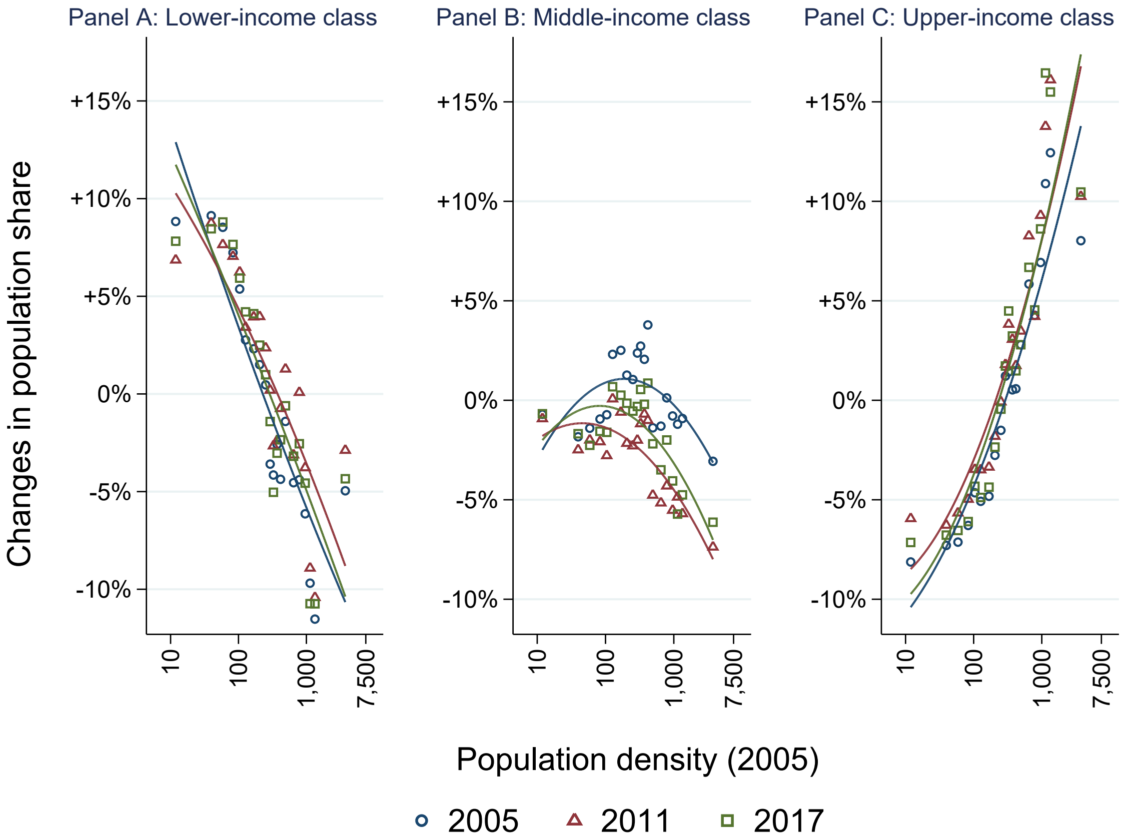
This section

Figure 28: Changes in Population Shares from 2005



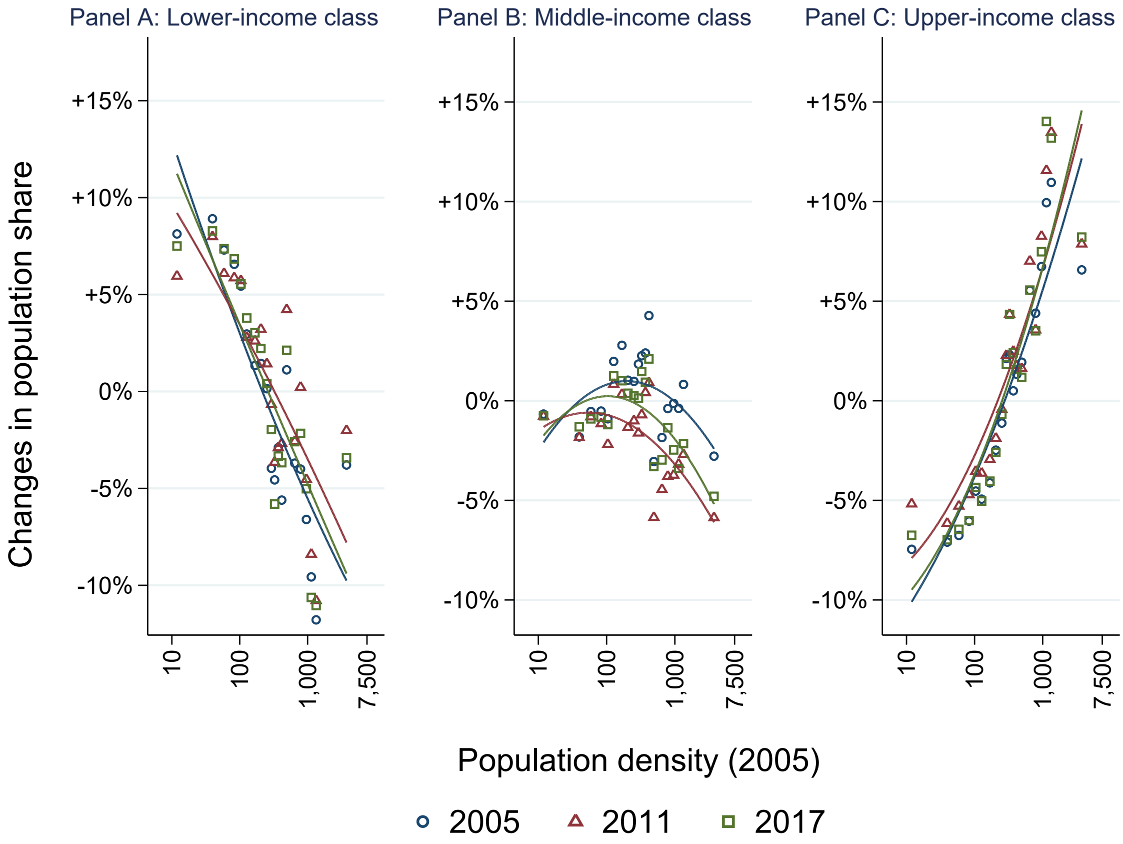

*Notes:* Each plotted point represents the cumulative change of population share in percentage points of a particular income category from 2005 to that year.

Figure 29: Changes in Population Shares from 2005




*Notes:* Each plotted point represents the difference between the bin cumulative change of population share in percentage points of a particular income category from 2005 to that year and that of the national average.

Figure 30: Changes in Population Shares from 2005




*Notes:* Each plotted point represents the difference between the bin cumulative change of population share in percentage points of a particular income category from 2005 to that year and that of the national average.

## B.3 Robustness Checks: Estimation Specifications

Table 21: Market Shares of Different Price Range: Commuting Zones

| Dep. Var. | Below33 | Middle | Above66 | Below33 | Middle | Above66 |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Income Share | 0.181*** | 0.024*** | 0.471*** | 0.170*** | 0.027*** | 0.462*** |
| | (0.011) | (0.010) | (0.018) | (0.010) | (0.010) | (0.015) |
| Share×Quintile 2 | -0.180*** | 0.134*** | -0.011 | -0.181*** | 0.137*** | -0.047 |
| | (0.029) | (0.026) | (0.048) | (0.023) | (0.023) | (0.035) |
| Share×Quintile 3 | -0.240*** | 0.299*** | 0.259*** | -0.220*** | 0.287*** | 0.245*** |
| | (0.048) | (0.041) | (0.077) | (0.039) | (0.036) | (0.059) |
| Share×Quintile 4 | -0.337*** | 0.387*** | 0.246* | -0.317*** | 0.375*** | 0.234** |
| | (0.080) | (0.058) | (0.128) | (0.064) | (0.052) | (0.095) |
| Share×Quintile 5 | -0.464*** | 0.393*** | 0.087 | -0.445*** | 0.378*** | 0.104 |
| | (0.109) | (0.060) | (0.107) | (0.091) | (0.055) | (0.079) |
| Robust SE | Yes | Yes | Yes | Yes | Yes | Yes |
| CZone FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Product FE | No | No | No | Yes | Yes | Yes |
| $R^2$ | 0.261 | 0.088 | 0.115 | 0.502 | 0.279 | 0.464 |
| $N$ | 653,068 | 653,068 | 653,068 | 653,068 | 653,068 | 653,068 |

*Notes:* Market shares measures are constructed from Nielsen Retail Scanner Data. Population share measures are constructed from IPUMS ACS Data. We include all the contiguous commuting zones in the United States whenever data is available. Time is from 2006 to 2017. Population density is computed based on population and land area data from NBER and US Census Bureau. All regressions include a constant term. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 22: Market Shares of Different Price Range: Commuting Zones

| Dep. Var. | Below25 | Middle | Above75 | Below25 | Middle | Above75 |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Income Share | 0.164*** | 0.046*** | 0.456*** | 0.153*** | 0.052*** | 0.451*** |
| | (0.011) | (0.012) | (0.017) | (0.010) | (0.011) | (0.014) |
| Share×Quintile 2 | -0.157*** | 0.169*** | -0.012 | -0.156*** | 0.172*** | -0.051 |
| | (0.027) | (0.029) | (0.046) | (0.022) | (0.024) | (0.033) |
| Share×Quintile 3 | -0.177*** | 0.405*** | 0.269*** | -0.160*** | 0.388*** | 0.254*** |
| | (0.045) | (0.045) | (0.073) | (0.022) | (0.038) | (0.053) |
| Share×Quintile 4 | -0.248*** | 0.461*** | 0.263** | -0.232*** | 0.443*** | 0.248*** |
| | (0.074) | (0.064) | (0.123) | (0.060) | (0.055) | (0.089) |
| Share×Quintile 5 | -0.319*** | 0.472*** | 0.112 | -0.306*** | 0.450*** | 0.125* |
| | (0.099) | (0.064) | (0.099) | (0.083) | (0.058) | (0.072) |
| Robust SE | Yes | Yes | Yes | Yes | Yes | Yes |
| CZone FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Product FE | No | No | No | Yes | Yes | Yes |
| $R^2$ | 0.321 | 0.150 | 0.091 | 0.528 | 0.373 | 0.467 |
| $N$ | 653,068 | 653,068 | 653,068 | 653,068 | 653,068 | 653,068 |

*Notes:* Market shares measures are constructed from Nielsen Retail Scanner Data. Population share measures are constructed from IPUMS ACS Data. We include all the contiguous commuting zones in the United States whenever data is available. Time is from 2006 to 2017. Population density is computed based on population and land area data from NBER and US Census Bureau. All regressions include a constant term. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 23: Market Shares of Different Price Range: Commuting Zones

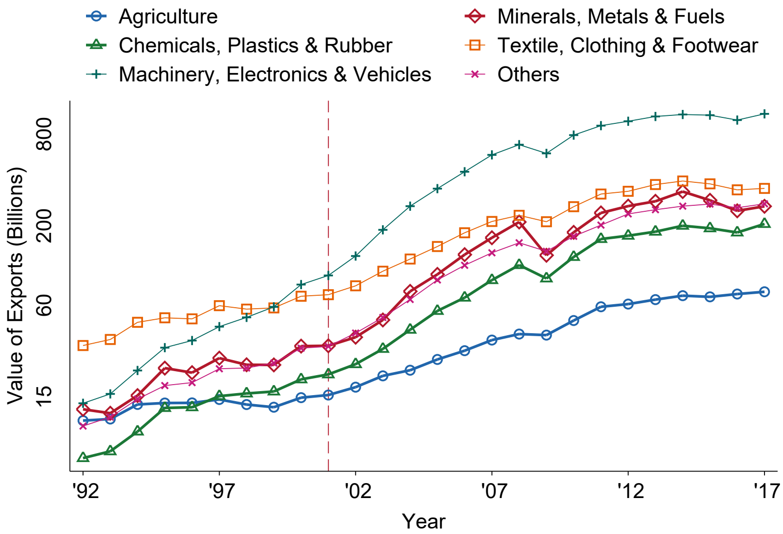
| Dep. Var. | Below10 | Middle | Above90 | Below10 | Middle | Above90 |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Income Share | 0.088*** | 0.040*** | 0.326*** | 0.081*** | 0.049*** | 0.327*** |
| | (0.009) | (0.012) | (0.013) | (0.008) | (0.010) | (0.011) |
| Share×Quintile 2 | -0.052*** | 0.146*** | -0.058 | -0.051** | 0.149*** | -0.094*** |
| | (0.022) | (0.028) | (0.036) | (0.019) | (0.021) | (0.022) |
| Share×Quintile 3 | -0.035*** | 0.348*** | 0.143** | -0.025 | 0.333*** | 0.149*** |
| | (0.034) | (0.044) | (0.059) | (0.029) | (0.033) | (0.035) |
| Share×Quintile 4 | -0.027*** | 0.355*** | 0.100 | -0.021 | 0.342*** | 0.102* |
| | (0.058) | (0.066) | (0.102) | (0.052) | (0.049) | (0.058) |
| Share×Quintile 5 | -0.041*** | 0.427*** | 0.072 | -0.042 | 0.411*** | 0.105** |
| | (0.073) | (0.064) | (0.080) | (0.065) | (0.048) | (0.046) |
| Robust SE | Yes | Yes | Yes | Yes | Yes | Yes |
| CZone FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Product FE | No | No | No | Yes | Yes | Yes |
| $R^2$ | 0.486 | 0.327 | 0.062 | 0.598 | 0.543 | 0.506 |
| $N$ | 653,068 | 653,068 | 653,068 | 653,068 | 653,068 | 653,068 |

*Notes:* Market shares measures are constructed from Nielsen Retail Scanner Data. Population share measures are constructed from IPUMS ACS Data. We include all the contiguous commuting zones in the United States whenever data is available. Time is from 2006 to 2017. Population density is computed based on population and land area data from NBER and US Census Bureau. All regressions include a constant term. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

# C Appendix to Chapter 3

## C.1 Trade Dynamics of China

Figure 31: Time-Series of Chinese Exports, 1992–2017




*Notes:* HS 2-digit trade flow data are collected from Comtrade. The aggregation is done using similar classifications as those in the WITS database. The y-axis is in log scale.

## C.2 Solution to Workers' Dynamic Problem

### C.2.1 Emax Operator

We solve the occupational choice problem following the conventional backward recursion method as in Keane and Wolpin (1994, 1997). First, we can write the worker's problem in recursive form as follows,

$$\begin{aligned} E_{a,t}\left[\max_{d_a^s} V_{a,t}^s(x_{a,t}^s, \epsilon_{d,a}^s)\right] = \max_{d_a^s} \quad & u(x_{a,t}^s, d_a^s) - \lambda(d_a^s, d_{a-1}^s) + \epsilon_{d,a}^s \\ & + \beta E_{a,t}\left[\max_{d_{a+1}^s} V_{a+1,t+1}^s(x_{a+1,t+1}^s, \epsilon_{d,a+1}^s)\right] \end{aligned} \tag{17}$$

For simplicity reason, I will suppress the notation for individual $s$. Solving the problem from backwards, at terminal period of an individual professional career $a = T$ and $t = t'$

for some $t'$, the value function then becomes

$$V_{T,t'}(x_{T,t'}) = \max_{d_T} \; u(x_{T,t'}, d_T) - \lambda(d_T, d_{T-1}) + \epsilon_{d,T}.$$

Next, we define the "Emax" operator as those in Keane and Wolpin (1994, 1997) as

$$\text{Emax}(x_{a,t}, \epsilon_{a,t}) \equiv E[V_{a,t}(x_{a,t}, \epsilon_{d,t})].$$

Then, for an individual at his terminal period $T$ and calendar time $t = t'$ for some $t'$, the Emax operator is such that

$$\begin{aligned}\text{Emax}(x_{T,t'}, \epsilon_{T,t'}) &\equiv E[V_{T,t'}(x_{T,t'}, \epsilon_{d,T})] \\ &= E[\max_{d_T} u(x_{T,t'}, d_T) - \lambda(d_T, d_{T-1}) + \epsilon_{d,T}].\end{aligned}$$

Substituting this expression into the previous period where $a = T - 1$, the Emax operator becomes

$$\begin{aligned}\text{Emax}(x_{T-1,t'-1}, \epsilon_{T-1,t'-1}) &\equiv E[V_{T-1,t'-1}(x_{T-1,t'-1}, \epsilon_{d,T-1})] \\ &= E\left[\max_{d_{T-1}} \; \tilde{u}_{T-1} + \beta \text{Emax}(x_{T,t'}, \epsilon_{d,T})\right]\end{aligned}$$

where $\tilde{u}_{T-1}$ is a short notation for the flow utility

$$u(x_{T-1,t'-1}, d_{T-1}) - \lambda(d_{T-1}, d_{T-2}) + \epsilon_{d,T-1}.$$

Iterating this logic backward, we can solve the decision problem in every period given the idiosyncratic shocks.

### C.2.2 Interpolation

In addition, we follow Keane and Wolpin (1994) to handle two computational issues. First, we circumvent the multiple numerical integrals associated with Emax operator using Monte Carlo integration. Second, we reduce the dimensionality of the state space by using the interpolating regressions where necessary (code available upon request for an extended model with larger state space).