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

Pao-Li Chang† Angdi Lu‡ Xin Yi§

October 30, 2019

Click Here to Download Latest Version

Abstract

We quantify the supply-side determinants of quality specialization across space. Specifically, we complement the quality specialization literature in international trade and study how larger cities specialize in higher-quality goods within a country. In our general equilibrium model, firms in larger cities produce goods with higher quality, because agglomeration benefits accrue more to skilled workers who are also more efficient in upgrading quality. Two channels are at work in our model. The first channel is through the treatment effect of agglomeration, such that firms become more productive if they locate in a larger city. The second channel works through sorting, in that more productive firms receive higher agglomeration benefits and endogenously sort into larger cities. These two effects are further mitigated by the increasing skill premium with respect to city size, though the latter is dominated in the spatial equilibrium. Using firm-level data from China, we structurally estimate the model and find that product quality is on average 23% higher in big cities than that of small cities. We further find that agglomeration forces account for half of the quality difference in big cities while sorting of firms accounts for another half. A counterfactual policy to relax land use regulation in housing production raises the quality of goods produced in big cities by 5.5% and (indirect) welfare of all residents by 6.2% through reallocation of economic activities across space.

Keywords: Agglomeration, Quality Upgrading, Firm Heterogeneity, Sorting

JEL Codes: D22, F12, R12, R32

*The paper is currently under major revision. A new draft with updated estimates and more results will be provided soon. For their constructive discussions, we thank Yao Amber Li, Zhao Chen, Yuan Mei, Xin Wang, Yong Wang, Lianming Zhu, and participants at the Asia Pacific Trade Seminars 2019, BCCDS 2019, SER Conference 2019, and IOG Conference 2019. All errors are ours.

†Associate Professor, School of Economics, Singapore Management University. Address: 90 Stamford Road, Singapore 178903. Email: plchang@smu.edu.sg. Tel: +65-68280830. Fax: +65-68280833.

‡PhD Student, School of Economics, Singapore Management University. Email: angdi.lu.2015@phdecons.smu.edu.sg

§PhD Student, School of Economics, Singapore Management University. Email: xinyi.2015@phdecons.smu.edu.sg

1. Introduction

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

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

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

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

tion 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.

2. Related Literature

The present study is related to several strands of literature in urban economics and international trade. First, our work is related to the spatial literature on the benefits of agglomeration (Davis and Dingel, 2019; Gaubert, 2018; Tian, 2018; Handbury and Weinstein, 2015; Behrens 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, 2013; 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.

3. The Model

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

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

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

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

3.3 Housing Sector and Wage Premium

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

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

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

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

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

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

$$L_s \left[(1-\alpha)\frac{w_s(L_s, L_u) - p_H^D(L_s, L_u)\bar{h}}{p_H^D(L_s, L_u)} + \bar{h}\right] = \left[\frac{p_H^D(L_s, L_u)}{w_u(L_s, L_u)}\right]^{\frac{1-h}{h}}, \tag{3.1}$$

$$L_u \left[(1-\alpha)\frac{w_u(L_s, L_u) - p_H^S(L_s, L_u)\bar{h}}{p_H^S(L_s, L_u)} + \bar{h}\right] = \left[\frac{p_H^S(L_s, L_u)}{w_u(L_s, L_u)}\right]^{\frac{1-h}{h}}. \tag{3.2}$$

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

$$\bar{U}_s = \left(\frac{w_s - p_H^D \bar{h}}{P}\right)^{\alpha} \left(\frac{w_s - p_H^D \bar{h}}{p_H^D}\right)^{1-\alpha}, \tag{3.3}$$

$$\bar{U}_u = \left(\frac{w_u - p_H^S \bar{h}}{P}\right)^{\alpha} \left(\frac{w_u - p_H^S \bar{h}}{p_H^S}\right)^{1-\alpha}. \tag{3.4}$$

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 (3.1), (3.2), (3.3), and (3.4) 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.

3.4 Production and Quality

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

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

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

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

The interpretation of our specification of the production function is as follows. $\sigma_L > 1$ measures the degree of substitution between skilled labor and unskilled labor. λ denotes the relative importance of effective skilled labor in the production. $\chi_u(q, \varphi)$ and $\chi_s(q, \varphi)$ capture the productivity of workers in a firm of productivity $\varphi(z, L, q; s_j)$ to produce outputs with quality q. In particular, we assume that $\partial\chi_\zeta(q, \varphi)/\partial q < 0$ so that firms find it costly to upgrade product quality. We also assume $\partial\chi_\zeta(q, \varphi)/\partial\varphi > 0$ and $\chi_s(q, \varphi)/\chi_u(q, \varphi)$ is increasing in φ, so that more productive firms face lower marignal cost and also employ more skilled labor.[2] 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 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 φ, and decreasing in skill premium in the located city (L_s, L_u). Holding everything else constant and without considering the agglomeration effect on productivity, firms tend to choose a lower skill intensity in a larger city since skill premium is higher in big cites.

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

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

3.5 Productivity and Agglomeration

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

Assumption 1. *$\varphi(z, L, q; s_j)$ is strictly log-supermodular in the size of labor $L = L_s + L_u$ and firm innate efficiency z, and is twice differentiable such that*

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

Our assumption that φ is only related to the total labor size $L = L_s + L_u$ but not the skill composition is too strong and ad-hoc. However, we are only able to do this because there is no prior structural estimates on the traditional agglomeration parameters *and* the log-supermodular forces for *both* skilled and unskilled population size in the literature. In addition, we lack the city information to implement a proper structural estimation for these parameters. Nevertheless, we will also

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

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.

3.6 Entry and Location Choice

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

3.7 Firm's Problem

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

In practice, we will only be able to solve for the optimal quality choices numerically if $f_q = 0$. To see this, note that one must know all the general equilibrium quantities in order to solve for the individual optimal choices above. However, the general equilibrium quantities can only be known after solving for the individual choices. This poses an insurmountable computational burden. To circumvent this issue, we set $f_q \approx 0$ which is supported by the empirical estimate of 4.7×10^{-5} in Fieler 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 π with respect to q, we can assess the impact of changes in firm efficiency z on the quality choice q^*. Proposition 2 summarizes our findings.

Proposition 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,*

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

[Supply more discussions on the intuition.]

3.8 Firm Sorting to Cities

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

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

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

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

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

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 $L_s^{H*} \geq L_s^{L*}$ and $L_u^{H*} \geq L_u^{L*}$ 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 $z_H \geq z_L$. In addition, $\tilde{r}_j^*(z_H) \geq \tilde{r}_j^*(z_L)$ and $\pi_j^*(z_H) \geq \pi_j^*(z_L)$.*

3.9 General Equilibrium

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

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

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

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

4. Goods markets clear. That is in each sector j, aggregate demand is equal to the aggregate sectoral outputs
$$X_j = \sigma_j \Upsilon_j P_j^{\sigma_j - 1} X_j M_j \int_z \frac{\Phi_j(z,q)}{w(q,\varphi,L_s,L_u)^{(1-\gamma_j)(\sigma_j-1)}} dF_j(z).$$

5. Local labor markets clear. That is in each city (L_s, L_u), the markets for skilled and unskilled labor clear,
$$\int_{L_{0s}}^{L_s} n f_{L_s}(n) dn = \sum_{j=1}^{S} M_j \int_0^\infty \mathbf{1}_j(L_s, L_u, z) l_s(z) dF_j(z), \quad \forall L_s > L_{0s}.$$
$$\int_{L_{0u}}^{L_u} n f_{L_u}(n) dn = \sum_{j=1}^{S} M_j \int_0^\infty \mathbf{1}_j(L_s, L_u, z) l_u(z) dF_j(z), \quad \forall L_u > L_{0u}.$$

6. National labor markets for skilled and unskilled labor clear. That is,
$$\bar{L}_s = \sum_{j=1}^{S} \Upsilon_{2j} P_j^{\sigma_j - 1} X_j M_j \int_z \frac{\lambda \chi_s(q,\varphi) \Phi_j(z,q)}{w(q,\varphi,L_s,L_u)^{(\sigma_j-1)(1-\gamma_j)+1-\sigma_L} w_s^{\sigma_L}} dF_j(z).$$
$$\bar{L}_u = \sum_{j=1}^{S} \Upsilon_{2j} P_j^{\sigma_j - 1} X_j M_j \int_z \frac{\chi_u(q,\varphi) \Phi_j(z,q)}{w(q,\varphi,L_s,L_u)^{(\sigma_j-1)(1-\gamma_j)+1-\sigma_L} w_u^{\sigma_L}} dF_j(z) + \bar{L}_u(1-h)(1-\alpha).$$

7. Capital market clears by Walras's Law.

8. The ex-ante expected profit of a firm is zero in each sector j, due to free entry,
$$f_E P = \Upsilon_{1j} P_j^{\sigma_j - 1} X_j \int_z \frac{\Phi_j(z,q)}{w(q,\varphi,L_s,L_u)^{(1-\gamma_j)(\sigma_j-1)}} dF_j(z).$$

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

4. 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),

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

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 $\nu_{\phi,j}$.[3] [4] 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, λ_{1s} and λ_{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, λ_{2s} and λ_{2u} define how costly it is to produce higher-quality good using each type of labor. We expect the sign and magnitude of these parameters to be negative and that $\lambda_{2s} > \lambda_{2u}$. Given the exponential functional form, this implies that production of quality will reduce the productivity of workers, and this productivity-dampening effect is stronger for unskilled workers than for skilled workers. Intuitively, it takes longer time and more effort for workers to produce goods with higher quality, and this is more so for unskilled workers. We choose the exponential functional form because it would generate a skill intensity distribution that is close to the data, as similarly noted in Fieler et al. (2018).[5]

4.1 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.

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

[4]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.

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

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

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

$$1 = \sigma_j \Upsilon_{1j} P_j^{\sigma_j-1} M_j \tilde{S}_j, \text{ for all } j \in S,$$

$$\bar{N}_u = \sum_{j=1}^{S} \Upsilon_{2j} P_j^{\sigma_j-1} \beta_j X M_j \tilde{E}_{u,j} + \bar{N}_u(1-h)(1-\alpha),$$

$$\bar{N}_s = \sum_{j=1}^{S} \Upsilon_{2j} P_j^{\sigma_j-1} \beta_j X M_j \tilde{E}_{s,j},$$

$$f_E P = \Upsilon_{1j} P_j^{\sigma_j-1} \beta_j X \tilde{S}_j, \text{ for all } j \in S$$

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.

5. Quantifying the Model

5.1 Data

The dataset we use for our structural estimation is the Annual Survey of Industrial Firms collected by the National Bureau of Statistics of China (NBSC). In particular, we use the data in year 2004 in our baseline quantitative analysis. The universe of firms covered in this dataset spans over all manufacturing firms, which include both state and non-state enterprises, that generate more than 5 million RMB in revenue each year. The dataset reports information on the location, capital, output, taxation, revenue, and education level of the workers in each firm. All firms are codified in 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 1. The concordance of sectors from CSIC to our definition of sectors is detailed in Table A.1 in the appendix.

We obtain the geographic location of firms using postal code reported in the data. A "city" is defined as the prefecture 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 1. It is evident firms from sectors such as medical, machinery, transport and automotive, electrical, and computer are more likely to hire a high proportion skilled workers and also more likely to locate in big cities. To further substantiate this observation, we also plot the city size against the average skill intensity in the city in Figure 1. 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.

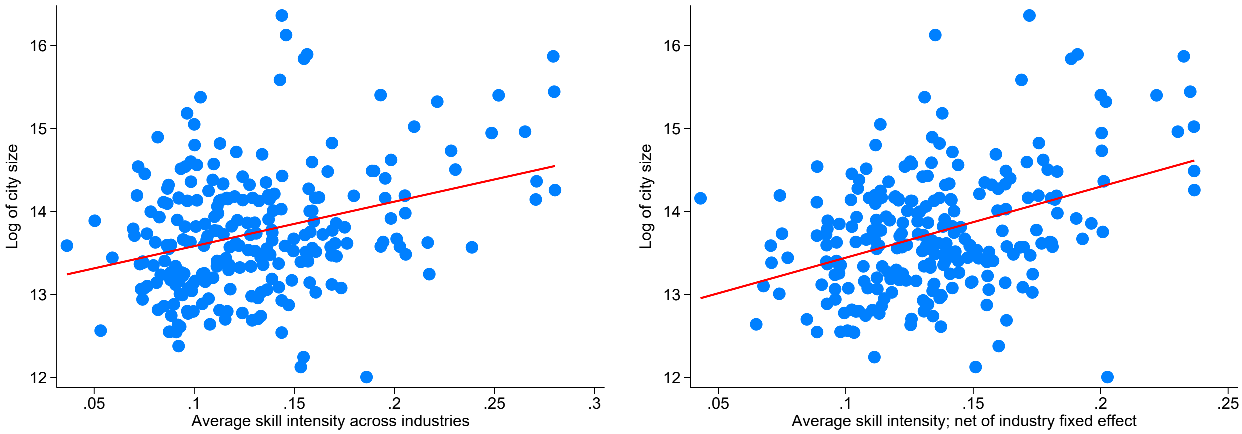
Table 1: Summary Statistics

	log Value Added			log Employment			% Skilled Workers				
Sector	Mean	$Q1$	$Q4$	Mean	$Q1$	$Q4$	Mean	$Q1$	$Q4$	% in Big Cities	N
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

5.2 Moments and Identification

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

Figure 1: Correlation of skill intensity and city size



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. The sensitivity of the moments to the change in parameters is reported in the Jacobian matrix in Table [Insert the no. here].

5.3 Structural Estimates

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

We now discuss the estimates for $\{\lambda_{1s,j}, \lambda_{1u,j}, \lambda_{2s,j}, \lambda_{2u,j}\}$ which is the set of parameters new in our model in comparison to the literature. Our estimates suggest that the productivity advantage of big cities is skill-biased, as the estimates are positive and $\hat{\lambda}_{1s}$ is greater than $\hat{\lambda}_{1u}$ in all sectors. This echos a strand of literature which argues that agglomeration forces benefit skilled workers more, for example because high-ability individuals learn better from idea exchange (Davis and Dingel, 2019). In particular, our estimates suggest that agglomeration disproportionately benefit skilled workers in medical and computer sectors, partly due to the fact that these sectors require extensive idea exchange among engineers and professionals. Finally, our estimates of $\hat{\lambda}_{2s}$ and $\hat{\lambda}_{2u}$ suggest that production of higher-quality good is costly and requires employing more labor, since both $\hat{\lambda}_{2s}$ and $\hat{\lambda}_{2u}$ are negative. Our estimates also imply that production of higher quality good is intensive in skilled labor, since $\hat{\lambda}_{2s} > \hat{\lambda}_{2u}$.[6]

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

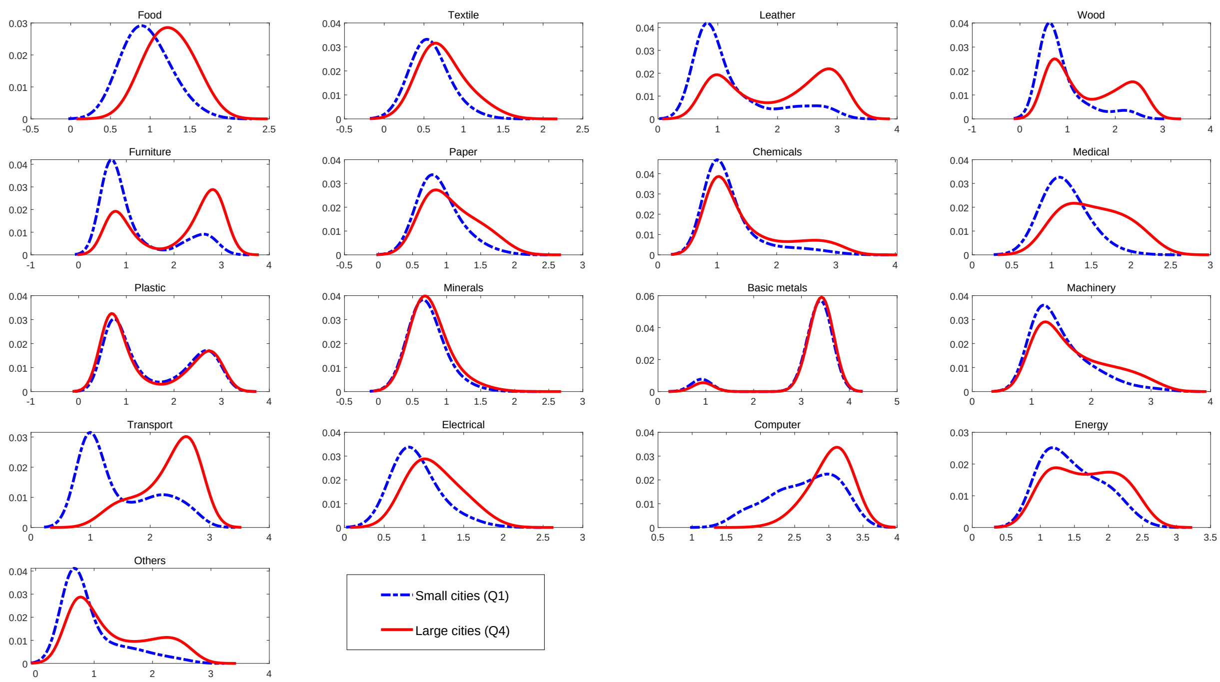
$$\frac{\ell_s^*(z, L_s, L_u)}{\ell_u^*(z, L_s, L_u)} = \lambda \frac{\chi_s(q, \varphi)}{\chi_u(q, \varphi)} \left[\frac{w_s(L_s, L_u)}{w_u(L_s, L_u)}\right]^{-\sigma_L} = \lambda \varphi^{\lambda_{1s} - \lambda_{1u}} e^{(\lambda_{2s} - \lambda_{2u})q} \left[\frac{w_s(L_s, L_u)}{w_u(L_s, L_u)}\right]^{-\sigma_L}.$$

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

5.4 Quantitative Results

We feed our parameter estimates into the model and extract average choice of product quality of the simulated firms by city quartiles. We find that our model generates significant quality differences across space. On average, product quality in the big cities (the 4th quartile) is 22.9% higher than that of the smallest cities (the 1st quartile). There is also significant sectoral heterogeneity in the quality specialization across space. For manufacturing sectors such as medical equipment, transport and automotive, food, and furniture, the average product quality difference between big and small cities can be as high as 27.4% to 59.5%. We report the entire distribution of quality choices across all firms located in different city quartiles in Figure 2.

Figure 2: Quality distribution in big vs. small cities, sector by sector



To further assess the contribution of firm sorting and traditional agglomeration benefit in determining the quality differences across space, we follow Gaubert (2018) and consider the following regression. We regress each simulated firm's choice of quality on the size of city that it locates in with industry fixed effect. We then repeat the exercise in a counterfactual where we shut down the sorting of firms by setting the efficiency of every firm to the average efficiency in the benchmark model and compute the reallocation of economic activities across space. The results are reported in Table 2. 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 2: 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

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

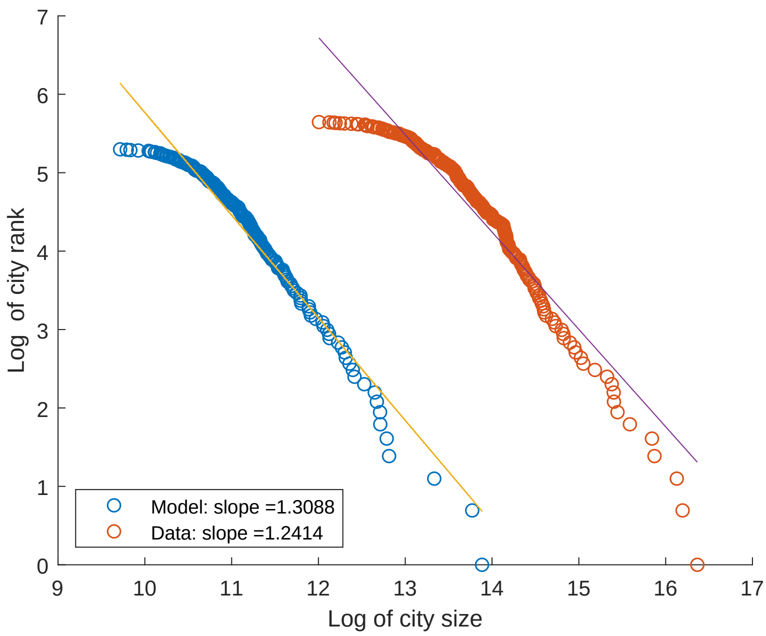
We first evaluate the fit of our model by comparing the simulated moments in the model to the empirical moments in the data. A summary of the results is reported in Table 3, where we aggregated moments across sectors. We also report the goodness of fit sector by sector in Appendix D. In general, our 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 3, 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 3: 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

Figure 3: City size distribution, model and data



6. Counterfactuals

We now evaluate the general equilibrium impact of a spatial policy that is frequently employed in developing economies such as China. The policies that we aim to evaluate are policies that regulate the use of land in a city such as zoning restrictions. Matching these policies to our model counterparts, land use regulation is approximated by the land use intensity in the production of housing. Whenever there are relatively few land use regulations, the land use intensity coefficient should be smaller as it is easier for developers to acquire land in their housing production. In the counterfactual, we shock the coefficient such that the coefficient for the high-end housing market becomes 20% smaller than the original value. The resulting changes are reported in Table 4. In overall, average quality across cities has increased by 5.5% while the aggregate welfare of all residents has increased by 20%.

Table 4: Counterfactual of relaxing land use regulations by 20%

	Change in variables (%)							
City quartile	Δq	Δw_s	Δw_u	Δp_H^D	Δp_H^S	ΔP	ΔW	$\Delta \tilde{W}$
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

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

7. Conclusion

In this paper, we study the pattern of quality specialization across Chinese cities through the lens of firm sorting and agglomeration. Extending the framework in Gaubert (2018) with two skill types and quality choices, we show theoretically both firm sorting and productivity advantage of agglomeration will induce quality upgrading. We structurally estimate and quantify the model using a plant-level dataset spanning the universe of manufacturing firms. We find that on average, product quality in big cities is 23% higher than that of small cities. A decomposition analysis shows that sorting and agglomeration each explains half of the quality pattern. Armed with the structural estimates, we then evaluate a potential policy that reduces land use regulations. We find that a 20% relaxation of land intensity induces a 5.5% increase in quality across cities and a 6.2% indirect welfare benefit. For future work, one could further quantify the relative magnitude of both the demand and supply-side explanations, as well as incorporating input-output linkages in the present model.

References

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.

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.

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

Baum-Snow, Nathaniel and Ronni Pavan, "Inequality and City Size," *Review of Economics and Statistics*, 2013, *95* (5), 1535–1548.

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.

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.

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.

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

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

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

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

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

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.*

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.

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

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

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

Fan, Haichao, 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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

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

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

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

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

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

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

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.

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

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

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

Appendices

A. Data and Measurement

A.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.

A.1.1 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.

A.1.2 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.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 A.1 .

Table A.1: Concordance of Sectors

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

B. A Microfoundation for Within-City Worker Sorting

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

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

where v is a random utility component that is drawn from a Frechet distribution with a shape parameter θ and a scale parameter normalized to 1. The budget constraint of a worker of type $\zeta \in \{s, u\}$ who lives in location $n \in \{\text{D}, \text{S}\}$ is $PC^n + p_H^n H^n = w_\zeta$. Without loss of generality, we can label the areas such that $p_H^D \geq p_H^S$. 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 ζ-type worker living in n is

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

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

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

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

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

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

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

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

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

C. Proofs and Derivations

C.1 Proof of Proposition 1

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

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

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

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

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

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

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

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

The partial derivatives can be computed as

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

$$\frac{\partial F_2}{\partial w_u} = \frac{1}{P^\alpha}\frac{1}{(p_H^S)^{1-\alpha}} > 0$$

And we further have

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

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

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

$$= \underbrace{\left[\frac{L_u(1-\alpha)}{p_H^S} + \frac{1-h}{h}\left(\frac{p_H^S}{w_u}\right)^{\frac{1-2h}{h}}\left(\frac{p_H^S}{(w_u)^2}\right)\right]}_{>0}\underbrace{\frac{\alpha(w_u - p_H^S\bar{h})}{p_H^S}}_{>0}\underbrace{\frac{1}{P^\alpha}\frac{1}{(p_H^S)^{1-\alpha}}}_{>0} > 0$$

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

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

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

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

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

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

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

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

which can be further evaluated to

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

and the relevant partial derivatives are

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

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

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

Writing it in matrix form, we have that

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

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

C.2 Proof of Proposition 2

We can rewrite the reduced first-order condition as

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

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

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

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

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

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

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

where individual components of this expression evaluate to the following.

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

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

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

and that

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

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

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

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

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

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

Hence, the second component in our targeted expression evaluates to

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

and the first component evaluates to

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

Therefore, the targeted expression evaluates to the following.

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

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

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

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

C.3 Proof of Proposition 3

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

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

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

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

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

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

which implies that

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

The individual components of this expression can be evaluated as

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

It follows that

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

and that

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

Hence, we can show that

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

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

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

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

Similarly,

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

It follows that

$$
\frac{1}{w}\frac{\partial w}{\partial q}\frac{\partial w}{\partial L}-\frac{\partial^2 w}{\partial L\partial q}=A+B-C-D
$$

$$
\begin{aligned}
=&\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,

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

C.4 Proof of Proposition 4

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

C.5 Expression for Computing Wages

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

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

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

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

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

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

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

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

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

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

Plugging equation (C.4) into (C.2) 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 (C.4). Plugging w_u and equation (C.3) into (C.1) yields the equation that implicitly determines housing price for skilled labor p_H^D,

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

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

C.6 Cost Function

Recall the production function of a firm is

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

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

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

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

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

The Lagrangian is

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

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

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

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

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

Solve for k and l in the expression of y,

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

The costs function can be expressed as

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

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

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

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

The Lagrangian is

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

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

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

Substituting ℓ_s into the constraint gives

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

The cost function for producing ℓ is such that

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

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

The cost of producing one unit of ℓ is

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

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

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

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

The cost function for production is

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

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

D. Model Fit

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

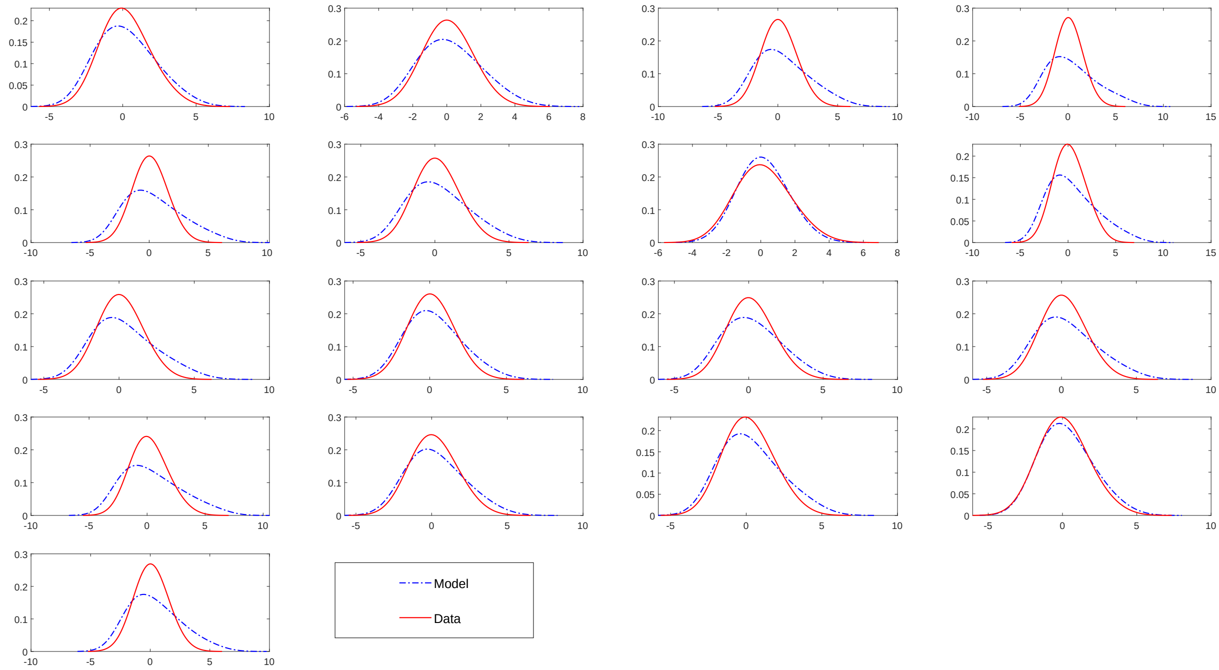


Figure D.2: Share of value added, sector by sector

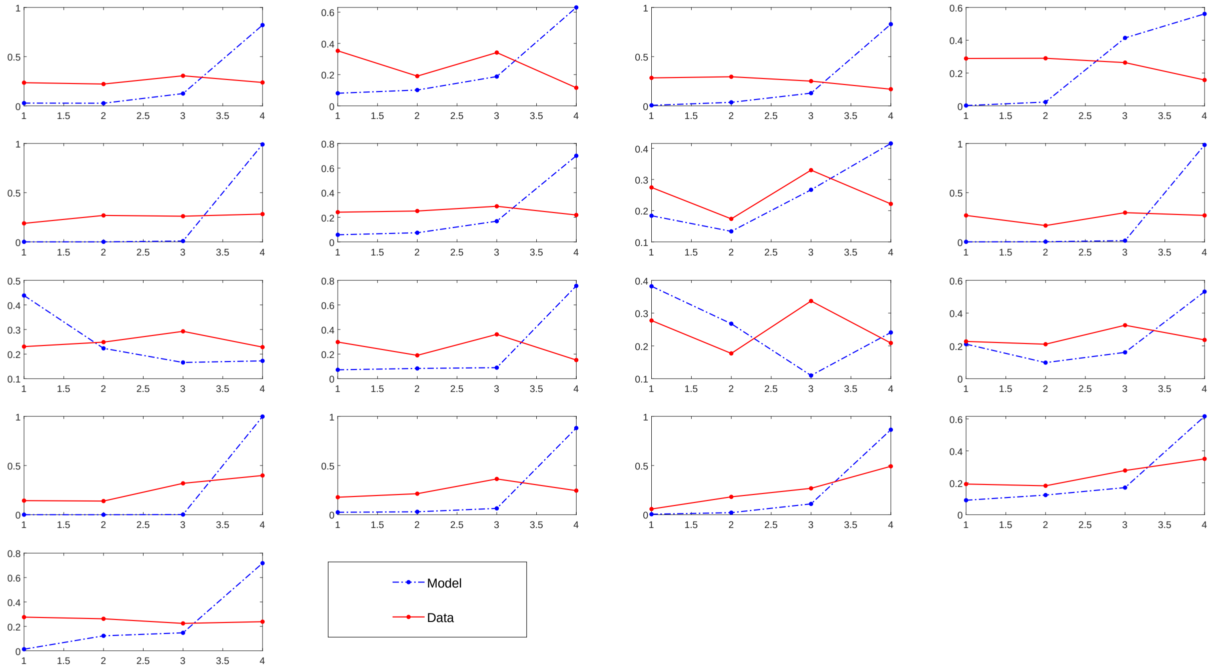


Figure D.3: Average value added, sector by sector

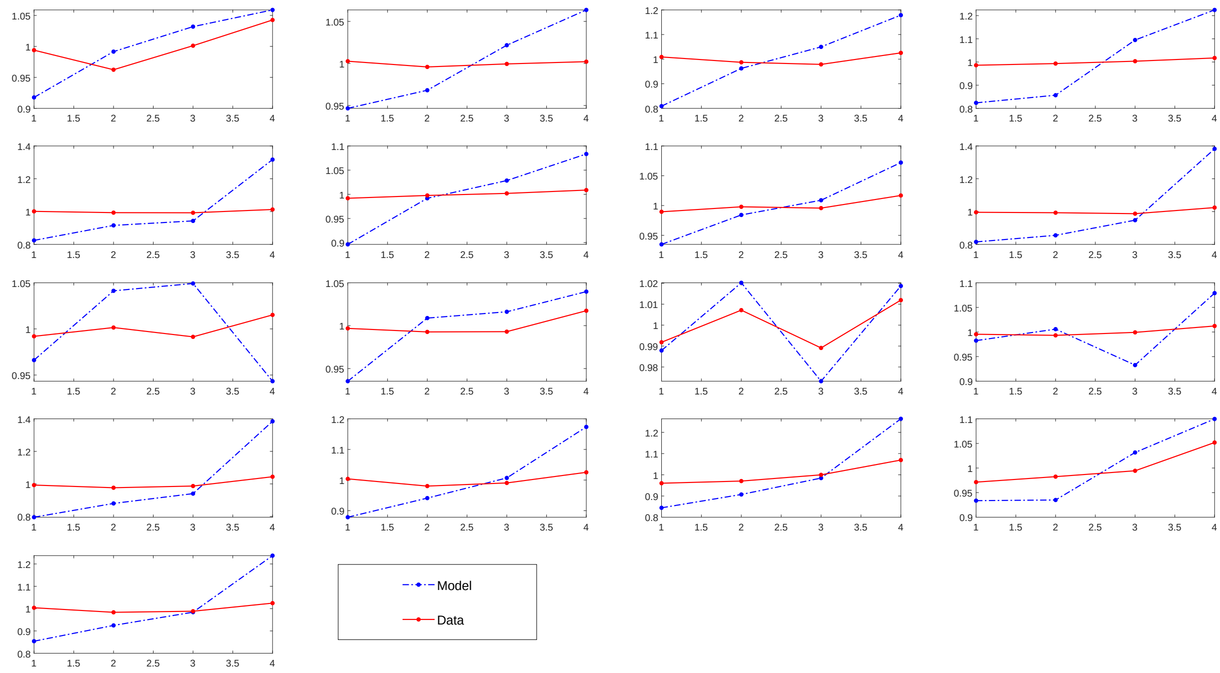
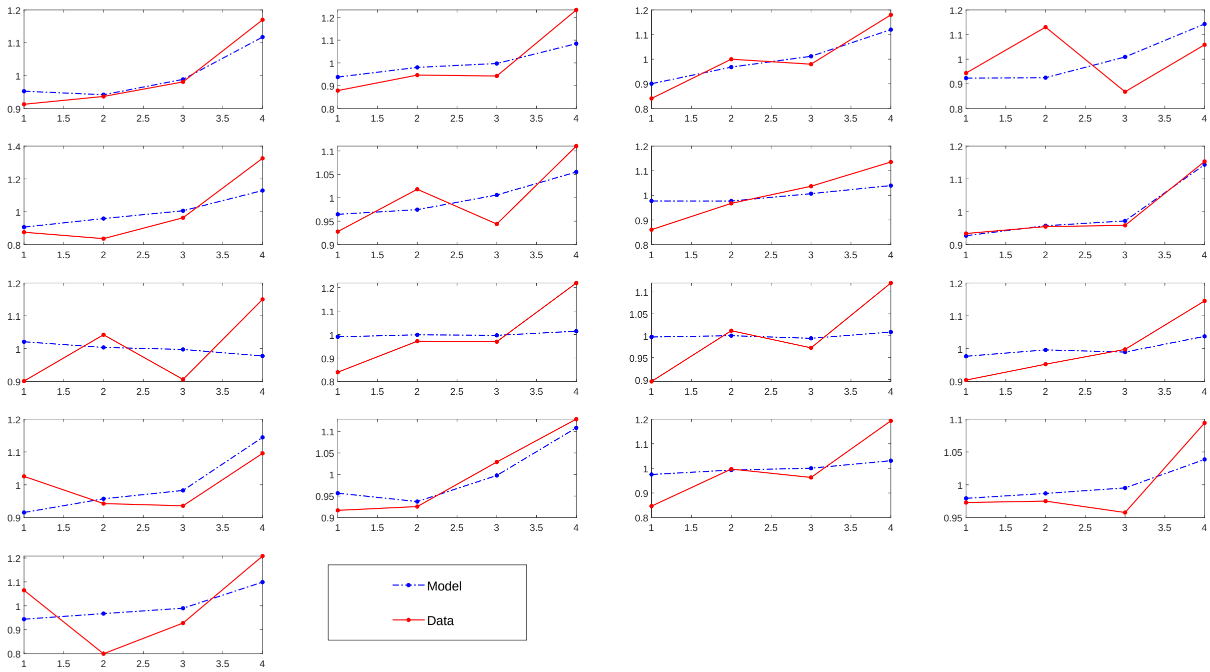


Figure D.4: Average skill intensity, sector by sector



E. Sensitivity Analysis

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