A Simple Model of Functional Specialization of Cities

Kohei Nagamachi

Kagawa University

21. January 2015

A Simple Model of Functional Specialization of Cities*

Kohei Nagamachi[1]

Kagawa University

January 21, 2015

Abstract

Resorting to the method proposed by Matsuyama (2013), this paper develops a static equilibrium model of a system of cities in which ex ante identical locations specialize in stages of production different in the degree of dependence on routine and nonroutine local services sectors, the latter of which is tied to an agglomeration force due to monopolistic competition á la Dixit and Stiglitz (1977). The model is simple in that the system is summarized by a second-order differential equation, which has a unique non-degenerate city size distribution with the comovement of income, population, factor prices, and urban diversity as observed for the U.S. cities. Two examples of use of the model are then illustrated: analyses of welfare gain from functional specialization and optimal income redistribution, the latter of which provides an important implication of an increasing importance of interactive activities in a modern developed economy for income redistribution. Although extending the model makes the model analytically intractable, it is still characterized by a differential equation easily solved with a numerical method and thus useful for further analyses.

Keywords: functional specialization, system of cities, optimal income redistribution policy
JEL classification: F12, R12, R13

*The former title is "Task Trade and the Size Distribution of Cities."

[1]We would like to thank Takatoshi Tabuchi, Kiminori Matsuyama, Takaaki Takahashi, Hajime Takatsuka, Rafael Gonzáles-Val, Shota Fujishima, Daisuke Oyama, Kenichi Amaya, Tadashi Morita, Daichi Shirai, and Dao-Zhi Zheng for their helpful and insightful comments and suggestions. We would also like to thank comments from participants of the 8th Meeting of the Urban Economics Association, 2013 Autumn Meeting of the Japanese Economic Association, 2013 Asia Meeting of the Econometric Society, 2014 25th Annual Meeting of the Applied Regional Science Conference, and September 11th Regional Science Seminar at Kagawa University. All remaining errors are ours. This research is financially supported by the JSPS Grant-in-Aid for Research Activity Start-up No.24830025 and the Project "Spatial Economic Analysis on Regional Growth" undertaken at Research Institute of Economy, Trade and Industry (RIETI). E-mail: nagamachi@gsm.kagawa-u.ac.jp

1 Introduction

Urban economics is now undergoing a refinement of traditional thinking, recognizing the increasing importance of functional specialization as a modern form of urban specialization (Duranton and Puga, 2005). A similar phenomenon has also been observed in trade literatures which stress task trade between countries (Grossman and Rossi-Hansberg, 2008). However, what distinguishes this refinement in urban economics from that in trade literatures is the focus of economists on the relation of functions with urban diversity. In their seminal work, Duranton and Puga (2001) discuss functional specialization of cities. Some cities specialize in the development of new products and their appropriate production technologies, whereas the others specialize in stylized production based on the methods developed in the former cities.[2] Urban diversity, which generally involves a network of people with different ideas, helps to provide solutions to particular problems or develop new ideas, acting as an agglomeration force or so-called urbanization economies (Jacobs, 1969). Due to this agglomeration force, cities can exist even if they face substantial congestion diseconomies.[3] The specialization of cities by function then results in regional disparities reflecting variations in the relative magnitudes of these counteracting forces.

The purpose of this paper is to formalize the above modern story of urban specialization into a model which is simple enough to serve as a tool for analyses of important issues to be analyzed in a system of cities framework. To this end, we extend Matsuyama's (2013) international trade model to a spatial equilibrium model of a system of cities with a continuum of stages of production in which ex ante identical locations specialize ex post in different sets of stages of production. Each stage of production differs in the degree of dependence on routine and nonroutine local services sectors, characterized by perfect competition and monopolistic competition á la Dixit and Stiglitz (1977), respectively, and those with higher skill intensity, higher cost share of the nonroutine local services, benefit more from varieties of nonroutine local services or urban diversity facilitating complex problem-solving activities. Therefore, urban diversity acts as an agglomeration force, that ensures that concentration into a particular location is sustainable despite being associated with a higher rental price of land which constitutes production inputs together with labor and is also consumed by households. Specialization of cities then, by generating differences in the average skill intensity among cities, results in a non-degenerate city size distribution. As in Matsuyama (2013), an equilibrium is characterized as a single, second-order difference or differential equation which we call the fundamental equation.

Using the model, we first show that an equilibrium with functional specialization of cities exists

[2] More precisely, the model provided by Duranton and Puga (2001) is the one of process innovation. However, the same mechanism seems to work in the case of product innovation. Empirical studies such as Feldman and Audretsch (1999) suggest this type of mechanism.

[3] Empirical studies such as Feldman and Audretsch (1999) and Davis and Henderson (2008) support this line of thinking by veri fying the importance of urban diversity and differentiated local service supplies in promoting innovative activities and enhancing the productivity of firms. Furthermore, the increasing importance of interactive tasks empirically shown by Michaels et al. (2013) indirectly suggests urban diversity as a clue to understanding modern urban agglomeration.

and is unique in the sense that there exists a unique, non-degenerate size distribution of cities. Furthermore, this equilibrium is shown to be characterized by the comovement of income, population, the wage rate, the land rent, the average establishment size (in the monopolistic competition sector), and the number of varieties of local nonroutine services, being actually observed for the U.S. metropolitan areas. The analytical tractability of the model also allows us to derive the necessary and sufficient condition for the size distribution of cities to obey a power law including Zipf's law as a special case.

We then illustrate two examples of use of the model: analyses of welfare gains from functional specialization and income redistribution, the latter of which focuses on a second-best allocation achieved under a lump-sum transfer scheme and the assumption that firms in the monopolistically competitive sector can set their own prices. Although we need to resort to numerical methods, computation is easy because an equilibrium of an extended model is still characterized as a single, second-order differential equation.

In the first example, it is shown that, for an empirically plausible range of parameter values, the (static) welfare gain from functional specialization is large, e.g., more than about 25% increase in life-time utility. Although negative welfare effects of higher land rents in top cities due to the concentration of economic activity is severe, functional specialization generates a strong positive welfare effect of urban diversity, which dominates all welfare effects and thus implies a positive correlation between the overall welfare gain from functional specialization and the welfare gain due to urban diversity. The welfare gain is strengthened when the elasticity of substitution decreases and when the distribution of skill intensity of production process is diverse because that magnifies benefits from specialization.

The second example provides an important implication of the increasing importance of interactive activities in a modern developed economy, reported by Michaels et al. (2013), for income redistribution: the desirable income redistribution depends on types of the change in the nature of production technology. More specifically, in our model, the increasing importance of interactiveness can be interpreted as a shift of the distribution of skill intensity to the right or a lower elasticity of substitution. On one hand, the former always implies a decrease in the optimal income tax rate simply because an increasing dominance of monopolistically competition makes the difference between the laissez-faire and the second-best allocations smaller by definition. On the other hand, the effect of the latter depends on cases because the optimal tax rate is monotonically increasing in the elasticity of substitution if the distribution of skill intensity is sufficiently skewed to the right, otherwise there is an inversed-U-shaped relationship. When the distribution of skill intensity is highly skewed to the right, the economy is dominated by the monopolistically competitive sector, implying less room of income redistribution motivated by the composition of constant and increasing returns. In this case, an increase in the elasticity of substitution favors more dispersed economic activity simply because of limited supply of land. However, if the distribution of skill intensity is not so skewed to the right, the laissez-faire is distant from the second-best allocation, implying more room for income redistribution. This is especially relevant for a lower elasticity of substitution which is associated with a greater concentration of economic activity and thus sever negative welfare effects of higher land rents. The associated welfare gain from income redistribution is limited, e.g., less than about 2.3% increase in life-time utility in the most of an empirically relevant range of parameters, consistent with Desmet and Rossi-Hansberg (2013).

This paper is related to three strands of literature. The first strand is the literature on the international trade theory of Ricardian comparative advantage. We extend the Matsuyama's (2013) model to a regional context and also go a step further by conducting a welfare analysis of income redistribution. He introduces monopolistic competition à la Dixit and Stiglitz (1977) into Dornbusch et al. (1977), resulting in a model with symmetry break through endogenous comparative advantage. He also develops a new method allowing us to characterize an equilibrium of the world with multiple and arbitrarily large number of countries as a second-order difference or differential equation, the latter of which is analytically solvable. Our model has the same degree of analytical tractability as in Matsuyama's (2013) international case, even if we include migration across cities and an immobile factor of land, that is used not only for production but also for consumption as in Pflüger and Tabuchi (2010). The current urban application of Ricard's theory with endogenous comparative advantage brings us a clear relationship between comparative advantage and agglomeration and dispersion forces of economic activity.

The second strand of literature is the research on a system of cities. To our knowledge, this is the first paper that, using a system of cities model with functional specialization, conducts a welfare analysis of an optimal income redistribution policy. Economic theories of the size distribution of cities range from stochastic growth models to static deterministic ones. Some models of the former type such as Eeckhout (2004) do not focus on the specialization of cities, whereas models such as Rossi-Hansberg and Wright (2007) and Duranton (2007) (of the former type) and Hsu (2012) (of the latter type) focus on cross-city variation in industries. Recent studies by Behrens et al. (2014a) and Behrens et al. (2014b) advance our understanding on the effects of heterogeneous agents and urban amenities on the size distribution of cities, respectively. However, the main focus of the above studies is on the positive side. While providing a model of a system of cities with functional specialization, Duranton and Puga (2005) preclude the possibility of an income redistribution policy by introducing competitive developers á la Henderson (1974). From a different perspective, Desmet and Rossi-Hansberg (2013) advances our understanding on welfare effects of efficiency, amenities, and frictions by applying the Business Cycle Accounting method in the macroeconomics literature to the urban context, but an optimal redistribution is not their scope of analysis.

The third is trade models which specify the production of goods as a continuum of intermediate goods, stages of production or tasks. The current model does not include trade costs of stages of production for analytical tractability and thus heterogeneous trade costs, introduced by Grossman and Rossi-Hansberg (2008). In this sense, the model specification is somewhat similar to previous international trade models.[4] Instead of heterogeneous trade costs, heterogeneity in skill intensity allows the fraction of stages of production is determined endogenously, reflecting their skill intensities and the comparative advantages of cities, the latter of which are in turn determined endogenously, i.e., a circular causation process.

The remainder of this paper is organized as follows. We first provide the model in Section 2. In Section 3, we discuss the equilibrium properties including the welfare gain from functional specialization. The model is then extended by introducing an income redistribution policy using an income tax and lump-sum transfers across cities to analyze the optimal income redistribution policy in Section 4. Finally, we conclude this paper in Section 5.

2 The Model

In this section, we provide a simple spatial equilibrium model with a continuum of stages of production with heterogeneity in skill intensity that are fragmented and outputs of which are traded across cities and within firms. The economic agents consist of mobile workers, final good firms, and city-specific local services firms, the last of which include monopolistically competitive and perfectly competitive firms. Here, local services firms provide outsourcing services that are necessary for conducting stages of production, and the variety effect introduced via monopolistic competition á la Dixit and Stiglitz (1977) captures urban diversity. In the latter sense, we call the monopolistically competitive sector the nonroutine sector and the other the routine sector. This stylized specification makes clear the distinction between the two sectors in a way that the nonroutine sector conducts nonroutine tasks that often require complex problem-solving activities which could be facilitated with the help of local network. The model is essentially an application of Matsuyama's (2013) framework to the urban context.

In the following, we explain the optimization problems of the agents in order. For the sake of convenience, we first assume that there are $J \in \mathbb{N}$ ex ante identical locations in the economy,[5] each of which is endowed with one unit of land. We subsequently modify this assumption by making J diverge to infinity but making the measure of each location converges to zero in such a way that the total measure of locations is equal to unity, which allows us to consider the size distribution of cities.[6]

[4] The current model differs from these studies in more respects. Unlike Grossman and Rossi-Hansberg (2008), each task is not related to a particular production factor. Rather, all tasks use the same set of production inputs; the skill intensity of each task is different from each other; and there is a continuous distribution of such skill intensity. In this sense, except for the use of labor and land instead of labor and capital as production inputs, the specification of our model is close to those of Dixit and Grossman (1982), Yi (2003), and Kohler (2004). However, the current model also shares the same assumption with Grossman and Rossi-Hansberg (2008) as well as Feenstra and Hanson (1996) in that there is no vertical linkage between different tasks or between intermediate inputs. Importantly, the current model differs from all these studies in that it deals with an arbitrarily large number of locations rather than just two countries or a single small open economy.

[5] In this paper, we do not distinguish cities, regions, and locations and use these words interchangeably.

[6] Whether the distribution corresponding to a discrete J converges to that with $J = \infty$ as J increases itself is an important research topic. However, answering such a question is beyond the scope of this paper. In the following, we simply assume that the case of $J = \infty$ can approximate the one with a large J, or we focus only on the limiting case.

2.1 Workers

There is a unit mass of identical workers in the economy. Each worker is freely mobile across locations and thus decides her location as well as consumption of goods and services.

Suppose that a worker had already chosen her residence $j \in \{1, 2, \cdots, J\}$, which is also her workplace. Then, she solves the following utility maximization problem:

$$U_j = \max_{c_j, h_j \geq 0} c_j^{1-\alpha} h_j^{\alpha} \quad s.t. \quad Pc_j + R_j h_j = W_j + \bar{R}_j, \qquad 0 < \alpha < 1.$$

Her income consists of labor income W_j and subsidy $\bar{R}_j$.[7] Assuming that the land rents in a city are received by the residents of that city as in Behrens et al. (2014b) and Michaels et al. (2013), $\bar{R}_j$ is given by

$$\bar{R}_j = \frac{R_j}{N_j} \quad \forall j. \tag{1}$$

She uses these sources of income to consume homogeneous tradeable good c_j and housing h_j, the prices of which are P and R_j, respectively.

The associated indirect utility function, together with free migration, then implies

$$\frac{W_j + \bar{R}_j}{W_{j'} + \bar{R}_{j'}} = \left(\frac{R_j}{R_{j'}}\right)^{\alpha} \qquad \forall j \neq j', \tag{2}$$

which imposes a restriction on the relationship between income and land rent differentials.

2.2 Final Good Sector

The tradeable homogeneous final good is produced using a constant-returns-to-scale (CRS) Cobb-Douglas production technology. More specifically, the production of one unit of the final good requires to process a continuum of stages of production (hereafter, stages) $\{t : 0 \leq t \leq 1\}$:[8]

$$Y = \exp\left[\int_0^1 \ln(y(t))dt\right], \tag{3}$$

where Y and $y(t)$ denote outputs of the final good and stage t, respectively. The equal weights and unit mass of the stages imply that the total sales, which are equal to the economy-wide income $E =$

[7] We assume that each worker is endowed with one unit of time and supplies it inelastically. As discussed later, there are two different sectors, the nonroutine and routine sectors, that mainly conduct nonroutine and routine tasks, respectively. For simplicity, we assume that each worker has an equal ability to perform each type of task. Thus, in an equilibrium where both sectors locate at almost every locations, wages rates within each location are equalized.

[8] In Matsuyama (2013), what we consider to be a continuum of stages of production here is interpreted as a continuum of sectors, and the technology specified by (3) directly enters the utility function. However, in the regional context, this interpretation is not favorable given the transition from sectoral specialization of cities to functional one argued by Duranton and Puga (2005).

$\sum_{j=1}^{J}(W_jN_j+R_j)$ times the expenditure share $(1-\alpha)$ of the final good, are distributed to each stage t.

Firms decide where each of these stages is performed. For each fixed stage $t \in [0,1]$, a typical firm decides the quantity $y_j(t)$ of production of stage t at a location $j \in \{1,2,\cdots,J\}$. Once the stage has been processed at each location, outputs $\{y_j(t)\}_{j=1}^{J}$ are aggregated and transported without cost to use them as a production input of the final good:[9]

$$y(t) = \sum_{j=1}^{J} y_j(t) \quad \forall t. \tag{4}$$

Furthermore, stage t is processed at location j with the help of the outsourcing services by local nonroutine services $X_{n,j}(t)$ and local routine services $X_{r,j}(t)$. Its output $y_j(t)$ is given by[10]

$$y_j(t) = X_{n,j}(t)^{\gamma(t)} X_{r,j}(t)^{1-\gamma(t)} \quad \forall j,t, \tag{5}$$

where $\gamma(t) \in [0,1]$ represents the skill intensity of stage t. In the following, we assume that $\gamma(t)$ is strictly monotonically increasing. For analytical convenience, we also assume that $\gamma(t)$ is continuously differentiable, i.e., $\gamma'(t) > 0$ for all t. We also assume that $\gamma(0) = 0$ and $\gamma(1) = 1$.

Therefore, letting $P_{n,j}$ and $P_{r,j}$ denote the prices of the nonroutine and routine services, respectively, we can write the profit maximization of the final good firm as follows:

$$\max PY - \sum_{j=1}^{J} \int_{\mathbb{T}_j} [P_{n,j}X_{n,j}(t) + P_{r,j}X_{r,j}(t)]\,dt \quad s.t. \quad (3)-(5), \text{ and } \mathbb{T}_j \subseteq [0,1],$$

where the control variables consist of the measurable set $\mathbb{T}_j$ of stages processed at location j as well as quantities $(Y, \{y(t)\}_t, \{y_j(t), X_{n,j}(t), X_{r,j}(t)\}_{j,t})$. Defining $|\mathbb{T}_j|$ as the Lebesgue measure of $\mathbb{T}_j$, we can see that if $\mathbb{T}_j$'s are all mutually exclusive, which is indeed the case, $|\mathbb{T}_j|$ fraction of the total sales $(1-\alpha)E$ is distributed to location j. In addition, $\gamma(t)$ and $1-\gamma(t)$ fractions of the distribution $(1-\alpha)E$ to stage t are distributed to the nonroutine and routine sectors, respectively. Thus, the nonroutine sector at location j receives $\int_{\mathbb{T}_j}\gamma(t)dt(1-\alpha)E \equiv \Gamma_j|\mathbb{T}_j|(1-\alpha)E$, where $\Gamma_j \equiv |\mathbb{T}_j|^{-1}\int_{\mathbb{T}_j}\gamma(t)dt$ is the average skill-intensity of location j.

In addition, as we discuss in the next two subsections, we assume that the market structures of the nonroutine and routine sectors are monopolistically and perfectly competitive, respectively.[11] More

[9] Transportation costs can be introduced by dividing goods and services into tradeable and non-tradeable parts with CRS Cobb-Douglas composite technology, where the share parameter corresponding to the tradeable is interpreted as the freeness of trade.

[10] The implicit assumption here is that either type of local services alone cannot be traded across locations. This exactly corresponds to what we do when communicating with people in different locations. For example, sending an interesting idea to your friend via e-mail breaks down to two stages: developing or formulating the idea, a nonroutine task, and writing or sending an e-mail, a routine task. Although each task is not tradeable itself, the output, i.e., the combination of those tasks, is now tradeable.

[11] This stylized specification reflects our view of the nature of nonroutine services such as management, research and

specifically, $X_{n,j}(t)$ denotes the composite of a continuum of varieties $\{x_{n,j}(v,t)\}_{v\in[0,D_j]}$. The quantity of each variety is also a control variable, with the following technology:

$$X_{n,j}(t) = \left[\int_0^{D_j} x_{n,j}(v,t)^{\frac{\sigma-1}{\sigma}} dv\right]^{\frac{\sigma}{\sigma-1}} \quad \forall j, t \in \mathbb{T}_j,$$

where D_j is the number of varieties produced at location j, which we call *urban diversity*, and $\sigma > 1$ is the elasticity of substitution between any two different varieties.

The profit maximization then implies the following demand for variety v from stage t processed at location j:

$$x_{n,j}(v,t) = \left[\frac{p_{n,j}(v)}{P_{n,j}}\right]^{-\sigma} X_{n,j}(t) \quad \forall v, j, t \in \mathbb{T}_j,$$

where $P_{n,j}$ is the price index of the nonroutine services at location j given by

$$P_{n,j} = \left[\int_0^{D_j} p_{n,j}(v)^{-\frac{1}{\theta}} dv\right]^{-\theta} \quad \forall j. \tag{6}$$

Here, $\theta \equiv 1/(\sigma-1)$.

2.3 Nonroutine Local Service Sector

As mentioned in the previous subsection, the nonroutine local service sector is characterized by monopolistic competition á la Dixit and Stiglitz (1977), and each firm produces one variety. In addition, as in Pflüger and Tabuchi (2010), we assume that production inputs consist of both labor and land. More specifically, the fixed and marginal costs of production are both measured in terms of their Cobb-Douglas composite with a cost share parameter $\beta \in (0,1)$ for land.

Formally, variety-v firm at location j solves

$$\pi_j(v) = \max_{p_{n,j}(v), q_j(v)} \left[p_{n,j}(v) - mR_j^{\beta} W_j^{1-\beta}\right] q_j(v) - R_j^{\beta} W_j^{1-\beta} f \quad s.t. \quad q_j(v) = \int_{\mathbb{T}_j} x_{n,j}(v,t) dt,$$

where $q_j(v)$ is the output of variety-v firm at location j, and m and f denote the shift parameters of marginal and fixed costs, respectively.

The profit maximization then implies the optimal pricing rule specified by $p_{n,j}(v) = (1+\theta) m R_j^{\beta} W_j^{1-\beta}$, and substituting this into (6) and taking the ratio across two different locations, j and j', we obtain

$$\frac{P_{n,j}}{P_{n,j'}} = \left(\frac{R_j}{R_{j'}}\right)^{\beta} \left(\frac{W_j}{W_{j'}}\right)^{1-\beta} \left(\frac{D_j}{D_{j'}}\right)^{-\theta} \quad \forall j \neq j'. \tag{7}$$

development, and legal services as well as routine services such as line production based on previously developed blueprints.

2.4 Routine Local Service Sector

Unlike the nonroutine local service sector, the routine sector is characterized by perfect competition with a CRS Cobb-Douglas production technology:

$$\max_{H_{r,j},\ L_{r,j}} P_{r,j} H_{r,j}^{\beta} L_{r,j}^{1-\beta} - R_j H_{r,j} - W_j L_{r,j}.$$

Note that the production of routine services also requires both labor and land. For simplicity, we assume the same cost share parameter β as in the nonroutine sector.[12]

The profit maximization then implies

$$\frac{P_{r,j}}{P_{r,j'}} = \left(\frac{R_j}{R_{j'}}\right)^{\beta} \left(\frac{W_j}{W_{j'}}\right)^{1-\beta} \quad \forall j \neq j'. \tag{8}$$

2.5 Equilibrium

We now define a market equilibrium. Since the locations are ex ante identical by assumption, the symmetric configuration always exists. However, this configuration does not seem robust to exogenous shocks to the economy, and thus, we focus on another type of equilibrium, which is the only equilibrium other than the symmetric one and which is unique at least in the limiting case $J \to \infty$, which is of general interest.

Specifically, we define an equilibrium with rankings. Without loss of generality, assume that $0 < |\mathbb{T}_1| < |\mathbb{T}_2| < \cdots < |\mathbb{T}_{j-1}| < |\mathbb{T}_j| < \cdots < |\mathbb{T}_J|$, i.e., as j increases, the associated market share increases. It is then immediately demonstrated that there must exist an increasing sequence $\{T_j\}_{j=0}^{J}$ of thresholds such that $\mathbb{T}_j = (T_{j-1}, T_j]$ for all $j \in \{1, 2, \cdots, J\}$; $T_0 = 0$; and $T_J = 1$ under free migration and free entry into the nonroutine sector. That is, if cities are different, we observe a perfect sorting of stages of production,[13] and the higher the location index j is, the higher the average skill intensity Γ_j is. Note that this configuration of the equilibrium is consistent with the argument advanced by Duranton and Puga (2001) that cities sort themselves into specialized cities, some of which host the research and development sectors testing prototype products, while others host the production sectors producing goods in a stylized manner. It should, however, be noted that in our model, there is no perfect specialization with respect to the service sectors in the sense that every city hosts both nonroutine and routine service sectors. Stated differently, the important characteristic that distinguishes one city from another is its average skill-intensity.

Therefore, we call the equilibrium on which we focus a *sorting equilibrium* and define it as follows:

---

[12] More precisely, there are two reasons why we introduce this assumption. The first is for simplicity. Without this assumption, we cannot obtain the analytical solution discussed later. The second is to make clear the distinction between the two local services sectors. With this assumption, the nonroutine sector differs from the routine one in one way, market structure.

[13] Section A explains why the economy exhibits a perfect sorting in equilibrium.

Definition 1. *A sorting equilibrium is a set of prices* $(P, \{P_{n,j}, P_{r,j}, R_j, W_j\}_j, \{p_{n,j}(v)\}_{v,j})$*, quantities* $(Y, \{c_j, h_j, H_{r,j}, L_{r,j}\}_j, \{y(t)\}_t \{y_j(t), X_{n,j}(t), X_{r,j}(t)\}_{j,t}, \{x_{n,j}(v,t)\}_{v,t,j}, \{q_j(v)\}_v)$*, transfers* $\{\bar{R}_j\}_j$*, populations* $\{N_j\}_j$*, diversities* $\{D_j\}_j$*, and a sequence* $\{T_j\}_{j=0}^J$ *of thresholds such that*

1. *workers maximize their utility by choosing quantities and locations;*
2. *firms maximize their profits;*
3. *markets clear:*

$$R_j = (1-\alpha)\beta|\mathbb{T}_j|E + \alpha(W_jN_j + \bar{R}_jN_j), \quad \text{(Land)} \tag{9}$$

$$W_jN_j = (1-\alpha)(1-\beta)|\mathbb{T}_j|E; \quad \text{(Labor)} \tag{10}$$

4. *there is free entry into the nonroutine local service sectors; and*
5. *thresholds* $\{T_j\}_{j=0}^J$ *are consistent with comparative advantage:*

$$\left(\frac{P_{n,j+1}}{P_{n,j}}\right)^{\gamma(T_j)} \left(\frac{P_{r,j+1}}{P_{r,j}}\right)^{1-\gamma(T_j)} = 1, \tag{11}$$

where $P_{n,j+1}/P_{n,j} < 1$ *and* $P_{r,j+1}/P_{r,j} > 1$ *for all* j*.*

Here, the market clearing conditions, i.e., (9) and (10), are evident from the specifications of the utility and production technologies presented in the previous subsections.[14] The conditions that $P_{n,j+1}/P_{n,j} < 1$ and $P_{r,j+1}/P_{r,j} > 1$ imply that location $j+1$, compared to location j, has a comparative advantage in the nonroutine service sector and thus has a comparative advantage in stages with higher skill intensity. The threshold T_j here is the stage for which locations j and $j+1$ have the same comparative advantage, i.e., both locations j and $j+1$ have the same unit cost of producing stage T_j output.

3 Equilibrium Properties

In this section, we first consolidate the equilibrium conditions presented in the previous section to obtain an equation governing the equilibrium of the economy in Subsection 3.1. We then prove the existence and uniqueness of a sorting equilibrium in Subsection 3.2, which has clear and empirically valid predictions about the relationships between the city size and some variables. The relation of our model to the size distribution of cities is also derived in Subsection 3.3.

[14] The market clearing condition for the final good is $PY = E$, where P denotes the price index of the final good. We omit this condition from the definition because we do not use it in deriving analytical results presented in the next section.

3.1 Fundamental Equation

For a given J, we first show that the equilibrium system of the economy reduces to the following *fundamental equation*:

$$\left(\frac{T_{j+1}-T_j}{T_j-T_{j-1}}\right)^{\alpha(1-\beta)+\beta-(1-\alpha)(1-\beta)\theta\gamma(T_j)}=\left(\frac{\Gamma_{j+1}}{\Gamma_j}\right)^{\theta\gamma(T_j)}\qquad \forall j\in\{1,2,\cdots,J-1\},$$

with $T_0=0$ and $T_J=1$. Given the definition of Γ_j, i.e., $\Gamma_j=|T_j-T_{j-1}|^{-1}\int_{T_{j-1}}^{T_j}\gamma(t)dt$, the fundamental equation is a second-order difference equation with two boundary conditions.

For this purpose, we start with the result that consolidating market clearing conditions (9) and (10) along with the distribution of land rents, (1), yields

$$\text{(12)}\qquad \frac{R_{j+1}}{R_j}=\frac{W_{j+1}N_{j+1}+R_{j+1}}{W_jN_j+R_j}=\frac{|\mathbb{T}_{j+1}|}{|\mathbb{T}_j|}\quad \forall j.$$

That is, given the ordering of $|\mathbb{T}_j|$, the higher the market share $|\mathbb{T}_j|$ is, the higher the land rent R_j and regional income $W_jN_j+R_j\equiv E_j$ are. In addition, it is also implied that differentials of land rent R_j and market share $|\mathbb{T}_j|$ are equal. That $R_j<R_{j+1}$ is simply because both locations have the same area of land while income concentrates in location $j+1$.

This result is then combined with the free-migration condition (2) to obtain

$$\text{(13)}\qquad \frac{N_{j+1}}{N_j}=\left(\frac{|\mathbb{T}_{j+1}|}{|\mathbb{T}_j|}\right)^{1-\alpha}\quad \forall j,$$

which states that the ordering of population N_j is the same as the market share $|\mathbb{T}_j|$, and that the differential of population N_j is less than proportional to that of the market share $|\mathbb{T}_j|$ or, using (12), the land rent R_j or local income E_j. The latter is interpreted as the spatial equilibrium imposing some upper bound on the population differential. As a result, the ordering of the wage rate W_j is also the same as $|\mathbb{T}_j|$ because the labor market clearing condition (10) implies that the differential of labor compensation W_jN_j is equal to that of the market share $|\mathbb{T}_j|$:

$$\text{(14)}\qquad \frac{W_{j+1}}{W_j}=\left(\frac{|\mathbb{T}_{j+1}|}{|\mathbb{T}_j|}\right)^{\alpha}\quad \forall j.$$

That $W_j<W_{j+1}$ is interpreted as a differential compensating the disparity in the cost of living, i.e., $R_j<R_{j+1}$. The spatial equilibrium (2) limits the degree of the wage differential, allowing the co-movement between the wage rate and population.

These results immediately imply that a higher market share $|\mathbb{T}_j|$ is associated with more severe market competition, which is a *dispersion force*, in the sense that the unit production cost and thus the price $P_{r,j}$ of the routine local services is higher in that location. More specifically, substituting (12) and (14) into (8), we obtain

$$\frac{P_{r,j+1}}{P_{r,j}} = \left(\frac{|\mathbb{T}_{j+1}|}{|\mathbb{T}_j|}\right)^{\alpha(1-\beta)+\beta} \quad \forall j. \tag{15}$$

Another important implication of this result is that a location with a higher market share $|\mathbb{T}_j|$ exhibits comparative advantage in the production of nonroutine services. This can be seen from the determination of urban diversity, a factor that generates comparative advantage in nonroutine services. Substituting (7), (12), (14) and (15) into (11), we obtain

$$\left(\frac{R_{j+1}}{R_j}\right)^{\beta}\left(\frac{W_{j+1}}{W_j}\right)^{1-\beta} = \left(\frac{D_{j+1}}{D_j}\right)^{\theta\gamma(T_j)}, \text{ or } \left(\frac{|\mathbb{T}_{j+1}|}{|\mathbb{T}_j|}\right)^{\alpha(1-\beta)+\beta} = \left(\frac{D_{j+1}}{D_j}\right)^{\theta\gamma(T_j)}$$

for all j. That is, if a sorting equilibrium exists, more severe market competition, i.e., higher wage rage and land rent, in a location with a higher market share $|\mathbb{T}_j|$ are associated with greater urban diversity D_j, leading the location to exhibit comparative advantage in the production of nonroutine services. We here observe an *agglomeration force* represented by urban diversity D_j. Therefore, we can see that in this model, there is a close relationship between these agglomeration and dispersion forces through regional comparative advantage, which is a result of the fragmentation of production stages across locations. In a sorting equilibrium, if it were to exist, regional disparities would emerge as a balance between the agglomeration and dispersion forces,[15] and this balance would be affected by functional specialization of locations reflecting the function $\gamma(t)$ and, thus, the distribution of the skill intensities in the economy.

Finally, by utilizing the free-entry condition for the nonroutine sector, we can derive the fundamental equation. First, note that the free entry, $\pi_j(v) = 0$ for all v and j, together with the optimal pricing rule, implies that the output $q_j(v)$ of each variety v at location j is constant, i.e., $q_j(v) = f/(\theta m) \equiv q$ for all v and j. Then, the market clearing condition for the nonroutine sector yields $D_j p_{n,j} q = (1-\alpha)\Gamma_j |\mathbb{T}_j| E$, or taking the ratio of this equation, we obtain

$$\frac{D_{j+1}}{D_j} = \frac{\Gamma_{j+1}}{\Gamma_j}\frac{|\mathbb{T}_{j+1}|}{|\mathbb{T}_j|}\left(\frac{p_{n,j+1}}{p_{n,j}}\right)^{-1} = \frac{\Gamma_{j+1}}{\Gamma_j}\left(\frac{|\mathbb{T}_{j+1}|}{|\mathbb{T}_j|}\right)^{(1-\alpha)(1-\beta)} \quad \forall j.$$

To avoid this arbitrariness of the finiteness of J, we investigate the distribution of cities in a sorting equilibrium. We accomplish this by making J diverge to infinity while limiting the total measure of cities to unity.[16] Then, applying the Matsuyama's (2013) method to the fundamental equation,[17]

[15] Note that the equality between agglomeration and dispersion forces holds at a threshold T_j. Since $R_{j+1}/R_j, W_{j+1}/W_j, D_{j+1}/D_j > 1$ and $\gamma(t)$ is increasing in t, compared with location j, location $j+1$ has comparative advantage in stage $t > T_j$ due to dominance of agglomeration force over dispersion force implying lower unit costs of conducting more skill-intensive tasks.

[16] Note that the unit endowment of land is fixed for each location.

[17] Essentially, the method involves interpreting $1/J$ as a differential dt when J is sufficiently large and then applying the asymptotic expansion. Here, it is crucial that as J diverges to infinity, each city hosts only one stage of production in the

we obtain the following boundary value problem for a second-order ordinary differential equation (ODE):

$$\frac{\Phi''}{\Phi'} = \frac{\theta\gamma'(\Phi)\Phi'}{G(\Phi)} \quad \text{with } \Phi(0) = 0 \text{ and } \Phi(1) = 1, \tag{16}$$

where

$$G(\Phi) \equiv \alpha(1-\beta) + \beta - (1-\alpha)(1-\beta)\theta\gamma(\Phi).$$

Each city is characterized by the stage t that it hosts because as J diverges to infinity, $|\mathbb{T}_j|$ converges to zero or, stated differently, $\mathbb{T}_j$ converges to a point that characterizes one of the cities. Here, $\Phi(t)$ is the Lorenz curve of the market share that corresponds to $\sum_{k=1}^{j} |\mathbb{T}_k|$ for some j. Thus, given the uniformity of stage t, $\Phi'(t)$ corresponds to $|\mathbb{T}_j|$, the market share of a location. In the following, given the one-to-one correspondence between a city and a stage, we call the city that hosts stage t city t. We assume that $G(1) > 0$ in order to focus on a meaningful case.[18]

3.2 Existence and Uniqueness of Endogenous Rankings

In order to prove the existence of a sorting equilibrium in the limiting case, it suffices to show that there exists a solution to the fundamental equation (16). Importantly, we can obtain a unique solution to the fundamental equation analytically, which also implies the uniqueness of a sorting equilibrium. The economic interpretation of this result is as follows: although cities may differ, the associated variations in city characteristics are limited to a range that is consistent with the unique distribution. Since cities are ex ante identical, we cannot identify which stage of production each city specializes in ex post.[19]

More specifically, we obtain the inverse function of the Lorenz curve (let $H : z \to t$ denote the function, i.e., $H(z) \equiv \Phi^{-1}(z)$):

$$H(z) = \frac{\int_0^z G(u)^{\frac{1}{(1-\alpha)(1-\beta)}} du}{\int_0^1 G(u)^{\frac{1}{(1-\alpha)(1-\beta)}} du} \quad \forall z \in [0,1]. \tag{17}$$

For a given $\gamma(t)$, this equation yields a unique inverse Lorenz curve $H(z)$. Given that $H'(z) > 0$ and $H''(z) < 0$ for all t, $\Phi(t)$ is unique and has a property: $\Phi'(t) > 0$ and $\Phi''(t) > 0$.

limit and thus is characterized by t.

[18] Intuitively, this assumption implies that the magnitude of market competition, i.e., the power $\alpha(1-\beta)+\beta$ appearing in the fundamental equation, is larger than that of the agglomeration force, $(1-\alpha)(1-\beta)\theta\gamma(t)$, net of the effect of the average skill intensity Γ_j, thus resulting in "bounded" city sizes. It can be easily shown that as $G(1)$ converges to zero from above, the max-min ratio of population diverges to infinity.

[19] This resembles the implication of a stochastic process that is specified by a Markov chain with a non-degenerate unique invariant distribution. That is, the realization of a random variable varies randomly in a manner that is consistent with the invariant distribution. Importantly, this uniqueness of the sorting equilibrium allows us to conduct the numerical exercises discussed later. If we have a cross-sectional dataset of cities, we can calibrate the model.

Then, using this result and normalizing the economy-wide income E to unity without loss of generality, we can establish the following proposition:

Proposition 1. *Suppose that $\gamma'(t) > 0$ and $G(1) > 0$. Then, a sorting equilibrium, characterized by a Lorenz curve of the market size $\Phi(t)$, exists and is unique. In addition, this equilibrium has the following properties: The market share $E(t)$ of city t is given by $\Phi'(t)$, and population $N(t)$, land rent $R(t)$, wage rate $W(t)$, diversity $D(t)$ and the average establishment size $\zeta(t)$ (in the nonroutine sector) in city t are given by*

$$\begin{aligned}
N(t) &= \frac{\Phi'(t)^{1-\alpha}}{\int_0^1 \Phi'(t)^{1-\alpha}dt}, \\
R(t) &= [\alpha(1-\beta)+\beta]\Phi'(t), \\
W(t) &= (1-\alpha)(1-\beta)\int_0^1 \Phi'(t)^{1-\alpha}dt\Phi'(t)^{\alpha}, \\
D(t) &= \frac{\theta}{(1+\theta)f}\frac{1-\alpha}{[\alpha(1-\beta)+\beta]^{\beta}}\left[(1-\alpha)(1-\beta)\int_0^1 \Phi'(t)^{1-\alpha}dt\right]^{-(1-\beta)}\gamma(\Phi(t))\Phi'(t)^{(1-\alpha)(1-\beta)}, \\
\zeta(t) &= \frac{(1+\theta)f}{\theta}(1-\beta)[\alpha(1-\beta)+\beta]^{\beta}\left[(1-\alpha)(1-\beta)\int_0^1 \Phi'(t)^{1-\alpha}dt\right]^{-\beta}\Phi'(t)^{(1-\alpha)\beta}
\end{aligned}$$

for all $t \in [0,1]$. Therefore, as t increases, i.e., as a location specializes in a more nonroutine stage of production, the values of all of these variables increase.

Proof. The market share $\Phi'(t)$ is simply a definition. Population function $N(t)$ follows if we apply the asymptotic expansion to (13):

$$\frac{N(t+\Delta t)}{N(t)} = \left[1+\frac{\Phi''(t)}{\Phi'(t)}\Delta t + o(|\Delta t|)\right]^{1-\alpha} = 1+(1-\alpha)\frac{\Phi''(t)}{\Phi'(t)}\Delta t + o(|\Delta t|).$$

Arranging this result and making Δt converge to zero, we obtain

$$\frac{N'(t)}{N(t)} = (1-\alpha)\frac{\Phi''(t)}{\Phi'(t)} \quad \forall t,$$

which, together with the normalization, i.e., $\int_0^1 N(t)dt = 1$, implies the desired result of $N(t)$. The land rent function $R(t)$ follows immediately from the land and labor market clearing conditions, where $|\mathbb{T}_j|$ is now replaced with $\Phi'(t)$. The wage function $W(t)$ also follows from the labor market clearing condition with $|\mathbb{T}_j|$ replaced with $\Phi'(t)$ and the population function $N(t)$. The diversity function $D(t)$ derives from the market clearing condition for nonroutine services, i.e., $D_j p_{n,j} q = (1-\alpha)\Gamma_j|\mathbb{T}_j| = (1-\alpha)\int_{T_{j-1}}^{T_j}\gamma(t)dt$, together with the results for the land rent and wage rate functions. Here, $\int_{T_{j-1}}^{T_j}\gamma(t)dt$ is replaced with $\gamma(\Phi(t))\Phi'(t)$. The establishment size function $\zeta(t)$ is given by the consistency, i.e., the labor compensation calculated by $W(t)D(t)\zeta(t)$ must be equal to the market size $(1-\alpha)\Phi'(t)$ times the skill-intensity $\gamma(\Phi(t))$ times the labor share $1-\beta$. The statement that all of these functions are increasing in t follows from the result that $\gamma'(t), \Phi'(t), \Phi''(t) > 0$. $\square$

As argued in the previous subsection, the result that $R'(t) > 0$ derives from competition in the land rental markets. The comovement between $W(t)$ and $N(t)$ is an implication of spatial equilibrium, and comparative advantage leads to the result that $D'(t) > 0$. $\zeta'(t) > 0$ because the differential of the wage rate times the number of firms in the nonroutine sector is less than proportional to that of the market size of the sector.[20]

Figure 1 confirms the prediction of the model that the variables appearing in Proposition 1 are all positively correlated. Panel (a)-(f) in the figure plot skill intensity, the market size, the wage rate, the land value (as a proxy for the land rent), urban diversity, and the average establishment size of the nonroutine sector, respectively, against the total employment size, which corresponds to population in the model, for Metropolitan Statistical Areas (hereafter, MSAs) using the data explained in the next two paragraphs. Each of the panels also depicts a red line corresponding to the linear regression. We observe that there are positive correlations between the variables in Proposition 1. Spearman's rank correlation coefficients of the panels are 0.650, 0.994, 0.503, 0.318, 0.937, and 0.471, respectively.

The data on the variables except for the last three are based on the *May 2011 Occupational Employment Statistics* (hereafter, OES). Skill intensity of an MSA is defined as the share of detailed occupations listed in the *Standard Occupational Classification System* (hereafter, SOC), more than 20% of employments of which have a master's degree or a higher one, in the labor compensation of all detailed occupations reported by OES. The market size of an MSA is approximated by the total labor compensation using the fact that the two are proportional in the model. The wage rate of an MSA is the median of annual wage rates of detailed occupations in the MSA. The land value (per acre) is taken from Table A3 in Albouy and Ehrlich (2012). As for urban diversity and the establishment size, we calculate these as the total number of establishments and the number of employments per establishment, respectively, of "Professional, Scientific, and Technical Services" (in the *North American Industry Classification System*), using the *2011 Statistics U.S. Businesses* (hereafter, SUSB).

Here, we focus on the MSAs for which OES report non-military detailed occupations that cover more than 75% of the total employments in order to reduce a sample selection bias in the correlation between the total employment size and skill intensity.[21] This results in the sample size of 303 in the

[20] This is easily understood in the discrete version. Using the market clearing condition $W_j D_j \zeta_j = (1-\alpha)(1-\beta)\Gamma_j |\mathbb{T}_j| E$, we get

$$\frac{W_{j+1}}{W_j}\frac{D_{j+1}}{D_j}\frac{\zeta_{j+1}}{\zeta_j} = \frac{\Gamma_{j+1}}{\Gamma_j}\frac{|\mathbb{T}_{j+1}|}{|\mathbb{T}_j|}.$$

Then, using the differentials derived in the previous subsection, we obtain

$$\frac{\zeta_{j+1}}{\zeta_j} = \left(\frac{|\mathbb{T}_{j+1}|}{|\mathbb{T}_j|}\right)^{(1-\alpha)\beta} \quad \forall j.$$

[21] It should be noted that the use of the employment data of detailed occupations reported by OES might result in a selection bias in the relationship between the total employment size and skill intensity of MSAs. This is because functional

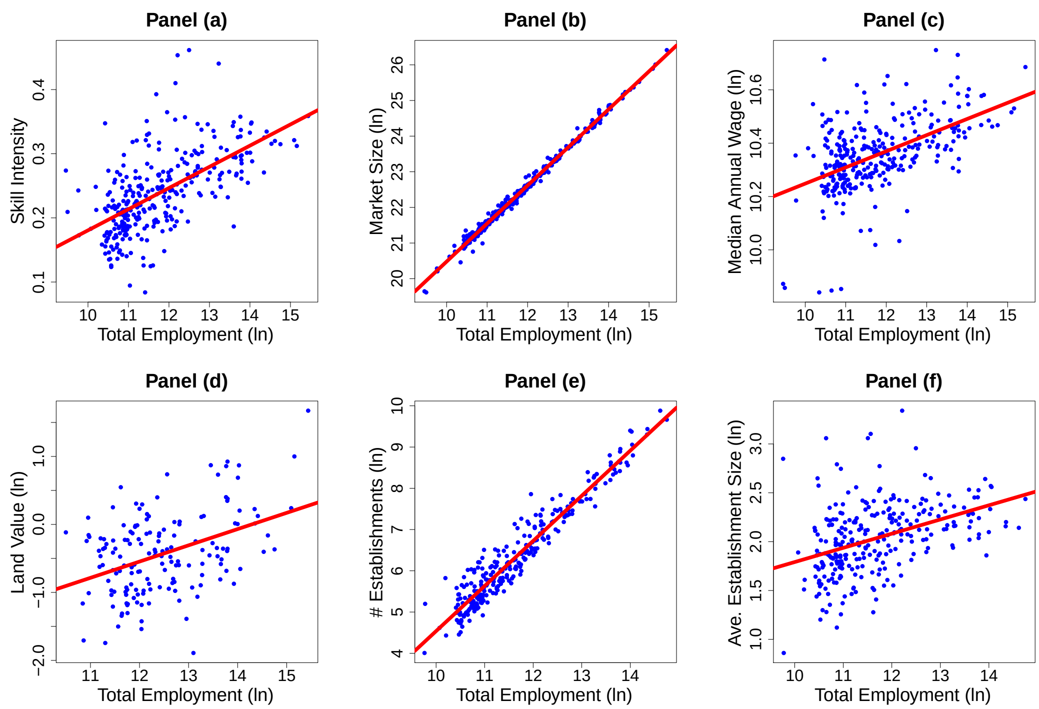

Figure 1: Correlations

cases which use OES only, i.e., Panel (a)-(c). The sample size of Panel (d) is reduced to 160, because the definitions of MSAs in Albouy and Ehrlich (2012) are not exactly the same as in our study, which focuses on MSAs listed in OES.[22] The sample sizes for the case of Panel (e) and (f), which use both OES and SUSB, are in the middle of the above two cases, 279 and 278, respectively.

3.3 Size Distribution of Cities

As has been argued in the literature, the upper tail of the population size of cities in the United States is well approximated by a power law or, more specifically, a Pareto distribution with a coefficient of one, the so-called Zipf's law (Gabaix and Ioannides, 2004; Gabaix, 2009). Economic mechanisms resulting in Zipf's law have also been proposed, ranging from random growth models such as Rossi-Hansberg and Wright (2007) and Duranton (2007) to static models such as Hsu (2012) and **?**.

The purpose of this subsection is therefore to relate our model to the size distribution of cities.

specialization across cities implies that smaller cities have a smaller share of workers working for the nonroutine sector, i.e., detailed occupations among which 20% of employments have a master's or a higher degree by definition, and thus OES might tend not to report the employment or wage data on those occupations in smaller MSAs more than proportionally to their city sizes relative to larger MSAs. If this is the case, skill intensity is biased downward for small-sized MSAs, leading to a seemingly positive correlation between the total employment size and skill intensity. Another approach of reducing this type of selection bias is to use the data of major (not detailed) occupations in SOC. In this case, OES report the employment and wage data of all occupations for all MSAs. Using this more aggregated data, we still observe a positive relationship between the total employment size and skill intensity, which is consistent with the prediction of the model.

[22] We match MSAs appearing in OES with those in Albouy and Ehrlich (2012) comparing the full names of MSAs.

More specifically, we derive the necessary and sufficient condition under which the associated sorting equilibrium exhibits a power law including Zipf's law as a special case. The next proposition states that this is equivalent to identifying the functional form of $\gamma(t)$ guaranteeing that the size distribution of cities obeys a power law.[23]

Proposition 2. *The size distribution of cities in the sorting equilibrium is characterized by a truncated Pareto distribution if and only if $\gamma(t)$ is given by*

$$\gamma(t) = \begin{cases} a - \{a^{\eta} - [a^{\eta} - (a-1)^{\eta}]t\}^{\frac{1}{\eta}} & \text{if } \eta \in \left(-\frac{1}{(1-\alpha)(1-\beta)}, 0\right) \cup (0, +\infty) \\ a - \exp\{\ln a - [\ln a - \ln(a-1)]t\} & \text{if } \eta = 0, \end{cases}$$

where

$$a \equiv \frac{\alpha(1-\beta)+\beta}{(1-\alpha)(1-\beta)\theta} > 1.$$

Furthermore, if $\eta = -\alpha/[(1-\alpha)(1-\beta)]$, the size distribution of cities is consistent with Zipf's law.

Proof. Only the essence is discussed here. The first part of the proposition is proven in four steps: The first step is to notice, from Proposition 1, that $N(t)$ obeys a power law if and only if $\Phi'(t)$ obeys a power law.

The second step is to show that $\Phi'(t)$ follows a power law if and only if λ defined by $\lambda \equiv G(\Phi(t))^{-1}$ obeys a power law. In order to prove this statement, we begin by differentiating $t = H[\Phi(t)]$ with respect to t to obtain $1 = H'[\Phi(t)]\Phi'(t)$. Using (17), we then obtain

$$\Phi'(t) \propto G(\Phi(t))^{-\frac{1}{(1-\alpha)(1-\beta)}}.$$

The third step is to prove that λ obeys a power law if and only if

$$\gamma'\left[\gamma^{-1}(B(\lambda))\right] \propto \lambda^{\tilde{\eta}}, \quad B(\lambda) \equiv \frac{\alpha(1-\beta)+\beta-\lambda^{-1}}{(1-\alpha)(1-\beta)\theta}. \tag{18}$$

Here, $\tilde{\eta}$ is defined by $\tilde{\eta} = \eta - 1$. In order to validate this statement, we first note that the density function $f_Z(z)$ of $z = \Phi(t)$ is given by $f_Z(z) = f_T[H(z)]H'(z) = H'(z)$, where the first equality uses the relationship between t and z, i.e., $t = H(z)$, and the second uses the uniformity of stage t. Then, using the relationship between λ and z given by $\lambda = G(z)^{-1}$ and the density function $f_Z(z)$, we obtain the density function $f_\Lambda(\lambda)$ of λ as follows:

$$f_\Lambda(\lambda) = f_Z\left[\gamma^{-1}(B(\lambda))\right] \frac{\partial}{\partial\lambda}\gamma^{-1}(B(\lambda)) \propto \frac{\lambda^{-\left[\frac{1}{(1-\alpha)(1-\beta)}+2\right]}}{\gamma'\left[\gamma^{-1}(B(\lambda))\right]}.$$

[23] One might interpret this kind of approach as searching for a knife-edge condition and thus conclude that the result presented below is not robust compared with previous theories. Testing the model, however, is not a purpose of this paper. Proposition is simply used for calibration of the model (Subsection 3.4).

α	β	σ	$\tilde{a}$	η
0.095	0.085	5.877	1.022	-0.561

Table 1: Calibrated Parameters

The final step is to show that (18) holds if and only if $\gamma(t)$ is given by the one specified in the proposition.

The second part of the proposition is demonstrated using the results discussed above and the fact that a random variable X_1 given by $X_1 = X_2^{\omega}$ ($\omega > 0$), where X_2 follows a Pareto distribution with coefficient $\delta > 0$, obeys a Pareto distribution with coefficient δ/ω. $\square$

3.4 Welfare Gain from Functional Specialization

Benchmark

As a numerical exercise, we compute the sorting equilibrium and the welfare gain from functional specialization for an example of parameter values, where the welfare gain is defined as the difference in log-utility between the laissez-faire case described in the previous subsections and the case where there are continuum of locations all of which hosts the whole range of stages of production *and* the same number of workers.

For this purpose, we specify the functional form of $\gamma(t)$ as follows:

$$\gamma(t) = \begin{cases} \tilde{a} - \{\tilde{a}^{\eta} - [\tilde{a}^{\eta} - (\tilde{a}-1)^{\eta}]\, t\}^{\frac{1}{\eta}} & \text{if } \eta \neq 0, \\ \tilde{a} - \exp\{\ln \tilde{a} - [\ln \tilde{a} - \ln(\tilde{a}-1)]\, t\} & \text{if } \eta = 0, \end{cases}$$

where $\tilde{a} > 1$ is a parameter that is different from a in Proposition 2. Thus, instead of assuming an exact power law, here we consider a slightly general $\gamma(t)$ that includes an exact power law as a special case. Based on a guess that the upper tail of market sizes $E(t) = \Phi'(t)$ of MSAs is well approximated by a power law, we expect that the above specification would successfully match the data.[24] A benchmark set of parameter values we use is presented in Table 1, which is obtained by the calibration described in Appendix G.

Figure 2 depicts the equilibrium Lorenz curve $\Phi(t)$ and the associated profiles of population $N(t)$, the wage rate $W(t)$, the land rent $R(t)$, urban diversity $D(t)$, and the average establishment $\zeta(t)$. As predicted by Proposition 1, all functions are monotonically increasing in t.

Letting $\bar{U}$ and $\bar{U}_s$ denote utilities in the laissez-faire and symmetric cases, respectively, the welfare

[24] Note that since the population $N(t)$ of city t is proportional to the market size $E(t)$ of city t (Proposition 1), if the distribution of population is closed to a power law, then the market size is also close to the power law.

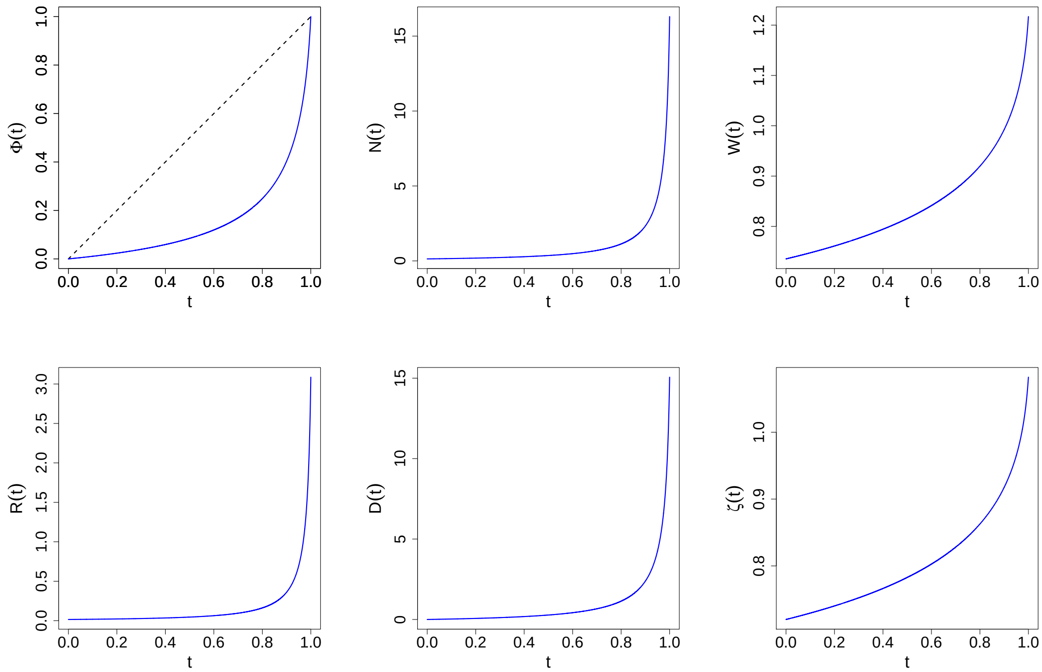

Figure 2: Lorenz Curve $\Phi(t)$ and Profiles, $N(t)$, $W(t)$, $R(t)$, $D(t)$, $\zeta(t)$

gain is given by[25]

$$\ln \bar{U} - \ln \bar{U}_s = \left\{ \ln \left[\frac{E(t)/N(t)}{R(t)^{\alpha}} \right] - \ln \left(\frac{1}{R_s^{\alpha}} \right) \right\} - (1-\alpha)\left(\ln P - \ln P_s\right),$$

where the first and the second terms represent the effects of the differences in the land-rent-adjusted average income and the price index of the final good, respectively. Variables with subscript A correspond to those in the symmetric case. Note that spatial equilibrium implies that the rent-adjusted income $E(t)/[N(t)R(t)^{\alpha}]$ is independent of t. Here, the price index is further decomposed as follows:

$$\begin{aligned} \ln P - \ln P_s \quad &= \quad \int_0^1 \beta[\ln R(t) - \ln R_s]\Phi'(t)dt + \int_0^1 (1-\beta)[\ln W(t) - \ln W_s]\Phi'(t)dt \\ &\quad - \int_0^1 \theta\left[\gamma(\Phi(t))\ln D(t) - \Gamma_s \ln D_s\right]\Phi'(t)dt, \end{aligned}$$

[25] Since both the total population and the measure of the continuum of locations are normalized to unity, the population size of each location in the symmetric case is equal to unity. In addition, given the normalization of the economy-wide income, i.e., $E = 1$, the per capita income of each location is also equal to one. Therefore, the symmetric outcome is equivalent to full agglomeration, i.e., workers and stages of production all concentrate in a single location.

Item	Value (%)
Welfare Gain	1.6
Rent-adjusted Income	-8.7
Final Good Price	10.3
Land Rent	-7.8
Wage Rate	-0.7
Urban Diversity	18.8

Table 2: Welfare Gain from Functional Specialization: Benchmark

Note: The welfare effect of each item is measured in terms of % increase in utility from the symmetric case.

where

$$R_s = \alpha(1-\beta)+\beta, \quad W_s = (1-\alpha)(1-\beta),$$
$$D_s = \frac{\theta(1-\alpha)\Gamma_s}{(1+\theta)f} R_s^{-\beta} W_s^{-(1-\beta)}, \quad \Gamma_s \equiv \int_0^1 \gamma(z)dz.$$

Table 2 reports that the welfare gain from functional specialization is about 1.6% of the utility level in the symmetric case, implying about 40% increase in life-time utility with log utility and the discount factor of 0.96.

The decomposition shows that the effect of the final good price slightly dominates that of the rent-adjusted income. In each of the two effects, increases in land rents due to agglomeration play important roles, i.e., reduce the rent-adjusted income and increase the price index of the final good. However, a much larger positive effect, 18.8% increase in utility, is generated by urban diversity, i.e., agglomeration increases the number of varieties of professional services.

Robustness

How does this result depend on the elasticity σ of substitution and the distribution of skill intensities of stages of production? In order to answer this question, fixing the share parameters, α and β, and $\tilde{a}$, a parameter less crucial than η in determining the distribution of skill intensity, we compute the welfare gain for each pair (σ,η) or (σ,Γ_s) given the one-to-one correspondence between η and Γ_s. Since $G(1)>0$ in Proposition 1 must hold, we need the following restriction to σ:

$$\sigma > \max\left\{1, \frac{1}{\alpha(1-\beta)+\beta}\right\} \approx 5.817.$$

The upper bound for σ is set to 10 which corresponds to the upper bound of the typical range of the elasticity of substitution in the literatures (Anderson and van Wincoop, 2004). The lower and upper bounds, -0.685 and 9.210, for η are set in such a way that the average skill intensity Γ_s belongs to a

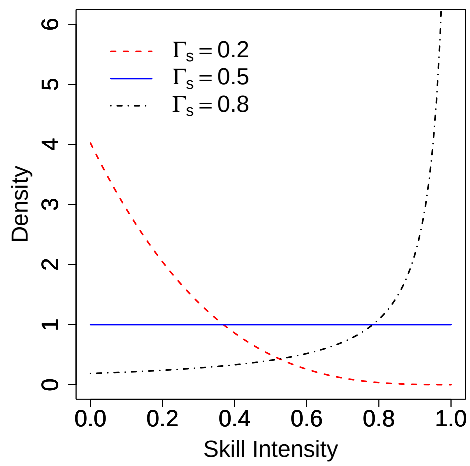

Figure 3: Density Function of Skill Intensity: $\Gamma_s = 0.2, 0.5, 0.8$

range $[0.1, 0.9]$, which covers the benchmark case, 0.882. Figure 3 depicts the density of skill intensity for three values of the average skill intensity Γ_s, showing that as Γ_s increases, the composition of production processes shifts toward those with higher skill intensity.

Figure 4 depicts the welfare gain and its decomposition for each pair (σ, Γ_s) of the elasticity of substitution and the average skill intensity as well as the benchmark case represented as a point $(5.877, 0.882)$. Although the absolute magnitude of each decomposed term changes significantly depending on parameter values, its relative magnitude is the same as in the benchmark case, and the magnitude of the overall welfare gain is robustly high, e.g., in a large region of (σ, Γ_s), the total welfare gain is more than 1%, implying 25% increase in life-time utility.

Numerical comparative statics also shows that for a fixed average skill intensity Γ_s, an increase in the elasticity σ of substitution reduces the welfare gain from functional specialization; and that for a fixed elasticity σ of substitution, there is an inversed-U-shaped relationship between the average skill intensity Γ_s and the overall welfare gain. The former result is interpreted that as an increase in the elasticity σ of substitution weakens the urban diversity effect, resulting in a lower overall welfare gain. The other effects, i.e., rent-adjusted income, land rent, and wage rate, are all in the opposite direction reflecting weaker market competition. The latter result is simply because as the density of skill intensity concentrates on either the upper or lower corners, i.e., $\Gamma_s \to 1$ or $\Gamma_s \to 0$, heterogeneity across production processes with respect to skill intensity disappears, leading to less benefit from specialization or trade. This interpretation is supported by the result that each decomposed term has the same inversed-U-shaped relationship. However, we note that the absolute magnitude of each

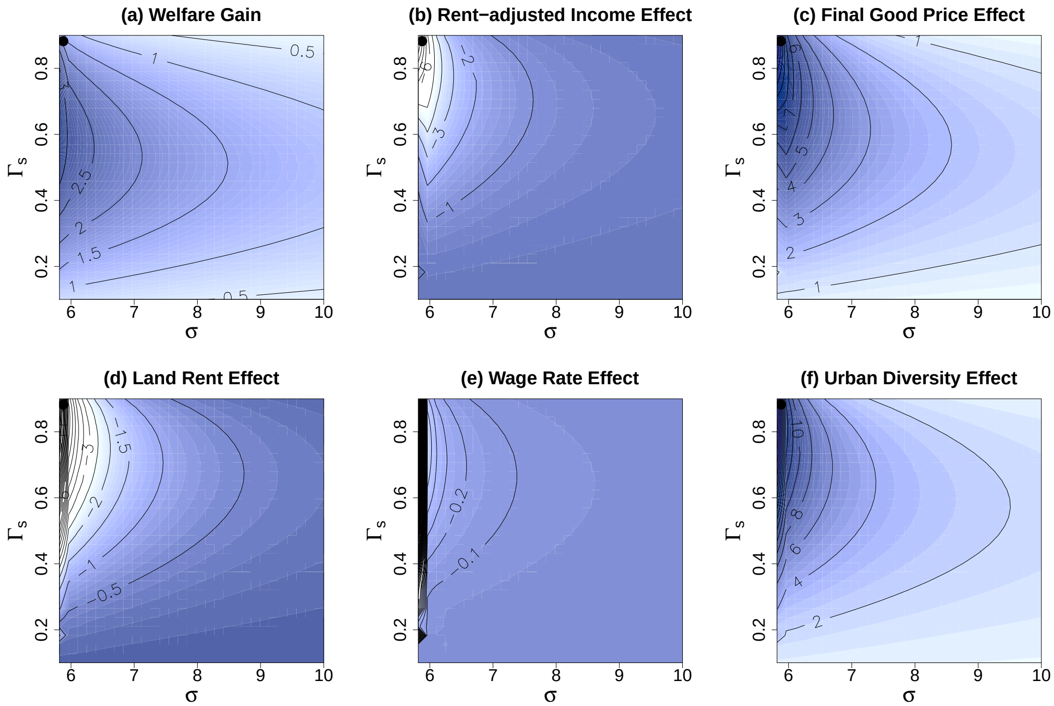

Figure 4: Welfare Gain from Functional Specialization

Note: Panels show the welfare effects in terms of % increase in utility from the symmetric case. A darker color within a fixed panel means a higher welfare gain. Colors in different panels are not comparable. The point $(5.877, 0.882)$ represents the benchmark case. For a given triplet $(\alpha, \beta, \tilde{a})$, there is a negative relationship between η and the average skill intensity Γ_s.

effect is strengthened when the density of skill intensity is skewed to the right.

4 Welfare Analysis of Income Redistribution Policy

Finally, in order to provide an example of use of the model, we conduct a welfare analysis of an income redistribution policy focusing on a second-best allocation, i.e., allocation achieved under the assumption that monopolistically competitive firms can set their own price levels, as in Pflüger and Tabuchi (2010) who follow Helpman (1998).[26] This kind of analysis allows us to investigate the

[26] The possible source of inefficiency here is the composition of the routine and the nonroutine local services sectors. As specified in Section 2, the latter sector is characterized by monopolistic competition á la Dixit and Stiglitz (1977), and if the final good firm uses the nonroutine local services only, i.e., $\gamma(t) \to 1$ for all t, the associated laissez-faire outcome is efficient (See the review in Dhingra and Morrow (2012) and their extension to the environments with heterogeneous firms.) One important future direction of research could therefore be to consider a class of utility or production functions with variable elasticity of substitution. Since it is immediate to introduce preference and production externalities into the model, developing a method also applicable to environments with variable elasticity of substitution allows us to more deeply understand welfare gains from optimal policy in a system of cities and the interactions among various sources of inefficiency.

relationship between room of policy intervention and technological environment in a framework of a system of cities.[27] For this purpose, we describe a lump-sum transfer scheme in Subsection 4.1 and derive the modified fundamental equation in Subsection 4.2, respectively. We then conduct numerical comparative statics to obtain implications of production technology, specifically, the pair (σ, Γ_s) of the elasticity of substitution and the average skill intensity, for income redistribution in Subsection 4.3.

4.1 Income Redistribution Policy Rule

We begin with a finite number of locations J. Let $E_{b,j}$ and $E_{a,j}$ denote before- and after- tax regional income, respectively, i.e., $E_{b,j} = N_j(W_j + \bar{R}_j) = W_j N_j + R_j$. Then, let us introduce a government that implements the following income redistribution policy rule:[28]

$$E_{a,j} = (1-\tau)E_{b,j} + \frac{\tau E}{J} \quad \forall j,$$

where $\tau \in [0,1]$ is the proportional income tax rate; the case where $\tau = 0$ corresponds to the laissez-faire economy described and analyzed in the previous sections. That is, by levying a tax on individuals' incomes $E_{b,j}$ that are distributed by the market, the policy makes the equilibrium outcome more equalized than in the laissez-faire case. In addition, the lump-sum transfers among cities make this dispersion effect more effective because as individuals concentrate in a city, the *per capita lump-sum transfer within the city*, i.e., $\tau E/(JN_j)$, decreases.

In order to reflect this government policy rule, the land market clearing and free-migration conditions should be modified as

$$\begin{aligned} R_j &= (1-\alpha)\beta|\mathbb{T}_j|E + \alpha E_{a,j}, \\ \frac{E_{a,j+1}/N_{j+1}}{E_{a,j}/N_j} &= \left(\frac{R_{j+1}}{R_j}\right)^{\alpha}, \end{aligned}$$

where $E_{a,j}/N_j$ represents the after-tax per capita income at location j, implying

$$\begin{aligned} \frac{R_{j+1}}{R_j} &= \frac{(1-\alpha)[\alpha(1-\beta)(1-\tau)+\beta]|\mathbb{T}_{j+1}| + \alpha\tau J^{-1}}{(1-\alpha)[\alpha(1-\beta)(1-\tau)+\beta]|\mathbb{T}_j| + \alpha\tau J^{-1}}, \\ \frac{N_{j+1}}{N_j} &= \frac{(1-\alpha)(1-\tau)|\mathbb{T}_{j+1}| + \tau J^{-1}}{(1-\alpha)(1-\tau)|\mathbb{T}_j| + \tau J^{-1}} \left\{ \frac{(1-\alpha)[\alpha(1-\beta)(1-\tau)+\beta]|\mathbb{T}_{j+1}| + \alpha\tau J^{-1}}{(1-\alpha)[\alpha(1-\beta)(1-\tau)+\beta]|\mathbb{T}_j| + \alpha\tau J^{-1}} \right\}^{-\alpha}, \quad \forall j. \end{aligned}$$

[27] Although not pursued in this paper, technological externalities in both preference and technology are easily introduced into the model with the same degree of analytical tractability, allowing us to understand the relationship between the composition of externalities and the efficiency of market allocation.

[28] The government here cannot control the distribution of population across cities directly. Instead, the choice variable of the government is the income tax rate τ only. For a given tax rate, the population distribution of cities is implicitly determined by free migration expressed as the equalization of utility across cities. This is in contrast to the analysis in Pflüger and Tabuchi (2010) where the government controls the population distribution across (two) locations, and lump-sum transfers are automatically implied by the free-migration condition.

4.2 Modified Fundamental Equation

Using the modified equilibrium conditions, we obtain the following modified fundamental equation:

$$\left(\frac{|\mathbb{T}_{j+1}|}{|\mathbb{T}_j|}\right)^{1-\beta[1+\theta\gamma(T_j)]}\left\{\frac{(1-\alpha)[\alpha(1-\beta)(1-\tau)+\beta]|\mathbb{T}_{j+1}|+\alpha\tau J^{-1}}{(1-\alpha)[\alpha(1-\beta)(1-\tau)+\beta]|\mathbb{T}_j|+\alpha\tau J^{-1}}\right\}^{[\alpha(1-\beta)+\beta][1+\theta\gamma(T_j)]}$$
$$\times\left[\frac{(1-\alpha)(1-\tau)|\mathbb{T}_{j+1}|+\tau J^{-1}}{(1-\alpha)(1-\tau)|\mathbb{T}_j|+\tau J^{-1}}\right]^{-(1-\beta)[1+\theta\gamma(T_j)]}=\left(\frac{\Gamma_{j+1}}{\Gamma_j}\right)^{\theta\gamma(T_j)}\qquad\forall j\in\{1,2,\cdots,J-1\}.$$

Replacing J^{-1} with Δt as J becomes sufficiently large and using the method of asymptotic expansion, we obtain the fundamental equation in the limiting case:

$$g(\Phi,\Phi')\Phi''=\theta\gamma'(\Phi)\Phi',$$

where

$$\begin{aligned}
g(\Phi,\Phi') &\equiv \frac{1-\beta[1+\theta\gamma(\Phi)]}{\Phi'}+\frac{[\alpha(1-\beta)+\beta][1+\theta\gamma(\Phi)]}{\Phi'+\tilde{\tau}_1}-\frac{(1-\beta)[1+\theta\gamma(\Phi)]}{\Phi'+\tilde{\tau}_2},\\
\tilde{\tau}_1 &\equiv \frac{\alpha\tau}{(1-\alpha)[\beta+\alpha(1-\beta)(1-\tau)]},\qquad \tilde{\tau}_2\equiv\frac{\tau}{(1-\alpha)(1-\tau)}.
\end{aligned}$$

With $\tau=0$, this fundamental equation reduces to the one in the laissez-faire case.

This case is clearly complicated relative to the previous one, and thus we must resort to a numerical method such as the fourth-order Runge-Kutta method. It should also be noted that a solution to this ODE is not necessarily a sorting equilibrium because, unlike in the previous case, we do not have any analytical characterization stating that $\Phi(t)$ is a positive, strictly convex function. Instead, after solving the fundamental equation numerically, we need to check whether the function $\Phi(t)$ has this property.[29] However, an advantageous property of this ODE is that it is very easy to check the uniqueness of a solution for a given set of parameters. This is because we know that $\Phi(t)$ is a Lorenz curve, and thus, $\Phi'(0)$ must be less than one. In addition, we have boundary conditions, $\Phi(0)=0$ and $\Phi(1)=1$. Therefore, by simply discretizing the interval $(0,1)$ and using each point as an initial guess for $\Phi'(0)$, we can obtain all the possible solutions to the ODE with the help of forward shooting.

[29] Using the land rent and wage rate functions stated in the next subsection, we have

$$P_L(t)\propto\left[(1-\alpha)(1-\tau)+\tau/\Phi'(t)\right]^{-(1-\beta)}\left\{(1-\alpha)[\alpha(1-\beta)(1-\tau)+\beta]\Phi'(t)+\alpha\tau\right\}^{\alpha(1-\beta)+\beta}\quad\text{for all } t\in[0,1],$$

which implies that the price $P_L(t)$ of the routine good and thus market competition are increasing in t if ($\Phi'(t)>0$ and) $\Phi''(t)>0$ for all $t\in[0,1]$. Then, the argument in Subsection 3.1 suggests that if a solution to the fundamental equation exists, we should have a $D(t)$ that is increasing in t, which is the implication of endogenous comparative advantage of cities with higher t in the nonroutine sector. That is, if a solution to the fundamental equation exists and if $\Phi'(t)>0$ and $\Phi''(t)>0$ for all $t\in[0,1]$, the solution is actually a sorting equilibrium.

4.3 Optimal Income Tax Rate

Benchmark

For the benchmark set of parameter values in Table 1, it was verified that for each $\tau \in [0,1]$, there is a unique solution to the fundamental equation that has the property $\Phi'(t), \Phi''(t) > 0$ for all t, implying that the solution is actually a sorting equilibrium.

Given the unique solution, the equilibrium utility is then calculated by

$$\ln \bar{U} = \ln\left[\frac{E_a(t)}{N(t)R(t)^{\alpha}}\right] - (1-\alpha)\ln P \qquad \forall t \in [0,1],$$

where

$$\begin{aligned}
E_a(t) &= [1-\alpha(1-\tau)]^{-1}[(1-\alpha)(1-\tau)\Phi'(t)+\tau], \\
\ln P &= \int_0^1 [\beta \ln R(t) + (1-\beta)\ln W(t) - \theta\gamma(\Phi(t))\ln D(t)]\,\Phi'(t)dt, \\
R(t) &= [1-\alpha(1-\tau)]^{-1}\left\{(1-\alpha)[\alpha(1-\beta)(1-\tau)+\beta]\Phi'(t)+\alpha\tau\right\}, \\
W(t) &= (1-\alpha)(1-\beta)\frac{\Phi'(t)}{N(t)}, \\
N(t) &= e^{c_0}\left[(1-\alpha)(1-\tau)\Phi'(t)+\tau\right]\left\{(1-\alpha)\left[\alpha(1-\beta)(1-\tau)+\beta\right]\Phi'(t)+\alpha\tau\right\}^{-\alpha}, \\
D(t) &= \frac{\theta}{(1+\theta)f}(1-\alpha)\gamma(\Phi(t))\Phi'(t)R(t)^{-\beta}W(t)^{-(1-\beta)}
\end{aligned}$$

for all $t \in [0,1]$, where e^{c_0} is computed by integrating $N(t)$ over $[0,1]$ and using the normalization condition, i.e., $\int_0^1 N(t)dt = 1$. The implication of spatial equilibrium is that the first term in the above equation is independent of location or task t such that utility is equalized across locations.[30]

Panel (a) in Figure 5 then depicts the consequent relationship between tax rate τ and the natural logarithm of the equilibrium utility $\ln \bar{U}$, which contains a unique peak at the optimal income tax rate of about 0.6%.[31] Therefore, the laissez-faire outcome, i.e., the case of $\tau = 0$, is not efficient. Its allocation is Pareto-dominated by the sorting equilibrium with a positive but sufficiently low income tax rate. The optimal tax rate achieves about 0.09% higher equilibrium utility as reported in Table 3, which corresponds to about 2.25% increase in life-time utility with log utility and the discount factor of 0.96.

The unique peak is a result of the combination of two relationships: (i) the welfare effect of the land-rent-adjusted income is monotonically increasing in the tax rate τ; and (ii) the welfare effect of the final good price is monotonically decreasing in the tax rate τ. Panel (a) and (b) in Figure 5

[30] Using the normalization condition on the population size, i.e., $\int_0^1 N(t)dt = 1$, it is easily shown that $E_a(t)/[N(t)R(t)^{\alpha}] = [1-\alpha(1-\tau)]^{-(1-\alpha)}e^{-c_0}$ for all $t \in [0,1]$.

[31] This small number is not a computational error as suggested by Panel (a) in Figure 5 and Figure 6, the latter of which shows the optimal tax rate for each pair (σ, Γ_s) of the elasticity of substitution and the average skill intensity.

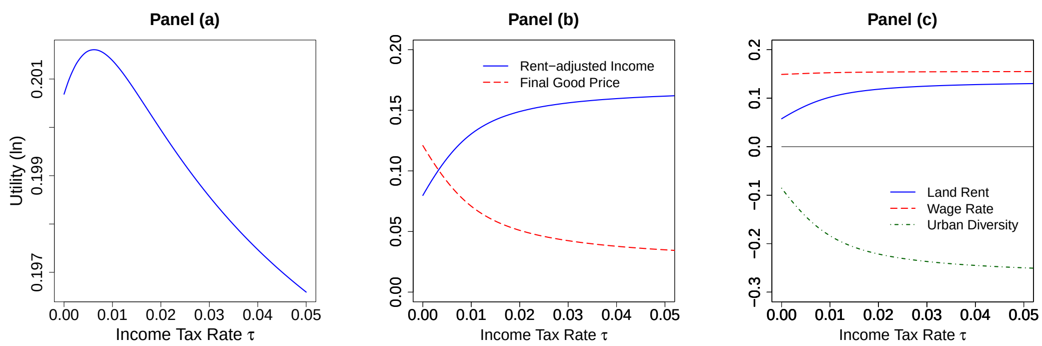

Figure 5: Income Tax τ and the Equilibrium Utility (ln): Benchmark

Note: The vertical axes of all panels are measured in terms of utility. For ease of visualization, the horizontal axis covers only a part of τ, i.e., $[0, 0.05]$. For higher τ's, the relative magnitude of each term does not change.

Item	Value (%)
Welfare Gain	0.09
Rent-adjusted Income	3.63
Final Good Price	-3.53
Land Rent	3.18
Wage Rate	0.24
Urban Diversity	-6.95

Table 3: Welfare Gain from Optimal Income Redistribution Policy: Benchmark

Note: The welfare effect of each item is measured in terms of % increase in utility from the laissez-faire case.

suggest that for a sufficiently low tax rate, the former effect dominates, while the latter does for a sufficiently high tax rate. As the decomposition of the welfare effect of the final good price shows (Panel (c) in Figure 5), the latter relationship derives from the fact that the welfare effect of urban diversity increases as the spatial distribution of economic activity concentrates, i.e., as the tax rate τ decreases. The wage rate and the land rent both become lower when economic activity is dispersed, i.e., τ is higher, suggesting that the former relationship simply reflects the relationship between the welfare effect of the land rent and the tax rate. The decomposition of the welfare gain from the optimal income tax reported in Table 3 confirms such relative magnitudes of welfare effects.

Robustness

Numerical comparative statics shows that the laissez-fair outcome is inefficient in the most of an empirically plausible parameter range. The left panel in Figure 6 depicts the optimal income tax rate corresponding to each pair (σ, Γ_s) of the elasticity of substitution and the average skill intensity in percentage terms.[32] It shows that the optimal tax rate is positive in the range, implying that the laissez-fair is characterized as excess agglomeration. This result agree with the one reported by Pflüger and Tabuchi (2010), who, using a New Economic Geography model, argue that with land use in both consumption and production, the spatial configuration is characterized by either efficient or excess agglomeration.

Although the optimal tax rate changes depending on parameter values, we observe systematic variations: For a fixed elasticity σ of substitution, the optimal tax rate is monotonically decreasing in the average skill intensity Γ_s and approaching to zero as Γ_s increases to one; and, for a fixed average skill intensity Γ_s, the optimal tax rate is monotonically increasing in the elasticity σ of substitution if Γ_s is sufficiently high and has an inversed-U-shaped relationship otherwise (the right panel in Figure 6). The former result is interpreted as the combination of two key results. On one hand, if the average skill intensity Γ_s is sufficiently close to unity, the density of skill intensity is highly skewed to the right (Figure 3), making almost every production processes to use inputs produced by the monopolistically competitive sector and thus suggesting that the laissez-faire is close to the second-best allocation by definition. On the other hand, as the average skill intensity Γ_s decreases, production technology exhibits constant returns more, and, given limited land areas generating a dispersion force, it is implied that the optimal allocation is more dispersed than the laissez faire.

The first part of the latter result is interpreted in a similar way to the former case. That is, as the elasticity σ of substitution increases, production technology exhibits constant returns to a greater extent, favoring more dispersed economic activity as the optimal. In the second part, where the average skill intensity Γ_s is not close to unity, the laissez-faire outcome is distant from the second-best allocation by definition, implying more room for income redistribution. This is especially relevant for lower σ which is associated with a greater concentration of economic activity and thus sever negative effects of higher land rents on welfare. Then income redistribution can mitigate those negative effects even if losing opportunities urban diversity provides. This interpretation is consistent with the result that the region of the elasticity of substitution, where the optimal tax rate is decreasing in σ, expands as the average skill intensity Γ_s decreases (the left panel of Figure 6). The analysis here suggests that the optimal income tax rate is determined by the interplay between two opposing components: the distance of production technology from the dominance of monopolistic competition in terms of the distribution of skill intensity and that from the dominance of constant returns to scale in terms of the elasticity of substitution.

The above results then provide an important implication for a modern developed economy: As Michaels et al. (2013) shows, we observe an increasing importance of interactive activities in developed countries such as the United States. In our model, this change in the nature of economic activity

[32] We note that the analysis here is not exhaustive in the sense that we focus on a range of (σ, Γ_s), where computational errors associated with numerical solutions to the modified fundamental equation and numerical integration used when calculating equilibrium utility are sufficiently small such that the objective function, i.e., equilibrium (log) utility as a function of the income tax rate τ, is relatively well approximated with cubic spline interpolation, an approximation method used together with the golden section search when computing the optimal tax rate.

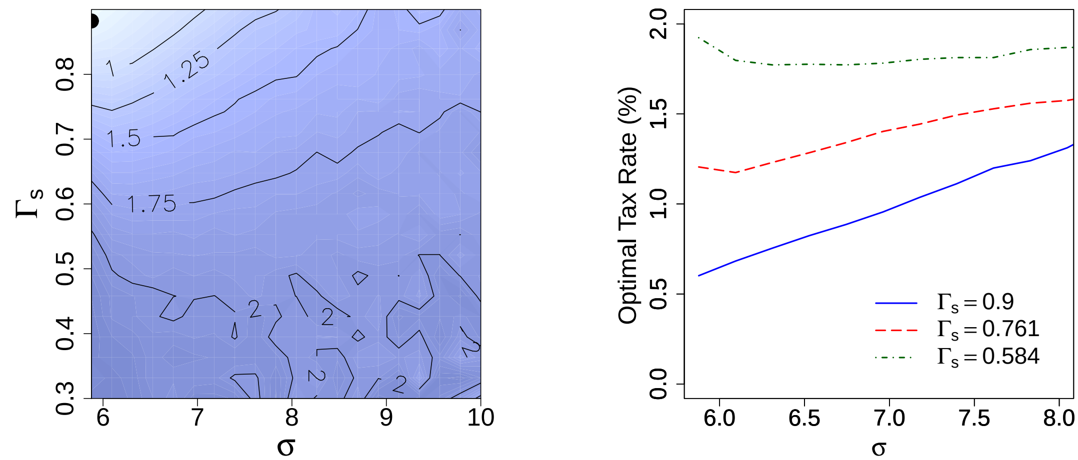

Figure 6: Optimal Income Tax Rate, $100 \times \tau$ (%)

Note: Left panel: A darker color means a higher tax rate; The point $(5.877, 0.882)$ represents the benchmark case in which case the optimal tax rate is about 0.6%; In the south east region, computational errors are relatively large due to a smaller role of monopolistic competition implying equilibrium utility is less elastic with respect to parameters (σ, Γ_s). Right panel: The range of σ above 8 is excluded because computational errors are relatively large.

can be interpreted in two ways, i.e., a right shift of the density of skill intensity or an increase in the degree of product differentiation. The comparative statics then suggests that the desirable income redistribution policies are not necessarily the same in both case.

As for the welfare gain from income redistribution, numerical results show that the relative magnitude of each decomposed effect is the same as in the benchmark case, implying that although the absolute magnitude of each effect varies depending on parameter values, a positive overall welfare gain from income redistribution is a result of weakened negative effects of land rents compensating lower urban diversity. Panel (a) in Figure 7 shows the overall welfare gain from the optimal income redistribution policy, measured in terms of the increase in the equilibrium utility from the laissez-faire, is positive but limited to a similar degree as the benchmark case,[33] being in contrast to decomposed effects, Panel (b)-(f), which vary widely depending on parameter values. However, signs and relative magnitudes of those effects are the same as in the benchmark case. We also note that the welfare gain from income redistribution has the same property as that from functional specialization, i.e., gains tend to be larger when the distribution of skill intensity is close to a uniform.

[33]This result is consistent with that in Desmet and Rossi-Hansberg (2013) who use a totally different, neo-classical framework with Marshallian externalities in both production and preference, showing that eliminating efficiency differences across cities results in 1.2% increase in the U.S.

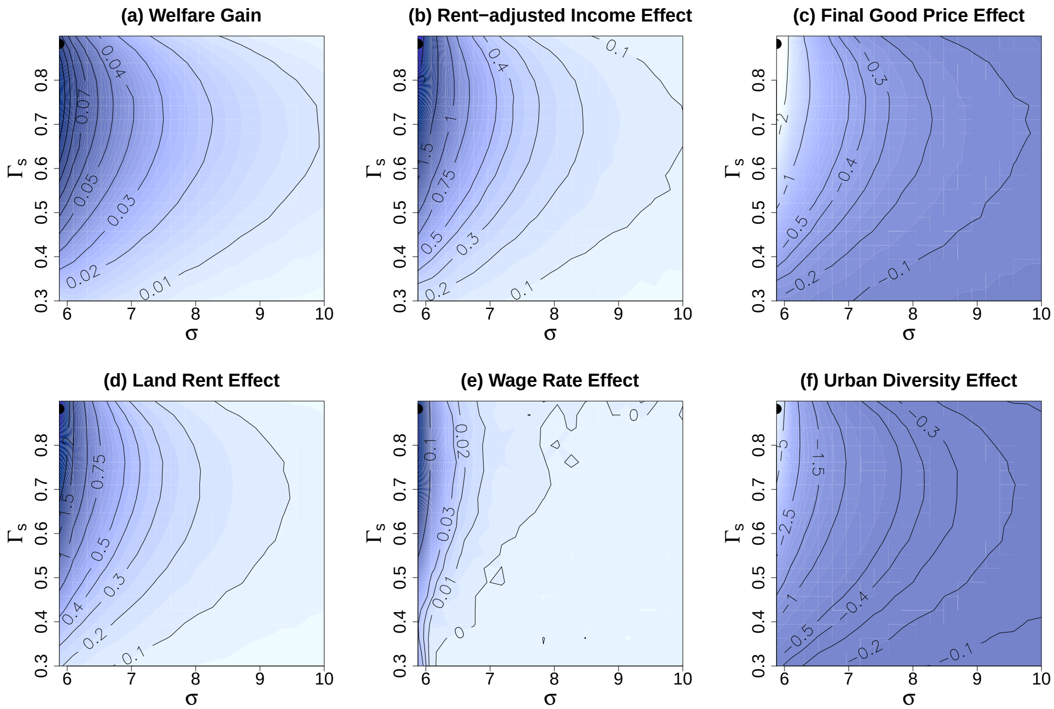

Figure 7: Welfare Gain from Optimal Income Redistribution Policy and Optimal Tax Rate τ

Note: Panels show the welfare effects of an optimal tax in terms of % increase in utility from the laissez-faire. A darker color within a fixed panel means a higher welfare gain or a higher tax rate. Colors in different panels are not comparable. The point $(5.877, 0.882)$ represents the benchmark case. For a given triplet $(\alpha, \beta, \tilde{a})$, there is a negative relationship between η and the average skill intensity Γ_s.

5 Conclusion

To formalize functional specialization of cities, this paper develops a static equilibrium model of a system of cities in which ex ante identical locations specialize ex post in different sets of stages of production, resulting in a unique, non-degenerate size distribution of cities with the comovement of income, population, the wage rate, the land rent, the average establishment size in the the nonroutine local services sector, and urban diversity as observed for the U.S. cities. The model is fairly tractable in that the necessary and sufficient condition for the city size distribution to obey a power law is analytically obtained. The analysis then takes a step further to analyses of welfare gains from functional specialization and optimal income redistribution. Even if focusing on a static environment, the welfare gain from functional specialization is large, but that from income redistribution is limited. The latter analysis also provides an important implication of an increasing importance of interactive activities in a modern developed economy for desirable income redistribution.

Given its simplicity, we hope that the model serves as a tool for further analyses of issues related to specialization in a system of cities. Although not pursued in this paper, the model seems to be easily extended to introduce externalities both in preference and production, elements often employed by recent quantitative studies such as Desmet and Rossi-Hansberg (2013). In addition, the model can also accommodate exogenous amenity or productivity differentials, which are also assumed in previous studies, at least in two extreme cases: the ordering of cities in terms of skill intensity is exactly the same as in the case of exogenous differentials; or it is exactly the contrary. The former case is used as an approximation if exogenous amenities are important in determining the city size distribution, e.g., Behrens et al. (2014b), while the latter could be used if interested in the "first nature v.s. second nature" type of argument. Intermediate cases are complicated to analyze because there are multiple distribution in equilibrium. In any case, extensions allow us to more deeply understand the relationship between optimal income redistribution and the above elements as well as that between comparative advantage and agglomeration and dispersion forces.

References

ALBOUY, D. AND G. EHRLICH, "Metropolitan Land Values and Housing Productivity," NBER Working Papers 18110, National Bureau of Economic Research, Inc, May 2012.

ANDERSON, J. E. AND E. VAN WINCOOP, "Trade Costs," *Journal of Economic Literature* 42 (September 2004), 691–751.

BEHRENS, K., G. DURANTON AND F. ROBERT-NICOUD, "Productive Cities: Sorting, Selection, and Agglomeration," *Journal of Political Economy* 122 (2014a), 507 – 553.

BEHRENS, K., G. MION, Y. MURATA AND J. SUEDEKUM, "Spatial frictions," Technical Report, 2014b.

DAVIS, J. C. AND J. V. HENDERSON, "The agglomeration of headquarters," *Regional Science and Urban Economics* 38 (September 2008), 445–460.

DAVIS, M. A. AND F. ORTALO-MAGNE, "Household Expenditures, Wages, Rents," *Review of Economic Dynamics* 14 (April 2011), 248–261.

DESMET, K. AND E. ROSSI-HANSBERG, "Urban Accounting and Welfare," *American Economic Review* (2013), forthcoming.

DHINGRA, S. AND J. MORROW, "Monopolistic competition and optimum product diversity under firm heterogeneity," *London School of Economics, mimeograph* (2012).

DIXIT, A. K. AND G. M. GROSSMAN, "Trade and Protection with Multistage Production," *Review of Economic Studies* 49 (October 1982), 583–94.

DIXIT, A. K. AND J. E. STIGLITZ, “Monopolistic Competition and Optimum Product Diversity,” *American Economic Review* 67 (June 1977), 297–308.

DORNBUSCH, R., S. FISCHER AND P. A. SAMUELSON, “Comparative Advantage, Trade, and Payments in a Ricardian Model with a Continuum of Goods,” *American Economic Review* 67 (December 1977), 823–839.

DURANTON, G., “Urban Evolutions: The Fast, the Slow, and the Still,” *American Economic Review* 97 (March 2007), 197–221.

DURANTON, G. AND D. PUGA, “Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products,” *American Economic Review* 91 (December 2001), 1454–1477.

———, “From sectoral to functional urban specialisation,” *Journal of Urban Economics* 57 (March 2005), 343–370.

EECKHOUT, J., “Gibrat’s Law for (All) Cities,” *American Economic Review* 94 (December 2004), 1429–1451.

FEENSTRA, R. C. AND G. H. HANSON, “Foreign Investment, Outsourcing, and Relative Wages,” in R. C. Feenstra, G. M. Grossman and D. A. Irwin, eds., *The Political Economy of Trade Policy: Papers in Honor of Jagdish Bhagwati* (Cambridge, MA: MIT Press, 1996), 89–127.

FELDMAN, M. P. AND D. B. AUDRETSCH, “Innovation in cities: Science-based diversity, specialization and localized competition,” *European Economic Review* 43 (February 1999), 409–429.

GABAIX, X., “Power Laws in Economics and Finance,” *Annual Review of Economics* 1 (05 2009), 255–294.

GABAIX, X. AND Y. M. IOANNIDES, “The evolution of city size distributions,” in J. V. Henderson and J. F. Thisse, eds., *Handbook of Regional and Urban Economics*volume 4 of *Handbook of Regional and Urban Economics*, chapter 53 (Elsevier, 2004), 2341–2378.

GROSSMAN, G. M. AND E. ROSSI-HANSBERG, “Trading Tasks: A Simple Theory of Offshoring,” *American Economic Review* 98 (December 2008), 1978–97.

HELPMAN, E., “The size of regions,” in D. Pines, E. Sadka and I. Zilcha, eds., *Topics in Public Economics Theoretical and Applied Analysis* (Cambridge: Cambridge University Press, 1998), 33–54.

HENDERSON, J. V., “The Sizes and Types of Cities,” *American Economic Review* 64 (September 1974), 640–56.

HSU, W., “Central Place Theory and City Size Distribution,” *Economic Journal* 122 (09 2012), 903–932.

JACOBS, J., *The economy of cities*, Vintage international (Random House, 1969).

KOHLER, W., “International outsourcing and factor prices with multistage production,” *Economic Journal* 114 (03 2004), C166–C185.

MATSUYAMA, K., “Endogenous Ranking and Equilibrium Lorenz Curve Across (ex ante) Identical Countries,” *Econometrica* 81 (09 2013), 2009–2031.

MICHAELS, G., F. RAUCH AND S. J. REDDING, “Task Specialization in U.S. Cities from 1880-2000,” CEP Discussion Papers dp1186, Centre for Economic Performance, LSE, February 2013.

PFLÜGER, M. AND T. TABUCHI, “The size of regions with land use for production,” *Regional Science and Urban Economics* 40 (November 2010), 481–489.

ROSSI-HANSBERG, E. AND M. L. J. WRIGHT, “Urban Structure and Growth,” *Review of Economic Studies* 74 (04 2007), 597–624.

VALENTINYI, A. AND B. HERRENDORF, “Measuring Factor Income Shares at the Sector Level,” *Review of Economic Dynamics* 11 (October 2008), 820–835.

YI, K.-M., “Can Vertical Specialization Explain the Growth of World Trade?,” *Journal of Political Economy* 111 (February 2003), 52–102.

A Why the ordering, $0 = T_1 < T_2 < \cdots < T_J = 1$, holds

The following argument shows that the ordering of thresholds assumed in the text is only the relevant case.

1. Suppose without loss of generality that $|\mathbb{T}_j| < |\mathbb{T}_{j+1}|$ for all j.[34] Then, we can show that $\mathbb{T}_j \cap \mathbb{T}_{j'} = \emptyset$ for all $j \neq j'$. This is because if $\mathbb{T}_j \cap \mathbb{T}_{j'} \neq \emptyset$ for some $j \neq j'$, then it must hold that

$$\frac{P_j(t)}{P_{j'}(t)} = \left(\frac{P_{n,j}}{P_{n,j'}}\right)^{\gamma(t)} \left(\frac{P_{r,j}}{P_{r,j'}}\right)^{1-\gamma(t)} = 1 \qquad \forall t \in \mathbb{T}_j \cap \mathbb{T}_{j'}.$$

Since $\gamma(t)$ is strictly increasing, this is possible only if

$$\frac{P_{n,j}}{P_{n,j'}} = \frac{P_{r,j}}{P_{r,j'}} = 1.$$

[34] Of course, it is possible that $|\mathbb{T}_j| = |\mathbb{T}_{j'}|$ for some j and j'. However, if a shock to the economy is drawn from some continuous distribution, such a case occurs with probability zero.

However, this contradicts the fact that $|\mathbb{T}_j| \neq |\mathbb{T}_{j'}|$, given that

$$\frac{P_{r,j}}{P_{r,j'}} = \left(\frac{R_j}{R_{j'}}\right)^{\beta}\left(\frac{W_j}{W_{j'}}\right)^{1-\beta} = \left(\frac{|\mathbb{T}_j|}{|\mathbb{T}_{j'}|}\right)^{\alpha(1-\beta)+\beta}.$$

2. Furthermore, it holds that $\cup_j \mathbb{T}_j = [0,1]$. This is simply due to the fact that for any fixed $t \in [0,1]$, there exists j that minimizes the costs of processing stage t.

3. Therefore, we can say that there exists a sequence of thresholds $\{T_j\}$, implying that we can use the condition for comparative advantage, i.e., Condition 5 in the definition of an equilibrium. It is then implied that $\Gamma_j < \Gamma_{j+1}$ for all j.

4. Since $\mathbb{T}_j \cap \mathbb{T}_{j'} = \emptyset$ for any $j \neq j'$ and it must hold that $|\mathbb{T}_j| < |\mathbb{T}_{j+1}|$ and $\Gamma_j < \Gamma_{j+1}$ for all j, the sequence $\{T_j\}$ must be increasing. Otherwise, there exists $j < j'$, i.e., $|\mathbb{T}_j| < |\mathbb{T}_{j'}|$, such that $\Gamma_j > \Gamma_{j'}$, which contradicts $|\mathbb{T}_j| < |\mathbb{T}_{j'}|$ under the assumption that $(1-\alpha)(1-\beta)\theta\gamma(t) < \alpha(1-\beta)+\beta$ for all $t \in [0,1]$.

B Derivation of the Fundamental Equation

Reproduce the fundamental equation corresponding to the discrete case:

$$\left(\frac{T_{j+1}-T_j}{T_j-T_{j-1}}\right)^{\alpha(1-\beta)+\beta-(1-\alpha)(1-\beta)\theta\gamma(T_j)} = \left(\frac{\Gamma_{j+1}}{\Gamma_j}\right)^{\theta\gamma(T_j)}, \tag{19}$$

where

$$\Gamma_j \equiv \frac{1}{T_j - T_{j-1}}\int_{T_{j-1}}^{T_j} \gamma(t)dt.$$

In the limit, i.e., $J \to \infty$, each location specializes in a single stage of production. Therefore, each location is characterized by its stage $t \in [0,1]$. So, let $\Phi(t)$ denote the income Lorenz curve, i.e., $\Phi(t)$ is equal to the accumulated income of locations that belong to $[0,t]$.

Using this Lorenz curve, we then have the following approximation:

$$\frac{T_{j+1}-T_j}{T_j-T_{j-1}} = \frac{\Phi(t+\Delta t)-\Phi(t)}{\Phi(t)-\Phi(t-\Delta t)}, \tag{20}$$

where t and Δt correspond to j/J and $1/J$, respectively. Using the asymptotic expansion, the right hand side is expressed by

$$\frac{\Phi(t+\Delta t)-\Phi(t)}{\Phi(t)-\Phi(t-\Delta t)} = 1 + \frac{\Phi''(t)}{\Phi'(t)}\Delta t + o(|\Delta t|), \tag{21}$$

where $o(\cdot)$ denotes the *little o*, i.e., as $|\Delta t| \to 0$, $o(|\Delta t|)/|\Delta t| \to 0$.

We also have the following approximation:

$$\begin{aligned}
\Gamma_{j+1} &= \frac{\int_{\Phi(t)}^{\Phi(t+\Delta t)} \gamma(\tau)d\tau}{\Phi(t+\Delta t)-\Phi(t)} = \gamma(\Phi(t)) + \frac{1}{2}\gamma'(\Phi(t))\Phi'(t)\Delta t + o(|\Delta t|), \\
\Gamma_{j} &= \frac{\int_{\Phi(t-\Delta t)}^{\Phi(t)} \gamma(\tau)d\tau}{\Phi(t)-\Phi(t-\Delta t)} = \gamma(\Phi(t)) - \frac{1}{2}\gamma'(\Phi(t))\Phi'(t)\Delta t + o(|\Delta t|).
\end{aligned}$$

Then,

$$\frac{\Gamma_{j+1}}{\Gamma_j} = 1 + \frac{\gamma'(\Phi(t))\Phi'(t)}{\gamma(\Phi(t))}\Delta t + o(|\Delta t|). \tag{22}$$

Thus, substituting (20)-(22) into (19), we obtain

$$\begin{aligned}
& \left[1+\frac{\Phi''(t)}{\Phi'(t)}\Delta t + o(|\Delta t|)\right]^{\alpha(1-\beta)+\beta-(1-\alpha)(1-\beta)\theta\gamma(\Phi(t))} = \left[1+\frac{\gamma'(\Phi(t))\Phi'(t)}{\gamma(\Phi(t))}\Delta t + o(|\Delta t|)\right]^{\theta\gamma(\Phi(t))} \\
\Longrightarrow \quad & 1+[\alpha(1-\beta)+\beta-(1-\alpha)(1-\beta)\theta\gamma(\Phi(t))]\frac{\Phi''(t)}{\Phi'(t)}\Delta t + o(|\Delta t|) \\
& \qquad = 1+\theta\gamma(\Phi(t))\frac{\gamma'(\Phi(t))\Phi'(t)}{\gamma(\Phi(t))}\Delta t + o(|\Delta t|) \\
\Longrightarrow \quad & [\alpha(1-\beta)+\beta-(1-\alpha)(1-\beta)\theta\gamma(\Phi(t))]\frac{\Phi''(t)}{\Phi'(t)} + \frac{o(|\Delta t|)}{\Delta t} = \theta\gamma'(\Phi(t))\Phi'(t) + \frac{o(|\Delta t|)}{\Delta t} \\
\Longrightarrow \quad & [\alpha(1-\beta)+\beta-(1-\alpha)(1-\beta)\theta\gamma(\Phi(t))]\frac{\Phi''(t)}{\Phi'(t)} = \theta\gamma'(\Phi(t))\Phi'(t) \quad (\Delta t \to 0)
\end{aligned}$$

C Derivation of the Inverse Lorenz Curve H

As in the text, let $H : z \to t$ denote the inverse Lorenz curve, i.e., $H(z) \equiv \Phi^{-1}(z)$. By definition, it satisfies

$$t = H(\Phi(t)),$$

which in turn implies that

$$1 = H'(\Phi(t))\Phi'(t) \equiv H'(z)\Phi'(t). \tag{23}$$

Meanwhile, integrating the fundamental equation, we obtain

$$\Phi'(t) \propto [\alpha(1-\beta)+\beta-(1-\alpha)(1-\beta)\theta\gamma(z)]^{-\frac{1}{(1-\alpha)(1-\beta)}}. \tag{24}$$

Substituting (24) into (23), we then obtain

$$
\begin{aligned}
&& H'(z) &\propto [\alpha(1-\beta)+\beta-(1-\alpha)(1-\beta)\theta\gamma(z)]^{\frac{1}{(1-\alpha)(1-\beta)}} \\
\text{(25)} \qquad & \Longrightarrow & H(z) &= c_0 + c_1 \int_0^z G(u)^{\frac{1}{(1-\alpha)(1-\beta)}} du,
\end{aligned}
$$

where c_0 and c_1 are undetermined constants. Using the boundary conditions, $H(0) = \Phi(0) = 0$ and $H(1) = \Phi(1) = 1$, c_0 and c_1 are then determined by

$$
\begin{aligned}
c_0 &= 0, \\
c_1 &= \left[\int_0^1 G(u)^{\frac{1}{(1-\alpha)(1-\beta)}} du \right]^{-1},
\end{aligned}
$$

implying the desired expression:

$$
H(z) = \frac{\int_0^z G(u)^{\frac{1}{(1-\alpha)(1-\beta)}} du}{\int_0^1 G(u)^{\frac{1}{(1-\alpha)(1-\beta)}} du} \qquad \forall z \in [0,1].
$$

D Proof of Proposition 2

Here, we only show that (i) $\gamma'[\gamma^{-1}(B(\lambda))] \propto \lambda^{\tilde{\eta}}$ if and only if $\gamma(t)$ is given as in Proposition 2; and that (ii) the size distribution of cities is consistent with Zipf's law if and only if $\eta = -\alpha/[(1-\alpha)(1-\beta)]$.

D.1 Determination of the Functional Form of $\gamma(t)$

Define x by

$$
x \equiv \frac{\alpha(1-\beta)+\beta-\lambda^{-1}}{(1-\alpha)(1-\beta)\theta},
$$

with which we can rewrite $\gamma'[\gamma^{-1}(B(\lambda))] \propto \lambda^{\tilde{\eta}}$ as follows:

$$
\gamma'[\gamma^{-1}(x)] \propto (a-x)^{-\tilde{\eta}}, \quad \text{or} \quad \frac{1}{\gamma'[\gamma^{-1}(x)]} \propto (a-x)^{\tilde{\eta}}.
$$

Defining a function g by $t = g(x) = \gamma^{-1}(x)$, we can interpret this as

$$
g'(x) = c_1(a-x)^{\tilde{\eta}}, \qquad c_1 > 0. \tag{26}
$$

Note that the function g must satisfy

$$
g(0) = 0 \quad \text{and} \quad g(1) = 1 \tag{27}
$$

since $\gamma(0) = 0$ and $\gamma(1) = 1$.

Now, suppose that $\tilde{\eta} = -1$. We then obtain

$$t = g(x) = c_0 - c_1 \ln(a - x) \quad \text{and thus} \quad \gamma(t) = a - \exp\left(\frac{c_0 - t}{c_1}\right). \tag{28}$$

Using the terminal condition (27), it is immediate to obtain

$$c_0 = \frac{\ln a}{\ln a - \ln(a-1)} \quad \text{and} \quad c_1 = [\ln a - \ln(a-1)]^{-1},$$

which together with (28) imply the desired result for $\eta = \tilde{\eta} + 1 = 0$.

As for the case of $\tilde{\eta} \neq -1$, we can apply the similar method and obtain the desired result.

D.2 Determination of the Constraint on the Parameters for Zipf's Law

Under the assumption that $\eta \neq 0$, the density function of λ is given by

$$f_\Lambda(\lambda) \propto \lambda^{-\left[\frac{1}{(1-\alpha)(1-\beta)} + \eta + 1\right]}.$$

This implies that λ obeys a Pareto distribution with coefficient of $1/[(1-\alpha)(1-\beta)] + \eta$.

Then, since $\Phi'(t)$ is related with λ by

$$\Phi'(t) \propto \lambda^{\frac{1}{(1-\alpha)(1-\beta)}},$$

$\Phi'(t)$ obeys a Pareto distribution with coefficient of $(1-\alpha)(1-\beta)\eta + 1$.

Furthermore, since $N(t) \propto \Phi'(t)^{1-\alpha}$, $N(t)$ obeys a Pareto distribution with coefficient of $(1-\beta)\eta + 1/(1-\alpha)$.

Finally, the definition of Zipf's law, i.e., a Pareto distribution with unit coefficient, pins down the desired value of η.

E Derivation of the Modified Fundamental Equation

Reproduce the modified fundamental equation:

$$\left(\frac{|\mathbb{T}_{j+1}|}{|\mathbb{T}_j|}\right)^{1-\beta[1+\theta\gamma(T_j)]} \left\{\frac{(1-\alpha)[\alpha(1-\beta)(1-\tau)+\beta]|\mathbb{T}_{j+1}| + \alpha\tau J^{-1}}{(1-\alpha)[\alpha(1-\beta)(1-\tau)+\beta]|\mathbb{T}_j| + \alpha\tau J^{-1}}\right\}^{[\alpha(1-\beta)+\beta][1+\theta\gamma(T_j)]}$$
$$\times \left[\frac{(1-\alpha)(1-\tau)|\mathbb{T}_{j+1}| + \tau J^{-1}}{(1-\alpha)(1-\tau)|\mathbb{T}_j| + \tau J^{-1}}\right]^{-(1-\beta)[1+\theta\gamma(T_j)]} = \left(\frac{\Gamma_{j+1}}{\Gamma_j}\right)^{\theta\gamma(T_j)} \qquad \forall j \in \{1, 2, \cdots, J-1\},$$

which contains

$$\frac{|\mathbb{T}_{j+1}|+\tilde{\tau}_i J^{-1}}{|\mathbb{T}_j|+\tilde{\tau}_i J^{-1}}, \qquad i=1,2.$$

Because, replacing J^{-1} with Δt, we have

$$\begin{aligned}
|\mathbb{T}_{j+1}|+\tilde{\tau}_i J^{-1} &= \Phi(t+\Delta t)-\Phi(t)+\tilde{\tau}_i\Delta t = \left[\Phi'(t)+\tilde{\tau}_i\right]\Delta t+\frac{1}{2}\Phi''(t)\Delta t^2+o(|\Delta t|)\\
|\mathbb{T}_j|+\tilde{\tau}_i J^{-1} &= \Phi(t)-\Phi(t-\Delta t)+\tilde{\tau}_i\Delta t = \left[\Phi'(t)+\tilde{\tau}_i\right]\Delta t-\frac{1}{2}\Phi''(t)\Delta t^2+o(|\Delta t|),
\end{aligned}$$

with the help of the asymptotic expansion, we can approximate the above term as follows:

$$\frac{|\mathbb{T}_{j+1}|+\tilde{\tau}_i J^{-1}}{|\mathbb{T}_j|+\tilde{\tau}_i J^{-1}} = 1+\frac{\Phi''(t)}{\Phi'(t)+\tilde{\tau}_i}\Delta t+o(|\Delta t|).$$

Then, substituting the following approximations into the modified fundamental equation,

$$\begin{aligned}
\left(\frac{|\mathbb{T}_{j+1}|}{|\mathbb{T}_j|}\right)^{1-\beta[1+\theta\gamma(T_j)]} &= \left[1+\frac{\Phi''(t)}{\Phi'(t)}\Delta t+o(|\Delta t|)\right]^{1-\beta[1+\theta\gamma(\Phi(t))]}\\
&= 1+\{1-\beta[1+\theta\gamma(\Phi(t))]\}\frac{\Phi''(t)}{\Phi'(t)}\Delta t+o(|\Delta t|),\\
\left(\frac{|\mathbb{T}_{j+1}|+\tilde{\tau}_1 J^{-1}}{|\mathbb{T}_j|+\tilde{\tau}_1 J^{-1}}\right)^{[\alpha(1-\beta)+\beta][1+\theta\gamma(T_j)]} &= \left[1+\frac{\Phi''(t)}{\Phi'(t)+\tilde{\tau}_1}\Delta t+o(|\Delta t|)\right]^{[\alpha(1-\beta)+\beta][1+\theta\gamma(\Phi(t))]}\\
&= 1+[\alpha(1-\beta)+\beta][1+\theta\gamma(\Phi(t))]\frac{\Phi''(t)}{\Phi'(t)+\tilde{\tau}_1}\Delta t+o(|\Delta t|),\\
\left(\frac{|\mathbb{T}_{j+1}|+\tilde{\tau}_2 J^{-1}}{|\mathbb{T}_j|+\tilde{\tau}_2 J^{-1}}\right)^{-(1-\beta)[1+\theta\gamma(T_j)]} &= \left[1+\frac{\Phi''(t)}{\Phi'(t)+\tilde{\tau}_2}\Delta t+o(|\Delta t|)\right]^{-(1-\beta)[1+\theta\gamma(\Phi(t))]}\\
&= 1-(1-\beta)[1+\theta\gamma(\Phi(t))]\frac{\Phi''(t)}{\Phi'(t)+\tilde{\tau}_2}\Delta t+o(|\Delta t|),\\
\left(\frac{\Gamma_{j+1}}{\Gamma_j}\right)^{\theta\gamma(T_j)} &= \left[1+\frac{\gamma'(\Phi(t))\Phi'(t)}{\gamma(\Phi(t))}\Delta t+o(|\Delta t|)\right]^{\theta\gamma(\Phi(t))}\\
&= 1+\theta\gamma(\Phi(t))\frac{\gamma'(\Phi(t))\Phi'(t)}{\gamma(\Phi(t))}\Delta t+o(|\Delta t|),
\end{aligned}$$

and rearranging the result, we obtain the modified fundamental equation.

F Derivation of Profiles

F.1 Land Rent $R(t)$

From the text, we have

$$
\begin{aligned}
\frac{R_{j+1}}{R_j} &= \frac{|\mathbb{T}_{j+1}|+\tilde{\tau}_1 J^{-1}}{|\mathbb{T}_j|+\tilde{\tau}_1 J^{-1}}, \\
\sum_{j=1}^{J} R_j &= \frac{(1-\alpha)[\alpha(1-\beta)(1-\tau)+\beta]+\alpha\tau}{1-\alpha(1-\tau)},
\end{aligned}
$$

where we invoke the normalization of the total income, i.e., $E=1$, as in the text.

Using the asymptotic expansion, the former can be written as follows:

$$
\begin{aligned}
& \frac{R(t+\Delta t)}{R(t)} = 1+\frac{\Phi''(t)}{\Phi'(t)+\tilde{\tau}_1}\Delta t + o(|\Delta t|) \\
\Longrightarrow \quad & \frac{1}{R(t)}\frac{R(t+\Delta t)-R(t)}{\Delta t} = \frac{\Phi''(t)}{\Phi'(t)+\tilde{\tau}_1} + \frac{o(|\Delta t|)}{\Delta t} \\
\Longrightarrow \quad & \frac{R'(t)}{R(t)} = \frac{\Phi''(t)}{\Phi'(t)+\tilde{\tau}_1} \quad (\Delta t \to 0) \\
\Longrightarrow \quad & \ln R(t) = c_0 + \ln[\Phi'(t)+\tilde{\tau}_1] \\
\Longrightarrow \quad & R(t) = \tilde{c}_0[\Phi'(t)+\tilde{\tau}_1].
\end{aligned}
$$

Integrating the last expression, we then obtain

$$
\begin{aligned}
& \int_0^1 R(t)dt = \sum_{j=1}^{J} R_j = \frac{(1-\alpha)[\alpha(1-\beta)(1-\tau)+\beta]+\alpha\tau}{1-\alpha(1-\tau)} = \tilde{c}_0\left[\int_0^1 \Phi'(t)dt+\tilde{\tau}_1\right] = \tilde{c}_0(1+\tilde{\tau}_1) \\
& \tilde{c}_0 = \frac{(1-\alpha)[\alpha(1-\beta)(1-\tau)+\beta]}{1-\alpha(1-\tau)},
\end{aligned}
$$

where the last equality of the first line uses the fact that $\Phi(0)=0$ and $\Phi(1)=1$.

Therefore, we have

$$
\begin{aligned}
R(t) &= \frac{(1-\alpha)[\alpha(1-\beta)(1-\tau)+\beta]}{1-\alpha(1-\tau)}[\Phi'(t)+\tilde{\tau}_1] \\
&= [1-\alpha(1-\tau)]^{-1}\left\{(1-\alpha)[\alpha(1-\beta)(1-\tau)+\beta]\Phi'(t)+\alpha\tau\right\}.
\end{aligned}
$$

F.2 Population $N(t)$

From the text, we have

$$
\frac{N_{j+1}}{N_j} = \frac{|\mathbb{T}_{j+1}|+\tilde{\tau}_2 J^{-1}}{|\mathbb{T}_j|+\tilde{\tau}_2 J^{-1}}\left(\frac{|\mathbb{T}_{j+1}|+\tilde{\tau}_1 J^{-1}}{|\mathbb{T}_j|+\tilde{\tau}_1 J^{-1}}\right)^{-\alpha},
$$

which is in turn approximated as follows:

$$\begin{aligned}\frac{N(t+\Delta t)}{N(t)} &= \left[1+\frac{\Phi''(t)}{\Phi'(t)+\tilde{\tau}_2}\Delta t+o(|\Delta t|)\right]\left[1+\frac{\Phi''(t)}{\Phi'(t)+\tilde{\tau}_1}\Delta t+o(|\Delta t|)\right]^{-\alpha} \\ &= \left[1+\frac{\Phi''(t)}{\Phi'(t)+\tilde{\tau}_2}\Delta t+o(|\Delta t|)\right]\left[1-\alpha\frac{\Phi''(t)}{\Phi'(t)+\tilde{\tau}_1}\Delta t+o(|\Delta t|)\right] \\ &= 1+\left[\frac{\Phi''(t)}{\Phi'(t)+\tilde{\tau}_2}-\alpha\frac{\Phi''(t)}{\Phi'(t)+\tilde{\tau}_1}\right]\Delta t+o(|\Delta t|).\end{aligned}$$

Integrating both sides, arranging the result and taking the limit ($\Delta t \to 0$), we then obtain

$$\begin{aligned} & \frac{N'(t)}{N(t)} = \frac{\Phi''(t)}{\Phi'(t)+\tilde{\tau}_2}-\alpha\frac{\Phi''(t)}{\Phi'(t)+\tilde{\tau}_1} \\ \Longrightarrow\quad & \ln N(t) = \hat{c}_0+\ln[\Phi'(t)+\tilde{\tau}_2]-\alpha\ln[\Phi(t)+\tilde{\tau}_1] \\ \Longrightarrow\quad & N(t) = e^{\hat{c}_0}\left[\Phi'(t)+\tilde{\tau_2}\right]\left[\Phi'(t)+\tilde{\tau}_1\right]^{-\alpha}, \quad \text{or} \\ & N(t) = e^{c_0}\left[(1-\alpha)(1-\tau)\Phi'(t)+\tau\right]\left\{(1-\alpha)[\alpha(1-\beta)(1-\tau)+\beta]\Phi'(t)+\alpha\tau\right\}^{-\alpha},\end{aligned}$$

where c_0 satisfies

$$e^{\hat{c}_0} = e^{c_0}(1-\alpha)(1-\tau)\left\{(1-\alpha)[\alpha(1-\beta)(1-\tau)+\beta]\right\}^{-\alpha}.$$

Here, c_0 should be consistent with the normalized total population, i.e., $\int_0^1 N(t)dt = 1$.

F.3 Wage Rate $W(t)$

From the labor market clearing condition in the text, we have

$$\begin{aligned}\frac{W_{j+1}N_{j+1}}{W_jN_j} &= \frac{|\mathbb{T}_{j+1}|}{|\mathbb{T}_j|}, \\ \sum_{j=1}^{J} W_jN_j &= (1-\alpha)(1-\beta),\end{aligned}$$

where we invoked the normalization of the total income, i.e., $E = 1$.

Defining the local labor income M_j by $M_j = W_jN_j$, we can approximate the above first equation as follows:

$$\frac{M(t+\Delta t)}{M(t)} = 1+\frac{\Phi''(t)}{\Phi'(t)}\Delta t+o(|\Delta t|).$$

Therefore, we have

$$\begin{aligned}
& \frac{M'(t)}{M(t)} = \frac{\Phi''(t)}{\Phi'(t)} \\
\Longrightarrow \quad & \ln M(t) = c_0 + \ln \Phi'(t) \\
\Longrightarrow \quad & M(t) = \tilde{c}_0 \Phi'(t),
\end{aligned}$$

where $\tilde{c}_0$ is determined by integrating both sides and using the above result:

$$\int_0^1 M(t)dt = \sum_{j=1}^{J} W_j N_j = (1-\alpha)(1-\beta) = \tilde{c}_0 \int_0^1 \Phi'(t)dt = \tilde{c}_0.$$

This implies that

$$M(t) = W(t)N(t) = (1-\alpha)(1-\beta)\Phi'(t) \quad \Longrightarrow \quad W(t) = (1-\alpha)(1-\beta)\frac{\Phi'(t)}{N(t)}.$$

F.4 Urban Diversity $D(t)$

Reproduce the market clearing condition for the nonroutine service sector:

$$D_j p_{n,j} q = (1-\alpha)\Gamma_j |\mathbb{T}_j| = (1-\alpha)\int_{T_{j-1}}^{T_j} \gamma(t)dt,$$

where

$$p_{n,j} = (1+\theta) m R_j^{\beta} W_j^{1-\alpha}, \qquad q = \frac{f}{\theta m}.$$

This suggests that

$$\begin{aligned}
& D(t)\frac{(1+\theta)f}{\theta} R(t)^{\beta} W(t)^{1-\beta} dt = (1-\alpha)\gamma(\Phi(t))\Phi'(t)dt \\
\Longrightarrow \quad & D(t) = \frac{\theta}{(1+\theta)f}(1-\alpha)\gamma(\Phi(t))\Phi'(t)R(t)^{-\beta}W(t)^{-(1-\beta)}.
\end{aligned}$$

F.5 After-tax Local Income $E_a(t)$

From the text, we have

$$E_{a,j} = (1-\tau)E_{b,j} + \tau J^{-1} = (1-\tau)(W_j N_j + R_j) + \tau J^{-1},$$

where the normalization of the total income, i.e., $E = 1$, is used.

This then suggests that

$$
\begin{aligned}
E_a(t)dt &= (1-\tau)[W(t)N(t)+R(t)]dt+\tau dt \\
&= (1-\tau)[M(t)+R(t)]dt+\tau dt \\
\implies \quad E_a(t) &= (1-\tau)[M(t)+R(t)]+\tau.
\end{aligned}
$$

Substituting

$$
\begin{aligned}
M(t)+R(t) &= (1-\alpha)(1-\beta)\Phi'(t)+[1-\alpha(1-\tau)]^{-1}\left\{(1-\alpha)[\alpha(1-\beta)(1-\tau)+\beta]\Phi'(t)+\alpha\tau\right\} \\
&= [1-\alpha(1-\tau)]^{-1}[(1-\alpha)\Phi'(t)+\alpha\tau]
\end{aligned}
$$

into the above equation, we obtain

$$
E_a(t)=[1-\alpha(1-\tau)]^{-1}[(1-\alpha)(1-\tau)\Phi'(t)+\tau].
$$

F.6 Natural Logarithm of the Price Index $\ln P$ of the Final Good

In a sorting equilibrium, each location t specializes in a stage of production with skill intensity of $z=\Phi(t)$. Then, the Cobb-Douglas technology implies that

$$
\ln P=\int_0^1 \ln[P(\Phi^{-1}(z))]dz,
$$

where $P(t)$ denotes the location-t unit cost of processing stage with $z=\Phi(t)$. The change of the variable then implies

$$
\ln P=\int_0^1 \ln[P(t)]\Phi'(t)dt.
$$

Meanwhile, the unit cost $P(t)$ of processing a stage with skill intensity of $\Phi(t)$ is given by

$$
P(t)=P_n(t)^{\gamma(\Phi(t))}P_r(t)^{1-\gamma(\Phi(t))}.
$$

Substituting the unit costs, $P_n(t)$ and $P_r(t)$, of local service sectors

$$
\begin{aligned}
P_n(t) &= (1+\theta)mR(t)^{\beta}W(t)^{1-\beta}D(t)^{-\theta}, \\
P_r(t) &= R(t)^{\beta}W(t)^{1-\beta}
\end{aligned}
$$

into the equation, it follows that

$$
P(t)=[(1+\theta)m]^{\gamma(\Phi(t))}R(t)^{\beta}W(t)^{1-\beta}D(t)^{-\theta\gamma(\Phi(t))},
$$

which, with the help of normalization, $(1+\theta)m = 1$, becomes

$$P(t) = R(t)^{\beta} W(t)^{1-\beta} D(t)^{-\theta\gamma(\Phi(t))}.$$

Therefore, we have

$$\ln P = \int_0^1 \left[\beta \ln R(t) + (1-\beta) W(t) - \theta\gamma(\Phi(t)) D(t)\right] \Phi'(t) dt.$$

G Calibration

In this section, we calibrate the model to the U.S. data. We describe our procedure and the data we use in Subsection G.1 and G.2, respectively. The result of the calibration is then presented in Subsection G.3. Our purpose is not testing the model. Instead, we simply intend to pick up an example of values of parameters used as a benchmark in numerical exercises conducted in Subsections 3.4 and 4.3 in such a way that the model roughly matches the data.

G.1 Procedure

In order to compute the equilibrium numerically, we need to specify the values of the parameters $(\alpha, \beta, \sigma, \tilde{a}, \eta)$. The first two are calibrated independently, and the other by matching the model with the U.S. data in terms of the upper tail of the distribution of market sizes of MSAs. Without loss of generality, $f = 1$ and $(1+\theta)m = 1$.

Expenditure Share α of Land

Let A, B, and C denote the expenditure share of land, housing (excluding land), and goods and services, respectively, i.e., $A+B+C = 1$. According to Davis and Ortalo-Magne (2011), the housing expenditure (including land) $A+B = 0.24$. In addition, Albouy and Ehrlich (2012) report that the one-third of housing costs are due to land, implying $A = 1/3 \times (A+B)$. Since there is no housing in the model, the expenditure share α of land is then calculated by $\alpha = A/(A+C) \approx 0.095$.

Cost Share β of Land

Valentinyi and Herrendorf (2008) report sectoral income shares of land, structures, and equipments, the summation of which corresponds to the capital share. Let $\hat{A}$, $\hat{B}$, and $\hat{C}$ denote the income shares of land, structures and equipments, and labor, respectively, i.e., $\hat{A}+\hat{B}+\hat{C} = 1$. If we focus on the services sector, $\hat{A} = 0.06$, and $\hat{C} = 0.65$ according to Table 5 in Valentinyi and Herrendorf (2008). Since production factors in our model consist of land and labor only, we can calculate β by $\beta = \hat{A}/(\hat{A}+\hat{C}) \approx 0.085$.

Parameter	Meaning	Restriction	Target	Source
α	Expenditure share of land	$(0,1)$	Expenditure share of land	Albouy and Ehrlich (2012)
				Davis and Ortalo-Magne (2011)
β	Cost share of land	$(0,1)$	Cost share of land	Valentinyi and Herrendorf (2008)
σ	Elasticity of substitution	$(\max\{1,[\alpha(1-\beta)+\beta]^{-1}\},+\infty)$	Max-Min ratio of labor compensation	*Occupational Employment Statistics*
$\tilde{a}$	Parameter of $\gamma(t)$	$(1,+\infty)$	Distribution of labor compensations	*Occupational Employment Statistics*
η	Parameter of $\gamma(t)$	$(-\infty,+\infty)$	do.	do.

Table 4: Restrictions on and Targets of the Calibrated Parameters

Other Parameters $(\sigma,\tilde{a},\eta)$

The elasticity $\sigma = 1/\theta + 1$ of substitution is calibrated in such a way that the model can exactly match the natural logarithm of the observed max-min ratio of the market size, which is given by

$$\ln\left[\frac{\Phi'(1)}{\Phi'(0)}\right]\Bigg|_{\text{data}} = \frac{1}{(1-\alpha)(1-\beta)}\ln\left[\frac{\alpha(1-\beta)+\beta}{\alpha(1-\beta)+\beta-(1-\alpha)(1-\beta)\theta}\right].$$

As for $(\tilde{a},\eta)$, we solve the following constrained minimization problem in a brute force manner with a discretized parameter space:

$$\min_{\tilde{a},\eta}\ \max_{i\in\{1,2,\cdots,\hat{N}\}}\left|\hat{F}(m_i)-F(m_i;\tilde{a},\eta)\right| \quad s.t. \quad \tilde{a}>1,$$

where $\hat{F}$ and F are the distribution functions of the natural logarithm of the market size for the data and the model, respectively. $\hat{N}$ is the sample size of MSAs, and m_i is the actual normalized market size of ith MSA. Table 4 summarizes the restrictions on and targets of the calibrated parameters, where the lower bound for the elasticity σ of substitution takes account of the assumption that $G(1) > 0$ in Proposition 1.

G.2 Data

The data that we use in the calibration are taken from the *May 2011 Occupational Employment Statistics* compiled by the Bureau of Labor Statistics, which reports the number of employments and the average annual wage rates for occupations listed in the *2010 Standard Occupational Classification System* for each of the Metropolitan Statistical Areas (MSAs). We then exploit the equivalence between the normalized market size of an MSA and the normalized total labor compensation due to the constancy of the labor cost share $1-\beta$ in order to construct the distribution of the natural logarithm of the market size from the data on labor compensation. Focusing on the upper tail of the distribution results in the sample size of 342 MSAs which is not particularly different from those in previous studies such as Rossi-Hansberg and Wright (2007). The implied max-min ration of market sizes is 5.304.

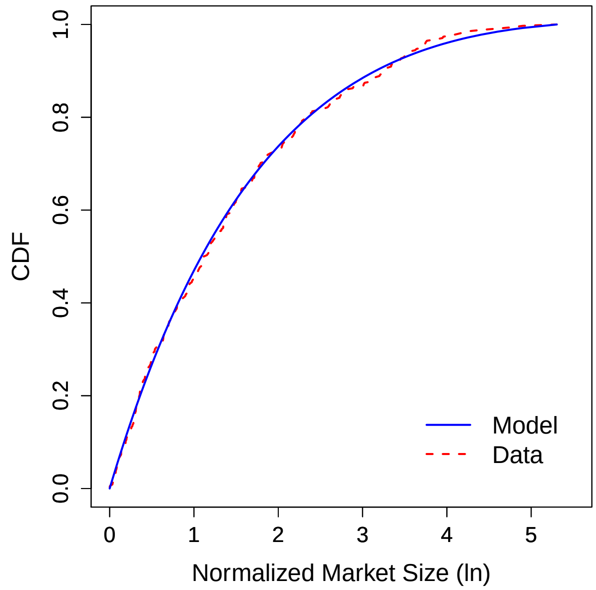

Figure 8: Cumulative Distribution Function of the Normalized Market Size

Note: The cumulative distribution of the data is constructed using labor compensations of the top 342 MSAs.

Source: May 2011 Occupational Employment Statistics.

G.3 Result

The results of the calibration are reported in Table 1. The elasticity σ of substitution falls in the typical range, $[5, 10]$, in the (international) literatures such as Anderson and van Wincoop (2004). As shown in Figure 8, the model replicates the observed distribution of the natural logarithm of market sizes with the maximal deviation of 2.0%.