The Size Distribution of Cities with Distance-Bound Households

Watanabe, Axel

Concordia University and CIREQ

17 April 2020

The Size Distribution of Cities
with Distance-Bound Households

Axel Watanabe*

April 13, 2020

Abstract

There has been a long tradition of presumed perfect mobility in urban economics. Workers switch their locations in direct response to differences in local economic performance. Recent empirical observations prove otherwise. The number of movers rapidly declines with distance moved while there is a positive correlation between distance moved and skill level. I build a general equilibrium model of a system of cities to explain the city-size distribution as a result of reduced mobility. Workers with a heterogeneous skill level have a corresponding distance-tolerance level and self-sort into select cities. The resulting size distribution reflects the trade-off between the distance moved and earning opportunities enhanced by agglomeration. I extrapolate consumers' tolerance towards distance and skill level from US Census data on city size and intercity migration.

Keywords: labor mobility, internal migration, city-size distribution
JEL classification: J61, R12

1 Introduction

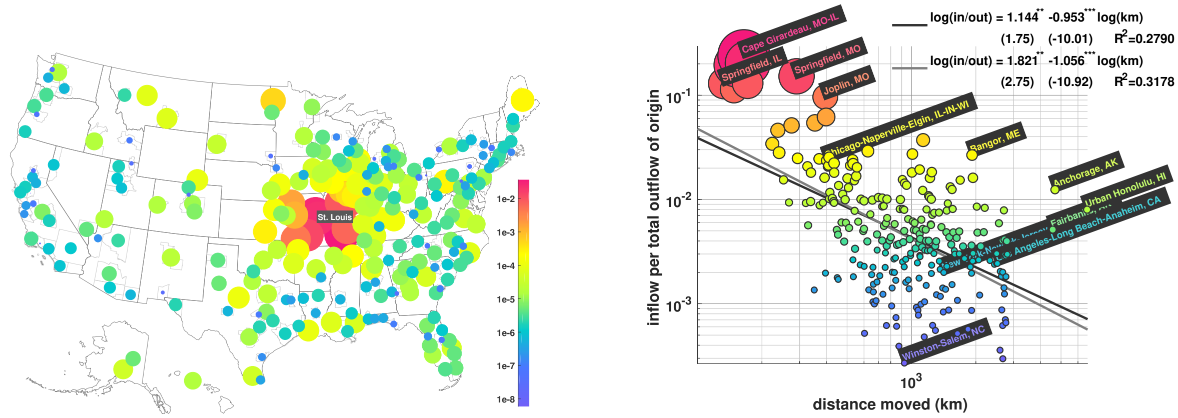
1.1 Consumers Are Not Footloose

Labor mobility exhibits recognizable geographical patterns. There is a log linear relationship between intercity migration and distance moved. An exceeding share of migration occurs within a close proximity and there are only a few who move coast to coast. The US does not impose any statutory restrictions on relocation of households, and yet, consumers behave as if there were some in place. Take St. Louis for example. Figure 1 represents the inflow into the city from other metropolitan statistical areas (MSA). The vast majority of people in the city are from Missouri and Illinois when in fact consumers are free to move anywhere in the country. The inflow drops at an exponential rate with distance. When the distance increases by 1%, the inflow from that area drops by about 1%. It is known that the city-size distribution has a fat tail (cf. figure 2, Gabaix and Ioannides [GI04], and Duranton [Dur07]). The intercity migration itself follows the same rank-size pattern as shown in figure 1(b).

The city-size distribution is a result of household relocation. Any city size is the sum of the inflows into, less the outflows out of, the city over time. It does not take a big leap to imagine that the city-size distribution hinges on the degree of ease of movement. However, mobility has been overlooked in the literature, and probably for good reasons (to be discussed below). The models of the city-size distribution traditionally assume perfect mobility. Workers move to another city in direct response to local economic conditions regardless of how far the distance moved is. The resulting equilibrium size distribution is independent from where workers were in the period before. While the path dependence takes place on the side of productivity at the city level, at the individual level, workers behave as if they do not remember or care which cities they lived in preceding periods.

This paper is tasked with explaining why there is a correlation between inflow frequency and distance moved in order to describe the city-size distribution as a result of intercity migration. I identify four causes, among many, as critical factors behind the geographically bounded labor mobility.

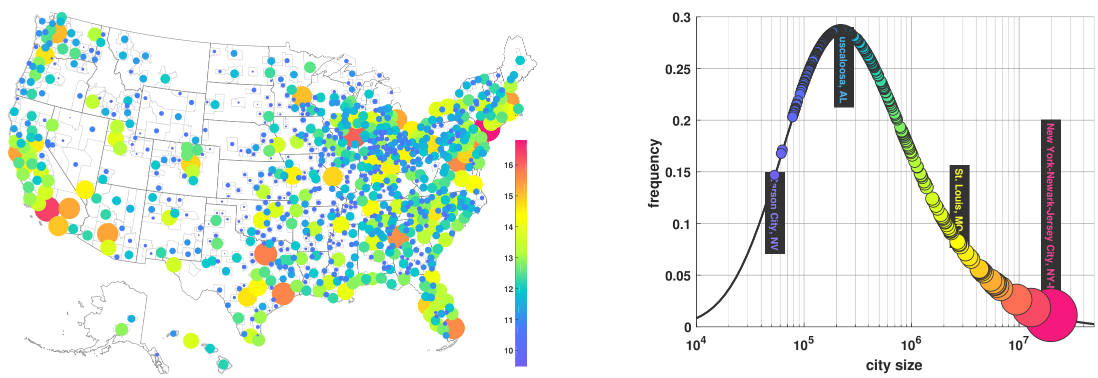
*Department of Economics, Concordia University, and CIREQ. Address: 1455 boul. de Maisonneuve O, Montréal (Québec) H3G 1M8. E-mail: axel.watanabe@concordia.ca. I am grateful for valuable inputs from Xintong Han and Prosper Dovonon. The usual disclaimer applies.



(a) The inflow into St. Louis normalized by the total outflow of origin.

(b) Black line includes Alaska and Hawaii. Gray line is for lower 48 only.

Figure 1. *Data source*: US Census Bureau, 2009-2013 American Community Survey.



(a)

(b)

Figure 2. US city-size distribution (population in log scale). *Data source*: US Census

The first conjecture is, all else equal, that the geographical extent of job search expands with the level of skill. Ph.D. students on the job market fly everywhere, whereas it is unlikely to see a high school student looking through fast food restaurant chain's vacancy notices all over the country. In fact, employment-related reasons list among the primary factors of relocation. As figure 3 shows, employment-related reasons become more predominant as the level of educational attainment steps up.

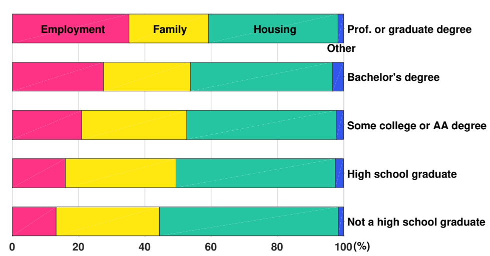


Figure 3. Reason for relocation by educational attainment. *Data source*: Geographical mobility between 2014 and 2015, US Census.

Davis and Dingel [DD12] show that skilled labor tend to move more frequently as a result of spatial sorting triggered by idea exchange. In their model, Skilled labor has different equilibrium utility levels depending on their skill level. They move to the city where they can make full use of their skill. Consequently, they tend to move farther and more frequently compared to unskilled labor. They suspect that their findings are set off by difference in the search behavior of workers of varying skill levels. Rauch [Rau13] shares the same supposition. See Molloy et al. [MSW11] for other reasons behind relocation.

Regarding heterogeneous skills and spatial sorting, Behrens et al. [BDRN14] show that workers sort into a city and select their occupation according to their skill level. Along with skill levels, location-variant serendipity determines the productivity and in turn the degree of agglomeration in each city. Eeckhout et al. [EPS14] find evidence in support of extreme skill complementarity where the co-presence of workers from top- and bottom-tier skill levels does not undermine but rather enhances their productivity. In either model, relocation is costless or does not depend on the distance moved.

Mobility is a topic of interest in labor economics too, but not in the same sense as I discuss in the current paper. Market imperfections lead to reduced mobility in terms of type-matched industry, but not geographical mismatch (cf. Manning [Man10] and Hirsch et al. [HJO16]).

Along with the heterogeneous skill levels, uncertainty aversion may deter long-distance relocation. A resident in Duluth, MN is likely to know more about the local economy of Minneapolis than that of Houston for example. If he receives identical job offers from both locations, a move to Minneapolis is easier as he does not have to learn about the city as much as he would about Houston. I expect that consumers become exponentially less knowledgeable about the local economy of a city as it gets farther from his birthplace.

Thirdly, I assume that all else equal, a worker prefers to work near her birthplace for social reasons. Whereas a worker may be mobile per se, it is prohibitively costly to move the entire network of people she meets in her daily personal or social life. In addition, she may not know about the urban life in a distant city as much as her home town's. While job-related reasons are an important determinant in location choice as discussed above, she does not live for her job alone. She might also enjoy the social life, which is intrinsically chained to the location it takes place and she cannot take it with her to another city. Thus, presented with two identical job opportunities, one in her birth city and the other from elsewhere, she prefers to take the former.

Together with the cost of being far from their social network, workers may incur a lingering region-specific cost due to cultural differences within the same country. If a Québécois educated mostly in French moves to Vancouver for a job in an Anglophone firm, his productivity, and by extension, his lifetime earnings may be lower than what he could have made in a comparable but Francophone firm in Montréal. A liberal-minded Minnesota expat in the South may find the life there tormenting, compared to a Texan of a similar educational background in the South. Even if he could go back to Minnesota, the experience in the South may turn out to be a traumatic life event that stays with him for the rest of his life, and thus very costly to him. These regional differences in societal norms act as a deterrent against free mobility. Indeed, Falck et al. [FHLS12] document Germans' reluctance to move outside of their area of shared regional dialect. Woodard [Woo11] suggests similar cultural divides in the US, which reduces mobility.

1.2 Mobility in Urban Economics

Nevertheless, perfect mobility is a sensible assumption to make when analyzing the city-size distribution for two reasons. For the most part, researchers regard a city as a placeholder to host households. They leave off geographical features when conducting their analysis and focus on a size and other economic variables that are not tied to the location of the city. The equilibrium is unique only up to the size distribution but not to the location of cities.

On balance, urban economists have been good at deriving the true-to-life size distribution but not as good at identifying where cities are. And rightly so because correspondence between the size and the location is not the salient features in the existing models. Workers can switch their locations at no cost and thus **do not sense the distance** when they relocate. It matters little which city hosts which point in the predicted size distribution. One can move all New Yorkers to St. Louis, and drive all St. Louisans out into Minneapolis (or any permutation of size and location thereof) without making any difference in the equilibrium outcome. This is justifiable because the city-size distribution arises as a result of depleted spatial arbitrage opportunities. Any difference by location, be it the utility level or real wage, will be eliminated in the end as consumers move (freely) to exhaust any arbitrage opportunities left. This greatly reduces the number of the equilibria. Any worker of the same attributes achieves the same utility level regardless of how far they had to move to realize it.

In addition, one could argue that relocation cost (and by extension the distance moved) is negligible in the grand scheme of things. While a consumer incurs the cost of relocation only so often, the benefits of relocation (e.g., higher wage, better quality of life etc.) keep accruing over time once relocation is complete. Thus the difference in distance moved is easily dwarfed by the lifetime increase in economic welfare.

However, the assumption could be expedient. While it is true that the pecuniary cost of relocation is a one-time expense, out-of-towners may pay other implicit prices over a long period of time, be it personal or regional as I explored above. If workers have young children in their household, they cannot ask their parents or other family members to look after them once they move out of the city. And day care services take up a significant portion of their income. Workers may want to visit their family or close friends from high school every holiday season, and those trips cost them more (both in terms of time and money) if they live farther away from home.

1.3 Tolerance to Distance

While I cannot observe consumers' aversion to distance moved per se, the resulting distribution figure 1(b) clearly suggests its existence. In this light, the current paper breaks away from the convention and incorporates consumers' tolerance to distance as described above. I listed four factors that contribute to consumers' reluctance to move, but I cannot tell them apart because they can be observed only indirectly via the geographical distribution of inflow. Instead of representing them one by one, I model distance-bound household's behavior with a single, general variable that measures the degree of tolerance towards distance moved. Each consumer will draw her tolerance level from a given distribution and makes a location choice based on it. Distance tolerance is interpreted broadly to represent the aforementioned skill level, affinity to the birthplace and region-specific characters rather than a mere distance. For instance, a consumer raised in a liberal household in California may find Texas very "far" not only geographically but also psychologically, resulting in a low tolerance towards distance.

In the context of Starrett's spatial impossibility theorem (cf. Starrett [Sta78], and Boyd and Conley [BC97]), I am violating two assumptions in this paper; one majorly and the other only trivially. First, free mobility is the very assumption that I would like to forgo. Consumers do assess relocation cost and it makes a sizable difference in their welfare. The overwhelming majority of models of the city-size distribution keep to the assumption of free mobility instead. Second, I violate the assumption of homogeneous space to break the symmetry in the size distribution. Each industry receives a productivity boost only in a city, but not in a rural area. However, I keep the violation to a bare minimum: I do not introduce city-variant productivity differences. All the industries share the same productivity regardless of which city they are housed in. I will discuss more on productivity in (6). Regarding the remaining assumptions, unlike New Economic Geography models (cf. Fujita et al. [FKV99]), I maintain the assumption of perfect and complete markets.

The rest of the paper is organized as follows: In the upcoming section I will lay out the model and uncover the relationship among distance, inflow and city size. I will validate my theoretical predictions in section 3 using US Census data, and section 4 concludes.

2 Model

2.1 Landscape

The economy is assumed to be closed. Some cities near the border may receive an economic boost for its proximity to the country's trading partners, or on the contrary, lag behind because they are far from the center of domestic economic activities. In order to stay focused on reduced mobility, I take the country to be a sphere X of size one. Since there is no end point on a sphere, if there is any asymmetry in equilibrium, I know definitively that it is not caused by the terrestrial restrictions such as the proximity to the border or the center of the country, but rather by imperfect mobility.

Consider a production economy of a system of cities. There are I cities indexed by i, and $N\,(\in \mathbb{R}_+)$ consumers in the economy. The model rolls out in two stages. Initially, consumers are uniformly distributed across the country. Each consumer is endowed with a pair $(t,\ y) \in \{1,\ \cdots,\ T\} \times Y\,(\subseteq \mathbb{N} \times \mathbb{R})$. The first entry is type t, which represents her skill and identifies her best suited industry to work in. There are n_t of type-t

consumers. Along with the type, she also draws her distance-tolerance factor y from the distribution with probability density function (pdf) $f_t(y)$ and cumulative distribution function (CDF) $F_t(y)$. A high y implies that relocation is not costly and that she does not mind moving far. Note that $\int_Y f_t(y)dy = n_t$ for all t, totaling up to $\sum_t \int_Y f_t(y)dy = N$ nationwide.

The type distribution $f_t(y)$ may possibly depend not only on t but also on birthplace $x \in X$. However, since consumers cannot chose a place to be born at, it is safer and more reasonable to assume that $f_t(y)$ takes the same form regardless of the location.

In the first stage, consumer's birthplace and type are revealed as above. However, for the reasons to be discussed later in (14), consumers only know their value of $(t,\ y)$ but do not know from what distribution their y is drawn.

In the second stage, consumers of type t make simultaneous and uncoordinated decision on their location. A type-t consumer can either stay at her initial location x or move to one of the cities. For simplicity, assume that each city attracts and/or hosts at most one type, in which case, $T \leq I$. Furthermore, I set $T = I$ so that there is a one-to-one correspondence between the set of types and the set of cities.[1] For each type, there exists exactly one city conducive exclusively to that type. Empirically speaking, this assumption is not much of a restriction to impose because of flexibility in T. If T falls short of I, I may split "prospective doctors" into "prospective doctors who speak French" and "prospective doctors who do not" to increase T till it matches I (the latter cannot move to Québec unless they are willing to change their vocation). From here on I refer to type t by its corresponding city i and use the term "city", "type" and "industry" interchangeably where applicable.[2] I will review below three choices of location that a type-i consumer can make and their consequences:

1. move to type-matched city i (as portrayed by Chloe, the city dweller),
2. stay put at x (by Ryan, the rural resident), or
3. move to type-discordant city $j(\neq i)$ (by Diane, the disoriented).

2.2 Chloe the City Dweller

First, take a look at Chloe, a type-i consumer who becomes a resident of city i.

2.2.1 Consumption and Location Choice

Her preferences over a numéraire composite consumption good c_i and housing h_i are represented by

$$u(c_i,\ h_i;\ x,\ y) = c_i(x,\ y) + \eta \log h_i(x,\ y), \tag{1}$$

where η measures the portion of her expenditure on housing. See appendix A.1 for other preference specifications and interpretation on $c_i(x,\ y)$. She is endowed with a unit of time, which she converts into $c_i(x,\ y)$ to earn wage w_i. Her budget constraint is

$$w_i \geq c_i(x,\ y) + p_i h_i(x,\ y) + \rho\left[l_i(x),\ y\right], \tag{2}$$

where $l_i(x)$ is the geodesic (the shortest path on the sphere) between x and city i, and $\rho(\cdot)$ measures the lifelong opportunity cost of relocation.[3] [4] For any give y the opportunity cost of relocation $\rho(\cdot)$ is increasing and concave in geodesic. Consumers' nonlinear perception of distance gives grounds for this assumption. A

[1] I could set $T < I$, which would become superfluous: As I will establish later, any city $i > T$ will not have any inflow of workers and fall off the list of cities anyway.

[2] I will discuss this assumption further in section 4.

[3] There are three ways to factor in aversion to relocation: 1) Put y directly into the utility function with $\frac{\partial u(\cdot,\ \cdot,\ y)}{\partial y} > 0$. 2) Put y into the budget constraint. 3) Combine the two above. I went for 2) because it is the most interpretive way to disentangle the role of tolerance from the overall end result. Since the utility level does **not** equalize in equilibrium, it is hard to keep track of $\rho(\cdot)$ if I leave it nested in the utility function. Nevertheless, with appropriate parameterization and some added complications, 1) and 3) will also lead to the same results to follow.

[4] I take η to be less than $w_i - \rho(\cdot)$ for any $i,\ x$ and y to exclude corner solutions.

St. Louis native will find a move from St. Louis to Chicago more draining than a move from Fairbanks to Anchorage (roughly the same distance apart). The additional cost increase wears out with the distance.

Coupled with this assumption, I also implement $\partial \rho[l_i(x),\, y]/\partial y < 0$, i.e., the higher the distance tolerance is, the lower the relocation cost will be. Note that w_i is constant across x and y, and preferences over $(c_i(x),\, h_i(x))$ are the same across y according to the current specifications: Her productivity will neither increase or decrease regardless of how far she moved to reach i or how much she was willing to do so. In addition, her preferences for the bundle are independent of y (though her Marshallian demand $c_i(x,\, y)$ and by extension, her indirect utility function will differ by tolerance level in equilibrium).

In addition to finding the optimal consumption bundle, a consumer also needs to make her location choice. Denote location choice by a mapping $g_i(x,\, y) : X \times Y \to \{0,\, 1, \cdots,\, I\}$. If a consumer who drew $(x,\, y)$ decides to become a Chloe, $g_i(x,\, y) = i$. Similarly, for Diane, $g_i(x,\, y) = j (\neq i)$. I tack 0 to the set of cities for Ryan: $g_i(x,\, y) = 0$. In preparation for definition 2.1 of feasibility to follow, define an indicator function $\mathbb{1}_{\{i\}}[g_i(x,\, y)]$ that takes the value of 1 if $g_i(x,\, y) = i$ and 0 otherwise. For instance, Chloe takes the value of 1 whereas Ryan and Diane take 0.

2.2.2 Feasibility

I will make a quick note here on the shape of the country and the geographical distribution of consumers. When the surface area is normalized to unity, the radius of the ball is $\frac{1}{2\sqrt{\pi}}$ (cf. figure 4).

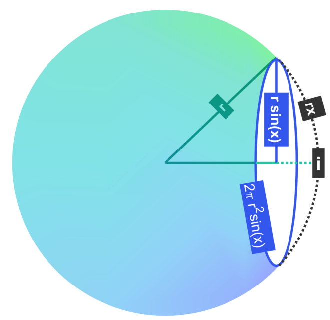


Figure 4. Consumers are uniformly distributed on the ball with radius $r = \frac{1}{2\sqrt{\pi}}$.

Since each type is uniquely associated with their type-concordant city, in what follows, I will identify x by the geodesic length between the type city and consumer's birthplace. (I will nevertheless call back $l_i(x)$ if the distinction between a generic location and geodesic location is necessary for identification purposes, where applicable). Land supply x radian away from city i is $\frac{\sin x}{2}$ (the perimeter of the cut surface in figure 4). Accordingly, a measure of type-i residents born x radian away from city i is $n_i \frac{\sin x}{2}$. The number of type-i residents in city i is then

$$s_i = \int_0^{\pi} n_i \frac{\sin x}{2} \int_Y \mathbb{1}_{\{i\}}[g_i(x,\, y)]\, dF_i(y) dx \tag{3}$$

With all the necessary variables and functions in hand, I formally define the feasible allocation in this economy as follows:

<u>Definition 2.1 Feasible Allocation</u>:
An allocation is a list of functions $[c_i(x,\, y),\, h_i(x,\, y),\, g_i(x,\, y)]_{i=1}^{I}$ with $c_i : X \times Y \to \mathbb{R}_+$ and $h_i : X \times Y \to \mathbb{R}_+$, location choice $g_i : X \times Y \to \{0,\, 1,\, \cdots,\, I\}$, and output $(z_i)_{i=1}^{I} \in \mathbb{R}_+^I$. Given type-size distribution $(n_i)_{i=1}^{I}$ and distance-tolerance distribution $[f_i(y)]_{i=1}^{I}$, an allocation is feasible if

$$\begin{aligned} s_i z_i &= \int_0^{\pi} n_i \frac{\sin x}{2} \int_Y \mathbb{1}_{\{i\}}[g_i(x,\, y)]\, c_i(x,\, y) dF_i(y) dx, && (4) \\ H &= \int_0^{\pi} n_i \frac{\sin x}{2} \int_Y \mathbb{1}_{\{i\}}[g_i(x,\, y)]\, h_i(x,\, y) dF_i(y) dx, \quad \textit{and} && (5) \\ s_i &\leq n_i \end{aligned}$$

for any i, where s_i is defined by (3).[5]

[5] Strictly speaking, I need to define Ryan's allocation in definition 2.1 as well. I did not do so because the topic of interest, the city-size distribution, does not depend on the feasibility with respect to Ryan-like consumers. The only role Ryan plays in the determination of the size distribution is his absence, which I already made accounted for through the indicator function in (4) and (5). The same goes for Diane.

As explained earlier, $\rho(\cdot)$ entails both tangible and intangible costs. It represents the airfare to visit Chloe's family every Thanksgiving for example. However, it goes beyond these out-of-pocket expenses. Along with (4) and (5), I could as well hypothesize the third market for "consolation for relocation". In other words, Chloe or Diane "purchases" a unit of solace for being an expat or condolences for her virtual loss of friends and families at the price of $\rho(\cdot)$. This third market would make it easier to interpret and place the present model in the grand scheme of general equilibrium. While it is true that I seek to build a **general** equilibrium model, I will leave out this market, because, hometown-bound airplanes aside, these "goods" do not come in material form, and it is not clear who actually supplies them nor does there seem to be any resource constraint on them to warrant an explicit condition. Instead, I simply take $\rho(\cdot)$ as a function of $(x,\ y)$ drawn from the exogenous distribution and fold it into the budget constraint as an unavoidable dead load inflicted upon her when she leaves her birthplace behind.

2.2.3 Production

Turning to production, as mentioned earlier, workers supply one unit of (perfectly inelastic) labor[6] to produce the composite goods with a constant returns to scale technology. In particular, τ units of labor produces $z_i = A_i(s_i)\tau$ units of composite goods. As opposed to what is conventionally assumed, I assume that $A_i(s_i)$ does not vary with industry i or city size s_i (unless it is zero). In particular

$$A(s_i) = \begin{cases} 1 & \text{if} \quad s_i = 0 \\ a(>1) & \text{if} \quad s_i > 0. \end{cases} \tag{6}$$

In the current model, I do not rely on productivity differences to break the otherwise uniform distribution of workers. I shut off the channel through which productivity differences bring in variations in city sizes in order to isolate the role that distance tolerance plays, or else I will not be able to tell how much of the size differences is the result of imperfect mobility. However, I still do need to secure some incentive for residents to clump together in one location (a city). Absent economies of localization, no one will move to a city (cf. Glaeser et al. [GKS01]). Specification (6) is the minimally invasive way to do so without introducing added complications from type-dependent productivity.

Firms are a price taker and earn zero profit in equilibrium. Thus, each worker earns

$$w_i = a. \tag{7}$$

Evidently, the model does not pick up on the urban wage premium (cf. Yankow [Yan06]), which I can easily incorporate by having a dependent on s_i in (6). Besides, my focus is imperfect mobility, not urban wage premium. Note also that disposable income, i.e., wage w_i exclusive of the cost of relocation $\rho(\cdot)$, still varies city to city, as y is drawn from type-specific $f_i(y)$, which in turn changes the realized value of $\rho\left[l_i(x),\ y\right]$ by type, and subsequently, the disposable wage $w_i - \rho(\cdot)$ by city.[7]

2.3 Ryan the Rural Resident

Let us turn to Ryan. He is a type-i consumer who stays put. He becomes a Robinson Crusoe-type rural resident to lead a life under the backyard capitalism. His marginal product gets pushed back to $A(s_i = 0) = 1(< a)$ according to (6), housing consumption becomes independent of the city size, and the cost of relocation turns $\rho(0,\ y) = 0$. I mark his maximum utility level by $\underline{v}_i$, which, by construction, is independent of x. It still does depend on the housing units available in a rural area and how costly $\rho(\cdot)$ is to Chloe (and thus how cost-saving it is to stay in the countryside by comparison). In order to keep the model on point, assume that the land in the rural area is abundant enough and the number of people who do not move out of the birthplace will not affect the value of $\underline{v}_i$.

[6] Labor supply is not presumed to depend on the distance moved. An out-of-town worker may work **more** to compensate for the lifetime cost of relocation. She may as well work **less** to spare time in rebuilding her social network in her newly acquainted foreign soil. In the absence of available evidence, I shall take a neutral stance on labor supply.

[7] See appendix A.1 for more on the effective wage.

Furthermore, assume $\underline{v}_i$ is the same across the types so as to remove arbitrariness. Variations in city size are already captured by $f_i(y)$ and I do not need two sources of variations. I will discuss this further when I determine the type composition between Chloe and Ryan in (14).

2.4 Diane the Disoriented

As a third option, another type-i consumer, Diane, could move to a type-incompatible city $j(\neq i)$. In this case, she will not receive a type-specific productivity boost either, i.e., $A(s_i = 0) = 1$. Worse yet, her housing consumption will be smaller than the rural value above due to the likely presence of type-j residents choking up the rent, whom Diane splits the land supply H with. As I will show later in (13), Ryan's utility level is at most Chloe's. Thus, Diane in city j makes the same income (i.e., 1) as Ryan while having a smaller house than Ryan's. She will then achieve an even lower utility level than his. As such, I will safely rule out this last option in equilibrium.[8]

2.5 Trans-Tolerance Value

For any given i,

$$c_i(x,\ y) \ = \ w_i - \rho\,[l_i(x),\ y] - \eta, \quad \text{and} \tag{8}$$

$$h_i(x,\ y) \ = \ \frac{\eta}{p_i} \tag{9}$$

maximize (1) subject to (2). Chloe's indirect utility function is then

$$v_i\,(p_i,\ w_i(x);\ x,\ y,\ s_i) = w_i - \rho\,[l_i(x),\ y] - \eta + \eta\,(\log\eta - \log p_i)\,. \tag{10}$$

Housing supply is H in each city and the housing market clears when

$$s_i h_i = H, \tag{11}$$

from which I obtain the equilibrium rent $p_i = \dfrac{\eta s_i}{H}$, i.e, the more crowded the city becomes, the more expensive the rent per unit will be. Note that the expenditure on housing is always $p_i h_i = \eta$ regardless of the city size: Chloe copes with an increasing city size by reducing her lot size without changing her expenditure share of housing.[9]

Firm's first order condition (7) and housing market clearance (11) further simplify her indirect utility function (10) to

$$v_i\,(a,\ x,\ y,\ s_i) = a - \rho\,[l_i(x),\ y] - \eta\log s_i + \eta(\log H - 1). \tag{12}$$

Unsurprisingly, Chloe's utility level drops with distance x, holding everything else constant. Notice the trade-off among the economies of agglomeration a, diseconomies of agglomeration $-\eta\log s_i$ and distance tolerance y. Holding the value of $v_i(\cdot)$ constant, if the destination city becomes crowded or the productivity boost a gets smaller, the only residents with high enough tolerance y would become a Chloe, or else they are better off becoming a Ryan.

Now consider who becomes a Chloe and who becomes a Ryan. A type-i consumer will become a Chloe if her utility level (12) is greater than Ryan's $\underline{v}_i$:

$$v_i\,(a,\ x,\ y,\ s_i) = a - \rho\,[l_i(x),\ y] - \eta\log s_i + \eta(\log H - 1) \geq \underline{v}_i. \tag{13}$$

A resident at the margin meets (13) with equality. Since $\rho(\cdot,\ y)$ is strictly monotone decreasing in y, one can solve (13) with equality for y to define a **trans-tolerance** function

$$y_i(x) := \rho^{-1}\left[l_i(x),\ a - \eta\log s_i + \eta(\log H - 1) - \underline{v}_i\right]. \tag{14}$$

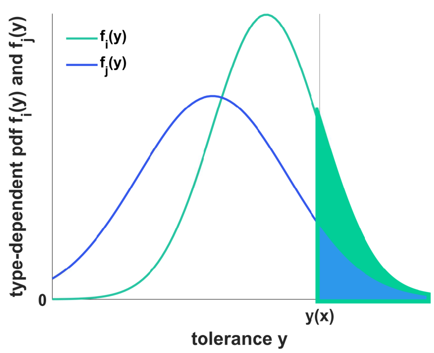
[8] At the intraurban level, there is an evidence of residents, albeit reluctantly, going for this option. See Coulson et al. [CLW01].

[9] Note also that as long as $s_i > 0$, any additional increase in s_i will not benefit Chloe in contrast to models that feature agglomeration. In particular, agglomeration only swaps 1 in (6) with a that appears in (12) but nothing more regardless of the city size. Thus, for Chloe, the best-case scenario is to have an infinitesimally small population in city i to unlock urban productivity (but she has no control over the population).

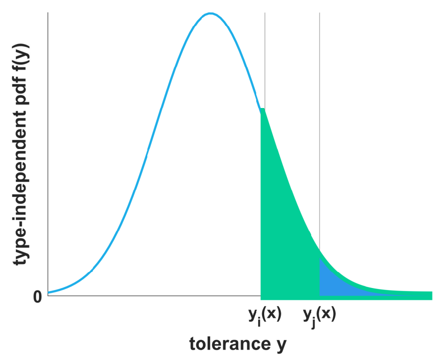
A couple of observations on (14) are in order. First off, $y_i(x)$ determines the fraction of people to become a Chloe. Anyone who drew $y \geq y_i(x)$ does not show much affinity to her birthplace or her opportunity cost of staying put is too high and moves out, whereas anyone with $y \leq y_i(x)$ has a lot to lose by relocation and stays in. Thus, the higher the trans-tolerance is, the higher the ratio of Ryan's will be. In this regard, trans-tolerance can be thought of as the minimum tolerance level required to make consumers want to move out of their birthplace, or analogously, the maximum tolerance level permissible to have consumers stay in, making it the threshold value of tolerance where the phase switches.

Secondly, $y_i(x)$ is increasing in x because $\rho(\cdot)$ is increasing in $l_i(x)$. In the vicinity of city i, the number of Ryan's is very small because it does not take much to turn residents a Chloe. As a result, the borderline tolerance is pretty low. However, as the distance to i increases, the cost of relocation bears down on the consumers and they will not become a Chloe as easily as before unless their tolerance is high, making the borderline tolerance high as well.

There are in fact two ways to go about the trans-tolerance value. One is to assume that $y_i(x) = y(x)$ for all i. The other is to allow $y_i(x)$ to take different values depending on the type. I will explain the difference between them below.



(a) Type-independent trans-tolerance.



(b) Type-dependent trans-tolerance.

Figure 5. Shaded areas mark a measure of Chloe's. In either case, type i has more Chloe's than type j.

First suppose that $y_i(x) = y(x)$ for all i at any $x \in X$. If $y_i(x)$ will be the same across the types, then y should have been drawn from different distributions depending on the type as in figure 5(a) (or else the city-size distribution will be uniform). In this case, if $f_i(y)$ first-order stochastically dominates $f_j(y)$, then $s_i > s_j$ (cf. (16) below). Type i should be more distance-tolerant than type j so that at any given x, more of type i must have drawn $y \geq y(x)$ than type j. In this case, y can be interpreted as a skill level that indicates the favorable degree of concentration of workers. Industry j features low-skill labor that does not benefit from concentration of workers within the same industry. Consequently, their distance tolerance is drawn from the distribution with a low mean. By contrast, industry i involves a type of workers who capitalize on interactions among them at a large scale. Type-i workers have a lower barrier to relocation because their opportunity cost of staying put and not tapping into their urban productivity $A(s_i)(> 1)$ is substantially high.

Now suppose instead that $y_i(x)$ can differ from $y_j(x)$ but y itself is drawn from the identical distribution regardless of the type as in figure 5(b). In this case, if $y_i(x) < y_j(x)$, then $s_i > s_j$. The variation in city size comes directly from (14), rather than the distribution from which y is drawn. City i has a larger influx of people because the net effect of agglomeration $a - \eta \log s_i$ is large and/or the fallback utility level v_i is low. Once again, type i is likely to be a high-skill type whereas industry j does not call for much concentration of workers.

Empirically speaking, I cannot tell which one is at work because I do not have direct observations of $f_i(y)$ or $\rho^{-1}(\cdot)$. To cover all bases, I will consider both type-dependent and -independent trans-tolerance for the empirical analysis in section 3. For the theoretical analysis to follow, I will pick type-independent trans-tolerance (figure 5(a)) for the following reasons.

I make two simplifications in order to implement the first way in a consistent manner. First one is what I already made in section 2.3. The fallback value $\underline{v}$ is shared across the type so that, holding everything else constant, trans-tolerance will not change by type because of the utility level Ryan achieves. Second, I take s_i that appears in (14) to be an expected value of s_i for technical reasons. In the current model, after y is drawn, consumers move on to the next stage, where everyone makes a simultaneous decision on his/her location. Consumers only know their draw of y but presumably do not know what distribution it is drawn from (neither do we in reality). Consequently, they cannot accurately compute the value of s_i before making decision. For this reason, consider s_i in (14) to be an expected value of s_i, shared across the type. In this way, I keep $y_i(x)$ constant with respect to i. Nevertheless, the type-difference will be factored into the equilibrium through a different channel, i.e., through a type-variant $f_i(y)$.

It is ideal to have the expected city size constant across the type. In this way I can ascribe all the variance in size to imperfect mobility rather than to the existing city size. However, in reality, location choice is not simultaneous and consumers make sequential decisions on location knowing the existing city size. I can easily model this by letting the estimated value of s_i in the first stage match the realized value in the second stage, which effectively turns the model into figure 5(b). That said, even if I opt instead for type-independent $y(x)$, the model is still consistent in that it concludes in two stages. Once consumers make their location choice, they cannot take back their decision and readjust their location upon the realization of s_i, as there are no subsequent periods to do so. Furthermore, since I forgo the utility equalization in equilibrium (cf. definition 2.2), workers are not set to readjust their location anyway.[10]

While the **ratio** between Chloe and Ryan declines with distance, the **measure** of city residents flowing in from location x,

$$m_i(x) := n_i \frac{\sin x}{2} \int_{y(x)}^{\infty} f_i(y) dy, \tag{15}$$

is not necessarily monotone in x (I will come back to this shortly). Consequently, the city size is

$$s_i = \int_0^{\pi} m_i(x) dx = \int_0^{\pi} n_i \frac{\sin x}{2} \int_{y(x)}^{\infty} f_i(y) dy dx. \tag{16}$$

Note that $s_i \leq n_i$, i.e., not all the type-i residents become a Chloe unless $y(x) \to -\infty$. Thus, the current model could be regarded as a variant of Harris-Todaro model [HT70] with migration decision made on the basis of urban-rural **utility** differential rather than expected **income** differential. Consequently, in the present model, the urban population $S = \sum_{i=1}^{I} s_i$ is at most equal to the total population $N = \sum_{i=1}^{I} n_i$ (and probably less).

In fact, most of the models of city-size distributions can be thought of as a limiting case of the present model where trans-tolerance tends to infinity ($y(x) \to -\infty$) so that everyone becomes a Chloe. This can happen in a couple of different ways. Looking at (14), if I remove the concept of distance, i.e., $l_i(x) = 0$ so that any $x \in X$ is equidistant and effectively 0 km away from city i, then no one bears the cost of relocation $\rho(0, \cdot)$ so that for sufficiently low $\underline{v}_i$ everyone becomes a Chloe and the size distribution turns uniform. Similarly, if a or H becomes overwhelmingly dominant and/or $\underline{v}_i \to -\infty$, even in the presence of sensible distance $l_i(x)$, everyone moves to city i. The resulting city-size distribution $(s_i)_{i=1}^{I}$ becomes the (exogenous) type distribution $(m_i)_{i=1}^{I}$ itself. Existing models endogenously derive the city-size distribution using other factors of choice than imperfect mobility to frame agglomeration (the one more complicated than (6)).

2.6 Inflow as a Function of Distance Moved

Turning back to the aforementioned monotonicity, I establish

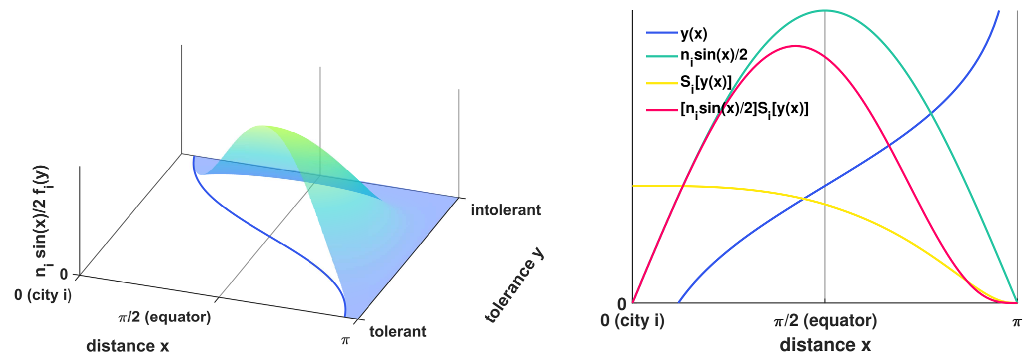
Proposition 2.1 Immigration and Distance Moved

For any type i, the inflow of workers $m_i(x)$ declines with distance moved for $x \in \left[\frac{\pi}{2}, \pi\right]$.

Proof. Both $\int_{y(x)}^{\infty} f_i(y) dy$ and $\sin x$ decrease within the range of distance in question, making $m_i(x)$ decline with x. $\square$

[10] I will lay off subscript i in $y_i(x)$ until section 3.

Remark. While this proposition sounds plausible and desirable, it is partly an artifact of having a ball for a country. In figure 6(b), an unconditional mass of Chloe (the yellow line) drops with distance, whereas the baseline population (the green line) only does so past the equator. Thus, the product of the two (the red line) will decline past the equator for certain. What is important is the part that I **cannot** establish a proposition for, i.e., what happens between 0 (city i) and $\frac{\pi}{2}$ (the equator). I expect that the percentage of Chloe's naturally declines with distance for increasing cost of relocation. However, conditional on y, the base population itself will increase with x. The size of the locations equidistant from city i is $\frac{\sin x}{2}$. With the assumption of a uniform[11] distribution of consumers, there are $n_i \frac{\sin x}{2}$ of potential Chloe's and Ryan's, which increase till x hits the equator. I do not know, when combined, how these two factors play out together. Between $x = 0$ and



(a) (b)

Figure 6. The blue line is trans-tolerance $y(x)$ in both figures. The part sliced off in figure 6(a) is the mass of Chloe's and the remaining portion is Ryan's. $S_i(y)$ denotes their survival function of $f_i(y)$. In figure 6(b), the yellow line is decreasing if the blue line is increasing and vice versa. The red line is the product of green and yellow lines, i.e., the volume of the part sliced off in figure 6(a). It is monotone decreasing in the lower hemisphere $x \geq \frac{\pi}{2}$ (proposition 2.1) but may not be so in the remaining hemisphere.

$\frac{\pi}{2}$, the migration rate (the red line in figure 6(b)) actually picks up first for increasing land size $n_i \frac{\sin x}{2}$ before it eventually starts to wind down as reduction from $\int_{y(x)}^{\infty} f_i(y)dy$ eventually overwhelms the increasing land size.

In view of this, one must be careful when interpreting estimation results in section 3. Just by the look of $m_i(x)$, the real effect of distance intolerance may be offset by a mere geometric (and not economic) fact that the perimeter of concentric circles gets longer as one moves far from center i towards $x = \frac{\pi}{2}$.

Lastly, $\rho^{-1}(\cdot)$ is decreasing in y as does $\rho(\cdot)$. Then holding else constant, the higher the in-city boost a is, the lower the trans-tolerance will be, and consequently, the more residents will become a Chloe. The opposite is true for diseconomies of agglomeration $-\eta \log s_i$ or the fallback value $\underline{v}_i$. Thus, there is a trade-off between a and s_i or $\underline{v}_i$ for given $y(x)$. To keep to the same $y(x)$ while a increases, s_i and/or $\underline{v}_i$ must be large as well.

As I alluded to in section 2.4, nobody becomes a Diane. Living in a wrong city, her composite good consumption evens out with Ryan's but her housing consumption falls behind his. She will be better off staying put (to become a Ryan) or moving to her type-concordant city (to become a Chloe) than moving into a city she is not cut out for.

Now that I know all consumers' location choice, I can define the equilibrium.

[11] I can easily bypass this problem by banning Chloe and Ryan's parents from giving birth in the upper hemisphere and making city i an isolated island. This is indeed a viable scenario for workers moving to Alaskan or Hawaiian cities. For workers with their type-matched cities found in the 48 contiguous states, the more likely scenario is that the initial distribution is not uniform but rather the opening population declined with distance from potential city sites to begin with. Since I do not have a way to know what the real initial distribution was, this scenario does not apply here.

2.7 Competitive Equilibrium

<u>Definition 2.2 Competitive Equilibrium</u>:
An equilibrium is a feasible allocation $[c_i(x, y), h_i(x, y), g_i(x, y)]_{i=1}^I$ *and* $(z_i)_{i=1}^I$*, and price system* $(p_i)_{i=1}^I \in \mathbb{R}_+^I$*, such that* $[c_i(x, y), h_i(x, y), g_i(x, y)]_{i=1}^I$ *maximizes the utility level and* $(z_i)_{i=1}^I$ *maximizes the profit under* $(p_i)_{i=1}^I$ *for any* $(x, y) \in X \times Y$ *and* $i \in \{1, \cdots, I\}$

Note that since Diane fails to maximize her utility level and thus such a consumer cannot exist in equilibrium, (4) and (5) simplify to

$$s_i A(s_i) = \int_0^\pi n_i \frac{\sin x}{2} \int_{y(x)}^\infty c_i(x, y) dF_i(y) dx, \tag{17}$$

$$H = \int_0^\pi n_i \frac{\sin x}{2} \int_{y(x)}^\infty h_i(x, y) dF_i(y) dx, \quad \text{and} \tag{18}$$

in equilibrium.

Furthermore, using Chloe's demand for goods (8) and the first order condition (7), and (17), the city size can be written as

$$s_i = \frac{1}{a} \int_0^\pi n_i \frac{\sin x}{2} \int_{y(x)}^\infty \{a - \rho[l_i(x), y] - \eta\} \, dF_i(y) dx. \tag{19}$$

The first item, $\frac{1}{a}$ on the right-hand side of (19), may seem odd at first. How come the city size declines with urban productivity? The answer is threefold. The first two reasons are commonplace among the models of the city-size distribution, whereas the last one is specific to the current model.

First, it simply means that if urbanites are productive, it takes less of them to meet the same demand, ceteris paribus. Moreover, this adverse effect is actually offset by two forces. To begin with, increased productivity raises wage, which in turn increases demand as can be seen in (8). The city then needs more workers to cater for growing demand, increasing s_i in the end. In most models, this is the end of the story, and usually the second force (increase in demand) outweighs the first force (reduction in employment due to better technology to lower required labor), because of the usual size-dependent urban productivity as opposed to the size-independent productivity (6). Consequently, urban productivity does not reduce the city size but rather increases it.

Furthermore, unique to the current model, better urban productivity lowers trans-tolerance so that those who would otherwise stay put will become a Chloe upon realizing an increase in a. The overall effect depends on the exact shape of $f_i(\cdot)$. If $f_i(\cdot)$ is low near $y(x)$, then an increase in a may backfire because $A(s_i)$ is only binary and does not have the city size built into it. In this case, immigration is not sensitive to a change in $y(x)$ and by extension, in a. The consumer base does not grow much and the city may not need as many workers after all.

Notice the curious lack of oft-cited utility equalization in definition 2.2, namely, I do **not** require $v_i(a, x, y, s_i) = v_j(a, x, y, s_j)$ for all i, j (compare this to, for example, the first equation on the right column on page 1446 of Eeckhout [Eec04] or equation (8) in Berliant and Watanabe [BW18]). The utility level can and does vary by type and by the initial location of a consumer in equilibrium. The utility level varies by type in Eeckhout et al. (equation (3) in [EPS14]) as well but mine further varies by birthplace.

By and large, a Chloe who happened to be born close to her type destination achieves a higher utility level than those who did not. There is no way to arbitrage this differential among Chloe's in the current model model as they **cannot** choose or optimize their place to be born into. **Neither can we in reality**.

The only utility equalization that is guaranteed to take place in the present model is that in the continuum of distance tolerance, a Chloe who happens to draw $y = y(x)$ achieves the same utility level as Ryan's at each $x \in X$. Recall that trans-tolerance $y(x)$ is so defined to pin down the ratio between Chloe's and Ryan's. Therefore, the equilibrium utility level does not equate between Ryan and Chloe either, except at the margin.

Note also that the competitive equilibrium, if exists, is also Pareto optimal. I did not use externalities of any form, be it positive or negative, to break the otherwise uniform distribution of economic activities. Congestion, a usual source of negative externalities, exists here but it is priced through the housing market and thus it is modeled not as real externalities but only as pecuniary externalities. Agglomerations, a usual source of positive externalities, take place too but in its uncomplicated form (6), adding another Chloe to a city does not increase an incumbent Chloe's productivity, which already took a jump from 1 to a when the first Chloe moved in and has hung at a since. Thus, the first fundamental theorem of welfare economics applies here.

2.8 City-Size Distribution

Since I cannot derive the city-size distribution without specifying the form of $f_i(y)$, I will narrow down some candidate forms. First of all, I need to be able to tell types apart using only $f_i(\cdot)$. I did not introduce heterogeneous productivity in (6) in order to keep the focus on geographical mobility. Thus, any difference in city size needs to arise out of heterogeneous distance tolerance: if $f_i(\cdot)$ and $f_j(\cdot)$ are identical, then, barring any other source of heterogeneity, s_i and s_j become identical too as can be seen from (3).

That said, I may introduce heterogeneity via the cost of relocation $\rho(\cdot)$ instead of $f_i(\cdot)$ to the same effect. Put differently, I can define $\rho_i(\cdot)$ by type. Therefore, I have some degree of freedom. In total there are three possible ways to move forward from here:

1. Assume $f_i(\cdot)$ and $\rho(\cdot)$: Type differs only in distance tolerance but once y is drawn, consumers recognize it in the same way regardless of their type.

2. Assume $f(\cdot)$ and $\rho_i(\cdot)$: Distance tolerance is identically distributed but the way consumers recognize it differs by type.

3. Assume $f_i(\cdot)$ and $\rho_i(\cdot)$: A combination of both.

(Note that at least one of the two has to be type dependent or else people will end up in a uniform size distribution). Researchers are free to pick whichever depending on the data availability. Unfortunately, there are no data available on either one of them as far as I am aware. Since $\rho(\cdot)$ is bivariate while $f_i(\cdot)$ is univariate, I will take the first option for tractability reasons.

Along with these two functions, one may also make n_i vary with i. However, n_i is orthogonal to mobility. To turn off its explanatory power, I assume n_i is constant across i. Bear in mind that it is not useful to collect the data on the size of an individual industry to estimate n_i. The size of the industry only represents the people who **did** move to the city, but not the potential number of people who are best suited for the industry but did **not** make it to the city due to their low distance tolerance. Nevertheless, I will keep the subscript for comparative statics later to examine the trade-off between n_i and $f_i(\cdot)$.

Given this, I can rewrite city size (16) as:

$$s_i = \int_0^\pi n_i \frac{\sin x}{2}\, S_i[y(x)]dx, \tag{20}$$

where $S_i(y) := 1 - F_i(y)$ is the survival function of y.

I will differentiate distance-tolerance distributions by some criteria to make testable predictions out of (16). There are various ways to rank density functions. I propose two of them below and discuss their implications on the city-size distribution.

PROPOSITION 2.2 FIRST-ORDER STOCHASTIC DOMINANCE AND CITY SIZE
If $f_i(\cdot)$ first-order stochastically dominates $f_j(\cdot)$, $s_i \geq s_j$ in equilibrium.

Proof. Suppose that $f_i(\cdot)$ first-order stochastically dominates $f_j(\cdot)$. For any given location x, $S_i[y(x)] \geq S_j[y(x)]$. Integrating both sides of the inequality over the country with the density of $n_i \frac{\sin x}{2} = n_j \frac{\sin x}{2}$ at each x, one obtains

$$s_i = \int_0^\pi n_i \frac{\sin x}{2}\, S_i[y(x)]dx \geq \int_0^\pi n_j \frac{\sin x}{2}\, S_j[y(x)]dx = s_j \tag{21}$$

from (20). □

Remark. Furthermore, if $S_i(y) > S_j(y)$ at $y = y(x)$, (21) holds with strict inequality. Observe that the inequality will be flipped if n_j is sufficiently larger than n_i. That is, there is a trade-off between the distance-tolerance distribution and the number of potential city residents. Even when type i is overall tolerant towards relocation, its corresponding city size may be trumped by more intolerant type j if type i is outnumbered by type j in the hinterland to begin with.

It should also be noted that proposition 2.2 will not hold if estimated s_i has to match the realized value, and/or Ryan's utility level depends on his type attribute. In that case there could be a rank reversal. For instance, if type i's default value $\underline{v}_i$ is higher than type j's, or the expected value of s_i is higher than that of s_j (city i is deemed to be overcrowded), then s_i may fall behind s_j because $y(x)$ of type i will be higher than type j's.

The next proposition is useful for the empirical analysis to follow:

PROPOSITION 2.3 DISTANCE ELASTICITY OF INFLOW AND CITY SIZE

Suppose that the inflow can be written as

$$n_i \frac{\sin x}{2} S_i\left[y(x)\right] = \beta_{1i} + \beta_{2i} x.$$

If $\beta_{1i} = \beta_{1j}$ and $\beta_{2i} > \beta_{2j}$, $s_i > s_j$ in equilibrium. Similarly, if $\beta_{1i} > \beta_{1j}$ and $\beta_{2i} = \beta_{2j}$, $s_i > s_j$.

Proof. Suppose $\beta_{1i} = \beta_{1j}$ and $\beta_{2i} > \beta_{2j}$. The equilibrium city size (20) will be $s_i = \pi b_{1i} + \frac{1}{2}\pi^2 \beta_{2i} > \pi b_{1j} + \frac{1}{2}\pi^2 \beta_{2j} = s_j$. The similar argument goes for the second part of the proposition. □

Remark. I will discuss the meaning of β_2 in section 3.

Rearrange (15) and one can estimate trans-tolerance by

$$y(x) = S_i^{-1}\left[\frac{m_i(x)}{n_i \frac{\sin x}{2}}\right]. \tag{22}$$

3 Empirical Testing

3.1 Data Employed

I use US Census Bureau's American Community Survey 2009-2013.[12] The questionnaire asks which MSA a responder lived a year prior to the survey. A total of 381 MSA's report in- and out-migration[13] so that there are 381×381 entries of inflow and outflow recorded between each pair of cities.

I associate the migration data above with the estimated population in 2013[14] for two reasons. The most recent census taken before 2013 was 2010. However, since I frame the model with people moving first to settle the city size, it is not plausible to regress the result between 2009 and 2013 on 2010 data. The next census is in 2020, which is still underway at the time of writing.

As can be seen in figure 1(b), exclusion of Alaska and Hawaii improves the fit from $R^2 = .2185$ to .2752. However, the objective is not to find the best fit for this particular regression. I do not have any theoretical basis on which to precludes cities in these states. The only difference between the two states and the remaining 48 states is that there are either Canadian cities (not included) or the Pacific (not habitable) lying between them. However, the western part of the contiguous 48 states is not densely populated either (cf. figure 2(a)). If I exclude the two states, I may as well have to exclude California, which lists 4 entries on the roster of the 20 largest MSA's. I will keep them intact throughout this section.

In all the figures and tables to follow, ***, **, and * denote coefficient significant at 1%, 5% and 10% respectively.

[12] Data available at https://www.census.gov/data/tables/2013/demo/geographic-mobility/metro-to-metro-migration.html.

[13] There were 383 MSA's as of 2013, two of which were recently promoted from a micro statistical area to MSA. Their size are reported but inflow data were not yet recorded in the 2009-2013 survey and thus excluded from the study.

[14] https://www.census.gov/data/tables/time-series/demo/popest/2010s-total-metro-and-micro-statistical-areas.html.

3.2 Translation from Theory to Empirical Analysis

I will make several adjustments to the data above in order to test theoretical implications of section 2.

In theory, the initial distribution is uniform whereas in reality, all locations are pre-populated with a number of consumers inherited from the previous period. To make the initial distribution as close to a uniform distribution as possible and eliminate the initial heterogeneity, I normalize the inflow by the total outflow from the city of origin.

Turning to the geographic difference between the theory and the actual US, while a city is assumed to be a point in theory, an actual MSA takes up an expanse of land. I will use its centroid to compute its geodesic to incoming locations. In addition, the present model features a sphere and the maximum moving distance possible is $\sqrt{\pi}$ regardless of the destination. The actual US does not stretch over the entire sphere but rather cuts off at $.21\sqrt{\pi}$. Consequently, the maximum distance differs city to city. Among 381 MSA's, Carson City, NV[15] has the shortest maximum distance possible of 4,187 km, from Bangor, ME. In turn, Honolulu and Bangor have the longest maximum distance possible of 8,293 km, between each other. While the gap between the top and bottom of the maximum range is mitigated by the fact that Alaska and Hawaii are included, this may nevertheless contaminate the estimation results: I may inadvertently underestimate Carson City's size for the reason other than distance tolerance, i.e., even if there were someone willing to move to the city from 8,293 km away, that worker will not show up in the data because the country cuts off at 4,187 km. I may overshoot Honolulu and Bangor vice versa.

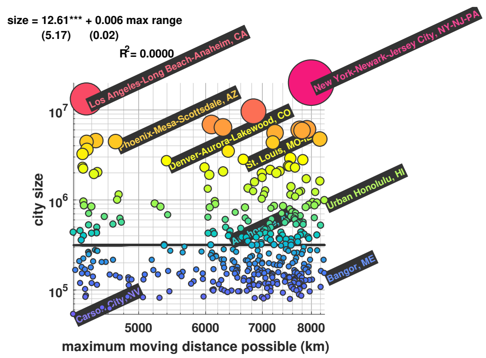


Figure 7.

That being said, I do not detect any systemic interaction between the maximum range and city size in figure 7.[16] The cap on the distance (unsurprisingly) does not affect the city size. While the longest cutoff is about twice as long as the shortest cutoff, consumers perceive the distance in a logarithmic scale. The perceived gap is thus much smaller than twofold as a linear scale implies. I will nevertheless regress city size on inflow **and** the maximum range in section 3.4. The latter captures the said non-economic constraints so that the coefficient on the former will not be watered down by their presence.

Some city pairs report a flow of zero. However, I need to take a log of the regressand as a linear scale is useless in the city-size context (cf. Limpert et al. [LSA01]). As with the city-size distribution itself, the vast majority of incoming cities are small. The difference between a flow of zero and one is marginal and should not be a determining factor in the estimations to follow.

In order to stabilize the readings, I impute the flow by raising them by one across the board. One may instead drop the cities with an inflow of zero. However, doing so will discount the fact that no one moved from those cities, the observation that is equally as important as the fact that someone moved from other cities. Besides, some estimates are sensitive to this removal, when they should not be. To compensate for this distortion, I will conduct a separate regression using quantiles in section 3.5, where I do not augment the

[15] Very coincidentally, Carson City was also the smallest MSA in 2013.

[16] To be consistent with the remainder of the current section, I measure the maximum distance in log scale. Since maximum distances are already quite long thanks in particular to inclusion of two Hawaiian MSA's, taking a log of them barely changes the estimation.

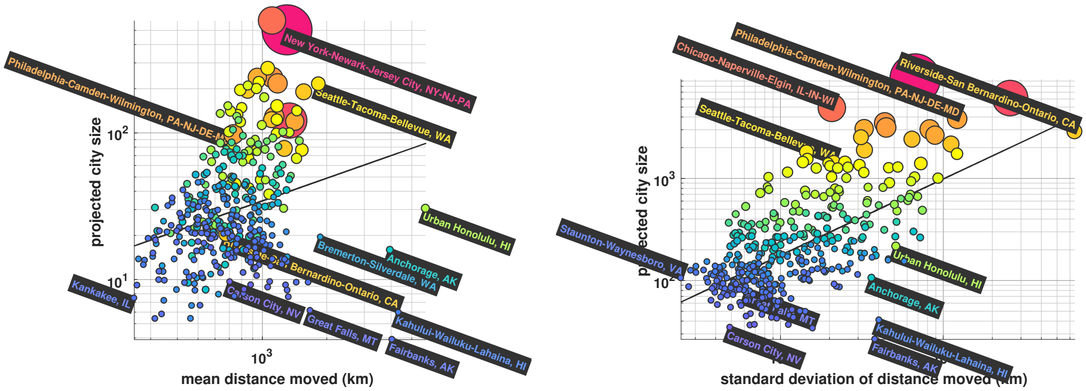
flow and leave the flow of zero as is.

I carry out three sets of empirical analysis to examine the nature of interaction between the distance tolerance and the city size by industry. In section 3.3 I regress the city size on the moments of $m_i(x)$ by industry. Section 3.4 examines the relationship between the size and the rate of decline of $m_i(x)$. Finally in section 3.5 I regress the size on quantiles to ensure robustness of the preceding two analyses.

Having done all the three estimations above, I will then extrapolate the trans-tolerance value for each type from the existing city size in section 3.6.

3.3 Regression on Moments

First, I regress the city size on the mean and standard deviation of its inflow in table 1. When the mean or the standard deviation of inflow inflates by 1%, the destination's size grows by .5% and .7% respectively. The findings indicates that the pdf flattens out and shifts to the right with the city size in line with proposition 2.2. For any given x, large cities have a high mean of inflow than small cities, which in turn implies that the former first-order stochastically dominates the latter in y. Note that if one assumes perfect mobility instead, the mean and variance of distance moved would be the same for any i and thus orthogonal to the city size, which is unlikely according to figure 8.



(a) (b)

Figure 8. Color and line width are size proportionate.

	intercept	mean	standard deviation	R^2	adjusted R^2
coefficient	7.365***	0.8059***		0.1079	0.1056
t-statistic	9.39	6.77			
coefficient	6.975***		0.8024***	0.4005	0.3989
t-statistic	19.38		15.91		
coefficient	3.988***	0.5123***	0.7487***	0.4423	0.4394
t-statistic	6.02	5.32	15.05		

Table 1.

There are some notable outliers. Four CDF's in Alaska and Hawaii (two each in each state) do not take off until later because they only have one city nearby (the one and only other MSA in the same state) and the next hike in value needs to wait till they cross the Pacific or Canada. They are largely a geographical artifact and do not necessarily mean that they gather high-skilled labor. Aside from them, among large cities, Philadelphia consists mostly of locally sourced labor, whereas Seattle and Riverside take in more globally oriented workers. However, conditional on the mean, Philadelphia and Riverside find residents of more diverse origin than Seattle does for their size.

While both of them are significant, table 1 indicates that the standard deviation exerts more influence on the size than the mean does. The size responds more to how widespread the cities of origin are than to how far people moved on average. Consequently, the aforementioned geographical artifact does not distort the projected city size in Alaska and Hawaii as much.

3.4 Regression on Distance Elasticity of Inflow

There are two more ways to summarize the inflow distribution than the moments examined above. In this section, I regress the city size on its inflow over x and compute coefficient β_{1i} on the constant and β_{2i} on x as in figure 1(b), not only for St. Louis but for all 381 MSA's. Then I further regress the city sizes on β_{1i} and β_{2i}. Table 2 and figure 9 report the results.

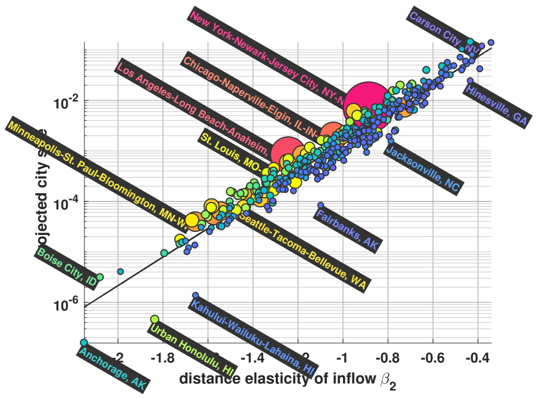


Figure 9.

	intercept	β_1	β_2	$\beta_1\beta_2$	max range	R^2	adjusted R^2
coefficient	12.79***	.2183***				.3641	.3624
t-statistic	285.53	14.73					
coefficient	11.22***		-1.394***			.1736	.1714
t-statistic	66.37		-8.92				
coefficient	20.35***	.9360***	6.914***			.7000	.6985
t-statistic	55.15	25.76	20.58				
coefficient	20.69***	1.101***	7.106***	.1287***		.7224	.7202
t-statistic	57.37	23.92	21.83	5.51			
coefficient	7.532***	.9224***	6.345***		1.394***	.7484	.7464
t-statistic	4.88	27.64	20.13		8.52		
coefficient	7.054***	1.109***	6.525***	.1462***	1.487***	.7771	.7747
t-statistic	4.84	26.85	21.88	6.95	9.60		

Table 2. Note that the reported values are the coefficients **of the coefficients** β_1 and β_2 **themselves**.

Since I take a log of x and s_i, β_2 measures the percentage increase in inflow against a 1% increase in distnace, i.e., the distance elasticity of inflow. I included the interactive term $\beta_1\beta_2$ as a regressor. While I cannot fully predict the city size by β_1 or β_2 alone, the product of the two tends to be high among large cities than among small cities.

Having controlled for the maximum range and other city-specific terms, the city size increases by 6.5% when the distance elasticity of inflow grows by 1%. That is, it is unlikely that a city has a large population **and** is unable to bring in people from afar. In turn, an industry with a low trans-tolerance value boasts a large city.

In figure 9, all four entries from Alaska and Hawaii cut below the estimated size. It is likely a systemic pattern. Given the size of the city, these cities should have lower β_2 and they would have if they were surrounded by other cities nearby. In reality, they are surrounded by Canada or the Pacific, neither one of them provides an inflow.

None of these would matter if one assumed perfect mobility. New York City may be composed exclusively of people from New England or of people from California, with no difference in its size in the end. In contrast, the present model anticipates that New York City cannot have the size it has unless it gathers workers from across the country. Conversely, Beaumont, TX cannot have many people moving in from California, which would otherwise indicate that high-skilled labor would move to Beaumont, running counter to the fact that the city is actually small. Unlike New York City, most of its inflow should and does originate from within its vicinity as its low β_2 value suggests.

3.5 Regression on Quantiles

Lastly, I regress the city size on quantiles. Unlike previous sections 3.3 and 3.4 I do not have to impute or alter an inflow of 0 here because I only take a log of distance rather than inflow. The results still point to the same direction as sections 3.3 and 3.4 do.

A 2-quantile estimation (the first row in table 3) indicates that the farther the median is, the larger the city will be, which meshes with figure 8(a) in section 3.3 and figure 9 in section 3.4. For 3 quantiles and above, regardless of the number of quantiles, the farther quantiles are always positively correlated with the city size. On the other hand, the nearest quantile is negatively correlated with the size only if the distance is split in 3 to 5 ways. Above 5, the quantiles are too fine and start to pick up noises.

Large cities' trans-tolerance grows slowly with x and consequently, $m_i(x)$ falls slower than small cities'. If the majority of the residents are moving in from remote locations, the last quantile should be far. The resulting city size should be large according to proposition 2.2. Figure 10(b) seems to support the proposition. By the same token, if the majority of residents are moving in from the nearby locations, the first quantile is short and according to proposition 2.2, the resulting city size should be small. Although not as strongly as the previous case does, figure 10(a) echoes the proposition as well.

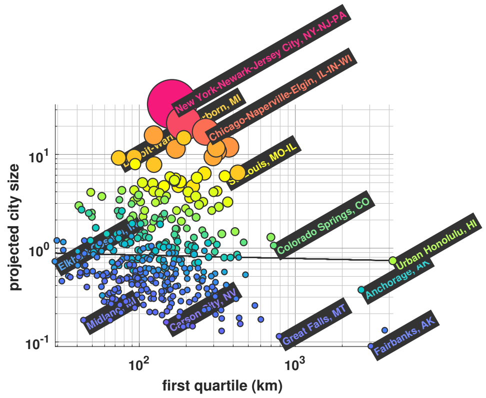
While R^2 values are lower than those found in tables 1 and 2 due to the lack of control for other variables, this analysis complements sections 3.3 and 3.4 and ensures its robustness.

	intercept	1st quantile	2nd quantile	3rd quantile	4th quantile	5th quantile	6th quantile	R^2	adjusted R^2
coefficient	1.217	1.332***						0.1372	0.1350
t-statistic	0.82	7.77							
coefficient	10.09***	-0.03591	.4353***					0.0875	0.0827
t-statistic	22.83	-0.33	4.23						
coefficient	9.748***	-0.116	0.2254*	.3015**				0.0968	0.0896
t-statistic	18.37	-1.01	1.96	2.48					
coefficient	9.260***	-0.03657	0.06634	0.1557	.3264**			0.0983	0.0887
t-statistic	15.3	-0.29	0.4	0.98	2.3				
coefficient	9.280***	0.1062	-0.3106*	0.3752*	0.06757	0.2694*		0.1034	0.0915
t-statistic	13.59	0.8	-1.76	1.95	0.37	1.7			
coefficient	8.895***	0.1801	-.5429***	.4859**	0.0266	0.07047	0.03307*	0.1118	0.0975
t-statistic	11.52	1.23	-2.59	2.27	0.12	0.32	1.76		

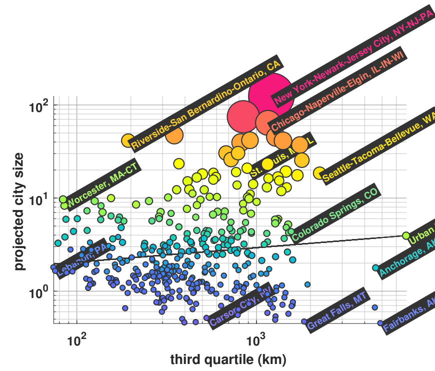
Table 3.

3.6 Estimation of the City's Type

By examining the distribution of the distance moved of incoming residence, one can infer which city corresponds to which type. For the time being, I will assume that y follows the normal distribution with mean μ_i and variance σ_i for expository purposes. (In practice, any distribution with the property in proposition 2.2 will do). I set μ_i equal to the log of the total inflow the destination receives, and shift it upwards by μ_{171} across the board so that the city of geometric mean size (Tuscaloosa, AL, 171st in rank in this case) will have



(a) First quartile



(b) Thrid quartile

Figure 10.

the mean of zero. The variance is set to unity. The actual tolerance distribution can be imputed from other observable variables such as the distribution of wage, the number of children or educational attainment, which I will leave for future research. I ran kernel density estimation on inflow first to filter out the noise.

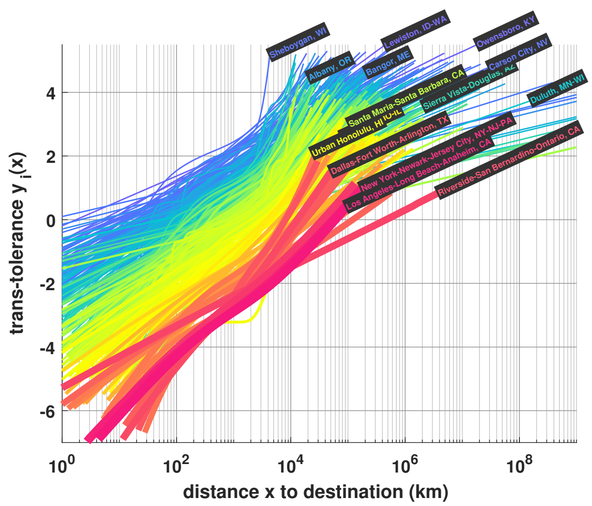


Figure 11. Trans-tolerance by distance. Line color and width are proportional to the total inflow into each destination.

Figure 11 plots $y_i(x)$ for all the destinations and table 4 lists $y_i(x)$ for select destinations at $x = 10^0, 10^1, \cdots, 10^5$ km.[17] All in all, large cities register lower trans-tolerance values than small cities. For instance, a resident $x = 10^3$ km away from her type destination has to have only $y = -2.87$ or above to move to Los Angeles, whereas that of Lewiston, ID and WA needs to exceed 2.37. If $y_i(x)$ is assumed (and likely) to be negatively correlated with the skill level for instance, then those who move to Los Angeles are more skilled than those who move to Lewiston, ID and WA.

Among cities listed in table 4, Riverside seems to attract workers from a wider range of locations than its cohort. Conversely, Albany, OR is more locally oriented compared to MSA's of similar size and thus less productive than what its size alone suggests.

[17]Some destinations record $-\infty$ in their immediate vicinity due to scant data points I have to obtain robust readings. These represents trans-tolerance values for residents who were born only 10^0 or 10^1 km away from their type destination, which are expectedly low. While x starts from 0 in theory, unfortunately, the majority of MSA's are more than 10^2 km apart from each other and I do not have enough data density from which to extrapolate those residents' accurate trans-tolerance value. They are likely to be recorded as movers **within** the same MSA in practice. By contrast MSA's are treated as a point in section 2 and consumers cannot conceivably move within a point.

rank	remark	MSA	total inflow	$y_i(10^0)$	$y_i(10^1)$	$y_i(10^2)$	$y_i(10^3)$	$y_i(10^4)$	$y_i(10^5)$
1		Los Angeles-Long Beach-Anaheim, CA	244,099	$-\infty$	-10.46	-4.84	-2.87	-0.24	4.13
2		New York-Newark-Jersey City, NY-NJ-PA	228,599	$-\infty$	-9.59	-4.51	-2.89	-0.37	3.86
3		Washington-Arlington-Alexandria, DC-VA-MD-WV	196,434	$-\infty$	-10.51	-4.26	-2.52	-0.19	4.23
4		Riverside-San Bernardino-Ontario, CA	178,510	-10.35	-6.21	-3.82	-2.02	-0.60	1.49
5		Dallas-Fort Worth-Arlington, TX	172,896	$-\infty$	$-\infty$	-5.15	-2.75	0.75	5.16
34		St. Louis, MO-IL	52,944	$-\infty$	$-\infty$	-3.99	-1.23	1.54	5.32
44	max range	Urban Honolulu, HI	41,804	$-\infty$	$-\infty$	-7.51	-3.22	0.85	5.95
95	1st quarter	Santa Maria-Santa Barbara, CA	22,928	$-\infty$	-7.51	-2.02	0.04	2.34	6.68
171	geometric mean	Tuscaloosa, AL	11,911	-7.11	-3.22	-0.95	0.65	2.14	4.25
191	2nd quarter	Sierra Vista-Douglas, AZ	10,576	-7.78	-4.12	-1.57	0.14	2.07	4.86
286	3rd quarter	Albany, OR	5,658	$-\infty$	$-\infty$	0.78	1.59	6.48	7.87
314	max range	Bangor, ME	4,540	$-\infty$	-6.34	-1.19	1.36	3.55	8.08
366	min range and size	Carson City, NV	3,062	-3.37	-1.29	0.39	1.85	3.22	4.75
381	last quarter	Lewiston, ID-WA	1,732	$-\infty$	-3.95	0.10	2.37	4.52	8.25

Table 4.

4 Conclusion and Extensions

I examined the role tolerance to distance moved plays in determining the city-size distribution. Each worker draws a distance tolerance level from the distribution unique to her type. She then makes a decision on whether to stay put or move to a city to tap into urban productivity that the city has to offer. She compares urban productivity with an urban housing market, a fallback value of her utility level when she stays, and of course, how far the city is from her birthplace when making a location choice. I derive the trans-tolerance value at which the workers splits between movers and non-movers, which is a function of the aforementioned location choice factors. The city-size distribution arises as a result of the trans-tolerance value specific to each industry and city.

I regressed the city size on several aspects of the underlying distribution of distance tolerance. The empirical data are in accordance with the predictions from the model.

I assumed that each city hosts at most one type. In reality, cities host multiple types. Assuming co-location of different types in the same city in in the same way as Eeckhout et al. [EPS14] would yield a finer result than above, provided that relocation data are recorded by industry. I do not know of such data.

In order to stay focused on the city size, I left urban productivity as plain as possible. In reality, the distance moved may reveal the immigrants' productivity levels, the aggregate of which determines the citywide productivity. It will be useful to relax the current assumption on urban productivity and have the distance moved explain it.

As examined in section 2.6, no one becomes a Diane in equilibrium. In reality, it is not easy to know in advance where the type-match city is located. Skill compatibility is not fully understood until workers actually start working at their destination, which may or may not be the right destination. One may introduce some uncertainty in matching between type and industry in order to analyze when and where Diane-type workers can exist in equilibrium.

I assumed that consumers live for two periods: a youth period in the city of birth and an adult period after they made a relocation decision. Some workers switch their cities multiple times in the course of their lifetime in reality. The current model can be extended to incorporate a longer time horizon.

As I alluded to when I estimated the trans-tolerance value from the migration data, ideally, I would like to start with the observed data rather than estimating from the migration data. While there are no data on distance tolerance per se, there are ways to infer its distribution. For instance, the number of children in the household may reveal their opportunity cost of labor. That implies their skill level, which in turn, can be used to gauge their trans-tolerance level as these two variables are deemed to be negatively correlated.

A Appendix

A.1 Selection of Utility Function

With the quasi-linear utility function, the difference in disposable income, i.e., wage exclusive of the cost of relocation $\rho(\cdot)$, is simply absorbed by $c_i(x)$, while $h_i(x)$ remains constant. That is, $h_i(x)$ is the same no matter how long the distance moved is, while $c_i(x)$ will adjust to accommodate the differences in $\rho(\cdot)$.

On the other hand, the counteracting force against agglomeration is captured solely via $h_i(x)$. The rent grows with city size s_i and with the fixed land supply of H, $h_i(x)$ declines with s_i. However, the city size itself will not affect $c_i(x)$.

Thus, with the quasi-linear utility function, I have a clean separation of variation in distance moved (registered exclusively through $c_i(x)$) and the diseconomies of agglomeration (registered exclusively through $h_i(x)$). As an added bonus, $c_i(x)$ becomes linear in disposable income in the end. I can regard $c_i(x)$ simply as a leftover income after deducting housing expenses and $\rho(\cdot)$.

This does not preclude the use of other specifications such as Cobb-Douglas or by extension, a CES utility function, which will violate one more assumption of Starett's theorem. I opt for the quasi-linear utility function to leave out complications not essential for my analysis.

References

[BC97] John H Boyd, III and John P Conley. Fundamental nonconvexities in Arrovian markets and a Coasian solution to the problem of externalities. *Journal of Economic Theory*, 72(2):388–407, 1997.

[BDRN14] Kristian Behrens, Gilles Duranton, and Frédéric Robert-Nicoud. Productive cities: Sorting, selection, and agglomeration. *Journal of Political Economy*, 122(3):507–553, 2014.

[BW18] Marcus Berliant and Axel H Watanabe. A scale-free transportation network explains the city-size distribution. *Quantitative Economics*, 9(3):1419–1451, 2018.

[CLW01] N Edward Coulson, Derek Laing, and Ping Wang. Spatial mismatch in search equilibrium. *Journal of Labor Economics*, 19(4):949–972, 2001.

[DD12] Donald R. Davis and Jonathan I. Dingel. A spatial knowledge economy. *Working paper*, 2012.

[Dur07] Gilles Duranton. Urban evolutions: The fast, the slow, and the still. *The American Economic Review*, 97(1):197–221, 2007.

[Eec04] Jan Eeckhout. Gibrat's law for (all) cities. *The American Economic Review*, 94(5):1429–1451, 2004.

[EPS14] Jan Eeckhout, Roberto Pinheiro, and Kurt Schmidheiny. Spatial sorting. *Journal of Political Economy*, 122(3):554–620, 2014.

[FHLS12] Oliver Falck, Stephan Heblich, Alfred Lameli, and Jens Südekum. Dialects, cultural identity, and economic exchange. *Journal of urban economics*, 72(2-3):225–239, 2012.

[FKV99] Masahisa Fujita, Paul Krugman, and Anthony J. Venables. *The Spatial Economy*. The MIT Press, 1999.

[GI04] Xavier Gabaix and Yannis M Ioannides. The evolution of city size distributions. *Handbook of Regional and Urban Economics*, 4:2341–2378, 2004.

[GKS01] Edward L Glaeser, Jed Kolko, and Albert Saiz. Consumer city. *Journal of economic geography*, 1(1):27–50, 2001.

[HJO16] Boris Hirsch, Elke J Jahn, and Michael Oberfichtner. The urban wage premium in imperfect labour markets. *working paper*, 2016.

[HT70] John R Harris and Michael P Todaro. Migration, unemployment and development: a two-sector analysis. *The American economic review*, 60(1):126–142, 1970.

[LSA01] Eckhard Limpert, Werner A Stahel, and Markus Abbt. Log-normal distributions across the sciences: Keys and clues. *BioScience*, 51(5):341–352, 2001.

[Man10] Alan Manning. The plant size-place effect: agglomeration and monopsony in labour markets. *Journal of economic geography*, 10(5):717–744, 2010.

[MSW11] Raven Molloy, Christopher L Smith, and Abigail Wozniak. Internal migration in the united states. *Journal of Economic perspectives*, 25(3):173–96, 2011.

[Rau13] Ferdinand Rauch. Cities as spatial clusters. *Journal of Economic Geography*, 14(4):759–773, 2013.

[Sta78] David Starrett. Market allocations of location choice in a model with free mobility. *Journal of Economic Theory*, 17(1):21–37, 1978.

[Woo11] Colin Woodard. *American nations: A history of the eleven rival regional cultures of North America*. Penguin, 2011.

[Yan06] Jeffrey J Yankow. Why do cities pay more? an empirical examination of some competing theories of the urban wage premium. *Journal of Urban Economics*, 60(2):139–161, 2006.