Skill, Agglomeration, and Inequality in the Spatial Economy*

Farid Farrokhi

Purdue

October 2018

Abstract

This paper develops a spatial equilibrium model with skill heterogeneity and endogenous agglomeration to study distributional welfare consequences of spatial policies. I show empirically that the relationship between log worker productivity and log city population is nonlinear in city size and in worker's skill. The model predicts these nonlinearities through local idea exchange between workers. I structurally estimate the model to match US Census employment and wage data, and use the estimates for decomposition and counterfactual exercises. A policy, with zero aggregate welfare effect, that favors smaller cities at the expense of larger cities, would notably reduce welfare inequality.

*I am thankful to Lorenzo Caliendo, Jonathan Dingel, Jonathan Eaton, Michal Fabinger, Russ Hillberry, David Hummels, David Jinkins, Ahmad Lashkaripour, Andres Rodriguez-Clare, and Chong Xiang for helpful comments and suggestions. All errors are my own. Correspondence: ffarrokh@purdue.edu.

1 Introduction

Economic activity and wage inequality largely vary across cities. Within-city wage inequality in the United States is strongly correlated with city population, and this correlation has strengthened over time. Several studies attribute these relationships to skill-biased agglomeration economies (Baum-Snow and Pavan, 2013; Baum-Snow et al., 2017). While these scale economies make room for policies that favor certain places or workers, little is known about their distributional welfare implications. The right implications, even for aggregate outcomes, crucially depend on the functional form of agglomeration and its interaction with other forces in a spatial economy (Glaeser and Gottlieb, 2008; Kline and Moretti, 2014b). For how scale affects productivity along the dimension of scale or skill, the spatial economics literature relies on strong assumptions on the empirical side, and have not delivered operational predictions on the theoretical side. The task of building worker heterogeneity into a spatial framework to characterize scale elasticities and to use the structure for counterfactual policy analysis remains incomplete.

This paper develops a spatial equilibrium framework to address the effects of place-based or skill-biased changes in policy and technology on the distribution of nominal and real wages across cities and across skills. The model features a continuum of skill levels, imperfect labor mobility across cities, and an endogenous formation of agglomeration economies through local idea exchange. The resulting scale economies vary by city size and across skill, although the model nests the commonly used power law specification. National supply and private productivity of skills are given, and the equilibrium determines employment, wage, and production of each skill in each city. The model is designed to maintain sufficient tractability to be estimated and used for comparative statics analysis.

I begin with a new observation by putting together three facts. It is well documented, and I confirm in the U.S. Census data, that (i) larger cities are more skill-abundant, (ii) worker-level agglomeration elasticity, i.e. elasticity of a worker's productivity w.r.t. city size, increases in the worker's skill, (iii) city-level agglomeration elasticity, i.e. elasticity of city-level productivity w.r.t. city size, appears to be invariant with city size. If the worker-level agglomeration elasticity was also invariant with city size, then facts (i)-(ii) would imply a higher city-level agglomeration elasticity for larger cities, contradicting fact (iii). To square with the facts, worker-level agglomeration elasticity must be decreasing in city size.

I design a framework that accommodates all these patterns and several other well-documented facts on employments and wages across and within cities. The model is built on the Marshallian insight that returns to scale can arise due to local idea exchange and the inability of individual producers to appropriate their contribution to knowledge accumulation (Marshall, 1920; Ethier, 1982; Davis and Dingel, 2018). Every worker allocates her time to two productive activities: private production and idea exchange with other workers. The opportunity cost of every minute a worker allocates to idea exchange is forgone private production; and, the benefit crucially depends on the extent to which others participate in idea exchange.

This collective interaction creates knowledge spillovers within a city from every worker to the others, making workers' productivities rise with city size.

The model delivers the right predictions under the complementarity between private returns to skill and social returns to idea exchange. Under this complementarity, an increase in skill raises the marginal benefit from idea exchange, encouraging more participation in idea exchange. Hence, given city size, worker-level agglomeration elasticity is increasing in the worker's skill. In addition, an increase in city size causes the available pool of ideas to expand. This raises the benefit from idea exchange although every worker will allocate less resources to idea exchange. These responses are implied by the complementary under which private production should keep up with the increase in the benefit from idea exchange. Because an individual's resources are limited, any further increase of time allocated to private production requires further decrease of that to idea exchange, dampening the effect of knowledge spillovers on productivity. Hence, given worker skill, worker-level agglomeration elasticity is decreasing in city size.

I bring the model to U.S. Census data by interpreting metropolitan areas as the units in which benefits from agglomeration and disutilities from congestion arise. I use equilibrium conditions to recover local productivity and amenity shifters given model parameters and observed population and income of cities. I estimate model parameters using the method of moments subject to the recovered productivity and amenity shifters. The moments consist of deciles of wage distribution within every metropolitan area, a condition that imposes a flat city-level agglomeration elasticity, and a condition that imposes an orthogonality between productivity shifters and variables that presumably do not vary systematically with labor demand shocks.

While parameters are identified jointly, each parameter matters more with respect to one of the moments. The slope at which skill premiums rise with city size is crucial to the extent of complementarity between private returns to skill and social returns to scale, with a higher observed slope implying a higher complementarity. Controlling for selection of skills to cities, the variance of wages within cities is governed by the dispersion of the distribution of national skill. The flat city-level agglomeration elasticity matters for a parameter that controls the size of knowledge pool given idea exchange participations. When this aggregation parameter is too large, the complementarity between private and social returns boosts agglomeration forces, and does so more in smaller cities. Hence, the aggregation parameter must be small enough for the model to generate flat city-level agglomeration elasticities. Still, different combinations of parameters can match all moments except the last, with each combination implies a different overall magnitude of the city-level agglomeration elasticity. The orthogonality condition identifies the right combination by determining this magnitude.

Using the estimates and model-implied equations, I decompose the city-level agglomeration elasticity into the contribution of (i) worker-level agglomeration elasticity and (ii) cities' skill composition. The estimation yields a city-level elasticity of 0.082, out of which 62% (= 0.051/0.082) is accounted for by work-level elasticities conditional on holding city skill composition fixed, and 38% (= 0.031/0.082) by endogenous changes to within-city skill distribution when city population changes.

I use my estimated model to perform a few key counterfactual experiments. First, I explore distributional welfare implications of place-based policies. I consider a policy that does not affect aggregate welfare in a special case of the model with homogeneous workers and a constant agglomeration elasticity. Specifically, this policy subsidizes cities below the median size at the expense of a tax equivalent rate of 5% on per capita income of residents of larger cities. Similar to the special case, I find that this policy does not change aggregate welfare. However, it reduces welfare inequality across skill groups. Expected welfare rises for workers in the 1st to 8th deciles of national skill distribution, and it falls only for the two top deciles. In particular, welfare ratio of top to bottom decile of national skill distribution decreases by 0.64%. Higher skill workers reside more than proportionately in larger cities, so on average they lose more (or gain less) from a policy that favors small cities. In addition, higher skill workers benefit more from a concentration of economic activity, a benefit that diminishes when population across cities becomes more equal.

This result has an important policy implication. Benchmark models in the literature leave little justification for place-based policies (Kline and Moretti, 2014a). In these models, labor is homogeneous and if city-level agglomeration elasticity is constant (empirically, the relevant case), then the gains to receiving areas are canceled out by the losses to other areas (See Proposition 3 in Section 4.4). By incorporating skill heterogeneity and flexible scale elasticities, the results in this paper shift our attention to the distributional welfare effects of these policies.

In addition, I examine the welfare implications of skill-biased changes in technology. Specifically, I study what real wage inequality across skills would have been in 2000, if private returns to skill were such that skill wage premiums had remained at their 1980 levels. To do so, I re-estimate private returns to skill using skill wage premiums in 1980. I find a notable rise in welfare inequality, highlighting the importance of complementarities that amplify returns to skill by endogenous sorting of higher skills to larger cities. In addition, the results highlight the importance of taking the tails of skill distribution into account. In particular, while the aggregate welfare effect on deciles 1-5 is negligible, the expected welfare of the 1st decile largely falls and that of the 5th decile notably rises. This result implies important disaggregated welfare effects that could remain confounded with predictions of frameworks that feature only two worker types.

Related Literature

I complement a literature that microfounds scale economies, particularly Davis and Dingel (2018), in two important ways. In my modeling I am inspired by Davis and Dingel (2018) to combine workers' choices of where to reside and how much to participate in local learning opportunities. However, they do not address (i) what these learning externalities imply for the functional form of agglomeration economies and for welfare effects of spatial policies, and (ii) how a model of this style can be connected to data for quantitative policy analysis. In this respect my contribution is two-fold: First, in a deliberately simpler setting, I characterize scale elasticities, conditions under which they exhibit non-linearities, and implications for distributional effects of place-based policies. Second, I take the model to data for estimation and quantitative analysis. Given these developments, I step forward to fill the gap between a theoretical style to microfound scale economies and the recent literature on quantitative spatial economics, as surveyed by Redding and Rossi-Hansberg (2017).

The relevance of idea exchange in generating local external economies has been emphasized by several generations of economists. Marshall (1920, IV.X.7) writes: "When an industry has thus chosen a locality for itself, it is likely to stay there long [...] if one man starts a new idea, it is taken up by others and combined with suggestions of their own, and thus it becomes the source of further new ideas."[1]. As the source of agglomeration in modern economies, a large body of evidence supports that city-level human capital predicts a steeper earning profile (Glaeser and Gottlieb, 2009; De la Roca and Puga, 2017).

This paper contributes to a body of literature that studies systems of cities, their determinants, and their implications for spatial policy (Henderson, 1974; Desmet and Rossi-hansberg, 2013; Combes et al., 2008; Behrens et al., 2014; Gaubert, 2017). Given large resources allocated to place-based policies, several studies examine their aggregate impact. Glaeser and Gottlieb (2008) summarize that "the mere existence of agglomeration externalities does not indicate which places should be subsidized. Without a better understanding of nonlinearities in these externalities, any government spatial policy is as likely to reduce as to increase welfare." Related studies have found no evidence for nonlinear agglomeration elasticities, implying that "where to subsidize" does not matter for aggregate welfare (Glaeser and Gottlieb, 2008, 2009; Notowidigdo, 2011; Kline and Moretti, 2014a) I show that the available evidence for linear city-level externalities implies nonlinear worker-level externalities. I present a microfoundation for this observation, then embed it into a spatial equilibrium framework to examine the effects of spatial policies on welfare inequality.

The growing literature on quantitative spatial economics has largely focused on homogeneous workers (Allen and Arkolakis, 2014; Redding and Rossi-Hansberg, 2017). My paper contributes to this literature by addressing how spatial development policies affect not only welfare at the aggregate but also the distribution of welfare. In addition, this study relates to a literature that identifies agglomeration economies (Rosenthal and Strange, 2004). My identification strategy is similar to (Ahlfeldt et al., 2015) who impose orthogonality conditions on location fundamentals. Since my model incorporates endogenous sorting of skills to cities, the estimates are obtained structurally by controlling for selection and skill composition. This structural approach complements reduced-form studies that are based on instrumental variables (among others, as a

[1] See Chipman (1965, Part 2.8) and Ethier (1982) on the role of external economies in international trade theory. In addition, the literature on economic growth shows how to generate external economies through a process of searching and learning (Lucas and Moll, 2014; Perla and Tonetti, 2014).

recent study, see De la Roca and Puga (2017)).

This study relates to a literature that employs tools from the assignment theory to study inequality. For example, Costinot and Vogel (2010) study the effects of globalization on inequality through the assignment of skills to tasks. In comparison, the analysis of the assignment of skills to *cities* inevitably departs from constant-returns-to-scale production technology as *scale economies* are ubiquitous in urban development.

Lastly, this study complements the empirical literature on spatial inequality. Moretti (2013) shows that in the United States skilled workers, due to sorting to larger cities, experienced higher housing costs, implying a lower welfare inequality relative to observed wage inequality. Diamond (2016) allows for a richer specification and accounts for skill-biased endogenous local amenities, implying a higher welfare inequality than suggested otherwise. Farrokhi and Jinkins (2017) embed domestic trade costs into a spatial framework to study how domestic trading infrastructure affects inequality. In a recent study, Fajgelbaum and Gaubert (2018) characterize the optimal spatial policy in a framework that features trade and multiple worker types, and find that relative to the efficient allocation the US system of cities features too much concentration of skills. In this paper, I build a microfoundation for agglomeration into a spatial equilibrium framework, which helps me derive the functional form of agglomeration as a result rather than an assumption. In particular, I show that the typical power law specification for scale elasticities (which is commonly used in economic geography and international trade) is only a special case of the productivity outcome of local idea exchange activities, and at the disaggregated worker level is not necessarily consistent with the data. Furthermore, I depart from specifying skill groups into college and non-college graduates. By incorporating a continuum of skills, my examination takes into account the tails of wage distribution, which is extensively shown to be an important dimension of inequality.

2 Data and Empirical Patterns

2.1 Data

I exploit wage and employment data of US Census 2000. The unit of observation is consolidated metropolitan statistical area (CMSA). A metropolitan area consists of "a large urban core together with surrounding communities that have a high degree of economic and social integration with the urban core." For large metro areas, a consolidated metro area consolidates neighboring areas to that large area in order for the entire unit to be even more representative of an integrated local labor market. I intentionally do not treat the non-metropolitan parts of a state as a unit of observation. These non-metropolitan areas within a state are in some cases far from each other, and so, they hardly reflect local agglomeration forces. The sample consists of 243 CMSAs. I use the word "city" interchangeably with CMSA.

I follow the standard practice in the literature to construct a sample from Census data. The sample covers all workers between age 25 to 55, who worked at least 40 weeks per year and 35 hours per week, and their hourly wage is at least 75 percent of minimum wage in 1999. Throughout my statistical analysis, I use person weights in order for the sample to be representative. A worker is said to be a college graduate if she has four or more years of college education. For every city, college employment ratio is the ratio of the employment of college graduates relative to that of others, and college wage premium is the ratio of average wage of college graduates relative to that of others. My focus, however, will not be on the college cut, rather on the entire wage distribution. To do so, I construct wage deciles for every city. Mean wage of decile $k = 1, ..., 10$ in a city is the average wage of all workers within the k-th decile of wage distribution within that city.

Table 1 reports the summary of statistics. The largest city is New York metro area[2] with a working full-time employment of 5.37 million. Mean hourly wage varies between 12.73 and 25.40. College wage premium ranges between 1.30 and 2.10, close to the range of the hourly wage ratio of decile 7 to 4. The lowest ratio of the 9th to 2nd decile is 2.41 for Wausau, WI with a working population of 37 thousand; and, the largest is 3.68 for Los Angeles metro area with a working population of 3.8 million. The lowest ratio of 10th to 1st decile is 5.22 for St. Cloud, MN with a working population of 45 thousand; and, the largest is 11.30 for New York metro area.

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**
working population	237,012	537,864	19,977	5,366,497
hourly wage	17.06	1.97	12.73	25.4
decile 10 to 1 hourly wage ratio	7.39	1.00	5.22	11.30
decile 9 to 2 hourly wage ratio	2.93	0.23	2.41	3.68
decile 8 to 3 hourly wage ratio	2.00	0.11	1.73	2.36
decile 7 to 4 hourly wage ratio	1.49	0.05	1.36	1.66
decile 6 to 5 hourly wage ratio	1.14	0.02	1.09	1.19
college employment ratio	0.42	0.16	0.15	1.02
college wage premium	1.61	0.12	1.30	2.10
Unit of observation	Consolidated Metropolitan Statistical Area			
Number of observations	243			

Table 1: Summary of Statistics, Census 2000

2.2 Empirical Patterns

I report several key observations on employments and wages across and within cities. Every of these observations has been documented in the previous literature. Therefore, in addition to my regression results, I

[2] In fact, it is a consolidated MSA called New York-Northern New Jersey-Long Island, NY-NJ-CT-PA that covers a total of ten MSAs reflecting an integrated regional labor market. The same consideration applies to other metro names.

summarize results from related studies. What is new here is the way I put these observations together to motivate a theoretical framework.

Pattern 1. *In more populated cities, (a) wages are higher, (b) wage inequality is higher, (c) skilled to unskilled employment ratio is higher.*

Figure 1 plots mean hourly wage as well as the ratio of top to bottom decile of wage distribution against population (employment size). The elasticity of city-level wage with respect to city population is 0.079. The elasticity for wage ratio of 10 to 1 decile with respect to population is even slightly higher, equal to 0.089.

Tables 2 and 3 report results from regressions that support Pattern 1. College wage premium is positively and strongly correlated with city population. In addition, wage ratio of decile 10 to 1, 9 to 2, 8 to 3, 7 to 4, and 6 to 5 are all positively and strongly correlated with city population. Given the positive correlation between wage inequality and city size, it is indeed expected that larger cities are skill abundant. College employment ratio increases with city population at an elasticity of 0.125. Taking college as a proxy for skill, this positive relation reflects that larger cities have a higher ratio of skilled to unskilled employment.

Among many studies, see Glaeser et al. (2009), Glaeser and Gottlieb (2009), and Davis and Dingel (2018) as further evidence for Pattern 1.

Pattern 2. *The relationship between wage inequality and city size is substantially more pronounced in the full distribution of wages than in college wage premium.*

College wage premium conceals important information in the entire wage distribution within a city. In particular, the relationship between wage inequality and city size is substantially stronger if we take the tails of wage distribution into account. The elasticity of college wage premium w.r.t. city population is four times smaller than that of the top to bottom decile wage ratio (0.023 compared to 0.089).

In addition, changes to the tails of wage distribution could be crucial for developments in wage inequality. It is well-known that the relationship between skill wage premium and city size has strengthened between 1980 and 2000. Tables A.2 and A.3 report results from the same regressions for year 1980. Between 1980 and 2000, the elasticity of college wage premium and city size has increased from to 0.016 to 0.023. However, important developments are confounded there. The elasticity of 10 to 1 decile wage ratio increased by three times from 0.031 in 1980 to 0.089 in 2000, and that of decile 9 to 2 increased by four times from 0.010 to 0.041. This elasticity for decile 8 to 3, 7 to 4, and 6 to 5 were not even statistically significant in 1980.[3]

At another level, educational attainment is only an imperfect indicator of skill particularly for higher-than-median segment of wage distribution.[4] (Baum-Snow and Pavan, 2013) rigorously document that a

[3] To make an example for putting the importance of the tails of wage distribution in a clearer perspective consider that in 1980 the ratio of top to bottom decile of within-city wage distribution was 6.3 in New York compared to 4.7 in St. Cloud. In 2000, those ratios were 11.3 in New York and only 5.2 in St. Cloud.

[4] For example, in 2000 Census data, I observe that mean wage of workers with doctorate degree is higher than that of workers with master's in only 54% of observations across American cities.

great part of the recent increase in US wage inequality stems from the increase in the variance of wage inequality *within* educational groups.

Pattern 3. *Returns from city size are greater for higher skilled workers.*
To provide some preliminary evidence, Table A.4 reports results from individual-level regressions of log wage against log population, college dummy, and the interaction term between log population and college, controlling for worker gender, race, and cubic years of experience. The elasticity of hourly wage with respect to population is 0.056 for workers with no college degree whereas it is 0.080 for college graduates (column 2). The difference between these coefficients reflects the larger returns from city size to higher skilled workers.

Many studies unanimously confirm this empirical pattern with more empirical rigor and using more detailed datasets. Bacolod et al. (2009) find that productivity returns from city size are larger for workers with stronger cognitive skills such as scientists, engineers, physicians, and lawyers, or stronger social skills such as social workers, therapists, and sales persons. De la Roca and Puga (2017) find that returns from city size are larger for workers with higher unobserved ability. Their estimation uses a panel of worker-level wages and location choices across the geography of Spain where worker fixed effects reflect unobserved ability.[5] Baum-Snow et al. (2017) find that at least 80% of the larger increase in skill wage premiums of larger cities are due to factor biases of agglomeration benefits. Similar conclusions are made by Gould (2007), Glaeser and Resseger (2010), and Combes et al. (2008).

Pattern 4. *City-level agglomeration elasticity appears to be invariant with city size.*
To examine whether the elasticity of city-level wages or productivities with respect to city size feature nonlinearities, I consider this specification,

$$\ln y(i) = \alpha_0 + \alpha_1 \ln L(i) + \alpha_2 (\ln L(i))^2 + \epsilon(i),$$

where $y(i)$ is either wage or a measure of labor productivity in city i, $L(i)$ is total employment there, and $\epsilon(i)$ is the residual. I obtain city-level productivities by inverting a standard model of economic geography. Specifically, I use the discrete version of Allen and Arkolakis (2014) which nests several other related frameworks and allows for the calculation of composite productivities without imposing a functional form.[6] I explore whether α_2 is nonzero, and if so, what its sign is.

Table 4 reports the OLS and IV regression results. The OLS estimate of α_2 is not statistically different from zero. However, OLS estimates could be biased because city population might be correlated with unobserved city productivity. A number of instrumental variables have been used in the literature to address this endogeneity. These include variables that are related to (i) natural amenities such as climate, (ii) historical activity such as past population, and (iii) housing amenities such as housing supply elasticity

[5] They are also able to identify dynamic returns from city size which they interpret as greater learning in larger cities.

[6] I am happy to provide this productivity calculation upon request.

or land regulations. In addition, if z is an exogenous variable that can be used as an instrument for $\ln L$, then other functions of z, such as z^2, can be added to the list of instruments for identifying the nonlinear effect of $\ln L$.[7] I continue to report IV regression results using these instruments, discuss the limitations of these regressions, and provide further evidence from the literature.

Across columns (2)-(5) I use the following instruments. In column (2), I use housing supply elasticity and its squared value. These elasticity estimates are borrowed from Saiz (2010). In column (3), I use log population in 1900 and its squared value. I obtain these data from the Census of 1900 where I can match them only partially to the CMSAs of 2000. In column (4), I use an index of land unavailability, an index of land regulation, and their squared values. The land unavailability index is taken from Saiz (2010) and the other is the Wharton Residential Land Use Regulatory Index (Gyourko et al., 2008). In column (5), I use housing supply elasticity and an index of natural amenity, taken from the US department of agriculture.[8]

Across these regressions, I find no statistically significant nonlinear effect of city population on city wage or productivity, as the coefficient of $(\ln L)^2$ is not statistically different from zero. In addition, although not reported here to save space, I have found no evidence for nonlinearities using alternative instruments, including other climate-related variables[9], interaction terms between climate and housing variables, and other geographic variables such as elevation, terrain ruggedness, and percentage of urban fringe overlaying aquifers, and other historical variables such as the extent of streetcars in 1902.[10]

These variables cover a wide range of instruments across numerous papers, but they are admittedly not ideal. For example, a better climate might be correlated with proximity to coastal areas which in turn might be correlated with international trade shocks to productivity. For climate variables, I typically find a somewhat weak first stage, whereas past population and housing variables show strong correlation with population.[11] See Combes and Gobillon (2015) for a detailed discussion on merits and limitations of these

[7] See Wooldridge (2010), Section 9.5, for a detailed discussion on using instruments to estimate nonlinear effects of an explanatory variable.

[8] This index is a weighted average of temperature in January, hours of sunlight in January, temperature in July, humidity in July, a measure of the desirability of land surface, and percent of water area, all as average values between 1941 and 1970.

[9] I obtained very similar results by using several other measures of natural amenities based on Climate Normals, which are three-decade averages of climatological variables. I used: number of days per year where maximum temperature $\geq$ 90 F (and 100 F), number of days per year where minimum temperature $\leq$ 20 F (and 32 F), annual cooling degree days with base 70 F, annual heating degree days with base 60 F, number of days during the year with precipitation $\geq$ 0.50 inches, number of days during the year with snow depth $\geq$ 1 inch (and 10 inches), probability of precipitation $\geq$ 0.50 inches for 29 day windows centered on each day of the year.

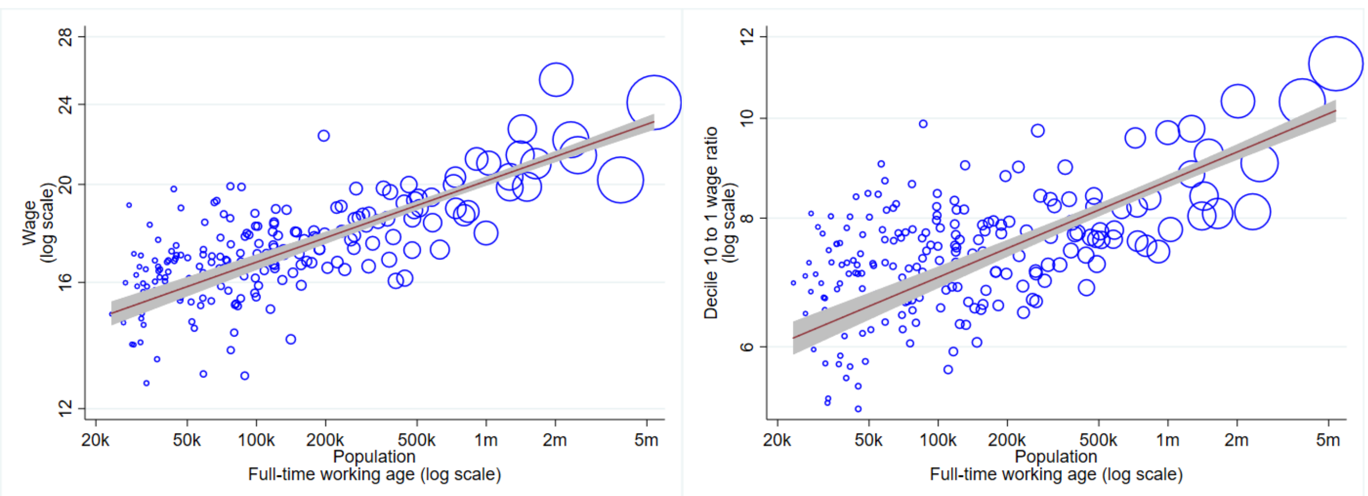
[10] I have taken these geographic variables from Burchfield et al. (2006) and streetcar data from Cutler et al. (1999).

[11] As for magnitude of the estimates compared to the available ones in the literature, an important consideration must be taken into account. Across models, one may allow for differentiated goods (by mobile firms or locations), or assume no differentiation. When goods are more differentiated, producers in smaller cities are in a better position to compete with those in larger cities. That is, when goods are more differentiated, demand is less sensitive to price, and so remains higher for lower productive producers in smaller cities. Hence, if the elasticity of substitution is lower, the estimation generates a higher correlation between productivity and city size to give firms in larger cities a productivity advantage. Here, I follow Allen and Arkolakis (2014) and set the elasticity of substitution at 9. This implies a correlation of 0.20 (Table 4, Panel B, column 1). As I set the elasticity of substitution at a lower value, I get a larger estimate. As the elasticity goes to infinity, and if we assume no

instruments. Despite the limitations, however, the fact that across all these regressions no evidence for nonlinearity appears is suggestive of their absence.

The regression results presented here are well supported by the literature. Glaeser and Gottlieb (2008) use US Census data and a subset of instruments I use here to search for potential nonlinearities. They find that "the impact of population on productivity seems to be the same for both smaller and larger metropolitan areas." Notowidigdo (2011) finds no evidence of nonlinear wage responses to shocks to US local labor markets. In an influential study, Kline and Moretti (2014a) use historical data of the Tennessee Valley Authority Program to estimate its effects on wages and employment. They, too, search for nonlinear (city-level) agglomeration elasticities using a flexible specification of productivity spillovers. However, they find no evidence for these nonlinearities.[12]

Figure 1: Wages, Wage Inequality, and City Size



(a) Mean hourly wage

(b) Ratio of top to bottom decile of hourly wage distribution

trade costs, then productivity is wage and the estimate is around 0.08 (Table 4, Panel A, column 1) consistent with the range reported in Rosenthal and Strange (2004).

[12] As for other countries, Combes et al. (2012) report supporting evidence for linear (city-level) scale elasticities in France, and De la Roca and Puga (2017) do not report any nonlinear city-level elasticity across their specifications for Spain. Au and Henderson (2006) find inverted U-shape for "net" agglomeration in Chinese cities, but two caveats apply there. First, the inverted U-shape arises due to commuting costs in their setting, second they emphasize migration policy barriers specific to Chinese economy that may effectively keep the city size too small.

	(1)	(2)	(3)
	log hourly wage	log college wage premium	log college employment ratio
log population	0.0794***	0.0233***	0.125***
	(0.00841)	(0.00333)	(0.0245)
Observations	243	243	243
R-squared	0.701	0.288	0.349

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 2: Wages, College Wage Premium, and College Employment Ratio versus City Population

	log hourly wage ratio				
	(1)	(2)	(3)	(4)	(5)
	decile 10 to 1	decile 9 to 2	decile 8 to 3	decile 7 to 4	decile 6 to 5
log population	0.0895***	0.0406***	0.0226***	0.0123***	0.00420***
	(0.0124)	(0.00714)	(0.00433)	(0.00243)	(0.00100)
Observations	243	243	243	243	243
R-squared	0.610	0.427	0.351	0.300	0.253

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 3: Wage Inequality and City Population

	OLS		IV		
	(1)	(2)	(3)	(4)	(5)
Panel A, dependent variable: log wage					
$\log(L)$	0.08***	0.08***	0.13***	0.09***	0.07***
	(0.01)	(0.01)	(0.04)	(0.01)	(0.03)
$\log(L)^2$	0.01	-0.00	-0.01	-0.00	-0.08
	(0.00)	(0.01)	(0.02)	(0.01)	(0.13)
Kleibergen-Paap (p-value)		0.00	0.00	0.01	0.62
Panel B, dependent variable: log productivity					
$\log(L)$	0.20***	0.21***	0.23***	0.22***	0.20***
	(0.01)	(0.02)	(0.04)	(0.01)	(0.02)
$\log(L)^2$	0.00	-0.00	-0.01	-0.01	-0.04
	(0.00)	(0.01)	(0.02)	(0.01)	(0.09)
Kleibergen-Paap (p-value)		0.00	0.00	0.01	0.62
Obs	243	209	55	209	200

Table 4: Wage and Productivity versus City Population

Notes: See the text under Pattern 4 for instruments used across columns. City productivity is obtained by calibrating the discrete version of Allen and Arkolakis (2014) with CMSAs as units of geography. See the text, and footnote 11.

2.3 Discussion: From the Observations to the Model

Putting together Patterns 1.c, 3, and 4 gives an interesting implication for the functional form of agglomeration economies. The subtlety is to distinguish between measures of city-level and worker-level productivity. Define city-level agglomeration elasticity as the elasticity of city productivity with respect to city population, and worker-level agglomeration elasticity as the elasticity of individual worker productivity with respect to the population of city where she works. By construction, city-level agglomeration elasticity reflects both worker-level agglomeration elasticity and city skill composition.

According to Patterns 1.c and 3, larger cities are skill-abundant, and worker-level agglomeration elasticity is larger for higher skilled workers. If worker-level agglomeration elasticity was constant w.r.t. city size, then Patterns 1.c and 3 would imply that city-level agglomeration elasticity was increasing in city size, contradicting Pattern 4.[13] To square with the facts, the worker-level agglomeration elasticity, given the worker skill, must be decreasing in city size.[14]

[13] Note that a model in which worker-level agglomeration has a power law shape, but the power law parameter is greater for a higher skilled worker, implies an increasing city-level agglomeration elasticity.

[14] With the same logic, a production function that allows for many worker types with constant worker-level scale elasticities

Motivated by these observations, I design a model in which worker-level agglomeration elasticity is endogenously increasing in worker skill (given city size), and it is endogenously decreasing in city size (given worker skill). The former allows the model also to accommodate Pattern 1.b (positive relationship between city size and wage inequality). Similar to a large body of literature, Pattern 1.a is simply accounted for by the existence of agglomeration economies. Given Pattern 2, the importance of the full wage distribution to examine wage inequality, I depart from college versus non-college cut, and allow for a continuum of skills.

I report a number of other regressions in Appendix A. Several empirical observations are reported for 1980, including a less strong relationship between skill premium and city size in 1980 (see Fig. 9), a pattern to which I will return in Section 6. In addition, regressions in Table A.4 show that returns from city size are largely accounted for by the impact on hourly wage rather than hours worked. Accordingly, in my model I abstract away from the effect of agglomeration on hours worked and focus on its impact on hourly wages. In addition, all the empirical patterns presented here hold if I replace log wage by residual log wage, as the residual of a regression of log wage against observed worker characteristics including cubic polynomial of years of experience, gender, and race.

For a clear exposition, I first present a single block of the model (which I call the basic model), then I present the full model. I deliberately incorporate the minimal requirements to square with the facts and to address questions stated in the Introduction. The full model can be extended to incorporate other features. I leave some of these extensions to Appendix D. Proofs and mathematical derivations are contained in Appendix C.

3 Basic Model

I present a simple model of production in which returns to scale arise due to local idea exchange. This "basic model" incorporates only one city and one skill level. In Section 4, I present the "full model" that features an arbitrary number of cities and a continuum of skills.

I model the production technology by envisioning the metaphor of a research institution. Every researcher allocates her time to work in her private office and to exchange ideas in seminars, workshops, and face-to-face interactions. The researcher's productivity rises in her private production and social returns from her idea exchange activity. The tradeoff she faces is as follows. The opportunity cost of every minute she allocates to idea exchange is forgone private production; and, the benefit depends on the extent to which others participate in idea exchange. This collective interaction creates knowledge spillovers, making

for each type necessarily overpredicts the city-level scale elasticity. This overprediction of city-level scale elasticity makes spatial policies biased toward benefiting higher skill groups that reside more than proportionately in larger locations. This is a straightforward result as an extension to Proposition 3 in Section 4.4. Therefore, a takeaway from my discussion in this section is encouraging a more careful examination of the commonly-used power law assumption for worker-level scale elasticities in the literature such as Diamond (2016) or the empirical section of Fajgelbaum and Gaubert (2018).

productivity of every individual worker to increase with the scale of production.

3.1 Setup

Consider an economy with one city populated by a measure L of homogeneous workers. Every worker ω is endowed by one unit of time, and allocates her time into two productive activities: (a) idea exchange, and (b) private production. These two activities are inputs to production function of individual ω,

$$y_\omega = \left[(Aa_\omega)^{\frac{\rho-1}{\rho}} + (Bb_\omega)^{\frac{\rho-1}{\rho}}\right]^{\frac{\rho}{\rho-1}} \tag{1}$$

Here, a_ω and b_ω are the amounts of time allocated to idea exchange and private production, respectively, with the resource constraint $a_\omega + b_\omega \leq 1$. Since an individual's endowment of time is normalized to one, y_ω is both individual ω's production as well as her productivity. B is exogenous productivity of labor in private production. A is endogenous knowledge capacity of the city as the potential for benefit from idea exchange activity. $\rho \geq 0$ is the elasticity of substitution between knowledge-based input $A \times a$ and privately-produced input $B \times b$. If $0 \leq \rho < 1$, idea exchange and private production are complements, and otherwise the two are substitutes. The production function collapses to Leontief if $\rho \to 0$, to Cobb-Douglas if $\rho \to 1$, and to a linear production if $\rho \to \infty$.

Knowledge capacity, A, depends on all individuals' participations in idea exchange activity. Specifically, let A aggregate all workers' time allocation to idea exchange,

$$A = \left[\int_{\omega\in\Omega} a_\omega \; d\omega\right]^\gamma, \tag{2}$$

where Ω denote the set of workers in the city, and A is a homogeneous function of degree $\gamma \geq 0$ in arguments $[a_\omega]_{\omega\in\Omega}$. A higher aggregation parameter γ means a greater accumulation of knowledge from the same idea exchange participations. Every individual ω takes A as given, and optimally allocates $1 - a_\omega$ of her time to private production and a_ω to idea exchange to maximize her productivity y_ω.

Henceforth I drop ω since all workers will have the same choice. Given an exogenous private productivity B, *production outcome* is a triple (a, A, y) as a function of city population L. The optimal choice of a is[15]

$$a = \frac{A^{\rho-1}}{A^{\rho-1} + B^{\rho-1}} \equiv I(A) \tag{3}$$

[15] I restrict my analysis to cases of positive ρ where this interior solution is guaranteed. If otherwise $\rho < 0$, there will be a corner solution with optimal a equal one if $A > B$ and zero if $A < B$.

In addition, by aggregation of ideas (2)

$$A = (La)^{\gamma} \equiv K(a) \tag{4}$$

Equation (3) describes **I**dea exchange activity a as a function of knowledge capacity A, and equation (4) describes **K**nowledge capacity A as a function of a, hence two mappings I and K in the space of (a, A). The production outcome is where these two mappings jointly intersect. An outcome (a, A, y) is locally *stable* if there exists an open set around a such that for all a' in that open set, the sequence of $I(K...(I(K(a'))))$ converges to a.

Mapping I implies that the time allocated to idea exchange, a, rises with knowledge capacity A, if idea exchange and private production are substitutes, $\rho > 1$, and falls if they are complements, $0 < \rho < 1$. According to mapping K, $a = L^{-1}A^{1/\gamma}$, implying that a increases with knowledge capacity A, and a larger city size L shifts mapping K downward.

Figures (2)-(3) illustrate a pictorial analysis of the production outcome when γ is set to one. In case of $\rho \in (0, 1)$, idea exchange and private production are complements as reflected in the downward sloping mapping I in Figure (2). The intersection O gives a unique stable outcome (a, A). The figure also shows how a and A endogenously change in response to an increase in population of workers L. An increase in L lowers the slope of mapping K, represented by the new mapping K'. The new outcome O' is associated with a smaller a and a larger A. Figure 3–Panel I illustrates the production outcome if $\rho \in (1, 2)$ where knowledge and private production are mildly substitutes. Here again there is a unique stable outcome (a, A) shown by O. Figure 3–Panel II shows the case of $\rho \in (2, \infty)$ where knowledge and private production are highly substitutes. In this case $(a, A) = (0, 0)$ is a stable outcome, and if population is large enough there is another stable outcome shown by O_2.

3.2 Agglomeration Elasticity

I first characterize the effect of city size L on the agglomeration elasticity. Then, I explain how changes in private productivity B affect this elasticity.

3.2.1 Agglomeration Elasticity and City Size

The model generates non-constant elasticity of productivity with respect to population. Let α denote *agglomeration elasticity* as change in log productivity with respect to change in log population.[16] By fully

[16] Here, because labor is homogeneous, city-level and worker-level agglomeration elasticities are the same objects. I make the distinction in the full model presented in the next section.

differentiating the system of equations 1, 3, and 4, the agglomeration elasticity is given by

$$\alpha \equiv \frac{d \ln y(L)}{d \ln L} = \frac{a}{1/\gamma + (1-\rho)(1-a)} \tag{5}$$

According to equation (5), agglomeration elasticity, α, is always positive along stable outcomes, but it is typically *non-constant* in L. The following proposition fully characterizes production outcome, and whether agglomeration elasticity increases, remains unchanged, or decreases in population.

Proposition 1. *The production outcome (a, A, y), and the agglomeration elasticity α as functions of population L are characterized as follows.*

- *Case 1: "Complementarity", $\rho \in (0, 1)$. There exists a unique stable outcome. If L rises, then a falls; A, $a \times A$, and y rise; and α is decreasing in L.*
- *Case 2: "Cobb-Douglas", $\rho = 1$. There exists a unique stable outcome. If L rises, then a remains constant but A and y rise; and $\alpha = \frac{\gamma}{2}$, is a constant.*
- *Case 3: "Mild Substitutability", $\rho \in (1, 1 + 1/\gamma)$. There are two outcomes. Outcome $(a, A) = (0, 0)$ is not stable. The other outcome with positive values of a and A is stable. In the stable outcome, if L rises, then a, A, and y increase; and α is increasing in L.*
- *Case 4: "In-Between Substitutability", $\rho = 1 + 1/\gamma$. There exists no nonzero outcome if $L < B^{1/\gamma}$. Otherwise, there exists a unique stable outcome in which by an increase in L, both a and A rise; and $\alpha = \gamma$, is a constant.*
- *Case 5: "High Substitutability", $\rho \in (1 + 1/\gamma, \infty)$. There exists no nonzero outcome if $L < (\rho - 1)\Big(\rho - 1 - 1/\gamma\Big)^{\frac{1-\gamma(\rho-1)}{\gamma(\rho-1)}} B^{1/\gamma}$. Otherwise there are three outcomes of which two of them are stable. The trivial stable outcome is $(a, A) = (0, 0)$. In the nontrivial stable outcome with positive values of a and A, if L rises, a, A, and y increase; and α is decreasing in L.*

When idea-exchange and private production are complements, i.e. $\rho \in (0, 1)$, an increase in population causes every individual worker to allocate less time to idea exchange activity but the overall knowledge capacity expands. Since the fall in idea exchange activity is less proportional than the increase in knowledge capacity, knowledge-based input available to each individual, $a \times A$, also increases. The reduction in a with the expansion in $a \times A$ causes an increase in private production, $(1-a)B$. These responses are implied by the complementary under which private production should keep up with the increase in knowledge-based input. Because an individual's resources are limited, any further increase in private production requires further decrease in idea exchange activity. The decrease in per capita time allocated to idea-exchange activity dampens the effect of knowledge spillovers on productivity. As a result, the elasticity of productivity with respect to population, α, decreases in L.

In case $\rho \in (1 + 1/\gamma, \infty)$, private production and idea-exchange activities are highly substitutable. The rise in idea-exchange activity substitutes private production, so much so, that by creating knowledge spillovers private production largely shrinks. For exactly the opposite reason of the complementary case, the agglomeration elasticity, α, decreases in L. The only case where α increases in L is when private production and idea-exchange activities are only mildly substitutes, i.e. $\rho \in (1, 1+1/\gamma)$. The model collapses to constant agglomeration elasticities if $\rho = 1$ or $1 + 1/\gamma$. Since at $\rho = 1 + 1/\gamma$ the nonzero stable outcome requires a lower bound on population, it is only the case of $\rho = 1$ where the model nests the power law specification for agglomeration.[17] [18]

Discussion. To examine the generalizability of the particular functional form adopted here, let equation (2) be replaced by $A = f(a, L)$, where $f'_a \equiv \partial f/\partial a > 0$ and $f'_L \equiv \partial f/\partial L > 0$. The agglomeration elasticity is then given by

$$\alpha = \frac{a(Lf'_L/f)}{1 + (af'_a/f)(1-\rho)(1-a)}$$

Consider the case where private and social returns are complements, i.e. $\rho \in (0, 1)$. This case is of my particular interest as detailed in Section 3.2.2. In this case, if (i) Lf'_L/f is a nonzero constant or increasing in L, and (ii) af'_L/f is a nonzero constant or decreasing in a, then α is decreasing in L. These two are sufficient conditions. For a given functional form that does not satisfy these sufficient conditions, we can still use the above equation to directly check whether α is decreasing in L.

The functional form used here delivers an α that decreases in L at a wide range of curvatures depending on the aggregation parameter γ. This flexibility is one advantage of the chosen functional form, which together with its tractability proves helpful for bringing the model to data.[19,20]

[17] The other case, $\rho = 1 + 1/\gamma$ and more generally the case of $\rho \geq 1 + 1/\gamma$ is of independent interest for a model in which a range of locations (possibly on a continuum of space) can have zero population as a stable outcome, and other locations can have positive population as an alternative stable outcome.

[18] Although the model does not admit a closed-form solution, it is extremely straightforward to compute the production outcome numerically, using the following updating rule: Start with a guess of A, then use equation (3) to calculate a, then use equation (4) to update A, and similarly iterate over A until convergence. The stability of production outcome ensures the convergence . Results from a numerical simulation are shown by Figure A.1 to illustrate the model behavior more clearly.

[19] For instance, $f(a, L) = 1 - \exp(-\kappa a L)$, used in Davis and Dingel (2018), also delivers a decreasing α. But with this functional form, the resulting α falls too sharply with respect to L.

[20] While I use a functional form to maintain tractability for the empirical analysis, an alternative microfoundation of learning opportunities might be built on yet another particular mechanism. For example, a literature on economic growth microfounds external effects through a process of searching to match with a new idea to improve one's own productivity (Lucas and Moll, 2014; Perla and Tonetti, 2014). In contrast to these mechanisms, the way a producer learns from others might not necessarily be through the matching of one producer with one idea. It could rather be the learning of one producer simultaneously from many ideas available in a local environment, which is in fact, more in line with the insight in Marshall (1920).

3.2.2 Agglomeration Elasticity and Private Productivity

The agglomeration elasticity α is increasing in idea exchange participation a (equation 5). In turn, a is increasing in private productivity B only when $\rho \in (0, 1)$ (equation 3). Hence, holding population L fixed, agglomeration elasticity α is increasing in private productivity B only if $\rho \in (0, 1)$. In this case, the production outcome simultaneously satisfies two desirable properties:

1. The agglomeration elasticity falls with city population; and,

2. ... rises with workers' private productivity.

Figure 4 depicts agglomeration elasticity in the basic model for $\rho \in (0, 1)$ and two values of B. The elasticity falls along each curve as population L increases, and it rises by an upward shift across the curves as private productivity B increases. If a higher L comes with a higher B, then a constant city-level agglomeration elasticity means a fall in worker-level agglomeration elasticity (along the curve) together with an increase in the strength of agglomeration due to the increase in B (upward shift of the curve). Hence, the model delivers the right predictions only when knowledge and private production are complements, i.e. $\rho \in (0, 1)$.

In the next section, I extend the setup by incorporating many cities and many skills. I maintain the relevant range of $\rho \in (0, 1)$ for the complementarity between knowledge and private production. In this full model, a higher skilled worker is endowed by a higher private productivity. Under the complementarity, a higher skilled worker finds it optimal to allocate more time to idea exchange activity, hence she contributes more to local knowledge accumulation. Because returns to scale are larger for a worker when private productivity is greater, higher skilled workers on average sort to larger cities, making larger cities more abundant in skill. This positive feedback endogenously creates a supermodular relationship between skill and city size.

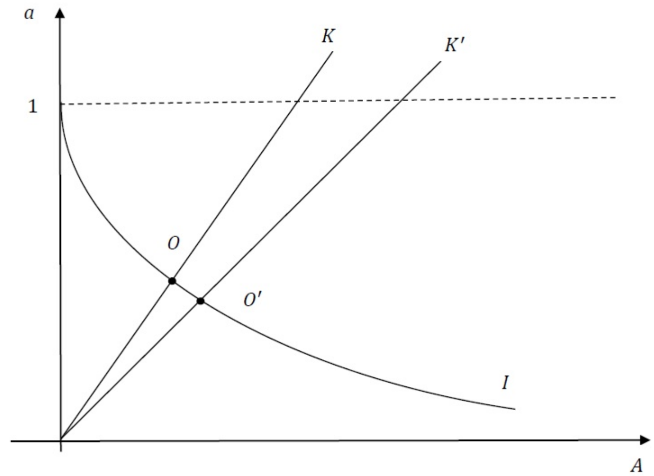


Figure 2: Production outcome when knowledge and private production are complements.

Notes. Here, γ is set to one. The effect of an increase in population L on a and A is illustrated by moving from point O to point O'.

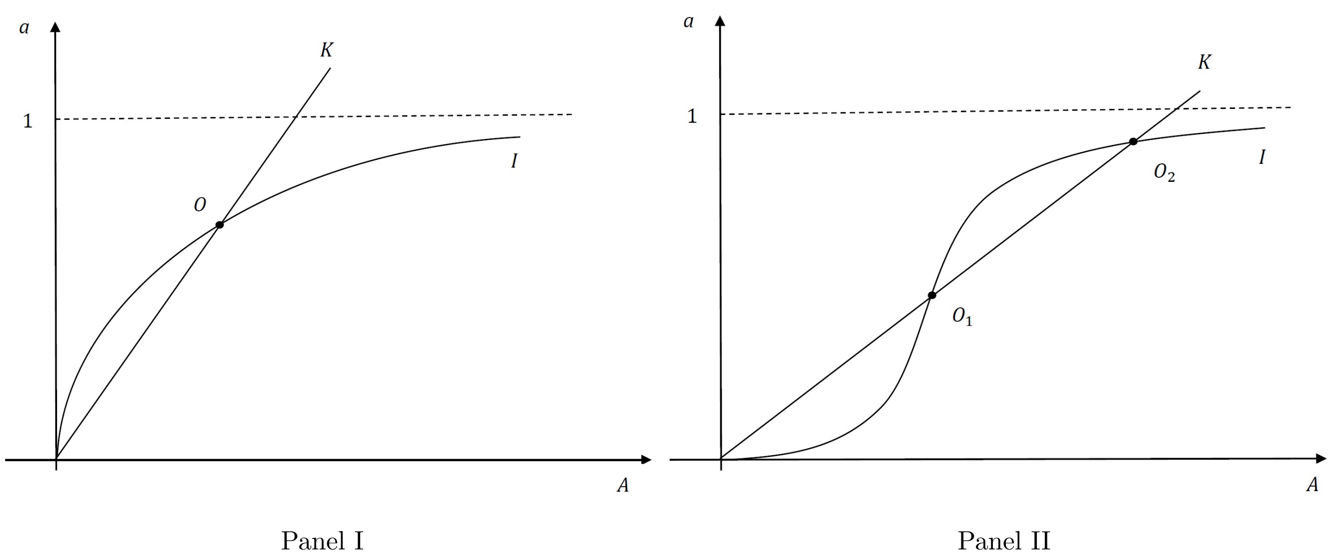
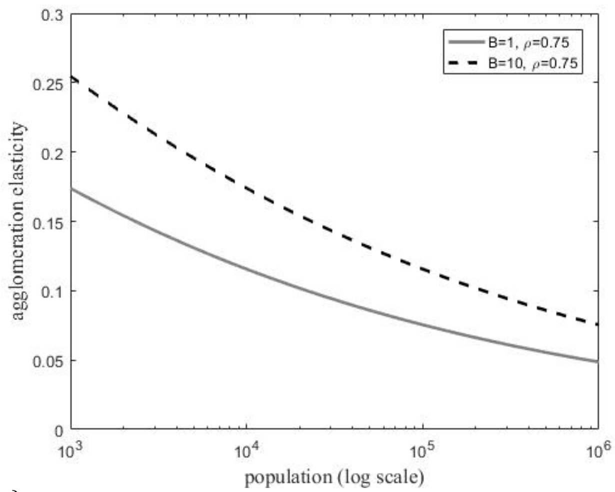


Figure 3: Production outcome when knowledge and private production are substitutes.

Notes. Here, γ is set to one. Left: $\rho \in (1, 2)$ where knowledge and private production are modestly substitutable. Right: $\rho \in (2, +\infty)$ where knowledge and private production are highly substitutable. Here O_1 is unstable and O_2 is stable.

Figure 4: Private productivity and agglomeration elasticity



Notes. $\rho = 0.75$, $\gamma = 1$, $B = \{1, 10\}$

4 Full Model

Consider an economy consisting of I cities, a continuum of skills, and perfectly competitive markets.

4.1 Labor Demand

The source of worker heterogeneity is skill z, reflected in private productivity, $B(z)$, to be an increasing function of z. Let $V(z)$ be exogenous national supply of skill z, $L(i,z)$ be endogenous population of skill z in location i, and $L(i)$ be endogenous total population in location i. Let $v(i,z) \equiv \frac{L(i,z)}{V(z)}$ denote distribution of workers within skill z across locations, and $\ell(i,z) \equiv \frac{L(i,z)}{L(i)}$ denote distribution of skill within location i with support $\Omega(i)$. Knowledge and private production are complements, i.e. $\rho \in (0,1)$. A worker of skill z in location i allocates her time to idea exchange and private production to maximize her productivity,

$$y(i,z) = \left[\Big(A(i)a(i,z) \Big)^{\frac{\rho-1}{\rho}} + \Big(B(z)(1-a(i,z)) \Big)^{\frac{\rho-1}{\rho}} \right]^{\frac{\rho}{\rho-1}}$$

by taking knowledge capacity in location i as given,

$$A(i) = \left[\int_{z \in \Omega(i)} L(i,z)a(i,z)\, dz \right]^{\gamma} \tag{6}$$

Optimal time allocated to idea exchange for an individual of skill z in location i is

$$a(i,z) = \frac{A(i)^{\rho-1}}{A(i)^{\rho-1} + B(z)^{\rho-1}} \tag{7}$$

which then implies that

$$y(i,z) = \left[A(i)^{\rho-1} + B(z)^{\rho-1}\right]^{\frac{1}{\rho-1}} \tag{8}$$

Holding location i fixed, since knowledge complements individual capability, a higher skilled worker allocates more resources to idea exchange activity. That is, $a(i,z)$ is increasing in z. In addition, the elasticity of productivity with respect to knowledge capacity is greater for a higher skill.[21] That is, the productivity benefit from knowledge spillovers is greater for a higher skill.

Each location has a measure one of homogeneous firms with the following production technology,

$$Q(i) = \bar{y}(i) \int_z y(i,z) L(i,z) dz \tag{9}$$

Here, $\bar{y}(i)$ is exogenous factor productivity of location i. Let $w(i,z)$ denote the wage paid to skill z in location i. Since profits can not be positive, profit maximization implies

$$w(i,z) \quad \geq \quad \bar{y}(i) y(i,z) \quad \text{with equality if } L(i,z) > 0 \tag{10}$$

Each location produces a final good which is perfectly substitutable with final goods produced by other locations. This final good is the numeraire. There are no trade costs.

4.2 Labor Supply

Individual ω with skill z receives utility from wages, amenities, and her unobserved location preference shock, and chooses where to reside,

$$\max_i \ w(i,z) u(i) \zeta_\omega(i,z)$$

Here, $u(i)$ is amenity value in location i, specified as

$$u(i) = \bar{u}(i) L(i)^\beta \tag{11}$$

[21] The elasticity of y with respect to A is $\frac{\partial \ln y(i,z)}{\partial \ln A(i)} = \left(\frac{y(i,z)}{A(i)}\right)^{1-\rho} = a(i,z)$, which is increasing in z when knowledge and private production are complements, i.e. $\rho \in (0,1)$.

where $\bar{u}(i)$ is exogenous amenity shifter of location i, and β governs the degree of externalities in local amenities. When $\beta < 0$, then these externalities reflect congestion forces. The unobserved location preference shock, ζ, is independent across workers and locations, and follows a Fréchet distribution, $P(\zeta \leq x) = \exp(-x^{-\theta})$. Here, θ governs the dispersion of location preference shocks. A larger θ implies a smaller variance of these shocks, and if $\theta \to \infty$, the model collapses to perfect labor mobility.

The discrete choices of workers imply the expected welfare of workers of skill z, denoted by $W(z)$,

$$W(z) = \bar{\gamma}\left[\sum_j \Big(w(j,z)u(j)\Big)^{\theta}\right]^{\frac{1}{\theta}} \tag{12}$$

where $\bar{\gamma} \equiv \Gamma\Big(1+\frac{1}{\theta}\Big)$. Inside $W(.)$, agglomeration benefits are implicit in wages w, and congestion disutilities are captured by amenities u. Holding skill z fixed, share of workers who choose location i is given by

$$v(i,z) \equiv \frac{L(i,z)}{V(z)} = \frac{\Big(w(i,z)\bar{u}(i)L(i)^{\beta}\Big)^{\theta}}{\sum_j \Big(w(j,z)\bar{u}(j)L(j)^{\beta}\Big)^{\theta}} \tag{13}$$

Workers tend to reside in locations where wages are higher, amenity fundamentals are better, and congestion disutilities are lower, with the extent of the relationship governed by θ. In addition, the model has the following desirable property: Mobility rate is higher for higher skilled workers, because they are endogenously more sensitive in their location decisions to external shocks. This is because higher skill workers benefit more from agglomeration.

As long as θ is finite, $L(i,z) > 0$, and supply of skill z relative to z_0, holding location i fixed, is[22]

$$\frac{L(i,z)}{L(i,z_0)} = \Big(\frac{V(z)}{V(z_0)}\Big)\Big(\frac{W(z)}{W(z_0)}\Big)^{-\theta}\Big(\frac{w(i,z)}{w(i,z_0)}\Big)^{\theta} \tag{14}$$

4.3 Spatial Equilibrium

Take workers with skill z and hold z_0 as the reference skill. Combining relative labor demand (equation 10) and relative labor supply (equation 14) delivers skill wage premium, $\widetilde{w}(i,z)$, and skill employment ratio, $\widetilde{L}(i,z)$, each as a function of relative productivities,

$$\widetilde{w}(i,z) \equiv \frac{w(i,z)}{w(i,z_0)} = \frac{y(i,z)}{y(i,z_0)} \tag{15}$$

$$\widetilde{L}(i,z) \equiv \frac{L(i,z)}{L(i,z_0)} = \Big(\frac{V(z)}{V(z_0)}\Big)\Big(\frac{W(z)}{W(z_0)}\Big)^{-\theta}\Big(\frac{y(i,z)}{y(i,z_0)}\Big)^{\theta} \tag{16}$$

[22] See Appendix C for derivation of this equation.

The model generates skill premiums that vary endogenously across cities. The intuition is similar to that in Farrokhi and Jinkins (2017). Since returns from city size are larger to higher skilled workers, the productivity premium, $y(i,z)/y(i,z_0)$, is larger if z is greater. Since workers have heterogeneous preference locations, an increase in the productivity premium will not be fully arbitraged away by the increase in supply of skill z to city i. Accordingly, all else being equal, the model generates a higher dispersion in skill premiums across cities, the greater the cross-skill differences in returns to scale and the more dispersed the location preference shocks. The former is shut down if $\rho = 1$, and the latter vanishes if $\theta \to \infty$.

In addition, total sales of every location equals total payments to workers there,

$$\bar{y}(i) \int_z y(i,z) L(i,z) dz = \int_z w(i,z) L(i,z) dz \tag{17}$$

Equations 15-16 describe market clearing conditions in relative terms, and equation 17 does so in levels.

Definition. An equilibrium consists of wages $w(i,z)$ and populations $L(i,z)$ such that equations (7)-(17) hold: (i) workers optimally allocate their labor supply to locations, and their time to idea exchange and private production, (ii) firms optimally choose labor demand, (iii) labor markets clear for all skills and locations, and (iv) the labor allocation is feasible.

Holding z_0 as the reference skill, skill wage premium and skill employment ratio of skill z are given by equations (15)-(16). The following proposition informs the equilibrium assignment of skills to cities.

Proposition 2.

(i) Skill wage premium and skill employment ratio are higher in location i than j if and only if $A(i) > A(j)$.

(ii) If $L(i) > L(j)$ and $L(i,z)/L(i) = L(j,z)/L(j)$. Then $A(i) > A(j)$.

(iii) If $L(i) = L(j)$ and $L(i,z)/L(j,z)$ is increasing in z. Then $A(i) > A(j)$.

In a city with a larger knowledge pool, there are relatively more skilled workers paid at relatively higher premiums. Knowledge pool, in turn, is larger in city i than j, if population of i is greater than j given skill distribution; or if, controlling for population, city i is more skill abundant than j (i.e. skill distribution of i first order statistically dominates that of j).[23]

[23] Here, a more skill-abundant city offers a greater knowledge pool since higher skilled workers allocate more time to idea exchange. In addition to this horizontal dimension, it is straightforward to allow for a vertical dimension, e.g. in equation (6) replace $a(i,z)$ by $a(i,z)z^{\varsigma}$ with ς governing the vertical contribution of skill to local knowledge pool. However, for identification we will then require to observe workers' resource allocation to learning. To my knowledge, available surveys such as the American Time Use Survey do not incorporate such variables. Upon such data availability, a more extended version of the model could be estimated. I leave the discussion about which variations in the available data identify the model for Section 5.3.

In addition, I present several other extensions to the model by allowing for tradeable goods and nontradeable housing, costly trade between cities, finite substitution across tradeable goods, and finite substitution across workers of different skills. I also show conditions for existence and uniqueness for a simplified version of the model. See Appendix D.

4.4 Place-Based Policies

Place-based policies can be defined under a general formulation, even if government is not explicitly modeled. Let $\bar{x}(i)$ be the product of exogenous productivity and amenity shifters in location i,

$$\bar{x}(i) \equiv \bar{y}(i)\bar{u}(i) \tag{18}$$

I define a place-based policy, then give the intuition behind it:

Definition. *A Place-Based Policy consists of changes to amenities and productivities shifters, $[d\bar{x}(i)/\bar{x}(i)\ \forall i]$, and a constant Δ, denoted by PBP$(d\bar{x}(i)/\bar{x}(i)\ \forall i, \Delta)$, such that*

$$\sum_{i=1}^{I} \frac{L(i)}{\bar{L}} \frac{d\bar{x}_i}{\bar{x}_i} = \Delta \tag{19}$$

Proposition 3 makes it clear below that a place-based policy may achieve the same outcome by improving either productivities $\bar{y}$ or amenities $\bar{u}$. For this reason, I define a PBP with respect to $\bar{y}\bar{u}$. Here, Δ reflects government resources used to fund a project. In a special case in which $\Delta = 0$, weighted average of positive changes to the infrastructure of a set of locations and negative changes to the infrastructure of other locations is zero (with weights as population shares).[24]

The following proposition is a new formulation that summarizes a well-known result in the literature (Glaeser and Gottlieb, 2008; Kline and Moretti, 2014b).

Proposition 3. *Suppose there is one skill type, and the agglomeration elasticity is constant ($\rho = 1$). Then, the effect of a place-based policy, PBP$(d\bar{x}(i)/\bar{x}(i)\ \forall i, \Delta)$, on expected welfare is:*

$$\frac{dW}{W} = \Delta$$

The direct effect of a place-based policy is that resources spent to improve any set of local infrastructures will increase national welfare. The subtlety is in indirect effects which are neutral if gains in receiving areas

[24] Here, a practical interpretation is that subsidizing one location is funded by taxes imposed on other locations. While we can explicitly add a tax scheme to the model, I found it not necessary to formulate a place-based policy. For quantitative analysis of the effects of spatial taxes on labor allocation and productivity, see Fajgelbaum et al. (2016) and Ossa (2017).

are offset by losses elsewhere. The proposition states that these gains and losses exactly cancel out each other if labor is homogeneous and the agglomeration elasticity is constant.[25]

For example, consider two distinct policies: PBP_1 characterized by $d\bar{x}(i)/\bar{x}(i) = \delta_1(i)$ and $\Delta_1 = \Delta$, compared to PBP_2 characterized by $d\bar{x}(i)/\bar{x}(i) = \delta_2(i)$ and $\Delta_2 = \Delta$. PBP_1 improves the infrastructure in small cities while PBP_2 does so in large cities, but the two policies are associated with the same Δ. Then, under no skill heterogeneity and constant agglomeration elasticity, the two policies PBP_1 and PBP_2 will have exactly the same effect on aggregate welfare. In other words, it does not matter where to subsidize.

This proposition summarizes an important benchmark in the literature that points to the neutrality of place-based policies. My analysis goes beyond this benchmark, as my model features skill heterogeneity and flexible agglomeration elasticities. In Section 6, I use my model to show that even if place-based policies generate no aggregate indirect effect, they have important consequences for welfare inequality.

4.5 Agglomeration Elasticities

For a worker with skill z in city i, worker-level agglomeration elasticity, denoted by $\alpha(i,z)$, is the change to that worker's productivity with respect to a change to population of city i, given the distribution of skill in the city $\ell(i,z) \equiv L(i,z)/L(i)$,

$$\alpha(i,z) \equiv \left[\frac{d\ln y(i,z)}{d\ln L(i)}\bigg|\ \ell(i,z)\ \forall z\right] = \frac{a(i,z)}{1/\gamma + (1-\rho)(1-a_{avg}(i))} \tag{20}$$

where $a_{avg}(i) \equiv \frac{\int_z \ell(i,z)a(i,z)^2dz}{\int_z \ell(i,z)a(i,z)dz}$ is a weighted of average $a(i,z)$ in city i, with weights equal $\ell(i,z)a(i,z)$. With total production of city i, $Q(i)$, given by equation (9), let city productivity be

$$q(i) \equiv Q(i)/L(i).$$

We denote "city-level agglomeration elasticity conditional on within-city skill distribution" by $\alpha(i)$, as the change to log productivity of city i, $q(i)$, with respect to a change to log population of city i, $L(i)$, given the distribution of skill in the city $\ell(i,z) \equiv L(i,z)/L(i)$,

$$\alpha(i) \equiv \left[\frac{d\ln q(i)}{d\ln L(i)}\bigg|\ \ell(i,z)\ \forall z\right] = \frac{\int_z y(i,z)\ell(i,z)\alpha(i,z)dz}{\int_z y(i,z)\ell(i,z)dz} \tag{21}$$

$\alpha(i)$ is the weighted average of $\alpha(i,z)$ with weights being equal to $y(i,z)\ell(i,z)$. See Appendix C for derivations of (20)-(21). Note that both $\alpha(i)$ and $\alpha(i,z)$ are conditional on holding fixed the within-city skill

[25] It should be understood that this result holds locally. My numerical exercises show that for large enough changes in $\bar{x}$, this result is not a good approximation anymore.

distribution. In the spatial equilibrium, however, a change in $L(i)$ comes with changes to the distribution of $\ell(i,z)$ $\forall z$ in city i. We denote the "unconditional city-level agglomeration elasticity" by $\alpha^{unc}(i)$,

$$\alpha^{unc}(i) \equiv \frac{d \ln q(i)}{d \ln L(i)} \tag{22}$$

Since the equilibrium assignment of skills to cities does not admit an analytical solution, $\alpha^{unc}(i)$ does not have a closed-form expression. However, as I solve the model numerically, I will be able to compute $\alpha^{unc}(i)$.

In the empirically relevant case, an increase in city population $L(i)$ shifts the within-city skill distribution $\ell(i,z)$ to the right. As a result, $\alpha^{unc}(i)$ is larger than $\alpha(i)$. The gap $\alpha^{unc}(i) - \alpha(i)$ is the extent to which the skill increase in a growing city contributes to the unconditional scale elasticity.

4.6 Recovering Location Characteristics

Let Θ summarizes model parameters, consisting of $(\rho, \gamma, \theta, \beta)$ and parameters that specify $V(z)$ and $B(z)$. Given Θ, the model maps population and income of cities to fundamentals of productivity and amenity. The following calibration problem which recovers productivity and amenity shifters $\bar{y}(\Theta) = [\bar{y}(i;\Theta)]_{i=1}^{I}$ and $\bar{u} = [\bar{u}(i;\Theta)]_{i=1}^{I}$ such that predicted population and income of cities are equal to those in the data,

$$\begin{aligned} L(i; \bar{y}(\Theta), \bar{u}(\Theta), \Theta) &= L^{data}(i) \;\; \forall i = 1, ..., I \\ Q(i; \bar{y}(\Theta), \bar{u}(\Theta), \Theta) &= Q^{data}(i) \;\; \forall i = 1, ..., I \end{aligned}$$

The calibration algorithm takes $L(i)$ and $Q(i)$ as given, and solves for productivity shifters $\bar{y}(i)$ and amenity shifters $\bar{u}(i)$ such that all equilibrium conditions hold. I present this algorithm in details in Appendix B.2.

4.7 Solution Algorithm

I extend the solution algorithm for the basic model (see footnote 18) by incorporating equilibrium relations that describe distribution of skill across and within locations. Specifically, the algorithm iterates over $A(i)$ given a current guess for population distribution $L(i,z)$ and expected welfare $W(z)$, and within each iteration it uses information on skill participation in idea exchange, and conditions that describe equilibrium in relative terms and levels. The solution algorithm takes national skill distribution $V(z)$, private return to skill $B(z)$, productivity shifter $\bar{y}(i)$, and amenity shifter $\bar{u}(i)$ as exogenously given, and solves for wages $w(i,z)$ and populations $L(i,z)$. Appendix B.1 presents the solution algorithm in details.

5 Estimation

I complete the model specification, then estimate model parameters using simulated method of moments.

5.1 Specification

National supply of skill, $V(z)$ for $z \in [0, \infty)$, has a Fréchet distribution with shape parameter μ,

$$\frac{1}{\bar{L}} \int_0^z V(z')dz' = -\exp(-z^{-\mu}) \tag{23}$$

I explain in Section 5.3 why I choose this specification compared to other alternatives such as Pareto or log normal distribution. Here, $V(z)$ reflects how abundant skill z is at the national level. A lower μ means a larger dispersion of national skill distribution.

Private returns to skill, $B(z)$, is specified as

$$B(z) = \bar{B}z^{\varphi} \tag{24}$$

where $\bar{B}, \varphi > 0$. $B(z)$ represents the private productivity of skill z. An increase in φ reflects a skill-biased change in technology.

5.2 Estimation Procedure

The model is isomorphic to the one in which preferences are Cobb-Douglas combination of final goods and housing, with $\beta = -$housing share/$(1 -$ housing share). I follow Albouy (2008) and set housing share at 0.32, which implies $\beta = -0.47$. Following several available estimates, I set dispersion parameter of location preferences, θ, at 2. Suárez Serrato and Zidar (2016) find a range between $\frac{1}{0.65}$ and $\frac{1}{0.84}$ in their baseline specification. Their estimates are somewhat larger for alternative specifications (See Panel B of their Table 6).[26] Diamond (2016) finds a similar range. Hsieh et al. (2016), Galle et al. (2017), Adao et al. (2017) and Burstein et al. (2017) also find similar values for the dispersion across regions, occupations, or industries. These estimates, which typically lie around 1.5 and 2.5, reflect a relatively large variance of location preferences, implying a relatively low elasticity of labor supply to a location. In my benchmark specification, I set $\varphi = 1$ which serves as a normalization as detailed below. Let $\Theta = [\rho, \gamma, \mu, \bar{B}]$ denote the vector of parameters to be estimated.

Moments. I estimate model parameters using simulated method of moments, subject to the mapping between fundamentals of productivity and amenity to observed population and income of cities. I match

[26] This is based on the equivalence of standard deviation of log of draws between the extreme value type I in their work and of type II here.

model predictions to six sets of moments. The first four sets of moments, denoted by m_1, ..., m_4, consist of mean wages in the 1st, 2nd, 9th, and 10th deciles of wage distribution within every city, summing up to 972 moments (4 deciles $\times$ 243 cities). The fifth moment, m_5, is the ratio of maximum to minimum city-level agglomeration elasticity across all cities, $\widetilde{\alpha} \equiv \max \alpha(i)/\min \alpha(i)$, where $\alpha(i)$ is given by equation (21). This moment does not set the magnitude of $\alpha(i)$. It only restricts $\alpha(i)$ to be constant across cities $i = 1, ..., I$, meaning that I set $m_5^{data} = \widetilde{\alpha}^{data} = 1$. The last moment, m^6, puts an orthogonality condition between the deviations of exogenous productivity shifters from their mean to deviations of variable Z from its mean,

$$\mathbb{E}\Big[\Big(\ln \bar{y} - \mathbb{E}[\ln \bar{y}]\Big)\Big(\ln Z - \mathbb{E}[\ln Z]\Big)\Big] = 0$$

Here, Z is housing supply elasticity, taken from Saiz (2010). The predicted moment is

$$\sum_{i=1}^{I}\Big((\ln \bar{y}(i;\Theta) - \sum_{i=1}^{I} \ln \bar{y}(i;\Theta)/I)(Z(i) - \sum_{i=1}^{I} Z(i)/I)\Big),$$

with $m_6^{data} = 0$. This orthogonality condition is imposed based on the presumption that exogenous productivity shifters, as labor demand shocks, do not vary systematically with housing supply variables that directly affect labor supply.

Let $\mathbf{m}$ stack all these moments,

$$g(\Theta) = \mathbf{m}(\Theta) - \mathbf{m}^{data} = \begin{bmatrix} m_1(\Theta) - m_1^{data} \\ \cdots \\ m_6(\Theta) - m_6^{data} \end{bmatrix}$$

The moment condition can be formulated as $\mathbb{E}[g(\Theta_0)] = 0$, where Θ_0 is the true parameter vector. The parameter estimate $\widehat{\Theta}$ is the solution to the following constrained optimization problem,

$$\begin{aligned} \widehat{\Theta} \quad &= \quad \arg\min_{\Theta} g^T(\Theta)\Lambda g(\Theta) \\ & \quad\quad s.t. \quad \bar{y} = \bar{y}(\Theta), \quad \bar{u} = \bar{u}(\Theta) \end{aligned}$$

where $\bar{y}(\Theta)$ and $\bar{u}(\Theta)$ are the vectors of recovered productivity and amenity shifters described in the calibration problem, and Λ is a weighting matrix. I follow the tradition in the urban literature by weighting wage moments with city population share (normalized by observed mean wage of each decile). The weight on m^5 and m^6 are set to one, and the off-diagonal weights are set to zero.

5.3 Discussion and Identification

The residuals in the estimation can be interpreted in two ways. First, suppose in equation (9) we replace $y(i,z)$ by $\bar{y}(i,z)y(i,z)$ where $\bar{y}(i,z)$ is an exogenous city-skill-specific shifter. Suppose further that by the timing of decisions, workers choose where to reside before they learn about $\bar{y}(i,z)$. Then, $\bar{y}(i,z)$ will be the residual of wage $w(i,z)$ after controlling for endogenous $y(i,z)$ and city-specific component $\bar{y}(i)$. The model then perfectly matches any sets of wage moments at appropriate values of $\bar{y}(i,z)$. The other interpretation is that the residuals are measurement errors of wages, particularly because wage records in the Census data are self-reported and subject to mismeasurement. Then, true wage $w^{\text{true}}(i,z) = w(i,z)\epsilon_w(i,z)$ with $w(i,z)$ as observed wage and $\epsilon_w(i,z)$ as error term. Taking $\bar{y}(i)$, $\bar{y}(i,z)$ and $\epsilon_w(i,z)$ into account,

$$\ln w(i,z) = \ln \bar{y}(i) + \ln y(i,z) + \underbrace{\ln \bar{y}(i,z) - \ln \epsilon_w(i,z)}_{\text{error term}}$$

Accordingly, the wage moments, m^1-m^4, impose orthogonality conditions on these error terms, interpreted as city-skill-specific productivity shocks or measurement errors.[27]

While the four parameters $(\rho, \gamma, \mu, \bar{B})$ are identified jointly, each is associated in a more intuitive way with one of the moments. As discussed in Section 3 the model requires ρ to be between zero and one to generate worker-level agglomeration elasticities that increase in skill and decrease in city size. Consider Figure 1-(b) that plots the ratio of top to bottom decile of within-city wage distribution against city population (in logs). In the data, this ratio rises in population at a slope around 0.09. This slope, when generated by the model, is smaller if ρ is higher in its admissible range. The model generates a slope that is too large compared to 0.09 as ρ approaches zero, and generates a flat plot as ρ approaches one. This slope, reflecting the skill-bias of wages in larger cities, is key to the identification of ρ.[28]

It is important for the national distribution of skill to have a fat right tail. For example, in an alternative specification, a log normal distribution performs less satisfactorily to generate sufficiently large skill premiums. Since Fréchet distribution belongs to a family of distributions with a thicker right tail than log normal, it performs better in generating sufficiently large skill premiums. In addition, although a Pareto distribution has a right tail with the same properties as Fréchet, it preforms less satisfactorily in capturing wages in the

[27] In other words, my specification implies that the comparative advantage of higher skilled workers to work in larger cities arises endogenously. In this sense, I abstract away from heterogeneity across sectors, occupations, tasks, or equipments (such as computer intensity), e.g. higher skilled workers may have a comparative advantage in sectors, occupations, tasks, or equipments which are concentrated in bigger cities. For such a setting with sectors in an urban setting, see Davis and Dingel (2017), and for occupations and equipments in an international setting, see Burstein et al. (2018).

[28] As explained earlier under equations (15)-(16), a lower dispersion of location preferences, θ, can also contribute to higher predicted skill premiums in larger cities. When I estimate θ together with other parameters, I obtain an estimate of θ similar to those in the literature, but the identification between ρ and θ becomes less transparent. Since several studies have estimated θ and found a tight range around 2, I have taken θ from the literature. Upon request, I am happy to report my estimates when I fix θ at other values, or when I include it in the estimation.

bottom decile of wage distribution. Fréchet distribution has the advantage of covering close-to-zero draws which improves the model fit to wages in lower deciles. Again consider Figure 1-(b). When Fréchet shape parameter μ is too large, the model generates lower premiums than the observed ones (not only the slope in the plot but also the intercept). When μ is too small, the model does the reverse. The identification of μ relies on the interaction of skill distribution with other mechanisms in the model to match the observed *overall* magnitude of skill premiums.

Parameter φ in private returns to skill, $B(z) = \bar{B}z^{\varphi}$, is not identified because μ is already responsible for the dispersion in private returns across the national skill distribution. Since we observe wages rather than private productivities, a high skill premium within a city could reflect a high dispersion in the skill distribution given private returns, or a high dispersion in private returns given the skill distribution. For this reason I normalize $\varphi = 1$. In the next section, however, I re-estimate the model by letting φ change, given the same skill distribution to match skill premiums in years earlier than 2000. Such a counterfactual change in private returns reflects a skill-biased change in technology.

The model performs more satisfactorily by allowing parameter $\bar{B}$ to be estimated. I have estimated the model alternatively by normalizing $\bar{B}$ to a constant such as unity. In this alternative, the model overestimates wages in the 1st and 2nd deciles of wage distribution for most cities. Given the complementarity between private and social returns, workers require to keep a balance between $B(z)$ and $A(z)$ to maximize their productivity. When $\bar{B}$ decreases, the level of $B(z)$ decreases at all z, but the balance between $B(z)$ and $A(z)$ will diminish more for a lower z. This flexibility allows the model to generate lower wages in the bottom deciles.

In addition, I have estimated the model alternatively by normalizing the aggregation parameter $\gamma = 1$. Moments that this alternative estimation fails to match show variations that identify γ. With $\gamma = 1$, the estimated model generates city-level agglomeration elasticities that are largely decreasing in city size (failing to match the fifth moment). The reason is that when γ is too large, agglomeration forces amplify the complementarity between private and social returns, and does so more in smaller cities. Accordingly, a lower γ is required for the model to generate flat city-level agglomeration elasticities.

The last moment condition matters for the magnitude of agglomeration elasticities. The recovered $\bar{y}(i)$ is given by $\bar{y}(i) = Q(i)/\Big(\int_z y(i,z)L(i,z)dz\Big)$. Given total income $Q(i) = w^{avg}(i)L(i)$, and $\ell(i,z) \equiv L(i,z)/L(i)$,

$$\ln w^{avg}(i) = \ln \int_z y(i,z)\ell(i,z)dz + \ln \bar{y}(i), \tag{25}$$

where $y(i,z)$, is an implicit function of $L(i)$ and the distribution of $\ell(i,z)$. The orthogonality between $\ln \bar{y}(i)$ and $Z(i)$ gives a structural counterpart of the reduced-form, IV regression of $\log w^{avg}(i)$ against $\log L(i)$

using $Z(i)$ as an instrument. Consider a special case in which $y(i,z) = L(i)^{\alpha_0}$, then,

$$\ln w^{avg}(i) = \alpha_0 \ln L(i) + \ln \bar{y}(i) \tag{26}$$

As discussed in Section 2, we require an instrument to identify α_0 (The reduced-form, IV regression of equation (26), with housing supply elasticity as the instrument, sets α_0 at 0.084, with KP F-stat equal 38).[29] In the structural estimation, without the last moment condition, it is possible for the model to match the other moments at different combinations of ρ and γ. The orthogonality identifies the right combination. In particular, a higher γ and a lower ρ make room for larger agglomeration elasticities. The right combination of γ and ρ generates a flat $\alpha(i)$ (5th moment), the right slope of skill premiums (implicit in the 1st-4th moments), and the right magnitude of city-level agglomeration elasticity (imposed by the 6th moment).

The structural estimation not only guides counterfactual exercises, but also makes room for decomposition exercises. The reduced-form estimate α_0 reflects the combined effect of (i) worker-level agglomeration elasticity, and (ii) cities' skill composition. I use the structure, with parameter estimates, to decompose the city-level elasticity into these two components, a result that I discuss in Section 5.4.

5.4 Results

The estimation delivers these estimates with bootstrap standard errors reported in parentheses:

ρ	γ	μ	$\bar{B}$
0.62	0.17	1.41	0.69
(0.07)	(0.02)	(0.10)	(0.61)

Table 5: Parameter Estimates

As explained in Section 5.3, the identification of $\bar{B}$ depends on the variations in the bottom deciles of wage distribution. The relationship between wages and city size is more cloudy in the bottom deciles (Fig. 5). This feature of data is reflected in somewhat large standard error of $\bar{B}$.[30]

I turn to evaluate the performance of the estimated model. To do so, I simulate the model, evaluated

[29] Any other variable that can serve as an instrument here can be used as a moment condition in the structural estimation. Here I use housing supply variables for two reasons. First, in reduced form the first stage is somewhat weak for most variables used in Section 2, Pattern 4, except for housing supply variables, and historical variables such as past population or extent of streetcar use in 1902. Second, these historical variables are more likely to be correlated with current productivity shocks as these shocks can be persistent over time. Therefore, although not perfect, housing supply variables seem to provide the most admissible instrument among other variables.

[30] The standard errors are calculated based on the estimates for random samples with replacement. The 95% confidence interval of $\bar{B}$ is $[0.45, 3.27]$, which concentrates around 0.69, whereas larger values belong to samples in which small cities with higher bottom decile wages are over-representative.

at the parameter estimates, then compare predicted with observed moments.

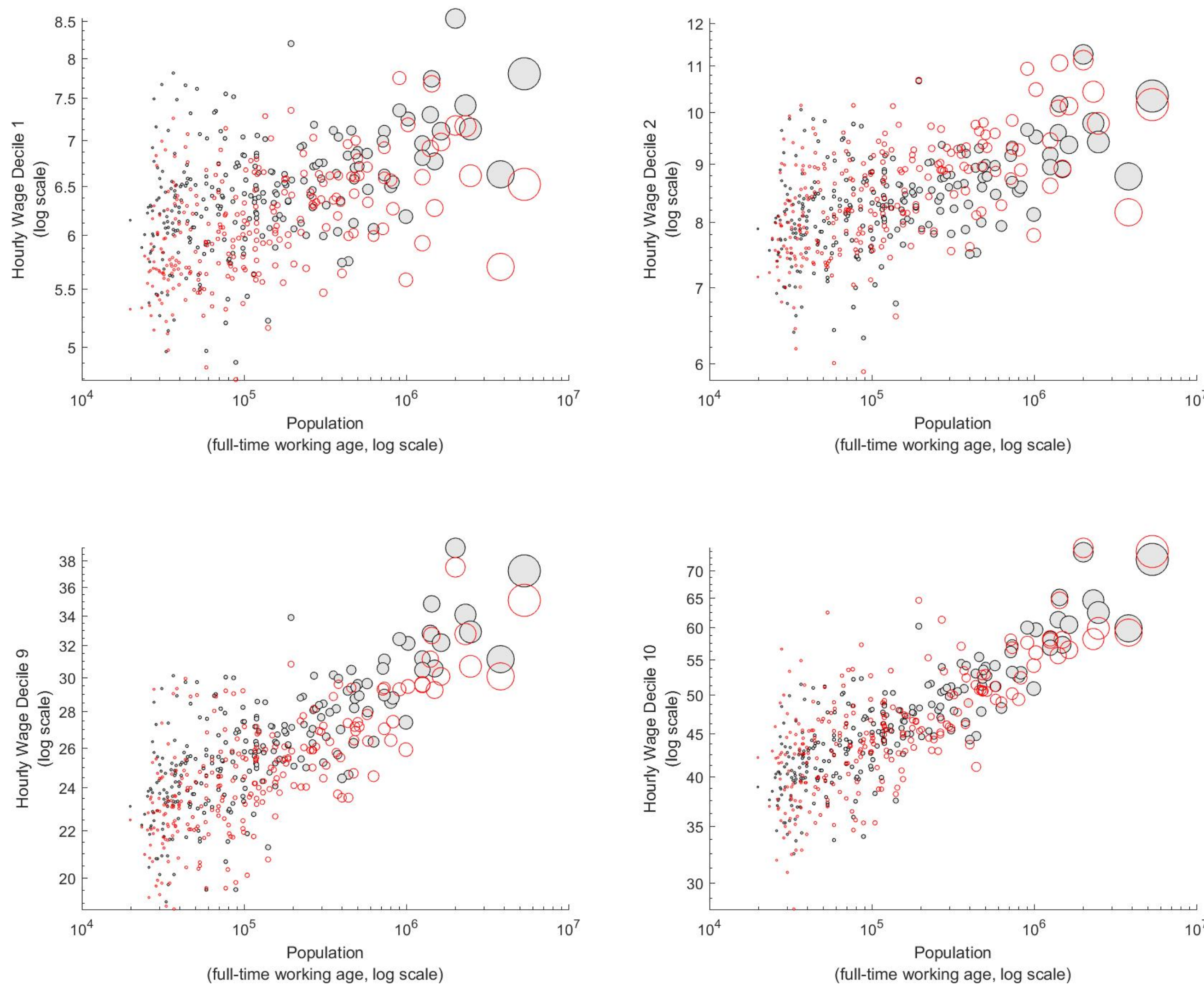
Figure 5 shows predicted and actual wages of the 1st, 2nd, 9th, and 10th deciles of within skill distribution across all cities. Figure A.2 shows the model fit with respect to non-targeted wage deciles. Overall, predicted wages across skills and cities are tightly close to the observed wages. Figure 6 shows predicted skill premiums across city size compared to the observed ones in the data. The presented skill premiums are the ratio of decile 10 to 1 of within city wage distribution, the ratio of decile 9 to 2, the ratio of decile 8 to 3, etc. The model closely predicts the intercepts and slopes of the scatter plots of the skill premiums against city size.

Figure 7 depicts the model-implied private returns to skill, $B(z)$, (left panel), and worker-level agglomeration elasticity in the median size city $\alpha(.,z)$ (right panel) across the skill continuum. Both $B(z)$ and $\alpha(.,z)$ sharply rise at the top decile of the skill distribution. The worker-level agglomeration elasticity decreases in city size, but it increases in skill. The resulting city-level agglomeration elasticity conditional on within-city skill distribution, $\alpha(i)$ given by eq. (21), remains virtually constant across cities (see Figure A.3). Estimated $\alpha(i)$ on average equals 0.051, it equals 0.050 at the 20th percentile and 0.053 at the 80th percentile of estimated αs.

Figure (A.4) plots the endogenous part of city productivity, $\ln \int_z y(i,z)\ell(i,z)dz$, against city population, $L(i)$, in log terms. The slope of this plot is the unconditional city-level agglomeration elasticity, α^{unc}, defined in eq. (22). This slope is virtually constant (due to the 5th moment) and equals 0.082. This estimate is effectively the same as the reduced-form, IV estimate of city-level agglomeration elasticity in regression (26).

The gap $\alpha^{unc} - \alpha$ equals $0.031 (= 0.082 - 0.051)$. That is, 38% $(= 0.031/0.082)$ of the scale elasticity at the level of cities is accounted for by endogenous changes to within-city skill distribution as city population changes. The other 62% is a weighted average of the agglomeration elasticity at the level of workers, controlling for within-city skill distribution, with weights given by equation (21).

Figure 5: Wages of Deciles 1, 2, 9, 10 in All Cities, Model vs Data



Notes: Red circle: data, Gray filled circle: model

Figure 6: Skill Premiums, Model vs Data

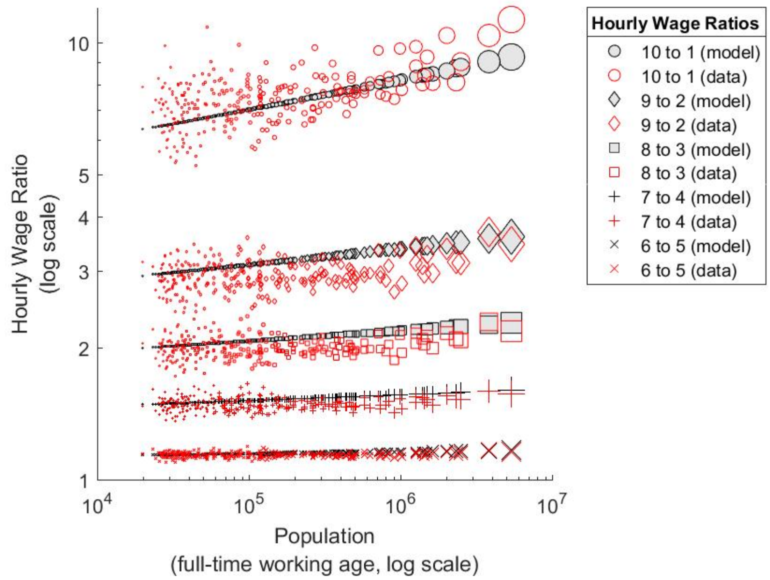
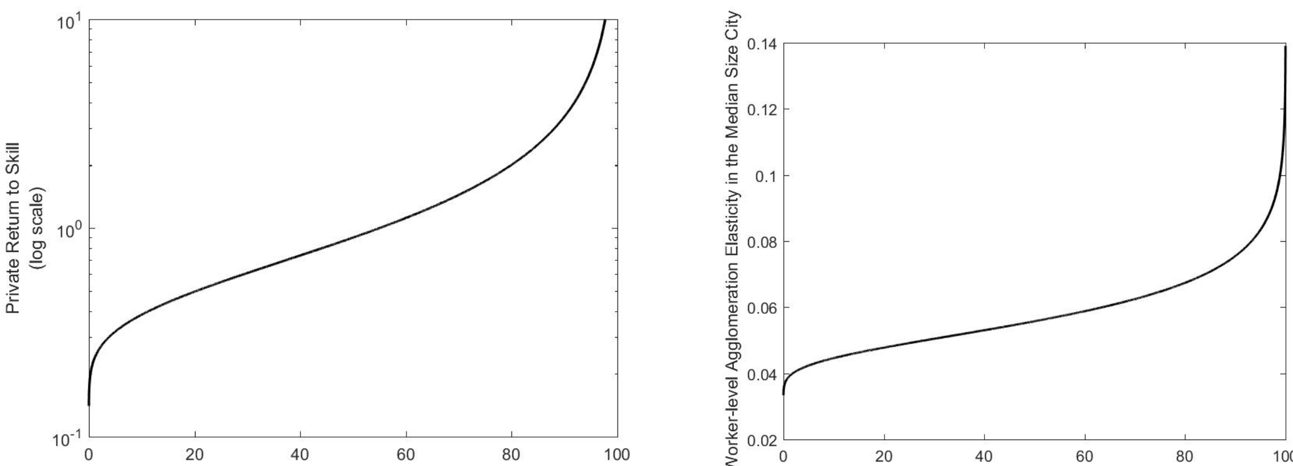


Figure 7: Private Returns and Worker-level Agglomeration Elasticity across Skills



Private Returns to Skill

Worker-Level Agglomeration Elasticity in the Median City

6 Quantitative Exercises

I use the estimated model to examine the welfare implications of place-based policies and skill-biased changes in technology.

6.1 Place-based Policies

I conduct a quantitative exercise to examine welfare implications of place-based policies. The importance of this exercise is better understood in comparison with the benchmarks in the literature, summarized in Section 4.4. In a model with homogeneous labor, where to subsidize does not matter for aggregate welfare, if the agglomeration elasticity is constant (which empirically appears to be the relevant case, as discussed in Section 2). I take a step forward by addressing: "What are the implications of place-based policies for welfare inequality when we take skill heterogeneity into account?"

To address this question, I consider a policy which has no aggregate welfare effect in a special case of my model in which there is one skill type and ρ is set to one ensuring that agglomeration elasticity is constant (and equals $\gamma/2$, as shown in Proposition 1). In order for this policy to have zero welfare effect in this special case, I set $\Delta = 0$, as illustrated by Proposition 3, Section 4.4. Let the set of small cities be all cities whose population is below the median size city, and put other cities in the set of large cities. I consider a counterfactual policy by which $\bar{u}(i)$ increases by δ^+ percent for all small cities, and decreases by $\delta^- = 5$ percent for all large cities, such that equation (19) holds:

$$\text{(population of small cities)} \times \delta^+ + \text{(population of large cities)} \times (-\delta^-) = 0$$

Here, $\delta^- = -0.05$ may reflect a tax equivalent rate of 5% on per capita income on residents of large cities. The above equation implies $\delta^+ = 0.52$, reflecting an equivalent of 52% subsidy on per capita income of residents of small cities (with a national employment share of 9%). I use expected welfare, $W(z)$, given by equation 12, as the measure of welfare in my analysis. I compute average $W(z)$ for all workers z within every decile of the national skill distribution to report average welfare of that decile.

Figure 8 shows changes to average welfare across skills from the baseline to the counterfactual (post-policy) equilibrium. Strikingly, average welfare rises for all deciles 1-8 of the national skill distribution, and it falls only for the two top deciles. Accordingly, welfare inequality decreases across skill deciles. The welfare ratio of top to bottom decile, in particular, falls by 0.64%. Although expected welfare increases for 8 out of 10 deciles, aggregate expected welfare remains virtually unchanged. The reason lies in relatively large welfare values of workers in the two top deciles. In sum, although this policy —that subsidies small cities at the expense of large ones— keeps aggregate welfare unchanged, it notably reduces welfare inequality across skills.

Why does welfare inequality fall? Higher skilled workers reside more than proportionately in larger cities. As a result they lose more from taxing large cities, or they gain less from subsidizing small cities. In addition, higher skilled workers benefit more from the concentration of economic activity. A policy that favors smaller cities, to a certain extent, makes the distribution of population across cities more equal. Hence, higher skilled workers lose some of their agglomeration advantage that arises due to the complementarity between skill and scale.

These results highlight the importance of distributional impact of policies that favor smaller areas. From the perspective of policy-makers regarding distributional objectives, it might be impractical to target skill groups directly. If so, spatial polices can be a practical alternative. The results here shift our attention from the neutrality of policy effects on aggregate welfare to their sizable effect on welfare inequality.

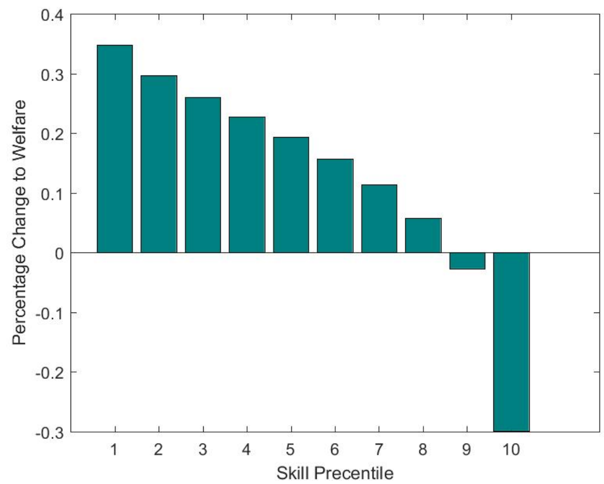


Figure 8: Percentage change to welfare from baseline to counterfactual across skill deciles

6.2 Skill-biased Technological Change

Wage inequality in the U.S. sharply rose between 1980 and 2000. The national rise in wage inequality was accompanied by a growth in the relationship between within-city wage inequality and city population. Figure 9 plots the ratio of top to bottom decile of wage distribution across American cities in 1980 and 2000. Both the intercept (reflecting the overall premium) and the slope (reflecting the relationship between the premium and city size) are higher in 2000. Several influential studies have asserted that this development reflects a skill-biased change in technology, meaning that returns to skill grew more for higher skilled workers. I use my estimated model to explore the implications of such a technological change for welfare inequality. Specifically, I ask: "What would have been counterfactual welfare inequality across skills in 2000 if skill wage premiums had remained at their 1980 levels?"

To perform this counterfactual, I fix ρ, γ, μ and $\bar{B}$ as well as productivity and amenity shifters, $\bar{y}$ and $\bar{u}$, and re-estimate parameter φ, in $B(z) = \bar{B}z^{\varphi}$, by matching model predictions to skill wage premiums in 1980. The moments consist of wage ratios of decile 10 to 1, 9 to 2, 8 to 3, 7 to 4, and 6 to 5 within every city in 1980. The resulting estimate of φ is 0.81, compared to 1 in the benchmark for 2000. The re-estimated model fits well to the premiums in 1980, as shown in Figure A.5. The increase in φ from 0.81 to 1.00

represents a skill-biased technological change between 1980 and 2000 with which the model matches the observed changes to skill premiums. This exercise remains a counterfactual as I keep the skill distribution and location shifters constant.

Figure 10 shows percentage change to average welfare from the counterfactual case (related to 1980) to the baseline (related to 2000) across the skill distribution. The average welfare of the 1st, 2nd, and 3dr deciles falls, and that of the other deciles rises. Workers in the bottom decile suffer a loss of 9.6% and in the top decile enjoy 28.2% increase in expected welfare. The notable welfare loss of the 1st and 2nd deciles will not be visible in the change to the average welfare of an aggregate group consisting of the five bottom deciles. This in part highlights the importance of including the tails of skill distribution into the analysis.

These significant changes to welfare inequality highlight the importance of sorting and complementarities between skill and scale. An exogenous skill-biased change in private returns to skill is amplified by the increase in skill-biased social returns that stem from the endogenous sorting of higher skills to larger cities.

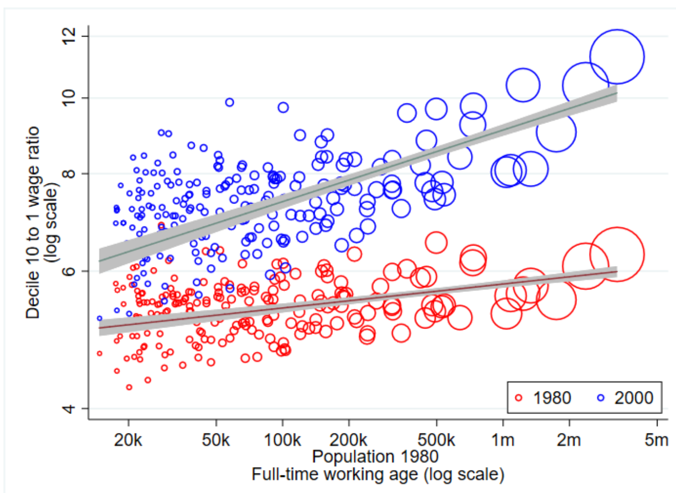


Figure 9: Ratio of the observed top to bottom decile of hourly wage distribution in 1980 and 2000

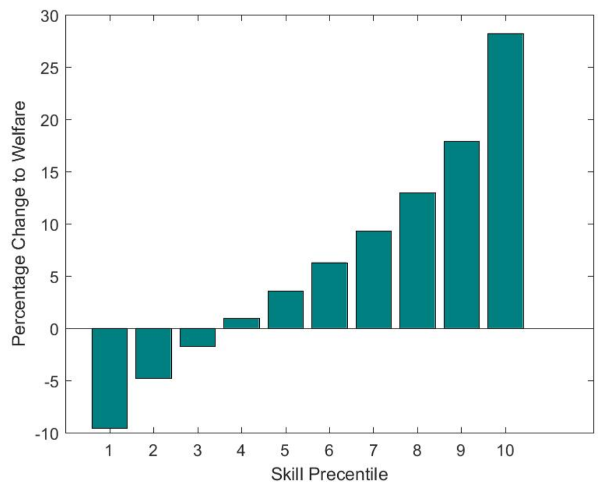


Figure 10: Percentage change to welfare from counterfactual to baseline across skill deciles

7 Conclusion

The recent literature that addresses scale economies in economic geography and trade has been, more or less, agnostic about specific sources at work behind empirical relationships that reflect returns to scale. This paper complements that approach by focusing on a particular microfoundation, motivated by a long tradition that attributes external returns to knowledge spillovers. This alternative approach helps explain nonlinearities in scale economies, and can be used to address distributional welfare implications of policies.

The tools developed here can be used in other applications and extensions. I have deliberately abstracted away from several dimensions that are studied elsewhere. For example, Hsieh and Moretti (2015) emphasize on the spatial misallocation of economic activity due to housing constraints, Diamond (2016) shows that local amenities may improve with the skill composition of a city, Fajgelbaum et al. (2016) and Ossa (2017) explicitly allow for a tax system in a spatial framework. It is straightforward to embed these dimensions into the current model. A particular direction for future research is to collect micro-level data to illustrate competing microfoundations that underly the limits of local learning, such as frictions in knowledge spillovers across geographic space and production tasks, or due to the human capacity for assimilating new knowledge.

References

Adao, R., Arkolakis, C., and Esposito, F. (2017). Trade, Agglomeration Effects, and Labor Markets: Theory and Evidence. *Working Paper.*

Ahlfeldt, G. M., Redding, S. J., Sturm, D. M., and Wolf, N. (2015). The Economics of Density: Evidence From the Berlin Wall. *Econometrica*, 83(6):2127–2189.

Albouy, D. (2008). Are Big Cities Bad Places to Live? Estimating Quality of Life across Metropolitan Areas. *National Bureau of Economic Research Working Paper Series*, No. 14472.

Allen, T. and Arkolakis, C. (2014). Trade and the topography of the spatial economy. *Quarterly Journal of Economics*, 129(3):1085–1139.

Au, C.-C. and Henderson, J. V. (2006). Are Chinese cities really too small? *Review of Economic Studies*, 73:549–576.

Bacolod, M., Blum, B. S., and Strange, W. C. (2009). Skills in the city. *Journal of Urban Economics*, 65(2):136–153.

Baum-Snow, N., Freedman, M., and Pavan, R. (2017). Why Has Urban Inequality Increased? *American Economic Journal: Applied Economics (Forthcoming)*.

Baum-Snow, N. and Pavan, R. (2013). Inequality and City Size. *Review of Economics and Statistics*, 95(5):1535–1548.

Behrens, K., Duranton, G., and Robert-Nicoud, F. (2014). Productive Cities: Sorting, Selection, and Agglomeration. *Journal of Political Economy*, 122(3):507–553.

Burchfield, M., Overman, H. G., Puga, D., and Turner, M. A. (2006). Causes of sprawl: A portrait from space. *Quarterly Journal of Economics*, 121(2):587–633.

Burstein, A., Morales, E., and Vogel, J. (2017). Changes in between-group inequality : computers , occupations , and international trade. *Working Paper*.

Burstein, A., Morales, E., and Vogel, J. (2018). Changes in between-group inequality: computers, occupations, and international trade.

Chipman, J. S. (1965). A Survey of the Theory of International Trade : Part 2, The Neo-Classical Theory. *Econometrica*, 33(4):685–760.

Combes, P. P., Duranton, G., and Gobillon, L. (2008). Spatial wage disparities: Sorting matters! *Journal of Urban Economics*, 63(2):723–742.

Combes, P.-P., Duranton, G., Gobillon, L., Puga, D., and Roux, S. (2012). The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection. *Econometrica*, 80(6):2543–2594.

Combes, P.-P. and Gobillon, L. (2015). The Empirics of Agglomeration Economies. *Handbook of Regional and Urban Economics*, 5:247–348.

Costinot, A. and Vogel, J. (2010). Matching and Inequality in the World Economy. *Journal of Political Economy*, 118(4):747–786.

Cutler, D. M., Glaeser, E. L., and Vigdor, J. L. (1999). The Rise and Decline of the American Ghetto. 107(3):455–506.

Davis, D. R. and Dingel, J. (2017). The Comparative Advantage of Cities. *Working Paper*.

Davis, D. R. and Dingel, J. I. (2018). A Spatial Knowledge Economy. *Working Paper*.

De la Roca, J. and Puga, D. (2017). Learning byworking in big cities. *Review of Economic Studies*, 84(1):106–142.

Desmet, B. K. and Rossi-hansberg, E. (2013). Urban Accounting and Welfare. *American Economic Review*, 103(6):2296–2327.

Diamond, R. (2016). Determinants and Welfare Implications of US Workers Diverging Location Choices by Skill. *American Economic Review*, 106(3):479–524.

Ethier, W. J. (1982). Decreasing Costs in International Trade and Frank Graham's Argument for Protection. *Econometrica*, 50(5):1243–1268.

Fajgelbaum, P., Morales, E., Serrato, J. C. S., and Zidar, O. (2016). State Taxes and Spatial Misallocation. *Working Paper*.

Fajgelbaum, P. D. and Gaubert, C. (2018). Optimal Spatial Policies , Geography and Sorting.

Farrokhi, F. and Jinkins, D. (2017). Wage Inequality and the Location of Cities. *Working Paper*.

Galle, S., Rodriguez-Clare, A., and Yi, M. (2017). Slicing the Pie : Quantifying the Aggregate and Distributional Effects of Trade. *Working Paper*.

Gaubert, C. (2017). Firm Sorting and Agglomeration. *Working Paper*, (December):1–61.

Glaeser, E. L. and Gottlieb, J. D. (2008). The Economics of Place-Making Policies. *Brookings Papers on Economic Activity*, 2008(1):155–253.

Glaeser, E. L. and Gottlieb, J. D. (2009). The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States. *Journal of Economic Literature*, 47(4):983–1028.

Glaeser, E. L., Resseger, M., and Tobio, K. (2009). Inequality in cities. *Journal of Regional Science*, 49(4):617–646.

Glaeser, E. L. and Resseger, M. G. (2010). The Complementarity between Cities and Skills. *Journal of Regional Sciencs*, 50(1):221–244.

Gould, E. D. (2007). Cities, workers, and wages: A structural analysis of the urban wage premium. *Review of Economic Studies*, 74(2):477–506.

Gyourko, J., Saiz, A., and Summers, A. A. (2008). A New Measure of the Local Regulatory Environment for Housing Markets: The Wharton Residential Land Use Regulatory Index. *Urban Studies*, 45:693–729.

Henderson, J. V. (1974). The Sizes and Types of Cities. *American Economic Review*, 64(4):640–656.

Hsieh, C.-t., Hurst, E., Jones, C. I., and Klenow, P. J. (2016). The Allocation of Talent and U.S. Economic Growth. *Working Paper*.

Hsieh, C.-T. and Moretti, E. (2015). Housing Constraints and Spatial Misallocation. *NBER Working Papers 21154*.

Kline, P. and Moretti, E. (2014a). Local economic development, agglomeration economies, and the big push: 100 years of evidence from the Tennessee Valley Authority. *Quarterly Journal of Economics*, 129(1):275–331.

Kline, P. and Moretti, E. (2014b). *People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs*, volume 6.

Lucas, R. E. and Moll, B. (2014). Knowledge Growth and the Allocation of Time. *Journal of Political Economy*, 122(1):1–51.

Marshall, A. (1920). *Principles of Economics*. London: Macmillan and Co., Ltd. Available at the online Library of Economics and Liberty.

Moretti, E. (2013). Real wage inequality appendix. *American Economic Journal: Applied Economics*, 123(6):257–259.

Notowidigdo, M. J. (2011). The Incidence of Local Labor Demand Schocks. *National Bureau of Economic Research*, No. w17167.

Ossa, R. (2017). A Quantitative Analysis of Subsidy Competition in the U . S . *Working Paper*.

Perla, J. and Tonetti, C. (2014). Equilibrium Imitation and Growth. *Journal of Political Economy*, 122(1):52–76.

Redding, S. J. and Rossi-Hansberg, E. (2017). Quantitative Spatial Economics. *Annual Review of Economics*, (9.1).

Rosenthal, S. S. and Strange, W. C. (2004). Evidence on the Nature and Sources of Agglomeration Economies. *Handbook of Regional and Urban Economics*, 4:2119–2171.

Saiz, A. (2010). The Geographic Determinants of Housing Supply *. *Quarterly Journal of Economics*, 125(3):1253–1296.

Suárez Serrato, J. C. and Zidar, O. (2016). Who benefits from state corporate tax cuts? A local labor markets approach with heterogeneous firms. *American Economic Review*, 106(9):2582–2624.

Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.

Appendix A Additional Tables and Figures

Variable	Mean	Std. Dev.	Min.	Max.
working population	157903.32	351374.82	14740	3280260
hourly wage	7.58	0.8	5.75	11.1
decile 10 to 1 wage ratio	5.39	0.45	4.26	6.88
decile 9 to 2 wage ratio	2.71	0.17	2.14	3.19
decile 8 to 3 wage ratio	1.95	0.1	1.61	2.27
decile 7 to 4 wage ratio	1.47	0.06	1.3	1.62
decile 6 to 5 wage ratio	1.14	0.02	1.08	1.18
college employment ratio	0.38	0.12	0.15	0.89
college wage premium	1.36	0.1	1.1	1.72
Unit of observation	Consolidated Metropolitan Statistical Area			
Number of observations	205			

Table A.1: Summary of Statistics, Census 1980

	(1)	(2)	(3)
VARIABLES	log wage	log college wage premium	log college employment ratio
log population	0.0459***	0.0165**	0.0837***
	(0.00519)	(0.00732)	(0.0148)
Observations	205	205	205
R-squared	0.467	0.124	0.240

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A.2: Wages, College Wage Premium, and College Employment Ratio versus City Population, 1980

	log hourly wage ratio				
	(1)	(2)	(3)	(4)	(5)
	decile 10 to 1	decile 9 to 2	decile 8 to 3	decile 7 to 4	decile 6 to 5
log population	0.0307***	0.00976**	0.00261	0.00220	0.00163
	(0.00651)	(0.00410)	(0.00278)	(0.00297)	(0.00127)
Constant	1.329***	0.881***	0.632***	0.360***	0.108***
	(0.0781)	(0.0495)	(0.0337)	(0.0353)	(0.0151)
Observations	205	205	205	205	205
R-squared	0.304	0.076	0.009	0.010	0.034

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table A.3: Wage Inequality versus Population, 1980

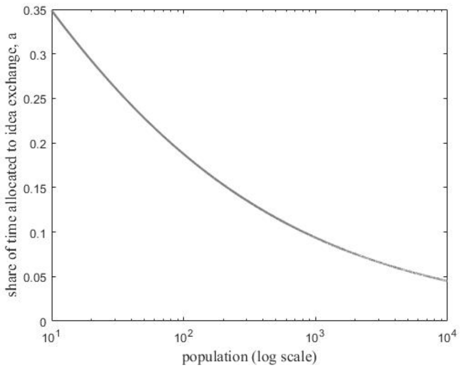
Note: As shown by Table A.4, the effect of city size on hours worked is not statistically significant for non-college workers. Total wage, hourly wage, and hours worked of college graduates increase, on average, by elasticities of 0.0823, 0.0802, and 0.00281, respectively. For college graduates, variations in hours worked account for only 3.5% of their higher wages in larger cities, with the remaining 96.5% explained by variations in hourly wage.

	(1)	(2)	(3)
	log hours worked	log hourly wage	log wage
log population	-0.000642	0.0563***	0.0556***
	(0.00110)	(0.00503)	(0.00510)
college	-0.00275	0.175***	0.172***
	(0.0118)	(0.0311)	(0.0299)
college x log population	0.00281***	0.0239***	0.0267***
	(0.000953)	(0.00240)	(0.00233)
Constant	7.696***	1.546***	9.242***
	(0.0150)	(0.0703)	(0.0703)
Observations	2,611,777	2,611,777	2,611,777
R-squared	0.072	0.259	0.281

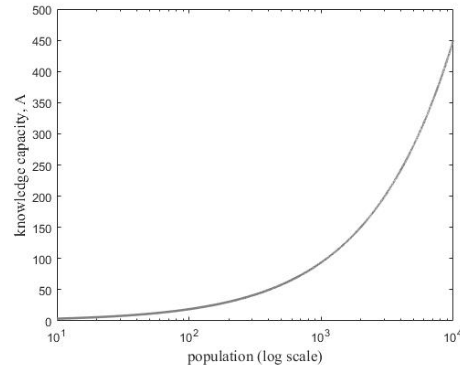
Standard errors, reported in parentheses, are clustered by CMSA.

*** p<0.01, ** p<0.05, * p<0.1

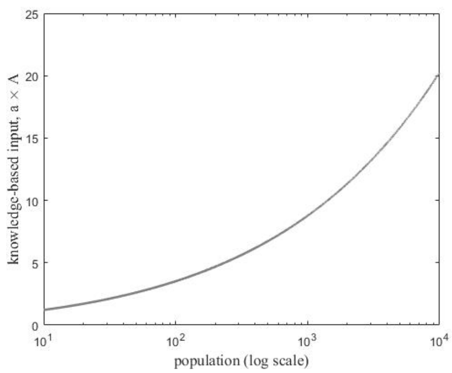
Table A.4: Hours Worked and Hourly Wage versus Population and College



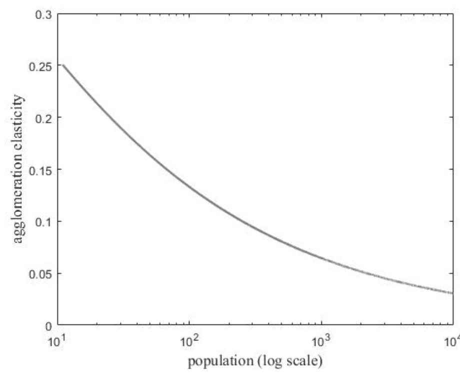
Idea exchange participation



Knowledge capacity



Knowledge-based input

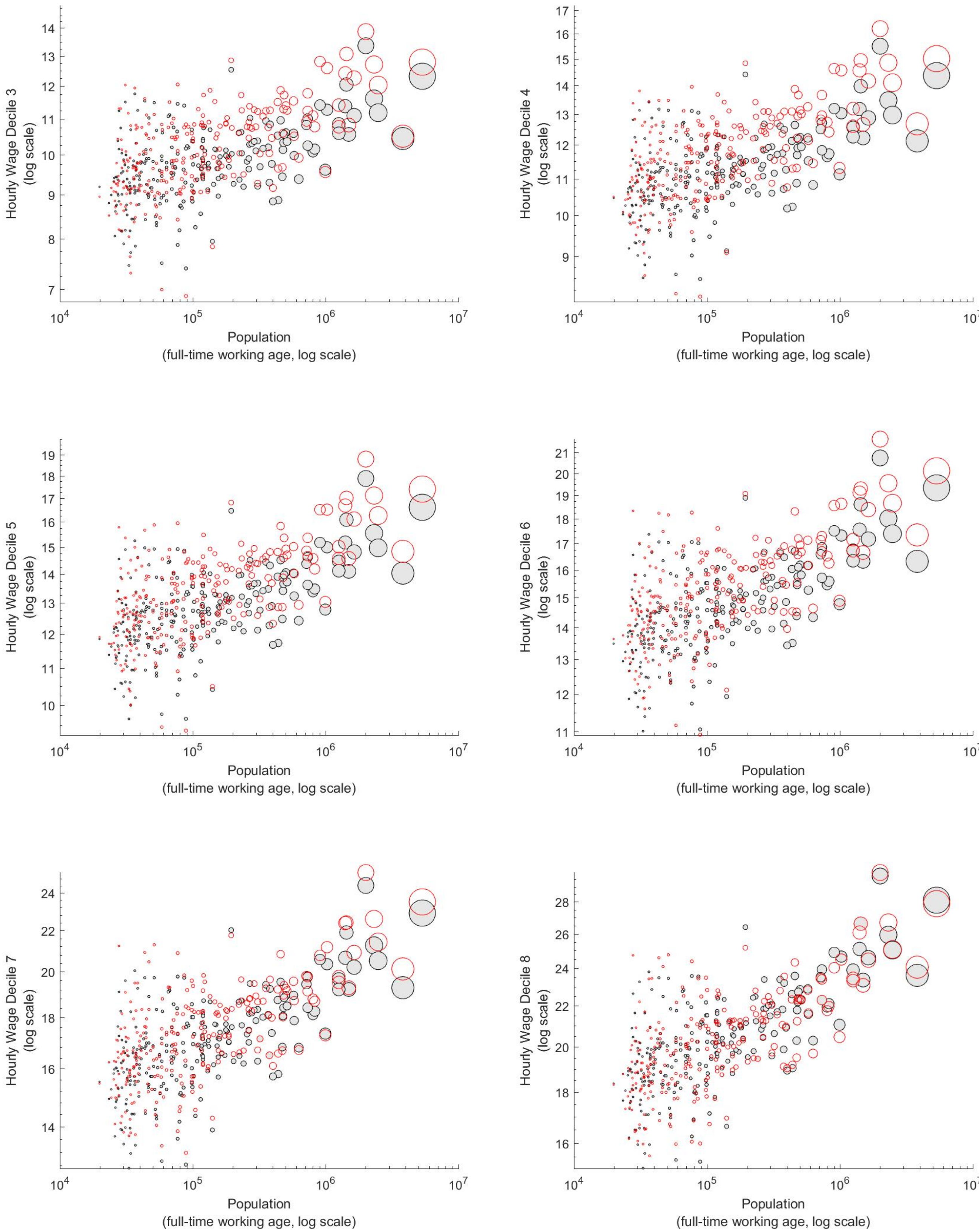


Agglomeration elasticity

Figure A.1: Productivity versus population, $\rho \in (0, 1)$

Notes: Simulated based on $\rho = 0.5$, $\gamma = 1$, $B = 1$

Figure A.2: Wages of Skill Deciles in All Cities, Model vs Data



Notes: Red circle: data, Gray filled circle: model

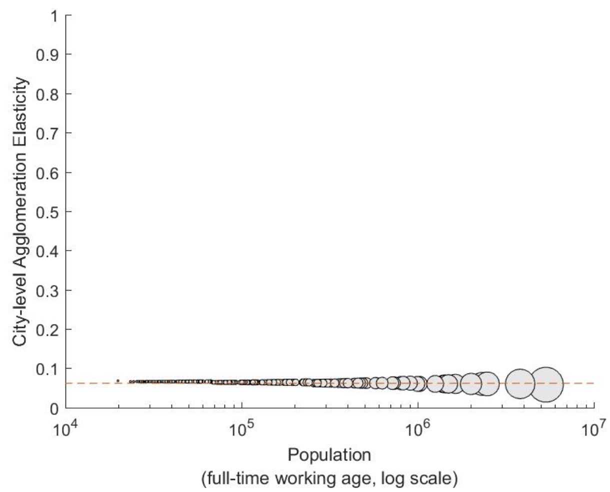


Figure A.3: Conditional City-Level Agglomeration Elasticity

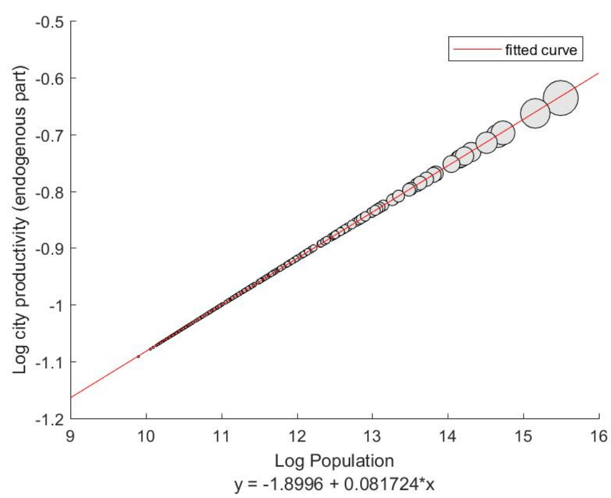


Figure A.4: City Productivity (endogenous part) against City Population

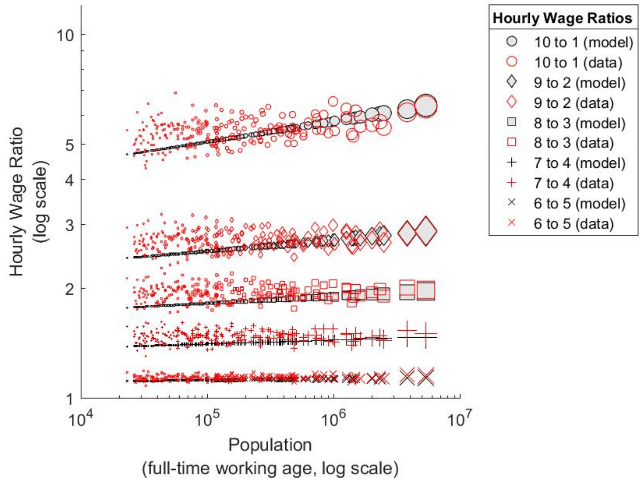


Figure A.5: Wage Skill Premiums against City Size in 1980, Model vs Data

Appendix B Numerical Algorithms

B.1 Solution Algorithm

The solution algorithm takes $V(z)$, $B(z)$, $\bar{y}(i)$, and $\bar{u}(i)$ as given, and solves for wages $w(i,z)$ and populations $L(i,z)$.

For the exposition, it is useful to fix a skill level z_0, and define $\widetilde{W}(z) \equiv \frac{W(z)}{W(z_0)}$, and $\widetilde{w}(i,z) \equiv \frac{w(i,z)}{w(i,z_0)}$.

1. Guess $L(i,z)$, $A(i)$, and $\widetilde{W}(z)$ for all i and z.

2. Calculate $y(i,z)$ according to equation (8), and $a(i,z)$ according to equation (7).

3. Calculate $Q(i)$ according to equation (9), and $\widetilde{w}(i,z)$ according to equation (15)

4. Using the equilibrium condition (17), and $w(i,z) = \widetilde{w}(i,z)w(i,z_0)$, calculate

$$w(i,z) = \frac{\widetilde{w}(i,z)Q(i)}{\int_z \widetilde{w}(i,z)L(i,z)dz}$$

5. Given $L(i) = \int_z L(i,z)dz$, calculate $v(i,z)$ according to equation (13).

6. Update $L(i,z)$

$$L^{new}(i,z) = v(i,z)V(z)$$

Update $\widetilde{W}(z)$ according to equation (12),

$$\widetilde{W}^{new}(z) = \left[\frac{\sum_j \left(w(j,z)\bar{u}(j)L(j)^{\beta}\right)^{\theta}}{\sum_j \left(w(j,z_0)\bar{u}(j)L(j)^{\beta}\right)^{\theta}}\right]^{\frac{1}{\theta}}$$

Update $A(i)$ according to equation (6),

$$A^{new}(i) = \left[\int_z a(i,z)L(i,z)dz\right]^{\gamma}$$

Stop if $\max\left|\frac{L^{new}(i,z)}{L(i,z)} - 1\right| < \epsilon$, $\max\left|\frac{A^{new}(i)}{A(i)} - 1\right| < \epsilon$, $\max\left|\frac{\widetilde{W}^{new}(z)}{\widetilde{W}(z)} - 1\right| < \epsilon$ for a sufficiently small ϵ. Otherwise, set $L(i,z) = L^{new}(i,z)$, $\widetilde{W}(z) = \widetilde{W}^{new}(z)$, and $A(i) = A^{new}(i)$ then go to Step 2.

B.2 Calibration Algorithm

The calibration algorithm takes $L(i)$ and $w_{avg}(i)$ as given, and solves for productivity shifters $\bar{y}(i)$ and amenity shifters $\bar{u}(i)$ such that all equilibrium conditions hold.

1. Guess $L(i,z)$, $A(i)$, and $\widetilde{W}(z)$ for all i and z.

2. Calculate $y(i,z)$ according to equation (8), and $a(i,z)$ according to equation (7).

3. Calculate $\widetilde{w}(i,z)$ according to equation (15).

4. Using $L(i)w^{avg}(i) = \int_z w(i,z)L(i,z)dz$, and $w(i,z) = \widetilde{w}(i,z)w(i,z_0)$, calculate

$$w(i,z) = \frac{\widetilde{w}(i,z)L(i)w^{avg}(i)}{\int_z \widetilde{w}(i,z)L(i,z)dz}$$

5. Calibrate $\bar{y}(i)$ using equilibrium condition (17),

$$\bar{y}(i) = \frac{w^{avg}(i)L(i)}{\int_z y(i,z)L(i,z)dz}$$

6. Calibrate $\bar{u}(i)$ according to the following. By labor market clearing,

$$\begin{aligned} L(i) &= \int_z L(i,z)dz \\ &= \int_z \frac{\left(w(i,z)\bar{u}(i)L(i)^{\beta}\right)^{\theta}}{(W(z)/\bar{\gamma})^{\theta}} V(z)dz \\ &= \bar{u}(i)^{\theta} L(i)^{\beta\theta} \int_z \frac{w(i,z)^{\theta}}{(W(z)/\bar{\gamma})^{\theta}} V(z)dz \end{aligned}$$

which implies that

$$\bar{u}(i) = \bar{c}\left[\int_z \frac{w(i,z)^{\theta}}{\widetilde{W}(z)^{\theta}} V(z)dz\right]^{-\frac{1}{\theta}} L(i)^{\frac{1-\beta\theta}{\theta}}$$

where $\bar{c}$ is a constant which ensures a normalization that $\bar{u}(i_0) = 1$,

$$\bar{c} \equiv \left[\int_z \frac{w(i,z)^{\theta}}{\widetilde{W}(z)^{\theta}} V(z)dz\right]^{\frac{1}{\theta}} L(i_0)^{-\frac{1-\beta\theta}{\theta}}$$

7. Calculate $W(z) = \widetilde{W}(z)W(z_0)$ using equation (12). The normalization of $\bar{u}(i_0) = 1$ pins downs $W(z_0) = \bar{c}\bar{\gamma}$.

8. Update $L(i,z)$ according to (13); $\widetilde{W}(z)$ according to (12); and, $A(i)$ according to (6).

$$L^{new}(i,z) = \frac{\left(w(i,z)\bar{u}(i)L(i)^{\beta}\right)^{\theta}}{(W(z)/\bar{\gamma})^{\theta}} V(z)$$

$$\widetilde{W}^{new}(z) = \left[\frac{\sum_j \left(w(j,z)\bar{u}(j)L(j)^{\beta}\right)^{\theta}}{\sum_j \left(w(j,z_0)\bar{u}(j)L(j)^{\beta}\right)^{\theta}}\right]^{\frac{1}{\theta}}$$

$$A^{new}(i) = \left[\int_z a(i,z)L(i,z)dz\right]^{\gamma}$$

Notice that calibrated $\bar{u}(i)$'s from Step 6 ensure that $\int_z L^{new}(i,z)dz = L(i)$, and by construction $\sum_i L(i,z) = V(z)$.

Stop if $\max\left|\frac{L^{new}(i,z)}{L(i,z)} - 1\right| < \epsilon$, $\max\left|\frac{A^{new}(i)}{A(i)} - 1\right| < \epsilon$, $\max\left|\frac{\widetilde{W}^{new}(z)}{\widetilde{W}(z)} - 1\right| < \epsilon$ for a sufficiently small ϵ. Otherwise, set $L(i,z) = L^{new}(i,z)$, $\widetilde{W}(z) = \widetilde{W}^{new}(z)$, and $A(i) = A^{new}(i)$ then go to Step 2.

Appendix C Proofs and Mathematical Derivations

C.1 Derivations

Derivation of Equation (5). Using equation (3), $\frac{\partial a}{\partial A}\frac{A}{a} = (\rho-1)(1-a)$. Using this and equation (4),

$$\begin{aligned}\frac{1}{\gamma}A^{1/\gamma-1}dA &= adL + Lda \\ &= adL + L(\rho-1)(1-a)aA^{-1}dA\end{aligned}$$

By rearranging the terms, and given $A^{1/\gamma} = aL$,

$$\frac{dA}{dL}\frac{L}{A} = \frac{1}{1/\gamma + (1-\rho)(1-a)}$$

Using equations (3)-(4), $y = \left(A^{1-\rho} + B^{1-\rho}\right)^{1/(1-\rho)}$. So, $\frac{\partial y}{\partial A}\frac{A}{y} = (A/y)^{\rho-1} = a$. Finally,

$$\alpha \equiv \frac{d\ln y}{d\ln L} = \frac{d\ln y}{d\ln A} \times \frac{d\ln A}{d\ln L} = \frac{a}{1/\gamma + (1-\rho)(1-a)}$$

Derivation of Equation (16). In Appendix D, I allow for a finite elasticity of substitution across workers of different skills, denoted by ε. There, I derive equation (A.17). That is, with a finite elasticity of substitution ε, we directly check that employment is nonzero. If θ is finite, as $\varepsilon \to \infty$, we get equation (16).

Derivation of Equation (20). Holding city i fixed, I drop i. Taking the derivative of $a(z)$ w.r.t. A using optimal time allocation equation (7), gives

$$\frac{da(z)}{a(z)} = (\rho-1)(1-a(z))\frac{dA}{A}$$

Using the above when taking the derivative of A w.r.t. L, given $\ell(z)$, in equation (6),

$$\begin{aligned}\frac{1}{\gamma}A^{1/\gamma-1}dA &= \Big[\int_z \ell(z)a(z)dz\Big]dL + \Big[\int_z \ell(z)a(z)(\rho-1)(1-a(z))\frac{dA}{A}dz\Big]L \\ \frac{1}{\gamma}A^{1/\gamma-1}dA &= A^{1/\gamma}\frac{dL}{L} + \Big[\int_z \ell(z)a(z)(\rho-1)(1-a(z))\frac{dA}{A}dz\Big]L \\ \frac{1}{\gamma}dA &= A\frac{dL}{L} + A^{1-1/\gamma}\Big[\int_z \ell(z)a(z)(\rho-1)(1-a(z))\frac{dA}{A}dz\Big]L\end{aligned}$$

$$\frac{dA}{dL}\frac{L}{A} = \frac{1}{\frac{1}{\gamma} - (\rho-1)A^{-1/\gamma}L\Big[\int_z \ell(z)a(z)dz - \int_z \ell(z)a(z)^2dz\Big]}$$

$$\frac{dA}{dL}\frac{L}{A} = \frac{1}{\frac{1}{\gamma} - (\rho-1)A^{-1/\gamma}L\Big[A^{1/\gamma}L^{-1} - \int_z \ell(z)a(z)^2dz\Big]}$$

$$\frac{dA}{dL}\frac{L}{A} = \frac{1}{\frac{1}{\gamma} - \rho + 1 + (\rho-1)\frac{\int_z \ell(z)a(z)^2dz}{\int_z \ell(z)a(z)dz}}$$

Given that $\frac{\partial y(z)}{\partial A}\frac{A}{y(z)} = a(z)$,

$$\alpha(z) = \frac{d\ln y(z)}{d\ln A} \times \frac{d\ln A}{d\ln L} = \frac{a(z)}{\frac{1}{\gamma} + (1-\rho)\Big(1 - \frac{\int_z \ell(z)a(z)^2dz}{\int_z \ell(z)a(z)dz}\Big)}$$

Derivation of Equation (21). City productivity, $q(i) = Q(i)/L(i)$, is given by

$$q(i) = \bar{y}(i)\int_z \ell(i,z)y(i,z)dz$$

Fully differentiate this equation given the distribution of $\ell(i,z)$ for all z in city i,

$$q(i)d\ln q(i) = \bar{y}(i)\int_z \ell(i,z)y(i,z)d\ln y(i,z)$$

$$d\ln q(i) = \frac{\bar{y}(i)\int_z \ell(i,z)y(i,z)d\ln y(i,z)}{\bar{y}(i)\int_z \ell(i,z)y(i,z)dz}$$

$$\frac{d\ln q(i)}{d\ln L(i)} = \frac{\int_z \ell(i,z)y(i,z)\frac{d\ln y(i,z)}{d\ln L(i)}}{\int_z \ell(i,z)y(i,z)dz}$$

C.2 Proofs of Propositions

Proposition 1.

Define $H(A) = I(A) - K^{-1}(A)$ where $I(.)$ and $K(.)$ are defined by equations (3) and (4).

Case 1. Since $H'(A) < 0$, $H(0) = 1$, $H(\infty) = -\infty$, the Intermediate Value Theorem implies that there exists a unique outcome (a, A). The stability is ensured by the fact that I intersects K from above (as can be seen in Fig. 2). By an increase in L, a decreases. This is immediate as graphically shown by Fig. 2 or mathematically derived by differentiating $H = 0$ with respect to L. In equation (21) the numerator of α (which equals a) decreases, but the denominator increases as $(1-\rho) > 0$.

Case 2. When $\rho = 1$, $a = 0.5$ The rest is immediate as in Case 1.

Case 3. Since $I(0) = 0$ and $K^{-1}(0) = 0$, $H(0) = 0$. For $\epsilon > 0$, $H(\epsilon) > 0$, $H(\infty) = -\infty$, and since $\gamma \leq 1$, $H'(A) < 0$. The Intermediate Value Theorem then implies that there exists a unique outcome (a, A) in the range of nonzero A. The stability of the nonzero outcome is ensured as I intersects K from above. Here, by an increase in L, a increases, again as graphically shown by Fig. 2 or mathematically derived by differentiating $H = 0$ w.r.t. L. Take the partial derivative of α w.r.t. a in equation (21),

$$\frac{\partial \alpha}{\partial a} = \frac{1/\gamma + 1 - \rho}{(1/\gamma + (1-\rho)(1-a))^2}, \quad \text{implying:} \quad \frac{\partial \alpha}{\partial a} > 0 \Leftrightarrow \rho < 1 + 1/\gamma \tag{A.1}$$

So, when $\rho \in (1, 1 + 1/\gamma)$, as L increases, a increases which then implies that α increases.

Case 4 & 5. It is easy to check $H(0) = 0$. The tricky part is finding the threshold $L^\star$ beyond which there are two nonzero outcomes, below which there is a single nonzero outcome. $L^\star$ has a unique feature that schedule I is tangent to schedule K,

$$I(A; L^\star) = K^{-1}(A; L^\star) \quad \text{and} \quad \partial I(A; L^\star)/\partial A = \partial K^{-1}(A; L^\star)/\partial A$$

Replacing for I, K, $\partial I/\partial A$, and $\partial K/\partial A$,

$$\frac{A^{\rho-1}}{A^{\rho-1}B^{\rho-1}} = (L^\star)^{-1}A^{1/\gamma} \quad \text{and} \quad \frac{(1-\rho)A^{\rho-2}B^{\rho-1}}{(A^{\rho-1}B^{\rho-1})^2} = \frac{1}{\gamma}(L^\star)^{-1}A^{1/\gamma-1}$$

Solving for these two equations gives $a^\star = a(L^\star)$ as

$$a^\star = \frac{(\rho-1)\gamma - 1}{(\rho-1)\gamma}$$

An interesting observation is that $1/\gamma + (1-\rho)(1-a^\star) = 0$. That is, $a^\star$ is such that the agglomeration elasticity α equals infinity as a limiting case. $a^\star$ determines $A^\star = A(L^\star)$ and the corresponding population,

$$L^\star = (\rho - 1)\Big(\rho - 1 - 1/\gamma\Big)^{\frac{1-\gamma(\rho-1)}{\gamma(\rho-1)}} B^{1/\gamma}$$

When $L > L^\star$, there are two nonzero outcomes. For the outcome with a lower a, since I intersects K from below, the outcome is unstable. For the opposite reason the outcome with a higher a is stable. To see that α is decreasing in L, follow the same argument as in Case 3, but this time $\partial\alpha/\partial a < 0$ as implied by equation (A.1). Finally, Case 4 with $\rho = 1 + 1/\gamma$ is just a special, limiting case of Case 5.

Proposition 2.

(i) Consider two skill levels $z > z_0$. The skill wage premium of skill z with respect to z_0 in location i is

$\widetilde{w}(i,z) = w(i,z)/w(i,z_0)$ described by equation (15). According to equation (8),

$$
\begin{aligned}
\widetilde{w}(i,z) \geq \widetilde{w}(j,z) \quad &\Leftrightarrow \quad \frac{y(i,z)}{y(i,z_0)} \geq \frac{y(j,z)}{y(j,z_0)} \\
&\Leftrightarrow \quad \Big[\frac{A(i)^{\rho-1} + B(z)^{\rho-1}}{A(i)^{\rho-1} + B(z_0)^{\rho-1}}\Big]^{\frac{1}{\rho-1}} \geq \Big[\frac{A(j)^{\rho-1} + B(z)^{\rho-1}}{A(j)^{\rho-1} + B(z_0)^{\rho-1}}\Big]^{\frac{1}{\rho-1}} \\
&\Leftrightarrow \quad \frac{A(i)^{\rho-1} + B(z)^{\rho-1}}{A(i)^{\rho-1} + B(z_0)^{\rho-1}} \leq \frac{A(j)^{\rho-1} + B(z)^{\rho-1}}{A(j)^{\rho-1} + B(z_0)^{\rho-1}} \\
&\Leftrightarrow \quad (A(i)^{\rho-1} + B(z)^{\rho-1})(A(j)^{\rho-1} + B(z_0)^{\rho-1}) \leq (A(j)^{\rho-1} + B(z)^{\rho-1})(A(i)^{\rho-1} + B(z_0)^{\rho-1}) \\
&\Leftrightarrow \quad (B(z)^{\rho-1} - B(z_0)^{\rho-1})A(j)^{\rho-1} \leq (B(z)^{\rho-1} - B(z_0)^{\rho-1})A(i)^{\rho-1} \\
&\Leftrightarrow \quad A(j)^{\rho-1} \geq A(i)^{\rho-1} \\
&\Leftrightarrow \quad A(j) \leq A(i)
\end{aligned}
$$

(ii) Using equations (13) and (15), if $L(i) = L(j)$, then

$$
\frac{L(i,z)}{L(j,z)} = \Big(\frac{y(i,z)}{y(j,z)}\Big)^\theta \Big(\frac{\bar{u}(i)}{\bar{u}(j)}\Big)^\theta
$$

Given the relative productivity of skill z in locations i and j,

$$
\frac{y(i,z)}{y(j,z)} = \Big[\frac{A(i)^{\rho-1} + B(z)^{\rho-1}}{A(j)^{\rho-1} + B(z)^{\rho-1}}\Big]^{\frac{1}{\rho-1}},
$$

it is straightforward to check that $\frac{L(i,z)}{L(j,z)}$ is increasing in z if and only if $A(i) > A(j)$.

(iii) By way of contradiction suppose $A(i) < A(j)$. Then,

$$
a(i,z) = \frac{A(i)^{\rho-1}}{A(i)^{\rho-1} + B(z)^{\rho-1}} > \frac{A(j)^{\rho-1}}{A(j)^{\rho-1} + B(z)^{\rho-1}} = a(j,z)
$$

Since $L(i) > L(j)$, $\ell(i,z) = \ell(j,z)$, and $a(i,z) > a(j,z)$,

$$
A(i)^{1/\gamma} = L(i)\int \ell(i,z)a(i,z)dz > L(j)\int \ell(j,z)a(j,z)dz = A(j)^{1/\gamma} \quad \Rightarrow \quad A(i) > A(j)
$$

which is a contradiction.

Proposition 3. Given $L(i) = \bar{L}\Big(w(i)u(i)\Big)^\theta \Big/ W^\theta$, amenities $u(i) = \bar{u}(i)L(i)^{-\beta}$, and wages under constant scale elasticity $w(i) = \bar{y}(i)L(i)^\alpha$, equilibrium labor allocation is given by

$$
L(i) = \bar{L}\frac{\Big(\bar{x}(i)L(i)^{\alpha-\beta}\Big)^\theta}{W^\theta}
$$

By fully differentiating the equilibrium labor allocation,

$$dL(i) = \frac{dW - \bar{L}^{1/\theta} L(i)^{\lambda} d\bar{x}(i)}{\lambda \bar{L}^{1/\theta} \bar{x}(i) L(i)^{\lambda-1}}$$

where $\lambda = (\theta(\alpha - \beta) - 1)/\theta$. Since $\sum_i dL(i) = 0$,

$$\sum_i \frac{L(i)}{\lambda} \frac{dW}{W} - \frac{L(i)}{\lambda} \frac{d\bar{x}(i)}{\bar{x}(i)} = 0$$

which implies that

$$\frac{dW}{W} = \sum_i \frac{L(i)}{\bar{L}} \frac{d\bar{x}(i)}{\bar{x}(i)}$$

Appendix D Extensions

I present an extended version of the model by allowing for housing, costly trade between cities, finite substitution across tradeable goods, and finite substitution across workers of different skills. The assumptions and notation are otherwise the same, and so, I do not repeat them. For completeness, however, I present both old and new equations.

D.1 Labor Demand

The following three equations are the same as equations (6), (7), and (8) describing $A(i)$, $a(i,z)$, and $y(i,z)$,

$$A(i) = \left[\int_{z \in \Omega_z} L(i,z) a(i,z)\ dz \right]^{\gamma} \tag{A.2}$$

$$a(i,z) = \frac{A(i)^{\rho-1}}{A(i)^{\rho-1} + B(z)^{\rho-1}} \tag{A.3}$$

$$y(i,z) = \left[A(i)^{\rho-1} + B(z)^{\rho-1} \right]^{\frac{1}{\rho-1}} \tag{A.4}$$

I replace equation (9) with the following CES production function,

$$Q(i) = \bar{y}(i) \left[\int_z \Big(y(i,z) L(i,z) \Big)^{\frac{\varepsilon-1}{\varepsilon}} dz \right]^{\frac{\varepsilon}{\varepsilon-1}} \tag{A.5}$$

where $\varepsilon > 0$ is the elasticity of substitution across workers of different skills. Unit cost of production equals

$$\frac{c(i)}{\bar{y}(i)}, \quad \text{where} \quad c(i) = \left[\int_z \Big(\frac{w(i,z)}{y(i,z)} \Big)^{1-\varepsilon} dz \right]^{\frac{1}{1-\varepsilon}}$$

Let $d(i,j)$ be the iceberg trade cost of shipping a good from i to j. Price of a good produced in location i and consumed in location j equals the marginal cost of producing a good in i and shipping it to j,

$$p(i,j) = \frac{c(i)d(i,j)}{\bar{y}(i)} \tag{A.6}$$

Share of spending on skill z in location i, denoted by $b(i,z)$, is

$$b(i,z) \equiv \frac{w(i,z)L(i,z)}{\int w(i,z)L(i,z)dz} = \Big(\frac{w(i,z)}{c(i)y(i,z)}\Big)^{1-\varepsilon} \tag{A.7}$$

Holding location i fixed, demand for skill z relative to demand for skill z' is

$$\frac{L(i,z)}{L(i,z')} = \Big(\frac{w(i,z)}{w(i,z')}\Big)^{-\varepsilon}\Big(\frac{y(i,z)}{y(i,z')}\Big)^{\varepsilon-1} \tag{A.8}$$

D.2 Labor Supply

The utility of a worker ω with skill z in location i is a Cobb-Douglas combination of a bundle of tradeable goods, $T(i,z)$, and nontradeable housing, $H(i,z)$, augmented with utility from local amenities, $u(i)$ and a location preference shock $\zeta_\omega(i,z)$,

$$\Big(\frac{T(i,z)}{\delta}\Big)^{\delta}\Big(\frac{H(i,z)}{1-\delta}\Big)^{1-\delta} u(i)\zeta_\omega(i,z)$$

$\delta \in (0,1)$ is the share of expenditures on tradeables, and ζ is i.i.d. draws from a Fréchet distribution with location parameter normalized to one, dispersion parameter θ. The tradeable goods are differentiated by the location of production. The bundle $T(i,z)$ aggregates quantities of consumption for a worker of skill z in location i from goods produced in j, $t(j,i,z)$, under a constant elasticity of substitution $\sigma > 0$,

$$T(i,z) = \Bigg[\sum_j t(j,i,z)^{\frac{\sigma-1}{\sigma}}\Bigg]^{\frac{\sigma}{\sigma-1}}$$

Local amenities in location i, $u(i)$, is

$$u(i) = \bar{u}(i)L(i)^{\beta} \tag{A.9}$$

where β governs local externalities in consumption (other than housing). A worker with skill z who resides in location i faces the following budget constraint,

$$w(i,z) = R(i)H(i,z) + \sum_j p(j,i)t(j,i,z),$$

where $R(i)$ is price per unit of housing in i, and $p(j,i)$ is price of good j in destination i. Land is owned by immobile landlords who receive housing rents as their only source of income, and like local workers, decides how much of each tradeable goods and housing to consume. The supply of residential land is inelastically given and normalized to one. The land market clearing condition pins down the price per unit of housing,

$$R(i) = \Big(\frac{1-\delta}{\delta}\Big) \int_z w(i,z) L(i,z)\ dz, \tag{A.10}$$

Total income (wages plus housing rents) equals $\frac{1}{\delta}\int_z w(i,z)L(i,z)\ dz$. Both workers and landlords spend δ share of their income on tradeables and the rest on housing. Thus, location i's total spending on tradeables equals $\int_z w(i,z)L(i,z)\ dz$, and aggregate consumption of location j on goods produced in i, denoted by $X(i,j)$, is

$$X(i,j) = \Big(\frac{p(i,j)}{P(j)}\Big)^{1-\sigma} \int_z w(j,z)L(j,z)\ dz \tag{A.11}$$

where $p(i,j)$ is given by equation (A.6) and $P(j)$ is the CES price index of tradeables,

$$P(j) = \Bigg[\sum_i p(i,j)^{1-\sigma}\Bigg]^{\frac{1}{1-\sigma}} \tag{A.12}$$

The discrete choices of workers imply the expected welfare of workers of skill z, denoted by $W(z)$,

$$W(z) = \bar{\gamma}\Bigg[\sum_i \Big(\frac{w(i,z)}{P(i)^{\delta}R(i)^{1-\delta}}u(i)\Big)^{\theta}\Bigg]^{\frac{1}{\theta}} \tag{A.13}$$

where $\bar{\gamma} \equiv \Gamma\Big(1+\frac{1}{\theta}\Big)$. Inside $W(.)$, agglomeration benefits are implicit in wages w, and consumption externalities are captured by amenities u and housing costs R. Holding skill z fixed, share of workers who choose location i is given by

$$v(i,z) \equiv \frac{L(i,z)}{V(z)} = \frac{\Big(\frac{w(i,z)}{P(i)^{\delta}R(i)^{1-\delta}}u(i)\Big)^{\theta}}{(W(z)/\bar{\gamma})^{\theta}} \tag{A.14}$$

Holding location i fixed, supply of skill z relative to supply of skill z' is

$$\frac{L(i,z)}{L(i,z')} = \Big(\frac{V(z)}{V(z')}\Big)\Big(\frac{W(z)}{W(z')}\Big)^{-\theta}\Big(\frac{w(i,z)}{w(i,z')}\Big)^{\theta} \tag{A.15}$$

D.3 Equilibrium

Combining relative labor demand (equation A.8) and relative labor supply (equation A.15) delivers skill wage premium and skill employment ratio as a function of relative productivities,

$$\frac{w(i,z)}{w(i,z')} = \gamma_W(z,z')\Big(\frac{y(i,z)}{y(i,z')}\Big)^{\frac{\varepsilon-1}{\theta+\varepsilon}} \tag{A.16}$$

$$\frac{L(i,z)}{L(i,z')} = \gamma_L(z,z')\Big(\frac{y(i,z)}{y(i,z')}\Big)^{\frac{\theta(\varepsilon-1)}{\theta+\varepsilon}} \tag{A.17}$$

where

$$\gamma_W(z,z') \equiv \Big(\frac{V(z)}{V(z')}\Big)^{-\frac{1}{\theta+\varepsilon}}\Big(\frac{W(z)}{W(z')}\Big)^{\frac{\theta}{\theta+\varepsilon}}$$

$$\gamma_L(z,z') \equiv \Big(\frac{V(z)}{V(z')}\Big)^{\frac{\varepsilon}{\theta+\varepsilon}}\Big(\frac{W(z)}{W(z')}\Big)^{-\frac{\theta\varepsilon}{\theta+\varepsilon}}$$

Labor market clearing requires total wages received by workers to be equal total payments to workers in every location. Using equation (A.11),

$$\int_z w(i,z)L(i,z)\ dz = \sum_j \Big(\frac{p(i,j)}{P(j)}\Big)^{1-\sigma}\int_z w(j,z)L(j,z)\ dz \tag{A.18}$$

Definition. Given $V(z)$, $B(z)$, $\bar{u}(i)$, $\bar{y}(i)$, $d(i,j)$ for all i, j, and z, a spatial equilibrium consists of $w(i,z)$ and $L(i,z)$ for all i and z such that equations (A.2)-(A.18) hold.

Rewriting the Equilibrium Equations Equations (A.2), (A.3), (A.4) define a mapping $y:\mathbf{L}\rightarrow y(\mathbf{L})$, from the distribution of employment across skills $\mathbf{L}=[L(.,z),\ \forall z]$ to productivity $y(\mathbf{L})$ (holding a city fixed). Equation (A.12), and combining (A.14) and (A.18) give two systems in two unknowns $w(i,z)$ and $L(i,z)$,

$$P(j)^{1-\sigma} = \sum_i d(i,j)^{1-\sigma}c(i)^{1-\sigma}\bar{y}(i)^{\sigma-1} \tag{A.19}$$

$$\int_z w(i,z)L(i,z)\ dz = \int_z\sum_j d(i,j)^{1-\sigma}c(i)^{1-\sigma}\bar{y}(i)^{\sigma-1}P(j)^{\sigma-1}w(j,z)L(j,z)\ dz \tag{A.20}$$

where $c(i)$, $P(i)$, and $R(i)$ are themselves functions of $w(i,z)$ and $L(i,z)$,

$$c(i) = \left[\int_z \left[\frac{w(i,z)}{y\Big(\mathbf{L}(i)\Big)}\right]^{1-\varepsilon} dz\right]^{\frac{1}{1-\varepsilon}}, \qquad \mathbf{L}(i) \equiv \Big[L(i,z),\ \forall z\Big] \tag{A.21}$$

$$P(i) = \bar{\gamma}^{\frac{-1}{\delta}} V(z_0)^{\frac{1}{\delta\theta}} W(z_0)^{\frac{-1}{\delta}} \bar{u}(i)^{\frac{1}{\delta}} L(i)^{\frac{\gamma}{\delta}} R(i)^{\frac{\delta-1}{\delta}} w(i,z_0)^{\frac{-1}{\delta}} L(i,z_0)^{\frac{-1}{\delta\theta}} \tag{A.22}$$

$$R(i) = \Big(\frac{1-\delta}{\delta}\Big) \int_z w(i,z)L(i,z)\ dz \tag{A.23}$$

An equilibrium is a solution $w(i,z)$ and $L(i,z)$, $\forall i,z$, to the two systems of equations (A.19)-(A.20), given A.21, A.22, A.23, and the mapping from $\mathbf{L}(i)$ to $y(\mathbf{L}(i))$.

Special Cases. The extended model collapses to the model in Section 4 when

$$\varepsilon \to \infty,\ \sigma \to \infty,\ d(i,j) = 1,\ \delta \to 1$$

That is, when there are zero shipping costs and final goods are perfectly substitutable, workers are perfectly substitutable, and housing is not explicitly modeled. In particular, the combined externality in local consumption is given by $\frac{\beta+\delta-1}{\delta}$. In addition, the extended model nests many other frameworks in the literature. For example, it nests Allen and Arkolakis (2014) when

$$\rho \to 1,\ B(z) = \text{constant},\ \varepsilon \to \infty,\ \delta \to 1$$

As Proposition 1 shows when $\rho = 1$, then agglomeration elasticity is constant, and as $B(z)$ is a constant, there will be no skill heterogeneity.

Existence and Uniqueness in a Simplified Version of the Model. Consider a version of the model with homogeneous workers who have Cobb-Douglas preferences over the final good and housing with δ as the share of the former; final goods are perfectly substitutable and there are no shipping costs,

$$\varepsilon \to \infty,\ \sigma \to \infty,\ d(i,j) = 1,\ \rho \to 1,\ B(z) = \text{constant}$$

Rental price of housing is

$$r(i) = \frac{1-\delta}{\delta} w(i)L(i)$$

Since $p(i) = 1$, due to perfect competition and external returns to scale, labor demand implies that

$$w(i) = \bar{y}(i)y(L(i))$$

where $\bar{y}(i)$ is exogenous productivity shifter, and $y(L(i))$ governs the relationship between productivity and local population. Labor supply to location i is given by

$$L(i) = \bar{L}\frac{\Big(w(i)R(i)^{\delta-1}u(i)\Big)^{\theta}}{\sum_j \Big(w(j)R(j)^{\delta-1}u(j)\Big)^{\theta}}$$

Proposition D.1. *If $y(L(i)) = L(i)^{\alpha}$, then equilibrium is stable and unique if $\alpha < \frac{1-\delta-\beta}{\delta}$.*

Proof. Given $u(i) = \bar{u}(i)L(i)^{\beta}$, $w(i) = \bar{y}(i)L(i)^{\alpha}$, and replacing for $R(i)$ from above,

$$\begin{aligned}
\frac{1}{\bar{L}}\sum_j \Big(w(j)\Big(\frac{1-\delta}{\delta}\Big)^{\delta-1} w(j)^{\delta-1}L(j)^{\delta-1}\bar{u}(j)L(j)^{\beta}\Big)^{\theta} &= \Big(w(i)\Big(\frac{1-\delta}{\delta}\Big)^{\delta-1} w(i)^{\delta-1}L(i)^{\delta-1}\bar{u}(i)L(i)^{\beta}\Big)^{\theta} L(i)^{-1} \\
\frac{1}{\bar{L}}\sum_j \Big(\frac{1-\delta}{\delta}\Big)^{\theta(\delta-1)} \bar{u}(j)^{\theta}w(j)^{\theta\delta}L(j)^{\theta(\beta+\delta-1)} &= \Big(\frac{1-\delta}{\delta}\Big)^{\theta(\delta-1)} \bar{u}(i)^{\theta}w(i)^{\theta\delta}L(i)^{\theta(\beta+\delta-1)-1}
\end{aligned}$$

Replacing $w(i) = \bar{y}(i)L(i)^{\alpha}$,

$$L(i)^{\lambda_1} = \bar{u}(i)^{-\theta}\bar{y}(i)^{-\theta\alpha}\bar{L}^{-1}\sum_j \bar{u}(j)^{\theta}\bar{y}(j)^{\theta\alpha}\Big(L(j)^{\lambda_1}\Big)^{\frac{\lambda_2}{\lambda_1}}$$

where

$$\lambda_1 \equiv \theta(\alpha\delta + \beta + (\delta-1)) - 1$$

$$\lambda_2 \equiv \theta(\alpha\delta + \beta + (\delta-1))$$

Equilibrium is unique if $\frac{\lambda_2}{\lambda_1} \in [-1, 1]$. The equilibrium is stable and unique if $0 > \lambda_2 > \lambda_1$, or equivalently if

$$\alpha < \frac{1-\delta-\beta}{\delta}$$

When share of housing $1-\delta = 0$, this condition reduces to $\alpha < -\beta$.