Wage Inequality and the Location of Cities

Farid Farrokhi[1] and David Jinkins[*2]

[1]Purdue
[2]Copenhagen Business School

January 30, 2019

Abstract

We document that isolated cities have less wage inequality in American census data. To explain this correlation and other correlations between population and wages, we build an equilibrium empirical model that incorporates high and low-skill labor, costly trade, and both agglomeration and congestion forces. Our paper bridges the gap between the economic geography literature which abstracts from inequality, and the spatial inequality literature which abstracts from geography. We find that geographical location explains 16.5% of observed variation in wage inequality across American cities. We use our model to simulate counterfactual trade and technology shocks. Reductions in domestic trade costs benefit both skill groups but low-skill workers benefit more.

1 Introduction

Inequality has long fascinated economists, and growing income inequality has been recently and heatedly discussed in public forums.[1] This public discussion has been complemented by a number of academic studies highlighting the spatial distribution of wage inequality. We have learned that there is a strong and increasing positive relationship between wage inequality and city size (Baum-Snow and Pavan, 2012; Moretti, 2013; Lindley and Machin, 2014), and that high and low-skill workers are increasingly segregated across cities (Diamond, 2015). In this paper, we add a further attribute of a city to this discussion: spatial position. We first document the relationship between inequality and geography in the American data. Then we build and estimate an equilibrium model to measure the importance of geography for wage inequality, and to study the effects of trade and productivity shocks on welfare and inequality.

Using American census data, we show that geographical location has significant power in explaining observed skill wage premia. This result holds across a wide variety of specifications and weighting strategies. In a word, the closer a city is to the ocean and the nearer it is to other cities, the more unequal it tends to be.[2] For example, Minneapolis is around one standard deviation more isolated than Miami, and has wage inequality around two standard deviations lower than Miami. In order to explain this correlation together with previously documented facts on population and wages, we develop an estimable equilibrium model of domestic trade and inequality.

While our research speaks to several literatures, our primary contribution is in developing an estimable equilibrium model of spatial wage inequality in which *geography matters*. Following and contributing to the popular debate on inequality, several authors have expanded our understanding of wage and welfare inequality in American data (Baum-Snow and Pavan, 2012; Combes et al., 2012b; Moretti, 2013; Davis and Dingel, 2014; Diamond, 2015; Farrokhi, 2018). As a shorthand, we refer to these papers as the spatial inequality literature. To date, the spatial inequality literature has abstracted from geography. Either cities are unable to trade with each other, or able to trade with each other costlessly. In both of these extremes, the geographic location of a city relative to other cities is irrelevant, so questions about the interaction of geography with inequality cannot be addressed. By including costly trade between cities in a model of mobile heterogeneous labor, we can measure the contribution of geography to inequality.

In order to solve an equilibrium model of inequality, we use tools recently introduced to the economic geography literature by Allen and Arkolakis (2014). We follow a growing body of literature estimating structural economic geography models to evaluate the effects of economic policy on migration and welfare (Bartelme, 2015; Desmet et al., 2016; Allen et al., 2016). The economic geography literature as a whole has typically focused on welfare at the aggregate

---

[1] The literature on the causes of the rise in American wage inequality in the United States is large. For an extensive treatment, see Goldin and Katz (2009). There is also a growing body of literature on consequences of inequality. For example some studies link income inequality to the recent rise of populism in the United States (McCarty et al., 2016), others to adverse health outcomes (Wilkinson and Pickett, 2006).

[2] These concepts will be defined precisely in Section 2.2.

(Krugman, 1991b; Fujita et al., 2001; Fajgelbaum et al., 2015; Monte et al., 2015).[3] We complement this literature by studying the effects of policy not only on aggregate welfare but also on welfare inequality.

Our modeling approach allows us to fully solve for counterfactual outcomes taking general equilibrium effects into account. In contrast, the spatial inequality literature has often used equilibrium models without solving for equilibrium. Recent spatial inequality contributions employ instrumental variables and equilibrium relationships to identify a handful of parameters of interest (Moretti, 2013; Baum-Snow et al., 2014). This methodology is sufficient to test alternative hypotheses about sources of inequality, but it limits a researcher's ability to run counterfactual policy experiments. The closest paper in this recent literature to ours is Diamond (2015), who estimates a rich structural spatial inequality model based on discrete choices of workers over where to live. While Diamond (2015) allows for a more flexible specification, we adopt a more stylized model. The advantage of our stylized model is that we can fully solve our model for equilibrium population, wage, and inequality levels at a wide range of counterfactual parameter values.

In our model, we have a continuum of locations. In each location, there are immobile landlords, immobile firms, and perfectly mobile workers. Workers come in two types, high-skill and low-skill, and each worker has an idiosyncratic utility from living in each location. A worker decides where to live taking prices and wages as given. A firm also takes local wages as given, and produces a tradeable good using high-skill and low-skill labor as inputs. The key difference between high and low-skill workers is that high-skill workers benefit more from agglomeration.[4] In equilibrium, welfare of marginal workers in each skill group equalizes across space.

We require a model that generates higher skill wage premia in less remote cities. The interplay between two critical features of our model deliver the required relationship. These two features are stronger agglomeration forces for high-skill workers, and heterogeneous location preferences. The intuition behind this interaction can be described in a few sentences. Consider a city near other cities, a centrally-located city. Its access to cheap tradeable goods and nearby markets make this city attractive to live in. This leads the city, all else equal, to have a relatively high population of both high and low-skill workers compared with a remote city. Due to agglomeration forces, high-skill workers are relatively more productive in the centrally-located city. If the ratio of high to low-skill wages in the centrally-located city were the same as in the remote city, firms would demand a larger ratio of high to low-skill workers in the centrally-located city. In order for the demand for high-skill labor to equal its supply, in equilibrium the high to low-skill wage ratio must be higher in the centrally-located city. Because location preferences matter, high-skill workers elsewhere do not fully arbitrage the higher wages in the

---

[3]One notable exception is Fujita and Thisse (2006), which focuses on inequality and costly trade in an international trade context with only two regions and only high-skill workers mobile. In addition, Fajgelbaum and Gaubert (2018) characterize optimal spatial policy in a setting with trade and skill groups where they compare the observed concentration of skill compared to the efficient outcome in the United States.

[4]Davis and Dingel (2019) microfound a mechanism for this assumption related to complementary between idea exchange and ability.

3

centrally-located city away.

We interpret American census data in 2000 as the equilibrium outcome of our model, and Core Based Statistical Areas as our cities or geographical units of observation. We estimate our model parameters using equilibrium relationships that describe labor supply and demand across these cities. In addition, we estimate costs of trading goods between cities in a similar way to Allen and Arkolakis (2014).

We use our estimated model to perform several quantitative exercises. First, we use our estimated amenities, productivities, and trade costs to decompose sources of variation in observed wage premia across American cities. We find that geographical position explains 16.5% of the variation in wage premia across cities.

In addition, we simulate the model equilibrium when domestic trade costs change. We find that reductions in domestic trade costs benefit both types of labor, but low-skill labor gains more than high-skill labor. This result is in contrast to a number of papers that study the effects of international trade on inequality (Antràs et al., 2006; Hummels et al., 2014).[5] In our exercise, better trading infrastructure tends to spread out the population in the United States so that high-skill workers lose some of their agglomeration advantage over low-skill workers. The negative effect of trade on wage inequality in the international context is reversed when labor is mobile in the presence of agglomeration economies in the national context.[6]

Lastly, we simulate the equilibrium effects of the rise of Silicon Valley by implementing a counterfactual productivity shock to all cities in California such that our model generates actual changes to the share of high and low-skill population in California between 1980 to 2000. We find that this skill-biased technology shock increased the expected welfare of high-skill workers nationally by 1.3% and of low-skill workers nationally by 0.3%.

2   Documenting inequality and geography

In this section, we describe our data sources, give our definitions of measures of geography and inequality, and present the empirical findings which motivate our modeling exercise.

2.1   Data sources

Our empirical section is largely based on the IPUMS 5% sample of the 2000 American census. In this cut of the data, we use full-time workers older than 24 and younger than 55 with reported income.

---

[5]A large body of research in international trade has focused on the effect of trade on inequality. The traditional result is the Stolper-Samuelson Theorem, which says that trade increases inequality in countries abundant in high-skill labor, and decreases inequality in countries abundant in low-skill labor (Davis and Mishra, 2007). Of course, an important part of trade models is the inability of factors to cross borders, so the analogy between our work and the trade literature should not be taken too far.

[6]In recent work Fan (2019) finds that domestic reallocation of labor tends to mitigate the increase in inequality caused by an international trade liberalization.

We want to compare inequality in different locations. As agglomeration will be an important component of our model, the size of a location will be critical for our analysis. Different authors in the literature have used different regions as units of analysis. For our purposes, a location will be either a Core Based Statistical Area (CBSA) or the non-CBSA part of a census area known as a Public Use Microdata Area (PUMA). As shorthand, we will sometimes refer to these areas as "cities". A CBSA is a set of counties with a high degree of social and economic ties to a central urbanized area as measured by commuting ties (Census, 2012). PUMAs are drawn to completely cover the United States. In order to comply with census disclosure rules, each PUMA contains between 100,000 and 300,000 residents. By including the non-CBSA parts of PUMAs in our analysis, we widen the scope of our study to the entire continental United States.

In addition to the IPUMs data, we need information on the geographical position of each location as well as information on trade flows between locations. We use geographical position data from the Missouri Census Data Center. For trade flows we use publicly available data from the U.S. Commodity Flow Survey (CFS). Our data on trade flows is from 2007, as this is the first year in which data is available at the required level of disaggregation. The 2007 CFS covers business establishments with paid employees in mining, manufacturing, wholesale trade, and selected retail and services trade industries. In the survey, a total sample size of approximately 102,000 establishments are selected from a universe of 754,000 establishments. For further discussion of data sources and manipulation, see Appendix A.

2.2 Location specific variables

We calculate the ratio of high-skill to low-skill wages across cities. Skill is unobservable, so we follow the literature in comparing the wages of workers with a 4-year college degree to workers with only a high-school degree. We calculate this skill wage premium using the method described in Acemoglu and Autor (2011). This method flexibly controls for differences in the distribution of experience and gender across cities. First we calculate the total hours worked nationally by gender and five levels of experience (up to 9 years, 10-19 years, 20-29 years, 30-39 years, and 40 and above). We then regress log wages of full-time workers separately in each city on a gender dummy, a set of education dummies, race dummies, a quadratic in experience, and all interactions between the education dummies, gender dummy, and experience. Using our regression estimates, we predict wages in each city for only whites in cells of gender, experience, and education level. We then use the midpoint for each experience group to construct wage predictions for each cell. For example, we create a wage prediction in Atlanta for a white high school educated woman with 5 years of experience, for a white college educated man with 15 years of experience, and so on. We then use the national hours worked by each gender and experience cell as weights along with our predicted log wages to calculate the average log wage by city for each education level. Finally we subtract our average log wage for high school graduates from the average log wage for 4-year college graduates in each city. This is the skill

wage premium.[7]

Next we assign to each location several measures of isolation from other locations. None of these simple, atheoretical measures is completely satisfying alone. We simply aim to use these measures of isolation to build a case that geography matters, and motivate our subsequent structural analysis rooted in preferences, production technology, and trade costs. To this end, we assign each location two primary measures of isolation: distance to ocean and remoteness.[8] We measure a location's domestic isolation using remoteness, a concept we borrow from the international trade literature (Head, 2003). Each location is labeled with a number $i = 1 \ldots N$. The distance between location i and location j is d_{ij}. The distance we use here is structurally estimated later in this paper, and captures the iceberg trade cost between every pair of locations given the network of transportation infrastructure in the United States. The remoteness of location i, Rem_i, is the weighted, generalized mean of the distances between location i and all other locations:

$$Rem_i = \left(\sum_j \omega_j d_{ij}^{1-\sigma} \right)^{\frac{1}{1-\sigma}}$$

In words, a location with low transport costs to other locations will have low remoteness. In a standard trade model with a CES demand system, the price index of tradeable goods in location i follows a similar expression, with weights ω_i related to economic size, and σ interpretable as the elasticity of substitution in utility of differentiated goods. We set ω_j to the population of location j, and set $\sigma = 4$, following recent estimates in the international trade literature (Broda and Weinstein, 2006; Simonovska and Waugh, 2014).

In addition, we use distance from coastlines to proxy for a location's isolation from international trade.[9] We measure nearest distance to the ocean as the crow flies using data from *Natural Earth*.[10] This data comes at a very fine level of disaggregation. To aggregate up to the level of our locations, we assign each location the mean distance to the ocean within its borders.[11]

Table 1 reports descriptive statistics, and Figure 1 shows how our measures vary across the United States. The borders in this map are the intersection of PUMAs and CBSAs, but are colored based on the geographical unit of analysis described in Section 2.1. Remoteness is highest in the North and North-West of the United States. Distance from the ocean is highest

---

[7]Our results are robust to alternative measures of wages such as a direct measure of mean log hourly wage as well as log wage residuals obtained from regressing log hourly wage against gender, experience, and race. Our results are also robust to defining a high-skill worker as a worker with 4 years of college or more, and low-skill as all other workers.

[8]To be clear here about terminology, remoteness is one specific measure of isolation. By isolation, we mean aggregated distance from economically important locations.

[9]Coşar and Fajgelbaum (2016) show how distance from the ocean affects trade patterns within a country.

[10]*Natural Earth* is a free source of physical geographical data in the public domain maintained by the North American Cartographic Information Society. More information at naturalearthdata.com .

[11]Our data comes projected in spherical coordinates. For ease of interpretation, we convert our spherical distances to approximate kilometers using the rule of thumb that one spherical degree in the United States is approximately equal to 100 km. All of our analysis is in logarithms, so scaling errors will only affect the constant.

in the Center-North of the United States.[12] The skill wage premium is higher in the parts of the country which are less isolated and have higher population.

Below, we report correlations between our reduced-form measures of isolation and skill premium to motivate that geography matters for inequality. In Section 5, we will use a structurally-estimated price index to quantify precisely how much of the observed variations in skill premium can be explained by geography.

Statistic	Mean	Std Dev	Min	Max
Distance from coast	533	346	0.39	1586
Remoteness	1.30	0.18	0.91	1.89
Population	57.48 k	223.96 k	410	4.75 m
College wage premium	1.40	0.09	1.15	1.83
Worker-level observations	3.06 m			
Location-level observations	1267			

Notes: Following Acemoglu and Autor college wage premium is calculated based on predicted wages, and we restrict our sample to high-school graduates and college graduates.

Table 1: Data summary statistics

---

[12]State borders can sometimes be seen in our measures of remoteness and distance from the coast. This is because one of our measures of geographical location, the PUMA, is always contained within a state, and our other measure, the MSA, is made up of counties which are always contained within a state. Thus the areas we are averaging over often end at state borders. Continuous variation appears, therefore, to stop at the borders of states. For example, the border of Montana and the Dakotas can be clearly seen in the distance from the coast map. This is because Montana is on average closer to the ocean than the Dakotas.

(a) Remoteness

(b) Distance from Coast

(c) Population
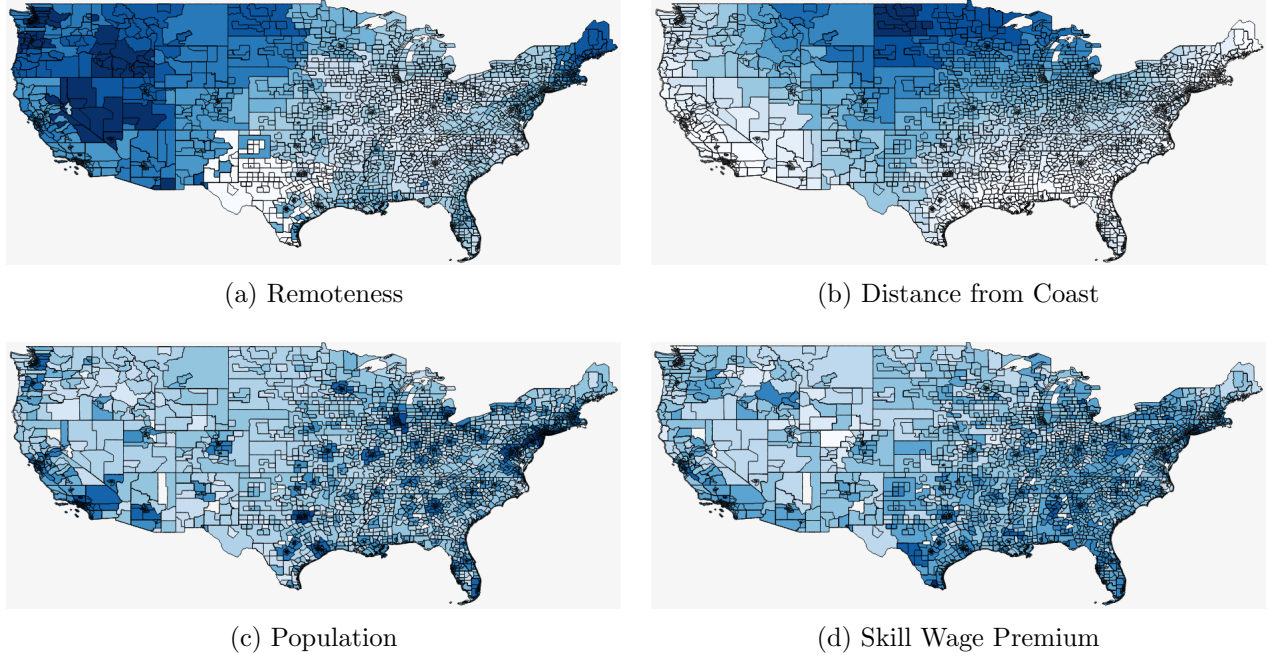
(d) Skill Wage Premium

Figure 1: Locations colored by attribute

2.3 Skill premium and measures of geography

We document the covariance of our measures of geography with the skill premium. The literature has documented that wages, skill premium, and skill ratio are highly and positively correlated with population (Baum-Snow and Pavan, 2012; Moretti, 2013; Lindley and Machin, 2014). In Appendix B we confirm all these well-established facts in our data. Here, we add our measures of geography in standard regressions of individual-level wages and city-level skill premia.

Individual-level observations. Table 2 reports the results from regressing individual-level wages against city-level remoteness allowing for different slopes across high-school and college graduates. We control for individuals' characteristics including gender, race, and years of experience. We include city population or city population of college graduates as well as state fixed effects across our specifications. Three findings stand out. First, remoteness is negatively correlated with wages. Second, controlling for other variables, wages of college graduates compared to high-school graduates fall more with remoteness. This finding suggests that, after controlling for other observed characteristics, college graduates have a higher wage premium in less remote cities. Third, controlling for city population (or city population of college graduates) lowers the magnitude and statistical significance of the remoteness coefficient.

City-level observations. Columns (1)-(4) of Table 3 report estimates from city-level regressions of skill wage premium on our measures of geography. Columns (5)-(6) report estimates of regressions of population measures against remoteness. In Panel A, we do not include state fixed effects and standard errors are not clustered at the state level. In Panel B, we include state fixed effects and standard errors are clustered at the state level. There are benefits and disadvantages of controlling for state fixed effects. The benefit is the ability to control for taxes

and other policies implemented at the level of states. The disadvantage is that by including state fixed effects, our main right-hand-side variable of interest, geography, might not vary sufficiently within states. We weight all regressions by population, because our dependent variable is itself composed of data means. Removing these weights does not affect the signs or statistical significance of our estimates. We report results from additional specifications in Appendix B.

In columns (1)-(4), we find that locations that are more remote within the United States or more distant from coastlines appear to have less wage inequality. This relationship is statistically significant in all regressions in Panel A, but it drops in absolute value, from 0.167 to 0.074, when population is controlled for. In Panel B, with state fixed effects, the remoteness coefficient loses statistical significance when population is included. Results remain the same when we replace population with college population. The effect of geography on the skill wage premium is mitigated when population is controlled for, and in some specifications the effect is not significantly distinguishable from zero.

In columns (5)-(6), we find that remoteness is negatively correlated with population or with college population across cities. This negative correlation together with the positive association between city size and wage premium are consistent with a hypothesis in which remoteness affects wage premium through the population channel.

Overall, these regressions demonstrate that geographic features of cities correlate with wage inequality across a wide range of specifications, but this effect is mitigated or loses its statistical significance when population (or college population) is controlled for.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dependent variable: Log wage of individual workers						
Log remoteness	-0.539***	-0.118***	-0.136***	-0.0663*	-0.0836**	-0.110*	-0.118*
	(0.0947)	(0.0356)	(0.0336)	(0.0375)	(0.0364)	(0.0641)	(0.0619)
College		0.474***	0.472***	0.297***	0.317***	0.300***	0.325***
		(0.00645)	(0.00648)	(0.0361)	(0.0261)	(0.0321)	(0.0242)
College X Log remoteness				-0.117***	-0.114***	-0.113***	-0.112***
				(0.0365)	(0.0345)	(0.0321)	(0.0310)
Log population		0.0500***		0.0456***		0.0380***	
		(0.00262)		(0.00276)		(0.00214)	
College X Log population				0.0155***		0.0151***	
				(0.00259)		(0.00236)	
Log College population			0.0450***		0.0406***		0.0341***
			(0.00218)		(0.00231)		(0.00177)
College X Log College population					0.0157***		0.0148***
					(0.00214)		(0.00203)
R-squared	0.094	0.267	0.268	0.268	0.269	0.276	0.277
State FE	N	N	N	N	N	Y	Y

Notes: Standard errors, clustered at city level, are reported in parentheses. In all regressions, there are 3,050,723 observations, we weight individuals based on census sampling weights, and we include individual-level gender and race dummies, a cubic polynomial of years of experience, and state fixed effects. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 2: Wages and remoteness at the level of individual workers

	(1)	(2)	(3)	(4)	(5)	(6)
		Log college wage premium			Log pop	Log col pop

Panel A: Not controlling for state effects

Log remoteness	-0.191***	-0.167***	-0.0743***	-0.0794***	-6.995***	-7.432***
	(0.0181)	(0.0176)	(0.0254)	(0.0241)	(1.226)	(1.348)
Log dist from coast		-0.0135***	-0.0126***	-0.0127***	-0.0683	-0.0627
		(0.00273)	(0.00299)	(0.00293)	(0.130)	(0.143)
Log population			0.0132***			
			(0.00169)			
Log college population				0.0118***		
				(0.00140)		
R-squared	0.235	0.289	0.414	0.417	0.315	0.285

Panel B: Controlling for state effects

Log remoteness	-0.180***	-0.176***	0.0290	0.0229	-12.08***	-13.18***
	(0.0371)	(0.0388)	(0.0525)	(0.0492)	(1.540)	(1.697)
Log dist from coast		-0.00386	-0.00470	-0.00491	0.0492	0.0695
		(0.00371)	(0.00336)	(0.00332)	(0.192)	(0.217)
Log population			0.0170***			
			(0.00161)			
Log college population				0.0151***		
				(0.00139)		
R-squared	0.449	0.449	0.597	0.602	0.509	0.482

Notes: In all regressions, there are 1267 observations, and we assign population weights to observations. In Panel A, we do not include state fixed effects and robust standard errors are reported in parentheses. In Panel B, we include state fixed effects and standard errors are clustered at the state level.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 3: Skill premium vs geography measures at the level of cities

3 Theory

In the last section, we presented evidence that skill wage premium tends to be lower in more isolated locations. To explain this correlation, we build a model that incorporates high- and low-skill labor, costly trade, and both agglomeration and congestion forces. The model helps us examine the equilibrium responses of inequality to shocks that stem from trade or technology.

3.1 Setup

The model is static, with a continuum of locations $j \in J$, a continuum of high-skill workers labeled as H, and a continuum of low-skill workers labeled as L. The set of locations J, and total population of skill groups, N_L and N_H, are given. Workers can choose to reside and work in any single location. Firms in each location produce a location-specific variety of a tradeable final good using the two types of labor as inputs into a constant elasticity of substitution production function. Each location produces a single tradeable location-specific final good. Consumers cannot perfectly substitute across these location-specific final goods. That is, trade is Armington. Both workers and firms are price takers in perfectly competitive markets.

3.1.1 The worker's problem and labor supply

The utility of worker ω in skill group s in location i is a Cobb-Douglas combination of a bundle of tradeable goods, $Q_\omega(i)$, and residential land use, $Z_\omega(i)$, augmented with utility from local amenities, $\bar{u}_s(i)$, and location preference shocks, $\varepsilon_\omega(i)$,

$$U_\omega(i) = \left(\frac{Q_\omega(i)}{\delta}\right)^\delta \left(\frac{Z_\omega(i)}{1-\delta}\right)^{1-\delta} \bar{u}_s(i)\varepsilon_\omega(i). \tag{1}$$

Here, $\delta \in (0,1)$ is the share of expenditures on tradeables.[13] The tradeable goods are differentiated by the location of production. The bundle $Q(i)$ aggregates quantities of consumption in location i from goods produced in j, $q(j,i)$, under a constant elasticity of substitution $\sigma > 0$,

$$Q(i) = \left[\int_J q(j,i)^{\frac{\sigma-1}{\sigma}}\, dj\right]^{\frac{\sigma}{\sigma-1}}$$

A worker with skill s who resides in location i earns wages $w_s(i)$, and faces the following budget constraint,

$$w_s(i) = C(i)Z(i) + \int_J p(j,i)q(j,i)\, dj, \tag{2}$$

where $C(i)$ is price per unit of housing in i, and $p(j,i)$ is price of good j in destination i. While the system of preferences is homothetic, we capture potential heterogeneity across skill groups by letting them value local amenities differently. The idiosyncratic preference shock, ε, is independent across workers and locations, and follows a Fréchet distribution, $\Pr(\varepsilon \leq x) = \exp(-x^{-\theta})$, where θ governs the dispersion of the location preference shocks.

A worker has two decisions to make. She decides where to live, and how much to consume. Given a choice of location, the second problem is standard. Utility maximization implies that a worker spend δ share of her income on tradeable goods and the rest on housing. A worker of

---

[13]While housing services is usually estimated to be a weak necessity good (Aguiar and Bils (2015) report an income elasticity of 0.92), for simplicity we follow the recent spatial inequality literature in assuming constant expenditure shares (Moretti, 2013; Diamond, 2015).

type s in location i spends $x_s(j,i)$ on goods produced in j,

$$x_s(j,i) = \left[\frac{p(j,i)}{P(i)}\right]^{1-\sigma}\delta w_s(i) \tag{3}$$

where $P(i)$ is the CES price index of tradeables,

$$P(i) = \left[\int_J p(j,i)^{1-\sigma}\,dj\right]^{\frac{1}{1-\sigma}}. \tag{4}$$

Land is owned by immobile landlords who receive housing rents as their income, and like local workers, decides how much of each good and residential land to consume. The supply of residential land, denoted by $\bar{Z}(i)$, is inelastically given. The land market clearing condition immediately pins down the price per unit of housing,

$$C(i) \;\;=\;\; \frac{1-\delta}{\delta\bar{Z}(i)}\Big(n_L(i)w_L(i) + n_H(i)w_H(i)\Big), \tag{5}$$

where $n_s(i)$ denotes the population of skill group s in location i. The price index in location i, combines prices of tradeable goods and of housing, given by $P(i)^\delta C(i)^{1-\delta}$. Total income in location i, equals total wages plus housing rents, given by $\frac{1}{\delta}(n_L(i)w_L(i) + n_H(i)w_H(i))$.

The second decision a worker makes is where to live. A worker ω with skill level s faces the following discrete choice problem of where to reside:

$$\max_{i\in J}\;\frac{w_s(i)}{P(i)^\delta C(i)^{1-\delta}}\bar{u}_s(i)\varepsilon_\omega(i)$$

Using the properties of the Fréchet distribution, the supply of type s labor in location i relative to j is given by:

$$\frac{n_s(i)}{n_s(j)} = \left(\frac{w_s(i)\bar{u}_s(i)/(P(i)^\delta C(i)^{1-\delta})}{w_s(j)\bar{u}_s(j)/(P(j)^\delta C(j)^{1-\delta})}\right)^\theta.$$

The elasticity of relative labor supply to relative wages equals θ. The variance of ε across both workers and locations is decreasing in θ. When θ is large, unobserved location preferences are similar across locations. Thus, small changes to wages, prices, or amenities induce large movements of workers. That is, the supply curve of workers to a location is flat. When θ is small, workers have widely varying preferences over locations, so that large changes in wages, prices, or amenities are necessary to induce movement.

We define the well-being index, denoted by W_s, for population of skill s:

$$W_s \equiv \left[\int_{j\in J}\left(\frac{w_s(j)\bar{u}_s(j)}{P(i)^\delta C(i)^{1-\delta}}\right)^\theta dj\right]^{\frac{1}{\theta}}$$

This index is proportional to the expected welfare of a worker of type s before she draws her

13

location preferences.[14] The share of workers of type s in location i is given by:

$$\frac{n_s(i)}{N_s} = \left(\frac{w_s(i)\bar{u}_s(i)/\left(P(i)^\delta C(i)^{1-\delta}\right)}{W_s} \right)^\theta \tag{6}$$

If a location offers higher wages, better amenities, lower prices of tradeables, and lower housing rents, it will attract more population, with the extent of the relationship governed by θ.

3.1.2 The firm's problem and labor demand

Each location has a measure one of homogeneous firms with a CES production under constant returns to scale,

$$A(i)\left[\beta_H(i)n_H(i)^{\frac{\rho-1}{\rho}} + \beta_L(i)n_L(i)^{\frac{\rho-1}{\rho}} \right]^{\frac{\rho}{\rho-1}},$$

where $A(i)$ is total factor productivity in location i. $\rho > 0$ is the elasticity of substitution between high- and low-skill workers. $\beta_H(i) > 0$ and $\beta_L(i) > 0$ are factor intensities. We incorporate agglomeration forces by distinguishing two sources of productivity externalities. First, we specify total factor productivity as:

$$A(i) = \bar{A}(i)n(i)^\alpha, \tag{7}$$

with $\alpha > 0$. This agglomeration force changes productivity of both low and high-skill workers. A standard Krugman-type economic geography model with monopolistic competition and free entry generates the same relation through endogenous measure of firms, with the exact relation if $\alpha = 1/(1+\sigma)$.[15]

In addition, the empirical literature on urban and labor economics substantiates that agglomeration forces are stronger for high-skill workers.[16] On the theoretical side, the literature explains this fact by modeling spillovers through the exchange of ideas within high-skill workers (Davis and Dingel, 2019). To capture this mechanism in our empirical model, we let skilled worker's productivity covary positively with the population of skilled workers in a location,

$$\begin{aligned} \beta_H(i) &= \bar{\beta}_H(i)n_H(i)^\varphi \\ \beta_L(i) &= \bar{\beta}_L(i) \end{aligned} \tag{8}$$

where $\varphi > 0$ governs the agglomeration advantage that is specific to high-skill workers. By a nor-

---

[14]To get the expected welfare, we must multiply W_s by $\Gamma(1 + \frac{1}{\theta})$ where Γ is the gamma function. This scaling term depends only upon θ, an exogenous preference parameter.

[15]The literature has considered alternate sources of aggregate productivity externalities, for example the sorting of firms as in Gaubert (2014) and Ziv (2017).

[16]For example, Glaeser and Resseger (2010) find that "productivity increases with area population for skilled places, but not for low-skill places," and Bacolod et al. (2009) find that workers with stronger cognitive skills experience stronger agglomeration. See Gould (2007), Matano and Naticchioni (2011), and Combes et al. (2012a) for more details on stronger agglomeration gains for workers with higher skills and wages.

malization, we assign no further agglomeration benefit to low-skill labor. By cost minimization, the unit cost of production equals

$$\frac{\nu(i)}{A(i)}, \text{ where } \nu(i) = \left[\beta_H(i)^\rho w_H(i)^{1-\rho} + \beta_L(i)^\rho w_L(i)^{1-\rho}\right]^{\frac{1}{1-\rho}} \tag{9}$$

The share of spending of producers on high-skill workers, denoted by $b(i)$, is given by

$$b(i) = \frac{\beta_H(i)^\rho w_H(i)^{1-\rho}}{\beta_H(i)^\rho w_H(i)^{1-\rho} + \beta_L(i)^\rho w_L(i)^{1-\rho}} \tag{10}$$

Lastly, as markets are perfectly competitive, price equals marginal cost. Let $d(i,j)$ be the trade cost of shipping a good from i to j. The price of a good produced in location i and consumed in location j is:

$$p(i,j) = \frac{\nu(i)d(i,j)}{A(i)} \tag{11}$$

3.1.3 Spatial equilibrium

On the demand side of the labor market, payments to skilled labor equal firms' spending on skill labor, $w_H(i)n_H(i) = b(i)(w_H(i)n_H(i) + w_L(i)n_L(i))$. Substituting (10) and (8) into this relation gives us the relative labor demand function

$$\frac{n_H(i)}{n_L(i)} = \left(\frac{\bar{\beta}_H(i)}{\bar{\beta}_L(i)}\right)^\rho \left(\frac{w_H(i)}{w_L(i)}\right)^{-\rho} n_H(i)^{\varphi\rho} \tag{12}$$

On the supply side of the labor market, employment shares described by equation (6) imply

$$\frac{n_H(i)}{n_L(i)} = \frac{N_H}{N_L}\left(\frac{W_H}{W_L}\right)^{-\theta} \left(\frac{\bar{u}_H(i)}{\bar{u}_L(i)}\right)^\theta \left(\frac{w_H(i)}{w_L(i)}\right)^\theta \tag{13}$$

A necessary condition for labor market clearing is that skill premia simultaneously satisfy the pairs of relative demand (12) and relative supply (13). Combining, we get:

$$\frac{w_H(i)}{w_L(i)} = \left(\frac{W_H}{W_L}\right)^{\frac{\theta}{\theta+\rho}} \left(\frac{\bar{u}_H(i)}{\bar{u}_L(i)}\right)^{\frac{-\theta}{\theta+\rho}} \left(\frac{N_H}{N_L}\right)^{\frac{-1}{\theta+\rho}} \left(\frac{\bar{\beta}_H(i)}{\bar{\beta}_L(i)}\right)^{\frac{\rho}{\theta+\rho}} n_H(i)^{\frac{\varphi\rho}{\theta+\rho}} \tag{14}$$

Labor market clearing also requires total wages received by all workers to be equal to total payments to them,[17]

$$\frac{w_H(i)n_H(i)}{b(i)} = \int_J \frac{w_H(j)n_H(j)}{b(j)}\left[\frac{p(i,j)}{P(j)}\right]^{1-\sigma} dj \tag{15}$$

---

[17]Total wages in location i equal $\frac{w_H(i)n_H(i)}{b(i)}$, and total income (wages plus rents) equals $\frac{w_H(i)n_H(i)}{\delta b(i)}$. Both workers and landlords spend δ share of their income on tradeables and the rest on housing. Thus, total wages in i equal $\int_J \delta\left[\text{total income in } j\right] \left[\frac{p(i,j)}{P(j)}\right]^{1-\sigma} dj$.

Equations 14 and 15 describe labor market clearing in relative terms and in levels. (Equivalently, equation 15 describes the goods market clearing condition.)

A "spatial equilibrium" consists of $w_H(i)$, $w_L(i)$, $n_H(i)$, and $n_L(i)$ such that: (1) firms optimize their labor demand, (2) workers optimize their labor supply, (3) markets clear, and (4) the labor allocation is feasible.[18] This completes our description of the economy.

3.1.4 Discussion

Suppose we were to shut down preference heterogeneity, $\theta \to \infty$. From (14) we see that skill wage premia will be constant across locations. Alternatively, suppose there is no agglomeration advantage for high-skill workers, $\varphi = 0$. Then skill premia can vary between destinations only due to exogenous differences in tastes and productivities between skill groups. To have equilibria with endogenously varying skill premia, we need both heterogeneity in unobserved location preferences (finite θ), and an agglomeration advantage for high-skill workers ($\varphi > 0$). That is, large cities demand relatively more high-skill workers due to agglomeration, but since unobserved location preferences matter, high-skill workers do not fully arbitrage the wage increase away.[19]

To provide further intuition, suppose trade costs to and from a remote city fall. This shock decreases the price of incoming tradeables, hence the supply of workers to the city rises. In addition, the shock increases outgoing sales, hence labor demand in the city rises. If the employment of low- and high-skill workers increase proportionately, agglomeration advantages will make high-skill workers relatively more productive. That is, firms demand a higher ratio of high to low-skill workers than their relative supply in the city. Equilibrium is restored only by raising skill wage premium in the city.

This relationship between trade costs and inequality does not depend on exogenous differences across skill groups. The exogenous differences are residuals in the relation between skill population ratio and skill wage premium in equations of relative demand (12) and relative supply (13). These residuals reflect factors we do not model such as state and local tax incidence, provision of welfare, and non-labor factor endowments.

Our model implies that the spatial distribution of workers contributes to welfare inequality. Specifically, writing the distribution of low-skill labor as a function of that of high-skill labor,

---

[18]That is simply $\int_J n_H(j)\, dj = N_H$ and $\int_J n_L(j)\, dj = N_L$.

[19]Our model relies on differential agglomeration forces between high and low-skill workers and an upward sloping labor supply curve to generate the observed negative relationship between remoteness and the skill premium. Alternatively, such a relationship could potentially be generated by non-homothetic preferences. In particular, suppose that (poorer) low-skill workers consume a higher share of tradeables and a lower share of housing services. Then, in a spatial equilibrium, low-skill workers would need to be compensated more for living in remote areas, delivering a lower skill premium in remote areas.

Empirically, however, it is *high-skill* workers that consume a higher share of tradeables and a lower share of housing services. As mentioned in footnote 14, studies typically find housing services to be a slight necessity good, and recent research has found that higher income Americans consume a higher share of tradeables (Hummels and Lee, 2017). Since non-homothetic preferences predict the opposite of the observed relationship between remoteness and the skill premium, by assuming homothetic preferences we may be underestimating the strength of the agglomeration advantage of high-skill workers.

and after some algebra, we decompose three forces behind average welfare inequality,

$$\frac{W_H}{W_L} = \underbrace{\left(\frac{N_H}{N_L}\right)^{-\frac{1}{\rho}}}_{\text{aggregate scarcity}} \times \underbrace{(N_H)^{\varphi}}_{\text{aggregate agglom.}} \times \underbrace{\left[\int_J \pi(i)\,di\right]^{-\frac{\theta+\rho}{\theta\rho}}}_{\text{distributional effect}} \tag{16}$$

The first term reflects aggregate scarcity of high- to low-skill workers; the second term represents aggregate agglomeration advantage of high-skill workers; and the last term summarizes dispersion forces.[20] This last term depends on the entire distribution of population which in turn endogenously changes with geography.

3.2 Solving for spatial equilibrium

We characterize model equilibria, normalizing wages of high-skill workers such that $\int_J w_H(j)dj = 1$. Using equations (8), (9), and (10) we can write $\nu(i)$ as a function of employment, input expenditure share, and wages of high-skill workers:

$$\nu(i) = \tilde{\nu}(i)w_H(i), \quad \text{where} \quad \tilde{\nu}(i) \equiv \left[\bar{\beta}_H(i)n_H(i)^{\varphi}\right]^{\frac{\rho}{1-\rho}}\left[b(i)\right]^{\frac{-1}{1-\rho}} \tag{17}$$

In addition, normalizing the land supply to one, we can write housing rents as a function of employment and wages of high-skill workers:

$$C(i) = \tilde{C}(i)w_H(i), \quad \text{where} \quad \tilde{C}(i) \equiv \frac{(1-\delta)n_H(i)}{\delta b(i)} \tag{18}$$

First, replacing the price index of tradeables $P(j)$ from employment share (6) into the goods market clearing condition (15) results in:

$$A(i)^{1-\sigma}\nu(i)^{\sigma-1}n_H(i)w_H(i)b(i)^{-1}$$
$$= W_H^{\frac{1-\sigma}{\delta}}N_H^{\frac{\sigma-1}{\delta\theta}}\int_J d(i,j)^{1-\sigma}\bar{u}_H(j)^{\frac{\sigma-1}{\delta}}C(j)^{\frac{(\sigma-1)(\delta-1)}{\delta}}n_H(j)^{\frac{1-\sigma+\delta\theta}{\delta\theta}}w_H(j)^{\frac{\sigma-1+\delta}{\delta}}b(j)^{-1}\,dj \tag{19}$$

Second, substituting the price index of tradeables $P(j)$ from employment share (6) into the CES price formula (4), results in:

$$\bar{u}_H(i)^{\frac{1-\sigma}{\delta}}C(i)^{\frac{(\sigma-1)(1-\delta)}{\delta}}n_H(i)^{\frac{\sigma-1}{\delta\theta}}w_H(i)^{\frac{1-\sigma}{\delta}} = W_H^{\frac{1-\sigma}{\delta}}N_H^{\frac{\sigma-1}{\delta\theta}}\int_J d(j,i)^{1-\sigma}A(j)^{\sigma-1}\nu(j)^{1-\sigma}\,dj \tag{20}$$

The pair of 19–20 gives us two systems of integral equations. Assuming that trade costs are symmetric, we can reduce the two systems into one using a method from Allen and Arkolakis (2014). If either of integral equations hold along with the following relation, both systems of

---

[20]Specifically, $\pi(i) = \left(\frac{\bar{\beta}_H(i)\bar{u}_H(i)}{\bar{\beta}_L(i)\bar{u}_L(i)}\right)^{\frac{-\theta\rho}{\theta+\rho}}\left(\frac{n_H(i)}{N_H}\right)^{\frac{\theta(1-\rho\varphi)+\rho}{\theta+\rho}}$

integral equations must hold:

$$A(i)^{1-\sigma}\nu(i)^{\sigma-1}n_H(i)w_H(i)b(i)^{-1} = \lambda\bar{u}_H(i)^{\frac{1-\sigma}{\delta}}n_H(i)^{\frac{\sigma-1}{\delta\theta}}w_H(i)^{\frac{1-\sigma}{\delta}}C(i)^{\frac{(\sigma-1)(1-\delta)}{\delta}} \qquad (21)$$

where $\lambda > 0$ is a constant.

Relationship with existing models. Our analysis relates to two styles of spatial models. First, as mentioned earlier we extend a stylized spatial inequality model by incorporating costly trade between cities. Conversely, we extend an empirical model of economic geography by incorporating skill groups. Our model, in particular, nests Allen and Arkolakis (2014) if (i) there is no heterogeneity in location preferences, (ii) there is no agglomeration advantage for high-skill workers, (iii) workers with different skills are perfectly substitutable. That is, if $\varphi = 0$, $\theta = \infty$, and $\rho = \infty$.[21]

Uniqueness. The standard proof of equilibrium uniqueness in the existing related literature in economic geography depends on a specification that allows logarithmic relationships between a subset of endogenous variables (Allen and Arkolakis, 2014). Our model deviates from such logarithmic relationships. For example, instead of $A = \bar{A}n^\gamma$ where one must solve for n, we have $A = \bar{A}(n_L + n_H)^\gamma$ where we must solve for n_L and n_H. Here, the relationship between high-skill population n_H and productivity A is not logarithmic. For this reason, the standard proof in this recent literature can not be directly used in our setting. We can show that for the special case in which our model collapses to Allen and Arkolakis, uniqueness is achieved at our preferred parameter estimates. In addition, we have solved our model at our parameter estimates (to be reported in the next section) using different initial values, and found no evidence of multiplicity.

Solution algorithm. We solve our system of integral equations using an iterative method. A feature of our model is that given exogenous parameters, every endogenous variable can be written as a function of $n_H(i)$. Using this feature, our solution algorithm updates our guess for population of high-skill workers $n_H(i)$ in each iteration. In checking existence and uniqueness we confirm that these iterations converge to one solution for a wide variety of initial guesses. In Appendix D we describe our solution algorithm in detail.

4   Estimation

In this section we estimate our structural model. Our data consists of four vectors: high and low-skill populations in each location, and high and low-skill wages in each location. Using our model structure, we invert these four vectors of data to recover four vectors of exogenous shifters: high-skill factor intensity $\bar{\beta}_H$ (with $\bar{\beta}_L = 1 - \bar{\beta}_H$), total factor productivity shifter $\bar{A}(i)$, and amenity values to low and high-skill workers $\bar{u}_L(i)$ and $\bar{u}_H(i)$.

The inversion of the data into these exogenous shifters depend on the matrix of trade costs as

---

[21]The way we model congestion is a little different than in Allen and Arkolakis (2014), but our models are isomorphic once the conditions (i), (ii), and (iii) are fulfilled. We interpret the source of congestion as limited land for housing. Allen and Arkolakis are agnostic about the source of congestion, only assuming that amenities are reduced by population.

well as six key parameters: (i) high-skill agglomeration advantage φ, (ii) elasticity of substitution across skill groups ρ, (iii) labor supply elasticity θ, (iv) common agglomeration parameter α, (v) share of expenditures on housing $1 - \delta$, and (vi) elasticity of substitution across goods σ. We estimate trade costs between American cities in a similar way to Allen and Arkolakis (2014). We calculate housing share, $1 - \delta = 0.355$, based on the Consumer Expenditure Survey 2000.[22] We set the elasticity of substitution across goods $\sigma = 4$, in line with the empirical literature using international trade data (Broda and Weinstein, 2006; Simonovska and Waugh, 2014). Following a large literature, we use instrumental variables and equilibrium relationships to estimate the other four parameters (Moretti, 2013; Desmet et al., 2016; Allen et al., 2016).

Since our estimation procedure contains several sequential steps, we present intermediate results directly after we describe intermediate estimation steps. Trade costs are estimated first. Next key elasticities are estimated from equilibrium labor demand and supply relationships. We then invert a set of equilibrium integral equations to recover exogenous location-specific productivities and amenities.

4.1 Estimation of trade costs

In many countries, the largest cities are on coastlines or near major rivers. The United States is no exception, with the East and West coasts containing the majority of the population. If domestic trade costs were simply quadratic in distance, then Lebanon, Kansas (with population size of 218) would be the center of gravity in the continental United States. A wide range of geographical features in addition to distance affect the cost of trading between any two locations. It is often easier to go around a mountain even if the geodesic between two locations goes through one. New York and Miami are about as far apart as New York and Lebanon, Kansas, but shipping a container to Miami is cheaper because of the possibility of using a ship. To capture these nontrivial features of geography, we estimate trade costs by using a method from Allen and Arkolakis (2014) which takes geographic features into account.

We provide a short overview here, with more details contained in the original Allen and Arkolakis paper. There are three steps to the estimation process. In the first step, we use three separate image files each containing a map of the United States. On one of the maps is the road network, on the second is the railway network, and on the last is the waterway network.[23] We consider four possible methods for moving goods: road, rail, water, and air. For each of these methods separately, we assign a cost of traveling over each pixel of the relevant image file. Then, we calculate the lowest possible cost of using each method to move goods between all pairs of locations using a fast marching algorithm.

---

[22]Specifically, housing expenditures consist of (i) shelter, (ii) utilities, fuels, and public service, (iii) household operations, (iv) housekeeping supplies, and (v) house-furnishings and equipment. We exclude personal insurance and pensions from total expenditures. Share of housing is 0.40 in Monte et al, 0.42 in Moretti and Diamond, 0.19-0.25 in Allen and Arkolakis.

[23]Following Allen and Arkolakis (2014), we take the road, rail, and water shipping network of the United States as fixed in our counterfactuals. We think of our counterfactuals as pertaining to the medium-run. That is, labor is mobile, but basic productivity, amenities, and the broad outline of the transportation network is fixed.

After we finish the first step, we know how much it costs to move goods on the road between two locations, but only in terms of the units we assigned to road travel. We cannot compare the cost of road travel to the cost of water transport because we do not know the exchange rate, as it were, of road travel to water transport. The second step is to use a discrete choice framework and data on trade flows via each mode between each pair of locations in order to back out these cost ratios. Shippers have idiosyncratic, extreme value distributed costs for each mode of transportation. If a large share of transport is via road, then it must be that road is on average a cheaper mode of transportation.

The discrete choice model will only give us the cost ratio between any two modes of transportation, but we still need to pin down the level of costs. To do so, we use the gravity specification implied by our model. Consistent with our later structural estimation, we set the elasticity of substitution across goods equal to four. Estimating the gravity equation gives the scale of trade costs. With the scaling parameter in hand, we can then calculate expected trade costs between every pairs of locations.

Our estimates for trade costs are summarized in Table 4. Road, by normalization, has no fixed cost, and according to the estimation, has a mid-level marginal cost. Rail has a significant fixed cost, but lower marginal cost than road transport. Water has both high fixed and marginal cost, reflecting that little shipment within the United States is done by water. Air has a high fixed cost, but a low marginal cost. To be more concrete, we estimate that the average iceberg cost of shipping from Chicago to New York City and the average cost of shipping from Chicago to Fargo, Minnesota are almost the same (1.27 and 1.26 respectively). Even though Chicago is closer to Fargo (569 miles) than it is to New York City (714 miles) as the bird flies, the highway system connecting Chicago to New York City is both higher quality and more direct.

Readers familiar with Allen and Arkolakis (2014) or Desmet et al. (2016) will notice that our estimates are quantitatively somewhat different than those of these earlier studies, although the ranking of variable and fixed costs is similar. One reason for the difference is that we set a lower trade elasticity in our structural estimation, $\sigma = 4$ rather than $\sigma = 9$.[24] Our trade costs are likely higher in absolute terms than in Allen and Arkolakis (2014), as our products are more differentiated.[25]

---

[24]Regarding the difference between our estimates and those in Allen and Arkolakis (2014), even if we use $\sigma = 9$ we get somewhat different results, even though we implement the same algorithm on the same data. We discuss reasons for these differences in Appendix C.

[25]A further technical issue is that 4.7% of our iceberg trade costs are estimated to be less than one. In the structural estimation below, we normalize trade costs by scaling up all trade costs proportionally until the lowest iceberg trade cost has a value of one.

	Road	Rail	Water	Air
Variable cost	1.2526	1.1165	2.1261	0.4866
Fixed cost	0	0.9766	1.2210	1.7748

Note: Distance costs between locations for a particular shipment mode are calculated as exp(variable cost * distance + fixed cost), where distance is mode-specific and normalized so that the width of the United States is one.

Table 4: Estimated Trade Costs

4.2 Estimation of labor demand and supply

4.2.1 Relative demand and supply

We estimate high-skill agglomeration advantage φ, the elasticity of substitution across skill groups ρ, and labor supply elasticity θ, using the equilibrium conditions (12) and (13) derived in Section 3.1.3. We write these equations in log relative terms as follows,

$$\tilde{w}(i) = \tilde{\kappa} + \frac{1}{\theta}\tilde{n}(i) - \tilde{u}(i) \tag{22}$$

$$\tilde{n}(i) = -\rho\tilde{w}(i) + \rho\varphi\log n_H(i) + \rho\tilde{\beta}(i) \tag{23}$$

where

$$\tilde{n}(i) = \log\left[\frac{n_H(i)}{n_L(i)}\right], \ \tilde{w}(i) = \log\left[\frac{w_H(i)}{w_L(i)}\right], \ \tilde{\beta}(i) = \log\left[\frac{\bar{\beta}_H(i)}{\bar{\beta}_L(i)}\right], \ \tilde{u}(i) = \log\left[\frac{\bar{u}_H(i)}{\bar{u}_L(i)}\right]$$

and, $\tilde{\kappa}$ is a constant.[26] Estimating these equations using OLS can be problematic due to correlations between error terms and regressors. In equation (22), the skill population ratio, $\tilde{n}$, is expected to be higher in locations where the ratio of amenity values for high-skill relative to low-skill, $\tilde{u}$, are greater. This correlation means that OLS presumably underestimates $1/\theta$. In addition, in equation (23), skill premium, $\tilde{w}$, and high skill population, n_H, are presumably higher in locations where the ratio of high-skill to low-skill productivity, $\tilde{\beta}$ are larger. This correlation implies that OLS underestimates ρ and overestimates φ.

We use instrumental variables to estimate equations (22) and (23). To estimate θ in the relative supply function (22), we instrument skill population ratio $\tilde{n}$ using a variable that is meant to exclusively capture shifts from the demand side. We are inspired by a large urban and spatial inequality literature in constructing our exogenous shock using industry-level variation across locations (Bartik (1991); Moretti (2013); Diamond (2015)). Let d index industry, $E_d(i)$ be the employment share of industry d in location i with $\sum_d E_d(i) = 1$, and $\frac{N_{H,d}(-i)}{N_{L,d}(-i)}$ be the

---

[26] $\tilde{\kappa} = -\frac{1}{\theta}\log\left[\frac{N_H}{N_L}\left(\frac{W_H}{W_L}\right)^{-\theta}\right]$

national skill population share in industry d excluding location i itself. Our instrument is

$$\sum_d E_d(i) \log \left(\frac{N_{H,d}(-i)}{N_{L,d}(-i)} \right)$$

We assume that industry composition only affects the wage premium through its effect on the skill population ratio, and is uncorrelated with relative amenities. Suppose relative employment of high-skill workers is greater nationwide in certain industries. Then, cities with larger employment shares in those certain industries will have more demand for high-skill relative to low-skill workers. This creates a shift in demand for high-skill workers, which is presumably uncorrelated with supply factors (amenities) in a location.

The exogeneity assumption is that our instrument is uncorrelated with relative amenities– the amenities assigned to a location by high skill workers relative to the amenities assigned to a location by low skilled workers. In our model, we assume that amenities are immutable features of locations. Our exogeneity assumption would be violated if skill-intensive industries chose to locate in places relatively preferred by skilled workers. We believe that industry location is more driven by other factors such as proximity to markets as in Krugman (1980), proximity to natural resource endowments as in Ellison and Glaeser (1999), or simply historical accident as in Krugman (1991a).

Since we do not have strong priors about what sort of amenities high skill workers prefer relative to low skill workers, the exogeneity condition is difficult to test directly. We can however show that the instrument is uncorrelated with a number of different measures of amenity levels. Figure 2 contains scatter plots of the instrument against air pollution, humidity in the summer, temperature in the winter, distance from the coast, log remoteness and the quality of life index described in the following paragraph.[27] The instrument is uncorrelated with all of these measures except a weak negative correlation with the quality of life index which is only measured for MSAs.

To estimate ρ and φ in the relative demand function (23), we use the residuals of the relative supply function, $\tilde{u}$, as an instrument for skill premium, $\tilde{w}$. The orthogonality between this instrument and the error terms is based on the assumption that the relative amenity valuation, $\tilde{u}$, as a supply factor is uncorrelated with relative factor intensities, $\tilde{\beta}$, as a demand factor. In addition, we instrument high-skill population $n_H(i)$ using an extended quality of life index that we borrow from Albouy (2012).[28] This index is only reported for MSAs. We extend the index to our broader set of geographical units by regressing the index on a large set of observables, and predicting missing values. Our estimates remain virtually the same if we restrict our sample to only MSAs. This quality of life index is by construction uncorrelated with prices and wages in a location, but as Albouy shows, it strongly correlates with a wide range of natural and artificial amenities in a location. The orthogonality between this instrument and error terms is based on the assumption that this measure of quality of life is not correlated with relative factor intensity.

---

[27]Information on air quality is from Agency (2019), and maps of the US by temperature and humidity are from Oceanic and Administration (2016).

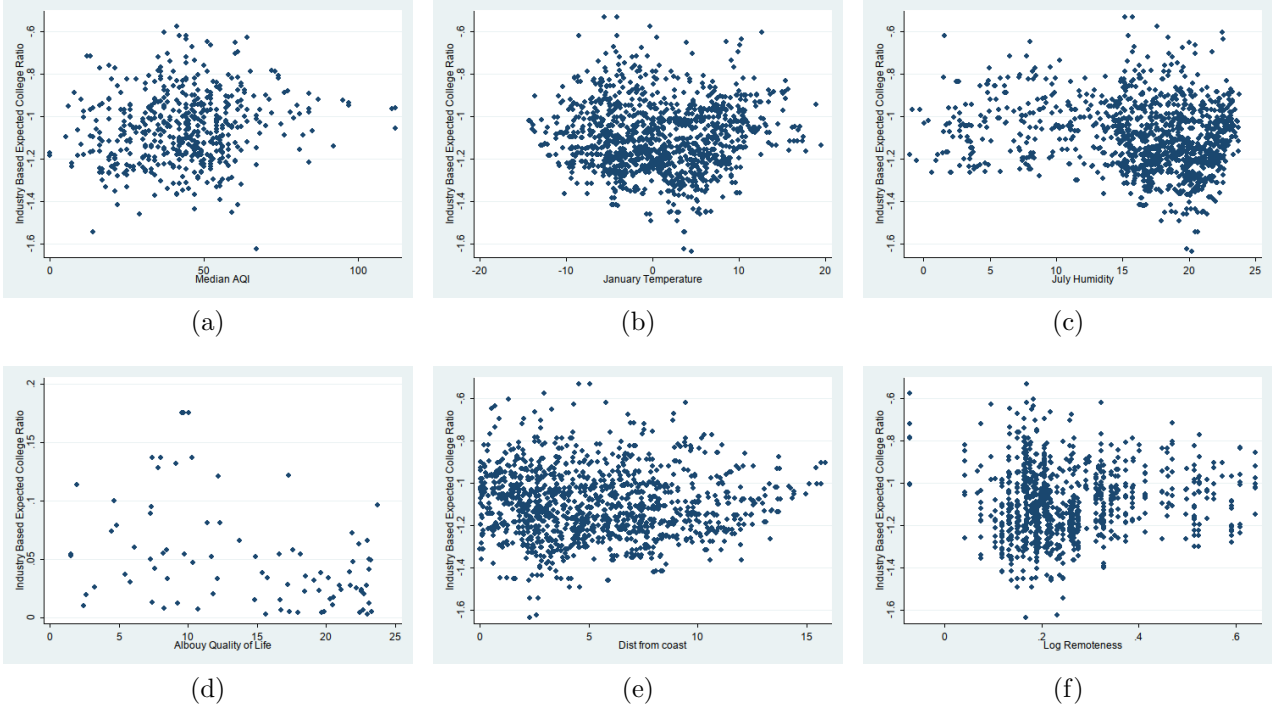[28]Specifically, we use Albouy's "adjusted" measure of quality of life.

Figure 2: Industry Skill Instrument vs Amenity Level Measures

Estimation results are summarized in Table 5. The Cragg-Donald F-statistics for the first stage strongly reject that the instruments are weak. The more conservative heteroskedacticity-robust Kleibergen-Paap F-statistics are somewhat lower, especially the F-statistic on the labor demand regression of 10.0. For all parameters the IV regressions push the OLS estimates in directions consistent with our priors explained above. According to our estimates, the dispersion of location preferences $\theta = \frac{1}{.072} = 13.8$, the elasticity of substitution in production between high and low-skill labor $\rho = 3.276$, and the agglomeration advantage of high-skill labor $\varphi = 0.368/3.276 = 0.112$.[29] In addition, the residuals in equations (22) and (23) give us the exogenous shifters of relative productivities and amenities $\tilde{\beta}$ and $\tilde{u}$.

_____

[29]Our estimate of the elasticity of substitution between high skill and low skill labor ρ is a bit higher than estimates reported by the literature. In a literature review, Katz et al. (1999) reports values for this elasticity between 1.40 to 1.70. Ciccone and Peri (2006) come up with estimates between 1.3 and 2, Diamond (2015) estimates $\rho = 1.6$, and Card (2009) finds that $\rho = 2.5$. There is a shorter literature estimating the dispersion of location preferences θ. Our estimate of θ is close to the point estimate of 11.7 in Allen and Donaldson (2018) and higher than what others have found in the literature. Monte et al. (2015) estimate a preference dispersion parameter of 3.30, and Serrato and Zidar (2016) estimate a parameter between one and two.

	log skill premium, Eq. (22)		log population ratio, Eq. (23)	
	OLS	IV	OLS	IV
log population ratio	0.055***	0.072 ***		
log skill premium			-0.092	-3.276***
log high skill population			0.182***	0.368***
constant	0.416***	0.426 ***	-2.568***	-3.389***
1st stage F (CD)		4956		100.7
1st stage F (KP)		1123		10.0

Note: Robust standard errors are in parentheses. Number of observations is 1267 across all columns. All observations are weighted by city population. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Estimating relative labor demand and supply

4.2.2 Productivities and amenities

With key parameter estimates in hand, we next solve for total factor productivity $A(i)$ and high-skill base utility from amenities $\bar{u}_H(i)$. Similar to spatial quantitative models, the identification of these shifters relies on the combination of observed population and wages across cities.

To invert the population and wage data, we use equilibrium integral equations derived by our model. Our estimation procedure consists of two steps:

Step 1. We first estimate total factor productivity A inclusive of spillovers as well as high skill amenity values $\bar{u}_H$. To do so, we rewrite the two systems of integral equations as follows:

$$
\begin{aligned}
A(i)^{1-\sigma} &= W_H^{\frac{1-\sigma}{\delta}} N_H^{\frac{\sigma-1}{\delta\theta}} \nu(i)^{1-\sigma} n_H(i)^{-1} w_H(i)^{-1} b(i) \\
&\times \int_J d(i,j)^{1-\sigma} \bar{u}_H(j)^{\frac{\sigma-1}{\delta}} C(j)^{\frac{(\sigma-1)(\delta-1)}{\delta}} n_H(j)^{\frac{1-\sigma+\delta\theta}{\delta\theta}} w_H(j)^{\frac{\sigma-1+\delta}{\delta}} b(j)^{-1} \; dj \quad (24) \\
\bar{u}_H(i)^{\frac{1-\sigma}{\delta}} &= W_H^{\frac{1-\sigma}{\delta}} N_H^{\frac{\sigma-1}{\delta\theta}} n_H(i)^{\frac{1-\sigma}{\delta\theta}} w_H(i)^{\frac{\sigma-1}{\delta}} C(i)^{\frac{(\sigma-1)(\delta-1)}{\delta}} \\
&\times \int_J d(j,i)^{1-\sigma} A(j)^{\sigma-1} \nu(j)^{1-\sigma} \; dj \quad (25)
\end{aligned}
$$

Here, $A(i)$ and $\bar{u}_H(i)$ are unknown variables, whereas population and wages are known. As long as trade costs are symmetric $d(i,j) = d(j,i)$, we can further reduce the two systems of equation into one. If either of above integral equations hold along with the following relation, then both systems will hold:

$$
\bar{u}_H(i)^{\frac{\sigma-1}{\delta}} C(i)^{\frac{(\sigma-1)(\delta-1)}{\delta}} n_H(i)^{\frac{1-\sigma+\delta\theta}{\delta\theta}} w_H(i)^{\frac{\sigma-1+\delta}{\delta}} b(i)^{-1} = \lambda A(i)^{\sigma-1} \nu(i)^{1-\sigma}, \quad (26)
$$

where $\lambda > 0$ is a constant. The numerical algorithm by which we solve these equations is described in detail in Appendix D.

Step 2. We use our recovered productivities $A(i)$ to estimate common agglomeration parameter α and to recover base productivities $\bar{A}(i)$. Taking logs of (7) we get:

$$\log A(i) = \alpha \log n(i) + \log \bar{A}(i) \tag{27}$$

We regress recovered log total factor productivity on log population, instrumenting population with our estimated high-skill amenity values $\bar{u}_H(i)$. Results are reported in Table 6. We find that the elasticity of Hicks-neutral productivity with respect to population is 0.305. The IV and OLS results are similar. While not reported, removing population weights barely changes these estimates.

Our estimate is broadly in line with the estimates of other recent quantitative economic geography models, for example, Giannone (2017) finds a common agglomeration elasticity of 0.31. Estimates in the urban and macro literature, on the other hand, are typically less than 0.10 (see the survey by Rosenthal and Strange (2004)). The lower agglomeration elasticity in the urban and macro literature is caused by the assumption that goods produced in different locations are perfect substitutes, i.e. σ is infinity. In contrast, we assume a finite elasticity of substitution across goods differentiated by location. If goods produced in different locations are imperfect substitutes, demand falls less when productivity is low. Thus, relative to the case of perfect substitutability, producers in a smaller city are in a better position to compete with producers in larger cities. To explain the data, our model must therefore assign a larger productivity advantage to larger cities.

Dependent variable: Log productivity		
	OLS	IV
Log population	0.319***	0.321***
Constant	-3.049***	-3.065 ***
1st stage F (CD)		59000
1st stage F (KP)		3434
Obs	1267	1267

Note: Robust standard errors. All observations are weighted by population. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Estimating agglomeration

At the calibrated values of $\bar{A}$ and $\bar{u}_H$, our solution algorithm reproduces the exact data on wages and population of low- and high- skill workers. This check confirms the accuracy of both our calibration and simulation algorithms.

4.2.3 Results for the productivity and amenity shifters

In Figure 3, we present the estimated geographical distribution of the four *exogenous shifters*: base productivity, high-skill amenities, relative productivity, and relative amenity valuation.

We estimate that common base productivity is higher in the coastal regions of the United States as well as the Rocky Mountains. It is worth pointing out that, unlike Allen and Arkolakis (2014), we do not find that cities are fundamentally more productive than other regions.[30] Here we avoid to some degree the critique of the new economic geography literature that cities are exogenously more productive than nearby, naturally similar areas. Instead, we find that areas of the United States which are either near coastline, or areas such as the Rocky Mountain region that have relatively low humidity are more fundamentally productive.[31] We do find, however, that exogenous high-skill amenities $\bar{u}_H$ are strongly correlated with city size. Our results are consistent with those of Albouy (2012) who shows that in many ways cities are attractive places to live for reasons not related to productivity.

(a) Base common TFP $\bar{A}(i)$

(b) Relative high-skill productivity $\frac{\bar{\beta}_H}{\bar{\beta}_L}(i)$

(c) Base high-skill amenities $\bar{u}_H(i)$

(d) Relative high-skill amenities $\frac{\bar{u}_H}{\bar{u}_L}(i)$
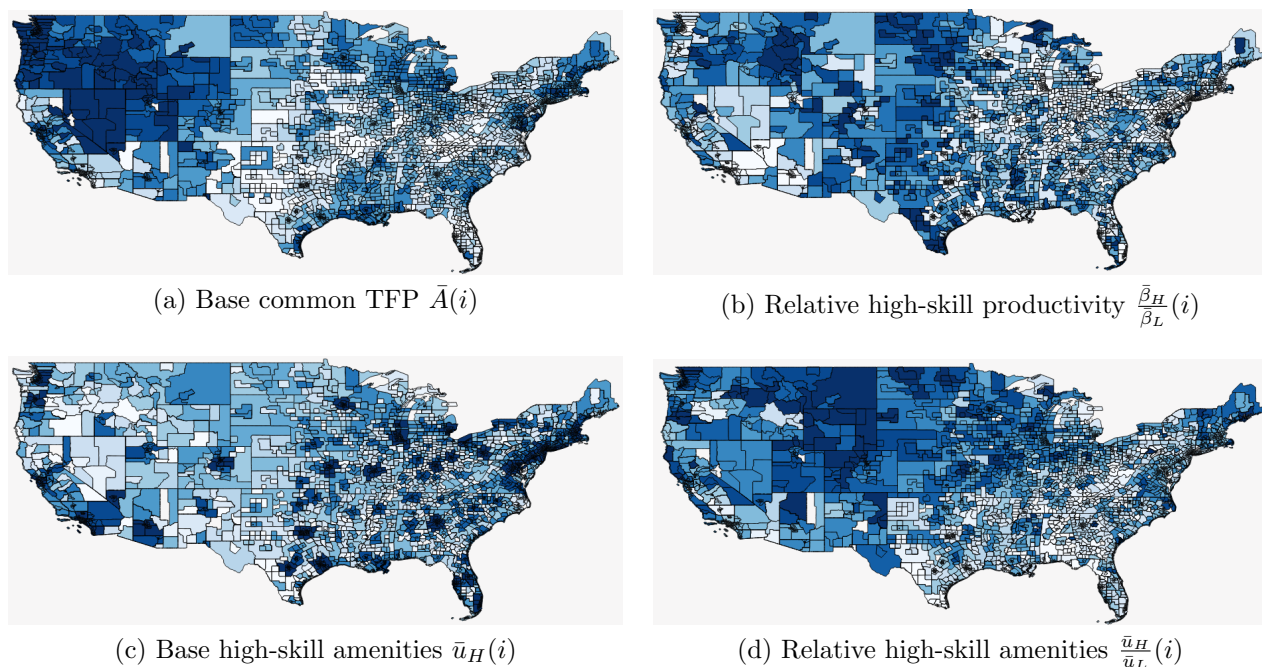
Figure 3: Locations colored by estimate

Turning to the relative measures, we find that both are reasonably smooth across geography. We find that in relative terms, low-skill people prefer to live in the South, and tend to be exogenously more productive in the Lower Midwest region, possibly reflecting the relatively high soil quality in that region. High-skill people prefer to live in the Upper Midwest, Mountain regions, and Northwest.

---

[30]Neither do we find them consistently less productive than other regions.

[31]The reader should keep in mind that these estimates are neither *observed* productivity nor amenities. Those objects are functions of the distribution of population which in turn is an equilibrium object in our model.

5 Quantitative exercises

5.1 Role of geography in wage inequality

We motivated our modeling exercise in part as adding geography into a spatial inequality model. To measure the contribution of geography to wage inequality, we decompose observed variation in wage premia into variations in exogenous base productivities and amenities in absolute and relative terms, as well as geographic position. Consider the following relation:

$$\log\left(\frac{w_H(i)}{w_L(i)}\right) = \gamma_1 \log\left(\frac{\bar{\beta}_H(i)}{\bar{\beta}_L(i)}\right) + \gamma_2 \log\left(\frac{\bar{u}_H(i)}{\bar{u}_L(i)}\right) + \gamma_3 \log \bar{A}(i) + \gamma_4 \log \bar{u}_H(i) + \gamma_5 \log P(i) + \zeta(i)$$

The first four terms on the right hand side are the four exogenous shifters in our model. The fifth term is the tradeables price index P. The price index of tradeables in a location exclusively embodies the geography of a location with respect to all other locations because it is the only term that incorporates bilateral trade costs. Lastly, as our model does not imply the above relation in closed form, we include an error term ζ.

	Notation	Sk. prem	Sk. prem	Shp. R^2	Sk. prem	Shp. R^2	Sk. prem	Shp. R^2
Log tradeable price	P	-0.052***	-0.008***	16.5%				
Log remoteness	Rem				-0.053***	9.8%		
Log unweighted remoteness							-0.005***	1.4%
Log amenity level	$\bar{u}_H$		0.063***	15.7%	0.067***	26.7%	0.07 ***	30.4%
Log base productivity	$\bar{A}$		0.002	2.4%	0.033***	2.0%	0.004**	2.1%
Log relative productivity	$\frac{\bar{\beta}_H}{\bar{\beta}_L}$		0.228***	10.0%	0.242***	8.1%	0.227***	7.5%
Log relative amenities	$\frac{\bar{u}_H}{\bar{u}_L}$		-0.826***	55.3%	-0.800***	53.3%	-0.830***	58.6 %
Observations		1267	1267		1267		1267	
R-squared		0.303	0.992		1.000		0.991	

Note: Regressions report robust standard errors. All observations are weighted by population.
*** p<0.01, ** p<0.05, * p<0.1.

Table 7: Decomposition

We use this relation to quantify how much observed variation in geographic features across American cities explain variation in their wage premia. In the first column of Table 7, we report R^2 for a simple regression of the log skill wage premium on the log price index of tradeables. We find that the price index alone can explain 30% of the variation in the wage premium.

In the third column of Table 7, we report results from the full decomposition. Our five shifters explain 99% of the variation in observed wage premia. Using the Shapley decomposition method, we find that 16.5% of observed variation in the skill wage premium are due to the variation in geographic features across American cities. Geographic features explain more of the variation in wage premia than relative productivity and productivity levels combined. While both geography and productivities contribute measurably to wage inequality across space, we find that the largest part of the variation in wage premia is explained by variation in relative amenities. The signs of each factor in the regression is as expected. We expect more productive

and nicer places, all else equal to have higher population and thus more wage inequality. We expect more remote places to have lower wage inequality. We also expect places with higher relative productivity to have more inequality. Finally, we expect places which high-skill workers value more to have lower wage inequality, since high-skill workers will be relatively attracted to these places even if their wages there are relatively low.

In the last four columns of Table 7, we report two similar decompositions, but with the price index replaced by the remoteness measure from Section 2.2 and an unweighted version of the remoteness measure, which is simply an unweighted average of iceberg trade costs. Results are similar across decompositions, although geography explains less of the variation in skill premium when we use remoteness, and even less when we use the unweighted version of remoteness. This is expected, because the price index is the model consistent weighted average of trade costs. The population weights in the remoteness measure do not fully reflect relevant productivity differences between cities. The unweighted remoteness measure weights all trade costs the same way, but having low trade costs with a productive city is important, while having low trade costs with an unproductive city is irrelevant.

5.2   Domestic trade and inequality

We examine how welfare inequality reacts to changes in domestic trade costs. To highlight the forces at work, we increase trade costs to five times the estimated level. This experiment approximates the standard exercise in the international trade literature of measuring the welfare changes of moving from the observed level of trade to autarky. Although this counterfactual experiment is extreme, it allows us to compare our results with those in the literature. In this exercise, we ask not only how much aggregate welfare decreases, but also how much relative welfare of high to low-skill workers changes.

(a) High-skill welfare

(b) Low-skill welfare

(c) Relative high to low-skill welfare
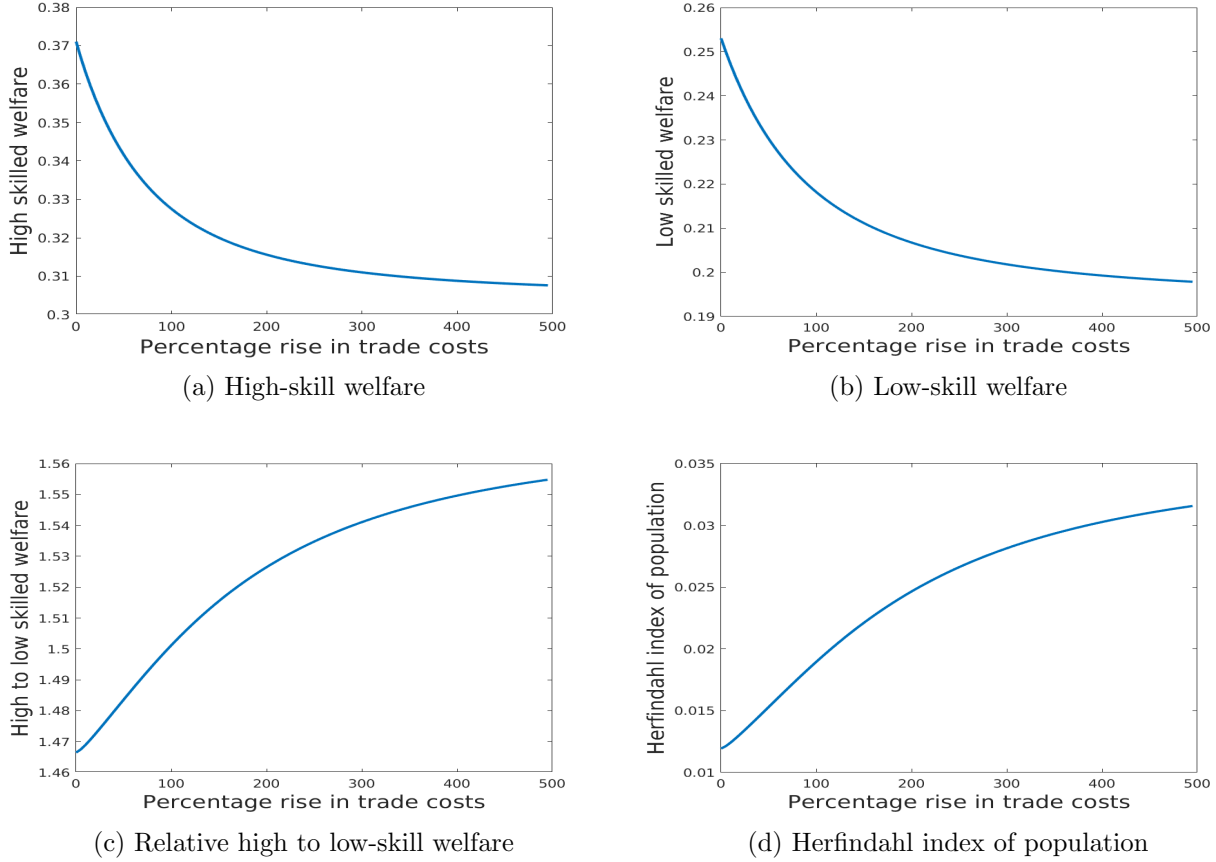
(d) Herfindahl index of population

Figure 4: Trade cost experiments

It is not immediately obvious how we should expect population to change after an increase in trade costs. One force causes additional concentration. After trade costs increase, cheap goods from productive cities are no longer cheap in small towns, and conversely, already expensive goods produced in small towns become even more expensive in cities. This mechanism encourages workers to move to cities in order to access cheap goods and find lucrative jobs, leading to an overall concentration of population and thus an increase in agglomeration forces. Since high-skill workers benefit relatively more from agglomeration, the concentration in population raises welfare inequality. A second force tends to make population diffuse. If trade costs approach infinity, all cities are equally isolated. The decrease in the cost advantage of formerly well-connected cities encourages workers to move to formerly isolated cities to benefit from lower housing costs there. This diffusion leads to a lessening of the strength of agglomeration forces and drives welfare inequality down.

We find that the force causing concentration dominates. Figure 4 summarizes our results from a large number of counterfactual experiments. In each experiment we increase all trade costs from their baseline values uniformly by $1, 2, ..., 500$ percent. Our basic finding is that both high and low-skill welfare fall with increases in trade costs, but low-skill welfare falls more. In the extreme case of five times measured trade cost, high and low-skill welfare decrease by 17.1%

and 21.8% respectively. Accordingly, the ratio of high to low-skill welfare increases by 6.0%. To make a connection to the intuition we provided above, we also report changes to a Herfindahl index in population, that is, the sum of squared population shares of American cities. As shown in Figure 4, the Herfindahl index in population monotonically increases with trade costs.

In addition to overall effects on welfare, our model allows us to analyze exactly which areas of the United States grow and shrink as trade costs rise.[32] Figure 5 contains percentage population changes relative to our data when all trade costs are 500 percentage higher. Blue indicates an increase in population (light greater than 5%, and dark greater than 25%), white indicates no change (-5% to 5% population growth), and red indicates a decrease in population (light greater than 5%, and dark greater than 25%). Population concentrates in a small number of cities and their surrounding areas across the United States.[33]

We find that population moves away from formerly well-connected areas in the South and Midwest regions of the United States to the coasts. In particular, the formerly less connected cities Seattle and Portland in the Pacific Northwest expand markedly. Conversely, formerly highly connected cities with lower amenity levels, such as St. Louis or Cincinnati, largely shrink. The five cities with the highest and the five cities with the lowest changes in population are listed in Table 8.[34] Also in Table 8, we calculate the elasticities of high and low-skill population to a uniform increase in trade costs at the initial level. As a rule, high-skill population is more sensitive to changes in trade costs. As cities grow skilled workers become relatively more productive, inducing more skilled in-migration. As cities shrink skilled workers become relatively less productive, inducing more skilled out-migration.

We have argued that rising trade costs both lead to concentration in large cities, and dispersion away from well-connected cities. Since people prefer to live in well-connected cities, these two variables are highly positively correlated. We provide evidence for these two mechanisms by first regressing log initial population on the log initial price index of tradeables. A city with a low residual from this regression is more well-connected than we might expect given its population. We would expect the dispersion force to dominate in such a city, and the population to fall. The opposite is true for a city with a high residual. We would expect the concentration force to dominate and the city to grow. Figure 6 is a scatter plot of log population change against the residual from the regression described above. The circle size is proportional to the initial population of the city. As predicted we see a positive slope, indicating that population falls more on average in cities with low residuals.

These results complement a literature that studies the effects of international trade on inequality in developing countries (Antràs et al., 2006; Hummels et al., 2014). Indeed, and in contrast to the Stolper-Samuelson theorem, globalization has increased inequality even in developing countries (Davis and Mishra, 2007). We find instead that domestic trade costs and

---

[32]We can also examine changes to population ratios and skill premia across cities. These closely track changes to population so that the maps do not qualitatively differ from that presented here on total population changes.

[33]Surrounding areas grow because although trade costs are very high, they are not infinite.

[34]Locations are only included in this table if their total population of high school graduates and 4-year college degree is greater than 300,000 in the data.

inequality are positively correlated. The key difference in our context is that workers are mobile, and thus agglomeration economies change endogenously with market integration. The negative effect of trade on wage inequality in the international context is reversed when labor is mobile across locations within a nation.
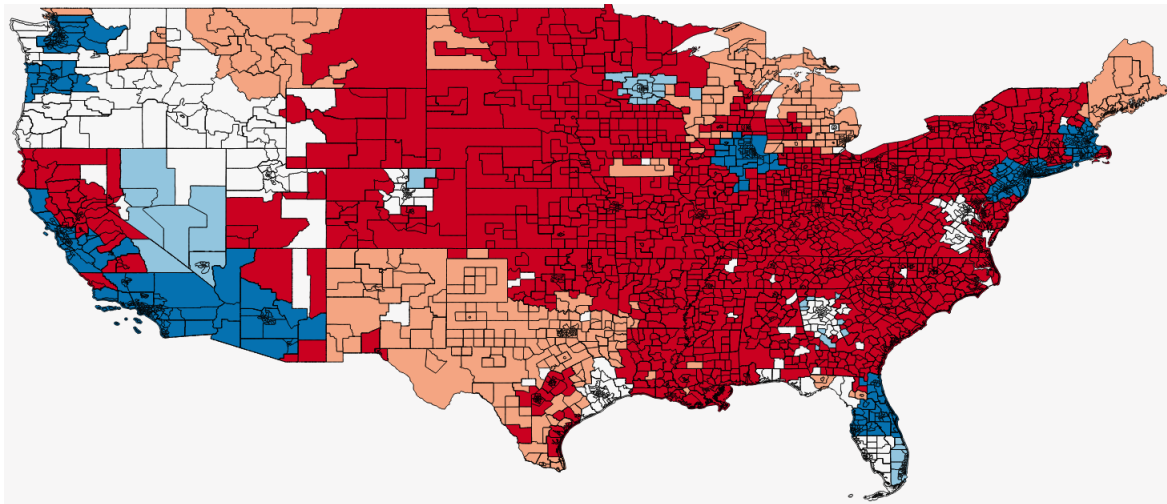

Figure 5: Population changes when all trade costs are five times larger

Note: Blue: increase in population, White: no or little growth, Red: decrease in population.

| | 500% trade cost | | | Elasticity at Initial Pop | |
City	baseline working population (million)	500% t.c. working population (million)	% change	high-skill pop elasticity	low-skill pop elasticity
Philadelphia	1.0	2.0	105%	0.34	0.27
New York City	2.7	5.5	105%	0.33	0.27
Seattle	0.5	1.0	103%	0.33	0.27
San Bernardino	0.4	0.7	83%	0.33	0.27
Los Angeles	1.5	2.7	82%	0.32	0.26
Cleveland	0.4	0.2	-33%	-0.12	-0.04
St. Louis	0.4	0.2	-45%	-0.27	-0.15
Pittsburgh	0.4	0.2	-46%	-0.22	-0.12
Kansas City	0.3	0.2	-47%	-0.30	-0.18
Cincinnati	0.3	0.2	-51%	-0.19	-0.09

Table 8: The five top regions with the highest and the top five with the lowest counterfactual population growth (reported for locations initially larger than 300,000 high school and college workers)
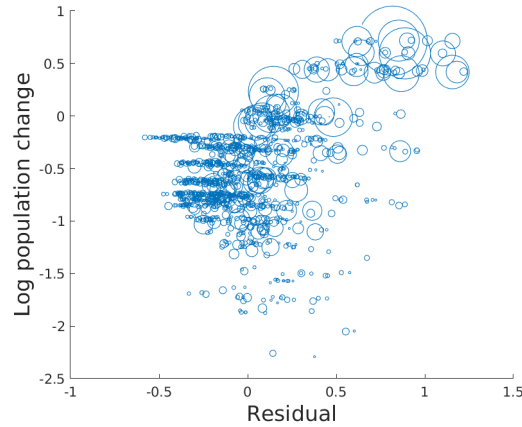
Figure 6: Log population change against residual of log initial population on log price index

5.3 Californian productivity shocks

In the 20 years leading up to the turn of the 21st century, California's share of the US population increased by 1.8 percentage points. In the same period, the college population ratio in California grew by 14 percentage points.[35] The growth in California's population and its biased growth in highly educated workers were the outcome of nontrivial interactions between productivity, demographics, housing regulations and other factors both in California and in other states. That being said, one particularly important factor behind Californian growth in this period was the expansion of the computing and high technology sectors. This period saw the rise of Silicon Valley during the lead up to the Dot-Com Bubble.

---

[35]The US population share of California grew from 10.2% in 1980 to 12.0% in 2000. Put another way, California's population increased by 43.1% from 1980 to 2000, while the total population of the United States increased by only 19.3%. The college population share in California was 37% in 1980 and 51% in 2000.

Average national high-skill welfare	1.3	
Average national low-skill welfare	0.3	
Average national welfare ratio	1.0	
	California	Rest of the United States
High-skill wages	11.4	-0.4
Low-skill wages	7.4	-1.1
Skill premium	3.4	0.7
High-skill population	42.9	-4.0
Low-skill population	3.9	-0.4
Price of tradeables	-5.0	-0.7
Total price index	6.9	-1.4
High-skill real wages	4.0	1.0
Low-skill real wages	0.6	0.3

Table 9: The effects of California's productivity shocks on welfare, prices, and wages (percentage change)

We perform a counterfactual exercise to study how technological progress in California contributed to welfare and inequality in California and across the United States. Holding the rest of the United States constant, we scale up the total factor productivity $\bar{A}$ and the skill bias of technology $\frac{\bar{\beta}_H}{\bar{\beta}_H}$ in all regions in California proportionally to match the observed growth in Californian population share and the growth in the college population ratio. We find that the skill bias of Californian technology rose by 8.8% and Californian total factor productivity increased by 9.0% from 1980 to 2000.

We report other results of this counterfactual exercise as percent changes from the counterfactual case to our baseline in Table 9. We find that national expected welfare of high-skill workers increased by 1.3%, national expected welfare of low-skill workers increased by 0.3%, and welfare inequality rose by 1.0%.

Furthermore, we examine changes to wages and prices across locations within a skill group, and across skill groups within a location. Table 9 reports changes to the population-weighted mean wages, prices, and skill premium in California and the rest of the United States. Overall, the skill premium rose an average by 3.4% across Californian cities, and 0.7% on average elsewhere. The skill premium in California increased less than the skill bias of technology. We might have expected the opposite, since the effect of exogenous high-skill productivity increases on wage inequality could be amplified through population growth and the accompanying high-skill agglomeration advantage. On the other hand, general equilibrium effects act to dampen the effect of productivity changes by increasing the supply of high-skill workers in California. In addition, the overall price index including both housing and tradeables rose by 6.9% in California while it fell by 1.4% elsewhere. In California, the higher price index is due to a dramatic

increase in housing price which dominates the fall in the price of tradeables, while in the rest of the US cheaper tradeables are the main driver of the lower cost of living.

6 Extensions

In this section, we discuss a few ways we might extend our results.

Sorting into industries and occupations. We have shown that more centrally located cities tend to have more wage inequality. We look into how geographic location would matter for inequality through sorting of skills into industries and occupations. To provide suggestive evidence, we add industry- and occupation-related controls to our individual-level regressions in Section 2. We classify all industries into 30 groups and all occupations into 23 groups based on IPUM classifications (see Table 11 in the appendix). We then define skill intensity for an industry (or occupation) as the share of college employment in that industry (or occupation) at the national level. Table 10 reports our regressions results. In column (1) we include industry and occupation fixed effects. In column (2) we instead control for industry and occupation skill intensity. In column (3), we reproduce column (5) of Table 2 where we allow for college dummy interacted with remoteness and population without controlling for industry or occupation. In column (4), we also allow industry and occupation skill intensity to interact with population and remoteness. As in Table 2, we are mainly interested in the coefficient on the interaction of college dummy with remoteness, which reflects how large wages of college graduates are compared to high-school graduates in more remote cities.

Geographic location of a city remains correlated with skill premia across all specifications. The coefficient of remoteness interacted with college slightly decreases from -0.113 in column (3) to -0.094 in column (4) where we add industry and occupation controls.[36] We take this as suggestive evidence that sorting to industries and occupations may mediate the relationship between remoteness and skill wage premia, but only marginally.

---

[36]Furthermore, the coefficient of remoteness for high-school graduates is not statistically significant in column (4), suggesting that once industry and occupation are controlled for, the remoteness effect operates primarily through college graduates.

	dependent variable: log wage of individual workers			
	(1)	(2)	(3)	(4)
Log remoteness	-0.168***	-0.170***	-0.110*	0.0333
	(0.0593)	(0.0616)	(0.0641)	(0.0737)
College	0.344***	0.364***	0.300***	0.347***
	(0.00382)	(0.00397)	(0.0321)	(0.0270)
Log population	0.0361***	0.0393***	0.0380***	0.00848**
	(0.00205)	(0.00221)	(0.00214)	(0.00385)
Ind skill intensity		0.0921*		-0.925***
		(0.0507)		(0.154)
Occ skill intensity		0.629***		-0.244***
		(0.0142)		(0.0684)
College X Log remoteness			-0.113***	-0.0941***
			(0.0321)	(0.0261)
Ind skill intensity X Log remoteness				-0.810***
				(0.216)
Occ skill intensity X Log remoteness				0.144**
				(0.0729)
College X Log population			0.0151***	0.00246
			(0.00236)	(0.00199)
Ind skill intensity X Log population				0.0933***
				(0.0146)
Occ skill intensity X Log population				0.0682***
				(0.00568)
R-squared	0.357	0.293	0.276	0.296
Ind & Occ FE	Y	N	N	N

Notes: Standard errors, clustered at city level, are reported in parentheses. In all regressions, there are 3,050,723 observations, we weight individuals based on census sampling weights, and we include individual-level gender and race dummies, a cubic polynomial of years of experience, and state fixed effects. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 10: Wages and remoteness controlling for industry and occupation, at the level of individual workers

Endogenous skill-biased amenities. Amenities might endogenously respond to skill composition of a city, as discussed in Diamond (2015), and low and high skill workers may value these amenities differently. In an extension of our model, we may specify amenities as $u_s(i) = \bar{u}_s(i) \left[\frac{n_H(i)}{n_L(i)} \right]^{\eta_s}$ where η_s captures the degree to which skill composition affects the supply of amenities. Including this specification into relative supply equation (13), and combining with relative demand

equation (12) gives

$$\frac{w_H(i)}{w_L(i)} = \left(\frac{W_H}{W_L}\right)^{\frac{\theta}{\theta+\tilde{\rho}}} \left(\frac{\bar{u}_H(i)}{\bar{u}_L(i)}\right)^{\frac{-\theta}{\theta+\tilde{\rho}}} \left(\frac{N_H}{N_L}\right)^{\frac{-1}{\theta+\tilde{\rho}}} \left(\frac{\bar{\beta}_H(i)}{\bar{\beta}_L(i)}\right)^{\frac{\tilde{\rho}}{\theta+\tilde{\rho}}} n_H(i)^{\frac{\varphi\tilde{\rho}}{\theta+\tilde{\rho}}} \tag{28}$$

where $\tilde{\rho} \equiv \rho\left(1 - \theta(\eta_H - \eta_L)\right)$. This equation, as an equilibrium relationship in relative terms, collapses to equation (14) if $\eta_H - \eta_L = 0$. We examine how our parameter estimates may change in this extended model. On the one hand, we rewrite our estimable equation (22),

$$\tilde{w}(i) = \tilde{\kappa} + \frac{1 - \theta(\eta_H - \eta_L)}{\theta}\tilde{n}(i) - \tilde{u}(i) \tag{29}$$

Our instrumental variables strategy used to estimate the dispersion of location preferences the relative supply equation (22) should still remain valid for estimating (29). In this endogenous amenities version of the model, however, we interpret the estimate differently. In particular, the estimated coefficient $\tilde{n}(i)$, 0.072, implies that $\theta = \frac{1}{0.072+(\eta_H-\eta_L)}$. Suppose $\eta_H > \eta_L$ meaning that high-skill workers attach a higher valuation to amenities derived from relative supply of high-skill workers. Then, relative to our baseline estimates, we would infer a smaller θ. That is, we would estimate more dispersion in unobserved location preferences. Intuitively, the labor supply equations describe the way that the relative supply of high skill workers reacts to the relative wage level. This is informative about worker preferences over locations, because the more indifferent they are between locations, the more population will react to wage level. Since endogenous amenities increase the incentive for high skill workers to live in cities with a high relative supply of high skill workers, to explain the observed relationship between relative wages and relative population we do not need workers to be as indifferent between locations as in our baseline model.

Since η_s does not enter into relative demand equation (23), we will have the same estimates of ρ and φ as in our baseline. Substituting these expressions into equation (28), we find that the exponent of n_H in the extended model is exactly the same as that in equation (14) in our baseline model. Hence, the estimated elasticity of the wage premium to high-skill employment will remain unchanged. On the other hand, our estimates of the fundamental productivity and amenity shifters would be somewhat different in a version of the model with endogenous amenities. The key challenge in such an exercise would be to devise a model-consistent estimation method to separate the elasticity parameters η_s from the location dispersion parameter θ.

Migration costs. In this paper we have developed a medium-run static model building on an empirical spatial equilibrium literature. Geography enters our model purely through trade costs. Another potential way in which geography might affect the distribution of wages is through migration costs. For example, if it is more costly to move far away, then initial placement will be an important determinant of final location choice. Since our model is static and we are considering medium-run outcomes, in our baseline we abstract from dynamic frictions like

36

moving costs.[37] Several studies on China using similar static models in the spatial equilibrium tradition have developed methods by which to include some sense of moving cost (Fan, 2019; Tombe and Zhu, 2019). This exercise is of critical importance in studying China, as the *hukou* system permanently reduces the public services available to migrants, particularly people born in the countryside wishing to move to the city. Consistent with the Chinese context, migration costs in Tombe and Zhu (2019) are modeled not as a one-time cost of relocation, but rather as a flow cost which substantially scales down welfare according to their estimates.[38] If relocation costs are less substantial and paid once at the time of moving, as we might presume about the American context we study, then in the medium-run they are likely less important than they are in China.

Although we do not include moving costs in our baseline model, we can speculate how large flow costs as in Tombe and Zhu (2019) might affect our results. It is easiest to consider the case in which moving costs are uniform across skill groups, and are paid only if a worker currently lives in a location other than where he is born. There will be an intuitive trade off between the dispersion of preferences over cities and moving costs. The higher are moving costs, the higher will θ need to be in order to rationalize the observed spatial distribution of workers. Since moving costs and preference dispersion both discourage movement, counterfactual population flows in response to productivity or policy shocks may be similar in such an extension to our baseline results.

7 Conclusion

We document that isolated cities tend to have less wage inequality. We develop a theory in which the higher cost of tradeables in isolated cities makes them less attractive to live in, and high-skill workers are less productive in smaller cities. We build a quantitative model to understand and measure this mechanism. Our model bridges the gap between the spatial inequality literature which abstracts from geography, and the economic geography literature, which abstracts from inequality. We find that 16.5% of observed variations in skill wage premium is due to the geographic location of cities. In addition, we find that a uniform increase in domestic trade costs causes inequality to rise due to the interaction between a higher concentration of population and the agglomeration advantage of high-skill labor. In a counterfactual experiment, we find that the rise of Silicon Valley increased skill wage premium in California by 3.4% and welfare inequality across the United States by 1.0%.

---

[37]A recent literature has developed dynamic spatial models more suited to studying dynamic frictions. Kennan and Walker (2011) use a dynamic discrete choice model to understand the role of learning and moving costs in US interstate migration. Caliendo et al. (2017) develop a dynamic equilibrium model to understand how trade reforms interacted with population movement in the European Union.

[38]Tombe and Zhu (2019) specify migration costs as proportional to destination welfare, analogously to how iceberg trade costs are specified as proportional to origin price. They estimate that these iceberg welfare costs of migration are on average 2.8 in 2000 across Chinese province pairs. To get a sense of the magnitude of these costs we can compare them with their trade counterpart. Anderson and Van Wincoop (2004) report iceberg trade costs across US states to be around 1.7 on average.

References

Acemoglu, D. and Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of labor economics*, volume 4, pages 1043–1171. Elsevier.

Agency, U. E. P. (2019). Air quality data. https:/www.epa.gov/outdoor-air-quality-data. Accessed: 2019-01.

Aguiar, M. and Bils, M. (2015). Has consumption inequality mirrored income inequality. *American Economic Review*, 105:2725–2756.

Albouy, D. (2012). Are big cities bad places to live? estimating quality of life across metropolitan areas. *mimeograph.*

Allen, T. and Arkolakis, C. (2014). Trade and the topography of the spatial economy. *The Quarterly Journal of Economics*, 129(3):1085–1140.

Allen, T., Arkolakis, C., and Li, X. (2016). Optimal city structure. *mimeograph.*

Allen, T. and Donaldson, D. (2018). The geography of path dependence. Technical report, Working Paper.

Anderson, J. E. and Van Wincoop, E. (2004). Trade costs. *Journal of Economic literature*, 42(3):691–751.

Antràs, P., Garicano, L., and Rossi-Hansberg, E. (2006). Offshoring in a knowledge economy. *Quarterly Journal of Economics*, 121(1).

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

Bartelme, D. (2015). Trade costs and economic geography: evidence from the us. *Ann Arbor, MI: Department of Economics, University of Michigan, https://drive. google. com/file/d/0B_fRktLO_ V0ncHNNcFowMlF0Yms/view.*

Bartik, T. J. (1991). Boon or boondoggle? the debate over state and local economic development policies.

Baum-Snow, N., Freedman, M., and Pavan, R. (2014). Why has urban inequality increased? *Working paper.*

Baum-Snow, N. and Pavan, R. (2012). Understanding the city size wage gap. *The Review of economic studies*, 79(1):88–127.

Broda, C. and Weinstein, D. E. (2006). Globalization and the gains from variety. *The Quarterly journal of economics*, 121(2):541–585.

Caliendo, L., Opromolla, L. D., Parro, F., and Sforza, A. (2017). Goods and factor market integration: a quantitative assessment of the eu enlargement. Technical report, National Bureau of Economic Research.

Card, D. (2009). Immigration and inequality. *The American Economic Review*, 99(2):1.

Census, U. (2012). 2010 census summary file 1: Technical documentation. *http://www.census.gov/prod/cen2010/doc/sf1.pdf*.

Ciccone, A. and Peri, G. (2006). Identifying human-capital externalities: Theory with applications. *The Review of Economic Studies*, 73(2):381–412.

Combes, P.-P., Duranton, G., Gobillon, L., Puga, D., and Roux, S. (2012a). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica*, 80(6):2543–2594.

Combes, P.-P., Duranton, G., Gobillon, L., and Roux, S. (2012b). Sorting and local wage and skill distributions in france. *Regional Science and Urban Economics*, 42(6):913–930.

Coşar, A. K. and Fajgelbaum, P. D. (2016). Internal geography, international trade, and regional specialization. *American Economic Journal: Microeconomics*, 8(1):24–56.

Davis, D. R. and Dingel, J. I. (2014). The comparative advantage of cities. Technical report, National Bureau of Economic Research.

Davis, D. R. and Dingel, J. I. (2019). A Spatial Knowledge Economy. *American Economic Review*, 109(1):153–170.

Davis, D. R. and Mishra, P. (2007). Stolper-samuelson is dead: And other crimes of both theory and data. In *Globalization and poverty*, pages 87–108. University of Chicago Press.

Desmet, K., Nagy, D. K., and Rossi-Hansberg, E. (2016). The geography of development. *Journal of Political Economy*.

Diamond, R. (2015). The determinants and welfare implications of us workers diverging location choices by skill: 1980-2000. *American Economic Review*.

Ellison, G. and Glaeser, E. L. (1999). The geographic concentration of industry: does natural advantage explain agglomeration? *American Economic Review*, 89(2):311–316.

Fajgelbaum, P. and Gaubert, C. (2018). Optimal spatial policies, geography and sorting. Technical report, National Bureau of Economic Research.

Fajgelbaum, P., Morales, E., Surez-Serrato, J. C., and Zidar, O. (2015). State taxes and spatial misallocation. Technical report.

Fan, J. (2019). Internal geography, labor mobility, and the distributional impacts of trade. *American Economic Journal: Macroeconomics, forthcoming*.

Farrokhi, F. (2018). Skill, agglomeration, and inequality in the spatial economy. Technical report.

Fujita, M., Krugman, P. R., and Venables, A. (2001). *The spatial economy: Cities, regions, and international trade*. MIT press.

Fujita, M. and Thisse, J.-F. (2006). Globalization and the evolution of the supply chain: Who gains and who loses? *International Economic Review*, 47(3):811–836.

Gaubert, C. (2014). Firm sorting and agglomeration. *Mimeo*.

Giannone, E. (2017). Skilled-biased technical change and regional convergence. Technical report, University of Chicago Working Paper, available at: http://home. uchicago. edu/~ elisagiannone/files/JMP_ElisaG. pdf.

Glaeser, E. L. and Resseger, M. G. (2010). The complementarity between cities and skills*. *Journal of Regional Science*, 50(1):221–244.

Goldin, C. D. and Katz, L. F. (2009). *The race between education and technology*. Harvard University Press.

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

Head, K. (2003). Gravity for beginners. *University of British Columbia*, 2053.

Hummels, D., Jørgensen, R., Munch, J., and Xiang, C. (2014). The wage effects of offshoring: Evidence from danish matched worker-firm data. *The American Economic Review*, 104(6):1597–1629.

Hummels, D. and Lee, K. Y. (2017). The income elasticity of import demand: Micro evidence and an application. Technical report, National Bureau of Economic Research.

Katz, L. F. et al. (1999). Changes in the wage structure and earnings inequality. *Handbook of labor economics*, 3:1463–1555.

Kennan, J. and Walker, J. R. (2011). The effect of expected income on individual migration decisions. *Econometrica*, 79(1):211–251.

Krugman, P. (1980). Scale economies, product differentiation, and the pattern of trade. *The American Economic Review*, 70(5):950–959.

Krugman, P. (1991a). History and industry location: the case of the manufacturing belt. *The American Economic Review*, 81(2):80–83.

Krugman, P. (1991b). Increasing returns and economic geography. *The Journal of Political Economy*, 99(3):483–499.

Lindley, J. and Machin, S. (2014). Spatial changes in labour market inequality. *Journal of Urban Economics*, 79:121–138.

Matano, A. and Naticchioni, P. (2011). Wage distribution and the spatial sorting of workers. *Journal of Economic Geography*, 12(2):379–408.

McCarty, N., Poole, K. T., and Rosenthal, H. (2016). *Polarized America: The dance of ideology and unequal riches*. mit Press.

Monte, F., Rossi-Hansberg, E., and Redding, S. J. (2015). Commuting, migration, and local employment elasticities.

Moretti, E. (2013). Real wage inequality. *American Economic Journal: Applied Economics*, 5(1):65–103.

Oceanic, U. N. and Administration, A. (2016). Us temperature and humidity raster maps. https://www.climate.gov/maps-data. Accessed: 2016-09.

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.

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

Simonovska, I. and Waugh, M. E. (2014). The elasticity of trade: Estimates and evidence. *Journal of international Economics*, 92(1):34–50.

Tombe, T. and Zhu, X. (2019). Trade, migration and productivity: A quantitative analysis of china. *American Economic Review, forthcoming*.

Wilkinson, R. G. and Pickett, K. E. (2006). Income inequality and population health: a review and explanation of the evidence. *Social science & medicine*, 62(7):1768–1784.

Ziv, O. (2017). Geography in reduced form.

Appendix For Online Publication

A   Data appendix

In this appendix, we describe in detail the data we used, where we got it, and how we processed it. The goal is that a researcher wishing to replicate our analysis will be able to use this section and code available on our website to exactly replicate and understand our results.

A.1   Some conceptual issues on the geographical unit

Many authors in the urban economics literature have used the same IPUMs 5% sample. IPUMs data only reliably report a PUMA for each individual. An individual's MSA and county are only reported when there is no ambiguity about her location. If an individual resides in a PUMA which straddles the border of a MSA, then she will be reported without an MSA. Of all observations potentially in an MSA, only 80% can be determined to actually live inside the metro area. The problem is even larger with counties. We can only unambiguously place individuals in 423 of the 3007 American counties. Observations with non-PUMA identifiers in IPUMs data are likely unrepresentative of the true populations in those locations. On the other hand, while we can reliably place census observations into PUMAs, PUMAs are undesirable as a unit of analysis. PUMAs are not economically meaningful, and the area of a PUMA varies widely with population density.

In light of these data issues, we follow the methodology proposed in a recent working paper to recover CBSA data aggregates (Baum-Snow et al., 2014). To construct aggregates, we weight census observations based on 2003 PUMA populations and the fraction of each PUMAs population residing in each CBSA. This information is available from the Missouri Census Data Center. The strong assumption required for this method to be valid is that population within a PUMA is distributed uniformly with respect to the data aggregates in which we are interested.

A.2   Commodity Flow Survey coverage

Data from the 2007 Commodity Flow Survey is widely used by transportation researchers. The goal of the Commodity Flow Survey is to estimate the volume and mode of domestic shipments by commodity at various levels of geographic aggregation. The survey is mandatory and given to a sample of American manufacturing, wholesale, and certain types of retail establishments. A surveyed establishment is required to fill out a quarterly questionnaire for one year about shipments it sent out over the last week. Surveys are combined with sampling weights to produce estimates of total trade flows by commodity. For more information on this survey, see the official Commodity Flow Survey website https:www.census.goveconcfs.

A.3 Industry and Occupation Classification

Industries
1. Agriculture, Forestery, and Fisheries
2. Mining
3. Construction
4. Food and kindred products:
5. Textile mill products:
6. Apparel and other finished textile products
7. Paper and allied products
8. Printing, publishing, and allied industries
9. Chemicals and allied products
10. Petroleum and coal products
11. Rubber and miscellaneous plastics products
12. Leather and leather products
13. Lumber and wood products, except furniture
14. Stone, clay, glass, and concrete products
15. Metal industries
16. Machinery and computing equipment
17. Electrical machinery, equipment, and supplies
18. Transportation equipment
19. Professional and photographic equipment, and watches
20. Transportation
21. Communications
22. Utilities and sanitary services
23. Wholesale Trade
24. Retail Trade
25. Finance, Insurance, and Real Estate
26. Business and Repair Services
27. Personal Services
28. Entertainment and Recreation Services
29. Professional and Related Services
30. Public Administration

Occupations
1. Executive, Administrative, and Managerial Occupations
2. Management Related Occupations
3. Engineers, Architects, and Surveyors
4. Technical, Sales, and Administrative Support Occupations
5. Sales Occupations
6. Administrative Support Occupations, Including Clerical
7. Private Household Occupations
8. Protective Service Occupations
9. Service Occupations, Except Protective and Household
10. Farm Operators and Managers
11. Other Agricultural and Related Occupations
12. Mechanics and Repairers
13. Mechanics and Repairers, Except Supervisors
14. Construction Trades
15. Extractive Occupations
16. Precision Production Occupations
17. Machine Operators, Assemblers, and Inspectors
18. Transportation and Material Moving Occupations
19. Math and Computer Scientists, Natural Scientists, Teachers(Postsecondary), Social Scientists and Urban Planners
20. Health Diagnosing Occupations, Health Assessment and Treating Occupations, Therapists
21. Teachers(Except Postsecondary), Librarians, Archivists, and Curators, Social, Recreation, and Religious Workers$
22. Lawyers and Judges
23. Writers, Artists, Entertainers, and Athletes

Table 11: Industry and Occupation Classification

A.4 Detailed replication instructions

Researchers can find a full replication data set as well as detailed replication instructions on both authors' personal websites.

B Additional Regression tables

The following three tables document the correlations presented in scatterplots in Section 2.3. Table 12 contains regressions documenting facts from the literature on the positive correlation between wages, skill wage premia, and skill population ratio with population at the level of cities. The skill wage premium with skill population ratio are both positively correlated with

population. Table 13 documents the positive relationship between the skill wage premium and the skill population ratio. The point estimates vary significantly, but the relationship is positive across a number of specifications.

Table 14 documents the relationship between skill wage premium and remoteness. The relationship is negative and statistically significant in the simple regression of skill premium against remoteness (columns 1-2). The correlation is smaller in size, but still significant, when we control for city population (columns 3-4). However, the coefficient on remoteness loses its significance when in addition to city population we include state fixed effects (columns 5-6).

	Log wage	Log skill wage prem	Log skill pop ratio
Log population	0.0540***	0.0165***	0.207***
	(0.00355)	(0.000772)	(0.00814)
Constant	2.002***	0.242***	-3.214***
	(0.0414)	(0.0150)	(0.122)
Observations	1,267	1,267	1,267
R-squared	0.796	0.595	0.714
State FE	YES	YES	YES

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

Table 12: Regressions documenting the relationship between wage, skill premium, skill ratio and population

Dependent variable: Log skill wage premium		
Log skill pop ratio	0.0553***	0.0592***
	(0.00710)	(0.00406)
Constant	0.416***	0.479***
	(0.00777)	(0.0134)
Observations	1,267	1,267
R-squared	0.200	0.578

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

Table 13: Regressions documenting the relationship between skill wage premium and skill population ratio

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: Log skill wage premium					
Log remoteness	-0.191***	-0.126***	-0.0942***	-0.103***	0.0242	-0.103***
	(0.0181)	(0.0140)	(0.0262)	(0.0143)	(0.0447)	(0.0143)
Log population			0.0136***	0.0112***	0.0170***	0.0112***
			(0.00187)	(0.00136)	(0.00113)	(0.00136)
Constant	0.413***	0.367***	0.226***	0.251***	0.230***	0.251***
	(0.00515)	(0.00403)	(0.0245)	(0.0152)	(0.0250)	(0.0152)
Observations	1,267	1,267	1,267	1,267	1,267	1,267
R-squared	0.235	0.064	0.367	0.103	0.610	0.103
State FE	N	N	N	N	Y	N
Pop Weight	Y	N	Y	N	Y	N

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

Table 14: Regressions documenting the relationship between remoteness and the skill wage premium

C   Comparing our Trade Cost Estimates to Allen and Arkolakis (2014)

Our baseline estimates for distance costs vary quantitatively from those estimated using the same data and methodology developed in Allen and Arkolakis (2014). As mentioned in the main text, because of differences in scaling, we shouldn't expect our estimates to be at the same absolute level, but we might expect proportionality. Allen and Arkolakis were kind enough to provide a main Matlab estimation file to us soon after we began working on this project. We downloaded and cleaned the input files ourselves from the same sources as the previous study, and wrote several small functions which were omitted from the original code provided to us. Thus, we were quite surprised to find that our estimates vary from the original. In particular, we estimate a much higher variable cost for water and air than do Allen and Arkolakis. The rank of our estimates for fixed costs are the same as Allen and Arkolakis. The rank of our variable costs are the same, except for water transport which we estimate to be the most expensive form of transport. See Table 15 columns (1) and (2) below.

Allen and Arkolakis later released full replication code for their paper. Below we compare our estimates in more detail. While we were not able to make our results match theirs exactly, we can find several differences which explain part of the gap. We classify these differences into three categories:

1. **Differences in specification**: In our estimation we use a value of $\sigma = 4$ to be consistent

with our structural model. Allen and Arkolakis (2014) use $\sigma = 9$. This difference alone does not explain the different estimates. In this section, all reported estimates are for $\sigma = 9$ whether it be our estimation or those of Allen and Arkolakis. Furthermore, in our estimation because our water map is somewhat different from that used in Allen and Arkolakis (discussed below), we penalize off water transport when shipping by water less than Allen and Arkolakis. Allen and Arkolakis assume that when shipping via water, it is ten times as expensive to transverse a non-water pixel on the map than a water pixel. In our baseline we assume it is only 3.5 times as expensive. For the purposes of comparison, in this section all reported estimates use the Allen Arkolakis value of ten.

2. **Differences in data**: The input value of truck transport in Allen and Arkolakis' replication data is exactly twice what is reported in the 2007 Commodity Flow Survey data we downloaded. It appears this is a bug. In column (3) in Table 15, we run the code of Allen and Arkolakis with half the value of truck transportation in their replication data. This change increases the estimated value of the variable cost of water and air transport, bringing their estimates closer to ours (though still significantly lower).

   In the Commodity Flow Survey data, pure water transport and pure rail transport are separated from transport via water and truck and rail and truck.[39] Allen and Arkolakis use only pure water and rail transport in their input data, whereas we count both categories. In column (4) we run our code using only pure water and pure rail figures. This changes our estimates, but does not bring them significantly closer to those in Allen and Arkolakis.

   The fast marching algorithm used to compute distances between locations uses maps of the United States. The maps we use to compute distances for road and rail are visually nearly exactly the same as those in Allen and Arkolakis. The water maps differ, however. We allow (cheap) water transport only along common shipping routes in the ocean. Allen and Arkolakis allow water transport along any part of the ocean. Column (5) reports the results when run our baseline code with a water map similar to that of Allen and Arkolakis. That is, we use our map, but also allow cheap movement in the coastal waters around the United States. This change hardly affects our baseline estimates. A difference we did not examine, but could potentially affect estimates is that our maps and those of Allen and Arkolakis use different projections. Ours use the projection NAD83:4269.[40]

3. **Differences in code**: Allen and Arkolakis estimate the parameters for the shippers' discrete choice of mode of transport minimizing the following loss function. Let $\varepsilon(\beta)_{od}^m$ be the difference between the predicted and observed fraction of shipments of mode m

---

[39]Explicit category definitions for CFS data can be found here: www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/co

[40]In addition to these differences, the coordinates used by Allen and Arkolakis for CFS areas appear to be rounded as is typical when exporting data from Stata. Their coordinates range from -2.2 to 2.1 million on the x-axis and from -1.2 to 1.4 million on the y-axis. All coordinates with absolute value above one million have the final five digits rounded to zero. As we were unable to precisely link our data sets by CFS region, the extent that this affects estimates is not clear.

between origin o and destination d evaluated at parameter vector β. Let N be the total number of bilateral pairs:

$$\sum_m |\frac{1}{N} \sum_o \sum_d \varepsilon(\beta)^m_{od}|$$

Our algorithm minimizes the squared residual:

$$\sum_m \sum_o \sum_d (\varepsilon(\beta)^m_{od})^2$$

The two loss functions deliver qualitatively different solutions to the problem both in Allen and Arkolakis' code and in ours. We show how this affects our results by first running Allen and Arkolakis' baseline code using our data, and then running their code using our data as well as our objective function.[41] In column (6) we see results that are more similar to Allen and Arkolakis than in the baseline. Using our objective function in column (7), we move the results much closer to our baseline. In column (8) we run Allen and Arkolakis' code with their data but with our objective.[42] Column (9) is our code and data, with the Allen and Arkolakis objective. In all of these exercises, we see substantial, but not complete convergence in our respective results. One caveat is that in column (8) the air transport variable costs become even smaller than those estimated in Allen and Arkolakis.

To sum up, we find that the coefficients on all the variable costs except roads are quite sensitive to changes in the exact data and specification used to perform the Allen and Arkolakis bilateral trade cost estimation. Across specifications in Table 15, our estimates of the variable cost of truck transport are fairly similar, at least of the same order of magnitude, as those estimated by Allen and Arkolakis. Since almost all shipping in the continental United States is by truck (more than 97% of the value in our CFS data), the final bilateral trade costs produced by the original Allen and Arkolakis code and our code are quite similar. Table 16 presents summary statistics for our estimates. We conjecture that the estimates of the cost of shipping for other modes of transport are sensitive to specification and inputs, but ultimately matter little for the final bilateral trade costs estimates. To thoroughly examine this conjecture would require a more careful analysis which is out of the scope of this paper.

As a final comment, the results in our paper continue to be based on our baseline estimates. We believe that it is proper to count water and truck as a water shipment and rail and truck as a rail shipment since around 50% of the value of rail shipments in our data also involve trucking, and around 30% of the value of water shipments involve trucking. We also believe that forcing water shipments to be along trade routes to ports is also a realistic assumption, since loading and unloading cargo without a port is costly. Finally, we prefer the smoother least squares loss

---

[41]Because our input data on demographics was not in the same format as in Allen and Arkolakis, in our final gravity regressions we altered Allen and Arkolakis' code to omit demographic similarity between locations. This may be driving some of the results we report here.

[42]Demographic similarity variables are included in the gravity regression here.

Transport Type	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Road var	0.5636	0.4702	0.5675	0.4764	0.4702	0.4760	0.4287	0.4377	0.4654
Rail var	0.1434	0.4174	0.1426	0.4529	0.4174	0.0599	0.3614	0.3541	0.3936
Water var	0.0779	0.7736	0.2153	0.7153	0.7799	0.1770	0.4322	0.6364	0.6049
Air var	0.0026	0.1744	0.0354	0.1200	0.1747	0.0059	0.2622	0.0000	0.4106
Rail fixed	0.4219	0.3729	0.3986	0.4719	0.3726	0.3564	0.1772	0.4217	0.6115
Water fixed	0.5407	0.4126	0.3986	0.5695	0.4175	0.3564	0.2480	0.4853	0.8737
Air fixed	0.5734	0.6769	0.5315	0.7843	0.6763	0.4603	0.3189	0.6691	0.8800

Table 15: Comparing distance estimates in several models

	mean	median	min	max
Allen and Arkolakis	1.1842	1.1531	0.9635	1.5967
Farrokhi and Jinkins	1.2025	1.1702	0.9376	1.7368

Table 16: Summary statistics for geographic component of bilateral trade costs (T_g)

function for estimating the relative shipment costs by mode.

1. AA Baseline

2. FJ baseline

3. AA Half Truck

4. FJ Double Truck / Only water,only rail

5. FJ with AA-like water map

6. AA code FJ data

7. AA code FJ data FJ mode obj

8. AA code/data FJ mode obj

9. FJ code/data AA mode obj

D   Numerical algorithms

D.1   Solving for productivities and amenities

We treat data on wages and employment, w_H, w_L, n_H, and n_L as the outcome of a spatial equilibrium. Given these data and our recovered productivity ratios $\bar{\beta}_H/\bar{\beta}_L$ and amenity ratios $\bar{u}_H/\bar{u}_L$ from the residuals of relative labor demand and supply, the following algorithm solves for productivities inclusive of spillovers A's and amenities of high-skill workers $\bar{u}_H$'s. At these calibrated values of A's and $\bar{u}_H$'s, the model exactly predicts the data on w_H, w_L, n_H, and n_L.

Given that trade costs are symmetric, we reduce the two systems of equations described by (25) using relation (26),

$$A(i)^{1-\sigma} = \lambda W_H^{\frac{1-\sigma}{\delta}} N_H^{\frac{\sigma-1}{\delta\theta}} \nu(i)^{1-\sigma} n_H(i)^{-1} w_H(i)^{-1} b(i) \int_J d(i,j)^{1-\sigma} \nu(j)^{1-\sigma} A(j)^{\sigma-1} \, dj \quad (30)$$

As long as (26) holds, the solution to (30) will be the solution to both systems of equations (25). The following algorithm solves for amenities and productivities up to scale.

1. Start with an initial guess for productivity, $A^{(0)}(i)$.

2. Compute the kernel,

$$K(j,i) \equiv \nu(i)^{1-\sigma} n_H(i)^{-1} w_H(i)^{-1} b(i) d(i,j)^{1-\sigma} \nu(j)^{1-\sigma}$$

3. Define $\kappa \equiv \lambda W_H^{\frac{1-\sigma}{\delta}} N_H^{\frac{\sigma-1}{\delta\theta}}$, and in iteration t, $f(i) \equiv A(i)^{1-\sigma}$. Define

$$\tilde{f}(i) \equiv \frac{f(i)}{\int_J f(i) di}$$

as a normalization that sets the integral over $\tilde{f}$ to one. Then, the system of integral equations described by (30) is equivalent to:

$$\tilde{f}(i) = \kappa \int_J K(j,i) \tilde{f}(j)^{-1} \, dj \quad (31)$$

Initial guess equals $f^{(0)}(i) = A^{(0)}(i)^{1-\sigma}$. In iteration $t \geq 1$, update $f^{(t)}(i)$ according to this updating rule:

$$\tilde{f}^{(t+1)}(i) = \frac{\int_J K(j,i) \tilde{f}^{(t)}(j)^{-1} \, dj}{\int_J \int_J K(j,i) \tilde{f}^{(t)}(j)^{-1} \, dj di} \quad (32)$$

Since we divide integrals in (32), we do not need to know κ to update our guess. If at iteration t, $|\tilde{f}^{(t)}(i) - \tilde{f}^{(t-1)}(i)| < 10^{-12}$ for all i, stop updating and go to the next step. Otherwise, continue iterating using the updating rule (32).

The output of this step is a vector of $\tilde{f}(i)$'s that satisfy (31), and so (30), and so the two systems of equations (25).

4. As a check that the solutions are correct, the following must be a constant equal to κ for all i,

$$\kappa = \frac{\tilde{f}(i)}{\int_J K(j,i) \tilde{f}(j)^{-1} \, dj} = \frac{\tilde{f}(i')}{\int_J K(j,i') \tilde{f}(j)^{-1} \, dj}$$

According to the definitions in Step 3,

$$\frac{A(i)}{A(j)} = \left(\frac{f(i)}{f(j)} \right)^{\frac{1}{1-\sigma}} = \left(\frac{\tilde{f}(i)}{\tilde{f}(j)} \right)^{\frac{1}{1-\sigma}}$$

Normalize $A(i_0) = 1$ for city i_0, and calculate all other $A(i)$'s.

5. Using equation (26),

$$\frac{\bar{u}_H(i)}{\bar{u}_H(j)} = \frac{C(i)^{1-\delta} n_H(i)^{\frac{\sigma-1-\delta\theta}{(\sigma-1)\theta}} w_H(i)^{\frac{1-\sigma-\delta}{\sigma-1}} b(i)^{\frac{\delta}{\sigma-1}} A(i)^{\delta} \nu(i)^{-\delta}}{C(j)^{1-\delta} n_H(j)^{\frac{\sigma-1-\delta\theta}{(\sigma-1)\theta}} w_H(j)^{\frac{1-\sigma-\delta}{\sigma-1}} b(j)^{\frac{\delta}{\sigma-1}} A(j)^{\delta} \nu(j)^{-\delta}}$$

Normalize the amenity value of city i_0, $\bar{u}_H(i_0) = 1$, and calculate all other $\bar{u}_H(i)$'s.

D.2   Solving for wages and employment

Given model parameters and the four shifters $\bar{A}, \bar{\beta}_H, \bar{u}_H, \bar{u}_L$ we solve for equilibrium wages and employment. First, we write the distribution of low-skill labor as a function of the distribution of high-skill labor. By plugging skill wage premium from (14) into relative labor supply (13),

$$n_L(i) = \left(\frac{W_H}{W_L}\right)^{\frac{\theta\rho}{\theta+\rho}} \left(\frac{N_H}{N_L}\right)^{\frac{-\rho}{\theta+\rho}} \left(\frac{\bar{\beta}_H(i)}{\bar{\beta}_L(i)}\right)^{\frac{-\theta\rho}{\theta+\rho}} \left(\frac{\bar{u}_H(i)}{\bar{u}_L(i)}\right)^{\frac{-\theta\rho}{\theta+\rho}} \left(n_H(i)\right)^{\frac{\theta(1-\rho\varphi)+\rho}{\theta+\rho}} \tag{33}$$

In addition, integral equations (19) and (20) could be equivalently written as

$$A(i)^{1-\sigma} \tilde{\nu}(i)^{\sigma-1} n_H(i) w_H(i)^{\sigma} b(i)^{-1}$$
$$= W_H^{\frac{1-\sigma}{\delta}} N_H^{\frac{\sigma-1}{\delta\theta}} \int_J d(i,j)^{1-\sigma} \bar{u}_H(j)^{\frac{\sigma-1}{\delta}} \tilde{C}(j)^{\frac{(\sigma-1)(\delta-1)}{\delta}} n_H(j)^{\frac{1-\sigma+\delta\theta}{\delta\theta}} w_H(j)^{\sigma} b(j)^{-1} \, dj \tag{34}$$

$$\bar{u}_H(i)^{\frac{1-\sigma}{\delta}} \tilde{C}(i)^{\frac{(\sigma-1)(1-\delta)}{\delta}} n_H(i)^{\frac{\sigma-1}{\delta\theta}} w_H(i)^{1-\sigma}$$
$$= W_H^{\frac{1-\sigma}{\delta}} N_H^{\frac{\sigma-1}{\delta\theta}} \int_J d(j,i)^{1-\sigma} A(j)^{\sigma-1} \tilde{\nu}(j)^{1-\sigma} w_H(j)^{1-\sigma} \, dj \tag{35}$$

where $\tilde{\nu}$ and $\tilde{C}$ are replaced from equations 17-18. The pair of 34–35 (or equivalently the pair of 19–20) give us two integral equations. The two systems can be reduced to one using the following relation, that is equivalent to equation (21),

$$A(i)^{1-\sigma} \tilde{\nu}(i)^{\sigma-1} n_H(i) w_H(i)^{\sigma} b(i)^{-1} = \lambda \bar{u}_H(i)^{\frac{1-\sigma}{\delta}} n_H(i)^{\frac{\sigma-1}{\delta\theta}} w_H(i)^{1-\sigma} \tilde{C}(i)^{\frac{(\sigma-1)(1-\delta)}{\delta}} \tag{36}$$

Given exogenous parameters we can write every endogenous variable as a function of population of high-skill workers $n_H(i)$. Our solution algorithm takes advantage of this feature of the model to update our guess for $n_H(i)$ in each iteration. The algorithm is as follows:

1. Guess $n_H(i)$ for all i.

2. Compute W_H/W_L according to (16). Then plug it in (33) to find $n_L(i)$.

3. Calculate skill premia, $w_H(i)/w_L(i)$, according to (14).

4. Compute $b(i) = 1/(1 + \frac{n_L(i)}{n_H(i)} \frac{w_L(i)}{w_H(i)})$

5. Calculate $\tilde{\nu}(i)$ according to (17) and $\tilde{C}(i)$ according to (18).

6. Let
$$\tilde{w}_H(i) \equiv \lambda^{\frac{-1}{2\sigma-1}} w_H(i)$$

hen, according to (36), $\tilde{w}_H(i)$ is given by:

$$\tilde{w}_H(i) = b(i)^{\frac{1}{2\sigma-1}} A(i)^{\frac{\sigma-1}{2\sigma-1}} \tilde{\nu}(i)^{\frac{1-\sigma}{2\sigma-1}} \bar{u}_H(i)^{\frac{1-\sigma}{\delta(2\sigma-1)}} \tilde{C}(i)^{\frac{(\sigma-1)(1-\delta)}{\delta(2\sigma-1)}} n_H(i)^{\frac{\sigma-1-\delta\theta}{\delta\theta(2\sigma-1)}}$$

7. Let $f(i) \equiv \tilde{w}_H(i)^{1-\sigma}$, $\kappa \equiv W_H^{\frac{1-\sigma}{\delta}} N_H^{\frac{\sigma-1}{\delta\theta}}$, and

$$K(j,i) \equiv \bar{u}_H(i)^{\frac{\sigma-1}{\delta}} \tilde{C}(i)^{\frac{(1-\sigma)(1-\delta)}{\delta}} n_H(i)^{\frac{1-\sigma}{\delta\theta}} d(j,i)^{1-\sigma} A(j)^{\sigma-1} \tilde{\nu}(j)^{1-\sigma}$$

Then, system of integral equations (35) can be written as follows (notice that the scale parameter λ cancels out):

$$f(i) = \kappa \int_J K(j,i) f(j) \ dj \tag{37}$$

The solution to (37) is equivalently the solution to the pair of systems of equations 34–35. In iteration t, update $f^{(t)}(i)$ according to

$$f^{(t+1)}(i) = \frac{\int_J K(j,i) f^{(t)}(j) \ dj}{\int_J \int_J K(j,i) f^{(t)}(j) \ dj di} \tag{38}$$

Equation (38) is our updating rule. Note that we do not need to know κ to update our guess. If $f^{(t+1)}(i)$ is not close enough to $f^{(t)}(i)$, go to step 2 in order to continue iterations. Otherwise, go to the next step.

The output of this step is a vector of $f(i)$'s that satisfy (37) and equivalently the systems of equations 34–35.

8. As $\int_J w_H(j) dj = 1$ (the normalization defined in equilibrium), calculate wages:

$$w_H(i) = \frac{\tilde{w}_H(i)}{\int_J \tilde{w}_H(j) dj}$$

9. Calculate λ:

$$1 = \int_J w_H(j) dj = \lambda^{\frac{1}{2\sigma-1}} \int_J \tilde{w}_H(j) dj$$

So,

$$\lambda = \left[\int_J \tilde{w}_H(j) dj \right]^{-(2\sigma-1)}$$

10. Find κ,

$$\kappa = \frac{f(i)}{\int_J K(j,i) f(j) \ dj} = \frac{f(\ell)}{\int_J K(j,\ell) f(j) \ dj}$$

The above should hold for all i and ℓ. This step, thus, is also a check that the solutions to

integral equations are correct. Then, calculate:

$$W_H = N_H^{\frac{1}{\theta}} \kappa^{\frac{\delta}{1-\sigma}}$$

Once $w_H(i)$ and W_H are known, it is straightforward to calculate all other equilibrium objects.