Wage Inequality and the Location of Cities

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

[1]Purdue
[2]Copenhagen Business School

January 10, 2017

Abstract

In cross-sectional American census data, we document that isolated cities tend to have less wage inequality. 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. The 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 geographical location explains 9.2% of observed variation in wage inequality across American cities. In counterfactual experiments, we find that reductions in domestic trade costs benefit all American workers and decrease welfare inequality. We also examine the effects on inequality and welfare of both regional and national skill-biased technology shocks. We find that in larger cities wage inequality grows more than welfare inequality.

*An earlier version of this paper was circulated with the title "Trade and Inequality in the Spatial Economy". We thank Sina Smid and Karolina Stachlewska for excellent research assistance. We thank Treb Allen, Nathanial Baum-Snow, Jonathan Dingel, Jeff Lin, Tobias Seidal, and participants in seminars at Cardiff University, Copenhagen Business School, Copenhagen University, the Nordic International Trade Seminars, the European Trade Study Group, the North American Regional Science Association, Penn State, and Purdue for helpful suggestions.

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 complimented 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 wage premia. This result holds when controlling for population and 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 (Davis and Dingel, 2014; Baum-Snow and Pavan, 2012; Combes et al., 2012; Moretti, 2013; Davis and Dingel, 2014; Diamond, 2015). 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 (Desmet et al., 2016; Allen et al., 2016). The economic geography literature as a whole has typically focused on welfare at the aggregate (Krugman, 1991; Fujita

[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.

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 average 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. However, while her model equilibrium can not be solved, we can fully solve our model equilibrium at a wide range of counterfactual parameters.

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 greater 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 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 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

[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.

[4]Davis and Dingel (2012) microfound a mechanism for this assumption related to complementary between idea exchange and ability.

these cities. In addition, we estimate costs of trading goods between cities in a similar way to Allen and Arkolakis (2014).

Using our estimated model, we decompose the variation in observed wage premia across American cities. We find that geographical position explains 9.2% of the variation in wage premia across cities. By simulating counterfactual exercises, 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 population out 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 at the presence of agglomeration economies in the national context.[6]

We use our model to perform several counterfactuals. 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 productivity shock would increase the expected welfare of high-skill workers nationally by 1.7% and of low-skill workers nationally by 0.5%.

In a second counterfactual we simulate skill-biased technological change in order to match the observed change in skill wage premia between 1980 and 2000 across the United States. Even though we only increase the productivity of high-skill workers, low skill workers gain as well. We find that the rise in observed high-skill wage premia in larger cities overstates the rise in welfare inequality caused by skill-biased technological change.

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 65 with reported

[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 (2015) finds that domestic reallocation of labor tends to mitigate the increase in inequality caused by an international trade liberalization.

income, giving us observations on over four million workers distributed across the United States.[7]

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). 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). PUMA's 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.

Many authors in the urban economics literature have used the same IPUM's 5% sample. In the course of cleaning and understanding these data for our project we discovered some important data issues which have received little discussion in the literature. In particular, IPUM's 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 IPUM's data are likely unrepresentative of the true populations in those locations. On the other hand, while we can reliably place census observations into PUMA's, PUMA's are undesirable as a unit of analysis. PUMA's 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 PUMA's 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.

In addition to the IPUM's 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 US Commodity Flow Survey. 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. For a more complete discussion of data sources and manipulation, see the data appendix.

[7]We clean the data using modified replication code from Baum-Snow and Pavan (2013). For more information on how the data were cleaned, see the data appendix.

2.2 Location specific variables

We assign each location two geographic measures: distance to ocean and remoteness. 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 $i = 1 \dots N$. 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, R_i, is the weighted, generalized mean of the distances between location i and all other locations:

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

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 w_i related to economic size. The theoretical model we develop features such a price index, and σ in our model is interpretable as the elasticity of substitution in utility of goods produced in different locations. We set $\sigma = 4$ to remain consistent with our benchmark calibration of the full structural model. We use both unweighted and population-weighted measures of remoteness in the motivation for our modeling exercise below. After estimating our model, we will summarize the geographic position of a city with respect to other cities based on the structurally derived price index of tradeables.

The remoteness index is meant to measure a location's isolation from other domestic locations. We use distance from the ocean to proxy for a location's isolation from international trade.[8] We measure nearest distance to the ocean as the crow flies using data from *Natural Earth*.[9] This data comes at a very fine level. To aggregate up to the level of our locations, we assign each location the mean distance to the ocean within its borders.[10]

The high-skill wage premium (or college wage premium) is measured as is standard in the labor literature, and is calculated independently for each location. A worker is high-skill if he has at least a four-year college degree. The high-skill wage premium is the mean wage of high-skill workers in a location divided by the mean wage of other workers. The high-skill population ratio (or college population ratio) is the ratio of high-skill population to other workers' population in a location. We use census population weights when calculating all means.

Table 1 contains some descriptive statistics, and Figure 1 shows how our measures vary across the United States. The borders in this map are the intersection of PUMA's and CBSA's, but are colored based on the geographical unit of analysis described in Section 2.1. Remoteness

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

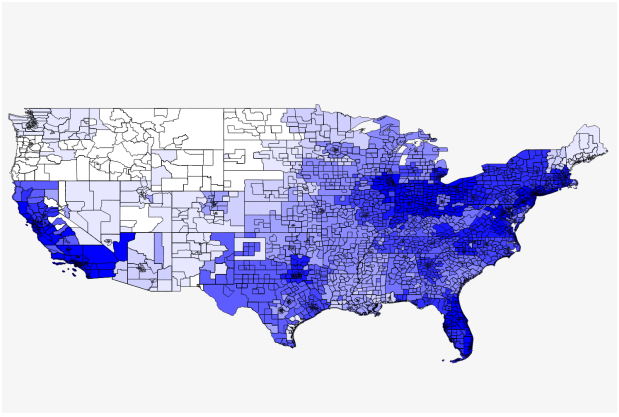
[9] *Natural Earth* is a free source of physical geographical data in the public domain maintained by the North American Cartographic Information Society. More information at naturalearthdata.com .

[10] 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 logorithms, so scaling errors will only affect the constant.

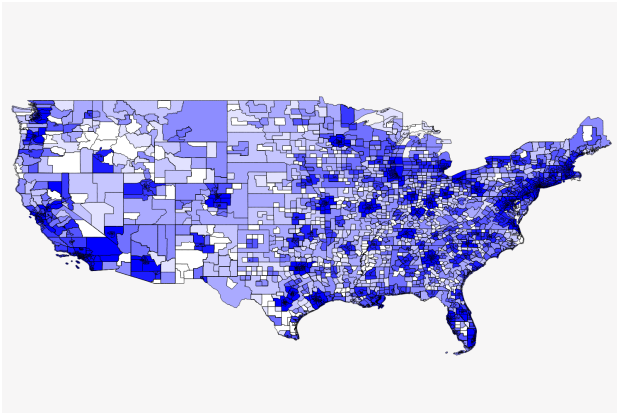
is highest in the North and North-West of the United States. Population is highest in CBSA's. The high-skill wage premium is higher in the parts of the country which are relatively less remote and with higher population. The high-skill population ratio is relatively high in CBSA's as well as the Rocky Mountain states such as Montana and Idaho.

Statistic	Mean	Standard Dev	Min	Max
Distance to the ocean	533	346	0.39	1586
Remoteness	1.26	0.17	1.04	1.75
Remoteness (pop wts)	1.30	0.18	0.93	1.90
Population	51k	196k	363	4.1m
High-skill wage premium	1.54	0.13	1.20	2.21
Census observations	4m			
Location observations	1267			

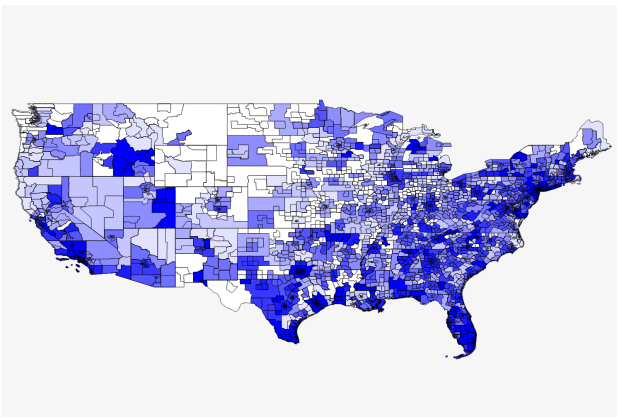
Table 1: Data summary statistics



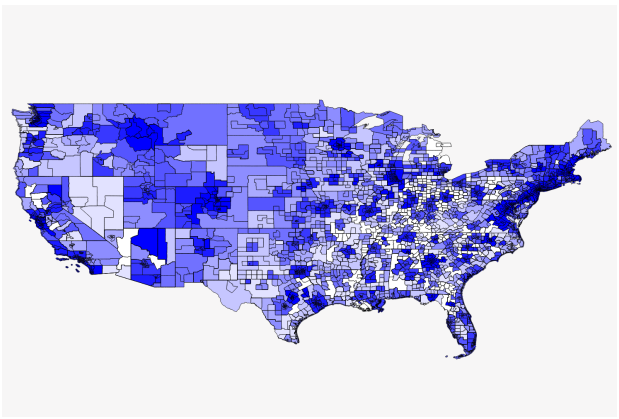
(a) Remoteness (weighted)



(b) Population



(c) High-skill Wage Premium



(d) High-skill Population Ratio

Figure 1: Locations colored by attribute

2.3 Patterns in the Data

In this section we document the covariance of our measures of geography with the high-skill wage premium, as well as other correlations that emerge from the data. Table 2 contains coefficient estimates from regressions of high-skill wage premium on various measures of geography. We find that locations that are more remote within the United States or more distant from coastlines have less wage inequality, even after controlling for population. This relationship is significantly different from zero in all but one of the specifications presented in the table. 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, except that unweighted remoteness becomes significant at the 10% level in column (2). We cluster our robust standard errors at the state level.

We report some additional specifications in Appendix B. In particular, we also use individual-level data as observations. We show that even after controlling for a wide range of individuals' characteristics as well as population of their location, at a highly statistically significant level remoteness lowers wages, and lowers wages of college-educated workers more than non-college educated workers. The overall message of these regressions is that the geographic features of cities correlate with wage inequality across a wide range of specifications and even after controlling for population.

We present other correlations in our data using raw data scatter plots in Figure 2. Each scatter plot contains a fitted regression line weighted by population. We would like our model to be able to generate the following relationships we see in the data:

1. High-kill wage premium and high-skill population ratio are positively correlated.
2. High-skill population ratio and high-skill wage premium are positively correlated with population.
3. High-skill wage premium and high-skill population ratio are negatively correlated with measures of geographical isolation.
4. Remoteness and population are negatively correlated

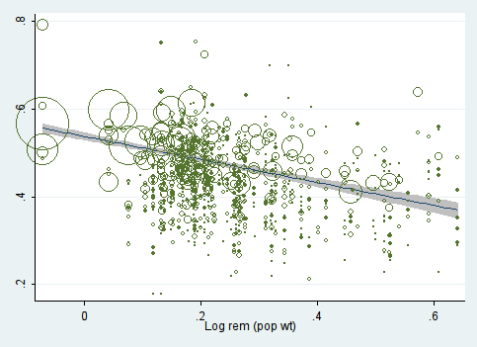
These relationships are all statistically significant at conventional levels. If we remove the population weights, the relationships do not change save for the negative correlation between the high-skill population ratio and our geographical measures. The correlations with both remoteness and distance to ocean become positive rather than negative. The positive relationship between population-weighted remoteness and high-skill population ratio is statistically significant. Some of this is explained by remote locations containing land grant universities, and partly it may be sparsely populated areas in the mountain region of the United States. If we restrict our analysis to locations with population greater than 15,000 there are no significant relationships.[11]

[11]For additional analysis using regression tables, see Appendix B.

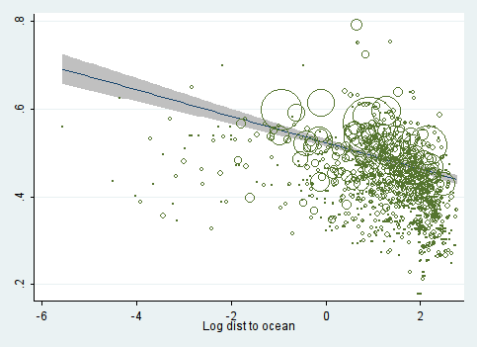
VARIABLES	(1) Wage prem	(2) Wage prem	(3) Wage prem	(4) Wage prem	(5) Wage prem	(6) Wage prem
Log dist to ocean	-0.0302*** (0.00462)		-0.0347*** (0.00552)		-0.0296*** (0.00343)	
Log rem		-0.0294 (0.0734)	-0.127*** (0.0451)		-0.175*** (0.0340)	
Log rem (pop wt)				-0.262*** (0.0317)		-0.131** (0.0505)
Log pop					0.0232*** (0.00139)	0.0188*** (0.00307)
Constant	0.523*** (0.00997)	0.498*** (0.0197)	0.560*** (0.0175)	0.537*** (0.00904)	0.279*** (0.0191)	0.280*** (0.0426)
Observations	1,267	1,267	1,267	1,267	1,267	1,267
R-squared	0.164	0.002	0.198	0.229	0.498	0.379
State Clust	YES	YES	YES	YES	YES	YES
Pop Weight	YES	YES	YES	YES	YES	YES

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

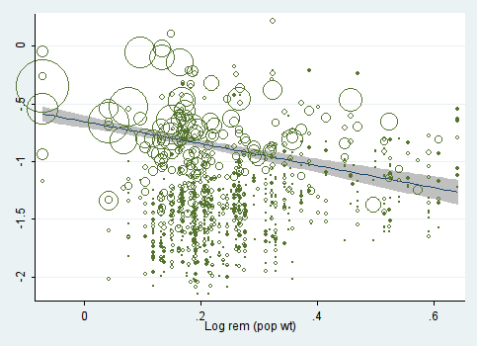
Table 2: Log high-skill wage premium vs geography regressions



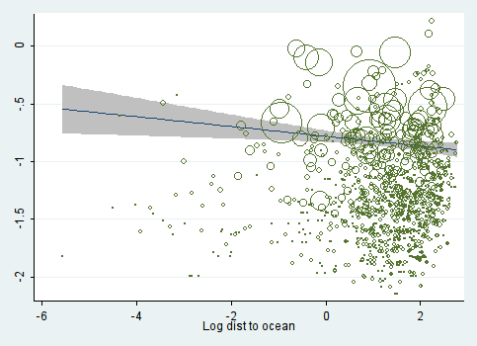
(a) High-skill wage prem. vs remoteness (pop wt)



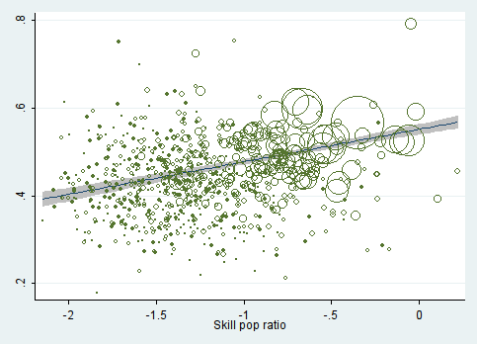
(b) High-skill wage prem. vs distance to ocean



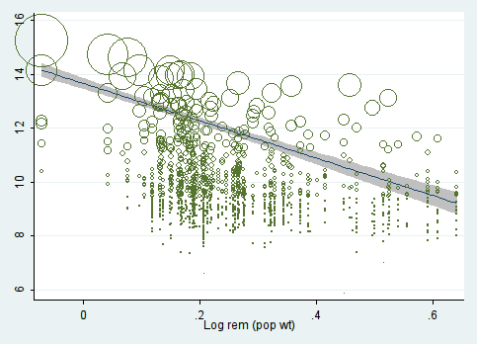
(c) High-skill pop. ratio vs remoteness (pop wt)



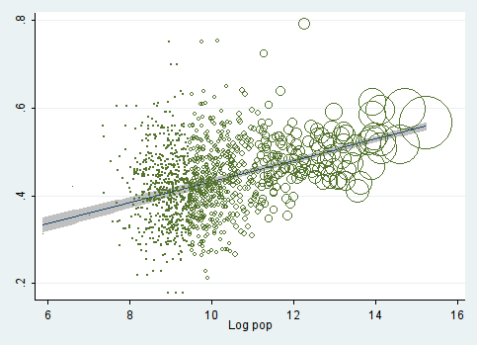
(d) High-skill pop. ratio vs distance to ocean



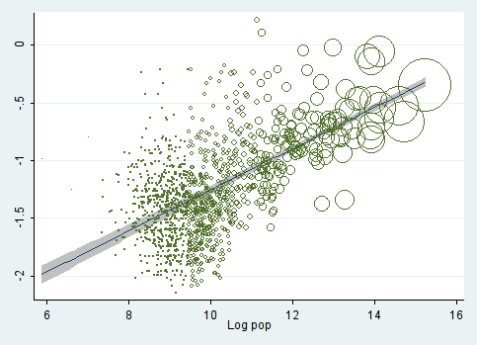
(e) High-skill wage prem. vs high-skill pop. ratio



(f) Population vs remoteness (pop wt)



(g) High-skill wage prem. vs population



(h) High-skill pop. ratio vs population

Figure 2: Data correlations

3 Theory

In the last section, we presented evidence that geography plays a role in shaping the college wage premia. In order to introduce a mechanism that can explain our findings, in this section we embed inequality in an empirical economic geography framework. That is, we develop a model with costly trade between locations. We are interested in the equilibrium responses of inequality to economic shocks, so it is important that our model will be in general equilibrium, with both heterogenous labor and goods mobile across locations. After developing the model, we discuss the intuition for how our model generates spatial patterns of inequality in Section 3.1.4.

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 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. Both workers and firms are price takers in perfectly competitive markets. In equilibrium, welfare of each skill group equalizes across space for marginal workers.

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, $\varepsilon_\omega(i)$,

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

All workers with the same skill who reside in the same location earn the same wages and consume the same amount of tradeables and housing. $\delta \in (0,1)$ is the share of expenditures on tradeables. 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,

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

where $R(i)$ is price per unit of land 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 worker-specific idiosyncratic preference shock, ε, is independent across workers and locations. Its inverse, $\tilde{\varepsilon} \equiv 1/\varepsilon$, has a Fréchet distribution given by

$$\Pr(\tilde{\varepsilon} \leq x) = \exp(-x^{-\theta}).$$

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

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

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 pins down the price per unit of land,

$$\begin{aligned} R(i)\bar{Z}(i) &= (1-\delta)(n_L(i)w_L(i) + n_H(i)w_H(i)) + (1-\delta)R(i)\bar{Z}(i) \\ R(i) &= \frac{1-\delta}{\delta \bar{Z}(i)}\Big(n_L(i)w_L(i) + n_H(i)w_H(i)\Big), \end{aligned} \tag{5}$$

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 housing, given by $P(i)^{\delta}R(i)^{1-\delta}$. 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:

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

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

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

The elasticity of relative labor supply to relative wages is:

$$\frac{\partial \log\Big(n_s(i)/n_s(j)\Big)}{\partial \log\Big(w_s(i)/w_s(j)\Big)} = \theta.$$

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. Another way of putting it is that 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:

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

This index is proportional to the expected welfare of a worker of type s before she draws her location preferences.[12] The share of workers of type s in location i is given by:

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

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

Workers are low or high-skill. Each location has a measure one of homogeneous firms with a CES production under constant returns to scale,

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

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 of economic geography model with monopolistic competition and free

[12] 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.

entry generates the same relation through endogenous measure of firms, with the exact relation if $\alpha = 1/(1+\sigma)$.

In addition, the empirical literature on urban and labor economics substantiates that agglomeration forces are stronger for high skill workers.[13] On the theoretical side, the literature explains this fact by modeling spillovers thorough the exchange of ideas within high-skill workers (Davis and Dingel, 2012). In order to capture this mechanism in our empirical model, we let skill worker's productivity covary positively with the population of skill 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 normalization, we assign no further agglomeration benefit to low-skill labor. By cost minimization, the unit cost of production equals

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

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 $d(j,i)$ be the trade cost of shipping a good from j to i. The price of a good produced in location j and consumed in location i is:

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

3.1.3 Spatial equilibrium

We derive equations that characterize equilibria, define a spatial equilibrium, and provide intuition behind our equations.

On the demand side of labor market, payments to skill labor equals firms' spendings 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) in this relation, delivers the relative labor demand,

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

On the supply side, employment shares described by equation (6) result in the relative labor

[13] For example, Glaeser and Resseger (2010) states 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.

supply,

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

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

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

Total wages to labor in location i equal $\frac{w_H(i)n_H(i)}{b(i)}$, and total income (wages plus housing rents) equals $\frac{w_H(i)n_H(i)}{\delta b(i)}$. Both workers and landlords spend δ share of their income on tradeable goods and the rest on the residential land. The labor market clearing also requires total wages received by labor to be equal to total payments to labor,

$$\begin{aligned} \text{total wages in } i &= \int_J \delta\Big[\text{total income in } j\Big]\Big[\frac{p(i,j)}{P(j)}\Big]^{1-\sigma} dj \\ \frac{w_H(i)n_H(i)}{b(i)} &= \int_J \frac{w_H(j)n_H(j)}{b(j)}\Big[\frac{p(i,j)}{P(j)}\Big]^{1-\sigma} dj \end{aligned} \tag{15}$$

Equations 14 and 15 together describe labor market clearing (in relative terms and in levels). In addition, equation 15 also describes the goods market clearing.

Definition. A "spatial equilibrium" is a set of $w_H(i)$, $w_L(i)$, $n_H(i)$, and $n_L(i)$ such that:

1. Share of workers of each skill group in each location is give by (6).
2. Demand and supply of high and low-skill workers are equal in each location, according to (12), (13), and (14).
3. Expenditures of each skill group in location i from goods produced in location j, $x_s(j,i)$, price of good i in j, $p(j,i)$, price index of tradeables $P(i)$, and housing rents $R(i)$, are respectively given by (3), (11), (4), and (5).
4. Share of spending on high-skill workers is given by (10).
5. Allocation of labor is feasible, $\int_J n_H(j)\,dj = N_H$ and $\int_J n_L(j)\,dj = N_L$. Wages of high-skill workers are normalized such that $\int_J w_H(j)dj = 1$.

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. Besides, 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 more skill workers due to agglomeration, but since unobserved location preferences matter, high-skill workers do not fully arbitrage the wage increase away.

To provide further intuition on the mechanism through which our model generates spatial variation in wage premia, suppose we build a new airport in a remote city. As a result of reductions in trade costs, the price of incoming tradeables falls so the supply of workers of both types to the city increases. In addition, outgoing sales to other cities rise, so the demand of firms in the city for workers of both types increases. Suppose the population of low and high-skill workers increase proportionately. Due to agglomeration advantages, however, high-skill workers become relatively more productive. That is, firms demand a larger ratio of high to low-skill workers than their relative supply in the city. Equilibrium is restored by raising high-skill wage premium in the city.

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

The distribution of low-skill labor can be written as a function of the distribution of high-skill labor. By plugging high-skill wage premium from (14) into relative labor supply (13),

$$n_L(i) = \Big(\frac{W_H}{W_L}\Big)^{\frac{\theta\rho}{\theta+\rho}} \Big(\frac{N_H}{N_L}\Big)^{\frac{-\rho}{\theta+\rho}} \Big(\frac{\bar{\beta}_H(i)}{\bar{\beta}_L(i)}\Big)^{\frac{-\theta\rho}{\theta+\rho}} \Big(\frac{\bar{u}_H(i)}{\bar{u}_L(i)}\Big)^{\frac{-\theta\rho}{\theta+\rho}} \Big(n_H(i)\Big)^{\frac{\theta(1-\rho\varphi)+\rho}{\theta+\rho}} \tag{16}$$

Using the above relation, we show that the spatial distribution of workers contributes to welfare inequality. Equation (16) and $N_L = \int_J n_L(i)\, di$ pin down the relative average well-being of high to low-skill workers, W_H/W_L, as a function of $n_H(i)$,

$$\begin{aligned} \frac{W_H}{W_L} \;=\; & \underbrace{\Big(\frac{N_H}{N_L}\Big)^{-\frac{1}{\rho}}}_{\text{aggergate scarcity}} \times \underbrace{(N_H)^{\varphi}}_{\text{aggregate agglom.}} \times \\ & \underbrace{\left[\int_J \Big(\frac{\bar{\beta}_H(i)\bar{u}_H(i)}{\bar{\beta}_L(i)\bar{u}_L(i)}\Big)^{\frac{-\theta\rho}{\theta+\rho}} \Big(\frac{n_H(i)}{N_H}\Big)^{\frac{\theta(1-\rho\varphi)+\rho}{\theta+\rho}} di\right]^{-\frac{\theta+\rho}{\theta\rho}}}_{\text{distributional effect}} \end{aligned} \tag{17}$$

Equation (17) decomposes the three forces behind real well-being inequality: 1) Aggregate scarcity of high-skill relative to low-skill workers, 2) Aggregate agglomeration advantage of high-skill workers, 3) A term that summarizes dispersion forces. This term is weighted average of relative exogenous amenity values and productivities in which weights are determined by the distribution of the density of high-skill workers. While the first two behave at the aggregate, the third relates to the spatial distribution of workers. In this sense, Equation (17) decomposes

an index of welfare inequality into the three forces: scarcity, agglomeration, and dispersion.

3.2 Solving for spatial equilibrium

In this section we characterize model equilibria with two integral equations. These integral equations together with equations 12–17 will be the basis of our empirical exercise.

Using equations (8), (9), and (10) we can write $c(i)$ as a function of employment, input expenditure share, and wages of high-skill workers:

$$c(i) = \tilde{c}(i) w_H(i), \quad \text{where} \quad \tilde{c}(i) \equiv \left[\bar{\beta}_H(i) n_H(i)^{\varphi}\right]^{\frac{\rho}{1-\rho}} \left[b(i)\right]^{\frac{-1}{1-\rho}} \tag{18}$$

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

$$R(i) = \tilde{R}(i) w_H(i), \quad \text{where} \quad \tilde{R}(i) \equiv \frac{(1-\delta) n_H(i)}{\delta b(i)} \tag{19}$$

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

$$\begin{aligned} &A(i)^{1-\sigma} c(i)^{\sigma-1} n_H(i) w_H(i) b(i)^{-1} \\ &= W_H^{\frac{1-\sigma}{\delta}} N_H^{\frac{\sigma-1}{\delta\theta}} \int_J d(i,j)^{1-\sigma} u(j)^{\frac{\sigma-1}{\delta}} R(j)^{\frac{(\sigma-1)(\delta-1)}{\delta}} n_H(j)^{\frac{1-\sigma+\delta\theta}{\delta\theta}} w_H(j)^{\frac{\sigma-1+\delta}{\delta}} b(j)^{-1} \, dj \end{aligned} \tag{20}$$

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

$$\bar{u}_H(i)^{\frac{1-\sigma}{\delta}} R(i)^{\frac{(\sigma-1)(1-\delta)}{\delta}} n_H(i)^{\frac{\sigma-1}{\delta\theta}} w_H(i)^{\frac{1-\sigma}{\delta}} = W_H^{\frac{1-\sigma}{\delta}} N_H^{\frac{\sigma-1}{\delta\theta}} \int_J d(j,i)^{1-\sigma} A(j)^{\sigma-1} c(j)^{1-\sigma} \, dj \tag{21}$$

Integral equations 20 and 21 could be equivalently written as

$$\begin{aligned} &A(i)^{1-\sigma} \tilde{c}(i)^{\sigma-1} n_H(i) w_H(i)^{\sigma} b(i)^{-1} \\ &= W_H^{\frac{1-\sigma}{\delta}} N_H^{\frac{\sigma-1}{\delta\theta}} \int_J d(i,j)^{1-\sigma} u(j)^{\frac{\sigma-1}{\delta}} \tilde{R}(j)^{\frac{(\sigma-1)(\delta-1)}{\delta}} n_H(j)^{\frac{1-\sigma+\delta\theta}{\delta\theta}} w_H(j)^{\sigma} b(j)^{-1} \, dj \end{aligned} \tag{22}$$

$$\begin{aligned} &\bar{u}_H(i)^{\frac{1-\sigma}{\delta}} \tilde{R}(i)^{\frac{(\sigma-1)(1-\delta)}{\delta}} n_H(i)^{\frac{\sigma-1}{\delta\theta}} w_H(i)^{1-\sigma} \\ &= W_H^{\frac{1-\sigma}{\delta}} N_H^{\frac{\sigma-1}{\delta\theta}} \int_J d(j,i)^{1-\sigma} A(j)^{\sigma-1} \tilde{c}(j)^{1-\sigma} w_H(j)^{1-\sigma} \, dj \end{aligned} \tag{23}$$

where $\tilde{c}$ and $\tilde{R}$ are replaced from equations 18-19. The pair of 20–21, or equivalently 22–23 give 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 integral equations must hold:

$$A(i)^{1-\sigma}c(i)^{\sigma-1}n_H(i)w_H(i)b(i)^{-1} = \lambda\bar{u}_H(i)^{\frac{1-\sigma}{\delta}}n_H(i)^{\frac{\sigma-1}{\delta\theta}}w_H(i)^{\frac{1-\sigma}{\delta}}R(i)^{\frac{(\sigma-1)(1-\delta)}{\delta}} \tag{24}$$

Or, equivalently,

$$A(i)^{1-\sigma}\tilde{c}(i)^{\sigma-1}n_H(i)w_H(i)^{\sigma}b(i)^{-1} = \lambda\bar{u}_H(i)^{\frac{1-\sigma}{\delta}}n_H(i)^{\frac{\sigma-1}{\delta\theta}}w_H(i)^{1-\sigma}\tilde{R}(i)^{\frac{(\sigma-1)(1-\delta)}{\delta}} \tag{25}$$

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 $\rho = \infty$.[14]

Uniqueness. The standard proof of equilibrium uniqueness in the existing literature depends on a specification that allows logarithmic relationships between a subset of endogenous variables (Allen and Arkolakis, 2014). For example, $A = \bar{A}n^{\gamma}$ where we mean to solve for n. In our model, such logarithmic relation violates more than once. For example, since $A = \bar{A}(n_L + n_H)^{\gamma}$, there is no longer a logarithmic relationship between population of high-skill n_H and productivity A. 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 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 mean wages in each location. Using

[14]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.

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 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.[15] We set the elasticity of substitution across goods $\sigma = 4$, in line with the empirical literature using international trade data.[16] Following a large literature, we use instrumental variables and equilibrium relationships to estimate all other 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 (pop. 218) would be the center of gravity in the continental United States. In fact, 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 data appendix and 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. We consider four possible methods for moving goods – road, rail, water, and air. For each of these methods

[15] 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 Arkoalkis.

[16] See Broda and Weinstein (2006) and Simonovska and Waugh (2014).

separately, we assign a cost of traveling over each pixel of the relevant image file. For example, if we are considering water transport, we assign a low cost to each water pixel, and a high cost to all other pixels. Then, we calculate the lowest possible cost of using each method to move goods between all pairs of locations. The algorithm we use to find this lowest cost path for each transport method is called the fast marching algorithm.

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. We consider four possible modes of 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. For example, if we are considering water transport, we assign a low cost to each water pixel, and a high cost to all other pixels. Then, we calculate the lowest possible cost of using each method to move goods between all pairs of locations. We use the fast marching algorithm to find the lowest cost path for each mode of transportation.

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 exchange rates. The idea is that 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 rate of exchange between any two modes of transportation, but we still need to pin down the cost per mile of using one of the modes. That is, it is not enough for our structural model to know that it costs twice as much to move goods by road as it does to move them by air. We need to pin down the level of iceberg trade costs as well. To do so, we return to a classic gravity specification that is implied by our model. To be consistent with our later structural estimation, we set the elasticity of substitution across goods equal to four rather than nine as in the Allen and Arkolakis (2014). Estimating the gravity equation gives us an estimate for the unknown scaling parameter. 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 3. Road by assumption 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 the average iceberg cost of shipping from San Francisco to Portland is 1.35, while the average iceberg cost of shipping from San Francisco to Chicago is 2.3.

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. Because the resolution of the maps we use may be different than those used in the previous papers, we should not expect the levels of our estimates to be the same, but we might expect proportionality. One reason for the difference is that we set a lower trade elasticity in our structural estimation, $\sigma = 4$ rather than $\sigma = 9$.[17] Our trade costs are likely higher in absolute terms than in Allen and Arkolakis (2014), as our products are more differentiated. We need higher trade costs to explain the low volume of trade between distant locations.[18]

	Road	Rail	Water	Air
Variable cost	1.2427	1.0965	2.1383	0.4270
Fixed cost	0	1.0117	1.2734	1.8307

Table 3: 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,

$$\tilde{w}(i) = \tilde{\kappa} + \frac{1}{\theta}\tilde{n}(i) - \tilde{u}(i) \tag{26}$$

$$\tilde{n}(i) = -\rho\tilde{w}(i) + \rho\varphi \log n_H(i) + \rho\tilde{\beta}(i) \tag{27}$$

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.[19] Estimating these equations using OLS can be problematic due to correlations between error terms and regressors. In equation (26), the skill population ratio, $\tilde{n}$, is expected to be higher in locations where the ratio of amenity values for high-skill relative

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

[18] 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. In light of the equation for the price index (4), scaling trade costs simply affects the estimated level of TFP. We have estimated the model with the unnormalized trade cost matrix. The trade cost normalization does not affect our results, except for the decomposition of wage inequality. In the version with unnormalized trade costs, the share of variance in the wage premium explained by geography is slightly higher.

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

to low-skill, $\tilde{u}$, are greater. This correlation means that OLS presumably underestimates $1/\theta$. In addition, in equation (27), 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 (26) and (27). To estimate θ in the relative supply function (26), we instrument skill population ratio $\tilde{n}$ using a variable that is meant to exclusively capture shifts from the demand side. Motivated by Bartik (1991) and Moretti (2013), we construct this instrumental variable as follows. 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 national skill population share in industry d excluding location i itself. Our instrument is

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

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.

To estimate ρ and φ in the relative demand function (27), 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).[20] This index is only reported for MSA's. 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. As a robustness check, our results virtually do not change if we restrict our sample to only MSA's. 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.

Estimation results are summarized in Table 4. The F-statistics for the first stage in our IV regressions are large enough rejecting that our instruments are weak. Further, 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 $\theta = 12.195$, the elasticity of substitution in production between high and low-skill labor $\rho = 2.409$, and the agglomeration advantage of high-skill labor $\varphi = 0.316/2.409 = 0.131$.[21] In addition, the

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

[21] Our estimate of the elasticity of substitution between high skill and low skill labor ρ is within the range of estimates reported by the literature, if a bit on the high side. 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 $\rho = 2.5$.

residuals in equations (26) and (27) give us the exogenous shifters of relative productivities and amenities $\tilde{\beta}$ and $\tilde{u}$.

	log skill premium, Eq. (26)		log population ratio, Eq. (27)	
	OLS	IV	OLS	IV
log population ratio	0.074***	0.082 ***		
log skill premium			-0.385	-2.409***
log high skill population			0.178***	0.316***
constant	0.551***	0.558 ***	-2.629***	-3.178***
1st stage F (KP)		1150		12
number of observations	1267	1267	1267	1267

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

Table 4: 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 $\bar{u}_H(i)$. Here we use equilibrium integral equations derived by our model. Our estimation procedure, inspired by Allen et al. (2016), 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:

$$\begin{aligned} A(i)^{1-\sigma} &= W_H^{\frac{1-\sigma}{\delta}} N_H^{\frac{\sigma-1}{\delta\theta}} c(i)^{1-\sigma} n_H(i)^{-1} w_H(i)^{-1} b(i) \\ &\quad \times \int_J d(i,j)^{1-\sigma} \bar{u}_H(j)^{\frac{\sigma-1}{\delta}} R(j)^{\frac{(\sigma-1)(\delta-1)}{\delta}} n_H(j)^{\frac{1-\sigma+\delta\theta}{\delta\theta}} w_H(j)^{\frac{\sigma-1+\delta}{\delta}} b(j)^{-1} \, dj \quad (28) \\ \bar{u}_H(i)^{\frac{1-\sigma}{\delta}} &= W_H^{\frac{1-\sigma}{\delta}} N_H^{\frac{\sigma-1}{\delta\theta}} n_H(i)^{\frac{1-\sigma}{\delta\theta}} w_H(i)^{\frac{\sigma-1}{\delta}} R(i)^{\frac{(\sigma-1)(\delta-1)}{\delta}} \\ &\quad \times \int_J d(j,i)^{1-\sigma} A(j)^{\sigma-1} c(j)^{1-\sigma} \, dj \quad (29) \end{aligned}$$

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:

$$\bar{u}_H(i)^{\frac{\sigma-1}{\delta}} R(i)^{\frac{(\sigma-1)(\delta-1)}{\delta}} n_H(i)^{\frac{1-\sigma+\delta\theta}{\delta\theta}} w_H(i)^{\frac{\sigma-1+\delta}{\delta}} b(i)^{-1} = \lambda A(i)^{\sigma-1} c(i)^{1-\sigma}, \quad (30)$$

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

$$\log A(i) = \alpha \log n(i) + \log \bar{A}(i) \tag{31}$$

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 5. We find that the elasticity of hicks-neutral productivity with respect to population is 0.305. The IV and OLS results are very similar. While not reported, removing population weights barely changes these estimates. Our estimate is a bit higher than agglomeration elasticities of 0.03-0.27 reported in the survey by Rosenthal and Strange (2004).

Dependent variable: Log productivity		
	OLS	IV
Log population	0.303***	0.305***
Constant	-2.757***	-2.779 ***
1st stage F (KP)		4119
Obs	1267	1267

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

Table 5: Estimating agglomeration

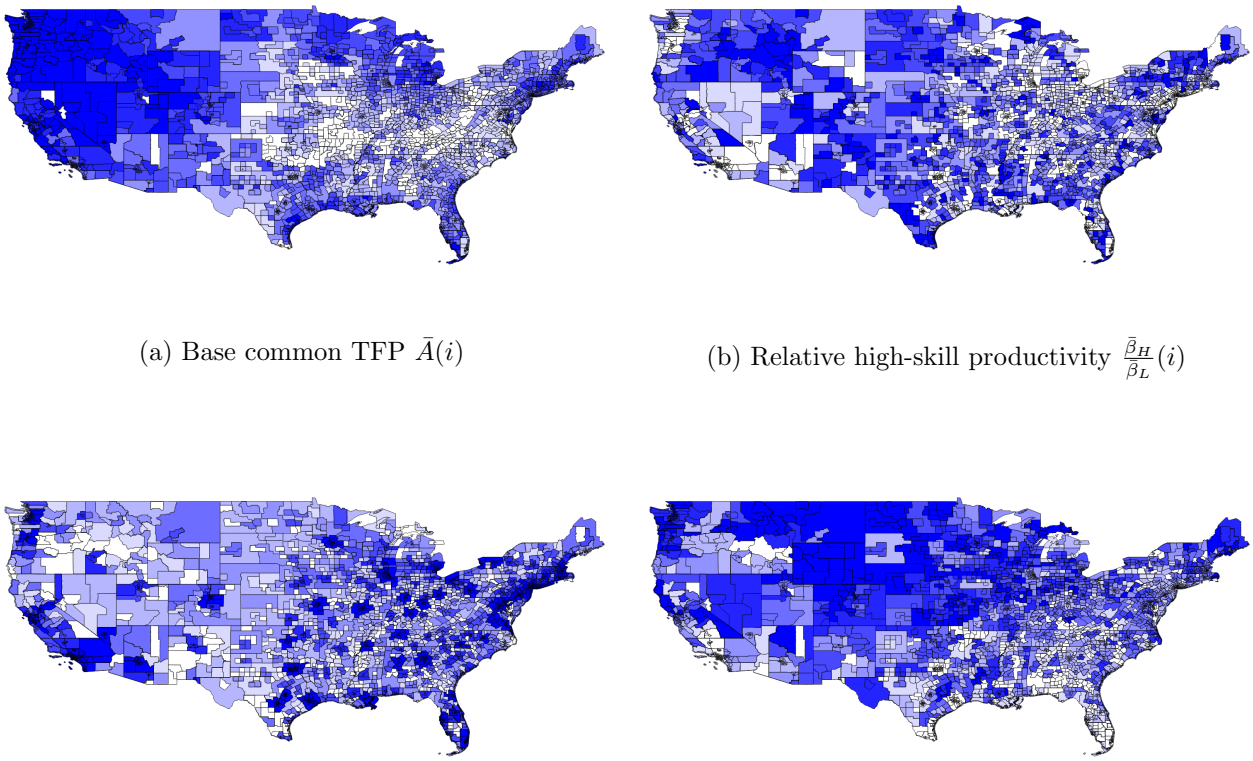
We find that the elasticity of shared TFP with respect to population is 0.305. The IV and OLS results are comparable. While not reported, removing population weights barely changes the estimates. Our estimate is a bit higher than agglomeration elasticities of 0.03-0.27 reported in the survey by Rosenthal and Strange (2004).

4.2.3 Results for the productivity and amenity shifters

In Figure 3, we present the geographical distribution of our four estimated vectors of fundamentals: exogenous productivity, exogenous high-skill amenities, exogenous relative productivity, and exogenous 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.[22] 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. We do not avoid that critique for exogenous high-skill amenities $\bar{u}_H$, however, which are strongly correlated with city size. Our results are consistent with those of

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

Albouy (2012) who shows that in many ways cities are attractive places to live for reasons not related to productivity.[23]



(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 low-skill people differentially prefer to live in the South, Florida, and Southern California, and high-skill people prefer to live in the Upper Midwest, Mountain regions, and Northwest. This pattern seems intuitive, and as one might predict exogenous relative amenities are correlated with skill population share but uncorrelated with overall population. The pattern is less stark for exogenous relative high-skill base productivity. High-skill workers may be exogenously less productive in the Lower Midwest region. Due to the skill agglomeration advantage, observed relative high-skill productivity will be higher in cities.

[23] An alternative strategy would be to allow for congestion effects outside of housing which we model explicitly. If we regress $\bar{u}_H$ on population with a valid instrument (possibly the presence of a land-grant university, for example), we can estimate the elasticity of non-housing related amenities with respect to city size. We have done this, and get an elasticity of 0.7. This elasticity is quite high, and would function in our model like an additional agglomeration force, potentially the force which is highlighted in Diamond (2015). The residuals of this regression would then be our exogenous amenities, and they do not correlate with population. Instead of going in this direction, in our baseline we interpret amenities as medium-run exogenous infrastructure like sports stadiums and theaters.

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 variation in exogenous base productivity and amenities absolute and relative, 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 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	Log wage prem	Log wage prem	ShapelyR^2
Log tradeable price	P	-0.366***	-0.071***	9.2%
Log amenity level	$\bar{u}_H$		0.064***	21.1%
Log base productivity	$\bar{A}$		0.030***	3.2%
Log relative productivity	$\frac{\bar{\beta}_H}{\bar{\beta}_L}$		0.208***	5.6%
Log relative amenities	$\frac{\bar{u}_H}{\bar{u}_L}$		-0.840***	60.9%
Observations		1267	1267	
R-squared		0.240	1.000	

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

Table 6: 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 6, we report the R^2 for a simple regression of the log high-skill wage premium on the log price index. We find that geography alone can explain 24% of the variation in the wage premium.

In the third column of Table 6, we report results from the full decomposition. Our five shifters explain 100% of the variation in observed wage premia. We find that 9.2% of observed variation in skill wage premium are due to variations in geographic features across American cities. The largest part of the variation in wage premia, 60.9%, is explained by variation in relative amenities. We also note that 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 the wages there are low.[24]

5.2 Domestic trade and inequality

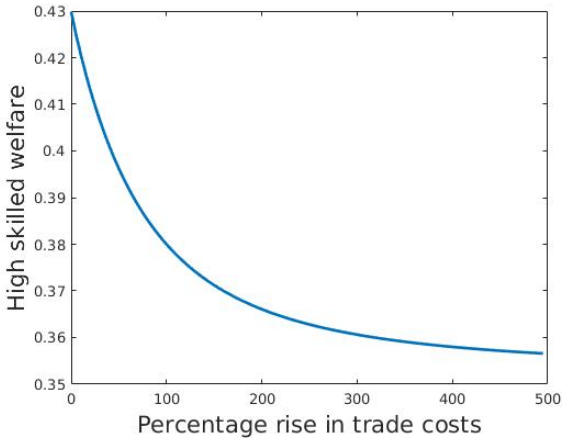
We examine how welfare inequality reacts to changes in domestic trade costs. To do so, we follow a standard exercise in the trade literature by increasing trade costs to the autarchy level. Although this counterfactual experiment is extreme, it allows us to compare our results with those in the literature. 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.

Although there are competing forces at work, we expect that once trade is shut down people will prefer to live in exogenously more productive cities. The cheap goods produced in big cities will no longer be sold in small towns, and conversely, there will no longer be a market in big cities for the goods produced in small towns. This mechanism leads to an overall increase in the concentration of population, and thus an increase in agglomeration forces. Since high-skill workers benefit relatively more from agglomeration, we expect welfare inequality to increase.

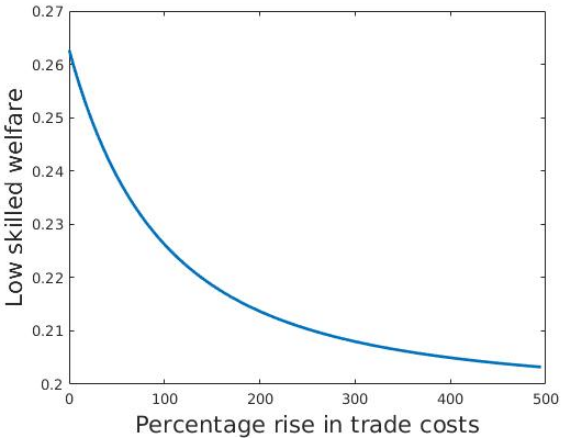
We find that this mechanism is indeed the dominant force. Figure 4 summarizes our results from a large number of counterfactual experiments. In each experiment we increase all trade costs from their baseline values proportionally. Our basic finding is that both high and low-skill welfare fall with increases in trade costs, but low-skill welfare falls more. In the extreme case of moving to autarchy, high and low-skill welfare decreases by 32.7% and 40.1% respectively. Accordingly, the ratio of high to low-skill welfare increases by 12.3%. 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.

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

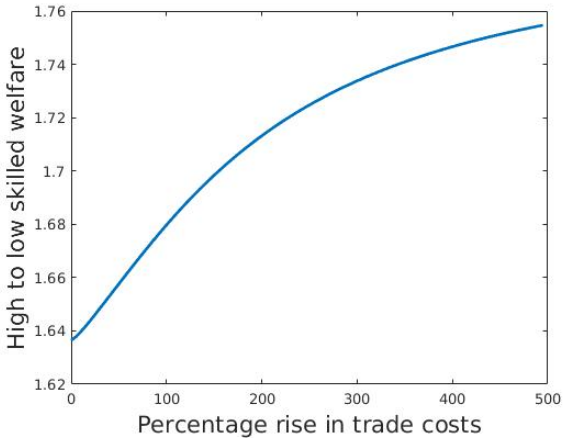
[24]For comparison with Allen and Arkolakis (2014), we also decompose variation in income into exogenous productivities, exogenous (high-skill) amenity levels, and geography. We find that 20.5% of variation in income across cities is due to variation in geography. This number is at the bottom of the range reported in Allen and Arkolakis (2014).



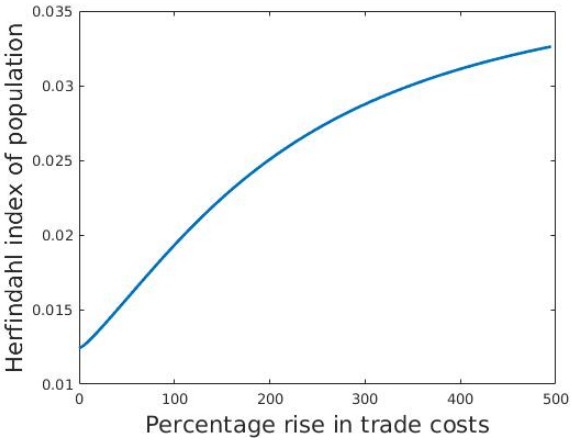
(a) High-skill welfare



(b) Low-skill welfare



(c) Relative high to low-skill welfare



(d) Herfindahl index of population

Figure 4: Trade cost experiments

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 17.6%. In the same period, the college population ratio in California grew by 37.5%.[25] 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.

[25]The overall population share 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		1.5
Average national low-skill welfare		0.3
Average national welfare ratio		1.2
	California	Rest of the United States
High-skill wages	12.7	-0.5
Low-skill wages	7.8	-1.4
Skill premium	3.9	0.9
High-skill population	46.8	-4.1
Low-skill population	6.8	-0.7
Price of tradeables	-5.3	-0.8
Price index	7.1	-1.6
High-skill real wages	4.7	1.1
Low-skill real wages	0.9	0.3

Table 7: The effects of California's productivity shocks on welfare, prices, and wages (percentage change)

In this section, we perform a counterfactual exercise to study how technological progress in California contributed to welfare and inequality in California and across the United States. To perform our counterfactual, we hold constant exogenous total factor productivity $\bar{A}$ and high-skill productivity $\bar{\beta}_H$ in the rest of the United States, and we alter the exogenous productivities of all regions in California to match the observed 17.6% lower 1980 Californian population share compared with our baseline, and the 37.5% lower college population ratio.

We report the results of this counterfactual exercise as percent changes from the counterfactual case to our baseline in Table 7. In order to match overall population growth and growth in college population share, we find that exogenous Californian high-skill productivity rose by 13.6% and exogenous low-skill productivity increased by 2.1% on average across Californian cities from 1980 to 2000. This represents an average 11.7% increase in the relative exogenous productivity of high-skill workers.[26] We find that expected welfare of high-skill workers increased by 1.5%, expected welfare of low-skill workers increased by 0.3%, and welfare inequality rose by 1.2%. Expected welfare growth is for workers *across the United States*, and can alternatively thought of as a social welfare measure.

Furthermore, we examine changes to wages and prices across locations within a skill group, and across skill groups within a location. Table 7 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 of 3.9% across Californian cities, and 0.9% on average else-

[26] We numerically solve for $\bar{A}$ and $\bar{\beta}_H$ of regions in California such that our model generates the desired fall in college population share and overall population share. The reported changes in exogenous high and low-skill productivities correspond to, on average, -8.0% change to $\bar{A}$ and -9.2% change to $\bar{\beta}_H$ of regions in California.

where. The skill premium in California increased less than the relative productivity increase of high-skill workers. We might have expected the opposite, since the effect of exogenous high-skill productivity increases on wage inequality is amplified through population growth and the accompanying high-skill agglomeration advantage. On the other hand, we have simulated a counterfactual equilibrium in our model. General equilibrium effects act to dampen the effect of productivity changes. The effect of the increase in high-skill productivity on high-skill wages is offset by the relative growth in supply of high-skill workers in California. In addition, the overall price index including both housing and tradeables rose by 7.1% in California while it fell by 1.6% 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.

5.4 Growth in the wage premium and American welfare inequality 1980-2000

All over the United States, the college wage premium significantly increased from 1980 to 2000. The rise in the skill premium has not, however, been uniform across American cities. For example, the skill premium rose by 33% in San Francisco, CA, by 17% in Detroit, MI, and only by 6% in Fort Wayne, IN. Figure 5 shows the distribution of changes in the skill premium across cities, with a median of 0.151 and standard deviation of 0.096. In this section we use our model to link these observed changes in the skill premium to what we ultimately care about, changes in American well-being inequality.

We focus on the counterfactual in which observed changes to college wage premium between 1980-2000 are driven entirely by skill-biased technological change. That is, high-skill workers' productivity increases while the productivity of low-skill workers is held constant. More specifically, we match the observed 1980 wage premia in all American cities by reducing exogenous high-skill productivities $\bar{A}(i)^{\frac{\rho-1}{\rho}}\bar{\beta}_H(i)$ holding low-skill productivity $\bar{A}(i)^{\frac{\rho-1}{\rho}}(1-\bar{\beta}_H(i))$ fixed. We report changes from the counterfactual to the baseline equilibrium.

We find that expected welfare of high-skill workers increases by 15.3%, and expected welfare of low-skill workers increases by 0.2%. Accordingly, welfare inequality increases by 15.1%. The change to welfare inequality is comparable to the median change in the college wage premium. While the sharp rise of the skill premium in San Francisco overstates the change in welfare inequality, the modest rise of the skill premium in Fort Wayne understates the change in welfare inequality. The main reason is that high-skill workers increasingly sort to cities that experience higher costs of living due to housing prices. To illustrate this relationship, we plot changes to the college population ratio against changes to housing rents in Figure 6. This sorting mechanism works entirely through general equilibrium effects on wages and prices, which we calculate through fully solving our model at counterfactual levels of exogenous parameters. While we believe our ability to solve for equilibrium at arbitrary parameter values is novel in this literature, our findings are consistent with Moretti (2013) who uses a more empirical

approach to argue that in large cities the rise in real wage inequality between 1980-2000 was less than the rise in observed skill premia.

A final comment is to contrast the implications of our model for welfare trends to those of Diamond (2015) who performs a similar exercise, decomposing changes in wage inequality for the years 1980-2000. In our model, the agglomeration forces for high-skill workers raise the welfare of low-skill workers through general equilibrium price and wage effects. Similar forces are present in Diamond (2015), but in her paper she focuses on how welfare inequality is amplified by the agglomeration force in amenities.[27] Except for different agglomeration mechanisms, the main elements of our models are fairly analogous. For example, both models generate congestion through limited land supply, and both have wages influenced by the supply of high and low-skill labor in a location (in our model a CES production function, in Diamond a log-linear approximation to a wage equation). The key difference is that Diamond (2015) models agglomeration through amenities, with amenities scaling up with the high-skill population ratio. We model agglomeration in productivity.[28] In our model productivity gains in one location spill over to other locations through price and wage effects. Prime among these is the lower cost of tradeables everywhere. Thus, even when productivity growth is exclusively skill-biased and firms may substitute away from low-skill workers, the real wage of low-skill workers tends to rise every where. In Diamond (2015), amenity spillovers in a location are only enjoyed locally.

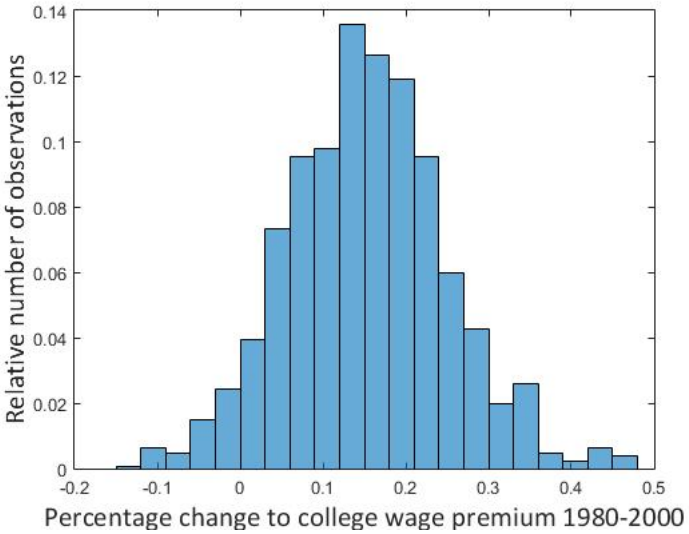


Figure 5: Histogram of observed changes to college wage premium across American cities 1980-2000

[27]Diamond carefully writes that she is only measuring the changes in welfare due to the rise in the college population ratio.

[28]A second important difference is that Diamond estimates high-skill labor to be less sensitive to housing prices, so that low-skill labor is forced out of expensive, high-amenity cities as housing prices rise. We abstract from this preference-driven mechanism by assuming that expenditure shares on housing are the same for both types.

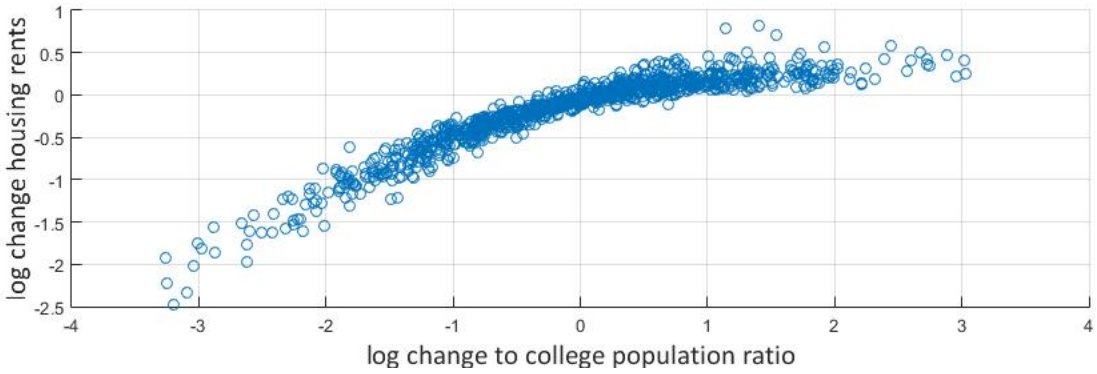


Figure 6: Log counterfactual change in college population ratio versus Log counterfactual change in housing rents, across American cities

6 Conclusion

We document that isolated cities tend to have less wage inequality. We develop a theory showing that 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 9.2% of observed wage inequality is due to the geographic location of cities. In a counterfactual experiment, we find that the rise of Silicon Valley increased skill wage premium in California by 3.9% and welfare inequality across the United States by 1.2%. We also find that the rise in the college wage ratio from 1980 to 2000 in large American cities overstated changes to welfare inequality in that period.

References

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*.

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.

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.

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

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., and Roux, S. (2012). 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. (2012). A spatial knowledge economy. Technical report, National Bureau of Economic Research.

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

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*.

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

Fan, J. (2015). Internal geography, labor mobility, and the distributional impacts of trade. Technical report, Unpublished working paper.

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.

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.

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.

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

Krugman, P. (1991). 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.

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.

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.

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.

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 Cleaning and adding geography to census data

As mentioned in the body of the paper, the main data source is the IPUMS 5% sample. The data was downloaded with the interface available on the IPUMS website.[29] Table 8 described the variables we downloaded in the initial sample.

VARIABLE NAME	Description
YEAR	Census year DATANUM & Data set number
SERIAL	Household serial number
HHWT	Household weight
STATEFIP	State (FIPS code)
COUNTY	County
METRO	Metropolitan status
METAREA	(general) Metropolitan area [general version]
METAREAD	(detailed) Metropolitan area [detailed version]
PUMA	Public Use Microdata Area
GQ	Group quarters status
PERNUM	Person number in sample unit
PERWT	Person weight
SEX	Sex
AGE	Age
RACE	(general) Race [general version]
RACED	(detailed) Race [detailed version]
EDUC	(general) Educational attainment [general version]
EDUCD	(detailed) Educational attainment [detailed version]
WKSWORK1	Weeks worked last year
UHRSWORK	Usual hours worked per week
INCWAGE	Wage and salary income
INCBUS00	Business and farm income, 2000.

Table 8: Variables from IPUMS 5% sample

Using these variables, we cleaned the data by modifying replication code for Baum-Snow and Pavan (2013) from Nathanial Baum-Snow's website.[30] Cleaning involves dropping observations with imputed characteristics, dropping ages less than 25 and greater than 64, dropping those who worked less than 40 weeks in the year or less than 35 hours in a week, those which made less than minimum wage in 1999, and active duty military. Finally anyone with positive business income was dropped. This last restriction is because we think wages are a poor measure of income for owners of businesses. The cleaning was done in Stata with the file "census_prep.do", and cleaned data is saved as "census00.dta".

[29] https:usa.ipums.orgusa-actionvariablesgroup

[30] http://www.econ.brown.edu/fac/nathaniel_baum-snow/ineq-citysize.zip

After cleaning the data, we merge in the . This operation is done in Stata using the file "msa_puma_geography.do". We calculate the population weighted locations of PUMA's and MSA's using the excellent Missouri Census Data Center website.[31]. Selecting all states and both "source" and "target" equal to "PUMA for 5 Pct Samples (2000)" generates a csv file, which we copy into an excel sheet to get "PUMAs.xlsx". Selecting all states and both "source" and "target" equal to "Metro Area:MSA or CMSA (2000)" generates a csv file. This csv gives MSAs four digit codes to match CMSA codes, while in the census data we have only three digit codes. For the most deleting the last digit of the four digit codes is all that is necessary, but in five instances there are two four digit codes with the first three digits identical. We select the region which matches the census MSA three digit code. These five changes are:

1. 233: 2330/2335 exclude both
2. 265: 2650/2655 exclude 2655 (Florence,SC)
3. 298: 2980/2985 exclude 2985 (Grand Forks, ND-MN)
4. 328: 3280/3285 exclude 3285 (Hattiesburg, MS)
5. 360: 3600/3605 exclude 3600 (Jacksonville FL)

After these changes, results are stored in the file "MSAs-change.xlsx"

Next we add the areas of PUMAs to complete the geographical features we need. We download an IPUMS file containing all intersections between 2000 PUMAs and 2010 PUMAs.[32] We then collapse this file by 2000 PUMA to get areas in square kilometers.

Finally, the file "census_prep.do" merges all the geographical features into the census data, and creates a new variable "fj_region" which is an MSA if the census observation is classified in an MSA, and PUMA otherwise. It is important to point out that just because an observation is not classified in an MSA, that it is not in fact part of an MSA. Moreover, the population observed in an MSA may not be representative of the true MSA population. Here is what is said about the issue on the IPUMS website:[33]

> The most detailed geographic information available is for 1980 county groups or for 1990 or 2000 PUMAs, areas which occasionally straddle official metro area boundaries. If any portion of a straddling area's population resided outside a single metro area, the METAREA variable uses a conservative assignment strategy and identifies no metro area for all residents of the straddling area.
>
> Users should not assume that the identified portion of a partly identified metro area is a representative sample of the entire metro area. In fact, because the unidentified population is located in areas that straddle the metro area boundaries, the identified population will often skew toward core populations and omit outlying communities. Also, weighted population counts for incompletely identified metro areas

[31] http://mcdc2.missouri.edu/websas/geocorr2k.html
[32] https://usa.ipums.org/usa/resources/volii/puma00_puma10_spatial_crosswalk.xlsx
[33] https://usa.ipums.org/usa-action/variables/METAREA#description_section

will be low by amounts ranging from 1 to 69% (since the unidentified individuals will not be counted as living in the metro area).

A.2 Constructing CFS area distances

In order to calculate distances between Commodity Flow Survey (CFS) areas, we need two types of information. One is information about the size and nature of commodity flows themselves, and the second is the physical locations of roads, waterways, and railways in the United States. Information on commodity flows was downloaded from the US Census website.[34] We used flows from the year 2007, because this was the first year in which tables breaking down commodity flows by mode of transportation was available. These raw data come in so-called "long" format, with each row a origin-destination-mode observation. We find it more convenient to work with data in the "wide" format, with each row an origin and destination, but with separate columns for the value of each mode of transportation. We do this using the python script "pivot_cfs_mode.py".

There are two input files necessary to run "pivot_cfs_mode.py". The first, "Origin_by_Destination_by_Mode_pyt is simply the downloaded census file saved as a csv. We also need the centroid of each CFS area in order to calculate physical distance between CFS areas. We use QGIS software to do this. Our data comes from a shapefile available from a US census website.[35] We manipulate the data in the QGIS project "calc_centroids.qgs". Specifically, after loading the downloaded shapefile, we create a new variable "st_cfs_area" to make CFS areas unique. We then use the "dissolve" command to eliminate counties, and the "mean coordinates" command to get centroids. We save the calculated centroids as "cfs_2007_centroids.csv".

The output of "pivot_cfs_mode.py" is "replication_data'_no_ethnic.csv". As the name implies, in order to run the gravity equation in our distance cost estimation we need to add information on the correlation in ethnic composition between all CFS areas. We separately downloaded the variables in Table 9 from IPUMS. We save these new variables as the stata file "demographic_data_2000.dta".

The Stata script "merge_in_demographics.do" combines the new and the old census data, and then creates the correlation matrices "lang_corr_matrix.csv", "race_corr_matrix.csv", "birth_pl_corr_matrix.csv". We want the correlation matrices listed by origin destination pair to match our CFS data, so we reshape and combine the data in this format in the python script "append_ethnic_var.py", which creates a file "combined_stacked.dta". Finally, we add the ethnic variables to our CFS mode information using the do file "append_stacked.do", which outputs the file "replication_data.csv".

The file "replication_data.csv" now contains everything we need to run the distance cost estimation except the map files. The last step, however, is to put the data in a format which Matlab can understand.[36] This means that we strip off all text and put the CFS area coordinate

[34] http://www2.census.gov/econ2007/CF/sector00/special_tabs/Origin_by_Destination_by_Mode.zip

[35] http://www.census.gov/econ/census/shapefiles/CFS_AREA_shapefile_010215.zip

[36] Allen and Arkolakis wrote their estimation in Matlab and kindly made the files available to us. We use a modified version of their code to calculate our distance costs.

VARIABLE NAME	Description
YEAR	Census year
DATANUM	Data set number
SERIAL	Household serial number
HHWT	Household weight
GQ	Group quarters status
PERNUM	Person number in sample unit
PERWT	Person weight
RACE	(general) Race [general version]
RACED	(detailed) Race [detailed version]
BPL	(general) Birthplace [general version]
BPLD	(detailed) Birthplace [detailed version]
LANGUAGE	(general) Language spoken [general version]
LANGUAGED	(detailed) Language spoken [detailed version]
RACESING	(general) Race: Single race identification [general version]
RACESINGD	(detailed) Race: Single race identification [detailed version]

Table 9: Demographic variables from IPUMS 5% sample

columns into a file called "cfs_coor.csv", the trade value columns into "cfs_trade.csv", and the demographic correlations into "cfs_eth.csv".

The Matlab script "allen_arkolakis_estimation.m" takes the csv files described in the last paragraph as inputs, and outputs the file "cfs_areas_trade_cost_list.csv". We next copy this list as a column into the file "replication_data.csv". This is necessary because we need origin and destination names attached to the trade costs to merge with the census data.

At this point we have recovered distance costs between all CFS areas. For the structural estimation, however, we need to know the distance costs between all fj_regions. Typically, fj_regions are completely included in a single CFS area. A small number of fj_regions stradle two or more CFS areas. In these cases, we prefer first non "rest of state" CFS areas, and then the CFS area which contains the highest population.[37] This calculation is done in the Stata script "link.do". The script goes on to recreate the distance cost matrix, but with fj_regions rather than CFS areas. If the distance cost is missing (i.e. there is no trade between the relevant CFS areas), then the highest observed trade cost is substituted. The final output is a distance cost matrix with both rows and columns fj_regions. This is one input into our structural estimation.

B Regression tables for motivating facts

The following three tables document the correlations presented in scatterplots in Section 2.3. Table 10 contains regressions documenting facts from the literature. The skill wage premium and skill population ratio are both positively correlated with population. Remoteness is negatively correlated with population, not surprisingly since population is used as a weight. Finally skill population ratio is negatively correlated with remoteness, but only when using analytic

[37]In most states there are separate CFS areas for large cities and then a single larger CFS area encompassing the rest of the state

population weights.

Table 11 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. Some results instrument the skill wage premium with the presence of a land grant university. The exogeneity assumption is that the selection of a location for a land grant university only affects the skill population share through its effect on the skill wage premium.

Table 12 documents the negative relationship between the skill wage premium and remoteness. The relationship is negative and significant in all specifications, save for including state fixed effects and removing analytic population weights.

Table 13 documents the relationship between wages and remoteness across skill groups at the level of individual workers. Controlling for a wide range of individuals' characteristics as well as population of their location, at a statistically significant level, remoteness lowers wages, and lowers wages of college-educated workers more than non-college educated workers.

VARIABLES	Skill wage prem	Skill pop ratio	Log pop	Skill pop ratio	Skill pop ratio
Log pop	0.0215***	0.180***			
	(0.00151)	(0.00846)			
Log rem			-16.08***	-0.409	-2.336***
			(2.128)	(0.273)	(0.462)
Constant	0.272***	-3.100***	13.72***	-1.424***	-0.731***
	(0.0199)	(0.119)	(0.519)	(0.0892)	(0.144)
Observations	1,267	1,267	1,267	1,267	1,267
R-squared	0.611	0.698	0.509	0.292	0.387
State FE	YES	YES	YES	YES	YES
Pop Weight	YES	YES	YES	NO	YES

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

Table 10: Regressions documenting facts from the literature

VARIABLES	pop ratio	pop ratio	pop ratio	pop ratio	pop ratio	pop ratio
Skill wage prem	0.284**	2.334***	14.00**	0.848***	2.744***	29.78
	(0.142)	(0.362)	(6.735)	(0.137)	(0.287)	(19.23)
Constant	-1.445***	-1.982***	-7.730**	-1.912***	-2.526***	-16.26*
	(0.0608)	(0.161)	(3.305)	(0.101)	(0.171)	(9.777)
Observations	1,267	1,267	1,267	1,267	1,267	1,267
R-squared	0.004	0.175		0.314	0.473	
State FE	NO	NO	NO	YES	YES	YES
Pop Weight	NO	YES	YES	NO	YES	YES
Land Grant IV	NO	NO	YES	NO	NO	YES

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

Table 11: Regressions documenting the positive relationship between skill wage premium and skill population ratio

VARIABLES	Skill wage prem	Skill wage prem	Skill wage prem	Skill wage prem
Log rem	-0.200***	-0.356***	-0.0330	-0.435***
	(0.0259)	(0.0471)	(0.0633)	(0.112)
Constant	0.467***	0.536***	0.501***	0.583***
	(0.00526)	(0.00964)	(0.0166)	(0.0249)
Observations	1,267	1,267	1,267	1,267
R-squared	0.053	0.231	0.314	0.501
State FE	NO	NO	YES	YES
Pop Weight	NO	YES	NO	YES

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

Table 12: Regressions documenting the negative relationship between remoteness and the skill wage premium

	(1) log wage	(2) log wage	(3) log wage	(4) log wage
log remoteness	-0.341***	-0.390***	-0.308***	-0.568***
	(0.002)	(0.002)	(0.003)	(0.005)
log population	0.081***	0.092***	0.091***	0.070***
	(0.001)	(0.000)	(0.001)	(0.001)
College dummy	0.471***	0.519***	0.490***	0.377***
	(0.001)	(0.001)	(0.015)	(0.014)
College X log remoteness			-0.250***	-0.226***
			(0.005)	(0.005)
College X log population			0.006***	0.015***
			(0.001)	(0.001)
Gender dummies	NO	YES	YES	YES
Experience cubic	NO	YES	YES	YES
Race dummies	NO	YES	YES	YES
State FE	NO	NO	NO	YES
R-squared	0.164	0.263	0.263	0.280
Observations	4,569,335	4,569,335	4,569,335	4,569,335

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors in parentheses.

Table 13: Regressions documenting the relationship between remoteness and wages across skill groups at the level of individual workers

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 14 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:

1. 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.

2. 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. Column (3) in Table 14 shows that this does not drive the difference between our estimates. We doubled the value of truck transport in our data, and our estimates remained qualitatively the same. We speculate the estimates do not change much because road transport is already the dominant form of shipment in the domestic United States. Increasing the dominance does not have a qualitative effect on the estimates.

3. 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.[38] 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. Our estimates of air and rail transport then move substantially closer to the numbers estimated in Allen Arkolakis, although they are still somewhat different.

4. The maps we use to compute distances for road and rail are 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. Because of different coordinate systems hardwired into the code, it is hard to directly analyze how much this factors in the analysis.[39]

5. Allen and Arkolakis estimate the parameters for ther shippers' discrete choice of mode of transport minimizing the following loss function. Let $\varepsilon(\beta)^m_{od}$ be the difference between the predicted and observed fraction of shipments of mode m 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 minimization algorithm.[40] In Column (5) we see results that are much

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

[39] 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, the extent that this affects estimates is not clear.

[40] 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

Transport Type	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Road var	0.5636	0.4065	0.4414	0.3094	0.4235	0.3944	0.5445	0.4430
Rail var	0.1434	0.3661	0.3935	0.2965	0.0016	0.3414	0.4405	0.3986
Water var	0.0779	0.6265	0.6439	0.4556	0.0454	0.4597	0.7915	0.6806
Air var	0.0026	0.2233	0.3187	0.1050	0.0506	0.0000	0.0000	0.3157
Rail fixed	0.4219	0.2995	0.3274	0.3308	0.3687	0.4178	0.5246	0.3106
Water fixed	0.5407	0.3428	0.3600	0.3907	0.3887	0.5938	0.6037	0.3384
Air fixed	0.5734	0.5254	0.4963	0.4935	0.3630	0.6313	0.8323	0.4904

Table 14: Comparing distance estimates in several models

more similar to Allen and Arkolakis than in the baseline. Using our algorithm in Column (6), we move the results much closer to our baseline. Finally, in Column (7) we run Allen and Arkolakis' code with their data and with our minimization algorithm.[41] Here we see substantial convergence toward our baseline estimates. One caveat is that air transport variable costs become even smaller than those estimated in Allen and Arkolakis.

We conclude that one of the main drivers of the difference in our estimates may be the loss function in the estimation algorithm for the mode of transport problem. Map differences may also be playing an important role, driving differences in our air cost estimates, along with the choice of input data for water and rail described above.

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 function for estimating the relative shipment costs by mode. In sum, although we have not been able to understand exactly what causes the difference between our estimates, we believe we have a few leads which could be investigated further. Since accounting for the differences is not the primary goal of our study, we leave the discussion here.

1. AA Baseline
2. FJ baseline[42]
3. FJ Double Truck
4. FJ Double Truck / Only water,only rail
5. AA code FJ data
6. AA code FJ data FJ mode obj

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

[42] There was a small bug which we fixed *after* running the AA comparisons below. In particular we were missing several small locations in our estimation code. Adding these locations did not change our estimates much, which can be seen by comparing column (8) and column (2).

7. AA code/data FJ mode obj
8. FJ baseline before bug fix (omitting several small locations)

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, productivity ratios $\bar{\beta}_H/\bar{\beta}_L$, and amenity ratios $\bar{u}_H/\bar{u}_L$, the following algorithm solves for productivities inclusive of spillovers A's and amenities of high-skill workers $\bar{u}_H$'s.

Given that trade costs are symmetric, we reduce the two systems of equations 29 and 29 using relation 30,

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

The algorithm solves for amenities and productivities up to a scale.

1. Start with an initial guess for $A(i)$.
2. Compute the kernel,

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

3. Define $\kappa \equiv \lambda W_H^{\frac{1-\sigma}{\delta}} N_H^{\frac{\sigma-1}{\delta\theta}}$, and $f(i) = A(i)^{1-\sigma}$. Define $A_0 \equiv \int_J f(i)di$. Let

$$\tilde{f}(i) \equiv \frac{f(i)}{\int_J f(i)di} = \frac{f(i)}{A_0},$$

as a normalization that sets the integral over $\tilde{f}$ to one. Then, the system of integral equations described above can be written as follows:

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

In iteration t, update $f^{(t)}(i)$ according to

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

Note that as 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 step 5.

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

all i,

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

By construction $\int_J \tilde{f}(i)di = 1$, therefore $A(i)^{1-\sigma} = \tilde{f}(i)A_0$. For $A_0 = 1$, calculate A:

$$A(i) = \tilde{f}(i)^{\frac{1}{1-\sigma}} A_0^{\frac{1}{1-\sigma}}$$

Now, go to step 2.

5. Using equation (30), calculate $\bar{u}_H(i)$.

$$\bar{u}_H(i) = \lambda^{\frac{\delta}{\sigma-1}} R(i)^{1-\delta} n_H(i)^{\frac{\sigma-1-\delta\theta}{(\sigma-1)\theta}} w_H(i)^{\frac{1-\sigma-\delta}{\sigma-1}} b(i)^{\frac{\delta}{\sigma-1}} A(i)^{\delta} c(i)^{-\delta}$$

We normalize amenities and productivities such that for the first city in our list that we call i_0, $\bar{u}_H(i_0) = 1$ and $A(i_0) = 1$.

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.

1. Guess $n_H(i)$ for all i.
2. Compute W_H/W_L according to (17). Then plug it in (16) to find $n_L(i)$.
3. Calculate skill premia, $\omega(i) \equiv w_H(i)/w_L(i)$, 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 $\tilde{c}(i)$ according to (18) and $\tilde{R}(i)$ according to (19).
6. Let

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

where, according to (24), $\tilde{w}_H(i)$ is given by:

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

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

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

Then, system of integral equations (21) can be written as follows (note that the scale parameter cancels out):

$$f(i) = \kappa \int_J K(j,i) f(j)\ dj$$

In iteration t, update $f^{(t)}(i)$ according to

$$f^{(t+1)}(i) = \frac{\int_J K(j,i) f^{(t)}(j)\ dj}{\int_J \int_J K(j,i) f^{(t)}(j)\ djdi} \tag{33}$$

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. Otherwise, go to the next step.

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,

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

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.