Trade and Inequality in the Spatial Economy

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

[1]Penn State

[2]Copenhagen Business School

November 8, 2015

PRELIMINARY

Inequality has long fascinated economists, and growing income inequality has been recently and heatedly discussed in public forums.[1] A remarkable fact emerging from this discussion is the strong positive relationship between wage inequality and city size (Baum-Snow and Pavan, 2012). In this paper, we add to the study of inequality and distribution of economic activity in two ways. First, we document new facts on the interaction of geography with inequality. Second, we develop a quantifiable model which explains our findings and their implications. In particular, we study the effects of improving trade infrastructure on wage and welfare inequality.

Our first contribution is to document facts about inequality and geography. We assign to each American city a measure of remoteness meant to capture its distance from all other cities. We then show that this measure correlates negatively with the skill premium, the ratio of the mean wage of college degree to the mean wage of non-college degree workers. That is, wage inequality is lower in remote cities. To our knowledge, our paper is the first to document this fact.

*We thank Sina Smid for research assistance, and participants in seminars at Cardiff University, Copenhagen Business School, Copenhagen University, the European Trade Study Group, and Penn State for helpful suggestions.

[1]Inequality is one of the key issues in the current US presidential race.

The literature has developed little in the way of a unified framework that incorporates both geography and inequality in the spatial economy. On one hand, economic geography examines the role of trade costs in determining spatial patterns of economic activity. This literature does not address inequality but welfare at the aggregate. A classic example is (Krugman, 1991), and a more recent one is Allen and Arkolakis (2014). On the other hand, recent models of spatial inequality often treat cities as isolated. Workers in a city interact with each other but cities in a nation either do not trade or trade costlessly (Davis and Dingel, 2014). Since the geographic location of a city relative to other cities remains irrelevant, this literature cannot address the interaction of geography with inequality. By including costly trade between cities in a model of mobile heterogeneous labor, we bridge the gap.

In our model, we have a continuum of locations, workers, and firms. Workers come in two types, skilled and unskilled, and each worker has an ideosyncratic utility from living in each location. A worker decides where to live taking wages as given. A firm also takes local wages as given, and produces a tradable good using skilled and unskilled labor as inputs. The sole difference between the two types of labor is that skilled labor benefits more from agglomeration.[2]

We require a model which generates skill premia differing across locations. In particular, to match our empirical findings we want less remote cities to have higher skill premia. We develop a model with two critical features which delivers the required relationship. The two critical features are stronger agglomeration forces for skilled workers, and heterogenous location preferences.

The intuition behind the model can be described in a few sentences. Consider a city near other cities, a centrally-located city. Its access to cheap tradable goods and nearby markets make it attractive to live in. This leads the city, all else equal, to have a relatively high population compared with a remote city. Due to agglomeration forces, skilled workers become relatively more productive in the centrally-located city. Thus, all else equal, firms there demand a relatively high share of skilled workers as inputs. In order for the relative share of skilled workers on the supply side to equal the firm's demand for relative share of skilled workers as inputs, there must be a higher skill premium in the centrally-located city.

Our results show that improving trade infrastructure benefits both types of labor, but low skill labor may benefit more. Better infrastructure tends to spread populations out, so that skilled workers lose some of their agglomeration advantage over unskilled workers. Typically

[2] We allow there to be a Hicks-neutral agglomeration force as well.

models of inequality and trade predict that skilled workers gain more than unskilled workers (Antràs et al., 2006), a result that may be reversed in our model.

The literature on the skill premium has found a number of patterns. To the extent that we are able to measure the relevant variables, all of these facts are consistent with our data. The literature has shown convincingly that the skill premium is higher in cities with larger populations (Davis and Dingel, 2012). The relationship between the skill premium and city size has become stronger over time (Baum-Snow and Pavan, 2013; Lindley and Machin, 2014). In addition, larger cities have been shown to have a higher share of college-educated worker, another pattern which has strengthened over time (Moretti, 2008; Lindley and Machin, 2014). Areas of denser economic activity have higher skill premiums (Combes et al., 2012).

These stylized facts have inspired a number of theories (Davis and Dingel, 2012, 2014; Baum-Snow and Pavan, 2012; Combes et al., 2012). These theories abstract from costly trade, treating trade between cities as either non-existant or frictionless.[3] The style of our modeling exercise below has more in common with recent forays of trade economists into economic geography (Allen and Arkolakis, 2014; Desmet et al., 2014; Fajgelbaum et al., 2015). These theories model costly trade, and focus on the spatial location of economic activity. They have, however, only one type of labor, so they cannot analyze the interaction between geography and inequality. In an international trade context, Fujita and Thisse (2006) study inequality with costly trade, but their model has immobile unskilled workers.

A recent working paper, Fan (2015) analyzes the impact of an international trade liberalization on inequality using a spatial equilibrium framework. As Fan's focus is on aggregate welfare effects rather than on the forces determining the distribution of the skill premium, he relies on skilled and unskilled workers differentially valuing amenities to generate the observed distribution of the skill premium.[4] Since we are interested in the geographic distribution of inequality, we have a simpler model in which skilled workers only differ from the unskilled in terms of strength of agglomeration effects in the production function.

[3] Davis and Dingel (2012) does contain an extension where trade costs are treated in the limit as they go to zero.

[4] Skilled and unskilled workers in Fan (2015) differ in terms of valuation of amenities, draws from productivity distributions, costs of migration, and initial allocations across space. The initial allocations are important as migration costs depend on the location a worker starts from.

1 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.

1.1 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 individuals older than 24 and younger than 65 with reported income, giving us observations on over four million workers distributed across the United States.[5]

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 a metropolitan statistical area (MSA) when possible, and otherwise a census-related area known as a public use microdata area (PUMA). States choose PUMAs to contain between 100,000 and 300,000 residents. PUMAs are diverse in geographical size as the population density of American regions vary widely.

Many authors in the urban economics literature have used the same IPUM's 5% sample, and in the course of cleaning and understanding the 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, CBSA, and county are only reported when there is no ambiguity about her location. If an individual is in a PUMA which straddles the border of two counties, then she will be reported without a county. In practice, only 66% of our cleaned observations have a reported county, and we observe individuals in only 423 of the 3007 American counties. Particularly for sparsely populated and smaller geographic areas, the IPUM's identifiers are likely unrepresentative. For MSA's, it is likely that our analysis will pick up the core of the urban area, but miss those on the perimeter.

In addition to the IPUMs data, we need information on the geographical position of each location as well as information on trade flows between locations. We use geographical position data from the Missouri Census Data Center. For trade flows we use publicly-available data

[5]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.

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.

1.2 Important concepts

Our main measure of geography is remoteness, a concept we borrow from the international trade literature.[6] We label each location with a number $i = 1 \dots N$. The distance between location i and location j is $d_{i,j}$. The distance we use here is structurally estimated later in this paper, and captures the trade cost between every pair of locations given the current state of transportation infrastructure. The remoteness of location i R_i is the harmonic mean of the distances between location i and the other locations:

$$R_i = \frac{1}{\sum_{j \neq i} \frac{1}{d_{i,j}}}$$

In words, a location which is close to other locations will have low remoteness. An alternative formulation a bit closer to the trade literature would be to weight the distances by population (Head, 2003). A weighted measure is more appropriate in the trade context where population is immobile to a first approximation. In our context, however, population in endogenous and depends upon trade costs, so we stick to the unweighted version.[7]

We have two measures of inequality, the college wage premium (or skill premium) and the college population share (or skill share). Both are measured in a way standard in the labor literature, and are calculated in each location. A worker is highly educated if he has at four-year college degree. The college premium is the mean wage of highly-educated workers in an location divided by the mean wage of other workers. The college population share is the population of highly-educated workers in an location divided by the population of less-educated workers. We use census population weights when calculating all means.

[6]It is a famous empirical regularity that the trade of two countries is proportional to the product of their national products divided by the physical distance between them. This relationship is known as the naive gravity equation. The adjective naive makes an appearance because such a gravity equation implies that trade between two countries is unaffected by what takes place in a third country. For example, the trade between the United States and Mexico is unaffected by the rapid increase in trade between China and the United States. Remoteness in an international trade context is the national-product-weighted sum of the distance between a country and all other countries. Multiplying the gravity equation by a remoteness term allows third countries to influence bilateral trade relationships.

[7]We have done the analysis with weighted versions of this measure, and the results are similar.

Both in this motivating empirical section and the structural model, we will be focusing on population density rather than population as the driver of agglomeration and congestion. This is necessary because of the census requirement that the PUMA, which is the smallest area we can observe, contains between 100,000 and 300,000 people. Thus the population of every PUMA is nearly. The population density, on the other hand, varies by several orders of magnitude in a meaningful way. For example one PUMA in Montana has 0.17 people per square km. in a PUMA in Montana, while another in Connecticut has 311 people per square km, similar to the New York City MSA.

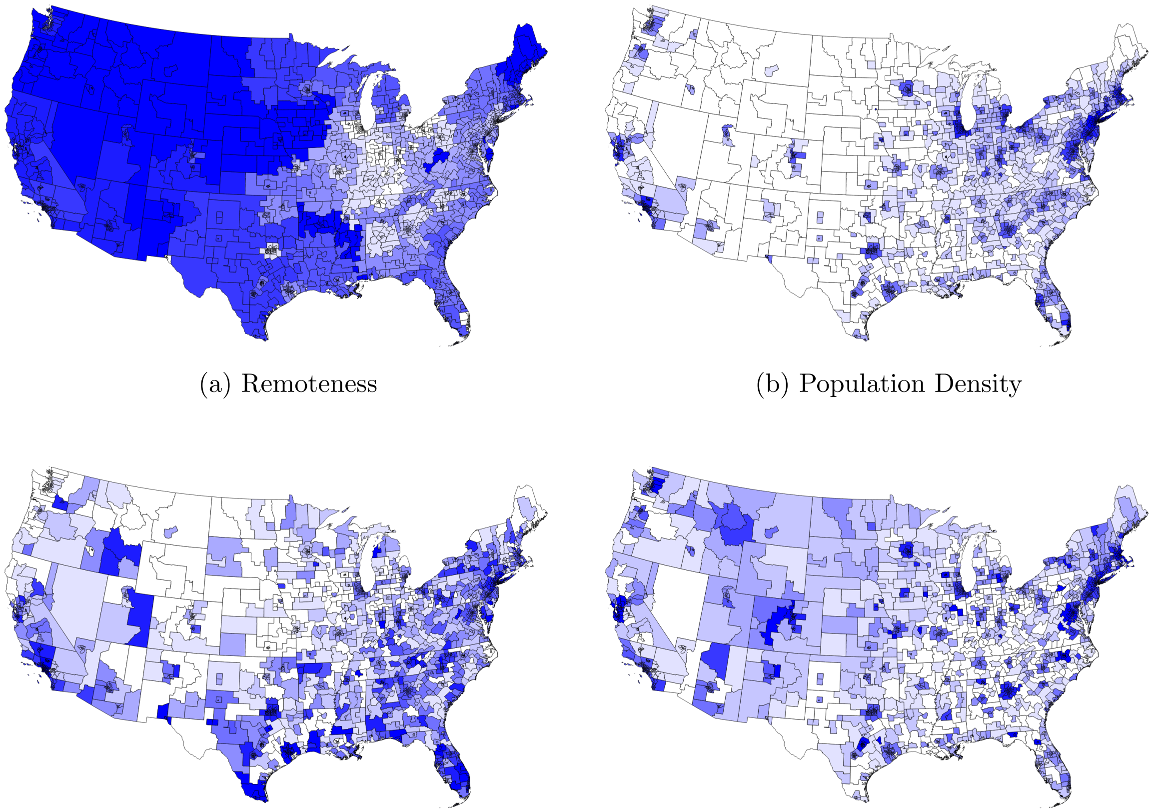
Table 1 contains some descriptive statistics, and the series of figures Figure 1 shows how our measures vary across the United States. The borders on this map are PUMAs. PUMAs within an MSA are just all assigned the data value for the MSA. The more intense the blue, the higher the measure. Remoteness is highest in the West, and particular the North West of the United States. Population density is highest in metropolitan areas. The college wage premium seems to be higher in the parts of the country which are relatively less remote, a finding we will pick up in our regressions in the next section. The college population share, however, is relatively high in the Rocky Mountain states such as Montana and Idaho, and not clearly correlated with remoteness. This will also be our finding in the next section.

Statistic	Mean	Standard Dev	Min	Max
Remoteness	2.5×10^{-3}	3.3×10^{-4}	2.0×10^{-3}	3.2×10^{-3}
Population Density	25.7	48.2	0.17	408
Skill premium	1.58	0.14	1.20	2.26
College share	0.32	0.16	0.11	1.34
Census observations	4.3×10^{6}			
Location observations	775			

Table 1: Data summary statistics

1.3 Motivating empirical results

In this section we document the covariance of our measure of geography, remoteness, with our measures of inequality, the college wage premium and college share. Regressions of the skill premium on remoteness and other variables are found in Table 2. In column one, we see that remoteness has a strong negative relationship with population density. Doubling a locations remoteness is associated with a fourfold decrease in its population density.



(a) Remoteness

(b) Population Density

(c) Skill Premium

(d) College Share

Figure 1: Counties by data feature

Remoteness also covaries negatively with the skill premium, with an elasticity of around 12%. To put this in context, we find that the population density of a location varies positively with the skill premium with an elasticity of around 2%. The positive relationship between city size and the skill premium has been emphasized in other studies. We find that both remoteness and population density are significantly and positively associated with college population share. The correlation of remoteness with college population share is not, however, robust to different regression specifications. For example, in Table 3 we use simple geodesic distance rather than our structurally estimated distance. The correlation between population share and remoteness remains positive, but loses statistical significance.

We suspect that a few land-grant university outliers are behind the unstable positive correlation between remoteness and college population share. The most remote locations in

VARIABLES	(1) L pop dens	(2) L wage prem	(3) L wage prem	(4) L coll shr	(5) L coll shr
L remote	-3.681***	-0.124***		0.569***	
	(0.450)	(0.0255)		(0.116)	
L pop dens			0.0211***		0.151***
			(0.00203)		(0.0105)
Constant	-19.87***	-0.292*	0.405***	2.177***	-1.591***
	(2.723)	(0.153)	(0.00554)	(0.695)	(0.0280)
Observations	775	775	775	775	775
R^2	0.101	0.031	0.120	0.025	0.231

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

Table 2: Regressions of log college premium on log geography variables

our sample include Iowa City, IA; Santa Fe, NM; Ames, IA; Lincoln, NE; Albuquerque, NM; and Missoula, MT. Iowa City is the location with the highest college population share in the United States, and the other locations on this list also contain high college population shares. Farid's affiliation, Penn State, is clearly visible as the college population share outlier in the center of Pennsylvania. These relationships are not surprising because each of these locations contains a large university or national laboratory. Remote universities are vestiges of a 19th century US policy encouraging each state to fund its own public university, and the Sandia and Los Alamos laboratories at Santa Fe were deliberately located in remote locations due to secret research. Since these exogenous shocks to high skilled productivity in remote locations are outside of our model, we do not try to capture the possible positive correlation between remoteness and college population share in our theory.[8]

The takeaway from the empirical section is that geographical position seems to have a strong effect on inequality. In particular, relatively remote locations in the United States have relatively low population density and low college wage premia. These two facts motivate the theory developed in the next section.

2 Theory

In the last section, we presented evidence that geography plays a role in shaping the college wage premia. In this section we generalize spatial models of trade to explain our findings as well as the previously documented fact that college population shares and college wage

[8]There is more work to be done excluding these public university outliers and reestimating.

VARIABLES	(1) L pop dens	(2) L wage prem	(3) L wage prem	(4) L coll shr	(5) L coll shr
L remote	-0.703***	-0.0115*		0.0428	
	(0.0957)	(0.00603)		(0.0321)	
L pop dens			0.0211***		0.151***
			(0.00203)		(0.0105)
Constant	-3.659***	0.356***	0.405***	-0.887***	-1.591***
	(0.821)	(0.0512)	(0.00554)	(0.270)	(0.0280)
Observations	775	775	775	775	775
R^2	0.066	0.005	0.120	0.002	0.231

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

Table 3: Regressions of log college premium on log geography variables, geodesic distance

premia are positively correlated with population size. In particular, the model introduces a mechanism to explain why the college wage premium and population decreases with the remoteness of a location.

2.1 Environment

The model is static, with a continuum of locations $j \in J$, a continuum of skilled workers with total population N_H, and a continuum of unskilled workers with total population N_L. Workers can choose to reside and work in any (single) location. Firms in each location can produce a location-specific variety of a tradable final good with a constant elasticity production function using the two types of labor as inputs. Both workers and firms are price takers in perfectly competitive markets.

2.2 Worker's problem

The utility of a worker in location i is:

$$Q(i)u(i)\varepsilon(i) \tag{1}$$

where $Q(i)$ is utility from tradeables, $u(i)$ is the utility from exogenous amenities, and $\varepsilon(i)$ is worker-specific preference for location i. Tradeable goods are differentiated by the origin of production, and $q(j,i)$ is consumer's consumption in i from goods originated in location j.

The aggregator $Q(i)$ in the utility function with elasticity of substitution $\sigma > 0$ is:

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

We assume that worker location prefrences $\varepsilon(i)$ are independent across workers and locations, and have a Type II Extreme Value distribution,

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

Workers are endowed with exogenous skill $s \in \{H, L\}$. A worker's wage in location i is denoted by $w_s(i)$, and her budget constraint is

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

where $p(j,i)$ is price of good j in destination i, and J is the given set of locations. The *only* place skill type appears in the worker's problem, is through the wage's effect on the budget constraint. Otherwise workers of both types have the same preferences.

A worker has two types of decision to make. She decides where to live, and how much of each good to consume. Given a choice of location, the second problem is standard. A worker of type s in location i spends $x_s(j,i)$ on goods produced in j,

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

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

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

The second decision a worker makes is where to live. In order to characterize this decision, we introduce congestion forces as in Allen and Arkolakis (2014):

$$u(i) = \bar{u}(i) n(i)^{\gamma} \tag{5}$$

Here, $n(i)$ is total population in location i and $\gamma < 0$ is the degree of congestion effect. This specification is isomorphic to a setting where preferences are Cobb-Douglas between tradeables and housing in which housing supply is inelastic. A worker with skill level s faces the following discrete choice problem over locations:

$$\max_{i \in J} \frac{w_s(i)}{P(i)} u(i)\varepsilon(i)$$

Using the properties of the Type-II extreme value distribution, we can characterize the share of type-s labor in location i as:[9]

$$\frac{n_s(i)}{n_s(j)} = \left(\frac{w_s(i)u(i)/P(i)}{w_s(j)u(j)/P(j)} \right)^{\theta} \tag{6}$$

where $n_s(i)$ is population of workers with type s in location i. The elasticity of relative labor supply to relative wages is:

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

The variance of $\varepsilon(i)$ across both workers and locations is decreasing in θ. A large θ implies that individual unobserved preferences for location are similar across locations. Small changes in 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. That is, the supply curve of workers to a location is steep.

A location with a higher relative wage for skilled workers will attract relatively more skilled workers, but it won't attract all the skilled workers. Small changes in relative wages will induce small changes in relative populations. In a model without preference heterogeneity,

[9]Incidentally this equation gives the model its first testable prediction:

$$\frac{\frac{w_H(i)}{w_L(i)}}{\frac{n_H(i)}{n_L(i)}^{\frac{1}{\theta}}} = \frac{\frac{w_H(j)}{w_L(j)}}{\frac{n_H(j)}{n_L(j)}^{\frac{1}{\theta}}}$$

A higher ratio of skilled workers to unskilled workers implies a higher skill premium in wages. In the data, the college wage premium and college population share are strongly positively correlated. A simple regression implies a θ value of 1.5.

in equilibrium all populated locations must offer the same utility within skill type. This means that a location which deviates and offers a slightly higher wage will attract *all* workers.

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

$$W_s \equiv \left[\int_{j=0}^{\|J\|} \left(w_s(j) P(j)^{-1} u(j) \right)^{\theta} dj \right]^{\frac{1}{\theta}}$$

This index is proportional to the expected welfare of a worker of type s before she draws her location preferences.[10] Again using properties of the Type-II extreme value distribution, we derive the relationship for all $i \in J$:

$$\frac{n_s(i)}{N_s} = \left(\frac{w_s(i) P(i)^{-1} u(i)}{W_s} \right)^{\theta} \tag{7}$$

Here N_s is the exogenously given total population of type-s workers. If a location offers relatively high welfare, it will have a relatively high population, with the extent of the relationship governed by θ.

2.3 Firm's problem

Each location has a representative firm with a CES production function using skilled and unskilled workers,

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

where $A(i)$ is total factor productivity in location i. $\varepsilon > 0$ is the elasticity of substitution between high and low skill workers. $\beta_H(i) > 0$ and $\beta_L(i) > 0$ are factor intensities. Differences in factor productivities (and agglomeration, to be discussed shortly) are the *only* differences between skilled and unskilled labor. Following the literature, we distinguish two agglomeration forces. First

$$A(i) = \bar{A}(i) n(i)^{\alpha} \tag{8}$$

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

with $\alpha > 0$. This agglomeration force changes productivity of both low and high skill workers. A standard Krugman type monopolistic competition with free entry generates the same relation through endogenous measure of firms (if $\alpha = 1/(1+\sigma)$). In addition, there is strong evidence that agglomeration forces are stronger for high skill workers (Glaeser and Resseger, 2010). Motivated by these findings, we make a skilled worker's productivity covary positively with the population of skilled workers in a city:[11]

$$\begin{aligned} \beta_H(i) &= \bar{\beta}_H(i) n_H(i)^{\varphi} \\ \beta_L(i) &= \bar{\beta}_L(i) \end{aligned} \tag{9}$$

Here, $\varphi > 0$ governs the skilled bias technological change for high skilled labor. By cost minimization, the unit cost of production equals

$$\frac{c(i)}{A(i)}, \text{ where } c(i) = \left[\beta_H(i)^{\varepsilon} w_H(i)^{1-\varepsilon} + \beta_L(i)^{\varepsilon} w_L(i)^{1-\varepsilon}\right]^{\frac{1}{1-\varepsilon}} \tag{10}$$

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

$$b(i) = \frac{\beta_H(i)^{\varepsilon} w_H(i)^{1-\varepsilon}}{\beta_H(i)^{\varepsilon} w_H(i)^{1-\varepsilon} + \beta_L(i)^{\varepsilon} w_L(i)^{1-\varepsilon}} \tag{11}$$

As $\beta_H(i)$ is increasing in the population of skilled workers in a city, so is $b(i)$.

Now the intuition for our result is available. Suppose a new highway is built to a city. As a result of reductions in trade costs, the city will benefit from a lower price index and higher sales abroad, so firms in the city will attract workers of both types. As the city grows, productivity rises. However, due to agglomoreation advantages, skilled workers become relatively more productive. Firms will demand relatively more skilled workers. There will be a disequilibrium as firms demand relatively more skilled workers than the skill share in the population. Equilibrium is restored by raising skilled wages, which will both attract skilled workers and discourage firms from hiring them.

As mentioned above, markets are perfectly competitive. Let $d(j,i)$ be the trade costs of shipping a good from j to i. The price of a good produced in location j and consumed in location i is:

[11] For both productivities and amenities, we have been talking about population while in our empirical exercise we used population density. From (8) and (9), it can be seen that the area is just a constant which will be absorbed by the levels and could in principle be easily adjusted for.

$$p(j,i) = c(j)d(j,i)/A(j) \tag{12}$$

2.4 Spatial Equilibrium

A *spatial equilibrium* is a set of $w_H(i)$, $w_L(i)$, $n_H(i)$, and $n_L(i)$ such that:[12]

1. Labor shares in locations are according to (7).

2. Demand and supply of skilled and unskilled labor are equal in each location.

3. Goods market clear

$$w_H(i)n_H(i) = b(i)\int_J \sum_{s\in\{H,L\}} n_s(i)x_s(j,i)\ dj \tag{13}$$

   where $x_s(j,i)$ is given by (3), and $b(i)$ is given by (11).

4. Allocation of labor is feasible, $\int_J n_H(j)\ dj = N_H$ and $\int_J n_L(j)\ dj = N_L$. Wages are normalized $\int_J w_H(j)dj = 1$.

2.5 Equilibrium Analysis

Before we derive a system of equations characterizing the equilibrium, we first discuss some of its features. In particular, we show that without both heterogeneous location preferences and skill-specific agglomeration, our model would not generate location-specific skill premia. However, we need no other type of skill-specific preferences, productivity, or mobility to generate this result.

Perfect competition gives us that the income share of skilled labor equals firm's optimal input share on skilled labor:

$$\frac{w_H(i)n_H(i)}{w_H(i)n_H(i) + w_L(i)n_L(i)} = \frac{\beta_H(i)^\varepsilon w_H(i)^{1-\varepsilon}}{\beta_H(i)^\varepsilon w_H(i)^{1-\varepsilon} + \beta_L(i)^\varepsilon w_L(i)^{1-\varepsilon}}$$

[12] There is a bit hidden in this definition, as we have already used perfect competition to give us goods prices $p(j,i)$ given wages, the consumer's optimization problem to give us demand $x(j,i)$ given prices and wages, and the firm's optimization problem to give us relative labor demand (11)

With some algebra, and substituting from (9), we get:

$$\frac{n_H(i)}{n_L(i)} = \left(\frac{\bar{\beta}_H(i)}{\bar{\beta}_L(i)}\right)^{\varepsilon} \left(\frac{w_H(i)}{w_L(i)}\right)^{-\varepsilon} n_H(i)^{\varphi\varepsilon} \tag{14}$$

Using equation (7), the supply side of labor markets imply:[13]

$$\frac{n_H(i)}{n_L(i)} = \frac{N_H}{N_L}\left(\frac{W_H}{W_L}\right)^{-\theta} \left(\frac{w_H(i)}{w_L(i)}\right)^{\theta} \tag{15}$$

Labor markets clear when skill premia simultaneously satisfy the pairs of demand (14) and supply (15). Combining, we get:

$$\frac{w_H(i)}{w_L(i)} = \left(\frac{W_H}{W_L}\right)^{\frac{\theta}{\theta+\varepsilon}} \left(\frac{N_H}{N_L}\right)^{\frac{-1}{\theta+\varepsilon}} \left(\frac{\bar{\beta}_H(i)}{\bar{\beta}_L(i)}\right)^{\frac{\varepsilon}{\theta+\varepsilon}} n_H(i)^{\frac{\varphi\varepsilon}{\theta+\varepsilon}} \tag{16}$$

Suppose we were to shut down preference heterogeneity, $\theta \to \infty$. From (16) we see that the skill premium will be constant across locations (and, surprise, proportional to ex-ante expected welfare). On the other hand, suppose there is no agglomeration for skilled workers, $\varphi = 0$. Then the skill premium can vary between destinations only due to exogenous differences in returns to skilled and unskilled labor. In order to have equilibria with (endogenously) varying skill premia, we need *both* θ to be finite and $\varphi > 0$. In words, we need both variance in unobserved location preferences and the agglomeration force for skilled workers. Large cities demand more skilled workers due to agglomeration, but when unobserved location preferences matter skilled workers do not fully arbitrage the a wage increase away. In this respect, our model departs from a standard Rosen-Roback (or Allen-Arkolakis) model by highlighting the interaction of agglomeration and the labor supply elasticity.

Next we show that the distribution of workers across space has a real effect on welfare inequality. Plugging equilibrium skill premium from (16) into relative labor supply (15), distribution of low skilled labor can be written as a function of distribution of high skilled labor,

$$n_L(i) = \left(\frac{\bar{\beta}_H(i)}{\bar{\beta}_L(i)}\right)^{\frac{-\theta\varepsilon}{\theta+\varepsilon}} \left(\frac{W_H}{W_L}\right)^{\frac{\theta\varepsilon}{\theta+\varepsilon}} \left(\frac{N_H}{N_L}\right)^{\frac{-\varepsilon}{\theta+\varepsilon}} \Big(n_H(i)\Big)^{\frac{\theta(1-\varepsilon\varphi)+\varepsilon}{\theta+\varepsilon}} \tag{17}$$

[13] As far as we are aware, other models of inequality in the literature assume fundamental differences in preferences between skilled and unskilled workers. To derive (15) we are using the fact that skilled and unskilled workers value amenities in the same way, and that they face the same prices.

Define $\rho_H(i) = n_H(i)/N_H$ as the density of high skill labor. Equation (17) and $N_L = \int_J n_L(i)\, di$ pin down the relative average well-being of skilled and unskilled workers, W_H/W_L, as a function of $\rho_H(i)$,

$$\frac{W_H}{W_L} = \underbrace{\left(\frac{N_H}{N_L}\right)^{-\frac{1}{\varepsilon}}}_{\text{aggergate scarcity}} \times \underbrace{(N_H)^{\varphi}}_{\text{aggregate agglom.}} \times \underbrace{\left[\int_J \left(\frac{\bar{\beta}_H(i)}{\bar{\beta}_L(i)}\right)^{\frac{-\theta\varepsilon}{\theta+\varepsilon}} \Big(\rho_H(i)\Big)^{\frac{\theta(1-\varphi)+\varepsilon}{\theta+\varepsilon}} di\right]^{-\frac{\theta+\varepsilon}{\theta\varepsilon}}}_{\text{distributional effect}} \tag{18}$$

Equation (18) decomposes the three forces behind real well-being inequality: 1) Aggregate scarcity of high skill labor, 2) Aggregate agglomeration advantage of high skill workers, 3) A weighted average of relative exogenous productivities in which weights are determined by the the distribution of the density of high skill labor. While the first two behave at the aggregate, the third relates to the distribution of skills. In this sense, Equation (18) relates an index of real inequality at the national level to the distribution of high skill workers across cities within the nation.

2.6 Equilibrium equations

In this section we characterize our equilibrium with only two integral equations. These integral equations will be the basis of our empirical exercise.

By simply inverting the definition (11) of input share $b(i)$, we get total income in i equal to $\frac{1}{b(i)} w_H(i) n_H(i)$. Using the goods market clearing condition, equation (13), we derive:

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

Equations (10) and (11) further imply that

$$\begin{aligned} c(i) &= \tilde{c}(i) w_H(i) \\ \text{where } \tilde{c}(i) &= \Big[\bar{\beta}_H(i) n_H(i)^{\varphi}\Big]^{\frac{\varepsilon}{1-\varepsilon}} \Big[b(i)\Big]^{\frac{-1}{1-\varepsilon}} \end{aligned} \tag{19}$$

Replacing $c(i)$ with $\tilde{c}(i)$ in the above integral equation and also replacing the price index using

(7), after some algebra we get:

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

Again substituting the price index from (7) and $\tilde{c}(i)$ into (4), we get another integral equation:

$$\begin{aligned}
&u(i)^{1-\sigma}n_H(i)^{(\sigma-1)/\theta}w_H(i)^{1-\sigma} = \\
&W_H^{1-\sigma}N_H^{\frac{\sigma-1}{\theta}}\int_J d(j,i)^{1-\sigma}A(j)^{\sigma-1}\tilde{c}(j)^{1-\sigma}w_H(j)^{1-\sigma}\,dj
\end{aligned} \tag{21}$$

Equations 25–26 give us two systems of integral equations. If $\varphi = 0$ and $\theta = \infty$, the integral equations collapse to those in Allen and Arkolakis (2014).

We can reduce our system further using a method from Allen and Arkolakis (2014). If either of the systems of integral equation hold along with the following relation, both systems of integral equations must hold:

$$b(i)^{-1}A(i)^{1-\sigma}\tilde{c}(i)^{\sigma-1}n_H(i)w_H(i)^{\sigma} = \lambda u(i)^{1-\sigma}n_H(i)^{(\sigma-1)/\theta}w_H(i)^{1-\sigma} \tag{22}$$

where $\lambda > 0$ is some constant. The solution algorithm developed below is based on successive iterations over $n_H(i)$.

There are two "endogenous" variables hidden in (25) - (22), $b(i)$ and $\tilde{c}$. Endogenous here means that they depend upon $n_H(i)$. Thus solving our model involves an inner loop to update these endogenous variables. Details of our solution algorithm are contained in Appendix C.

Allen and Arkolakis (2014) have a slick proof of uniqueness, which depends only on the relationship between parameters governing agglomeration and parameters governing congestion. Unfortunately the same proof cannot be used in our setting. Intuitively, there should be a similar condition in our model relating the agglomeration forces to the congestion force will be that the parameter governing joint agglomeration α and the parameter governing skilled agglomeration φ cannot be too large relative to the parameter governing congestion γ.

3 Estimation

In this section we discuss the estimation of our structural model. Trade costs are estimated first, followed by a loop which estimates the base productivities and amenities of each location given observed wage premia and college population shares.

3.1 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 one thinks exogenous trade costs are simply quadratic in distance, then one might think a location in the center of the United States would have the lowest trade cost. The geographical center of the United States is just outside of Lebanon, Kansas, population 218.

The flaw with this line of thinking is that the geography affects 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 it is natural to think that shipping bagels to Miami is cheaper because of the possibility of using a ship. To this end, 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 detail in the data appendix and in the origin Allen and Arkolakis paper.

There are three steps to the estimation process. In the first step, we use three separate image files each showing 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 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.

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 don't know the exchange rate, as it were, of road travel to water transport. The second step is to use a discrete choice framework and the fraction of observed trade flows via each mode between each pair of locations in order to back out these exchange rates. The idea is that if a large share of transport is via road, then it must be that road is a relatively cheap form of transportation.

At the end of the second step, we have recovered the exchange rates, or if you like relative costs, between each of the forms of transport. It is not enough for our structural model, however, 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 costs as well. To do this we return to a trade classic, the gravity model. The version we use is given by combining (3) and (12) to get exports from location j to location i:

$$X(j,i) = \Big[\frac{c(j)d(j,i)}{A(j)P(i)}\Big]^{1-\sigma} W_i \tag{23}$$

Here W_i is the total income of workers in location i. Take logs, and the only bilateral term in (23) is the distance cost. Using the discrete choice model, we can get the average cost of transport up to an unknown parameter. Running a regression of the logged version of (23) with origin and destination fixed effects and using the average cost of transport for $d(j,i)$ pins down the scale of trade costs.

3.2 Estimation of other parameters

This section is concerned with estimating and calibrating all universal and location specific parameters, which are listed in Table 4). First we discuss the parameters we calibrate from the literature, then parameters we get from reduced form estimates, and then finally parameters we estimate from the equilibrium integral equations of our structural model.

3.2.1 Parameters taken from the literature

Two critical parameters for our model are the consumer's elasticity of substitution across goods σ and the non-skill specific agglomeration force α. Allen and Arkolakis (2014) set $\sigma = 9$, but Monte et al. (2015) set $\sigma = 4$. Allen and Arkolakis justify their relatively high choice with the intuition that goods produced within a country are more substitutable than than goods produced in different countries. The Allen and Arkolakis model is isomorphic to Monte et. al if $\alpha = 1/(\sigma - 1)$ (see appendix 2 of Monte et. al). In Monte et. al, then, the implied $\alpha = 1/(4 - 1) = 0.33$, while Allen and Arkolakis choose $\alpha = 0.1$ citing Rosenthal and Strange (2004) who report that productivity rises by 3-8% when population doubles.

universal parameters:	
σ	elasticity of substitution across goods
ε	elasticity of substitution across low- and high-skilled labor
γ	congestion strength
α	common agglomeration strength
θ	dispersion parameter of unobserved preferences for locations
φ	high-skill agglomeration advantage
location-specific parameters:	
$\bar{\beta}_H(i)$	basic high-skill factor intensity in location i, $\bar{\beta}_L(i) = 1 - \bar{\beta}_H(i)$
$\bar{A}(i)$	basic TFP in location i
$\bar{u}(i)$	basic amenity in location i

Table 4: List and decription of all model parameters

Gottlieb and Glaeser (2009) classify agglomeration forces in the literature with three categories: (1) urban concentration brings about a reduction in transport costs; (2) urban concentration facilitates labor movement across firms; (3) urban concentration encourages the exchange of ideas and innovation. The first two forces apply to both types of labor, while the third is more important for high skilled labor. Recent studies of the wage premium have emphasized the exchange of ideas (Davis and Dingel, 2012; Baum-Snow and Pavan, 2013).

Since σ and α will affect welfare, we choose a relatively high $\sigma = 9$ and a low $\alpha = 0$ within the ranges suggested by the literature because this will tend to bias our welfare estimates down.

Value of γ. In Mante et al., preferences are Cobb-Douglas over differentiated goods and housing, with the share of housing be equal δ. Again, an isomorphism with Allen and Arkolakis requires that congestion $\gamma = \delta/(1-\delta)$. Allen and Arkolakis report that, according to the Bureau of Labor Statistics (BLS), the housing expenditure γ in the year 2000 was in the range of 19%–25% depending on which items you count as housing. This implies that congestion γ in the range of -0.23 to -0.33. Allen and Arkolakis choose $\gamma = -0.30$. On the other hand, Monte et al. report that $\delta = 0.40$ according to the Bureau of Economic Analysis, so $\gamma = -0.67$. In Diamond (2015), housing share in expenditures is in the range of 0.39 to 0.46 according to the year 2000 Consumer Expenditure Survey. In Moretti (2013) housing share is 42.7% according to the BLS (counting fuel and utilities as housing). The bottom line is that $\delta \in [0.35, 0.40]$ is in line with the literature, so we choose $\gamma = -0.6$.

Value of ε. Katz et al. (1999) provide a literature review where they report values for the elasticity between high and low skill labor in the production function of 1.40 to 1.70. Ciccone and Peri (2006) come up with estimated values of elasticity of substitution between unskilled and skilled labor between 1.3 and 2. Diamond (2015) estimates $\varepsilon = 1.6$. Card (2009) using data on MSA's finds that $\varepsilon = 2.5$. There are many other similar estimates.

One notable exception is Baum-Snow et al. (2014). They use CBSAs, and their elasticity ε is the elasticity between capital-augmented high-skilled labor and low-skilled labor. In their reduced form estimation they find ε roughly in the range of 1.5 to 2.5. But in their structural estimation, they find $\varepsilon \in [6.7, 10]$. To be consistent with a large amount of the literature, however, we choose $\varepsilon = 2$ which lies smack in the middle of the range of 1 to 3.

3.2.2 Parameters pinned down by reduced-form estimations

We can directly read a number of parameters through reduced form estimates.

Value of θ. Let $\omega(i) \equiv w_H(i)/w_L(i)$ be the skill premium in location i. The model implies,

$$\log\left(\frac{\omega(i)}{\omega(j)}\right) = \frac{1}{\theta}\log\left(\frac{n_H(i)/n_L(i)}{n_H(j)/n_L(j)}\right),$$

which delivers the following stochastic form:

$$\log\omega(i) = \alpha_0 + \alpha_1 \log\Big(n_H(i)/n_L(i)\Big) + error_{ij} \tag{24}$$

where in the model, $\alpha_0 = \log\omega(1) - \alpha_1 \log\Big(n_H(1)/n_L(1)\Big)$, $\alpha_1 = 1/\theta$; with $j = 1$ as the reference, since we have only $J - 1$ independent observations.

Value of φ and β_H. At the equilibrium of labor market (equation 15),

$$\frac{w_H(i)}{w_L(i)} = \left(\frac{W_H}{W_L}\right)^{\frac{\theta}{\theta+\varepsilon}} \left(\frac{N_H}{N_L}\right)^{\frac{-1}{\theta+\varepsilon}} \left(\frac{\bar{\beta}_H(i)}{1-\bar{\beta}_H(i)}\right)^{\frac{\varepsilon}{\theta+\varepsilon}} n_H(i)^{\frac{\varphi\varepsilon}{\theta+\varepsilon}},$$

which delivers the exact following regression:

$$\log\omega(i) = \alpha_0 + \alpha_1 \log n_H(i) + \zeta(i).$$

Given ε and θ, which are already known by previous sections,

$$\begin{aligned} \varphi &= \frac{\alpha_1(\theta+\varepsilon)}{\varepsilon} \\ \frac{\beta_H(i)}{1-\beta_H(i)} &= \exp\left[\frac{\zeta(i)(\theta+\varepsilon)}{\varepsilon}\right] \end{aligned}$$

3.2.3 Parameters pinned down by the system of equilibrium equations

Finally, the inherent, base values of productivity and amenities will be estimated from the equilibrium integral equations implied by our model. We rewrite the two systems of integral equations:

$$\bar{A}(i)^{1-\sigma} = W_H^{1-\sigma} N_H^{\frac{\sigma-1}{\theta}} \tilde{c}(i)^{1-\sigma} n_H(i)^{-1} w_H(i)^{-\sigma} b(i) n(i)^{(\sigma-1)\alpha} \int_J d(i,j)^{1-\sigma} \bar{u}(j)^{\sigma-1} n(j)^{(\sigma-1)\gamma} n_H(j)^{(\theta+1-\sigma)/\theta} w_H(j)^{\sigma} b(j)^{-1} dj \quad (25$$

$$\bar{u}(i)^{1-\sigma} = W_H^{1-\sigma} N_H^{\frac{\sigma-1}{\theta}} n_H(i)^{(1-\sigma)/\theta} w_H(i)^{\sigma-1} n(i)^{(\sigma-1)\gamma} \int_J d(j,i)^{1-\sigma} \bar{A}(j)^{\sigma-1} n(j)^{(\sigma-1)\alpha} \tilde{c}(j)^{1-\sigma} w_H(j)^{1-\sigma} dj \quad (26)$$

Here, $\bar{A}(i)$ and $\bar{u}(i)$ are the only unknowns, while all parameters are already estimated/calibrated and all other variables are given by data. We can further reduce the two systems of equation into one. As in Allen and Arkolakis (2014), it is necessary that trade costs are symmetric $d(i,j) = d(j,i)$. Consider the following equation,

$$\bar{u}(i)^{\sigma-1} n(i)^{(\sigma-1)\gamma} n_H(i)^{(\theta+1-\sigma)/\theta} w_H(i)^{\sigma} b(i)^{-1} = \lambda \bar{A}(i)^{\sigma-1} n(i)^{(\sigma-1)\alpha} \tilde{c}(i)^{1-\sigma} w_H(i)^{1-\sigma} \tag{27}$$

Along with (27), system (25) of integral equations delivers (26). $\lambda > 0$ is a constant. The numerical algorithm by which we solve these equations is described in Appendix B.

4 Results

The trade cost estimates we get are reported in Table 5, and our other calibrated and estimated parameters are given in Table 6. Our value for preference dispersion θ is lower than that found in Allen and Arkolakis (2014), two versus ten. This means that our individuals will have stronger opinions about location, and wage differentials will have to be higher to induce movement. Our trade costs are also qualitatively a bit different, especially water has a higher variable cost in our estimation.

0.44	Road variable cost
0.31	Rail fixed cost
0.40	Rail variable cost
0.34	Water fixed cost
0.68	Water variable cost
0.49	Air fixed cost
0.32	Air variable cost

Table 5: Trade cost estimates

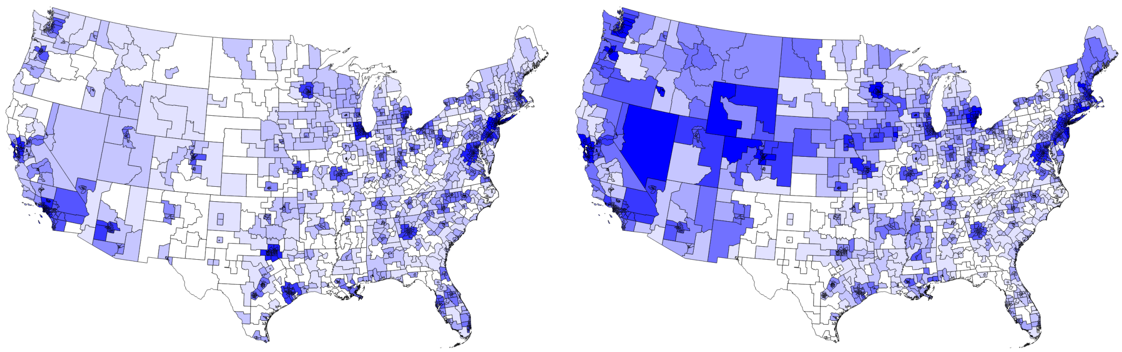
universal parameters:		
σ	9	elasticity of substitution across goods
ε	2	elasticity of substitution across low- and high-skilled labor
γ	-0.6	congestion strength
α	0	common agglomeration strength
θ	2.62	dispersion parameter of unobserved preferences for locations
φ	0.085	high-skill agglomeration advantage
location-specific parameter means:		
$\bar{\beta}_H(i)$	0.45	basic high-skill factor intensity in location i
$\bar{A}(i)$	2.52	basic TFP in location i
$\bar{u}(i)$	1,309	basic amenity in location i

Table 6: Other estimated model parameters

Figure 2 contains a visualization of our estimated base amenities and productivities. The more intense the blue, the higher the measure is. Productivities are relatively high in the Western part of the United States, as the model needs to induce people to live in those regions with relatively high trade costs. Amenities follows a similar pattern, with high population density areas estimated to have high amenities. This result is also found in **?**, as the model needs to justify why cities are located exactly where they are located rather than nearby. It is reassuring that the model chooses to attract people to the Northern and Rocky Mountain regions with productivity rather than amenities.

In our baseline estimation, the ratio of expected high skill welfare to low skill welfare is 1.6907. The welfare is expected because an individual's realized welfare depends on the particular location preference draw he receives. Suppose that the government were to start a campaign to build highways to relatively remote parts of the United States. How would such a policy affect the distribution of wage inequality and overall welfare inequality?

In order to answer this question, we reestimate our model after reducing the highest 30% of trade costs, trade costs of at least two. In Figure 3, we see that welfare inequality falls with



(a) Amenities (b) Productivities

Figure 2: Locations by estimated value

the reduction in trade costs. In Figure 4 the largest cities have decreases in population, and the smallest, more remote cities have increases in population.

In Figure 5, we see the geography of changes in population and wage premium after a 10% reduction in the highest trade costs. The map for change in college share looks identical. The maps are the same, because the forces which increase population also increase the wage premium and college share.

In many trade models, reductions in trade barriers leads to relative losses for low skill labor. In our model, it is possible that a decrease in barriers gives low skill workers relative gains. If we reduce the trade cost to the smallest, most remote cities, those cities become more attractive to live in. On the margin, people move to those cities. Population inequality decreases, which causes a loss of gains from agglomeration accruing to high skilled workers, so low skill people gain relatively more from a reduction in trade costs.

5 Conclusion

We find that remote cities are less dense and have a relatively low college wage premium. These stylized facts motivate the development of our spatial equilibrium model. We show that both heterogenous location preferences among workers and stronger agglomeration forces for skilled workers deliver the moments we find in the data. While our model delivers simple intuition, it is rich enough to estimate using American data. After estimating our model, we find that reducing the costs of trade for the most remote cities reduces inequality in city size, and leads

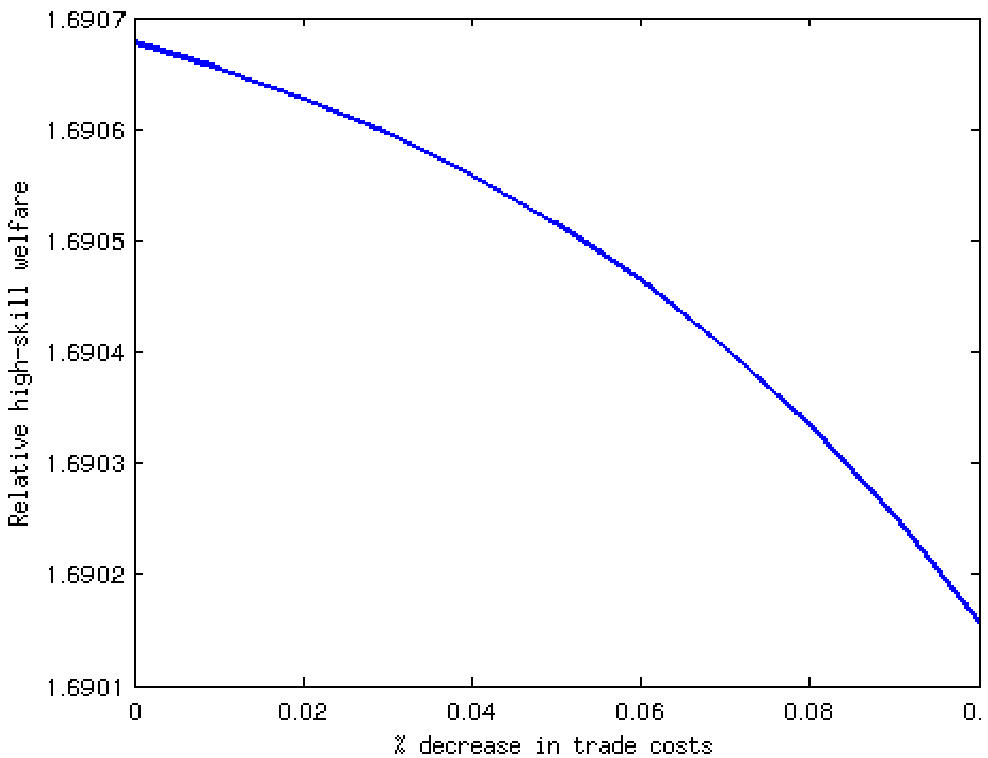


Figure 3: Relative welfare change reducing high trade costs

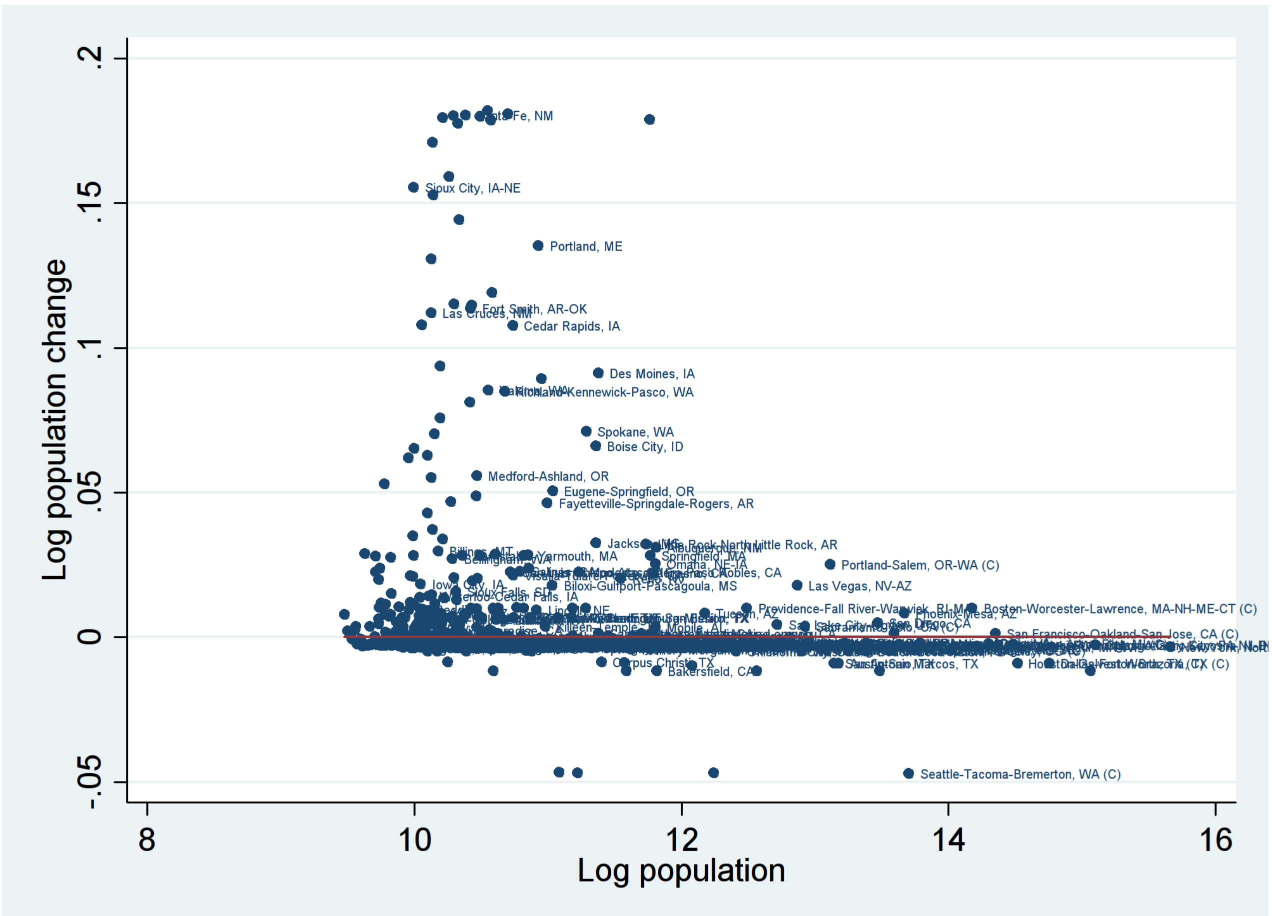
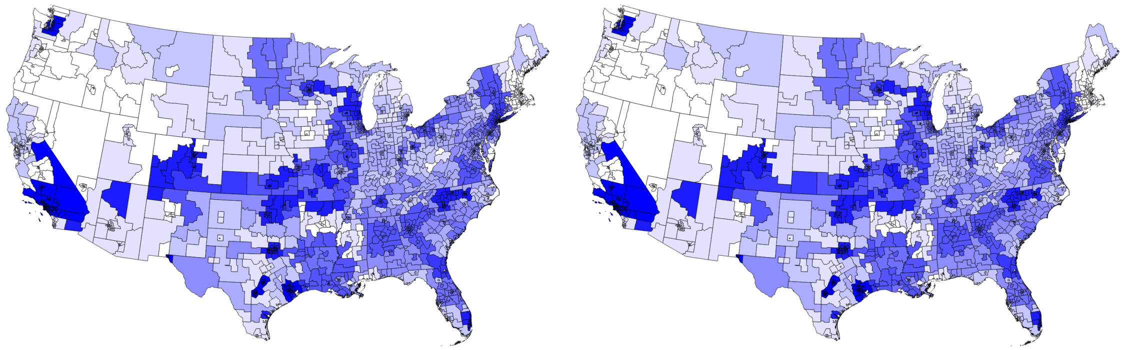


Figure 4: Population change after 10% reduction in high trade costs



(a) Population

(b) Wage premium

Figure 5: Locations by change after trade cost reduction

to a reduction in the overall welfare inequality.

References

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

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

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

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.

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.

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.

Desmet, K., Nagy, D. K., and Rossi-Hansberg, E. (2014). The geography of development: Evaluating migration restrictions and coastal flooding.

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. 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.

Gottlieb, J. D. and Glaeser, E. L. (2009). The wealth of cities: Agglomeration economies and spatial equilibrium in the united states. *Journal of Economic Literature*, 47(4):983–1028.

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

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.

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

Moretti, E. (2008). Real wage inequality. Technical report, National Bureau of Economic Research.

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.

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 7: Variables from IPUMS 5% sample

A Data appendix

In this appendix, we describe exactly what 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.[14] Table **??** described the variables we downloaded in the initial sample.

Using these variables, we cleaned the data by modifying replication code for Baum-Snow and Pavan (2013) from Nathanial Baum-Snow's website.[15] 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, those which made less than minimum wage

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

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

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".

After cleaning the data, the next step is to merge in the latitude, longitude, and physical areas of PUMAs and MSA's. 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.[16]. 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.[17] 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

[16] http://mcdc2.missouri.edu/websas/geocorr2k.html
[17] https://usa.ipums.org/usa/resources/volii/puma00_puma10_spatial_crosswalk.xlsx

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:[18]

> 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 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.[19] 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_Mod 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.[20] We manipulate

[18] https://usa.ipums.org/usa-action/variables/METAREA#description_section

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

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

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 8: Demographic variables from IPUMS 5% sample

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 8 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_matr
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.[21] This means that we strip off all text and put the CFS area coor-

[21] Allen and Arkolakis wrote their estimation in Matlab and kindly made the files available to us. We use a

dinate 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.[22] 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 Algorithm for solving integral equations

Up to scale such that $\int_J \bar{A}(i)^{\sigma-1}di = \mathcal{A}$.

1. Guess $\bar{A}(i)$.
2. Define $f(i) = \lambda^{-1/2}\bar{A}(i)^{1-\sigma}$, $\kappa = W_H^{1-\sigma}N_H^{\frac{\sigma-1}{\theta}}$, and

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

   Then, the system of integral equations described by (25) can be written as follows:

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

modified version of their code to calculate our distance costs.

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

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

$$f^{(t+1)}(i) = \frac{\int_J K(j,i) f^{(t)}(j)^{-1} \, dj}{\int_J \int_J K(j,i) f^{(t)}(j)^{-1} \, djdi} \tag{28}$$

Note that as we divide integrals in 28, we do not need to know κ to update our guess. If at some iteration t, $f^{(t)}(i)$ is close enough to $f^{(t-1)}(i)$ for all i, stop updating.

3. Find the scale parameters λ.

$$1 = \int_J f(j) = \lambda^{-1/2} \int_J \bar{A}(j)^{1-\sigma} dj = \lambda^{-1/2} \mathcal{A} \;\Rightarrow\; \lambda = \mathcal{A}^{1/2} \tag{29}$$

While by construction $\int_J f(j) d_j = 1$, the scale of productivity is set by $\mathcal{A}$,

$$\bar{A}(i)^{\sigma-1} = f(i)\mathcal{A}$$

4. Find κ. For each i, the following is a constant equal to κ

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

So,

$$W_H = \kappa^{\frac{1}{1-\sigma}} N_H^{\frac{1}{\theta}} \tag{30}$$

5. Using equation (27), calculate $\bar{u}(i)$.

C Counterfactual Simulation Algorithm

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

2. Compute W_H/W_L according to (18). Then plug it in (17) to find $n_L(i)$.

3. Calculate skill premia, $\omega(i) \equiv w_H(i)/w_L(i)$, according to (16).

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

5. Find $\tilde{c}(i)$ according to (19).

6. Calculate $w_H(i)$ according to (22) up to scale parameter λ,

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

   where

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

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

$$K(j,i) = n_H(i)^{(1-\sigma)/\theta} u(i)^{\sigma-1} d(j,i)^{1-\sigma} A(j)^{\sigma-1} \tilde{c}(j)^{1-\sigma}$$

   Then, system of integral equations (26) 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

$$f^{(t+1)}(i) = \frac{\int_J K(j,i) f^{(t)}(j)\ dj}{\int_J \int_J K(j,i) f^{(t)}(j)\ djdi} \tag{31}$$

   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. Find λ by $\int_J w_H(j)dj = 1$ (the normalization defined in equilibrium),

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

   So,

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

   From here, find $w_H(i)$.

9. 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:

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

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