Cognitive Hubs and Spatial Redistribution*

Esteban Rossi-Hansberg
Princeton University

Pierre-Daniel Sarte
Federal Reserve Bank of Richmond

Felipe Schwartzman
Federal Reserve Bank of Richmond

July 10, 2019

Abstract

We study the allocation of workers with different occupations across U.S. cities. We propose and quantify a spatial equilibrium model with multiple industries that employ cognitive-non-routine (CNR) and alternative (non-CNR) occupations. We allow for city-industry-occupation specific productivity levels that are partly determined by externalities across local workers. We estimate the relevant parameters that determine these externalities within and across occupations using a number of instruments based on past migration patterns and the location of land-grant colleges. The productivity of CNR workers in a city depends significantly on its share of CNR workers and total employment size. Heterogeneous preferences for locations and the estimated externalities imply that the equilibrium allocation is not efficient. In equilibrium, the social value of CNR workers exceeds their private value by more than 70%. An optimal policy that benefits workers in both occupations equally, incentivizes the formation of creative hubs with larger fractions of CNR workers in some of today's largest cities. These cities should become smaller, so non-CNR workers receive transfers that incentivize them to move to small cities that have the appropriate industrial composition. We show that the spatial redistribution of CNR and non-CNR workers implied by the optimal policy reinforces equilibrium trends observed since 1980. These trends were generated by a nation-wide increase in the share of CNR workers combined with the local CNR specific externalities and low real-estate productivity in CNR-intensive cities.

*Preliminary draft. erossi@princeton.edu, pierre.sarte@rich.frb.org, felipe.schwartzman@rich.frb.org. We thank Mike Finnegan, Daniel Ober-Reynolds and Jackson Evert for excellent research assistance. The views expressed here are those of the authors and do not reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

1 Introduction

"*Most of what we know we learn from other people (...) most of it we get for free.*" Robert E. Lucas Jr.

Workers capable of doing the complicated cognitive non-routine tasks required in a modern economy are scarce. Acquiring the expertise to work as a doctor, manager, lawyer, computer scientist or researcher requires many years of schooling, sustained effort, and individual ability. These workers are a valuable input and so their allocation across industries and locations is important for the overall efficiency and welfare in an economy. The marginal productivity of a worker depends on the local productivity of the industry where she works, but also on the set of workers that work in the same city. Larger cities with a large fraction of workers in cognitive non-routine (CNR) occupations offer learning and collaboration opportunities that enhance the productivity of other workers. Of course, the abundance of CNR workers also lowers their marginal product, particularly in industries that are not intensive in these occupations. The interaction of these forces in equilibrium determines the segregation of workers and the specialization of industries in space. Can the economy allocate scarce CNR workers in a way that improves the lives of all workers? Our aim is to study the allocation of industries and occupations across cities in the U.S. and to design and quantify optimal spatial policies.

The need for optimal spatial policy is the direct implication of the presence of urban externalities. Externalities that enhance the productivity of workers in larger cities have been discussed, analyzed, and measured at least since Marshall (1920).[1] It is natural to hypothesize that these production externalities depend on the occupational composition of a city. After all, CNR occupations require more interactions between knowledgeable workers. As we show in detail in the next section, the patterns of occupational segregation and wages across space also suggest that this is the case. First, in the absence of technological differences or externalities across locations, decreasing returns to workers in an occupation imply that relative CNR to non-CNR wages should decline with the share of CNR workers. We find a large positive relationship, even after controlling for a number of observable worker characteristics.[2] Why are CNR workers making relatively more in locations where they are abundant? It could be that these locations specialize in industries intensive in these occupations. The evidence, however, suggest that firms in CNR abundant cities are more intensive in CNR workers. What makes demand for these workers high in these cities? Our take is that abundance of CNR workers itself makes them more productive: an occupation specific externality. Estimating the strength of these externalities is a central part of our analysis.

Proposing and quantifying optimal spatial policies requires a number of contributions. First, it requires us to develop a spatial equilibrium model with multiple industries and occupations, as well as occupation specific externalities. Multiple industries, costly trade,

[1] See Duranton and Puga (2004) for a review of the literature on externalities in cities.

[2] We also show that general amenities for both occupations cannot be attracting the best workers to CNR abundant cities, since although the real wages of CNR workers increase with a city's CNR intensity, the real wages of non-CNR workers do not, indicating that the amenities of these cities are not particularly desirable. For an alternative view, see Couture et al. (2018)

and input-output linkages are important since demand for different occupations depends on the occupation intensity of the industries in each location. The model also features occupation specific externalities that are allowed to depend on the share of workers in CNR occupations and the total workforce in the city. Heterogeneous preferences for locations act as a form of migration cost. Second we need to derive the optimal allocation in this setup. We choose to study efficient allocations that benefit both occupations equally. The optimal allocation exhibits transfers between locations and occupations. The characteristics of these transfers depend on the quantification of the model. Third, the quantification of this model requires us to estimate the parameters that determine the endogenous part of the city-industry-occupation specific productivity. We first recover the productivity levels that make our model match the observed equilibrium. We then parameterize the relationship between these productivity levels and the occupational composition and size of the city, and estimate this equation using an instrumental variable approach. As proposed in the empirical literature (e.g. Card (2001) and Moretti (2004a)), we use past migration flows of particular immigrant groups and the location of land-grant colleges as instruments. Our strategy yields a robust set of results that are consistent with the existing literature, although measured directly on the productivity recovered using our general equilibrium framework rather than wages, which is an advantage. They imply that the productivity of CNR and non-CNR workers depends similarly on city size, but that the productivity of CNR workers also strongly and significantly depends on the share of CNR workers. We find less evidence that the productivity of non-CNR workers depends on the composition of occupations. Fourth, we need to compute the optimal allocation, characterize it, and discuss its implementation with particular policy tools.

Our results propose a new approach to spatial policy. They indicate that we can improve the spatial allocation by making large CNR intensive cities smaller but even more CNR abundant than they currently are. These "cognitive hubs" take advantage of the scarce CNR input in the economy by grouping these occupations and maximizing the externalities among them. In equilibrium, the social value of CNR workers is 72% larger than their private value. Cognitive hubs should not emerge in all large cities, since some of these cities are productive in industries where CNR workers are less useful (e.g. Miami or Las Vegas), but they should emerge in many of them, scattered geographically across the country. In order to increase the share of CNR workers and to decrease congestion, it is optimal to move non-CNR workers to smaller cities with lower CNR shares that operate in industries where CNR workers are relatively less useful. The result is that many small cities should grow in size and become more non-CNR abundant, while many large cities should decline in size but become more CNR abundant. Hence, contrary to some of the previous literature and much of the public discourse, the economics of the problem indicate that small industrial cities in the U.S. should attract non-CNR workers, not try to become the next San Jose. Furthermore, the concentration of CNR workers in a few cities should be encouraged, not scorned. Everyone can benefit from using CNR workers in the most productive way possible, as long as the right transfers are implemented.

Naturally, achieving the optimal allocation requires a number of transfers and taxes that depend on the location and occupation of an agent. As Fajgelbaum and Gaubert (2018) discuss in their excellent paper on spatial policy, the optimal allocation requires a flat wage tax on all individuals to correct for the differences in the marginal utility of consumption generated by heterogeneous preferences for location. On top of this tax, in both theirs and our frameworks, implementing the optimal allocation requires a set of transfers. In our analysis, the base transfers for non-CNR workers amount to \$15,242 (in 2013 dollars) while CNR workers, who earn substantially more, end up paying a base transfer of \$11,573. One interpretation of this base transfer is as a "universal basic income" paid to all non-CNRs to make them benefit equally from the policy. CNR workers still needs to be incentivized to go to CNR intensive cities, and non-CNR workers to go to non-CNR intensive cities, so occupation and city-specific transfers are positively correlated with city size for CNR workers and negatively correlated with CNR share and size for non-CNR workers. Their exact value, however, also depends on the particular location and industrial composition of the city. Ultimately, the policy amounts to a subsidy to non-CNR workers to move to smaller cities with low CNR shares, and incentives to CNR workers to form even more intensive cognitive hubs in today's largest cities.

Perhaps surprisingly, when we compare the current spatial equilibrium allocation to that in 1980, we find that the pattern of spatial allocation of workers has approached the one implied by the optimal policy (with current fundamentals). Namely, since the 80s, CNR workers have become more abundant but also increasingly concentrated in CNR intensive hubs, many of which are large cities. This formation of cognitive hubs occurred in parallel to the increase in wage inequality across space and occupations. Our quantitative model implies that both processes were linked through local occupation-specific externalities. Without these spillovers, there would have been less spatial polarization and the welfare gains received by CNR workers would have been smaller than those of non-CNRs, since CNRs became more abundant. The quantitative model also allows us to examine the role of housing regulations, which it captures through changes in the productivity of the real estate sector. The cost of these regulations has been emphasized by many authors like Glaeser and Gyourko (2018), Herkenhoff et al. (2018) and Hsieh and Moretti (2019). Since the 80s, the relatively low real-estate productivity growth in CNR-intensive cities increased housing prices and led to more segregated CNR hubs. These changes brought the spatial distribution of occupations closer to the optimal allocation. Unfortunately, since they resulted from reductions in the measured relative productivity of real-estate in cognitive hubs, it also led to declines in welfare.

Relationship to the Literature A substantial literature has pointed to increasing spatial concentration of skilled workers (Berry and Glaeser (2005), Diamond (2016), and Giannone et al. (2017)), as well as increasing wage inequality across space and within cities (Baum-Snow and Pavan (2013), and Autor et al. (2019)), with the skill premium increasing the most in large cities. Our paper speaks to the optimal policy reaction to those trends.

We focus on production externalities as a key driving force behind those spatial patterns. The estimation of those externalities is a central theme in urban and spatial economics. A robust finding is the existence of a relationship between city size and productivity (see Melo et al. (2009) for a meta analysis). While we also allow for such agglomeration externalities, our main focus is on externalities tied to the occupational composition of the city. This is compatible with empirical evidence by Ellison et al. (2010) that industries with similar occupational make-up tend to be spatially proximate. Given the correlation between occupational types and skill levels, our finding of strong spill-overs stemming from the occupational composition of cities mirrors findings by Moretti (2004a; 2004b) regarding the local external effects of human capital.

There has been ample research on the extent of spatial misallocation in the U.S. economy and the degree to which it corresponds to heterogeneity in taxation policy (or its local incidence), zoning laws or other unspecified sources of distortions. Examples of papers in that vein are Albouy (2009), Desmet and Rossi-Hansberg (2013), Ossa (2015), Fajgelbaum et al. (2018), Colas et al. (2017), Hsieh and Moretti (2019) **a**nd, most recently Herkenhoff et al. (2018).

Our paper sheds light on place based policies, specifically to the extent that it focuses on stimulus to particular industries. A summary of the related literature can be found in Neumark and Simpson (2015). Rather than evaluating ad-hoc policies, we derive the optimal allocation in a quantitative spatial model with local externalities. Our derivation of optimal policy generalizes that of Fajgelbaum and Gaubert (2018)'s to an environment with multiple industries and input-output linkages. Two other recent papers that discuss the optimal distribution of city sizes are Eeckhout and Guner (2015) and Albouy et al. (2019).

We integrate industrial, occupational and spatial heterogeneity in a single coherent framework. Other recent work that has emphasized the joint distribution of industrial and skill composition within the US are Hendricks (2011) and Brinkman (2014). Like Caliendo et al. (2017), we allow for trade costs, which include an explicitly spatial dimension, but add to that framework by also allowing for occupational heterogeneity and local production externalities. Finally, on a more technical note, our paper adds to the rapidly expanding 'quantitative spatial economics' literature that uses general equilibrium models to address issues related to international, regional and urban economics. Redding and Rossi-Hansberg (2017) provide a review of the main ingredients in these models.

The rest of the paper is organized as follows. Section 2 presents a stylized facts that strongly suggest the presence of externalities among CNR workers within cities. Section 3 presents our multi-industry model with occupation specific externalities within cities. Section 4 presents the quantification of the model, including our estimation of the externality parameters, and discusses the role of externalities in the equilibrium allocation. Section 5 presents the optimal allocation and the resulting transfers and their implementation. Section 6 provides a decomposition of the impact of fundamentals changes in national CNR employment share and technology across sectors and cities between 1980 and the recent data. Section 7 concludes. We relegate many details of the model and additional robustness exercises and counterfactuals to the Appendix.

2 Stylized Facts

The main question we aim to answer is whether there is a role for policy in altering the observed spatial polarization of employment. We now provide some basic facts regarding the joint spatial distribution of wages and employment for workers of different occupations that point to the existence of important occupational externalities driving observed labor market polarization patterns. Those constitute prima-facie evidence that the optimal policy may in fact involve reinforcing existing patterns rather than attenuating those.

We separate workers from 2011 to 2015 in two large occupational groups: those that are intensive in cognitive non-routine (CNR) tasks and the others (non-CNR).[3] We calculate the average residual wages of workers in each occupation and each city after controlling for observable worker socio-economic characteristics.[4] This classification builds on the observation by Acemoglu and Autor (2011) that one can best understand wage inequality trends through such a task based approach.

Figure 1 shows that across U.S. cities, wages of workers employed in Cognitive Non-Routine (CNR) occupations, relative to those of workers in other (non-CNR) occupations, increase with the corresponding share of CNR workers in total employment. This suggests that differences in relative wages across cities are, to a large degree, driven by differences in relative demand for CNR workers.[5] The size of the scatterplot markers captures the city sizes. They indicate that large cities appear to also be CNR intensive.

Focusing on CNR workers, the top panel of Figure 2 indeed shows that real wages of CNR workers increase with the intensity of CNR employment across cities. Moreover, some of the high real wage cities include places like San Francisco and New York, that on average may provide higher amenities to CNR workers (see Diamond (2016)). In those cities, therefore, labor demand forces are seemingly pronounced enough to more than make up for the labor supply inducing effects of local amenities, such as the variety of retail and entertainment options. If workers differ in their preferences for where to live, the real wages depicted in the top panel of Figure 2 reflect the compensating differential to the marginal CNR worker in a given city.

Differences in the relative demand for CNR workers across cities can arise for several reasons. First, differences in relative demand for CNR workers may arise from exogenous (or historically determined) differences in industrial composition or regulations. Suppose that the industry make up of a city, n, is the main determinant of its demand for CNR workers relative to other types. Then its wage bill share for CNR workers would be (approximately) $\sum_j \delta^{CNR,j}\sigma_n^j$, where $\delta^{CNR,j}$ is the national wage bill share of CNR workers in industry j, and σ_n^j is the wage bill of industry j as a share of that city's total wage bill. Figure 3 compares the wage-bill shares of CNR workers implied by the different industrial composition of U.S. cities relative to those observed in the data. The black line is a 45 degree line. The observed wage bill shares differ from, and in fact increase more than one-for-one with, those implied by differences in industrial mix alone, thus ruling out industrial composition as a sole determinant of labor demand across cities.[6]

Differences in the relative demand for CNR workers across cities can also arise endogenously if more productive workers within occupational types sort themselves into particular cities. Baum-Snow and Pavan (2013) indeed argue that observable worker characteristics

[3]Specifically, we follow Jaimovich and Siu (2018), and define CNR occupations include occupations with SOC-2 classifications 11 to 29 and non-CNR occupations are those with SOC-2 classifications 35 to 55.

[4]We include as control variables education, potential experience, race, gender, English proficiency, number of years in the U.S., marital status, having had a child in the last year, citizenship status, and veteran status.

[5]In particular, suppose that technologies were similar across cities, and that the share of CNR workers were driven by the supply of those workers. Then, with decreasing marginal returns to worker type, increases in the relative supply of CNR workers would lower their relative wages.

[6]The figure also rules out the production technology for different industries being well characterized by Cobb-Douglas (i.e. the elasticity of substitution across worker types is likely not equal to 1).

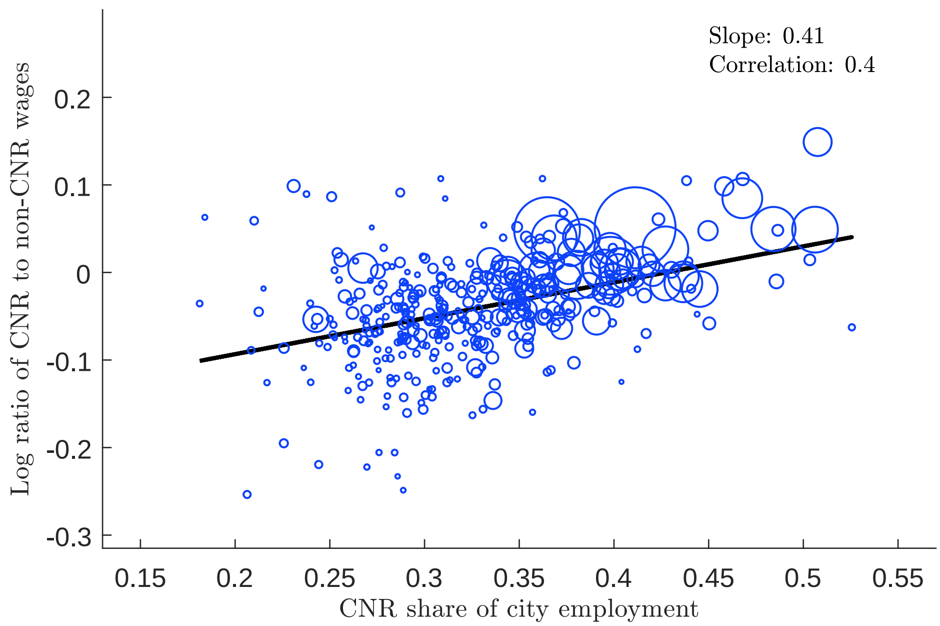


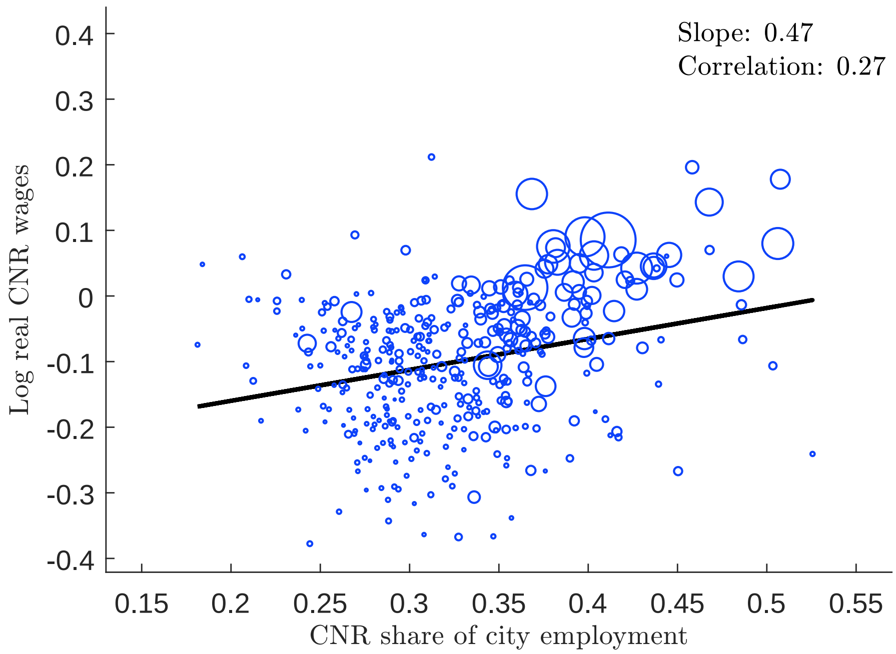
Figure 1: Occupational employment share and wage premium

See text for details on definitions of the wage premium and occupation classification. Log wage premium is depicted as a deviation from employment weighted mean. Each observation refers to a CBSA. Marker sizes are proportional to total employment.

are an important determinant of the city size wage premium. However, the fact that relative wages in Figure 1 are computed from residuals after controlling for observable worker characteristics suggests that sorting along these characteristics is not the only driving force underlying that figure. Thus, if sorting is nevertheless part of an explanation driving the positive relationship between the wage premia of CNR workers and the employment share of those workers, it must be taking place along dimensions that are not easily observed. However, assuming that differences in amenities are experienced in similar ways by CNR and non-CNR workers, high productivity non-CNR workers would then also sort themselves into cities with a high share of CNR workers. The bottom panel of Figure 2 suggests that this is not, in fact, the case.[7]

Finally, differences in relative demand for CNR workers can be rationalized by endogenous differences in productivity, even when not from sorting, if these differences arise from production externalities that predominantly affect CNR workers. First, to the extent that production externalities also increase with the concentration of CNR workers in a given city, it is then naturally the case that the demand for CNR workers would increase with the share of employment in CNR occupations, as suggested by Figure 1. Second, if production externalities mainly enhance the productivity of CNR workers, then real wages of CNR workers would increase with the share of CNR employment within cities, as in the

[7]The small relevance of sorting to explain differences in wages across cities has in fact been recently verified in empirical work by Baum-Snow and Pavan (2011) and Roca and Puga (2017).



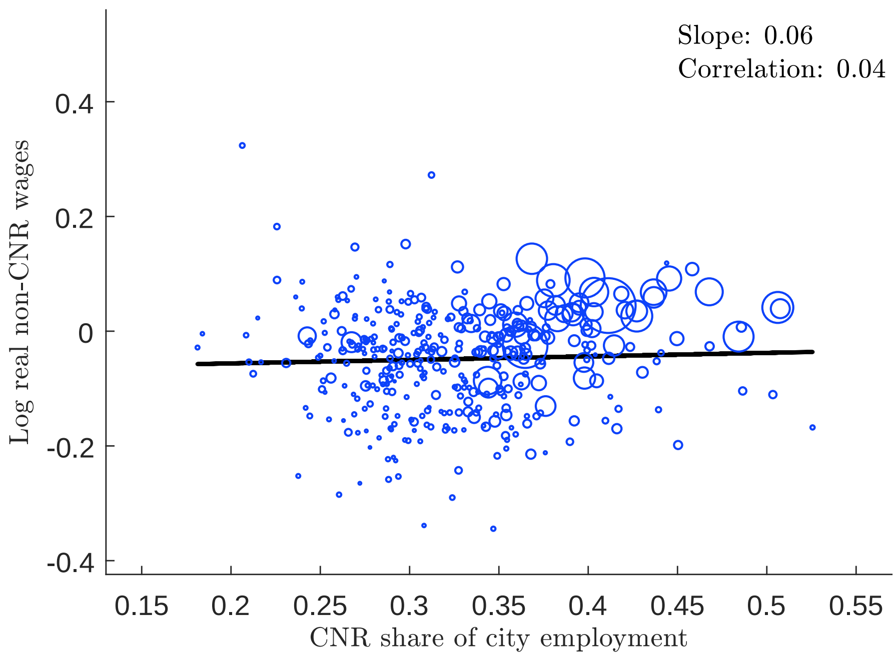


Figure 2: Occupational employment share and real wages

See text for details on definitions of the wage premium and occupation classification. Real wages are calculated using consumption price indices obtained from the model quantification (See Section 4). Log real wages are depicted as a deviation from employment weighted mean. Each observation refers to a CBSA. Marker sizes are proportional to total employment.

top panel of Figure 2, but no such effect would be expected among non-CNR workers, as suggested by the bottom panel of Figure 2. Third, and most importantly, Figure 3 shows that observed wage bill shares of CNR workers increase more than one-for-one with those implied by differences in industrial composition alone. This observation would be expected in an environment where production externalities intensify the implications of industrial mix. Specifically, CNR workers tend to concentrate in cities whose industrial composition is tilted towards industries intensive in CNR workers. In the presence of production externalities, therefore, this concentration would lead to increases in the productivity of CNR workers. If the elasticity of substitution between worker types is higher than 1, one would then expect to observe higher wage shares for CNR workers in those cities relative to those given by industrial composition alone.

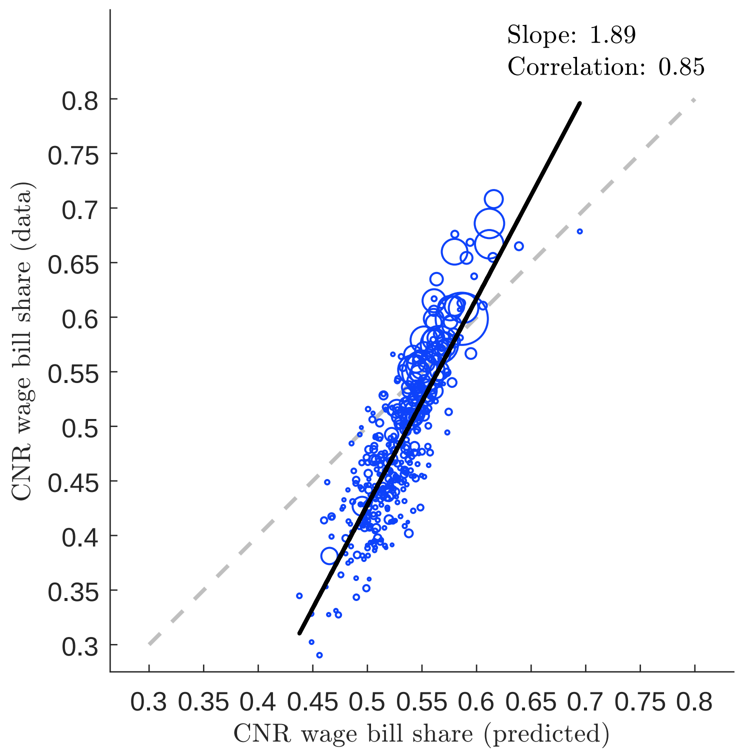


Figure 3: Occupational wage bill share: predicted vs actual
See text for details on definitions of the wages and occupation classification. The predicted wage bill-shares are obtained by assuming that within-industry wage bill shares were equal to national averages (see text for details). Each observation refers to a CBSA. Marker sizes are proportional to total employment.

3 Model

The economy has N cities and J sectors. We denote a particular city by $n \in \{1, ..., N\}$ (or n') and a particular sector by $j \in \{1, ..., J\}$ (or j'). Individuals are endowed with an occupational type and cannot switch types. There are K occupational types, denoted by

$k \in \{1, ..., K\}$ (or k'), with aggregate number of workers L^k per type (total employment in occupation k aggregated across industries and cities). Firms in all cities use multiple types of labor but in potentially different proportions depending on the industry and the city. Aggregate regional land and structures in region n are denoted by H_n. Labor of all types moves freely across regions and sectors, while structures are region-specific. Sectors can be tradable or non-tradable.

In general, various quantities in the economy can be associated with industries, cities, or occupations. For notational convenience, we denote aggregates across a given dimension by omitting the corresponding index. Thus, for example, L_n^{kj} is the number of workers employed in occupation k in industry j in city n, $L_n^k = \sum_j L_n^{kj}$ represents the number of workers employed in occupation k in city n, $L^k = \sum_n L_n^k$ represents all workers in occupation k, and $L = \sum_k L^k$ is simply total employment in the economy.

3.1 Individuals

Workers in each location $n \in \{1, ..., N\}$, are endowed with labor of type $k \in \{1, ..., K\}$, and order consumption baskets according to Cobb-Douglas preferences with shares α^j over their consumption of final domestic goods, C_n^{kj}:

$$C_n^k = \prod_j \left(C_n^{kj}\right)^{\alpha^j},$$

where C_n^k is a consumption aggregator. Consumption goods consumed in city n are purchased at prices P_n^j in sectors $j \in \{1, ..., J\}$. Utility is homogeneous of degree one, so that $\sum_j \alpha^j = 1$.

Workers supply one unit of labor inelastically. The income of a worker of type k residing in city n is

$$I_n^k = w_n^k + \chi^k, \tag{1}$$

where w_n^k is the wage earned by a worker in occupation k in city n. The term χ^k represents the return per household on a national portfolio of land and structures from all cities,

$$\chi^k = b^k \frac{\sum_{n'} r_{n'} H_{n'}}{L^k},$$

where r_n is the rental rate on land and structures in that city and b^k denotes the share of the national portfolio appropriated by CNR workers. In what follows, we assume that CNR workers appropriate a share of the national portfolio proportional to their share of wages in total wage bill.

Agents of a given occupational type differ in how much they value living in different cities. Those differences are summarized by a vector $\mathbf{a} = \{a_1, a_2, ..., a_N\}$, with each entry a_n scaling the utility value that an individual receives from living in city n. We associate the elements a_n with the particular way in which different workers experience the amenities of given cities. Conditional on living in city n, the problem of an agent employed in occupation k and characterized by amenity vector $\mathbf{a}$ is

$$v_n^k(\mathbf{a}) \equiv \max_{\left\{C_n^{kj}(\mathbf{a})\right\}_{j=1}^{J}} a_n A_n^k \prod_j \left(C_n^{kj}(\mathbf{a})\right)^{\alpha^j}, \text{ subject to } \sum_j P_n^j C_n^{kj}(\mathbf{a}) = I_n^k,$$

where A_n^k denotes an exogenous component of city-specific utility common to all individuals of type k living in city n. All workers in a given occupation living in a given city will choose the same consumption basket. It follows that $C_n^{kj}(\mathbf{a}) = C_n^{kj}$ for all $\mathbf{a}$.

Agents move freely across cities. The value, $v_n^k(\mathbf{a})$, of locating in a particular city n for an individual employed in occupation k, with idiosyncratic preference vector $\mathbf{a}$,

$$v_n^k(\mathbf{a}) = \frac{a_n A_n^k I_n^k}{P_n} = a_n A_n^k C_n^k.$$

In equilibrium, workers move to the location where they receive the highest utility so that

$$v^k(\mathbf{a}) = \max_n v_n^k(\mathbf{a}),$$

where $v^k(\mathbf{a})$ now denotes the equilibrium utility of an individual in occupation k with amenity vector $\mathbf{a}$. We assume that a_n is drawn from a Frechet distribution, independently across cities. We denote by Ψ^k the joint *cdf* for the elements of $\mathbf{a}$ across workers in occupation k, so that

$$\Psi^k(\mathbf{a}) = \exp\left\{-\sum_n (a_n)^{-\nu}\right\},$$

where the shape parameter ν reflects the extent of preference heterogeneity across workers employed in occupation k. Higher values of ν imply less heterogeneity, with all workers ordering cities in the same way when $\nu \to \infty$. The assumption of a Frechet distribution for idiosyncratic amenity parameters implies closed form expressions for the fraction of workers in each city:

$$L_n^k = \Pr\left(v_n^k(\mathbf{a}) > \max_{n' \neq n} v_{n'}^k(\mathbf{a})\right) = \frac{\left(A_n^k C_n^k\right)^\nu}{\sum_{n'}\left(A_{n'}^k C_{n'}^k\right)^\nu} L^k. \tag{2}$$

3.2 Firms

There are two types of firms: those producing intermediate goods and those producing final goods. There is a continuum of varieties of intermediate goods which are aggregated into a finite number of final goods corresponding to J sectors. Varieties of intermediate goods are characterized by the sector in which they are produced, and by a vector of city-specific productivity parameters, $\mathbf{z} = \{z_1, z_2, ..., z_N\}$, with each element z_n scaling the productivity of firms in city n producing that variety.

Final goods are sold in the city where they are produced. Varieties of intermediate goods are traded across cities. Because of transportation costs, the price earned by intermediate goods producers need not be the same as the price paid by final goods producers. Intermediate goods producers operating in city n, sector j, producing a variety indexed by $\mathbf{z}$, produce a quantity, $q_n^j(\mathbf{z})$, for which they earn a price $p_n^j(\mathbf{z})$. Final goods producers operating in city n, sector j, purchase a quantity $Q_n^j(\mathbf{z})$ of the variety of intermediate goods indexed by $\mathbf{z}$.

3.2.1 Intermediate Goods

Idiosyncratic productivity draws, $\mathbf{z}$, arise from a Frechet distribution with shape parameter θ. Draws are independent across goods, sectors, and regions. Specifically, if we let Φ^j be the joint *cdf* of variety-specific productivity parameters across firms in industry j, then

$$\Phi^j(\mathbf{z}) = \exp\left\{-\sum_n (z_n)^{-\theta}\right\}.$$

Production of intermediate goods a variety indexed by $\mathbf{z}$, in city n, and industry j, takes place using the technology,

$$q_n^j(\mathbf{z}) = z_n \left[H_n^j(\mathbf{z})^{\beta_n^j}\left[\sum_k \left(\lambda_n^{kj} L_n^{kj}(\mathbf{z})\right)^{\frac{\epsilon-1}{\epsilon}}\right]^{\frac{\epsilon}{\epsilon-1}(1-\beta_n^j)}\right]^{\gamma_n^j} \prod_{j'} M_n^{j'j}(\mathbf{z})^{\gamma_n^{j'j}} \tag{3}$$

where $\gamma_n^{j'j} \geq 0$ is the share of sector j input expenditures spent on materials from sector j' in city n, $\gamma_n^j \geq 0$ is the share of value added in gross output in sector j, and β_n^j is the share of land and structures in value added in that sector. The production function has constant returns to scale, $\sum_{j'=1}^{J} \gamma_n^{j'j} = 1 - \gamma_n^j$. Also, λ_n^{kj} denotes a labor augmenting component which is city, industry, and occupation specific. We denote by $H_n^j(\mathbf{z})$ the quantity of structures used by a firm producing a variety $\mathbf{z}$ in industry j operating in city n, by $M_n^{j'j}(\mathbf{z})$ the quantity of material goods this firm uses from sector j', and by $L_n^{kj}(\mathbf{z})$ the workers of type k it employs.

Finally, we allow λ_n^{kj} to reflect externalities stemming from the composition of the labor force, that is,

$$\lambda_n^{kj} = \lambda_n^{kj}(\mathbf{L_n}),$$

where $\mathbf{L}_n = \{L_n^1, ..., L_n^K\}$ summarizes the occupational make up of the labor force in city n.

3.2.2 Final Goods

A final goods firm operating in industry j in city n produces the quantity Q_n^j according to the technology,

$$Q_n^j = \left[\int \left[\sum_{n'} Q_{nn'}^j(\mathbf{z})\right]^{\frac{\eta-1}{\eta}} d\Phi^j(\mathbf{z})\right]^{\frac{\eta}{\eta-1}},$$

where $Q_{nn'}^j(\mathbf{z})$ represents its use of intermediate goods of variety $\mathbf{z}$ produced in city n'.

One unit of any intermediate good in sector j shipped from region n' to region n requires producing $\kappa_{nn'}^j \geq 1$ units in the origin n'. Therefore, producers of final goods in each sector solve

$$\max_{Q_{nn'}^j(\mathbf{z})} P_n^j Q_n^j - \sum_{n'} \int \kappa_{nn'}^j p_{n'}^j(\mathbf{z}) Q_{nn'}^j(\mathbf{z})\, d\Phi^j(\mathbf{z}),$$

subject to $Q^j_{nn'}(\mathbf{z}) \geq 0$, where P^j_n is the price of the final good in sector j, city n. Intermediate goods in non-tradable sectors cannot be shipped between cities.

Final goods firms purchase intermediate goods from the location for which the acquisition cost, including transportation costs, is the least. Denote by X^j_n the total expenditures on final goods j by city n, which must equal of the value of final goods in that sector, $X^j_n = P^j_n Q^j_n$. Because of zero profits in the final goods sector, total expenditures on intermediate goods in a given sector are then also equal to the cost of inputs used in that sector. Following usual derivation in Eaton and Kortum (2002), one can show that, for a given a final good j produced in city n, the share of intermediate inputs imported from location n' is

$$\pi^j_{nn'} \quad = \quad \frac{\left[\kappa^j_{nn'} x^j_{n'}\right]^{-\theta}}{\sum_{n''=1}^{N} \left[\kappa^j_{nn''} x^j_{n''}\right]^{-\theta}}$$

where

$$x^j_n = \left\{ \left(\frac{r_n^{\beta^j_n}}{\beta^j_n} \right) \left[\frac{1}{1-\beta^j_n} \sum_k \left(\frac{w^k_n}{\lambda^{kj}_n} \right)^{1-\epsilon} \right]^{\frac{1-\beta^j_n}{1-\epsilon}} \right\}^{\gamma^j_n} \prod_{j'=1}^{J} \left(\frac{P^{j'}_n}{\gamma^{j'j}_n} \right)^{\gamma^{j'j}} \tag{4}$$

is a cost index for production of varieties in sector j, city n. In our quantification of the model we also allow for two non-tradable sectors, for which by definition $\pi^j_{nn} = 1$.

3.3 Market Clearing Conditions

Within each city n, the number of workers employed in occupation k must equal the number of those workers who choose to live in that city. Put alternatively,

$$\sum_j \int L^{kj}_n(\mathbf{z}) d\Phi^j(\mathbf{z}) = \int \zeta^k_n(\mathbf{a}) d\Psi^k(\mathbf{a}), \; \forall \; n = 1, ..., N, \; k = 1, ..., K. \tag{5}$$

where $\zeta^k_n(\mathbf{a}) \in \{0,1\}$ denotes the location choice of households as a function of their type. Market clearing for land and structures in each region imply that

$$\sum_j \int H^j_n\left(\mathbf{z}\right) d\Phi^j\left(\mathbf{z}\right) = H_n, \; n = 1, ..., N. \tag{6}$$

Final goods market clearing implies that

$$\sum_k L^k_n C^{kj}_n + \sum_{j'} \int M^{jj'}_n\left(\mathbf{z}\right) d\Phi^{j'}\left(\mathbf{z}\right) = Q^j_n. \tag{7}$$

Finally, intermediate goods market clearing requires that

$$q^j_n(\mathbf{z}) = \sum_{n'} \kappa^j_{n'n} Q^j_{n'n}(\mathbf{z}) \tag{8}$$

3.4 The Planner's Problem

This section describes the solution to the planner's problem taking as given that workers in different occupations can freely choose in which city to live. Under this assumption, the expected utility of a worker of type k is given by

$$v^k = \Gamma\left(\frac{\nu-1}{\nu}\right)\left(\sum_n \left(A_n^k C_n^k\right)^\nu\right)^{\frac{1}{\nu}},$$

Given welfare weights for each occupation ϕ^k, the utilitarian planner then solves

$$\mathcal{W} = \sum_k \phi^k \Gamma\left(\frac{\nu-1}{\nu}\right)\left(\sum_{n=1}^N \left(A_n^k C_n^k\right)^\nu\right)^{\frac{1}{\nu}} L^k, \tag{9}$$

The planner maximizes (9) subject to the resource constraints for final goods in each city and sector (7), the resource constraints for intermediate goods of all varieties $\mathbf{z}$ and industries j produced in all cities n (8), the resource constraints for local labor markets (5), the resource constraints for the use of land and structures (6), non-negativity constraints on both household consumption of different goods and input flows and the labor supply conditions (2).

The main difference between the optimal allocation and the equilibrium one is that the planner takes into account the wedge between the private and the social marginal product of labor. Lemma 1 below characterizes that wedge:

Lemma 1. *Let Δ_n^k denote the wedge between the private and the social marginal value of a worker of occupation k in city n. Then*

$$\Delta_n^k = \sum_{k'j} w_n^{k'} \frac{L_n^{k'j}}{L_n^k} \frac{\partial \ln \lambda_n^{k'j}\left(\mathbf{L}_n^k\right)}{\partial \ln L_n^k} \tag{10}$$

The expression for the wedge provides some clues as to the distortions that the planner will seek to correct. To focus ideas, consider the leading case in which there are two occupations and spill-over elasticities are the same in all sector. In that special case we have

$$\Delta_n^k = w_n^k \frac{\partial \ln \lambda_n^{kj}\left(\mathbf{L}_n^k\right)}{\partial \ln L_n^k} + w_n^{k'} \frac{L_n^{k'}}{L_n^k} \frac{\partial \ln \lambda_n^{k'j}\left(\mathbf{L}_n^k\right)}{\partial \ln L_n^k}$$

For given wages, the wedge associated with workers from some occupation k increases with the proportion of workers from the other occupation k' if $\frac{\partial \ln \lambda_n^{k'j}\left(\mathbf{L}_n^k\right)}{\partial \ln L_n^k} > 0$ and will decrease otherwise. Those elasticities are therefore going to be key to determine whether the planner will seek to attenuate or reinforce spatial polarization trends.

Given those wedges, the optimal policy is then most intuitively characterized by the set of taxes and subsidies that allow for it to be implemented.

Proposition 1 provides the details:

Proposition 1. *If the planner's problem is globally concave, the optimal allocation can be achieved by a set of transfers such that*

$$P_n C_n^k = (1 - t_L^k)(w_n^k + \Delta_n^k) + \chi^k + R^k$$

where

$$t_L^k = \frac{1}{1+\nu}$$

and R^k is such that

$$\phi^k v^k L^k = \sum_n P_n C_n^k L_n^k$$

The proposition generalizes proposition 2 in Fajgelbaum and Gaubert (2018) to a multi-industry environment. The industrial composition matters to the extent that spill-over elasticities are industry-specific. Even then, the employment shares of different industries in different occupations are a sufficient statistic for the role of the industrial composition, independently of the details of the input-output linkages. If those elasticities are common for all industries, then, given wages, the optimal policy does not depend on the industry composition at all and is exactly equal to the one derived by Fajgelbaum and Gaubert (2018).

The planner's solution for household consumption differs from that implied by their budget constraint in two main ways. First, the planner's solution depends on the social marginal product of labor, given by $w_n^k + \Delta_n^k$, rather than its private counterpart. Second, in the planner's solution, consumption increases less than one-for-one with the (social) marginal product of labor. This is optimal because, given heterogeneity in preferences for locations, households that choose to live in lower wage cities do so because they enjoy higher marginal utility of consumption in those cities.[8]

4 Quantifying the Model

We quantify the model with 22 industries and the two large occupational groups (cognitive non-routine and others) emphasized in Section 2. Of the 22 industries, two are non-tradable, meaning that all local output is also used locally. Tradable industries include 10 manufacturing sectors and 10 service sectors. Of the two non-tradable sectors, one is identified with the real-estate services, and is the single user of land in each city, and the other is a residual non-tradable sector including retail, construction and utilities. For our quantification, we focus on the period between 2011 to 2015.

[8]Fajgelbaum and Gaubert (2018) show that the heterogeneity in preferences induces the same optimal tax as an isoelastic negative spillover in amenities.

The set of parameters needed to quantify our framework fall into broadly two types: i) parameters that are constant across cities (but may vary across occupations and/or industries) and ii) parameters that vary at a more granular level and require using all of the model's equations (i.e. by way of model inversion) to match data that vary across cities, industries, and occupations.

To obtain an initial calibration for the share parameters γ_n^j, $\gamma_n^{jj'}$ and α^j, we use an average of the 2011 to 2015 BEA Use Tables, each adjusted by the same year's total gross output, where we assume that tradable sectors have a $\gamma_n^j = \gamma^j$, constant across cities and similarly for $\gamma_n^{jj'}$'s.[9]

As mentioned above, we adopt the convention that all land and structures are managed by firms in the real estate sector, which then sell their services to other sectors. Accordingly, for all sectors other than real estate, we reassign the gross operating surplus remaining, after deducting equipment investment, to purchases from the real estate Sector. These surpluses are in turn added to the gross operating surplus of real estate.[10] This convention implies that the share of land and structures β_n^j in the production of all sectors other than real estate is equal to zero.

We choose θ, the Frechet parameter governing trade elasticities, to 10. In general, $\theta = 10$ is well within the range of estimates for trade elasticity in the literature, between 3 to 17 (see Footnote 44 in Caliendo and Parro (2015), as well as Head and Mayer (2014), section 4.2 for comprehensive summaries of estimates). While estimates of θ have been carried out at various levels of disaggregation, these can vary somewhat widely for a given sector or commodity across studies.[11] For our purposes, this uncertainty is further compounded by the fact that trade elasticities that are relevant for trade between countries may not be appropriate for trade between regions or cities.

In order to calibrate trade costs, we assume that two of the sectors ("real estate," as well "retail, construction, and utilities") are non-tradable, so that their transportation costs are treated as infinite. For the tradable sectors, we follow Anderson et al. (2014) and assume that trade costs increase with distance, $\kappa_{nn'}^j = (d_{nn'})^{t^j}$, where $d_{nn'}$ is the distance between city n and city n' in miles. The parameter t^j is industry specific.[12] For commodities, we directly estimate t^j from the Commodity Flow Survey synthetic microdata using standard gravity regressions based on trade-shares, $\pi_{n'n}^j$, in the model. In the tradable services, we use the values obtained by Anderson et al. (2014) using Canadian data. Ciccone and Peri (2005) summarize estimates for the elasticity of substitution between skilled and unskilled labor in the literature as ranging between 1.36 and 2. Card (2001) estimates elasticity of substitution between occupations to be closer to 10. We adopt $\epsilon = 2$ as a benchmark. Finally, we set ν so that the average elasticity of employment with respect to real wages in our model matches

[9] Since the model does not allow for foreign trade, we adjust the Use Table by deducting purchases from international producers from the input purchases and, for accounting consistency, from the definition of gross output for the sector.

[10] One can verify that those reassignments do not affect aggregate operational surplus (net of equipment investment), aggregate labor compensation, and aggregate value added (net of equipment investment).

[11] For example, while Caliendo and Parro (2015) estimate an elasticity of 7.99 for Basic Metals and 4.75 for Chemicals, Feenstra et al. (2018) estimate a elasticities of, respectively, 1.16 and 1.46 for those two categories.

[12] We assume that within city distance is equal to 20 miles.

the estimate of 1.36 in Table A.11, column 4 of Fajgelbaum et al. (2018). This implies $\nu = 2.02$.[13]

Given the parameters above, we use data on wages by occupation and location (w_n^k), and on employment by occupation, industry and location (L_n^{kj}) to back out the equilibrium values of productivity and amenity shifters, λ_n^{kj} and A_n^k. Data is available from the ACS pertaining to w_n^K, and $\frac{L_n^{kj}}{\sum_{k'} L_n^{k'j}}$. The ACS also allows us to adjust wages for individual characteristics, so that our data captures city wage premia for each occupation. The Census County Business Patterns (CBP) provide us measures of total employment $\sum_{k'} L_n^{k'j}$ that better match BEA industry-level counts. We combine total employment from the CBP with ACS data on employment shares to obtain L_n^{kj}.

As we will see, the existence of non-tradable sectors necessitate the introduction of additional degrees of freedom in the model. In particular, we allow the share of intermediate input in the production of residual non-tradable sectors ($\gamma_n^{\text{retail, construction and utilities}}$) and the share of land and structures in the production of real estate ($\beta_n^{\text{real estate}}$) to be city-specific. These additional degrees of freedom then require the introduction of additional data constraints. From the BEA, we obtain regional price indices P_n which include imputed prices for housing services, $P_n^{\text{real estate}}$. An important caveat in relying on RPP prices is that, as carefully demonstrated by Handbury and Weinstein (2014), they may not account properly for product buyer and retailer heterogeneity or the number of varieties of goods available in different cities, thus providing biased estimates of the relationship between city size and prices. Thus, rather than using the full RPP price levels, we use instead the price of services (other than real estate) relative to tradable goods. Assuming that the biases identified by Handbury and Weinstein (2014) are similar enough in goods and services, this approach mitigates the effects of these biases on our measure of non-tradable prices. Furthermore, those biases are somewhat less of a concern in the real estate sector since the BEA calculates RPP prices in that sector using considerably more disaggregated data at the local level than is otherwise available for other goods. Thus, we also use $P_n^{\text{real estate}}$ from the RPP estimates of the price of real estate services in different cities.

The main step of the quantification uses the production structure of the model to obtain local production costs. Specifically, we use wage bill data to obtain local gross output by sector:

$$\sum_{n'} \pi_{n'n}^j X_{n'}^j = \frac{\sum_k w_n^k L_n^{kj}}{\left(1 - \beta_n^j\right) \gamma_n^j}. \tag{11}$$

From the household and firm optimal demands for final goods, we can obtain local expenditures as a function of local income:

[13] Our value is larger than the value for the corresponding parameter that Fajgelbaum et al. (2018) obtain from the same estimates. This reflects the facts that it is the elasticity of labor supply with respect to consumption rather than wages. In comparison, they do not allow for non-wage income of workers, estimate values for the analogous parameter in their model between 0.75 and 2.25 depending on identification assumptions.

$$X_n^j = \alpha^j \left[\sum_k w_n^k L_n^k + \chi^k L_n \right] + \sum_{j'} \frac{\gamma_n^{jj'}}{\gamma_n^{j'} (1 - \beta_n^{j'})} \sum_k w_n^k L_n^{kj'}, \tag{12}$$

where

$$\chi^k \equiv b^k \frac{1}{L^k} \sum_{n'} \left(\frac{\beta_n^{\text{Real Estate}}}{1 - \beta_n^{\text{Real Estate}}} \sum_k w_{n'}^k L_{n'}^{k,\text{Real Estate}} \right)$$

For non-tradable sectors $\pi_{nn}^j = 1$ and $\pi_{n'n}^j = 0$ so that the equations (11) and (12) above provide restrictions that we use to pin down city specific share parameters $\beta_n^{\text{real estate}}$ and $\gamma_n^{\text{retail, construction and utilities}}$ after assuming that the proportions of materials used in each city by nontradable sectors $\frac{\gamma_n^{j'j}}{1-\gamma_n^j}$ is the same for all cities.

Given the share parameters, wage and employment data, we can pin down expenditures X_n^j. Recall that the trade shares $\pi_{n'n}^j$ are functions of local cost indices x_n^j defined in equation (4). For tradable sectors we can then solve for x_n^j from the system of equations:

$$\frac{\sum_k w_n^k L_n^{kj}}{\gamma_n^j} = \sum_{n'=1}^{N} \pi_{n'n}^j \left(\mathbf{x}^j \right) X_{n'}^j.$$

We can further pin-down the values of x_n^j for the non-tradable sectors by matching BEA regional price data. Using the obtained values for x_n^j, share parameters and wages we can obtain sector specific prices P_n^j and equilibrium productivity shifters λ_n^{kj}. Note that those do not depend in any way on determinants of labor supply, since they are calculated for given wages and employment.

Finally, given wages and rental incomes, we can calculate consumption in each city. We then use the labor supply equation (2) to back out the values of city and occupation-specific amenity shifters compatible with observed equilibrium.

4.1 Model Quantification and Previous Literature

We now outline the how data and results in various steps of our model inversion compare to previous findings where available.

4.1.1 Wages, City Size, and City Composition

Table 1 compares the relationships between wages, employment, and employment composition across different cities highlighted in previous work relative to the data used in our model inversion. The first three rows of the table show regression coefficients of log wages for CNR workers, non-CNR workers, and the CNR wage premium, on different measures of city employment and employment composition. The subsequent rows show similar regression coefficients obtained in previous literature. The data we use implies relationships that are consistent with those in other work. In particular, all wages increase with city size, more so for skilled workers. A similar relationship holds for wages and city composition, where proportionally more skilled cities exhibit higher wages for all workers, more so for skilled workers.[14]

Table 1: Wages, Employment, and City Composition

Dependent Variable	$\ln(L_n)$	$\ln\left(\frac{L_n^{\text{CNR}}}{L_n^{\text{nCNR}}}\right)$	$\frac{L_n^{\text{CNR}}}{L_n}$
$\ln\left(w_n^{\text{CNR}}\right)$	0.070 (0.003)	0.332 (0.021)	1.474 (0.089)
$\ln\left(w_n^{\text{nCNR}}\right)$	0.049 (0.002)	0.218 (0.017)	0.973 (0.072)
$\ln\left(\frac{w_n^{\text{CNR}}}{w_n^{\text{nCNR}}}\right)$	0.021 (0.001)	0.114 (0.008)	0.501 (0.036)
Moretti HS[1]	—	—	0.85 (0.06)
Moretti Some College	—	—	0.86 (0.06)
Moretti College +	—	—	0.74 (0.06)
Roca & Puga wage log wage constant[2]	0.0455 (0.0080)	—	—
Diamond log college wage[3]	—	0.26 (0.11)	—
Diamond log non-college wage[4]	—	0.18 (0.01)	—
Baum-Snow et al. log wage, 2005-2007[5]	0.065 (< 0.01)	—	—
Baum-Snow et al. log wage ratio[6]	0.029 (< 0.003)	—	—

1. Moretti (2004a) "Estimate the social return to higher education: evidence from longitudinal and repeated cross-sectional data", Table 5.
2. Roca and Puga (2017) "Learning by Working in Big Cities", Table 1.
3. Diamond (2016) "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000", Figure 4.
4. Diamond (2016), Figure 3
5. Baum-Snow et al. (2018) "Why Has Urban Inequality Increased?", Table 1. Standard error reported as less than 0.01.
6. Baum-Snow et al. (2018), Table 2. Standard error reported as less than 0.003.

4.1.2 Tradable Goods Prices and City Size

Recent work by Handbury and Weinstein (2014), using Nielsen home-scanned data on tradable goods bought in grocery stores, highlights that tradable consumer prices decrease with city size. Prior to that study, the consensus view, based on more aggregated prices, was

[14] An exception is Moretti (2004a) who finds no statistically significant differences in the way that wages of college educated workers and non-college educated workers vary with employment composition across cities. Our findings, however, rely on a more recent time period where other work has found an increasingly pronounced relationship between skill and city size (see Baum-Snow and Pavan (2013)).

that such prices instead increased with city size.[15] Given the detailed nature of Nielsen home-scanned prices, Handbury and Weinstein (2014) are able to control for product buyer and retailer heterogeneity in a way that is not easily achieved with more aggregate prices. Allowing for those controls reduces the elasticity of tradable goods bought in grocery stores with respect to city size to zero. When Handbury and Weinstein (2014) further adjust local price indices to reflect differences in the number of varieties of goods available in different cities, they find that the price of tradable goods bought in grocery stores actually decreases with city size, with an elasticity equal to -0.011 (Table 6). In addition, when calculating this elasticity after purging the effect of local rents on retail costs, they obtain -0.017 (Table 9 in Working Paper Version).

Table 2 below summarizes the relationship between prices, obtained in our model inversion, and city size. Similar to Handbury and Weinstein (2014), our model inversion reveals a decreasing relationship between prices and city size and, in fact, across all tradable sectors with an average elasticity of -0.019. In the Food and Beverage sector, our model inversion reveals an elasticity of -0.10, virtually identical to Handbury and Weinstein (2014) for grocery products. Remarkably, our finding arises without direct observation of prices, but by the way of a supply and demand relationships within a structural trade model where cities produce different goods and where trade across space is costly.

4.1.3 Total Factor Productivity

A wide literature in urban economics has addressed the relationship between productivity and city size (i.e. "agglomeration economies"), as well as that between productivity and employment composition. Baseline estimates of real Total Factor Productivity (TFP) typically rely on Cobb-Douglas production functions that allow for different types of labor to enter separately. Within the context of our model, we follow Caliendo et al. (2017) and express measured TFP as[16]

$$\ln TFP_n^j = \ln\left(\sum_{n'} \pi_{n'n} X_{n'}^j\right) - \ln P_n^j - \gamma^j \beta^j \ln H_n^j - \gamma^j (1-\beta^j) \sum_k \delta^{kj} \ln L_n^{kj} - \sum_{j'} \gamma^{j'j} \ln M_n^{j'j}.$$

where δ^{kj} is the share occupation k wages in sector j wage bill. Up to a first-order approximation (and abstracting from selection effects induced by trade), we have that for tradable sectors[17]

$$\ln TFP_n^j \simeq \sum_k \delta^{kj} \gamma^j \ln \lambda_n^{kj}.$$

[15] In principle, given that rents generally increase with city-size, tradable consumer prices might indeed follow the same pattern to the degree that they are partially influenced by local rents as an input cost.

[16] we now omit the city index n from the share parameters since those are assumed to be constant across cities for all tradable sectors.

[17] For non-tradable sectors the city varying share parameters make it hard to do a direct comparison across-cities. Furthermore, our data does not allow us to separate the productivity of the real estate sector from the stock of housing.

where we can interpret $\left(\lambda_n^{kj}\right)^{\gamma^j}$ as the component of TFP in sector j and city n associated with occupation k.[18]

Table 2: Elasticities of Final Goods Prices, P_n^j, w.r.t. L_n

Sector	Elasticity
Food and Beverage	**−0.010**
Textiles	−0.022
Wood, Paper, and Printing	−0.011
Oil, Chemicals, and Nonmetallic Minerals	−0.016
Metals	−0.014
Machinery	−0.005
Computer and Electric	−0.013
Electrical Equipment	−0.003
Motor Vehicles (Air, Cars, and Rail)	−0.008
Furniture and Fixtures	−0.009
Miscellaneous Manufacturing	−0.013
Wholesale Trade	−0.018
Transportation and Storage	−0.015
Professional and Business Services	−0.040
Other	−0.025
Communication	−0.005
Finance and Insurance	−0.025
Education	−0.049
Health	−0.050
Accommodation	−0.023
Real Estate	0.132
Retail, Construction and Utilities	0.090
Average	−0.007
Tradable Average	−0.019
Manufacturing Average	−0.011
Tradable Services Average	−0.028

The model inversion produces estimates for productivity and amenities in different cities. Table 3 below shows the city that has the highest TFP for each industry. The results largely conform to intuition. San Jose, CA has the highest productivity in the production of Computers and Electric equipment; Honolulu stands out for Accommodation and for Transportation and Storage; Anchorage for Oil, Chemicals and Nonmetallic Minerals; and Seattle for Motor Vehicles (which includes aircrafts). It is also interesting to note that the two largest cities in the country are also top cities for several industries, with New York dominating in most service sectors, and Los Angeles standing out in several manufacturing sectors.

Table 4 below compares estimates of productivity elasticities with respect to city size and employment composition obtained from our model inversion to those found in previous work. In particular, we report elasticities with respect to city size from the meta analysis carried out in Melo et al. (2009). As reported in Table 4, all results point to a positive relationship between TFP and city size. Moreover, our findings fall within the range of reduced form

[18] In Caliendo et al. (2018), abstracting from the selection of effect induced by trade, $\ln TFP_n^j = \ln \gamma_n^j \ln \lambda_n^j$. The first-order approximation and interpretation would be exact in the Cobb-Douglas case.

Table 3: City with Top TFP for each Industry

Industry	MSA
Food and Beverage	Los Angeles, CA
Textiles	Los Angeles, CA
Wood, Paper, and Printing	Minneapolis, MN
Oil, Chemicals, and Nonmetallic Minerals	Anchorage, AK
Metals	Chicago, IL
Machinery	Houston, TX
Computer and Electric	San Jose, CA
Electrical Equipment	Los Angeles, CA
Motor Vehicles (Air, Cars, and Rail)	Seattle, WA
Furniture and Fixtures	Los Angeles, CA
Miscellaneous Manufacturing	Los Angeles, CA
Wholesale Trade	New York, NY
Transportation and Storage	Honolulu, HI
Professional and Business Services	San Jose, CA
Other	Los Angeles, CA
Communication	New York, NY
Finance and Insurance	New York, NY
Education	New York, NY
Health	New York, NY
Accommodation	Honolulu, HI

estimates found in the literature, with the possible exception of services. However, as in previous literature, the elasticity of (tradable) services productivity with respect to city size is several times larger than that of manufacturing. To the extent that regional prices are not readily available, the relationship between TFP and city size estimated in some of the existing literature captures variations in nominal TFP, that is $\ln TFP_n^j + \ln P_n^j$. In other words, while our model inversion produces measures of P_n^j, the absence of local price data can otherwise bias downward empirically estimated elasticities of TFP with respect to city size. Indeed, our findings indicate that elasticities of real TFP with respect to city-size are somewhat larger than those of nominal TFP.

We also compare regressions coefficients of TFP on the share of skilled workers to coefficients estimated by Moretti (2004b) using a panel of firms (we use the CNR share of employment, whereas he uses the college educated share of employment). Again, we find semi-elasticities that are of the same sign and comparable in magnitude to those in Moretti (2004b). As before, the regression coefficients become larger when deflated by the model-consistent regional price index.

Table 5 shows the coefficients in Table 4 for tradable sectors disaggregated by industry. We find that the positive relationship between TFP and city size holds uniformly across all tradable sectors. In addition, we find that the semi-elasticity of TFP in Computer and Electronics with respect to employment composition across cities is more than twice as large as the average for manufacturing, replicating the finding by Moretti (2004b) for high tech. sectors.

Table 4: Elasticities of TFP with respect to City Size and Employment Composition

	$\ln(L_n)$		$\frac{L_n^{CNR}}{L_n}$	
	Real	Nominal	Real	Nominal
Average[1]	0.041	0.023	0.864	0.559
Manufacturing Average	0.025	0.014	0.530	0.407
Tradable Services Average	0.061	0.034	1.271	0.744
Melo et. al. Economy[2]	0.031 (0.099)		—	
Melo et. al. Manufacturing	0.040 (0.095)		—	
Melo et. al. Services	0.148 (0.148)		—	
Moretti College Share[3] (Manufacturing)	—		0.846 (0.102)	

1. Excludes non-tradables
2. Melo et al. (2009) "A Meta-analysis of estimates of urban agglomeration economies", Table 2. "By type of response variable" and "By industry group."
3. Moretti (2004b) "Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions", Table 2. College share in other industries, Cobb-Douglass production, 1992.

Table 5: Sectoral Elasticities of TFP with respect to city size and employment composition

	$\ln(L_n)$		$\frac{L_n^{CNR}}{L_n}$	
	Real	Nominal	Real	Nominal
Food and Beverage	0.021	0.011	0.396	0.279
Textiles	0.023	0.002	0.191	0.161
Wood, Paper, and Printing	0.021	0.010	0.555	0.264
Oil, Chemicals, and Nonmetallic Minerals	0.045	0.029	0.912	0.772
Metals	0.021	0.007	0.453	0.223
Machinery	0.016	0.011	0.383	0.318
Computer and Electric	0.053	0.040	1.413	1.159
Electrical Equipment	0.013	0.011	0.366	0.316
Motor Vehicles (Air, Cars, and Rail)	0.019	0.011	0.464	0.316
Furniture and Fixtures	0.014	0.005	0.130	0.239
Miscellaneous Manufacturing	0.029	0.016	0.569	0.434
Wholesale Trade	0.049	0.031	0.923	0.705
Transportation and Storage	0.037	0.021	0.646	0.447
Professional and Business Services	0.069	0.029	1.476	0.700
Other	0.066	0.041	1.093	0.853
Communication	0.036	0.031	0.826	0.733
Finance and Insurance	0.070	0.046	1.532	1.052
Education	0.100	0.050	2.373	1.101
Health	0.081	0.031	2.159	0.680
Accommodation	0.046	0.022	0.412	0.425

4.1.4 Amenities

We now turn to the amenities implied by the model inversion. The relationship between relative amenities parameters for CNR and non-CNR workers and the size and composition of cities is depicted in Figure 4. Our findings conform to Diamond (2016), in that we find that cities with more CNR workers are also relatively more amenable to those same workers. At the same time, we also find that larger cities are relatively more amenable to CNR workers, helping account for the concentration of CNR workers in large cities.

Diamond (2016) provides evidence for a causal impact of local population composition on amenities. One natural question is the extent to which the observed relationships between city composition, city size and relative amenities would survive in the absence of those endogenous effects. The panels in Figure 5 provide an answer to that question after filtering out the part of amenities which Diamond's (2016) estimates imply are endogenous to the local labor composition.[19] While suppressing those endogenous effects eliminates the positive relationship between CNR share and relative amenities, the relationship between the relative residual amenities and city size becomes stronger. This reflects the fact that, given the estimates by Diamond (2016), large non-CNR populations generate larger congestion effects on the CNR than on the non-CNR.

4.2 Estimating Production Externalities by Worker Type

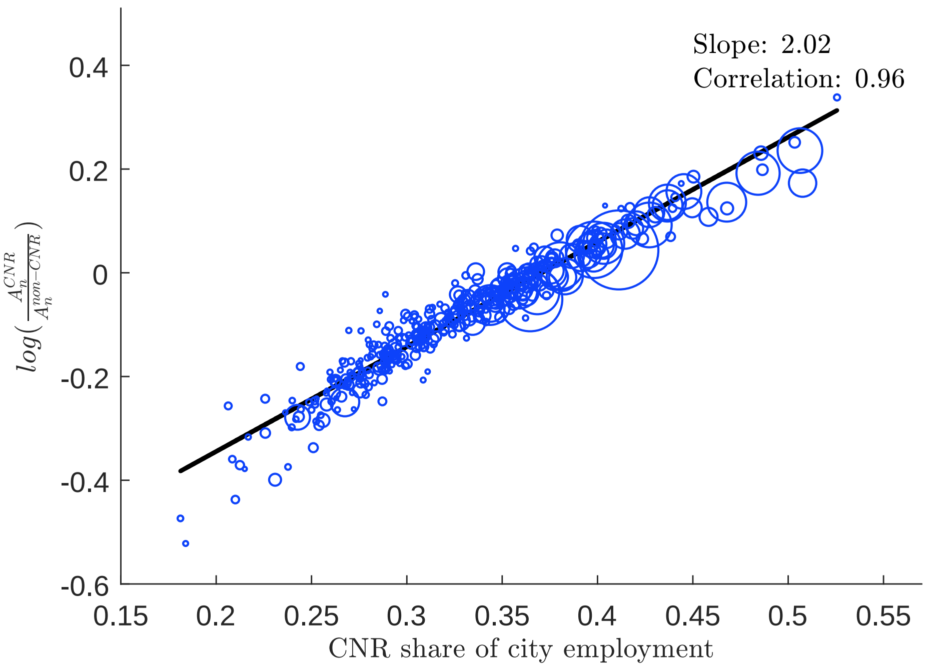
So far we have described the equilibrium levels of productivity compatible with observed data on wages and sectoral employment. We now turn to estimating how those levels are determined by the population composition of cities. In particular, we assume that spill-overs have the same labor augmenting effect in all sectors. This implies the following functional form:

$$\ln T_n^{kj} = \alpha^k + \tau^{R,k}(1-\beta_n^j)\gamma_n^j \ln\left(\frac{L_n^k}{L_n}\right) + \tau^{L,k}(1-\beta_n^j)\gamma_n^j \ln\left(L_n^k\right) + d^{kj} + u_n^{kj}, \qquad (13)$$

where d^{kj} denotes a set of industry dummies and u_n^{kj} denotes city-specific sources of natural advantages in production of sector k with workers of type k. Our functional form allows for an agglomeration effect that independends on city composition, and an effect of the share of each worker type.

The first column of Table 6 reports the coefficients from a simple OLS regression where we allow for two-way clustered standard errors by city and industry. We only use productivity parameters extracted for tradable industries, which we measure with the help of demand and supply relationships. The estimated coefficients are positive and significant. They indicate that individual productivity is enhanced by the presence of other workers of the same occupational group. The cross-occupational effects depend on the estimated parameter values.

[19]We use the parametrization that Fajgelbaum and Gaubert (2018) obtain based on the estimates by Diamond (2016). See appendix for details.



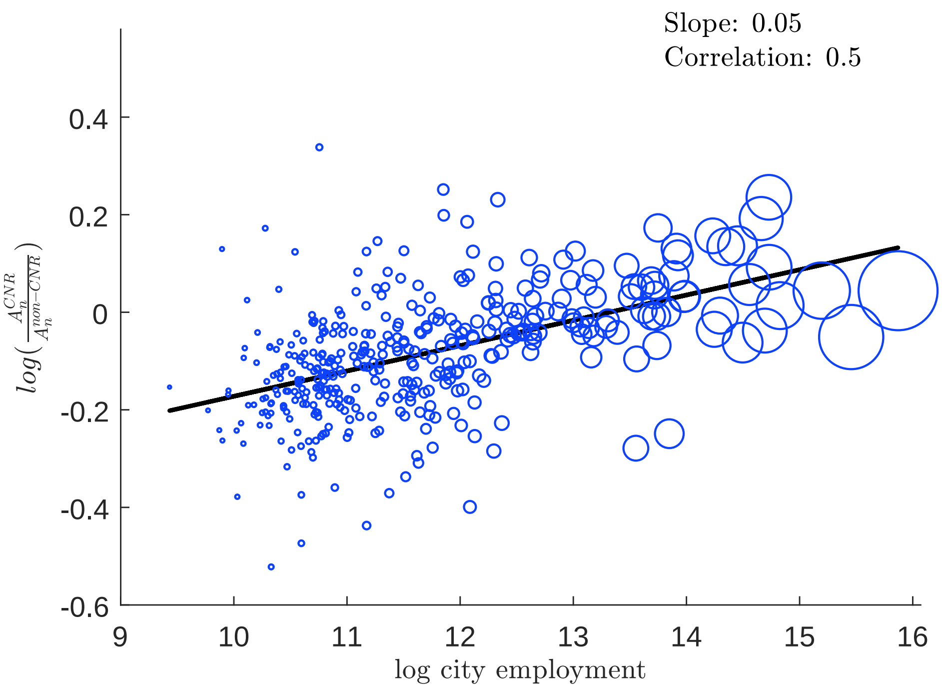


Figure 4: Relative amenities and city size and composition

Ratio of occupational-specific amenity parameters for each city obtained from the model quantification against employment share of CNR workers and log total employment. Each observation refers to a CBSA. Marker sizes are proportional to total city employment.

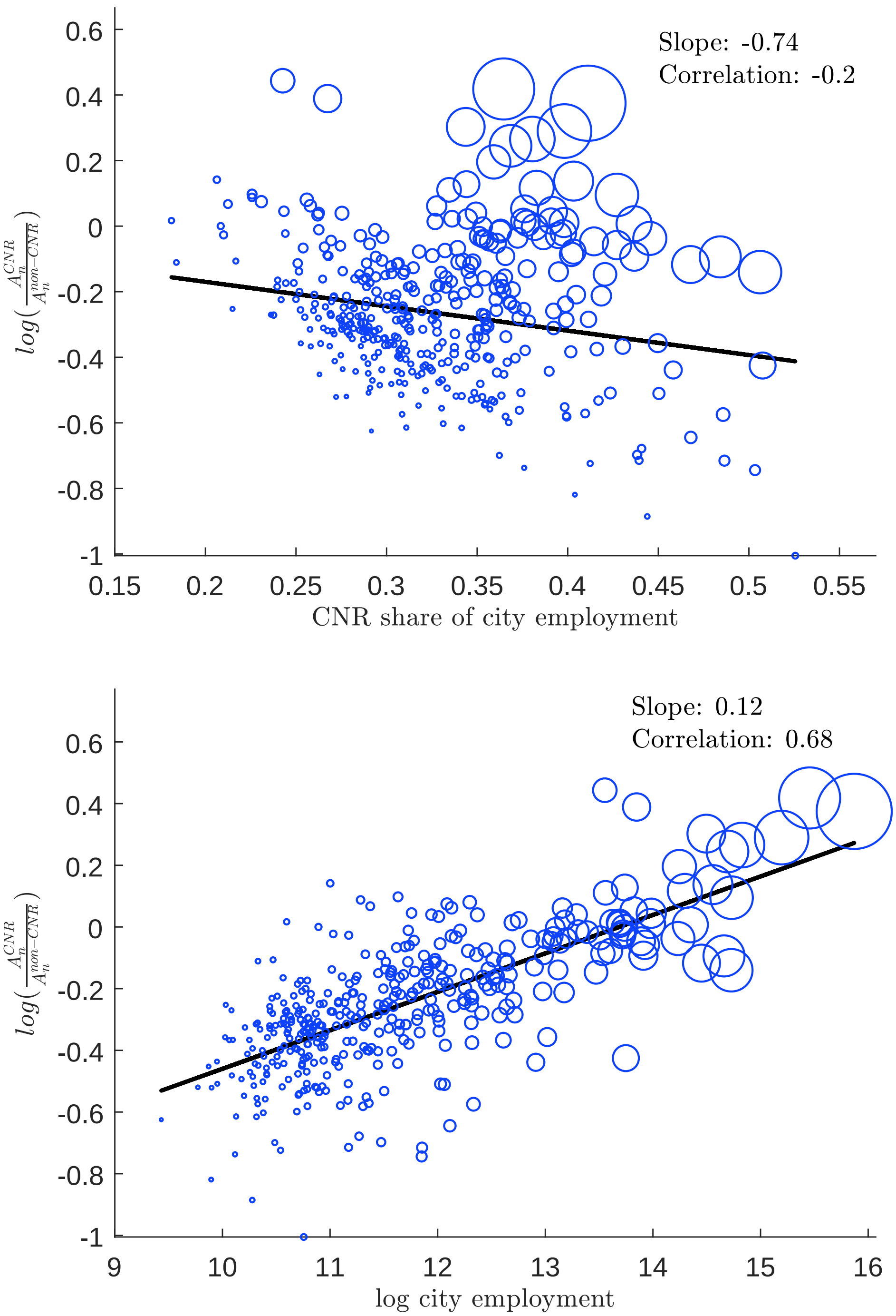


Figure 5: Relative amenities and city size and composition (exogenous part)

Ratio of occupational-specific amenity parameters for each city obtained after extracting the endogenous part of amenities implied by the parametrization used by Fajgelbaum and Gaubert (2018). Each observation refers to a CBSA. Marker sizes are proportional to total city employment.

Specifically, for $k \neq k'$,

$$\frac{\partial \ln \lambda_n^k}{\partial \ln L_n^{k'}} = -\left(\tau^{R,k} - \tau^{Lk,}\right) \frac{L_n^{k'}}{L_n} \ ,$$

so that the cross-occupational externalities are negative if $\tau^{L,k} < \tau^{R,k}$. The estimates indicate that this is the case, so that cross-occupational externalities are always negative. Given the discussion in Section 3.4, this implies that individuals of a given occupation have their highest external impact in cities where those individuals are proportionately more represented.

Those estimates are conceivably biased since, as implied by the model, workers of a given type may choose to live in cities where they are relatively most productive. This induces a correlation between the exogenous component of worker productivity, u_n^{kj}, and the share of each type of worker in a given city. Also, the estimates might be biased by the presence of omitted variables which are correlated with both u_n^{kj} and the occupational ratio.

Table 6 explores the effects of adding various controls to our basic OLS regression. Column 2 includes dummies for 9 census divisions.[20] Those should absorb many of the geographical and historical components that may jointly determine amenities and productivity in different places. Column 3 introduces the share of manufacturing workers in 1920 as a control. This aims to extract long standing factors that may influence the industrial composition in individual places. Column 4 introduces geographic amenities constructed by the United States Department of Agriculture (USDA) that include measures of climate, topography and water area.[21] Those controls allow for the possibility that the same geographic characteristics that may lead workers to choose certain cities may also influence their productivity. Finally, column 5 adds controls for demographic characteristics of different cities, including racial composition, gender split, the fraction of immigrant population, and age structure.[22] Those help narrow down the identification of the externality coefficients to the extent that more productive cities attract individuals of certain demographic make-up.

The addition of all these controls changes the point estimates for the coefficients on CNR workers by only a small amount, and enhances the effect of labor market composition on non-CNR workers. While those controls may help extract exogenous sources of productivity variation that affect individual location decisions, any residual variation in productivity may still be correlated with population levels and composition, leading to biased coefficient estimates. In order to further account for those residual effects, we adopt an instrumental variable strategy, drawing on the existing empirical literature for candidate sources of exogenous variation. The main difference from our approach is that we seek to explain productivity measures extracted from a structural model directly.

[20] Those are 1. New England, 2. Mid-Atlantic, 3. East North Central, 4. West North Central, 5. South Atlantic, 6. East South Central, 8. Mountain and 9. Pacific

[21] Geographic controls include average temperature for January and July, hours of sunlight in January, humidity in July from 1941 to 1970, variation in topography, and percent of water area.

[22] Demographic controls are, by city, the percent female, black, hispanic, and percent in the age bins 16-25 and 26-65 (observations related to the younger than 16 population are dropped from the sample, and the age bin 66+ is omitted from the regression).

Table 6: OLS Estimates

	(1)		(2)		(3)		(4)		(5)	
VARIABLES	CNR	non-CNR	CNR	non-CNR	CNR	non-CNR	CNR	non-CNR	CNR	non-CNR
$\gamma_n^j log(\frac{L_n^k}{L_n})$	0.863***	0.575**	0.819***	0.679***	0.829***	0.721***	0.822***	0.705***	0.885***	0.714***
	(0.12)	(0.23)	(0.11)	(0.19)	(0.12)	(0.21)	(0.12)	(0.21)	(0.12)	(0.22)
$\gamma_n^j log(L_n)$	0.395***	0.319***	0.394***	0.319***	0.386***	0.323***	0.388***	0.321***	0.369***	0.303***
	(0.05)	(0.04)	(0.05)	(0.04)	(0.05)	(0.04)	(0.05)	(0.04)	(0.05)	(0.04)
Jan. Temp					0.00772	-0.0695***	0.00160	-0.0705***	-0.00785	-0.0615**
					(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)
Jan. Hrs Sun					0.0716***	0.0454***	0.0689***	0.0459***	0.0606***	0.0520***
					(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)
July Temp					-0.0784***	-0.0232	-0.0743***	-0.0195	-0.0678***	-0.0370*
					(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
July Humid					-0.0189	-0.0119	-0.0137	-0.00986	0.00351	-0.0218
					(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)
Topology					-0.0283**	-0.0255**	-0.0273**	-0.0250**	-0.0353***	-0.0179
					(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
% Water Area					0.0364***	0.00254	0.0372***	0.00287	0.0340***	0.00674
					(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
% 1920 Mfg Workers							-0.0100	0.00384	-0.00564	0.00226
							(0.01)	(0.01)	(0.01)	(0.01)
% female									-0.0406***	-0.0415***
									(0.01)	(0.01)
% Black									-0.000468	0.0396***
									(0.02)	(0.01)
% Hispanic									0.0299*	0.00754
									(0.02)	(0.02)
% Age 16-25									0.00143	0.0136
									(0.03)	(0.04)
% Age 26-65									0.0177	0.0269
									(0.03)	(0.04)
Industry FE	X	X	X	X	X	X	X	X	X	X
Census Division FE			X	X	X	X	X	X	X	X
Observations	7,640	7,640	7,640	7,640	7,560	7,560	7,460	7,460	7,459	7,459
R-squared	0.578	0.749	0.634	0.788	0.645	0.796	0.646	0.799	0.648	0.801

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

4.2.1 Instrumenting for Employment Level and Composition

In order to isolate the residual simultaneity between exogenous productivity variation and labor allocation, we resort to variants of instruments proposed in the literature. Specifically, we follow Ciccone and Hall (1996) and use population a century before to capture historical determinants of current population, and we follow Card (2001) and Moretti (2004a), and use variation in early immigrant population and the presence of land-grant colleges to capture historical determinants of skill composition of cities. We now discuss the particular instruments in more detail.

Population in 1920 Ciccone and Hall (1996) argue for the validity of historical variables as instruments under the assumption that, after allowing for the controls described above, original sources of agglomeration only affect current population patterns through the preferences of workers, and not through their effect on the residual component of productivity. This reasoning motivates using population almost one hundred years before our data period as an instrument and will also serve as motivation for the other instruments, described below.

Irish Immigration in 1920 Next, we use the fraction of Irish immigrants in the population of each city in 1920. This instrument is motivated by Card (2001), who uses the location of immigrant communities as an instrument for labor supply in different occupations. For our purposes, we focus on the location of Irish immigrants following evidence reviewed by Neal (1997), and further studied by Altonji et al. (2005), showing that attending catholic schools substantially increases the likelihood of completing high school and college education. We use as an instrument the fraction of Irish immigrants, rather than the overall catholic population, because Irish immigrants represented the first wave of catholic immigration to the U.S. and, therefore, historically were the first to invest in education. As additional validation for this instrument, we compile data on the current location of catholic colleges, and observe that MSAs in which catholic colleges are present had in 1920 more than three times the fraction of Irish immigrants as other locations.

The Presence of Land-Grant Colleges Lastly, following Moretti (2004a), we also use as an instrument the presence of a land-grant college within the city. Land-grant colleges were established as a result of the Morrill Act of 1862, and extended in 1890, a federal act that sought to give states the opportunity to establish colleges in engineering and other sciences. Since the act is more than a century old, the presence of a land-grant college in the city is unlikely to be related to unobservable factors affecting productivity in different cities over our base period, $2011-2015$. At the same time, as shown in Moretti (2004a), the presence of land-grant colleges is generally correlated with the composition of skills across cities.

IV Estimates Table 7 shows the estimation results with instrumental variables and all the controls. The first column repeats the OLS results in the last column of Table 6, and the second column shows the corresponding two-stage-least-squares estimates. Those are similar to the OLS estimates, being well within one standard error from one another. To evaluate the strength of the instrumental variables, we follow the procedure in Sanderson and Windmeijer (2016) to obtain separate first-stage F statistics for each of the endogenous variables.[23] Since the F-statistics are below 10, the critical value that Staiger et al. (1997) recommend as indicative of strong instruments, the estimates may be biased and have incorrect standard errors. The literature on weak IVs recommends resorting to limited information maximum likelihood (LIML) estimators, since those are unbiased and have more accurate standard errors for lower first-stage F statistics. The third column performs the estimation using continuously updated GMM estimator (GMM-CUE), which has similar properties as a limited information maximum likelihood estimator but allows for clustered and heteroskedastic standard errors. The Stock and Yogo (2005) critical values in the LIML model for a p-value of 5% to be 10% or better is 6.46, at or close to our obtained values.

We adopt the GMM-CUE coefficients in the last column as our benchmark. The coefficients on the former are clearly larger than on the latter, indicating negative cross-occupational externalities. The externality effects are substantial. In particular, they imply that, all else constant, moving from a city with a share of CNR employment of 32 percent

[23] This follows largely the intuition laid out by Angrist and Pischke (2008), and account for the fact that strong IVs have to be able to predict the two endogenous variables independently from one another.

(corresponding approximately to the median value), such as Lancaster, PA to one with a share of CNR employment of 36 percent (corresponding to the 75$^{\text{th}}$ percentile of the share distribution), such as Phoenix, AZ, increases the productivity of CNR workers by 14.9 percent and reduces that of non-CNR workers by 5.7 percent. Agglomeration externalities are similarly important. Moving from a city near the median of the size distribution, such as Midland, TX with 80 thousand employed workers to one near the third quartile of the city size distribution such as Trenton (NJ) with 182 thousand would imply a gain of 27.3 percent for CNR workers and 28.2 percent for non-CNR.

Table 7: Instrumental Variables Estimate

	(1) **OLS**		(2) **2SLS**		(3) **CUE**	
VARIABLES	CNR	non-CNR	CNR	non-CNR	CNR	non-CNR
$\gamma_n^j log(\frac{L_n^k}{L_n})$	0.885***	0.714***	1.190***	0.330	1.265***	0.938*
	(0.12)	(0.22)	(0.37)	(0.52)	(0.37)	(0.52)
$\gamma_n^j log(L_n)$	0.369***	0.303***	0.315***	0.269***	0.333***	0.343***
	(0.05)	(0.04)	(0.06)	(0.05)	(0.06)	(0.04)
Observations	7,459	7,459	7,459	7,459	7,459	7,459
K.P. F			4.180	5.748	4.180	5.748
S.W.F. L_n^k Share			6.429	8.964	6.429	8.964
S.W.F. L_n			6.384	9.040	6.384	9.040

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

4.3 The Role of Externalities in the Spatial Allocation

In Section 2 we conjectured that occupational externalities account for some salient patterns in the data. We now verify whether this basic intuition holds given the model's parametrization and estimated parameters.

Towards this end, we compute the counter-factual equilibrium allocation in the absence of externalities (i.e., if we set $\tau^{R,k} = \tau^{L,k} = 0$ for both k). The results are presented in Figures 6 and 7. Figures 6 shows that, absent externalities, the relationship between the share of CNR workers and the wage premium becomes negative, indicating that the composition of the labor-force is now determined by differences in labor supply. Figure 7 shows that, without externalities, the equilibrium wage bill share increases less than the wage bill share predicted by the industrial composition of the cities. Absent externalities, productivity is pinned down exogenously, and so a city that has a comparative advantage in the production of CNR intensive goods exhibits higher wages for CNR workers. Hence, firms in that city will substitute CNR for non-CNR workers and, given an elasticity of substitution between occupations greater than 1, reduce the CNR wage bill share.

Overall, these exercises indicate that the patterns we identified in Section 2 can in fact be generated by the set of occupation-specific elasticities we have identified and measured. Given the significance of these externalities, the optimal and equilibrium allocations differ, and so there is room for optimal spatial policy. We now describe the optimal allocation and the corresponding policies to implement it.

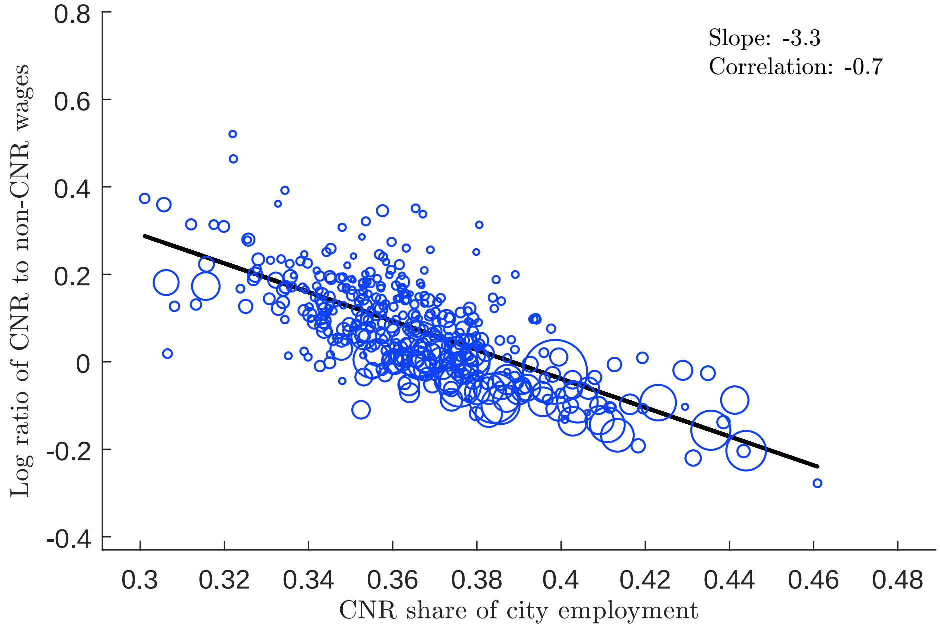


Figure 6: Occupational share and wage premium - no externalities

Counterfactual values obtained from assuming no externalities ($\tau^{R,k} = \tau^{L,k} = 0$), while keeping the exogenous part of productivity as originally quantified.

5 Optimal Allocation

Our aim is to characterize the optimal allocation and the policies to implement it. In the equilibrium allocation there is a 'wedge', for each city-occupation, between the total social marginal product of labor (including externalities) and the private value, internalized by firms and workers. We have denoted this wedge by Δ_n^k, and its value for each city-occupation can be calculated using equation 10. Figures 8 and 9 show the deviations of those wedges from their (employment weighted) means for CNR and non-CNR workers, respectively. The average wedge for CNR workers is itself fairly large, at \$51,572 dollars per worker, or 72 percent of the mean CNR wage. The average wedge for non-CNR workers is more modest and negative, at -\$5,053. The wedge of non-CNR workers can be negative because their presence reduces the share of CNR's in cities, which reduces the productivity of CNR workers. Together, these values imply an average gain of \$56,625 from switching a non-CNR worker for a CNR. This large value indicates that CNR workers are very scarce in the economy, and so using them productively is important. Furthermore, they indicate that education and migration policies that form and attract CNR workers have potentially, depending on their cost, high social value. In our analysis, however, we take the supply of CNR and non-CNR workers as given.

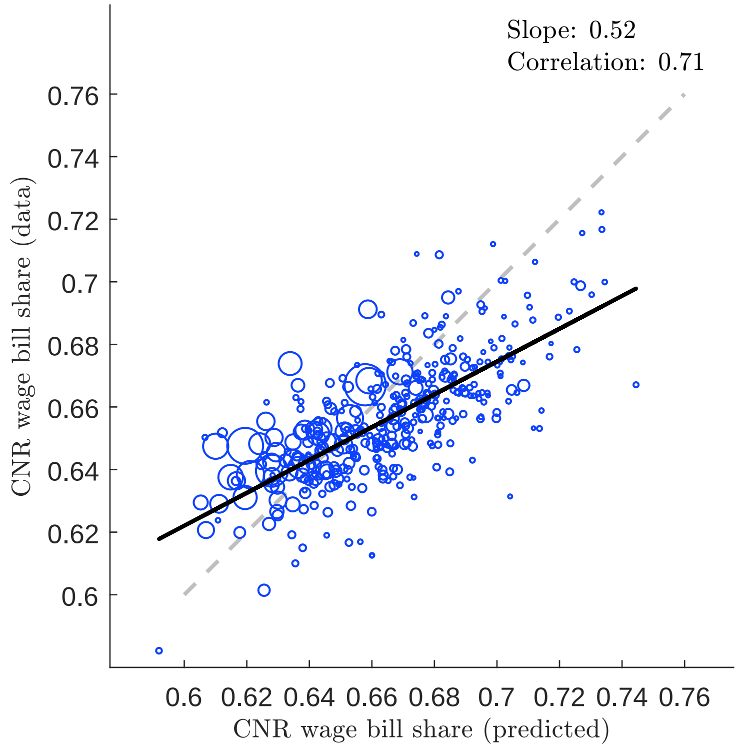


Figure 7: Occupational wage bill share: predicted vs equilibrium - no externalities

Counterfactual values obtained from assuming no externalities ($\tau^{R,k} = \tau^{L,k} = 0$), while keeping the exogenous part of productivity as originally quantified.

The wedge between the social and private value of CNR workers correlates positively and strongly with city size (0.43). The CNR wedge, in contrast, has a positive and mild correlation with the CNR share in the city (0.06). This is natural given the size externality and the fact that high CNR cities have already exploited CNR externalities, external effects are concave, and there are diminishing returns to each factor. The correlation becomes stronger (0.36) when weighted by city size, indicating that there are still relative gains to be exploited in larger, CNR intensive cities. Externalities from CNR workers appear to be particularly high in California, New York and Houston, and particularly low in Florida and, more broadly, in the South and Mid-West.

The overall pattern for non-CNR workers in Figure 9 is more pronounced, with a clearly negative relationship between the non-CNR wedge and the size and CNR share of cities. The social value of non-CNR workers relative to their private value is large in many small cities, but also in some large cities in Florida, as well as Las Vegas, and Phoenix. Clearly, these results indicate that non-CNR should be encouraged to move to small non-CNR abundant cities. It is in those cities where they can make their largest contribution.

We compute the optimal allocation under the assumption that welfare weights are such that, in the planned solution, the welfare gains are proportionately equal for both types of workers. Figures 10 and 11 show the percentage change in employment in the optimal

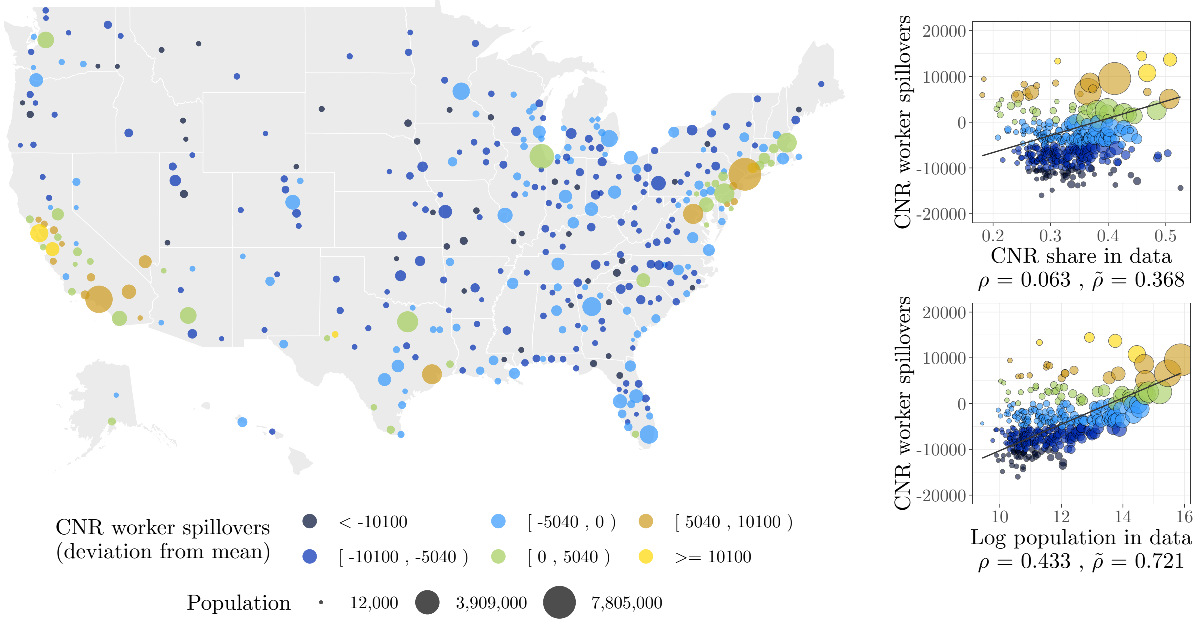


Figure 8: Δ_n^{CNR} (equilibrium values)

Δ_n^{CNR} captures the wedge between the social and the private marginal product of labor for CNR workers, as described in Lemma 1. Figure depicts deviation from employment weighted average (US$ 51,572). Each marker in the map refers to a CBSA. Marker sizes are proportional to total employment in each city. ρ and $\tilde{\rho}$ are unweighted and population weighted correlations respectively.

allocation relative to the equilibrium allocation for CNR and non-CNR workers, respectively. The results indicate that it is optimal to move CNR workers to larger cities and non-CNR workers to smaller cities, thereby exacerbating the spatial polarization of employment. This increased polarization follows naturally from the spill-over coefficient estimates in Section 4.2, which point out that both types of workers (but particularly those in CNR occupations) become more productive when clustered with other workers of their own type.

As Figures 10 shows, the increases in CNR workers are particularly large in cities like New York, San Francisco or San Jose, where the wedge between social and private marginal product of labor for CNR workers is particularly high. These cities, together with other large cities like Chicago, Dallas, and Los Angeles, which are less CNR intensive, become the cognitive hubs in the optimal allocation. More generally, cognitive hubs should be created in larger cities that are more CNR abundant in the equilibrium allocation. Cities that gain CNR workers are somewhat uniformly distributed in space according to overall economic activity. They are cognitive hubs in that they absorb CNR workers with smaller cities around them losing CNR workers.

The counterpart of new cognitive hubs are larger non-CNR abundant small and medium sized cities. Figure 11 illustrates that while the planner chooses to generally remove non-

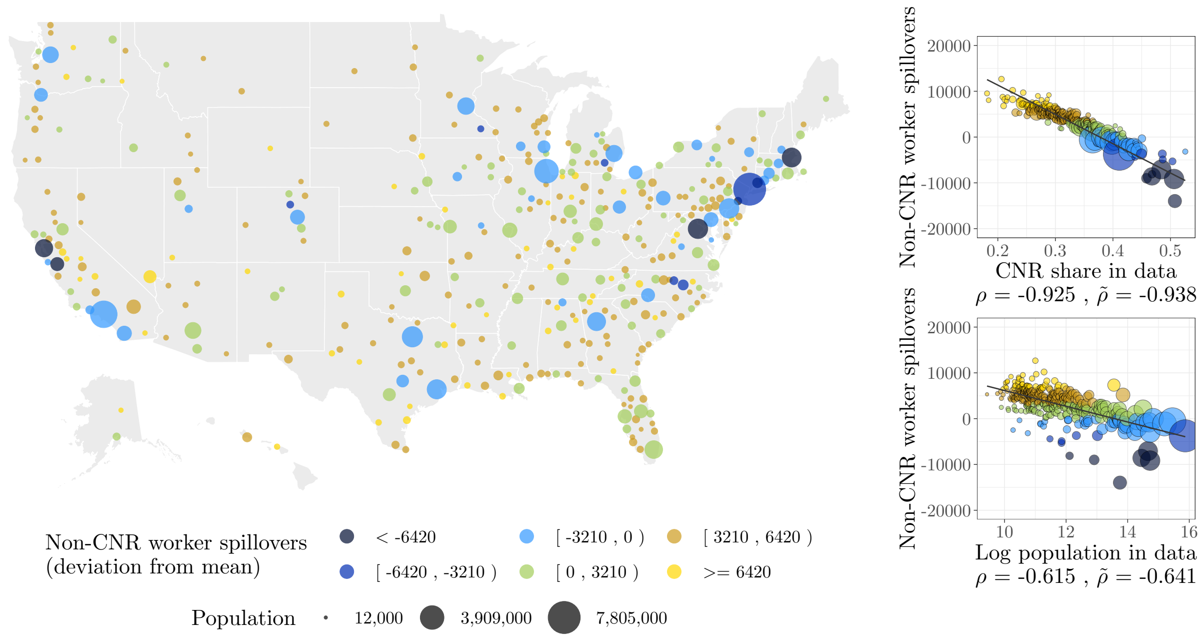


Figure 9: Δ_n^{nCNR} (equilibrium values)

Δ_n^{nCNR} captures the wedge between the social and the private marginal product of labor for non-CNR workers, as described in Lemma 1. Figure depicts deviation from employment weighted average (US$ -5,053). Each marker in the map refers to a CBSA. Marker sizes are proportional to total employment in each city. ρ and $\tilde{\rho}$ are unweighted and population weighted correlations respectively.

CNR workers from large cities, in the optimal allocation some large cities become more non-CNR abundant. This is the case for cities like Miami, Las Vegas, Phoenix, and San Antonio where non-CNR workers have, in equilibrium, a social marginal product that is larger than their private marginal product. These cities become the non-CNR center. They specialize in non-CNR intensive industries and grow in size since the inflow of non-CNR workers is larger than the exodus of CNR workers specified by the optimal allocation.

Implementing the optimal allocation involves location and occupation specific transfers. These transfers imply that in the optimal allocation total consumption is equal to total income in a city. These transfers perform several roles. First, they incentivize agents to move as described above. Namely, they incentivize CNR workers to move to cognitive hubs and non-CNR workers to move to smaller towns. Second, they guarantee that relative welfare gains are the same across occupations and locations. Overall, the planner compensates non-CNR workers for moving to smaller and less productive and amenable cities by performing transfers from large to small cities. These transfers are mostly financed by CNR workers in larger cities. CNR workers are happy to pay, since they benefit from the policy as well. The net flow of resources between received or paid by cities (the trade balance) is shown in Figure 13, where transfers are calculated by the difference between local nominal per capita

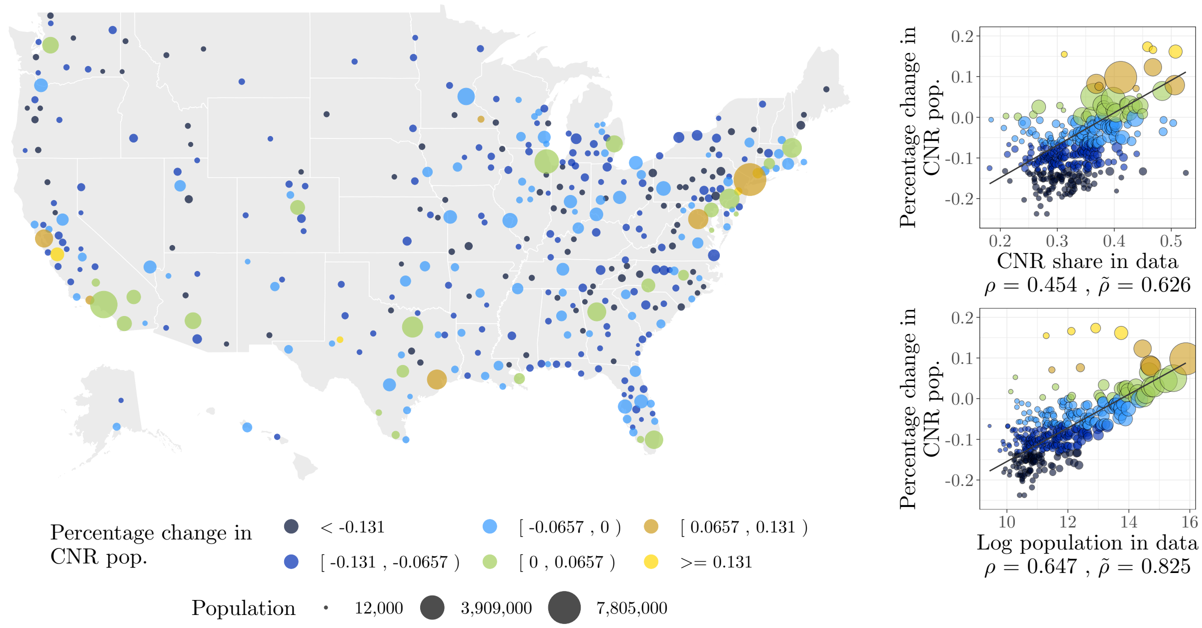


Figure 10: L_n^{CNR} (percentage change from data equilibrium)

Percentage change in employment of CNR workers between equilibrium and optimal values. Each marker in the map refers to a CBSA. Marker sizes are proportional to total equilibrium employment in each city. ρ and $\tilde{\rho}$ are unweighted and population weighted correlations respectively.

consumption and output $(\sum_k P_n C_n^k L_n^k - \sum_k w_n^k L_n^k - r_n H_n)$.

The trade balance is large and negative for the new cognitive hubs. Cities like San Francisco and San Jose that are relatively large and very CNR intensive send net payments of as much as \$23,000 per resident, while some of the smaller cities receive net transfers of the same magnitude. These net transfers in some of the smaller cities amount to a form of basic income paid to all the lowest earning workers (non-CNR occupations in small cities) and financed (in net) by agents living in the cognitive hubs. Again, some relatively large cities like Las Vegas end up paying less or receiving net transfers since they become even larger centers for non-CNR employment.

The pattern of spill-over coefficients we estimated, together with the rest of the quantification of our model, implies that the optimal allocation exacerbates the extent of labor market polarization. However, it does not reveal why labor markets ended up being as polarized as they already are in equilibrium. As indicated in Figure 4, larger cities are relatively more amenable to CNR workers, a pattern that survives even if we exclude the part of amenities that work by Diamond (2016) indicates is potentially endogenous. In effect, in the appendix we show that, even if that part is excluded, the planner chooses to concentrate CNR workers in large cities.

The overall gains in welfare from implementing the optimal allocation are 0.38% for

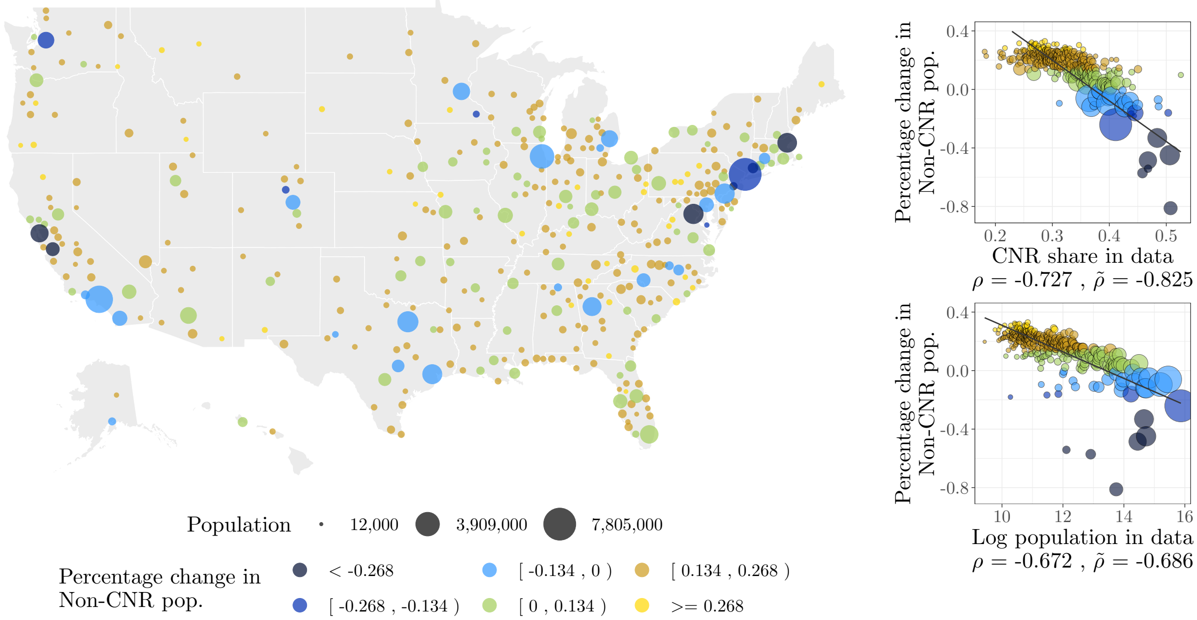


Figure 11: L_n^{nCNR} (percentage change from data equilibrium)

Percentage change in employment of non-CNR workers between equilibrium and optimal values. Each marker in the map refers to a CBSA. Marker sizes are proportional to total equilibrium employment in each city. ρ and $\tilde{\rho}$ are unweighted and population weighted correlations respectively.

workers in both occupations. The gains seem perhaps small, but they guarantee that both types of workers benefit. One reason for the relatively small gains is that the observed equilibrium allocation is already fairly polarized, so potential gains from incentivizing an increase in polarization are small. In fact, in Section 6 we show that the distribution of occupations has been approaching today's optimal allocation since 1980.

In order to achieve equal welfare gains the optimal transfer scheme has two components. One that incentivizes agents to go to the 'right' location and is related to the differences in Δ_n^k across locations. The other is a fixed transfer by occupation (R^k). This fixed transfer guarantees that all workers obtain equal gains from moving to the planner allocation. This fixed transfer amount to a negative transfer (a payment) of -\$11,573 for CNR workers and a positive transfer of \$15,242 to all non-CNR workers. The latter can be implemented as a universal basic income that is paid by CNR workers. This transfer can be viewed as the redistribution that CNR workers are willing to accept to form the cognitive hubs where they can thrive.

The optimal transfers also involve a component that incentivizes CNR worker to move to large, CNR abundant, cities and for non-CNR workers to move to smaller cities with a smaller fraction of CNR workers. This is achieved by giving large incentives to non-CNR workers to move out of large, and more markedly, CNR abundant cities. Due to the occupation specific

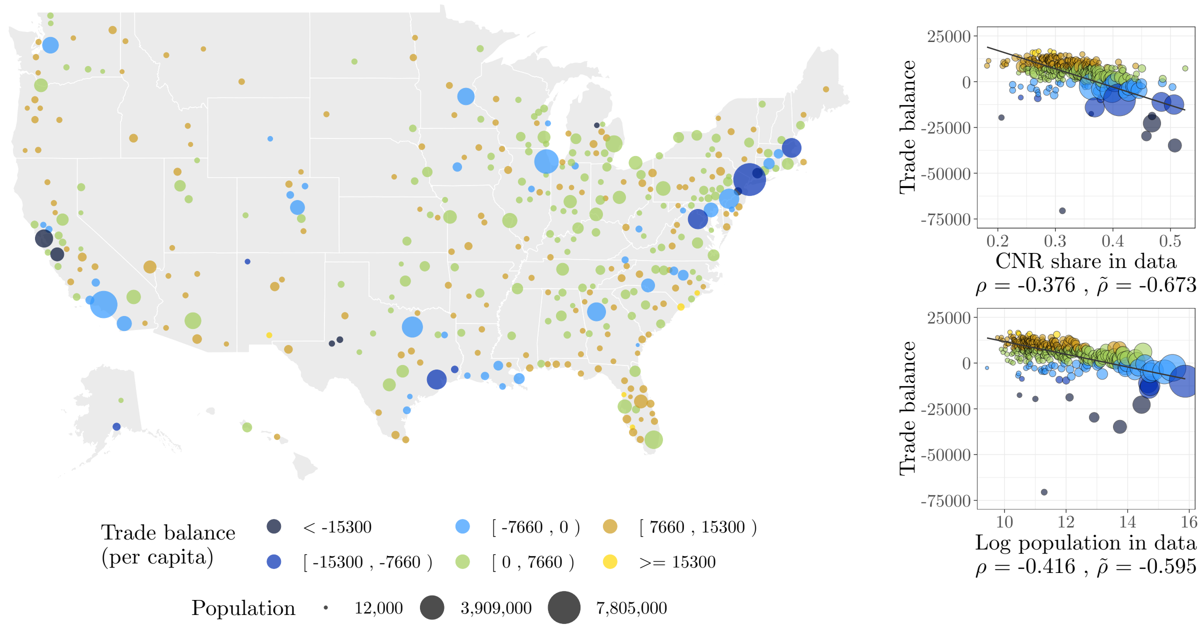


Figure 12: Trade balance per capita

Trade balance is defined as the difference in the optimal allocation between the value consumed and value added in each city ($\sum_k P_n C_n^k L_n^k - \sum_k w_n^k L_n^k - r_n H_n$). Trade balance per capita are given by those values divided by L_n. Each marker in the map refers to a CBSA. Marker sizes are proportional to total equilibrium employment in each city. ρ and $\tilde{\rho}$ are unweighted and population weighted correlations respectively.

externalities, this reallocation yields larger CNR productivity increases in CNR abundant cities, which attracts CNR workers to these cities and eliminates the need for large net transfers to theses workers. In fact, in a handful of the most CNR abundant cities, the effect is so strong that the planner prefers to balance it with modest negative transfers to avoid congestion.

The optimal transfers are, of course, related to the Δ_n^k wedges described in Figures 8 and 9, and are depicted in Figures 13 (for the CNR) and 14 (for the non-CNR). Once the incentive-based transfers are taken into account, non-CNR workers in the median city receive a net transfer of \$6,552, ranging between as little as \$925 (in the 10th percentile city) and as much as \$8,496 (in the 90th percentile city). Note that non-CNR workers have to be incentivized to move to smaller cities. However, because cognitive hubs offer a high wage to non-CNR workers (since those are productive cities and have few non-CNR workers), non-CNR workers in these cities need to pay a transfer too, in order to prevent other non-CNR workers from joining them. This explains the wide range in non-CNR transfers and reduces the average net burden on CNR workers.[24]

[24] If we weight by population, transfers and contributions to non-CNR workers range from a net contribu-

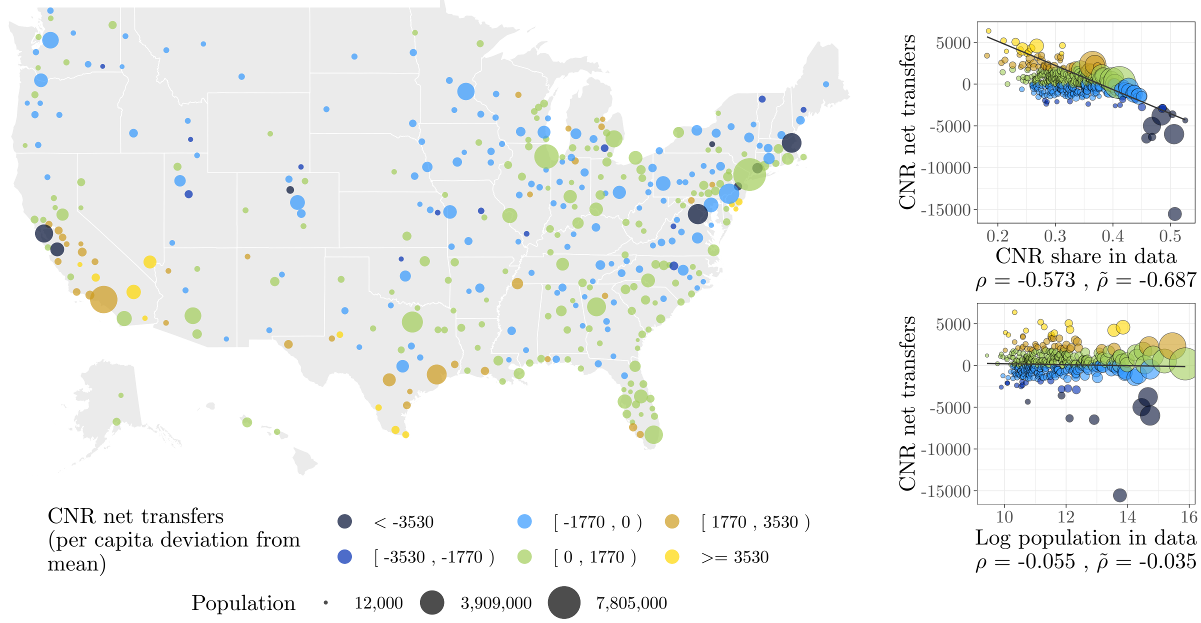


Figure 13: Optimal transfers to CNR workers (per CNR worker)

Optimal transfers per CNR workers are defined as the difference in the optimal allocation between the value consumed and the income they would receive given optimal wages and rents but absent the transfers ($P_n C_n^{CNR} - w_n^{CNR} - \chi^k$). Each marker in the map refers to a CBSA. Each marker in the map refers to a CBSA. Marker sizes are proportional to total equilibrium employment in each city. ρ and $\tilde{\rho}$ are unweighted and population weighted correlations respectively.

In the median city, after incentives are taken into account, CNR workers contribute \$1,477. This number is modest since CNR workers are very socially valuable and need to be incentivized to stay in the large, CNR abundant, cities. The results yield a range of net transfers to CNR workers that is significantly smaller than the range for non-CNRs, since most of the location incentives for this group are achieved via production externalities. CNR's are net-recipients of transfers in close to 15% of the locations, but net payers in 85%. Overall, payments to and from CNR workers range from a net receipt of US\$357 (in the 10th percentile city) to a net payment of US\$2,999 (in the 90th percentile).[25]

tion of about \$5,600 (10th percentile) to net receipts of \$7,340 (90th percentile), with a median net receipt of about \$2,148.

[25] Weighing by population net transfers to or from CNR workers range from a receipt of close to \$630 (10th percentile) to a contribution of about \$5,500 (90th percentile), with the median net contribution being close to \$1,460.

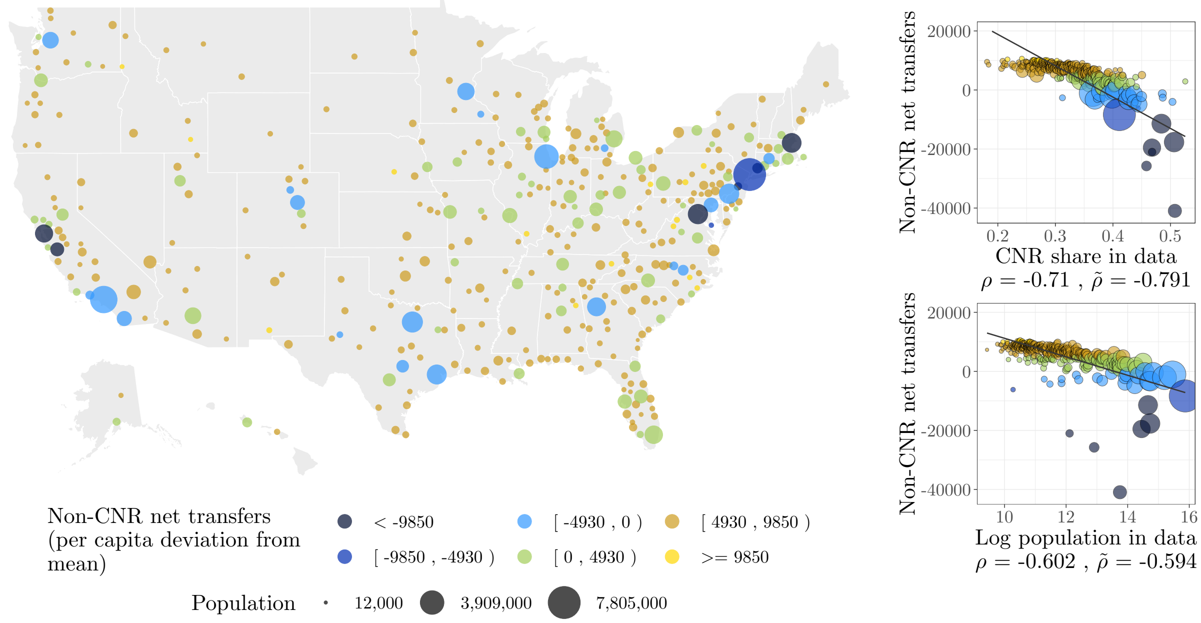


Figure 14: Optimal transfers to non-CNR workers (per non-CNR worker)

Optimal transfers per non-CNR workers are defined as the difference in the optimal allocation between the value consumed and value added in each city attributed to non-CNR workers ($P_n C_n^{nCNR} - w_n^{nCNR} - \chi^k$). Each marker in the map refers to a CBSA. Marker sizes are proportional to total equilibrium employment in each city. ρ and $\tilde{\rho}$ are unweighted and population weighted correlations respectively.

6 The Formation of Cognitive Hubs after 1980

The US economy has evolved towards the formation of cognitive hubs at least since the 1980. Quantifying our model to 1980 data yields a set of fundamental characteristics of the economy that allows us to study this phenomenon in detail. In order to compare the spatial structure of the economy in 1980 to the one in 2015, we want to abstract from aggregate technology trends. Thus, we first build a 'Baseline' economy that adds only *aggregate* changes in technology, population, and input shares to the 1980 economy.[26] The Baseline economy does not include any location-specific change in productivities across industries and occupations or in amenities. It also does not include changes in aggregate CNR shares of population. We study the role of these components by adding them gradually. As shown in Figure 15, in this Baseline 1980 economy, US population is concentrated in cities in which the CNR share of employment was close to average. In the 2011-15 data, however, there is greater dispersion around the (now larger) average. Figure 15 further shows the distribution of CNR workers implied by the 2011-2015 planner's solution calculated above. Together,

[26]For details on the construction of this Baseline economy and other counterfactuals, see Appendix E.

the histograms imply that the increasing concentration of CNR workers was in the direction implied by today's optimal allocation.

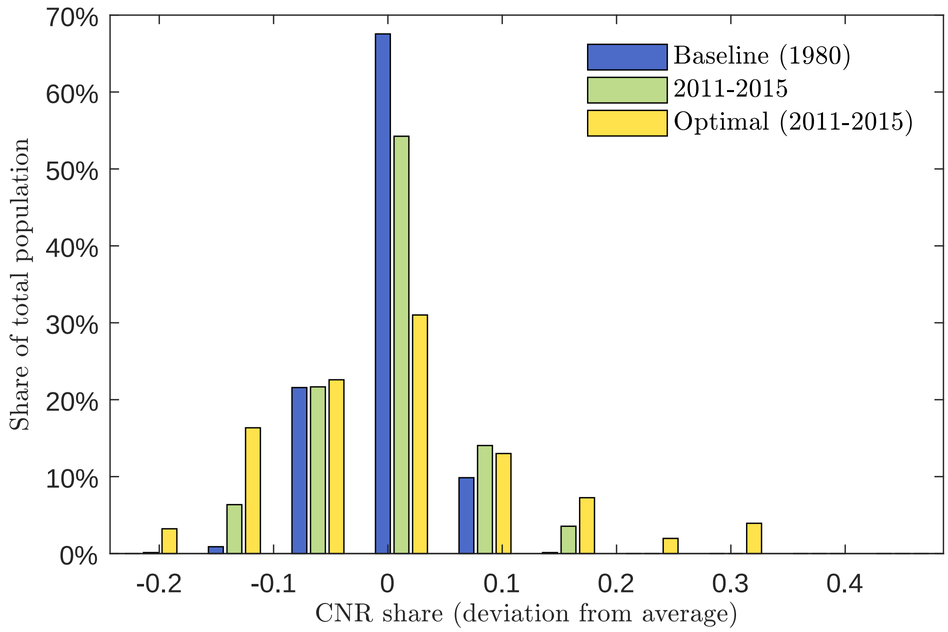


Figure 15: Distribution of population by CNR share in city of employment

Share of population in cities with different ratio of CNR workers to total population. Bins refer to deviation from population weighted average.

Our model allows us to examine the forces underlying the formation of cognitive hubs and obtain a quantitative assessment of their welfare relevance. We do this with a series of counter-factual exercises that clarify the importance of different forces in driving national trends. Thanks to the structural model, we can also do the same decomposition in the absence of externalities. Those exercises then provide us with a measure of the relevance of local spill-overs for observed spatial trends. Table 8 shows how such a decomposition affects the welfare of CNR and non-CNR workers. The columns depict welfare levels relative to the Baseline 1980 economy for each occupation. The lower panel repeats the exercises for a world without externalities (i.e., where we set the externality elasticity parameters to zero).

The top row of the table shows that welfare in 1980 was lower for both groups and in all scenarios, as one would expect given underlying technological trends. The second row shows the effects of changing input shares. As is well known, CNR intensive industries have become a larger part of the US economy, leading to relative gains for CNRs as compared to non-CNRs. The next step (rows 3 and 14) brings total population and average technology in each city/industry to 2011-15 levels, while keeping the relative productivity of CNRs and non-CNRs at 1980 levels. Relative to baseline (rows 4 and 15), CNRs are worse off and non-CNRs better off. The difference is accounted for skill-biased technical change. Note that externalities amplify the effect of skill biased technical change, although the quantitative effect is modest.

Table 8: Welfare Comparison, Relative to Baseline

	CNR	non-CNR	CNR-to-non-CNR Ratio
Full Model			
1. 1980 parameters	0.586	0.804	0.729
2. (1) + current input shares	0.623	0.725	0.859
3. (2) + national trends in technology and population	0.927	1.081	0.857
4. (3) + national skill biased technical change (Baseline)	1.000	1.000	1.000
ratio of CNR to non-CNR welfare in Baseline		1.907	
5. (4) + change in occupation share in employment	1.037	1.049	0.989
6. (5) + change in local technology ex real estate	1.056	1.067	0.990
7. (6) + change in real estate productivity	1.031	1.048	0.984
8. (7) + change in amenities (2011-15 parameters)	1.030	1.050	0.981
9. Optimal Allocation	1.034	1.054	0.981
10. 2011-15 parameters ex real estate technology	1.044	1.060	0.986
11. Optimal allocation with parameters in (10)	1.051	1.066	0.986
Model without externalities			
12. 1980 parameters	0.503	0.665	0.757
13. (12) + current input shares	0.692	0.783	0.883
14. (13) + national trends in technology and population	0.935	1.059	0.883
15. (14) + national skill biased technical change (Baseline)	1.000	1.000	1.000
ratio of CNR to non-CNR welfare in Baseline		3.664	
16. (15) + change in occupation share in employment	0.858	1.108	0.774
17. (16) + change in local technology ex real estate	0.852	1.114	0.765
18. (17) + change in real estate productivity	0.862	1.128	0.764
19. (18) + change in amenities (2011-15 parameters)	0.862	1.129	0.764
20. Optimal allocation	0.864	1.132	0.764
21. 2011-15 parameters ex real estate technology	0.846	1.106	0.765
22. Optimal allocation with parameters in (21)	0.848	1.109	0.765

Rows 5 and 15 change the composition of employment to 2011-15 levels, with more CNR workers and fewer non-CNR workers. Here, externalities play their largest role. Absent externalities, the model would imply significant losses for CNR workers, as they become more abundant, whereas standard neo-classical arguments would imply a reduction in their relative wage. In effect, absent externalities, CNR workers would end up with welfare 14% below the baseline counterfactual, while non-CNR worker's welfare would grow by 11%. In contrast, with externalities, both CNR and non-CNR workers end up gaining about the same, since CNRs segregate which improves their productivity.

Rows 6 and 16 add exogenous changes to local technology of non-real estate sectors over and above what is implied by average national trends. It captures, for example, the fact that computer and electronics output became particularly more productive in San Jose and finance particularly more productive in New York. As shown in Figure 16 those gains were larger in cities that had high CNR share in 1980. These location specific technological changes interact with externalities to increase the welfare of both types of workers.

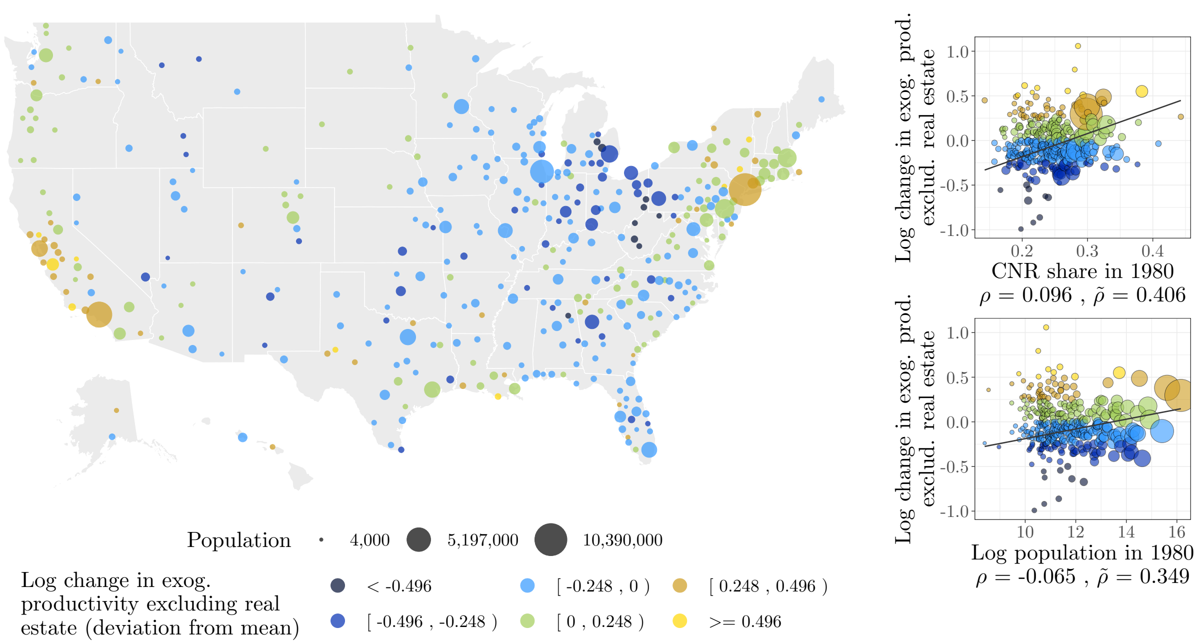


Figure 16: Change in the exogenous part of technology (all non-real estate sectors)

Changes are relative to baseline counterfactual, averaged across all sectors except real estate. Each observation refers to a CBSA. Marker sizes are proportional to total city employment. Averages are taken with value added weights. ρ and $\tilde{\rho}$ are unweighted and population weighted correlations respectively.

Rows 7 and 17 add changes in the productivity of the real estate sector. This exercise encompasses the effects of two different underlying processes. On the one hand, real estate productivity increased more in fast growing cities, as the stock of housing was increased in order to accommodate rising populations. On the other hand, as has been increasingly recognized (Glaeser and Gyourko (2018), Hsieh and Moretti (2019) and Herkenhoff et al. (2018)), housing regulations have impeded development in some very productive areas. The table shows that the net effect of those two forces would have been positive in the absence of externalities, as housing development may have accommodated increasing population in high growing locations. At the same time, their effect is negative once externalities are accounted for, since, as shown in Figure 17, housing productivity also lagged behind in CNR intensive cities.

Finally, Rows 8 and 18 add the changes in amenities. In particular, Row 8 corresponds to the 2011-15 equilibrium allocation. Changes in the spatial distribution of amenities appear to add little to total welfare. Of course, in considering the results presented in Table 8 is important to be aware that the particular sequence in which we added the changes between 1980 and 2011-15 can have an effect on our results. We chose to present a sequence that is intuitive to us, but the main findings highlighted above are robust to other sequences.

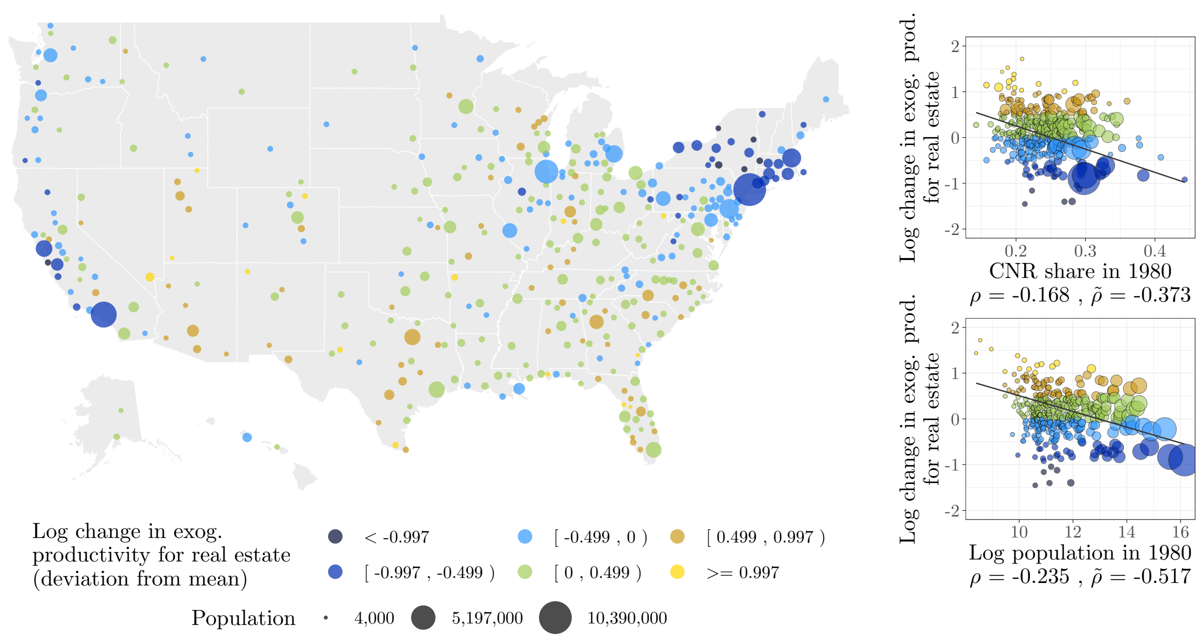


Figure 17: Change in the exogenous component of technology in real estate.

Changes are relative to baseline counterfactual. ρ and $\tilde{\rho}$ are unweighted and population weighted correlations respectively.

We calculate optimal allocation under two scenarios. The first, depicted in Rows 9 and 20, corresponds, when externalities are included, to the one depicted in Section 5 above. The second assesses the role of housing policy in impeding optimal policy. In particular, we calculate the optimal policy under the assumption that housing productivity was distributed as in 1980 (the corresponding equilibrium counterfactual welfare is presented in Rows 10 and 21 and the corresponding optimal allocation in Rows 11 and 22). We find that the increment in welfare is more than 50% larger in that scenario than when starting from the actual equilibrium. This implies that policy can be less effective as a result of the observed changes in housing supply restrictions in some of the cognitive hubs.

7 Conclusion

Our aim is to understand the extent to which workers are misallocated in space and the policies to improve the observed allocation. The main culprit of spatial misallocation is the existence of large occupation-specific externalities, combined with potential distortions due to land use regulations. Our quantified spatial model allowed us to measure occupation specific local productivities by industry and use them, together with a relatively standard instrumental variable approach, to estimate these externalities for CNR and non-CNR occupations. Our estimates suggest that both CNR and non-CNR workers become more productive in large cities, but CNR productivity improves particularly when CNR workers are surrounded by other CNR workers. These estimates, together with estimated local amenities by occupation, as well as exogenous productivity differences across industries and locations, and the full input-output linkages and transport cost in the U.S. economy, determine the current allocation of economic activity. Policy can improve this allocation for both occupations, although only by 0.38%. Housing and optimal transfer policies reinforce each other. Hence, combining them (by reverting the spatial distribution of real estate productivity to that of 1980) leads to welfare gains of 2% for CNRs and 1.6% for non-CNRs.

Since the 80's the U.S. economy has experienced increased skill and occupational polarization. Large cities increasingly have more highly educated CNR workers that earn more. In contrast, many medium and small cities have suffered an exodus of skilled workers and experienced persistent population declines. These trends, amplified by local externalities, were also associated with a rise in income inequality between occupations. This growing gap between top and medium and small-sized cities has motivated policymakers and city governments to advocate policies to attract CNR workers to smaller towns in order to reverse their fortunes. Our analysis shows that these efforts are problematic.

CNR workers are extremely useful, but scarce. Furthermore, their productivity is tremendously enhanced by living with other CNR workers. So attracting them to smaller towns with more mixed populations is a waste of resources. CNR workers are too valuable for society to be used in this way. A better policy is to reinforce existing trends and let them concentrate in cognitive hubs with only a few non-CNR workers. Of course, some non-CNR workers will always be needed due to imperfect substitutability in production. The result is smaller, more CNR intensive, cognitive hubs in the largest cities. The resulting migration of non-CNR workers to small towns, makes these towns grow. We show that this policy can be implemented with a baseline transfer to non-CNR workers, reminiscent of a universal basic income, and a set of occupation-location specific transfers that grow with city size for CNR workers and decline with city size for non-CNRs. Overall, CNR workers transfer resources to non-CNR workers to generate equal welfare gains.

Our analysis suggest that efforts to stop spatial occupational segregation are misguided. In fact, encouraging it further can yield benefits for everyone in the presence of the necessary transfers. Implementing this transfers is important, though. Otherwise, cognitive hubs might use other indirect means to push out non-CNR workers like, for example, housing supply constrains and zoning restrictions, or lack of investments in transportation networks to aid commuting. Such efforts can generate occupational polarization without Pareto gains for all workers. Implementing the necessary transfers will not only help avoid those inefficient policies and benefit CNR workers, but it will also improve the lives of non-CNR workers and the many small and medium sized cities where they will end up living, working, and producing.

References

Acemoglu, D. and D. Autor (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of labor economics*, Volume 4, pp. 1043–1171. Elsevier. 2

Albouy, D. (2009). The unequal geographic burden of federal taxation. *Journal of Political Economy 117*(4), 635–667. 1

Albouy, D., K. Behrens, F. Robert-Nicoud, and N. Seegert (2019). The optimal distribution of population across cities. *Journal of Urban Economics 110*, 102–113. 1

Altonji, J. G., T. E. Elder, and C. R. Taber (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of political economy 113*(1), 151–184. 4.2.1

Anderson, J. E., C. A. Milot, and Y. V. Yotov (2014). How much does geography deflect services trade? canadian answers. *International Economic Review 55*(3), 791–818. 4

Angrist, J. D. and J.-S. Pischke (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press. 23

Autor, D. H. et al. (2019). Work of the past, work of the future. In *Richard T. Ely Lecture given at Annual Meeting of American Economics Association, Atlanta, GA, January*, Volume 4. 1

Baum-Snow, N., M. Freedman, and R. Pavan (2018). Why has urban inequality increased? *American Economic Journal: Applied Economics 10*(4), 1–42. 1

Baum-Snow, N. and R. Pavan (2011). Understanding the city size wage gap. *The Review of economic studies 79*(1), 88–127. 7

Baum-Snow, N. and R. Pavan (2013). Inequality and city size. *Review of Economics and Statistics 95*(5), 1535–1548. 1, 2, 14

Berry, C. R. and E. L. Glaeser (2005). The divergence of human capital levels across cities. *Papers in regional science 84*(3), 407–444. 1

Brinkman, J. (2014). The supply and demand of skilled workers in cities and the role of industry composition. 1

Caliendo, L. and F. Parro (2015). Estimates of the trade and welfare effects of nafta. *The Review of Economic Studies 82*(1), 1–44. 4, 11

Caliendo, L., F. Parro, E. Rossi-Hansberg, and P.-D. Sarte (2017). The impact of regional and sectoral productivity changes on the us economy. *The Review of economic studies 85*(4), 2042–2096. 1, 4.1.3, A.2.3

Card, D. (2001). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics 19*(1), 22–64. 1, 4, 4.2.1, 4.2.1

Ciccone, A. and R. E. Hall (1996). Productivity and the density of economic activity. *The American Economic Review 86*(1), 54. 4.2.1, 4.2.1

Ciccone, A. and G. Peri (2005). Long-run substitutability between more and less educated workers: evidence from us states, 1950–1990. *Review of Economics and statistics 87*(4), 652–663. 4

Colas, M., K. Hutchinson, et al. (2017). Heterogeneous workers and federal income taxes in a spatial equilibrium. Technical report. 1

Couture, V., C. Gaubert, J. Handbury, and E. Hurst (2018). Income growth and the distributional effects of urban spatial sorting. Technical report, Working Paper University of Chicago Booth. 2

Desmet, K. and E. Rossi-Hansberg (2013). Urban accounting and welfare. *American Economic Review 103*(6), 2296–2327. 1

Diamond, R. (2016). The determinants and welfare implications of us workers' diverging location choices by skill: 1980-2000. *American Economic Review 106*(3), 479–524. 1, 2, 1, 4.1.4, 19, 5, F

Duranton, G. and D. Puga (2004). Micro-foundations of urban agglomeration economies. In *Handbook of regional and urban economics*, Volume 4, pp. 2063–2117. Elsevier. 1

Eaton, J. and S. Kortum (2002). Technology, geography, and trade. *Econometrica 70*(5), 1741–1779. 3.2.2, A.2.3

Eeckhout, J. and N. Guner (2015). Optimal spatial taxation: Are big cities too small? 1

Ellison, G., E. L. Glaeser, and W. R. Kerr (2010). What causes industry agglomeration? evidence from coagglomeration patterns. *American Economic Review 100*(3), 1195–1213. 1

Fajgelbaum, P. and C. Gaubert (2018). Optimal spatial policies, geography and sorting. Technical report, National Bureau of Economic Research. 1, 1, 3.4, 8, 19, 5, F

Fajgelbaum, P. D., E. Morales, J. C. Suárez Serrato, and O. Zidar (2018). State taxes and spatial misallocation. *The Review of Economic Studies 86*(1), 333–376. 1, 4, 13

Feenstra, R. C., P. Luck, M. Obstfeld, and K. N. Russ (2018). In search of the armington elasticity. *Review of Economics and Statistics 100*(1), 135–150. 11

Giannone, E. et al. (2017). Skill-biased technical change and regional convergence. In *2017 Meeting Papers*, Number 190. Society for Economic Dynamics. 1

Glaeser, E. and J. Gyourko (2018, February). The economic implications of housing supply. *Journal of Economic Perspectives 32*(1), 3–30. 1, 6

Handbury, J. and D. E. Weinstein (2014). Goods prices and availability in cities. *The Review of Economic Studies 82*(1), 258–296. 4, 4.1.2

Head, K. and T. Mayer (2014). Gravity equations: Workhorse, toolkit, and cookbook. In *Handbook of international economics*, Volume 4, pp. 131–195. Elsevier. 4

Hendricks, L. (2011). The skill composition of us cities. *International Economic Review 52*(1), 1–32. 1

Herkenhoff, K. F., L. E. Ohanian, and E. C. Prescott (2018). Tarnishing the golden and empire states: Land-use restrictions and the us economic slowdown. *Journal of Monetary Economics 93*, 89–109. 1, 1, 6

Hsieh, C.-T. and E. Moretti (2019). Housing constraints and spatial misallocation. *American Economic Journal: Macroeconomics*. 1, 1, 6

Jaimovich, N. and H. E. Siu (2018). The trend is the cycle: Job polarization and jobless recoveries. *The Review of Economics and Statistics*. 3

Marshall, A. (1920). *Principles of Economics (Revised ed.)*. London: Macmillan; reprinted by Prometheus Books. 1

Melo, P. C., D. J. Graham, and R. B. Noland (2009). A meta-analysis of estimates of urban agglomeration economies. *Regional science and urban Economics 39*(3), 332–342. 1, 4.1.3, 4

Moretti, E. (2004a). Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data. *Journal of econometrics 121*(1-2), 175–212. 1, 1, 1, 14, 4.2.1, 4.2.1

Moretti, E. (2004b). Workers' education, spillovers, and productivity: evidence from plant-level production functions. *American Economic Review 94*(3), 656–690. 1, 4.1.3, 4

Neal, D. (1997). The effects of catholic secondary schooling on educational achievement. *Journal of Labor Economics 15*(1, Part 1), 98–123. 4.2.1

Neumark, D. and H. Simpson (2015). Place-based policies. In *Handbook of regional and urban economics*, Volume 5, pp. 1197–1287. Elsevier. 1

Ossa, R. (2015). A quantitative analysis of subsidy competition in the us. Technical report, National Bureau of Economic Research. 1

Redding, S. J. and E. Rossi-Hansberg (2017). Quantitative spatial economics. *Annual Review of Economics 9*, 21–58. 1

Roca, J. D. L. and D. Puga (2017). Learning by working in big cities. *The Review of Economic Studies 84*(1), 106–142. 7, 1

Sanderson, E. and F. Windmeijer (2016). A weak instrument f-test in linear iv models with multiple endogenous variables. *Journal of Econometrics 190*(2), 212–221. 4.2.1

Staiger, D., J. H. Stock, et al. (1997). Instrumental variables regression with weak instruments. *Econometrica 65*(3), 557–586. 4.2.1

Stock, J. H. and M. Yogo (2005). Testing for weak instruments in linear iv regression. *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, 80. 4.2.1

Appendix

A Economic Environment

A.1 Individuals

All workers in a given occupation living in a given city will choose the same consumption basket. It follows that $C_n^{kj}(\mathbf{a}) = C_n^{kj}$ for all $\mathbf{a}$. Moreover, the demand for good j by workers in occupation k living in city n is then

$$C_n^{kj} = \alpha^j \frac{P_n}{P_n^j} C_n^k \tag{14}$$

where $P_n = \prod_{j=1}^{J} \left(\frac{P_n^j}{\alpha^j}\right)^{\alpha^j}$ is the ideal price index in city n.

Agents move freely across cities. The value, $v_n^k(\mathbf{a})$, of locating in a particular city n for an individual employed in occupation k, with idiosyncratic preference vector $\mathbf{a}$,

$$v_n^k(\mathbf{a}) = \frac{a_n A_n^k I_n^k}{P_n} = a_n A_n^k C_n^k.$$

In equilibrium, workers move to the location where they receive the highest utility so that

$$v^k(\mathbf{a}) = \max_n v_n^k(\mathbf{a}),$$

where $v^k(\mathbf{a})$ now denotes the equilibrium utility of an individual in occupation k with amenity vector $\mathbf{a}$. We assume that a_n is drawn from a Frechet distribution. Draws are independent across cities. We denote by Ψ^k the joint *cdf* for the elements of $\mathbf{a}$ across workers in occupation k, so that

$$\Psi^k(\mathbf{a}) = \exp\left\{-\sum_n (a_n)^{-\nu}\right\},$$

where the shape parameter ν reflects the extent of preference heterogeneity across workers employed in occupation k. Higher values of ν imply less heterogeneity, with all workers ordering cities in the same way when $\nu \to \infty$.

Assuming that workers of different types can freely move between cities, the average utility of a worker of type k is given by

$$v^k = \Gamma\left(\frac{\nu-1}{\nu}\right)\left(\sum_n \left(A_n^k C_n^k\right)^{\nu}\right)^{\frac{1}{\nu}}, \tag{15}$$

where $\Gamma(.)$ is the Gamma function.

A.2 Firms

A.2.1 Intermediate Goods Producers

Cost minimization implies that input demand satisfies:

$$\frac{r_n H_n^j(\mathbf{z})}{x_n^j(\mathbf{z}) q_n^j(\mathbf{z})} = \gamma_n^j \beta^j, \tag{16}$$

$$\frac{w_n^k L_n^{kj}(\mathbf{z})}{x_n^j(\mathbf{z}) q_n^j(\mathbf{z})} = \frac{\left(\frac{w_n^k}{\lambda_n^{kj}}\right)^{1-\epsilon}}{\sum_{k'}\left(\frac{w_n^{k'}}{\lambda_n^{k'j}}\right)^{1-\epsilon}} \gamma_n^j \left(1-\beta^j\right), \tag{17}$$

$$\frac{P_n^{j'} M_n^{j'j}(\mathbf{z})}{x_n^j(\mathbf{z}) q_n^j(\mathbf{z})} = \gamma_n^{j'j}, \tag{18}$$

where $x_n^j(\mathbf{z})$ is the Lagrange multiplier which in this case reflects the unit cost of production. We can solve for $x_n^j(\mathbf{z})$ by substituting optimal factor choices into the production function,

$$x_n^j(\mathbf{z}) \equiv \frac{x_n^j}{z_n} = \frac{B_n^j}{z_n}\left\{\frac{r_n^{\beta_n^j}}{Z_n^j}\left[\sum_k \left(\frac{w_n^k}{\lambda_n^{kj}}\right)^{1-\epsilon}\right]^{\frac{1-\beta^j}{1-\epsilon}}\right\}^{\gamma_n^j} \prod_{j'=1}^{J}\left(P_n^{j'}\right)^{\gamma_n^{j'j}} \tag{19}$$

where x_n^j is a city and industry specific unit cost index such that

$$B_n^j = \left[\left(1-\beta_n^j\right)^{\beta_n^j - 1}\left(\beta_n^j\right)^{-\beta_n^j}\right]^{\gamma_n^j}\left[\prod_{j'}\left(\gamma_n^{j'j}\right)^{-\gamma_n^{j'j}}\right]\left(\gamma_n^j\right)^{-\gamma_n^j}.$$

Given constant returns to scale and competitive intermediate goods markets, firms produce positive but finite amounts of a variety if its price is equal to its unit production cost,

$$p_n^j(\mathbf{z}) = x_n^j(\mathbf{z}) = \frac{x_n^j}{z_n}. \tag{20}$$

A.2.2 Final Goods

Let $Q_n^j(\mathbf{z}) = \sum_{n'} Q_{nn'}^j(\mathbf{z})$ denote the total amount of intermediate goods of variety $\mathbf{z}$ purchased from different cities by a final goods producer in city n, sector j. Given that intermediate goods of a given variety produced in different cities are perfect substitutes, final goods producers purchase varieties only from cities that offer the lowest unit cost,

$$Q_{nn'}^j(\mathbf{z}) = \begin{cases} Q_n^j(\mathbf{z}) & if\ \kappa_{nn'}^j p_{n'}^j(\mathbf{z}) < \min_{n'' \neq n'} \kappa_{nn''}^j p_{n''}^j(\mathbf{z}) \\ 0 & \text{otherwise} \end{cases},$$

where we abstract from the case where $\kappa_{nn'}^j p_{n'}^j(\mathbf{z}) = \min_{n'' \neq n'} \kappa_{nn''}^j p_{n''}^j(\mathbf{z})$ since, given the distributional assumption on $\mathbf{z}$, this event only occurs on a set of measure zero.

Denote by $P_n^j(\mathbf{z})$ the unit cost paid by a final good producer in city n and sector j for a particular variety whose vector of productivity draws is $\mathbf{z}$. Given that final goods firms only purchase intermediate goods from the lowest cost supplier,

$$P_n^j(\mathbf{z}) = \min_{n'} \left\{ \kappa_{nn'}^j p_{n'}^j(\mathbf{z}) \right\} = \min_{n'} \left\{ \frac{\kappa_{nn'}^j x_{n'}^j}{z_{n'}} \right\}. \tag{21}$$

For non-tradable intermediate goods, firms must buy those goods locally, so that if j is non-tradable,

$$P_n^j(\mathbf{z}) = \frac{x_n^j}{z_n}. \tag{22}$$

Then, the demand function for intermediate goods of variety $\mathbf{z}$ in industry j and city n is given by

$$Q_n^j(\mathbf{z}) = \left(\frac{P_n^j(\mathbf{z})}{\tilde{P}_n^j} \right)^{-\eta} Q_n^j, \tag{23}$$

where $\tilde{P}_n^j$ the ideal cost index for final goods produced in sector j in city n,

$$\tilde{P}_n^j = \left[\int P_n^j(\mathbf{z})^{1-\eta} \, d\Phi^j(\mathbf{z}) \right]^{\frac{1}{1-\eta}}. \tag{24}$$

Since the production function for final goods is constant returns to scale and the market for final goods is competitive, final goods firms will produce positive but finite quantities of a final good if its price is equal to its cost index, that is if $P_n^j = \tilde{P}_n^j$.

A.2.3 Derivation of Prices

We follow Eaton and Kortum (2002) in solving for the distribution of prices. Given this distribution and zero profits for final goods producers, when sector j is tradable, the price of final goods in sector j in region n solves

$$\left(P_n^j\right)^{1-\eta} = \int P_n^j(\mathbf{z})^{1-\eta} \, d\Phi^j(\mathbf{z}) \, d\mathbf{z},$$

which is the expected value of the random variable $P_n^j(\mathbf{z})^{1-\eta}$.

Let $P_{nn'}^j(\mathbf{z}) = \frac{\kappa_{nn'}^j x_{n'}^j}{z_{n'}}$ denote the unit cost of a variety indexed by $\mathbf{z}$ produced in city n' and sold in n. Following the steps described in Caliendo et al. (2017), we have that

$$\Pr\left[P_{nn'}^j(\mathbf{z}) \le p\right] = 1 - e^{-\omega_{nn'}^j p^\theta}$$

where $\omega_{nn'}^j = \left[\kappa_{nn'}^j x_{n'}^j\right]^{-\theta_j}$. The price of variety $\mathbf{z}$ in city n and industry j, $P_n^j(\mathbf{z})$, is the minimum across $P_{nn'}^j(\mathbf{z})$. Its *cdf* is,

$$\Pr\left[P_n^j(\mathbf{z}) \le p\right] = 1 - e^{-\Omega_n^j p^\theta},$$

where $\Omega_n^j = \sum_{n'} \omega_{nn'}^j = \sum_{n'} \left[\kappa_{nn'}^j x_{n'}^j\right]^{-\theta}$ (Ω_n^j does not depend on n' because we are integrating out the city dimension).

Let $F_{P_n^j}(p)$ denote the *cdf* of $P_n^j(\mathbf{z})$, $\Pr\left[P_n^j(\mathbf{z}) \leq p\right]$. Then, its associated *pdf*, denoted $f_{P_n^j}(p)$, is $\Omega_n^j \theta p^{\theta-1} e^{-\Omega_n^j p^\theta}$. As in Caliendo et al. (2017), we have that

$$P_n^j = \Gamma(\xi)^{\frac{1}{1-\eta}} \left(\Omega_n^j\right)^{-\frac{1}{\theta}},$$

where $\Gamma(\xi)$ is the Gamma function evaluated at $\xi = 1 + \frac{1-\eta}{\theta}$. The price of goods in tradable sector j may then also be expressed as

$$P_n^j = \Gamma(\xi)^{\frac{1}{1-\eta}} \left[\sum_{n'=1}^{N} \left[\kappa_{nn'}^j x_{n'}^j\right]^{-\theta}\right]^{-\frac{1}{\theta}}.$$

In a given non-tradable sector j, $\kappa_{nn'}^j = \infty \; \forall n' \neq n$, so that the equation reduces to

$$P_n^j = \Gamma(\xi)^{\frac{1}{1-\eta}} x_n^j.$$

A.2.4 Trade Shares

Let X_n^j denote *total expenditures* on final goods j by city n, which must equal of the value of final goods in that sector, $X_n^j = P_n^j Q_n^j$. Recall that because of zero profits in the final goods sector, total expenditures on intermediate goods in a given sector are then also equal to the cost of inputs used in that sector, so that $P_n^j Q_n^j = \int P_n^j(\mathbf{z}) Q_n^j(\mathbf{z}) d\Phi^j(\mathbf{z})$. Let $X_{nn'}^j = \int \kappa_{nn'}^j p_{n'}^j(\mathbf{z}) Q_{nn'}^j(\mathbf{z}) d\Phi^j(\mathbf{z})$ denote the value spent by city n on intermediate goods of sector j produced in city n'. Further, let $\pi_{nn'}^j$ denote the share of city n's expenditures on sector j goods purchased from region n'. Then,

$$\pi_{nn'}^j = \frac{X_{nn'}^j}{X_n^j}.$$

Observe that, since there is a continuum of varieties of intermediate goods, the fraction of goods that firms in city n purchase from firms in city n' is given by

$$\tilde{\pi}_{nn'}^j \equiv \Pr\left[P_{nn'}^j(\mathbf{z}) \leq \min_{n'' \neq n'} \left\{P_{nn''}^j(\mathbf{z})\right\}\right].$$

Following the steps described in Caliendo et al. (2017), we have that

$$\begin{aligned}
\tilde{\pi}_{nn'}^j &= \frac{\omega_{nn'}^j}{\Omega_n^j} \\
&= \frac{\left[\kappa_{nn'}^j x_{n'}^j\right]^{-\theta}}{\sum_{n''=1}^{N} \left[\kappa_{nn''}^j x_{n''}^j\right]^{-\theta}}
\end{aligned}$$

We can verify that $\tilde{\pi}_{nn'}^j = \pi_{nn'}^j$, that is, the share of goods that firms in city n purchase from city n' is equal to the share of the *value* of goods produced in city n' in the bundle purchased by firms in city n (see Eaton and Kortum (2002), Footnote 17). Observe also that $\sum_{n'=1}^{N}\left[\kappa_{nn'}^{j}x_{n'}^{j}\right]^{-\theta} = \left(P_{n}^{j}\right)^{-\theta}\Gamma\left(\xi\right)^{\frac{\eta}{1-\eta}}$. Therefore, we may alternatively write the trade share $\pi_{nn'}^{j}$ as

$$\pi_{nn'}^{j} = \frac{X_{nn'}^{j}}{X_{n}^{j}} = \left[\frac{\kappa_{nn'}^{j}x_{n'}^{j}\Gamma\left(\xi\right)^{\frac{1}{1-\eta}}}{P_{n}^{j}}\right]^{-\theta}$$

In non-tradable sectors, $\pi_{nn}^{j} = 1$.

A.3 Market Clearing and Aggregation at the Industry and City Level

Given the labor supply equation (2) and the definition $L_{n}^{kj} = \int L_{n}^{kj}(\mathbf{z})d\Phi^{j}(\mathbf{z})$, the labor market clearing equation (5) may be rewritten as

$$\sum_{j} L_{n}^{kj} = L^{k}\frac{\left(A_{n}^{k}C_{n}^{k}\right)^{\nu}}{\sum_{n'}\left(A_{n'}^{k}C_{n'}^{k}\right)^{\nu^{k}}},\ \forall\ n = 1,...,N,\ k = 1,...,K.$$

Given the definition $H_{n}^{j} = \int H_{n}^{j}(\mathbf{z})d\Phi^{j}(\mathbf{z})$, the market clearing equation for structures in each city (6) may be rewritten as

$$\sum_{j} H_{n}^{j} = H_{n},\ n = 1,...,N.$$

Given our definition of total final expenditures, $X_{n}^{j} = P_{n}^{j}Q_{n}^{j}$, and the demand function for consumption goods of sector j (14), the market clearing condition for final goods in each city n and sector j (7) may be expressed in terms of sectoral and city aggregates,

$$\sum_{k} L_{n}^{k}\left(\alpha^{j}P_{n}C_{n}^{k}\right) + P_{n}^{j}\sum_{j'} M_{n}^{jj'} = X_{n}^{j}.$$

Finally, given that $\pi_{n'n}^{j}X_{n'}^{j} = X_{n'n}^{j} = \int p_{n}^{j}(\mathbf{z})\kappa_{n'n}^{j}Q_{n'n}^{j}(\mathbf{z})d\Phi^{j}(\mathbf{z})$, the market clearing condition for intermediate inputs (8) may be rewritten in terms of sectoral city aggregates as

$$\underbrace{\int p_{n}^{j}(\mathbf{z})q_{n}^{j}(\mathbf{z})d\Phi^{j}(\mathbf{z})}_{\text{Total value of intermediate goods produced in city } n} = \sum_{n'} \pi_{n'n}^{j}X_{n'}^{j},$$

where $\sum_{n'}\pi_{n'n}^{j}X_{n'}^{j}$ is the total value of expenditures across all cities spent on intermediate goods produced in city n.

We can use this last aggregation relationship to obtain aggregate factor input demand equations as follows,

$$
\begin{aligned}
w_n^k L_n^{kj} &= \gamma_n^j \left(1-\beta^j\right) \frac{\left(\frac{w_n^k}{\lambda_n^{kj}}\right)^{1-\epsilon}}{\sum_{k'=1}^{K}\left(\frac{w_n^{k'}}{\lambda_n^{k'j}}\right)^{1-\epsilon}} \sum_{n'} \left(\pi_{n'n}^j X_{n'}^j\right), \\
r_n H_n^j &= \gamma_n^j \beta_n^j \sum_{n'} \left(\pi_{n'n}^j X_{n'}^j\right), \\
P_n^{j'} M_n^{jj'} &= \gamma_n^{j'j} \sum_{n'} \left(\pi_{n'n}^j X_{n'}^j\right).
\end{aligned}
$$

Finally, combining these factor demand equations yields the aggregate production function,

$$
\sum_{n'} \pi_{n'n}^j X_{n'}^j = x_n^j \left[Z_n^j \left(\sum_k \left(\lambda_n^{kj} L_n^{kj}\right)^{1-\frac{1}{\epsilon}} \right)^{\frac{\epsilon}{\epsilon-1}(1-\beta_n^j)} \left(H_n^j\right)^{\beta_n^j} \right]^{\gamma_n^j} \prod_{j'} \left(M_n^{j'j}\right)^{\gamma_n^{j'j}}.
$$

A.4 Definition of Equilibrium

Equilibrium for this system of cities is given by a set of final goods prices P_n^j, wages to different occupations, w_n^k, rental rates, r_n, intermediate goods prices paid by final goods producers, $P_n^j(\mathbf{z})$, intermediate goods prices received by intermediate goods producers, $p_n^j(\mathbf{z})$, consumption choices, C_n^{kj}, intermediate input choices, $Q_n^j(\mathbf{z})$, intermediate input production, $q_n^j(\mathbf{z})$, demand for materials, $M_n^{jj'}(\mathbf{z})$, labor demand, $L_n^{kj}(\mathbf{z})$, demand for structures, $H_n(\mathbf{z})$, and location decisions, $\zeta_n^k(\mathbf{a})$, such that:

i) Workers choose consumption of each final good optimally, as implied by equation (14) and the budget constraint, $\sum_j P_n^j C_n^{kj} = P_n C_n^k = I_n^k$, where $P_n = \prod_j \left(\frac{P_n^j}{\alpha^j}\right)^{\alpha^j}$ and I_n^k is given by equation (1).

ii) Workers choose optimally where to live as implied by equation (2).

iii) Intermediate input producers choose their demand for materials, labor and structures optimally (as implied by factor demand equations (16), (17) and (18)), and produce positive but finite amounts only if (20) holds, where x_n^j in that equation is given by (19).

iv) Final goods producers choose the origin of intermediate inputs optimally, implying that a producer in city n and industry j imports a variety $\mathbf{z}$ from city n' if and only if $\kappa_{nn'}^j p_n^j(\mathbf{z}) = \min_{n''}\left\{\kappa_{nn''}^j p_{n''}^j(\mathbf{z})\right\}$. The price that they pay for intermediate goods satisfies (21) if the good is tradable and (22) if it is non-tradable.

v) Final goods producers choose their intermediate input use optimally according to (23) and produce positive but finite amounts only if (24) holds.

vi) Market clearing conditions for employment (equation 5), land and structures (equation 6), final goods (equation 7), and intermediate goods (equation 8) hold.

A.5 Aggregate Equilibrium

At the aggregate level, equilibrium is given by values for the prices P_n, P_n^j, x_n^j, r_n, w_n^k, aggregate quantities C_n^k, L_n^{kj}, H_n^j, $M_n^{j'j}$, expenditures, X_n^j, and expenditure shares, $\pi_{nn'}^j$, that satisfy the following equations

$$\sum_{k,j'} L_n^{kj'}\left(\alpha^j P_n C_n^k\right) + \sum_{j'} P_n^j M_n^{jj'} = X_n^j \ \ (NJ \text{ eqs.}) \tag{25}$$

$$L_n^k = \sum_j L_n^{kj} = \frac{\left(A_n^k C_n^k\right)^{\nu^k}}{\sum_{n'}\left(A_{n'}^k C_{n'}^k\right)^{\nu}} L^k \ \ (NK \text{ eqs.}) \tag{26}$$

$$\sum_j H_n^j = H_n \ \ (N \text{ eqs.}) \tag{27}$$

$$P_n = \prod_j \left(\frac{P_n^j}{\alpha^j}\right)^{\alpha^j} \ \ (N \text{ eqs.}) \tag{28}$$

$$P_n C_n^k = w_n^k + \frac{\sum_{n'} r_{n'} H_{n'}}{\sum_{k',n',j} L_{n'}^{k'j}} \ \ (NK \text{ eqs.}) \tag{29}$$

$$w_n^k L_n^{kj} = \frac{\left(\frac{w_n^k}{\lambda_n^{kj}(\mathbf{L}_n)}\right)^{1-\epsilon}}{\sum_{k'}\left(\frac{w_n^{k'}}{\lambda_n^{k'j}(\mathbf{L}_n)}\right)^{1-\epsilon}} \gamma_n^j \left(1-\beta_n^j\right) \sum_{n'} \pi_{n'n}^j X_{n'}^j \ (NKJ \text{ eqs.}) \tag{30}$$

$$r_n H_n^j = \gamma_n^j \beta^j \sum_{n'} \pi_{n'n}^j X_{n'}^j \ \ (NJ \text{ eqs.}) \tag{31}$$

$$P_n^{j'} M_n^{j'j} = \gamma_n^{j'j} \sum_{n'} \pi_{n'n}^j X_{n'}^j \ \ \left(NJ^2 \text{ eqs.}\right) \tag{32}$$

$$P_n^j = \begin{cases} \Gamma(\xi)^{\frac{1}{1-\eta}} \left(\sum_{n'}\left[\kappa_{nn'}^j x_{n'}^j\right]^{-\theta}\right)^{-\frac{1}{\theta}} & \text{if } j \text{ is tradable} \\ \Gamma(\xi)^{\frac{1}{1-\eta}} x_n^j & \text{if } j \text{ is non-tradable} \end{cases} \quad (NJ \text{ eqs.}) \tag{33}$$

$$\begin{aligned} &\sum_{n'} \pi_{n'n}^j X_{n'}^j \\ = \ & x_n^j \left[Z_n^j \left(\sum_k \left(\lambda_n^{kj}(\mathbf{L}_n) L_n^{kj}\right)^{1-\frac{1}{\epsilon}} \right)^{\frac{\epsilon}{\epsilon-1}(1-\beta_n^j)} \left(H_n^j\right)^{\beta_n^j} \right]^{\gamma_n^j} \prod_{j'} \left(M_n^{j'j}\right)^{\gamma_n^{j'j}} \ \ (NJ \text{ eqs.}) \end{aligned} \tag{34}$$

$$\pi_{nn'}^j = \frac{\left[\kappa_{nn'}^j x_{n'}^j\right]^{-\theta}}{\sum_{n''}\left[\kappa_{nn''}^j x_{n''}^j\right]^{-\theta}} \ \ (N^2 J \text{ eqs.}) \tag{35}$$

This system of equations comprises $2N + 2NK + 4NJ + 2NKJ + NJ^2 + N^2J$ equations in the same number of unknowns.

By substituting equation (32) into equation (25), adding over all industries (j) and all cities (n) and rearranging, we arrive at the National Accounting identity stating that aggregate value added is equal to aggregate consumption expenditures in the economy,

$$\sum_{n,k,j} L_n^{kj} P_n C_n^k = \sum_{n,j} \gamma_n^j X_n^j \tag{36}$$

At the same time, multiplying both sides of equation (29) by L_n^k, adding over city (n) and occupation (k), and substituting out w_n^k and $r_n H_n$ using equations (27), (30) and (31), yields the same national accounting identity. The fact that we can arrive at that same identity by manipulating different sets of equations implies that there is one redundant equation in the system, leading to one too many unknowns relative to the number of equations. The presence of a redundant equation is a feature of Walrasian systems. In order to pin down the price level, therefore, we need to amend the system with an additional equation defining the numeraire. Specifically, we set a weighted average of final goods prices to 1,

$$\sum_{n,j} \varpi_n^j P_n^j = 1, \tag{37}$$

where ϖ_n^j is a set of weights.

Finally, observe that if we substitute the factor demand equations (30), (31), (32) into (34), we obtain the expression for the unit cost index,

$$x_n^j = B_n^j \left\{ \frac{r_n^{\beta_n^j}}{Z_n^j} \left[\sum_{k=1}^{K} \left(\frac{w_n^k}{\lambda_n^{kj}} \right)^{1-\epsilon} \right]^{\frac{1-\beta_n{}^j}{1-\epsilon}} \right\}^{\gamma_n^j} \prod_{j'=1}^{J} \left(P_n^{j'} \right)^{\gamma_n^{j'j}}. \tag{38}$$

B Quantifying the Model and Inversion

We now provide additional detail on how we perform the model quantification. The set of parameters needed to quantify our framework fall into broadly two types: i) parameters that are constant across cities (but may vary across occupations and/or industries) and that are directly available from statistical agencies, or that may be chosen to match national or citywide averages, and ii) parameters that vary at a more granular level and require using all of the model's equations (i.e. by way of model inversion) to match data that vary across cities, industries, and occupations.

B.1 Parameters That Are Constant Across Cities

1. γ_n^j, $\gamma_n^{jj'}$, *Input use shares in gross output*: To obtain an initial calibration for these share parameters, we use an average of the 2011 to 2015 BEA Use Tables, each adjusted by the same year's total gross output. The Use Table divides the value of the output in each sector j, $\sum_{n',n} \pi_{n'n}^j X_{n'}^j = \sum_n X_n^j$, into the value of input purchases from other sectors j', $\sum_{n,j'} P_n^{j'} M_n^{j'j}$, labor compensation, $\sum_{n,k} w_n^k L_n^{kj}$, operational surplus, $\sum_n r_n H_n^j$, and taxes on production and imports, $-\sum_n s_n^j X_n^j$,

$$\sum_n X_n^j = \sum_{n,j'} P_n^{j'} M_n^{j'j} + \sum_{n,k} w_n^k L_n^{kj} + \sum_n r_n H_n^j - \sum_n s_n^j X_n^j. \tag{39}$$

Input purchases from other sectors are separated into purchases from domestic producers and purchases from international producers. Since the model does not allow for foreign trade, we adjust the Use Table by deducting purchases from international producers from the input purchases and, for accounting consistency, from the definition of gross output for the sector.

As in Caliendo et al. (2018), for all sectors, we augment material purchases to include the purchases of equipment. Specifically, we subtract from the operational surplus of each sector 17 percent of their value added and then add the same value back to materials.[27] This 17 percent value is estimated by Greenwood et al. (1997) as the equipment share in output. We then pro-rate the equipment share of value added to different materials in proportion to their use within each sector.

We interpret the remaining part of the gross operational surplus in a given sector as compensation for services provided by real estate. We adopt the convention that all land and structures are managed by firms in the real estate sector, which then sell their services to other sectors. Accordingly, for all sectors other than real estate, we reassign the gross operating surplus remaining, after deducting equipment investment, to purchases from the real estate sector. These surpluses are in turn added to the gross operating surplus of Real Estate.

It follows that, for all sectors j other than real estate,

$$\sum_n r_n H_n^j = 0.$$

and in each of those sectors,

$$\begin{aligned} P_n^{\text{Real Estate}} M_n^{\text{Real Estate},j} &= \text{Purchases from real estate by } j \\ &\quad + \text{Operational Surplus of } j \\ &\quad - \text{Equipment Investment by } j. \end{aligned}$$

In contrast, in the real estate sector,

$$\begin{aligned} \sum_n r_n H_n^{\text{real estate}} &= \text{Total Operational Surplus across all } j \\ &\quad - \text{Total Equipment Investment across all } j. \end{aligned}$$

One can verify that those reassignments do not affect aggregate operational surplus (net of equipment investment), aggregate labor compensation, and aggregate value added (net of equipment investment).

We assume that tradable sectors have a $\gamma_n^j = \gamma^j$, constant across cities and similarly for $\gamma_n^{jj'}$'s. The two non-tradable sectors have city specific parameters. Given these adjustments to the Use Table, the share parameters for tradable sectors follow immediately,

[27] When gross operation surplus amounts to less than 17 percent of value added, the entire operational surplus is deducted.

$$\gamma^j = \frac{\sum_{n,k} w_n^k L_n^{kj} + \sum_n r_n H_n^j}{\sum_n (1+s_n^j) X_n^j}, \ \gamma^{j'j} = \frac{\sum_n P_n^{j'} M_n^{j'j}}{\sum_n (1+s_n^j) X_n^j}. \tag{40}$$

Furthermore, we have that, for the non-tradable sectors,

$$\gamma_n^j = \frac{\sum_k w_n^k L_n^{kj} + r_n H_n^j}{(1+s_n^j) X_n^j}, \tag{41}$$

where s_n^j is an ad-valorem subsidy for city n, sector j, which we introduce to account for the fact that part of the sectoral value added calculated by the BEA is in fact paid out in indirect taxes. Finally, since we do not observe use of materials by individual sectors in each city, we assume that, the proportions of materials used in each city by nontradable sectors if fixed at the national averages $\frac{\gamma_n^{j'j}}{1-\gamma_n^j}$ is the same for all cities and satisfies equals $\frac{\sum_n P_n^{j'} M_n^{j'j}}{\sum_{n,j'} P_n^{j'} M_n^{j'j}}$

$$\frac{\gamma_n^{j'j}}{1-\gamma_n^j} = \hat{\gamma}^{j'j} = \frac{\sum_n P_n^{j'} M_n^{j'j}}{\sum_{n,j'} P_n^{j'} M_n^{j'j}}$$

The calibration of γ_n^j and, therefore of $\gamma_n^{j'j}$, will require choosing additional parameters as described below but consistent with the above equations.

B.2 Model Inversion for the Granular Parameters

Data is available from the ACS pertaining to w_n^K, and $\frac{L_n^{kj}}{\sum_{k'} L_n^{k'j}}$. The Census County Business Patterns (CBP) provide us measures of total employment $\sum_{k'} L_n^{k'j}$ that better match BEA industry-level counts, which we combine with the ACS data to obtain L_n^{kj}. From the BEA, we obtain regional price indices P_n (of dimension $N-1$ since these are only defined in relative terms), which include imputed prices for housing services, $P_n^{\text{real estate}}$ (of dimension $N-1$), and sectoral price indices, P^j (of dimension $J-1$).

The data above consists of $NK + NKJ + 2(N-1) + J - 1$ independent observations that allows us to solve for NK values for amenity parameters, A_n^k, NKJ scaling factors in production, $T_n^{kj} \equiv (H_n^j)^{\gamma_n^j \beta_n^j} \left(\lambda_n^{kj}\right)^{\gamma_n^j (1-\beta_n^j}$, $(N-1)$ shares of non-residential structures in value added in the real estate sector, $\beta_n^{\text{Real Estate}}$, $(N-1)$ shares of value added in non-tradable output, and $J-1$ independent values for consumption share parameters, α^j.[28]

The steps below describe the model inversion.

1. *Computing consumption shares,* α^j**.** We first add up equation (25) across n and j. We then use the factor demand equations (30) and (31) to obtain $\gamma_n^j X_n^j = \left(\sum_n r_n H_n^{j'} + \sum_{n,k} w_n^k L_n^{kj'}\right)$, and the national accounting identity (36) to substitute

[28] In principle, this leaves us two equations short of the number necessary to estimate all of the parameters. One additional restriction, or equation, that must be satisfied is that the national share of income from land and structures in the production of real estate be equal to that implied by the Use Tables. Furthermore, additional restrictions imposed on α^j and β^j, specifically that $\alpha_j \in [0,1]$, and $\beta^j \in [0,1]$, actually imply some overidentifying restrictions.

out X_n^j's and C_n^k's from the aggregated equation (25):

$$\alpha^j \sum_{j'} \left(\sum_n r_n H_n^{j'} + \sum_{n,k} w_n^k L_n^{kj'} \right) + \sum_{j'} P_n^j M_n^{jj'} = \left(\sum_n r_n H_n^j + \sum_{n,k} w_n^k L_n^{kj} + \sum_{j'} P_n^{j'} M_n^{j'j} \right),$$

The ACS does not provide data on $r_n H_n^j$, $\sum_{j'} P_n^j M_n^{jj'}$ or their sum across cities. While the BEA provides data on sectoral aggregates, those cover the whole country, as opposed to only MSA's. Thus, we rely instead on ratios, $\frac{\sum_n r_n H_n^j}{\sum_{n,k} w_n^k L_n^{kj}}$ and $\frac{\sum_{n,j'} P_n^{j'} M_n^{jj'}}{\sum_{n,k} w_n^k L_n^{kj}}$ obtained from the Use Tables, which we can then combine with data on $\sum_{n,k} w_n^k L_n^{kj}$ and the above equation. Specifically,

$$\begin{aligned} & \alpha^j \sum_{n,j',k} w_n^k L_n^{kj'} \left(\frac{\sum_n r_n H_n^{j'}}{\sum_{n,k} w_n^k L_n^{kj'}} + 1 \right) + \sum_{j',n,k} \frac{\sum_{n,j'} P_n^{j'} M_n^{jj'}}{\sum_{n,k} w_n^k L_n^{kj}} w_n^k L_n^{kj'} \\ = & \sum_{n,k} w_n^k L_n^{kj} \left(\frac{\sum_n r_n H_n^j}{\sum_{n,k'} w_n^{k'} L_n^{k'j}} + \frac{\sum_{n,j'} P_n^{j'} M_n^{j'j}}{\sum_{n,k} w_n^{k'} L_n^{k'j}} + 1 \right), \end{aligned}$$

or

$$\begin{aligned} & \alpha^j \sum_{n,j',k} w_n^k L_n^{kj'} \left(\varrho_H^j + 1 \right) + \sum_{j',n,k} w_n^k L_n^{kj'} \varrho_M^{jj'} \\ = & \sum_{n,k} w_n^k L_n^{kj} \left(\varrho_H^j + \varrho_M^{j'j} + 1 \right), \end{aligned}$$

where ϱ_H^j and $\varrho_M^{j'j}$ denote, respectively, the ratio of national aggregate rental income and the ratio of national aggregate material inputs usage from sector j' to national aggregate wage income in sector j which are consistent with the Use Tables. The J equations above can be solved for J values of α^j . One can verify that any value of α^j obtained from those equations will satisfy $\sum_j \alpha^j = 1$. One complicating factor is that in each sector j, α^j must live in $[0, 1]$. However, because of measurement inconsistencies between ACS and BEA data, the procedure generates negative values of α_j in three out of 22 sectors. One of those sectors ("Oil, Chemicals, and Nonmetallic Minerals") indeed has much of its employment located outside of urban areas. We use information from the Use Tables to calibrate α^j in that sector, setting it equal to 5.57 percent. The other two sectors ("Wood, Paper, and Printing", and "Metals") are to a large degree producers of inputs for other industries, so that we set their consumption shares to 0. To ensure that all equations hold while satisfying those restrictions, we allow $\varrho_M^{j'j}$'s to deviate somewhat from those obtained from the Use Tables. This will require adjusting γ^j and $\gamma^{j'j}$ for the tradable sectors, since those satisfy $\gamma^j = \frac{1+\varrho_H^j}{1+\varrho_H^j + \sum^{j'} \varrho_M^{j'j}}$, $\gamma^{j'j} = \frac{\varrho_M^{j'j}}{1+\varrho_H^j + \sum^{j'} \varrho_M^{j'j}}$.

2. *Expressing gross output and rental income from each sector and city as functions of share parameters and wage bills.* Using the labor demand equations (30), we obtain

$$\sum_{n'} \pi^j_{n'n} X^j_{n'} = \frac{\sum_k w^k_n L^{kj}_n}{\left(1-\beta^j_n\right)\gamma^j_n}. \tag{42}$$

In the non-tradable sectors, $\pi^j_{nn} = 1$ and $\pi^j_{n'n} = 0$ for $n' \neq n$ so that

$$X^j_n = \frac{\sum_k w^k_n L^{kj}_n}{\left(1-\beta^j_n\right)\gamma^j_n}.$$

For all sectors other than real estate, we have that $\beta^j_n = 0$, so that $r_n H^j_n = 0$. For the real estate sector, we have from the first-order conditions in that sector that

$$r_n H^{\text{Real Estate}}_n = \frac{\beta^{\text{Real Estate}}_n}{1-\beta^{\text{Real Estate}}_n} \sum_k w^k_n L^{k \text{ real estate}}_n.$$

Since real estate services are the only sector with positive rental income, this is also equal to the total rental income in each city.

3. *Computing the shares of land and structures in value added for the real estate sector,* $\beta^{\text{real estate}}_n$. We use equations (29) to substitute out $P_n C^k_n$ in equations (25). We then apply the relationships from equation (42) to substitute out gross output in (30) to (32), and use the resulting equations to substitute out factor demands in (25). Given that in the non-tradable sectors ("real estate" and "retail, construction, and utilities"), expenditures are equal to gross output, this implies that, for $j \in \{$"real estate", "retail, construction, and utilities"$\}$, we have that

$$\begin{aligned}
&\frac{1}{1-\beta^j_n} \frac{\sum_k w^k_n L^{kj}_n}{\gamma^j_n} \\
&= \alpha^j \sum_k w^k_n L^k_n \\
&+ \alpha^j \frac{L_n}{L} \sum_{n'} \left(\frac{\beta^{\text{Real Estate}}_n}{1-\beta^{\text{Real Estate}}_n} \sum_k w^k_{n'} L^{k,\text{Real Estate}}_{n'} \right) \\
&+ \sum_{j'} \frac{1-\gamma^{j'}_n}{\gamma^{j'}_n \left(1-\beta^{j'}_n\right)} \hat{\gamma}^{jj'} \sum_k w^k_n L^{kj'}_n,
\end{aligned}$$

where we are using the fact that $\beta^j_n = 0$ for all sectors other than real estate. Given that we have two non-tradable sectors, this is a system of $2N$ equations, in N values for γ^j_n and N values of $\beta^{\text{real estate}}_n$.

4. *Computing individual values for nominal expenditures, X^j_n, in tradable sectors.* We use equations (29) to substitute out $P_n C^k_n$ in equations (25). We then apply the relationships from equation (42) to substitute out gross output in (30) to (32), and use the resulting equations to substitute out factor demands in (25). In the tradable sectors, this gives us

$$X_n^j = \alpha^j \sum_k w_n^k L_n^k + \alpha^j \frac{L_n^k}{\sum_{k'} L^{k'}} \sum_{n'} \left(\sum_{j'} \frac{\beta_n^{j'}}{1-\beta_n^{j'}} \sum_k w_{n'}^k L_{n'}^{kj'} \right) + \sum_{j'} \gamma_n^{jj'} \frac{\sum_k w_n^k L_n^{kj'}}{\left(1-\beta_n^{j'}\right)\gamma_n^{j'}}$$

Given values for β_n^j from the previous step, values for X_n^j are then immediately determined from the data.

5. *Computing relative cost indices for tradable goods, $\tilde{x}_n^j$.* For $N(J-2)$ tradable sectors (all but "real estate," as well as "retail, construction, and utilities"), we now solve for $(N-1)(J-2)$ values of the cost index, $\frac{x_n^j}{\sum_{n'} x_{n'}^j}$, for each $j \in \{1, ..., J\}$ from the system of $(N-1)(J-2)$ independent equations,

$$\frac{\sum_k w_n^k L_n^{kj}}{\left(1-\beta_n^j\right)\gamma_n^j} = \sum_{n'=1}^{N} \pi_{n'n}^j \left(\mathbf{x}^j\right) X_{n'}^j,$$

where $\mathbf{x}^j = \{x_1^j, ..., x_N^j\}$ is the vector of unit production costs. This system comprises only $(N-1)(J-2)$ independent equations since, for each j, adding up the right-hand-side and left-hand-side over n gives the same result on both sides irrespective of $\mathbf{x}^j$. At the same time $\pi_{n'n}^j(\mathbf{x}^j)$ is homogeneous of degree 0 in $\mathbf{x}^j$ for each j in equation (35), so that we can still solve for the ratio, $\tilde{x}_n^j \equiv \frac{x_n^j}{\sum_{n'} x_{n'}^j}$.[29]

6. *Computing relative tradable consumer prices, $\widetilde{P}_n^j$, in every sector and city.* Substituting $\tilde{x}_n^j$ from the previous step into equation(33) and rearranging, we have that for the tradable sectors,

$$P_n^j = \Gamma(\xi)^{\frac{1}{1-\eta}} \left(\sum_{n'} \left[\kappa_{nn'}^j \tilde{x}_n^j\right]^{-\theta} \right)^{-\frac{1}{\theta}} \times \sum_{n'} x_{n'}^j,$$

which gives a system of $N(J-2)$ equations. We can thus determine

$$\Xi_P^j \equiv \Gamma(\xi)^{\frac{1}{1-\eta}} \frac{\sum_{n'} x_{n'}^j}{P^j} = \left[\sum \varpi_n^j \left(\sum_{n'} \left[\kappa_{nn'}^j \tilde{x}_n^j\right]^{-\theta} \right)^{-\frac{1}{\theta}} \right]^{-1}$$

[29]Numerically, the system is easier to solve for $\frac{\left(x_n^j\right)^{\theta^j}}{\sum_{n'}\left(x_{n'}^j\right)^{\theta^j}}$ from which we can easily obtain values for x_n^j's.

for each j by imposing $\sum_n \varpi_n^j P_n^j = P^j$, where ϖ_n^j are model-consistent expenditure weights given by $X_n^j / \sum_{n'} X_{n'}^j$ obtained in step 4. We may then then obtain for all tradable j's

$$\tilde{P}_n^j \equiv \frac{P_n^j}{P^j} = \Xi_P^j \left(\sum_{n'} \left[\kappa_{nn'}^j \tilde{x}_n^j \right]^{-\theta} \right)^{-\frac{1}{\theta}}.$$

Note that data on P^j is only available in changes from a base period. Thus, we define the base period to be 2011-2015, our benchmark period, and set $P^j = 1$ in all sectors in that period. For the remainder of the analysis, therefore, $\tilde{P}_n^j = P_n^j$.

7. *Computing non-tradable consumer prices.* In the non-tradable sectors, we have that $P_n^j = \Gamma(\xi)^{\frac{1}{1-\eta}} x_n^j$ for all n and j, and for those sectors, we determine prices based on data from the Regional Price Parity (RPP) indices calculated by the BEA. We directly obtain values for $\Gamma(\xi)^{\frac{1}{1-\eta}} x_n^{\text{real estate}} \equiv P_n^{\text{real estate}}$ from the RPP estimates of the price of real estate services in different cities. For the other non-tradables ("retail, construction, and utilities"), we choose $P_n^{\text{retail, etc.}} = \Gamma(\xi)^{\frac{1}{1-\eta}} x_n^{\text{retail, etc.}}$ so that the price of services (other than real estate) relative to tradable goods in the model to match its counterpart in the RPP. To carry out this calculation, observe that the price index for services can be defined by $P_n^{\text{Services}} = \Pi_{j \in \text{Services}} \left(\frac{\Sigma_{j' \in \text{Services}} \alpha^{j'} P_n^j}{\alpha_j} \right)^{\frac{\alpha^j}{\Sigma_{j' \in \text{Services}} \alpha^{j'}}}$ where the service sectors include retail, etc., wholesale trade, transportation and storage, professional and business services, other, communication, finance and insurance, education, health, and accommodation. The price index for goods can be defined analogously where the goods sector includes all remaining sectors other than real estate.

8. *Computing firm productivity in different sectors, j, and cities, n, associated with occupation k, λ_n^{kj}.* From equations (30), we have that

$$w_n^k L_n^{kj} = \frac{\left(\frac{w_n^k}{\lambda_n^{kj}} \right)^{1-\epsilon}}{\sum_{k'} \left(\frac{w_n^{k'}}{\lambda_n^{k'j}} \right)^{1-\epsilon}} \sum_k w_n^k L_n^{kj},$$

which can rewrite as

$$w_n^k L_n^{kj} = \frac{\left(\frac{\sum_{k'} \lambda_n^{k'j}}{\lambda_n^{kj}} w_n^k \right)^{1-\epsilon}}{\sum_{k'} \left(\frac{\sum_{k'} \lambda_n^{k'j}}{\lambda_n^{k'j}} w_n^{k'} \right)^{1-\epsilon}} \sum_k w_n^k L_n^{kj}.$$

Thus, for each city n and industry j, we can use $K-1$ of those equations to solve for $K-1$ ratios, $\tilde{\lambda}_n^{kj} = \frac{\lambda_n^{kj}}{\sum_{k'} \lambda_n^{k'j}}$. With some rearrangement, we those are equal to

$$\tilde{\lambda}_n^{kj} = \frac{\left(w_n^k\right)^{\frac{\epsilon}{\epsilon-1}} \left(L_n^{kj}\right)^{\frac{1}{\epsilon-1}}}{\sum_{k'} \left(w_n^{k'}\right)^{\frac{\epsilon}{\epsilon-1}} \left(L_n^{k'j}\right)^{\frac{1}{\epsilon-1}}}.$$

From equations (43) (obtained by substituting the factor demand equations (30), (31) and (32) into equations (34)), and the value for $r_n H_n$ obtained in step 2, we obtain

$$H_n^{\gamma_n^j \beta^j} \left(\sum_{k'} \lambda_n^{k' j} \right)^{\gamma_n^j (1-\beta_n^j)} \Gamma(\xi)^{-\frac{1}{1-\eta}} \tag{43}$$
$$= \frac{B_n^j}{\tilde{x}_n^j \Xi_P^j} \left\{ \left(\sum_j \frac{\beta_n^j}{1-\beta_n^j} \sum_k w_n^k L_n^{kj} \right)^{\beta_n^j} \left[\sum_{k=1}^{K} \left(\frac{w_n^k}{\tilde{\lambda}_n^{kj}} \right)^{1-\epsilon} \right]^{\frac{1-\beta_n^j}{1-\epsilon}} \right\}^{\gamma_n^j} \prod_{j'=1}^{J} \left(P_n^{j'} \right)^{\gamma_n^{j' j}},$$

where we set $\Gamma(\xi)^{-\frac{1}{1-\eta}} = 1$ since it is common to all sectors and cities and thus immaterial in any counterfactual exercise. Recall that the use of land and structures as inputs has been folded in the real estate sector that then sells real estate services to all other sectors (i.e. $\beta^j = 0$ in all but real estate). Then, multiplying both sides of equation (43) by the ratios $\left(\tilde{\lambda}_n^{kj}\right)^{\gamma_n^j}$ gives $NK(J-1)$ values for the productivity of firms in different sectors, j, and cities, n, associated with occupation k, $T_n^{kj} \equiv (H_n^j)^{\gamma_n^j \beta_n^j} \left(\lambda_n^{kj}\right)^{\gamma_n^j (1-\beta_n^j)}$, which, in the special case where one abstracts from differences in occupational composition across cities, reproduces measured regional and sectoral productivity in Caliendo et al. (2018).

C The Planner's Problem

This section describes the solution to the planner's problem taking as given that workers in different occupations can freely choose in which city to live. Under this assumption, the expected utility of a worker of type k is given by equation (15). Given welfare weights for each occupation ϕ^k, the utilitarian planner then solves

$$\mathcal{W} = \sum_k \phi^k \Gamma \left(\frac{\nu - 1}{\nu} \right) \left(\sum_{n=1}^{N} \left(A_n^k C_n^k \right)^{\nu} \right)^{\frac{1}{\nu}} L^k, \tag{44}$$

where recall that C_n^k aggregates final goods from different sectors:

$$C_n^k = \prod_j \left(C_n^{kj} \right)^{\alpha^j}, \tag{45}$$

The planner maximizes (44) subject to the resource constraints for final goods,

$$\sum_k L_n^k C_n^{kj} + \sum_{j'} \int M_n^{jj'}(\mathbf{z}) d\Phi(\mathbf{z}) = \left(\int \left[\sum_{n'} Q_{nn'}^j(\mathbf{z}) \right]^{\frac{\eta-1}{\eta}} d\Phi(\mathbf{z}) \right)^{\frac{\eta}{\eta-1}}, \tag{46}$$

where $Q_{nn'}^j(\mathbf{z})$ are the purchases of intermediate goods produced in city n' by final goods firms in city n; the resource constraints for intermediate goods of all varieties $\mathbf{z}$ and industries j produced in all cities n

$$\sum_{n'} Q^j_{n'n}(\mathbf{z})\kappa^j_{n'n} = q^j_n(\mathbf{z}), \quad \forall \mathbf{z} \in \mathbb{R}^+_n, \tag{47}$$

where

$$q^j_n(\mathbf{z}) = z_n \left[H^j_n(\mathbf{z})^{\beta^j_n} \left[\sum_k \left(\lambda^{kj}_n \left(\mathbf{L}_n\right) L^{kj}_n\left(\mathbf{z}\right)\right)^{\frac{\epsilon-1}{\epsilon}} \right]^{\frac{\epsilon}{\epsilon-1}\left(1-\beta^j_n\right)} \right]^{\gamma^j_n} \prod_{j'} M^{j'j}_n(\mathbf{z})^{\gamma^{j'j}_n},$$

the resource constraints for local labor markets,

$$\sum_j \int L^{kj}_n(\mathbf{z}) d\Phi^j(\mathbf{z}) = L^k_n, \tag{48}$$

where labor supply in each city, L^k_n, satisfies

$$L^k_n = \frac{\left(A^k_n C^k_n\right)^\nu}{\sum_{n'} \left(A^k_{n'} C^k_{n'}\right)^\nu} L^k, \tag{49}$$

the resource constraints for the use of land and structures,

$$\sum_j \int H^j_n(\mathbf{z}) d\Phi^j(\mathbf{z}) = H_n, \tag{50}$$

and non-negativity constraints on both household consumption of different goods and input flows:

$$C^{kj}_n \geq 0 \text{ and } Q^j_{n'n}(\mathbf{z}) \geq 0.$$

From the resource constraint on local labor markets (48), and the labor supply condition (49), it follows immediately that national labor markets clear (i.e., $\sum_{n,j} \int L^{kj}_n(\mathbf{z}) d\Phi^j(\mathbf{z}) = L^k$).

C.1 Solving the Planner's Problem

We solve the Planner's problem for interior allocations, (i.e., where C^k_n and L^k_n are strictly greater than zero for all n and k). For each city n and sector j, let P^j_n be the Lagrange multiplier corresponding to the final goods resource constraint in city n, sector j (46), $\widetilde{P}_n$ the multiplier corresponding to the aggregation of sectoral goods for city (45) and $\widetilde{p}^j_n(\mathbf{z})$ the multiplier corresponding to the intermediate goods resource constraints (47). For each city n and occupation k, let w^k_n be the multiplier corresponding to regional labor market clearing (48), W^k_n the multiplier corresponding to the definitions of employment in each occupation and sector (49). Finally, for each city n, let r_n the multiplier corresponding to market clearing for structures (50).

The first-order conditions of the planner's problem are:

$$\partial C^{kj}_n : \ \widetilde{P}_n \alpha^j \frac{C^k_n}{C^{kj}_n} = P^j_n L^k_n, \tag{51}$$

which also defines an ideal price index,

$$P_n = \frac{\widetilde{P}_n}{L_n^k} = \prod_j \left(\frac{P_n^j}{\alpha^j}\right)^{\alpha^j}. \tag{52}$$

In addition,

$$\begin{aligned}
\partial C_n^k \quad &: \quad \phi^k \Gamma\left(\frac{\nu-1}{\nu}\right)\left(\sum_{n'}\left(A_{n'}^k C_{n'}^k\right)^\nu\right)^{\frac{1}{\nu}} \frac{\left(A_n^k C_n^k\right)^\nu}{\sum_{n'}\left(A_{n'}^k C_{n'}^k\right)^\nu} \frac{1}{C_n^k} L^k \\
&= \quad L_n^k P_n - \sum_{n'=1}^N \frac{\partial \zeta_{n'}^k\left(\mathbf{C}^k\right)}{\partial C_n^k} W_{n'}^k,
\end{aligned} \tag{53}$$

where

$$\frac{\partial \zeta_{n'}^k\left(\mathbf{C}^k\right)}{\partial C_n^k} = \begin{cases} \left(\frac{\nu}{C_n^k}\right)\left(1-\frac{L_n^k}{L^k}\right) L_n^k & \text{if } n' = n \\ -\left(\frac{\nu}{C_n^k}\right)\left(\frac{L_{n'}^k}{L^k}\right) L_n^k & \text{if } n' \neq n \end{cases}.$$

Also

$$\partial L_n^k : \sum_{j=1}^J P_n^j C_n^{kj} - \widetilde{w}_n^k + W_n^k = 0. \tag{54}$$

where

$$\begin{aligned}
\widetilde{w}_n^k \quad &= \quad w_n^k \\
&+ \sum_j \int \frac{\partial z_n \left[H_n^j(\mathbf{z})^{\beta_n^j}\left[\sum_{k''}\left(\lambda_n^{k''j}\left(\mathbf{L}_n\right) L_n^{k''j}\left(\mathbf{z}\right)\right)^{\frac{\epsilon-1}{\epsilon}}\right]^{\frac{\epsilon}{\epsilon-1}\left(1-\beta_n^j\right)}\right]^{\gamma_n^j} \prod_{j'=1}^J M_n^{j'j}(\mathbf{z})^{\gamma_n^{j'j}}}{\partial L_n^k} \widetilde{p}_n^j\left(\mathbf{z}\right) d\mathbf{z}
\end{aligned} \tag{55}$$

denotes the total social marginal value of an extra worker of type k in city n. On the production side, efficient allocations dictate

$$\partial Q_{nn'}^j(\mathbf{z}) : \begin{cases} Q_{nn'}^j(\mathbf{z}) > 0 & \text{if } \kappa_{nn'}^j \widetilde{p}_{n'}^j(\mathbf{z}) = P_n^j \left(Q_n^j\right)^{\frac{1}{\eta}}\left[\sum_{n'=1}^N Q_{nn'}^j(\mathbf{z})\right]^{-\frac{1}{\eta}} d\Phi(\mathbf{z}) \\ Q_{nn'}^j(\mathbf{z}) = 0 & \text{if } \kappa_{nn'}^j \widetilde{p}_{n'}^j(\mathbf{z}) > P_n^j \left(Q_n^j\right)^{\frac{1}{\eta}}\left[\sum_{n'=1}^N Q_{nn'}^j(\mathbf{z})\right]^{-\frac{1}{\eta}} d\Phi(\mathbf{z}) \end{cases}. \tag{56}$$

This last equation delivers efficient trade shares, $\pi_{nn'}^j$, and prices, P_n^j, using the usual Eaton and Kortum derivations. Also,

$$\partial L_n^{kj}(\mathbf{z}) : \gamma_n^j(1-\beta_n^j)\frac{q_n^j(\mathbf{z})}{L_n^{kj}\left(\mathbf{z}\right)} \frac{\left(\frac{w_n^k}{\left(\lambda_n^{kj}(\mathbf{L}_n)\right)}\right)^{1-\epsilon}}{\sum_{k'}\left(\frac{w_n^{k'}}{\left(\lambda_n^{kj}(\mathbf{L}_n)\right)}\right)^{1-\epsilon}} \widetilde{p}_n^j(\mathbf{z}) = w_n^k d\Phi(\mathbf{z}), \tag{57}$$

$$\partial H_n^j(\mathbf{z}) : \gamma_n^j \beta_n^j \frac{q_n^j(\mathbf{z})}{H_n^j(\mathbf{z})} \widetilde{p}_n^j(\mathbf{z}) = r_n d\Phi(\mathbf{z}), \tag{58}$$

$$\partial M_n^{j'j}(\mathbf{z}) : \gamma_n^{j'j} \frac{q_n^j(\mathbf{z})}{M_n^{j'j}(\mathbf{z})} \widetilde{p}_n^j(\mathbf{z}) = P_n^{j'} d\Phi(\mathbf{z}). \tag{59}$$

With the usual manipulations of these equations, we can obtain

$$\widetilde{p}_n^j(\mathbf{z}) \equiv p_n^j(\mathbf{z}) d\Phi^j(\mathbf{z}) = \frac{x_n^j d\Phi(\mathbf{z})}{z_n}, \tag{60}$$

where

$$x_n^j = B_n^j \left[r_n^{\beta_n^j} \left[\sum_k \left(\frac{w_n^k}{\left(\lambda_n^{kj}\left(\mathbf{L}_n\right)\right)} \right)^{1-\epsilon} \right]^{\frac{1-\beta_n^j}{1-\epsilon}} \right]^{\gamma_n^j} \prod_{j'} \left(P_n^{j'}\right)^{\gamma_n^{j'j}}, \tag{61}$$

and B_n^j is defined as above.

D Characterization of the Planner's Solution

In the decentralized equilibrium, the budget constraint of household of type k in city n satisfies

$$P_n C_n^k = w_n^k + \chi^k,$$

where $\chi^k = b^k \frac{\sum_{n'} r_{n'} H_{n'}}{\sum_{n',j} L_{n'}^{k,j}}$. In contrast, we now show that the consumption of household k in city n implied by the planner's solution satisfies

$$P_n C_n^k = \frac{\nu}{1+\nu} \widetilde{w}_n^k + \chi^k,$$

where $\chi^k = \frac{v^k - \sum_{n'} \nu \left(\frac{L_{n'}^k}{L^k}\right) W_{n'}^k}{1+\nu}$, and recall that $\widetilde{w}_n^k$ is the social marginal product of labor for occupation k in city n.

Proof:
Equation (53) may alternatively be expressed as

$$\phi^k v^k \frac{L_n^k}{C_n^k} = L_n^k P_n - \left(\frac{\nu}{C_n^k}\right) L_n^k W_n^k + \sum_{n'=1}^{N} \left(\frac{\nu}{C_n^k}\right) \left(\frac{L_{n'}^k}{L^k}\right) L_n^k W_{n'}^k,$$

where v^k is defined in equation (15). Alternatively, we have that

$$\underbrace{\phi^k v^k - \sum_{n'} \nu \left(\frac{L_{n'}^k}{L^k}\right) W_{n'}^k}_{(1+\nu)\chi^k} = P_n C_n^k - \nu W_n^k.$$

Substituting for W_n^k from (54) in this last expression gives

$$P_n C_n^k = \nu \left(\widetilde{w}_n^k - P_n C_n^k \right) + (1 + \nu)\, \chi^k$$

or

$$P_n C_n^k = \frac{\nu}{1 + \nu} \widetilde{w}_n^k + \chi^k. \tag{62}$$

□

Observe that we can also use (54) to write χ^k as a function of prices, $\widetilde{w}_n^k$, P_n, and consumption, C_n^k. In particular,

$$\chi^k = \frac{\phi^k v^k}{1 + \nu} - \frac{\nu}{1 + \nu} \sum_{n'} \left(\frac{L_{n'}^k}{L^k} \right) \left(\widetilde{w}_n^k - P_{n'} C_{n'}^k \right).$$

We can then obtain an expression for the total consumption expenditures of households of type k by adding (62) across cities n, with the expression for χ^k substituted in,

$$\phi^k v^k L^k = \sum_n P_n C_n^k L_n^k \tag{63}$$

Substituting out $\phi^k v^k$ back into the expression for χ^k and rearranging, we obtain

$$\chi^k = \frac{\sum_n P_n C_n^k L_n^k}{L^k} - \sum_n \frac{\nu^k}{1 + \nu} \left(\frac{L_n^k}{L^k} \right) \widetilde{w}_n^k.$$

Finally, note that $\sum_{n,k} P_n C_n^k L_n^k = \sum_{n,k} \left(w_n^k L_n^k + r_n H_n \right)$, so that

$$\sum_k L^k \chi^k = \sum_{n,k} \frac{1}{1 + \nu} w_n^k L_n^k - \sum_{n,k} \frac{\nu}{1 + \nu} \left(\widetilde{w}_n^k - w_n^k \right) L_n^k + \sum_n r_n H_n \tag{64}$$

The individual values for χ^k are determined to be such that equation (64) is satisfied.

D.1 The Social and Private Marginal Value of Workers of type k in city n (Proof of Lemma 1)

Solving the derivative in the equation defining the social value of workers of type k in city n (55), we obtain

$$\begin{aligned} &\tilde{w}_n^k - w_n^k \\ &= \sum_j \int \frac{\partial z_n^j \left[H_n^j(\mathbf{z})^{\beta_n^j} \left[\sum_{k'} \left(\lambda_n^{k'j} \left(\mathbf{L}_n \right) L_n^{k'j} \left(\mathbf{z} \right) \right)^{1-\frac{1}{\epsilon}} \right]^{\frac{\epsilon}{1-\epsilon}\left(1-\beta_n^j\right)} \right]^{\gamma_n^j} \prod_{j'=1}^{J} M_n^{j'j}(\mathbf{z})^{\gamma^{j'j}}}{\partial L_n^k} p_n^j(\mathbf{z}) d\Phi(\mathbf{z}) \end{aligned}$$

where $p_n^j(\mathbf{z}) d\Phi^j(\mathbf{z}) = \tilde{p}_n^j(\mathbf{z})$. This expression is equivalent to

$$\tilde{w}_n^k - w_n^k = \sum_{j,k'} (1-\beta_n^j)\gamma_n^j \frac{\left(\frac{w_n^{k'}}{\lambda_n^{k'j}(\mathbf{L}_n)}\right)^{1-\epsilon}}{\sum_{k''}\left(\frac{w_n^{k''}}{\lambda_n^{k'j}(\mathbf{L}_n)}\right)^{1-\epsilon}} \frac{1}{\lambda_n^{k'j}(\mathbf{L}_n)} \frac{\partial \lambda_n^{k'j}(\mathbf{L}_n)}{\partial L_n^k} q_n^j(\mathbf{z}) p_n^j(\mathbf{z}) d\Phi(\mathbf{z}).$$

Rearranging and integrating equation (57) yields

$$w_n^k L_n^{kj} = (1-\beta_n^j)\gamma_n^j \frac{\left(\frac{w_n^{k}}{\lambda_n^{kj}(\mathbf{L}_n)}\right)^{1-\epsilon}}{\sum_{k'}\left(\frac{w_n^{k'}}{\lambda_n^{k'j}(\mathbf{L}_n)}\right)^{1-\epsilon}} \int q_n^j(\mathbf{z}) p_n^j(\mathbf{z}) d\Phi(\mathbf{z}),$$

so that the expression for the deviation of private from social marginal product of labor simplifies further to

$$\tilde{w}_n^k - w_n^k = \sum_{j,k'} w_n^{k'} \frac{L_n^{k'j}}{L_n^k} \frac{\partial \ln \lambda_n^{k'j}(\mathbf{L}_n)}{\partial \ln L_n^k}. \tag{65}$$

D.2 Implementation (Proof of Proposition 1)

We now discuss the implementation of the optimal policy. One possible implementation is to combine a direct employment subsidy to firms that is specific to cities and occupations (Δ_n^k), a linear occupation-specific labor income tax (t_L^k), combined with occupation-specific transfers (R^k).

With externalities in occupations, the social and private marginal products of labor differ. The first step in the implementation of optimal allocations, therefore, is to subsidize firms in different locations to hire different occupation types. We define $\tilde{w}_n^k$ to be the after-subsidy wage associated with workers in occupation k living in city n such that

$$\tilde{w}_n^k = w_n^k + \Delta_n^k,$$

where Δ_n^k is a per-worker subsidy offered to firms in city n hiring workers of type k. With these subsidized wages in place, we take advantage of various additional taxes and transfers to implement optimal allocations. In particular, equation (1) becomes

$$I_n^k = (1-t_L^k)\tilde{w}_n^k + \chi^k + R^k, \; (NK \text{ eqs.}) \,. \tag{66}$$

where transfers have to be such that the government budget balances,

$$\sum_{n,k} L_n^k R^k = \sum_{n,k} t_L^k w_n^k L_n^k - \sum_{n,k} (1-t_L^k)\Delta_n^k L_n^k. \tag{67}$$

We also have that labor demand depend only on pre-subsidy wages, w_n^k,

$$w_n^k L_n^{kj}(\mathbf{z}) = \frac{\left(\frac{w_n^{k}}{\lambda_n^{kj}(\mathbf{L_n})}\right)^{1-\epsilon}}{\sum_{k'}\left(\frac{w_n^{k}}{\lambda_n^{k'j}(\mathbf{L_n})}\right)^{1-\epsilon}} \gamma_n^j \left(1-\beta_n^j\right) p_n^j(\mathbf{z}) q_n^j(\mathbf{z}), \; (NKJ \; eqs.) \tag{68}$$

$$x_n^j = B^j \left[r_n^{\beta_n^j} \left[\sum_k \left(\frac{w_n^k}{\lambda_n^{kj} \left(\mathbf{L}_n^k \right)} \right)^{1-\epsilon} \right]^{\frac{1-\beta_n^j}{1-\epsilon}} \right]^{\gamma_n^j} \prod_{j'} \left(P_n^{j'} \right)^{\gamma_n^{j'j}}, \tag{69}$$

Definition 1. An equilibrium with taxes and transfers is defined as the equilibrium without taxes and transfers but with the additional conditions that i) I_n^k is given by equation (66), ii) the first-order condition describing intermediate goods producers' labor demand is given by (68), iii) the cost index x_n^j is given by (69), and iv) the government budget constraint (67) is satisfied.

Proposition. *Let*

$$t_L^k = \frac{1}{1+\nu}$$

$$\Delta_n^k = \sum_{k'j} w_n^{k'} \frac{L_n^{k'j}}{L_n^k} \frac{\partial \ln \lambda_n^{k'j} \left(\mathbf{L}_n^k \right)}{\partial \ln L_n^k}$$

and R^k is such that

$$\phi^k v^k L^k = \sum_n P_n C_n^k L_n^k$$

Then, if the planner's problem is globally concave, the equilibrium with taxes and transfers implements the optimal allocation.

Proof. 1) The first order condition for household consumption choice (14) is identical to the first order condition for consumption in the planner's problem, (51). The modified budget constraint for the household (66) implies a relationship between consumption and prices identical to equation (62), which is derived from the first order conditions (53) and (54) in the planner's problem. At the same time, the optimal location decision for the household, (2) is identical to the free mobility constraint in the planner's choice (49) for a given set of consumption C_n^k.

2) The first order condition for factor demand for intermediate input producers, (16), (18) and (68) are identical to the first order conditions for the planner's problem (57), (58) and (59) once one uses equation (65) to substitute $\tilde{w}_n^k$ out of (57).

3) The condition that a producer in city n and industry j imports a variety $\mathbf{z}$ from city n' if and only if $\kappa_{nn'}^j p_n^j(\mathbf{z}) = \min_{n''} \kappa_{nn''}^j p_{n''}^j(\mathbf{z})$ is implied by the first order condition for the planner's problem (56), given that $Q_n^j(\mathbf{z}) = \sum_{n'} Q_{nn'}^j(\mathbf{z})$, $\tilde{p}_n^j(\mathbf{z}) d\Phi^j(\mathbf{z}) = p_n^j(\mathbf{z})$.

4) The first order condition for use of different varieties by final goods producers (23) is implied by (56) given that $Q_n^j(\mathbf{z}) = \sum_{n'} Q_{nn'}^j(\mathbf{z})$, $\tilde{p}_n^j(\mathbf{z}) d\Phi^j(\mathbf{z}) = p_n^j(\mathbf{z})$, and $P_n^j(\mathbf{z}) = \min_{n'} \kappa_{nn'}^j p_{n'}^j(\mathbf{z})$.

5) The market clearing conditions for employment (equation 5), structures (equation 6), final goods (equation 7) and intermediate goods (8) are identical to the resource constraints faced by the planner, respectively, (48) combined with (49), (50), (46) and (47). □

E Model Quantification for 1980 and Counterfactual Exercises

E.1 Details of Model Quantification for 1980

In order to quantify the model for 1980 we follow similar steps as described in Section B, with modifications to accommodate data constraints.

Regional Price Parities data are not available for 1980. We use those in order to calculate the productivity of the non-tradable sectors. To obtain the productivity of the real estate sector in 1980 we match instead changes in CoreLogic housing price index data, available by county. As for the productivity of the non-tradable sector, we assume that its spatial distribution does not change. Also, the model inversion exercise done for the 2011-15 does not pin down national average level of productivity for each industry, only its occupational and spatial variation. In order to obtain the time variation in time of those levels, we choose average 1980 productivity levels to match national level sectoral price data series made available by the BEA.

In order to obtain wages and the occupational composition of cities and industries, we use the 5% sample of the 1980 Census data, which is comparable to the ACS. The 1980 Census has data for 213 MSA's, accounting for approximately 85% of US employment in that year. For the remaining MSA's we impute wages and employment by occupation and by sector by taking the predicted values of a regression of those variables on 1980 CBP employment by sector and housing prices.

E.2 Details of Counterfactual Exercises

In the counterfactual exercises described in Section 6, we separate average changes in productivity or amenities from their geographical and occupational dispersion. Productivity changes refer to changes in value added productivity, given by $\left(T_n^{kj}\right)^{\frac{1}{\gamma_n^j}} = \left(H_n^j\right)^{\beta_n^j} \left(\lambda_n^{kj}\right)^{\gamma_n^j(1-\beta_n^j)}$. The average change in productivity between 1980 and 2011-15 for a given industry is a Tornqvist type index: a geometric weighted average of the changes in productivity across cities, with the weights given by the value added by each city/industry as a fraction of total industry/city value added. Those shares are first calculated separately for 1980 and 2011-15 period, and the weights correspond to the arithmetic average of those shares.[30]

The model does not allow us to pin down an aggregate trend in amenities, since changing amenities in all cities by a common scaling parameter leaves the equilibrium unchanged. We thus assume that there was no such trend, so as to focus on the welfare implications of endogenous changes equilibrium variables. For the baseline economy, this implies keeping a Tornqvist type index of amenities constant relative to the 2011-15 period: specifically, we keep a weighted geometric average of changes in amenities equal to 1, with the weights given by employment shares by city (again the shares are taken for the baseline and 2011-15 periods separately and the weights are given by an arithmetic average).

[30]We do a similar calculation in order to obtain productivity trends by city/industry/occupation

F Counterfactual Economy with no Endogenous Amenities

We now verify whether the planner solution would be likely to change if one were to adjust local amenities to remove the parts that Diamond (2016) argues are likely to be endogenous. For that purpose we do two counterfactual exercises. For both exercises, we first extract the exogenous part of amenities as implied by the mapping of Diamond's (2016) estimates into amenity spill-overs performed by Fajgelbaum and Gaubert (2018). Specifically, we calculate a value of $A_n^{k,exo}$ such that $A_n^{k,exo} \prod \left(L_n^{k'}\right)^{\tau_a^{k'k}} C_n^k = 1$, with $\tau_a^{CNR,CNR} = 0.77$, $\tau_a^{nCNR,CNR} = -1.24$, $\tau_a^{CNR,nCNR} = 0.18$ and $\tau_a^{nCNR,nCNR} = -0.43$. In the first exercise, we calculate a counterfactual equilibrium where the labor supply equations are given by $L_n^k = \frac{\left(A_n^{k,exo} C_n^k\right)^{-\nu^k}}{\sum_{n'} \left(A_{n'}^{k,exo} C_{n'}^k\right)^{-\nu^k}}$. In the second exercise, we calculate the optimal allocation in that counterfactually quantified environment.

Figure 18 shows how the distribution of CNR workers in the optimal allocation compares with the counter-factual equilibrium. As before, the planner has an incentive to increase labor market polarization by packing proportionately more CNR workers in larger cities. Figure 19 shows that, as before, this increased polarization is matched by transfers from the large cities to the small ones.

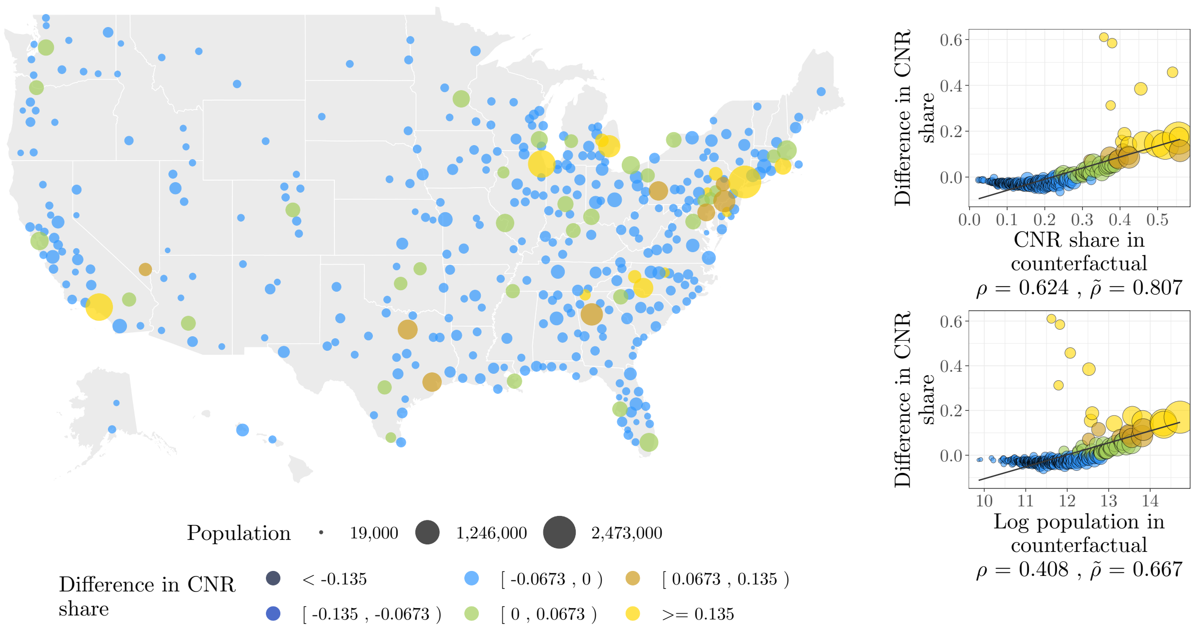


Figure 18: Optimal L_n^{CNR}/L_n with counterfactual amenities (change from counterfactual equilibrium)

Each marker in the map refers to a CBSA. Marker sizes are proportional to total equilibrium employment in each city. ρ and $\tilde{\rho}$ are unweighted and population weighted correlations respectively.

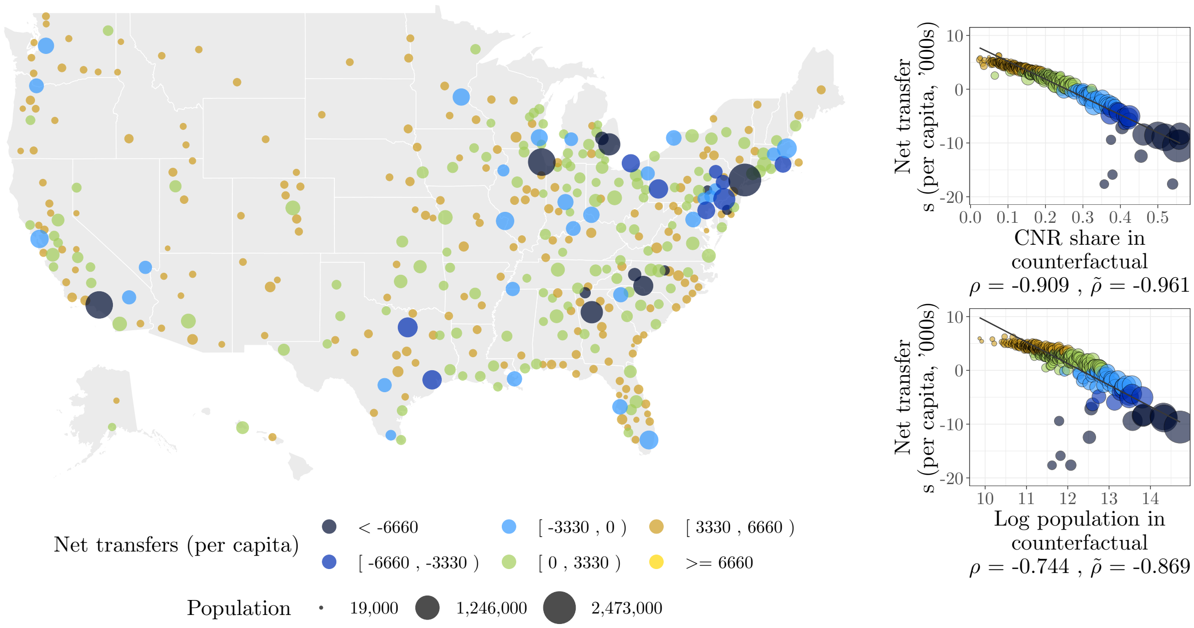


Figure 19: Optimal transfers in with counterfactual amenities

Optimal transfers defined as the difference in the optimal allocation between the value consumed and value added in each city ($\sum_k P_n C_n^k - \sum_k w - n^k L_n^k - r_n H_n$). Each marker in the map refers to a CBSA. Marker sizes are proportional to total equilibrium employment in each city. ρ and $\tilde{\rho}$ are unweighted and population weighted correlations respectively.