Commuting, Migration and Local Employment Elasticities*

Ferdinando Monte† — *Georgetown University*
Stephen J. Redding‡ — *Princeton University*
Esteban Rossi-Hansberg§ — *Princeton University*

September 19, 2016

Abstract

Many changes to the economic environment are local, including policy changes and infrastructure investments. The effect of these changes depends crucially on the ability of factors to move in response. Therefore a key object of interest is the *elasticity of local employment* with respect to the change in the local environment. To understand the determinants of this elasticity, we develop a quantitative general equilibrium model that incorporates spatial linkages between locations in goods markets (trade) and factor markets (commuting and migration). We show that local employment elasticities differ substantially across U.S. counties and commuting zones (CZs) in ways that are not well explained by standard empirical controls. We propose a simple summary statistic based on the share of residents who work locally that goes a long way towards explaining this heterogeneity. Using a natural experiment from the location of million dollar plants (MDPs), we find strong evidence of heterogeneous local employment elasticities that take exactly the form predicted by our model.
JEL CLASSIFICATION: F12, F14, R13, R23

1 Introduction

Agents spend about 8% of their workday commuting to and from work.[1] They make this significant daily investment, to live and work in different locations, so as to balance their living costs and residential amenities with the wage they can obtain at their place of employment. The ability of firms in a location to attract workers depends, therefore, not only on the ability to attract local residents through migration, but also on the ability to attract commuters from other, nearby, locations. Together, migration and commuting determine the response of local employment to a local labor demand shock, which we term the *local employment elasticity*. This elasticity is of great policy interest since it determines the impact of local

*Much of this research was undertaken while Ferdinando Monte was visiting the International Economics Section (IES) at Princeton. We are grateful to the IES and Princeton more generally for research support. We are also grateful to conference and seminar participants at various intitutions for helpful comments and suggestions.
†McDonough School of Business, 37th and O Streets, NW, Washington, DC 20057. ferdinando.monte@georgetown.edu.
‡Dept. Economics and WWS, Fisher Hall, Princeton, NJ 08544. 609 258 4016. reddings@princeton.edu.
§Dept. Economics and WWS, Fisher Hall, Princeton, NJ 08544. 609 258 4024. erossi@princeton.edu.
[1] See for example Redding and Turner (2015).

policies, such as transport infrastructure investments, local taxation and regional development programs. Estimating its magnitude has been the subject of a large empirical literature on local labor markets, which has considered a variety of sources of local labor demand shocks, including sectoral composition (Bartik shocks), productivity, international trade, natural resource abundance and business cycle fluctuations, as discussed further below.[2] In this paper we explore the determinants and characteristics of the local employment elasticity (and the corresponding *local residents elasticity*) using a detailed quantitative spatial equilibrium theory.

We develop a quantitative spatial general equilibrium model that incorporates spatial linkages between locations in both goods markets (trade) and factor markets (commuting and migration). We show that there is no single local employment elasticity. Instead the local employment elasticity is an endogenous variable that differs across locations depending on their linkages to one another in goods and factor markets. Calibrating our model to county-level data for the United States, we find that the elasticity of local employment with respect to local productivity shocks varies from around 0.5 to 2.5. Therefore an average local employment elasticity estimated from cross-section data can be quite misleading when used to predict the impact of a local shock on any individual county and can lead to substantial under or overprediction of the effect of the shock. We use our quantitative model to understand the systematic determinants of the local employment elasticity and show that a large part of the variation results from differences in commuting links between a location and its neighbors. This allows us to propose variables that can be included in reduced-form regressions to improve their ability to predict the heterogeneity in local employment responses without imposing the full structure of our model.

Our theoretical framework allows for an arbitrary number of locations that can differ in productivity, amenities and geographical relationship to one another. The spatial distribution of economic activity is driven by a tension between productivity differences and home market effects (forces for the concentration of economic activity) and an inelastic supply of land and commuting costs (dispersion forces). Commuting allows workers to access high productivity employment locations without having to live there and hence alleviates the congestion effect in such high productivity locations. We show that the resulting commuting flows between locations exhibit a gravity equation relationship with a much higher distance elasticity than for goods flows, suggesting that moving people is more costly than moving goods across geographic space. We discipline our quantitative spatial model to match the observed gravity equation relationships for trade in goods and commuting flows as well as the observed cross-section distributions of employment, residents and wages across U.S. counties. Given the observed data on wages, employment by workplace, commuting flows and land area, and a parameterization of trade and commuting costs, we show that our model can be used to recover unique values of the unobserved location fundamentals (productivity and amenities) that exactly rationalize the observed data as an equilibrium of the model. We show how the values of these observed variables in an initial equilibrium can be used to undertake counterfactuals for the impact of local labor demand shocks (captured by productivity shocks in our model) and for the impact of changes in trade or commuting costs.

[2] For a survey of this empirical literature, see Moretti (2011).

An advantage of our explicitly modeling the spatial linkages between locations is that our framework can be taken to data on local economic activity at different levels of spatial aggregation. In contrast, existing research that does not explicitly model the spatial linkages between locations is faced with a trade-off when studying local labor markets. On the one hand, larger spatial units have the advantage of reducing the unmodeled spatial linkages between locations. On the other hand, larger spatial units have the disadvantage of reducing the ability to make inferences about local labor markets. Furthermore, there exists no choice of boundaries that eliminates commuting between spatial units. For example, is Princeton part of New York's or Philadelphia's commuting zone (CZ)? Some of its residents commute to New York, but others commute to Philadelphia. Our approach overcomes these problems by explicitly modeling the geographic linkages between the spatial units. In our baseline specification, we report results for counties, because this is the finest level of geographical detail at which commuting data are reported in the American Community Survey and Decennial Census, and several influential papers in the local labor markets literature have used county data (such as Greenstone, Hornbeck and Moretti 2010).[3] In robustness tests, we also report results for CZs, which are aggregations of counties by the Bureau of Economic Analysis designed to minimize commuting flows between locations.

We demonstrate both theoretically and empirically the robustness of our findings of heterogeneous local employment elasticities. From a theoretical perspective, we show that heterogeneous local employment elasticities are not specific to our theoretical model, but rather are a generic prediction of an entire class of theoretical models consistent with a gravity equation for commuting flows. Therefore this prediction holds across a range of different versions of our model, including incorporating heterogeneous land supply elasticities across locations, non-traded goods, congestion in commuting costs, heterogeneity in effective units of labor rather than in amenities, and different assumptions about the ownership of land. These different theoretical specifications can affect the elasticity of wages with respect to productivity, but as long as these specifications yield a gravity equation for commuting flows, they imply the same elasticity of employment with respect to wages. This elasticity can be derived directly from the gravity equation for commuting flows, which we show to be a strong empirical feature of the data.

From an empirical perspective, we show that we continue to find substantial heterogeneity in these local employment elasticities when we incorporate the variable land supply elasticities from Saiz (2010). Introducing this second source of heterogeneity generates more variation in local residents elasticities but does not reduce the variation in local employment elasticities. Furthermore, we continue to find substantial differences between local employment and local residents elasticities, where the only reason that these two elasticities can differ from one another is commuting. We also continue to find substantial heterogeneity in local employment elasticities when we replicate our analysis for CZs rather than counties. The reason is that are substantial differences across CZs in the extent to which there remain commuting links with other CZs, and it is these differences that generate the heterogeneity in local employment elasticities.

To provide further evidence of heterogeneous local employment elasticities that is independent of

[3]The LEHD Origin-Destination Employment Statistics (LODES) reports commuting data for more disaggregated spatial units than counties, but there is substantial interpolation, and data are missing for some state-year combinations.

our model, we use the natural experiment of million dollar plants (MDP) from Greenstone, Hornbeck and Moretti (2010), one of the most influential papers in the local labor markets literature. We compare “winner" and “runner-up" counties that are similar to one another, except that the winner counties were ultimately successful in attracting a MDP. We confirm the findings of a positive average treatment effect of MDPs from Greenstone, Hornbeck and Moretti (2010). Additionally, we find that this average treatment effect masks considerable heterogeneity across counties, which takes exactly the form predicted by our model. We find that winner counties that are more open in commuting linkages experience substantially and statistically significantly larger increases in employment than other winner counties.[4] To provide additional evidence, independent of our model, of the role of commuting in understanding observed employment changes, we show that over the period 1990-2007 changes in net commuting patterns accounted for a substantial proportion of the observed changes in employment. Furthermore, we find substantial heterogeneity across counties in the relative importance of commuting.

Having established the importance of commuting for the local employment response to local labor demand shocks, we next show that our model provides a platform for evaluating the counterfactual effects of changes in trade and commuting costs. Building on approaches in the international trade literature (e.g. Head and Ries 2001), we show how observed data on commuting flows can be used to back out implied bilateral commuting costs. We use these distributions to undertake counterfactuals for empirically-plausible changes in commuting costs. For example, reducing commuting costs between all counties by the difference between the 25th and 75th percentiles of distribution of these costs across counties (a reduction of 12%), we find an increase in welfare of 3.3 percent. The commuting technology facilitates a separation of workplace and residence, enabling people to work in relatively high productivity locations and live in relatively high amenity locations. Therefore reducing commuting costs increases the concentration of employment in locations that were net importers of commuters in the initial equilibrium (e.g. Manhattan) and the clustering of residents in locations that initially were net exporters of commuters (e.g. Brooklyn and parts of New Jersery). This logic seems to suggest that commuting might be important only for larger cities in the U.S., but this is in fact not the case. Although the changes in employment as a result of eliminating commuting are well explained by initial commuting intensity, this intensity cannot be easily proxied for using standard empirical controls, such as land area, size or housing supply elasticities. These results again underscore the relevant information embedded in commuting links.

Our paper is related to several existing literatures. In international trade, our work relates to quantitative models of costly trade in goods following Eaton and Kortum (2002) and extensions. Our research also contributes to the economic geography literature on costly trade in goods and factor mobility, which typically uses variation *across regions or systems of cities*, including Krugman (1991), Hanson (1996, 2005), Helpman (1998), Fujita et al. (1999), Rossi-Hansberg (2005), Redding and Sturm (2008), Redding (2016), Moretti and Klein (2014), Allen and Arkolakis (2014a,b), Caliendo, et al. (2014) and Desmet and Rossi-Hansberg (2014). Our work also contributes to the urban economics literature on the costly move-

[4]These results are consistent with the findings of Manning and Petrongolo (2011) that local development policies are fairly ineffective in raising local unemployment outflows, because labor markets overlap, and the associated ripple effects in applications largely dilute the impact of local stimulus across space.

ment of people (commuting), which typically uses variation *within cities*, including Alonso (1964), Mills (1967), Muth (1969), Lucas and Rossi-Hansberg (2002), Desmet and Rossi-Hansberg (2013), Behrens, et al. (2014), Ahlfeldt, et al. (2015), Monte (2015) and Allen, Arkolakis and Li (2015). In contrast, we develop a framework in which an arbitrary set of regions are connected in both goods markets (through costly trade) and labor markets (through migration and commuting), and which encompasses both within and across-city interactions. Although incorporating costly goods trade and commuting is a natural idea, our first main contribution is to develop a tractable framework that is amenable to both analytic and quantitative analysis, and for which we provide general results for the existence and uniqueness of the spatial equilibrium. Our second main contribution is to quantify this framework using disaggregated data on trade and commuting for the United States and to show how it provides a platform for evaluating a range of counterfactual interventions. Our third main contribution is to establish the importance of spatial interactions between locations (in particular through commuting) in determining the local economic effects of local labor demand shocks.

Our paper is also related to the large empirical literature on local labor markets, which has estimated the local effects of a range of local labor demand shocks: (a) Greenstone, Hornbeck and Moretti (2010)'s analysis of million dollar plants (MDPs); (b) Autor, Dorn and Hanson (2013), which examine the local economic effects from the international trade shock provided by China's emergence into global markets; (c) the many empirical studies that use the Bartik (1991) instrument, which interacts aggregate industry shocks with locations' industry employment shares, including Diamond (2016) and Notowidigdo (2013); (d) research on the geographic incidence of macroeconomic shocks, such as the 2008 Financial Crisis and Great Recession, such as Mian and Sufi (2014) and Yagan (2016); and (e) work on the impact of natural resource discoveries on the spatial distribution of economic activity, as in Michaels (2011) and Feyrer, Mansur and Sacerdote (2015).[5] Each of these papers is concerned with evaluating the local impact of economic shocks using data on finely-detailed spatial units. However, these spatial units are typically treated as independent observations in reduced-form regressions, with little attention paid to the linkages between these spatial units in goods and labor markets, and hence with little consideration of the distinction between employment and residents introduced by endogenous commuting decisions. A key contribution of our paper to show that understanding these spatial linkages is central to evaluating the local impact of these and other economic shocks.

The remainder of the paper is structured as follows. Section 2 develops our theoretical framework. Section 3 discusses the quantification of the model using U.S. data and reports summary statistics on commuting between counties. Section 4 shows both theoretically and empirically the heterogeneity of local employment elasticities. Section 5 studies the effect of changes in commuting costs and Section 6 summarizes our conclusions. A web appendix contains the derivations of theoretical results, the proofs of propositions, additional robustness tests, and a description of the data sources and manipulations.

[5]Other related research on local labor demand shocks includes Blanchard and Katz (1992), Bound and Holzer (2000), and Busso, Gregory and Kline (2013).

2 The Model

We develop a spatial general equilibrium model in which locations are linked in goods markets through trade and in factor markets through migration and commuting. The economy consists of a set of locations $n, i \in N$. Each location n is endowed with a supply of land (H_n). Following the new economic geography literature, we begin by interpreting land as geographical land area, which is necessarily in perfectly inelastic supply. We later extend our analysis to interpret land as developed land area, which has a positive supply elasticity that we allow to differ across locations. The economy as a whole is populated by a measure $\bar{L}$ of workers, each of whom is endowed with one unit of labor that is supplied inelastically.

2.1 Preferences and Endowments

Workers are geographical mobile and have heterogeneous preferences for locations. Each worker chooses a pair of residence and workplace locations to maximize their utility taking as given the choices of other firms and workers.[6] The preferences of a worker ω who lives and consumes in region n and works in region i are defined over final goods consumption ($C_{n\omega}$), residential land use ($H_{n\omega}$), an idiosyncratic amenities shock ($b_{ni\omega}$) and commuting costs (κ_{ni}), according to the Cobb-Douglas form,[7]

$$U_{ni\omega} = \frac{b_{ni\omega}}{\kappa_{ni}} \left(\frac{C_{n\omega}}{\alpha} \right)^{\alpha} \left(\frac{H_{n\omega}}{1-\alpha} \right)^{1-\alpha}, \tag{1}$$

where $\kappa_{ni} \in [1, \infty)$ is an iceberg commuting cost in terms of utility. The idiosyncratic amenities shock ($b_{ni\omega}$) captures the idea that individual workers can have idiosyncratic reasons for living and working in different locations. We model this heterogeneity in amenities following McFadden (1974) and Eaton and Kortum (2002).[8] For each worker ω living in location n and working in location i, idiosyncratic amenities ($b_{ni\omega}$) are drawn from an independent Fréchet distribution,

$$G_{ni}(b) = e^{-B_{ni} b^{-\epsilon}}, \qquad B_{ni} > 0, \epsilon > 1, \tag{2}$$

where the scale parameter B_{ni} determines the average amenities from living in location n and working in location i, and the shape parameter $\epsilon > 1$ controls the dispersion of amenities. This idiosyncratic amenities shock implies that different workers make different choices about their workplace and residence locations when faced with the same prices and wages. All workers ω residing in region n and working in region i receive the same wage and make the same consumption and residential land choices. Hence we suppress

[6]Throughout the following we use n to denote a worker's location of residence and consumption and i to denote a worker's location of employment and production, unless otherwise indicated.

[7]For empirical evidence using U.S. data in support of the constant housing expenditure share implied by the Cobb-Douglas functional form, see Davis and Ortalo-Magne (2011).

[8]A long line of research models location decisions using preference heterogeneity, as in Artuc, Chaudhuri and McClaren (2010), Kennan and Walker (2011), Grogger and Hanson (2011), Moretti (2011) and Busso, Gregory and Kline (2013).

the implicit dependence on ω except where important.[9]

To isolate the implications of introducing commuting, we model goods consumption and production as in the new economic geography literature. The goods consumption index in location n is a constant elasticity of substitution (CES) function of consumption of a continuum of tradable varieties sourced from each location i,

$$C_n = \left[\sum_{i \in N} \int_0^{M_i} c_{ni}(j)^{\rho} dj \right]^{\frac{1}{\rho}}, \qquad \sigma = \frac{1}{1-\rho} > 1. \tag{3}$$

Utility maximization implies that equilibrium consumption in location n of each variety sourced from location i is given by $c_{ni}(j) = \alpha X_n P_n^{\sigma-1} p_{ni}(j)^{-\sigma}$, where X_n is aggregate expenditure in location n; P_n is the price index dual to (3), and $p_{ni}(j)$ is the "cost inclusive of freight" price of a variety produced in location i and consumed in location n.[10]

Utility maximization also implies that a fraction $(1-\alpha)$ of worker income is spent on residential land. We assume that this land is owned by immobile landlords, who receive worker expenditure on residential land as income, and consume goods where they live. This assumption allows us to incorporate general equilibrium effects from changes in the value of land, without introducing a mechanical externality into workers' location decisions from the local redistribution of land rents.[11] Using this assumption, total expenditure on consumption goods equals the fraction α of the total income of residents plus the entire income of landlords (which equals the fraction $(1-\alpha)$ of the total income of residents):

$$P_n C_n = \alpha \bar{v}_n R_n + (1-\alpha)\, \bar{v}_n R_n = \bar{v}_n R_n \tag{4}$$

where $\bar{v}_n$ is the average labor income of residents across employment locations; and R_n is the measure of residents. Land market clearing determines the land price (Q_n) as a function of the supply of land (H_n):

$$Q_n = (1-\alpha)\, \frac{\bar{v}_n R_n}{H_n}. \tag{5}$$

2.2 Production

Tradeable varieties are produced under monopolistic competition and increasing returns to scale. Again to isolate the implications of introducing commuting, we assume that labor is the sole factor of production. To produce a variety, a firm must incur a fixed cost of F units of labor and a constant variable cost in terms

[9]Our baseline specification focuses on a single worker type with a Fréchet distribution of idiosyncratic preferences for tractability. In subsection B.9 of the web appendix, we generalize our analysis to multiple worker types z that differ in their average valuation of amenities (as determined by B_{ni}^z) and degree of heterogeneity in idiosyncratic preferences (ϵ^z). While our baseline specification results in choice probabilities that take a similar form as in the logit model, this extension results in choice probabilities that take a similar form as in the mixed logit model of McFadden and Train (2000).

[10]In subsection B.11 of the web appendix, we show how this standard specification can be further generalized to introduce non-traded consumption goods.

[11]In subsection B.12 of the web appendix, we show that the model has similar properties if landlords consume both consumption goods and residential land, although expressions are less clean. In the web appendix, we also report the results of a robustness test, in which we instead assume that all land is owned by a global portfolio that redistributes land rents to workers throughout the economy (as in Caliendo et al. 2015).

of labor that depends on a region's productivity A_i.[12] Therefore the total amount of labor ($l_i(j)$) required to produce $x_i(j)$ units of a variety j in region i is $l_i(j) = F + x_i(j)/A_i$.[13]

Profit maximization implies that equilibrium prices are a constant mark-up over marginal cost,

$$p_{ni}(j) = \left(\frac{\sigma}{\sigma - 1}\right) \frac{d_{ni} w_i}{A_i}, \tag{6}$$

where w_i is the wage in region i. Profit maximization and zero profits imply that equilibrium output of each variety is equal to $x_i(j) = A_i F (\sigma - 1)$. A constant equilibrium output of each variety and labor market clearing then imply that the total measure of produced varieties (M_i) is proportional to the measure of employed workers (L_i),

$$M_i = \frac{L_i}{\sigma F}. \tag{7}$$

2.3 Goods Trade

The model implies a gravity equation for bilateral trade between regions. Using the CES expenditure function, equilibrium prices (6) and the measure of firms in (7), the share of region n's expenditure on goods produced in region i is

$$\pi_{ni} = \frac{M_i p_{ni}^{1-\sigma}}{\sum_{k \in N} M_k p_{nk}^{1-\sigma}} = \frac{L_i \left(d_{ni} w_i / A_i\right)^{1-\sigma}}{\sum_{k \in N} L_k \left(d_{nk} w_k / A_k\right)^{1-\sigma}}. \tag{8}$$

Therefore trade between regions n and i depends on bilateral trade costs (d_{ni}) in the numerator ("bilateral resistance") and on trade costs to all possible sources of supply k in the denominator ("multilateral resistance"). Equating revenue and expenditure and using zero profits, workplace income in each location equals total expenditure on goods produced in that location, namely,[14]

$$w_i L_i = \sum_{n \in N} \pi_{ni} \bar{v}_n R_n. \tag{9}$$

[12] We assume a representative firm within each location. However, it is straightforward to generalize the analysis to introduce firm heterogeneity following Melitz (2003), where all firms entering in location i draw a productivity φ from an untruncated Pareto distribution $g(\varphi)$ that can be used to produce varieties in location i.

[13] In subsection B.13 of the web appendix, we generalize the production technology to incorporate the use of intermediate inputs (as in Krugman and Venables 1995 and Eaton and Kortum 2002), commercial land use and physical capital. We show that the model continues to predict heterogeneous local employment elasticities across locations.

[14] Although total workplace income equals total expenditure on goods produced in a location, total residential income can differ from total workplace income (because of commuting). Therefore total workplace income need not equal total residential expenditure, which implies that total exports need not equal total imports. When we take the model to the data, we also allow total residential expenditure to differ from total residential income, which provides an additional source for trade deficits. Within the model, these two variables can diverge if landlords own land in different locations from where they consume. This is how we interpret trade deficits in the empirical section.

Using equilibrium prices (6) and labor market clearing (7), the price index dual to the consumption index (3) can be expressed as

$$P_n = \frac{\sigma}{\sigma-1}\left(\frac{1}{\sigma F}\right)^{\frac{1}{1-\sigma}}\left[\sum_{i\in N} L_i\left(d_{ni}w_i/A_i\right)^{1-\sigma}\right]^{\frac{1}{1-\sigma}} = \frac{\sigma}{\sigma-1}\left(\frac{L_n}{\sigma F\pi_{nn}}\right)^{\frac{1}{1-\sigma}}\frac{d_{nn}w_n}{A_n}. \tag{10}$$

where the second equality uses (8) to write the price index (10) as in the class of models considered by Arkolakis and Allen (2014b). This class of models includes increasing returns models such as Krugman (1991) and constant returns models such as versions of Armington (1969) and Eaton and Kortum (2002).

One difference between the increasing returns model of new economic geography considered here and the constant returns versions of Armington (1969) and Eaton and Kortum (2002) is that here the measure of employed workers in each region (L_i) enters the trade share (π_{ni}) and price index (P_n). This role for the measure of employed workers reflects consumer love of variety and the endogenous measure of varieties. As more agents choose to work in a region, this increases the measure of varieties produced in that region, which increases the share of consumer expenditure allocated to that region and reduces the consumer price index (this is the pecuniary externality highlighted by the New Economic Geography Literature). A similar role for the measure of employed workers arises in versions of Armington (1969) and Eaton and Kortum (2002) augmented to include external economies of scale, as considered further below.

2.4 Labor Mobility and Commuting

Workers are geographically mobile and choose their pair of residence and workplace locations to maximize their utility. Given our specification of preferences (1), the indirect utility function for a worker ω residing in region n and working in region i is

$$U_{ni\omega} = \frac{b_{ni\omega}w_i}{\kappa_{ni}P_n^{\alpha}Q_n^{1-\alpha}}. \tag{11}$$

Indirect utility is a monotonic function of idiosyncratic amenities ($b_{ni\omega}$) and these amenities have a Fréchet distribution. Therefore, the indirect utility for a worker living in region n and working in region i also has a Fréchet distribution given by

$$G_{ni}(U) = e^{-\Psi_{ni}U^{-\epsilon}}, \qquad \Psi_{ni} = B_{ni}\left(\kappa_{ni}P_n^{\alpha}Q_n^{1-\alpha}\right)^{-\epsilon}w_i^{\epsilon}. \tag{12}$$

Each worker selects the bilateral commute that offers her the maximum utility, where the maximum of Fréchet distributed random variables is itself Fréchet distributed. Using these distributions of utility, the probability that a worker chooses to live in location n and work in location i is

$$\lambda_{ni} = \frac{B_{ni}\left(\kappa_{ni}P_n^{\alpha}Q_n^{1-\alpha}\right)^{-\epsilon}w_i^{\epsilon}}{\sum_{r\in N}\sum_{s\in N}B_{rs}\left(\kappa_{rs}P_r^{\alpha}Q_r^{1-\alpha}\right)^{-\epsilon}w_s^{\epsilon}} \equiv \frac{\Phi_{ni}}{\Phi}. \tag{13}$$

The idiosyncratic shock to preferences $b_{ni\omega}$ implies that individual workers choose different bilateral commutes when faced with the same prices (P_n, Q_n, w_i), commuting costs (κ_{ni}) and location characteristics (B_{ni}). Other things equal, workers are more likely to live in location n and work in location i, the lower the consumption goods price index (P_n) and land prices (Q_n) in n, the higher the wages (w_i) in i , the more attractive the average amenities for this bilateral commute (B_{ni}), and the lower the commuting costs (κ_{ni}).

Summing these probabilities across workplaces i for a given residence n, we obtain the overall probability that a worker resides in location n (λ_{Rn}). Similarly, summing across residences n for a given workplace i, we obtain the overall probability that a worker works in location i (λ_{Li}). So,

$$\lambda_{Rn} = \frac{R_n}{\bar{L}} = \sum_{i \in N} \lambda_{ni} = \sum_{i \in N} \frac{\Phi_{ni}}{\Phi}, \qquad \text{and} \qquad \lambda_{Li} = \frac{L_n}{\bar{L}} = \sum_{n \in N} \lambda_{ni} = \sum_{n \in N} \frac{\Phi_{ni}}{\Phi}, \tag{14}$$

where national labor market clearing corresponds to $\sum_n \lambda_{Rn} = \sum_i \lambda_{Li} = 1$.

The average income of a worker living in n depends on the wages in all the nearby employment locations. To construct this average income of residents, note first that the probability that a worker commutes to location i conditional on living in location n is

$$\lambda_{ni|n} = \frac{B_{ni} \left(w_i / \kappa_{ni} \right)^{\epsilon}}{\sum_{s \in N} B_{ns} \left(w_s / \kappa_{ns} \right)^{\epsilon}}. \tag{15}$$

Note that equation (15) implies that, as with trade flows, commuting flows exhibit a gravity equation with an elasticity of commuting flows with respect to commuting cost (κ_{ni}) given by $-\epsilon$. The probability of commuting to location i conditional on living in location n depends on the wage (w_i), amenities (B_{ni}) and commuting costs for workplace i in the numerator ("bilateral resistance") as well as the wage (w_s), amenities (B_{ns}) and commuting costs (κ_{ns}) for all other possible workplaces s in the denominator ("multilateral resistance"). In subsection B.8 of the web appendix, we show that heterogeneous local employment elasticities are a generic prediction of the class of theoretical models consistent with a gravity equation for commuting flows.

Using these conditional commuting probabilities, we obtain the following labor market clearing condition that equates the measure of workers employed in location i (L_i) with the measure of workers choosing to commute to location i, namely,

$$L_i = \sum_{n \in N} \lambda_{ni|n} R_n, \tag{16}$$

where R_n is the measure of residents in location n. Expected worker income conditional on living in location n is then equal to the wages in all possible workplaces weighted by the probabilities of commuting to those workplaces conditional on living in n, or

$$\bar{v}_n = \sum_{i \in N} \lambda_{ni|n} w_i. \tag{17}$$

Expected worker income ($\bar{v}_n$) is high in locations that have low commuting costs (low κ_{ni}) to high-wage employment locations.

Finally, population mobility implies that expected utility is the same for all pairs of residence and workplace and equal to expected utility for the economy as a whole. That is,

$$\bar{U} = \mathbb{E}\left[U_{ni\omega}\right] = \Gamma\left(\frac{\epsilon-1}{\epsilon}\right)\left[\sum_{r\in N}\sum_{s\in N} B_{rs}\left(\kappa_{rs}P_r^{\alpha}Q_r^{1-\alpha}\right)^{-\epsilon} w_s^{\epsilon}\right]^{\frac{1}{\epsilon}} \quad \text{all } n, i \in N, \tag{18}$$

where $\mathbb{E}$ is the expectations operator and the expectation is taken over the distribution for the idiosyncratic component of utility and $\Gamma(\cdot)$ is the Gamma function.

Although expected utility is equalized across all pairs of residence and workplace locations, real wages differ as a result of preference heterogeneity. Workplaces and residences face upward-sloping supply functions for workers and residents respectively (the choice probabilities (14)). Each workplace must pay higher wages to increase commuters' real income and attract additional workers with lower idiosyncratic amenities for that workplace. Similarly, each residential location must offer a lower cost of living to increase commuters' real income and attract additional residents with lower idiosyncratic amenities for that residence. Bilateral commutes with attractive characteristics (high workplace wages and low residence cost of living) attract additional commuters with lower idiosyncratic amenities until expected utility (taking into account idiosyncratic amenities) is the same across all bilateral commutes.

2.5 General Equilibrium

The general equilibrium of the model can be referenced by the following vector of six variables $\{w_n, \bar{v}_n, Q_n, L_n, R_n, P_n\}_{n=1}^{N}$ and a scalar $\bar{U}$. Given this equilibrium vector and scalar, all other endogenous variables of the model can be determined. This equilibrium vector solves the following six sets of equations: income equals expenditure (9), average residential income (17), land market clearing (5), workplace choice probabilities ((14) for L_n), residence choice probabilities ((14) for R_n), and price indices (10). The last condition needed to determine the scalar $\bar{U}$ is the labor market clearing condition, $\bar{L} = \sum_{n\in N} R_n = \sum_{n\in N} L_n$.

Proposition 1 ***(Existence and Uniqueness)*** *If $1+\epsilon < \sigma\left(1+(1-\alpha)\epsilon\right)$ there exists a unique general equilibrium of this economy.*

All the proofs of propositions are relegated to the appendix. The condition for the existence of a unique general equilibrium in Proposition 1 is a generalization of the condition in the Helpman (1998) model to incorporate commuting and heterogeneity in worker preferences over locations. Defining $\tilde{\alpha} = \alpha/(1+1/\epsilon)$, this condition for a unique general equilibrium can be written as $\sigma(1-\tilde{\alpha}) > 1$. Assuming prohibitive commuting costs ($\kappa_{ni} \to \infty$ for $n \neq i$) and taking the limit of no heterogeneity in worker preference over locations ($\epsilon \to \infty$), this reduces to the Helpman (1998) condition for a unique general equilibrium of $\sigma(1-\alpha) > 1$.

We follow the new economic geography literature in modeling agglomeration forces through love of variety and increasing returns to scale. But the system of equations for general equilibrium in our new economic geography model is isomorphic to a version of Eaton and Kortum (2002) and Redding (2016) with commuting and external economies of scale or a version of Armington (1969) model with commuting and external economies of scale (as in Arkolakis and Allen 2014a and Arkolakis, Allen and Li 2015), as summarized in the following proposition.

Proposition 2 ***(Isomorphisms)*** *The system of equations for general equilibrium in our new economic geography model with commuting and agglomeration forces through love of variety and increasing returns to scale is isomorphic to that in a version of the Eaton and Kortum (2002) model with commuting and external economies of scale or that in a version of the Armington (1969) model with commuting and external economies of scale.*

Following a long line of research in spatial economics, we focus on versions of these models that capture congestion forces through residential land use and an inelastic supply of land. But the same system of equations for general equilibrium can be generated by models that instead directly assume a congestion force (e.g. by assuming that utility in each location depends negatively on the measure of residents in that location).

2.6 Computing Counterfactuals

We use our quantitative framework to solve for the counterfactual effects of changes in the exogenous variables of the model (productivity A_n, amenities B_{ni}, commuting costs κ_{ni}, and trade costs d_{ni}) without having to necessarily determine the unobserved values of these exogenous variables. Instead in the web appendix we show that the system of equations for the counterfactual changes in the endogenous variables of the model can be written solely in terms of the observed values of variables in an initial equilibrium (employment L_i, residents R_i, workplace wages w_n, average residential income $\bar{v}_n$, trade shares π_{ni}, and commuting probabilities λ_{ni}). This approach uses observed bilateral commuting probabilities to capture unobserved bilateral commuting costs and amenities. Similarly, if bilateral trade shares between locations are available, they can be used to capture unobserved bilateral trade frictions (as in Dekle, Eaton and Kortum 2007). However, since bilateral trade data are only available at a higher level of aggregation than the counties we consider in our data, we make some additional parametric assumptions to solve for implied bilateral trade shares between counties, as discussed below. In the model, we assume that trade is balanced so income equals expenditure. However, when taking the model to the data, we allow for intertemporal trade deficits that are treated as exogenous in our counterfactuals, as in Dekle, Eaton and Kortum (2007) and Caliendo and Parro (2015), as discussed further in the quantitative section below.

3 Data and Measurement

Our empirical analysis combines data from a number of different sources for the United States. From the Commodity Flow Survey (CFS), we use data on bilateral trade and distances shipped for 123 CFS regions. From the American Community Survey (ACS), we use data on commuting probabilities between counties. From the Bureau of Economic Analysis (BEA), we use data on employment and wages by workplace. We combine these data sources with a variety of other Geographical Information Systems (GIS) data. We use our data on employment and commuting to calculate the implied number of residents and their average income by county. First, from commuting market clearing (16), we obtain the number of residents (R_n) using data on the number of workers (L_n) and commuting probabilities conditional on living in each location ($\lambda_{ni|n}$). Second, we use these conditional commuting probabilities, together with county wages, to obtain average residential income ($\bar{v}_n$) as defined in (17).

3.1 Gravity in Goods Trade

In the Commodity Flow Survey (CFS) data, we observe bilateral trade flows and distances shipped between 123 CFS regions and trade deficits for each these CFS regions.[15] To quantify the model at the county level, we allocate the deficit for each CFS region across the counties within that region according to their shares of CFS residential income (as measured by $\bar{v}_i R_i$). Using the resulting trade deficits for each county (D_i), we solve the equality between income and expenditure (9) for unobserved county productivities (A_i):

$$w_i L_i - \sum_{n \in N} \frac{L_i \left(d_{ni} w_i / A_i\right)^{1-\sigma}}{\sum_{k \in N} L_k \left(d_{nk} w_k / A_k\right)^{1-\sigma}} \left[\bar{v}_n R_n + D_n\right] = 0, \tag{19}$$

where we observe (or have solved for) wages (w_i), employment (L_i), average residential income ($\bar{v}_i$), residents (R_i) and trade deficits (D_i).

Given the elasticity of substitution (σ), our measures for (w_i, L_i, $\bar{v}_i$, R_i, D_i) and a parameterization of trade costs ($d_{ni}^{1-\sigma}$), equation (19) provides a system of N equations that can be solved for a unique vector of N unobserved productivities (A_i). We prove this formally in the next proposition.

Proposition 3 ***(Productivity Inversion)*** *Given the elasticity of substitution (σ), our measures of wages, employment, average residential income, residents and trade deficits $\{w_i, L_i, \bar{v}_i, R_i, D_i\}$, and a parameterization of trade costs ($d_{ni}^{1-\sigma}$), there exist unique values of the unobserved productivities (A_i) for each location i that are consistent with the data being an equilibrium of the model.*

The resulting solutions for productivities (A_i) capture characteristics (e.g. natural resources) that make a location more or less attractive for employment conditional on the observed data and the parameterized

[15] Other recent studies using the CFS data include Caliendo et. al (2014), Duranton, Morrow and Turner (2014) and Dingel (2015). The CFS is a random sample of plant shipments within the United States (foreign trade shipments are not included). CFS regions are the smallest geographical units for which this random sample is representative, which precludes constructing bilateral trade flows between smaller geographical units using the sampled shipments.

values of trade costs. These characteristics include access to international markets. To the extent that such international market access raises employment (L_i), and international trade flows are not captured in the CFS, this will be reflected in the model in higher productivity (A_i) to rationalize the higher observed employment. Having recovered for these unique unobserved productivities (A_i), we can solve for the implied bilateral trade flows between counties using equation (8) and $X_{ni} = \pi_{ni}\bar{v}_n R_n$. We use these solutions for bilateral trade between counties in our counterfactuals for changes in the model's exogenous variables, as discussed above.

To parameterize trade costs ($d_{ni}^{1-\sigma}$), we assume a central value for the elasticity of substitution between varieties from the existing empirical literature of $\sigma = 4$, which is in line with the estimates of this parameter using international trade data in Broda and Weinstein (2006).[16] We model bilateral trade costs (d_{ni}) as a function of distance. For bilateral pairs with positive trade, we assume that bilateral trade costs are a constant elasticity function of distance and a stochastic error ($d_{ni} = dist_{ni}^{\psi}\tilde{e}_{ni}$). For bilateral pairs with zero trade, the model implies prohibitive trade costs ($d_{ni} \to \infty$).[17] Taking logarithms in the trade share (8) for pairs with positive trade, we obtain that the value of bilateral trade between source i and destination n (X_{ni}) is given by

$$\log X_{ni} = \zeta_n + \chi_i - (\sigma - 1)\,\psi \log dist_{ni} + \log e_{ni}, \tag{20}$$

where the source fixed effect (χ_i) controls for employment, wages and productivity (L_i, w_i, A_i) in the source location i; the destination fixed effect (ζ_n) controls for average income, $\bar{v}_n$, residents, R_n, and multilateral resistance (as captured in the denominator of equation (8)) in the destination location n; and $\log e_{ni} = (1-\sigma)\log \tilde{e}_{ni}$.

Estimating the gravity equation (20) for all bilateral pairs with positive trade using OLS, we find a regression R-squared of 0.83. In Figure 1, we display the conditional relationship between the log value of trade and log distance, after removing source and destination fixed effects from both log trade and log distance. We find that the log linear functional form provides a good approximation to the data, with a tight and approximately linear relationship between the two variables. We estimate a coefficient on log distance of $-(\sigma-1)\,\psi = -1.29$. For our assumed value of $\sigma = 4$, this implies an elasticity of trade costs with respect to distance of $\psi = 0.43$. The tight linear relationship in Figure 1, makes us confident in this parametrization of trade costs as $d_{ni}^{1-\sigma} = dist_{ni}^{-1.29}$ as a way of using equation (19) to solve for unobserved productivities (A_i).

As an alternative check on our specification, in Figure 2, we aggregate the model's predictions for trade between counties within pairs of CFS regions and compare these predictions to the corresponding values in the data. The only way in which we used the data on trade between CFS regions was to estimate the distance elasticity $-(\sigma-1)\,\psi = -1.29$. Given this distance elasticity, we use the goods market

[16] This assumed value implies an elasticity of trade with respect to trade costs of $-(\sigma-1) = 3$, which is close to the central estimate of this parameter of 4.12 in Simonovska and Waugh (2014).

[17] One interpretation is that trade requires prior investments in transport infrastructure that are not modeled here. For bilateral pairs for which these investments have been made, trade can occur subject to finite costs. For other bilateral pairs for which they have not been made, trade is prohibitively costly. We adopt our specification for tractability, but other rationalizations for zero trade flows include non-CES preferences or granularity.

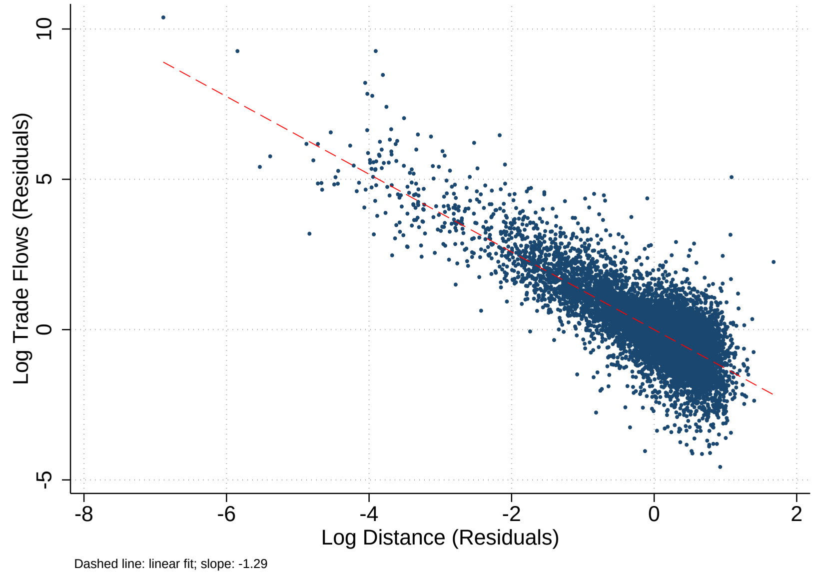


Figure 1: Gravity in Goods Trade Between CFS Regions

Figure 2: Bilateral Trade Shares in the Model and Data

clearing condition (19) to solve for productivities and generate the model's predictions for bilateral trade between counties and hence CFS regions. Therefore, the model's predictions and the data can differ from one another. Nonetheless, we find a strong and approximately log linear relationship between the model's predictions and the data, which is tighter for the larger trade values that account for most of aggregate trade.

3.2 The Magnitude and Gravity of Commuting Flows

We start by providing evidence on the quantitative relevance of commuting as a source of spatial linkages between counties and CZ's. To do so, we use data from the American Community Survey (ACS), which reports county-to-county worker flows for 2006-2010. To abstract from business trips that are not between a worker's usual place of residence and workplace, we define commuting flows as those of less than 120 kilometers in each direction (a round trip of 240 kilometers).[18]

In Table 1, we report some descriptive statistics for these commuting flows. We find that commuting beyond county boundaries is substantial and varies in importance across locations. For the median county, around 27 percent of its residents work outside the county (first row, fifth column) and around 20 percent of its workers live outside the county (second row, fifth column). For the county at the 90th percentile, these two figures rise to 53 and 37 percent respectively (seventh column, first and second rows respectively). Consequently, we find substantial dispersion across counties in the ratio of employment to residents (L/R), which captures the extent to which a county is an importer of commuters ($L/R < 1$) or an exporter of commuters ($L/R > 1$). This ratio ranges from 0.67 at the 10th percentile to 1.11 at the 90th percentile (third row, columns three and seven respectively). In subsection C.2 of the web appendix, we show that commuting linkages are not only heterogeneous across counties, but are also hard to explain with standard empirical controls, such as land area, size or supply elasticities for developed land.

	Min	p5	p10	p25	p50	p75	p90	p95	Max	Mean
Commuters from Residence	0.00	0.03	0.06	0.14	0.27	0.42	0.53	0.59	0.82	0.29
Commuters to Workplace	0.00	0.03	0.07	0.14	0.20	0.28	0.37	0.43	0.81	0.22
Employment/Residents	0.26	0.60	0.67	0.79	0.92	1.02	1.11	1.18	3.88	0.91
Outside CZ | Total (Res)	0.00	0.02	0.04	0.14	0.33	0.58	0.79	0.89	1.00	0.37
Outside CZ | Total (Work)	0.00	0.03	0.08	0.19	0.37	0.55	0.73	0.82	1.00	0.39
CZ Employment/Residents	0.63	0.87	0.91	0.97	1.00	1.01	1.03	1.04	1.12	0.98

Tabulations on 3,111 counties and 709 commuting zones. The first row shows the fraction of residents that work outside the county. The second row shows the fraction of workers who live outside the county. The third row shows the ratio of county employment to county residents. The forth row shows the fraction of residents that work outside the county who also work outside the county's CZ. The fifth row shows the fraction of workers that live outside the county who also live outside the county's CZ. The sixth row shows the ratio of CZ employment to CZ residents across all 709 CZ. p5, p10 etc refer to the 5th, 10th etc percentiles of the distribution.

Table 1: Commuting Across Counties and Commuting Zones

One might think that using commuting zones circumvents the need to incorporate commuting in the analysis since these areas are designed to encompass most commuting flows between counties. Nevertheless, we find that CZ's provide an imperfect measure of local labor markets, with substantial commuting beyond CZ boundaries that again varies in importance across locations. For the median county, around 33 percent of the workers who commute outside their county of residence also commute outside their CZ of residence (fourth row, fifth column), while around 37 percent of the workers who commute outside their

[18]The majority of commutes are less than 45 minutes in each direction (Duranton and Turner 2011), with commutes of 120 minutes in each direction rare. In our analysis, we measure distance between counties' centroids. We choose the 120 kilometers threshold based on a change in slope of the relationship between log commuters and log distance at this distance threshold. See the web appendix for further discussion.

county of workplace also commute outside their CZ of workplace (fifth row, fifth column). For the CZ at the 90th percentile, these two figures rise to 79 and 73 percent respectively (seventh column). Although the ratio of employment to residents (L/R) by construction varies less across CZs than across counties, we still find substantial variation from 0.63 to 1.12, which we show below is sufficient to generate substantial heterogeneity in local employment elasticities.

To provide further evidence on commuting that is independent of our model, we decompose changes in employment in each county over the period 1990-2007 into the percentage contributions of migration and commuting. For the median county, around 39 percent of the observed changes in employment are due to changes in commuting patterns, with this percentage varying substantially across counties from close to zero to close to one. For more than one third of counties, the contribution from commuting is larger than that from migration.[19] Taken together, these results highlight the quantitative relevance of commuting as a source of spatial linkages.

Using land market clearing (5) and the price index (10), gravity equation for the commuting probability (13) in the model can be written as

$$\lambda_{ni} - \frac{B_{ni}\kappa_{ni}^{-\epsilon}\left(\frac{L_n}{\pi_{nn}}\right)^{-\frac{\alpha\epsilon}{\sigma-1}} A_n^{\alpha\epsilon} w_n^{-\alpha\epsilon} \bar{v}_n^{-\epsilon(1-\alpha)} \left(\frac{R_n}{H_n}\right)^{-\epsilon(1-\alpha)} w_i^{\epsilon}}{\sum_{r\in N}\sum_{s\in N} B_{rs}\kappa_{rs}^{-\epsilon}\left(\frac{L_r}{\pi_{rr}}\right)^{-\frac{\alpha\epsilon}{\sigma-1}} A_r^{\alpha\epsilon} w_r^{-\alpha\epsilon} \bar{v}_r^{-\epsilon(1-\alpha)} \left(\frac{R_r}{H_r}\right)^{-\epsilon(1-\alpha)} w_s^{\epsilon}} = 0. \tag{21}$$

The commuting probabilities (21) provide a system of $N \times N$ equations that can be solved for a unique matrix of $N \times N$ unobserved amenities (B_{ni}). The next proposition shows this formally.

Proposition 4 ***(Amenities Inversion)*** *Given the share of consumption goods in expenditure (α), the heterogeneity in location preferences (ϵ), the observed data on wages, employment, trade shares, average residential income, residents and land area {w_i, L_i, π_{ii}, $\bar{v}_i$, R_i, H_i}, and a parameterization of commuting costs ($\kappa_{ni}^{-\epsilon}$), there exist unique values of the unobserved amenities (B_{ni}) for each pair of locations n and i that are consistent with the data being an equilibrium of the model.*

The resulting solutions for amenities (B_{ni}) capture unobserved factors that make a pair of residence and workplace locations more or less attractive conditional on the observed wages, employment, trade shares, average residential income, residents, and land area, as well as the parameterized commuting costs (e.g. attractive scenery and differences in transport infrastructure that are not captured in the parameterized commuting costs). Together unobserved productivity (A_i) and amenities (B_{ni}) correspond to structural residuals that ensure that the model exactly replicates the observed data given the parameters and our assumptions for trade and commuting costs.

Bilateral commuting flows in the gravity equation (13) depend on a composite parameter that captures the ease of commuting $\tilde{B}_{ni} \equiv B_{ni}\kappa_{ni}^{-\epsilon}$. For bilateral pairs with positive commuting, we partition this ease

[19] In subsections subsection C.1 the web appendix describes in detail the methodology we use to compute these statistics as well as other moments of the distribution.

of commuting ($\tilde{B}_{ni}$) into four components: (i) a residence component ($\mathbb{B}_n$), (ii) a workplace component ($\mathbb{B}_i$), (iii) a component that is related to distance (dist_{ni}^{ϕ}), and (iv) an orthogonal component ($\mathbb{B}_{ni}$) such that:

$$\log \tilde{B}_{ni} = \log(B_{ni}\kappa_{ni}^{-\epsilon}) = \log \mathbb{B}_n + \log \mathbb{B}_i + \phi \log\left(\text{dist}_{ni}\right) + \log \mathbb{B}_{ni}. \tag{22}$$

We can always undertake this statistical decomposition of the ease of commuting ($\log \tilde{B}_{ni}$), where the error term ($\log \mathbb{B}_{ni}$) is orthogonal to distance by construction, because the reduced-form coefficient on log distance (ϕ) captures any correlation of either log bilateral amenities ($\log B_{ni}$) and/or log bilateral commuting costs ($\log(\kappa_{ni}^{-\epsilon})$) with log distance. For bilateral pairs with zero commuting, the model implies negligible amenities ($B_{ni} \rightarrow 0$) and/or prohibitive commuting costs ($\kappa_{ni} \rightarrow \infty$), which corresponds to $\mathbb{B}_{ni} \rightarrow 0$ in our decomposition (22).[20]

In the first step of our gravity equation estimation, we use this decomposition (22) and our expression for commuting flows (13) to estimate the reduced-form distance coefficient (ϕ):

$$\log \lambda_{ni} = g_0 + \eta_n + \mu_i - \phi \log dist_{ni} + \log \mathbb{B}_{ni}, \tag{23}$$

where the residence fixed effect (η_n) captures the consumption goods price index (P_n), the price of residential land (Q_n), and the residence component of the ease of commuting ($\mathbb{B}_n$); the workplace fixed effect (μ_i) captures the wage (w_i) and the workplace component of commuting frictions ($\mathbb{B}_i$); the constant g_0 captures the denominator of λ_{ni} and is separately identified because we normalize the residence and workplace fixed effects to sum to zero; and the error term ($\log \mathbb{B}_{ni}$) is orthogonal to log distance, because all effects of log distance on the composite ease of commuting are captured in the reduced-form distance coefficient (ϕ).[21]

Estimating the gravity equation (23) for all bilateral pairs with positive commuters using OLS, we find a regression R-squared of 0.80. In Figure 3, we display the conditional relationship between log commuters and log distance, after removing residence and workplace fixed effects from both log commuters and log distance. We find that the log linear functional form provides a good approximation to the data, with a tight and approximately linear relationship between the two variables, with an estimated coefficient on log distance of $-\phi = -4.43$. This estimated coefficient is substantially larger than the corresponding coefficient for trade in goods of $-(\sigma - 1)\psi = -1.29$, which is consistent with the view that transporting people is considerably more costly than transporting goods, in line with the substantial opportunity cost of time spent commuting.

To identify the Fréchet shape parameter (ϵ), the second step of our gravity equation estimation uses additional structure from the model, which implies that the workplace fixed effects μ_i depend on wages

[20]As for goods trade above, one interpretation is that commuting requires prior investments in transport infrastructure that are not modeled here. We adopt our specification for tractability, but other explanations for zero commuting flows include a support for the distribution of idiosyncratic preferences that is bounded from above or granularity.

[21]In subsection B.10 of the appendix, we generalize this specification to introduce congestion that is a power function of the volume of commuters. We show that this generalization affects the interpretation of the estimated coefficients in the gravity equation, but leaves the model's prediction of a heterogeneous local employment elasticities unchanged.

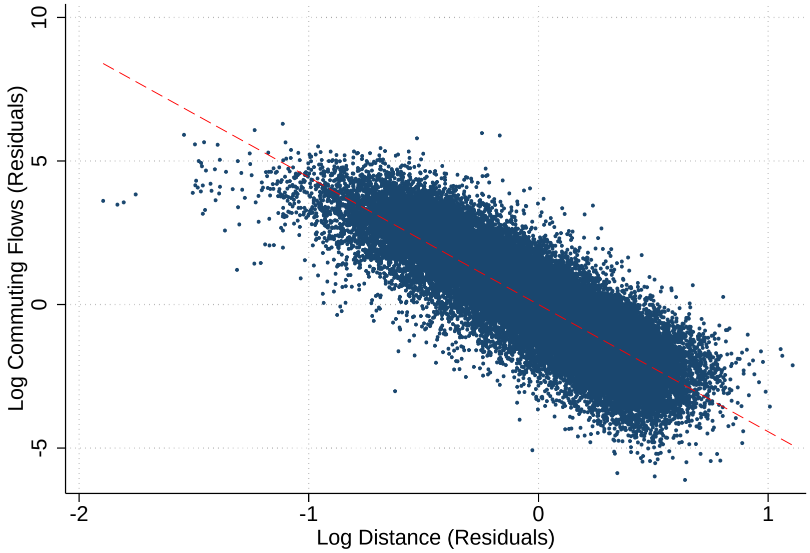


Figure 3: Gravity in Commuting Between Counties

(w_i) and the workplace component of the ease of commuting ($\mathbb{B}_i$):

$$\log \lambda_{ni} = g_0 + \eta_n + \epsilon \log w_i - \phi \log dist_{ni} + \log \mathbb{B}_i + \log \mathbb{B}_{ni} \tag{24}$$

We estimate the gravity equation (24) imposing $\phi = 4.43$ from our estimates above and identify ϵ from the coefficient on wages. Estimating (24) using OLS is potentially problematic, because workplace wages (w_i) depend on the supply of commuters, which in turn depends on amenities that appear in the error term ($\log \mathbb{B}_i + \log \mathbb{B}_{ni}$). Therefore we instrument $\log w_i$ with the log productivities $\log A_i$ that we recovered from the condition equating income and expenditure above, using the fact that the model implies that productivity satisfies the exclusion restriction of only affecting commuting flows through wages. Our Two-Stage-Least-Squares estimate of the Fréchet shape parameter for the heterogeneity of worker preferences is $\epsilon = 3.30$.[22] The tight fit shown in Figure 3, makes us confident that our parametrization of commuting costs and amenities in terms of distance fits the data quite well.

For the one remaining parameter of the model, the share of housing in consumer expenditure, we assume a central value from Bureau of Economic Analysis of $1 - \alpha = 0.40$ percent. Using our assumption of Cobb-Douglas utility and interpretation of land as geographical land area, in subsection C.3 of the web appendix, we show that the model's predictions for land prices are strongly positively correlated with observed county median housing values. In the next section, we also relax these assumptions to introduce a positive supply elasticity for developed land.

[22]We find that the Two-Stage-Least-Squares estimates are larger than the OLS estimates, consistent with the idea that bilateral commutes with attractive amenities have a higher supply of commuters and hence lower wages. The first-stage F-Statistic for productivity is 228.1, confirming that productivity is a powerful instrument for wages. Note that one could have estimated jointly ϕ and ϵ from the restricted equation (24) directly. Our approach, however, imposes only the minimal set of necessary restrictions at every step: we estimate a flexible gravity structure to identify ϕ in (23), and a slightly less general specification (where workplace fixed effects are restricted to capture only variation in workplace wages) to identify ϵ. Estimating the restricted equation (24) directly would yield very similar results: we find $\epsilon = 3.19$ and $\phi = 4.09$.

4 Local Employment Elasticities

Having quantified the model, we now explore its implications for local employment elasticities. In subsection 4.1, we undertake counterfactuals to evaluate the elasticity of local employment in each county with respect to a local labor demand shock (a productivity shock in the model). We find that the model predicts substantial heterogeneity in this elasticity across counties. In subsection 4.2, we show that this heterogeneity is not well explained by standard empirical controls, but is well explained by measures of connections to other counties in commuting networks. In subsection 4.3, we show that these results are robust to introducing heterogeneity in the supply of developed land across locations. In subsection 4.4, we demonstrate that these heterogeneous local employment elasticities correspond to heterogeneous treatment effects in reduced-form regressions for the impact of local labor demand shocks. In subsection 4.5, we provide independent evidence in support of these predictions of the model using the natural experiment of million dollar plants (MDPs), as examined in Greenstone, Hornbeck and Moretti (2010). We find heterogeneous treatment effects that take exactly the form implied by the model, such that the opening of MDPs has larger effects on employment in counties with more open commuting markets.

4.1 Heterogeneity in Local Employment Elasticities

To provide evidence on local employment elasticities, we compute 3,111 counterfactual exercises where we shock each county with a 5 percent productivity shock (holding productivity in all other counties and holding all other exogenous variables constant).[23] Figure 4 shows the estimated kernel density for the distribution of the general equilibrium elasticity of employment with respect to the productivity shock across these treated counties (black line). We also show the 95 percent confidence intervals around this estimated kernel density (gray shading). The mean estimated local employment elasticity of around 1.52 is greater than one because of home market effects and commuting. Around this mean, we find substantial heterogeneity in the predicted effects of the productivity shock, which vary from close to 0.5 to almost 2.5. This variation is surprisingly large. It implies that taking a local employment elasticity estimated for one group of counties and applying that elasticity to another group of counties can lead to substantial discrepancies between the true and predicted impacts of a productivity shock.

To provide a point of comparison, Figure 4 also includes the general equilibrium elasticity of residents in a county with respect to the same 5 percent productivity shock in that county (again holding other parameters constant). Again we show the estimated kernel density across the 3,111 treated counties (black line) and the 95 percent confidence intervals (gray shading). We find substantial differences between the employment and residents elasticity, with the residents elasticity having less dispersion and ranging from around 0.2 to 1.2. Since employment and residents can only differ through commuting, this by itself suggests that the heterogeneity in the local employment elasticity in response to the productivity shock is largely driven by commuting links between counties. In subsection C.5 of the web appendix, we show that this heterogeneity remains if we shock counties with patterns of spatially correlated shocks

[23] We have experimented with shocks of 1% and 10% as well, with essentially unchanged results.

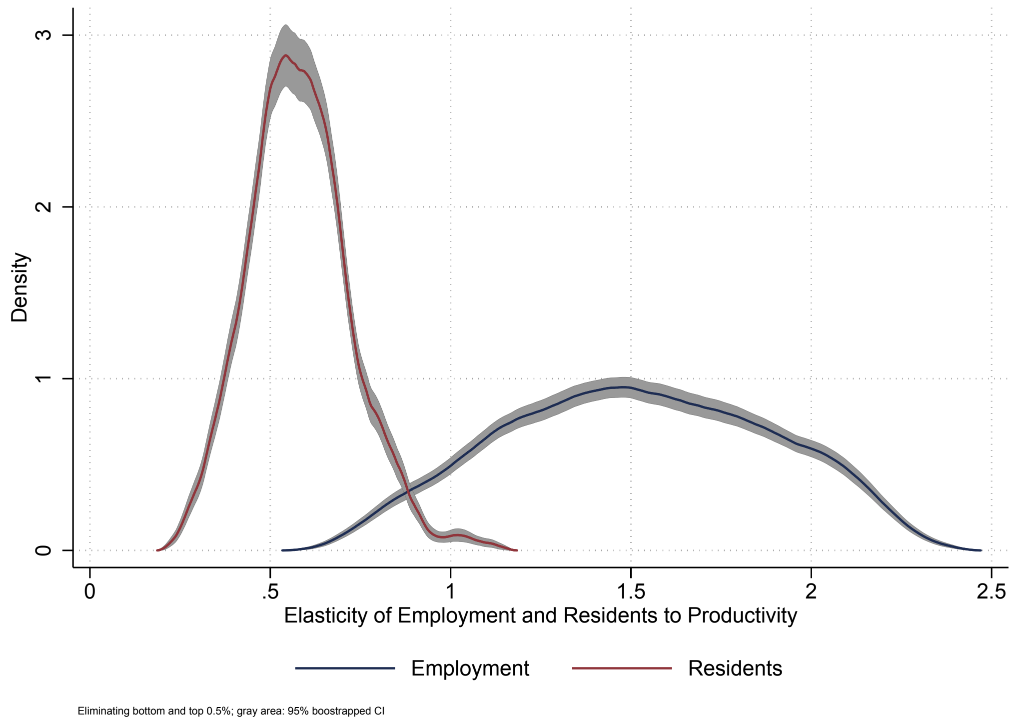


Figure 4: Kernel density for the distribution of employment and residents elasticities in response to a productivity shock across counties

reproducing the industrial composition of the U.S. economy. In subsection C.7 of the web appendix, we show that we continue to find substantial variation in local employment elasticities if we replicate our entire analysis for CZs rather than counties. Both sets of results are consistent with the fact that heterogeneous local employment elasticities are a generic prediction of a gravity equation for commuting. While CZ boundaries are drawn to minimize commuting, they inevitably cannot perfectly capture the rich geography of commuting flows implied by the gravity equation.

4.2 Explaining the Heterogeneity in Local Employment Elasticities

We now use the model to provide intuition on the determinants of the general equilibrium local employment elasticities, $\frac{dL_i}{dA_i}\frac{A_i}{L_i}$. We also use the structure of the model to determine a set of variables that can be used empirically to account for the estimated heterogeneity in the distribution of local employment elasticities. To do so, we compute partial equilibrium elasticities of own wages and own employment with respect to the productivity shock. These partial equilibrium elasticities capture the direct effect of a productivity shock on wages, employment and residents in the treated location, holding constant all other endogenous variables at their values in the initial equilibrium.[24] Hence, although potentially useful to provide intuition, or as empirical controls, they do not account for all the rich set of interactions in the model captured by the general equilibrium elasticities presented in Figure 4.

If we hold constant all variables except for w_i, L_i, and R_i in the treated county i, the partial elasticity

[24]See Section B.7 in the web appendix for the derivation of these partial equilibrium elasticities.

of employment with respect to the productivity shock is the product of the partial elasticity of employment with respect to wages and the partial elasticity of wages with respect to the productivity shock[25]

$$\frac{\partial L_i}{\partial A_i}\frac{A_i}{L_i} = \frac{\partial L_i}{\partial w_i}\frac{w_i}{L_i} \cdot \frac{\partial w_i}{\partial A_i}\frac{A_i}{w_i}. \tag{25}$$

From the equality between income and expenditure (9), the partial elasticity of wages with respect to the productivity is given by

$$\frac{\partial w_i}{\partial A_i}\frac{A_i}{w_i} = \frac{(\sigma-1)\sum_{n\in N}(1-\pi_{ni})\,\xi_{ni}}{\left[1+(\sigma-1)\sum_{n\in N}(1-\pi_{ni})\,\xi_{ni}\right]+\left[1-\sum_{n\in N}(1-\pi_{ni})\,\xi_{ni}\right]\frac{\partial L_i}{\partial w_i}\frac{w_i}{L_i}-\xi_{ii}\frac{\partial R_i}{\partial w_i}\frac{w_i}{R_i}}. \tag{26}$$

where $\xi_{ni} = \pi_{ni}\alpha\bar{v}_n R_n / w_i L_i$ is the share of location i's revenue from market n.

The intuition for the response of wages to the productivity shock can be seen most clearly for the case when $\frac{\partial L_i}{\partial w_i}\frac{w_i}{L_i} \approx 0$ and $\frac{\partial R_i}{\partial w_i}\frac{w_i}{R_i} \approx 0$. From the terms in $(\sigma-1)\sum_{n\in N}(1-\pi_{ni})\,\xi_{ni}$, the elasticity of wages with respect to productivity is high when location i accounts for a small share of expenditure (small π_{ni}) in markets n that account for a large share of its revenue (high ξ_{ni}). In these circumstances, the productivity shock reduces the prices of location i's goods and results in only a small reduction in the goods price indices (small π_{ni}) in its main markets (high ξ_{ni}).[26] Therefore the productivity shock leads to a large increase in the demand for location i's goods and hence in its wages. Thus $\sum_{n\in N}(1-\pi_{ni})\,\xi_{ni}$ provides a measure of location i's linkages to other locations in goods markets.

From the commuting market clearing condition (16), the partial elasticity of employment with respect to wages is

$$\frac{\partial L_i}{\partial w_i}\frac{w_i}{L_i} = \epsilon\sum_{n\in N}\left(1-\lambda_{ni|n}\right)\vartheta_{ni} + \vartheta_{ii}\left(\frac{\partial R_i}{\partial w_i}\frac{w_i}{R_i}\right), \tag{27}$$

where $\vartheta_{ni} = \lambda_{ni|n} R_n / L_i$ is the share of commuters from residence n in workplace i's employment.

The intuition for the response of employment to wages can be seen most clearly when we set $\frac{\partial R_i}{\partial w_i}\frac{w_i}{R_i} \approx 0$. From the term in $\epsilon\sum_{n\in N}\left(1-\lambda_{ni|n}\right)\vartheta_{ni}$, the elasticity of employment with respect to wages is high when workplace i employs a small share of commuters (small $\lambda_{ni|n}$) from residences n that supply a large fraction of its employment (high ϑ_{ni}). In these circumstances, location i's wage increase makes it a more attractive workplace and results in only a small increase in commuting market access (small $\lambda_{ni|n}$) in its main sources of commuters (high ϑ_{ni}).[27] Therefore the increase in wages leads to a large increase in commuters to workplace i and hence in its employment. Thus $\sum_{n\in N}\left(1-\lambda_{ni|n}\right)\vartheta_{ni}$ provides a measure of workplace i's linkages to other locations through commuting markets. In subsection B.8 of the web appendix, we show that this partial elasticity takes the same form in the entire class of theoretical models

[25] Note that we use the partial derivative symbol, $\frac{\partial L_i}{\partial A_i}\frac{A_i}{L_i}$, to denote the partial equilibrium elasticity when we allow w_i, L_i, and R_i to change but keep other variables in all other counties fixed.

[26] These price indices summarize the price of competing varieties in each market. Note that the elasticity of the price index (10) in location n with respect to wages in location i is $(\partial P_n/\partial w_i)(w_i/P_n) = \pi_{ni}$.

[27] Commuting market access appears in the numerator of the residential choice probabilities (λ_{Rn} in (14)) and summarizes access to employment opportunities: $W_n = \left[\sum_{s\in N} B_{ns}(w_s/\kappa_{ns})^{\epsilon}\right]^{1/\epsilon}$. Note that the elasticity of commuting market access in location n with respect to wages in location i is $(\partial W_n/\partial w_i)(w_i/W_n) = \lambda_{ni|n}$.

consistent with a gravity equation for commuting flows.

Using (14), the partial elasticity of residents with respect to wages is

$$\frac{\partial R_i}{\partial w_i}\frac{w_i}{R_i} = \epsilon\left(\frac{\lambda_{ii}}{\lambda_{Ri}} - \lambda_{Li}\right), \tag{28}$$

which also has an intuitive interpretation. A higher wage in location i makes it a more attractive workplace and increases its employment. Whether this increase in location i's employment leads to an increase in its share of residents depends on the fraction of residents who work locally ($\lambda_{ii}/\lambda_{Ri}$) relative to location i's overall share of employment (λ_{Li}). Thus $(\lambda_{ii}/\lambda_{Ri} - \lambda_{Li})$ provides a measure of location i's linkages to other locations through migration.

Combining the three elasticities (26), (27), and (28), we obtain closed-form solutions for the partial equilibrium elasticities of wages, employment and residents to a productivity shock. These partial equilibrium elasticities depend solely on the observed values of variables in the initial equilibrium: residential employment shares (λ_{Ri}), conditional commuting probabilities ($\lambda_{ni|n}$), employment shares (ϑ_{ni}), trade shares (π_{ni}) and revenue shares (ξ_{ni}).

When we undertake our counterfactuals, we solve for the full general equilibrium effect of the productivity shock to each county. But these partial equilibrium elasticities in terms of observed variables have substantial explanatory power in predicting the impact of the productivity shock across locations, as now shown in Table 2. In Column (1) we regress these elasticities on a constant, which captures the mean employment elasticity across the 3,111 treated counties. In Columns (2) through (4) we attempt to explain the heterogeneity in local employment elasticities using standard county controls. In Column (2) we include log county employment as a control for the size of economic activity in a county. In Column (3) we also include log county wages and log county land area. In Column (4) we also include the average wage and total employment in neighboring counties. Although these controls are all typically statistically significant, we find that they are not particularly successful in explaining the variation in employment elasticities. Adding a constant and all these controls yields an R-squared of only about 0.5 in Column (4). Clearly, there is substantial variation not captured by these controls.

In the remaining columns of the table we attempt to explain the heterogeneity in local employment elasticities using the partial equilibrium elasticities derived above. In Column (5) we first use the intuition (obtained by comparing the distributions in Figure 4) that commuting is essential to explain this elasticity. As a summary statistic of the lack of commuting links of a county we use $\lambda_{ii|i}$, namely, the share of workers that work in i conditional on living in i. The weaker the commuting links of a county, the higher $\lambda_{ii|i}$, which reduces the local employment elasticity of that county. This is exactly what we find in Column (5). Furthermore, this variable alone yields an R-squared of 0.89, nearly double the R-squared in the regression where we include all the standard controls.[28] Therefore, although the model incorporates several forms of spatial linkages (including trade and migration), we find that the heterogeneity in local

[28]To provide further evidence on the magnitude of these effects, Table 6 in Section C of the web appendix reports the same regressions as in Table 2 but using standardized coefficients. We find that a one standard deviation change in commuting ($\lambda_{ii|i}$) leads to around a one standard deviation change in the local employment elasticity.

employment elasticities is mainly explain by commuting linkages, which is consistent with our gravity equation estimates, where commuting is substantially more local (higher distance coefficient) than goods trade.

The partial equilibrium local elasticities computed above allow us to do better than just adding a summary measure of commuting links as the explanatory variable. In Column (6) we relate the variation in local employment elasticities to the measure of commuting linkages suggested by the model, $\sum_{n\in N}\left(1-\lambda_{ni|n}\right)\vartheta_{ni}$. We also add the measures of migration and trade linkages suggested by the model, $(\lambda_{ii}/\lambda_{Ri}-\lambda_{Li})$ and $\frac{\partial w_i}{\partial A_i}\frac{A_i}{w_i}$. Including these partial equilibrium measures of linkages further increases the R-squared to around 93 percent of the variation in the general equilibrium elasticity. Counties that account for a small share of commuters (small $\lambda_{ni|n}$) from their main suppliers of commuters (high ϑ_{ni}) have higher employment elasticities. In Column (7), we use the product of $\frac{\partial w_i}{\partial A_i}\frac{A_i}{w_i}$ and the first two terms rather than each term separately. This restriction yields similar results and confirms the importance of commuting linkages and, to a lesser extent, the interaction between migration and goods linkages. Finally, in the last two columns we combine these partial equilibrium elasticities with the standard controls we used in the first four columns. Clearly, although all variables are significant, these standard controls add little once we control for the partial equilibrium elasticities.

In sum, Table 2 shows that the heterogeneity in partial equilibrium elasticities is not well explained by standard county controls. In contrast, adding a summary statistic of commuting, or the partial equilibrium elasticities we propose above, can go a long way in explaining the heterogenous response of counties to productivity shocks.

4.3 Positive Developed Land Supply Elasticities

In the baseline version of the model, we interpret the non-traded amenity as simply land, which is in perfectly inelastic supply. In this section of the theoretical appendix, we develop an extension of the model, in which we interpret the non-traded amenity as "developed" land and allow for a positive developed land supply elasticity that can differ across locations.

We continue to assume the same specification of preferences, production and commuting decisions as in Section 2 of the main paper. We introduce a positive developed land supply elasticity by following Saiz (2010) in assuming that the supply of land (H_n) for each residence n depends on the endogenous price of land (Q_n) as well as on the exogenous characteristics of locations ($\bar{H}_n$):

$$H_n = \bar{H}_n Q_n^{\eta_n}, \tag{29}$$

where $\eta_n \geq 0$ is the developed land supply elasticity, which we allow to vary across locations; $\eta_n = 0$ corresponds to the special case of a perfectly inelastic land supply; and $\eta_n \to \infty$ represents the extreme case of a perfectly elastic land supply. This introduction of this positive and heterogeneous developed land supply elasticity only affects one of the conditions for general equilibrium in the model, namely the land market clearing condition. Using the supply function for land (29) in the land market clearing condition

	1	2	3	4	5	6	7	8	9
Dependent Variable:				Elasticity of Employment					
$\log L_i$		-0.003	0.009*	-0.054**				0.037**	0.033**
		(0.005)	(0.004)	(0.005)				(0.002)	(0.002)
$\log w_i$			-0.201**	-0.158**				-0.257**	-0.263**
			(0.028)	(0.027)				(0.009)	(0.009)
$\log H_i$			-0.288**	-0.172**				0.003	0.009**
			(0.006)	(0.009)				(0.003)	(0.003)
$\log L_{,-i}$				0.118**				-0.027**	-0.027**
				(0.007)				(0.003)	(0.003)
$\log \bar{w}_{-i}$				0.204**				0.163**	0.207**
				(0.047)				(0.015)	(0.015)
$\lambda_{ii\|i}$					-2.047**				
					(0.013)				
$\sum_{n\in N}\left(1-\lambda_{ni\|n}\right)\vartheta_{ni}$						2.784**		2.559**	
						(0.092)		(0.098)	
$\vartheta_{ii}\left(\frac{\lambda_{ii}}{\lambda_{Ri}}-\lambda_{Li}\right)$						0.915**		0.605**	
						(0.090)		(0.096)	
$\frac{\partial w_i}{\partial A_i}\frac{A_i}{w_i}$						-1.009**		-0.825**	
						(0.046)		(0.053)	
$\frac{\partial w_i}{\partial A_i}\frac{A_i}{w_i}\cdot\sum_{r\in N}\left(1-\lambda_{rn\|r}\right)\vartheta_{rn}$							1.038**		1.100**
							(0.036)		(0.048)
$\frac{\partial w_i}{\partial A_i}\frac{A_i}{w_i}\cdot\vartheta_{ii}\left(\frac{\lambda_{ii}}{\lambda_{Ri}}-\lambda_{Li}\right)$							-0.818**		-0.849**
							(0.036)		(0.047)
constant	1.515**	1.545**	5.683**	1.245**	2.976**	0.840**	1.553**	1.861**	2.064**
	(0.007)	(0.044)	(0.275)	(0.437)	(0.009)	(0.084)	(0.035)	(0.171)	(0.152)
R^2	0.00	0.00	0.40	0.51	0.89	0.93	0.93	0.95	0.95
N	3,111	3,111	3,111	3,081	3,111	3,111	3,111	3,081	3,081

In this table, $L_{,-i} \equiv \sum_{n:d_{ni}\leq 120, n\neq i} L_n$ is the total employment in i neighbors whose centroid is no more than 120km away; $\bar{w}_{-i} \equiv \sum_{n:d_{ni}\leq 120, n\neq i} \frac{L_n}{L_{,-i}} w_n$ is the weigthed average of their workplace wage. * $p < 0.05$; ** $p < 0.01$.

Table 2: Explaining the general equilibrium local employment elasticities to a 5 percent productivity shock

(5), we obtain the following generalization of our earlier expression for the equilibrium price of land (Q_n):

$$Q_n = \left((1-\alpha)\frac{\bar{v}_n R_n}{\bar{H}_n}\right)^{\frac{1}{1+\eta_n}}, \tag{30}$$

where the specification in the main paper corresponds to the special case in which $\eta = 0$.

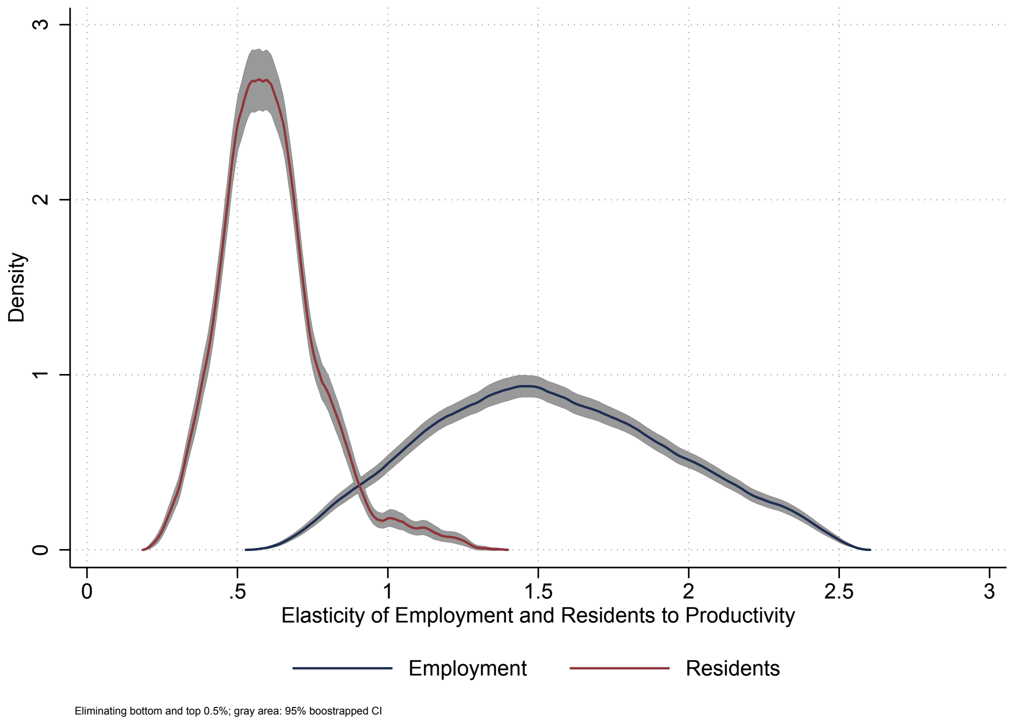
Therefore the sole implication of introducing a positive and heterogeneous developed land supply elasticity is to change one of the conditions for general equilibrium in the model. When undertaking counterfactuals, we replace the land market clearing condition from the main paper (5) with the modified land market clearing condition (30) in the system of equations for the counterfactual equilibrium.

We use the empirical estimates of developed land supply elasticities from Saiz (2010), which are based on physical and regulatory constraints to the geographical expansion of developed land area. Physical constraints are measured using Geographical Information Systems (GIS) data on the location of bodies of water (oceans and lakes) and wetlands and the elevation of terrain (the fraction of surrounding land that has a slope above 15 percent). Regulatory constraints are measured using the Wharton Residential Urban Land Regulation Index, which captures the stringency of residential growth controls. Using these data, Saiz (2010) estimates developed land supply elasticities for 95 Metropolitan Statistical Areas (MSAs) in the United States. The population-weighted average of these land supply elasticities is 1.75, and they range from 0.76 for the 10th ranked MSA (San Jose, CA) to 3.09 for the 85th ranked MSA (Charlotte-Gastonia-Rock Hill, NC-SC).

To incorporate these estimates into our quantitative analysis, we need to map these MSA-level estimates to our county-level data. Therefore we need to decide how to treat counties within multi-county MSAs and what to assume about counties that are not part of MSAs. We start with the most minimal departure from our baseline specification. We use the estimated Saiz elasticities where we have them (assuming the same elasticity for all counties within an MSA) and retain our baseline specification of a zero elasticity for all other counties. One limitation of this specification is that the Saiz estimates are based on the expansion of the geographical boundaries of developed land for the MSA as a whole. However, in MSAs that consist of multiple counties, central counties that are surrounded by other already-developed counties cannot expand this geographical frontier. Another limitation is that counties outside MSAs typically can expand this geographical frontier. To address these limitations, in subsection C.6 of the web appendix we consider a second specification in which we assume a land supply elasticity of zero for central counties within multi-county MSAs, the Saiz estimate for other counties in these MSA's as well as for single-county MSAs, and for all other counties the median Saiz estimate across MSAs of 1.67. In further robustness checks, we considered a range of alternative assumptions. Throughout we found similar pattern of results across these perturbations, particular with respect to the local employment elasticity.[29]

For each of these two specifications, we replicate our 3,111 counterfactual exercises where we shock each county with a 5 percent productivity shock (holding productivity in all other counties and holding all other exogenous variables constant). Figure 5 displays the results, and is analogous to Figure 4 in the main text above. The figure shows the estimated kernel density for the distribution of the general equilibrium elasticity of both employment (blue line) and residents (red line) with respect to the productivity shock across the treated counties. We also show the 95 percent confidence intervals around these estimated kernel densities (gray shading). Consistent with the results for our baseline specification, we continue to find substantial heterogeneity in local employment elasticities, confirming the robustness of our results to allowing for a variable developed land supply elasticity. Relative to our baseline specification, both the employment and residents elasticities are somewhat larger in magnitude, as the elastic land supply dampens the congestion effect from increased residents, and allows both employment and residents to increase more than with a perfectly inelastic land supply. Compared wth our baseline specification, we find

[29]For example, we modified the case presented in the text by assuming a zero housing supply elasticity for central counties in multi-county MSAs. We also perturbed the case in the web appendix by assuming that central counties in multi-county MSAs have the minimum housing supply elasticity across MSAs (0.6). In both these cases, and in our other experiments, we found a similar pattern of results.



Note: For counties in the 95 MSAs for which Saiz (2010) estimates a housing supply elasticity, we use the elasticity estimated by Saiz. For counties outside the 95 MSAs, we use an elasticity of zero as in our benchmark case.

Figure 5: Kernel density for the distribution of employment and residents elasticities in response to a productivity shock across counties (Positive housing supply elasticity)

that the heterogeneity in residents elasticity becomes larger relative to the heterogeneity in employment elasticities. This pattern of results is also intuitive. Our baseline specification focuses on one source of heterogeneity (commuting), which mainly affects local employment elasticities. In this robustness test, we introduce a second source of heterogeneity (through variable developed land supply elasticities), which mainly affects local residents elasticities (via residential land use). Therefore the main consequence of adding this second source of heterogeneity is to generate greater dispersion in residential elasticities. Nevertheless, there remains substantial dispersion in employment elasticities. Furthermore, we continue to find substantial differences between the distributions of elasticities for residents and employment, where the only way that these two distributions can differ from one another is through commuting.

4.4 Measuring the Incidence of Local Labor Demand Shocks

A large empirical literature is concerned with estimating the elasticity of local employment with respect to local demand shocks. The central specification in this empirical literature is a “differences-in-differences” specification across locations i and time t given by

$$\Delta \ln Y_{it} = a_0 + a_1 \mathbb{I}_{it} + a_2 X_{it} + u_{it}, \tag{31}$$

where Y_{it} is the outcome of interest (e.g. employment by workplace); $\mathbb{I}_{it}$ is a measure of the local demand shock (treatment); X_{it} are controls; and u_{it} is a stochastic error. The coefficient on the treatment a_1 corresponds to the log change in the outcome of interest with respect to the local demand shock. This specification has a differences-in-differences interpretation, because the first difference is over time (before and after the shock), and the second difference is between treated and control counties.

We now compare the general equilibrium elasticities of employment with respect to the productivity shock in the model to the results of this type of reduced-form "differences-in-differences" estimates of the local average treatment effects of the productivity shock. In particular, we construct a regression sample including both treated and untreated counties from our 3,111 counterfactuals in which we shock each county in turn with a 5 percent productivity shock ($3,111^2 = 9,678,321$ observations). We use these data to estimate a "differences-in-differences" specification of a similar form to (31):

$$\Delta \ln Y_{it} = a_0 + a_1 \mathbb{I}_{it} + a_2 X_{it} + a_3 \left(\mathbb{I}_{it} \times X_{it} \right) + u_{it},$$

where i denotes the 3,111 counties and t indexes the 3,111 counterfactuals; $\Delta \ln Y_{it}$ is the change in log employment between the counterfactual and actual equilibria; $\mathbb{I}_{it}$ is a (0,1) indicator for whether a county is treated with a productivity shock; and X_{it} are controls. We again consider two sets of controls (X_{it}): the model-suggested measures of linkages in goods and factor markets and more standard econometric controls (log employment, log wages and land area). We include both the main effects of these controls (captured by a_2) and their interactions with the treatment indicator to capture heterogeneity in the treatment effects (captured by a_3).

The key difference between this regression specification and the above results for the general equilibrium elasticities for the treated counties is that this regression specification *differences* relative to the untreated counties. The empirical literature has argued for the need to difference relative to untreated counties given the likely presence of other shocks or events, beyond the treatment, that can affect the treated counties simultaneously and can confound the true treatment effect. Of course, our synthetic dataset was generated without including any of these alternative shocks and so differencing is not needed in order to calculate the correct treatment effects.

In a specification without the controls ($a_2 = a_3 = 0$), the average effect of the productivity shock on the untreated counties is captured in the regression constant (a_0), and the local average treatment effect (a_1) corresponds to the difference in means between the treated and untreated counties. We compare estimating this regression specification including (i) a random untreated county in the control group, (ii) only the nearest untreated county in the control group, (iii) only neighboring counties within 120 kilometers of the treated county in the control group, (iv) only non-neighboring counties located from 120-240 kilometers from the treated county, and (v) all untreated counties in the control group.

We compare the predicted treatment effect from the "differences-in-differences" specification to the general equilibrium employment elasticity in the model by computing the following deviation term for the treated county:

$$\beta_i = \frac{a_1 + a_3 X_{it}}{0.05} - \frac{dL_i}{dA_i}\frac{A_i}{L_i}, \tag{32}$$

which corresponds to the difference between the predicted treatment effect, scaled by the size of the productivity shock, and the general equilibrium employment elasticity in the model. In Figure 6 we show kernel densities of the distribution of this deviation term across the 3,111 counterfactuals for productivity shocks to each county. We show the deviation term term using model-suggested controls (solid lines) and reduced-form controls (dashed lines). We display these results for each of the alternative control groups considered above ((i) to (v)).

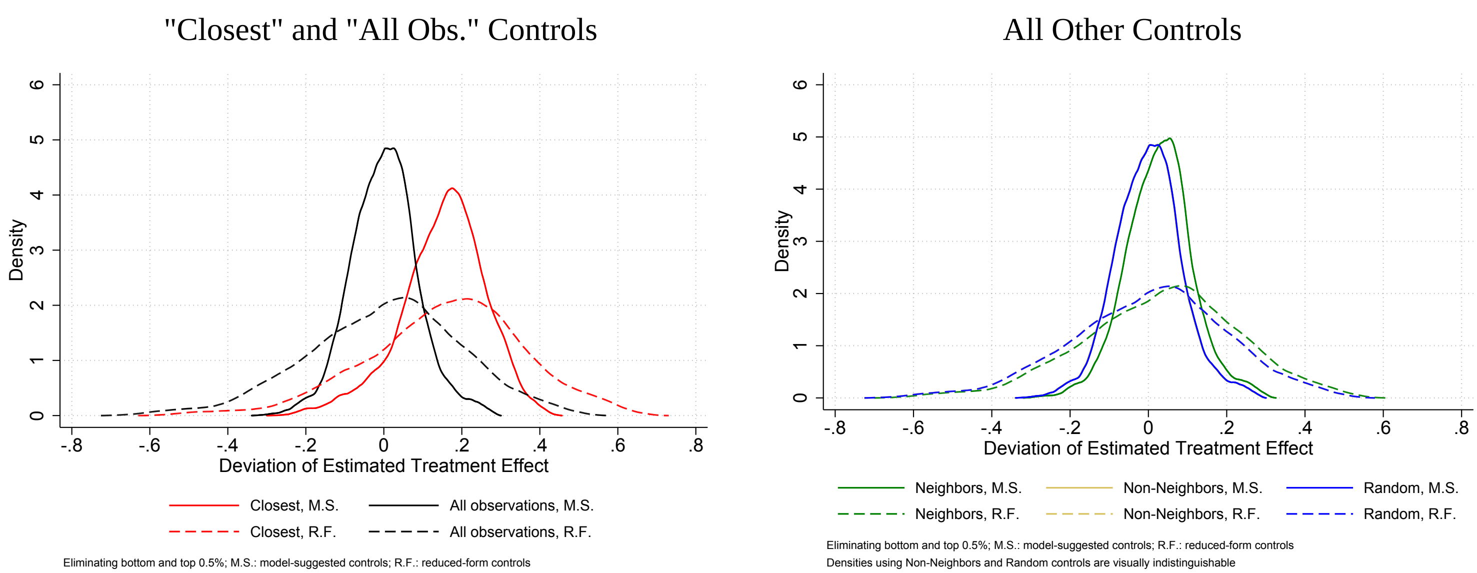


Figure 6: Distribution of the deviation term β_i across counties i, for different estimations

As shown in the figure, none of the “differences-in-differences” specifications completely captures the general equilibrium employment elasticity, as reflected in the substantial mass away from zero in these distributions. However, taking into account commuting linkages with the model-suggested controls substantially increases the predictive power of the “differences-in-differences” specification, as shown by substantial reduction in the mass away from zero using model-suggested rather than reduced-form controls. In general, we find similar results across the different control groups, with the results using random counties ((i) above) and non-neighbors ((iv) above) visually indistinguishable in the right-hand panel. However, we find a substantially larger deviation term using the nearest county as a control, because employment in the nearest untreated county is typically negatively affected by the increase in productivity in the treated county. While the use of contiguous locations as controls is often motivated based on similar unobservables (as in regression discontinuity designs), this pattern of results highlights that contiguous locations are also likely to be the most severely affected by spatial equilibrium linkages in goods and factor markets.

In subsection C.8 of the web appendix, we provide further evidence that the model-suggested controls are more successful in explaining the heterogeneity in treatment effects than the standard controls.

In subsection C.5, we show that we find a similar pattern of results whether we use spatially correlated shocks reproducing the industrial composition of the U.S. economy, and in subsection C.7, we show that we obtain the same pattern of findings if we replicate our entire analysis using CZs rather than counties. Therefore, while capturing the full general equilibrium effects of the productivity shocks requires solving the model-based counterfactuals, we find that augmenting "difference-in-difference" regressions with measures of commuting linkages substantially improves their ability to predict the heterogeneity in the estimated treatment effects.

4.5 Million Dollar Plants Natural Experiment

Having established the heterogeneity of local employment elasticities, due to differences in commuting linkages across counties, in our quantitative model, we now use evidence from a natural experiment to provide independent verification of this prediction in the data. In particular, we use the natural experiment of million dollar plants (MDP) from Greenstone, Hornbeck and Moretti (2010), one of the most influential papers in the local labor markets literature. The identification strategy compares employment in "winning" and "runner-up" counties before and after the opening of a MDP. As the runner-up counties are those that survived a long selection process, but narrowly lost the competition to winning counties, one would expect the two groups to have similar initial characteristics. Consistent with this, winning and runner-up counties are similar along a range of observables before a MDP opening. In contrast, the groups exhibit sharply different trajectories following a MDP opening. Winning counties experience larger plant-level increases in total factor productivity (TFP), and have larger county-level increases in the number of manufacturing plants and total manufacturing employment and output.

We start with a list of 82 MDP announcements ("cases") containing winner and runner-up counties from the corporate real estate journal *Site Selection*, as reported in Greenstone and Moretti (2004). Only a subset of 47 of these MDP announcements could be located in (confidential) census data by Greenstone, Hornbeck and Moretti (2010), in part because not all of the announced plants were necessarily ultimately opened. To be conservative, we use the full list of 82 announcements, where the fact that some of these plants may not have actually opened (or may have opened some time after the announcement date) will make it harder for us to find discernible effects. We begin by estimating the average treatment effect of a MDP announcement. Consistent with the reduced-form specification from the previous subsection, we estimate a "differences-in-differences" regression, where the first difference is before and after the announcement, and the second difference is between winning and runner-up counties. For each year $\tau > 0$ after the announcement of a MDP, we difference log county employment relative to its value in the year of the announcement ($\tau = 0$) and estimate a specification given by

$$\Delta \ln Y_{ik\tau} = \alpha_\tau \mathbb{I}_{ik\tau} + \mu_{k\tau} + u_{ik\tau} \tag{33}$$

where i again indexes counties and k denotes cases (instances of winner and runner-up counties); we estimate this regression separately for each year $\tau > 0$; $\Delta \ln Y_{ik\tau}$ is the change in log employment between the announcement and year τ; the first differencing over time has eliminated any county fixed effect in the level of log employment; $\mathbb{I}_{ik\tau}$ is a treatment indicator that equals one if a county i for case k has an announced MDP in year τ and zero otherwise; $\mu_{k\tau}$ is a case fixed effect that captures average employment growth up to year τ for both the winner and runner-up counties in case k; α_τ is the key coefficient of interest and corresponds to the average treatment effect of a MDP announcement.

We next examine the extent to which the treatment effect of a MDP announcement is heterogeneous across counties depending on their commuting linkages, as predicted by our quantitative model. Motivated by the results of our counterfactuals in the previous section, we measure these commuting linkages using each county's own commuting share. We augment the "differences-in-differences" regression in (33) with both the main effect of the own commuting share ($\lambda_{ii|i}$) and the interaction term between the MDP announcement and the own commuting share ($\lambda_{ii|i} \times \mathbb{I}_{ik\tau}$). So we estimate

$$\Delta \ln Y_{ik\tau} = \alpha_\tau \mathbb{I}_{ik\tau} + \beta_\tau \lambda_{ii|i} + \gamma_\tau \left(\lambda_{ii|i} \times \mathbb{I}_{ik\tau} \right) + \mu_{k\tau} + u_{ik\tau}, \tag{34}$$

where the coefficient β_τ on the main effect allows for the possibility that counties with open versus closed commuting markets could differ in employment growth for other reasons besides the MDP announcement; the key coefficient of interest is γ_t on the interaction term, which captures heterogeneity in the response of local employment to the MDP announcement. Based on our quantitative model, we expect γ_t to be negative and statistically statistically significant, such that counties with more closed commuting markets (higher own commuting shares $\lambda_{ii|i}$) have smaller local employment responses.

In Table 3, we report the estimation results. To provide a point for comparison, Panel A estimates equation (33) including only the case fixed effects (without the treatment indicator). Across periods ranging up to five years after the MDP announcement, we find that the case fixed effects alone explain around 18 to 29 percent of the variance in county employment growth. In Panel B, we estimate the average treatment effect of the MDP announcement using our baseline specification (33) (including the treatment indicator). We find strong confirmation of the empirical results in Greenstone, Hornbeck and Moretti (2010), with a positive and statistically significant average treatment effect. The magnitude of this treatment effect is substantial, ranging from 2.6 percent to more than 6 percent as we move from one to five years after the announcement. We also find that the MDP announcement has substantial explanatory power for county employment growth, with the regression R-squared rising to a range of 27 to 29 percent.

In Panel C, we estimate heterogeneous treatment effects using our augmented specification (34) (including the terms in the own commuting share). Confirming our model's predictions, the estimated coefficient on the interaction term (γ_τ) is negative and statistically significant, implying that the MDP announcement has smaller employment effects in counties with more closed commuting markets. The implied magnitude of this heterogeneity is large compared to the average treatment effect of 2 to 6 percent from Panel B. For a county with an own commuting share at the 75th percentile the predicted employment effect is between 23.6 and 26.6 percent (depending on the number of years after the treatment) smaller than for a county with a commuting share at the 25th percentile. In fact, in our benchmark model the same

Panel A : Case Fixed Effects Only

τ years after	Adj R^2
1	0.21
2	0.29
3	0.25
4	0.22
5	0.18

Panel B : Average Treatment Effect and Case Fixed Effects

τ years after	Adj R^2	Coefficient	Std. Error
1	0.31	0.026^{***}	0.006
2	0.39	0.038^{***}	0.008
3	0.36	0.050^{***}	0.010
4	0.31	0.053^{***}	0.013
5	0.27	0.062^{***}	0.015

Panel C : Heterogeneous Treatment Effects and Case Fixed Effects

		Treatment		Openness		Interaction	
τ years after	Adj R^2	Coefficient	Std. Error	Coefficient	Std Error	Coefficient	Std Error
1	0.40	0.117^{**}	0.028	-0.027	0.026	-0.124^{***}	0.036
2	0.50	0.133^{**}	0.037	-0.084^{***}	0.035	-0.133^{***}	0.048
3	0.46	0.149^{**}	0.048	-0.130^{***}	0.045	-0.140^{***}	0.062
4	0.41	0.159^{*}	0.059	-0.167^{***}	0.056	-0.151^{**}	0.077
5	0.39	0.213^{*}	0.071	-0.191^{***}	0.067	-0.213^{**}	0.092

This Table reports estimates of two versions of equation (33) and of equation (34). Case fixed effects are never reported. Panel A estimates equation (33) including only the case fixed effects $\mu_{k\tau}$ (i.e. without the treatment indicator $\mathbb{I}_{ik\tau}$). Panel B estimates the average treatment effect of the MDP announcement using specification (33). Panel C estimates heterogeneous treatment effects using (34). In Panel C, "Treatment" refers to the coefficient α_τ, "Openness" refers to the coefficient β_τ, and "Interaction" refers to the coefficient γ_τ as in equation (34). * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 3: Million Dollar Plants Natural Experient

exercise yields a slightly smaller reduction in the predicted employment effect (18.2 percent). If we compare the 90th and 10th percentile of own commuting shares the difference in employment effects ranges between 49% and 43% depending on the number of years after treatment. In our baseline model the same difference is 31.7%. We conclude that the impact of commuting linkages on the MDP announcement is statistically significant and quite large in economic terms. The reductions in the employment effect of the treatment that result from large own commuting shares are of the same magnitude, but somewhat larger, than the ones predicted by the model. So, if anything, our model seems to underplay the importance of commuting compared to the data.

We also find a negative and statistically significant main effect (β_τ) of the own commuting share, which suggests that counties with more closed commuting markets on average grew more slowly, consistent with their being on average more remote from other counties. Comparing Panels C and B, allowing for heterogeneous treatment effects leads to a substantial further increase in the regression R-squared to 40-50 percent. Therefore the increase in explanatory power from allowing for treatment heterogeneity depending on commuting links is around as large as that from allowing for average treatment effects (Panels A-B).

Finally, the results in Table 3 also indicate that the average treatment effects estimated by Greenstone, Hornbeck and Moretti (2010) depends on the average amount of commuting in the economy at the time of the treatment. An economy with larger average own commuting shares will result in smaller average treatment effects than an economy will small own commuting shares. This elasticity is governed by γ_{τ}, which is highly significant and declines from -0.124 to -0.213 over the first five years after treatment.

Taking the results of this section together, using the natural experiment of million dollar plants (MDP) as an independent source of variation, we find strong confirmation of our model's prediction of heterogeneous local employment elasticities depending on commuting networks. Comparing counties that are otherwise similar, except for whether they were ultimately fortunate enough to attract a MDP, we find substantially larger increases in employment in winner counties than in runner-up counties, and this increase in employment is considerably greater for those winner counties with more open commuting markets.

5 Changes in Commuting Costs

In the previous sections, we have shown the importance of commuting for the local effects of local economic shocks. We now show that it also matters in the aggregate for the spatial distribution of economic activity across locations and welfare. Commuting enables workers to access high productivity locations without having to pay the high cost of living in those locations. Increasing the cost of commuting restricts the opportunity set available to firms and workers and hence is expected to reduce welfare. Locations that were previously net exporters of commuters in the initial equilibrium become less attractive residences, while locations that were previously net importers of commuters in the initial equilibrium become less attractive workplaces. As agents relocate in response to the restricted opportunity set, we also expect decline in the specialization of counties as residential or business locations.

We begin by using the observed commuting data to back out implied values of the composite parameter capturing the ease of commuiting ($\tilde{B}_{ni} = B_{ni}\kappa_{ni}^{-\epsilon}$). Following Head and Ries (2001) in the international trade literature, we use the flows of commuters between locations n and i in both directions relative to their own commuting flows. Using the commuting gravity equation (13), and taking the geometric mean of these relative flows in both directions, we obtain the following measure of the average ease of commuting between locations n and i relative to the ease of commuting to themselves:

$$\tilde{\mathcal{B}}_{ni} \equiv \left(\frac{\tilde{B}_{ni}}{\tilde{B}_{nn}} \frac{\tilde{B}_{in}}{\tilde{B}_{ii}} \right)^{1/2} = \left(\frac{L_{ni}}{L_{nn}} \frac{L_{in}}{L_{ii}} \right)^{1/2}. \tag{35}$$

We compute this measure for both 1990 and 2007. Over this period, there was a substantial increase in both miles of paved roads and vehicle kilometers travelled.[30] Consistent with this increased intensity of transport use, we find a substantial increase in the relative ease of commuting from 0.96 at the 25th percentile to 0.88 at the median and 0.79 at the 75 percentile.

[30] Between 1990 and 2010, kilometers of paved public roads in the United States increased by over 20 percent (from 3.6 to 4.4 million), and vehicle kilometers travelled increased by more than 38 percent (from 3,451,016 to 4,775,352 million). For further discussion of this expansion in transport use, see for example Duranton and Turner (2011).

We use this distribution of implied changes in the relative ease of commuting to undertake counterfactuals for empirically-realistic changes in commuting costs. We assume a common reduction or increase in the relative ease of commuting for all counties equal to percentiles of this distribution (e.g. we assume that all counties experience the median increase in the ease of commuting). Given this assumption, we use the system of equations for general equilibrium in the model to solve for the new counterfactual equilibrium after the reduction in commuting costs, as discussed in Section 2.6 above. From the commuting probability (13) and expected utility (18), the change in the common level of welfare across all locations from the shock to commuting costs can be decomposed as follows:

$$\widehat{\overline{U}} = \left(\frac{1}{\widehat{\lambda}_{ii}}\right)^{\frac{1}{\epsilon}} \left(\frac{1}{\widehat{\pi}_{ii}}\right)^{\frac{\alpha}{\sigma-1}} \left(\frac{\widehat{w}_i}{\widehat{\bar{v}}_i}\right)^{1-\alpha} \frac{\widehat{L}_i^{\frac{\alpha}{\sigma-1}}}{\hat{R}_i^{1-\alpha}}, \tag{36}$$

where we have used the fact that $\{\kappa_{ii}, B_{ii}, A_i, d_{ii}\}$ are unchanged; the first term in $\widehat{\pi}_{ii}$ captures the impact through changes in openness to commuting; the second term in $\widehat{\pi}_{ii}$ captures the effect through adjustments in openness to goods trade; the remaining terms capture the influence of changes in the spatial distribution of wages ($\widehat{w}_i$), expected residential income ($\bar{v}_i$), employment ($\widehat{L}_i$) and residents ($\hat{R}_i$).

	Decrease by p75	Decrease by p50	Decrease by p25	Increase by p50
Implied $\tilde{\mathcal{B}}_{ni}$	0.79	0.88	0.96	1.13
Welfare Change	6.89%	3.26%	0.89%	-2.33%

This table shows the percentage change in welfare across different reductions in commuting costs exercises. Each column reports the percentile of the reduction (or increase) in commuting costs. The first row reports the implied $\tilde{\mathcal{B}}_{ni}$. The second row reports the percentage change in welfare for the corresponding exercise.

Table 4: Welfare Impacts for different Changes in Commuting Costs

As shown in Table 4, we find substantial effects of the reductions in commuting costs on aggregate welfare. Reducing commuting costs by the median proportional change observed over our time period from 1990 to 2007 is predicted to increase welfare by around 3.3 percent. In contrast, raising commuting costs by the same proportional amount decreases welfare by around 2.3 percent. As we scale up the reduction in commuting costs to the 75th percentile observed over our time period, we amplify the welfare gain to 6.9 percent. As we scale down the reduction in commuting costs to the 25th percentile, we diminish the welfare gain to 0.89 percent. The proportional changes in welfare from empirically-realistic changes in commuting costs are larger relative to standard empirical estimates of the welfare gains from opening the closed economy to international trade, which range from less than 1 percent for the United States to just over 10 percent for Belgium in the United States. While commuting flows typically occur at much smaller spatial scales than international trade flows, these results clearly highlight that commuting not only shapes the impact of local shocks but is also consequential for aggregate welfare.[31]

[31]Smaller values for the Fréchet shape parameter (ϵ) imply more heterogeneity in preferences for pairs of residence and

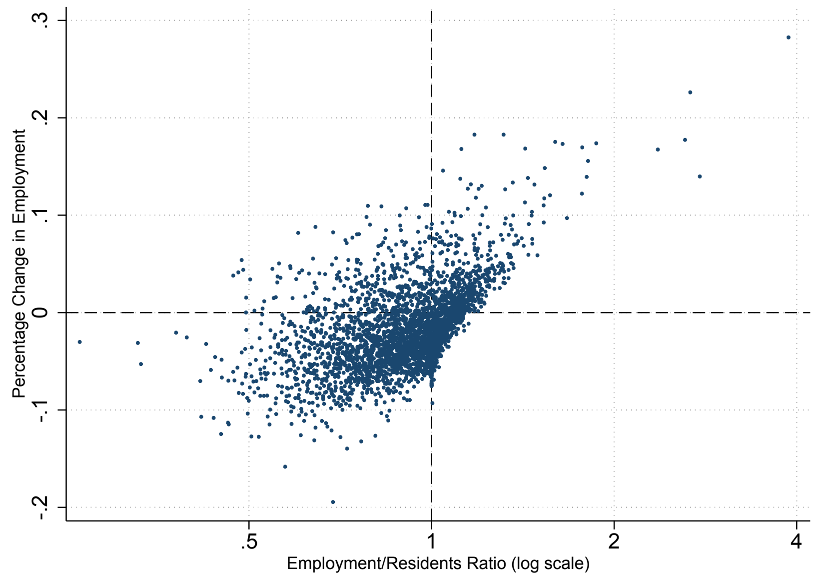


Figure 7: Counterfactual relative change in county employment ($\hat{L}$) for median decrease in commuting costs throughout U.S. against initial employment to residents ratio (L/R).

These empirically-realistic changes in commuting costs also result in substantial changes in the spatial distribution of employment and residents across locations. In Figure 7, we show the counterfactual change in employment in each county from reducing commuting costs by the median proportional change observed over our time period from 1990 to 2007. We display these counterfactual changes in employment against each county's initial commuting intensity L_i/R_i, where $L_i/R_i > 1$ implies that a county is a net importer of commuters and $L_i/R_i < 1$ implies that a county is a net exporter of commuters. We find substantial changes in employment for individual counties, which range from increases of 28 percent to reductions of 19 percent, and are well explained by initial commuting intensity. As discussed in subsection 3.2 above and shown in subsection C.2 of the web appendix, initial commuting intensity is itself hard to explain in terms of standard empirical controls such as land area, size and housing supply elasticities. In subsection C.9 of the web appendix, we undertake counterfactuals for reductions in commuting costs for CZs. We show that the counterfactual changes in CZ employment from reductions in commuting costs are well explained by measures of the extent to which the CZ uses the commuting technology in the initial equilibrium. But these counterfactual changes in CZ employment are not well explained by initial CZ employment or residents size. These results provide further confirmation that the importance of commuting is by no means restricted to large cities.

Given this importance of commuting links in shaping the distribution of economic activity across locations, it is natural to expect that these links also determine the magnitude of the impact of reductions in trade costs. In section C.10 of the web appendix, we explore this interaction between trade and commuting costs. We compare the counterfactual effects of a 20 percent reduction of trade costs in the actual world

workplace locations, which implies greater welfare losses from increases in commuting costs and smaller welfare gains from reductions in commuting costs. For example, in a world with a 50% lower value of ϵ, increasing commuting costs by the median proportional change reduces welfare by 4.77 percent, while reducing commuting costs by the same proportional amount increases welfare by 6.87 percent.

with commuting to the effects in a hypothetical world without commuting. In general, reductions in trade costs lead to a more dispersed spatial distribution of economic activity in the model. But this dispersal is smaller with commuting than without commuting. As trade costs fall, commuting increases the ability of the most productive locations to serve the national market by drawing workers from a suburban hinterland, without bidding up land prices as much as would otherwise occur. These results further underscore the prominence of commuting linkages in shaping the equilibrium spatial distribution of economic activity, and the necessity of incorporating them in models of economic geography.

6 Conclusions

Local technology, regulation, or infrastructure shocks can have far reaching economic effects through spatial linkages between locations. We have developed a spatial general equilibrium model that quantifies these spatial linkages in both goods markets (trade) and factor markets (commuting and migration). Our quantitative model uses the observed gravity equation relationships for goods and commuting flows to estimate the two key parameters of the model and matches exactly the observed cross-section distributions of employment, residents and incomes across U.S. counties.

We show that commuting flows are large and heterogeneous across counties and that commuting zones are imperfect in capturing these flows. The resulting differences in patterns of commuting lead to substantial variation across locations in elasticities of employment to productivity shocks, which have a mean of 1.52, but range from close to 0.5 to almost 2.5. These results question the generalizability of estimates of this elasticity that do not account for its large spatial heterogeneity. Furthermore, as our theoretical model incorporates multiple spatial linkages between locations (trade in goods and migration as well as commuting), it becomes an empirical question which of these linkages is more important given the observed patterns in the data in the initial equilibrium. We show that simple measures of commuting, or partial equilibrium measures of the spatial linkages derived from the model, are empirically successful in accounting for this variation in general equilibrium local employment elasticities.

We hope that our results are used to motivate the inclusion of these measures of commuting in future empirical estimations of local employment elasticities. Their inclusion is simple, as these measures depend only on observables in the initial equilibrium. Furthermore, these terms are not well accounted for by the inclusion of other variables like measures of employment or wages in the treated or neighboring locations. So including observed measures of these linkages is essential. Our results also question the use of empirical difference-in-differences strategies that use contiguous locations as the control group to measure the treatment effect of a policy or shock. In our counterfactual exercises, the closest untreated locations tend to be significantly affected by the shocks to the treated locations.

We find that commuting matters not only for the incidence of local shocks but also for aggregate welfare and the spatial distribution of economic activity. Increasing commuting costs by adding the median commuting cost to all counties reduces aggregate welfare by 2.33 percent, which is larger in magnitude than most estimates of the welfare gains from international trade. Commuting enables workers to live close to high productivity locations without having to pay the high land prices in those locations. Therefore increasing commuting costs to prohibitive levels redistributes economic activity away from areas that use commuting intensively (e.g. the New York region) towards areas that use commuting less intensively.

The theory we propose in this paper is, we believe, quite ambitious and rich. We view this theory as a reasonable framework to quantify the heterogeneity in local employment elasticities, because it is able to account for key observed relationships in the data (gravity in goods trade and commuting flows) and rationalizes the observed cross-section distribution of employment, wages, commuting and international trade flows in the initial equilibrium. In principle, the response to the economy to subsequent shocks could be different from its behavior in the initial equilibrium to which we calibrate. But since this initial equilibrium is itself the result of the accumulation of past shocks, we believe that it provides a reasonable benchmark against which to calibrate the model. Furthermore, our approach enables us to use bilateral patterns of goods trade and commuting flows in the initial equilibrium to capture the rich unobserved patterns of trade and commuting costs between locations.

Finally, we have extended quantitative general equilibrium models of economic geography to incorporate an additional margin that is empirically relevant (commuting between locations). But we acknowledge that there remain other margins that can also matter for local employment elasticities in response to shocks (e.g. labor force participation and unemployment). We view the incorporation of these additional margins into spatial general equilibrium models as an important part of the future research agenda. However, given the quantitative magnitudes of commuting flows in the data, we believe that commuting will continue to play an important role in shaping local employment elasticities even in such future models incorporating a wider range of adjustment margins.

References

[1] Ahlfeldt, Gabriel, Stephen Redding, Daniel Sturm and Nikolaus Wolf (2015) "The Economics of Density: Evidence from the Berlin Wall," *Econometrica*, forthcoming.

[2] Allen, T., and C. Arkolakis (2014a) "Trade and the Topography of the Spatial Economy," *Quarterly Journal of Economics*, 129(3), 1085-1140.

[3] Allen, T., and C. Arkolakis (2014b) "Universal Gravity," *NBER Working Paper*, 20787.

[4] Allen, T., C. Arkolakis and X. Li (2015) "Optimal City Structure," Yale University, mimeograph.

[5] Alonso, William (1964) *Location and Land Use*, Cambridge MA: Harvard.

[6] Arkolakis, Costas, Arnaud Costinot and Andres Rodriguez-Clare (2012) "New Trade Models, Same Old Gains," *American Economic Review*, 102(1), 94-130.

[7] Armington, Paul S. (1969) "A Theory of Demand for Products Distinguished by Place of Production," *IMF Staff Papers*, 16(1), 159-178.

[8] Artuc, E., S. Chaudhuri, and J. McLaren (2010) "Trade Shocks and Labor Adjustment: A Structural Empirical Approach," *American Economic Review*, 100(3), 1008-1045.

[9] Autor David and David Dorn (2013) “The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market,” *American Economic Review*, 103(5), 1553-1597.

[10] Autor David, David Dorn and Gordon Hanson (2013) “The China Syndrome: Local Labor Market Effects of Import Competition in the United States,” *American Economic Review*, 103(6), 2121-2168.

[11] Bartik, Timothy J. (1991) “Who Benefits from State and Local Economic Development Policies?” W.E. Upjohn Institute for Employment Research, Upjohn Press.

[12] Behrens, K., G. Mion, Y. Murata and J. Südekum, 2014. "Spatial frictions," DICE Discussion Papers 160, Heinrich-Heine-Universität Düsseldorf, Düsseldorf Institute for Competition Economics (DICE).

[13] Blanchard, Olivier Jean, and Lawrence F. Katz (1992) “Regional Evolutions,” *Brookings Papers on Economic Activity*, 23(1), 1-76.

[14] Bound, John and Holzer, Harry J (2000) “Demand Shifts, Population Adjustments, and Labor Market Outcomes during the 1980s,” *Journal of Labor Economics*, 18(1), 20-54.

[15] Broda, Christian and David E. Weinstein (2006) “Globalization and the Gains from Variety,” *Quarterly Journal of Economics*, 121(2), 541-86.

[16] Busso, M., J. Gregory, and P. Kline (2013) “Assessing the Incidence and Efficiency of a Prominent Place Based Policy,” *American Economic Review*, 103(2), 897-947.

[17] Caliendo, L. and F. Parro (2015) “Estimates of the Trade and Welfare Effects of NAFTA,” *Review of Economic Studies*, 82(1), 1-44.

[18] Caliendo, L., F. Parro, E. Rossi-Hansberg, and P.-D. Sarte (2014) “The Impact of Regional and Sectoral Productivity Changes on the U.S. Economy,” *NBER Working Paper*, 20168.

[19] Dekle, R., J. Eaton, and S. Kortum (2007) “Unbalanced Trade,” *American Economic Review*, 97(2), 351-355.

[20] Desmet, Klaus and Esteban Rossi-Hansberg (2013) “Urban Accounting and Welfare,” *American Economic Review*, 103:6, 2296-2327.

[21] Desmet, Klaus and Esteban Rossi-Hansberg (2014) “Spatial Development,” *American Economic Review*, 104(4), 1211-1243.

[22] Diamond, Rebecca (2016) “The Determinants and Welfare Implications of US Workers? Diverging Location Choices by Skill: 1980-2000,” *American Economic Review*, 106(3), 479-524.

[23] Dingel, Jonathan (2015) “The Determinants of Quality Specialization,” Chicago Booth School of Business, mimeograph.

[24] Duranton, Gilles and Matthew A. Turner (2011) “The Fundamental Law of Road Congestion: Evidence from U.S. Cities,” *American Economic Review*, 101(6), 2616-2652.

[25] Duranton, Gilles, Peter Morrow and Matthew A. Turner (2014) “Roads and Trade: Evidence from the U.S.,” *Review of Economic Studies*, 81(2), 681-724.

[26] Eaton, Jonathan and Kortum, Samuel (2002) "Technology, Geography, and Trade," *Econometrica*, 70(5), 1741-1779.

[27] Feyrer, James, Erin T. Mansur and Bruce Sacerdote (2015) "Geographic dispersion of economic shocks: Evidence from the fracking revolution" *NBER Working Paper*, 21624.

[28] Fujimoto, Takao and Krause, Ulrich (1985) "Strong Ergodicity for Strictly Increasing Nonlinear Operators," *Linear Algebra and its Applications*, 71(1), 101-112.

[29] Fujita, Masahisa and Hideaki Ogawa (1982) "Multiple Equilibria and Structural Transition of Non-Monocentric Urban Configurations,"*Regional Science and Urban Economics*, 12, 161-196.

[30] Fujita, Masahisa, Paul Krugman and Anthony J. Venables (1999) *The Spatial Economy: Cities, Regions and International Trade*, Cambridge: MIT Press.

[31] Greenstone, M. and E. Moretti (2004) "Bidding for Industrial Plants: Does Winning a 'Million Dollar Plant' Increase Welfare? " *NBER Working Paper*, 9844.

[32] Greenstone, M., R. Hornbeck, and E. Moretti (2010) "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings," *Journal of Political Economy*, 118(3), 536-598.

[33] Grogger, J., and G. Hanson (2011) "Income Maximization and the Selection and Sorting of International Migrants," *Journal of Development Economics*, 95(1), 42-57.

[34] Hanson, G. H. (1996) "Localization Economies, Vertical Organization, and Trade," *American Economic Review*, 86(5), 1266-1278.

[35] Hanson, Gordon H. (2005) "Market Potential, Increasing Returns, and Geographic Concentration," *Journal of International Economics*, 67(1), 1-24.

[36] Head, Keith and John Ries (2001) "Increasing Returns versus National Product Differentiation," *American Economic Review*, 91, 858-876.

[37] Helpman, Elhanan (1998) "The Size of Regions," In *Topics in Public Economics: Theoretical and Applied Analysis*, ed. David Pines, Efraim Sadka, and Itzhak Zilcha, Cambridge: Cambridge University Press.

[38] Kennan, J., and J. R. Walker (2011) "The Effect of Expected Income on Individual Migration Decisions," *Econometrica*, 79(1), 211-251.

[39] Krugman, Paul (1991) "Increasing Returns and Economic Geography," *Journal of Political Economy*, 99(3), 483-99.

[40] Krugman, Paul and Anthony J. Venables (1995) "Globalization and the Inequality of Nations," *Quarterly Journal of Economics*, 110(4), 857-880.

[41] Lucas, Robert E. and Esteban Rossi-Hansberg (2002) "On the Internal Structure of Cities", *Econometrica*, 70(4), 1445-76.

[42] Manning, Alan and Barbara Petrongolo (2011) "How Local are Labour Markets? Evidence from a Spatial Search Model", *CEP Discussion Paper*, No. 1101, London School of Economics.

[43] McFadden, D. (1974) "The Measurement of Urban Travel Demand," *Journal of Public Economics*, 3(4), 303-328.

[44] Melitz, Marc J. (2003) "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71(6), 1695-1725.

[45] Mian, Atif and Amir Sufi (2014) "What Explains the 2007-2009 Drop in Employment?" *Econometrica*, 82(6), 2197-2223.

[46] Michaels, Guy (2011) "The Long Term Consequences of Resource-Based Specialisation," *Economic Journal*, 121(551), 31-57.

[47] Mills, Edwin S. (1967) "An Aggregative Model of Resource Allocation in a Metropolitan Centre," *American Economic Review*, 57(2), 197-210.

[48] Monte, Ferdinando (2015) "The Local Incidence of Trade Shocks," Georgetown University, mimeograph.

[49] Moretti, Enrico (2011) "Local Labor Markets," in *Handbook of Labor Economics*, (ed.) D. Card, and O. Ashenfelter, Volume 4b, Amsterdam: Elsevier North Holland, 1238-1303.

[50] Moretti, Enrico and Patrick Klein (2011) "Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority," *Quarterly Journal of Economics*, 129(1), 275-331.

[51] Muth, Richard (1969) *Cities and Housing*, Chicago: University of Chicago Press.

[52] Notowidigdo, Matthew (2013) "The Incidence of Local Labor Demand Shocks," Northwestern University, mimeograph.

[53] Redding, Stephen J. and Daniel Sturm (2008) "The Costs of Remoteness: Evidence from German Division and Reunification," *American Economic Review*, 98(5), 1766-1797.

[54] Redding, Stephen J. and Matthew Turner (2015) "Transportation Costs and the Spatial Organization of Economic Activity," in (eds) Gilles Duranton, J. Vernon Henderson and William Strange, *Handbook of Urban and Regional Economics*, Chapter 20, 1339-1398.

[55] Redding, Stephen J. (2016) "Goods Trade, Factor Mobility and Welfare," *Journal of International Economics*, 101, 148-167.

[56] Rossi-Hansberg, Esteban (2005) "A Spatial Theory of Trade," *American Economic Review*, 95(5), 1464-1491.

[57] Saiz, Albert (2010) "The Geographic Determinants of Housing Supply," *Quarterly Journal of Economics*, 125(3), 1253-1296.

[58] Simonovska, Ina and Michael E. Waugh (2014) "The Elasticity of Trade: Estimates and Evidence," *Journal of International Economics*, 92(1), 34-50.

[59] Tolbert, Charles M. and Molly Sizer (1996) "U.S. Commuting Zones and Labor Market Areas: a 1990 Update," *U.S. Department of Agriculture Staff Paper*, AGES-9614.

[60] Yagan, Danny (2014) "Moving to Opportunity? Migratory Insurance over the Great Recession," University of California, Berkeley.

Web Appendix for "Commuting, Migration and Local Employment Elasticities" (Not for Publication)

Ferdinando Monte
Georgetown University

Stephen J. Redding
Princeton University

Esteban Rossi-Hansberg
Princeton University

A Introduction

Section B of this web appendix contains the proofs of the propositions in the paper, additional technical derivations of results reported in the paper, and further supplementary material discussed in the paper. Section C of this web appendix includes additional empirical results and robustness tests. Section D of this web appendix presents further information about the data definitions and sources.

B Theoretical Appendix

The first two sections of this theoretical part of the web appendix present additional technical derivations for the main paper. Section B.1 reports the derivations of expected utility and the commuting probabilities. Section B.2 shows how the equilibrium conditions of the model can be used to undertake counterfactuals using the observed values of variables in the initial equilibrium.

The next four sections give the proofs of the propositions in the paper. Section B.3 contains the proof of Proposition 1. Section B.4 includes the proof of Proposition 2. Section B.5 presents the proof of Proposition 3. Section B.6 incorporates the proof of Proposition 4.

The following four sections comprise supplementary material and extensions. Section B.7 reports the derivation of the partial equilibrium local employment elasticities discussed in the main paper. Section B.8 shows that the class of models consistent with a gravity equation for commuting flows implies heterogeneous local employment elasticities. Section B.9 introduces multiple worker types, and Section B.10 introduces congestion in commuting. Section B.11 develops an extension of the baseline model to incorporate non-traded consumption goods, which can be consumed either at a worker's residence or workplace. Section B.12 introduces an extension where landlords also use residential land and Section B.13 one in which we generalize the production function so that firms use intermediate goods. Finally, Section B.14 introduces heterogeneity in effective units of labor.

B.1 Commuting Decisions

We begin by reporting additional results for the characterization of worker commuting decisions.

B.1.1 Distribution of Utility

From all possible pairs of residence and employment locations, each worker chooses the bilateral commute that offers the maximum utility. Since the maximum of a sequence of Fréchet distributed random variables is itself Fréchet distributed, the distribution of utility across all possible pairs of residence and employment locations is:

$$1 - G(u) = 1 - \prod_{r=1}^{S}\prod_{s=1}^{S} e^{-\Psi_{rs}u^{-\epsilon}},$$

where the left-hand side is the probability that a worker has a utility greater than u, and the right-hand side is one minus the probability that the worker has a utility less than u for all possible pairs of residence and employment locations. Therefore we have:

$$G(u) = e^{-\Phi u^{-\epsilon}}, \qquad \Psi = \sum_{r=1}^{S}\sum_{s=1}^{S} \Psi_{rs}. \tag{B.1}$$

Given this Fréchet distribution for utility, expected utility is:

$$\mathbb{E}\left[u\right] = \int_{0}^{\infty} \epsilon \Psi u^{-\epsilon} e^{-\Psi u^{-\epsilon}} du. \tag{B.2}$$

Now define the following change of variables:

$$y = \Phi u^{-\epsilon}, \qquad dy = -\epsilon \Psi u^{-(\epsilon+1)} du. \tag{B.3}$$

Using this change of variables, expected utility can be written as:

$$\mathbb{E}\left[u\right] = \int_{0}^{\infty} \Psi^{1/\epsilon} y^{-1/\epsilon} e^{-y} dy, \tag{B.4}$$

which can be in turn written as:

$$\mathbb{E}\left[u\right] = \delta \Psi^{1/\epsilon}, \qquad \delta = \Gamma\left(\frac{\epsilon - 1}{\epsilon}\right), \tag{B.5}$$

where $\Gamma(\cdot)$ is the Gamma function. Therefore we have the expression in the main text above:

$$\mathbb{E}\left[u\right] = \delta \Psi^{1/\epsilon} = \delta \left[\sum_{r=1}^{S}\sum_{s=1}^{S} B_{rs}\left(\kappa_{rs} P_r^{\alpha} Q_r^{1-\alpha}\right)^{-\epsilon} w_s^{\epsilon}\right]^{1/\epsilon}. \tag{B.6}$$

B.1.2 Residence and Workplace Choices

Using the distribution of utility for pairs of residence and employment locations (12), the probability that a worker chooses the bilateral commute from n to i out of all possible bilateral commutes is:

$$\begin{aligned}\pi_{ni} &= \Pr\left[u_{ni} \geq \max\{u_{rs}\}; \forall r, s\right], \\ &= \int_0^\infty \prod_{s \neq i} G_{ns}(u) \left[\prod_{r \neq n} \prod_{s} G_{rs}(u)\right] g_{ni}(u) du, \\ &= \int_0^\infty \prod_{r=1}^{S} \prod_{s=1}^{S} \epsilon \Psi_{ni} u^{-(\epsilon+1)} e^{-\Psi_{rs} u^{-\epsilon}} du. \\ &= \int_0^\infty \epsilon \Psi_{ni} u^{-(\epsilon+1)} e^{-\Psi u^{-\epsilon}} du.\end{aligned}$$

Note that:

$$\frac{d}{du}\left[-\frac{1}{\Psi} e^{-\Psi u^{-\epsilon}}\right] = \epsilon u^{-(\epsilon+1)} e^{-\Psi u^{-\epsilon}}. \tag{B.7}$$

Using this result to evaluate the integral above, the probability that the worker chooses to live in location n and commute to work in location i is:

$$\lambda_{ni} = \frac{\Psi_{ni}}{\Psi} = \frac{B_{ni}\left(\kappa_{ni} P_n^{\alpha} Q_n^{1-\alpha}\right)^{-\epsilon} (w_i)^{\epsilon}}{\sum_{r=1}^{S} \sum_{s=1}^{S} B_{rs}\left(\kappa_{rs} P_r^{\alpha} Q_r^{1-\alpha}\right)^{-\epsilon} (w_s)^{\epsilon}}. \tag{B.8}$$

Summing across all possible workplaces s, we obtain the probability that a worker chooses to live in location n out of all possible locations is:

$$\lambda_n = \frac{R_n}{\bar{L}} = \frac{\Psi_n}{\Psi} = \frac{\sum_{s=1}^{S} B_{ns}\left(\kappa_{ns} P_n^{\alpha} Q_n^{1-\alpha}\right)^{-\epsilon} (w_s)^{\epsilon}}{\sum_{r=1}^{S} \sum_{s=1}^{S} B_{rs}\left(\kappa_{rs} P_r^{\alpha} Q_r^{1-\alpha}\right)^{-\epsilon} (w_s)^{\epsilon}}. \tag{B.9}$$

Similarly, summing across all possible residence locations r, we obtain the probability that a worker chooses to work in location i out of all possible locations is:

$$\lambda_i = \frac{L_i}{\bar{L}} = \frac{\Psi_i}{\Psi} = \frac{\sum_{r=1}^{S} B_{ri}\left(\kappa_{ri} P_r^{\alpha} Q_r^{1-\alpha}\right)^{-\epsilon} (w_i)^{\epsilon}}{\sum_{r=1}^{S} \sum_{s=1}^{S} B_{rs}\left(\kappa_{rs} P_r^{\alpha} Q_r^{1-\alpha}\right)^{-\epsilon} (w_s)^{\epsilon}}. \tag{B.10}$$

For the measure of workers in location i (L_i), we can evaluate the conditional probability that they commute from location n (conditional on having chosen to work in location i):

$$\begin{aligned}\lambda_{ni|i} &= \Pr\left[u_{ni} \geq \max\{u_{ri}\}; \forall r\right], \\ &= \int_0^\infty \prod_{r \neq n} G_{ri}(u) g_{ni}(u) du, \\ &= \int_0^\infty e^{-\Psi_i u^{-\epsilon}} \epsilon \Psi_{ni} u^{-(\epsilon+1)} du.\end{aligned}$$

Using the result (B.7) to evaluate the integral above, the probability that a worker commutes from location n conditional on having chosen to work in location i is:

$$\lambda_{ni|i} = \frac{\Psi_{ni}}{\Psi_i} = \frac{B_{ni}\left(\kappa_{ni}P_n^{\alpha}Q_n^{1-\alpha}\right)^{-\epsilon}(w_i)^{\epsilon}}{\sum_{r=1}^{S} B_{ri}\left(\kappa_{ri}P_r^{\alpha}Q_r^{1-\alpha}\right)^{-\epsilon}(w_i)^{\epsilon}},$$

which simplifies to:

$$\lambda_{ni|i} = \frac{B_{ni}\left(\kappa_{ni}P_n^{\alpha}Q_n^{1-\alpha}\right)^{-\epsilon}}{\sum_{r=1}^{S} B_{ri}\left(\kappa_{ri}P_r^{\alpha}Q_r^{1-\alpha}\right)^{-\epsilon}}. \tag{B.11}$$

For the measure of residents of location n (R_n), we can evaluate the conditional probability that they commute to location i (conditional on having chosen to live in location n):

$$\begin{aligned} \lambda_{ni|n} &= \Pr\left[u_{ni} \geq \max\{u_{ns}\}; \forall s\right], \\ &= \int_0^{\infty} \prod_{s \neq i} G_{ns}(u) g_{ni}(u) du, \\ &= \int_0^{\infty} e^{-\Psi_n u^{-\epsilon}} \epsilon \Psi_{ni} u^{-(\epsilon+1)} du. \end{aligned}$$

Using the result (B.7) to evaluate the integral above, the probability that a worker commutes to location i conditional on having chosen to live in location n is:

$$\lambda_{ni|n} = \frac{\Psi_{ni}}{\Psi_n} = \frac{B_{ni}\left(\kappa_{ni}P_n^{\alpha}Q_n^{1-\alpha}\right)^{-\epsilon}(w_i)^{\epsilon}}{\sum_{s=1}^{S} B_{ns}\left(\kappa_{ns}P_n^{\alpha}Q_n^{1-\alpha}\right)^{-\epsilon}(w_s)^{\epsilon}},$$

which simplifies to:

$$\lambda_{ni|n} = \frac{B_{ni}\left(w_i/\kappa_{ni}\right)^{\epsilon}}{\sum_{s=1}^{S} B_{ns}\left(w_s/\kappa_{ns}\right)^{\epsilon}}. \tag{B.12}$$

These conditional commuting probabilities provide microeconomic foundations for the reduced-form gravity equations estimated in the empirical literature on commuting patterns.[32] The probability that a resident of location n commutes to location i depends on the wage at i and the amenities and commuting costs from living in n and working in i in the numerator ("bilateral resistance"). But it also depends on the wage at all other workplaces s and the amenities and commuting costs from living in n and commuting to all other workplaces s in the denominator ("multilateral resistance").

Labor market clearing requires that the measure of workers employed in each location i (L_i) equals the sum across all locations n of their measures of residents (R_n) times their conditional probabilities of

[32] See also McFadden (1974). For reduced-form evidence of the explanatory power of a gravity equation for commuting flows, see for example Erlander and Stewart (1990) and Sen and Smith (1995).

commuting to i (λ_{ni}):

$$\begin{aligned} L_i &= \sum_{n=1}^{S} \lambda_{ni|n} R_n \\ &= \sum_{n=1}^{S} \frac{B_{ni}\left(w_i/\kappa_{ni}\right)^{\epsilon}}{\sum_{s=1}^{S} B_{ns}\left(w_s/\kappa_{ns}\right)^{\epsilon}} R_n, \end{aligned} \tag{B.13}$$

where, since there is a continuous measure of workers residing in each location, there is no uncertainty in the supply of workers to each employment location.

Expected worker income conditional on living in location n is equal to the wages in all possible workplace locations weighted by the probabilities of commuting to those locations conditional on living in n:

$$\begin{aligned} \bar{v}_n &= \mathbb{E}\left[w|n\right] \\ &= \sum_{i=1}^{S} \lambda_{ni|n} w_i, \\ &= \sum_{i=1}^{S} \frac{B_{ni}\left(w_i/\kappa_{ni}\right)^{\epsilon}}{\sum_{s=1}^{S} B_{ns}\left(w_s/\kappa_{ns}\right)^{\epsilon}} w_i, \end{aligned} \tag{B.14}$$

where $\mathbb{E}$ denotes the expectations operator and the expectation is taken over the distribution for idiosyncratic amenities. Intuitively, expected worker income is high in locations that have low commuting costs (low κ_{ns}) to high-wage employment locations.

Finally, another implication of the Fréchet distribution of utility is that the distribution of utility conditional on residing in location n and commuting to location i is the same across all bilateral pairs of locations with positive residents and employment, and is equal to the distribution of utility for the economy as a whole. To establish this result, note that the distribution of utility conditional on residing in location n and commuting to location i is given by:

$$= \frac{1}{\lambda_{ni}} \int_0^u \prod_{s \neq i} G_{ns}(u) \left[\prod_{r \neq n} \prod_{s} G_{rs}(u)\right] g_{ni}(u) du, \tag{B.15}$$

$$\begin{aligned} &= \frac{1}{\lambda_{ni}} \int_0^u \left[\prod_{r=1}^{S} \prod_{s=1}^{S} e^{-\Psi_{rs} u^{-\epsilon}}\right] \epsilon \Psi_{ni} u^{-(\epsilon+1)} du, \\ &= \frac{\Psi}{\Psi_{ni}} \int_0^u e^{-\Psi u^{-\epsilon}} \epsilon \Psi_{ni} u^{-(\epsilon+1)} du, \\ &= e^{-\Psi u^{\epsilon}}. \end{aligned}$$

On the one hand, lower land prices in location n or a higher wage in location i raise the utility of a worker with a given realization of idiosyncratic amenities b, and hence increase the expected utility of residing in n and working in i. On the other hand, lower land prices or a higher wage induce workers with lower realizations of idiosyncratic amenities b to reside in n and work in i, which reduces the expected utility of residing in n and working in i. With a Fréchet distribution of utility, these two effects exactly offset one another. Pairs of residence and employment locations with more attractive characteristics attract more commuters on the extensive margin until expected utility is the same across all pairs of residence and employment locations within the economy.

B.2 Computing Counterfactuals Using Changes

We now use the structure of the model to solve for a counterfactual equilibrium using the observed values of variables in an initial equilibrium. We denote the value of variables in the counterfactual equilibrium by a prime (x') and the relative change of a variable between the initial and the counterfactual equilibrium by a hat ($\widehat{x} = x'/x$). Given the model's parameters $\{\alpha,\ \sigma,\ \epsilon,\ \delta,\ \kappa\}$ and counterfactual changes in the model's exogenous variables $\{\hat{A}_n,\ \hat{B}_n,\ \hat{\kappa}_{ni},\ \hat{d}_{ni}\}$, we can solve for the counterfactual changes in the model's endogenous variables $\{\hat{w}_n,\ \widehat{\bar{v}}_n,\ \hat{Q}_n,\ \hat{\pi}_{ni},\ \hat{\lambda}_{ni},\ \hat{P}_n,\ \hat{R}_n,\ \hat{L}_n\}$ from the following system of eight equations (using the iterative algorithm outlined below):

$$\widehat{w}_i \widehat{L}_i w_i L_i = \sum_{n \in N} \pi_{ni} \widehat{\pi}_{ni} \widehat{\bar{v}}_n \widehat{R}_n \bar{v}_n R_n, \tag{B.16}$$

$$\widehat{\bar{v}}_n \bar{v}_n = \sum_{i \in N} \frac{\lambda_{ni} \widehat{B}_{ni} \left(\widehat{w}_i / \widehat{\kappa}_{ni}\right)^{\epsilon}}{\sum_{s \in N} \lambda_{ns} \widehat{B}_{ns} \left(\widehat{w}_s / \widehat{\kappa}_{ns}\right)^{\epsilon}} \widehat{w}_i w_i, \tag{B.17}$$

$$\widehat{Q}_n = \widehat{\bar{v}}_n \widehat{R}_n, \tag{B.18}$$

$$\widehat{\pi}_{ni} \pi_{ni} = \frac{\pi_{ni} \widehat{L}_i \left(\widehat{d}_{ni} \widehat{w}_i / \widehat{A}_i\right)^{1-\sigma}}{\sum_{k \in N} \pi_{nk} \widehat{L}_k \left(\widehat{d}_{nk} \widehat{w}_k / \widehat{A}_k\right)^{1-\sigma}}, \tag{B.19}$$

$$\widehat{\lambda}_{ni} \lambda_{ni} = \frac{\lambda_{ni} \widehat{B}_{ni} \left(\widehat{P}_n^{\alpha} \widehat{Q}_n^{1-\alpha}\right)^{-\epsilon} \left(\widehat{w}_i / \widehat{\kappa}_{ni}\right)^{\epsilon}}{\sum_{r \in N} \sum_{s \in N} \lambda_{rs} \widehat{B}_{rs} \left(\widehat{P}_r^{\alpha} \widehat{Q}_r^{1-\alpha}\right)^{-\epsilon} \left(\widehat{w}_s / \widehat{\kappa}_{rs}\right)^{\epsilon}}, \tag{B.20}$$

$$\widehat{P}_n = \left(\frac{\widehat{L}_n}{\widehat{\pi}_{nn}}\right)^{\frac{1}{1-\sigma}} \frac{\widehat{d}_{nn} \widehat{w}_n}{\widehat{A}_n}, \tag{B.21}$$

$$\hat{R}_n = \frac{\bar{L}}{R_n} \sum_i \lambda_{ni} \hat{\lambda}_{ni}, \tag{B.22}$$

$$\hat{L}_i = \frac{\bar{L}}{L_i} \sum_n \lambda_{ni} \hat{\lambda}_{ni}, \tag{B.23}$$

where these equations correspond to the equality between income and expenditure (B.16), expected worker income (B.17), the land market clearing condition (B.18), trade shares (B.19), commuting probabilities

(B.20), price indices (B.21), residential choice probabilities (B.22) and workplace choice probabilities (B.23).

We solve this system of equations using the following iterative algorithm for the counterfactual equilibrium. Given the model's parameters $\{\alpha, \sigma, \epsilon, \delta, \kappa\}$ and changes in the exogenous variables of the model $\{\hat{A}_n, \hat{B}_n, \hat{\kappa}_{ni}, \hat{d}_{ni}\}$, we can solve for the resulting counterfactual changes in the endogenous variables of the model $\{\hat{w}_n, \hat{\bar{v}}_n, \hat{Q}_n, \hat{\pi}_{ni}, \hat{\lambda}_{ni}, \hat{P}_n, \hat{R}_n, \hat{L}_n\}$ from the system of eight equations (B.16)-(B.23). We solve this system of equations using the following iterative algorithm. We first conjecture changes in workplace wages and commuting probabilities at iteration t, $\hat{w}_i^{(t)}$ and $\hat{\lambda}_{ni}^{(t)}$. We next update these conjectures to $\widehat{w}_i^{(t+1)}$ and $\widehat{\lambda}_{ni}^{(t+1)}$ using the current guesses and data. We start by computing:

$$\hat{\bar{v}}_n^{(t)} = \frac{1}{\bar{v}_n} \sum_{i \in N} \frac{\hat{B}_{ni} \lambda_{ni} \left(\hat{w}_i^{(t)} / \widehat{\kappa}_{ni} \right)^{\epsilon}}{\sum_{s \in N} \hat{B}_{ns} \lambda_{ns} \left(\hat{w}_s^{(t)} / \widehat{\kappa}_{ns} \right)^{\epsilon}} \hat{w}_i^{(t)} w_i, \tag{B.24}$$

$$\hat{L}_i^{(t)} = \frac{\bar{L}}{L_i} \sum_n \lambda_{ni} \hat{\lambda}_{ni}^{(t)}, \tag{B.25}$$

$$\hat{R}_n^{(t)} = \frac{\bar{L}}{R_n} \sum_i \lambda_{ni} \hat{\lambda}_{ni}^{(t)}, \tag{B.26}$$

which are only a function of data and current guesses. We use (B.24) and (B.26) in (B.18) to compute:

$$\widehat{Q}_n^{(t)} = \hat{\bar{v}}_n^{(t)} \widehat{R}_n^{(t)}. \tag{B.27}$$

We use (B.25) and (B.19) to compute:

$$\hat{\pi}_{ni}^{(t)} = \frac{\widehat{L}_i^{(t)} \left(\widehat{d}_{ni} \widehat{w}_i^{(t)} / \hat{A}_i \right)^{1-\sigma}}{\sum_{k \in N} \pi_{nk} \widehat{L}_k^{(t)} \left(\widehat{d}_{nk} \widehat{w}_k^{(t)} / \hat{A}_k \right)^{1-\sigma}}. \tag{B.28}$$

We use (B.25), (B.28) and (B.21) to compute:

$$\widehat{P}_n^{(t)} = \left(\frac{\widehat{L}_n^{(t)}}{\widehat{\pi}_{nn}^{(t)}} \right)^{\frac{1}{1-\sigma}} \frac{\widehat{w}_n^{(t)}}{\hat{A}_n}. \tag{B.29}$$

We use (B.24)-(B.29) to rewrite (B.16) and (B.20) as:

$$\tilde{w}_i^{(t+1)} = \frac{1}{Y_i \widehat{L}_i^{(t)}} \sum_{n \in N} \pi_{ni} \widehat{\pi}_{ni}^{(t)} \hat{\bar{v}}_n^{(t)} \widehat{R}_n^{(t)} Y_n, \tag{B.30}$$

$$\tilde{\lambda}_{ni}^{(t+1)} = \frac{\hat{B}_{ni} \left(\widehat{P}_n^{(t)\alpha} \widehat{Q}_n^{(t)1-\alpha} \right)^{-\epsilon} \left(\widehat{w}_i^{(t)} / \widehat{\kappa}_{ni} \right)^{\epsilon}}{\sum_{r \in N} \sum_{s \in N} \hat{B}_{rs} \lambda_{rs} \left(\widehat{P}_r^{(t)\alpha} \widehat{Q}_r^{(t)1-\alpha} \right)^{-\epsilon} \left(\widehat{w}_s^{(t)} / \widehat{\kappa}_{rs} \right)^{\epsilon}}. \tag{B.31}$$

Finally, we update our conjectures for wages and commuting probabilities using:

$$\hat{w}_i^{(t+1)} = \zeta \hat{w}_i^{(t)} + (1-\zeta)\,\tilde{w}_i^{(t+1)}, \tag{B.32}$$
$$\hat{\lambda}_i^{(t+1)} = \zeta \hat{\lambda}_i^{(t)} + (1-\zeta)\,\tilde{\lambda}_i^{(t+1)}, \tag{B.33}$$

where $\zeta \in (0,1)$ is an adjustment factor.

In Proposition 1, we provide conditions under which there exists a unique equilibrium in the model. Under these conditions, the above algorithm converges rapidly to the unique counterfactual equilibrium.

B.3 Proof of Proposition 1

Assume balanced trade so $w_i L_i = \sum_{n \in N} X_{ni}$, and so

$$w_i L_i = \sum_{n \in N} \frac{\frac{L_i}{\sigma F}\left(\frac{\sigma}{\sigma-1}\frac{d_{ni} w_i}{A_i}\right)^{1-\sigma}}{P_n^{1-\sigma}} \bar{v}_n R_n. \tag{B.34}$$

Using (10) to substitute for the price index, this balanced trade condition can be written as

$$w_i^{\sigma} A_i^{1-\sigma} = \sum_{n \in N} \pi_{nn} \frac{d_{ni}^{1-\sigma}}{d_{nn}^{1-\sigma}} \frac{R_n}{L_n} w_n^{\sigma-1} A_n^{1-\sigma} \bar{v}_n. \tag{B.35}$$

Note that balanced trade (B.34) also can be written as

$$w_i^{\sigma} A_i^{1-\sigma} = \sum_{n \in N} \frac{1}{\sigma F}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} d_{ni}^{1-\sigma} P_n^{\sigma-1} \bar{v}_n R_n. \tag{B.36}$$

Note that the residential choice probabilities (14) can be written as

$$\frac{R_n}{\bar{L}} = \frac{\bar{w}_n^{\epsilon}}{\left(\bar{U}/\delta\right)^{\epsilon} P_n^{\alpha\epsilon} Q_n^{(1-\alpha)\epsilon}}, \tag{B.37}$$

where $\bar{w}_n$ is a measure of commuting market access

$$\bar{w}_n = \left[\sum_{s \in N} B_{ns}\left(w_s/\kappa_{ns}\right)^{\epsilon}\right]^{\frac{1}{\epsilon}}. \tag{B.38}$$

Rearranging the residential choice probabilities (B.37), we obtain the following expression for the price index

$$P_n^{\alpha\epsilon} = \frac{\bar{w}_n^{\epsilon}}{\left(\bar{U}/\delta\right)^{\epsilon} Q_n^{(1-\alpha)\epsilon}\left(R_n/\bar{L}\right)}. \tag{B.39}$$

Using land market clearing (5), this expression for the price index can be re-written as

$$P_n^{1-\sigma} = \left(\frac{\bar{w}_n}{\bar{W}}\right)^{\frac{1-\sigma}{\alpha}} \bar{v}_n^{-(1-\sigma)\left(\frac{1-\alpha}{\alpha}\right)} H_n^{(1-\sigma)\left(\frac{1-\alpha}{\alpha}\right)} R_n^{-(1-\sigma)\left(\frac{1}{\alpha\epsilon}+\frac{1-\alpha}{\alpha}\right)}, \tag{B.40}$$

where

$$\bar{W} = \left(\bar{U}/\delta\right)(1-\alpha)^{1-\alpha}\bar{L}^{-1/\epsilon}.$$

Substituting price index expression (B.40) into (10), we obtain the following expression for the domestic trade share

$$\pi_{nn} = \bar{W}^{\frac{1-\sigma}{\alpha}} \frac{\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}}{\sigma F} \bar{w}_n^{\frac{\sigma-1}{\alpha}} w_n^{1-\sigma} L_n R_n^{(1-\sigma)\left(\frac{1}{\alpha\epsilon}+\frac{1-\alpha}{\alpha}\right)} \bar{v}_n^{(1-\sigma)\left(\frac{1-\alpha}{\alpha}\right)} H_n^{(\sigma-1)\left(\frac{1-\alpha}{\alpha}\right)} A_n^{\sigma-1} d_{nn}^{1-\sigma}. \tag{B.41}$$

Now substitute this expression for the domestic trade share in the wage equation (B.35) to obtain

$$w_n^{\sigma} A_n^{1-\sigma} = \bar{W}^{\frac{1-\sigma}{\alpha}} \frac{\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}}{\sigma F} \left[\sum_{k\in N} \bar{w}_k^{\frac{\sigma-1}{\alpha}} R_k^{1-(\sigma-1)\left(\frac{1}{\alpha\epsilon}+\frac{1-\alpha}{\alpha}\right)} \bar{v}_k^{1-(\sigma-1)\left(\frac{1-\alpha}{\alpha}\right)} H_k^{(\sigma-1)\left(\frac{1-\alpha}{\alpha}\right)} d_{kn}^{1-\sigma}\right]. \tag{B.42}$$

Now return to the price index (B.40) and use the definition of the price index from (10) and commuting market clearing (16)

$$\begin{aligned} &\bar{w}_n^{\frac{1-\sigma}{\alpha}} \bar{v}_n^{-(1-\sigma)\left(\frac{1-\alpha}{\alpha}\right)} H_n^{(1-\sigma)\left(\frac{1-\alpha}{\alpha}\right)} R_n^{-(1-\sigma)\left(\frac{1}{\alpha\epsilon}+\frac{1-\alpha}{\alpha}\right)} \\ &= \bar{W}^{\frac{1-\sigma}{\alpha}} \frac{\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}}{\sigma F} \left\{\sum_{k\in N}\left[\sum_{r\in N} \frac{B_{rk}\left(w_k/\kappa_{rk}\right)^{\epsilon}}{\sum_{s\in N} B_{rs}\left(w_s/\kappa_{rs}\right)^{\epsilon}} R_r\right]\left(\frac{d_{nk}w_k}{A_k}\right)^{1-\sigma}\right\}. \end{aligned} \tag{B.43}$$

Together (B.42) and (B.43) provide systems of $2N$ equations that determine the $2N$ unknown values of $\{\mathbf{w}, \mathbf{R}\}$. Consider first the system (B.42). Define the n^{th} element of the operator T_w by

$$T_w\left(\mathbf{w};\mathbf{R}\right)_n \equiv \left(\frac{\bar{W}^{\frac{1-\sigma}{\alpha}} \frac{\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}}{\sigma F} \left[\sum_{k\in N} \bar{w}_k^{\frac{\sigma-1}{\alpha}} R_k^{1-(\sigma-1)\left(\frac{1}{\alpha\epsilon}+\frac{1-\alpha}{\alpha}\right)} \bar{v}_k^{1-(\sigma-1)\left(\frac{1-\alpha}{\alpha}\right)} H_k^{(\sigma-1)\left(\frac{1-\alpha}{\alpha}\right)} d_{kn}^{1-\sigma}\right]}{A_n^{1-\sigma}}\right)^{\frac{1}{\sigma}}.$$

Note that $T_w\left(\mathbf{w};\mathbf{R}\right)$ satisfies the following properties:

1. Homogeneity in $\mathbf{R}$. For any $\mathbf{w}$, we have

$$T_w\left(\mathbf{w};\lambda\mathbf{R}\right) = \lambda^{\left[1-(\sigma-1)\left(\frac{1}{\alpha\epsilon}+\frac{1-\alpha}{\alpha}\right)\right]\frac{1}{\sigma}} T_w\left(\mathbf{w};\mathbf{R}\right)$$

2. Homogeneity in $\mathbf{w}$. For any $\lambda \geq 0$, we have

$$T_w(\lambda \mathbf{w}; \mathbf{R}) = \lambda^{\left[\frac{\sigma-1}{\alpha}+1-(\sigma-1)\left(\frac{1-\alpha}{\alpha}\right)\right]\frac{1}{\sigma}} T_w(\mathbf{w}; \mathbf{R}).$$

For this result, simply note that $\bar{w}_n(\lambda w_n) = \lambda \bar{w}_n(w_n)$ and $\bar{v}_n(\lambda w_n) = \lambda \bar{v}_n(w_n)$.
Note also that $\left[\frac{\sigma-1}{\alpha} + 1 - (\sigma-1)\left(\frac{1-\alpha}{\alpha}\right)\right]\frac{1}{\sigma} = 1$.

3. $T_w(\mathbf{w}; \mathbf{R})$ is monotone increasing in $\mathbf{w}$, namely, that for $\mathbf{w}, \mathbf{w}' \in B(X)$, $\mathbf{w} \leq \mathbf{w}'$ implies $T(\mathbf{w}) \leq T(\mathbf{w}')$. In fact:

$$\frac{d\bar{w}_n}{\bar{w}_n} = \sum_{s\in N} \frac{B_{ns}(w_s/\kappa_{ns})^{\epsilon}}{\sum_{k\in N} B_{nk}(w_k/\kappa_{nk})^{\epsilon}} \frac{dw_s}{w_s} > 0,$$

which establishes that $\bar{\mathbf{w}}$ is monotonic in $\mathbf{w}$. Now note that $\bar{v}_n$ can be written as:

$$\bar{v}_n = \sum_{s\in N} B_{ns}\kappa_{ns}^{-\epsilon} w_s^{1+\epsilon} \bar{w}_n^{-\epsilon},$$

so that:

$$d\bar{v}_n = (1+\epsilon)\sum_{s\in N} B_{ns}\kappa_{ns}^{-\epsilon} w_s^{1+\epsilon} \bar{w}_n^{-\epsilon} \frac{dw_s}{w_s} - \epsilon \sum_{s\in N} B_{ns}\kappa_{ns}^{-\epsilon} w_s^{1+\epsilon} \bar{w}_n^{-\epsilon} \sum_{k\in N} \left(\frac{d\bar{w}_n}{dw_k}\frac{w_k}{\bar{w}_n}\right)\frac{dw_k}{w_k} > 0.$$

Using $dw_s/w_s = d\ln w_s$, this expression can be re-written as:

$$d\bar{v}_n = (1+\epsilon)\sum_{s\in N} K_{ns}\ln\omega_s - \epsilon \sum_{s\in N} K_{ns} \sum_{k\in N} \xi_{nk}\ln\omega_k,$$

where

$$\begin{aligned}
K_{ns} &\equiv B_{ns}\kappa_{ns}^{-\epsilon} w_s^{1+\epsilon} \bar{w}_n^{-\epsilon}, \\
\ln\omega_s &\equiv d\ln w_s = \lim_{h\to 0} \ln(w_s + h) - \ln(w_s) = \lim_{h\to 0} \ln((w_s+h)/w_s), \\
\xi_{nk} &\equiv \frac{B_{nk}(w_k/\kappa_{nk})^{\epsilon}}{\sum_{s\in N} B_{ns}(w_s/\kappa_{ns})^{\epsilon}}, \\
\sum_{k\in N}\xi_{nk} &= 1.
\end{aligned}$$

Since $\ln\omega_s$ is concave in ω_s, it follows that:

$$\sum_{s\in N} K_{ns}\ln\omega_s > \sum_{s\in N} K_{ns} \sum_{k\in N} \xi_{nk}\ln\omega_k,$$

Given Properties 2 and 3, the results in Fujimoto and Krause (1985) guarantee that there exists a unique fixed point for $\mathbf{w}$, up to a normalization, conditional on $\mathbf{R}$,

$$\mathbf{w} = T_w(\mathbf{w};\mathbf{R}).$$

Denote this fixed point by $\mathbf{w}^{FP}(\mathbf{R})$.

4. Using the results from 1.-3. above, note that we can set $\mathbf{w}^{FP}(\lambda\mathbf{R}) = \mathbf{w}^{FP}(\mathbf{R})$ for any λ with an appropriate change in $\bar{U}$; in fact, given the Property 1, and since $\bar{U}$ shifts $T_w(\mathbf{w};\mathbf{R})$ proportionally by $\bar{U}^{\frac{1-\sigma}{\sigma\alpha}}$, we can always find an appropriate $\bar{U}'$, namely $\bar{U}'^{\frac{1-\sigma}{\sigma\alpha}}\lambda^{\left[1-(\sigma-1)\left(\frac{1}{\alpha\epsilon}+\frac{1-\alpha}{\alpha}\right)\right]\frac{1}{\sigma}} = 1$. We show that there is a $\bar{U}$ that clears the market at the end of the proof.

Now consider the system (B.43). Define the n^{th} element of an operator $\tilde{T}_L$ by

$$\tilde{T}_L(\mathbf{R};\mathbf{w})_n \equiv \left(\frac{\bar{W}^{\frac{1-\sigma}{\alpha}}\frac{\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}}{\sigma F}\left[\sum_{k\in N}\left[\sum_{r\in N}\frac{B_{rn}(w_n/\kappa_{rn})^{\epsilon}}{\sum_{s\in N}B_{rs}(w_s/\kappa_{rs})^{\epsilon}}R_r\right]\left(\frac{d_{nk}w_k}{A_k}\right)^{1-\sigma}\right]\times}{\bar{w}_n^{-\frac{1-\sigma}{\alpha}}\bar{v}_n^{(1-\sigma)\left(\frac{1-\alpha}{\alpha}\right)}H_n^{-(1-\sigma)\left(\frac{1-\alpha}{\alpha}\right)}}\right)^{\frac{1}{-(1-\sigma)\left(\frac{1}{\alpha\epsilon}+\frac{1-\alpha}{\alpha}\right)}}.$$

Note that $\tilde{T}_L(\mathbf{R};\mathbf{w})$ satisfies the following properties:

- Homogeneity in $\mathbf{R}$:

$$\tilde{T}_L(\lambda\mathbf{R};\mathbf{w}) = \lambda^{\frac{1}{-(1-\sigma)\left(\frac{1}{\alpha\epsilon}+\frac{1-\alpha}{\alpha}\right)}}\tilde{T}_L(\mathbf{R};\mathbf{w})$$

- Monotonicity in $\mathbf{R}$, immediate by inspection.

Define $T_L(\mathbf{R}) \equiv \tilde{T}_L\left(\mathbf{R},\mathbf{w}^{FP}(\mathbf{R})\right)$. Note that this operator satisfies the following properties:

5. Homogeneity:

$$\begin{aligned} T_L(\lambda\mathbf{R}) &\equiv \tilde{T}_L\left(\lambda\mathbf{R},\mathbf{w}^{FP}(\lambda\mathbf{R})\right) = \lambda^{\frac{1}{-(1-\sigma)\left(\frac{1}{\alpha\epsilon}+\frac{1-\alpha}{\alpha}\right)}}\tilde{T}_L\left(\mathbf{R},\mathbf{w}^{FP}(\lambda\mathbf{R})\right) = \\ &= \lambda^{\frac{1}{-(1-\sigma)\left(\frac{1}{\alpha\epsilon}+\frac{1-\alpha}{\alpha}\right)}}\tilde{T}_L\left(\mathbf{R};\mathbf{w}^{FP}(\mathbf{R})\right) \end{aligned}$$

where the first equality follows from the expression for $\tilde{T}_L(\mathbf{R};\mathbf{w})$ and the second equality from setting $\mathbf{w}^{FP}(\lambda\mathbf{R}) = \mathbf{w}^{FP}(\mathbf{R})$, as implied by Property 4 of $\mathbf{w}^{FP}(\mathbf{R}) = T_w\left(\mathbf{w}^{FP};\mathbf{R}\right)$ above;

6. Monotonicity in $\mathbf{R}$: note that $\tilde{T}_L(\mathbf{R};\mathbf{w})$ is homogeneous of degree 0 in $\mathbf{w}$, since $1-\sigma-\frac{1-\sigma}{\alpha}+(1-\sigma)\left(\frac{1-\alpha}{\alpha}\right)=0$. Hence, given any vector $d\mathbf{R}$, the total differential in $\tilde{T}_L\left(\mathbf{R},\mathbf{w}^{FP}(\mathbf{R})\right)$ induced through changes in $\mathbf{w}^{FP}(\mathbf{R})$ sums to zero by Euler's theorem. $\tilde{T}_L\left(\mathbf{R},\mathbf{w}^{FP}(\mathbf{R})\right)$ is then monotone in $\mathbf{R}$ by inspection of the expression of $\tilde{T}_L(\mathbf{R};\mathbf{w})$.

Then, if $\frac{1}{-(1-\sigma)\left(\frac{1}{\alpha\epsilon}+\frac{1-\alpha}{\alpha}\right)} \in (0,1]$, the results in Fujimoto and Krause (1985) guarantee then that there is a unique fixed point $\mathbf{R}^{FP}$ such that

$$\mathbf{R}^{FP} = T_L\left(\mathbf{R}^{FP}\right).$$

Since $-(1-\sigma)\left(\frac{1}{\alpha\epsilon}+\frac{1-\alpha}{\alpha}\right)>0$ by assumption, the binding constraint is that

$$-(1-\sigma)\left(\frac{1}{\alpha\epsilon}+\frac{1-\alpha}{\alpha}\right)>1,$$

or

$$\sigma>\frac{1+\epsilon}{1+\epsilon-\alpha\epsilon}. \tag{B.44}$$

Hence, under the condition imposed in the statement of the proposition, there exists a unique solution to the system of equations (B.42) and (B.43).

The only remaining task is to use the labor market clearing condition to guarantee that there exists a unique scalar $\bar{U}$ such that $\sum_{n\in N}R_n=\bar{L}$. First note that the above result and the definition of the operator $T_L(\mathbf{R})$ guarantees that for any $\lambda>0$ and any $\bar{U}$

$$\mathbf{R}_{\lambda\bar{U}}=T_L\left(\mathbf{R}_{\lambda\bar{U}};\lambda\bar{U}\right)=T_L\left(\mathbf{R}_{\lambda\bar{U}};\bar{U}\right)\lambda^{\frac{\frac{1-\sigma}{\alpha\sigma}}{-(1-\sigma)\left(\frac{1}{\alpha\epsilon}+\frac{1-\alpha}{\alpha}\right)}}=T_L\left(\mathbf{R}_{\lambda\bar{U}};\bar{U}\right)\lambda^{\frac{1}{-\sigma\left(\frac{1}{\epsilon}+1-\alpha\right)}}$$

where we have used the definition of $\bar{W}$. Then,

$$\Gamma\mathbf{R}_{\lambda\bar{U}}=T_L\left(\Gamma\mathbf{R}_{\lambda\bar{U}};\lambda\bar{U}\right)\Gamma^{1-\frac{1}{-(1-\sigma)\left(\frac{1}{\alpha\epsilon}+\frac{1-\alpha}{\alpha}\right)}}=T\left(\Gamma\mathbf{R}_{\lambda\bar{U}};\bar{U}\right)\Gamma^{1-\frac{1}{-(1-\sigma)\left(\frac{1}{\alpha\epsilon}+\frac{1-\alpha}{\alpha}\right)}}\lambda^{\frac{\frac{1-\sigma}{\alpha\sigma}}{-(1-\sigma)\left(\frac{1}{\alpha\epsilon}+\frac{1-\alpha}{\alpha}\right)}}$$

and so $\mathbf{R}_{\bar{U}}=\Gamma\mathbf{R}_{\lambda\bar{U}}$ if

$$\Gamma=\lambda^{\frac{\sigma-1}{\alpha\sigma}\left(1-(\sigma-1)\left(\frac{1}{\alpha\epsilon}+\frac{1-\alpha}{\alpha}\right)\right)}.$$

Note that since $1-(\sigma-1)\left(\frac{1}{\alpha\epsilon}+\frac{1-\alpha}{\alpha}\right)<0$, $\mathbf{R}_{\bar{U}}$ is monotone decreasing in $\bar{U}$ through this power function. Hence, there exists a unique $\bar{U}$ such that $\sum_{n\in N}R_n=\bar{L}$.

B.4 Proof of Proposition 2

B.4.1 New Economic Geography Model

We begin by considering our new economic geography model with agglomeration forces through love of variety and increasing returns to scale. The general equilibrium vector $\{w_n,\,\bar{v}_n,\,Q_n,\,L_n,\,R_n,\,P_n\}$ and scalar $\bar{U}$ solve the following system of equations. First, income equals expenditure on goods produced in each location:

$$w_iL_i=\sum_{n\in N}\frac{L_i\left(d_{ni}w_i/A_i\right)^{1-\sigma}}{\sum_{k\in N}L_k\left(d_{nk}w_k/A_k\right)^{1-\sigma}}\bar{v}_nR_n. \tag{B.45}$$

Second, expected worker income depends on wages:

$$\bar{v}_n=\sum_{i\in N}\frac{B_{ni}\left(w_i/\kappa_{ni}\right)^{\epsilon}}{\sum_{s\in N}B_{ns}\left(w_s/\kappa_{ns}\right)^{\epsilon}}w_i. \tag{B.46}$$

Third, land prices depend on expected worker income and the measure of residents:

$$Q_n = (1-\alpha)\frac{\bar{v}_n R_n}{H_n}. \tag{B.47}$$

Fourth, workplace choice probabilities solve:

$$\frac{L_n}{\bar{L}} = \frac{\sum_{r\in N} B_{rn}\left(\kappa_{rn} P_r^{\alpha} Q_r^{1-\alpha}\right)^{-\epsilon} w_n^{\epsilon}}{\sum_{r\in N}\sum_{s\in N} B_{rs}\left(\kappa_{rs} P_r^{\alpha} Q_r^{1-\alpha}\right)^{-\epsilon} w_s^{\epsilon}}. \tag{B.48}$$

Fifth, residential choice probabilities solve:

$$\frac{R_n}{\bar{L}} = \frac{\sum_{s\in N} B_{ns}\left(\kappa_{ns} P_n^{\alpha} Q_n^{1-\alpha}\right)^{-\epsilon} w_s^{\epsilon}}{\sum_{r\in N}\sum_{s\in N} B_{rs}\left(\kappa_{rs} P_r^{\alpha} Q_r^{1-\alpha}\right)^{-\epsilon} w_s^{\epsilon}}. \tag{B.49}$$

Sixth, price indices solve:

$$P_n = \frac{\sigma}{\sigma-1}\left(\frac{1}{\sigma F}\right)^{\frac{1}{1-\sigma}}\left[\sum_{i\in N} L_i\left(d_{ni} w_i / A_i\right)^{1-\sigma}\right]^{\frac{1}{1-\sigma}}. \tag{B.50}$$

Seventh, expected utility satisfies:

$$\bar{U} = \delta\left[\sum_{r\in N}\sum_{s\in N} B_{rs}\left(\kappa_{rs} P_r^{\alpha} Q_r^{1-\alpha}\right)^{-\epsilon} w_s^{\epsilon}\right]^{\frac{1}{\epsilon}}, \tag{B.51}$$

where $\delta = \Gamma\left(\frac{\epsilon-1}{\epsilon}\right)$ and $\Gamma(\cdot)$ is the Gamma function.

B.4.2 Eaton and Kortum (2002) with External Economies of Scale

We consider an Eaton and Kortum (2002) with external economies of scale augmented to incorporate heterogeneity in worker preferences over workplace and residence locations. Utility remains as specified in (1), except that the consumption index (C_n) is defined over a fixed interval of goods $j \in [0,1]$:

$$C_n = \left[\int_0^1 c_n(j)^{\rho}\, dj\right]^{\frac{1}{\rho}}.$$

Productivity for each good j in each location i is drawn from an independent Fréchet distribution:

$$F_i(z) = e^{-A_i z^{-\theta}}, \qquad A_i = \tilde{A}_i L_i^{\eta}, \qquad \theta > 1,$$

where the scale parameter of this distribution (A_i) depends on the measure of workers (L_i) and η parameterizes the strength of external economies of scale. The general equilibrium vector $\{w_n, \bar{v}_n, Q_n, L_n, R_n, P_n\}$ and scalar $\bar{U}$ solve the following system of equations. First, income equals expenditure on goods produced in each location:

$$w_i L_i = \sum_{n \in N} \frac{\tilde{A}_i L_i^{\eta} \left(d_{ni} w_i\right)^{-\theta}}{\sum_{k \in N} \tilde{A}_k L_k^{\eta} \left(d_{nk} w_k\right)^{-\theta}} \bar{v}_n R_n. \tag{B.52}$$

Second, expected worker income depends on wages:

$$\bar{v}_n = \sum_{i \in N} \frac{B_{ni} \left(w_i / \kappa_{ni}\right)^{\epsilon}}{\sum_{s \in N} B_{ns} \left(w_s / \kappa_{ns}\right)^{\epsilon}} w_i. \tag{B.53}$$

Third, land prices depend on expected worker income and the measure of residents:

$$Q_n = (1 - \alpha) \frac{\bar{v}_n R_n}{H_n}. \tag{B.54}$$

Fourth, workplace choice probabilities solve:

$$\frac{L_n}{\bar{L}} = \frac{\sum_{r \in N} B_{rn} \left(\kappa_{rn} P_r^{\alpha} Q_r^{1-\alpha}\right)^{-\epsilon} w_n^{\epsilon}}{\sum_{r \in N} \sum_{s \in N} B_{rs} \left(\kappa_{rs} P_r^{\alpha} Q_r^{1-\alpha}\right)^{-\epsilon} w_s^{\epsilon}}. \tag{B.55}$$

Fifth, residential choice probabilities solve:

$$\frac{R_n}{\bar{L}} = \frac{\sum_{s \in N} B_{ns} \left(\kappa_{ns} P_n^{\alpha} Q_n^{1-\alpha}\right)^{-\epsilon} w_s^{\epsilon}}{\sum_{r \in N} \sum_{s \in N} B_{rs} \left(\kappa_{rs} P_r^{\alpha} Q_r^{1-\alpha}\right)^{-\epsilon} w_s^{\epsilon}}. \tag{B.56}$$

Sixth, price indices solve:

$$P_n = \gamma \left[\sum_{i \in N} \tilde{A}_i L_i^{\eta} \left(d_{ni} w_i\right)^{-\theta}\right]^{-\frac{1}{\theta}}, \tag{B.57}$$

where $\gamma = \left[\Gamma\left(\frac{\theta - (\sigma - 1)}{\theta}\right)\right]^{\frac{1}{1-\sigma}}$ and $\Gamma(\cdot)$ denotes the Gamma function. Seventh, expected utility satisfies:

$$\bar{U} = \delta \left[\sum_{r \in N} \sum_{s \in N} B_{rs} \left(\kappa_{rs} P_r^{\alpha} Q_r^{1-\alpha}\right)^{-\epsilon} w_s^{\epsilon}\right]^{\frac{1}{\epsilon}}. \tag{B.58}$$

The system of equations (B.52)-(B.58) is isomorphic to the system of equations (B.45)-(B.51) under the following restrictions:

$$\begin{aligned}
\theta^{\text{EK}} &= \sigma^{\text{NEG}} - 1, \\
\eta^{\text{EK}} &= 1, \\
A_i^{\text{EK}} &= \left(A_i^{\text{NEG}}\right)^{\sigma^{\text{NEG}} - 1}, \\
\gamma^{\text{EK}} &= \frac{\sigma^{\text{NEG}}}{\sigma^{\text{NEG}} - 1} \left(\frac{1}{\sigma^{\text{NEG}} F^{\text{NEG}}}\right)^{\frac{1}{1-\sigma^{\text{NEG}}}}.
\end{aligned}$$

Under these parameter restrictions, both models generate the same general equilibrium vector $\{w_n,\ \bar{v}_n,\ Q_n,\ L_n,\ R_n,\ P_n\}$ and scalar $\bar{U}$.

B.4.3 Armington (1969) with External Economies of Scale

We consider an Armington (1969) model with external economies of scale augmented to incorporate heterogeneity in worker preferences over workplace and residence locations. Utility remains as specified in (1), except that the consumption index (C_n) is defined over goods that are horizontally differentiated by location of origin:

$$C_n = \left[\sum_{i \in N} C_i^{\rho}\right]^{\frac{1}{\rho}}.$$

The goods supplied by each location are produced under conditions of perfect competition and external economies of scale such that the "cost inclusive of freight" (cif) price of good produced in location i and consumed in location n is:

$$P_{ni} = \frac{d_{ni} w_i}{A_i}, \qquad A_i = \tilde{A}_i L_i^{\eta}.$$

The general equilibrium vector $\{w_n,\ \bar{v}_n,\ Q_n,\ L_n,\ R_n,\ P_n\}$ and scalar $\bar{U}$ solve the following system of equations. First, income equals expenditure on goods produced in each location:

$$w_i L_i = \sum_{n \in N} \frac{A_i^{\sigma-1} L_i^{\eta(\sigma-1)} \left(d_{ni} w_i\right)^{1-\sigma}}{\sum_{k \in N} A_k^{\sigma-1} L_k^{\eta(\sigma-1)} \left(d_{nk} w_k\right)^{1-\sigma}} \bar{v}_n R_n. \tag{B.59}$$

Second, expected worker income depends on wages:

$$\bar{v}_n = \sum_{i \in N} \frac{B_{ni} \left(w_i / \kappa_{ni}\right)^{\epsilon}}{\sum_{s \in N} B_{ns} \left(w_s / \kappa_{ns}\right)^{\epsilon}} w_i. \tag{B.60}$$

Third, land prices depend on expected worker income and the measure of residents:

$$Q_n = (1-\alpha) \frac{\bar{v}_n R_n}{H_n}. \tag{B.61}$$

Fourth, workplace choice probabilities solve:

$$\frac{L_n}{\bar{L}} = \frac{\sum_{r \in N} B_{rn} \left(\kappa_{rn} P_r^{\alpha} Q_r^{1-\alpha}\right)^{-\epsilon} w_n^{\epsilon}}{\sum_{r \in N} \sum_{s \in N} B_{rs} \left(\kappa_{rs} P_r^{\alpha} Q_r^{1-\alpha}\right)^{-\epsilon} w_s^{\epsilon}}. \tag{B.62}$$

Fifth, residential choice probabilities solve:

$$\frac{R_n}{\bar{L}} = \frac{\sum_{s \in N} B_{ns} \left(\kappa_{ns} P_n^{\alpha} Q_n^{1-\alpha}\right)^{-\epsilon} w_s^{\epsilon}}{\sum_{r \in N} \sum_{s \in N} B_{rs} \left(\kappa_{rs} P_r^{\alpha} Q_r^{1-\alpha}\right)^{-\epsilon} w_s^{\epsilon}}. \tag{B.63}$$

Sixth, price indices solve:

$$P_n = \left[\sum_{i \in N} A_i^{\sigma-1} L_i^{\eta(\sigma-1)} \left(d_{ni} w_i \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}}. \tag{B.64}$$

Seventh, expected utility satisfies:

$$\bar{U} = \delta \left[\sum_{r \in N} \sum_{s \in N} B_{rs} \left(\kappa_{rs} P_r^{\alpha} Q_r^{1-\alpha} \right)^{-\epsilon} w_s^{\epsilon} \right]^{\frac{1}{\epsilon}}. \tag{B.65}$$

The system of equations (B.59)-(B.65) is isomorphic to the system of equations (B.45)-(B.51) under the following restrictions:

$$\begin{aligned} \sigma^{\text{AR}} &= \sigma^{\text{NEG}}, \\ \eta^{\text{EK}} &= \frac{1}{\sigma^{\text{NEG}} - 1}, \\ A_i^{\text{AR}} &= A_i^{\text{NEG}}, \\ 1 &= \frac{\sigma^{\text{NEG}}}{\sigma^{\text{NEG}} - 1} \left(\frac{1}{\sigma^{\text{NEG}} F^{\text{NEG}}} \right)^{\frac{1}{1-\sigma^{\text{NEG}}}}. \end{aligned}$$

Under these parameter restrictions, both models generate the same general equilibrium vector $\{w_n, \bar{v}_n, Q_n, L_n, R_n, P_n\}$ and scalar $\bar{U}$.

B.5 Proof of Proposition 3

Note that the goods market clearing condition (19) can be written as the following excess demand system:

$$D_i(\tilde{\mathbf{A}}) = w_i L_i - \sum_{n \in N} \frac{\tilde{A}_i L_i \left(d_{ni} w_i \right)^{1-\sigma}}{\sum_{k \in N} \tilde{A}_k L_k \left(d_{nk} w_k \right)^{1-\sigma}} \bar{v}_n R_n = 0, \tag{B.66}$$

where $\tilde{A}_i = A_i^{\sigma-1}$ and $\{w_i, L_i, \bar{v}_n, R_n, d_{ni}\}$ have already been determined from the observed data or our parameterization of trade costs. This excess demand system exhibits the following properties in $\tilde{A}_i$:

Property (i): $D(\tilde{\mathbf{A}})$ is continuous, as follows immediately from inspection of (B.66).

Property (ii): $D(\tilde{\mathbf{A}})$ is homogenous of degree zero, as follows immediately from inspection of (B.66).

Property (iii): $\sum_{i \in N} D_i \left(\tilde{\mathbf{A}} \right) = 0$ for all $\tilde{\mathbf{A}} \in \Re_+^N$. This property can be established by noting:

$$\begin{aligned} \sum_{i \in N} D_i \left(\tilde{\mathbf{A}} \right) &= \sum_{i \in N} w_i L_i - \sum_{n \in N} \frac{\sum_{i \in N} \tilde{A}_i L_i \left(d_{ni} w_i \right)^{1-\sigma}}{\sum_{k \in N} \tilde{A}_k L_k \left(d_{nk} w_k \right)^{1-\sigma}} \bar{v}_n R_n, \\ &= \sum_{i \in N} w_i L_i - \sum_{n \in N} \bar{v}_n R_n, \\ &= 0. \end{aligned}$$

Property (iv): $D(\tilde{\mathbf{A}})$ exhibits gross substitution:

$$\frac{\partial D_i\left(\tilde{\mathbf{A}}\right)}{\partial \tilde{A}_r} > 0 \qquad \text{for all } i, r, \neq i, \qquad\qquad \text{for all } \tilde{\mathbf{A}} \in \Re_+^N,$$
$$\frac{\partial D_i\left(\tilde{\mathbf{A}}\right)}{\partial \tilde{A}_i} < 0 \qquad \text{for all } i, \qquad\qquad \text{for all } \tilde{\mathbf{A}} \in \Re_+^N.$$

This property can be established by noting:

$$\frac{\partial D_i\left(\tilde{\mathbf{A}}\right)}{\partial \tilde{A}_r} = \sum_{n \in N} \frac{L_r\left(d_{nr}w_r\right)^{1-\sigma} \tilde{A}_i L_i \left(d_{ni}w_i\right)^{1-\sigma}}{\left[\sum_{k\in N} \tilde{A}_k L_k \left(d_{nk}w_k\right)^{1-\sigma}\right]^2} \bar{v}_n R_n > 0.$$

and using homogeneity of degree zero, which implies:

$$\nabla D\left(\tilde{\mathbf{A}}\right)\tilde{\mathbf{A}} = 0,$$

and hence:

$$\frac{\partial D_i\left(\tilde{\mathbf{A}}\right)}{\partial \tilde{A}_i} < 0 \qquad\qquad \text{for all } \tilde{\mathbf{A}} \in \Re_+^N.$$

Therefore we have established gross substitution. We now use these five properties to establish that the system of equations (B.66) has at most one (normalized) solution. Gross substitution implies that $D\left(\tilde{\mathbf{A}}\right) = D\left(\tilde{\mathbf{A}}'\right)$ cannot occur whenever $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{A}}'$ are two technology vectors that are not colinear. By homogeneity of degree zero, we can assume $\tilde{\mathbf{A}}' \geq \tilde{\mathbf{A}}$ and $\tilde{A}_i = \tilde{A}'_i$ for some i. Now consider altering the productivity vector $\tilde{\mathbf{A}}'$ to obtain the productivity vector $\tilde{\mathbf{A}}$ in $N-1$ steps, lowering (or keeping unaltered) the productivity of all the other $N-1$ locations $n \neq i$ one at a time. By gross substitution, the excess demand in location i cannot decrease in any step, and because $\tilde{\mathbf{A}} \neq \tilde{\mathbf{A}}'$, it will actually increase in at least one step. Hence $D\left(\tilde{\mathbf{A}}\right) > D\left(\tilde{\mathbf{A}}'\right)$ and we have a contradiction.

We next establish that there exists a productivity vector $\tilde{\mathbf{A}}^* \in \Re_+^N$ such that $D(\tilde{\mathbf{A}}^*) = 0$. By homogeneity of degree zero, we can restrict our search for this productivity vector to the unit simplex $\Delta = \left\{\tilde{\mathbf{A}} \in \Re_+^N : \sum_{i\in N} \tilde{A}_i = 1\right\}$. Define on Δ the function $D^+(\cdot)$ by $D_i^+\left(\tilde{\mathbf{A}}\right) = \max\left\{D_i\left(\tilde{\mathbf{A}}\right), 0\right\}$. Note that $D^+(\cdot)$ is continuous. Denote $\alpha\left(\tilde{\mathbf{A}}\right) = \sum_{i\in N}\left[\tilde{A}_i + D_i^+\left(\tilde{A}_i\right)\right]$. We have $\alpha\left(\tilde{\mathbf{A}}\right) \geq 1$ for all $\tilde{\mathbf{A}}$.

Define a continuous function $f(\cdot)$ from the closed convex set Δ into itself by:

$$f\left(\tilde{\mathbf{A}}\right) = \left[1/\alpha\left(\tilde{\mathbf{A}}\right)\right]\left[\tilde{\mathbf{A}} + D^+\left(\tilde{\mathbf{A}}\right)\right].$$

Note that this fixed-point function tends to increase the productivities of locations with excess demand. By Brouwer's Fixed-point Theorem, there exists $\tilde{\mathbf{A}}^* \in \Delta$ such that $\tilde{\mathbf{A}}^* = f\left(\tilde{\mathbf{A}}^*\right)$.

Since $\sum_{i \in N} D_i\left(\tilde{\mathbf{A}}\right) = 0$, it cannot be the case that $D_i\left(\tilde{\mathbf{A}}\right) > 0$ for all $i \in N$ or $D_i\left(\tilde{\mathbf{A}}\right) < 0$ for all $i \in N$. Additionally, if $D_i\left(\tilde{\mathbf{A}}\right) > 0$ for some i and $D_r\left(\tilde{\mathbf{A}}\right) < 0$ for some $r \neq i$, $\tilde{\mathbf{A}} \neq f\left(\tilde{\mathbf{A}}\right)$. It follows that at the fixed point for productivity, $\tilde{\mathbf{A}}^* = f\left(\tilde{\mathbf{A}}^*\right)$, and $D_i\left(\tilde{\mathbf{A}}^*\right) = 0$ for all i. It follows that there exists a unique vector of unobserved productivities ($\tilde{\mathbf{A}}$) that solves the excess demand system (B.66).

B.6 Proof of Proposition 4

Note that the commuting probability (21) can be written as the following excess demand system:

$$D_i(\mathbf{B}) = \lambda_{ni} - \frac{B_{ni}\kappa_{ni}^{-\epsilon}\left(\frac{L_n}{\pi_{nn}}\right)^{-\frac{\alpha\epsilon}{\sigma-1}} A_n^{\alpha\epsilon} w_n^{-\alpha\epsilon} \bar{v}_n^{-\epsilon(1-\alpha)}\left(\frac{R_n}{H_n}\right)^{-\epsilon(1-\alpha)} w_i^{\epsilon}}{\sum_{r\in N}\sum_{s\in N} B_{rs}\kappa_{rs}^{-\epsilon}\left(\frac{L_r}{\pi_{rr}}\right)^{-\frac{\alpha\epsilon}{\sigma-1}} A_r^{\alpha\epsilon} w_r^{-\alpha\epsilon}\bar{v}_r^{-\epsilon(1-\alpha)}\left(\frac{R_r}{H_r}\right)^{-\epsilon(1-\alpha)} w_s^{\epsilon}} = 0, \qquad \text{(B.67)}$$

where $\{w_i,\ L_i,\ \bar{v}_n,\ R_n,\ \kappa_{ni},\ \pi_{nn},\ A_n,\ H_n\}$ have already been determined from the observed data or our parameterization of commuting costs. Note that the excess demand system (B.67) exhibits the same properties in $\mathbf{B}$ as the excess demand system (B.66) exhibits in $\tilde{\mathbf{A}}$. It follows that there exists a unique vector of unobserved productivities ($\tilde{\mathbf{A}}$) that solves the excess demand system (B.66).

B.7 Partial Equilibrium Elasticities

We now derive the partial equilibrium elasticities of the endogenous variables of the model with respect to a productivity shock.

Wage Elasticity: Totally differentiating the goods market clearing condition (9), we have:

$$\begin{aligned}
\frac{dw_i}{w_i}w_iL_i + \frac{dL_i}{L_i}w_iL_i = {} & \sum_{r\in N}(1-\pi_{ri})\,\pi_{ri}\bar{v}_rR_r\frac{dL_i}{L_i} - \sum_{r\in N}\sum_{s\in N}\pi_{rs}\pi_{ri}\bar{v}_rR_r\frac{dL_s}{L_s} \\
& - (\sigma-1)\sum_{r\in N}(1-\pi_{ri})\,\pi_{ri}\bar{v}_rR_r\frac{dw_n}{w_n} + (\sigma-1)\sum_{r\in N}\sum_{s\in N}\pi_{rs}\pi_{rn}\bar{v}_rR_r\frac{dw_s}{w_s} \\
& + (\sigma-1)\sum_{r\in N}(1-\pi_{ri})\,\pi_{ri}\bar{v}_rR_r\frac{dA_i}{A_i} - (\sigma-1)\sum_{r\in N}\sum_{s\in N}\pi_{rs}\pi_{ri}\bar{v}_rR_r\frac{dA_s}{A_s} \\
& + \sum_{r\in N}\pi_{ri}\bar{v}_rR_r\frac{d\bar{v}_r}{\bar{v}_r} + \sum_{r\in N}\pi_{ri}\bar{v}_rR_r\frac{dR_r}{R_r}.
\end{aligned}$$

To consider the direct effect of a productivity shock in location i on wages, employment and residents in that location, holding constant all other endogenous variables at their values in the initial equilibrium, we set $dA_s = dw_s = dL_s = dR_s = 0$ for $s \neq i$ and $d\bar{v}_r = 0$ for all r, which yields:

$$\begin{aligned}
\frac{dw_i}{w_i}w_iL_i + \frac{dL_i}{L_i}w_iL_i = {} & \sum_{r\in N}(1-\pi_{ri})\,\pi_{ri}\bar{v}_rR_r\frac{dL_i}{L_i} - (\sigma-1)\sum_{r\in N}(1-\pi_{ri})\,\pi_{ri}\bar{v}_rR_r\frac{dw_i}{w_i} \\
& + (\sigma-1)\sum_{r\in N}(1-\pi_{ri})\,\pi_{ri}\bar{v}_rR_r\frac{dA_n}{A_n} + \pi_{ii}\bar{v}_iR_i\frac{dR_i}{R_i}.
\end{aligned}$$

This implies:

$$\begin{aligned}
\frac{dw_i}{dA_i}\frac{A_i}{w_i} + \frac{dL_i}{dA_i}\frac{A_i}{L_i} = {} & \sum_{r\in N}(1-\pi_{ri})\,\frac{\pi_{ri}\bar{v}_rR_r}{w_iL_i}\left(\frac{dL_i}{dA_i}\frac{A_i}{L_i}\right) - (\sigma-1)\sum_{r\in N}(1-\pi_{ri})\,\frac{\pi_{ri}\bar{v}_rR_r}{w_iL_i}\left(\frac{dw_i}{dA_i}\frac{A_i}{w_i}\right) \\
& + (\sigma-1)\sum_{r\in N}(1-\pi_{ri})\,\frac{\pi_{ri}\bar{v}_rR_r}{w_iL_i} + \frac{\pi_{ii}\bar{v}_iR_i}{w_iL_i}\frac{dR_i}{dA_i}\frac{A_i}{R_i},
\end{aligned}$$

which can be re-written as:

$$\begin{aligned}\tfrac{dw_i}{dA_i}\tfrac{A_i}{w_i}+\left(\tfrac{dL_i}{dw_i}\tfrac{w_i}{L_i}\right)\left(\tfrac{dw_i}{dA_i}\tfrac{A_i}{w_i}\right)&=\textstyle\sum_{r\in N}\left(1-\pi_{ri}\right)\tfrac{\pi_{ri}\bar{v}_rR_r}{w_iL_i}\left(\tfrac{dL_i}{dw_i}\tfrac{w_i}{L_i}\right)\left(\tfrac{dw_i}{dA_i}\tfrac{A_i}{w_i}\right)\\ -\left(\sigma-1\right)\textstyle\sum_{r\in N}\left(1-\pi_{ri}\right)\tfrac{\pi_{ri}\bar{v}_rR_r}{w_iL_i}\left(\tfrac{dw_i}{dA_i}\tfrac{A_i}{w_i}\right)&+\left(\sigma-1\right)\textstyle\sum_{r\in N}\left(1-\pi_{ri}\right)\tfrac{\pi_{ri}\bar{v}_rR_r}{w_iL_i}\\ &+\tfrac{\pi_{ii}\bar{v}_iR_i}{w_iL_i}\left(\tfrac{dR_i}{dw_i}\tfrac{w_i}{R_i}\right)\left(\tfrac{dw_i}{dA_i}\tfrac{A_i}{w_i}\right),\end{aligned}$$

where we have used the fact that productivity does not directly enter the commuting market clearing condition (16) and the residential choice probabilities (14) and hence employment and residents only change to the extent that wages change as a result of the productivity shock. Rearranging this expression, we obtain the partial equilibrium elasticity in the main text above:

$$\frac{\partial w_i}{\partial A_i}\frac{A_i}{w_i}=\frac{\left(\sigma-1\right)\sum_{r\in N}\left(1-\pi_{ri}\right)\xi_{ri}}{\left[1+\left(\sigma-1\right)\sum_{r\in N}\left(1-\pi_{ri}\right)\xi_{ri}\right]+\left[1-\sum_{r\in N}\left(1-\pi_{ri}\right)\xi_{ri}\right]\frac{dL_i}{dw_i}\frac{w_i}{L_i}-\xi_{ii}\frac{dR_i}{dw_i}\frac{w_i}{R_i}}, \tag{B.68}$$

where $\xi_{ri}=\pi_{ri}\bar{v}_rR_r/w_iL_i$ is the share of location i's revenue from market r and we use the partial derivative symbol to clarify that this derivative is not the full general equilibrium one.

Employment Elasticity: Totally differentiating the commuting market clearing condition (16), we have:

$$\begin{aligned}\frac{dL_i}{L_i} &= \epsilon\sum_{r\in N}\left(1-\lambda_{ri|r}\right)\frac{dw_i}{w_i}\frac{\lambda_{ri|r}R_r}{L_n}-\epsilon\sum_{r\in N}\sum_{s\neq n}\lambda_{rs|r}\frac{dw_s}{w_s}\frac{L_{ri}}{L_i}\\ &\quad+\sum_{r}\frac{dR_r}{R_r}\frac{L_{ri}}{L_i}.\end{aligned}$$

To consider the direct effect of a productivity shock in location i on its employment and residents through a higher wage in that location, holding constant all other endogenous variables at their values in the initial equilibrium, we set $dw_s=dL_s=dR_s=0$ for $s\neq i$, which yields:

$$\frac{dL_i}{L_i}=\epsilon\sum_{r\in N}\left(1-\lambda_{ri|r}\right)\frac{\lambda_{ri|r}R_r}{L_i}\frac{dw_i}{w_i}+\frac{\lambda_{ii|i}R_i}{L_i}\frac{dR_i}{R_i}.$$

Rearranging this expression, we obtain the partial equilibrium elasticity in the main text above:

$$\frac{\partial L_i}{\partial w_i}\frac{w_i}{L_i}=\epsilon\sum_{r\in N}\left(1-\lambda_{ri|r}\right)\vartheta_{ri}+\vartheta_{ii}\left(\frac{dR_i}{dw_i}\frac{w_i}{R_i}\right), \tag{B.69}$$

where $\vartheta_{ri}=\lambda_{ri|r}R_r/L_i$ is the share of commuters from source r in location i's employment and we use the partial derivative symbol to clarify that this derivative is not the full general equilibrium one.

Residents Elasticity: Totally differentiating the residential choice probability (λ_{Rn} in (14)), we have:

$$
\begin{aligned}
\frac{dR_i}{R_i}\frac{R_i}{\bar{L}} \quad = \quad & -\epsilon\alpha\left(1-\lambda_{Ri}\right)\lambda_{Ri}\frac{dP_i}{P_i}+\epsilon\alpha\sum_{r\neq i}\lambda_{Rr}\lambda_{Ri}\frac{dP_r}{P_r} \\
& -\epsilon\left(1-\alpha\right)\left(1-\lambda_{Ri}\right)\lambda_{Ri}\frac{dQ_i}{Q_i}+\epsilon\left(1-\alpha\right)\sum_{r\neq i}\lambda_{Rr}\lambda_{Ri}\frac{dQ_r}{Q_r} \\
& +\epsilon\lambda_{ii}\frac{dw_i}{w_i}-\epsilon\lambda_{Li}\lambda_{Ri}\frac{dw_i}{w_i}-\epsilon\sum_{s\neq i}\lambda_{Ls}\lambda_{Ri}\frac{dw_s}{w_s}.
\end{aligned}
$$

To consider the direct effect of a productivity shock in location i on its residents through a higher wage in that location, holding constant all other endogenous variables at their values in the initial equilibrium, we set $dP_r = dQ_r = 0$ for all r and $dw_s = 0$ for $s \neq i$, which yields:

$$\frac{dR_i}{R_i}\frac{R_i}{\bar{L}} = \epsilon\left(\lambda_{ii}-\lambda_i\lambda_{Ri}\right)\frac{dw_i}{w_i}.$$

This implies:

$$\frac{\partial R_i}{\partial w_i}\frac{w_i}{R_i} = \epsilon\left(\frac{\lambda_{ii}}{\lambda_{Ri}}-\lambda_i\right), \tag{B.70}$$

which corresponds to the expression in the main text above and we use the partial derivative symbol to clarify that this derivative is not the full general equilibrium elasticity. Using the residents elasticity (B.70) in the employment elasticity (B.69), and using the residents and employment elasticities ((B.70) and (B.69) respectively) in the wage elasticity (B.68), we obtain the following partial equilibrium elasticities for the productivity shock,

$$
\begin{aligned}
&\tfrac{\partial R_i}{\partial w_i}\tfrac{w_i}{R_i} = \epsilon\left(\tfrac{\lambda_{ii}}{\lambda_{Ri}}-\lambda_{Li}\right), \\
&\tfrac{\partial L_i}{\partial w_i}\tfrac{w_i}{L_i} = \epsilon\textstyle\sum_{r\in N}\left(1-\lambda_{ri|r}\right)\vartheta_{ri}+\vartheta_{ii}\epsilon\left(\tfrac{\lambda_{ii}}{\lambda_{Ri}}-\lambda_{Li}\right), \\
&\tfrac{\partial w_i}{\partial A_i}\tfrac{A_i}{w_i} = \frac{(\sigma-1)\sum_{r\in N}(1-\pi_{ri})\xi_{ri}}{\left[1+(\sigma-1)\sum_{r\in N}(1-\pi_{ri})\xi_{ri}\right]+\left[1-\sum_{r\in N}(1-\pi_{ri})\xi_{ri}\right]\left[\epsilon\sum_{r\in N}\left(1-\lambda_{ri|r}\right)\vartheta_{ri}+\epsilon\vartheta_{ii}\left(\frac{\lambda_{ii}}{\lambda_{Ri}}-\lambda_{Li}\right)\right]-\xi_{ii}\epsilon\left(\frac{\lambda_{ii}}{\lambda_{Ri}}-\lambda_i\right)}.
\end{aligned}
$$

B.8 Gravity and Local Employment Elasticities

We now show that the class of models consistent with a gravity equation for commuting flows implies heterogeneous local employment elasticities. Assume that commuting flows satisfy the following gravity equation:

$$L_{ni} = \mathcal{R}_n\mathcal{K}_{ni}\mathcal{W}_i, \tag{B.71}$$

where L_{ni} are commuting flows from residence n to workplace i; $\mathcal{R}_n$ is a residence fixed effect; $\mathcal{W}_i$ is a workplace fixed effect; and $\mathcal{K}_{ni}$ is a measure of the ease of commuting (an inverse measure of bilateral commuting costs). This gravity equation (B.71) implies that the unconditional probability that a worker commutes from residence n to workplace i is:

$$\lambda_{ni} = \frac{L_{ni}}{\sum_{r\in N}\sum_{s\in N} L_{rs}} = \frac{\mathcal{R}_n \mathcal{K}_{ni} \mathcal{W}_i}{\sum_{r\in N}\sum_{s\in N} \mathcal{R}_r \mathcal{K}_{rs} \mathcal{W}_s}. \tag{B.72}$$

The corresponding the probability of working in location i is:

$$\lambda_{Li} = \frac{\sum_{r\in N} L_{ri}}{\sum_{r\in N}\sum_{s\in N} L_{rs}} = \frac{\sum_{r\in N} \mathcal{R}_r \mathcal{K}_{ri} \mathcal{W}_i}{\sum_{r\in N}\sum_{s\in N} \mathcal{R}_r \mathcal{K}_{rs} \mathcal{W}_s}, \tag{B.73}$$

and the probability of residing in location n is:

$$\lambda_{Rn} = \frac{\sum_{s\in N} L_{ns}}{\sum_{r\in N}\sum_{s\in N} L_{rs}} = \frac{\sum_{s\in N} \mathcal{R}_n \mathcal{K}_{ns} \mathcal{W}_s}{\sum_{r\in N}\sum_{s\in N} \mathcal{R}_r \mathcal{K}_{rs} \mathcal{W}_s}, \tag{B.74}$$

From the residential probability (B.74), the probability of commuting from residence n to workplace i conditional on residing in n is:

$$\lambda_{ni|n} = \frac{\lambda_{ni}}{\lambda_{Rn}} = \frac{\mathcal{R}_n \mathcal{K}_{ni} \mathcal{W}_i}{\sum_{s\in N} \mathcal{R}_n \mathcal{K}_{ns} \mathcal{W}_s} = \frac{\mathcal{K}_{ni} \mathcal{W}_i}{\sum_{s\in N} \mathcal{K}_{ns} \mathcal{W}_s}. \tag{B.75}$$

Using this conditional probability (B.75), the commuting market clearing condition can be written as:

$$L_i = \sum_{n\in N} \lambda_{ni|n} R_n = \sum_{n\in N} \frac{\mathcal{K}_{ni} \mathcal{W}_i}{\sum_{s\in N} \mathcal{K}_{ns} \mathcal{W}_s} R_n. \tag{B.76}$$

Totally differentiating this commuting market clearing condition (B.76) for a given commuting technology $\mathcal{K}_{ni}$, we have:

$$\frac{dL_i}{L_i} = \sum_{r\in N} \left(1 - \lambda_{ri|r}\right) \frac{d\mathcal{W}_i}{\mathcal{W}_i} \frac{\lambda_{ri|r} R_r}{L_i} - \sum_{r\in N} \sum_{s\neq i} \lambda_{rs|r} \frac{d\mathcal{W}_s}{\mathcal{W}_s} \frac{\lambda_{ri|r} R_r}{L_i} \tag{B.77}$$

$$+ \sum_{r\in N} \frac{dR_r}{R_r} \frac{\lambda_{ri|r} R_r}{L_i}. \tag{B.78}$$

Now consider the direct effect of a shock to the workplace fixed effect for location i ($d\mathcal{W}_i \neq 0$) evaluated at the values of the variables for all other locations $r \neq i$ from the initial equilibrium ($d\mathcal{W}_r = dL_r = dR_r = 0$ for $r \neq i$):

$$\frac{dL_i}{L_i} = \sum_{r\in N} \left(1 - \lambda_{ri|r}\right) \frac{\lambda_{ri|r} R_r}{L_i} \frac{d\mathcal{W}_i}{\mathcal{W}_i} + \frac{\lambda_{ii|i} R_i}{L_i} \frac{dR_i}{R_i}. \tag{B.79}$$

Rearranging this expression, we obtain the following partial equilibrium local employment elasticity:

$$\frac{\partial L_i}{\partial \mathcal{W}_i} \frac{\mathcal{W}_i}{L_i} = \sum_{r\in N} \left(1 - \lambda_{ri|r}\right) \vartheta_{ri} + \vartheta_{ii} \left(\frac{\partial R_i}{\partial \mathcal{W}_i} \frac{\mathcal{W}_i}{R_i} \right), \tag{B.80}$$

where $\vartheta_{ri} = \lambda_{ri|r} R_r / L_i$ is the share of commuters from residence r in workplace i's employment and we use the partial derivative symbol to clarify that this derivative is not the full general equilibrium one.

This partial equilibrium local employment elasticity (B.80) takes the same form as in the main paper (and in the previous section of this web appendix above), where in our model the shock to the workplace fixed effect for location i ($\mathcal{W}_i$) corresponds to a shock to the wage at that workplace, which in turn depends on the shock to productivity at that workplace. Therefore our result of a variable local employment elasticity that depends on access to commuters in surrounding locations is a generic feature of the class of models that are consistent with a gravity equation for commuting flows. We show in the main paper that observed commuting flows are characterized by a strong gravity equation relationship.

B.9 Commuting with Multiple Worker Types

In this section of the web appendix, we consider a generalization of our model to allow for multiple worker types, which differ in their valuation of amenities and the variance of their idiosyncratic preferences. These differences in variance in turn imply that the multiple types differ in the responsiveness of their migration and commuting decisions to economic characteristics of locations (such as wages). This extension of our Fréchet model to incorporate multiple worker types is analogous to the extension of the logit model to multiple types in the mixed logit model (see for example McFadden and Train 2000), which is in turn closely related to the random coefficients model of Berry, Levinsohn and Pakes (1995). We show that our prediction of heterogeneous local employment elasticities across locations is robust to this extension and that there is now an additional source of heterogeneity relative to our baseline specification.

In particular, suppose that there are multiple types of workers (e.g. skilled versus unskilled) indexed by $z = 1, \ldots, Z$. There is a separate labor market and a separate wage for each type of worker z in each workplace i (w_i^z). Workers of a given type have idiosyncratic preferences over workplace and residence locations. However, the distributions of these idiosyncratic preferences differ across types, in terms of both their average preferences for the amenities for each bilateral commute (as determined by B_{ni}^z) and the variance of their idiosyncratic preferences across these bilateral commutes (as determined by ϵ^z):

$$G_{ni}^z(b) = e^{-B_{ni}^z b^{-\epsilon^z}}. \tag{B.81}$$

B.9.1 Commuting Decisions for Each Worker Type

Under these assumptions, commuting decisions for each worker type are characterized by a gravity equation, which is analogous to that in our baseline specification with a single worker type. The probability that workers of type z choose to work in location i conditional on living in location n is:

$$\pi_{ni|n}^z = \frac{B_{ni}^z \left(w_i^z / \kappa_{ni}^z\right)^{\epsilon^z}}{\sum_{s \in N} B_{ns}^z \left(w_s^z / \kappa_{ns}^z\right)^{\epsilon^z}}. \tag{B.82}$$

The corresponding commuting market clearing condition for workers of type z is:

$$L_i^z = \sum_{r \in N} \frac{B_{ri}^z \left(w_i^z / \kappa_{ri}^z\right)^{\epsilon^z}}{\sum_{s \in N} B_{rs}^z \left(w_s^z / \kappa_{rs}^z\right)^{\epsilon^z}} R_r^z, \tag{B.83}$$

which yields a partial elasticity of employment for workers of type z with respect to their wage that takes a similar form as for our baseline specification with a single worker type:

$$\frac{\partial L_i^z}{\partial w_i^z} \frac{w_i^z}{L_i^z} = \epsilon^z \sum_{r \in N} \left(1 - \lambda_{ri|r}^z\right) \vartheta_{ri}^z + \vartheta_{ii}^z \left(\frac{\partial R_i^z}{\partial w_i^z} \frac{w_i^z}{R_i^z}\right), \tag{B.84}$$

where $\vartheta_{ri}^z = \lambda_{ri|r}^z R_r^z / L_i^z$ is the share of commuters from residence r in workplace i's employment.

B.9.2 Aggregate Commuting Decisions

Aggregating commuting decisions across worker types, the total number of workers that choose to work in location i is:

$$L_i = \sum_{z=1}^{Z} L_i^z. \tag{B.85}$$

Now consider the elasticity of total employment in location i (L_i) with respect to a common increase in the wages of all worker types in that location:

$$dw_i^z = dw_i^k = dw_i > 0, \qquad \forall z, k. \tag{B.86}$$

Differentiating with respect to wages in equation (B.85), we have:

$$dL_i = \sum_{z=1}^{Z} \frac{\partial L_i^z}{\partial w_i^z} dw_i^z, \tag{B.87}$$

which using our assumption (B.86) can be re-written as:

$$\frac{dL_i}{dw_i} = \sum_{z=1}^{Z} \frac{\partial L_i^z}{\partial w_i^z}, \tag{B.88}$$

which can be further re-written as:

$$\frac{dL_i}{dw_i} \frac{w_i}{L_i} = \sum_{z=1}^{Z} \left(\frac{\partial L_i^z}{\partial w_i^z} \frac{w_i^z}{L_i^z}\right) \left(\frac{L_i^z / L_i}{w_i^z / w_i}\right). \tag{B.89}$$

Combining (B.84) and (B.89), the local employment elasticity for each location is a weighted average of the local employment elasticities for each worker type for that location, where the weights depend on employment shares and relative wages. Therefore, local employment elasticities continue to be heterogeneous across locations in this extension of the model to incorporate multiple worker types, but there are is now an additional source of heterogeneity relative to our baseline specification. First, the local employment elasticity for a given worker type is heterogeneous across locations depending on commuting networks for that worker type (equation (B.84)). This first source of heterogeneity is the analogous to that in our baseline specification with a single worker type. Second, the composition of worker types and their relative wages can differ across locations, which provides an additional source of heterogeneity in local employment elasticities that is not present in our baseline specification (as in equation (B.89)). Taken together, this extension further reinforces our point that the local employment elasticity is not a structural parameter.

B.10 Congestion in Commuting

In this section of the web appendix, we generalize our baseline specification to allow for congestion in commuting. Assuming that congestion costs are a power function of the volume of commuters, we show that congestion affects the interpretation of the estimated parameters in our commuting gravity equation, but leaves the model's prediction of heterogeneous employment elasticities across locations unchanged. In particular, we assume that each worker draws idiosyncratic preferences for each pair of residence location n and workplace location i from the following distribution:

$$G_{ni}\left(b\right) = e^{-B_{ni}L_{ni}^{\chi}b^{-\epsilon}}, \tag{B.90}$$

where the scale parameter of this distribution ($B_{ni}L_{ni}^{\chi}$) is a power function of the volume of commuters. Our baseline specification corresponds to the special case in which $\chi = 0$; $\chi < 0$ corresponds to congestion in commuting decisions, such that the attractiveness of commuting from residence n to workplace i depends negatively on the volume of commuters. Under these assumptions, the probability that a worker commutes from residence n to workplace i is:

$$\pi_{ni} = \frac{L_{ni}}{\bar{L}} = \frac{B_{ni}L_{ni}^{\chi}\left(\kappa_{ni}P_n^{\alpha}Q_n^{1-\alpha}\right)^{-\epsilon}w_i^{\epsilon}}{\sum_{r\in N}\sum_{s\in N}B_{rs}L_{rs}^{\chi}\left(\kappa_{rs}P_r^{\alpha}Q_r^{1-\alpha}\right)^{-\epsilon}w_s^{\epsilon}}, \tag{B.91}$$

and expected utility conditional on choosing a given bilateral commute (which is the across all bilateral commutes) is equal to:

$$\bar{U} = \mathbb{E}\left[U_{ni\omega}\right] = \Gamma\left(\frac{\epsilon-1}{\epsilon}\right)\left[\sum_{r\in N}\sum_{s\in N}B_{rs}L_{rs}^{\chi}\left(\kappa_{rs}P_r^{\alpha}Q_r^{1-\alpha}\right)^{-\epsilon}w_s^{\epsilon}\right]^{\frac{1}{\epsilon}} \text{ all } n, i \in N. \tag{B.92}$$

Combining (B.91) and (B.92), the flow of workers that choose to commute from residence n to workplace i can be written as:

$$L_{ni} = \left(\frac{\bar{U}}{\Gamma}\right)^{-\epsilon}B_{ni}L_{ni}^{\chi}\left(\kappa_{ni}P_n^{\alpha}Q_n^{1-\alpha}\right)^{-\epsilon}w_i^{\epsilon}\bar{L}, \tag{B.93}$$

which can be in turn re-written as:

$$L_{ni} = \left(\frac{\bar{U}}{\Gamma}\right)^{-\frac{\epsilon}{1-\chi}} B_{ni}^{\frac{1}{1-\chi}} \left(\kappa_{ni} P_n^{\alpha} Q_n^{1-\alpha}\right)^{-\frac{\epsilon}{1-\chi}} w_i^{\frac{\epsilon}{1-\chi}} \bar{L}^{\frac{1}{1-\chi}}. \tag{B.94}$$

Therefore the probability that a worker commutes from residence n to workplace i can be equivalently expressed as:

$$\pi_{ni} = \frac{L_{ni}}{\sum_{r\in N}\sum_{s\in N} L_{rs}} = \frac{L_{ni}}{\bar{L}} = \frac{B_{ni}^{\frac{1}{1-\chi}} \left(\kappa_{ni} P_n^{\alpha} Q_n^{1-\alpha}\right)^{-\frac{\epsilon}{1-\chi}} w_i^{\frac{\epsilon}{1-\chi}}}{\sum_{r\in N}\sum_{s\in N} B_{rs}^{\frac{1}{1-\chi}} \left(\kappa_{rs} P_r^{\alpha} Q_r^{1-\alpha}\right)^{-\frac{\epsilon}{1-\chi}} w_s^{\frac{\epsilon}{1-\chi}}}, \tag{B.95}$$

which takes exactly the same form as in our baseline specification, except that the exponent on wages, which we interpret as ϵ in our baseline specification, should be interpreted as $\epsilon/(1-\chi)$ in this extended specification. Similarly, the values of bilateral amenities implied by this commuting probability, which we interpret as B_{ni} in our baseline specification, should be interpreted as $B_{ni}^{1/(1-\chi)}$ in this extended specification.

Using the unconditional commuting probabilities (B.95), we can also solve for the probability of commuting to workplace i conditional on living in residence n:

$$\pi_{ni|n} = \frac{B_{ni}^{\frac{1}{1-\chi}} \left(w_i/\kappa_{ni}\right)^{\frac{\epsilon}{1-\chi}}}{\sum_{s\in N} B_{ns}^{\frac{1}{1-\chi}} \left(w_s/\kappa_{ns}\right)^{\frac{\epsilon}{1-\chi}}}. \tag{B.96}$$

The corresponding commuting market clearing condition is:

$$L_i = \sum_{r\in N} \frac{B_{ri}^{\frac{1}{1-\chi}} \left(w_i/\kappa_{ri}\right)^{\frac{\epsilon}{1-\chi}}}{\sum_{s\in N} B_{rs}^{\frac{1}{1-\chi}} \left(w_s/\kappa_{rs}\right)^{\frac{\epsilon}{1-\chi}}} R_r, \tag{B.97}$$

which yields a partial elasticity of employment with respect to the wage that takes a similar form as for our baseline specification:

$$\frac{\partial L_i}{\partial w_i}\frac{w_i}{L_i} = \frac{\epsilon}{1-\chi}\sum_{r\in N}\left(1-\lambda_{ri|r}\right)\vartheta_{ri} + \vartheta_{ii}\left(\frac{\partial R_i}{\partial w_i}\frac{w_i}{R_i}\right), \tag{B.98}$$

where $\vartheta_{ri} = \lambda_{ri|r} R_r / L_i$ is the share of commuters from residence r in workplace i's employment. In this extended specification (B.98), the estimated coefficient on the first term on the right-hand side is again the exponent on wages from the gravity equation for commuting (B.95), but this estimated coefficient is now interpreted as $\epsilon/(1-\chi)$ rather than as ϵ.

Therefore, taking the results of this section together, the introduction of congestion in commuting affects the interpretation of the estimated parameters in our gravity equation for commuting, but leaves the model's prediction of heterogeneous elasticities of employment with respect to wages across locations unchanged.

B.11 Non-traded Goods

In the baseline version of the model in the paper, we consider a canonical new economic geography model with a single tradable consumption goods sector and land as the only non-traded good. A key focus of our analysis is the implications of introducing commuting into this standard framework for the elasticity of local employment with respect to local shocks. In this section of the web appendix, we generalize our analysis to incorporate non-traded consumption goods. We show that the commuting market clearing condition and the local elasticity of employment with respect to wages take the same form as in our baseline specification without non-traded goods.

The consumption index for worker ω residing at location n and working at location i is now assumed to take the following form:

$$U_{ni\omega} = \frac{b_{ni\omega}}{\kappa_{ni}} \left(\frac{C_{Nn\omega}}{\alpha_N} \right)^{\alpha_N} \left(\frac{C_{Tn\omega}}{\alpha_T} \right)^{\alpha_T} \left(\frac{H_{n\omega}}{1 - \alpha_N - \alpha_T} \right)^{1-\alpha_N-\alpha_T}, \tag{B.99}$$

$$\alpha_N, \alpha_T > 0, \qquad 0 < \alpha_N + \alpha_T < 1,$$

where $C_{Nn\omega}$ is consumption of the non-traded good; $C_{Tn\omega}$ is consumption of the traded good; and all other terms are defined in the same way as in our baseline specification. Total expenditure on consumption goods (traded plus non-traded) equals the fraction $\alpha^N + \alpha^T$ of the total income of residents plus the entire income of landlords (which equals the fraction $1 - \alpha^N - \alpha^T$) of the total income of residents):

$$P_n C_n = (\alpha_N + \alpha_T)\bar{v}_n R_n + (1 - \alpha_N - \alpha_T)\bar{v}_n R_n = \bar{v}_n R_n. \tag{B.100}$$

Utility maximization implies that a constant fraction $\alpha_N/(\alpha_N + \alpha_T)$ of total expenditure on consumption goods is allocated to the non-traded sector:

$$P_{Nn} C_{Nn} = \frac{\alpha_N}{\alpha_N + \alpha_T} P_n C_n = \frac{\alpha_N}{\alpha_N + \alpha_T} \bar{v}_n R_n, \tag{B.101}$$

and the remaining fraction is allocated to the traded sector:

$$P_{Tn} C_{Tn} = \frac{\alpha_T}{\alpha_N + \alpha_T} P_n C_n = \frac{\alpha_T}{\alpha_N + \alpha_T} \bar{v}_n R_n, \tag{B.102}$$

The non-traded good is assumed to be produced under conditions of perfect competition and according to a constant returns to scale technology with a unit labor requirement:

$$Y_{Nn} = L_{Nn}, \tag{B.103}$$

where Y_{Nn} is output of the non-traded good in location n and L_{Nn} is employment in the non-traded sector in that location. Perfect competition and constant returns to scale imply that the price of the non-traded good is equal to the wage:

$$P_{Nn} = w_n. \tag{B.104}$$

Combining this result with utility maximization (B.101), and using goods market clearing for the non-traded good ($C_{Nn} = Y_{Nn}$) and the production technology (B.103), we find that the wage bill in the non-traded sector is a constant share of residential income:

$$w_n L_{Nn} = \frac{\alpha_N}{\alpha_N + \alpha_T} \bar{v}_n R_n. \tag{B.105}$$

Using utility maximization and goods market clearing for tradeables, the wage bill in the traded sector is fraction of residential income across all locations:

$$w_n L_{Tn} = \frac{\alpha_T}{\alpha_N + \alpha_T} \sum_{r \in N} \pi_{rn} \bar{v}_r R_r. \tag{B.106}$$

Total employment equals the sum of employment in the non-traded and traded sectors:

$$L_n = L_{Tn} + L_{Nn} = \frac{\alpha_N}{\alpha_N + \alpha_T} \frac{\bar{v}_n R_n}{w_n} + \frac{\alpha_T}{\alpha_N + \alpha_T} \sum_{r \in N} \frac{\pi_{rn} \bar{v}_r R_r}{w_n}. \tag{B.107}$$

The commuting market clearing condition requires that total employment in each location equals the measure of workers that choose to commute to that location and takes the same form as in our baseline specification without the non-traded sector:

$$L_n = \sum_{r \in N} \frac{B_{rn} \left(w_n / \kappa_{rn}\right)^{\epsilon}}{\sum_{s \in N} B_{rs} \left(w_s / \kappa_{rs}\right)^{\epsilon}} R_r. \tag{B.108}$$

Given the same commuting market clearing condition, the partial elasticity of employment with respect to the wage that takes the same form as for our baseline specification:

$$\frac{\partial L_n}{\partial w_n} \frac{w_n}{L_n} = \epsilon \sum_{r \in N} \left(1 - \lambda_{rn|r}\right) \vartheta_{rn} + \vartheta_{nn} \left(\frac{\partial R_n}{\partial w_n} \frac{w_n}{R_n}\right), \tag{B.109}$$

where $\vartheta_{rn} = \lambda_{rn|r} R_r / L_n$ is the share of commuters from residence r in workplace n's employment.

Intuitively, when deciding where to work, workers care about the wage, and not whether this wage is paid in the traded or non-traded sector. Therefore, the gravity equation for commuting takes the same form as in our baseline specification without the non-traded sector, and hence the elasticity of local employment with respect to wages takes the same form as in our baseline specification without the non-traded sector. Depending on whether productivity shocks affect both the non-traded and traded sectors, the presence of non-traded goods can affect the elasticity of wages with respect to productivity, but it leaves unchanged the model's prediction of heterogeneous local employment elasticities with respect to wages.

B.12 Landlords Consume Residential Land

In this subsection of the web appendix, we show that allowing landlords to consume residential land as well as consumption goods is straightforward, and merely results in less clean expressions. Under this alternative assumption, consumption goods expenditure that was previously given by equation (4) is now instead given by:

$$P_n C_n = \alpha \left[1 + (1-\alpha)\right] \bar{v}_n R_n. \tag{B.110}$$

Using this relationship, the equality between income and expenditure that was previously given by equation (9) is now instead given by:

$$w_i L_i = \alpha \left[1 + (1-\alpha)\right] \sum_{n \in N} \pi_{ni} \bar{v}_n R_n, \tag{B.111}$$

and the land market clearing condition that was previously given by equation (5) is now instead given by:

$$Q_n = (1-\alpha) \left[1 + (1-\alpha)\right] \frac{\bar{v}_n R_n}{H_n}. \tag{B.112}$$

As in our baseline specification in which landlords consume only consumption goods, the general equilibrium of the model can be referenced by the following vector of six variables $\{w_n, \bar{v}_n, Q_n, L_n, R_n, P_n\}_{n=1}^N$ and a scalar $\bar{U}$. Given this equilibrium vector and scalar, all other endogenous variables of the model can be determined. This equilibrium vector solves the following six sets of equations: income equals expenditure (B.111), land market clearing (B.112), expected labor income (which remains as in equation (17)), workplace choice probabilities (which continue to equal (14) for L_n), residence choice probabilities (which are still equal (14) for R_n), price indices (again equal to (10), and the labor market clearing continue (which remains the same as $\bar{L} = \sum_{n \in N} R_n = \sum_{n \in N} L_n$). This system of equations for general equilibrium is exactly the same as in our baseline specification in which landlords consume only consumption goods, except for the terms in α that appear in equations (B.111) and (B.112). Therefore the properties of this version of the model in which landlords consume residential land as well as consumption goods are similar to those in our baseline specification. In particular, the model continues to predict heterogeneous local employment elasticities across locations.

B.13 Alternative Production Technology

In this subsection of the web appendix, we show how the production technology can be generalized to introduce intermediate inputs, commercial land use and physical capital. We show that the model continues to imply a gravity equation for commuting flows and hence continues to predict heterogeneous employment elasticities. Our baseline specification in the paper assumes the following production technology:

$$\Lambda_i = l_i(j) w_i = \left(F + \frac{x_i(j)}{A_i} \right) w_i. \tag{B.113}$$

We now consider a generalization of this production technology, in which total costs are a Cobb-Douglas function of labor (with wage w_i), intermediate inputs (with price P_i), commercial land (with rental rate Q_i) and physical capital (with rental rate $\mathbb{R}$). We follow Krugman and Venables (1995) and Eaton and Kortum (2002) in assuming that intermediate inputs enter the total cost function through the same CES aggregator as for final consumption. Perfect capital mobility ensures that the rental rate for physical capital is the same for all locations ($\mathbb{R}_i = \mathbb{R}$ for all i). Therefore the total cost function now becomes:

$$\Lambda_i = \left(F + \frac{x_i(j)}{A_i}\right) w_i^{\beta_L} Q_i^{\beta_Q} \mathbb{R}^{\beta_R} P_i^{1-\beta_L-\beta_Q-\beta_R}. \tag{B.114}$$

The probability that a worker chooses to live in location n and work in location i remains the same as in equation (13), which in turn implies that the commuting market clearing condition takes exactly the same form as in our baseline specification:

$$L_n = \sum_{r \in N} \frac{B_{rn} \left(w_n / \kappa_{rn}\right)^{\epsilon}}{\sum_{s \in N} B_{rs} \left(w_s / \kappa_{rs}\right)^{\epsilon}} R_r. \tag{B.115}$$

Given the same commuting market clearing condition, the partial elasticity of employment with respect to the wage takes the same form as for our baseline specification:

$$\frac{\partial L_n}{\partial w_n} \frac{w_n}{L_n} = \epsilon \sum_{r \in N} \left(1 - \lambda_{rn|r}\right) \vartheta_{rn} + \vartheta_{nn} \left(\frac{\partial R_n}{\partial w_n} \frac{w_n}{R_n}\right), \tag{B.116}$$

where $\vartheta_{rn} = \lambda_{rn|r} R_r / L_n$ is the share of commuters from residence r in workplace n's employment.

Therefore, although incorporating additional factors of production affects the partial elasticity of wages with respect to productivity, it leaves the partial elasticity of employment with respect to wages in equation (B.116) unchanged. The reason is that the model's prediction of heterogeneous local employment elasticities with respect wages is a generic implication of a gravity equation for commuting.

B.14 Heterogeneity in Effective Units of labor

In this section of the web appendix, we consider an alternative specification of the model, in which b corresponds to an idiosyncratic draw to effective units of labor instead of amenities. One interpretation of this alternative specification is that commuting costs are incurred in terms of effective units of labor and individual workers find some commutes more costly than others. Under this alternative specification, the idiosyncratic draw no longer enters the direct utility function, which is now:

$$U_{ni\omega} = \frac{1}{\kappa_{ni}} \left(\frac{C_{n\omega}}{\alpha}\right)^{\alpha} \left(\frac{H_{n\omega}}{1-\alpha}\right)^{1-\alpha}. \tag{B.117}$$

However, the idiosyncratic draw continues to enter the indirect utility function in exactly the same form as in our baseline specification, because worker income now depends on the wage per effective unit of labor

(w_i) times the realization for effective units of labor ($b_{ni\omega}$):

$$U_{ni\omega} = \frac{b_{ni\omega} w_i}{\kappa_{ni} P_n^{\alpha} Q_n^{1-\alpha}}. \tag{B.118}$$

Therefore the probability that a worker chooses to live in location n and work in location i takes exactly the same form as in our baseline specification:

$$\lambda_{ni} = \frac{B_{ni} \left(\kappa_{ni} P_n^{\alpha} Q_n^{1-\alpha}\right)^{-\epsilon} w_i^{\epsilon}}{\sum_{r \in N} \sum_{s \in N} B_{rs} \left(\kappa_{rs} P_r^{\alpha} Q_r^{1-\alpha}\right)^{-\epsilon} w_s^{\epsilon}} \equiv \frac{\Phi_{ni}}{\Phi}. \tag{B.119}$$

The main difference between the amenities and effective units of labor specifications is the interpretation of wages in the data. In the amenities specification, the observed wage for each location in the data corresponds directly to the wage in the model, and worker mobility ensures that expected utility is equalized across all workplace-residence pairs (but real wages without taking into account amenities differ). In contrast, in the effective units of labor specification, the observed wage for each location in the data corresponds to the wage per effective unit of labor times average effective units of labor in the model, and worker mobility ensures that expected real earnings after taking into account average effective units of labor are equalized across locations.

C Additional Empirical Results

In this section of the web appendix, we report additional empirical results and robustness tests.

C.1 Commuting vs. Migration

To assess the importance of commuting for local employment responses in the data (and not just within our quantitative model) we compute the following decomposition. Our approach uses the commuting market clearing condition (which necessarily holds in the data) to undertake decompositions for changes in employment over time. The commuting market clearing condition is given by

$$L_{Mi} = \sum_{n \in N} \lambda_{ni|n} L_{Rn},$$

which can be written as,

$$L_{Mi} = \underbrace{\lambda_{ii|i} L_{Ri}}_{\text{own residents}} + \underbrace{\sum_{n \neq i} \lambda_{ni|n} L_{Rn}}_{\text{commuters}}.$$

Taking differences, we obtain,

$$\Delta L_{Mi} = \Delta \left(\lambda_{ii|i} L_{Ri}\right) + \sum_{n \neq i} \Delta \left(\lambda_{ni|n} L_{Rn}\right).$$

Therefore one measure of the importance of commuting flows for employment changes in each county i is the following index,

$$\frac{\text{abs}\left[\sum_{n\neq i}\Delta\left(\lambda_{ni|n}L_{Rn}\right)\right]}{\text{abs}\left[\Delta\left(\lambda_{ii|i}L_{Ri}\right)\right]+\text{abs}\left[\sum_{n\neq i}\Delta\left(\lambda_{ni|n}L_{Rn}\right)\right]}.$$

Note that this measure is bounded between zero and one. When all changes in employment occur through commuting, this measure takes the value of one. When all changes in employment are achieved through migration, this measure takes the value of zero. If the change in residents, commuters, and total employment all have the same sign (which tends to be the case for counties that experience large increases or declines in employment), the index represent the fraction of the change in employment accounted for by changes in commuting patterns.

We perform the above calculation for observed changes between 1990 and 2007. The results show that for 34% of counties changes in employment due to commuting were larger than those due to migration. For the median county, 39% of the changes in employment were due to changes in commuting patterns. The mean index is 40%. For the county at the 75 (90) percentile the index is 58% (77%) and for the one at the 25% (10%) it is 21% (8%). This results provide direct evidence, independent of the model, that changes in commuting are an important margin through which changes in employment are achieved in the data.

C.2 Determinants of L/R

In this subsection of the web appendix, we show that commuting linkages are not only heterogeneous across counties, but are also hard to explain with standard empirical controls. In Table 5, we present the results of regressions that try to account for the cross-section of employment, residents, and the ratio of employment to residents (a measure of county commuting intensity). The first column shows that one can account for most of the variation in county employment using the number of residents and wages. Column (2) shows a similar result for the number of residents and Columns (3) and (4) show that the results are not affected when we add land area, developed-land supply elasticities, employment and wages in surrounding counties. So the first four columns show that it is relatively easy to explain the variation in employment and residents with standard variables. The following four columns demonstrate that this is not the case for commuting intensity. The level of residents, measures of wages, land area, developed-land supply elasticities and measures employment, residents and wages in surrounding counties do a poor job in accounting for the variation in commuting intensity. None of the R-squareds in the last four columns of Table 5 amounts to more than 0.3. Taken together, these results highlight the quantitative relevance of commuting as a source of spatial linkages between counties and CZ's within the United States, and the difficulty of explaining these links using standard empirical controls.

	1	2	3	4	5	6	7	8
Dependent Variable:	$\log L_i$	$\log R_i$	$\log L_i$	$\log R_i$	L_i/R_i	L_i/R_i	L_i/R_i	L_i/R_i
$\log R_i$	0.974**		1.001**		-0.000		0.020**	
	(0.003)		(0.004)		(0.003)		(0.004)	
$\log w_i$	0.460**		0.480**			0.341**		0.331**
	(0.018)		(0.018)			(0.018)		(0.018)
$\log L_i$		0.957**		0.922**		-0.001		0.028**
		(0.003)		(0.004)		(0.003)		(0.003)
$\log \bar{v}_i$		0.066**		0.019	0.171**		0.239**	
		(0.025)		(0.026)	(0.023)		(0.025)	
$\log H_i$			0.015*	0.037**			-0.011	-0.022**
			(0.006)	(0.006)			(0.006)	(0.006)
$\log R_{,-i}$			-0.020**				0.609**	0.389**
			(0.005)				(0.123)	(0.112)
$\log \bar{w}_{-i}$			-0.330**				0.247	0.070
			(0.032)				(0.240)	(0.218)
$\log L_{,-i}$				0.084**			-0.654**	-0.435**
				(0.005)			(0.122)	(0.111)
$\log \bar{v}_{-i}$				0.044			-0.347	-0.238
				(0.039)			(0.242)	(0.220)
Constant	-4.667**	-0.165	-1.485**	-1.199**	-0.881**	-2.647**	-0.057	-0.262
	(0.174)	(0.238)	(0.301)	(0.336)	(0.223)	(0.169)	(0.338)	(0.306)
R^2	0.98	0.98	0.99	0.98	0.03	0.16	0.15	0.30
N	3,111	3,111	3,081	3,081	3,111	3,111	3,081	3,081

In this table, $L_{,-i} \equiv \sum_{n:d_{ni} \leq 120, n \neq i} L_n$ is the total employment in i neighbors whose centroid is no more than 120km away; $\bar{w}_{-i} \equiv \sum_{n:d_{ni} \leq 120, n \neq i} \frac{L_n}{L_{,-i}} w_n$ is the weighted average of their workplace wage. Analogous definitions apply to $R_{,-i}$ and $\bar{v}_{-i}$.
* $p < 0.05$; ** $p < 0.01$.

Table 5: Explaining employment levels and commuting intensity

C.3 Land Prices

In this subsection of the web appendix, we show that the model's predictions for land prices are strongly positively correlated with observed median house prices. In our baseline specification, we assume Cobb-Douglas utility and interpret land area as geographical land area. In Figure 8, we show the predictions for land prices from this baseline specification against median house prices in the data. We find a strong and approximately log linear relationship, with a regression slope coefficient of 2.04 and R-squared of 0.26. Therefore, although our model is necessarily an abstraction, and there are a number of potential sources of differences between land prices in the model and house prices in the data, we find that the model has strong predictive power for the data. In section 4.3 of the paper, we generalize this baseline specification to allow for a positive supply elasticity for developed land that is heterogeneous across locations.

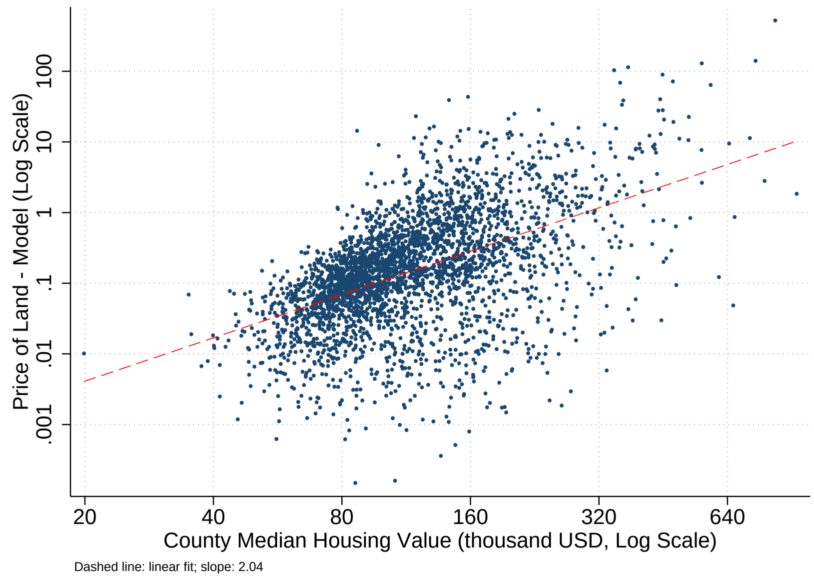


Figure 8: Land Prices in the Model and House Prices in the Data

C.4 Standardized Regression on Employment Elasticities

Table 6 presents the coefficient from the same set of regressions presented in Table 2 in the main text, after standardizing all variables to make their means zero and standard deviations one. Hence, all coefficients can be interpreted as the percentage change in the dependent variable related to changing the independent variable by one standard deviation.

C.5 Spatially Correlated Productivity Shocks

In Figures 9 to 12 we show that the heterogeneity in local employment elasticities persists if we simulate productivity shocks with a degree of spatial correlation based on observed industry composition. We construct spatially correlated shocks using aggregate productivity growth in manufacturing and non-manufacturing and the observed shares of these sectors within each county's employment. In particular, we proceed as follows. Data from BLS shows that between 2004 and 2010 TFP grew 6.2% for the manufacturing sector and 3.4% for the overall private business sector. Given a U.S. employment share in manufacturing of about 11% in 2007 (computed from the County Business Pattern, see Data Appendix below), we infer a growth in the non-manufacturing sector's TFP of 3.1%. We use the County Business Pattern 2007 to also compute the share of each county's manufacturing employment over total employment. Figure 9 shows a map of these shares across the United States.

We first show the consequences of a spatially correlated shock to manufacturing. We compute the equilibrium change in employment and residents in a single counterfactual exercise where each county's productivity is changed by 6.1% times the share of manufacturing employment in that county: hence, the spatial correlation in manufacturing shares induces a spatial correlation in productivity shocks. Figure 10 shows the resulting distribution of elasticities of employment and residents.

	1	2	3	4	5	6	7	8	9
Dependent Variable:				Elasticity of Employment					
$\log L_n$		-0.012	0.036*	-0.217**				0.147**	0.132**
		(0.018)	(0.018)	(0.019)				(0.008)	(0.008)
$\log w_n$			-0.126**	-0.100**				-0.162**	-0.166**
			(0.018)	(0.017)				(0.006)	(0.006)
$\log H_n$			-0.621**	-0.372**				0.007	0.020**
			(0.014)	(0.019)				(0.007)	(0.007)
$\log L_{,-n}$				0.429**				-0.097**	-0.097**
				(0.027)				(0.009)	(0.010)
$\log \bar{w}_{-n}$				0.090**				0.072**	0.091**
				(0.021)				(0.007)	(0.007)
$\lambda_{nn\|n}$					-0.945**				
					(0.006)				
$\sum_{r\in N}\left(1-\lambda_{rn\|r}\right)\vartheta_{rn}$						1.462**		1.343**	
						(0.048)		(0.051)	
$\vartheta_{nn}\left(\frac{\lambda_{nn}}{\lambda_{Rn}}-\lambda_{Ln}\right)$						0.487**		0.322**	
						(0.048)		(0.051)	
$\frac{\partial w_n}{\partial A_n}\frac{A_n}{w_n}$						-0.110**		-0.090**	
						(0.005)		(0.006)	
$\frac{\partial w_n}{\partial A_n}\frac{A_n}{w_n}\cdot\sum_{r\in N}\left(1-\lambda_{rn\|r}\right)\vartheta_{rn}$							0.544**		0.576**
							(0.019)		(0.025)
$\frac{\partial w_n}{\partial A_n}\frac{A_n}{w_n}\cdot\vartheta_{nn}\left(\frac{\lambda_{nn}}{\lambda_{Rn}}-\lambda_{Ln}\right)$							-0.428**		-0.444**
							(0.019)		(0.025)
constant	-0.000	-0.000	0.000	0.000	-0.000	0.000	-0.000	-0.006	-0.006
	(0.018)	(0.018)	(0.014)	(0.013)	(0.006)	(0.005)	(0.005)	(0.004)	(0.004)
R^2	0.00	0.00	0.40	0.51	0.89	0.93	0.93	0.95	0.95
N	3,111	3,111	3,111	3,081	3,111	3,111	3,111	3,081	3,081

In this table, $L_{,-n} \equiv \sum_{r:d_{rn}\leq 120, r\neq n} L_r$ is the total employment in n neighbors whose centroid is no more than 120km away; $\bar{w}_{-n} \equiv \sum_{r:d_{rn}\leq 120, r\neq n} \frac{L_r}{L_{,-n}} w_r$ is the weigthed average of their workplace wage. All variables are standardized. * $p < 0.05$; ** $p < 0.01$.

Table 6: Explaining the general equilibrium local employment elasticities to a 5 percent productivity shock

Figure 11 shows an analogous exercise for a shock to the non-manufacturing sector. Finally, Figure 12 shows the same elasticities when both sectors are shocked: in this case, each county's shock is a weighted average of the national increase in TFP in the manufacturing and non-manufacturing sectors, where the weights are the corresponding employment shares in the county.

C.6 Additional Results with Positive Developed Land Elasticities

In the main text of the paper we introduce the Saiz elasticities of developed land for all counties where they are available but keep the remaining elasticities at zero. Here we consider a second specification, in which we assume a land supply elasticity of zero for central counties within multi-county MSAs, the Saiz estimate for other counties in these MSA's as well as for single-county MSAs, and for all other counties

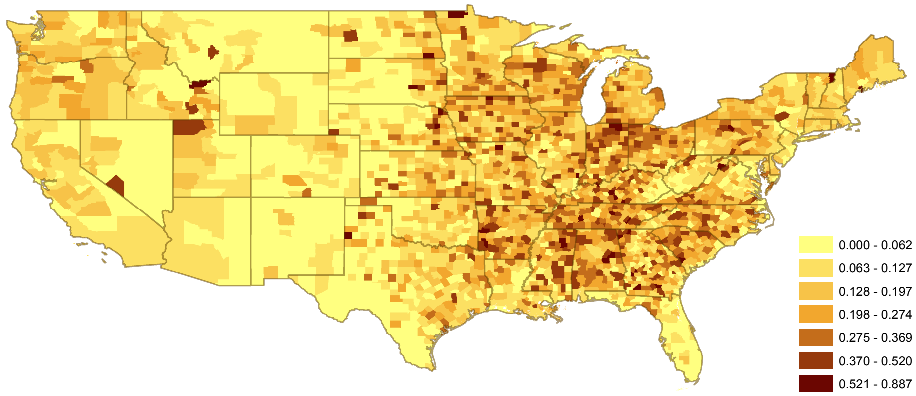


Figure 9: U.S. counties' share of employment in manufacturing, 2007.

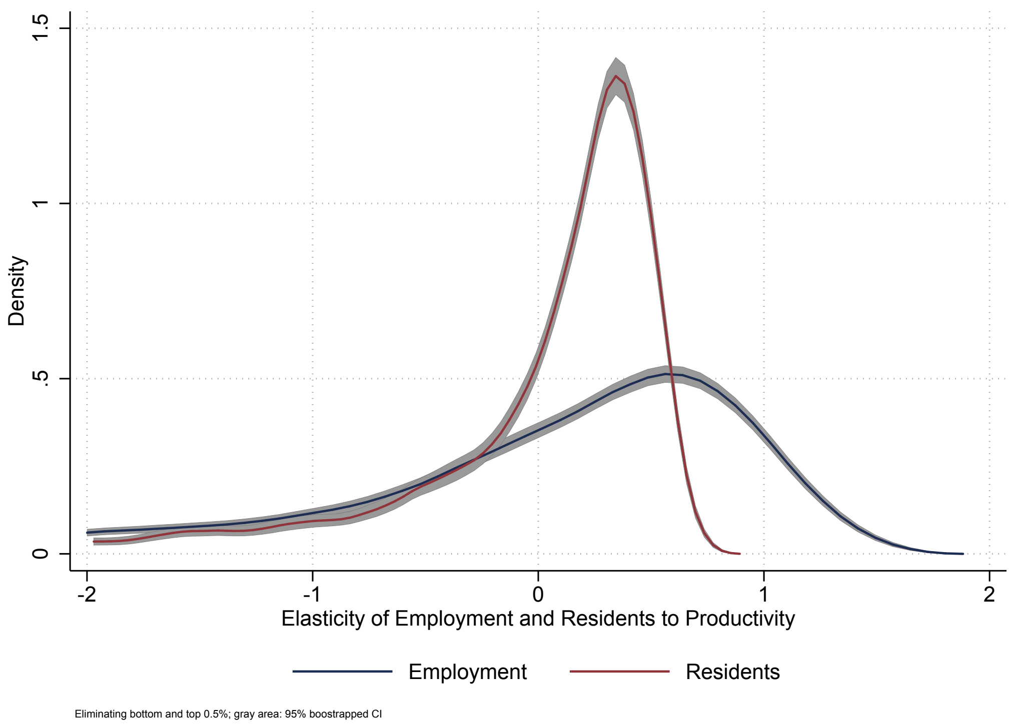


Figure 10: Kernel density for the distribution of employment and residents elasticities in response to a spatially correlated productivity shock in the manufacturing sector

we use the median Saiz estimate across MSAs of 1.67. In further robustness checks, we considered a range of alternative assumptions, and found the same pattern of results across these perturbations.

Figure 14 displays the results for this specification. Again we find a similar pattern of results as for our baseline specification and for the case in Section 4.3. First, both the employment and residents elasticities increase on average relative to the case in the main text, as counties outside MSAs now have higher developed land supply elasticities. This more elastic supply of developed land again dampens the congestion effect from increased residents, which allows both employment and residents to increase more

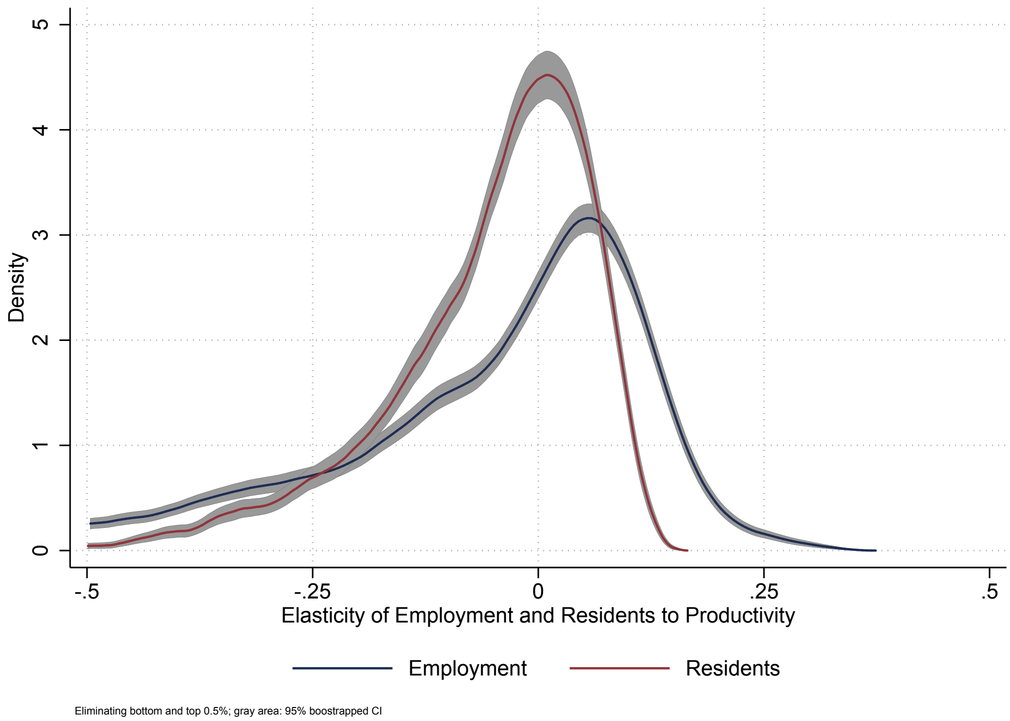


Figure 11: Kernel density for the distribution of employment and residents elasticities in response to a spatially correlated productivity shock in the non-manufacturing sector

than with a perfectly inelastic land supply. Second, there is a further increase in the dispersion in the residents elasticity relative to that in the employment elasticity. Again this is intuitive, as the positive land supply elasticity for counties outside MSAs magnifies the heterogeneity in the responses of residents. In those counties that experience larger increases in residents, there is a greater increase in the supply of developed land, which amplifies the differences in the responses of residents. Third, we continue to find heterogeneity in local employment elasticities that is around the same magnitude as in our baseline specification. Therefore adding an additional source of heterogeneity in residents elasticities does not diminish the heterogeneity in employment elasticities. Fourth, we continue to find substantial differences in the responses of employment and residents, which can only differ because of commuting.

To further reinforce this last point, Figure 14 scatters the residents elasticity against the employment elasticity for each individual county in our data in this robustness exercise. Counties with estimated Saiz elasticities (within MSAs) are shown in black, while counties without Saiz elasticities (outside MSAs) are shown in gray. For both sets of counties, we find that the two elasticities diverge substantially from the 45 degree line and exhibit a relatively weak correlation with one another. Clearly, the employment elasticity for a county is far from a perfect predictor of its residents elasticity, because commuting is a technology that allows for the separation of work and residence, and the extent to which this separation occurs in equilibrium varies across locations.[33]

[33]The relationship between the residence elasticity and the employment elasticity is, if anything, negative. So counties with

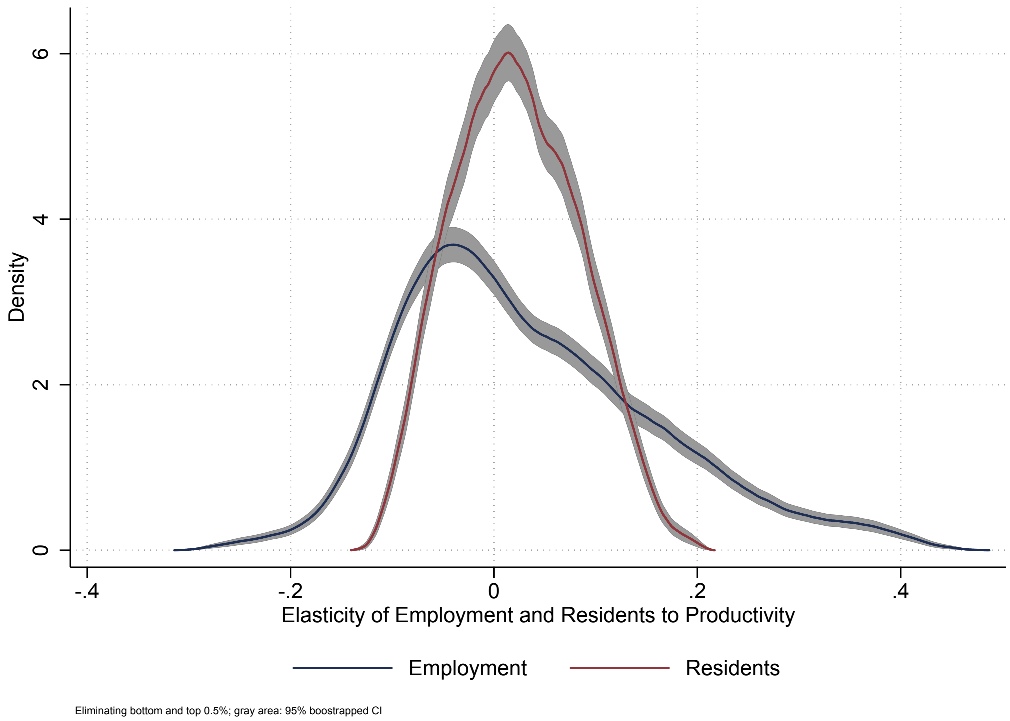


Figure 12: Kernel density for the distribution of employment and residents elasticities in response to a spatially correlated shock in the both sectors

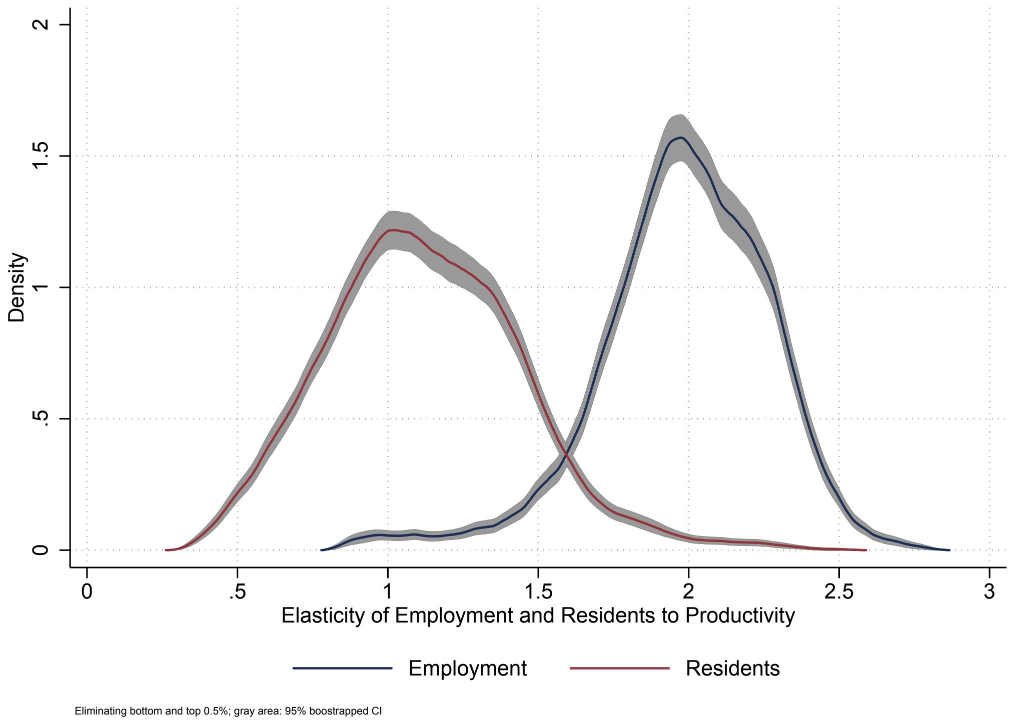
C.7 Commuting Zones

In Figure 15, we show that we continue to find heterogeneity in local employment elasticities if we use commuting zones (CZs) rather than counties. We compute 709 counterfactual exercises where we shock each CZ with a 5 percent productivity shock (holding productivity in all other CZs and holding all other exogenous variables constant). We display the estimated kernel density for the distribution of the general equilibrium elasticity of employment with respect to the productivity shock across these treated CZs (black line). We also show the 95 percent confidence intervals around this estimated kernel density (gray shading). Again we find substantial heterogeneity in the predicted effects of the productivity shock, although less than for counties. Furthermore, this heterogeneity in predicted effects is markedly larger for employment than residents. Since employment and residents can only differ through commuting, this suggests that the use of CZs is an imperfect control for commuting patterns between locations.

C.8 Additional Treatment Heterogeneity Results

In this subsection of the web appendix, we provide further evidence that the model-suggested controls are more successful in explaining the heterogeneity in treatment effects than the standard controls. In

large employment elasticities tend to exhibit low residential elasticities. This is natural, since counties with large employment elasticities are open counties that can attract residents from other counties instead of migrants.



Note: For counties in the 95 MSAs for which Saiz (2010) estimates a housing supply elasticity, we use an elasticity of zero for central city counties and the elasticity estimated by Saiz for the MSA as a whole for outlying counties. For counties outside the 95 MSAs, we use the median Saiz (2010) elasticity across the 95 MSAs.

Figure 13: Kernel density for the distribution of employment and residents elasticities in response to a productivity shock across counties (Aternative specification)

Figure 16, we show that the deviation term from the "difference-in-differences" specification (equation (32) from subsection 4.4 of the paper) is systematically related to the size of the general equilibrium employment elasticity in the model. For the specifications using reduced-form controls (left panel) and model-generated controls (right panel), we display the results of locally-linear weighted least squares regressions of the deviation term β_i against the general equilibrium employment elasticity $\frac{dL_i}{dA_i}\frac{A_i}{L_i}$, along with 95% confidence intervals. In each panel, we show the results of these regressions for each group of control counties, where the results using random county ((i) above), non-neighbors ((iv) above) and all counties ((v) above) are visually indistinguishable.

Using reduced-form controls (left panel) and all definitions of the control group except for the closest county (red line), we find that low elasticities are substantially over-estimated, while high elasticities are substantially under-estimated. This pattern of results is intuitive: low and high elasticities occur where commuting linkages are weak and strong respectively. A reduced-form specification that ignores commuting linkages cannot capture this variation and hence tends to overpredict for low elasticities and underpredict for high elasticities. This effect is still present for the closest county control group (red line), as reflected in the downward-sloping relationship between the deviation term and the general equilibrium elasticity. However, the closest county tends to be negatively affected by the productivity shock, which

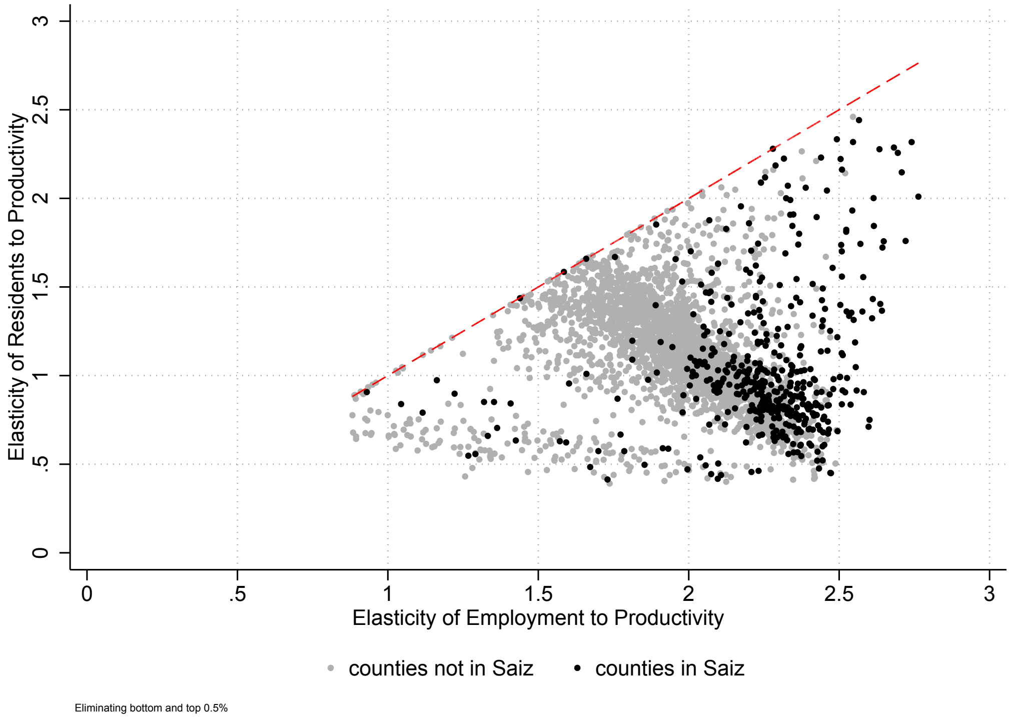
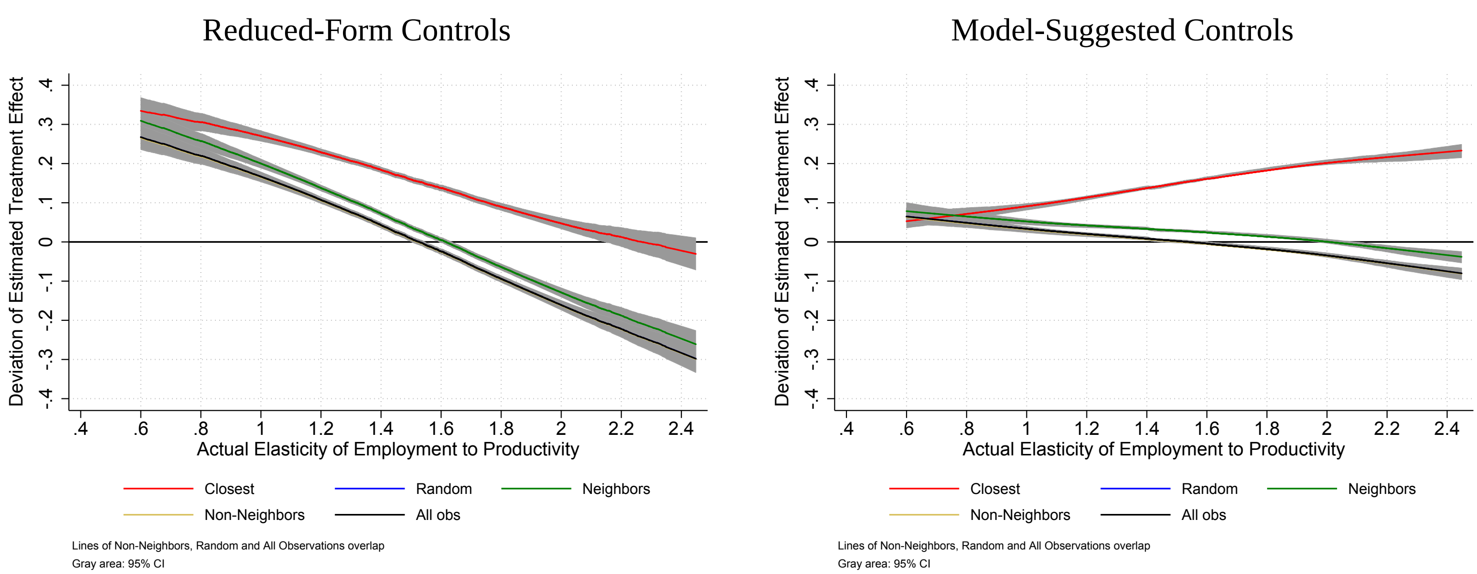


Figure 14: Scatter of the residents elasticity against the employment elasticity for each county in response to a productivity shock (Aternative specification)

shifts the distribution of predicted treatment effects (and hence the distribution of the deviation term) upwards.



Using model-suggested controls (right panel) and all definitions of the control group except for the closest county (red line), we find that the deviation term for the “differences-in-differences” predictions is

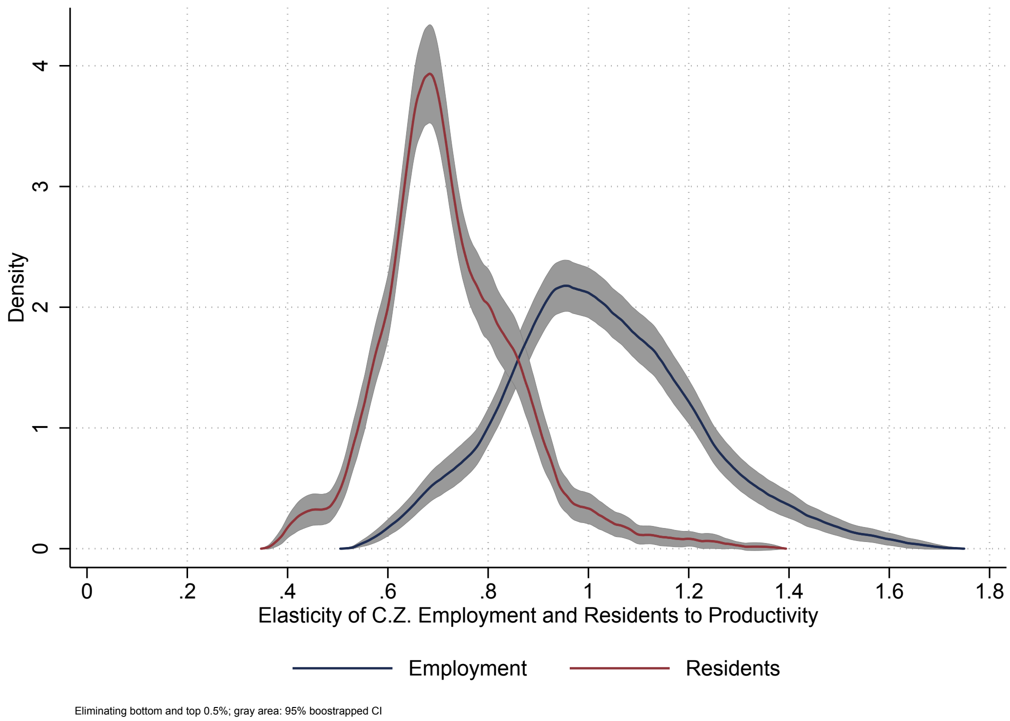


Figure 15: Kernel density for the distribution of employment and residents elasticities in response to a productivity shock across commuting zones

Figure 16: Average deviation term β_i vs. actual Employment Elasticity

close to zero and has a much weaker downward-sloping relationship with the general equilibrium elasticity in the model. The exception is the deviation term using the closest-county as a control, which has an upward-sloping relationship with the general equilibrium elasticity in the model and becomes large for high values of this elasticity. The reason is that the productivity shock to treated counties has larger negative effects on the closest county for higher values of the general equilibrium elasticity in the model, which leads to a larger upward shift in the distribution of the deviation term. This pattern of results again highlights the potentially large discrepancies from the general equilibrium elasticity from using contiguous locations as controls in the presence of spatial linkages in goods and factor markets.

C.9 Commuting Cost Reductions for CZs

In this subsection of the web appendix, we report the results of reductions in commuting costs for CZs. In Figures 17-18, we examine the implications of this counterfactual change in commuting technology for economic outcomes at the commuting zone (CZ) level. We aggregate our county data in both the initial equilibrium and the counterfactual equilibrium to the CZ level and compute counterfactual changes for each CZ in our data. In Figure 17, we show the counterfactual change in CZ employment from prohibitive commuting costs against a measure of CZ dependence on commuting. This measure is the average share of workers in a county within the CZ that live in the same county in which they work, which provides an inverse measure of dependence on commuting.

We find that aggregating up to the CZ level does not eliminate the effects of the change in commuting technology. Rather we find substantial changes in relative CZ employment from prohibitive commuting costs, ranging from declines of around 0.2 log points to rises of around 0.4 log points. Consistent with our earlier results for counties, we find that CZs that are more intensive users of commuting (smaller values on the horizontal axis) experience larger declines in employment, whereas CZs that are less intensive users of commuting (larger values on the horizontal axis) experience larger rises in employment.

Although the employment changes have a strong relationship with the initial intensity with which the CZ uses the commuting technology, perhaps surprisingly, they only have a weak relationship with the initial employment size of the CZ. Figure 18 displays the same change in employment from prohibitive commuting costs (same vertical axis) against initial CZ employment size (different horizontal axis). In line with our earlier findings for local employment elasticities, we find that employment size is a imperfect measure for commuting linkages in factor markets. We find little relationship between relative changes in employment and initial employment size in the CZ in Figure 18. Hence, the importance of commuting is by no means restricted to large cities. This explains why commuting could not be simply approximated by measures of county (or surrounding area) employment size in the previous section.

Taken together, the results in this section have shown that commuting linkages in factor markets are not only important for understanding the incidence of local labor demand shocks but also matter for the spatial distribution of economic activity and aggregate welfare. Increasing commuting costs to prohibitive levels involves a deterioration in the technology for organizing the spatial distribution of economic activity. This deterioration in technology implies less specialization across locations (in residents versus employment) and hence a less efficient sorting of workers and residents across locations in response to spatial variation in productivity and amenities. Furthermore, commuting links are not easily explained by measures of employment size either at the county or the commuting zone level. These results highlight further the importance of incorporating commuting data into regional analysis.

Given this importance of commuting links in shaping the distribution of economic activity across counties, it is natural to expect that these links also determine the magnitude of the impact of reductions in trade costs. In section C.10 of the web appendix, we explore this interaction between trade and commuting costs. We compare the counterfactual effects of a 20 percent reduction of trade costs in the actual world with commuting to the effects in a hypothetical world without commuting. In general, reductions in trade costs lead to a more dispersed spatial distribution of economic activity in the model. But this dispersal is smaller with commuting than without commuting. As trade costs fall, commuting increases the ability of the most productive locations to serve the national market by drawing workers from a suburban hinterland, without bidding up land prices as much as would otherwise occur. These results further underscore the prominence of commuting linkages in shaping the equilibrium spatial distribution of economic activity, and the necessity of incorporating them in models of economic geography.

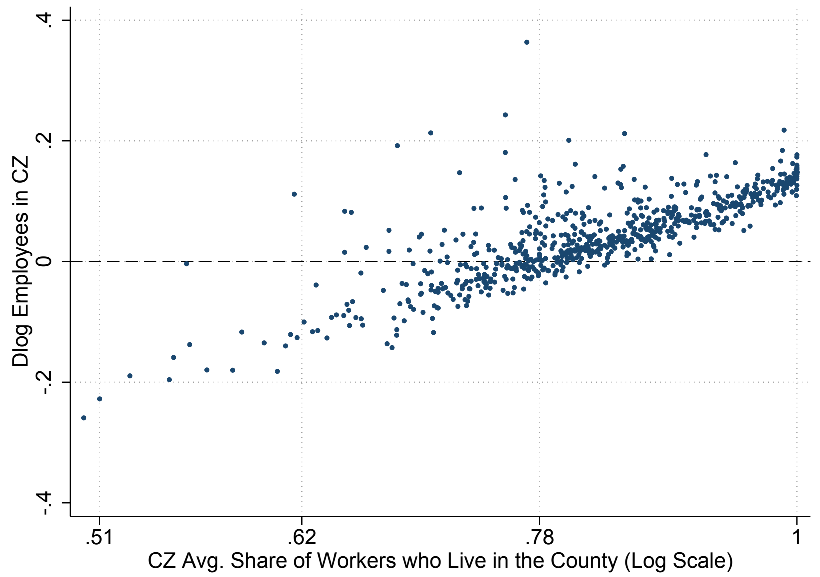


Figure 17: Counterfactual relative change in commuting zone employment ($\hat{L}$) from prohibitive commuting costs and initial dependence on commuting

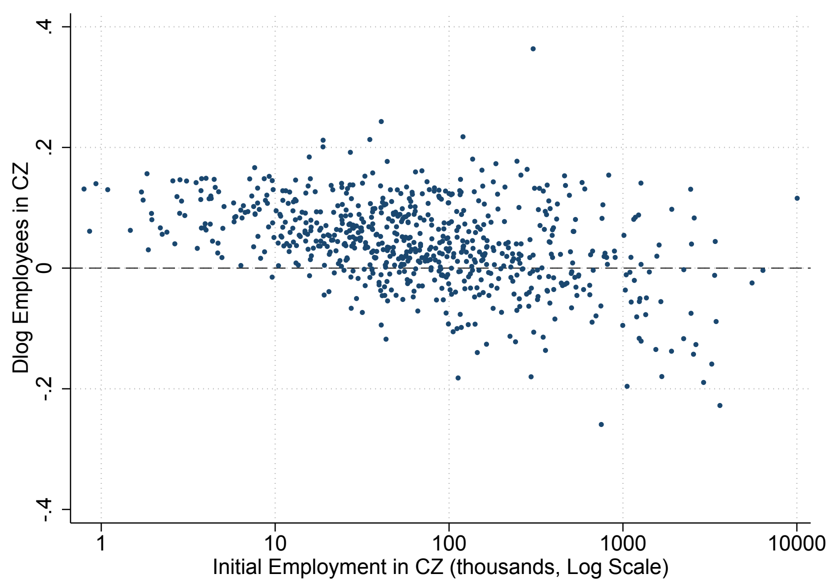


Figure 18: Counterfactual relative change in commuting zone employment ($\hat{L}$) from prohibitive commuting costs and initial employment size

C.10 Interaction Between Trade and Commuting Costs

In this subsection of the web appendix, we examine the extent to which trade and commuting costs interact in the model. As commuting costs shape both the local response to local shocks and have aggregate welfare effects, it is natural to expect that they also influence the effects of reductions in transport costs. To provide evidence on this interaction of spatial linkages in goods and factor markets, we compare the effects of reductions in trade costs, both with and without commuting. We first undertake a counterfactual for a 20 percent reduction in trade costs between locations ($\hat{d}_{ni} = 0.8$ for $n \neq i$ and $\hat{d}_{nn} = 1$) starting from the observed initial equilibrium with commuting (using the observed bilateral commuting shares to implicitly reveal the magnitude of bilateral commuting costs). We next undertake a counterfactual for the same 20 percent reduction in trade costs between locations from a counterfactual equilibrium with no commuting (starting from the observed equilibrium, we first undertake a counterfactual for no commuting, before then undertaking the counterfactual for the reduction in trade costs).

Population mobility again implies that expected utility is equalized across locations. From (18), the change in the common level of expected utility from the trade cost reduction can be decomposed as before into the contributions of changes in the domestic commuting share and real residential income. These changes in real residential income in turn can be further decomposed into the contributions of changes in the domestic trade share, wages, expected residential income, residents and workers. We thus obtain the decomposition of changes in welfare in equation (36), where we again use the fact that $\{\kappa_{nn}, B_{nn}, A_n, d_{nn}\}$ are unchanged.

We find that trade and commuting are weak complements in terms of aggregate welfare, in the sense that the welfare gains from reductions in trade costs are larger in the presence of commuting. Starting from the observed equilibrium, we find aggregate welfare gains from the trade cost reduction of 11.66 percent. In contrast, starting from the counterfactual equilibrium without commuting, we find aggregate welfare gains from the same trade cost reduction of 11.56 percent. Therefore the welfare gains from reducing the friction in goods markets (trade costs) are larger for lower values of the friction in factor markets (lower commuting costs).

We find substantially different effects of reductions in trade costs on the spatial distribution of economic activity with and without commuting. Figure 19 shows the relative change in employment from a 20 percent reduction in trade costs in the New York region (without commuting in the left panel and with commuting in the right panel). In general, reductions in trade costs lead to a more dispersed spatial distribution of economic activity in the model. But this dispersal is smaller with commuting than without commuting. As trade costs fall, commuting increases the ability of the most productive locations to serve the national market by drawing workers from a suburban hinterland, without bidding up land prices as much as would otherwise occur.

In the model lower trade costs and higher commuting costs are both forces for the dispersion of economic activity. On the one hand, lower trade costs weaken agglomeration forces by reducing the incentive for firms and workers to locate close to one another. On the other hand, higher commuting costs increase congestion forces by forcing workers to live where they work, thereby bidding up land prices in congested locations. However, these two sets of forces interact with one another, so that the impact of a reduction in trade costs depends on the level of commuting costs. While lower trade costs necessarily redistribute employment away from the most congested locations, this redistribution is smaller with commuting than without commuting.

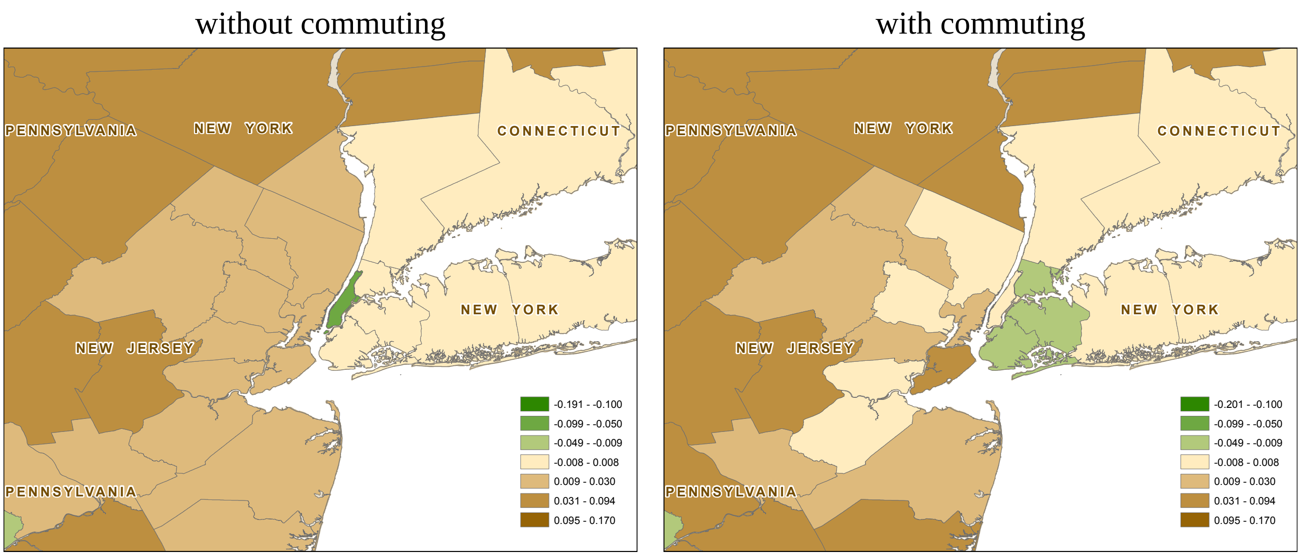


Figure 19: Relative change in employment ($\hat{L}$) from a 20 percent reduction in trade costs (with and without commuting) in the New York area

This exercise also illustrates more generally the role of commuting linkages in shaping the consequences of a reduction in trade costs. Figure 20 shows changes in county employment and real income following a reduction in trade costs in an economy without commuting (vertical axis) and with commuting (horizontal axis), alongside a 45-degree line. We find a relatively low correlation between changes in employment with and without commuting. In particular, commuting and trade tend to be complements in expanding areas: whenever employment increases with the reduction in trade costs, the commuting technology allows a larger expansion because it alleviates the increase in congestion (employment changes are below the diagonal in the left panel of Figure 20). But trade and commuting tend to be local substitutes from the perspective of real income: whenever real income increases with trade, the increase is larger without commuting because production is more spatially dispersed without commuting (real income changes are above the diagonal in the right panel of Figure 20). These results further underscore the prominence of commuting linkages in shaping the equilibrium spatial distribution of economic activity, and the necessity of incorporating them in models of economic geography.

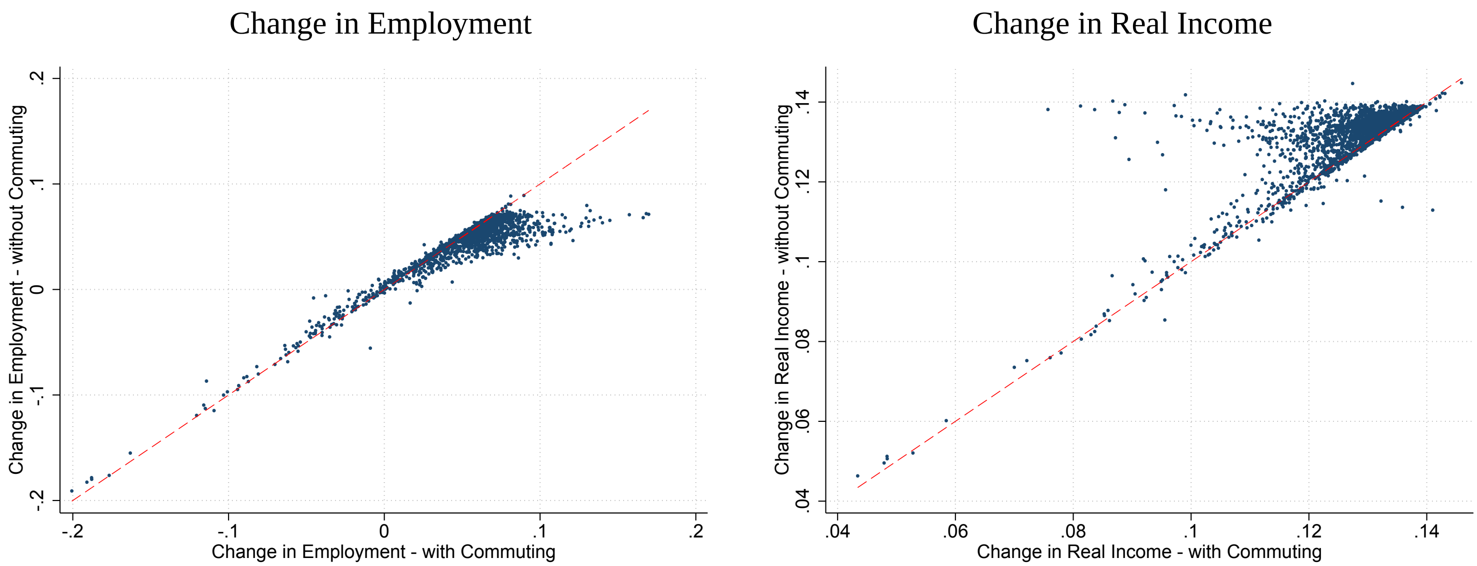


Figure 20: Relative change in employment ($\hat{L}$) and real income ($\widehat{\bar{v}}_n / \left(\hat{P}_n^{\alpha} \hat{Q}_n^{1-\alpha}\right)$) from a 20 percent reduction in trade costs (with and without commuting) across all counties

D Data Appendix

This section of the web appendix contains further information on the data sources and definitions and additional details about the construction of the figures and tables in the paper.

D.1 Data sources and definition

In what follows we list the sources and the variable definitions that we use. We consider them understood in the following section on data processing.

Earnings by Place of Work. This data is taken from the Bureau of Economic Analysis (BEA) website, under Regional Data, Economic Profiles for all U.S. counties. The BEA defines this variable as "the sum of Wages and Salaries, supplements to wages and salaries and proprietors' income. [...] Proprietor's income [...] is the current-production income (including income in kind) of sole proprietorships and partnerships and of tax-exempt cooperatives. Corporate directors' fees are included in proprietors' income, but the imputed net rental income of owner-occupants of all dwellings is included in rental income of persons. Proprietors' income excludes dividends and monetary interest received by nonfinancial business and rental incomes received by persons not primarily engaged in the real estate business." The BEA states that earnings by place of work "can be used in the analyses of regional economies as a proxy for the income that is generated from participation in current production". We use the year 2007.

Total Full-Time and Part-Time Employment (Number of Jobs). This data is taken from the BEA website, under Regional Data, Economic Profiles for all U.S. counties. The BEA defines this series as an estimate "of the number of jobs, full-time plus part-time, by place of work. Full-time and part-time jobs are counted at equal weight. Employees, sole proprietors, and active partners are included, but unpaid family workers and volunteers are not included. Proprietors employment consists of the number of sole proprietorships and the number of partners in partnerships. [...] The proprietors employment portion of the series [...] is more nearly by place of residence because, for nonfarm sole proprietorships, the estimates are based on IRS tax data that reflect the address from which the proprietor's individual tax return is filed, which is usually the proprietor's residence. The nonfarm partnership portion of the proprietors employment series reflects the tax-filing address of the partnership, which may be either the residence of one of the partners or the business address of the partnership." We use the year 2007.

County-to-County Worker Flows. This data contains county-level tabulations of the workforce "residence-to-workplace" commuting flows from the American Community Survey (ACS) 2006-2010 5-year file. The ACS asks respondents in the workforce about their principal workplace location during the reference week. People who worked at more than one location are asked to report the location at which they worked the greatest number of hours. We use data for all the 50 States and the District of Columbia.

County Land Area, County Centroids. This data comes from the 2010 Census Gazetteer Files. When we need to aggregate counties (see below), the land area is the total land area of the aggregated counties, and the centroid of the new county is computed using spatial analysis software.

County Median Housing Values. This data reports the county's median value of owner-occupied housing units from the American Community Survey 2009-2013 5-year file.

Commodity Flows among CFS Area. We use the 2007 Origin-Destination Files of the Commodity Flow Survey for internal trade flows of all merchandise among the 123 Commodity Flow Survey areas in the United States.

Share of counties' manufacturing employment. We use the County Business Pattern file for the year 2007. We use the information on total employment, and employment in manufacturing only. For some counties, employment is suppressed to preserve non-disclousure of individual information, and employment is only reported as a range. In those cases, we proceed as follow. We first use the information on the firm-size distribution, reported for all cases, to narrow the plausible employment range in the cell. We run these regressions separately for employment in manufacturing and total employment. We then use this estimated relation to predict the employment level where the data only reports information on the firm size-distribution. Whenever the predicted employment lies outside the range identified above, we use the employment at the relevant corner of the range.

D.2 Initial data processing

We start by assigning to each workplace county in the County-to-County Worker Flows data, information on the Earnings by Place of Work and the Number of Jobs. Note that the commuting data contains 3,143 counties while the BEA data contains 3,111 counties. This happens because, for example, some independent cities in Virginia for which we have separate data on commuting are included in the surrounding county in the BEA data. We make the two sources consistent by aggregating the relevant commuting flows by origin-destination, and so we always work with 3,111 counties.

The ACS data reports some unrealistically long commutes, which arise for example for itinerant professions. We call these flows "business trips" and we remove them as follow. We measure the distance between counties as the distance between their centroids computed using the Haversine formula. We start by assuming that no commute can be longer than 120km: hence, flows with distances longer than 120km are assumed to only be business trips, while flows with distances less than or equal to 120km are a mix business trips and actual commuting. We choose the 120km threshold based on a change in slope of the relationship between log commuters and log distance at this distance threshold. To split total travellers into commuters and business travellers, we write the identity $\tilde{\lambda}_{ij} = \psi_{ij}^B \tilde{\lambda}_{ij}^B$,where $\tilde{\lambda}_{ij}$ is total travellers, $\tilde{\lambda}_{ij}^B$ is business travellers, $\tilde{\lambda}_{ij}^C$ is commuters, and ψ_{ij} is defined as an identity as the ratio of total travellers to business travellers:

$$\psi_{ij} = \frac{\tilde{\lambda}_{ij}^C + \tilde{\lambda}_{ij}^B}{\tilde{\lambda}_{ij}^B}.$$

We assume that business travel follows the gravity equation $\tilde{\lambda}_{ij}^B = S_i M_j \text{dist}_{ij}^{\delta^B} u_{ij}$,where S_i is a residence fixed effect, M_j is a workplace fixed effect, dist_{ij} is bilateral distance, and u_{ij} is a stochastic error. We assume that ψ_{ij} takes the following form:

$$\psi_{ij} = \begin{cases} 1 & \text{dist}_{ij} > \bar{d} \\ \gamma \text{dist}_{ij}^{\delta^C} & \text{dist}_{ij} \leq \bar{d} \end{cases},$$

where we expect $\gamma > 1$ and $\delta_C < 0$. Therefore we have the following gravity equation for total travellers:

$$\ln \tilde{\lambda}_{ij} = \ln S_i + \ln M_j + \gamma \mathbb{I}_{ij} + (\delta_B + \delta_C \mathbb{I}_{ij}) \ln \text{dist}_{ij} + u_{ij}, \tag{D.1}$$

where $\mathbb{I}_{ij}$ is an indicator variable that is one if $\text{dist}_{ij} \leq \bar{d}$ and zero otherwise. Estimating the above equation for total travellers, we can generate the predicted share of commuters as:

$$\hat{s}_{ij}^C = 1 - \frac{\widehat{\tilde{\lambda}}_{ij}^B}{\widehat{\tilde{\lambda}}_{ij}} = 1 - \frac{\hat{S}_i \hat{M}_j \text{dist}_{ij}^{\hat{\delta}_B}}{\widehat{\tilde{\lambda}}_{ij}},$$

where $\widehat{\tilde{\lambda}}_{ij} = \exp\left(\ln \widehat{\tilde{\lambda}}_{ij}\right)$ are the fitted values from gravity (D.1). Note that this predicted share satisfies the requirements that (a) commuters are zero beyond the threshold $\bar{d}$, (b) the predicted share of commuters always lies in between zero and one, (c) commuters, business travellers and total travellers all satisfy gravity. Note also that since the regression cannot be run on flows internal to a county $\tilde{\lambda}_{ii}$, we set $\hat{s}_{ii}^C = 1$ (i.e., flows of agents who live and work in the same county are assumed to contain no business trips). Therefore we can construct commuting flows as:

$$\widehat{\tilde{\lambda}}_{ij}^C = \hat{s}_{ij}^C \tilde{\lambda}_{ij}.$$

The total business trips originating from residence i are then $\sum_j \left(1 - \hat{s}_{ij}^C\right) \tilde{\lambda}_{ij}$. For any residence i, we reimpute these business trips across destinations j in proportion to the estimated workplace composition of the residence i, $\widehat{\tilde{\lambda}}_{ij}^C / \sum_i \widehat{\tilde{\lambda}}_{ij}^C$. The total employment (and average wage) in a county in the initial equilibrium is taken from the BEA, while total residents (and average residential income) in a county are reconstructed using the estimated residence composition of each workplace. Table 1, Figure 3, and all the results in the paper are based on these "cleaned" commuting flows and initial equilibrium values.

Whenever necessary, we allow for expenditure imbalances across counties. We compute these imbalances as follows. We start from the CFS trade flows. The total sales of a CFS area anywhere must correspond, in a model with only labor (such as the one in this paper), to total payments to workers employed in the area. We rescale the total sales from a CFS area to the value of the total wage bill from the BEA data.[34] For any origin CFS, we keep the destination composition of sales as implied by the CFS bilateral flows. This procedure gives us, for any CFS, total expenditures and total sales consistent with the total labor payments in the economy. We compute the deficit of any CFS area by subtracting total sales from total expenditure. We apportion this deficit across all the counties in the CFS in proportion to the total residential income of the county, as computed above. The total expenditure of the county in the initial equilibrium is always total residential income plus deficit. In any counterfactual equilibrium, the dollar value of the deficit is kept fixed.

D.3 Further information on figures and tables

For some figures in the paper, the main text does not report some technical details related to data manipulation. We report those details here.

Table 1. The table reports statistics on the out-degree distribution (first and third row) and in-degree distribution of the fraction of commuters across counties. Commuting flows are cleaned with the procedure described above. The correspondence between counties and commuting zones is taken from the Economic Research Service of the United States Department of Agriculture.[35]

Figure 1. This figure reports a scatterplot of the log trade flows among CFS areas against log distance between these areas, after removing origin and destination fixed effects. The distance between CFS areas is the average distance travelled by shipments, computed dividing the total ton-miles travelled by the total tons shipped, as reported in the CFS data. Whenever this distance cannot be computed (in about 1/3 of the flows) we supplement it with an estimated distance as follows. We compute the centroids of CFS areas using the Freight Analysis Framework Regions shape-files provided by the Bureau of Transportation Statistics[36] and bilateral distances among these centroids using the Haversine formula. We then regress the actual distance shipped on these centroid-based distances, in logs, and find strong predictive power (slope

[34]For this step, we need a correspondence between CFS areas and counties that is provided by the Census at http://www.census.gov/econ/census/help/geography/cfs_areas.html.

[35]See http://www.ers.usda.gov/data-products/commuting-zones-and-labor-market-areas.aspx.

[36]See http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_atlas_database/2013 /polygon.html

of 1.012, $R^2 = 0.95$). We use the predicted distances from this regression for flows where the average distance shipped cannot be computed. If we restrict our sample to only flows for which the distance can be computed directly, we find a slope of -1.23, and R^2 of 0.82 (similar to the ones used in the main text of -1.29 and 0.83, respectively).

Figure 2. This figure reports a scatterplot of expenditure shares across CFS areas in the data and the model-implied expenditure shares after recovering the productivity of each county, with the procedure described in the main text. Both the estimated productivities and the implied trade shares are calculated using the expenditure of a county allowing for deficits computed as above.

Figure 3. This figure reports a scatterplot of log commuting flows against log distance between county's centroids after removing residence and workplace fixed effects. The commuting flows used in the regression are cleaned of the business trips as described above.

Figure 4. This figure reports a scatterplot of log of land price, as computed from the model, and the County Median Housing Value from the ACS. To compute the price of land in the model we use residents' expenditure allowing for trade deficits. For counties that are aggregated at the BEA level (see above), we compute the population weighted average of the median values.

References

Allen, Treb and Costas Arkolakis (2014) “Trade and the Topography of the Spatial Economy,” *Quarterly Journal of Economics*, forthcoming.

Armington, Paul S. (1969) “A Theory of Demand for Products Distinguished by Place of Production,” *IMF Staff Papers*, 16(1), 159-178.

Berry, Steven, James Levinsohn and Ariel Pakes (1995) “Automobile Prices in Market Equilibrium,” *Econometrica*, 63(4), 841-890.

Caliendo, Lorenzo, Fernando Parro, Esteban Rossi-Hansberg, and Pierre-David Sarte (2014) “The Impact of Regional and Sectoral Productivity Changes on the U.S. Economy,” *NBER Working Paper*, 20168.

Eaton, Jonathan and Kortum, Samuel (2002) “Technology, Geography, and Trade,” *Econometrica*, 70(5), 1741-1779.

Erlander, Stewart and Neil F. Stewart (1990) *The Gravity Model in Transportation Analysis: Theory and Extensions*, Utrecht: VSP.

Fujimoto, Takao and Krause, Ulrich (1985) “Strong Ergodicity for Strictly Increasing Nonlinear Operators,” *Linear Algebra and its Applications*, 71(1), 101-112

McFadden, Daniel (1974) “The Measurement of Urban Travel Demand,” *Journal of Public Economics*, 3(4), 303-328.

McFadden, Daniel and Kenneth Train (2000) “Mixed MNL Models of Discrete Response,” *Journal of Applied Econometrics*, 15, 447-470.

Redding, Stephen J. (2016) “Goods Trade, Factor Mobility and Welfare,” *Journal of International Economics*, 101, 148-167.

Sen, Ashish K. and Tony E. Smith (1995) *Gravity Models of Spatial Interaction Behavior*, New York: Springer Verlag.