Web Appendix for "Commuting, Migration and Local Employment Elasticities" (Not for Publication)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 includes additional empirical results and robustness tests. Section D 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 remaining 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. Section B.10 introduces congestion in commuting. Section B.11 develops an extension of the baseline model to incorporate non-traded consumption goods. Section B.12 considers the case where landlords use residential land. Section B.13 generalizes the production technology to incorporate intermediate inputs, commercial land use and capital. Section B.14 introduces heterogeneity in effective units of labor. Section B.15 considers the case where commuting costs are incurred in effective units of labor rather than in utility. Finally, Section B.16 considers a robustness test in which land is partially owned locally and partially owned by a national portfolio, where these ownership shares are chosen to rationalize measured trade deficits.

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 paper:

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

B.1.2 Residence and Workplace Choices

Using the distribution of utility for pairs of residence and employment locations, 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.[36] 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

[36] 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}):

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

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 equals the wages in all possible workplace locations weighted by the probabilities of commuting to those locations conditional on living in n:

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

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:

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

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

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), land market clearing (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 $\{\hat{w}_n, \widehat{\bar{v}}_n, \hat{Q}_n, \hat{\pi}_{ni}, \hat{\lambda}_{ni}, \hat{P}_n, \hat{R}_n, \hat{L}_n\}$ 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:

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

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

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

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

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

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

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

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:

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

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

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 (9) 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 (12) 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 commuter 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 (9), 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 (9) and commuter market clearing (14)

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

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

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

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

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

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

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

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

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 with Commuting

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 and Commuting

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 parameter 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 and Commuting

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 parameter restrictions:

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

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 (17) can be written as the following excess demand system:

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

where $\tilde{A}_i = A_i^{\sigma-1}$; $\{w_i, L_i, \bar{v}_n, R_n, d_{ni}\}$ have already been determined from the observed data or our parameterization of trade costs; and $\sum_{n \in N} D_n = 0$. This excess demand system exhibits the following properties in $\tilde{A}_i$:

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

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

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

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

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

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

This property can be established by noting:

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

and using homogeneity of degree zero, which implies:

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

and hence:

$$\frac{\partial \mathbb{D}_i\left(\tilde{\mathbf{A}}\right)}{\partial \tilde{A}_i} < 0 \qquad\qquad \text{for all } \tilde{\mathbf{A}} \in \Re^N_+.$$

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 $\mathbb{D}\left(\tilde{\mathbf{A}}\right) = \mathbb{D}\left(\tilde{\mathbf{A}}'\right)$ 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 $\mathbb{D}\left(\tilde{\mathbf{A}}\right) > \mathbb{D}\left(\tilde{\mathbf{A}}'\right)$ and we have a contradiction.

We next establish that there exists a productivity vector $\tilde{\mathbf{A}}^* \in \Re^N_+$ such that $\mathbb{D}(\tilde{\mathbf{A}}^*) = 0$. 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 $\mathbb{D}^+(\cdot)$ by $\mathbb{D}^+_i\left(\tilde{\mathbf{A}}\right) = \max\left\{\mathbb{D}_i\left(\tilde{\mathbf{A}}\right), 0\right\}$. Note that $\mathbb{D}^+(\cdot)$ is continuous. Denote $\alpha\left(\tilde{\mathbf{A}}\right) = \sum_{i\in N}\left[\tilde{A}_i + \mathbb{D}^+_i\left(\tilde{A}_i\right)\right]$. 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:

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

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 $\sum_{i\in N}\mathbb{D}_i\left(\tilde{\mathbf{A}}\right) = 0$, it cannot be the case that $\mathbb{D}_i\left(\tilde{\mathbf{A}}\right) > 0$ for all $i \in N$ or $\mathbb{D}_i\left(\tilde{\mathbf{A}}\right) < 0$ for all

$i \in N$. Additionally, if $\mathbb{D}_i\left(\tilde{\mathbf{A}}\right) > 0$ for some i and $\mathbb{D}_r\left(\tilde{\mathbf{A}}\right) < 0$ 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 $\mathbb{D}_i\left(\tilde{\mathbf{A}}^*\right) = 0$ 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 (19) can be written as the following excess demand system:

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

where $\{w_i, L_i, \bar{v}_n, R_n, \pi_{nn}, A_n, H_n\}$ 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 $\mathcal{B}$ as the excess demand system (B.66) exhibits in $\tilde{\mathbf{A}}$. It follows that there exists a unique vector of unobserved values of the ease of commuting ($\mathcal{B}$) that solves the excess demand system (B.67).

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 (8), we have:

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

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:

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

This implies:

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

which can be re-written as:

$$\frac{dw_i}{dA_i}\frac{A_i}{w_i}+\left(\frac{dL_i}{dw_i}\frac{w_i}{L_i}\right)\left(\frac{dw_i}{dA_i}\frac{A_i}{w_i}\right)=\sum_{r\in N}\left(1-\pi_{ri}\right)\frac{\pi_{ri}\bar{v}_r R_r}{w_i L_i}\left(\frac{dL_i}{dw_i}\frac{w_i}{L_i}\right)\left(\frac{dw_i}{dA_i}\frac{A_i}{w_i}\right)$$
$$-\left(\sigma-1\right)\sum_{r\in N}\left(1-\pi_{ri}\right)\frac{\pi_{ri}\bar{v}_r R_r}{w_i L_i}\left(\frac{dw_i}{dA_i}\frac{A_i}{w_i}\right)+\left(\sigma-1\right)\sum_{r\in N}\left(1-\pi_{ri}\right)\frac{\pi_{ri}\bar{v}_r R_r}{w_i L_i}$$
$$+\frac{\pi_{ii}\bar{v}_i R_i}{w_i L_i}\left(\frac{dR_i}{dw_i}\frac{w_i}{R_i}\right)\left(\frac{dw_i}{dA_i}\frac{A_i}{w_i}\right),$$

where we have used the fact that productivity does not directly enter the commuter market clearing condition (14) and the residential choice probabilities (12), 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 paper:

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

where $\xi_{ri}=\pi_{ri}\bar{v}_r R_r/w_i L_i$ 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 commuter market clearing condition (14), we have:

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

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 paper:

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

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 (12)), we have:

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

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 $\partial P_r = \partial Q_r = 0$ for all r and $\partial w_s = 0$ for $s \neq i$, which yields:

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

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 paper 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,

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

B.8 Gravity and Local Employment Elasticities

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

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

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 $\mathcal{B}_{ni}$ 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:

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

The corresponding the probability of working in location i is:

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

and the probability of residing in location n is:

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

From equations (B.72) and (B.74), the probability of commuting from residence n to workplace i conditional on residing in n is:

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

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

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

Totally differentiating this commuter market clearing condition (B.76) for a given commuting technology $\mathcal{B}_{ni}$, 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 ($\partial \mathcal{W}_i \neq 0$) evaluated at the values of the variables for all other locations from the initial equilibrium ($\partial \mathcal{W}_r = \partial L_r = \partial R_r = 0$ for $r \neq i$):

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

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

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

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. The first term on the right-hand side of equation (B.80) captures the impact of the shock to the workplace fixed effect ($\mathcal{W}_i$) on employment in location i through *commuting*. The second term on the right-hand side captures its impact on employment in location i through *migration*.

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. We show in the main paper that observed commuting flows are characterized by a strong gravity equation relationship.

To show empirically that the heterogeneity in local employment elasticities is a generic implication of the gravity equation, we compute the first term on the right-hand side of equation (B.80) that captures commuting. This first term depends solely on observed variables in the initial equilibrium: (i) the probability of commuting to workplace i conditional on living in residence r and (ii) the share of commuters from residence r in workplace i's employment. In Figure B.1, we show the estimated kernel density of the commuting component of the partial employment elasticity (black line) for counties, and the 95 percent confidence intervals (gray shading). As apparent from comparing Figure B.1 to Figure 4 in the paper, the heterogeneity in county local employment elasticities largely reflects the heterogeneity in this first commuting term, as confirmed in the regressions in Table 2 in the paper. Figure B.2 shows the same commuting component of the partial employment elasticity, but for CZs rather than counties. Comparing Figure B.2 to Figure C.13 in this web appendix, we find that the heterogeneity in CZ local employment elasticities also largely reflects the heterogeneity in this first commuting term.

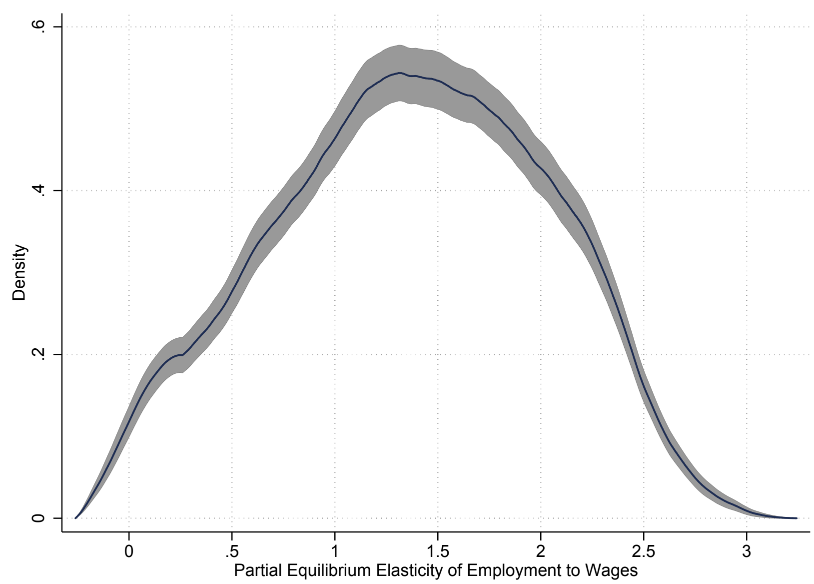


Figure B.1: Commuting component of partial employment elasticity for counties

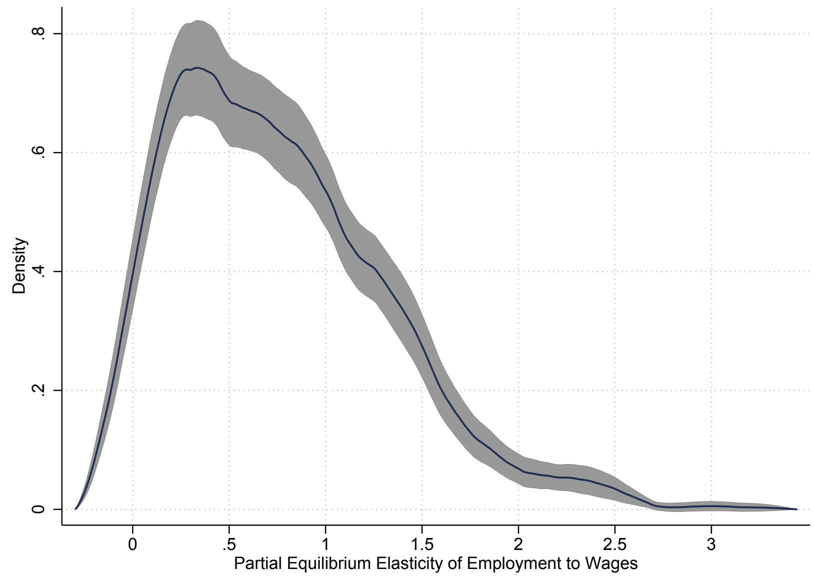


Figure B.2: Commuting component of partial employment elasticity for commuting zones (CZs)

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 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 commuter 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 for a common change in wages in equation (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 equations (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 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 n and workplace i from the following distribution:

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

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:

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

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:

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

Dividing equation (B.94) by its sum across all bilateral pairs, 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 exponents on commuting costs (κ_{ni}), consumption goods price indices (P_n) and land prices (Q_n) are all now multipied by $1/(1-\chi)$. Finally, 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 commuter 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 in 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 costs that are a power function of the volume of commuters 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 introduce commuting into 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 the introduction of commuting 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 commuter 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. As in our baseline specification, land is owned by immobile landlords, who receive worker expenditure on residential land as income, and consume only goods where they live. Therefore, 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 commuter 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 commuter market clearing condition, the partial elasticity of employment with respect to the wage takes the same form as in 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. Therefore, although the presence of non-traded goods can affect the elasticity of *wages* with respect to *productivity*, it leaves unchanged the model's prediction of heterogeneous local *employment* elasticities with respect to *wages*.

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, which in turn implies that the elasticity of local employment with respect to wages takes the same form as in our baseline specification without the non-traded sector.

B.12 Landlords Consume Residential Land

In this subsection of the web appendix, we show that allowing landlords to consume residential land in addition to consumption goods is straightforward, and merely results in less elegant 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 (8) 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 (15) in the paper), workplace choice probabilities (which continue to equal equation (12) in the paper for L_n), residence choice probabilities (which are still equal to equation (12) in the paper for R_n), price indices (again equal to equation (9) in the paper), and the labor market clearing condition (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 local employment elasticities. In our baseline specification in the paper, we assume the following total cost function for tradeable varieties:

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

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 common 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 capital rental rate is the same for all locations ($\mathbb{R}_i = \mathbb{R}$ for all i). Therefore the total cost function now becomes:

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

The probability that a worker chooses to live in location n and work in location i remains the same as in equation (11), which in turn implies that the commuter 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 commuter 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.

In general, incorporating additional factors of production affects the partial elasticity of *wages* with respect to *productivity*, but 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, with an idiosyncratic draw to effective units of labor instead of amenities. Under this alternative specification, the idiosyncratic draw ($b_{ni\omega}$) 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:

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

The main difference between our baseline specification and this alternative specification is the interpretation of wages in the data. In our baseline specification in terms of amenities, the observed wage for each workplace 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 this alternative specification in terms of effective units of labor, the observed wage for each workplace in the data corresponds to the wage per effective unit of labor times average effective units of labor conditional on choosing that workplace, and worker mobility ensures that expected real earnings after taking into account average effective units of labor are equalized across all workplace-residence pairs.

B.15 Commuting Costs in Terms of Labor

In this section of the web appendix, we consider an alternative specification of the model, in which commuting costs are modeled as a reduction in effective units of labor instead of as a reduction in utility. Under this alternative specification, the iceberg commuting cost (κ_{ni}) no longer enters the direct utility function, which is now:

$$U_{ni\omega} = b_{ni\omega} \left(\frac{C_{n\omega}}{\alpha} \right)^{\alpha} \left(\frac{H_{n\omega}}{1-\alpha} \right)^{1-\alpha}. \tag{B.120}$$

However, the iceberg commuting cost 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 effective units of labor net of commuting ($1/\kappa_{ni}$):

$$U_{ni\omega} = \frac{b_{ni\omega} w_i}{\kappa_{ni} P_n^{\alpha} Q_n^{1-\alpha}}. \tag{B.121}$$

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:

$$\lambda_{ni} = \frac{B_{ni} \left(\kappa_{ni} P_n^{\alpha} Q_n^{1-\alpha}\right)^{-\epsilon} w_i^{\epsilon}}{\sum_{r \in N} \sum_{s \in N} B_{rs} \left(\kappa_{rs} P_r^{\alpha} Q_r^{1-\alpha}\right)^{-\epsilon} w_s^{\epsilon}}. \tag{B.122}$$

The main difference between our baseline specification and this alternative specification is whether commuting reduces utility or the labor available for production. One way of interpreting this difference is whether workers absorb the commuting cost through reduced leisure or work time.

B.16 Partial Local and National Ownership of Land

In our baseline specification in the paper, we assume that land is owned by immobile landlords, who receive worker expenditure on residential land as income, and consume only goods where they live. This assumption allows us to incorporate general equilibrium effects from changes in the value of land, without introducing an externality into workers' location decisions from the local redistribution of land rents. In this section of the web appendix, we report a robustness test, in which we instead allow for partial local distribution of land rents (as in Caliendo et al. 2014). In particular, we assume that the share $(1 - \iota_n)$ of expenditure on residential land is redistributed lump sum to local residents, while the remaining share (ι_n) is paid into a national portfolio owned in equal shares by residents throughout the economy. We choose the land ownership share (ι_n) to rationalize the trade deficit for each county in the data. We show that our findings for heterogeneous local employment elasticities are robust to these alternative assumptions about the ownership of land.

B.16.1 Expenditure and Income

Let X_n denote the total expenditure of residents in location n. A fraction $(1 - \alpha)$ of this expenditure is allocated to land. Of this expenditure on land, we assume that a fraction $(1 - \iota_n)$ is redistributed lump sum to local residents, while the remaining fraction ι_n is paid into a national portfolio owned in equal shares by residents throughout the economy. The per capita return from the national land portfolio is given by

$$\xi \equiv \frac{\sum_{i \in N} \iota_i \left(1 - \alpha\right) X_i}{\sum_i R_i}. \tag{B.123}$$

Using this definition, expenditure in location n can be written as the sum of residential income, nationally-redistributed land rent, and locally-redistributed land rent

$$X_n = \bar{v}_n R_n + \xi R_n + (1 - \iota_n)(1 - \alpha) X_n, \tag{B.124}$$

and the trade deficit (equal to expenditure minus income) for each location can be expressed as

$$D_n \equiv X_n - (\bar{v}_n R_n + Q_n H_n) = \xi R_n - \iota_i (1 - \alpha) X_i. \tag{B.125}$$

Using equation (B.125) to substitute for ξR_n in equation (B.124), expenditure in location n can be equivalently written as

$$X_n = \frac{\bar{v}_n R_n + D_n}{\alpha}. \tag{B.126}$$

B.16.2 Calibrating ι to Rationalize the Observed Trade Deficits

We calibrate the land ownership shares (ι_n) for each location n to rationalize the observed trade deficits for each location in the initial equilibrium in the data. Using expenditure (B.124), and denoting the population share of each location in the initial equilibrium by $\rho_n \equiv R_n / \sum_i R_i$, we have

$$X_n = \bar{v}_n R_n + \rho_n \sum_{i \in N} \iota_i (1 - \alpha) X_i + (1 - \iota_n)(1 - \alpha) X_n. \tag{B.127}$$

Using equations (B.126) and (B.127), we have

$$D_n = \alpha X_n - \bar{v}_n R_n = \rho_n \sum_{i \in N} \iota_i (1 - \alpha) X_i - \iota_n (1 - \alpha) X_n, \tag{B.128}$$

which provides a linear system of equations for each location that can be solved for the unique values of ι_n that rationalize the observed trade deficits as an initial equilibrium of the model.

B.16.3 General Equilibrium

We now examine the implications of these alternative assumptions about land ownership for the system of equations that determines general equilibrium. First, workplace-residence choice probabilities (λ_{ni}) take a similar form as in our baseline specification in the paper

$$\lambda_{ni} = \frac{B_{ni} \left(\kappa_{ni} P_n^{\alpha} Q_n^{1-\alpha}\right)^{-\epsilon} w_i^{\epsilon}}{\sum_{r \in N} \sum_{s \in N} B_{rs} \left(\kappa_{rs} P_r^{\alpha} Q_r^{1-\alpha}\right)^{-\epsilon} w_s^{\epsilon}}. \tag{B.129}$$

Therefore the expressions for the number of residents (R_n) and workers (L_i) in each location take the same form as in our baseline specification in the paper

$$R_n = \bar{L} \sum_{i \in N} \lambda_{ni}, \tag{B.130}$$

$$L_i = \bar{L} \sum_{n \in N} \lambda_{ni}. \tag{B.131}$$

Residential expenditure and income are related through equation (B.124), as reproduced here

$$X_n = \bar{v}_n R_n + \xi R_n + (1 - \iota_n)(1 - \alpha) X_n, \tag{B.132}$$

where expected residential income ($\bar{v}_n$) is given by

$$\bar{v}_n = \sum_{i \in N} \lambda_{ni|n} w_i. \tag{B.133}$$

and nationally-redistributed rent per capita (ξ) in equation (B.123) can be written as

$$\xi = \frac{\sum_{i \in N} \iota_i (1 - \alpha) X_i}{\bar{L}}. \tag{B.134}$$

Workplace income equals expenditure on goods produced in that location

$$w_i L_i = \sum_{n \in N} \pi_{ni} \alpha X_n, \tag{B.135}$$

where the bilateral trade shares (π_{ni}) are given by

$$\pi_{ni} = \frac{L_i (d_{ni} w_i / A_i)^{1-\sigma}}{\sum_{k \in N} L_k (d_{nk} w_k / A_k)^{1-\sigma}}. \tag{B.136}$$

Finally, the land rent (Q_n) and price index for tradeables (P_n) are given by

$$Q_n = (1 - \alpha) \frac{X_n}{H_n}, \tag{B.137}$$

$$P_n = \frac{\sigma}{\sigma - 1} \left(\frac{L_n}{\sigma F \pi_{nn}} \right)^{\frac{1}{1-\sigma}} \frac{d_{nn} w_n}{A_n}. \tag{B.138}$$

B.16.4 Computational Algorithm for Counterfactual Changes

We now discuss the computational algorithm that we use to solve this system of equations for a counterfactual equilibrium given the model's parameters $\{\alpha, \sigma, \epsilon, \delta, \kappa\}$, our calibrated land ownership shares ι_n, and assumed changes in the exogenous variables of the model $\{\hat{A}_n, \hat{B}_n, \hat{\kappa}_{ni}, \hat{d}_{ni}\}$. We start with initial guesses for the proportional changes in commuting probabilities, wages and expenditure: $\{\hat{\lambda}_{ni},\, \hat{w}_i,\, \hat{X}_i\}$. Using these initial guesses in the system of equations for general equilibrium, we compute the following proportional changes in the endogenous variables of the model

$$\widehat{\bar{v}}_n^{(t)} = \frac{1}{\bar{v}_n} \sum_{i \in N} \frac{\hat{B}_{ni}\lambda_{ni}\left(\hat{w}_i^{(t)}/\widehat{\kappa}_{ni}\right)^{\epsilon}}{\sum_{s \in N} \hat{B}_{ns}\lambda_{ns}\left(\hat{w}_s^{(t)}/\widehat{\kappa}_{ns}\right)^{\epsilon}} \hat{w}_i^{(t)} w_i, \tag{B.139}$$

$$\hat{L}_i^{(t)} = \frac{\bar{L}}{L_i} \sum_n \lambda_{ni}\hat{\lambda}_{ni}^{(t)}, \tag{B.140}$$

$$\hat{R}_n^{(t)} = \frac{\bar{L}}{R_n} \sum_i \lambda_{ni}\hat{\lambda}_{ni}^{(t)}, \tag{B.141}$$

which are functions of the observed values of variables in the initial equilibrium and our guesses. From the land market clearing condition (B.137), the proportional change in land rents equals our guess for the proportional change in expenditure

$$\widehat{Q}_n^{(t)} = \widehat{X}_n^{(t)}. \tag{B.142}$$

Using the proportional change in employment from equation (B.140) and our guess for the proportional change in wages, we can solve for the proportional change in trade shares from equation (B.136)

$$\widehat{\pi}_{ni}^{(t)} = \frac{\widehat{L}_i^{(t)}\left(\widehat{d}_{ni}\widehat{w}_i^{(t)}/\hat{A}_i\right)^{1-\sigma}}{\sum_{k \in N} \pi_{nk}\widehat{L}_k^{(t)}\left(\widehat{d}_{nk}\widehat{w}_k^{(t)}/\hat{A}_k\right)^{1-\sigma}}. \tag{B.143}$$

Using the proportional change in employment from equation (B.140), the proportional change in trade shares from equation (B.143) and our guess for the proportional change in wages, we can solve for the proportional change in the tradeables price index from equation (B.138)

$$\widehat{P}_n^{(t)} = \left(\frac{\widehat{L}_n^{(t)}}{\widehat{\pi}_{nn}^{(t)}}\right)^{\frac{1}{1-\sigma}} \frac{\widehat{w}_n^{(t)}}{\hat{A}_n}. \tag{B.144}$$

Using our guess for the proportional change in expenditure ($\hat{X}_i^{(t)}$), we can also compute the counterfactual change in nationally-redistributed rent per capita from equation (B.134)

$$\hat{\xi}^{(t)} = \frac{1}{\xi} \frac{\sum_{i \in N} \iota_i \left(1-\alpha\right) X_i \hat{X}_i}{\bar{L}}. \tag{B.145}$$

Finally, we use (B.139)-(B.145) in the equality between income and expenditure (B.135), the workplace-residence choice probabilities (B.129) and expenditure (B.132) to solve for the implied proportional changes in wages, commuting probabilities and expenditure as

$$\tilde{w}_i^{(t+1)} = \frac{1}{w_i L_i \widehat{L}_i^{(t)}} \sum_{n \in N} \alpha \pi_{ni} \widehat{\pi}_{ni}^{(t)} X_n \hat{X}_n^{(t)}, \tag{B.146}$$

$$\tilde{\lambda}_{ni}^{(t+1)} = \frac{\hat{B}_{ni} \left(\widehat{P}_n^{(t)\alpha} \widehat{Q}_n^{(t)1-\alpha} \right)^{-\epsilon} \left(\widehat{w}_i^{(t)} / \widehat{\kappa}_{ni} \right)^{\epsilon}}{\sum_{r \in N} \sum_{s \in N} \hat{B}_{rs} \lambda_{rs} \left(\widehat{P}_r^{(t)\alpha} \widehat{Q}_r^{(t)1-\alpha} \right)^{-\epsilon} \left(\widehat{w}_s^{(t)} / \widehat{\kappa}_{rs} \right)^{\epsilon}}, \tag{B.147}$$

$$\tilde{X}_n^{(t+1)} = \frac{\bar{v}_n R_n \hat{v}_n \hat{R}_n + \xi R_n \hat{\xi}^{(t)} \hat{R}_n^{(t)} + (1 - \iota_n)(1 - \alpha) \hat{X}_n^{(t)}}{X_n}. \tag{B.148}$$

Using these solutions, we update our guesses for wages, commuting probabilities, and expenditures as

$$\hat{w}_i^{(t+1)} = \zeta \hat{w}_i^{(t)} + (1 - \zeta) \tilde{w}_i^{(t+1)}, \tag{B.149}$$

$$\hat{\lambda}_i^{(t+1)} = \zeta \hat{\lambda}_i^{(t)} + (1 - \zeta) \tilde{\lambda}_i^{(t+1)}, \tag{B.150}$$

$$\hat{X}_i^{(t+1)} = \zeta \hat{X}_i^{(t)} + (1 - \zeta) \tilde{X}_n^{(t+1)}, \tag{B.151}$$

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

B.16.5 Local Employment Elasticities

As in Subsection 4.1 in the paper, 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). Figure B.3 shows the estimated kernel densities for the distributions of the general equilibrium elasticities of employment (solid blue line) and residents (dashed red line) with respect to the productivity shock across these treated counties. We also show the 95 percent confidence intervals around these estimated kernel densities (gray shading). This figure is analogous to Figure 4 in the paper, but reports results for this robustness specification, in which the local rents from land are partially redistributed locally and partially contributed to a global portfolio. We continue to find substantial heterogeneity in local employment elasticities that is around the same magnitude as in our baseline specification. This pattern of results is consistent with the heterogeneity in local employment elasticities being a generic prediction of a gravity equation for commuting flows. As a result, our findings of heterogeneous local employment elasticities are robust across different assumptions about the ownership of land.

C Additional Empirical Results

In this section of the web appendix, we report additional empirical results and robustness tests. Subsection C.1 presents additional information on the magnitude and pattern of commuting flows, as a supplement to Subsection 3.2 in the paper. Subsection C.2 uses a time-series decomposition of employment changes to provide direct evidence (independent of our model) that commuting is an important margin through which

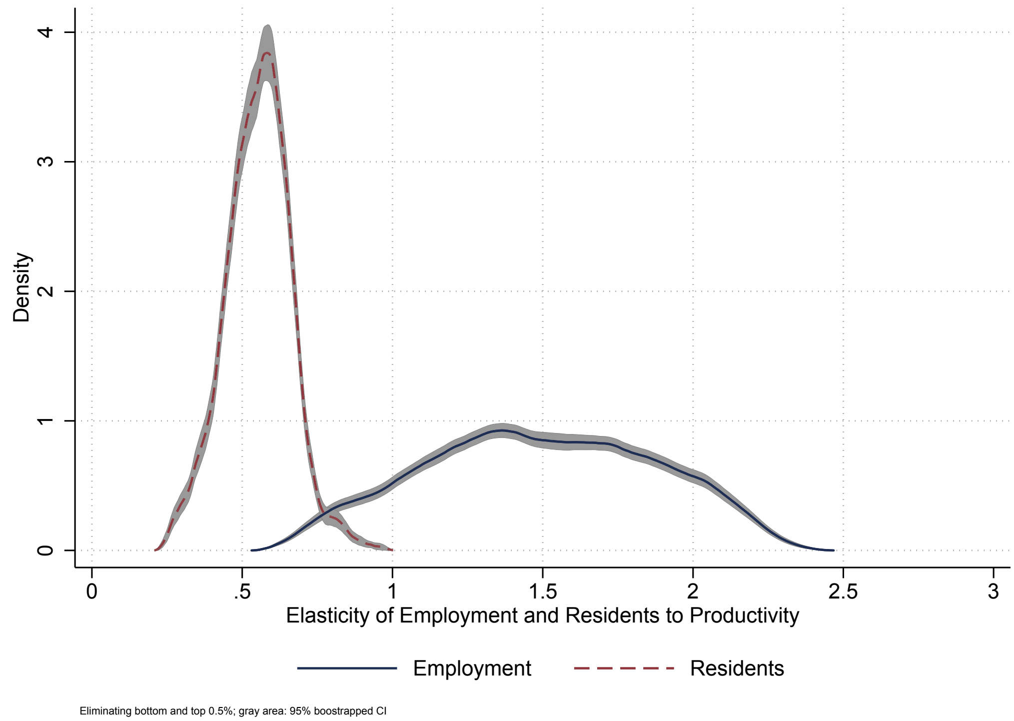


Figure B.3: Kernel density for the distribution of employment and residents elasticities in response to a oductivity shock across counties (partial local and national ownership of land)

changes in employment are achieved in the data.

Subsection C.3 shows that the model's predictions for land prices are strongly positively correlated with median house prices in the data. Subsection C.4 reports standardized coefficients for the regressions examining the determinants of the local employment elasticity in Table 2 in the paper. Subsection C.5 reports additional results from the reduced-form regressions with heterogeneous treatment effects from Subsection 4.4 of the paper. We show that the model-suggested controls are more successful in explaining the heterogeneous treatment effects than the standard empirical controls from the local labor markets literature. Subsection C.6 shows that the heterogeneity in local employment elasticities remains if we shock counties with spatially-correlated shocks reproducing the industrial composition of the U.S. economy.

Subsection C.7 reports additional results with heterogeneous positive supply elasticities for developed land following Saiz (2010), as a supplement to Subsection 4.3 in the paper. Subsection C.8 provides further evidence on the role of commuting in generating heterogeneity in local employment elasticities. We show that there is substantially less heterogeneity in these elasticities in a counterfactual world with no commuting between counties.

Subsection C.9 reports counterfactuals for a 20 percent reduction in the costs of trading costs, both starting from the initial equilibrium in the data with commuting, and starting from a counterfactual equilibrium with no commuting. Subsection C.10 shows that we continue to find substantial heterogeneity in local employment elasticities when we replicate our entire quantitative analysis for commuting zones

(CZs) rather than for counties.

C.1 Magnitude and Pattern of Commuting Flows

In Table 1 in Subsection 3.2 of the paper, we report three sets of descriptive statistics about commuting flows. First, we present the percentiles of the distributions of (i) the share of residents who work outside the county where they live and (ii) the share of workers who live outside the county where they work. Second, conditional on commuting across county boundaries, we report the percentiles of the distributions of the fraction of these commuters who (i) work outside the CZ where they live and (ii) live outside the CZ where they work. Third, we report the percentiles of the distribution of commuting intensity, as measured by the ratio of employment to residents (L_i/R_i) for counties and CZs, which captures whether counties and CZs are net importers ($L_i/R_i > 1$) or net exporters of commuters ($L_i/R_i < 1$).

In this subsection of the web appendix, we report two sets of additional results for commuting flows. First, we examine the determinants of the ratio of employment to residents (L_i/R_i). In Figure 7 in the paper, we show that the initial ratio of employment to residents is central to understanding the effects of counterfactual changes in commuting costs. In Table C.1 below, we show that it is not easy to explain this ratio of employment to residents with the standard empirical controls used in the local labor markets literature (such as various measures of size, area, income and housing supply elasticities). Therefore these results establish that this role of the initial ratio of employment to residents in understanding the effects of changes in commuting costs cannot be easily proxied for by these other controls.

In particular, Table C.1 reports the results of regressing log employment ($\log L_i$), log residents ($\log R_i$), and the ratio of employment to residents (L_i/R_i) on a number of standard empirical controls from the local labor markets literature. The first four columns show that the levels of either employment ($\log L_i$) or residents ($\log R_i$) are strongly related to these standard empirical controls. 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.[37] In contrast, the remaining four columns demonstrate that it is hard to explain the ratio of employment to residents (L_i/R_i) using these same empirical controls. The level of residents, wages, land area, developed-land supply elasticities, employment, and measures of economic activity in surrounding counties, do a poor job in accounting for the variation in this ratio. None of the R-squared's in the last four columns of Table C.1 amounts to more than one third. Taken together, these results confirm that the ratio of employment to residents (L_i/R_i) cannot be easily proxied for by the standard empirical controls used in the local labor markets literature.

Second, we examine the extent to which commuting flows are two-way using the Grubel-Lloyd index introduced in the international trade literature by Grubel and Lloyd (1971). In the context of commuting,

[37] In these specifications in Table C.1, we follow our baseline approach to incorporating the Saiz (2010) MSA-level estimates of developed land supply elasticities (as discussed in Subsection 4.3 of the paper), but we find almost identical results using our alternative approach (as discussed in Subsection C.7 of this web appendix).

the Grubel-Lloyd index captures the extent there is (i) one-way commuting, in which counties either only export or only import commuters, versus (ii) two-way commuting, in which counties simultaneously export and import commuters. Specifically, the Grubel-Lloyd index for county i is defined as

$$GL_i = 1 - \frac{\left| \sum_{n \neq i} L_{in} - \sum_{n \neq i} L_{ni} \right|}{\sum_{n \neq i} L_{in} + \sum_{n \neq i} L_{ni}}, \tag{C.1}$$

where the first subscript is the county of residence and the second subscript is the county of workplace. Therefore, $\sum_{n \neq i} L_{in}$ is county i's total exports of commuters to other counties $n \neq i$ and $\sum_{n \neq i} L_{ni}$ is county i's total imports of commuters from other counties $n \neq i$. If there is only one-way commuting, such that county i either only exports or only imports commuters, $GL_n = 0$. In contrast, if there is perfect two-way commuting, with county i's exports of commuters equal to its imports, $GL_n = 1$.

In Table C.2, we report the mean and percentiles of the distribution of the Grubel-Lloyd index from equation (C.1) across counties. We find pervasive two-way commuting, with the mean and median values of the Grubel-Lloyd index closer to perfect two-way commuting than to only one-way commuting. This pattern of results is consistent with the predictions of the model, in which workers' idiosyncratic preferences between pairs of residence and workplace in general induce two-way commuting. As discussed in Subsection 3.2 of the paper, the model rationalizes zero commuting flows from residence n to workplace i in terms of negligible amenities ($B_{ni} \to 0$) and/or prohibitive commuting costs ($\kappa_{ni} \to \infty$), which can be used to explain one-way commuting.

	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**		1.011**		-0.000		0.027**	
	(0.003)		(0.004)		(0.003)		(0.004)	
$\log w_i$	0.460**		0.473**			0.341**		0.323**
	(0.018)		(0.018)			(0.018)		(0.017)
$\log L_i$		0.957**		0.914**		-0.001		0.037**
		(0.003)		(0.004)		(0.003)		(0.003)
$\log \bar{v}_i$		0.066**		-0.018	0.171**		0.259**	
		(0.025)		(0.025)	(0.023)		(0.025)	
$\log H_i$			0.009	0.041**			-0.015*	-0.028**
			(0.006)	(0.006)			(0.006)	(0.006)
$\log R_{,-i}$			-0.022**				0.576**	0.353**
			(0.005)				(0.121)	(0.110)
$\log \bar{w}_{-i}$			-0.305**				0.218	0.042
			(0.032)				(0.237)	(0.215)
Saiz elasticity			-0.039**	0.057**			-0.039**	-0.041**
			(0.005)	(0.005)			(0.005)	(0.004)
$\log L_{,-i}$				0.082**			-0.621**	-0.400**
				(0.005)			(0.121)	(0.110)
$\log \bar{v}_{-i}$				0.037			-0.312	-0.184
				(0.038)			(0.240)	(0.217)
Constant	-4.667**	-0.165	-1.679**	-0.676*	-0.881**	-2.647**	-0.353	-0.467
	(0.174)	(0.238)	(0.298)	(0.331)	(0.223)	(0.169)	(0.336)	(0.302)
R^2	0.98	0.98	0.99	0.98	0.03	0.16	0.17	0.32
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 C.1: Explaining employment levels and commuting intensity

Statistic	Grubel-Lloyd Index for Commuting
p5	0.342
p10	0.414
p25	0.537
p50	0.696
p75	0.843
p90	0.937
p95	0.968
Mean	0.681

Mean and percentiles of the distribution of the Grubel-Lloyd index from equation (C.1) across counties.

Table C.2: Grubel-Lloyd Index for Commuting

C.2 Decomposing Employment Changes into Commuting and Migration

In this subsection of the web appendix, we assess the importance of commuting for local employment changes in the data (rather than just within our quantitative model). We use the commuter market clearing condition (which necessarily holds as an accounting identity in the data) to undertake decompositions of employment changes over time. The commuter market clearing condition is given by

$$L_i = \sum_{n \in N} \lambda_{ni|n} R_n, \tag{C.2}$$

which can be written as,

$$L_i = \underbrace{\lambda_{ii|i} R_i}_{\text{own residents}} + \underbrace{\sum_{n \neq i} \lambda_{ni|n} R_n}_{\text{commuters}}. \tag{C.3}$$

Taking differences over time, we obtain,

$$\Delta L_i = \underbrace{\Delta \left(\lambda_{ii|i} R_i \right)}_{\text{own residents}} + \underbrace{\sum_{n \neq i} \Delta \left(\lambda_{ni|n} R_n \right)}_{\text{commuters}}. \tag{C.4}$$

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

$$\frac{\text{abs} \left[\sum_{n \neq i} \Delta \left(\lambda_{ni|n} R_n \right) \right]}{\text{abs} \left[\Delta \left(\lambda_{ii|i} R_i \right) \right] + \text{abs} \left[\sum_{n \neq i} \Delta \left(\lambda_{ni|n} R_n \right) \right]}. \tag{C.5}$$

which is bounded between zero and one. When all changes in employment occur through commuting, this index takes the value of one. When all changes in employment are achieved through migration, this index 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), this index represents the fraction of the change in employment accounted for by commuting.

We implement this decomposition for changes in county employment from 1990-2010. For 34 percent of counties, the employment change due to commuting is larger than that due to migration. Across all counties, the mean value of the index (C.5) is 0.4, implying that on average 40 percent of the change in employment is the result of commuting. For the county at the median value of this index, 39 percent of the change in employment is due to the change in commuting. For the county at the 75th (90th) percentile, commuting accounts for 58 percent (77 percent) of the employment change. For the county at the 25th (10th) percentile, the contribution from commuting is 21 percent (8 percent). Taken together, these results provide direct evidence, independent of the model, that changes in commuting are an important and heterogeneous margin through which changes in employment are achieved in the data.

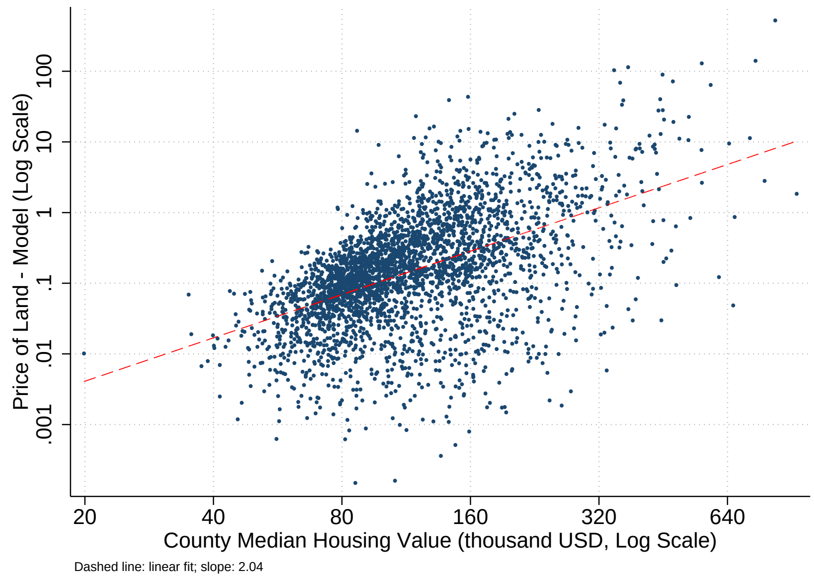


Figure C.1: Land Prices in the Model and House Prices in the Data

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 C.1, 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.

C.4 Standardized Employment Elasticities Regression

Table C.3 reports the coefficients from the same set of regressions presented in Table 2 in the paper, 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 a change in the independent variable by one standard deviation.

C.5 Additional Treatment Heterogeneity Results

In this subsection of the web appendix, we supplement the results reported in Subsection 4.4 of the paper, and provide further evidence that the model-suggested controls are more successful in explaining the heterogeneity in treatment effects than the standard empirical controls from the local labor markets literature. In Figure C.2, we show that the deviation term from the "difference-in-differences" specification (equa-

	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 weighted average of their workplace wage. All variables are standardized. * $p < 0.05$; ** $p < 0.01$.

Table C.3: Explaining the general equilibrium local employment elasticities to a 5 percent productivity shock (standardized regression)

tion (30) 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 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 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.

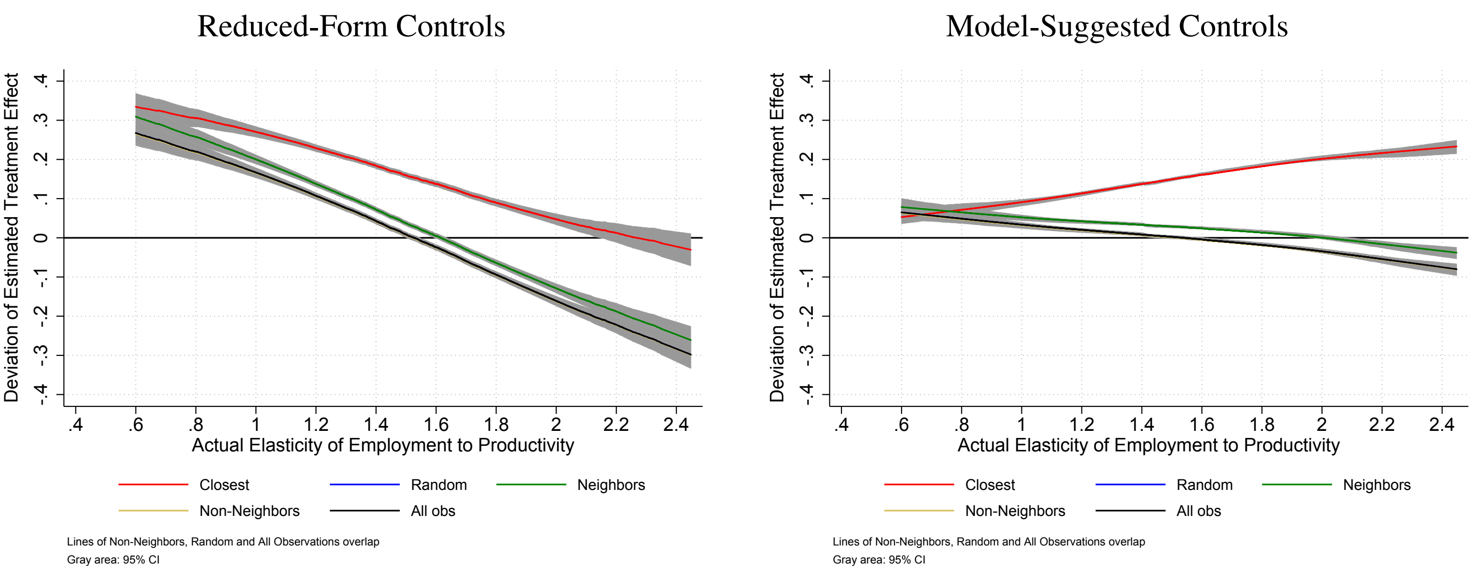


Figure C.2: Average deviation term β_i vs. general equilibrium employment elasticity

C.6 Spatially Correlated Productivity Shocks

In this section of the web appendix, we show that the heterogeneity in local employment elasticities remains if we shock counties with spatially-correlated productivity shocks reproducing the industrial composition of the US economy. We construct these 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 County Business Patterns; see Data Appendix below), we infer a growth in the non-manufacturing sector's TFP of 3.1%. We use the County Business Patterns 2007 data to also compute the share of each county's manufacturing employment over total employment. Figure C.3 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 C.4 shows the resulting distribution of elasticities of employment and residents.

Figure C.5 shows an analogous exercise for a shock to the non-manufacturing sector. Finally, Figure C.6 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. Across all of these specifications, we continue to find substantial heterogeneity in local employment elasticities.

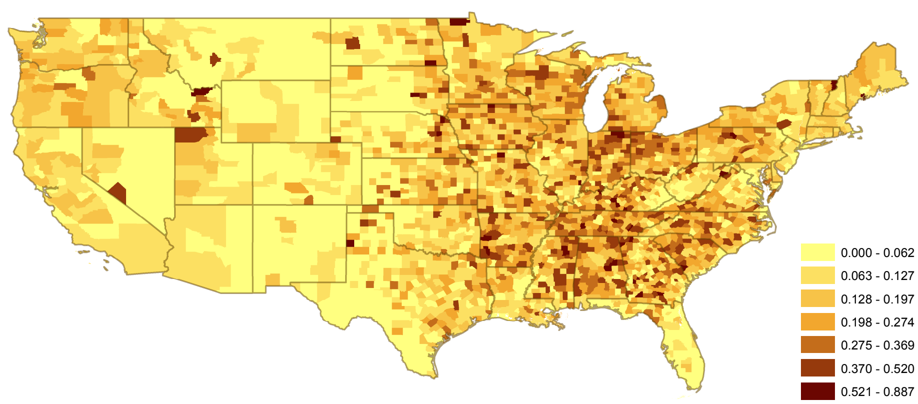


Figure C.3: U.S. counties' share of employment in manufacturing, 2007.

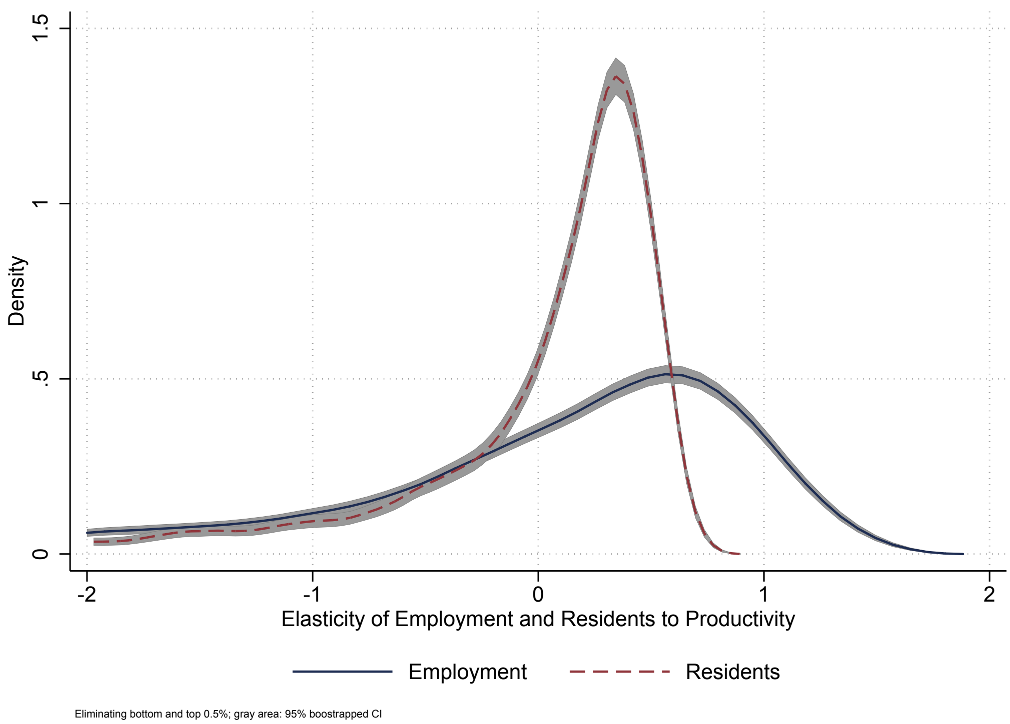


Figure C.4: Kernel density for the distribution of employment and residents elasticities in response to a spatially correlated productivity shock in the manufacturing sector

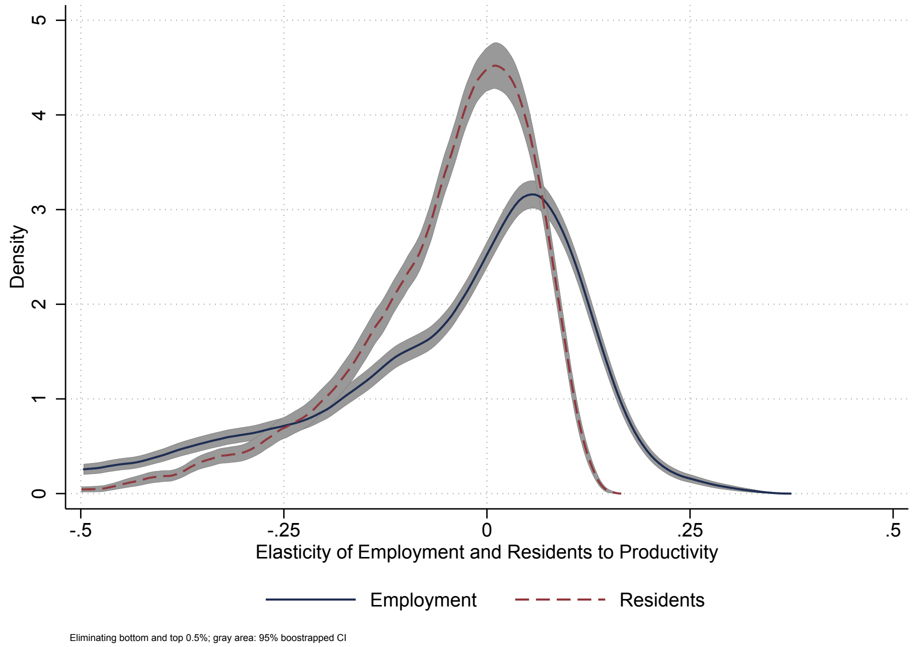


Figure C.5: Kernel density for the distribution of employment and residents elasticities in response to a spatially correlated productivity shock in the non-manufacturing sector

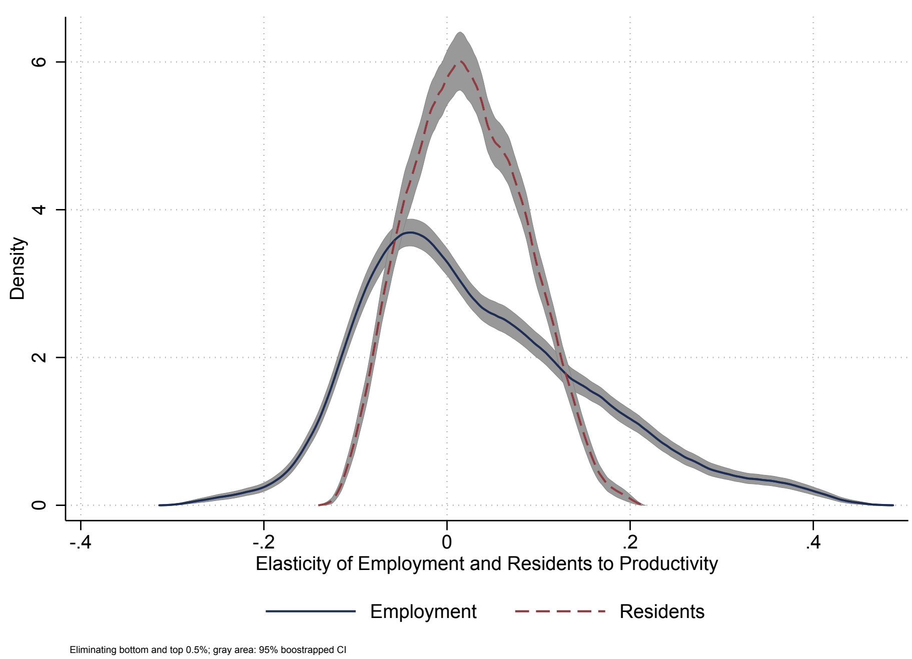


Figure C.6: Kernel density for the distribution of employment and residents elasticities in response to a spatially correlated shock in the both sectors

C.7 Additional Results with Positive Developed Land Elasticities

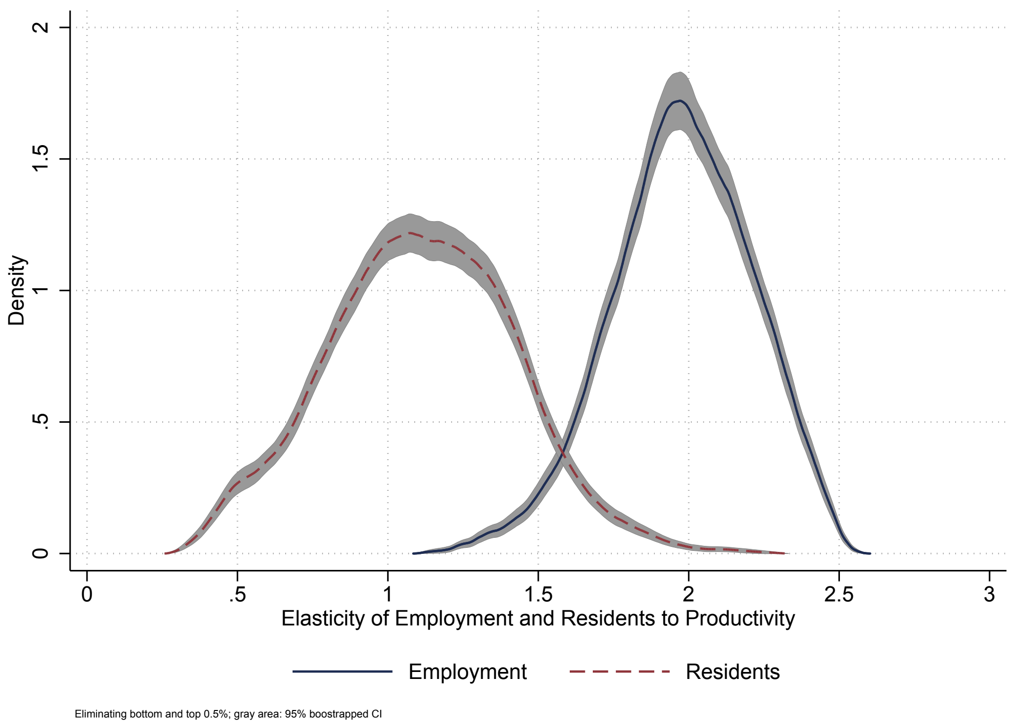
In Subsection 4.3 of the paper, we introduce the Saiz (2010) MSA-level elasticities of developed land where we have them (assuming the same elasticity for all counties within an MSA), and retain our baseline specification of a zero elasticity for all other counties. As discussed in the paper, one limitation of this first 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, we here 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 the median Saiz estimate across MSAs of 1.67 for all other counties. Using this second specification, we replicate our 3,111 counterfactual exercises where we shock each county with a 5 percent productivity shock (holding productivity fixed in all other counties and holding all other exogenous variables constant). Figure C.8 displays the results and is analogous to Figures 4 and 5 in the paper. This figure shows the estimated kernel density for the distribution of the general equilibrium elasticity of both employment (blue solid line) and residents (red dashed line). We also show the 95 percent confidence intervals around these estimated kernel densities (gray shading).

We find a similar pattern of results as for our baseline specification and first approach to introducing the Saiz elasticities in Section 4.3 of the paper. First, both the employment and residents elasticities increase on average relative to the specification in Section 4.3 of the paper, 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 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. 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 C.8 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 that are assigned the median Saiz estimate of 1.67 (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. Indeed, if anything, the relationship between the residence elasticity and the employment elasticity in Figure C.8

is negative. Therefore counties with larger employment elasticities, on average, have smaller residential elasticities. This pattern of results is consistent with the idea that counties with larger employment elasticities are more open to commuting, and hence more able to attract residents from other counties rather than from themselves.



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 C.7: Kernel density for the distribution of employment and residents elasticities in response to a productivity shock across counties (Alternative Saiz specification)

C.8 Additional Results with No Commuting Between Counties

In this subsection of the web appendix, we provide further evidence that the heterogeneity in local employment elasticities is driven by commuting, by reporting local employment elasticities for a counterfactual world with no commuting between counties. As in our counterfactuals in Section 4.1 in the paper, we start with the initial equilibrium in the observed data. We first undertake a counterfactual for prohibitive commuting costs between counties ($\kappa_{ni} \to \infty$ for $n \neq i$) and solve for the new spatial equilibrium distribution of economic activity. Starting from this counterfactual world with no commuting between counties, we next 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).

Figure C.9 shows the estimated kernel density for the distribution of the general equilibrium elasticity of employment with respect to the productivity shock across the treated counties (red dashed line). In

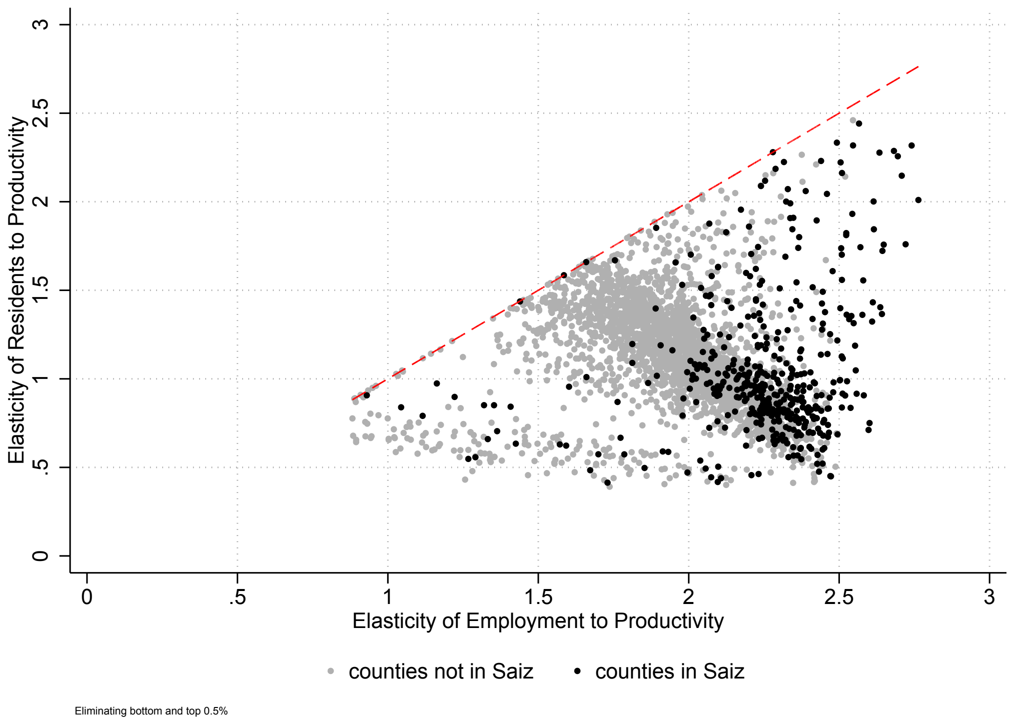


Figure C.8: Scatter of the residents elasticity against the employment elasticity for each county in response to a productivity shock (Alternative specification)

this counterfactual world with no commuting, the employment and residents elasticity are equal to one another. To provide a point of comparison, the figure also displays the estimated kernel density for the general equilibrium employment elasticity from our baseline specification in the paper (blue solid line), with commuting between counties. Even in the absence of commuting between counties, we expect local employment elasticities to be heterogeneous, because counties differ substantially from one another in terms of their initial shares of U.S. employment. Consistent with this, we find that local employment elasticities in the world with no commuting between counties range from around 0.5 to 1. However, this variation is substantially less than in our baseline specification with commuting between counties, where the local employment elasticities range from around 0.5 to 2.5. Therefore, these results provide further evidence that commuting indeed plays a central role in generating the heterogeneity in local employment elasticities. Comparing the two specifications in Figure C.9, local employment elasticities are also larger on average with commuting than in the counterfactual world without commuting. This pattern of results is consistent with commuting weakening congestion forces in the model. As a county experiences an increase in productivity, commuting enables it to increase employment by drawing residents from surrounding counties, thereby bidding up land prices less than otherwise would be the case in a world without commuting between counties.

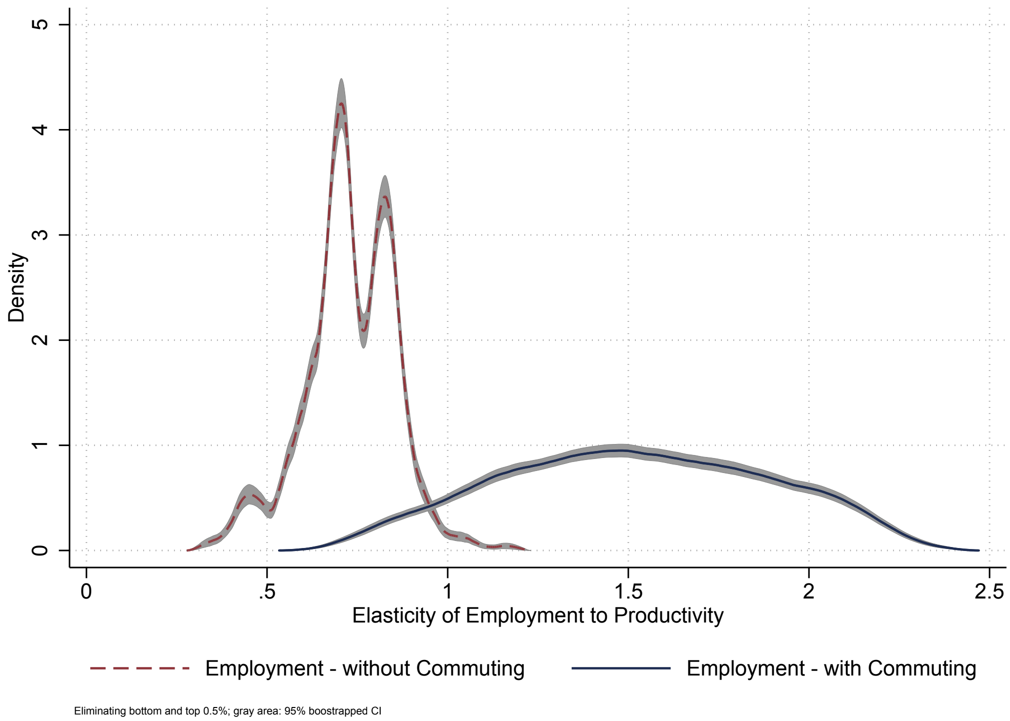


Figure C.9: Kernel density for the distribution of employment and resident elasticities in response to a productivity shock across counties (with and without commuting between counties)

C.9 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. To provide evidence on this interaction, we compare the effects of reductions in trade costs, both with and without commuting between counties. To do so, 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 between counties (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 between counties. That is, starting from the observed equilibrium, we first undertake a counterfactual for prohibitive commuting costs between counties ($\kappa_{ni} \rightarrow \infty$ for $n \neq i$), before then undertaking the counterfactual for the reduction in trade costs.

We find that commuting between counties has a relatively small impact on the welfare gains from trade cost reductions. 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 between counties, we find aggregate welfare gains from the same trade cost reduction of 11.56 percent. However, we find that commuting between counties plays a major role in influencing the impact of trade cost reductions on the spatial distribution of economic activity. Figure C.10 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 between counties than without it. 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.

Intuitively, lower trade costs and higher commuting costs are both forces for the dispersion of economic activity in the model. 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. 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 between counties than without it.

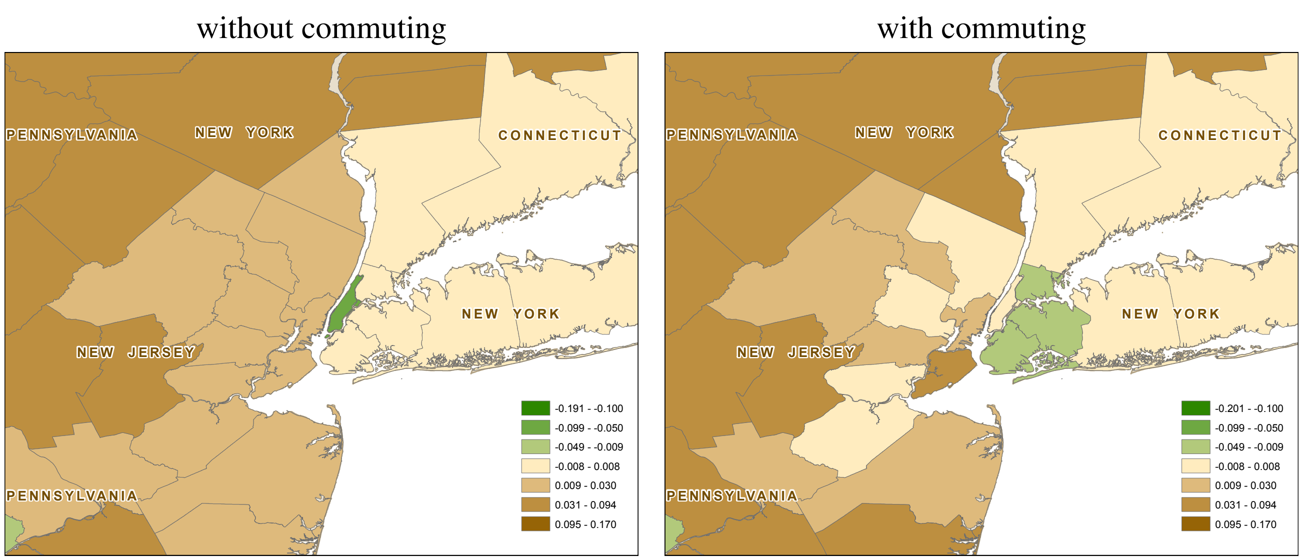


Figure C.10: Relative change in employment ($\hat{L}$) from a 20 percent reduction in trade costs (with and without commuting between counties) in the New York area

This exercise also illustrates more generally the role of commuting linkages in shaping the consequences of a reduction in trade costs. Figure C.11 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 between counties. 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 C.11). 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 C.11). 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.

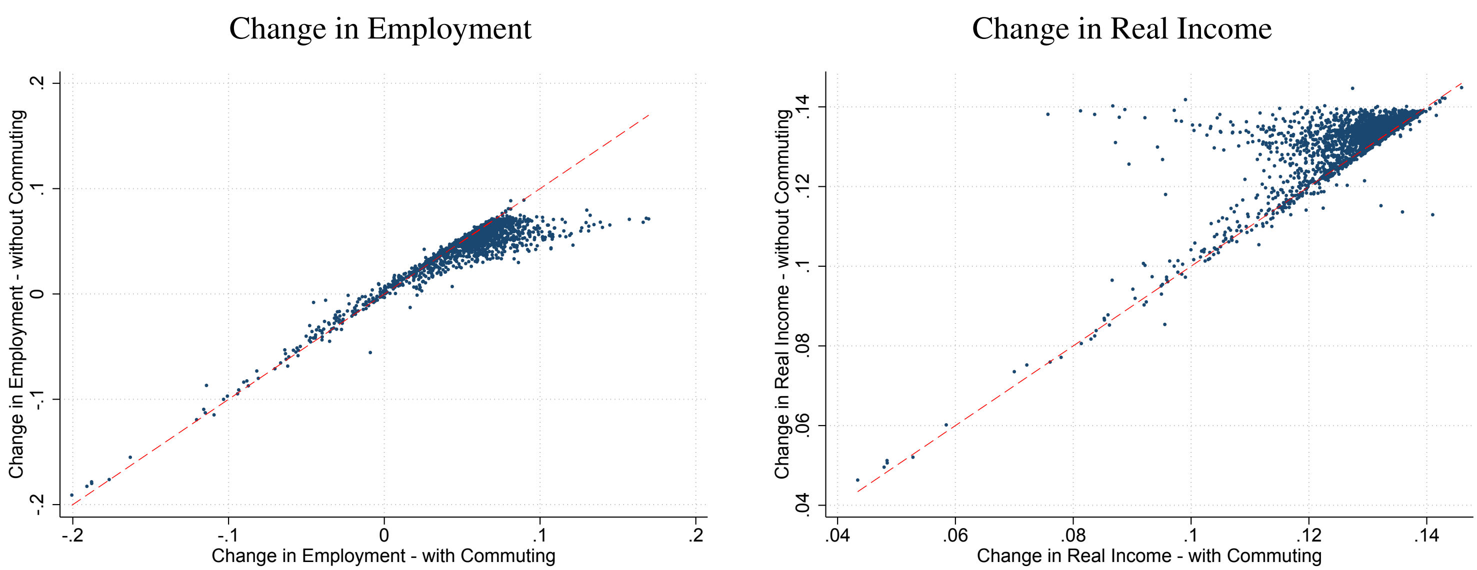


Figure C.11: Relative change in employment ($\hat{L}$) and real income ($\widehat{\bar{v}}_n / \left(\hat{P}_n^{\alpha} \hat{Q}_n^{1-\alpha}\right)$) from a 20 percent reduction in trade costs across all counties (with and without commuting between counties)

C.10 Commuting Zones (CZs)

As discussed in the paper, an important advantage of our theoretical framework is that it explicitly models the spatial linkages between locations in both goods and factor markets, and hence can be taken to data at alternative levels of spatial disaggregation. In contrast, existing research that does not explicitly model spatial linkages between locations faces a trade-off when studying local labor markets. On the one hand, using more disaggregated spatial units reduces the un-modeled spatial linkages between locations. On the other hand, using more aggregated spatial units reduces the ability to make inferences about local labor markets. Furthermore, no choice of boundaries for spatial units can be perfect for eliminating spatial linkages between locations. For example, should Princeton, NJ be considered part of New York's or Philadelphia's CZ? There is no perfect answer to this question, because some people commute from Princeton to New York, whereas other people commute from Princeton to Philadelphia. Therefore there exists no choice of boundaries for CZs that eliminates commuting. Our approach overcomes this problem by explicitly modeling the spatial linkages between the chosen spatial units.

In our baseline specification in the paper, we report results for counties, because this is the finest level of geographical detail at which commuting data are reported for the entire United States in the American Community Survey (ACS) and Census of Population, and a number of influential papers in the local labor markets literature have used county data (such as Greenstone, Hornbeck and Moretti 2010). In this section of the web appendix, we report the results of a robustness check, in which we replicate our entire analysis for Commuting Zones (CZs) (aggregations of counties). This replication involves undertaking the full quantitative analysis of the model at this higher level of spatial aggregation. First, we aggregate our employment and wage to the CZ level. Second, we aggregate our bilateral commuting data between pairs of counties to construct bilateral commuting flows between pairs of CZs. Third, we use our data on bilateral trade between CFS regions to solve for implied CZ productivity (A_i) and bilateral trade between CZs (π_{ni}), using the same approach as for counties in our baseline specification in Section 3.1 of the paper. Fourth, we use our data on bilateral commuting between pairs of CZs to solve for implied bilateral amenities (B_{ni}), using the same approach as in Section 3.2 of the paper. In Figure C.12, we show the conditional relationship between the log value of commuting flows and log distance between pairs of CZs, after removing workplace and residence fixed effects. This figure is analogous to Figure 3 in the paper, but uses CZs rather than counties. Again we find that the gravity equation provides a good approximation to the data, with a tight and approximately log linear relationship between the two variables.

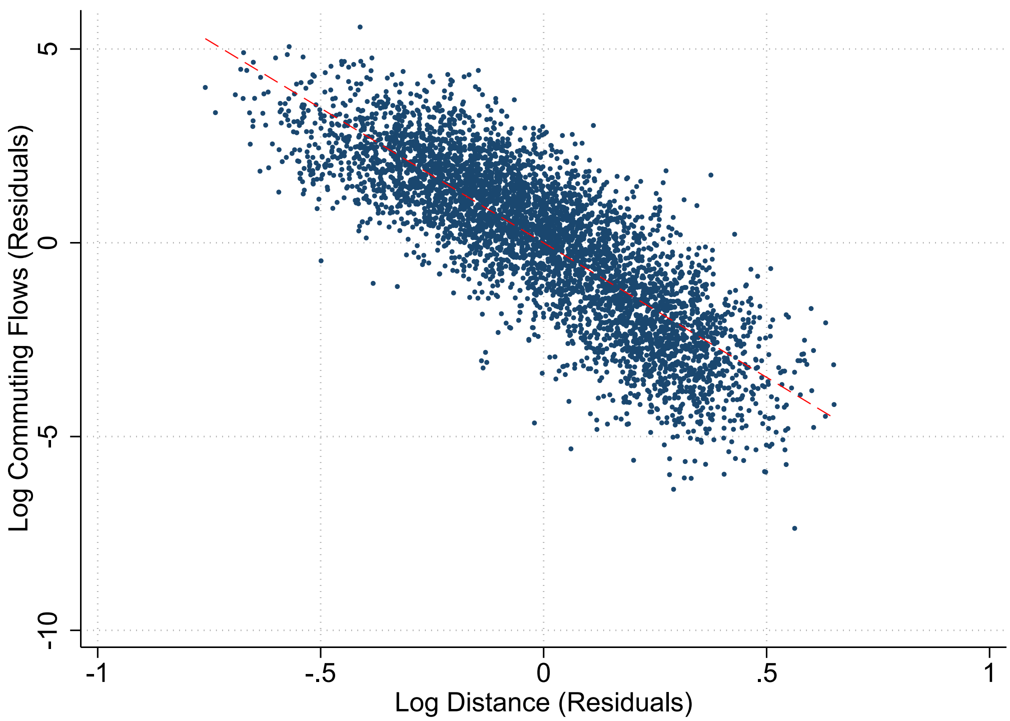


Figure C.12: Gravity in Commuting Between Commuting Zones (CZs)

Having calibrated the model to match the initial equilibrium in the observed data at the CZ level, we next shock each of the 709 CZs with a 5 percent productivity shock, following the same approach as for counties in Section 4.1 of the paper. Figure C.13 shows the estimated kernel density for the general equilibrium elasticities of employment and residents with respect to the productivity shock across the treated CZs (blue solid and red dashed lines). We also show the 95 percent confidence intervals around these estimated kernel densities (gray shading). As CZs are aggregations of counties, there is necessarily less

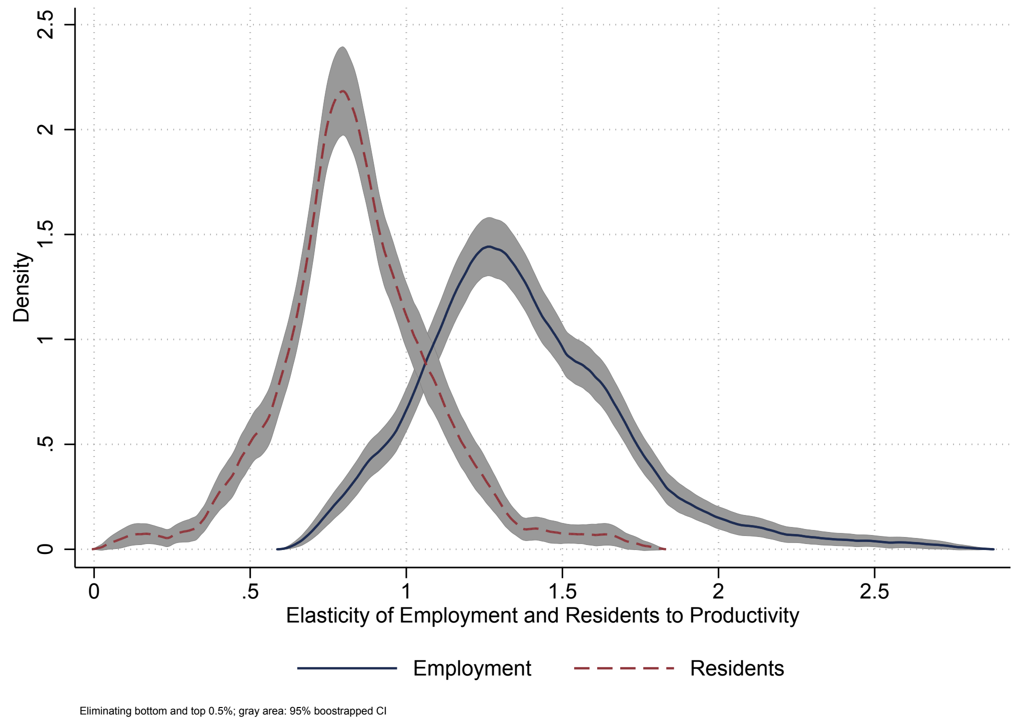


Figure C.13: Kernel density for the distribution of employment and residents elasticities in response to a productivity shock across CZs

commuting between pairs of CZs than between pairs of counties. Nonetheless, CZs differ substantially in the extent to which their boundaries capture commuting linkages. Therefore we find that there is sufficient variation in the importance of commuting networks across CZs to generate substantial heterogeneity in the local employment elasticity, which ranges from just above 0.5 to just over 2.5, a similar range as for the employment elasticity distribution across counties. Again we find substantial differences between the employment and residents elasticities, with the residents elasticity having less dispersion. Since employment and residents can only differ through commuting, these findings reinforce the importance of commuting in understanding the local response to local economic shocks, even at the more aggregated level of CZs.

In Table C.4, we provide further evidence on the role of commuting linkages in explaining the heterogeneity in employment elasticities across CZs. This table is analogous to Table 2 in the paper, but reports results for CZs rather than for counties. In Columns (1)-(4), we regress the local employment elasticity on standard empirical controls from the local labor markets literature. Although some of these controls are 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 just over one quarter in Column (4). Therefore there is considerable variation in local employment elasticities not explained by these standard empirical controls. In contrast, when we include the share of workers that work in i conditional on living in i ($\lambda_{ii|i}$) in Column (5) as as summary statistic for openness to commuting, we find that this variable is highly statistically significant, and results in an R-squared of over one half. Including the partial equilibrium elasticities that capture commuting linkages in the model further increases the R-squared to around 0.60, more than double that using the standard controls in Column (4). In the last two columns, we combine these partial equilibrium elasticities with the standard controls used in the first four columns. Although some of these standard controls are statistically significant, we find that they add little once we control for the partial equilibrium elasticities.

Taken together, these results confirm that the use of CZs is an imperfect control for commuting. There remains substantial heterogeneity in employment elasticities across CZs, because they differ in the extent to which their boundaries are successful in capturing commuting patterns. This heterogeneity in employment elasticities across CZs is not well explained by standard controls from the local labor markets literature. In contrast, consistent with our results for counties above, we find that adding a summary statistic of commuting, or the partial equilibrium elasticities from the model, can go a long way in explaining the heterogeneous responses of CZs to productivity shocks.

We next examine the impact of reductions in the costs of commuting between CZs on the spatial distribution of economic activity. We undertake a counterfactual in which we reduce commuting costs between CZs by the same proportional amount as for counties in our central exercise in Section 5 of the paper ($\hat{\mathcal{B}}_{ni} = 0.88$). In Figure C.14, we show the proportional change in employment for each CZ against its initial commuting intensity (L_i/R_i), where $L_i/R_i > 1$ implies that a CZ is a net importer of commuters and $L_i/R_i < 1$ implies that a CZ is a net exporter of commuters. We find substantial changes in employment for individual CZs, which range from increases of 10 percent to reductions of 20 percent. Furthermore, these changes in the distribution of employment across CZs are well explained by initial commuting intensity. In contrast, in Figure C.15, we show that the same proportionate change in employment for each CZ against its initial employment size. We find little relationship between the impact of the reduction in commuting costs on employment and initial CZ size. Therefore, these results confirm our findings for counties that the importance of commuting is by no means restricted to large cities.

More generally, in Table C.5, we show that it is not easy to proxy for CZ commuting intensity (L_i/R_i) using standard empirical controls from the local labor markets literature. This table is analogous to Table C.1 earlier in this web appendix, but reports results for CZs rather than for counties. The first four columns show that the levels of either employment ($\log L_i$) or residents ($\log R_i$) are strongly related to these standard empirical controls. The first column shows that one can account for most of the variation in CZ 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 CZs.[38] In contrast, the remaining four columns demonstrate that it is hard to explain the ratio of employment to residents (L_i/R_i) using these same empirical controls. The level of residents, wages, land area, developed-land supply elasticities, employment, and measures of economic activity in surrounding CZs, do a poor job in accounting for the variation in this ratio. None of the R-squared's in the last four columns of Table C.5 amounts to more than

[38]In these specifications in Table C.1, we follow our baseline approach to incorporating the Saiz (2010) MSA-level estimates of developed land supply elasticities (as discussed in Subsection 4.3 of the paper), but we find an almost identical results using our alternative approach (as discussed in Subsection C.7 of this web appendix).

	1	2	3	4	5	6	7	8	9
Dependent Variable:					Elasticity of Employment				
$\log L_i$		0.025**	0.044**	0.002				0.057**	0.055**
		(0.008)	(0.010)	(0.016)				(0.012)	
$\log w_i$			-0.037	-0.168				-0.002	-0.020
			(0.090)	(0.108)				(0.073)	
$\log H_i$			-0.166**	-0.087**				-0.010	-0.011
			(0.016)	(0.028)				(0.018)	
$\log L_{,-i}$				0.081**				-0.038**	-0.040**
				(0.017)					
$\log \bar{w}_{-i}$				0.107				0.012	0.036
				(0.121)					
$\lambda_{ii\mid i}$					-3.434**				
$\sum_{n \in N} \left(1 - \lambda_{ni\mid n}\right) \vartheta_{ni}$						8.815**		9.936**	
						(2.168)		(2.599)	
$\vartheta_{ii} \left(\frac{\lambda_{ii}}{\lambda_{Ri}} - \lambda_{Li}\right)$						6.044**		6.670**	
						(2.151)		(2.578)	
$\frac{\partial w_i}{\partial A_i} \frac{A_i}{w_i}$						-1.624**		-0.997**	
						(0.161)		(0.246)	
$\frac{\partial w_i}{\partial A_i} \frac{A_i}{w_i} \cdot \sum_{r \in N} \left(1 - \lambda_{rn\mid r}\right) \vartheta_{rn}$							1.345**		2.546**
							(0.129)		
$\frac{\partial w_i}{\partial A_i} \frac{A_i}{w_i} \cdot \vartheta_{ii} \left(\frac{\lambda_{ii}}{\lambda_{Ri}} - \lambda_{Li}\right)$							-1.391**		-0.680**
							(0.124)		
Constant	1.376**	1.098**	2.779**	1.747	4.522**	-3.459	2.347**	-4.959	1.387
	(0.013)	(0.091)	(0.850)	(1.249)	(0.111)	(2.054)	(0.118)	(2.649)	(0.901)
R^2	0.00	0.01	0.15	0.27	0.54	0.60	0.59	0.69	0.68
N	709	709	709	636	709	709	709	636	636

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 C.4: Explaining the general equilibrium local employment elasticities to a 5 percent productivity shock for commuting zones (CZs)

one third. Therefore, as with our earlier results for counties, we find that there is substantial additional information in patterns of commuting that is not captured by the standard empirical controls from the local labor markets literature.

Taking the results of this section as a whole, we find that the heterogeneity in commuting linkages across commuting zones (CZs) is sufficient to generate substantial heterogeneity in local employment elasticities, in response to either productivity shocks or reductions in commuting costs. This heterogeneity is hard to explain with the standard empirical controls from the local labor markets literature, but is well explained by measures of commuting linkages, highlighting the importance of incorporating this commuting information into the analysis of regional economies.

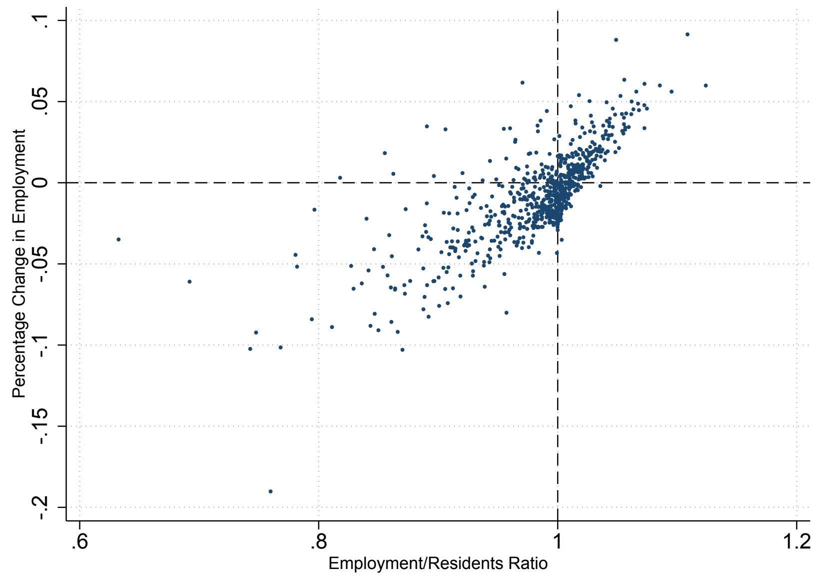


Figure C.14: Counterfactual relative change in commuting zone (CZ) employment ($\hat{L}$) from median proportional reduction in commuting costs ($\hat{\mathcal{B}}_{ni} = 0.88$) and initial dependence on commuting

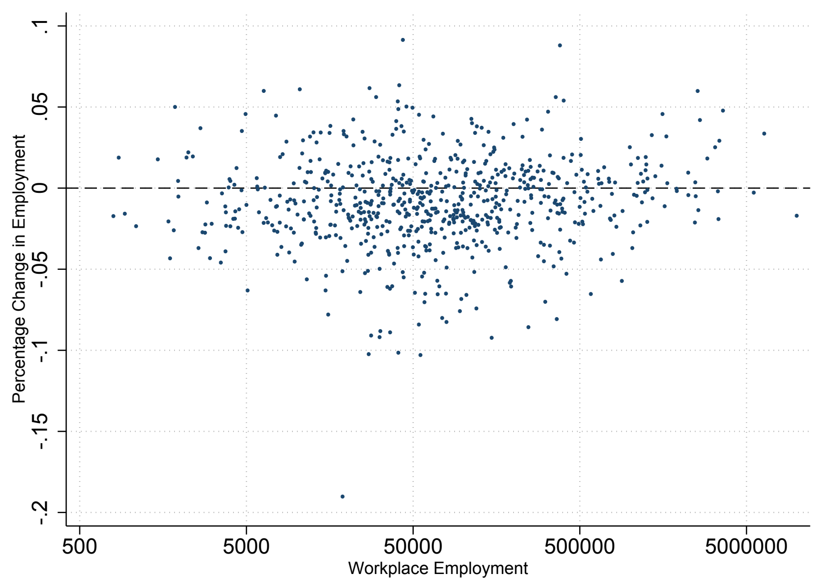


Figure C.15: Counterfactual relative change in commuting zone (CZ) employment ($\hat{L}$) from median proportional reduction in commuting costs ($\hat{\mathcal{B}}_{ni} = 0.88$) and initial employment size

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.

	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
$\log R_i$	0.991**		0.995**		-0.004*		0.002	
	(0.002)		(0.003)		(0.002)		(0.003)	
$\log w_i$	0.116**		0.167**			0.099**		0.128**
	(0.015)		(0.019)			(0.014)		(0.017)
$\log L_i$		1.001**		0.991**		-0.006**		0.001
		(0.002)		(0.003)		(0.002)		(0.003)
$\log \bar{v}_i$		-0.042**		-0.067**	0.057**		0.085**	
		(0.016)		(0.022)	(0.015)		(0.019)	
$\log H_i$			0.005	0.006			-0.000	-0.002
			(0.005)	(0.005)			(0.005)	(0.004)
$\log R_{,-i}$			-0.001				0.690**	0.633**
			(0.003)				(0.114)	(0.110)
$\log \bar{w}_{-i}$			-0.150**				0.982**	0.960**
			(0.021)				(0.218)	(0.209)
Saiz elasticity			-0.012**	0.012**			-0.009**	-0.011**
			(0.004)	(0.004)			(0.003)	(0.003)
$\log L_{,-i}$				0.011**			-0.697**	-0.639**
				(0.003)			(0.115)	(0.110)
$\log \bar{v}_{-i}$				0.107**			-1.076**	-1.069**
				(0.024)			(0.218)	(0.209)
Constant	-1.146**	0.450**	-0.167	-0.490*	0.425**	0.007	1.149**	0.891**
	(0.144)	(0.158)	(0.219)	(0.230)	(0.144)	(0.135)	(0.210)	(0.202)
R^2	0.99	0.99	0.99	0.99	0.02	0.07	0.27	0.32
N	709	709	636	636	709	709	636	636

* $p < 0.05$; ** $p < 0.01$

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 C.5: Explaining employment levels and commuting intensity for commuting zones (CZs)

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. Land area is geographical land area. When we need to aggregate counties (see below), the geographical land area is the sum of that for the aggregated counties, and the centroid of the new county formed by the aggregation is computed using spatial analysis software. In Subsection 4.3 of the paper, we develop an extension to allow for a heterogeneous positive supply elasticity for developed land following Saiz (2010).

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 county employment in manufacturing. 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-disclosure 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 relationship 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.[39] 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 paper 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.[40]

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

[39] 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.

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

Statistics[41] 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 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 paper 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 Section 3.1 of the paper. 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 C.1. 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.

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

References

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

Grubel, Herbert G. and Peter Lloyd (1975) *Intra-Industry Trade: The Theory and Measurement of International Trade in Differentiated Products*, London: MacMillan.

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

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.

Saiz, Albert (2010) “The Geographic Determinants of Housing Supply,” *Quarterly Journal of Economics*, 125(3), 1253-1296.

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