

Online Appendix for “Commuting, Migration and Local Employment Elasticities” (Not for Publication)*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 for the quantitative analysis of the model. Section C includes additional empirical results and robustness tests. Section D presents further information about the data definitions and sources.

B Quantitative Model Appendix

The first seven sections of this quantitative part of the web appendix present additional 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. Section B.3 provides conditions for the existence and uniqueness of the general equilibrium. Section B.4 derives isomorphisms to other trade models with commuting and external economies of scale. Section B.5 shows that unobserved productivity can be uniquely determined from the observed variables and reports additional evidence on gravity in goods trade. Section B.6 shows that unobserved amenities can be uniquely recovered from the observed data and reports additional evidence on gravity in commuting. Section B.7 uses the commuter market clearing condition to show the relationship between different measures of the openness of the local labor market to commuting.

The remaining sections comprise supplementary material and extensions. Section B.8 reports the derivation of the partial equilibrium local employment elasticities discussed in the main paper. Section B.9 shows that the class of models consistent with a gravity equation for commuting flows implies heterogeneous local employment elasticities. Section B.10 introduces multiple worker types. Section B.11 introduces

*Much of this research was undertaken while Ferdinando Monte was visiting the International Economics Section (IES) at Princeton. We are grateful to the IES and Princeton more generally for research support. We are also grateful to the editor, four anonymous referees, and conference and seminar participants for helpful comments and suggestions. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

[†]McDonough School of Business, 37th and O Streets, NW, Washington, DC 20057. ferdinando.monte@georgetown.edu.

[‡]Dept. Economics and WWS, Princeton, NJ 08544. 609 258 4016. reddings@princeton.edu.

[§]Dept. Economics and WWS, Princeton, NJ 08544. 609 258 4024. erossi@princeton.edu.

congestion in commuting. Section B.12 develops an extension of the baseline model to incorporate non-traded consumption goods. Section B.13 considers the case where landlords use residential land. Section B.14 generalizes the production technology to incorporate intermediate inputs, commercial land use and capital. Section B.15 introduces heterogeneity in effective units of labor. Section B.16 considers the case where commuting costs are incurred in effective units of labor rather than in utility. Finally, Section B.17 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:

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

Given this Fréchet distribution for utility, expected utility is:

$$\mathbb{E}[u] = \int_0^\infty \epsilon \Psi u^{-\epsilon} e^{-\Psi u^{-\epsilon}} du. \quad (\text{B.2})$$

Now define the following change of variables:

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

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

$$\mathbb{E}[u] = \int_0^\infty \Psi^{1/\epsilon} y^{-1/\epsilon} e^{-y} dy, \quad (\text{B.4})$$

which can be in turn written as:

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

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

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

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:

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

Note that:

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

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:

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

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:

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

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:

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

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):

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

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:

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

which simplifies to:

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

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):

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

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:

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

which simplifies to:

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

These conditional commuting probabilities provide microeconomic foundations for the reduced-form gravity equations estimated in the empirical literature on commuting patterns.¹ 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

¹See 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).

to all other workplaces s in the denominator (“multilateral resistance”).

Commuter 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 commuting to i ($\lambda_{ni|n}^R$):

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

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

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

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

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

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

$$\begin{aligned} &= \frac{1}{\lambda_{ni}} \int_0^u \prod_{s \neq i} G_{ns}(u) \left[\prod_{r \neq n} \prod_s G_{rs}(u) \right] g_{ni}(u) du, \\ &= \frac{1}{\lambda_{ni}} \int_0^u \left[\prod_{r=1}^S \prod_{s=1}^S e^{-\Psi_{rs} u^{-\epsilon}} \right] \epsilon \Psi_{ni} u^{-(\epsilon+1)} du, \\ &= \frac{\Psi}{\Psi_{ni}} \int_0^u e^{-\Psi u^{-\epsilon}} \epsilon \Psi_{ni} u^{-(\epsilon+1)} du, \\ &= e^{-\Psi u^\epsilon}. \end{aligned} \tag{B.15}$$

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 ($\hat{x} = x'/x$). 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 $\{\hat{w}_n, \hat{v}_n, \hat{Q}_n, \hat{\pi}_{ni}, \hat{\lambda}_{ni}, \hat{P}_n, \hat{R}_n, \hat{L}_n\}$ from the following system of eight equations (using the iterative algorithm outlined below):

$$\hat{w}_i \hat{L}_i w_i L_i = \sum_{n \in N} \pi_{ni} \hat{\pi}_{ni} \hat{v}_n \hat{R}_n \bar{v}_n R_n, \quad (\text{B.16})$$

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

$$\hat{Q}_n = \hat{v}_n \hat{R}_n, \quad (\text{B.18})$$

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

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

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

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

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

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 equilib-

rium. Given the model's parameters $\{\alpha, \sigma, \epsilon, \delta, \kappa\}$ and changes in the exogenous variables of the model $\{\hat{A}_n, \hat{B}_n, \hat{\kappa}_{ni}, \hat{d}_{ni}\}$, we can solve for the resulting counterfactual changes in the endogenous variables of the model $\{\hat{w}_n, \hat{v}_n, \hat{Q}_n, \hat{\pi}_{ni}, \hat{\lambda}_{ni}, \hat{P}_n, \hat{R}_n, \hat{L}_n\}$ 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 $\hat{w}_i^{(t+1)}$ and $\hat{\lambda}_{ni}^{(t+1)}$ using the current guesses and data. We start by computing:

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

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

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

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

$$\hat{Q}_n^{(t)} = \hat{v}_n^{(t)} \hat{R}_n^{(t)}. \quad (\text{B.27})$$

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

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

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

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

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

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

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

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

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

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

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

In Section B.3 below, we provide conditions under which the counterfactual equilibrium of this economy is unique.

B.3 Existence and Uniqueness

We now provide conditions for the existence and uniqueness of a general equilibrium of this economy.

B.3.1 Workplace and Residence Income

From the commuting probabilities in equation (10) in the paper, the labor income received by commuters from residence n to workplace i is:

$$w_i \lambda_{ni} \bar{L} = \left(\frac{\bar{U}}{\delta} \right)^{-\epsilon} \bar{L} B_{ni} \kappa_{ni}^{-\epsilon} (P_n^\alpha Q_n^{1-\alpha})^{-\epsilon} w_i^{1+\epsilon}. \quad (\text{B.34})$$

Summing across residences n , total workplace income in location i is:

$$Y_i = \sum_{n \in N} w_i \lambda_{ni} \bar{L} = \left(\frac{\bar{U}}{\delta} \right)^{-\epsilon} \bar{L} w_i^{1+\epsilon} \sum_{n \in N} B_{ni} \kappa_{ni}^{-\epsilon} (P_n^\alpha Q_n^{1-\alpha})^{-\epsilon}, \quad (\text{B.35})$$

Summing across workplaces i , total residence income in location n (equals total expenditure in residence n) is given by:

$$X_n = \sum_{i \in N} w_i \lambda_{ni} \bar{L} = \left(\frac{\bar{U}}{\delta} \right)^{-\epsilon} \bar{L} (P_n^\alpha Q_n^{1-\alpha})^{-\epsilon} \sum_{i \in N} B_{ni} \kappa_{ni}^{-\epsilon} w_i^{1+\epsilon}. \quad (\text{B.36})$$

Now note that land market clearing in equation (5) in the paper can be written as:

$$Q_n = (1 - \alpha) \frac{X_n}{H_n}. \quad (\text{B.37})$$

Using land market clearing (B.37), total workplace income in location i (B.35) can be re-written as:

$$Y_i = \left(\frac{\bar{U}}{\delta} \right)^{-\epsilon} \bar{L} (1 - \alpha)^{-(1-\alpha)\epsilon} w_i^{1+\epsilon} \sum_{n \in N} B_{ni} \kappa_{ni}^{-\epsilon} P_n^{-\alpha\epsilon} X_n^{-(1-\alpha)\epsilon}. \quad (\text{B.38})$$

Using land market clearing (B.37), total residence income in location n (B.36) can be re-written as:

$$X_n = \left(\frac{\bar{U}}{\delta} \right)^{-\epsilon} \bar{L} (1 - \alpha)^{-(1-\alpha)\epsilon} H_n^{(1-\alpha)\epsilon} P_n^{-\alpha\epsilon} X_n^{-\epsilon(1-\alpha)} \sum_{i \in N} B_{ni} \kappa_{ni}^{-\epsilon} w_i^{1+\epsilon}. \quad (\text{B.39})$$

B.3.2 Price Index and Goods Market Clearing

Using $Y_i = w_i L_i$, the price index in equation (8) in the paper can be re-written as:

$$P_n^{1-\sigma} = \left(\frac{\sigma}{\sigma - 1} \right)^{1-\sigma} \frac{1}{\sigma F} \left[\sum_{i \in N} Y_i \left(\frac{d_{ni}}{A_i} \right)^{1-\sigma} w_i^{-\sigma} \right]. \quad (\text{B.40})$$

Similarly, using $Y_i = w_i L_i$ and $X_n = \bar{v}_n R_n$, the goods market clearing condition in equation (7) in the paper can be re-written as:

$$Y_i = \sum_{n \in N} \frac{Y_i}{\sigma F} \left(\frac{\sigma}{\sigma - 1} \right)^{1-\sigma} \left(\frac{d_{ni}}{A_i} \right)^{1-\sigma} w_i^{-\sigma} P_n^{\sigma-1} X_n,$$

which simplifies to:

$$w_i^\sigma = \sum_{n \in N} \frac{1}{\sigma F} \left(\frac{\sigma}{\sigma - 1} \right)^{1-\sigma} \left(\frac{d_{ni}}{A_i} \right)^{1-\sigma} P_n^{\sigma-1} X_n. \quad (\text{B.41})$$

B.3.3 System of Equations

Combining workplace income (B.38), residence income (B.39), the price index (B.40), and goods market clearing (B.41), we obtain the following system of equations:

$$P_n^{1-\sigma} = \xi^P \sum_{i \in N} \mathcal{K}_{ni}^P Y_i w_i^{-\sigma}, \quad (\text{B.42})$$

$$w_n^\sigma = \xi^w \sum_{i \in N} \mathcal{K}_{ni}^w P_i^{\sigma-1} X_i, \quad (\text{B.43})$$

$$Y_n w_n^{-(1+\epsilon)} = \xi^Y \sum_{i \in N} \mathcal{K}_{ni}^Y P_i^{-\alpha\epsilon} X_i^{-(1-\alpha)\epsilon}, \quad (\text{B.44})$$

$$X_n^{1+\epsilon(1-\alpha)} P_n^{\alpha\epsilon} = \xi^X \sum_{i \in N} \mathcal{K}_{ni}^X w_i^{1+\epsilon}, \quad (\text{B.45})$$

where we have assumed symmetric trade costs ($d_{ni} = d_{in}$) and commuting costs ($B_{ni} \kappa_{ni}^{-\epsilon} = B_{in} \kappa_{in}^{-\epsilon}$); we have defined the following scalars:

$$\begin{aligned} \xi^P &\equiv \left(\frac{\sigma}{\sigma - 1} \right)^{1-\sigma} \frac{1}{\sigma F}, \\ \xi^w &\equiv \frac{1}{\sigma F} \left(\frac{\sigma}{\sigma - 1} \right)^{1-\sigma}, \\ \xi^Y &\equiv \left(\frac{\bar{U}}{\delta} \right)^{-\epsilon} \bar{L} (1 - \alpha)^{-(1-\alpha)\epsilon}, \\ \xi^X &\equiv \left(\frac{\bar{U}}{\delta} \right)^{-\epsilon} \bar{L} (1 - \alpha)^{-(1-\alpha)\epsilon}; \end{aligned}$$

and we have defined the following kernels:

$$\begin{aligned} \mathcal{K}_{ni}^P &\equiv \left(\frac{d_{ni}}{A_i} \right)^{1-\sigma}, \\ \mathcal{K}_{ni}^w &\equiv \left(\frac{d_{ni}}{A_i} \right)^{1-\sigma}, \\ \mathcal{K}_{ni}^Y &\equiv B_{ni} \kappa_{ni}^{-\epsilon} H_n^{(1-\alpha)\epsilon}, \end{aligned}$$

$$\mathcal{K}_{ni}^X \equiv B_{ni} \kappa_{ni}^{-\epsilon} H_n^{(1-\alpha)\epsilon}.$$

Note that equations (B.42)-(B.45) take the same form as the class of gravity equation models considered in Allen, Arkolakis and Li (2016). In particular, there are H vectors of endogenous variables $x^h \in \mathfrak{R}^N$, $h = 1, \dots, H$, and each vector, x^h , contains the endogenous variables for the I locations, $x_i^h \in \mathfrak{R}$, $i = 1, \dots, I$. Using this notation, and denoting the corresponding sets of endogenous variables and locations by Ω^H and Ω^N respectively, the system of equations (B.42)-(B.45) can be written as:

$$\prod_{h=1}^H \left(x_i^h\right)^{\beta_{kh}} = \xi^k \sum_{n=1}^I \mathcal{K}_{ni}^k \left[\prod_{h=1}^H \left(x_n^h\right)^{\gamma_{kh}} \right], \quad i \in \Omega^N, \quad k, h \in \Omega^H,$$

where the characteristic values $\xi^k \in \mathfrak{R}$ are endogenous scalars that balance the overall level of the two sides of the equations; the parameters are $\beta_{kh}, \gamma_{kh} \in \mathfrak{R}$; and \mathcal{K}_{ni}^k is the kernel that regulates interactions across locations, variables and equations.

We denote B and Γ as the $H \times H$ matrices, whose elements $(B)_{kh} = \beta_{kh}$ and $(\Gamma)_{kh} = \gamma_{kh}$ are the parameters from the left and right-hand sides of these equations, respectively. From equations (B.42)-(B.45), we have:

$$B = \begin{bmatrix} 1 - \sigma & 0 & 0 & 0 \\ 0 & \sigma & 0 & 0 \\ 0 & -(1 + \epsilon) & 1 & 0 \\ \alpha\epsilon & 0 & 0 & 1 + (1 - \alpha)\epsilon \end{bmatrix},$$

$$\Gamma = \begin{bmatrix} 0 & -\sigma & 1 & 0 \\ \sigma - 1 & 0 & 0 & 1 \\ -\alpha\epsilon & 0 & 0 & -(1 - \alpha)\epsilon \\ 0 & 1 + \epsilon & 0 & 0 \end{bmatrix}.$$

Note that all elements of the kernel \mathcal{K}_{ni}^k are strictly positive. Additionally, both B and Γ are invertible, and we denote A as the following composite matrix:

$$A = \Gamma B^{-1} = \begin{bmatrix} 0 & \frac{1}{\sigma}(\epsilon + 1) - 1 & 1 & 0 \\ \alpha \frac{\epsilon}{(\sigma-1)(\epsilon-\alpha\epsilon+1)} - 1 & 0 & 0 & \frac{1}{\epsilon-\alpha\epsilon+1} \\ \alpha \frac{\epsilon}{\sigma-1} + \alpha\epsilon^2 \frac{\alpha-1}{(\sigma-1)(\epsilon-\alpha\epsilon+1)} & 0 & 0 & \epsilon \frac{\alpha-1}{\epsilon-\alpha\epsilon+1} \\ 0 & \frac{1}{\sigma}(\epsilon + 1) & 0 & 0 \end{bmatrix}.$$

We also denote A^p as the matrix whose elements equal the absolute value of the elements of A , such that $(A^p)_{kh} = |(A)_{kh}|$, and define $\rho(A^p)$ as the largest eigenvalue of A^p . Applying Theorem 3 of Allen, Arkolakis and Li (2016), a sufficient condition for the equilibrium of the economy to be unique is $\rho(A^p) \leq 1$. Having pinned down unique equilibrium values of $\{P_n, w_n, Y_n, X_n\}$, all other endogenous variables of the model can be uniquely determined.

B.4 Isomorphisms

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:

$$w_i L_i = \sum_{n \in N} \frac{L_i (d_{ni} w_i / A_i)^{1-\sigma}}{\sum_{k \in N} L_k (d_{nk} w_k / A_k)^{1-\sigma}} \bar{v}_n R_n. \quad (\text{B.46})$$

Second, expected worker income depends on wages:

$$\bar{v}_n = \sum_{i \in N} \frac{B_{ni} (w_i / \kappa_{ni})^\epsilon}{\sum_{s \in N} B_{ns} (w_s / \kappa_{ns})^\epsilon} w_i. \quad (\text{B.47})$$

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

$$Q_n = (1 - \alpha) \frac{\bar{v}_n R_n}{H_n}. \quad (\text{B.48})$$

Fourth, workplace choice probabilities solve:

$$\frac{L_n}{\bar{L}} = \frac{\sum_{r \in N} B_{rn} (\kappa_{rn} P_r^\alpha Q_r^{1-\alpha})^{-\epsilon} w_n^\epsilon}{\sum_{r \in N} \sum_{s \in N} B_{rs} (\kappa_{rs} P_r^\alpha Q_r^{1-\alpha})^{-\epsilon} w_s^\epsilon}. \quad (\text{B.49})$$

Fifth, residential choice probabilities solve:

$$\frac{R_n}{\bar{L}} = \frac{\sum_{s \in N} B_{ns} (\kappa_{ns} P_n^\alpha Q_n^{1-\alpha})^{-\epsilon} w_s^\epsilon}{\sum_{r \in N} \sum_{s \in N} B_{rs} (\kappa_{rs} P_r^\alpha Q_r^{1-\alpha})^{-\epsilon} w_s^\epsilon}. \quad (\text{B.50})$$

Sixth, price indices solve:

$$P_n = \frac{\sigma}{\sigma - 1} \left(\frac{1}{\sigma F} \right)^{\frac{1}{1-\sigma}} \left[\sum_{i \in N} L_i (d_{ni} w_i / A_i)^{1-\sigma} \right]^{\frac{1}{1-\sigma}}. \quad (\text{B.51})$$

Seventh, expected utility satisfies:

$$\bar{U} = \delta \left[\sum_{r \in N} \sum_{s \in N} B_{rs} (\kappa_{rs} P_r^\alpha Q_r^{1-\alpha})^{-\epsilon} w_s^\epsilon \right]^{\frac{1}{\epsilon}}, \quad (\text{B.52})$$

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 equation (1) in the paper, 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:

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

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:

$$w_i L_i = \sum_{n \in N} \frac{\tilde{A}_i L_i^\eta (d_{ni} w_i)^{-\theta}}{\sum_{k \in N} \tilde{A}_k L_k^\eta (d_{nk} w_k)^{-\theta}} \bar{v}_n R_n. \quad (\text{B.53})$$

Second, expected worker income depends on wages:

$$\bar{v}_n = \sum_{i \in N} \frac{B_{ni} (w_i / \kappa_{ni})^\epsilon}{\sum_{s \in N} B_{ns} (w_s / \kappa_{ns})^\epsilon} w_i. \quad (\text{B.54})$$

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

$$Q_n = (1 - \alpha) \frac{\bar{v}_n R_n}{H_n}. \quad (\text{B.55})$$

Fourth, workplace choice probabilities solve:

$$\frac{L_n}{\bar{L}} = \frac{\sum_{r \in N} B_{rn} (\kappa_{rn} P_r^\alpha Q_r^{1-\alpha})^{-\epsilon} w_n^\epsilon}{\sum_{r \in N} \sum_{s \in N} B_{rs} (\kappa_{rs} P_r^\alpha Q_r^{1-\alpha})^{-\epsilon} w_s^\epsilon}. \quad (\text{B.56})$$

Fifth, residential choice probabilities solve:

$$\frac{R_n}{\bar{L}} = \frac{\sum_{s \in N} B_{ns} (\kappa_{ns} P_n^\alpha Q_n^{1-\alpha})^{-\epsilon} w_s^\epsilon}{\sum_{r \in N} \sum_{s \in N} B_{rs} (\kappa_{rs} P_r^\alpha Q_r^{1-\alpha})^{-\epsilon} w_s^\epsilon}. \quad (\text{B.57})$$

Sixth, price indices solve:

$$P_n = \gamma \left[\sum_{i \in N} \tilde{A}_i L_i^\eta (d_{ni} w_i)^{-\theta} \right]^{-\frac{1}{\theta}}, \quad (\text{B.58})$$

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:

$$\bar{U} = \delta \left[\sum_{r \in N} \sum_{s \in N} B_{rs} (\kappa_{rs} P_r^\alpha Q_r^{1-\alpha})^{-\epsilon} w_s^\epsilon \right]^{\frac{1}{\epsilon}}. \quad (\text{B.59})$$

The system of equations (B.53)-(B.59) is isomorphic to the system of equations (B.46)-(B.52) under the following parameter restrictions:

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

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 equation (1) in the paper, 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 a good produced in location i and consumed in location n is:

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

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:

$$w_i L_i = \sum_{n \in N} \frac{A_i^{\sigma-1} L_i^{\eta(\sigma-1)} (d_{ni} w_i)^{1-\sigma}}{\sum_{k \in N} A_k^{\sigma-1} L_k^{\eta(\sigma-1)} (d_{nk} w_k)^{1-\sigma}} \bar{v}_n R_n. \quad (\text{B.60})$$

Second, expected worker income depends on wages:

$$\bar{v}_n = \sum_{i \in N} \frac{B_{ni} (w_i / \kappa_{ni})^\epsilon}{\sum_{s \in N} B_{ns} (w_s / \kappa_{ns})^\epsilon} w_i. \quad (\text{B.61})$$

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

$$Q_n = (1 - \alpha) \frac{\bar{v}_n R_n}{H_n}. \quad (\text{B.62})$$

Fourth, workplace choice probabilities solve:

$$\frac{L_n}{\bar{L}} = \frac{\sum_{r \in N} B_{rn} (\kappa_{rn} P_r^\alpha Q_r^{1-\alpha})^{-\epsilon} w_n^\epsilon}{\sum_{r \in N} \sum_{s \in N} B_{rs} (\kappa_{rs} P_r^\alpha Q_r^{1-\alpha})^{-\epsilon} w_s^\epsilon}. \quad (\text{B.63})$$

Fifth, residential choice probabilities solve:

$$\frac{R_n}{\bar{L}} = \frac{\sum_{s \in N} B_{ns} (\kappa_{ns} P_n^\alpha Q_n^{1-\alpha})^{-\epsilon} w_s^\epsilon}{\sum_{r \in N} \sum_{s \in N} B_{rs} (\kappa_{rs} P_r^\alpha Q_r^{1-\alpha})^{-\epsilon} w_s^\epsilon}. \quad (\text{B.64})$$

Sixth, price indices solve:

$$P_n = \left[\sum_{i \in N} A_i^{\sigma-1} L_i^{\eta(\sigma-1)} (d_{ni} w_i)^{1-\sigma} \right]^{\frac{1}{1-\sigma}}. \quad (\text{B.65})$$

Seventh, expected utility satisfies:

$$\bar{U} = \delta \left[\sum_{r \in N} \sum_{s \in N} B_{rs} (\kappa_{rs} P_r^\alpha Q_r^{1-\alpha})^{-\epsilon} w_s^\epsilon \right]^{\frac{1}{\epsilon}}. \quad (\text{B.66})$$

The system of equations (B.60)-(B.66) is isomorphic to the system of equations (B.46)-(B.52) 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 Gravity in Goods Trade

As discussed in Section 3.1 of the paper, we use the equality between income and expenditure in equation (7) in the paper to solve for unobserved county productivities (A_i):

$$w_i L_i - \sum_{n \in N} \frac{L_i (d_{ni} w_i / A_i)^{1-\sigma}}{\sum_{k \in N} L_k (d_{nk} w_k / A_k)^{1-\sigma}} [\bar{v}_n R_n + D_n] = 0, \quad (\text{B.67})$$

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

Given the elasticity of substitution (σ), our measures for ($w_i, L_i, \bar{v}_i, R_i, D_i$) and a parameterization of trade costs ($d_{ni}^{1-\sigma}$), equation (B.67) provides a system of N equations that can be solved for a unique vector of N unobserved productivities (A_i), as summarized in the following proposition.

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

Proof. Note that the goods market clearing condition (B.67) can be written as the following excess demand system:

$$\mathbb{D}_i(\tilde{\mathbf{A}}) = w_i L_i - \sum_{n \in N} \frac{\tilde{A}_i L_i (d_{ni} w_i)^{1-\sigma}}{\sum_{k \in N} \tilde{A}_k L_k (d_{nk} w_k)^{1-\sigma}} [\bar{v}_n R_n + D_n] = 0, \quad (\text{B.68})$$

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.68).

Property (ii): $\mathbb{D}(\tilde{\mathbf{A}})$ is homogenous of degree zero, as follows immediately from inspection of (B.68).

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

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

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

$$\begin{aligned} \frac{\partial \mathbb{D}_i(\tilde{\mathbf{A}})}{\partial \tilde{A}_r} &> 0 \quad \text{for all } i, r, \neq i, & \text{for all } \tilde{\mathbf{A}} \in \mathfrak{R}_+^N, \\ \frac{\partial \mathbb{D}_i(\tilde{\mathbf{A}})}{\partial \tilde{A}_i} &< 0 \quad \text{for all } i, & \text{for all } \tilde{\mathbf{A}} \in \mathfrak{R}_+^N. \end{aligned}$$

This property can be established by noting:

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

and using homogeneity of degree zero, which implies:

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

and hence:

$$\frac{\partial \mathbb{D}_i(\tilde{\mathbf{A}})}{\partial \tilde{A}_i} < 0 \quad \text{for all } \tilde{\mathbf{A}} \in \mathfrak{R}_+^N.$$

Therefore we have established gross substitution. We now use these five properties to establish that

the system of equations (B.68) has at most one (normalized) solution. Gross substitution implies that $\mathbb{D}(\tilde{\mathbf{A}}) = \mathbb{D}(\tilde{\mathbf{A}}')$ 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 $\mathbb{D}(\tilde{\mathbf{A}}) > \mathbb{D}(\tilde{\mathbf{A}}')$ and we have a contradiction.

We next establish that there exists a productivity vector $\tilde{\mathbf{A}}^* \in \mathfrak{R}_+^N$ 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 $\Delta = \left\{ \tilde{\mathbf{A}} \in \mathfrak{R}_+^N : \sum_{i \in N} \tilde{A}_i = 1 \right\}$. Define on Δ the function $\mathbb{D}^+(\cdot)$ by $\mathbb{D}_i^+(\tilde{\mathbf{A}}) = \max \left\{ \mathbb{D}_i(\tilde{\mathbf{A}}), 0 \right\}$. Note that $\mathbb{D}^+(\cdot)$ is continuous. Denote $\alpha(\tilde{\mathbf{A}}) = \sum_{i \in N} \left[\tilde{A}_i + \mathbb{D}_i^+(\tilde{\mathbf{A}}) \right]$. We have $\alpha(\tilde{\mathbf{A}}) \geq 1$ for all $\tilde{\mathbf{A}}$.

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

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

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 $\tilde{\mathbf{A}}^* = f(\tilde{\mathbf{A}}^*)$.

Since $\sum_{i \in N} \mathbb{D}_i(\tilde{\mathbf{A}}) = 0$, it cannot be the case that $\mathbb{D}_i(\tilde{\mathbf{A}}) > 0$ for all $i \in N$ or $\mathbb{D}_i(\tilde{\mathbf{A}}) < 0$ for all $i \in N$. Additionally, if $\mathbb{D}_i(\tilde{\mathbf{A}}) > 0$ for some i and $\mathbb{D}_r(\tilde{\mathbf{A}}) < 0$ for some $r \neq i$, $\tilde{\mathbf{A}} \neq f(\tilde{\mathbf{A}})$. It follows that at the fixed point for productivity, $\tilde{\mathbf{A}}^* = f(\tilde{\mathbf{A}}^*)$, and $\mathbb{D}_i(\tilde{\mathbf{A}}^*) = 0$ for all i . It follows that there exists a unique vector of unobserved productivities ($\tilde{\mathbf{A}}$) that solves the excess demand system (B.68). ■

The resulting solutions for productivities (A_i) capture characteristics (e.g. natural resources) that make a location more or less attractive for employment conditional on the observed data and the parameterized values of trade costs. These characteristics include access to international markets. To the extent that such international market access raises employment (L_i), and international trade flows are not captured in the CFS, this will be reflected in the model in higher productivity (A_i) to rationalize the higher observed employment. Having recovered these unique unobserved productivities (A_i), we can solve for the implied bilateral trade flows between counties (X_{ni}) using equation (6) and $X_{ni} = \pi_{ni} \bar{v}_n R_n$. We use these solutions for bilateral trade between counties in our counterfactuals for changes in the model's exogenous variables, as discussed in the paper.

To parameterize trade costs ($d_{ni}^{1-\sigma}$), we assume a central value for the elasticity of substitution between varieties from the existing empirical literature of $\sigma = 4$, which is in line with the estimates of this parameter using price and expenditure data in Broda and Weinstein (2006).² We model bilateral trade costs (d_{ni}) as a function of distance. For bilateral pairs with positive trade, we assume that bilateral trade costs are a constant elasticity function of distance and a stochastic error ($d_{ni} = dist_{ni}^\psi \tilde{e}_{ni}$). For bilateral pairs with zero trade, the model implies prohibitive trade costs ($d_{ni} \rightarrow \infty$).³ Taking logarithms in the trade share in

²This assumed value implies an elasticity of trade with respect to trade costs of $-(\sigma - 1) = 3$, which is close to the central estimate of this parameter of 4.12 in Simonovska and Waugh (2014).

³One interpretation is that trade requires prior investments in transport infrastructure that are not modeled here. For bilateral pairs for which these investments have been made, trade can occur subject to finite costs. For other bilateral pairs

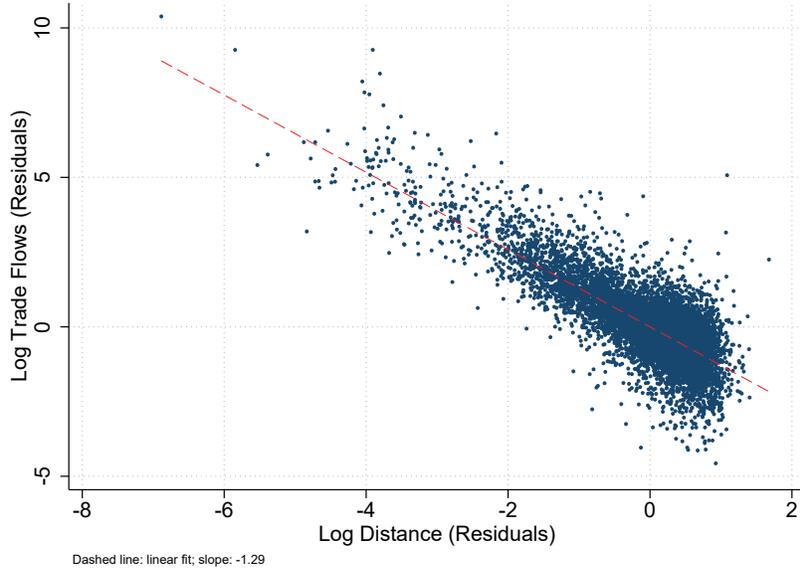


Figure B.1: Gravity in Goods Trade Between CFS Regions

equation (6) in the paper for pairs with positive trade, the value of bilateral trade between source i and destination n (X_{ni}) can be expressed as

$$\log X_{ni} = \zeta_n + \chi_i - (\sigma - 1) \psi \log dist_{ni} + \log e_{ni}, \quad (\text{B.69})$$

where the source fixed effect (χ_i) controls for employment, wages and productivity (L_i, w_i, A_i); the destination fixed effect (ζ_n) controls for average income, \bar{v}_n , residents, R_n , and multilateral resistance (as captured in the denominator of equation (6) in the paper); and $\log e_{ni} = (1 - \sigma) \log \tilde{e}_{ni}$.

Estimating the gravity equation (B.69) for all bilateral pairs with positive trade using OLS, we find a regression R-squared of 0.83. In Figure B.1, we display the conditional relationship between the log value of trade and log distance, after removing source and destination fixed effects from both log trade and log distance. Consistent with the existing empirical trade literature, we find that the log linear functional form provides a good approximation to the data, with a tight and approximately linear relationship between the two variables. We estimate a coefficient on log distance of $-(\sigma - 1) \psi = -1.29$. For our assumed value of $\sigma = 4$, this implies an elasticity of trade costs with respect to distance of $\psi = 0.43$. The tight linear relationship in Figure B.1, makes us confident in this parametrization of trade costs as $d_{ni}^{1-\sigma} = dist_{ni}^{-1.29}$ as a way of using equation (B.67) to solve for unobserved productivities (A_i).

To provide an alternative check on our specification, we aggregate the model's predictions for trade between counties within pairs of CFS regions, and compare these predictions to the data in Figure B.2. The only way in which we used the data on trade between CFS regions was to estimate the distance elasticity $-(\sigma - 1) \psi = -1.29$. Given this distance elasticity, we use the goods market clearing condition (B.67) to solve for productivities and generate predictions for bilateral trade between counties and hence CFS regions, as discussed above. Therefore, the model's predictions and the data can differ from one another. Nonetheless, we find a strong and approximately log linear relationship between the model's predictions

for which they have not been made, trade is prohibitively costly. We adopt our specification for tractability, but other rationalizations for zero trade flows include non-CES preferences or granularity.

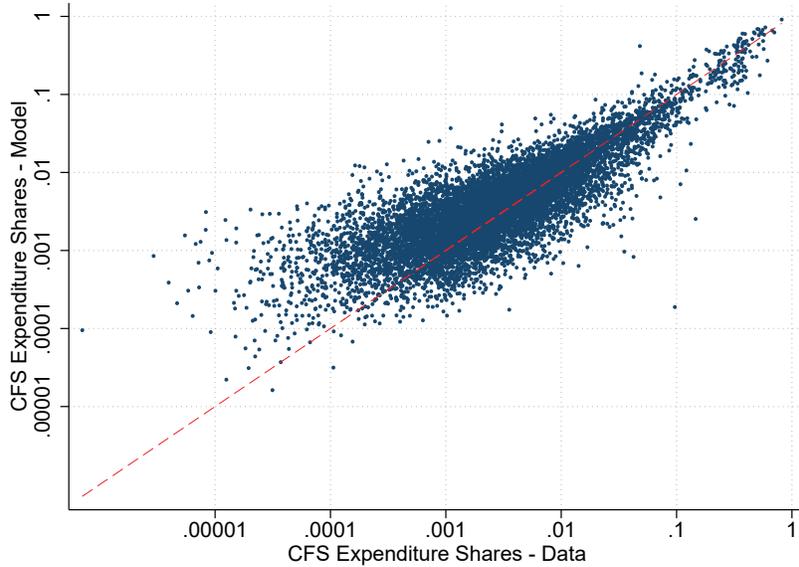


Figure B.2: Bilateral Trade Shares in the Model and Data

and the data, which is tighter for the larger trade values that account for most of aggregate trade.

B.6 Magnitude and Gravity of Commuting Flows

In this subsection of the web appendix, we provide additional evidence on the relevance of commuting as a source of spatial linkages between counties and CZs. In Figure 1 in the paper, we display unweighted kernel densities of the share of residents that work in the same county where they live (the “residence own commuting share”) over time. We focus on these unweighted kernel densities to capture heterogeneity across geographical locations (counties). As a robustness check, Figure B.3 in this web appendix displays analogous kernel densities that are weighted by the number of residents in each county. Therefore these weighted kernel densities capture heterogeneity across residents. As apparent from the two figures, we find a similar pattern of results whether we use the weighted or unweighted kernel densities. In both cases, we find a marked shift in density towards lower values of the residence own commuting share.

In Table 1 of the paper, we report unweighted descriptive statistics on commuting flows between counties and CZs from 2006-10. As a robustness check, Table B.1 in this web appendix displays reports analogous statistics that are weighted by the number of residents (or workers) in each county. Again the unweighted results capture heterogeneity across counties, while the weighted results capture heterogeneity across people. Whether we use the weighted or unweighted statistics, we find that commuting beyond county boundaries is both substantial and heterogeneous. For example, using the unweighted results, we find that for the median county around 27 percent of its residents work outside the county and around 20 percent of its workers live outside the county. By comparison, using the weighted results, we find that for the median county around 19 percent of its residents work outside the county and around 22 percent of its workers live outside the county.

In Section 3.2 of the paper, we discuss that these differences across counties in openness to commuting generate substantial variation in the ratio of employment to residents (L_i/R_i). In Table B.2 below, we show that this ratio of employment to residents is not only heterogeneous across counties, but is also hard

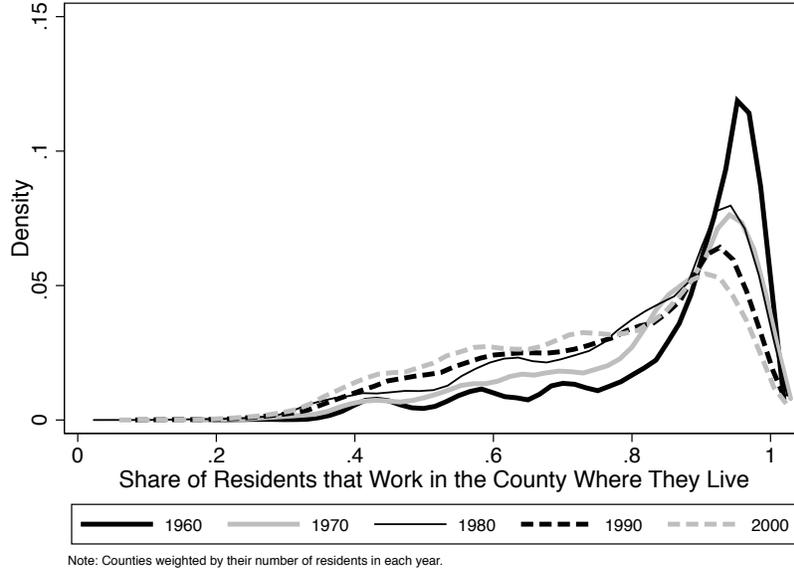


Figure B.3: Kernel densities of the share of residents that work in the county where they live (weighted by county residents)

	Min	p5	p10	p25	p50	p75	p90	p95	Max	Mean	N
Commuters from Residence County	0.00	0.00	0.04	0.08	0.19	0.38	0.51	0.56	0.82	0.24	3,111
Commuters to Workplace County	0.00	0.01	0.04	0.14	0.22	0.33	0.43	0.52	0.81	0.24	3,111
County Employment/Residents	0.26	0.64	0.73	0.88	1.00	1.10	1.21	1.30	3.88	1.00	3,111
Commuters from Residence CZ	0.00	0.00	0.00	0.02	0.03	0.07	0.14	0.18	0.49	0.06	709
Commuters to Employment CZ	0.00	0.00	0.00	0.02	0.05	0.08	0.11	0.13	0.25	0.06	709
CZ Employment/Residents	0.63	0.91	0.94	0.99	1.00	1.02	1.05	1.07	1.12	1.00	709

Tabulations on 3,111 counties and 709 commuting zones. The first row shows the fraction of residents that work outside the county. The second row shows the fraction of workers who live outside the county. The third row shows the ratio of county employment to county residents. The fourth row shows the fraction of a CZ's residents that work outside the CZ. The fifth row shows the fraction of a CZ's workers that live outside the CZ. The sixth row shows the ratio of CZ employment to CZ residents across all 709 CZs. p5, p10 etc refer to the 5th, 10th etc percentiles of the distribution. Results for commuters from residence are weighted by the number of residents. Results for commuters to workplace are weighted by the number of workers.

Table B.1: Commuting Across Counties and Commuting Zones (Weighted)

to explain 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 B.2 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. 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 B.2 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.

To examine the extent to which bilateral commuting flows are one-way versus two-way, we use the Grubel and Lloyd (1971) index from the international trade literature. In the context of commuting, this 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

$$GL_i = 1 - \frac{\left| \sum_{n \neq i} L_{in} - \sum_{n \neq i} L_{ni} \right|}{\sum_{n \neq i} L_{in} + \sum_{n \neq i} L_{ni}}, \quad (\text{B.70})$$

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 workplaces in other counties $n \neq i$ and $\sum_{n \neq i} L_{ni}$ is county i 's total imports of commuters from residences in other counties $n \neq i$. If there is only one-way commuting, $GL_n = 0$. 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 B.3, we report the mean and percentiles of the distribution of the Grubel-Lloyd index from equation (B.70) 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 ($B_{ni} \rightarrow 0$) and/or prohibitive commuting costs ($\kappa_{ni} \rightarrow \infty$), which can be used to explain one-way commuting.

	1	2	3	4	5	6	7	8	9	10	11	12
Dep. Variable:	$\log L_i$	$\log R_i$	$\log L_i$	$\log R_i$	L_i/R_i							
$\log R_i$	0.974** (0.005)		1.001** (0.009)			-0.000 (0.010)		0.020 (0.016)		0.064** (0.012)		0.050** (0.013)
$\log w_i$	0.460** (0.038)		0.480** (0.036)		0.341** (0.025)		0.331** (0.026)		0.468** (0.046)		0.479** (0.054)	
$\log L_i$		0.957** (0.013)		0.922** (0.015)	-0.001 (0.007)		0.028* (0.012)		0.044** (0.006)		0.049** (0.006)	
$\log \bar{v}_i$		0.066 (0.051)		0.019 (0.049)		0.171** (0.033)		0.239** (0.041)		0.287** (0.100)		0.273** (0.092)
$\log H_i$			0.015 (0.010)	0.037** (0.012)			-0.022 (0.011)	-0.011 (0.013)	-0.055** (0.012)	-0.067** (0.020)	-0.058** (0.013)	-0.059** (0.019)
$\log R_{,-i}$			-0.020* (0.008)				0.389* (0.160)	0.609** (0.171)	0.396 (0.407)	0.677 (0.524)	0.391 (0.406)	0.679 (0.516)
$\log \bar{w}_{-i}$			-0.330** (0.036)				0.070 (0.324)	0.247 (0.364)	-1.843 (1.326)	-2.619 (1.668)	-1.797 (1.308)	-2.654 (1.678)
$\log L_{,-i}$				0.084** (0.011)			-0.435** (0.155)	-0.654** (0.166)	-0.405 (0.408)	-0.694 (0.525)	-0.401 (0.408)	-0.691 (0.518)
$\log \bar{v}_{-i}$				0.044 (0.038)			-0.238 (0.315)	-0.347 (0.360)	1.523 (1.310)	2.410 (1.640)	1.482 (1.292)	2.431 (1.648)
Saiz elasticity											0.010 (0.008)	-0.022* (0.010)
Constant	-4.667** (0.413)	-0.165 (0.431)	-1.485** (0.323)	-1.199* (0.473)	-2.647** (0.282)	-0.881** (0.285)	-0.262 (0.422)	-0.057 (0.500)	-0.636 (0.560)	0.000 (1.010)	-0.839 (0.588)	0.373 (0.966)
R^2	0.98	0.98	0.99	0.98	0.16	0.03	0.30	0.15	0.68	0.53	0.68	0.54
N	3,111	3,111	3,081	3,081	3,111	3,111	3,081	3,081	457	457	457	457

Note: $L_{,-i} \equiv \sum_{n:d_{ni} \leq 120, n \neq i} L_n$ 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} . Columns 1-8 use the whole sample of counties. Columns 9 and 10 repeat the most complete specifications in columns 7 and 8 only for the subsample of counties where we have data on land supply elasticity. Columns 11 and 12 repeat columns 7 and 8 adding the Saiz land supply elasticity as a regressor. Standard errors are clustered by state. * denotes significance at the 5 percent level; ** denotes significance at the 1 percent level.

Table B.2: Explaining employment (L_i), residents (R_i), and the ratio of employment to residents (L_i/R_i)

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 (B.70) across counties.

Table B.3: Grubel-Lloyd Index for Commuting

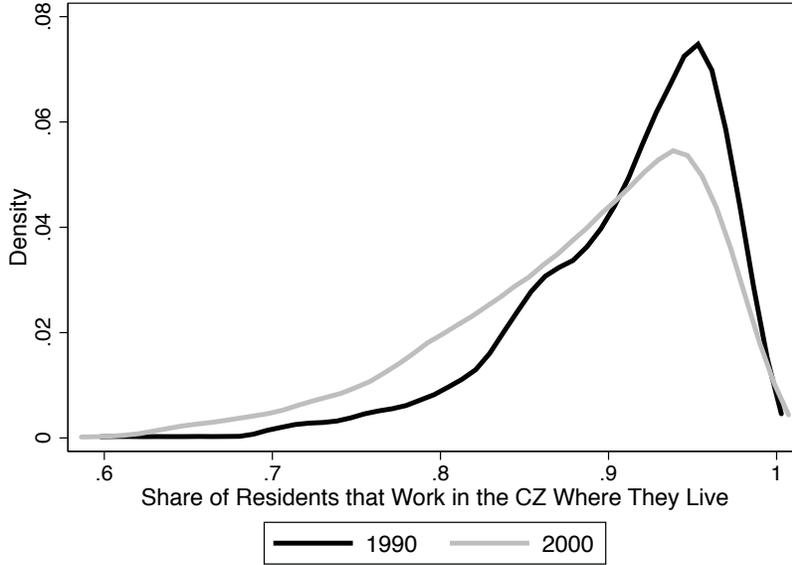


Figure B.4: Kernel densities of the share of residents that work in the CZ where they live

Additionally, to provide a point of comparison to Figure 1 in the paper for counties, Figure B.4 shows kernel densities of the share of residents that work in the same CZ where they live for 1990 and 2000. We construct these measures for CZs from the matrices of bilateral commuting probabilities between counties, which are only reported in the Population Census from 1990 onwards. We find the same pattern of an increase in commuting openness over time, with the increase between 1990 and 2000 for CZs in Figure B.4 larger than the increase over the same period for counties in Figure 1 in the paper.

As discussed in Section 3.2 of the paper, the gravity equation for the commuting probability in equation (10) in the paper can be written as

$$\lambda_{ni} - \frac{\mathcal{B}_{ni} \left(\frac{L_n}{\pi_{nn}}\right)^{-\frac{\alpha\epsilon}{\sigma-1}} A_n^{\alpha\epsilon} w_n^{-\alpha\epsilon} \bar{v}_n^{-\epsilon(1-\alpha)} \left(\frac{R_n}{H_n}\right)^{-\epsilon(1-\alpha)} w_i^\epsilon}{\sum_{r \in N} \sum_{s \in N} \mathcal{B}_{rs} \left(\frac{L_r}{\pi_{rr}}\right)^{-\frac{\alpha\epsilon}{\sigma-1}} A_r^{\alpha\epsilon} w_r^{-\alpha\epsilon} \bar{v}_r^{-\epsilon(1-\alpha)} \left(\frac{R_r}{H_r}\right)^{-\epsilon(1-\alpha)} w_s^\epsilon} = 0, \quad (\text{B.71})$$

where $\mathcal{B}_{ni} \equiv B_{ni} \kappa_{ni}^{-\epsilon}$ is a composite parameter that captures the ease of commuting. The commuting probabilities (B.71) provide a system of $N \times N$ equations that can be solved for a unique matrix of $N \times N$ values of the ease of commuting (\mathcal{B}_{ni}), as summarized in the following proposition.

Proposition B.2 (Amenities Inversion) *Given the share of consumption goods in expenditure (α), the heterogeneity in location preferences (ϵ), the observed data on wages, employment, trade shares, average residential income, residents and land area $\{w_i, L_i, \pi_{ii}, \bar{v}_i, R_i, H_i\}$, there exist unique values of the ease of commuting ($\mathcal{B}_{ni} \equiv B_{ni} \kappa_{ni}^{-\epsilon}$) for each pair of locations n and i that are consistent with the data being an equilibrium of the model.*

Proof. Note that the commuting probability (B.71) can be written as the following excess demand system:

$$\mathbb{D}_i(\mathcal{B}) = \lambda_{ni} - \frac{\mathcal{B}_{ni} \left(\frac{L_n}{\pi_{nn}}\right)^{-\frac{\alpha\epsilon}{\sigma-1}} A_n^{\alpha\epsilon} w_n^{-\alpha\epsilon} \bar{v}_n^{-\epsilon(1-\alpha)} \left(\frac{R_n}{H_n}\right)^{-\epsilon(1-\alpha)} w_i^\epsilon}{\sum_{r \in N} \sum_{s \in N} \mathcal{B}_{rs} \left(\frac{L_r}{\pi_{rr}}\right)^{-\frac{\alpha\epsilon}{\sigma-1}} A_r^{\alpha\epsilon} w_r^{-\alpha\epsilon} \bar{v}_r^{-\epsilon(1-\alpha)} \left(\frac{R_r}{H_r}\right)^{-\epsilon(1-\alpha)} w_s^\epsilon} = 0, \quad (\text{B.72})$$

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.72) exhibits the same properties in \mathcal{B} as the excess demand system (B.68) 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.72). ■

The resulting solutions for the ease of commuting (\mathcal{B}_{ni}) capture all factors that make a pair of residence and workplace locations more or less attractive conditional on the observed wages, employment, trade shares, average residential income, residents and land area (e.g. attractive scenery, distance and transport infrastructure). Together productivity (A_i) and the ease of commuting (\mathcal{B}_{ni}) correspond to structural residuals that ensure that the model exactly replicates the observed data given the parameters.

To estimate the heterogeneity in location preferences (ϵ), we model the determinants of the bilateral ease of commuting. For bilateral pairs with positive commuting flows, we partition the ease of commuting (\mathcal{B}_{ni}) into four components: (i) a residence component (\mathbb{B}_n), (ii) a workplace component (\mathbb{B}_i), (iii) a component that is related to distance ($\text{dist}_{ni}^{-\phi}$), and (iv) an orthogonal component (\mathbb{B}_{ni})

$$\log \mathcal{B}_{ni} \equiv \log(B_{ni}\kappa_{ni}^{-\epsilon}) = \log \mathbb{B}_n + \log \mathbb{B}_i - \phi \log(\text{dist}_{ni}) + \log \mathbb{B}_{ni}. \quad (\text{B.73})$$

We can always undertake this statistical decomposition of the ease of commuting ($\log \mathcal{B}_{ni}$), where the error term ($\log \mathbb{B}_{ni}$) is orthogonal to distance by construction, because the reduced-form coefficient on log distance ($-\phi$) captures any correlation of either log bilateral amenities ($\log B_{ni}$) and/or log bilateral commuting costs ($\log(\kappa_{ni}^{-\epsilon})$) with log distance. For bilateral pairs with zero commuting, the model implies negligible amenities ($B_{ni} \rightarrow 0$) and/or prohibitive commuting costs ($\kappa_{ni} \rightarrow \infty$).⁴

In the first step of our gravity equation estimation, we use this decomposition (B.73) and our expression for commuting flows (10) to estimate the reduced-form distance coefficient ($-\phi$):

$$\log \lambda_{ni} = g_0 + \eta_n + \mu_i - \phi \log \text{dist}_{ni} + \log \mathbb{B}_{ni}, \quad (\text{B.74})$$

where the residence fixed effect (η_n) captures the consumption goods price index (P_n), the price of residential land (Q_n), and the residence component of the ease of commuting (\mathbb{B}_n); the workplace fixed effect (μ_i) captures the wage (w_i) and the workplace component of the ease of commuting (\mathbb{B}_i); the constant g_0 captures the denominator of λ_{ni} and is separately identified because we normalize the residence and workplace fixed effects to sum to zero; and the error term ($\log \mathbb{B}_{ni}$) is orthogonal to log distance, because all effects of log distance on the composite ease of commuting are captured in the reduced-form distance coefficient ($-\phi$).⁵

Estimating the gravity equation (B.74) for all bilateral pairs with positive commuters using OLS, we find a regression R-squared of 0.80. In Figure B.5, we display the conditional relationship between log commuters and log distance, after removing residence and workplace fixed effects from both log commuters

⁴As for goods trade above, one interpretation is that commuting requires prior investments in transport infrastructure that are not modeled here. We adopt our specification for tractability, but other explanations for zero commuting flows include a support for the distribution of idiosyncratic preferences that is bounded from above or granularity.

⁵In Subsection B.11 of this web appendix, we generalize this specification to introduce congestion that is a power function of the volume of commuters. We show that this generalization affects the interpretation of the estimated coefficients in the gravity equation, but leaves the model's prediction of heterogeneous local employment elasticities unchanged.

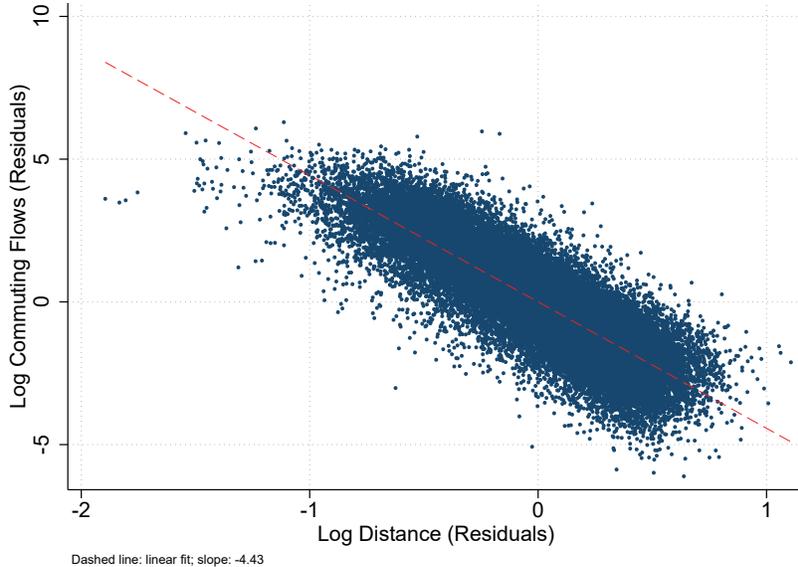


Figure B.5: Gravity in Commuting Between Counties

and log distance. Consistent with the existing empirical literature on commuting, we find that the log linear functional form provides a good approximation to the data, with a tight and approximately linear relationship between the two variables, and an estimated coefficient on log distance of $-\phi = -4.43$. This estimated coefficient is substantially larger than the corresponding coefficient for trade in goods of $-(\sigma - 1)\psi = -1.29$, which is consistent with the view that transporting people is considerably more costly than transporting goods, in line with the substantial opportunity cost of time spent commuting.

To identify the Fréchet shape parameter (ϵ), the second step of our gravity equation estimation uses additional structure from the model, which implies that the workplace fixed effects μ_i depend on wages (w_i) and the workplace component of the ease of commuting (\mathbb{B}_i):

$$\log \lambda_{ni} = g_0 + \eta_n + \epsilon \log w_i - \phi \log \text{dist}_{ni} + \log u_{ni}, \quad (\text{B.75})$$

where the error term is given by $\log u_{ni} \equiv \log \mathbb{B}_i + \log \mathbb{B}_{ni}$.

We estimate the gravity equation (B.75) imposing $\phi = 4.43$ from our estimates above and identify ϵ from the coefficient on wages. Estimating (B.75) using OLS is potentially problematic, because workplace wages (w_i) depend on the supply of commuters, which in turn depends on amenities that appear in the error term ($\log u_{ni}$). Therefore we instrument $\log w_i$ with the log productivities $\log A_i$ that we recovered from the condition (16) equating income and expenditure above, using the fact that the model implies that productivity satisfies the exclusion restriction of only affecting commuting flows through wages. Our Two-Stage-Least-Squares estimate of the Fréchet shape parameter for the heterogeneity of worker preferences is $\epsilon = 3.30$.⁶ The tight fit shown in Figure B.5 makes us confident that our parametrization of the composite

⁶We find that the Two-Stage-Least-Squares estimates are larger than the OLS estimates, consistent with the idea that bilateral commutes with attractive amenities have a higher supply of commuters and hence lower wages. The p-value for the first-stage F-Statistic for productivity is zero, confirming that productivity is a powerful instrument for wages. Note that one could have estimated jointly ϕ and ϵ from the restricted equation (B.75) directly. Our approach, however, imposes only the minimal set of necessary restrictions at every step: we estimate a flexible gravity structure to identify ϕ in (B.74), and a slightly less general specification (where workplace fixed effects are restricted to capture only variation in workplace wages) to

ease of commuting in terms of distance fits the data quite well.

B.7 Openness of the Local Labor Market to Commuting

In this section of the web appendix, we use the commuter market clearing condition to derive reduced-form measures of the openness of the local labor market to commuting. We show that the share of residents who work where they live (the “residence own commuting share,” $\lambda_{ii|i}^R$), the share of workers who live where they work (the “workplace own commuting share,” $\lambda_{ii|i}^L$), and the ratio of workers to residents (L_i/R_i) are all closely related to one another through the commuter market clearing condition.

Re-writing the commuter market clearing condition in equation (13) in the paper, we obtain

$$L_i = \lambda_{ii|i}^R R_i + \sum_{n \neq i} \lambda_{ni|n}^R R_n. \quad (\text{B.76})$$

Rearranging this commuter market clearing condition, the importance of commuting from other locations as a source of employment for location i can be written as

$$\frac{\sum_{n \neq i} \lambda_{ni|n}^R R_n}{L_i} = 1 - \frac{\lambda_{ii|i}^R R_i}{L_i}. \quad (\text{B.77})$$

We now use the definition of the conditional commuting probabilities ($\lambda_{ni|i}^L$ and $\lambda_{ni|n}^R$) in equations (B.11) and (B.12), which imply

$$\lambda_{ii|i}^R = \frac{\lambda_{ii}}{R_i/\bar{L}} = \frac{L_i}{R_i} \frac{\lambda_{ii}}{L_i/\bar{L}} = \frac{L_i}{R_i} \lambda_{ii|i}^L. \quad (\text{B.78})$$

Combining equations (B.77) and (B.78), we obtain:

$$\frac{\sum_{n \neq i} \lambda_{ni|n}^R R_n}{L_i} = 1 - \lambda_{ii|i}^L, \quad (\text{B.79})$$

where higher values of the workplace own commuting share ($\lambda_{ii|i}^L$) imply a local labor market that is more *closed* to commuting. Alternatively, the commuter market clearing condition can be written equivalently as

$$R_i = \lambda_{ii|i}^L L_i + \sum_{n \neq i} \lambda_{ni|n}^L L_n. \quad (\text{B.80})$$

Rearranging this expression, the importance of commuting from other locations as a source of residents for location i can be written as

$$\frac{\sum_{n \neq i} \lambda_{ni|n}^L L_n}{R_i} = 1 - \frac{\lambda_{ii|i}^L L_i}{R_i}. \quad (\text{B.81})$$

Combining equations (B.77) and (B.81), we obtain

$$\frac{\sum_{n \neq i} \lambda_{ni|n}^L L_n}{R_i} = 1 - \lambda_{ii|i}^R, \quad (\text{B.82})$$

where higher values of the residence own commuting share ($\lambda_{ii|i}^R$) again imply a local labor market that is

identify ϵ . Estimating the restricted equation (B.75) directly would yield very similar results: we find $\epsilon = 3.19$ and $\phi = 4.09$.

more *closed* to commuting.

Together, the residence and workplace own commuting shares ($\lambda_{ii|i}^R$ and $\lambda_{ii|i}^L$ respectively) are sufficient to recover the ratio of employment to residents (L_i/R_i). From equation (B.78), we have:

$$\frac{\lambda_{ii|i}^R}{\lambda_{ii|i}^L} = \frac{L_i}{R_i}. \quad (\text{B.83})$$

Therefore, knowing whether the minimum value of these two measures is equal to the residence commuting share ($\lambda_{ii|i}^R$) or the workplace commuting share ($\lambda_{ii|i}^L$) reveals whether a location is a net importer or net exporter of commuters:

$$\begin{aligned} \lambda_{ii|i}^R > \lambda_{ii|i}^L, & \quad \Leftrightarrow \quad L_i > R_i, \\ \lambda_{ii|i}^R < \lambda_{ii|i}^L, & \quad \Leftrightarrow \quad L_i < R_i. \end{aligned} \quad (\text{B.84})$$

We find that the residence and workplace own commuting shares ($\lambda_{ii|i}^R$ and $\lambda_{ii|i}^L$ respectively) are strongly positively correlated with one another, with a correlation of 0.60 from 2006-10 that is statistically significant at the 1 percent level. This positive correlation reflects in part the fact that gross commuting flows are large relative to net commuting flows, as explained by idiosyncratic preference draws in our model. We choose the residence own commuting share ($\lambda_{ii|i}^R$) as our baseline measure, because it is both model consistent and reported in the population census back to 1960. But we show that our results are robust to using either the residence or workplace commuting share or the average or minimum of these two measures.

B.8 Partial Equilibrium Elasticities

In this section of the web appendix, we use the model to derive partial equilibrium elasticities that capture the direct effect of a productivity shock on wages, employment and residents in the treated location, holding constant all other endogenous variables at their values in the initial equilibrium. Although these partial equilibrium elasticities do not incorporate the full set of interactions between locations that are captured in the general equilibrium elasticities in Figure 2 in the paper, we show in Section 4.1 of the paper that they explain some of the observed variation in these general equilibrium elasticities across locations.

We now derive these 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 in equation (7) in the paper, we have:

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

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:

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

This implies:

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

which can be re-written as:

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

where we have used the fact that productivity does not directly enter the commuter market clearing condition in equation (13) in the paper and the residential choice probabilities in equation (11) in the paper, 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 following partial equilibrium elasticity:

$$\frac{\partial w_i}{\partial A_i} \frac{A_i}{w_i} = \frac{(\sigma - 1) \sum_{r \in N} (1 - \pi_{ri}) \xi_{ri}}{\left[1 + (\sigma - 1) \sum_{r \in N} (1 - \pi_{ri}) \xi_{ri} \right] + \left[1 - \sum_{r \in N} (1 - \pi_{ri}) \xi_{ri} \right] \frac{dL_i}{dw_i} \frac{w_i}{L_i} - \xi_{ii} \frac{dR_i}{dw_i} \frac{w_i}{R_i}}, \quad (\text{B.85})$$

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 in equation (13) in the paper, we have:

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

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

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

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

$$\frac{\partial L_i}{\partial w_i} \frac{w_i}{L_i} = \epsilon \sum_{r \in N} \left(1 - \lambda_{ri|r}^R \right) \vartheta_{ri} + \vartheta_{ii} \left(\frac{dR_i}{dw_i} \frac{w_i}{R_i} \right), \quad (\text{B.86})$$

where $\vartheta_{ri} = \lambda_{ri|r}^R R_r / L_i$ 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.

Residents Elasticity: Totally differentiating the residential choice probability (λ_n^R in equation (11) in the paper), we have:

$$\begin{aligned} \frac{dR_i}{R_i} \frac{R_i}{\bar{L}} &= -\epsilon\alpha (1 - \lambda_i^R) \lambda_i^R \frac{dP_i}{P_i} + \epsilon\alpha \sum_{r \neq i} \lambda_r^R \lambda_i^R \frac{dP_r}{P_r} \\ &\quad - \epsilon(1 - \alpha) (1 - \lambda_i^R) \lambda_i^R \frac{dQ_i}{Q_i} + \epsilon(1 - \alpha) \sum_{r \neq i} \lambda_r^R \lambda_i^R \frac{dQ_r}{Q_r} \\ &\quad + \epsilon \lambda_{ii} \frac{dw_i}{w_i} - \epsilon \lambda_i^L \lambda_i^R \frac{dw_i}{w_i} - \epsilon \sum_{s \neq i} \lambda_s^L \lambda_i^R \frac{dw_s}{w_s}. \end{aligned}$$

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:

$$\frac{\partial R_i}{R_i} \frac{R_i}{\bar{L}} = \epsilon (\lambda_{ii} - \lambda_i^L \lambda_i^R) \frac{\partial w_i}{w_i}.$$

This implies the following partial equilibrium elasticity:

$$\frac{\partial R_i}{\partial w_i} \frac{w_i}{R_i} = \epsilon \left(\frac{\lambda_{ii}}{\lambda_i^R} - \lambda_i^L \right), \quad (\text{B.87})$$

where we use the partial derivative symbol to clarify that this derivative is not the full general equilibrium elasticity. Using the residents elasticity (B.87) in the employment elasticity (B.86), and using the residents and employment elasticities ((B.87) and (B.86) respectively) in the wage elasticity (B.85), we obtain the following partial equilibrium elasticities for the productivity shock,

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

B.9 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:

$$L_{ni} = \mathcal{R}_n \mathcal{B}_{ni} \mathcal{W}_i, \quad (\text{B.88})$$

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.88) implies that the unconditional probability that a worker

commutes from residence n to workplace i is:

$$\lambda_{ni} = \frac{L_{ni}}{\sum_{r \in N} \sum_{s \in N} L_{rs}} = \frac{\mathcal{R}_n \mathcal{B}_{ni} \mathcal{W}_i}{\sum_{r \in N} \sum_{s \in N} \mathcal{R}_r \mathcal{B}_{rs} \mathcal{W}_s}. \quad (\text{B.89})$$

The corresponding probability of working in location i is:

$$\lambda_i^L = \frac{\sum_{r \in N} L_{ri}}{\sum_{r \in N} \sum_{s \in N} L_{rs}} = \frac{\sum_{r \in N} \mathcal{R}_r \mathcal{B}_{ri} \mathcal{W}_i}{\sum_{r \in N} \sum_{s \in N} \mathcal{R}_r \mathcal{B}_{rs} \mathcal{W}_s}, \quad (\text{B.90})$$

and the probability of residing in location n is:

$$\lambda_n^R = \frac{\sum_{s \in N} L_{ns}}{\sum_{r \in N} \sum_{s \in N} L_{rs}} = \frac{\sum_{s \in N} \mathcal{R}_n \mathcal{B}_{ns} \mathcal{W}_s}{\sum_{r \in N} \sum_{s \in N} \mathcal{R}_r \mathcal{B}_{rs} \mathcal{W}_s}, \quad (\text{B.91})$$

From equations (B.89) and (B.91), the probability of commuting from residence n to workplace i conditional on residing in n is:

$$\lambda_{ni|n}^R = \frac{\lambda_{ni}}{\lambda_n^R} = \frac{\mathcal{R}_n \mathcal{B}_{ni} \mathcal{W}_i}{\sum_{s \in N} \mathcal{R}_n \mathcal{B}_{ns} \mathcal{W}_s} = \frac{\mathcal{B}_{ni} \mathcal{W}_i}{\sum_{s \in N} \mathcal{B}_{ns} \mathcal{W}_s}. \quad (\text{B.92})$$

Using this conditional probability (B.92), the commuter market clearing condition can be written as:

$$L_i = \sum_{n \in N} \lambda_{ni|n}^R R_n = \sum_{n \in N} \frac{\mathcal{B}_{ni} \mathcal{W}_i}{\sum_{s \in N} \mathcal{B}_{ns} \mathcal{W}_s} R_n. \quad (\text{B.93})$$

Totally differentiating this commuter market clearing condition (B.93) for a given commuting technology \mathcal{B}_{ni} , we have:

$$\frac{dL_i}{L_i} = \sum_{r \in N} \left(1 - \lambda_{ri|r}^R\right) \frac{d\mathcal{W}_i}{\mathcal{W}_i} \frac{\lambda_{ri|r}^R R_r}{L_i} - \sum_{r \in N} \sum_{s \neq i} \lambda_{rs|r}^R \frac{d\mathcal{W}_s}{\mathcal{W}_s} \frac{\lambda_{ri|r}^R R_r}{L_i} \quad (\text{B.94})$$

$$+ \sum_{r \in N} \frac{dR_r}{R_r} \frac{\lambda_{ri|r}^R R_r}{L_i}. \quad (\text{B.95})$$

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 $r \neq i$):

$$\frac{\partial L_i}{L_i} = \sum_{r \in N} \left(1 - \lambda_{ri|r}^R\right) \frac{\lambda_{ri|r}^R R_r}{L_i} \frac{\partial \mathcal{W}_i}{\mathcal{W}_i} + \frac{\lambda_{ii|i}^R R_i}{L_i} \frac{\partial R_i}{R_i}. \quad (\text{B.96})$$

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

$$\frac{\partial L_i}{\partial \mathcal{W}_i} \frac{\mathcal{W}_i}{L_i} = \underbrace{\sum_{r \in N} \left(1 - \lambda_{ri|r}^R\right) \vartheta_{ri}}_{\text{commuting}} + \underbrace{\vartheta_{ii} \left(\frac{\partial R_i}{\partial \mathcal{W}_i} \frac{\mathcal{W}_i}{R_i}\right)}_{\text{migration}}, \quad (\text{B.97})$$

where $\vartheta_{ri} = \lambda_{ri|r}^R R_r / L_i$ 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.97) 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.97) takes the same form as in the 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.97) 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.6, 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.6 to Figure 2 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.7 shows the same commuting component of the partial employment elasticity, but for CZs rather than counties. Comparing Figure B.7 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.

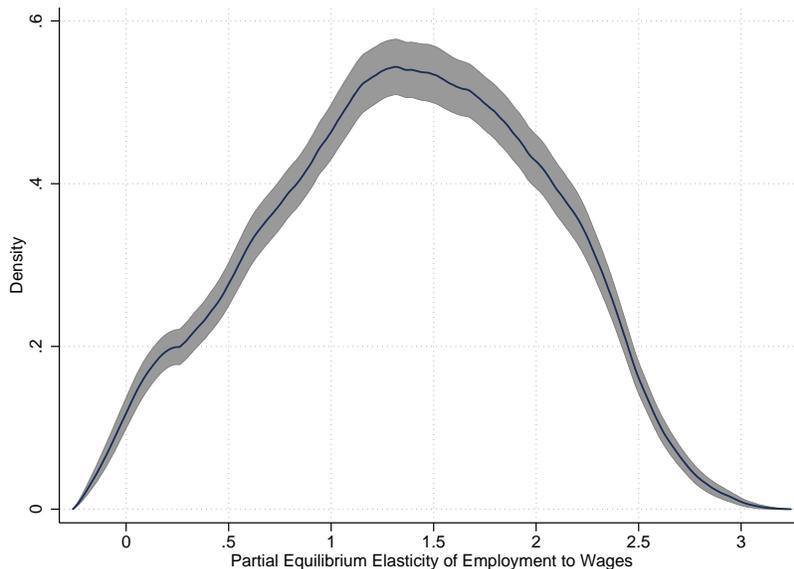


Figure B.6: Commuting component of partial employment elasticity for counties

B.10 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

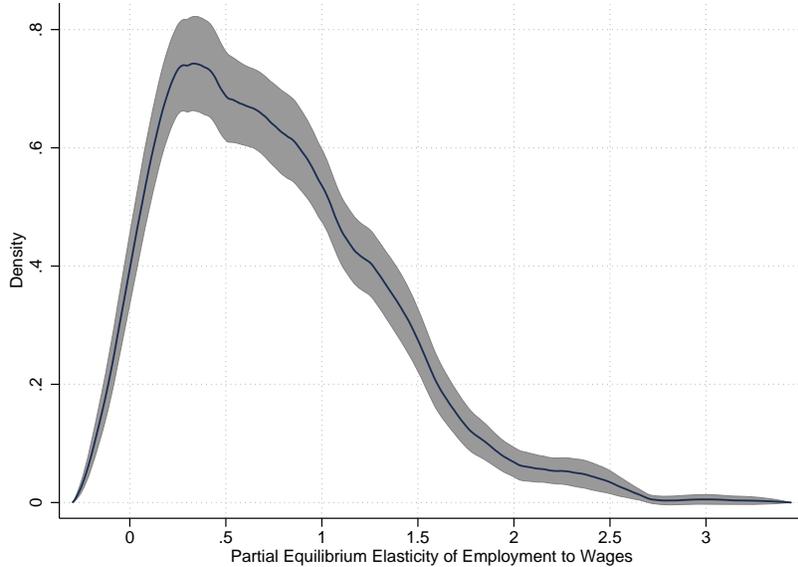


Figure B.7: Commuting component of partial employment elasticity for commuting zones (CZs)

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 $z = 1, \dots, Z$. 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):

$$G_{ni}^z(b) = e^{-B_{ni}^z b^{-\epsilon^z}}. \quad (\text{B.98})$$

B.10.1 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:

$$\pi_{ni|n}^z = \frac{B_{ni}^z (w_i^z / \kappa_{ni}^z)^{\epsilon^z}}{\sum_{s \in N} B_{ns}^z (w_s^z / \kappa_{ns}^z)^{\epsilon^z}}. \quad (\text{B.99})$$

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

$$L_i^z = \sum_{r \in N} \frac{B_{ri}^z (w_i^z / \kappa_{ri}^z)^{\epsilon^z}}{\sum_{s \in N} B_{rs}^z (w_s^z / \kappa_{rs}^z)^{\epsilon^z}} R_r^z, \quad (\text{B.100})$$

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:

$$\frac{\partial L_i^z}{\partial w_i^z} \frac{w_i^z}{L_i^z} = \epsilon^z \sum_{r \in N} \left(1 - \lambda_{ri|r}^{Rz}\right) \vartheta_{ri}^z + \vartheta_{ii}^z \left(\frac{\partial R_i^z}{\partial w_i^z} \frac{w_i^z}{R_i^z}\right), \quad (\text{B.101})$$

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

B.10.2 Aggregate Commuting Decisions

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

$$L_i = \sum_{z=1}^Z L_i^z. \quad (\text{B.102})$$

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:

$$dw_i^z = dw_i^k = dw_i > 0, \quad \forall z, k. \quad (\text{B.103})$$

Differentiating with respect to wages in equation (B.102), we have:

$$dL_i = \sum_{z=1}^Z \frac{\partial L_i^z}{\partial w_i^z} dw_i^z, \quad (\text{B.104})$$

which for a common change in wages in equation (B.103) can be re-written as:

$$\frac{dL_i}{dw_i} = \sum_{z=1}^Z \frac{\partial L_i^z}{\partial w_i^z}, \quad (\text{B.105})$$

which can be further re-written as:

$$\frac{dL_i}{dw_i} \frac{w_i}{L_i} = \sum_{z=1}^Z \left(\frac{\partial L_i^z}{\partial w_i^z} \frac{w_i^z}{L_i^z}\right) \left(\frac{L_i^z}{L_i} \frac{w_i}{w_i^z}\right). \quad (\text{B.106})$$

Combining equations (B.101) and (B.106), 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 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.101)). 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.106)). Taken together, this extension further reinforces our point that the local employment elasticity is not a structural parameter.

B.11 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:

$$G_{ni}(b) = e^{-B_{ni}L_{ni}^\chi b^{-\epsilon}}, \quad (\text{B.107})$$

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:

$$\lambda_{ni} = \frac{L_{ni}}{\bar{L}} = \frac{B_{ni}L_{ni}^\chi (\kappa_{ni}P_n^\alpha Q_n^{1-\alpha})^{-\epsilon} w_i^\epsilon}{\sum_{r \in N} \sum_{s \in N} B_{rs}L_{rs}^\chi (\kappa_{rs}P_r^\alpha Q_r^{1-\alpha})^{-\epsilon} w_s^\epsilon}, \quad (\text{B.108})$$

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

$$\bar{U} = \mathbb{E}[U_{ni\omega}] = \Gamma \left(\frac{\epsilon - 1}{\epsilon} \right) \left[\sum_{r \in N} \sum_{s \in N} B_{rs}L_{rs}^\chi (\kappa_{rs}P_r^\alpha Q_r^{1-\alpha})^{-\epsilon} w_s^\epsilon \right]^{\frac{1}{\epsilon}} \quad \text{all } n, i \in N. \quad (\text{B.109})$$

Combining equations (B.108) and (B.109), the flow of workers that choose to commute from residence n to workplace i can be written as:

$$L_{ni} = \left(\frac{\bar{U}}{\Gamma} \right)^{-\epsilon} B_{ni}L_{ni}^\chi (\kappa_{ni}P_n^\alpha Q_n^{1-\alpha})^{-\epsilon} w_i^\epsilon \bar{L}, \quad (\text{B.110})$$

which can be in turn re-written as:

$$L_{ni} = \left(\frac{\bar{U}}{\Gamma} \right)^{-\frac{\epsilon}{1-\chi}} B_{ni}^{\frac{1}{1-\chi}} (\kappa_{ni}P_n^\alpha Q_n^{1-\alpha})^{-\frac{\epsilon}{1-\chi}} w_i^{\frac{\epsilon}{1-\chi}} \bar{L}^{\frac{1}{1-\chi}}. \quad (\text{B.111})$$

Dividing equation (B.111) by its sum across all bilateral pairs, the probability that a worker commutes from residence n to workplace i can be equivalently expressed as:

$$\lambda_{ni} = \frac{L_{ni}}{\sum_{r \in N} \sum_{s \in N} L_{rs}} = \frac{L_{ni}}{\bar{L}} = \frac{B_{ni}^{\frac{1}{1-\chi}} (\kappa_{ni} P_n^\alpha Q_n^{1-\alpha})^{-\frac{\epsilon}{1-\chi}} w_i^{\frac{\epsilon}{1-\chi}}}{\sum_{r \in N} \sum_{s \in N} B_{rs}^{\frac{1}{1-\chi}} (\kappa_{rs} P_r^\alpha Q_r^{1-\alpha})^{-\frac{\epsilon}{1-\chi}} w_s^{\frac{\epsilon}{1-\chi}}}, \quad (\text{B.112})$$

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 multiplied 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.112), we can also solve for the probability of commuting to workplace i conditional on living in residence n :

$$\lambda_{ni|n}^R = \frac{B_{ni}^{\frac{1}{1-\chi}} (w_i/\kappa_{ni})^{\frac{\epsilon}{1-\chi}}}{\sum_{s \in N} B_{ns}^{\frac{1}{1-\chi}} (w_s/\kappa_{ns})^{\frac{\epsilon}{1-\chi}}}. \quad (\text{B.113})$$

The corresponding commuter market clearing condition is:

$$L_i = \sum_{r \in N} \frac{B_{ri}^{\frac{1}{1-\chi}} (w_i/\kappa_{ri})^{\frac{\epsilon}{1-\chi}}}{\sum_{s \in N} B_{rs}^{\frac{1}{1-\chi}} (w_s/\kappa_{rs})^{\frac{\epsilon}{1-\chi}}} R_r, \quad (\text{B.114})$$

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

$$\frac{\partial L_i}{\partial w_i} \frac{w_i}{L_i} = \frac{\epsilon}{1-\chi} \sum_{r \in N} \left(1 - \lambda_{ri|r}^R\right) \vartheta_{ri} + \vartheta_{ii} \left(\frac{\partial R_i}{\partial w_i} \frac{w_i}{R_i}\right), \quad (\text{B.115})$$

where $\vartheta_{ri} = \lambda_{ri|r}^R R_r/L_i$ is the share of commuters from residence r in workplace i 's employment. In this extended specification (B.115), 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.112), 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.12 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. We focus on the implications of introducing commuting into this canonical model for the elasticity of local employment with respect to local labor demand 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:

$$U_{ni\omega} = \frac{b_{ni\omega}}{\kappa_{ni}} \left(\frac{C_{Nn\omega}}{\alpha_N} \right)^{\alpha_N} \left(\frac{C_{Tn\omega}}{\alpha_T} \right)^{\alpha_T} \left(\frac{H_{n\omega}}{1 - \alpha_N - \alpha_T} \right)^{1 - \alpha_N - \alpha_T}, \quad (\text{B.116})$$

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

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):

$$P_n C_n = (\alpha_N + \alpha_T) \bar{v}_n R_n + (1 - \alpha_N - \alpha_T) \bar{v}_n R_n = \bar{v}_n R_n. \quad (\text{B.117})$$

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:

$$P_{Nn} C_{Nn} = \frac{\alpha_N}{\alpha_N + \alpha_T} P_n C_n = \frac{\alpha_N}{\alpha_N + \alpha_T} \bar{v}_n R_n, \quad (\text{B.118})$$

and the remaining fraction is allocated to the traded sector:

$$P_{Tn} C_{Tn} = \frac{\alpha_T}{\alpha_N + \alpha_T} P_n C_n = \frac{\alpha_T}{\alpha_N + \alpha_T} \bar{v}_n R_n, \quad (\text{B.119})$$

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:

$$Y_{Nn} = L_{Nn}, \quad (\text{B.120})$$

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:

$$P_{Nn} = w_n. \quad (\text{B.121})$$

Combining this result with utility maximization (B.118), and using goods market clearing for the non-traded good ($C_{Nn} = Y_{Nn}$) and the production technology (B.120), we find that the wage bill in the non-traded sector is a constant share of residential income:

$$w_n L_{Nn} = \frac{\alpha_N}{\alpha_N + \alpha_T} \bar{v}_n R_n. \quad (\text{B.122})$$

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

$$w_n L_{Tn} = \frac{\alpha_T}{\alpha_N + \alpha_T} \sum_{r \in N} \pi_{rn} \bar{v}_r R_r. \quad (\text{B.123})$$

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

$$L_n = L_{Tn} + L_{Nn} = \frac{\alpha_N}{\alpha_N + \alpha_T} \frac{\bar{v}_n R_n}{w_n} + \frac{\alpha_T}{\alpha_N + \alpha_T} \sum_{r \in N} \frac{\pi_{rn} \bar{v}_r R_r}{w_n}. \quad (\text{B.124})$$

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:

$$L_n = \sum_{r \in N} \frac{B_{rn} (w_n / \kappa_{rn})^\epsilon}{\sum_{s \in N} B_{rs} (w_s / \kappa_{rs})^\epsilon} R_r. \quad (\text{B.125})$$

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:

$$\frac{\partial L_n}{\partial w_n} \frac{w_n}{L_n} = \epsilon \sum_{r \in N} \left(1 - \lambda_{rn|r}^R\right) \vartheta_{rn} + \vartheta_{nn} \left(\frac{\partial R_n}{\partial w_n} \frac{w_n}{R_n}\right), \quad (\text{B.126})$$

where $\vartheta_{rn} = \lambda_{rn|r}^R R_r / L_n$ 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.13 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) in the paper is now instead given by:

$$P_n C_n = \alpha [1 + (1 - \alpha)] \bar{v}_n R_n. \quad (\text{B.127})$$

Using this relationship, the equality between income and expenditure that was previously given by equation (7) in the paper is now instead given by:

$$w_i L_i = \alpha [1 + (1 - \alpha)] \sum_{n \in N} \pi_{ni} \bar{v}_n R_n, \quad (\text{B.128})$$

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

$$Q_n = (1 - \alpha) [1 + (1 - \alpha)] \frac{\bar{v}_n R_n}{H_n}. \quad (\text{B.129})$$

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.128), land market clearing (B.129), expected labor income (which remains as in equation (14) in the paper), workplace choice probabilities (which continue to equal equation (11) in the paper for L_n), residence choice probabilities (which are still equal to equation (11) in the paper for R_n), price indices (again equal to equation (8) 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.128) and (B.129). 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.14 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:

$$\Lambda_i(j) = l_i(j)w_i = \left(F + \frac{x_i(j)}{A_i} \right) w_i. \quad (\text{B.130})$$

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:

$$\Lambda_i(j) = \left(F + \frac{x_i(j)}{A_i} \right) w_i^{\beta_L} Q_i^{\beta_Q} \mathbb{R}^{\beta_R} P_i^{1-\beta_L-\beta_Q-\beta_R}. \quad (\text{B.131})$$

The probability that a worker chooses to live in location n and work in location i remains the same as in equation (10) in the paper, which in turn implies that the commuter market clearing condition takes

exactly the same form as in our baseline specification:

$$L_n = \sum_{r \in N} \frac{B_{rn} (w_n / \kappa_{rn})^\epsilon}{\sum_{s \in N} B_{rs} (w_s / \kappa_{rs})^\epsilon} R_r. \quad (\text{B.132})$$

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:

$$\frac{\partial L_n}{\partial w_n} \frac{w_n}{L_n} = \epsilon \sum_{r \in N} \left(1 - \lambda_{rn|r}^R\right) \vartheta_{rn} + \vartheta_{nn} \left(\frac{\partial R_n}{\partial w_n} \frac{w_n}{R_n}\right), \quad (\text{B.133})$$

where $\vartheta_{rn} = \lambda_{rn|r}^R R_r / L_n$ 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.133) 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.15 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:

$$U_{ni\omega} = \frac{1}{\kappa_{ni}} \left(\frac{C_{n\omega}}{\alpha}\right)^\alpha \left(\frac{H_{n\omega}}{1-\alpha}\right)^{1-\alpha}. \quad (\text{B.134})$$

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 ($b_{ni\omega}$):

$$U_{ni\omega} = \frac{b_{ni\omega} w_i}{\kappa_{ni} P_n^\alpha Q_n^{1-\alpha}}. \quad (\text{B.135})$$

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:

$$\lambda_{ni} = \frac{B_{ni} (\kappa_{ni} P_n^\alpha Q_n^{1-\alpha})^{-\epsilon} w_i^\epsilon}{\sum_{r \in N} \sum_{s \in N} B_{rs} (\kappa_{rs} P_r^\alpha Q_r^{1-\alpha})^{-\epsilon} w_s^\epsilon}. \quad (\text{B.136})$$

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.16 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:

$$U_{ni\omega} = b_{ni\omega} \left(\frac{C_{n\omega}}{\alpha} \right)^\alpha \left(\frac{H_{n\omega}}{1-\alpha} \right)^{1-\alpha}. \quad (\text{B.137})$$

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 ($1/\kappa_{ni}$):

$$U_{ni\omega} = \frac{b_{ni\omega} w_i}{\kappa_{ni} P_n^\alpha Q_n^{1-\alpha}}. \quad (\text{B.138})$$

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:

$$\lambda_{ni} = \frac{B_{ni} (\kappa_{ni} P_n^\alpha Q_n^{1-\alpha})^{-\epsilon} w_i^\epsilon}{\sum_{r \in N} \sum_{s \in N} B_{rs} (\kappa_{rs} P_r^\alpha Q_r^{1-\alpha})^{-\epsilon} w_s^\epsilon}. \quad (\text{B.139})$$

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.17 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.17.1 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

$$\xi \equiv \frac{\sum_{i \in N} \iota_i (1 - \alpha) X_i}{\sum_i R_i}. \quad (\text{B.140})$$

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

$$X_n = \bar{v}_n R_n + \xi R_n + (1 - \iota_n) (1 - \alpha) X_n, \quad (\text{B.141})$$

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

$$D_n \equiv X_n - (\bar{v}_n R_n + Q_n H_n) = \xi R_n - \iota_i (1 - \alpha) X_i. \quad (\text{B.142})$$

Using equation (B.142) to substitute for ξR_n in equation (B.141), expenditure in location n can be equivalently written as

$$X_n = \frac{\bar{v}_n R_n + D_n}{\alpha}. \quad (\text{B.143})$$

B.17.2 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.141), and denoting the population share of each location in the initial equilibrium by $\rho_n \equiv R_n / \sum_i R_i$, we have

$$X_n = \bar{v}_n R_n + \rho_n \sum_{i \in N} \iota_i (1 - \alpha) X_i + (1 - \iota_n) (1 - \alpha) X_n. \quad (\text{B.144})$$

Using equations (B.143) and (B.144), we have

$$D_n = \alpha X_n - \bar{v}_n R_n = \rho_n \sum_{i \in N} \iota_i (1 - \alpha) X_i - \iota_n (1 - \alpha) X_n, \quad (\text{B.145})$$

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.17.3 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

$$\lambda_{ni} = \frac{B_{ni} (\kappa_{ni} P_n^\alpha Q_n^{1-\alpha})^{-\epsilon} w_i^\epsilon}{\sum_{r \in N} \sum_{s \in N} B_{rs} (\kappa_{rs} P_r^\alpha Q_r^{1-\alpha})^{-\epsilon} w_s^\epsilon}. \quad (\text{B.146})$$

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

$$R_n = \bar{L} \sum_{i \in N} \lambda_{ni}, \quad (\text{B.147})$$

$$L_i = \bar{L} \sum_{n \in N} \lambda_{ni}. \quad (\text{B.148})$$

Residential expenditure and income are related through equation (B.141), as reproduced here

$$X_n = \bar{v}_n R_n + \xi R_n + (1 - \iota_n) (1 - \alpha) X_n, \quad (\text{B.149})$$

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

$$\bar{v}_n = \sum_{i \in N} \lambda_{ni}^R w_i. \quad (\text{B.150})$$

and nationally-redistributed rent per capita (ξ) in equation (B.140) can be written as

$$\xi = \frac{\sum_{i \in N} \iota_i (1 - \alpha) X_i}{\bar{L}}. \quad (\text{B.151})$$

Workplace income equals expenditure on goods produced in that location

$$w_i L_i = \sum_{n \in N} \pi_{ni} \alpha X_n, \quad (\text{B.152})$$

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

$$\pi_{ni} = \frac{L_i (d_{ni} w_i / A_i)^{1-\sigma}}{\sum_{k \in N} L_k (d_{nk} w_k / A_k)^{1-\sigma}}. \quad (\text{B.153})$$

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

$$Q_n = (1 - \alpha) \frac{X_n}{H_n}, \quad (\text{B.154})$$

$$P_n = \frac{\sigma}{\sigma - 1} \left(\frac{L_n}{\sigma F \pi_{nn}} \right)^{\frac{1}{1-\sigma}} \frac{d_{nn} w_n}{A_n}. \quad (\text{B.155})$$

B.17.4 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

$$\hat{v}_n^{(t)} = \frac{1}{\bar{v}_n} \sum_{i \in N} \frac{\hat{B}_{ni} \lambda_{ni} \left(\hat{w}_i^{(t)} / \hat{\kappa}_{ni} \right)^\epsilon}{\sum_{s \in N} \hat{B}_{ns} \lambda_{ns} \left(\hat{w}_s^{(t)} / \hat{\kappa}_{ns} \right)^\epsilon} \hat{w}_i^{(t)} w_i, \quad (\text{B.156})$$

$$\hat{L}_i^{(t)} = \frac{\bar{L}}{L_i} \sum_n \lambda_{ni} \hat{\lambda}_{ni}^{(t)}, \quad (\text{B.157})$$

$$\hat{R}_n^{(t)} = \frac{\bar{L}}{R_n} \sum_i \lambda_{ni} \hat{\lambda}_{ni}^{(t)}, \quad (\text{B.158})$$

which are functions of the observed values of variables in the initial equilibrium and our guesses. From the land market clearing condition (B.154), the proportional change in land rents equals our guess for the proportional change in expenditure

$$\hat{Q}_n^{(t)} = \hat{X}_n^{(t)}. \quad (\text{B.159})$$

Using the proportional change in employment from equation (B.157) and our guess for the proportional change in wages, we can solve for the proportional change in trade shares from equation (B.153)

$$\hat{\pi}_{ni}^{(t)} = \frac{\hat{L}_i^{(t)} \left(\hat{d}_{ni} \hat{w}_i^{(t)} / \hat{A}_i \right)^{1-\sigma}}{\sum_{k \in N} \pi_{nk} \hat{L}_k^{(t)} \left(\hat{d}_{nk} \hat{w}_k^{(t)} / \hat{A}_k \right)^{1-\sigma}}. \quad (\text{B.160})$$

Using the proportional change in employment from equation (B.157), the proportional change in trade shares from equation (B.160) and our guess for the proportional change in wages, we can solve for the proportional change in the tradeables price index from equation (B.155)

$$\hat{P}_n^{(t)} = \left(\frac{\hat{L}_n^{(t)}}{\hat{\pi}_{nn}^{(t)}} \right)^{\frac{1}{1-\sigma}} \frac{\hat{w}_n^{(t)}}{\hat{A}_n}. \quad (\text{B.161})$$

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.151)

$$\hat{\xi}^{(t)} = \frac{1}{\xi} \frac{\sum_{i \in N} \iota_i (1-\alpha) X_i \hat{X}_i}{\bar{L}}. \quad (\text{B.162})$$

Finally, we use (B.156)-(B.162) in the equality between income and expenditure (B.152), the workplace-residence choice probabilities (B.146) and expenditure (B.149) to solve for the implied proportional changes

in wages, commuting probabilities and expenditure as

$$\tilde{w}_i^{(t+1)} = \frac{1}{w_i L_i \hat{L}_i^{(t)}} \sum_{n \in N} \alpha \pi_{ni} \hat{\pi}_{ni}^{(t)} X_n \hat{X}_n^{(t)}, \quad (\text{B.163})$$

$$\tilde{\lambda}_{ni}^{(t+1)} = \frac{\hat{B}_{ni} \left(\hat{P}_n^{(t)\alpha} \hat{Q}_n^{(t)1-\alpha} \right)^{-\epsilon} \left(\hat{w}_i^{(t)} / \hat{\kappa}_{ni} \right)^\epsilon}{\sum_{r \in N} \sum_{s \in N} \hat{B}_{rs} \lambda_{rs} \left(\hat{P}_r^{(t)\alpha} \hat{Q}_r^{(t)1-\alpha} \right)^{-\epsilon} \left(\hat{w}_s^{(t)} / \hat{\kappa}_{rs} \right)^\epsilon}, \quad (\text{B.164})$$

$$\tilde{X}_n^{(t+1)} = \frac{\bar{v}_n R_n \hat{v}_n \hat{R}_n + \xi R_n \hat{\xi}^{(t)} \hat{R}_n^{(t)} + (1 - \iota_n) (1 - \alpha) \hat{X}_n^{(t)}}{X_n}. \quad (\text{B.165})$$

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

$$\hat{w}_i^{(t+1)} = \zeta \hat{w}_i^{(t)} + (1 - \zeta) \tilde{w}_i^{(t+1)}, \quad (\text{B.166})$$

$$\hat{\lambda}_i^{(t+1)} = \zeta \hat{\lambda}_i^{(t)} + (1 - \zeta) \tilde{\lambda}_i^{(t+1)}, \quad (\text{B.167})$$

$$\hat{X}_i^{(t+1)} = \zeta \hat{X}_i^{(t)} + (1 - \zeta) \tilde{X}_i^{(t+1)}, \quad (\text{B.168})$$

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

B.17.5 Local Employment Elasticities

As in Section 4 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.8 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 2 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 part of the web appendix, we report additional empirical results and robustness tests. Subsection C.1 shows that the model's predictions for land prices are strongly positively correlated with median house prices in the data. Subsection C.2 reports standardized coefficients for the regressions examining the determinants of the local employment elasticity in Table 2 in the paper.

Subsection C.3 reports additional results from estimating “difference-in-differences” regressions using the counterfactuals from the model, as discussed in Subsection 4.1 of the paper. We show that the model-suggested controls are more successful in explaining the heterogeneous treatment effects than the standard

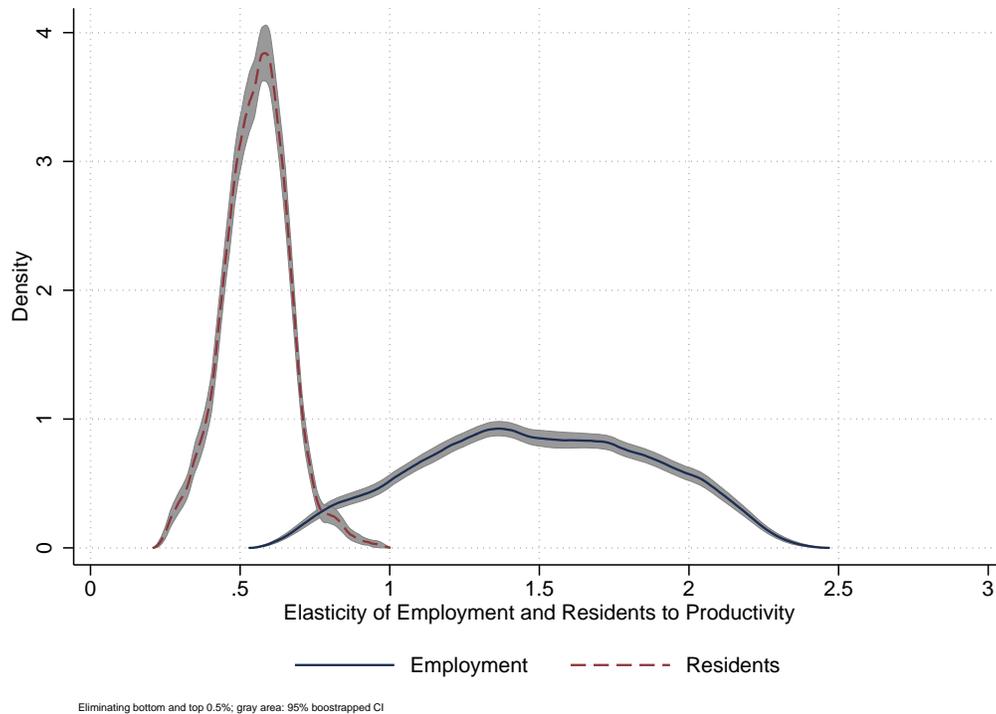


Figure B.8: Kernel density for the distribution of employment and residents elasticities in response to a productivity shock across counties (partial local and national ownership of land)

empirical controls from the local labor markets literature. Subsection C.4 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.5 reports additional results from the extension of the model to incorporate heterogeneous positive supply elasticities for developed land following Saiz (2010), as considered in Subsection 4.2 of the paper. Subsection C.6 provides further evidence on the role of commuting in generating heterogeneity in local employment elasticities in our quantitative model. We show that there is substantially less heterogeneity in these elasticities in a counterfactual world with no commuting between counties.

Subsection C.7 reports additional results for the MDP experiment from Section 5 of the paper. Subsection C.8 reports the shift-share decompositions of cross-section and time-series variation in employment discussed in Section 5 of the paper.

Subsection C.10 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.11 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 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 median house prices in the data, we find that the model has strong predictive power. In Section 4.2 of the paper, we generalize this baseline specification to allow for a positive supply elasticity for developed land that is heterogeneous across locations.

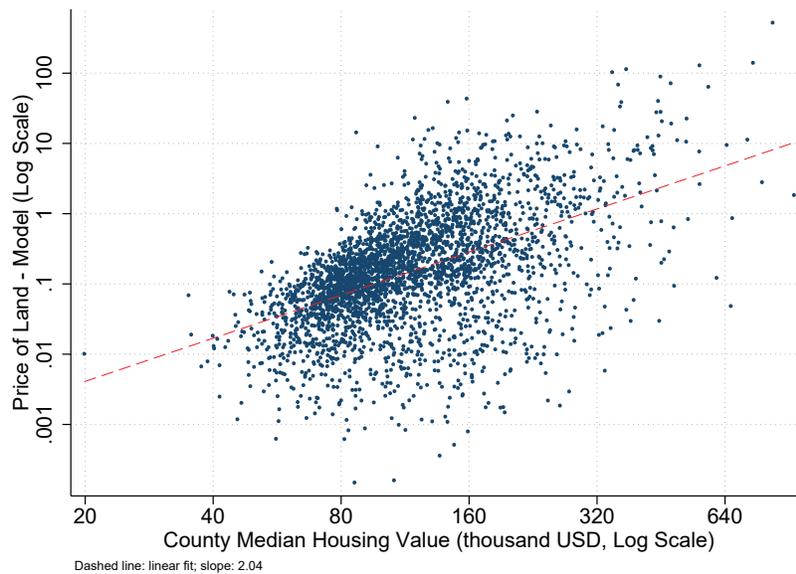


Figure C.1: Land Prices in the Model and House Prices in the Data

C.2 Standardized Employment Elasticities Regression

Table C.1 reports the estimated 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 fraction of standard deviations by which the dependent variable changes with a one standard deviation change in each independent variable.

Dependent Variable:	1	2	3	4	5	6	7	8	9
	Elasticity of Employment								
$\log L_i$		-0.012 (0.056)	0.036 (0.046)	-0.217** (0.025)				0.147** (0.018)	0.132** (0.017)
$\log w_i$			-0.126** (0.037)	-0.100** (0.024)				-0.162** (0.010)	-0.166** (0.010)
$\log H_i$			-0.621** (0.045)	-0.372** (0.033)				0.007 (0.020)	0.020 (0.020)
$\log L_{,-i}$				0.429** (0.061)				-0.097** (0.032)	-0.097** (0.033)
$\log \bar{w}_{-i}$				0.090* (0.037)				0.072** (0.016)	0.091** (0.017)
$\lambda_{ii i}^R$					-0.945** (0.019)				
$\sum_{n \in N} (1 - \lambda_{Rni}) \vartheta_{ni}$						1.462** (0.101)		1.343** (0.093)	
$\vartheta_{ii} \left(\frac{\lambda_{ii}}{\lambda_{Ri}} - \lambda_{Li} \right)$						0.487** (0.112)		0.322** (0.093)	
$\frac{\partial w_i}{\partial A_i} \frac{A_i}{w_i}$						-0.110** (0.013)		-0.090** (0.016)	
$\frac{\partial w_i}{\partial A_i} \frac{A_i}{w_i} \cdot \sum_{r \in N} (1 - \lambda_{rn r}) \vartheta_{rn}$							0.544** (0.047)		0.576** (0.048)
$\frac{\partial w_i}{\partial A_i} \frac{A_i}{w_i} \cdot \vartheta_{ii} \left(\frac{\lambda_{ii}}{\lambda_{Ri}} - \lambda_{Li} \right)$							-0.428** (0.051)		-0.444** (0.048)
Constant	-0.000 (0.090)	-0.000 (0.090)	0.000 (0.046)	0.000 (0.036)	-0.000 (0.031)	0.000 (0.028)	-0.000 (0.029)	-0.006 (0.026)	-0.006 (0.026)
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

Note: $L_{,-n} \equiv \sum_{r: d_{rn} \leq 120, r \neq n} L_r$ 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. Standard errors are clustered by state. * denotes significance at the 5 percent level; ** denotes significance at the 1 percent level.

Table C.1: Explaining the general equilibrium local employment elasticities to a 5 percent productivity shock (standardized regression)

C.3 Additional Treatment Heterogeneity Results

In this subsection of the web appendix, we supplement the results reported in Subsection 4.1 of the paper, and provide further evidence that the model-suggested controls are more successful in explaining the heterogeneity in treatment effects in our quantitative model than the standard empirical controls from the local labor markets literature. We compute the deviation between the general equilibrium elasticity in the model and the predicted elasticity from the reduced-form regression for each of the control groups (i)-(v) discussed in the paper

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

where we scale the regression estimates by size of the productivity shock.

In Figure C.2, we show that this deviation between the general equilibrium elasticity and the “differences-in-differences” prediction 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 “difference-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.

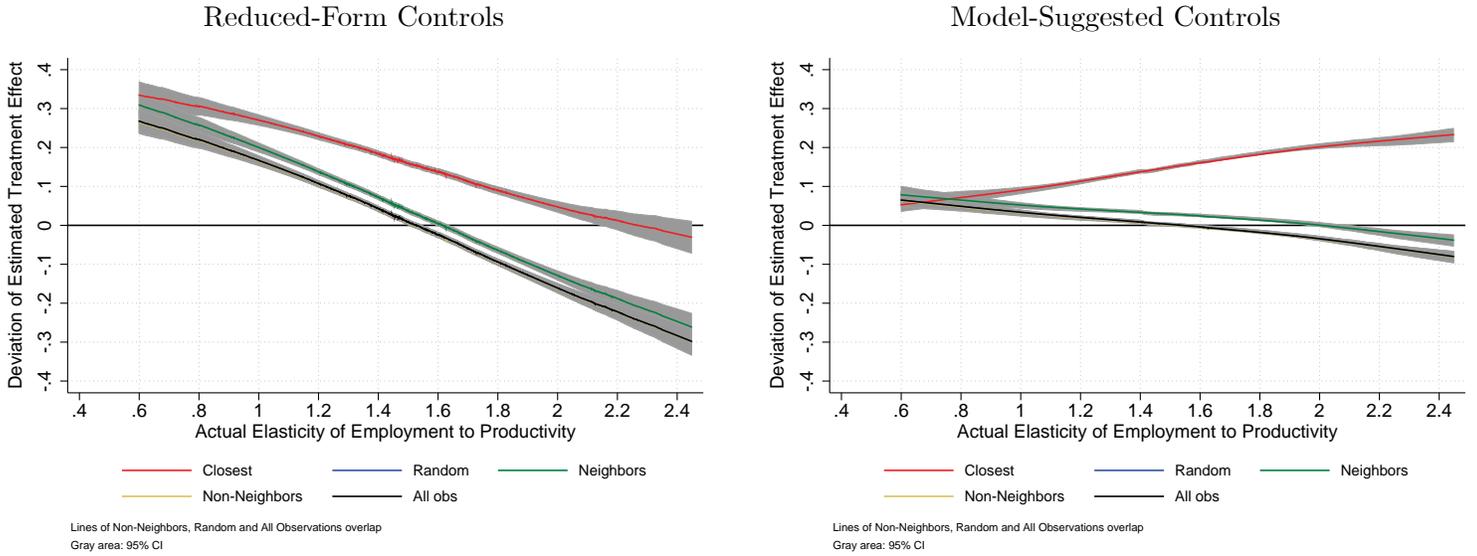


Figure C.2: Average deviation term β_i vs. general equilibrium employment elasticity

C.4 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.

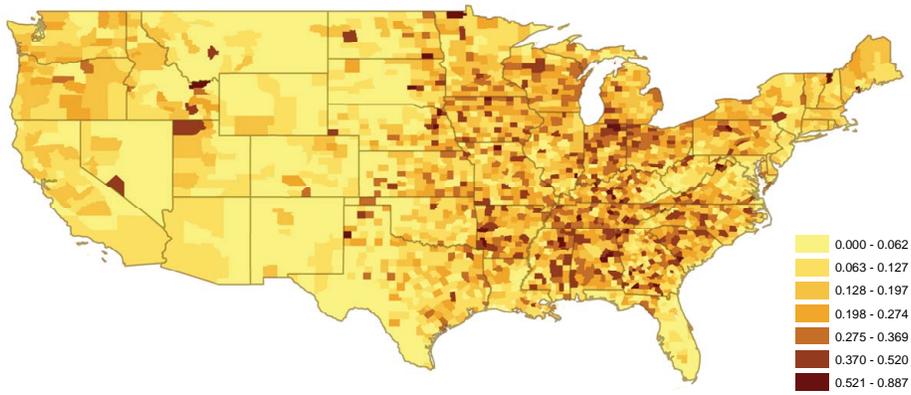


Figure C.3: U.S. counties' share of employment in manufacturing, 2007.

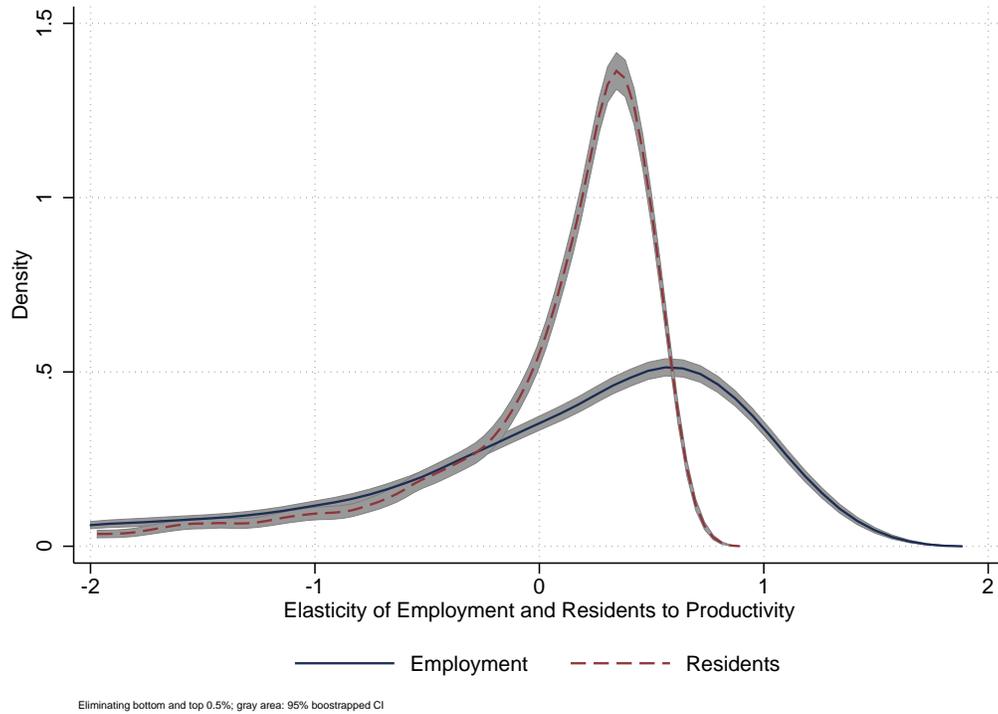


Figure C.4: Kernel density for the distribution of employment and residents elasticities in response to a spatially correlated productivity shock in the manufacturing sector

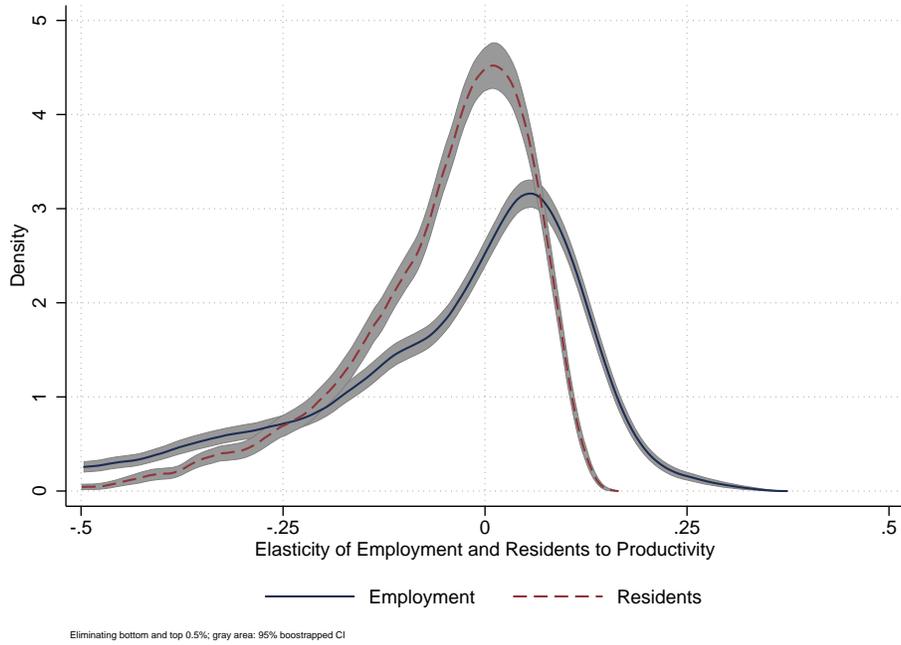


Figure C.5: Kernel density for the distribution of employment and residents elasticities in response to a spatially correlated productivity shock in the non-manufacturing sector

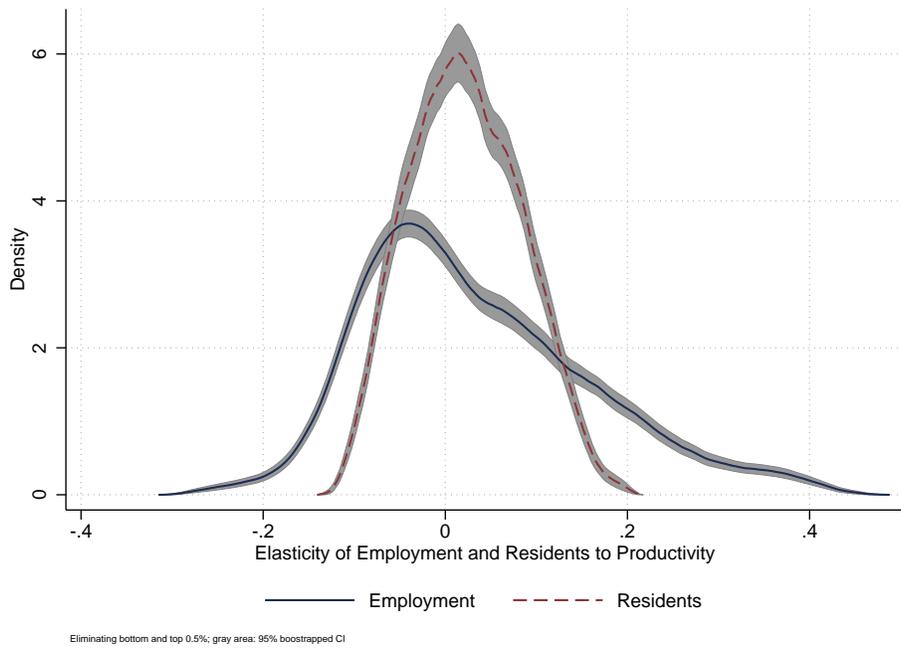


Figure C.6: Kernel density for the distribution of employment and residents elasticities in response to a spatially correlated shock in both sectors

C.5 Positive Developed Land Elasticities

In Subsection 4.2 of the paper, we develop an extension of the model in which we interpret the non-traded amenity as developed land and allow for a positive developed land supply elasticity that can differ across locations. In this subsection of the web appendix, we provide further details on this robustness check. We introduce a positive developed land supply elasticity by following Saiz (2010) in assuming that the supply of land (H_n) for each residence n depends on the endogenous price of land (Q_n) as well as on the exogenous characteristics of locations (\bar{H}_n):

$$H_n = \bar{H}_n Q_n^{\eta_n}, \quad (\text{C.1})$$

where $\eta_n \geq 0$ is the developed land supply elasticity, which we allow to vary across locations; $\eta_n = 0$ is our baseline specification of a perfectly inelastic land supply; and $\eta_n \rightarrow \infty$ is the special case of a perfectly elastic land supply.

Introducing a positive and heterogeneous developed land supply elasticity only affects one of the conditions for general equilibrium in the model, namely, the land market clearing condition. The rest of the model remains identical. Using the supply function for land (C.1) in the land market clearing condition (5), we obtain the following generalization of our earlier expression for the equilibrium price of land (Q_n):

$$Q_n = \left((1 - \alpha) \frac{\bar{v}_n R_n}{\bar{H}_n} \right)^{\frac{1}{1+\eta_n}}. \quad (\text{C.2})$$

We now show that our finding that commuting linkages are important for explaining differences in local employment elasticities is robust to controlling for variable housing supply elasticities. As in the main body of the paper, we focus on the subset of counties for which an estimate of the housing supply elasticity is available from Saiz (2010) and no imputation is required. We undertake counterfactuals for productivity shocks for these counties and undertake a horse race, in which we regress the general equilibrium employment elasticities in the model on our measures of commuting linkages, the Saiz housing supply elasticities, and other controls.

In Table C.2, we report the results from these regressions. In Columns 1-6, we begin by replicating the specifications from Table 2 in the paper for the subsample of counties for which Saiz housing supply elasticities are available. We find a similar pattern of results as for the full sample of counties in Table 2 in the paper. In particular, Column 2 shows that the residence own commuting share (λ_{ii}^R) alone explains 63 percent of the variation in local employment elasticities (compared with 89 percent for the full sample). Columns 7-12 of Table C.2 augment the specifications in Columns 1-6 with the Saiz housing supply elasticity. We find that both the estimated coefficient and statistical significance of our commuting measure are robust to the inclusion of the Saiz housing supply elasticity. The Saiz housing supply elasticity is statistically significant across specifications but contributes only to a minority of the explanatory power of the regression. This pattern of results is consistent with Figure 3 in the paper, where we show that introducing differences in housing supply elasticities increases the heterogeneity in the elasticity of residents with respect to the productivity shock, but has relatively little impact on the heterogeneity in the elasticity of employment with respect to the productivity shock.

This pattern of results is also consistent with existing research on housing supply elasticities. This

existing research has typically not distinguished between employment and residents (often focusing on population) and has typically been concerned on metropolitan statistical areas (MSAs) rather than counties. Therefore, the housing supply elasticity can be important for the response of the overall population of metropolitan areas to local labor demand shocks, but there can be considerable variation in the response of employment relative to residents across counties within these metropolitan areas. An important implication is that improvements in commuting technologies provide an alternative approach to relaxing housing supply elasticities in enabling individuals to access high productivity locations. While this possibility has been informally discussed in the existing literature on housing supply elasticities (as for example in Hsieh and Moretti 2017), our paper is the first study of which we are aware to provide quantitative empirical evidence on the relevance of commuting for local employment elasticities.

Dependent Variable:	Elasticity of Employment											
	1	2	3	4	5	6	7	8	9	10	11	12
	without housing supply elasticity						with housing supply elasticity					
$\log L_i$	-0.138** (0.021)				-0.042* (0.020)	-0.052* (0.022)	-0.058* (0.022)				0.054* (0.023)	0.052* (0.024)
$\log w_i$	-0.318* (0.129)				-0.359** (0.084)	-0.417** (0.103)	-0.183* (0.088)				-0.217** (0.053)	-0.249** (0.065)
$\log H_i$	-0.078** (0.019)				0.022 (0.025)	0.037 (0.024)	-0.117** (0.019)				-0.029 (0.026)	-0.021 (0.025)
$\log L_{,-i}$	-0.009 (0.036)				-0.031 (0.032)	-0.046 (0.037)	-0.034 (0.035)				-0.072* (0.033)	-0.085* (0.037)
$\log \bar{w}_{-i}$	0.516** (0.147)				0.137 (0.112)	0.384** (0.128)	0.667** (0.146)				0.324** (0.117)	0.507** (0.120)
$\lambda_{vi i}^R$		-1.738** (0.098)						-1.380** (0.108)				
$\sum_{n \in N} (1 - \lambda_{Rni}) \vartheta_{ni}$			5.500** (0.693)		3.653** (0.791)			3.266** (0.511)			3.227** (0.720)	
$\vartheta_{ii} \left(\frac{\lambda_{ii}}{\lambda_{Ri}} - \lambda_{Li} \right)$			4.014** (0.746)		2.181* (0.829)			1.932** (0.543)			1.822* (0.747)	
$\frac{\partial w_i}{\partial A_i} \frac{A_i}{w_i}$			-0.448 (0.304)		-1.275** (0.283)			-0.743** (0.273)			-0.527* (0.239)	
$\frac{\partial w_i}{\partial A_i} \frac{A_i}{w_i} \cdot \sum_{r \in N} (1 - \lambda_{rn r}) \vartheta_{rn}$				1.862** (0.216)	0.424 (0.283)			0.892** (0.206)			0.982** (0.185)	
$\frac{\partial w_i}{\partial A_i} \frac{A_i}{w_i} \cdot \vartheta_{ii} \left(\frac{\lambda_{ii}}{\lambda_{Ri}} - \lambda_{Li} \right)$				0.459 (0.251)	-0.992** (0.317)			-0.406* (0.189)			-0.405 (0.236)	
Saiz elasticity							0.158** (0.025)	0.140** (0.016)	0.148** (0.019)	0.154** (0.021)	0.152** (0.022)	0.158** (0.023)
Constant	2.088 (1.491)	3.042** (0.049)	-2.283** (0.469)	0.828** (0.207)	3.416** (1.227)	3.482** (1.218)	-1.418 (1.245)	2.678** (0.071)	-0.052 (0.340)	1.526** (0.170)	-0.741 (0.945)	-0.542 (0.980)
R^2	0.51	0.63	0.64	0.60	0.70	0.68	0.67	0.82	0.82	0.81	0.84	0.83
N	457	460	460	460	457	457	457	460	460	460	457	457

Note: $L_{,-i} \equiv \sum_{n: d_{ni} \leq 120, n \neq i} L_n$ 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. Columns 1-6 replicate estimates in columns 4-9 of Table 2 only using the subsample of counties for which estimates of the housing supply elasticity are available. Columns 7-12 replicate columns 1-6 introducing the housing supply elasticity. Standard errors are clustered by state. * denotes significance at the 5 percent level; ** denotes significance at the 1 percent level.

Table C.2: Explaining the general equilibrium local employment elasticities to a 5 percent productivity shock in the Saiz subsample

C.6 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 in the paper, we start with the initial equilibrium in the observed data. We first undertake a counterfactual for prohibitive commuting costs between counties ($\kappa_{ni} \rightarrow \infty$ 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.7 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 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 with commuting between counties (blue solid line). 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.7, 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.

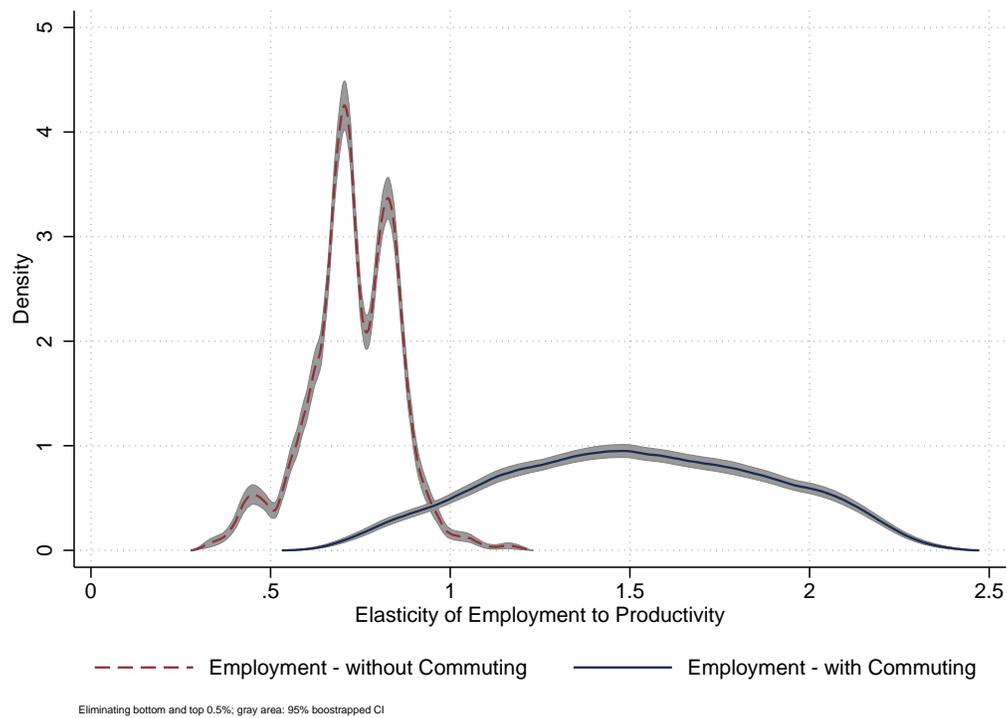


Figure C.7: Kernel density for the distribution of employment and resident elasticities in response to a productivity shock across counties (with and without commuting between counties)

C.7 Million Dollar Plants Natural Experiment

In this section of the web appendix, we report additional results for the MDP experiment from Section 5 of the paper.⁷ First, we report a balance table that compares the observed characteristics of winner and runner-up counties before a MDP announcement for the full set of 82 cases. Second, we consider the non-parametric specification that estimates separate treatment effects for each MDP case from equation (22) in the paper.

First, we compare the observed characteristics of winner and runner-up counties before a MDP announcement. Table C.3 reports the mean and standard error of the mean for employment, wages, land area and population density five years before a MDP announcement for these two groups of counties for the full set of 82 MDPs. We also report the same statistics for workplace and residence own commuting shares in 1990 as the closest Census year.⁸ We find that winner counties have somewhat lower prior values of levels of employment, wages, population and population density than runner-up counties. We also find that they have somewhat more open local labor markets in terms of workplace and residence own commuting shares. Despite these differences in individual observed characteristics, the fact that the firms selected these counties as winners and runners-up suggests that they have similar implied profitability for plant location. As a check on the identifying assumption that the losers form a valid counterfactual for the winners, we report an event-study specification following GHM in Section 5 of the paper.

Variable	Winner	Runners-up
Log employment	11.122 (0.176)	11.660 (0.116)
Log wages	2.758 (0.032)	2.802 (0.023)
Log land area	14.213 (0.085)	14.152 (0.063)
Log Population	11.999 (0.153)	12.446 (0.100)
Log population density	-2.214 (0.153)	-1.706 (0.115)
Workplace own commuting share	0.742 (0.015)	0.764 (0.011)
Residence own commuting share	0.737 (0.020)	0.786 (0.015)

Means and standard errors of the mean of observed characteristics; standard errors of the means are in parentheses; employment, wages, land area, population and population density for winner and runner-up counties in each case are measured five years before the MDP announcement; workplace and residence commuting shares are measured in 1990.

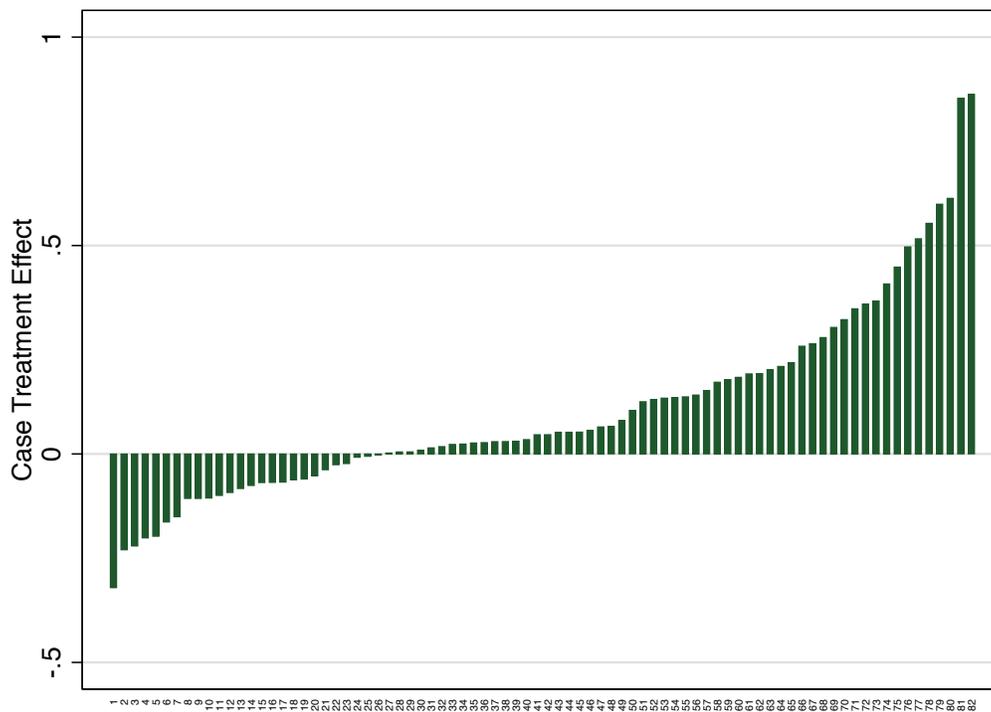
Table C.3: Characteristics of Winner and Runner-up Counties Before a MDP Announcement

Second, we turn to the heterogeneous treatment effects specification from equation (22) in the paper, which estimates a separate treatment effect for each of the 82 MDP cases. These heterogeneous treatment effects are identified as the mean change in employment in winner counties relative to control counties for each case (the excluded category is the runner-up counties for each case). In Figure C.8 below, we display these estimated treatment effects for each case. As apparent from the figure, we find substantial heterogeneity in these estimated treatment effects, which range from less than zero to just below one.

⁷See Section D.2 of this web appendix for further discussion of the data sources for this section.

⁸The MDP treatment years range from 1982 to 1993 with a median of 1989.

Therefore, although the average estimated treatment effect is positive, there is substantial variation around this average. We reject the null hypothesis that these estimated treatment effects all take the same value at conventional levels of statistical significance (p-value 0.000).



Note: Heterogeneous treatment effects for 82 cases. Specification includes county, case and year fixed effects, a post-MDP announcement dummy, interaction terms between the dummy for winner county and treatment year, interaction terms between commuting openness and treatment year, and three-way interaction terms between the winner dummy, commuting openness and treatment year (equation (22) in the paper).

Figure C.8: Heterogeneous Treatment Effects Across MDPs

C.8 Shift-Share Decomposition

In this subsection of the web appendix, we provide further evidence on the importance of commuting for employment changes using shift-share decompositions, as discussed in the paper. We undertake these decompositions for both cross-section and time-series variation in employment.

C.8.1 Cross-section Decomposition

We begin with our cross-section decomposition. We use the accounting identity provided by the commuter market clearing condition, which requires that employment in each county i equals the sum of commuting flows from all counties:

$$L_{it} = \sum_{n \in N} \lambda_{ni|nt}^R R_{nt}. \quad (\text{C.3})$$

Separating these commuting flows into those from the own county and those from other counties, this commuter market clearing condition can be re-written as:

$$L_{it} = \underbrace{\lambda_{ii|it}^R R_{it}}_{\text{(a) own residents}} + \underbrace{\sum_{n \neq i} \lambda_{ni|nt}^R R_{nt}}_{\text{(b) commuters}}. \quad (\text{C.4})$$

The same accounting also holds for the county with the median level of employment m :

$$L_{mt} = \underbrace{\lambda_{mm|mt}^R R_{mt}}_{\text{(a) own residents}} + \underbrace{\sum_{n \neq m} \lambda_{nm|nt}^R R_{nt}}_{\text{(b) commuters}}. \quad (\text{C.5})$$

Taking differences between equations (C.4) and (C.5), we obtain:

$$\Delta^I L_{it} = \left[\lambda_{ii|it}^R R_{it} - \lambda_{mm|mt}^R R_{mt} \right] + \left[\sum_{n \neq i} \lambda_{ni|nt}^R R_{nt} - \sum_{n \neq m} \lambda_{nm|nt}^R R_{nt} \right], \quad (\text{C.6})$$

where Δ^I is the cross-section difference operator between an individual county i and the county with the median level of employment m (such that $\Delta^I L_{it} = L_{it} - L_{mt}$). Subtracting and adding $\lambda_{ii|it}^R R_{mt}$ from the first term in square parentheses, and subtracting and adding $\sum_{n \neq m} \lambda_{ni|nt}^R R_{nt}$ from the second term in square parentheses, we have:

$$\begin{aligned} \Delta^I L_{it} &= \lambda_{ii|it}^R R_{it} - \lambda_{ii|it}^R R_{mt} - \lambda_{ii|mt}^R R_{mt} + \lambda_{ii|it}^R R_{mt} \\ &\quad + \sum_{n \neq i} \lambda_{ni|nt}^R R_{nt} - \sum_{n \neq m} \lambda_{ni|nt}^R R_{nt} - \sum_{n \neq m} \lambda_{nm|nt}^R R_{nt} + \sum_{n \neq m} \lambda_{ni|nt}^R R_{nt}. \end{aligned} \quad (\text{C.7})$$

which can be re-written as:

$$\Delta L_{it} = \underbrace{\lambda_{ii|it}^R \Delta^I R_{it}}_{\text{(i) own residents}} + \underbrace{R_{mt} \Delta^I \lambda_{ii|it}^R}_{\text{(ii) own commuting shares}} + \underbrace{\left(\sum_{n \neq i} \lambda_{ni|nt}^R R_{nt} - \sum_{n \neq m} \lambda_{ni|nt}^R R_{nt} \right)}_{\text{(iii) other residents}} + \underbrace{\sum_{n \neq m} \left(\lambda_{ni|nt}^R - \lambda_{nm|nt}^R \right) R_{nt}}_{\text{(iv) other commuting shares}} \quad (\text{C.8})$$

We thus obtain a decomposition of cross-section differences in employment between counties into the following four contributions: (i) differences in own residents holding own commuting shares constant; (ii) differences in own commuting shares holding own residents constant; (iii) differences in other residents holding other commuting shares constant; and (iv) differences in other commuting shares holding other residents constant. In the other residents term (iii), the only thing that varies between the two components of the term is the lower limit of the summation, which captures differences in the sets of other counties $n \neq i$ and $n \neq m$. In the other commuting term (iv), the only thing that varies between the two components of the term is the commuting shares with other counties: $\lambda_{ni|nt}^R \neq \lambda_{nm|nt}^R$ for $i \neq m$. All four terms are equal to zero for the county m with median employment, we report the distribution of results for all other counties $i \neq m$.

In interpreting this decomposition, several points are worth bearing in mind. First, we interpret any term involving workers commuting across county borders as capturing commuting, which implies that we view terms (ii), (iii) and (iv) as capturing commuting. In the special case of no commuting between counties, these final three terms are all necessarily equal to zero, because in this special case $\lambda_{ii|it}^R = 1$ and $\lambda_{ni|nt}^R = \lambda_{nm|nt}^R = 0$ for $n \neq i$ and $n \neq m$. Nevertheless, we acknowledge that other interpretations are possible, such as only viewing terms involving variation in commuting shares ($\lambda_{ni|nt}^R$) as capturing commuting (terms (ii) and (iv)). Second, we note that counties with similar total residents to the median county might have very different commuting links with other counties. We intentionally capture this variation in our decomposition, but we acknowledge that the importance of these commuting links can be affected by many idiosyncratic factors, such as the drawing of county boundaries.

Third, in principle, the relative importance of the different terms in the cross-section decomposition in equation (C.8) can vary depending on which county is chosen as the base. However, in practice, we find a similar qualitative and quantitative pattern of results for alternative choices of the base county other than the county with median employment. Fourth, we recognize that total employment and total residents are still strongly positively correlated across counties, because of third factors that affect both employment and residents (e.g. productivity and climate).

In Table C.4, we report the results of this cross-section decomposition (C.8) using our bilateral commuting data for 2006-10. As all four terms are equal to zero for the county with median employment, we report the distribution of results across all other counties. For each individual county, the four terms add up to the total difference in employment, which in turn implies that the mean of the four terms adds up to the mean total difference in employment (bottom row), because the mean is a linear operator. The same need not be true for the percentiles of the distribution of each contribution (other rows), because the county at a given percentile for one contribution may be different from the county at the same percentile for another contribution. Each individual term in the decomposition can be positive or negative, as reflected

in the negative values in a number of cells in the table.

As apparent from the table, we find substantial contributions from all four terms of the decomposition. On average, we find a difference in employment from the median county of 45,211, where the fact that this difference is positive reflects the fact that the distribution of employment across counties is skewed. Of this 45,211, we find that own residents contribute 32,712 (bottom row, second column), own commuting is responsible for 2,712 (bottom row, third column), other residents make a negative contribution of -44,119 (bottom row, fourth column), and other commuting accounts for the remaining 53,906 (bottom row, fifth column). We also find substantial heterogeneity across counties in the relative importance of these four terms. Between the 10th and 90th percentiles, these contributions range from -10,048 to 73,406 for own residents (second column), -1,145 to 6,350 for own commuting (third column), -86,445 to -2,063 for other residents (fourth column), and -1,339 to 108,003 for other commuting (fifth column).

Although the individual terms in these shift-share decompositions can be interpreted in different ways, we view these results as supporting the idea that commuting patterns are a quantitatively relevant margin for accounting for cross-sectional differences in employment across counties. In Figure 1 and Table 1 in the paper, we report results using a simpler and more intuitive measure of the relevance of commuting, given by the share of residents who work in the county where they live.

2006-10	(i) Changes Own Residents, Constant Own Commuting	(ii) Changes Own Commuting, Constant Own Residents	(iii) Changes Other Residents, Constant Other Commuting	(iv) Changes Other Commuting, Constant Other Residents	Sum (i)-(iv)
10th percentile	-10,048	-1,145	-86,445	-1,339	-
25th percentile	-6,242	613	-26,593	1,372	-
50th percentile	-1,077	2,917	-9,822.9	8,693	-
75th percentile	14,744	4,999	-4,157.7	30,616	-
90th percentile	73,406	6,350	-2,063	108,003	-
Mean	32,712	2,712	-44,119	53,906	45,211

Mean and percentiles of the distribution of the contributions to cross-section differences in employment between each county and the median county for 2006-10. The four terms are differences in (i) the number of residents holding own commuting shares constant; (ii) own commuting shares holding own residents constant; (iii) other residents holding other commuting shares constant; and (iv) other commuting shares holding other residents constant.

Table C.4: Cross-section Decomposition of Employment Differences across Counties for 2006-10

C.8.2 Time-series Decomposition

We next consider our time-series decomposition. Taking differences between equation (C.4) for time t and the analogous equation for time $t - 1$, we obtain:

$$\Delta^T L_{it} = \lambda_{ii|it}^R R_{it} - \lambda_{ii|it-1}^R R_{it-1} + \sum_{n \neq i} \lambda_{ni|nt}^R R_{nt} - \sum_{n \neq i} \lambda_{ni|nt-1}^R R_{nt-1}, \quad (\text{C.9})$$

where Δ^T is the time-series difference operator such that $\Delta^T L_{it} = L_{it} - L_{it-1}$. Subtracting and adding $\lambda_{ii|it}^R R_{it-1}$ from the first term in parentheses, and subtracting and adding $\sum_{n \neq i} \lambda_{ni|nt}^R R_{nt-1}$ from the second

term in parentheses, we obtain:

$$\begin{aligned} \Delta^T L_{it} &= \lambda_{ii|it}^R R_{it} - \lambda_{ii|it}^R R_{it-1} - \lambda_{ii|it-1}^R R_{it-1} + \lambda_{ii|it}^R R_{it-1} \\ &+ \sum_{n \neq i} \lambda_{ni|nt}^R R_{nt} - \sum_{n \neq i} \lambda_{ni|nt}^R R_{nt-1} - \sum_{n \neq i} \lambda_{ni|nt-1}^R R_{nt-1} + \sum_{n \neq i} \lambda_{ni|nt}^R R_{nt-1}. \end{aligned} \quad (\text{C.10})$$

which can be re-written as:

$$\Delta^T L_{it} = \underbrace{\lambda_{ii|it}^R \Delta^T R_{it}}_{\text{(i) own residents}} + \underbrace{R_{it-1} \Delta^T \lambda_{ii|it}^R}_{\text{(ii) own commuting shares}} + \underbrace{\sum_{n \neq i} \lambda_{ni|nt}^R \Delta^T R_{nt}}_{\text{(iii) other residents}} + \underbrace{\sum_{n \neq i} R_{nt-1} \Delta^T \lambda_{ni|nt}^R}_{\text{(iv) other commuting shares}}. \quad (\text{C.11})$$

In interpreting this decomposition, we again view any term involving workers commuting across county borders as capturing commuting (terms (ii), (iii) and (iv)). In the special case of no commuting between counties, the first term for changes in own residents ($\lambda_{ii|it}^R \Delta^T R_{it}$) is the only source of employment changes, because in this special case $\lambda_{ii|it}^R = 1$, $\lambda_{ni|nt}^R = 0$ for $n \neq i$, and $\Delta^T \lambda_{ni|nt}^R = 0$ for all n, i , which implies that the final three terms are necessarily all equal to zero. But we acknowledge that other interpretations are possible, such as only viewing terms that involve variation in commuting shares ($\lambda_{ni|nt}^R$) as capturing commuting (terms (ii) and (iv)).

In Table C.5, we report the results of this time-series decomposition (C.11) using the change in the bilateral commuting probabilities between 1990 and 2006-2010. As for our cross-section decomposition above, the four terms add up to the total change in employment for an individual county. However, the same need not be true for the percentiles of the distribution of each contribution (other rows), because the county at a given percentile for one contribution may be different from the county at the same percentile for another contribution. Each individual term in the decomposition can be positive or negative, as reflected in the negative values in a number of cells in the table.

1990 to 2006-10	(i) Changes Own Residents, Constant Own Commuting	(ii) Changes Own Commuting, Constant Own Residents	(iii) Changes Other Residents, Constant Other Commuting	(iv) Changes Other Commuting, Constant Other Residents
10th percentile	23	-2,619	28	-184
25th percentile	474	-1,155	130	32
50th percentile	1,728	-457	447	335
75th percentile	6,094	-82	1,517	1,109
90th percentile	21,170	181	5,626	3,268

Mean and percentiles of the distribution of the contributions to time-series changes in employment between 1990 and 2006-10 from (i) the number of residents holding own commuting shares constant; (ii) own commuting shares holding own residents constant; (iii) other residents holding other commuting shares constant; and (iv) other commuting shares holding other residents constant.

Table C.5: Time-series Decomposition of County Employment Changes between 1990 and 2006-10

We again find quantitatively relevant contributions from all four terms in the decomposition. For the median county, we find a change in employment of 1,981, of which own residents contribute 1,728 (fourth row, second column), own commuting is responsible for -457 (fourth row, third column), other residents make a contribution of 447 (fourth row, fourth column), and other commuting accounts for the remaining 335 (fourth row, fifth column). We also find substantial heterogeneity across counties in the

relative importance of these four terms. Between the 10th and 90th percentiles, these contributions range from 23 to 21,170 for own residents (second column), -2,619 to 181 for own commuting (third column), 28 to 5,626 for other residents (fourth column), and -184 to 3,268 for other commuting (fifth column).

In both the cross-section and over time, variation in county employment is ultimately driven by variation in productivity and other county characteristics. Therefore, notwithstanding the caveats discussed above, we view the findings of these cross-section and time-series decompositions as supporting the idea that response of employment to such county characteristics is shaped by heterogeneous patterns of commuting flows.

C.9 Changes in Commuting Costs

Figure C.9 presents the changes in local employment against the initial labor to resident ratio (L_i/R_i) for the counterfactual in which we reduce commuting cost by the median change between 1990 and 2010, as discussed in Section 6 of the paper.

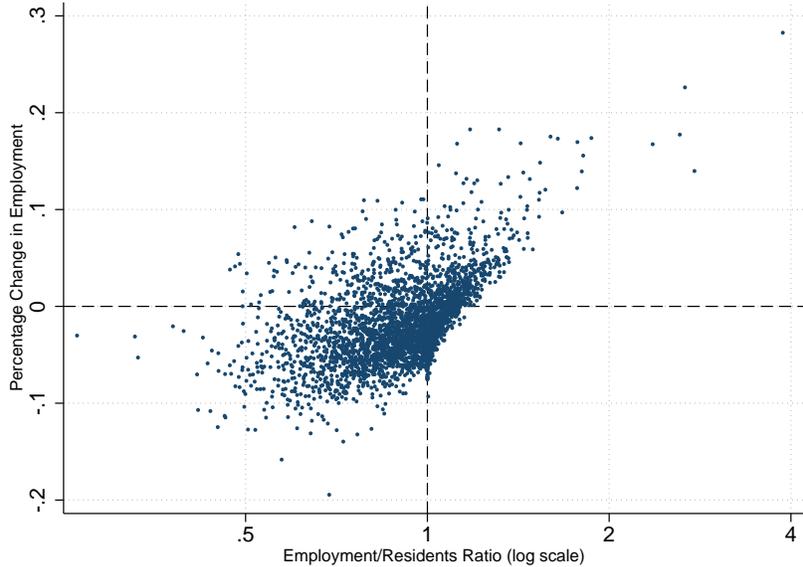


Figure C.9: Counterfactual relative change in county employment (\hat{L}) for median decrease in commuting costs throughout U.S. against initial employment to residents ratio (L/R).

C.10 Interaction Between Trade and Commuting Costs

In this subsection of the web appendix, we examine the extent to which trade and commuting costs interact in the model, as discussed in Section 6 of the paper. 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.

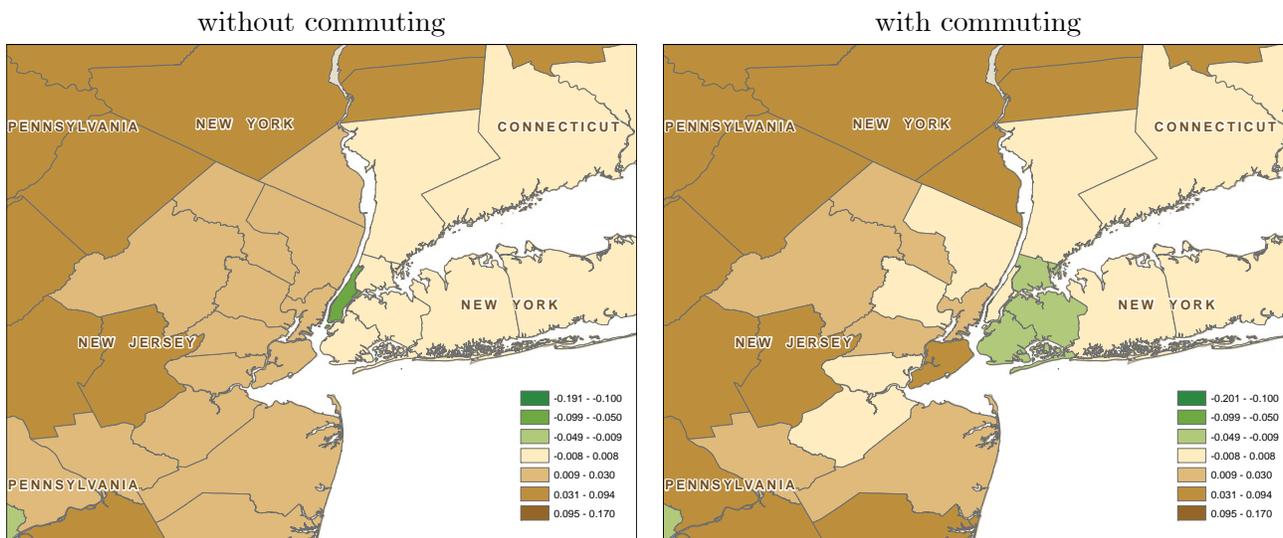


Figure C.10: Relative change in employment (\hat{L}) from a 20 percent reduction in trade costs (with and without commuting between counties) in the New York area

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.

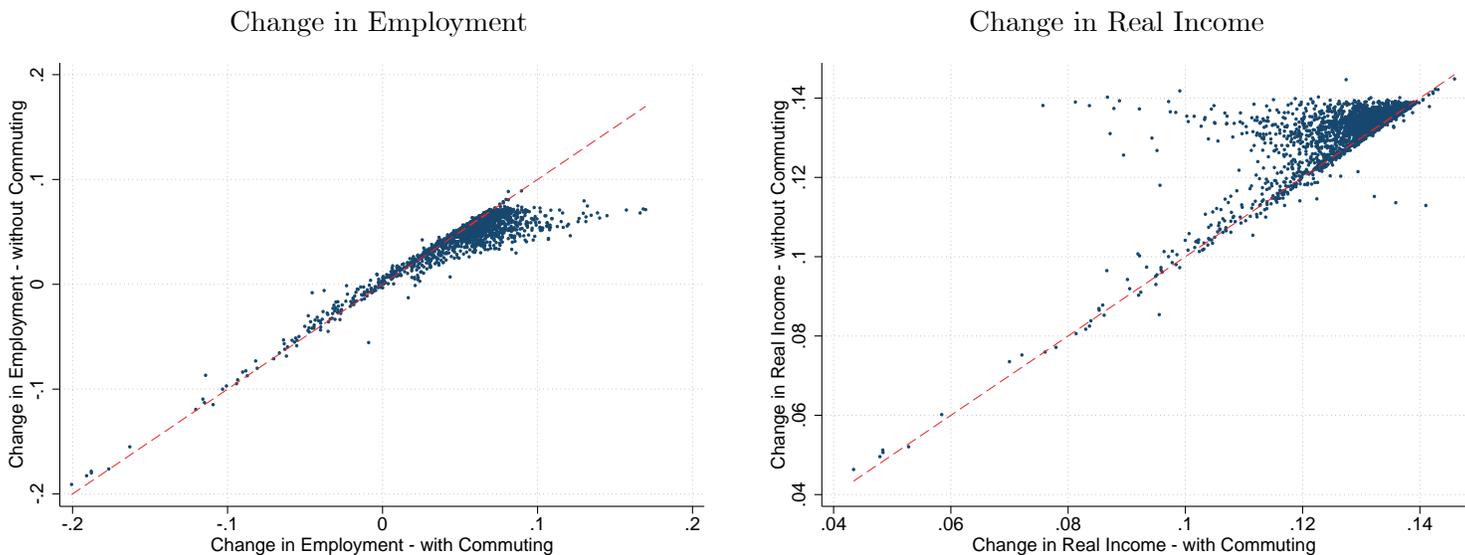


Figure C.11: Relative change in employment (\hat{L}) and real income ($\hat{v}_n / (\hat{P}_n^\alpha \hat{Q}_n^{1-\alpha})$) from a 20 percent reduction in trade costs across all counties (with and without commuting between counties)

C.11 Commuting Zones (CZs)

As discussed in the paper, previous research has often worked at relatively high levels of spatial aggregation (e.g. commuting zones (CZs)) to reduce commuting flows. In contrast, we explicitly model the spatial interactions between locations in goods and commuting markets, thereby providing a framework for examining the local impact of labor demand shocks at alternative spatial scales, including those finer than CZs.

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 and Section B.5 of this web appendix. 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 and Section B.6 of this web appendix. 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 B.5 in this web appendix, 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.

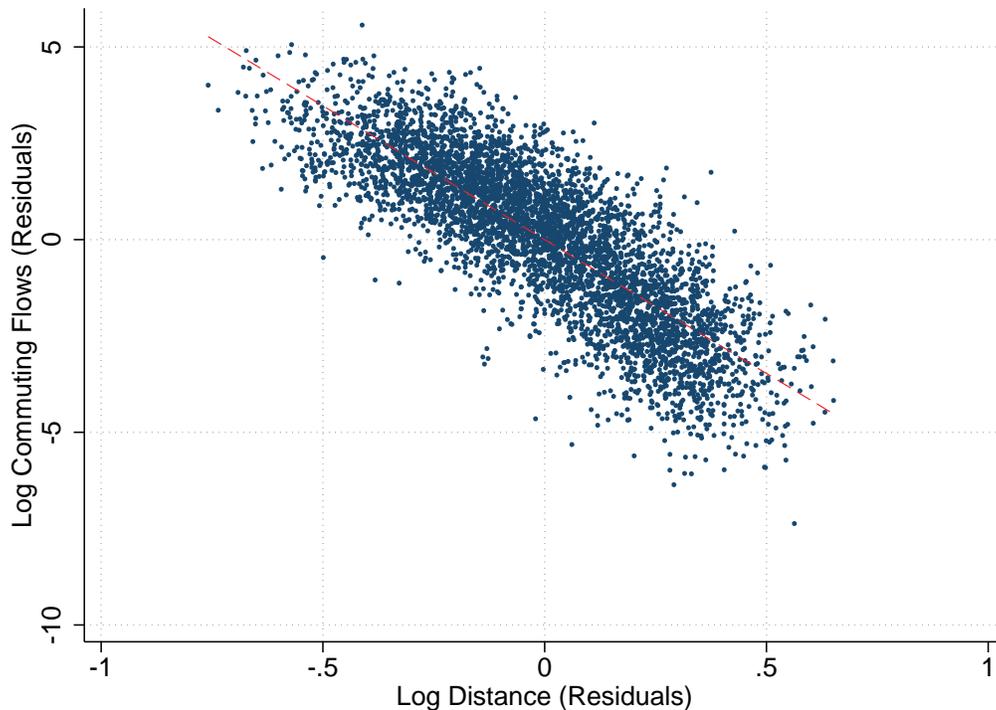


Figure C.12: Gravity in Commuting Between Commuting Zones (CZs)

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 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 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.6, 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

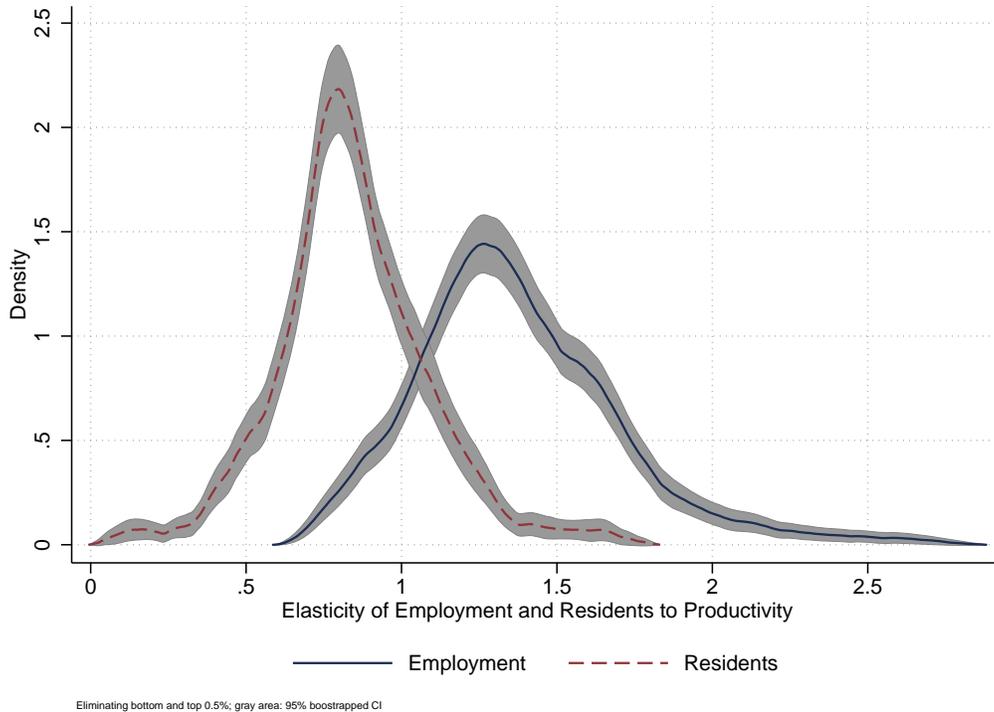


Figure C.13: Kernel density for the distribution of employment and residents elasticities in response to a productivity shock across CZs

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 (λ_{ii}^R) in Column (5) as a 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 6 of the

	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**
		(0.011)	(0.022)	(0.018)				(0.017)	(0.017)
$\log w_i$			-0.037	-0.168				-0.002	-0.020
			(0.176)	(0.136)				(0.088)	(0.089)
$\log H_i$			-0.166**	-0.087				-0.010	-0.011
			(0.042)	(0.049)				(0.023)	(0.023)
$\log L_{-i}$				0.081**				-0.038*	-0.040*
				(0.023)				(0.015)	(0.016)
$\log \bar{w}_{-i}$				0.107				0.012	0.036
				(0.146)				(0.107)	(0.108)
λ_{ii}^R					-3.434**				
					(0.216)				
$\sum_{n \in N} (1 - \lambda_{Rni}) \vartheta_{ni}$						8.815**		9.936*	
						(2.887)		(3.868)	
$\vartheta_{ii} \left(\frac{\lambda_{ii}}{\lambda_{Ri}} - \lambda_{Li} \right)$						6.044*		6.670	
						(2.916)		(3.836)	
$\frac{\partial w_i}{\partial A_i} \frac{A_i}{w_i}$						-1.624**		-0.997**	
						(0.194)		(0.333)	
$\frac{\partial w_i}{\partial A_i} \frac{A_i}{w_i} \cdot \sum_{r \in N} (1 - \lambda_{rn r}) \vartheta_{rn}$							1.345**		2.546**
							(0.210)		(0.323)
$\frac{\partial w_i}{\partial A_i} \frac{A_i}{w_i} \cdot \vartheta_{ii} \left(\frac{\lambda_{ii}}{\lambda_{Ri}} - \lambda_{Li} \right)$							-1.391**		-0.680*
							(0.196)		(0.307)
Constant	1.376**	1.098**	2.779	1.747	4.522**	-3.459	2.347**	-4.959	1.387
	(0.031)	(0.142)	(1.810)	(1.698)	(0.194)	(2.824)	(0.185)	(3.920)	(1.321)
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

Note: $L_{-i} \equiv \sum_{n: d_{ni} \leq 120, n \neq i} L_n$ 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. Standard errors are clustered by state; when a CZ overlaps different states, the state that accounts for most of the CZ population is assigned. * denotes significance at the 5 percent level; ** denotes significance at the 1 percent level.

Table C.6: Explaining the general equilibrium local employment elasticities to a 5 percent productivity shock for commuting zones (CZs)

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 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.7, 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

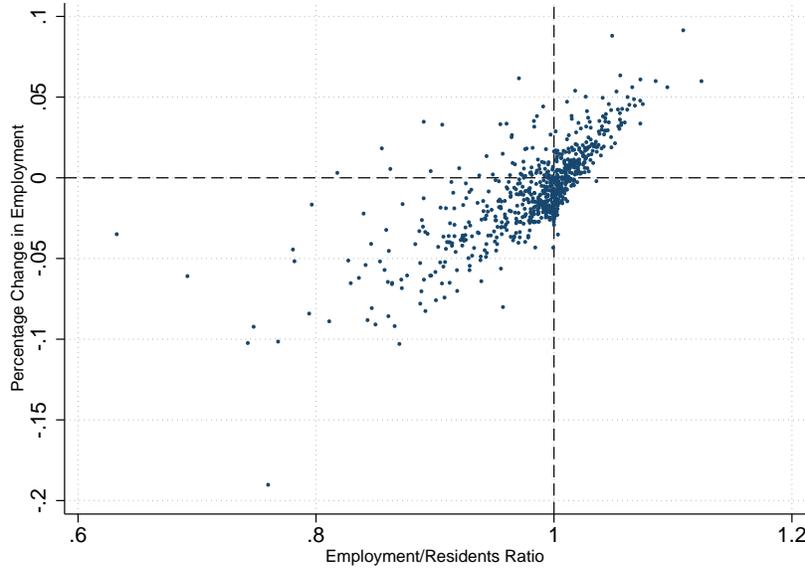


Figure C.14: Counterfactual relative change in commuting zone (CZ) employment (\hat{L}) from median proportional reduction in commuting costs ($\hat{B}_{ni} = 0.88$) and initial dependence on commuting

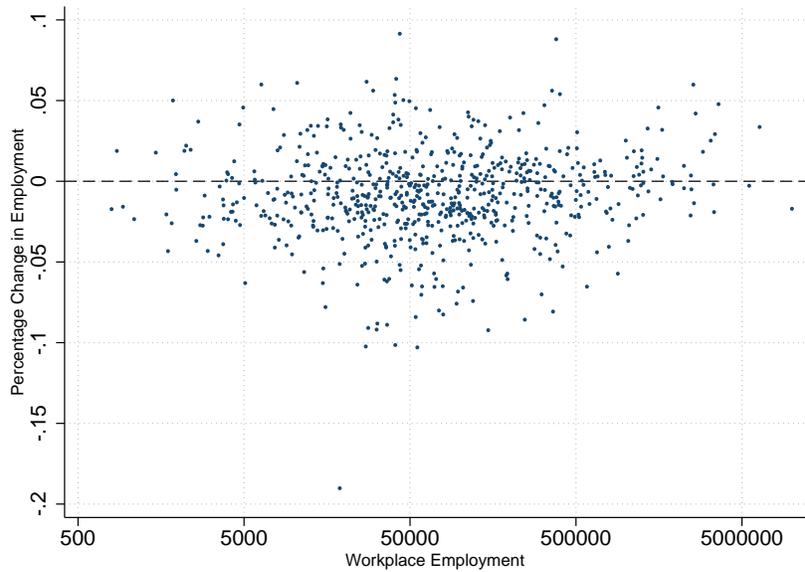


Figure C.15: Counterfactual relative change in commuting zone (CZ) employment (\hat{L}) from median proportional reduction in commuting costs ($\hat{B}_{ni} = 0.88$) and initial employment size

B.2 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. 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.7 amounts to more than 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.

	1	2	3	4	5	6	7	8	9	10	11	12
Dep. 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	L_i/R_i	L_i/R_i	L_i/R_i	L_i/R_i
$\log R_i$	0.991** (0.002)		0.992** (0.003)			-0.004** (0.001)		0.001 (0.002)		0.049** (0.010)		0.048** (0.015)
$\log w_i$	0.116** (0.024)		0.161** (0.027)		0.099** (0.020)		0.123** (0.022)		0.195* (0.074)		0.196* (0.074)	
$\log L_i$		1.001** (0.001)		0.993** (0.003)	-0.006** (0.002)		-0.002 (0.002)		0.025** (0.008)		0.029** (0.009)	
$\log \bar{v}_i$		-0.042** (0.015)		-0.057* (0.028)		0.057** (0.014)		0.078** (0.022)		0.014 (0.058)		0.014 (0.058)
$\log H_i$			0.006 (0.007)	0.005 (0.007)			-0.001 (0.006)	0.001 (0.006)	-0.005 (0.011)	-0.014 (0.015)	-0.009 (0.013)	-0.014 (0.018)
$\log R_{-i}$			-0.001 (0.003)				0.638** (0.107)	0.693** (0.113)	0.428* (0.168)	0.499* (0.196)	0.443* (0.188)	0.497* (0.195)
$\log \bar{w}_{-i}$			-0.149** (0.028)				1.007** (0.219)	1.017** (0.222)	0.934 (0.502)	0.890 (0.489)	0.963 (0.559)	0.887 (0.520)
$\log L_{-i}$				0.010** (0.003)			-0.644** (0.107)	-0.700** (0.114)	-0.414* (0.168)	-0.489* (0.195)	-0.431* (0.190)	-0.488* (0.193)
$\log \bar{v}_{-i}$				0.106** (0.029)			-1.119** (0.226)	-1.111** (0.230)	-1.176* (0.517)	-1.097* (0.500)	-1.207* (0.574)	-1.093* (0.535)
Saiz elasticity											0.005 (0.008)	-0.000 (0.008)
Constant	-1.146** (0.237)	0.450** (0.154)	-0.095 (0.159)	-0.591** (0.183)	0.007 (0.202)	0.425** (0.147)	0.974** (0.191)	1.227** (0.199)	0.993 (0.637)	2.389** (0.529)	0.993 (0.646)	2.388** (0.532)
R^2	1.00	1.00	1.00	1.00	0.07	0.02	0.31	0.26	0.61	0.52	0.61	0.52
N	709	709	636	636	709	709	636	636	110	110	110	110

Note: $L_{-i} \equiv \sum_{n:d_{ni} \leq 120, n \neq i} L_n$ 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} . Columns 1-8 are unweighted regressions. Columns 9 and 10 repeat the most complete specifications in columns 7 and 8 giving to each CZ a weight proportional to the fraction of the CZ's population living in counties where we have data on land supply elasticity; this process excludes from the regressions CZ for which no county has data on land supply elasticity. Columns 11 and 12 then repeat columns 9 and 10 adding the Saiz land supply elasticity as a regressor. Land supply elasticity for a CZ is the population-weighted average of its counties' land supply elasticities. Standard errors are clustered by state. * denotes significance at the 5 percent level; ** denotes significance at the 1 percent level.

Table C.7: Explaining employment levels and commuting intensity for commuting zones (CZs)

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.

D Data Appendix

This section of the web appendix contains further information on the data sources and definitions, as well as additional details about the construction of figures and tables. In Section D.1, we discuss the data used for the quantitative analysis of the model in Sections 3-4 of the paper. In Section D.2, we discuss the data used to provide independent evidence in support of the model's predictions in Section 5 of the paper and Sections C.7 and C.8 of this web appendix.

D.1 Quantitative Analysis of the Model (Sections 3-4 of the paper)

D.1.1 Data Sources and Definitions

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

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

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

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

County Land Area, County Centroids. This data comes from the 2010 Census Gazetteer Files. 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 4.2 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.1.2 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:

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

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:

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

where $\hat{\lambda}_{ij} = \exp(\ln \hat{\lambda}_{ij})$ 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:

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

The total business trips originating from residence i are then $\sum_j (1 - \hat{s}_{ij}^C) \tilde{\lambda}_{ij}$. For any residence i , we reimpute these business trips across destinations j in proportion to the estimated workplace composition of the residence i , $\hat{\lambda}_{ij}^C / \sum_i \hat{\lambda}_{ij}^C$. 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.⁹ 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.

⁹For 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.

D.1.3 Further Information on Figures and Tables

We now report additional technical details related to the data sources and manipulation for some of the tables and figures in the paper or this web appendix.

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.¹⁰

Figure 1. This figure reports kernel densities of the distribution of the share of a county’s residents working in their county of residence for 4 decades. Data on the share of residents working in the county are constructed from the ICPSR Study 7736¹¹, and the 1983, 1994 and 2000 editions of the “County and City Data Book” published by the U.S. Department of Commerce.

Figure B.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 Statistics¹² 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 B.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 B.3. This figure reports kernel densities analogous to Figure 1 that are weighted by the number of residents in each county, and it shares with that Figure the data source.

Figure B.5. 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

¹⁰See <http://www.ers.usda.gov/data-products/commuting-zones-and-labor-market-areas.aspx>.

¹¹United States Department of Commerce. Bureau of the Census. County and City Data Book [United States] Consolidated File: County Data, Bibliographic Citation: 1947-1977. ICPSR07736-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2012-09-18. <http://doi.org/10.3886/ICPSR07736.v2>

¹²See http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_atlas_database/2013/polygon.html

compute the population weighted average of the median values.

D.2 Additional Empirical Evidence

We now discuss the data sources and definitions for the independent evidence in support of the predictions of the model in Section 5 of the paper and Sections C.7 and C.8 of this web appendix.

Commuting Data (Section 5 of the paper and Sections C.7 and C.8 of this web appendix).

We construct three bilateral commuting matrices for 1990, 2000 and 2006-2010. We use these matrices for both Section 5 of the paper and C.8 of this web appendix. Our bilateral commuting data comes from the County-to-County Worker Flows tabulation files based on the U.S. Census (for years 1990 and 2000) and American Community Survey (for 2006-2010). We construct commuting flows following the same procedure indicated in Section D.1.2 for the contiguous United States. We compute distances between county centroids using the coordinates provided in the corresponding years of the Census Gazetteer files. To construct a balanced panel of counties over time, some aggregation of counties is needed, and we end up with a cross-section of 3,108 spatial units for all three years.

Million Dollar Plants (Section 5 of the paper and Section C.7 of this web appendix). We use the full list of 82 plants openings gathered by Greenstone and Moretti (2004) from the Journal *Site Selection*. For each county, yearly workplace employment is taken from the Bureau of Economic Analysis, County Economic Profiles (Table CA30). In particular, we use the measure of Wage and Salary Employment (data line 250). This measure includes “All jobs for which wages and salaries are paid are counted”, which cover all industries covered by Unemployment Insurance, plus adjustments for industries not fully covered by Unemployment Insurance as detailed in the “Local Area Personal Income Methodology” (November 2016) from the BEA. In weighted regressions, the population at the beginning of the sample for each county also comes from the same BEA source (line 100). For each county, the measured own commuting share is for the closest available year to the plant opening date from the commuting data discussed at the beginning of this subsection. For all 82 plant openings, the closest available year is 1990. To control for industry-year fixed effects, we assign industries to cases using the reported industry for each case from Appendix Table 2 in Greenstone and Moretti (2004). Cases are classified into 5 broad industries: Manufacturing (63 cases), Financial (1 case), Services (6 cases), Trade (4 cases), Transportation and Utilities (8 cases).

Shift-share Decompositions (Section C.8 of this web appendix). We use the bilateral matrix for 2006-2010 for the cross-section decomposition and the bilateral matrices for 1990 and 2006-2010 for the time-series decomposition.

References

- Allen, Treb, Costas Arkolakis and Xiangliang Li (2016) “On the Existence and Uniqueness of Trade Equilibria,” Yale University, mimeograph.
- 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.
- Grubel, Herbert G. and Peter Lloyd (1975) *Intra-Industry Trade: The Theory and Measurement of International Trade in Differentiated Products*, London: MacMillan.
- Hsieh, Chang-Tai and Enrico Moretti (2017) “Housing Constraints and Spatial Misallocation,” University of California, Berkeley, mimeograph.
- 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.