SUPPLEMENT TO 'SPATIAL ECONOMICS FOR GRANULAR SETTINGS'

JONATHAN I. DINGEL
Department of Economics, Columbia University, NBER, and CEPR

FELIX TINTELNOT
Department of Economics, Duke University, NBER, and CEPR

APPENDIX C: THEORY: PROOFS AND EXTENSIONS

Section C.1 provides a condition on parameter values sufficient for the existence and uniqueness of the continuum model's equilibrium. Section C.2 details the exact hat algebra for computing counterfactual outcomes. Section C.3 provides further details on the calibrated-shares procedure. Section C.4 presents a special case in which each counterfactual change is simply proportional to the corresponding baseline share. Section C.5 provides analytical expressions for the slope coefficients in this special case. Section C.6 derives expected squared errors for baseline shares. Section C.7 derives expressions for welfare changes. Section C.8 shows that the trade equilibrium of the model with a finite number of individuals is unique. Section C.9 shows that the continuum model is a limiting case of the model with a finite number of individuals. Section C.10 analytically characterizes the uncertainty in counterfactual changes caused by individuals' idiosyncratic preferences in a two-workplace economy. The following subsection presents extensions of the bare-bones model in Section 4 that introduce trade costs, the use of land in production, residential amenities, and local increasing returns.

C.1. *Existence and uniqueness of continuum model's equilibrium*

We derive a set of parameter values sufficient for the existence and uniqueness of the continuum model's equilibrium. As shown in Theorem 1 of Allen et al. (2023), the spectral radius of the elasticity matrix characterizes the existence and uniqueness properties of the system.

The equilibrium of the continuum model can be written as

$$\frac{r_k^{1+\alpha\epsilon} T_k}{\alpha} = \sum_n D_{kn} w_n^{1+\epsilon} \tag{C.1}$$

$$L_n w_n^{-\epsilon} = \sum_k D_{kn} r_k^{-\alpha\epsilon} \tag{C.2}$$

$$A_n^{1-\sigma} L_n w_n^{\sigma} = P^{\sigma-1} \sum_{n'} L_{n'} w_{n'}, \tag{C.3}$$

where, for notational convenience, we define

$$\Theta \equiv \sum_k \sum_n w_n^{\epsilon} (r_k^{\alpha} \delta_{kn})^{-\epsilon}$$
$$D_{kn} \equiv L \Theta^{-1} \delta_{kn}^{-\epsilon} \bar{\delta}_{kn}^{-1}.$$

Equations (C.1)–(C.3) are $3 \times N$ equations in $3 \times N$ unknowns $\{r_n, w_n, L_n\}_{n=1,2,\ldots,N}$ given values for Θ and P. Although Θ and P are endogenous, they do not vary across locations. Accordingly, they fall under the class of endogenous scalars discussed in Remark 4 of Allen et al. (2023), which leave the conclusion of their Theorem 1 Part ii.b unchanged.

Following Remark 5 of Allen et al. (2023), we construct the intermediate matrices $\mathbf{\Gamma}$ and $\mathbf{B}$:

$$\mathbf{\Gamma} = \begin{bmatrix} 1+\alpha\epsilon & 0 & 0 \\ 0 & -\epsilon & 1 \\ 0 & \sigma & 1 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 0 & 1+\epsilon & 0 \\ -\alpha\epsilon & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

Then we can construct the final matrix $\mathbf{A}$ defined by $(\mathbf{A})_{hh'} \equiv |(\mathbf{B}\mathbf{\Gamma}^{-1})_{hh'}|$:

$$\mathbf{B}\mathbf{\Gamma}^{-1} = \begin{bmatrix} 0 & 1+\epsilon & 0 \\ -\alpha\epsilon & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{1+\alpha\epsilon} & 0 & 0 \\ 0 & -\frac{1}{\sigma+\epsilon} & \frac{1}{\sigma+\epsilon} \\ 0 & \frac{\sigma}{\sigma+\epsilon} & \frac{\epsilon}{\sigma+\epsilon} \end{bmatrix}, \qquad \mathbf{A} = \begin{bmatrix} 0 & \frac{1+\epsilon}{\sigma+\epsilon} & \frac{1+\epsilon}{\sigma+\epsilon} \\ \frac{\alpha\epsilon}{1+\alpha\epsilon} & 0 & 0 \\ 0 & \frac{\sigma-1}{\sigma+\epsilon} & \frac{1+\epsilon}{\sigma+\epsilon} \end{bmatrix}$$

where $(\mathbf{A})_{hh'}$ denotes the $(h,h')^{\text{th}}$ entry of the matrix $\mathbf{A}$, which is the absolute value of the $(h,h')^{\text{th}}$ entry of the matrix $\mathbf{B}\mathbf{\Gamma}^{-1}$. We assume $\sigma > 1$.

We want to show that the spectral radius, the absolute value of the maximum eigenvalue, of $\mathbf{A}$ (denoted $\rho(\mathbf{A})$) equals one. The reasoning in Remark 6 in Appendix B.1.5 of Allen et al. (2023) demonstrates that $\rho(\mathbf{A}) \geq 1$. Thus, it remains to show that $\rho(\mathbf{A}) \leq 1$. To do so, we will use Lemma 1 in Appendix B.1.6 from Allen et al. (2023). Consider the matrix

$$\lambda I - \mathbf{A} = \begin{bmatrix} \lambda & -\frac{1+\epsilon}{\sigma+\epsilon} & -\frac{1+\epsilon}{\sigma+\epsilon} \\ -\frac{\alpha\epsilon}{1+\alpha\epsilon} & \lambda & 0 \\ 0 & -\frac{\sigma-1}{\sigma+\epsilon} & \lambda - \frac{1+\epsilon}{\sigma+\epsilon} \end{bmatrix}.$$

Define the function $f(\lambda) \equiv \det(\lambda I - \mathbf{A})$, and denote its kth derivative by $f_k(\lambda)$:

$$f(\lambda) = \lambda^3 - \lambda^2 \left(\frac{1+\epsilon}{\sigma+\epsilon}\right) - \left(\frac{1+\epsilon}{\sigma+\epsilon}\right)\left(\frac{\alpha\epsilon}{1+\alpha\epsilon}\right)\left(\lambda - \frac{1+\epsilon}{\sigma+\epsilon}\right) - \left(\frac{1+\epsilon}{\sigma+\epsilon}\right)\left(\frac{\alpha\epsilon}{1+\alpha\epsilon}\right)\left(\frac{\sigma-1}{\sigma+\epsilon}\right)$$
$$f_1(\lambda) = 3\lambda^2 - 2\lambda \left(\frac{1+\epsilon}{\sigma+\epsilon}\right) - \left(\frac{1+\epsilon}{\sigma+\epsilon}\right)\left(\frac{\alpha\epsilon}{1+\alpha\epsilon}\right)$$
$$f_2(\lambda) = 6\lambda - \frac{2(1+\epsilon)}{\sigma+\epsilon}.$$

Lemma 1 of Allen et al. (2023) states that $\rho(\mathbf{A}) \leq 1$ if and only if $f_k(1) \geq 0$ for $k = 0, 1, 2$.

$$f(1) = 1 - \frac{1+\epsilon}{\sigma+\epsilon} - \left(\frac{1+\epsilon}{\sigma+\epsilon}\right)\left(\frac{\alpha\epsilon}{1+\alpha\epsilon}\right) + \left(\frac{1+\epsilon}{\sigma+\epsilon}\right)\left(\frac{\alpha\epsilon}{1+\alpha\epsilon}\right)\left(\frac{2+\epsilon-\sigma}{\sigma+\epsilon}\right)$$
$$f_1(1) = 3 - 2\left(\frac{1+\epsilon}{\sigma+\epsilon}\right) - \left(\frac{1+\epsilon}{\sigma+\epsilon}\right)\left(\frac{\alpha\epsilon}{1+\alpha\epsilon}\right)$$
$$f_2(1) = 6 - \frac{2(1+\epsilon)}{\sigma+\epsilon}.$$

Both $f_1(1) \geq 0$ and $f_2(1) \geq 0$ for any $\sigma > 1$ and $\alpha, \epsilon > 0$. Inequality $f(1) \geq 0$ can be restated as

$$\left(\frac{1+\epsilon}{\sigma+\epsilon}\right)\left(\frac{\alpha\epsilon}{1+\alpha\epsilon}\right) \leq \frac{1}{2}. \tag{C.4}$$

Note that condition (C.4) is satisfied by our baseline parameter values ($\alpha = 0.24$, $\sigma = 4$, $\epsilon = 7.986$). Under condition (C.4), $\rho(\mathbf{A}) \leq 1$, which means that $\rho(\mathbf{A}) = 1$. Thus, by Theorem 1 Part ii.b of Allen et al. (2023), if condition (C.4) is true, the equilibrium exists and is unique up to scale.

C.2. *Exact hat algebra*

This section shows how to express a counterfactual equilibrium of the model in terms of counterfactual endogenous outcomes relative to baseline endogenous outcomes, counterfactual exogenous parameters relative to baseline exogenous parameters, constant elasticities, and baseline equilibrium shares. This method has been dubbed "exact hat algebra" in the international trade literature.

One can solve for the counterfactual equilibrium variables associated with combinations of counterfactual-to-baseline ratios of productivities $\hat{A}_n$, land endowments $\hat{T}_k$, and commuting costs $\hat{\lambda}_{kn}$ and $\hat{\bar{\delta}}_{kn}$.

The derivation of $\hat{w}_n$ starts from equation (3):

$$A_n \sum_k \frac{\ell_{kn}}{\bar{\delta}_{kn}} = \frac{(w_n/A_n)^{-\sigma}}{P^{1-\sigma}} Y \qquad \forall n.$$

Rearranging terms to isolate w_n and then taking ratios yields

$$\hat{w}_n = \hat{A}_n^{\frac{\sigma-1}{\sigma}} \underbrace{\left(\frac{\sum_k \frac{\ell'_{kn}}{\bar{\delta}'_{kn}}}{\sum_k \frac{\ell_{kn}}{\bar{\delta}_{kn}}} \right)^{\frac{-1}{\sigma}}}_{\equiv \hat{L}_n^{\frac{-1}{\sigma}}} \underbrace{\left(\frac{P'}{P} \right)^{\frac{\sigma-1}{\sigma}}}_{\equiv \hat{P}^{\frac{\sigma-1}{\sigma}}} \underbrace{\left(\frac{\sum_{k',n'} y'_{k'n'}}{\sum_{k',n'} y_{k'n'}} \right)^{\frac{1}{\sigma}}}_{\equiv \hat{Y}^{\frac{1}{\sigma}}}. \tag{C.5}$$

Next, we rewrite this in terms of $\frac{\hat{\ell}_{kn}}{\hat{\bar{\delta}}_{kn}}$ and baseline shares:

$$\hat{w}_n = \hat{A}_n^{\frac{\sigma-1}{\sigma}} \left(\frac{\sum_k \frac{\hat{\ell}_{kn}}{\hat{\bar{\delta}}_{kn}} \frac{\ell_{kn}}{\bar{\delta}_{kn}}}{\sum_k \frac{\ell_{kn}}{\bar{\delta}_{kn}}} \right)^{\frac{-1}{\sigma}} \hat{P}^{\frac{\sigma-1}{\sigma}} \hat{Y}^{\frac{1}{\sigma}}$$

$$= \hat{A}_n^{\frac{\sigma-1}{\sigma}} \left(\sum_k \frac{\hat{\ell}_{kn}}{\hat{\bar{\delta}}_{kn}} \frac{\frac{\ell_{kn}}{\bar{\delta}_{kn}}}{\sum_k \frac{\ell_{kn}}{\bar{\delta}_{kn}}} \right)^{\frac{-1}{\sigma}} \hat{P}^{\frac{\sigma-1}{\sigma}} \hat{Y}^{\frac{1}{\sigma}}$$

Multiply both sides by $\hat{w}_n^{-\frac{1}{\sigma}}$, simplify, and take both sides to the power $\frac{\sigma}{\sigma-1}$ to obtain

$$\hat{w}_n^{\frac{\sigma-1}{\sigma}} = \hat{A}_n^{\frac{\sigma-1}{\sigma}} \left(\sum_k \hat{y}_{kn} \frac{y_{kn}}{\sum_{k'} y_{k'n}} \right)^{-\frac{1}{\sigma}} \hat{P}^{\frac{\sigma-1}{\sigma}} \hat{Y}^{\frac{1}{\sigma}}$$

$$\hat{w}_n = \hat{A}_n \left(\sum_k \hat{y}_{kn} \frac{y_{kn}}{\sum_{k'} y_{k'n}} \right)^{\frac{1}{1-\sigma}} \hat{P} \hat{Y}^{\frac{1}{\sigma-1}},$$

which is equation (5).

To derive $\hat{P}$, rearrange terms and employ goods market clearing:

$$\hat{P}^{1-\sigma} = \sum_n \frac{\left(\frac{w'_n}{A'_n} \right)^{1-\sigma}}{\sum_{n'} \left(\frac{w_{n'}}{A_{n'}} \right)^{1-\sigma}} = \sum_n \left(\frac{\hat{w}_n}{\hat{A}_n} \right)^{1-\sigma} \frac{\left(\frac{w_n}{A_n} \right)^{1-\sigma}}{\sum_{n'} \left(\frac{w_{n'}}{A_{n'}} \right)^{1-\sigma}} = \sum_n \left(\frac{\hat{w}_n}{\hat{A}_n} \right)^{1-\sigma} \sum_k \frac{y_{kn}}{Y}.$$

This implies

$$\hat{P} = \left(\sum_n \left(\frac{\hat{w}_n}{\hat{A}_n} \right)^{1-\sigma} \sum_k \frac{y_{kn}}{Y} \right)^{\frac{1}{1-\sigma}}.$$

This shows that relative goods price index $\hat{P}$ can be written in terms of relative endogenous wages, relative exogenous productivities, and baseline (earnings) shares.

The ratio of counterfactual nominal output Y' to baseline nominal output Y is

$$\hat{Y} = \frac{Y'}{Y} = \frac{\sum_{k,n} y'_{kn}}{\sum_{k,n} y_{kn}} = \frac{\sum_{k,n} \hat{y}_{kn} y_{kn}}{\sum_{k,n} y_{kn}} = \sum_{k,n} \hat{y}_{kn} \frac{y_{kn}}{Y}$$

With the changes in output, changes in prices, and changes in wages, equation (5) is expressed entirely in terms of hats and shares.

Next, we derive $\hat{r}_k$. Dividing the counterfactual version of equation (4) by the baseline version yields

$$\hat{T}_k = \hat{r}_k^{-1} \frac{\sum_n y'_{kn}}{\sum_n y_{kn}}.$$

Rearranging this expression yields

$$\hat{r}_k = \hat{T}_k^{-1} \frac{\sum_n y'_{kn}}{\sum_n y_{kn}} = \hat{T}_k^{-1} \frac{\sum_n \hat{y}_{kn} y_{kn}}{\sum_n y_{kn}} = \hat{T}_k^{-1} \sum_n \hat{y}_{kn} \frac{y_{kn}}{\sum_{n'} y_{kn'}},$$

which is equation (6).

Finally, taking the ratio of counterfactual ℓ'_{kn} to baseline ℓ_{kn} using equation (2) yields, after considerable manipulation, equation (7):

$$\hat{\ell}_{kn} = \frac{\frac{w_n^{'\epsilon} (r_k^{'\alpha} \bar{\delta}'_{kn} \lambda'_{kn})^{-\epsilon}}{\sum_{k',n'} w_{n'}^{'\epsilon} (r_{k'}^{'\alpha} \bar{\delta}'_{k'n'} \lambda'_{k'n'})^{-\epsilon}}}{\frac{w_n^{\epsilon} (r_k^{\alpha} \bar{\delta}_{kn} \lambda_{kn})^{-\epsilon}}{\sum_{k',n'} w_{n'}^{\epsilon} (r_{k'}^{\alpha} \bar{\delta}_{k'n'} \lambda_{k'n'})^{-\epsilon}}}$$

$$= \frac{\hat{w}_n^{\epsilon}(\hat{r}_k^{\alpha}\hat{\bar{\delta}}_{kn}\hat{\lambda}_{kn})^{-\epsilon}}{\sum_{k',n'} w_{n'}^{'\epsilon}(r_{k'}^{'\alpha}\bar{\delta}_{k'n'}^{'}\lambda_{k'n'}^{'})^{-\epsilon}} \Big/ \sum_{k',n'} w_{n'}^{\epsilon}(r_{k'}^{\alpha}\bar{\delta}_{k'n'}\lambda_{k'n'})^{-\epsilon}$$

$$= \frac{\hat{w}_n^{\epsilon}(\hat{r}_k^{\alpha}\hat{\bar{\delta}}_{kn}\hat{\lambda}_{kn})^{-\epsilon}}{\sum_{k',n'} \frac{w_{n'}^{'\epsilon}(r_{k'}^{'\alpha}\bar{\delta}_{k'n'}^{'}\lambda_{k'n'}^{'})^{-\epsilon}}{w_{n'}^{\epsilon}(r_{k'}^{\alpha}\bar{\delta}_{k'n'}\lambda_{k'n'})^{-\epsilon}} \frac{w_{n'}^{\epsilon}(r_{k'}^{\alpha}\bar{\delta}_{k'n'}\lambda_{k'n'})^{-\epsilon}}{\sum_{k',n'} w_{n'}^{\epsilon}(r_{k'}^{\alpha}\bar{\delta}_{k'n'}\lambda_{k'n'})^{-\epsilon}}}$$

$$= \frac{\hat{w}_n^{\epsilon}(\hat{r}_k^{\alpha}\hat{\bar{\delta}}_{kn}\hat{\lambda}_{kn})^{-\epsilon}}{\sum_{k',n'} \hat{w}_{n'}^{\epsilon}(\hat{r}_{k'}^{\alpha}\hat{\bar{\delta}}_{k'n'}\hat{\lambda}_{k'n'})^{-\epsilon}\frac{\ell_{k'n'}}{L}}.$$

Together, equations (5), (6), and (7) deliver $\hat{w}_n$, $\hat{r}_k$, and $\hat{\ell}_{kn}$ given the elasticities σ, α, and ϵ, commuting costs $\bar{\delta}_{kn}$, baseline equilibrium values ℓ_{kn} and w_n, and counterfactual-to-baseline ratios $\hat{A}_n$, $\hat{T}_k$, $\hat{\lambda}_{kn}$, and $\hat{\bar{\delta}}_{kn}$.

C.2.1. *An example of baseline shares (and elasticities) as sufficient statistics*

Suppose that each residential location has an exogenous amenity level B_k such that i's indirect utility from residing in k and working in n is

$$U_{kn}^i = \epsilon \ln\left(\frac{B_k w_n}{r_k^{\alpha} P^{1-\alpha}\delta_{kn}}\right) + \nu_{kn}^i.$$

In the model presented in the main text, residential amenities do not vary across locations: $B_k = 1\ \forall k$. The probability of choosing residential-workplace pair kn is now

$$\Pr(U_{kn}^i > U_{k'n'}^i\ \forall (k',n') \neq (k,n)) = \frac{w_n^{\epsilon}\left(B_k^{-1} r_k^{\alpha}\delta_{kn}\right)^{-\epsilon}}{\sum_{k',n'} w_{n'}^{\epsilon}\left(B_{k'}^{-1} r_{k'}^{\alpha}\delta_{k'n'}\right)^{-\epsilon}}.$$

A residence is more attractive when it has higher amenities or greater land supply (lower rents).

Separating these amenities and land supplies is unnecessary for some counterfactual analysis. Infinitely many combinations of amenity and land-supply parameters could deliver the same baseline shares, but one does not need to identify all the model parameters in levels in order to compute counterfactual changes. Suppose that residential amenities vary across locations but are unchanged in the counterfactual scenario: $\hat{B}_k = 1\ \forall k$. Following the same steps as above, one can derive a system of equations characterizing the counterfactual equilibrium that is identical to equations (5)–(7). Given the baseline shares and elasticities (the sufficient

statistics), counterfactual changes do not depend on distinguishing between amenities B_k and land supply T_k.

C.3. *Calibrated-shares procedure*

The calibrated-shares procedure uses the observed ℓ_{kn} and $y_{kn} = w_n \ell_{kn} / \bar{\delta}_{kn}$ in equations (5)–(7). This implicitly calibrates combinations of model parameters. For example, this procedure implicitly rationalizes zero-commuters observations with infinite commuting costs, $\ell_{kn} = 0 \iff \delta_{kn} = \infty$. This procedure cannot characterize cases in which $\ell_{kn} = 0$ and $\ell'_{kn} \neq 0$ because the object $\hat{\delta}_{kn}^{-\epsilon} = \left(\frac{\delta_{kn}}{\delta'_{kn}}\right)^{\epsilon}$ is not sensibly defined if $\delta_{kn} = \infty$ and $\delta'_{kn} < \infty$.

This procedure does not identify the parameters A_n, T_k, and δ_{kn}. Given the elasticities σ, α, and ϵ and the baseline equilibrium shares ℓ_{kn} and y_{kn}, equations (2), (3), and (4) are insufficient to separately identify T_k and δ_{kn}. One would also need to observe land prices r_k.

When implementing this procedure in Sections 3.5 and 5.1, we compute tract-level workplace wages using LODES and ZIP Business Patterns data following Owens et al. (2020).

C.4. *Special case relating counterfactual errors to baseline-share errors*

In the continuum model with exogenous residential amenities (see Appendix C.2.1), no housing consumption ($\alpha = 0$), and perfectly elastic labor demand ($\sigma = \infty$), there is a linear mapping from baseline shares to predicted changes caused by a productivity change.

If individuals do not consume housing, $\alpha = 0$, the model simplifies. The endogenous variable r_k and the land-market-clearing condition are no longer part of the model. (For the counterfactual outcomes of interest, this is isomorphic to land being perfectly elastically supplied at a fixed rent r_k.) The price index is P in all residential locations. The labor-market-clearing condition is unchanged. The labor-allocation equation is now

$$\Pr(U_{kn}^i > U_{k'n'}^i \; \forall (k', n') \neq (k, n)) = \frac{(B_k w_n)^{\epsilon} \delta_{kn}^{-\epsilon}}{\sum_{k', n'} (B_{k'} w_{n'})^{\epsilon} \delta_{k'n'}^{-\epsilon}}.$$

Given this, the PPML estimator's residential and workplace fixed effects are proportionate to $\epsilon \ln B_k$ and $\epsilon \ln w_n$, respectively.

For this special case with perfectly elastic labor demand, the system of two equations defining counterfactual changes is

$$\hat{w}_n = \hat{A}_n \qquad \text{and} \qquad \hat{\ell}_{kn} = \frac{\hat{w}_n^{\epsilon} \hat{B}_k^{\epsilon} \left(\hat{\bar{\delta}}_{kn} \hat{\lambda}_{kn}\right)^{-\epsilon}}{\sum_{k', n'} \hat{w}_{n'}^{\epsilon} \hat{B}_{k'}^{\epsilon} \left(\hat{\bar{\delta}}_{k'n'} \hat{\lambda}_{k'n'}\right)^{-\epsilon} \frac{\ell_{k'n'}}{L}} \qquad \text{if } \ell_{kn} > 0.$$

If workplace n^* alone has a productivity increase ($\hat{A}_{n^*} > 1, \hat{A}_n = 1 \; \forall n \neq n^*$) and there are no changes to amenities nor commuting costs ($\hat{B}_k = 1 \; \forall k$, $\hat{\bar{\delta}}_{kn} = \hat{\lambda}_{kn} = 1 \; \forall kn$), then $\hat{\ell}_{kn^*}$ is constant across k and changes in commuters to workplace n^* vary only with baseline shares:

$$\hat{\ell}_{kn^*} = \frac{\hat{A}_{n^*}^{\epsilon}}{\hat{A}_{n^*}^{\epsilon} \frac{\sum_{k'} \ell_{k'n^*}}{L} + \frac{L - \sum_{k'} \ell_{k'n^*}}{L}} = \hat{L}_{n^*} \; \forall k$$

$$\Delta \ell_{kn^*} = \left(\hat{\ell}_{kn^*} - 1\right) \ell_{kn^*} = \left(\hat{L}_{n^*} - 1\right) \ell_{kn^*}.$$

Because the counterfactual change $\Delta \ell_{kn^*}$ is proportional to the baseline employment ℓ_{kn^*} in this simple case, if the baseline share ℓ_{kn^*}/L contains estimation error, a counterfactual change computed using that share will inherit that error proportionally.

The change in the number of workers employed at workplace n^* in this simple example is

$$\sum_k \Delta \ell_{kn^*} = \left(\hat{L}_{n^*} - 1\right) \sum_k \ell_{kn^*}.$$

If the estimation error in ℓ_{kn^*} is sampling error, workplace-level employment will be less noisy than a single residence-workplace pair: the coefficient of variation for a binomial distribution with I draws and probability p is $\sqrt{\frac{1-p}{pI}}$, which is decreasing in p, and the probability of choosing workplace n is greater than the probability of choosing residence-workplace pair kn ($p_n = \sum_k p_{kn} \geq p_{kn}$).

By contrast, changes in residents caused by a productivity shock are sensitive to pair-specific linkages. The change in the number of residents at residence k^* in this simple example is

$$\sum_n \Delta \ell_{k^*n} = \frac{\hat{L}_{n^*} - 1}{L - \sum_k \ell_{kn^*}} \sum_n \ell_{k^*n} \left[\frac{\ell_{k^*n^*}}{\sum_n \ell_{k^*n}} L - \sum_k \ell_{kn^*} \right].$$

This expression says that the change in residents is governed by the difference (in brackets) between a gross increase in residents, which is proportional to the share of residents who work at the treated workplace n^* in the baseline equilibrium, and a gross decrease in residents that is common across residential locations.

C.5. *Slope coefficients from regressing true probabilities on estimated shares*

This section discusses the slope coefficients for the calibrated-shares procedure and the covariates-based approach. In the special case ($\alpha = 0$, $\sigma = \infty$, and exogenous residential amenities), regressing the true changes in outcomes on the predicted changes is equivalent to regressing the true baseline shares on the estimated baseline shares (see Appendix C.4).

Thus, the object of interest is the slope coefficient β from regressing the true probability p_{kn} on the estimated probability $\hat{p}_{kn}$ in a given finite realization, where the true probability from the continuum model $p_{kn} \equiv \Pr(U^i_{kn} > U^i_{k'n'}\ \forall (k',n') \neq (k,n))$ is the right side of equation (2). Define the error $e_{kn} \equiv p_{kn} - \hat{p}_{kn}$. We denote the variance in errors across elements of the choice set (across residence-workplace pairs indexed by kn) within a finite realization by

$$\widehat{\text{Var}}(e_{kn}) \equiv \frac{1}{KN}\sum_{k,n} e_{kn}^2.$$

This definition exploits the fact that $\sum_{k,n} e_{kn} = 0$ in any finite realization. Thus, $\widehat{\text{Var}}(e_{kn})$ is the mean squared error (MSE). Similarly, denote the covariance of p_{kn} and e_{kn} across elements of the choice set within a finite realization by $\widehat{\text{Cov}}(p_{kn}, e_{kn}) \equiv \frac{1}{KN}\sum_{k,n} p_{kn}e_{kn}$ and the covariance of p_{kn} and $\hat{p}_{kn}$ by $\widehat{\text{Cov}}(p_{kn}, \hat{p}_{kn}) \equiv \frac{1}{KN}\sum_{k,n} p_{kn}\hat{p}_{kn} - \frac{1}{K^2N^2}\sum_{k,n} p_{kn}\sum_{k,n}\hat{p}_{kn}$. Finally, the variance of p_{kn} across kn pairs within a realization is $\widehat{\text{Var}}(p_{kn}) \equiv \frac{1}{KN}\sum_{k,n}\left(p_{kn} - \frac{1}{KN}\right)^2$, where we exploit the fact that the average p_{kn} must be $\frac{1}{KN}$ because $\sum_{k,n} p_{kn} = 1$. Note that p_{kn} is a fixed feature of the data-generating process that does not vary across realizations, so $\widehat{\text{Var}}(p_{kn}) = \text{Var}(p_{kn})$.

The slope coefficient of interest is

$$\hat{\beta} = \frac{\widehat{\text{Cov}}(p_{kn}, \hat{p}_{kn})}{\widehat{\text{Var}}(\hat{p}_{kn})} = \frac{\widehat{\text{Cov}}(p_{kn}, p_{kn} - e_{kn})}{\widehat{\text{Var}}(p_{kn} - e_{kn})} = \frac{\text{Var}(p_{kn}) - \widehat{\text{Cov}}(p_{kn}, e_{kn})}{\text{Var}(p_{kn}) + \widehat{\text{Var}}(e_{kn}) - 2\widehat{\text{Cov}}(p_{kn}, e_{kn})}.$$

First, consider the calibrated-shares procedure, which uses $\hat{p}_{kn} = s_{kn}$, the realized share in the observed sample. Because the calibrated-shares procedure is the frequency estimator that simply uses the observed shares as its estimate of the true probability, $\widehat{\text{Cov}}(p_{kn}, e_{kn}) \approx 0$ (this is true in expectation and typically close to zero in our simulations of finite realizations).[1] If the errors are uncorrelated with the true probabilities, $\widehat{\text{Cov}}(p_{kn}, e_{kn}) = 0$, then the slope coefficient is $\frac{\text{Var}(p_{kn})}{\text{Var}(p_{kn})+\text{MSE}}$. Given the data-generating process (given $\text{Var}(p_{kn})$) and orthogonal errors, the attenuation of the slope coefficient is increasing in the mean squared error. That is, here the calibrated-shares procedure's finite-sample errors manifest as both a lower slope coefficient and a higher mean squared error.

Second, consider the two-way fixed-effects estimator as an example of a covariates-based specification. Suppose that the data-generating process is $p_{kn} = p_k p_n \lambda_{kn}^{-\epsilon}$ with $\sum_n p_n \lambda_{kn}^{-\epsilon} = 1$ and $\sum_k p_k \lambda_{kn}^{-\epsilon} = 1$. The probabilities of choosing residence k and of choosing workplace n are p_k and p_n, respectively.[2] The two-way fixed-effects estimator is $\hat{p}_{kn} = s_k s_n$, where

[1] The population expectation of $\widehat{\text{Cov}}(p_{kn}, e_{kn}) \propto \sum_{kn} p_{kn}e_{kn} - \sum_{kn} p_{kn}\sum_{kn} e_{kn} = \sum_{kn} p_{kn}e_{kn}$ over realizations is zero for the calibrated-shares procedure, because $\mathbb{E}\left(\sum_{kn} p_{kn}e_{kn}\right) = \sum_{kn} p_{kn}\mathbb{E}\left(e_{kn}\right) = 0$ given that $\mathbb{E}(s_{kn}) = p_{kn}$.

[2] This requires $\sum_n p_n \lambda_{kn}^{-\epsilon} = 1\ \forall k$ so that $p_k = \sum_n p_{kn}\ \forall k$ and similarly $\sum_k p_k \lambda_{kn}^{-\epsilon} = 1\ \forall n$ so that $p_n = \sum_k p_{kn}\ \forall n$. These normalize λ_{kn} without loss of generality: common k- and n-level components ("fixed effects") are in p_k and p_n, not $\lambda_{kn}^{-\epsilon}$.

$s_k = \sum_n s_{kn}$ and $s_n = \sum_k s_{kn}$. The two-way fixed-effects estimator misses the contribution of $\lambda_{kn}^{-\epsilon}$ to the true probability p_{kn}, so typically $\widehat{\text{Cov}}(p_{kn}, e_{kn}) > 0$. This need not attenuate the slope coefficient, however. Intuitively, the estimator's omission of the (orthogonal) unobserved commuting cost reduces the R^2 of the regression, not the slope coefficient. Note that the slope coefficient can be written as

$$\hat{\beta} = \frac{\widehat{\text{Var}}(p_{kn}) - \widehat{\text{Cov}}(p_{kn}, e_{kn})}{\widehat{\text{Var}}(p_{kn}) - \widehat{\text{Cov}}(p_{kn}, e_{kn}) - \left(\widehat{\text{Cov}}(p_{kn}, e_{kn}) - \widehat{\text{Var}}(e_{kn})\right)}.$$

If the numerator $\text{Var}(p_{kn}) - \widehat{\text{Cov}}(p_{kn}, e_{kn})$ is much larger than $\widehat{\text{Cov}}(p_{kn}, e_{kn}) - \widehat{\text{Var}}(e_{kn})$, then the slope coefficient is close to one. Consider the case in which the fixed effects are reliably estimated (assume $s_k \approx p_k$, $s_n \approx p_n$, and $s_k s_n \approx p_k p_n$). (If s_k and s_n are not reliable estimators of p_k and p_n, then the frequency estimator s_{kn} will estimate p_{kn} very poorly.) In this case, the two differences of interest are approximately

$$\text{Var}(p_{kn}) - \widehat{\text{Cov}}(p_{kn}, e_{kn}) \approx \frac{1}{KN} \sum_{kn} p_k^2 p_n^2 \lambda_{kn}^{-\epsilon} - \frac{1}{K^2 N^2}$$

$$\widehat{\text{Cov}}(p_{kn}, e_{kn}) - \widehat{\text{Var}}(e_{kn}) \approx \frac{1}{KN} \sum_{kn} [p_k p_n]^2 (\lambda_{kn}^{-\epsilon} - 1).$$

Therefore, the numerator is much larger ($\gg$) than the other difference if

$$\begin{aligned} \text{Var}(p_{kn}) - \widehat{\text{Cov}}(p_{kn}, e_{kn}) &\gg \left[\widehat{\text{Cov}}(p_{kn}, e_{kn}) - \widehat{\text{Var}}(e_{kn})\right] \\ &\iff 0 \ll K^2 N^2 \text{Var}(p_k)\text{Var}(p_n) + K^2 \text{Var}(p_k) + N^2 \text{Var}(p_n) \end{aligned}$$

This implies that the slope coefficient for the two-way fixed-effects estimator will be close to one so long as there is substantial variation in the residence and employment shares ($\text{Var}(p_k) \gg 0, \text{Var}(p_n) \gg 0$).

If one ran this regression for a subset of location pairs, such as a single workplace n^*, the same expression for the slope coefficient would apply. Note that in this case, the sum of the prediction errors for a single workplace n^*, $\sum_k e_{kn^*}$, need not equal zero, so that $\widehat{\text{Var}}(e_{kn})$ equals the MSE plus the mean prediction error squared. Similarly, the definitions of $\widehat{\text{Cov}}(p_{kn}, e_{kn})$ and $\widehat{\text{Var}}(p_{kn})$ must be extended to allow for means not equal to zero and $\frac{1}{KN}$, respectively. In the Monte Carlo simulations of Section 3.4 and event studies of Section 3.5, we match the change in total employment in the workplace of interest, so that by construction the sum of the prediction errors is indeed zero.

C.6. *Simple example of expected squared error for baseline shares*

This section contrasts the expected squared errors for baseline shares produced by the calibrated-shares procedure and the covariates-based approach. Researchers face a familiar trade-off: a more flexible model with more parameters requires more data to be estimated precisely. The calibrated-shares procedure estimates a saturated model with an arbitrary commuting cost for each residence-workplace pair, while the covariates-based specification assumes that bilateral commuting costs depend on few observables. Their relative performance depends on how many commuters are observed per residence-workplace pair and the magnitudes of the unobserved commuting costs.

Consider an arbitrary data-generating process with true probabilities $p_{kn} \equiv \Pr(U^i_{kn} > U^i_{k'n'}\ \forall (k',n') \neq (k,n))$. The expected squared error for a baseline share $\frac{\ell_{kn}}{L}$ produced by an estimator $\hat{p}_{kn}$ is $\mathbb{E}\left(p_{kn} - \hat{p}_{kn}\right)^2$. The calibrated-shares procedure uses the frequency estimator $\hat{p}_{kn} = s_{kn}$, where s_{kn} is the realized share in the observed sample. The covariates-based approach imposes more structure. Consider a simple example in which the data-generating process is $p_{kn} = p_k p_n \lambda_{kn}^{-\epsilon}$, so bilateral commuting costs are entirely unobserved. The probabilities of choosing residence k and of choosing workplace n are p_k and p_n, respectively. The covariates-based specification for this example is the two-way-fixed-effects estimator: $\hat{p}_{kn} = s_k s_n$, where s_k is the share of residence k and s_n is the share of workplace n.

The calibrated-shares procedure is an unbiased estimator ($\mathbb{E}(s_{kn}) = p_{kn}$), so its expected squared error for baseline share $\frac{\ell_{kn}}{L}$ is its finite-sample variance. With I observed commuters, this is $\mathbb{E}\left(p_{kn} - s_{kn}\right)^2 = \text{var}\left(s_{kn}\right) = \frac{1}{I} p_{kn}(1 - p_{kn})$. This variance goes to zero as the number of observations goes to infinity, $\lim_{I\to\infty} \text{var}\left(s_{kn}\right) = 0$, so the only shortcoming of the calibrated-shares procedure here is its finite-sample performance.

By imposing more structure, the covariates-based approach delivers an estimator that typically has much lower variance but is biased to the extent that it omits unobserved commuting costs. In the simple example, because the estimator $\hat{p}_{kn} = s_k s_n$ uses residence- and workplace-level shares rather than the share of a single residence-workplace pair s_{kn}, the covariates-based specification has a much smaller variance in granular settings with finite individuals.[3] Unless $\lambda_{kn}^{-\epsilon} = 1$, the two-way-fixed-effects estimator is biased because $\mathbb{E}(s_k s_n) \neq p_{kn}$. Its expected squared error is

$$\mathbb{E}\left[(p_{kn} - \hat{p}_{kn})^2\right] = \left[\mathbb{E}(p_{kn} - s_k s_n)\right]^2 + \text{var}(s_k s_n) = \left[\frac{I-1}{I} p_k p_n (1 - \lambda_{kn}^{-\epsilon})\right]^2 + \text{var}(s_k s_n),$$

[3] Given thousands of residences and workplaces, $\text{var}(s_k s_n)$ is two or three orders of magnitude smaller than $\text{var}(s_{kn})$. It can be shown that $\text{var}(s_k s_n) = \frac{1}{I} p_k p_n (1 - p_k)(1 - p_n)\left(\frac{1}{I} + \frac{p_n}{1-p_n} + \frac{p_k}{1-p_k}\right) + \frac{1}{I^3} p_k p_n \left(\lambda_{kn}^{-\epsilon} - 1\right)\left[1 + 2(I-1)(p_k + p_n) + \left((I-2)(2I + \lambda_{kn}^{-\epsilon} - 3) - 2I\right) p_k p_n\right]$. For example, given the NYC setting with $I \approx 2.5 \times 10^6$ and $p_k \approx p_n \approx 0.5 \times 10^{-3}$, $\text{var}(s_k s_n)$ is on the order of 10^{-16} and $\text{var}(s_{kn})$ is on the order of 10^{-13}.

where the first term is the square of the bias of the estimator and the second term is its sample variance.[4] As $I \to \infty$, the bias term goes to $\left[p_k p_n (1 - \lambda_{kn}^{-\epsilon})\right]^2$ and the variance term goes to zero.

Which estimator has a smaller expected squared error depends on the contrast between the variance of the calibrated-shares procedure (governed by sample size I) and the bias of the covariates-based specification (governed by unobserved commuting cost $(1 - \lambda_{kn}^{-\epsilon})^2$). The variance of the covariates-based specification is unimportant because $\text{var}(s_k s_n)$ is orders of magnitude smaller than $\text{var}(s_{kn})$ (see footnote 3). The relevant comparison is that the squared bias of the covariates-based specification is smaller than the variance of the frequency estimator if

$$
\begin{aligned}
&\left[\mathbb{E}(s_k s_n - p_{kn})\right]^2 \leq \text{var}(s_{kn}) \\
\iff &\left[\frac{I-1}{I} p_k p_n (1 - \lambda_{kn}^{-\epsilon})\right]^2 \leq \frac{1}{I} \lambda_{kn}^{-\epsilon} p_k p_n (1 - \lambda_{kn}^{-\epsilon} p_k p_n) \\
\iff &\frac{(I-1)^2}{I} p_k p_n \leq \frac{\lambda_{kn}^{-\epsilon}}{\left(1 - \lambda_{kn}^{-\epsilon}\right)^2} (1 - \lambda_{kn}^{-\epsilon} p_k p_n).
\end{aligned}
$$

As I becomes arbitrarily large, the left side becomes arbitrarily large and the inequality is necessarily false when $\lambda_{kn}^{-\epsilon} \neq 1$. For a given I, as $\lambda_{kn}^{-\epsilon}$ approaches 1, the right side becomes arbitrarily large and the inequality is necessarily true. Denoting $\mathfrak{d}_{kn} \equiv \frac{(I-1)^2}{I} p_k p_n$, the roots of the quadratic equation associated with the right side of the inequality are $\frac{1+2\mathfrak{d}_{kn}}{2p_k p_n + 2\mathfrak{d}_{kn}} \pm \frac{\sqrt{1+4\mathfrak{d}_{kn}(1-p_k p_n)}}{2p_k p_n + 2\mathfrak{d}_{kn}}$, and the inequality above is true when $\lambda_{kn}^{-\epsilon}$ is close enough to one to lie between these roots.[5] Recall that the p_k- and p_n-weighted mean values of $\lambda_{kn}^{-\epsilon}$ equal one, so its distribution must have substantial variance to make the inequality false for many kn pairs.[6]

These results describe the expected squared error for a given residence-workplace pair kn. The Monte Carlo simulations of Section 3.4 and event studies of Section 3.5 examine errors in predicted changes in the number of commuters to a specific workplace of interest. There, we report the mean squared error for each simulation (each event), which is the average of the squared errors over all residences paired with the workplace of interest.

[4] Note that $\hat{p}_{kn} = s_k s_n$ is an unbiased estimator when $I = 1$ because with a single observation $s_k s_n$ equals s_{kn}.

[5] When the right side equals $\mathfrak{d}_{kn}$, the associated quadratic equation is $0 = (\mathfrak{d}_{kn} + p_k p_n)\left(\lambda_{kn}^{-\epsilon}\right)^2 - (1 + 2\mathfrak{d}_{kn})\lambda_{kn}^{-\epsilon} + \mathfrak{d}_{kn}$.

[6] For example, consider the NYC setting in which $I \approx 2.5 \times 10^6$, the average residential tract has $p_k \approx 0.5 \times 10^{-3}$, and the workplace tract containing 200 Fifth Avenue has $p_n \approx 0.5 \times 10^{-2}$, so that $\mathfrak{d}_{kn}$ is about 6.25. When $\mathfrak{d}_{kn} = 6.25$, the inequality is true when $\epsilon \ln \lambda_{kn} \in [-0.397, 0.397]$. For a log-normal distribution with unit mean, $\ln x \overset{\text{iid}}{\sim} \mathcal{N}(\frac{-\sigma^2}{2}, \sigma^2)$, to have half of its draws lie outside that interval, $\ln x \notin [-0.397, 0.397]$, its standard deviation must exceed approximately 0.56. For comparison, in our setting with NYC tracts, the observable $\epsilon \ln \bar{\delta}_{kn}$ has a standard deviation of 1.255.

C.7. *Welfare changes in the continuum model*

Recall from equation (1) that the indirect utility of an individual i residing in k and working in n is

$$U_{kn}^i = \epsilon \log\left(\frac{w_n}{r_k^\alpha P^{1-\alpha}\delta_{kn}}\right) + \nu_{kn}^i.$$

Following equation (3.10) in Train (2009), one can show that the expected utility $\bar{U}$ of each (ex ante identical) worker, each of whom selects her utility-maximizing residence-workplace pair, is

$$\bar{U} = \mathbb{E}_{\nu^i}\left[U_{kn}^i | U_{kn}^i \geq U_{k'n'}^i,\ \forall k', n'\right] = \log\left(\sum_{k',n'} w_{n'}^\epsilon \left(r_{k'}^\alpha P^{1-\alpha}\delta_{k'n'}\right)^{-\epsilon}\right) + c \quad \forall k, n, \quad \text{(C.6)}$$

where c is a constant. Note that c is unknown since the level of utility cannot be measured. Denote $C_{kn} = w_n / \left(r_k^\alpha P^{1-\alpha}\bar{\delta}_{kn}\right)$ as the real consumption for individuals residing in k and working in n. We can rewrite equation (C.6) in terms of real consumption:

$$\bar{U} = \log\left(\sum_{k,n} C_{kn}^\epsilon \lambda_{kn}^{-\epsilon}\right) + c,$$

where we use the decomposition of commuting costs δ_{kn} into time $\bar{\delta}_{kn}$ and disutility λ_{kn} with $\delta_{kn} = \bar{\delta}_{kn}\lambda_{kn}$.

Now, consider a counterfactual equilibrium. Recall that the counterfactual-to-baseline ratio of a variable $x > 0$ is denoted by $\hat{x} \equiv \frac{x'}{x}$. The expected utility in the counterfactual equilibrium is $\bar{U}' = \log\left(\sum_{k,n}(C'_{kn})^\epsilon(\lambda'_{kn})^{-\epsilon}\right) + c$. We define $1+\psi$ as the equivalent variation in consumption such that

$$\bar{U}' = \log\left(\sum_{k,n}(C'_{kn})^\epsilon(\lambda'_{kn})^{-\epsilon}\right) + c = \log\left(\sum_{k,n}\left[(1+\psi)C_{kn}\right]^\epsilon \lambda_{kn}^{-\epsilon}\right) + c = \epsilon\log(1+\psi) + \bar{U}.$$

In other words, the equivalent variation in consumption is defined such that multiplying the real consumption of individuals in all residence-workplace pairs by $1+\psi$ in the baseline equilibrium (holding prices and other allocations fixed) would generate the same welfare outcome as the counterfactual equilibrium with new prices and allocations. Re-arranging the equation above yields an expression for $1+\psi$ in terms of welfare difference $\bar{U}' - \bar{U}$:

$$1 + \psi = \exp\left(\frac{1}{\epsilon}\left[\bar{U}' - \bar{U}\right]\right),$$

which is our measure of equivalent variation.

In practice, $1+\psi$ can be calculated from changes in wages ($\hat{w}_n$), rents ($\hat{r}_k$), and price index ($\hat{P}$), given exogenous changes in productivities ($\hat{A}_n$), land endowments ($\hat{T}_k$), and commuting costs ($\hat{\delta}_{kn} = \hat{\bar{\delta}}_{kn}\hat{\lambda}_{kn}$), as well as initial wages (w_n) and commuting flows (ℓ_{kn}). One can show that

$$1+\psi = \left(\sum_{k,n} \frac{\ell_{kn}}{L} \hat{w}_n^{\epsilon} \left(\hat{r}_k^{\alpha} \hat{P}^{1-\alpha} \hat{\delta}_{kn} \right)^{-\epsilon} \right)^{\frac{1}{\epsilon}}, \tag{C.7}$$

where $\hat{w}_n$, $\hat{r}_k$, and $\hat{P}$ can be solved from equations (5), (6), and (7) (with detailed derivations described in Appendix C.2).

C.8. *The trade equilibrium is unique*

Recall that the trade equilibrium takes the labor allocation as given. Consequently, we just need to solve for wage and land rent vectors that satisfy equations (3) and (4) given arbitrary labor allocation ℓ_{kn}. Given a labor allocation and vector of wages, equation (4) determines a unique vector of land rents, so we can focus on solving for a unique vector of relative wages. Recall $L_n \equiv \sum_k \frac{\ell_{kn}}{\bar{\delta}_{kn}}$ is total labor supply in location n. Substituting L_n and $P = \left(\sum_n (w_n/A_n)^{1-\sigma}\right)^{1/(1-\sigma)}$ into equation (3) yields

$$A_n L_n = \frac{(w_n/A_n)^{-\sigma}}{\sum_{n'} (w_{n'}/A_{n'})^{1-\sigma}} Y \quad \forall n.$$

Note that this equation is homogeneous of degree zero in the wage vector. Selecting $w_1 = 1$ as the numeraire and writing the expression in terms of relative wages yields

$$w_n = \left(\frac{A_n}{A_1} \right)^{\frac{\sigma-1}{\sigma}} \left(\frac{L_n}{L_1} \right)^{-1/\sigma}.$$

This defines a unique vector of relative wages. Plugging those wages into equation (4) yields a unique vector of land rents.

C.9. *Continuum model as limiting case of the model with a finite number of individuals*

We derive the result that the equilibrium of the model with a finite number of individuals coincides with the equilibrium of the continuum model as the number of individuals becomes infinite, $I \to \infty$. Note that aggregate labor supply L is fixed, as each individual supplies L/I units of labor. The key step is to show that equation (2) holds as $I \to \infty$. Conditional on the labor allocation $\{\ell_{kn}\}$, the number of individuals plays no role in the trade equilibrium, which will coincide with that of a continuum model.

Definition 4.2 says that the share of labor supplied by residents of k who work in n is

$$\frac{\ell_{kn}}{L} = \frac{1}{I}\sum_{i=1}^{I} \mathbf{1}\left\{\tilde{U}^i_{kn}(\boldsymbol{\nu}^I) > \tilde{U}^i_{k'n'}(\boldsymbol{\nu}^I), \forall (k',n') \neq (k,n)\right\}.$$

Since the idiosyncratic preferences are drawn from an independent and identically distributed random variable, the law of large numbers implies that $\frac{1}{I}\sum_{i=1}^{I} \mathbf{1}\left\{\tilde{U}^i_{kn}(\boldsymbol{\nu}^I) > \tilde{U}^i_{k'n'}(\boldsymbol{\nu}^I), \forall (k',n') \neq (k,$ converges to the mean of a binary random variable that is equal to one with probability $\Pr(\tilde{U}^i_{kn} > \tilde{U}^i_{k'n'}, \forall (k',n') \neq (k,n))$. Thus,

$$\mathbb{E}\left(\mathbf{1}\left\{\tilde{U}^i_{kn} > \tilde{U}^i_{k'n'}, \forall (k',n') \neq (k,n)\right\}\right) = \Pr\left(\tilde{U}^i_{kn} > \tilde{U}^i_{k'n'}, \forall (k',n') \neq (k,n)\right).$$

Using the probability in equation (10),

$$\begin{aligned}\lim_{I\to\infty} \frac{1}{I}\sum_{i=1}^{I} \mathbf{1}\left\{\tilde{U}^i_{kn}(\boldsymbol{\nu}^I) > \tilde{U}^i_{k'n'}(\boldsymbol{\nu}^I), \forall (k',n') \neq (k,n)\right\} &= \mathbb{E}\left(\mathbf{1}\left\{\tilde{U}^i_{kn} > \tilde{U}^i_{k'n'}, \forall (k',n') \neq (k,n)\right\}\right) \\ &= \frac{\tilde{w}^{\epsilon}_n (\tilde{r}^{\alpha}_k \delta_{kn})^{-\epsilon}}{\sum_{k',n'} \tilde{w}^{\epsilon}_{n'} (\tilde{r}^{\alpha}_{k'} \delta_{k'n'})^{-\epsilon}}.\end{aligned}$$

As a result, as $I \to \infty$, $\frac{\ell_{kn}}{L} \to \frac{\tilde{w}^{\epsilon}_n (\tilde{r}^{\alpha}_k \delta_{kn})^{-\epsilon}}{\sum_{k',n'} \tilde{w}^{\epsilon}_{n'} (\tilde{r}^{\alpha}_{k'} \delta_{k'n'})^{-\epsilon}}$. Since the continuum-case rational expectations are the wages and land rents that clear markets for this labor allocation, the market-clearing prices of the model with a finite number of individuals converge to those values.

C.10. *Uncertainty about counterfactual changes due to idiosyncrasies: Analytical example*

To analytically characterize the uncertainty in counterfactual changes caused by individuals' idiosyncratic preferences, consider an economy with one residential location and two workplaces. We compute the change in quantities associated with a change in economic primitives from Υ to Υ' for a given realization of idiosyncratic preferences $\boldsymbol{\nu}^I$. An individual's utility from choosing workplace n is $\tilde{U}^i_{kn} = \tilde{U}_{kn} + \nu^i_{kn}$ in the baseline equilibrium and $\tilde{U}'^i_{kn} = \tilde{U}'_{kn} + \nu^i_{kn}$ in the counterfactual equilibrium. Because there is only one residential location, we suppress the k subscript for convenience (e.g., denote $\tilde{U}_{k1}$ by $\tilde{U}_1$).

Suppose the counterfactual change improves the attractiveness of workplace $n = 1$ relative to $n = 2$: $\tilde{U}'_1 - \tilde{U}'_2 > \tilde{U}_1 - \tilde{U}_2$. Individuals with $\nu^i_1 - \nu^i_2 < \tilde{U}'_2 - \tilde{U}'_1$ choose workplace 2 in both the baseline and counterfactual equilibria. Individuals with $\nu^i_1 - \nu^i_2 > \tilde{U}_2 - \tilde{U}_1$ choose workplace 1 in both the baseline and counterfactual equilibria. Those individuals who drew idiosyncratic shocks such that $\nu^i_1 - \nu^i_2 \in \left[\tilde{U}'_2 - \tilde{U}'_1, \tilde{U}_2 - \tilde{U}_1\right]$ will change their workplace from $n = 2$ in the baseline to $n = 1$ in the counterfactual. This interval is smaller if the counterfactual change is smaller.

The unconditional probability of leaving workplace 2 is $p \equiv \Pr\left(\tilde{U}_2' - \tilde{U}_1' < \nu_1^i - \nu_2^i < \tilde{U}_2 - \tilde{U}_1\right) = F\left(\tilde{U}_2 - \tilde{U}_1\right) - F\left(\tilde{U}_2' - \tilde{U}_1'\right)$ where $F()$ denotes the cumulative distribution function of the logistic distribution. This probability equals the share of switchers in the continuum model. In the model with a finite number of individuals, the unconditional number of switchers follows a binomial distribution with I independent outcomes and probability p of switching. The expected share of the population that switches is p. The variance of the share that switches is $\frac{p(1-p)}{I}$. With more independent draws, there is less uncertainty about the share who switch.

C.11. *Model extensions*

C.11.1. *Trade costs*

Relative to the model in Section 4, we now assume that goods trade is subject to iceberg trade costs: delivering a unit of the location-n variety to location k requires producing $\tau_{nk} \geq 1$ units in n. Individuals consume differentiated goods at their residences, and the price of location n's output for consumers in location k is $\tau_{nk} w_n / A_n$. Thus, the price index in location k is $r_k^\alpha P_k^{1-\alpha}$, where the local CES price index for goods is $P_k = \left[\sum_n \left(\tau_{nk} w_n / A_n\right)^{1-\sigma}\right]^{1/(1-\sigma)}$.

In this environment, the equivalent of equation (9) is such that, based on the beliefs $\{\tilde{w}_n\}$ and $\{\tilde{r}_k\}$, each worker chooses the residential location and the work location that maximize expected utility,

$$\tilde{U}_{kn}^i = \underbrace{\epsilon \ln\left(\frac{\tilde{w}_n}{\tilde{r}_k^\alpha \tilde{P}_k^{1-\alpha} \delta_{kn}}\right)}_{\equiv \tilde{U}_{kn}} + \nu_{kn}^i, \tag{C.8}$$

where $\tilde{P}_k = \left[\sum_n \left(\tau_{nk} \tilde{w}_n / A_n\right)^{1-\sigma}\right]^{1/(1-\sigma)}$.

With trade costs, the goods-market-clearing condition, which is the equivalent of equation (3), equates quantity supplied and quantity demanded:

$$A_n \sum_k \frac{\ell_{kn}}{\delta_{kn}} = \left(w_n / A_n\right)^{-\sigma} \sum_k \left[\left(\frac{\tau_{nk}}{P_k}\right)^{1-\sigma} \left(\sum_{n'} y_{kn'}\right)\right]. \tag{C.9}$$

Equation (4), which clears the land market, is unchanged by the introduction of trade costs.

The definition of a trade equilibrium is akin to Definition 4.1, with equations (C.9) and (4) serving as the relevant market-clearing conditions. The definition of a commuting equilibrium with finitely many individuals is akin to Definition 4.2 with $\{\tau_{kn}\}$ added to the list of economic primitives and $\tilde{U}_{kn}^i$ being defined in equation (C.8).

C.11.2. *Production uses land*

Relative to the model in Section 4, we now assume that production of goods uses a Cobb-Douglas combination of labor and land inputs. In particular, the quantity of output in workplace

location n is

$$q_n = A_n \left(\frac{L_n}{1-\beta} \right)^{1-\beta} \left(\frac{T_n^P}{\beta} \right)^{\beta},$$

where L_n is labor and T_n^P is land used in production. Therefore, the unit cost of production in location n is

$$c_n = \frac{1}{A_n} w_n^{1-\beta} r_n^{\beta}$$

and the CES price index is $P = (\sum_n c_n^{1-\sigma})^{1/(1-\sigma)}$.

Labor supplied in location n is given by $L_n = \sum_k \frac{\ell_{kn}}{\bar{\delta}_{kn}}$. From the first-order condition, land demanded for use in production in location n can be expressed as

$$T_n^P = \frac{\beta}{1-\beta} \left(\frac{w_n}{r_n} \right) \sum_k \frac{\ell_{kn}}{\bar{\delta}_{kn}}.$$

Substituting this expression into the production technology, output in location n is

$$q_n = \frac{1}{1-\beta} A_n \left(\frac{w_n}{r_n} \right)^{\beta} \sum_k \frac{\ell_{kn}}{\bar{\delta}_{kn}}.$$

As in Section 4, each individual devotes $1-\alpha$ of their expenditure to differentiated goods and α of their expenditure to land, while immobile landlords spend all of their income on differentiated goods, such that total expenditure on differentiated goods equals aggregate income. With land used in production, however, aggregate income now includes the income that accrues to each factor. Total income that accrues to labor is $\sum_n w_n L_n$. Total income that accrues to land used in production equals

$$\sum_{n'} r_{n'} T_{n'}^P = \frac{\beta}{1-\beta} \sum_{n'} w_{n'} \sum_{k'} \frac{\ell_{k'n'}}{\bar{\delta}_{k'n'}} = \frac{\beta}{1-\beta} \sum_{n'} w_{n'} L_{n'}.$$

The demand for each differentiated good stemming from CES preferences means that demand is proportional to aggregate income. Goods market clearing requires that quantity demanded equals quantity supplied in each location n:

$$\frac{1}{1-\beta} A_n \left(\frac{w_n}{r_n} \right)^{\beta} \sum_k \frac{\ell_{kn}}{\bar{\delta}_{kn}} = \frac{c_n^{-\sigma}}{P^{1-\sigma}} \left[\left(\frac{1}{1-\beta} \right) \sum_{n'} w_{n'} \sum_{k'} \frac{\ell_{k'n'}}{\bar{\delta}_{k'n'}} \right].$$

Writing the left-hand side in terms of the unit cost and simplifying yields

$$w_n \sum_k \frac{\ell_{kn}}{\bar{\delta}_{kn}} = \frac{c_n^{1-\sigma}}{P^{1-\sigma}} \left[\sum_{n'} w_{n'} \sum_{k'} \frac{\ell_{k'n'}}{\bar{\delta}_{k'n'}} \right]. \tag{C.10}$$

Residents spend a fraction α of their income on land, so their demand for land, denoted T_n^R, is given by

$$T_n^R = \alpha \sum_k \frac{\ell_{nn'}}{\bar{\delta}_{nn'}} \frac{w_{n'}}{r_n} = \frac{\alpha}{r_n} \sum_{n'} y_{nn'}.$$

Clearing the land market means equating the fixed land supply T_n to the sum of the quantities of land demanded by producers and residents:

$$T_n = T_n^P + T_n^R = \frac{1}{r_n} \left[\frac{\beta}{1-\beta} \sum_k \frac{\ell_{kn}}{\bar{\delta}_{kn}} w_n + \alpha \sum_{n'} \frac{\ell_{nn'}}{\bar{\delta}_{nn'}} w_{n'} \right]. \tag{C.11}$$

The definition of a trade equilibrium is akin to Definition 4.1, with equations (C.10) and (C.11) serving as the relevant market-clearing conditions. The definition of a commuting equilibrium with finitely many individuals is akin to Definition 4.2 with the Cobb-Douglas production parameter β added to the list of economic primitives.

C.11.3. *Residential amenities*

Relative to the model in Section 4, we now assume that each location is endowed with a residential amenity B_k. Individual i's utility from residing in k and working in n is now

$$\tilde{U}_{kn}^i = \epsilon \ln \left(\frac{B_k \tilde{w}_n}{\tilde{r}_k^\alpha P^{1-\alpha} \delta_{kn}} \right) + \nu_{kn}^i. \tag{C.12}$$

Note that local amenities can depend on local residential density as in Allen and Arkolakis (2014): $B_k = \bar{B}_k \left(\sum_n \ell_{kn} \right)^{-\theta}$, where $\bar{B}_k$ is the fundamental amenities in location k, $\sum_n \ell_{kn}$ is the total number of individuals residing in k, and $\theta \geq 0$ governs the strength of the congestion force. When $\theta = 0$, local amenities are exogenous.

The definition of a trade equilibrium in Definition 4.1 remains unchanged. The definition of a commuting equilibrium with finitely many individuals is akin to Definition 4.2 with the fundamental residential amenities $\{\bar{B}_k\}$ and θ added to the list of economic primitives and $\tilde{U}_{kn}^i$ being defined in equation (C.12).

C.11.4. *Local increasing returns*

Relative to the model in Section 4, we now assume that production exhibits local external economies of scale. In particular, in each location n the linear production technology is $q_n = A_n L_n$, where L_n is the labor supply of workers working in location n and $A_n \equiv \bar{A}_n L_n^\eta$. Thus, the CES price index is $P = \left[\sum_n \left(w_n / (\bar{A}_n L_n^\eta) \right)^{1-\sigma} \right]^{1/(1-\sigma)}$.

Since output in location n is $\bar{A}_n \left(\sum_k \ell_{kn}/\bar{\delta}_{kn}\right)^{1+\eta}$, equating quantity supplied and quantity demanded requires

$$\bar{A}_n \left(\sum_k \ell_{kn}/\bar{\delta}_{kn}\right)^{1+\eta} = \frac{{w_n}^{-\sigma} \bar{A}_n^{\sigma} \left(\sum_k \ell_{kn}/\bar{\delta}_{kn}\right)^{\eta\sigma}}{P^{1-\sigma}} Y. \tag{C.13}$$

The remaining results are unchanged after replacing equation (3) with equation (C.13). The definition of a trade equilibrium is akin to Definition 4.1, with equations (C.13) and (4) serving as the relevant market-clearing conditions. The definition of a commuting equilibrium with finitely many individuals is akin to Definition 4.2 with $\{\bar{A}_n\}$ and η added to the list of economic primitives.

Local increasing returns enter the sufficient conditions for the existence and uniqueness of the equilibrium in the continuum model and thereby the uniqueness of the continuum-case rational expectations. We assume that $\eta < \frac{1}{\sigma - 1}$. We define $\psi = 1 + \eta - \eta\sigma$. Following the same steps as in Appendix C.1, the relevant matrix whose spectral radius we need to evaluate is

$$\mathbf{A} = \begin{bmatrix} 0 & \frac{\psi(1+\epsilon)}{\sigma+\psi\epsilon} & \frac{1+\epsilon}{\sigma+\psi\epsilon} \\ \frac{\alpha\epsilon}{1+\alpha\epsilon} & 0 & 0 \\ 0 & \frac{|\sigma-\psi|}{\sigma+\psi\epsilon} & \frac{1+\epsilon}{\sigma+\psi\epsilon} \end{bmatrix}.$$

As in Appendix C.1, $\rho(\mathbf{A}) \geq 1$ and $\rho(\mathbf{A}) \leq 1$ if and only if $f_k(1) \geq 0$ for $k = 0, 1, 2$.

$$f(1) = 1 - \frac{1+\epsilon}{\sigma+\psi\epsilon} - \psi\left(\frac{1+\epsilon}{\sigma+\psi\epsilon}\right)\left(\frac{\alpha\epsilon}{1+\alpha\epsilon}\right)\left(1 - \frac{1+\epsilon}{\sigma+\psi\epsilon}\right) - \left(\frac{1+\epsilon}{\sigma+\psi\epsilon}\right)\left(\frac{\alpha\epsilon}{1+\alpha\epsilon}\right)\left(\frac{|\sigma-\psi|}{\sigma+\psi\epsilon}\right)$$

$$f_1(1) = 3 - 2\left(\frac{1+\epsilon}{\sigma+\psi\epsilon}\right) - \psi\left(\frac{1+\epsilon}{\sigma+\psi\epsilon}\right)\left(\frac{\alpha\epsilon}{1+\alpha\epsilon}\right)$$

$$f_2(1) = 6 - 2\left(\frac{1+\epsilon}{\sigma+\psi\epsilon}\right).$$

Notice that $f_2(1) > f_1(1)$ under the parametric assumptions. We proceed by numerically finding the values of η that satisfy $f(1) \geq 0$ and $f_1(1) \geq 0$ given values of α, σ, and ϵ. Under the baseline parameter values ($\alpha = 0.24, \sigma = 4, \epsilon = 7.986$), $f(1) \geq 0$ and $f_1(1) \geq 0$ are satisfied when $\eta \leq 0.0029$. The equilibrium exists and is unique up to scale when agglomeration forces are modest.

C.12. *Galton's fallacy for Poisson counts*

When a process is serially uncorrelated or mean reverting, changes will be negatively correlated with prior levels. Consider an AR(1) process such that

$$y_t = (1-\tilde{\rho})\tilde{\mu} + \tilde{\rho} y_{t-1} + \epsilon_t,$$

where $\tilde{\mu}$ is the unconditional mean of y_t, $\tilde{\rho} \in [0,1)$ is its serial correlation, and ϵ_t is iid white noise with mean zero and variance σ_ϵ^2. If $\tilde{\rho} = 0$, y_t is independently and identical distributed.

This process can be rewritten by iterating and then taking the difference:

$$\begin{aligned} y_t &= (1-\tilde{\rho})\tilde{\mu} + \tilde{\rho} y_{t-1} + \epsilon_t \\ &= \tilde{\mu} + \sum_{i=0}^{\infty} \tilde{\rho}^i \epsilon_{t-i} \\ y_{t+1} - y_t &= \epsilon_{t+1} + (\tilde{\rho}-1)\sum_{i=0}^{\infty} \tilde{\rho}^i \epsilon_{t-i}. \end{aligned}$$

As a result, the initial level and subsequent change are negatively correlated: $\text{Cov}\left(y_t, y_{t+1} - y_t\right) = \frac{-1}{1+\tilde{\rho}}\sigma_\epsilon^2 < 0$.

We now consider commuter counts, which, unlike the AR(1) process above, must be integers. Consider a Poisson process with an AR(1)-style mean:

$$\begin{aligned} y_{t+1} &\sim \text{Poisson}\left((1-\rho)\mu + \rho y_t\right) \\ y_0 &\sim \text{Poisson}\left(\mu\right), \end{aligned}$$

where $\rho \in [0,1)$ is the serial correlation coefficient and $\rho = 0$ is the iid case. We define the initial condition to define the unconditional mean.

The following identities are helpful in showing the relationship between a count and its subsequent change:

$$\mathbb{E}[y_{t+1}] = \mu, \qquad \mathbb{E}\left[y_t^2\right] = \mu^2 + \frac{\mu}{1-\rho^2}, \qquad \mathbb{E}[y_{t+1} y_t] = \mu^2 + \frac{\rho\mu}{1-\rho^2}.$$

Using these, we can show that the covariance is negative:

$$\text{Cov}\left(y_t, y_{t+1} - y_t\right) = \text{Cov}\left(y_t, y_{t+1}\right) - \text{Var}\left(y_t\right) = \frac{\rho}{1-\rho^2}\mu - \frac{\mu}{1-\rho^2} = \frac{\rho-1}{1-\rho^2}\mu = \frac{-1}{1+\rho}\mu < 0.$$

This matches the expression for the continuous-valued distributions above.

APPENDIX D: DESCRIPTIVE STATISTICS AND DATA SOURCES

Section D.1 presents descriptive statistics for Detroit and Minneapolis-St. Paul analogous to those for New York City presented in the main text. Section D.2 examines commuting flows between counties. Section D.3 shows that commuting flows are often impersistent. Section D.4 shows that zero commuting flows are often asymmetric. Section D.5 shows that small and zero flows are a large proportion of cells in the migration matrix and that these flows are impersistent over time. Section D.6 shows that estimates of workplace fixed effects are biased by dropping observations with zero commuters. Section D.7 describes the data sources we employ.

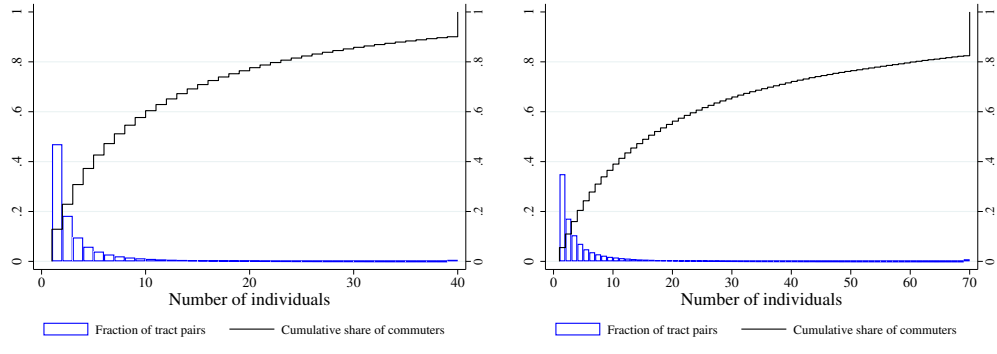
D.1. *Tract-level commuting flows for Detroit and Minneapolis-St. Paul*

This section presents summary statistics for tract-to-tract commuting counts in the Detroit urban area and the Minneapolis-St. Paul metropolitan area in 2014 analogous to those reported for New York City in Section 3.1. We chose these examples because Owens et al. (2020) study the Detroit urban area and Minnesota is the only state that reports employment by establishment rather than firm in the LODES data (Graham et al., 2014).

The Detroit urban area has about 1.3 million resident-employees and 1.4 million tract pairs, so the average cell in its commuting matrix is near one. Among the tract pairs, 74% have zero commuters between them. In Detroit, 42.6% of commuters have five or fewer commuters in their cell of the commuting matrix.

The Minneapolis-St. Paul metropolitan area has about 1.5 million resident-employees and 0.6 million tract pairs, so the average cell in its commuting matrix has less than three commuters. Among the tract pairs, 61% have zero commuters between them. In the Twin Cities, 24.3% of commuters have five or fewer commuters in their cell of the commuting matrix.

FIGURE D.1.—Number of commuters between pairs of tracts in Detroit and the Twin Cities



NOTES: These histograms report the number of tract pairs in the Detroit urban area (left panel) and the Minneapolis-St. Paul metropolitan area (right panel) by the number of individuals who reside in the origin tract and work in the destination tract in 2014 LODES data. These histograms restrict the samples to pairs of tracts with a strictly positive number of commuters. Observations larger than the 99th percentile are winsorized (40 commuters for Detroit urban area and 70 commuters for Minneapolis-St. Paul metropolitan area). Detroit has 1,166 residential census tracts. The Twin Cities has 789 residential census tracts.

The spatial concentration of employment contributes to the sparsity of these commuting matrices. The median tract in Detroit has 465 employees working in it. Since Detroit has 1,166 residential tracts, at least 60% of locations must have zero residents commuting to this workplace.

D.2. *County-level commuting flows*

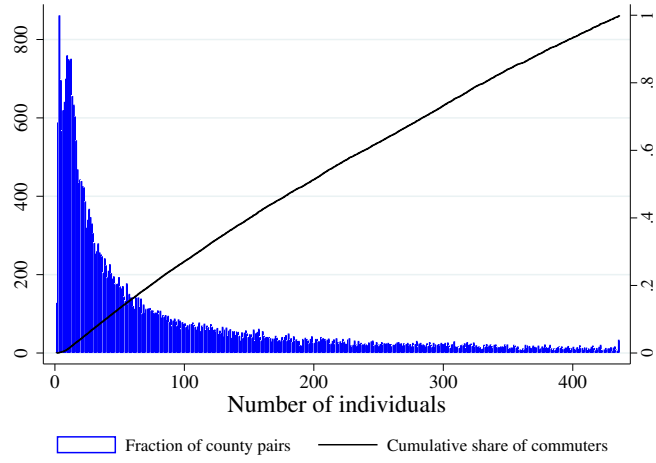
Even when studying larger geographic units, many pairs of locations may have small flows. Consider US counties, which vary greatly in population size. In the 2006–2010 American Community Survey (ACS) data, there are 136 million commuters (with commutes less than 120 kilometers).[7] Of those 136 million, 101 million live and work in the same county, so there are 35 million cross-county commuters between 79,188 pairs of counties. Thus, the average off-diagonal element of the county-to-county commuting matrix has 445 commuters. However, the distribution of commuters is extremely uneven. The top 10 county pairs account for more than 2 million commuters alone. For the bottom 90% of off-diagonal observations, the mean value is only 40 commuters. Almost half of the county pairs report zero commuters. Among those reporting positive values, Figure D.2 shows that this distribution is skewed, so that many thousands of county pairs have small numbers of commuters. Unlike census tracts, however, these small flows constitute a small share of the total number of commuters. That is, US counties with tens of thousands of residents do not exhibit the patterns we described for granular settings, but many county pairs have small commuting flows and show similar patterns.

In practice, the uncertainty accompanying these small flows is severely compounded by the fact that the ACS is a 1-in-20 sample of the population. In three different five-year waves of the ACS, nearly half of the county pairs within 120 kilometers of each other are reported to have zero commuters, as shown in Table D.1. More than half of the non-zero county pairs report fewer than 100 commuters, therefore representing the behavior of five or fewer respondents. As a consequence, as shown in the third column of Table D.1, for more than one-third of the county pairs with positive commuting flows, the Census-reported margin of error exceeds the reported number of commuters.

Commuting data from other countries show similar features of granular settings. For example, one-quarter of Germany's county (Kreisfreie Städte and Landkreise) pairs within 120 kilometers have fewer than 10 commuters (Krebs and Pflüger, 2023). In Brazil's 2010 Censo Demográfico, which reports estimates based on a 10% sample of the population, there are 81 million commuters (with commutes less than 60 kilometers).[8] There are 7 million cross-municipality commuters between 131,620 pairs of municipalities. Thus, the average off-

[7] We follow Monte et al. (2018) by restricting attention to county pairs that are less than 120 kilometers apart.
[8] Dingel et al. (2021) use these data to construct metropolitan areas based on commuting flows.

FIGURE D.2.—Number of commuters between US counties



NOTES: This histogram depicts the number of county pairs by the estimated number of commuters in the 2006–2010 ACS. The sample is restricted to pairs of distinct counties within 120 kilometers, the smallest 90% of estimated commuter counts among such counties, and only those pairs reporting a strictly positive number of commuters.

TABLE D.1

ZEROS IN US COUNTY-TO-COUNTY COMMUTING MATRIX

Dataset	Zero Pairs	Positive Pairs	MOE >X (%)
ACS 2006-2010	36,403	42,785	37
ACS 2009-2013	35,547	43,641	36
ACS 2011-2015	35,002	44,096	35

NOTES: This table reports the number of county pairs with zero commuters and non-zero commuters for three editions of the ACS. The sample is restricted to pairs of counties within 120 kilometers of each other. The final column reports the share of county pairs for which the Census-reported 90% margin of error exceeds the (strictly positive) reported number of commuters.

diagonal element of the municipality-to-municipality commuting matrix has only 56 commuters. About three-quarters of the cells in this commuting matrix are empty.

D.3. *Commuting counts are often impersistent*

The integer values appearing in commuting matrices are not very persistent. A residence-workplace pair may have three commuters one year and none the next. Another pair of locations may double its number of commuters, from one to two. While the conventional continuum approach interprets these changes as substantial economic shifts, the finite-sample perspective is that these changes are not very informative if they are merely a symptom of small counts. We document substantial churn in commuting counts from year to year, suggesting that there is considerable finite-sample noise in addition to signal in these commuting counts.

Table D.2, which presents the transition matrix for pairs of tracts in the Detroit urban area and New York City between the years 2013 and 2014, demonstrates considerable impersistence. In the Detroit urban area, for pairs with one to four commuters in 2013, the percentages appearing on the diagonal of the transition matrix are quite low. A pair of tracts with one commuter in 2013 was almost three times as likely to have zero commuters in 2014 than to have one commuter. A pair with four commuters in 2013 was more likely to appear in any other column in 2014 than to report four commuters again. The 86% of pairs that had zero commuters in both years may appear to suggest persistence, but this is primarily a symptom of the fact that in both years about three-quarters of observations are zero.[9] More than 130,000 pairs of tracts that had zero commuters in 2013 had at least one commuter in 2014. At the same time, 39% of Detroit tract pairs with positive flow in 2013 were zeros in 2014 . Thus, while zeros are pervasive in this commuting matrix, they are not very persistent. The results for New York City in Table D.2 are very similar to those for Detroit. The commuter counts are so impersistent that, for many tract pairs, a gravity-based estimate predicts a tract pair's commuter count in 2014 better than its observed count in 2013 does.

TABLE D.2

THE IMPERSISTENCE OF COMMUTING COUNTS FOR TRACT PAIRS IN DETROIT AND NYC

(a) Detroit

	2014					
2013	0	1	2	3	4	5+
0	0.86	0.10	0.02	0.01	0.00	0.00
1	0.60	0.22	0.10	0.04	0.02	0.02
2	0.37	0.25	0.16	0.09	0.06	0.08
3	0.23	0.22	0.18	0.13	0.08	0.16
4	0.15	0.17	0.17	0.14	0.11	0.26
5+	0.04	0.06	0.07	0.08	0.08	0.68

(b) NYC

	2014					
2013	0	1	2	3	4	5+
0	0.91	0.07	0.01	0.00	0.00	0.00
1	0.65	0.20	0.08	0.04	0.02	0.02
2	0.39	0.25	0.15	0.09	0.05	0.07
3	0.24	0.22	0.17	0.12	0.08	0.16
4	0.15	0.17	0.17	0.14	0.11	0.27
5+	0.03	0.05	0.06	0.07	0.07	0.71

NOTES: This table describes pairs of tracts in the Detroit urban area (left panel) and New York City (right panel) by reported number of commuters in the 2013 and 2014 LODES. It is a transition matrix, in which each cell lists the share of tract pairs in that row (number of commuters in 2013) that belong to that column (number of commuters in 2014). Each row sums to 100%, modulo rounding.

We note considerable churn even when using larger geographic units. Table D.3 shows that 22% of the county pairs reporting zero commuters in the 2006–2010 ACS reported a positive number of commuters in the following five-year interval. Conversely, about 18% of

[9] If $p \in [0, 1]$ of the pairs were randomly independently assigned zero in each period, then p^2 of those pairs would lie in the upper left cell of the transition matrix. Thus, even if zeros were randomly independently assigned to three-quarters of the tract pairs in each period, nine-sixteenths of the pairs would be zero in both periods. That would not be evidence of persistence.

pairs reporting a positive number of commuters had zero commuters in the following five-year interval. For pairs of counties with a strictly positive number of commuters smaller than 111 in 2006–2011, the diagonal elements of the transition matrix are less than half. For example, a pair of counties reported to have 71–90 commuters in 2006–2011 has only a 14% probability of appearing in the same bin in the following five-year interval.

TABLE D.3

THE IMPERSISTENCE OF COMMUTING COUNTS FOR PAIRS OF US COUNTIES

2006–2010	Initial Share(%)	2011–2015: 0	1–30	31–50	51–70	71–90	91–110	111–500	501–1,500	>1,500
0	45.88	0.78	0.18	0.02	0.01	0.00	0.00	0.00	0.00	0.00
1–30	19.64	0.35	0.46	0.10	0.05	0.02	0.01	0.02	0.00	0.00
31–50	5.47	0.16	0.36	0.19	0.12	0.07	0.04	0.07	0.00	0.00
51–70	3.42	0.08	0.26	0.18	0.15	0.10	0.08	0.15	0.00	0.00
71–90	2.50	0.05	0.16	0.15	0.16	0.14	0.12	0.23	0.00	0.00
91–110	1.86	0.02	0.11	0.12	0.15	0.12	0.13	0.35	0.00	0.00
111–500	12.02	0.00	0.02	0.03	0.04	0.05	0.05	0.74	0.07	0.00
501–1,500	5.01	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.81	0.06
>1,500	4.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.94

NOTES: This table presents a transition matrix for pairs of counties within 120 kilometers of each other by reported number of commuters in two editions of the ACS. The first column reports the percentage of county pairs that were in each row in the 2006–2010 data. The remaining columns report the share of county pairs within the row that appeared in the corresponding bin in the 2011–2015 data. By definition, the shaded cells in each row sum to one, modulo rounding. The bin boundaries are arbitrary, since Figure D.2 shows no obvious bunching. We found similar impersistence when using alternative boundaries.

To the extent that the observed impersistence of commuting counts is a small-sample problem, these findings caution against procedures that infer structural parameters from the relative magnitudes of these counts. The difference between one commuter and two commuters (or one respondent and two respondents in a finite sample) is little evidence that the latter outcome was twice as probable. Similarly, procedures that rationalize observations with zero commuters by imposing infinite commuting costs so that these are zero-probability events rule out potential margins of adjustment based on zeros that appear to be largely transitory.

Tract-pair-level commuter counts are so impersistent that a gravity model estimated using an observed bilateral characteristic like transit time or distance can sometimes predict future commuter counts better than the observed value. Table D.4 shows that, for tract pairs with fewer than ten commuters reported, a gravity-based estimate predicts the following year's value better than its current value does. The fitted values from a gravity model estimated using 2013 data have a higher R^2 for predicting observed 2014 values than the observed 2013 values in both Detroit and New York City. Using the observed 2013 values yields better predictions only for the tract pairs with the largest 2% of commuter counts.

TABLE D.4

GRAVITY-BASED ESTIMATES PREDICT 2014 VALUES BETTER THAN 2013 VALUES DO

# of commuters	Share	Gravity: time	2013 values	Gravity: distance	2013 values
Panel A: Detroit					
≤ 5	0.960	0.384	0.308	0.367	0.307
≤ 10	0.983	0.494	0.473	0.465	0.472
Panel B: NYC					
≤ 5	0.978	0.362	0.306	0.373	0.306
≤ 10	0.990	0.474	0.475	0.477	0.473

NOTES: Each row reports results for all tract pairs in Detroit (upper panel) or New York City (lower panel) with fewer than 5 or 10 commuters in the 2013 LODES data. The "Share" column reports the share of tract pairs covered by the row. The "Gravity" columns report the R^2 obtained by regressing the 2014 number of commuters on the number of commuters predicted by a gravity model estimated using 2013 data. The "Gravity: time" column estimates the gravity equation as described in Section 3.2. The "Gravity: distance" column uses $\ln \delta_{kn} = \ln \text{distance}_{kn}$ rather than commuting costs defined in Section 3.2. The "2013 values" columns report the R^2 obtained by regressing the 2014 number of commuters on the 2013 number of commuters. All estimated slope coefficients are positive. The two "2013 values" columns differ because the regression sample is within observations where predictions from the gravity model are available. In the distance specification, $k = n$ observations are dropped. For NYC, the commute is infeasible for two observations.

D.4. *Zeros are often asymmetric*

The zeros in commuting matrices are often asymmetric. Denoting the number of commuters living in residence k and working in workplace n by ℓ_{kn}, an observed zero is asymmetric when $\ell_{nk} = 0$ and $\ell_{kn} > 0$. For US counties, $\ell_{nk} = 0$ for 22% of county pairs with $\ell_{kn} > 0$. In Detroit, $\ell_{nk} = 0$ for 66% of tract pairs with $\ell_{kn} > 0$. For Brazilian municipalities, $\ell_{nk} = 0$ for 49% of municipio pairs with $\ell_{kn} > 0$. These asymmetric flows are not explained by asymmetric numbers of total workers or residents: in Detroit, $\ell_{nk} = 0$ for 47% of tract pairs for which $\ell_{kn} > 0$ and total employment in k and n differs by 10% or less.

The fact that commuting matrices' zeros are often asymmetric poses a puzzle for calibration procedures that rationalize zero-commuter observations by infinite commuting costs. This interpretation of zeros implies severely asymmetric commuting costs, even though daily commutes are round-trip journeys. If we believe that commuting from n to k is impossible because we observe $\ell_{nk} = 0$, how do the individuals who live in k and work in n commute home at the end of the day? The mechanisms generating prohibitive commuting costs would have to exhibit within-day variation. But the most plausible source of intraday variation in commuting costs — congestion caused by a large number of commuters — cannot explain residence-workplace pairs that have no commuters.

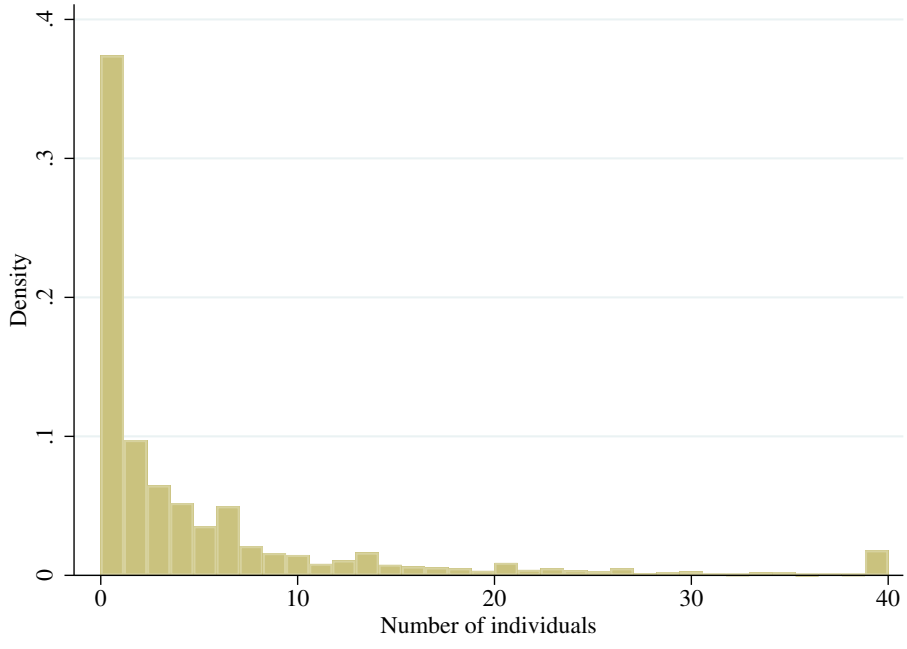
In many empirical settings, asymmetric zeros may simply reflect small flows rather than evidence of very asymmetric commuting costs. When most pairs of locations have zero commuters, and the modal pair with a positive number reflects the decision of only one

commuter or one survey respondent, the difference between zero and one isn't particularly informative.

D.5. *State-to-state migration flows in ACS data*

Using ACS data from 2001 and 2002, we show that small and zero flows are a large proportion of cells in the migration matrix and that these flows are impersistent over time. The ACS 2001 dataset has 644,427 prime-age individuals out of the total sample of 1,192,206 individuals. Of these, 73,101 individuals moved residences. Of the 73,101 individuals who moved, 80.6% migrated within their states. We observe a total of 14,215 between-state movers out of the 644,427 prime-age individuals in the 2001 ACS.

FIGURE D.3.—Histogram of migration flows between US states, 2001



Given the 14,125 between-state movers in 2001 across 2,550 state pairs, the average cell in the migration matrix has 5.6 movers. About one-third are zero. The histogram of the positive state-to-state migration counts in Figure D.3 exhibits features similar to the histogram of tract-to-tract commuting counts: a large number of very small flows and a long right tail. About 73% of cells have fewer migrants than the average value of 5.6. Values of 1 to 5 each account for about 10% of cells in the migration matrix.

The 2001–2002 transition matrix in Table D.5 shows considerable impersistence in migration flows. The particular value of a small state-to-state flow is not stable from year to year, similar

TABLE D.5

TRANSITION MATRIX FOR STATE MIGRATION FLOWS BETWEEN 2001 AND 2002

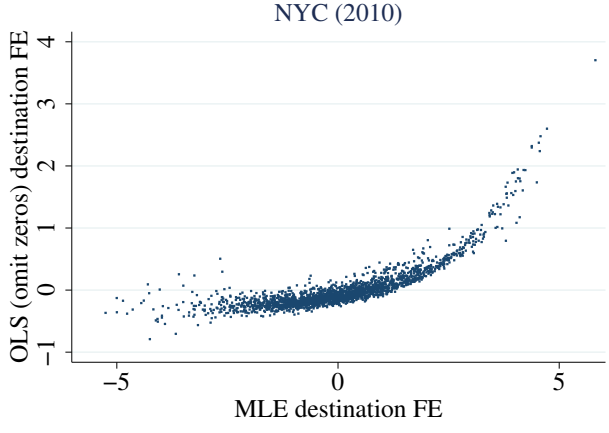
	2002						
2001	0	1	2	3	4	5	6+
0	0.59	0.16	0.13	0.05	0.03	0.01	0.02
1	0.40	0.17	0.17	0.10	0.06	0.05	0.06
2	0.32	0.16	0.21	0.09	0.07	0.07	0.09
3	0.23	0.16	0.21	0.09	0.08	0.09	0.14
4	0.26	0.13	0.15	0.10	0.06	0.06	0.25
5	0.13	0.09	0.12	0.14	0.14	0.06	0.31
6+	0.03	0.03	0.06	0.06	0.06	0.07	0.69

to the transition matrix for commuting counts. In Table D.5, migration flows are winsorized at 6 and values are given as a percentage of 2001 migration flows, so the rows sum to one. The diagonal elements, for which the number of migrants is the same across the two years, are quite small. A significant portion of the positive migration flows in 2001 are zero in 2002, as shown in the first column of the table.

Foschi et al. (2023) also report that the ACS migration data exhibit finite-sample problems and suggest using Internal Revenue Service migration counts instead.

D.6. *OLS estimation bias*

FIGURE D.4.—Destination fixed effects from tract-to-tract gravity regressions



NOTES: This plot depicts the destination fixed effects estimated in Table 1's column 1 (horizontal axis) and column 2 (vertical axis). See the notes to Table 1 for details.

Figure D.4 shows that estimates of workplace fixed effects are biased by dropping observations with zero commuters. When only using positive-commuter observations, the residence and workplace fixed effects characterize the average number of commuters for a residence-workplace pair, conditional on the number of commuters being strictly greater than zero. These conditional averages are necessarily greater than the unconditional averages, and the difference is larger for workplace locations with fewer workers.[10] Figure D.4 contrasts the OLS (vertical axis) and maximum likelihood (horizontal axis) estimates of the workplace fixed effects. The difference is stark: the range of the OLS estimates is half that of the maximum likelihood estimates because of considerable truncation from below. In essence, the popular practice of omitting zero-commuter observations attributes low-employment destinations' lower employment counts to infinite commuting costs, not lower wage beliefs (lower productivity). By contrast, our maximum likelihood estimator infers that these destinations are less attractive workplaces from the fact that many origins have zero residents working in these destinations.

D.7. *Data sources*

This subsection describes the data sources we employ.

Longitudinal Employer-Household Dynamics, Origin-Destination Employment Statistics (LODES). The block-level origin-destination data for New York City, Detroit, and Minneapolis-St. Paul are obtained from LODES version 7.4, published by the US Census Bureau.[11] Following Owens et al. (2020), we use the primary job counts.

The origin-destination data for NYC, Detroit, and Minneapolis-St. Paul are in the following files:

Geography	filepath within `https://lehd.ces.census.gov/data/lodes/LODES7/`
NYC	from `ny/od/ny_od_main_JT01_2002.csv.gz`
	to `ny/od/ny_od_main_JT01_2017.csv.gz`
Detroit	from `mi/od/mi_od_main_JT01_2009.csv.gz`
	to `mi/od/mi_od_main_JT01_2014.csv.gz`
Minneapolis-St. Paul	from `mn/od/mn_od_main_JT01_2010.csv.gz`
	to `mn/od/mn_od_main_JT01_2014.csv.gz`
	from `mn/od/mn_od_aux_JT01_2010.csv.gz`
	to `mn/od/mn_od_aux_JT01_2014.csv.gz`
	from `wi/od/wi_od_main_JT01_2010.csv.gz`
	to `wi/od/wi_od_main_JT01_2014.csv.gz`
	from `wi/od/wi_od_aux_JT01_2010.csv.gz`
	to `wi/od/wi_od_aux_JT01_2014.csv.gz`

[10] Census tracts are defined so that the number of residents is similar across tracts, while there is tremendous heterogeneity in total employment. Thus, the selection bias is evident in the workplace fixed effects. When estimating the analogous gravity regression for county-to-county commuting flows, we find that the selection bias manifests in both the origin and destination fixed effects.

[11] See the official documentation for details of the data structure.

The Minneapolis-St. Paul core-based statistical area includes counties in both Minnesota and Wisconsin. The LODES `aux` files describe jobs in which the workplace is in the state and the residence is outside the state.

In these files, the relevant variables are

- `w_geocode` 15-digit workplace census block code. We aggregate it into 11-digit FIPS census tract code as the workplace identifier.
- `h_geocode` 15-digit residence census block code. We aggregate it into 11-digit FIPS census tract code as the residence identifier.
- `S000` Total number of jobs.
- `SE01, SE02, SE03` Number of jobs with earnings $1250/month or less, $1,251/month to $3,333/month, and greater than $3,333/month. We use the three variables and the maximum wage obtained from ZIP Code Business Statistics to calculate the annual average payroll per employee in the workplace tract.

Five-Year American Community Survey Commuting Flows and Employment Data. We summarize 2006-2010 county-level commuting statistics using data from the American Community Survey. It can be downloaded from Table 1. Residence County to Workplace County Flows for the United States and Puerto Rico Sorted by Residence Geography: 2006-2010.

- `StateID_h, StateID_w` 2-digit FIPS state code for residence or workplace.
- `CountyID_h, CountyID_w` 3-digit FIPS county code for residence or workplace.
- `E` The tract-to-tract number of commuters.

NYC tract-to-tract transit time. We use data from Davis et al. (2019) describing pairs of New York City tracts in terms of the travel times by public transport. It can be downloaded from https://github.com/jdingel/DavisDingelMonrasMorales/raw/master/initialdata/input/tractpairs.dta.zip.

- `geoid11_orig` Origin 11-digit FIPS geographic identifier.
- `geoid11_dest` Destination 11-digit FIPS geographic identifier.
- `traveltime_public` Travel time by public transportation in minutes from Google Maps. We use it to compute the commuting cost $\bar{\delta}_{kn}$.

Detroit tract-to-tract transit time. Our data on tract-to-tract transit time in Detroit come from `Kij_GoogleTime.xlsx` and `Tract_Classification.xlsx` in the Owens et al. (2020) replication files. Note that the benchmark year in our analysis is 2014. Our analysis covers the 297 census tracts in Detroit, as well as the surrounding adjacent metro area (Wayne County, Oakland County, and Macomb County), which includes 866 additional tracts.

- `duration_minutes` Tract-to-tract travel time in minutes. We reshape the data and obtain the IDs for workplace and residence tracts. We then use the travel time to compute the commuting cost $\bar{\delta}_{kn}$.
- `work_ID` Workplace identifier.
- `tract_str` 11-digit FIPS geographic identifier that maps workplace and residence IDs to census tract code.

US census tract-level geographical information. The distance information for Detroit and NYC are retrieved from the US census tract-level geographical information.

- `geoid` 11-digit FIPS census tract ID.
- `intptlat, intptlong` Current latitude and longitude of the interior point, used to calculate geodesic distance between the centroids.

ZIP Code Business Statistics. We apply the method in Owens et al. (2020) to calculate a weighted wage from ZIP Code Business Patterns. The yearly employed data are from https://www2.census.gov/: `econ2006/CB/sector00/CB0800CZ1.zip` to `econ2010/CB/sector00/CB0600CZ1.zip` for data between 2008 and 2010, and from `econ2011/CB/sector00/CB1100CZ11.zip` to `econ2013/CB/sector00/CB1100CZ11.zip` for data between 2011 and 2013.

- `zipcode` 5-digit zip code.
- `emp_f` Flag for the range of the number of employees.
- `emp` Number of employees, replaced with the midpoint of the associated bins indicated by `emp_f` if censored.
- `payqtr1_f, payann_f` Flags for first-quarter payroll and annual payroll, used for dropping observations if classified as missing data.
- `payann` Annual payroll ($1,000).

ZIP-tract crosswork. To aggregate the zip-level weighted wage to the tract level, we join the ZIP data with the HUD-USPS Zip-to-Tract crosswalk. The annual ZIP-to-tract crosswalks are available from https://www.huduser.gov/portal/datasets/usps/: `ZIP_TRACT_122010.xlsx` to `ZIP_TRACT_122013.xlsx`.

- `ZIP` 5-digit zip code.
- `TRACT` 11-digit FIPS census tract code.
- `BUS_RATIO` The ratio of business addresses in the ZIP-tract intersection to the total number of business addresses in the entire ZIP. It is used as the proportion to reweight the employment and annual payroll data. We sum by census tract to arrive at the tract-level employment and annual payroll for each census tract. Then, we divide payroll by employment to obtain the tract-level average payroll per employee and its maximum value.

NYC Neighborhood Tabulation Areas (NTA)-tract crosswalk. To aggregate the NYC census tract level data to the Neighborhood Tabulation Areas defined by the New York City Department of City Planning, we use the 2010 census tract to 2010 NTA equivalency.

- `NeighborhoodTabulationAreaNT` 4-digit NTA code.
- `CensusBureauFIPSCountyC` 3-digit FIPS census county code.
- `CensusTract` 9-digit FIPS census tract code.

APPENDIX E: FINITE MODEL: PRICE DISPERSION AND "EX POST REGRET"

E.1. *Contrast with prices from continuum model*

The realized equilibrium rents and wages differ generically from the point-mass beliefs about rents and wages that govern individuals' choices of residences and workplaces. Because equilibrium prices solve a system of non-linear equations, the average realized equilibrium prices may differ from the point-mass beliefs. As described below, these differences are small at our baseline parameter values: the 95$^{\text{th}}$-percentile tract's absolute percent deviation of mean realized price from the continuum-case rational expectation is 0.72% for wages and 0.03% for rents.[12]

Equilibrium prices in the model with a finite number of individuals solve a system of non-linear equations, so their expected values are not necessarily equal to the continuum-case prices. The labor supply to workplace n is $L_n = \sum_k \frac{\ell_{kn}}{\bar{\delta}_{kn}}$. This is a random variable whose realized value depends on the realization of $\boldsymbol{\nu}^I$. Labor demand in workplace n, when there are local increasing returns, can be written as a rearrangement of equation (C.13):

$$w_n = \bar{A}_n^{\frac{\sigma-1}{\sigma}} L_n^{\eta - \frac{(1+\eta)}{\sigma}} \left(\frac{Y}{P^{1-\sigma}} \right)^{1/\sigma}.$$

For brevity, define the inverse labor demand elasticity $\tilde{\eta} \equiv \eta - \frac{(1+\eta)}{\sigma}$. Choose the numeraire so that $\frac{Y}{P^{1-\sigma}} = 1$.

The equilibrium wage is given by the intersection of the labor demand curve and the randomly realized quantity of labor supplied L_n. The equilibrium wage in workplace n as a function of this random variable L_n is

$$w_n\left(L_n\right) = L_n^{\tilde{\eta}} \bar{A}_n^{\frac{\sigma-1}{\sigma}}.$$

The equilibrium wage in the continuum model is $w_n\left(\mathbb{E}\left[L_n\right]\right)$, while the mean equilibrium wage in the model with a finite number of individuals is $\mathbb{E}\left[w_n\left(L_n\right)\right]$. These values coincide when demand is perfectly elastic ($\tilde{\eta} = 0$) or when increasing returns are so large that wages are linear in quantity supplied ($\tilde{\eta} = 1$), but otherwise $w_n()$ is not a linear operator and thus these values differ by Jensen's inequality. Similarly, rearranging equation (4) shows that equilibrium rents, $r_k = \frac{\alpha}{T_k} \sum_n y_{kn}$, are a non-linear function of realized labor quantities and wages.

Using $f_n(L_n)$ to denote the probability mass function of supplied labor, the difference between the continuum model's wage and the mean wage in the model with a finite number of individuals is

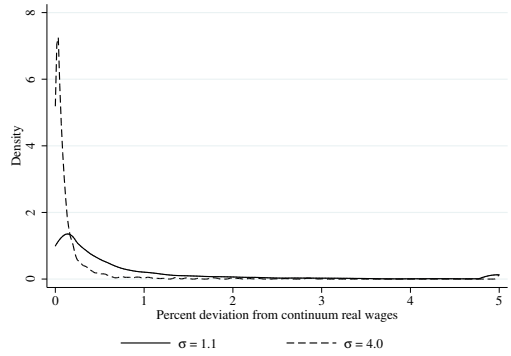
$$\frac{\mathbb{E}\left[w(L_n)\right]}{w(\mathbb{E}\left[L_n\right])} - 1 = \int_0^\infty \left[\left(\frac{L_n}{\mathbb{E}[L_n]} \right)^{\tilde{\eta}} - 1 \right] f(L_n) \mathrm{d}L_n.$$

[12] These differences are larger when the labor demand elasticity σ is lower: 3.61% for wages and 0.05% for rents when $\sigma = 1.1$.

As noted above, this difference is zero when labor demand is perfectly elastic ($\eta = 0$ & $\sigma = \infty$).

To quantify how the difference in wages varies with the labor demand elasticity, we compute 100,000 simulations of equilibrium outcomes in the model with a finite number of individuals for 100,000 realizations of ν^I when $\sigma = 4$ and when $\sigma = 1.1$. We compute the average wage in each tract across these 100,000 equilibria, excluding the rare instances in which a tract's equilibrium wage is undefined because its realized labor supply L_n is zero. The average realizations of wages and rents across 100,000 simulations of the model with a finite number of individuals are very close to the continuum-case rational expectation of these prices at our baseline parameter values. The gap between the two model's prices is larger when the labor demand elasticity is lower. Figure E.1 shows the two distributions of the wage difference across the 2,143 workplace tracts for the two labor demand elasticities. The median tract's absolute percentage point deviation of mean realized price from the continuum-case rational expectation is 0.06% for wages and 0.01% for rents. When $\sigma = 1.1$, these differences are 0.32% for wages and 0.03% for rents. The values of the 95$^{\text{th}}$-percentile differences are reported in Section 4.4.

FIGURE E.1.—Gap between mean wage and continuum-case expectation of wage



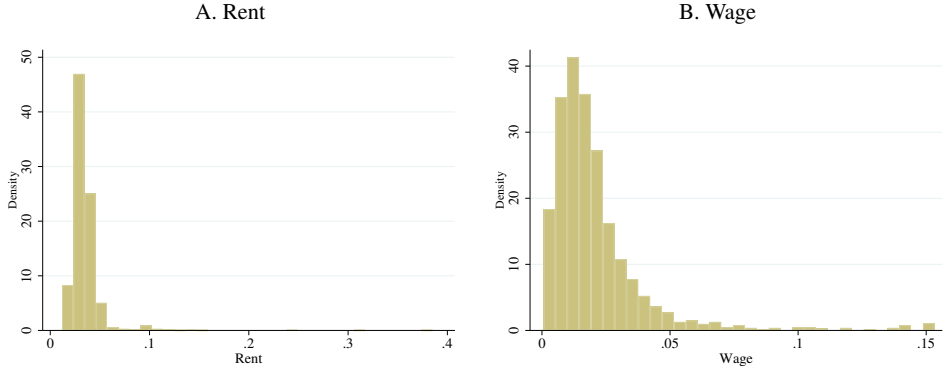
NOTES: This figure depicts the absolute percentage-point deviation of the mean realized wage from the continuum-case rational expectation of the wage across the 2,143 workplace tracts. We compute the mean wage using 100,000 simulations of the model with a finite number of individuals. This mean excludes the rare instances in which a tract's equilibrium wage is undefined because its realized labor supply L_n is zero. Deviations larger than five percentage points are winsorized to five.

E.2. *Dispersion of real wages and rents relative to beliefs*

Realized equilibrium prices vary with individual idiosyncrasies. Figure E.2 plots the dispersion of prices for each tract in New York City for the estimated model with a finite number of individuals. We define price dispersion as the standard deviation of the price in the commuting equilibrium with finitely many individuals across 100,000 simulations divided by the continuum-case rational expectation of that price. Across tracts, the median level of this measure of dispersion, akin to the coefficient of variation, is 0.016 for wages and is 0.032 for

rents. While this dispersion may be considered sizable, these differences between expectations and realized prices do not necessarily imply that individuals would move if we relaxed the irreversibility assumption. Individuals' decisions depend on their idiosyncratic preferences, so many are inframarginal.

FIGURE E.2.—Price dispersion across equilibria



NOTES: The plots depict the dispersion of prices (r_k/P or w_n/P) for each tract in New York City using the model with a finite number of individuals estimated on 2010 data. We define price dispersion as the standard deviation of the price in the commuting equilibrium with finitely many individuals across 100,000 simulations divided by the continuum-case rational expectation of that price. Panel A depicts the dispersion in tracts' rents, which have a median value of 0.032 (p5 = 0.021, p95 = 0.051). Panel B depicts the dispersion in tracts' wages, which have a median value of 0.016 (p5 = 0.004, p95 = 0.054).

E.3. *Ex post regret*

How often would individuals choose a different residence-workplace pair after observing the realized equilibrium prices if they were able? We compute the share of individuals with ex post regret and their magnitudes as follows. Realized equilibrium rents and wages are observed for locations with positive residents and positive employment, the sets $\mathcal{K}$ and $\mathcal{N}$, respectively. We define the magnitude of "ex post regret" as the increase in income an individual would require as compensation to not change their choice given these realized prices. In particular, at realized rents $\{r_k\}$ and wages $\{w_n\}$, for individual i who chose residence-workplace pair kn, ex post regret χ_i is implicitly defined such that χ_i solves

$$\max_{k' \in \mathcal{K}, n' \in \mathcal{N}} \left(\epsilon \ln \left(\frac{w_{n'}}{P^{1-\alpha} r_{k'}^{\alpha} \delta_{k'n'}} \right) + \nu^i_{k'n'} \right) = \left(\epsilon \ln \left(\frac{(1+\chi_i) w_n}{P^{1-\alpha} r_k^{\alpha} \delta_{kn}} \right) + \nu^i_{kn} \right),$$

where the left side is the individual's maximum utility at observed realized prices, the right side is the individual's utility from kn with their income multiplied by $(1+\chi_i)$, and by definition kn was the choice maximizing $\tilde{U}^i_{k,n}$. Denote $\underline{k}$ and $\underline{n}$ as the optimal choices of k' and n' for

individual i. Given kn, we can solve for the ex post regret χ_i based on the following equation:

$$\epsilon \ln\left(\frac{w_{\underline{n}}}{P^{1-\alpha} r_{\underline{k}}^{\alpha} \delta_{\underline{kn}}}\right) + \nu^i_{\underline{kn}} = \epsilon \ln\left(\frac{(1+\chi_i) w_n}{P^{1-\alpha} r_k^{\alpha} \delta_{kn}}\right) + \nu^i_{kn},$$

which implies that

$$\chi_i = \left(\frac{e^{\left(\frac{1}{\epsilon}\nu^i_{\underline{kn}}\right)} w_{\underline{n}} r_{\underline{k}}^{-\alpha} \delta_{\underline{kn}}^{-1}}{e^{\left(\frac{1}{\epsilon}\nu^i_{kn}\right)} w_n r_k^{-\alpha} \delta_{kn}^{-1}}\right) - 1. \tag{E.1}$$

Individual i's regret is zero if kn is their optimal choice at realized prices.

We simulate 2.5 million individuals to compute this statistic. Each of these individuals uses the point-mass beliefs from the continuum model. The realized wages and rents differ because of idiosyncratic preference shocks that do not wash out in the aggregate.

Table E.1 reports the distribution of ex post regrets. In this setting, we find that ex post regret is quantitatively modest. 96% of individuals would not want to change their residence-workplace choice. For the 4% who would want to switch, the median ex post regret χ_i is equal to 0.7%.

TABLE E.1

DISTRIBUTION OF EX POST REGRETS

		Unconditional distribution					Conditional distribution	
s	Share with regret	p95	p96	p97	p98	p99	Mean	Median
1	0.0442	0.0000	0.0011	0.0042	0.0082	0.0150	0.0106	0.0073
2	0.0433	0.0000	0.0009	0.0039	0.0078	0.0143	0.0102	0.0071
3	0.0446	0.0000	0.0012	0.0043	0.0083	0.0150	0.0106	0.0072
4	0.0446	0.0000	0.0012	0.0043	0.0084	0.0152	0.0106	0.0073
5	0.0437	0.0000	0.0010	0.0040	0.0079	0.0144	0.0103	0.0071
6	0.0444	0.0000	0.0012	0.0042	0.0083	0.0150	0.0107	0.0073
7	0.0447	0.0000	0.0013	0.0043	0.0083	0.0150	0.0105	0.0072
8	0.0445	0.0000	0.0012	0.0043	0.0084	0.0150	0.0106	0.0073
9	0.0452	0.0000	0.0014	0.0045	0.0086	0.0154	0.0109	0.0074
10	0.0444	0.0000	0.0011	0.0042	0.0082	0.0148	0.0106	0.0072
mean	0.0444	0.0000	0.0012	0.0042	0.0083	0.0149	0.0106	0.0072

NOTES: The table reports the share of individuals with ex post regret and the distribution of ex post regrets over their desired switches in simulations of our estimated model with a finite number of individuals. The first column identifies the simulation s. The second column reports the fraction of individuals who have ex post regret and therefore would prefer a different choice given realized prices. Columns under "Unconditional distribution" report the distribution of ex post regret based on the full sample ($I = 2,488,905$). Columns under "Conditional distribution" report the distribution of ex post regret among those who would want to switch. The "Mean" row reports the mean value across the 10 simulations.

Table E.2 compares the price dispersion in the 100,000 simulations depicted in Figure E.2 and the 10 simulations depicted in Table E.1. The simulations in Figure E.2 each require one

draw from a multinomial distribution with 4.6 million outcomes whose probabilities appear in equation (10). Generating the labor allocation for each simulation by this method requires less than one second of computing. The simulations in Table E.1 involve more than 11 trillion draws from the type-1 extreme value distribution. Generating the labor allocation for each simulation by this method requires about 100 hours of computing. Hence, the latter simulations are much more computationally expensive. It is heartening that the distributions of price dispersion summarized in Table E.2 are very similar.

TABLE E.2

PRICE DISPERSION ACROSS SIMULATION METHODS

Simulation count	mean	p5	p10	p25	p50	75	p90	p95
Wage								
100,000	0.021	0.004	0.006	0.010	0.016	0.025	0.039	0.054
10	0.023	0.004	0.005	0.009	0.015	0.025	0.039	0.057
Rent								
100,000	0.035	0.021	0.024	0.027	0.032	0.038	0.045	0.051
10	0.034	0.016	0.019	0.024	0.031	0.038	0.049	0.058

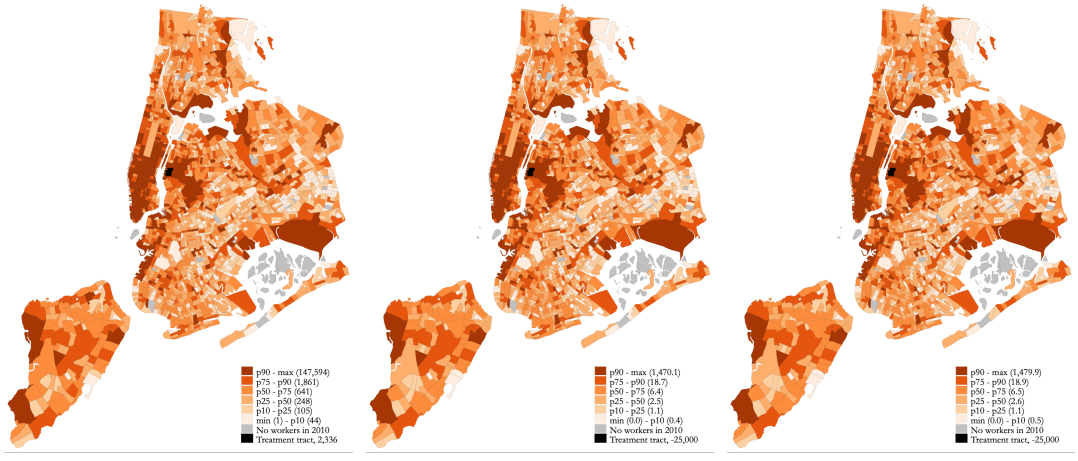
NOTES: This table compares the price (r_k/P or w_n/P) dispersion generated by the simulations used in Section E.2 (100,000 simulations) and Section E.3 (10 simulations). As described in the text, each tract's price dispersion is the standard deviation of the price in the commuting equilibrium with finitely many individuals across simulations divided by the continuum-case rational expectation of that price. The first column reports the average price dispersion across all tracts. The remaining columns report the level of price dispersion for tracts at selected percentiles.

APPENDIX F: AMAZON HQ2 COUNTERFACTUAL

F.1. *Geographic incidence of Amazon HQ2*

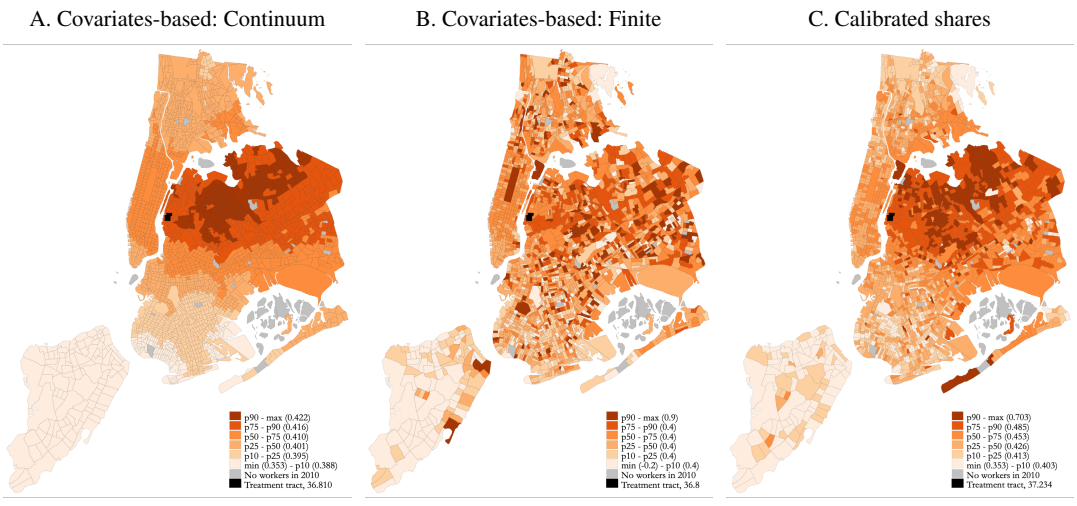
FIGURE F.1.—Amazon HQ2 counterfactual decreases in workers

A. Employment in 2010 | B. Covariates-based decreases | C. Calibrated-shares decreases



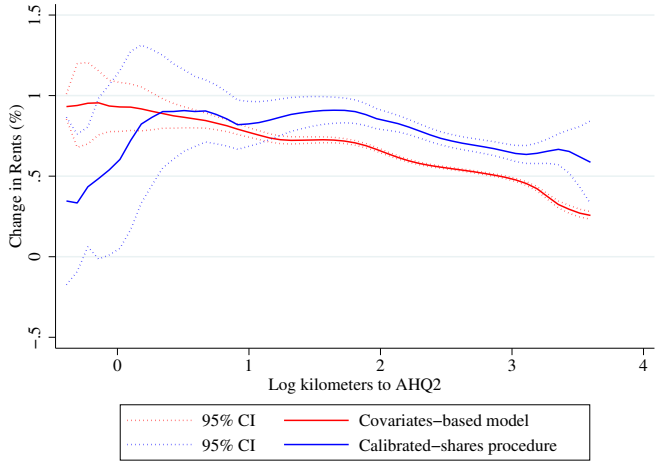
NOTES: Panel A depicts the number of workers employed in each tract in 2010. Panels B and C depict decreases in the number of workers predicted by the covariates-based model and calibrated-shares procedure, respectively. As total population is fixed, the number of workers decreases in all tracts except the Amazon HQ2 location. The covariates-based predictions for workers describe both the continuum model and the model with finite individuals because the change in the former equals the expected change in the latter.

FIGURE F.2.—Predicted changes in wages



NOTES: These maps depict percent changes in wages, $(\hat{w}_n/\hat{P} - 1) \times 100$. Panel A depicts the change in the covariates-based continuum model. Panel B depicts the mean change across 100 simulations of the model with finite individuals. Panel C depicts each tract's predicted wage change using the calibrated-shares procedure.

FIGURE F.3.—Rent changes by distance to Amazon HQ2



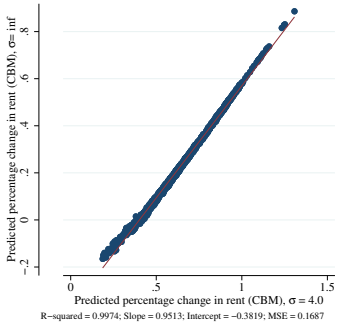
NOTES: This figure plots the changes in rents predicted by the covariates-based continuum model and the calibrated-shares procedure as functions of (log) distance to the Amazon HQ2 tract. The percent change in rents is $\left(\frac{\hat{r}_k}{\hat{P}} - 1\right) \times 100$.

As alternatives to $\sigma = 4$, we consider the values $\sigma = 1.1$ and $\sigma = \infty$. The predicted changes in commuters to the Amazon HQ2 tract from both the covariates-based model and the calibrated-shares procedure are invariant to the labor demand elasticity σ: regressing

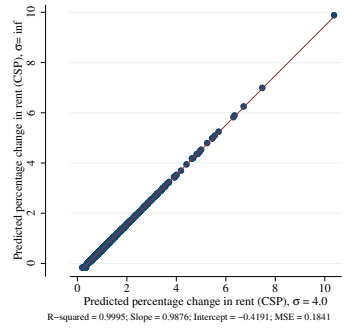
predictions for one value of σ on predictions from another yields a slope of one and an intercept of zero to the fifth or sixth decimal place. The predicted changes in real rents by residential tract from both the covariates-based model and the calibrated-shares procedure are almost perfectly correlated ($R^2 \geq 0.997$), but their scale changes with σ: a lower labor demand elasticity implies a higher average real rent increase (see Figure F.4 contrasting $\sigma = 4$ and $\sigma = \infty$). The predicted changes in wages by workplace tract necessarily vary with the labor demand elasticity: a lower labor demand elasticity implies a higher average wage increase and range of wage changes. When $\sigma = \infty$, there is no spatial variation in wage changes.

FIGURE F.4.—Comparison of predictions of the change in real rents for $\sigma = 4.0$ and $\sigma = \infty$

A. CBM comparison of real rents for $\sigma = 4.0$ and $\sigma = \infty$

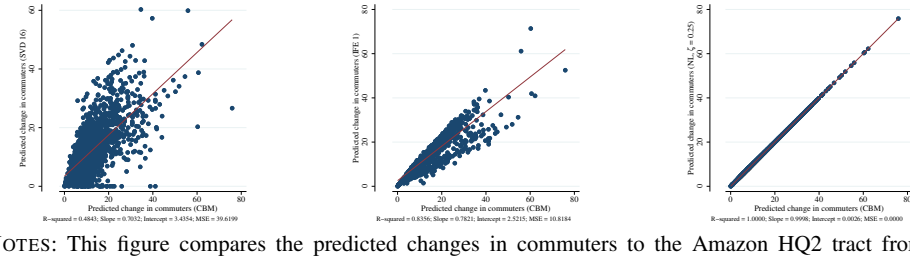


B. CSP comparison of real rents for $\sigma = 4.0$ and $\sigma = \infty$



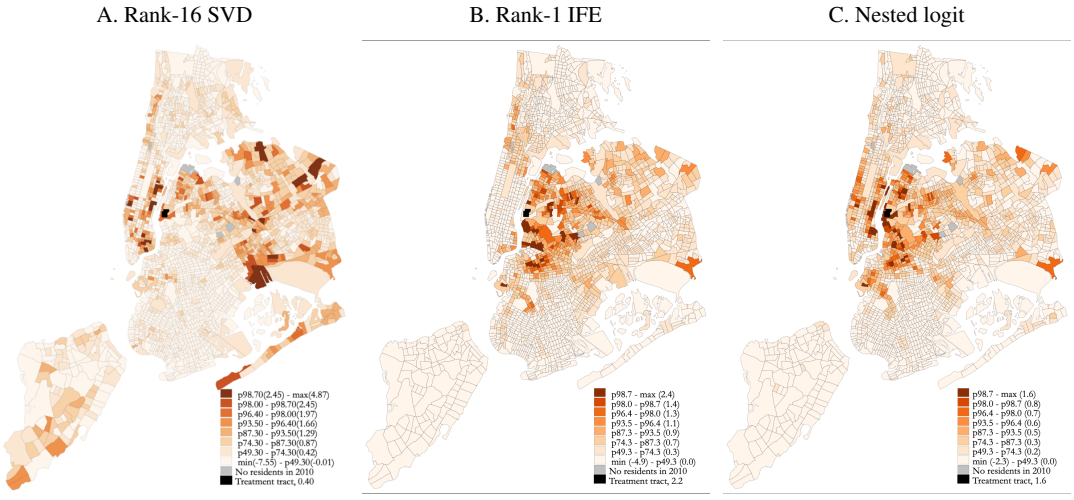
NOTES: These plots compare the percent change in real rents, $(\hat{r}/\hat{P} - 1) \times 100$, predicted by the covariates-based model (left panel) and calibrated-shares procedure (right panel) for the labor demand elasticities of $\sigma = 4.0$ and $\sigma = \infty$. In both panels, the predicted real rent changes associated with different values of σ are almost perfectly correlated and only differ by a constant.

FIGURE F.5.—Predicted changes in commuters to the AHQ2 tract from SVD, IFE, and nested-logit specifications



NOTES: This figure compares the predicted changes in commuters to the Amazon HQ2 tract from the approximated matrix derived from a rank-16 singular value decomposition (left panel), from the rank-1 interactive-fixed-effects specification (center panel), and from the nested-logit specification with $\zeta = 0.25$ (right panel) to the predictions from the covariates-based specification.

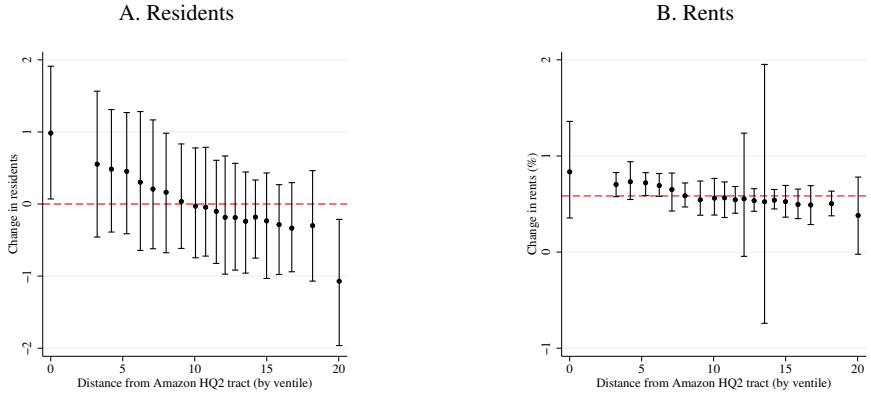
FIGURE F.6.—Predicted changes in residents from SVD, IFE, and nested-logit specifications



NOTES: These maps depict the changes in residents predicted by the approximated matrix derived from a rank-16 singular value decomposition (left panel), the rank-1 interactive-fixed-effects specification (center panel), and the nested-logit specification with $\zeta = 0.25$ (right panel). They should be compared to the predictions depicted in Figure 6B.

F.2. *Uncertainty about Amazon HQ2 predictions due to idiosyncrasies*

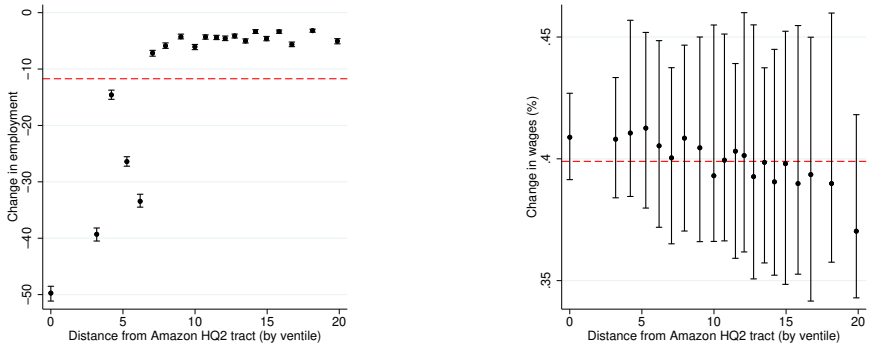
FIGURE F.7.—Changes in residents and rents by distance ventile



NOTES: This figure depicts the distribution of average tract-level outcomes by ventile of distance to the Amazon HQ2 tract. Each ventile contains about 108 residential tracts. Citywide average outcomes are shown with a dashed red line. Each point is the mean value of the ventile's outcomes across 100 simulations. The confidence intervals are constructed for each ventile from the 5th and 95th percentile outcomes of the 100 simulations. The percent change in real rents is defined as $\left(\frac{\hat{r}_k}{\bar{P}} - 1\right) \times 100$. Shorter-distance ventiles tend to have larger increases in both residents and rents on average. For all but the first and twentieth ventiles, the 90% confidence interval for the change in residents across 100 simulations includes zero. The rent increase in the first ventile in the mean simulation is about double that in the 20th ventile. The 90% confidence interval of every ventile includes 0.6%, the citywide average rent increase.

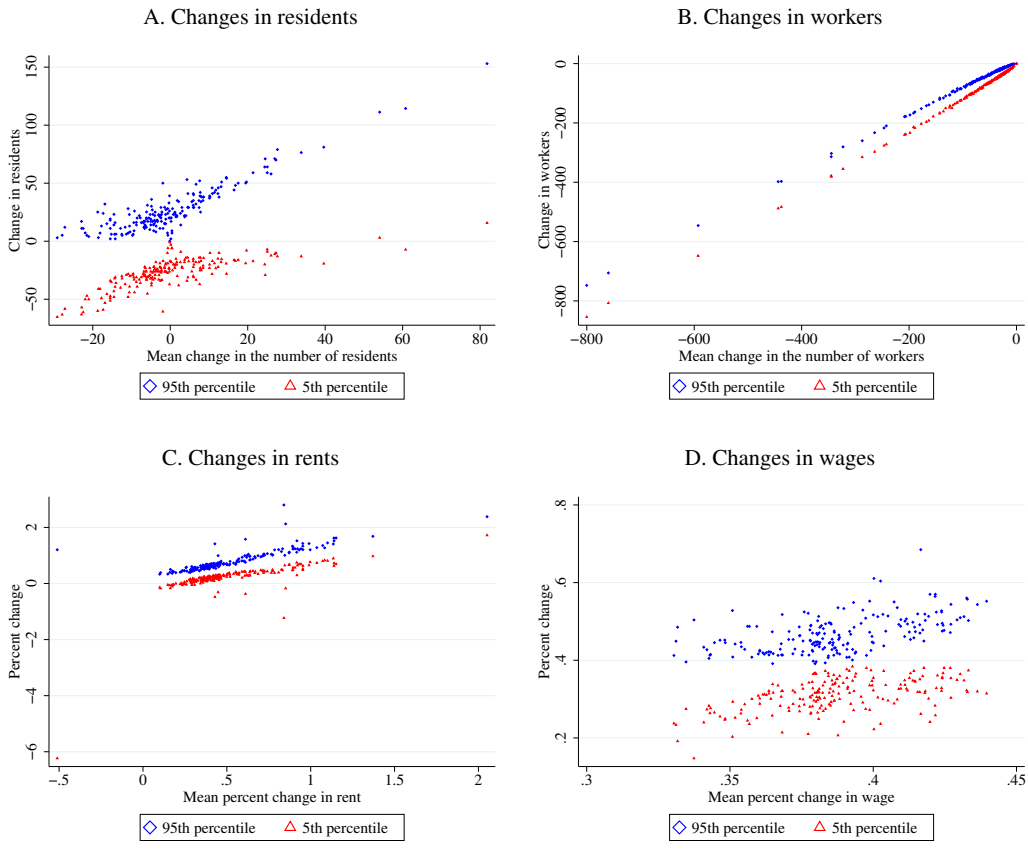
FIGURE F.8.—Changes in employment and wages by distance ventile

A. Employment B. Wages



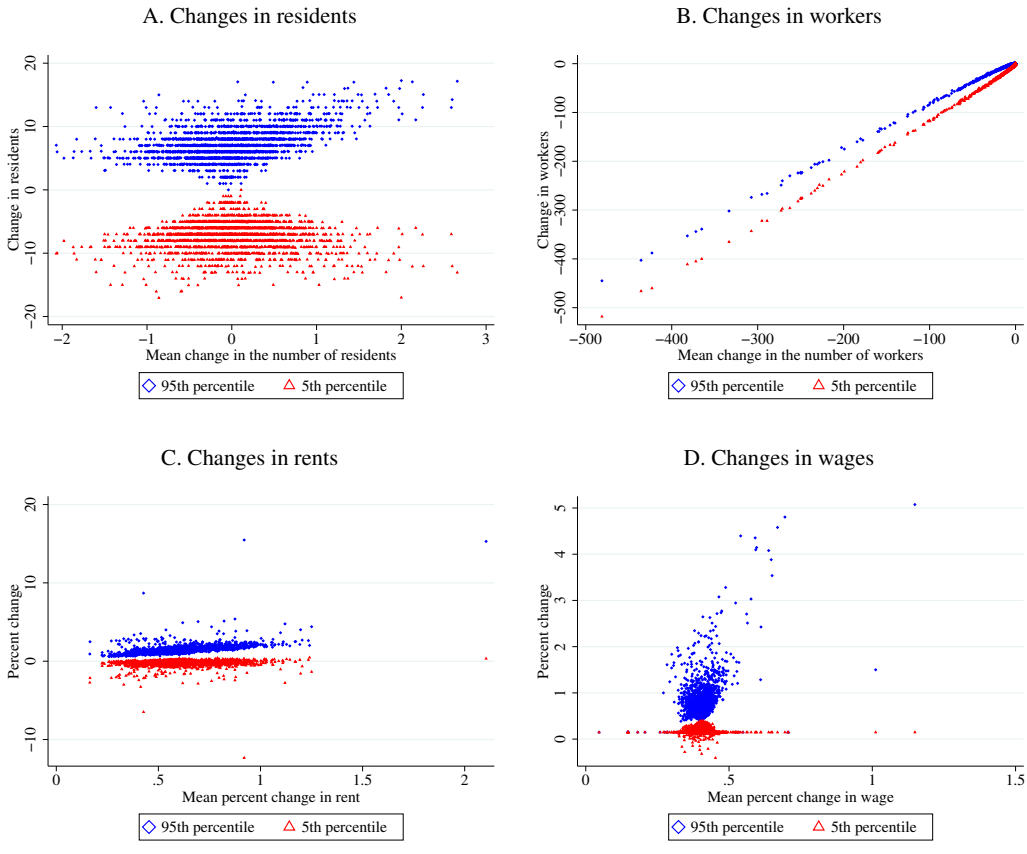
NOTES: This figure depicts the distribution of average tract-level outcomes by ventile of distance to the Amazon HQ2 tract. Each ventile contains about 107 workplace tracts. Citywide average outcomes are shown with a dashed red line. Each point is the mean value of the ventile's outcomes across 100 simulations. The confidence intervals are constructed for each ventile from the 5th and 95th percentile outcomes of the 100 simulations. The percent change in real wages is defined as $\left(\frac{\hat{w}_n}{\hat{P}} - 1\right) \times 100$.

FIGURE F.9.—Uncertainty about counterfactual changes induced by idiosyncrasies (NTAs)



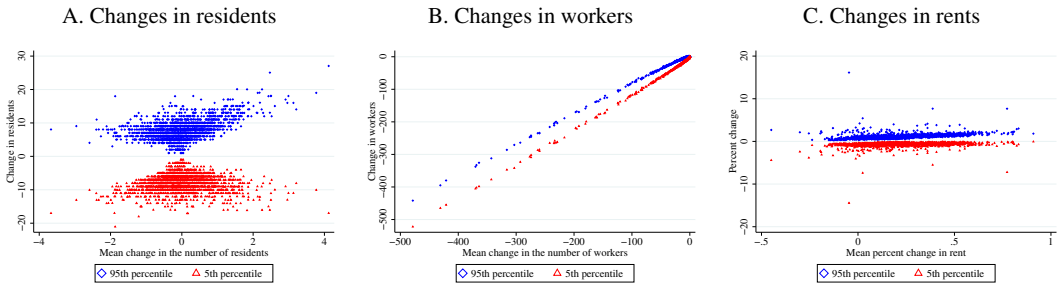
NOTES: These plots depict variation in counterfactual changes induced by individual idiosyncrasies in the NTA-level model with a finite number of individuals. The panel depicts the 5th and 95th percentiles of predicted changes in quantities and prices across 100 simulations of the NTA-level model with a finite number of individuals.

FIGURE F.10.—Uncertainty about counterfactual changes induced by idiosyncrasies (nested logit)



NOTES: These plots depict variation in counterfactual changes induced by individual idiosyncrasies in the nested-logit specification (with parameter $\zeta = 0.25$) of the model with a finite number of individuals. The panel depicts the 5$^{\text{th}}$ and 95$^{\text{th}}$ percentiles of predicted changes in quantities and prices across 100 simulations of the model with a finite number of individuals. To simulate draws from a nested-logit distribution $F(\mathring{\nu}^i)$, we first represent $\mathring{\nu}^i \sim F(\mathring{\nu}^i)$ using a standard Gumbel distribution $\mathcal{G}(0,1)$ and a positive stable distribution (PSD) $\mathcal{P}(\zeta)$, based on Galichon (2022): $\mathring{\nu}^i_{kn} = \zeta\left(\nu^i_{kn} + \log Z^i_{B_z}\right)$, where ν^i_{kn} follows $\mathcal{G}(0,1)$, and $Z^i_{B_z}$ follows $\mathcal{P}(\zeta)$. Second, we apply Ridout (2009)'s method to simulate $Z^i_{B_z} \sim \mathcal{P}(\zeta)$ for $\zeta = 0.25$. Finally, we combine the standard Gumbel draws and the PSD draws to construct the draws from the nested-logit distribution.

FIGURE F.11.—Uncertainty about counterfactual changes induced by individual idiosyncrasies ($\sigma = \infty$)



NOTES: The plots depict the 5th and 95th percentiles of predicted changes in quantities and prices across 100 simulations of the model with a finite number of individuals in which $\sigma = \infty$. In Panel A, the 90% confidence interval for the change in residents includes zero for all tracts. In Panel B, the 90% confidence interval for the change in employment includes zero for 490 of the 2142 non-Amazon workplaces. Panel B excludes the Amazon HQ2 workplace tract. The 90% confidence interval for the change in employment for the Amazon workplace tract is 24,728 to 25,259. In Panel C, real rent changes are $(\hat{r}_k / \hat{P} - 1) \times 100$, and 1 out of 2160 origin tracts have a positive change in rents at the 5th percentile.

REFERENCES

ALLEN, TREB AND COSTAS ARKOLAKIS (2014): "Trade and the Topography of the Spatial Economy," *The Quarterly Journal of Economics*, 129 (3), 1085–1140. [18]

ALLEN, TREB, COSTAS ARKOLAKIS, AND XIANGLIANG LI (2023): "On the Equilibrium Properties of Network Models with Heterogeneous Agents," Working Paper 27837, NBER. [1, 2, 3]

DAVIS, DONALD R., JONATHAN I. DINGEL, JOAN MONRAS, AND EDUARDO MORALES (2019): "How Segregated Is Urban Consumption?" *Journal of Political Economy*, 127 (4), 1684–1738. [30]

DINGEL, JONATHAN, ANTONIO MISCIO, AND DONALD DAVIS (2021): "Cities, Lights, and Skills in Developing Economies," *Journal of Urban Economics*, 125 (103174). [22]

FOSCHI, ANDREA, CHRISTOPHER L. HOUSE, CHRISTIAN PROEBSTING, AND LINDA L. TESAR (2023): "Labor Mobility and Unemployment over the Business Cycle," *AEA Papers and Proceedings*, 113, 590–96. [28]

GALICHON, ALFRED (2022): "On the representation of the nested logit model," *Econometric Theory*, 38 (2), 370–380. [43]

GRAHAM, MATTHEW R., MARK J. KUTZBACH, AND BRIAN MCKENZIE (2014): "Design Comparison of LODES and ACS Commuting Data Products," Working Papers 14-38, Center for Economic Studies, U.S. Census Bureau. [21]

KREBS, OLIVER AND MICHAEL PFLÜGER (2023): "On the road (again): Commuting and local employment elasticities in Germany," *Regional Science and Urban Economics*, 99, 103874. [22]

MONTE, FERDINANDO, STEPHEN J. REDDING, AND ESTEBAN ROSSI-HANSBERG (2018): "Commuting, Migration, and Local Employment Elasticities," *American Economic Review*, 108 (12), 3855–90. [22]

OWENS, RAYMOND, III, ESTEBAN ROSSI-HANSBERG, AND PIERRE-DANIEL SARTE (2020): "Rethinking Detroit," *American Economic Journal: Economic Policy*, 12 (2), 258–305. [7, 21, 29, 30, 31]

RIDOUT, MARTIN S (2009): "Generating random numbers from a distribution specified by its Laplace transform," *Statistics and Computing*, 19, 439–450. [43]

TRAIN, KENNETH (2009): *Discrete Choice Methods with Simulation*, Cambridge University Press. [13]