NBER WORKING PAPER SERIES

SPATIAL ECONOMICS FOR GRANULAR SETTINGS

Jonathan I. Dingel
Felix Tintelnot

Working Paper 27287
http://www.nber.org/papers/w27287

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2020, Revised June 2025

We thank four anonymous referees, Rodrigo Adão, Gabriel Ahlfeldt, Treb Allen, Stephane Bonhomme, Kirill Borusyak, Victor Couture, Teresa Fort, Lars Hansen, Elhanan Helpman, Thomas Holmes, John Huizinga, Erik Hurst, Kyle Mangum, Yuhei Miyauchi, Eduardo Morales, Stephen Redding, Jordan Rosenthal-Kay, Esteban Rossi- Hansberg, Chris Severen, Sebastian Sotelo, Daniel Sturm, Miguel Zerecero, Oren Ziv, and numerous conference and seminar participants for helpful feedback. We are grateful to Junbiao Chen, Levi Crews, Gianna Fenaroli, Reigner Kane, Xianglong Kong, Daniil Iurchenko, Leran Qi, John Ruf, Ye Sun, Linghui Wu, Shijian Yang, and Mingjie Zhu for excellent research assistance. Dingel thanks the James S. Kemper Foundation Faculty Research Fund at the University of Chicago Booth School of Business. This work was completed in part with resources provided by the University of Chicago Research Computing Center and Columbia University's Shared Research Computing Facility. We gratefully acknowledge support from the National Science Foundation under grant number 2018609. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2020 by Jonathan I. Dingel and Felix Tintelnot. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Spatial Economics for Granular Settings
Jonathan I. Dingel and Felix Tintelnot
NBER Working Paper No. 27287
May 2020, Revised June 2025
JEL No. C25, F16, R1, R13, R23

ABSTRACT

We examine the application of quantitative spatial models to the growing body of fine spatial data used to study local economic outcomes. In granular settings in which people choose from a large set of potential residence-workplace pairs, observed outcomes in part reflect idiosyncratic choices. Using analytical examples, Monte Carlo simulations, and event studies of neighborhood employment booms, we demonstrate that calibration procedures that equate observed shares and modeled probabilities perform very poorly in these high-dimensional settings. Parsimonious specifications of spatial linkages deliver better counterfactual predictions. To quantify the uncertainty about counterfactual outcomes induced by the idiosyncratic component of individuals' decisions, we introduce a quantitative spatial model with a finite number of individuals. Applying this model to Amazon's proposed second headquarters in New York City reveals that its predicted consequences for most neighborhoods vary substantially across realizations of the individual idiosyncrasies.

Jonathan I. Dingel
Columbia University
Department of Economics
and NBER
jid2106@columbia.edu

Felix Tintelnot
Duke University
Department of Economics
and NBER
felix.tintelnot@gmail.com

1 Introduction

Economists increasingly use quantitative models to evaluate urban policies such as infrastructure investments and land-use planning. The growing availability of economic data observed at fine spatial scales offers tremendous potential for new insights. However, to inform policy, these models need to produce credible counterfactual predictions when applied to granular settings.

Quantitative spatial models feature links between pairs of locations that shape the geography of economic outcomes (Redding and Rossi-Hansberg, 2017; Proost and Thisse, 2019). In many of these models, individuals choose a residence and a workplace based on locations' characteristics, the links between them, and an idiosyncratic component. A setting with granular geographic units has many locations and therefore a great number of links between pairs of locations. Two concerns arise when the number of links is large relative to the number of decision makers. First, if one uses a high-dimensional parameterization of spatial linkages, such as one parameter per pair of locations, the modeler runs the risk of fitting the model to the idiosyncratic components of decisions rather than capturing spatial linkages. Second, counterfactual changes in local economic outcomes are uncertain because they depend on individuals' idiosyncratic preferences.

In this paper, we examine the application of quantitative spatial models to granular settings. In doing so, we make three main contributions. First, we show that the calibration procedure most often used, which perfectly matches the observed baseline outcomes, predicts changes in outcomes poorly in granular settings. We show how to use simulations that require only the inputs typically used in this procedure to diagnose the problem. We confirm the poor predictive performance using event studies of neighborhood employment booms in New York City. Second, we demonstrate that both a parsimonious parameterization using only a transit-time covariate and parameterizations using matrix approximations to capture unobserved spatial linkages deliver better counterfactual predictions in this high-dimensional setting. Third, we develop a spatial model with a finite number of individuals that we use to quantify the variability of counterfactual outcomes induced by individual idiosyncrasies. Applying this model to Amazon's proposed second headquarters in New York City shows that counterfactual changes for most census tracts vary considerably.

Section 2 reviews the computation of counterfactual outcomes in quantitative spatial models. With a continuum of individuals, counterfactual changes can be characterized by a system of equations containing counterfactual endogenous outcomes relative to baseline endogenous outcomes, counterfactual exogenous parameters relative to baseline exogenous parameters, constant elasticities, and baseline equilibrium shares (a formulation known as "exact hat algebra"). When parameterizing the spatial links between residences and workplaces that determine the baseline equilibrium shares, researchers have used two starkly different approaches. The covariates-based approach makes the cost of commuting a function of pair-specific covariates, such as transit time. The calibrated-shares procedure calibrates a cost parameter for each pair of locations to match the observed commuting shares. The contrast in baseline equilibrium shares produces contrasting predictions about counterfactual outcomes.

Calibrating pair-specific commuting costs to match observed commuting shares offers advantages and disadvantages. The number of commuters between a pair of locations in a finite sample depends on three components: spatial links correlated with included covariates, orthogonal spatial links, and individual-specific idiosyncrasies. The potential advantage of calibrating pair-specific costs to match observed shares is that this can capture spatial links not predicted by covariates like transit times, such as many Columbia University employees living in nearby university-owned residences. The potential disadvantage is that parameterizing the model's spatial links by matching all observed variation means overfitting the model: the calibrated parameters conflate the spatial links common to all individuals and individual-specific idiosyncrasies in the observed data.

Section 3 examines the application of these models to granular empirical settings. The commuting matrices used in urban economics often have so many more pairs of locations than commuters that many of the cells must contain zero commuters. Small flows constitute a substantial share of total commuting: more than 40% of New York City commuters have a residence-workplace pairing populated by five or fewer individuals. Such settings can pose small-sample problems for the calibrated-shares procedure, which equates observed shares and underlying probabilities. This procedure's expected squared error for a baseline share is its finite-sample variance.

Errors in baseline shares generate errors in predicted changes in counterfactual scenarios. We illustrate this in a special case in which the error in the predicted change in commuters is proportional to the error in the corresponding baseline share. More generally, counterfactual changes are a non-linear function of all baseline shares. We use Monte Carlo simulations to examine the finite-sample behavior of the calibrated-shares procedure in a realistic setting. The data-generating process is the parsimonious covariates-based parameterization of New York City in 2010, which uses the same covariate employed to estimate the commuting elasticity for the calibrated-shares procedure. These simulations show that small-sample bias can seriously impair the calibrated-shares procedure when the model is otherwise correctly specified. While the procedure works well as the number of individuals becomes arbitrarily large, it produces much worse predictions in simulations using the actual number of individuals in New York City.

Next, we use event studies to examine how well the incumbent approaches predict changes in commuting flows, the key spatial linkage between neighborhoods in many models of cities. We study large employment increases in a single census tract that often stem from the arrival or expansion of a large employer. We increase the productivity parameters for these tracts to match their increases in total employment and compare the predicted changes in bilateral commuting flows to those observed in the data. When examining 83 tract-level employment booms in New York City, we find that the covariates-based approach——using transit times as the only bilateral covariate——predicts the observed changes in commuting flows better than the calibrated-shares procedure in 80 of these events. Regressing the observed change in commuters on the covariates-based approach's prediction typically yields a slope near one, whereas the predictions from the calibrated-shares procedure are *negatively* correlated with the observed changes in more than half of the events.

We explore two additional approaches that offer the advantage of a more flexible specification

while guarding against the disadvantage of overfitting. One specification augments the covariates-based approach to capture unobserved spatial linkages that have an interactive fixed effects structure (Bai, 2009; Chen, Fernández-Val, and Weidner, 2021). This more flexible specification modestly improves predictive performance. An alternative approach computes baseline shares used in the exact hat algebra from rank-restricted matrix approximations without using covariates. This simple, computationally cheap transformation delivers substantially better predictions than using the observed shares.[1] Both these approaches allow one to incorporate unobserved components when applying continuum models to granular settings.

The noisy-shares problem is relevant for a number of fields in which economists rely on a "shares inversion" to estimate or calibrate models. Market shares computed from consumer surveys, region-to-region trade shares computed from shipment surveys, and migration shares computed from household surveys that cover a fraction of the population can be noisy.[2] Shares computed using the universe of commuting flows, sectoral exports, firm-to-firm sales, or employee-employer matches are noisy when there are few decisions or transactions made among the many possibilities so that the observed shares contain a substantial idiosyncratic component.[3]

In Section 4, we depart from treating the data as a finite sample from the continuum model and introduce a model with a finite number of individuals. We use it to quantify the variability of counterfactual outcomes induced by individual idiosyncrasies. Where the continuum model yields a single equilibrium allocation, this model produces a *distribution* of economic outcomes. Notably, outcomes with strictly positive probabilities may not occur in a particular realization. The key modeling challenge is that individual decisions affect equilibrium wages and rents, but it is computationally infeasible to enumerate all the possible labor allocations and resulting prices. To overcome this combinatorial challenge, we assume that individuals make choices based on wage and rent beliefs that are the equilibrium prices from the continuum model with the same economic primitives.

Our model with finite individuals can be estimated using the same data used to estimate the continuum model. When individuals' price beliefs are the continuum-case equilibrium prices, the estimated parameters of this model coincide with those of the covariates-based continuum model. Given the same parameter values, the mean equilibrium labor allocation equals the allocation of the continuum model. By contrast, equilibrium prices solve a system of non-linear equations, so their mean values are not necessarily equal to the continuum-case prices. Crucially, the model with a finite number of individuals characterizes the dispersion in quantities and prices that arises from

[1] Smoothing the observed shares using rank-restricted matrix approximations before employing them in exact hat algebra is reminiscent of, yet distinct from, practices such as applying the Hodrick and Prescott (1997) filter before calibrating macroeconomic models, grouping firms into classes using k-means clustering before estimating a model of workers and firms (Bonhomme, Lamadon, and Manresa, 2019), and smoothing observations before inverting conditional choice probabilities in dynamic models (e.g., Kalouptsidi, 2014; Hsiao, 2022).

[2] The US Commodity Flow Survey, for example, samples about one in seven establishments in the industries it covers and asks them to report at most 40 of the shipments sent during each of four weeks within the year. See Appendix D.5 for the case of state-to-state migration shares.

[3] For example, Gaubert and Itskhoki (2021) report that sectoral trade flows dominated by a small number of firms exhibit greater mean reversion.

idiosyncrasies.

In Section 5, we contrast methods for evaluating the local economic effects of Amazon's proposed—but later abandoned—HQ2 in New York City. The tract-level changes in residents and land rents across New York City predicted by the covariates-based approach differ considerably from the predictions of the calibrated-shares procedure. We find considerable uncertainty generated by individual idiosyncrasies for these outcomes at the neighborhood level and meaningful uncertainty even for groups of tracts. The uncertainty about aggregate outcomes is negligible. The variation in local counterfactual outcomes due to individual idiosyncrasies in the model with a finite number of individuals is larger than that due to parameter uncertainty in the covariates-based continuum model. This sizable uncertainty regarding the local effects of a headline-grabbing potential employment boom suggests that detailed counterfactual predictions would be highly uncertain in many granular settings.

Our paper introduces a quantitative spatial model suitable for application to small spatial units, such as census tracts. Social scientists study neighborhoods to understand housing markets, intergenerational mobility, racial segregation, and many other phenomena. Spatially precise satellite imagery (Donaldson and Storeygard, 2016) and phone movement data (Couture et al., 2022; Kreindler and Miyauchi, 2023) allow even finer investigations where there are few decision makers, as anticipated by Holmes and Sieg (2015, p.106). Granular data are valuable in part because a growing body of evidence shows highly localized agglomeration economies (Rosenthal and Strange, 2020).[4] We provide tools for modeling these granular settings.

Our event studies of neighborhood employment booms contribute to a small literature assessing the predictive power of quantitative spatial models. These studies compare model-predicted changes to observed outcomes using plausibly exogenous shocks (Ahlfeldt et al., 2015; Monte, Redding, and Rossi-Hansberg, 2018; Adão, Arkolakis, and Esposito, 2022). More broadly, economists have evaluated the predictions of quantitative trade models about changes in trade policy (Kehoe, 2005; Kehoe, Pujolàs, and Rossbach, 2017; Adão, Costinot, and Donaldson, 2024). We show that predictions about the "ripple effects" of a local economic shock on other locations perform poorly when economic linkages are calibrated to match noisy shares. In granular settings, this noise stems from the idiosyncratic components of individual choices.

Finally, our work relates to prior studies of how idiosyncrasies affect economic outcomes and parameter estimation. A growing literature examines how firm-specific shocks influence aggregate fluctuations in macroeconomics (Gabaix, 2011; Carvalho and Grassi, 2019) and trade (di Giovanni, Levchenko, and Mejean, 2014; Gaubert and Itskhoki, 2021). Our paper joins research examining how individual idiosyncrasies affect economists' inferences. Ellison and Glaeser (1997) and Schoefer and Ziv (2022) address the role of individual manufacturing plants when computing the geographic concentration of industries and geographic variation in productivity, respectively. Mogstad et al.

[4] Arzaghi and Henderson (2008) estimate that productivity gains from interactions among advertising firms occur primarily within 500 meters. Ahlfeldt et al. (2015) estimate production and residential externalities that decay by more than half within two minutes of travel time. Rossi-Hansberg, Sarte, and Owens (2010) estimate that externalities from non-market interactions decline by half approximately every 1,000 feet.

(2024) raise concerns about inferring ranks of small geographic units. The finite numbers of firms and shipments have been studied as one explanation for zeros in international trade (Eaton, Kortum, and Sotelo, 2013; Armenter and Koren, 2014), and the finite number of consumers is one explanation for zeros in product-level sales data (Quan and Williams, 2018; Hortaçsu et al., 2023; Gandhi, Lu, and Shi, 2023). Similar to our framework, Panigrahi (2022) models a finite set of firms that form links based on beliefs from a limiting economy with a continuum of firms. Our paper makes two contributions to these various strands of the literature. First, we demonstrate the importance of avoiding overfitting when using continuum models to predict counterfactual outcomes. Second, we introduce a spatial model with a finite number of individuals that quantifies the uncertainty about counterfactual outcomes induced by individual idiosyncrasies.

2 Computing counterfactual outcomes in continuum models

This section describes the computation of counterfactual outcomes in quantitative spatial models. We use a baseline model with a continuum of individuals making residential and workplace location choices, as in Ahlfeldt et al. (2015) and Redding and Rossi-Hansberg (2017).

2.1 Primitives

We consider a closed economy populated by a measure of L individuals who each supply one unit of labor. Discrete locations are indexed by k or n. Each location has a fixed quantity of land, T_k, owned by immobile landlords who consume only goods.[5] Each location has the technology to produce a differentiated good (i.e., the Armington assumption). These goods are freely traded.

Individuals have Cobb-Douglas preferences over goods and land, devoting α of their expenditure to the latter. They have constant elasticity of substitution (CES) preferences over the differentiated goods, with elasticity of substitution $\sigma > 1$. Individuals (indexed by i) have idiosyncratic preferences for pairs of residential and workplace locations, such that i's indirect utility from residing in k and working in n is

$$U_{kn}^i = \underbrace{\epsilon \ln\left(\frac{w_n}{r_k^\alpha P^{1-\alpha}\delta_{kn}}\right)}_{\equiv U_{kn}} + \nu_{kn}^i, \tag{1}$$

where w_n is the wage in location n, r_k is the land rent in k, P is the common CES price index for goods, δ_{kn} is the commuting cost between k and n, U_{kn} is the mean utility of choice kn, and ν_{kn}^i is the idiosyncratic preference of individual i for residing in k and working in n. The commuting elasticity ϵ governs the importance of mean utility relative to the idiosyncratic preference, which is drawn from the standard Gumbel distribution.[6]

Commuting costs have two components: time ($\bar{\delta}_{kn}$) and disutility (λ_{kn}), so $\delta_{kn} = \bar{\delta}_{kn} \times \lambda_{kn}$. Time spent commuting is not spent working, so individuals residing in k and working in n earn

[5] This simplifying assumption follows Monte, Redding, and Rossi-Hansberg (2018).

[6] The cumulative distribution function of the Gumbel (type-1 extreme value) distribution with a location parameter of zero and scale parameter of one is $F(\nu_{kn}^i) = \exp\left(-\exp\left(-\nu_{kn}^i\right)\right)$.

only $w_n/\bar{\delta}_{kn}$ because they only spend $1/\bar{\delta}_{kn}$ of their time working.

Production of each location's differentiated good is linear in labor. Output in n is $q_n = A_n L_n$, where A_n is that location's productivity and L_n is the quantity of labor supplied by workers working in n. The goods market is competitive, so the price of location n's output is w_n/A_n. Thus, the CES price index is $P = \left[\sum_n (w_n/A_n)^{1-\sigma}\right]^{1/(1-\sigma)}$.

2.2 Equilibrium

Let ℓ_{kn} denote the measure of people residing in k and working in n. Using the Gumbel distribution assumption, one can integrate over the idiosyncratic preferences to obtain the fraction of people residing in k and working in n:

$$\frac{\ell_{kn}}{L} = \frac{w_n^{\epsilon} \left(r_k^{\alpha} \delta_{kn}\right)^{-\epsilon}}{\sum_{k',n'} w_{n'}^{\epsilon} \left(r_{k'}^{\alpha} \delta_{k'n'}\right)^{-\epsilon}}. \tag{2}$$

Goods market clearing equates each location's output to the quantity demanded. Labor supplied in location n is $L_n = \sum_k \ell_{kn}/\bar{\delta}_{kn}$, and thus output there is $A_n \sum_k \ell_{kn}/\bar{\delta}_{kn}$. 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. Aggregate income is $Y \equiv \sum_{k,n} y_{kn}$, where $y_{kn} \equiv w_n \ell_{kn}/\bar{\delta}_{kn}$ is labor income earned in workplace n by workers residing in k. The CES demand for each differentiated good means that equating quantity supplied and quantity demanded requires

$$A_n \sum_k \frac{\ell_{kn}}{\bar{\delta}_{kn}} = \frac{(w_n/A_n)^{-\sigma}}{P^{1-\sigma}} Y \qquad \forall n. \tag{3}$$

Note that goods market clearing implies labor market clearing.

Similarly, land market clearing equates the fixed land endowment T_k to the quantity demanded by individuals, who devote a constant fraction α of their expenditure to land:

$$T_k = \frac{\alpha}{r_k} \sum_n y_{kn} \qquad \forall k. \tag{4}$$

Equilibrium is a set of wages $\{w_n\}$, rents $\{r_k\}$, and labor allocation $\{\ell_{kn}\}$ such that equations (2), (3), and (4) hold. Appendix C.1 shows that if $\left(\frac{1+\epsilon}{\sigma+\epsilon}\right)\left(\frac{\alpha\epsilon}{1+\alpha\epsilon}\right) \leq \frac{1}{2}$, this equilibrium exists and is unique (by Theorem 1 of Allen, Arkolakis, and Li 2023).

2.3 Procedures for counterfactual predictions

In this model, counterfactual changes can be characterized by a system of equations containing counterfactual endogenous outcomes relative to baseline endogenous outcomes, counterfactual exogenous parameters relative to baseline exogenous parameters, constant elasticities, and baseline equilibrium shares. This is known as "exact hat algebra" in the trade literature (Dekle, Eaton, and

Kortum, 2008; Costinot and Rodríguez-Clare, 2014). Denote the counterfactual value of a variable x by x' and denote the counterfactual-baseline ratio of a variable $x > 0$ by $\hat{x} \equiv \frac{x'}{x}$ (and impose $\hat{x} = 0$ if $x = 0$). As shown in Appendix C.2, tedious manipulation of equations (2), (3), and (4) yields the following system of equations characterizing the counterfactual equilibrium:

$$\hat{w}_n = \hat{A}_n \left(\sum_k \hat{y}_{kn} \frac{y_{kn}}{\sum_{k'} y_{k'n}} \right)^{\frac{1}{1-\sigma}} \hat{P} \hat{Y}^{\frac{1}{\sigma-1}} \tag{5}$$

$$\hat{r}_k = \hat{T}_k^{-1} \sum_n \hat{y}_{kn} \frac{y_{kn}}{\sum_{n'} y_{kn'}} \tag{6}$$

$$\hat{\ell}_{kn} = \frac{\hat{w}_n^{\epsilon} \left(\hat{r}_k^{\alpha} \hat{\bar{\delta}}_{kn} \hat{\lambda}_{kn} \right)^{-\epsilon}}{\sum_{k',n'} \hat{w}_{n'}^{\epsilon} \left(\hat{r}_{k'}^{\alpha} \hat{\bar{\delta}}_{k'n'} \hat{\lambda}_{k'n'} \right)^{-\epsilon} \frac{\ell_{k'n'}}{L}} \qquad \text{if } \ell_{kn} > 0. \tag{7}$$

This system defines the relative endogenous outcomes $\hat{w}_n$, $\hat{r}_k$, and $\hat{\ell}_{kn}$ in terms of relative exogenous parameters $\hat{A}_n$, $\hat{T}_k$, $\hat{\bar{\delta}}_{kn}$, and $\hat{\lambda}_{kn}$, elasticities σ, α, and ϵ, and baseline shares $\frac{\ell_{kn}}{L}$ and $\frac{y_{kn}}{Y}$.

There are many ways to estimate or calibrate the model's parameters to deliver the baseline shares. At one end of the spectrum, one can parsimoniously parameterize the time and disutility components of commuting costs as functions of only observed covariates (e.g., Ahlfeldt et al. 2015).[7] To compute counterfactual equilibria, one would plug the *fitted* model's values of $\frac{\ell_{kn}}{L}$ and $\frac{y_{kn}}{Y}$ into equations (5)–(7).[8] At the other end of the spectrum, one can calibrate millions of commuting-cost parameters so that the model's baseline equilibrium matches the observed shares. In this case, one would plug the *observed* $\frac{\ell_{kn}}{L}$ and $\frac{y_{kn}}{Y}$ into equations (5)–(7) to compute counterfactual outcomes. The latter calibrated-shares procedure has been widely used in spatial economics recently, often in granular settings with as many residence-workplace pairs as decision makers.[9] Between these two extremes lie a host of strategies to determine the baseline equilibrium shares, some of which we discuss in detail below.

We emphasize the distinction between using the comparative statics defined by equations (5)–(7) to compute counterfactual outcomes and fitting the model's parameters. Because equations (5)–(7) show that computing counterfactual outcomes only requires knowing the model's parameters up to the point where the model delivers the shares $\frac{\ell_{kn}}{L}$ and $\frac{y_{kn}}{Y}$, others have used the phrase "exact hat algebra" to refer to both rewriting the equations in hats and calibrating combinations of model

[7] In this approach, equation (2) is generically not satisfied by the observed values of ℓ_{kn} and the parameterization of δ_{kn}. One justification for the model not fitting the observed data would be that the latter is a sample from the population. Section 3.2 describes estimation of model parameters in this covariates-based approach.

[8] Given a parameterized model, one could also solve the system of equations (2)–(4) at counterfactual parameter values. However, the system of equations (5)–(7) is instructive because it indicates that any parameter combination yielding the same baseline shares and elasticities would make the same counterfactual predictions. In that sense, these are sufficient statistics for the counterfactual outcomes. Appendix Section C.2.1 provides an example.

[9] The calibrated-shares procedure has decades-deep roots in computable general equilibrium models of international trade (Rutherford, 1995). Studies in spatial economics employing this technique include Allen, Arkolakis, and Li (2016); Perez-Cervantes (2016); Waddell and Sarte (2016); Monte, Redding, and Rossi-Hansberg (2018); Heblich, Redding, and Sturm (2020), Krebs and Pflüger (2019); Severen (2021); and Owens, Rossi-Hansberg, and Sarte (2020).

parameters to rationalize observed shares. In fact, the system of equations defines counterfactual outcomes regardless of how one estimates or calibrates the parameters of the baseline equilibrium. The key question is how to fit the model's parameters to data.

3 Counterfactual analysis in granular settings

We assess the predictive performance of quantitative spatial models in granular empirical settings. In Section 3.1, we introduce data on tract-to-tract commuting in New York City. This empirical setting, like many studied in prior research, has many location pairs relative to the number of decision makers. After Section 3.2 describes how we estimate model parameters, Section 3.3 contrasts the nature of the estimation errors made by the covariates-based and calibrated-shares procedures. The covariates-based specification omits unobserved commuting costs, whereas the calibrated-shares procedure includes finite-sample noise in parameter values. We examine this trade-off when modeling New York City using Monte Carlo simulations in Section 3.4 and event studies of neighborhood employment booms in Section 3.5. There is a stark contrast between the two incumbent approaches in predicting the changes in tract-to-tract commuting flows. While the covariates-based specification using transit times predicts quite well, the calibrated-shares procedure's predicted changes are in fact negatively correlated with the observed changes in the majority of events. We also propose and assess alternative specifications that can capture unobserved components of spatial linkages while guarding against overfitting.

3.1 Commuting flows in granular settings

Urban economists often study empirical settings with a modest number of decision makers relative to the number of pairs of locations. It is common for these numbers to be roughly equal.[10] At even finer resolution, Ahlfeldt et al. (2015) model about 3 million Berliners choosing among about 254 million pairs of city blocks. Here, we examine tract-to-tract commuting flows in New York City. Appendix D.1 reports comparable statistics for Detroit and Minneapolis-St. Paul.

We use data on commuting between census tracts in New York City taken from the Longitudinal Employer-Household Dynamics, Origin-Destination Employment Statistics (LODES).[11] Tracts are defined by the US Census Bureau to be relatively uniform in population size, such that the typical tract has 4,000 residents. One implication of granularity is that some data providers perturb observations to protect confidentiality. For example, the published LODES tract-level workplace employment counts are infused with noise. Moreover, the published LODES commuting matrices

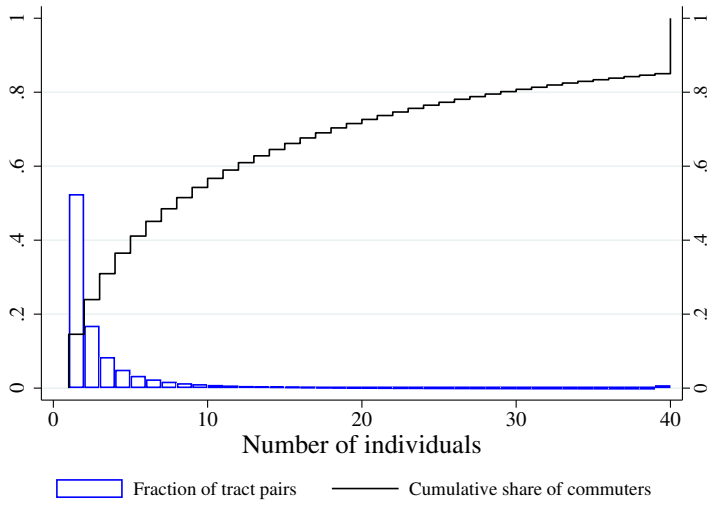
[10] Owens, Rossi-Hansberg, and Sarte (2020) study 1.3 million people commuting between 1.3 million pairs of tracts in the Detroit urban area. Severen (2021) studies the Los Angeles metropolitan area, which has 6.7 million commuters and more than 6 million pairs of tracts. Tsivanidis (2023) models Bogotá's 8 million residents commuting between almost 7 million pairs of tracts. Zárate (2023) examines Mexico City, which has 9 million people and 13 million pairs of tracts.

[11] Davis et al. (2019) use the 2010 wave to compute the joint distribution of residences and workplaces for New York City. Owens, Rossi-Hansberg, and Sarte (2020) use the 2014 wave of these data for the Detroit urban area to estimate a gravity model of commuting.

report synthetically generated residence-workplace pairs.[12] These confidentiality-protecting procedures are both a symptom of the granular setting and an additional cause for caution when interpreting the value of any individual observation.

Small flows constitute a large share of tract-to-tract commuting in New York City. The city has about 2.5 million resident-employees and 4.6 million tract pairs. Thus, its commuting matrix is almost necessarily sparse, as the average cell contains about one-half of a commuter. In fact, 85% of tract pairs have zero commuters between them, so the average positive cell has about three people. As shown in Figure 1, more than half of the tract pairs with a positive number of commuters have only one. Crucially, Figure 1 shows that small flows are a substantial share of the total: 41.1% of New York City commuters have five or fewer commuters in their cell of the commuting matrix. Furthermore, the commuting flows are impersistent over time and exhibit asymmetric zeros (see Appendices D.3 and D.4). Introducing multiple worker types would exacerbate these patterns by reducing the number of decision makers per choice for each type.[13]

Figure 1: Number of commuters between pairs of tracts in New York City



NOTES: This figure describes tract pairs in New York City by the number of individuals who reside in the origin tract and work in the destination tract in 2010 LODES data. The sample is restricted to the 15% of tract pairs that have a strictly positive number of commuters. The blue histogram depicts the fraction of tract pairs reporting each number of commuters, for each value from 1 to 39 and values of 40 or greater. The black increasing step function depicts the cumulative share of commuters by the number of commuters between the pair of tracts. New York City has 2,160 residential census tracts.

Even when studying larger geographic units, there may be few decision makers per pair of locations, particularly when the data describe only a sample of respondents. For example, while

[12] Graham, Kutzbach, and McKenzie (2014): "For each job in a workplace cell, LODES draws from a Dirichlet multinomial posterior distribution of possible residential locations. ... The prior adds uncertainty, so that even commutes with few or no observed flows may appear to have a job. Conversely, even when there are commuters from an origin in the likelihood, that residence may not be drawn and thus would not appear in LODES."

[13] For example, individuals in each of the four educational attainment categories reported in the LODES data live in all but one or two residential tracts and are employed in more than 99% of workplace tracts, so each type would have a choice set with more than 4.5 million elements.

some US counties have millions of residents, Appendix D.2 documents that more than half of the positive commuting counts between pairs of counties represent the behavior of five or fewer survey respondents. More broadly, small-sample bias may complicate the analysis of other spatial phenomena, such as migration. Appendix D.5 shows that state-to-state migration flows reported in American Community Survey (ACS) data also represent a small number of respondents, are frequently zero, and are impersistent across time (see Foschi et al. (2023) for related findings).

3.2 Estimation

We estimate the continuum model using the tract-to-tract commuting data for New York City in 2010. Following the literature, we set $\alpha = 0.24$ and $\sigma = 4$ prior to estimating the remaining parameters.[14]

We assume that the time component of commuting costs, $\bar{\delta}_{kn}$, is observed, while the disutility component, λ_{kn}, is unobserved. We construct the observed part of the commuting costs by assuming that each worker has H hours that are spent either working or commuting. We compute $\bar{\delta}_{kn} = \frac{H}{H - t_{kn} - t_{nk}}$, where $H = 9$ hours and t_{kn} is the transit time from k to n according to Google Maps.[15] Assumptions about the unobserved disutility component, λ_{kn}, vary widely across methods, as described below.

One way to estimate the continuum model using the covariates-based approach is to interpret observed data on residence-workplace choices as a finite sample from the continuum model. The covariates-based specification assumes there is no unobserved disutility component of commuting costs, $\lambda_{kn} = 1 \; \forall k, n$. Given observed values of commuting costs $\delta_{kn} = \bar{\delta}_{kn}$, the remaining parameters can be estimated by maximum likelihood. Interpreting the right side of equation (2) as a probability and denoting the number of individuals who chose the kn pair by ℓ_{kn}, the log likelihood function is

$$\mathcal{L} \equiv \sum_{k,n} \ell_{kn} \ln \left[\mathbb{P}(U^i_{kn} > U^i_{k'n'} \; \forall k'n' \neq kn) \right] = \sum_{k,n} \ell_{kn} \ln \left[\frac{w_n^{\epsilon} \left(r_k^{\alpha} \bar{\delta}_{kn} \right)^{-\epsilon}}{\sum_{k',n'} w_{n'}^{\epsilon} \left(r_{k'}^{\alpha} \bar{\delta}_{k'n'} \right)^{-\epsilon}} \right]. \tag{8}$$

This is the canonical conditional-logit likelihood of McFadden (1974) applied to location choices, as in McFadden (1978). The k- and n-specific terms are captured by residence and workplace fixed effects, respectively. Maximizing this likelihood function is numerically equivalent to a Poisson pseudo maximum likelihood (PPML) estimator that is available for a variety of software packages (Guimarães, Figueiredo, and Woodward, 2003).[16]

[14] Davis and Ortalo-Magne (2011) suggest $\alpha = 0.24$. Monte, Redding, and Rossi-Hansberg (2018) use $\sigma = 4$ when examining goods trade between US counties, citing estimates for international trade between countries. Absent estimates of tract-level labor demand elasticities, we also report results for $\sigma = 1.1$ and $\sigma = \infty$. None of our main conclusions are sensitive to the value of σ.

[15] Given H hours, $1/\bar{\delta}_{kn}$ is the share of that time spent working if the individual resides in k and works in n. The semi-elasticity of commuting flows with respect to transit time $\mathrm{t}_{kn} = t_{kn} + t_{nk}$ is $-\epsilon / (H - \mathrm{t}_{kn})$. Using $H = 8$ or $H = 10$ yields very little change in the model fit relative to the $H = 9$ results reported in Table 1. For New York City, we use Google Maps public-transit times collected by Davis et al. (2019). We impute missing observations for fewer than 4.1% of tract pairs in New York City by predicting transit times using physical distance.

[16] See Sotelo (2019) for a discussion of the relationship between the multinomial and PPML estimators in the

The calibrated-shares procedure, which calibrates λ_{kn}, also requires an estimate of the commuting elasticity ϵ. One might make two distinct assumptions about the unobserved component λ_{kn}. If $\mathbb{E}\left(\lambda_{kn}^{-\epsilon}|r_k, w_n, \bar{\delta}_{kn}\right) = 1$, the constant elasticity function (2) can be estimated using the PPML estimator (Silva and Tenreyro, 2006). This yields the same estimate of the commuting elasticity ϵ as maximizing equation (8). If one takes the logarithm of each side of equation (2), assumes $\mathbb{E}\left(\ln \lambda_{kn}|r_k, w_n, \bar{\delta}_{kn}\right) = 0$, and restricts the estimation sample to observations for which ℓ_{kn} is strictly positive, then the commuting elasticity can be estimated by ordinary least squares (OLS). As stressed by Silva and Tenreyro (2006), these two estimators can yield very different parameter estimates.

We apply both the PPML and OLS estimators to New York City in 2010. Our estimate of the commuting cost elasticity is presented in column 1 of Table 1. The estimate of $\epsilon \approx 8$, which is comparable to the value of 6.8 estimated by Ahlfeldt et al. (2015) for commuting within Berlin, implies that idiosyncratic preferences are modestly dispersed. Estimating the commuting elasticity using OLS yields a much lower elasticity, largely because of the well-understood selection bias associated with omitting four-fifths of observations from the estimation sample.[17] In what follows, we use the ϵ estimate reported in column 1 for both the covariates-based approach and the calibrated-shares procedure.

Table 1: Commuting elasticity estimates

	PPML/MLE	OLS
Commuting cost	-7.986	-2.307
	(0.307)	(0.00768)
Model fit (R^2 or pseudo-R^2)	0.662	0.561
Location pairs	4,628,878	690,673
Commuters	2,488,905	2,488,905

NOTES: All specifications include residence fixed effects and workplace fixed effects. The "PPML/MLE" column presents the results from maximum likelihood estimation of equation (8). The "OLS" column presents the results of estimating the log version of equation (2) by ordinary least squares, omitting observations in which $\ell_{kn} = 0$. The model-fit statistic is the pseudo-R^2 for MLE and R^2 for OLS. We report the PPML standard errors (clustered by k and by n), which are larger than the logit MLE standard errors associated with maximizing equation (8).

Given the elasticities σ, α, and ϵ, one only needs the model-implied baseline shares $\frac{\ell_{kn}}{L}$ and $\frac{y_{kn}}{Y}$ to compute counterfactual outcomes using equations (5)–(7). The calibrated-shares procedure sets the parameters $\{A_n\}$, $\{T_k\}$, and $\{\lambda_{kn}\}$ such that the model-implied baseline shares $\frac{\ell_{kn}}{L}$ and $\frac{y_{kn}}{Y}$ equal their observed values. For the covariates-based approach, the model parameters $\{A_n\}$ and $\{T_k\}$ can be computed from the estimated fixed effects and the market-clearing conditions. In particular, the residence and workplace fixed effects are proportional to $r_k^{-\alpha\epsilon}$ and w_n^{ϵ}, respectively.

context of gravity models. In practice, we use the Stata package `ppmlhdfe` of Correia, Guimarães, and Zylkin (2020).

[17] Applying the PPML estimator to the OLS sample of strictly positive flows yields an estimate of $\epsilon = 4.4$. Thus about two-thirds of the discrepancy between the OLS and MLE estimates reflects the omission of the zero flows. This selection bias also affects the fixed effects (see Appendix D.6).

Given α, σ, ϵ, $\{\delta_{kn}\}$, $\{r_k\}$, and $\{w_n\}$, equations (2), (3), and (4) can be solved to obtain $\{T_k\}$ and $\{A_n\}$, and one can compute the model-implied baseline shares $\frac{\ell_{kn}}{L}$ and $\frac{y_{kn}}{Y}$.

3.3 Errors in estimated baseline shares and prediction errors

In practice, each estimation method will imperfectly recover the data-generating process. Estimation error is important because errors in baseline shares translate into errors in counterfactual predictions through their roles in equations (5)–(7). For example, consider an economy in which $\sigma = \infty$ and $\alpha = 0$ with exogenous residential amenities that has a productivity change at one workplace. In this special case, the error in the predicted change in commuters $\Delta\ell_{kn}$ is simply proportional to the error in the baseline share ℓ_{kn}/L (see Appendix C.4).

The two approaches' estimation errors stem from different sources, so researchers face a trade-off. If there are unobserved links ($\lambda_{kn} \neq 1$), the covariates-based model is misspecified. If observed shares ℓ_{kn}/L are noisy, the calibrated-shares procedure will yield noisy estimates of λ_{kn}.[18] Thus, the trade-off depends on the magnitude of unobserved commuting costs and the number of decision makers per pair of locations. To justify applying the calibrated-shares procedure to a granular setting, the researcher must believe unobserved spatial links play a large role.

The potential advantage of the calibrated-shares procedure is its ability to capture links that covariates miss. For example, the largest commuting flow in the 2010 LODES data for New York City is 827 commuters who reside between 110$^{\text{th}}$ and 114$^{\text{th}}$ Streets in Morningside Heights and work at adjacent Columbia University. The specification that uses transit times predicts only 70 of the 827 observed commuters, failing to capture the effect of the university's dual role as employer and landlord.

The disadvantage of this approach is the sheer number of parameters to be estimated. If the economy has N locations, the calibrated-shares procedure estimates N^2 commuting-cost parameters, whereas the covariates-based specification only estimates $2N$ fixed effects (and the coefficients on the covariates).[19] In New York City, $2N \ll I \ll N^2$ (because I is 2.5 million and N^2 is 4.6 million), so the calibrated-shares procedure will produce estimates incorporating substantial finite-sample noise. The parsimonious covariates-based specification exhibits much less sampling error.[20]

We examine the predictions made by these two approaches using two metrics. One way to assess predictions is to regress the changes in outcomes on the predicted changes. An unbiased forecast would yield a slope coefficient of one. In an otherwise correctly specified model, finite-sample noise

[18] We focus on the finite-sample noise stemming from a granular setting. Sanders (2024) considers how other forms of measurement error affect counterfactual outcomes in quantitative spatial models.

[19] While we focus on finite-sample properties, this difference in dimensionality has implications for asymptotic properties like consistency as both the number of individuals and the number of locations grow. Suppose the number of locations N is an exponential function of the number of individuals I: $N \propto I^a$. The ratio of individuals to parameters grows at rate $1-a$ for the covariates-based approach and $1-2a$ for the calibrated-shares procedure. For $a \in (0.5, 1)$, $\frac{I}{N}$ goes to infinity and $\frac{I}{N^2}$ goes to zero. In the United States, the Census Bureau aims for Census tracts to have constant size of about 4,000 residents, so a is about one.

[20] Appendix C.6 presents a simple example for the NYC setting in which sampling errors differ by orders of magnitude.

will typically attenuate the slope coefficient on the calibrated-shares procedure's predicted changes towards zero.[21] The covariates-based specification's omission of the unobserved commuting cost does not necessarily attenuate its slope coefficient because the omitted residual is orthogonal to the covariates by construction. Appendix C.5 provides analytical expressions for the slope coefficient in the special case mentioned above in which predicted changes are proportional to baseline shares.

A second way to assess predictions is to compute their mean squared errors. In the special case, the expected squared error for a predicted change in commuters is proportional to the expected squared error for the corresponding baseline share. The calibrated-shares procedure's expected squared error for a baseline share is its finite-sample variance, which vanishes as I goes to infinity. By imposing more structure, a covariates-based approach typically has lower finite-sample variance but omits the unobserved commuting cost that contributes to its expected squared error for a baseline share. Their relative performance depends on how many commuters are observed per residence-workplace pair and the magnitudes of the unobserved commuting costs. Appendix C.6 derives this result in a simple example.

Errors in baseline shares generate errors in predicted changes in counterfactual scenarios. The magnitude of the resulting errors hinges on two key questions: how large are the errors in the baseline shares and how sensitive is the counterfactual outcome of interest to individual linkages? Appendix C.4 uses the special case to illustrate that residential responses to a workplace productivity shock are more sensitive to sampling errors in baseline shares ℓ_{kn}/L than employment responses. This is because the former depends on the baseline share of a single residence-workplace pair, whereas the latter depends only on the workplace's share of total employment. Beyond this special case, the predicted changes in these models are generally a non-linear function of the economy's baseline equilibrium quantities. To explore a more typical case, we turn to Monte Carlo simulations.

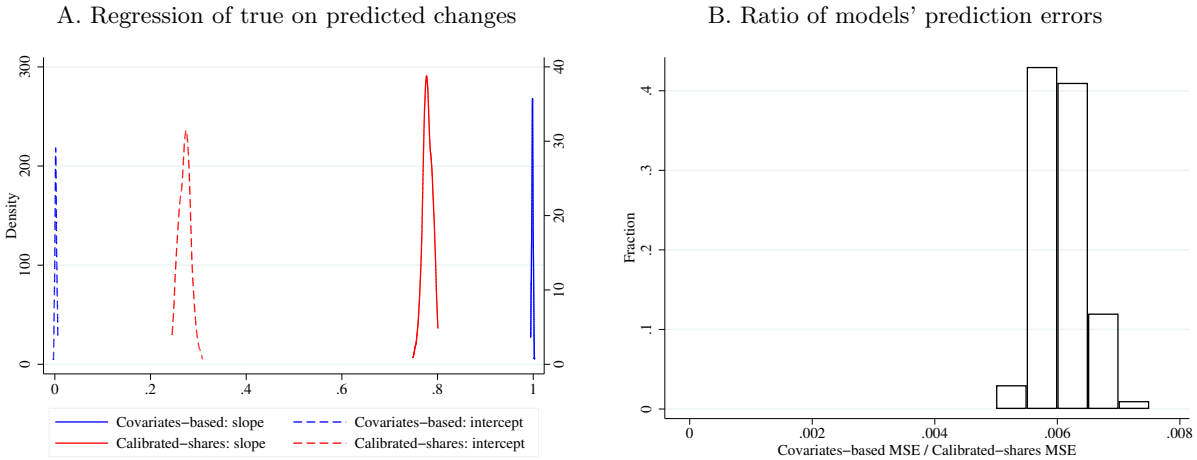
3.4 Monte Carlo: The calibrated-shares procedure in a granular setting

This section uses Monte Carlo simulations to assess how well the calibrated-shares procedure for computing counterfactual outcomes performs in a granular setting. In anticipation of examining New York City neighborhood employment booms in the next section, the data-generating process is the estimated covariates-based model of New York City in 2010. We imagine the researcher observes equilibrium wages and a finite-sample labor allocation drawn from a multinomial distribution defined by probabilities given by the right side of equation (2). The researcher then uses the calibrated-shares procedure to predict the consequences of a "counterfactual" productivity increase in one workplace tract. The "true" consequences of this productivity change are given by the counterfactual labor allocation and prices in the covariates-based continuum model. If the observed sample had an infinite number of individuals, the realized labor allocation would equal the true

[21] To gain intuition, consider a regression of the true probabilities on finite-sample shares. Because the frequency estimator is an unbiased estimator of the true probability, the finite-sample share is equal to the true probability plus finite-sample error. With infinite observations, the slope coefficient would be one, and in a typical finite sample, the slope coefficient would suffer attenuation bias, akin to classical measurement error.

shares (because the frequency estimator is consistent), and the calibrated-shares procedure would perfectly predict the changes caused by the productivity increase. In short, our simulations examine the finite-sample behavior of the calibrated-shares procedure using the actual number of individuals in New York City ($I \approx 2.5$ million).

Figure 2: Calibrated-shares procedure overfits in Monte Carlo simulations



NOTES: This figure depicts the regression coefficients and mean squared errors from 100 simulations in which $I = 2,488,905$, the data-generating process is the estimated covariates-based model of New York City in 2010, and the counterfactual change is a 9% increase in productivity in one workplace tract. See Appendix A.1 for a detailed description of the simulation procedure. "Covariates-based" means estimating the covariates-based parameterization described in Section 3.2 using the simulated data. "Calibrated-shares" means using the observed (simulated) shares in the exact hat algebra defined by equations (5)–(7). In Panel A, the densities for the covariates-based model are on the left vertical axis and the densities for the calibrated-shares procedure are on the right vertical axis.

We contrast the true and predicted changes in the number of commuters from each residential origin to the workplace with increased productivity. As detailed in Appendix A.1, the modeled productivity increases for both the covariates-based model and the calibrated-shares procedure are engineered so that the total employment increase in the "treated" tract is perfectly predicted. We assess the predictive power by contrasting true and predicted changes in bilateral commuting flows and rents. We do so by computing the mean squared error (MSE) and by regressing the observed changes on the predicted changes. With an unbiased forecast, the regression of observed on predicted changes would yield a slope coefficient of one and an intercept of zero. Of course, the estimated coefficients in a single simulation will also reflect finite-sample noise. Simulating this process 100 times yields a distribution of regression coefficients and MSEs, as shown in Figure 2.

Before assessing the calibrated-shares procedure's predictive performance, we note that the covariates-based model is well-behaved in the finite sample. The covariates-based model is correctly specified because $\lambda_{kn} = 1 \ \forall k, n$ in the data-generating process. It delivers accurate predictions when applied to a finite sample: its slope and intercept coefficients are centered on one and zero, respectively. It also accurately predicts the changes in rents (see Appendix Figure A.1).

The calibrated-shares procedure makes worse predictions. While it perfectly fits the commuting shares from the finite sample, it produces less accurate predictions about the changes in commuters and rents caused by the productivity increase. This illustrates an overfitting problem: a more flexible parameterization improves in-sample fit but worsens out-of-sample performance (Hastie, Tibshirani, and Friedman 2009, p. 221; Belloni, Chernozhukov, and Hansen 2014, p. 30). Consistent with the attenuation bias mentioned in Section 3.3, regressing the true change in commuters on the change predicted by the calibrated-shares procedure yields a slope of only 0.78 in the median simulation in Figure 2. Its forecast error is larger than that of the covariates-based model, which has a mean squared error that is 1% of the calibrated-shares forecast error in the median simulation. Its predictions about rents are similarly inferior: the regression slopes for rents are centered on 0.19, and the covariates-based model's mean squared error for rents is 0.003% of the calibrated-shares forecast error (see Appendix Figure A.1).

Table 2: Calibrated-shares procedure's finite-sample performance

A. Regressand is continuum change in commuters

I	2.5	5	12.5	25	50	125	250	2560
Calibrated-shares: slope	0.782	0.876	0.948	0.974	0.986	0.995	0.997	1.000
Calibrated-shares: intercept	0.269	0.153	0.064	0.032	0.017	0.007	0.004	0.000
Calibrated-shares: MSE	0.225	0.113	0.045	0.023	0.011	0.005	0.002	0.000

B. Regressand is finite-sample change in commuters

I	2.5	5	12.5	25	50	125	250	2560
Calibrated-shares: slope	-0.408	0.194	0.669	0.835	0.913	0.968	0.982	0.998
Calibrated-shares: intercept	1.724	0.982	0.404	0.202	0.106	0.040	0.022	0.002
Calibrated-shares: MSE	17.022	8.486	3.400	1.699	0.851	0.340	0.169	0.017

NOTES: This table reports the average values of the slope coefficient, intercept coefficient, and mean squared error from 100 simulations as we vary I, the number of individuals in the simulated economy. The column titles denote the number of individuals in millions. The "2.5" million case corresponds to $I = 2,488,905$, which is the number of individuals who reside and work in New York City in the 2010 LODES data. In Panel A, the regressand is the change in commuters with a continuum of individuals. This change does not vary across simulations. In Panel B, the regressand is the simulation-specific change in commuters from a realization drawn from the model using pre-shock parameter values to a realization drawn from the model using post-shock parameter values.

Since the only element of the data-generating process at odds with the assumptions of the calibrated-shares procedure is the finite number of individuals, these simulation results demonstrate that finite-sample bias can severely limit that procedure's predictive power. The procedure suffers from an overfitting problem when the number of individuals is small relative to the number of parameters. As shown in Panel A of Table 2, the calibrated-shares procedure performs better as we increase the number of individuals. Its mean squared error is inversely proportional to the number of people.[22] Only when this number is one to two orders of magnitude larger do

[22]The covariates-based model exhibits the same speed of convergence, as can be seen in Appendix Tables A.2 and A.3.

the calibrated-shares procedure's predictions become closely correlated with outcomes. Its rent predictions become closely correlated with outcomes only when the number of individuals is two to three orders of magnitude larger (Appendix Table A.1).

In any empirical application, the researcher does not observe changes in outcomes for a continuum of individuals. The observed changes in outcomes come from finite-sample draws from the pre- and post-shock data-generating processes. We simulate this setting by drawing two labor allocation realizations, the first from the estimated model of New York City in 2010 and the second from that model after the productivity increase in one workplace tract. Because the two realizations are independent draws, commuter counts are mean reverting: absent a shock, the changes are negatively correlated with the initial counts. This does not change the slope coefficient for the covariates-based model's predictions (see Appendix Figure A.2), but it severely degrades the predictive performance of the calibrated-shares procedure. For the smallest number of individuals in Panel B of Table 2, the calibrated-shares procedure's predicted changes are negatively correlated with the simulated changes. As explained in Appendix C.12, this particular symptom of overfitting is a case of Galton's fallacy.[23]

In the simulations described thus far, the covariates-based model is correctly specified. This neglects the calibrated-shares procedure's potential advantage over the covariates-based specification: it can capture spatial links that covariates do not. In Appendix A.3, we conduct simulations in which the covariates-based model is misspecified because $\lambda_{kn} \neq 1$ in the data-generating process. Tables A.2 and A.3 show that, for a given number of individuals, the calibrated-shares procedure's performance improves relative to the covariates-based model as the share of variation in δ_{kn} orthogonal to the included covariates increases. With sufficiently large unobserved commuting costs, the calibrated-shares procedure outperforms the covariates-based model. It achieves this by calibrating λ_{kn} for each pair of locations, which is highly data demanding. For the actual number of individuals in this empirical setting, the calibrated-shares procedure performs better when the unobserved component of commuting costs has a standard deviation at least half as large as that of the observed component (equivalent to a standard deviation of 15 minutes of transit time each way). As the number of individuals grows, the overfitting problem diminishes and the advantage of a more flexible parameterization is realized.

3.5 Event studies: Predicting commuting responses to a local shock

We now examine how various approaches are able to predict changes in commuting flows using neighborhood employment booms in New York City. As in the Monte Carlo simulations, we investigate changes in commuting patterns to workplace tracts that had large increases in employment. In particular, we conduct "event studies" in which we examine the commuting flows to the 83 workplace tracts in New York City that had a two-year increase in total employment

[23] One can use the model of Section 4 to take finite-sample draws from a pre- and post-shock data-generating process without inducing mean reversion. One does so by fixing the idiosyncratic preferences $\{\nu_{kn}^i\}$ to be the same in the two draws. A Monte Carlo exercise using this assumption shows that the calibrated-shares procedure still suffers from an overfitting problem and predicts outcomes worse than the covariates-based specification (see Appendix A.4).

from 2010 to 2012 of at least 400 employees and at least 12.5% from a 2010 level of at least 400 employees. We focus on local employment booms because these large changes are likely driven by workplace-specific shocks, such as new office openings or an expansion by a large employer, rather than resident-workplace-specific shocks. For example, in 2011, Tiffany & Co. moved its corporate headquarters to 260,000 square feet of office space at 200 Fifth Avenue. In late 2010, Google acquired a building of nearly 3 million square feet at 111 Eighth Avenue. These locations are two of the 83 workplace tracts we examine.[24]

To assess the predictive performance of a model parameterization, we use its baseline shares for New York City in 2010 and equations (5)–(7) to compute the increase in productivity required to match the observed 2010–2012 change in employment for the "treated" tracts.[25] Since different procedures fit the 2010 data differently, these productivity increases could vary, but in practice they are very similar. Since the productivity increases are defined so that both procedures match the observed increase in employment, we examine their predictions for the change in bilateral commuter counts.

3.5.1 Covariates-based model outperforms calibrated-shares procedure

Figure 3 contrasts the changes in commuter flows predicted by the covariates-based model and the calibrated-shares procedure for all 83 employment booms from 2010 to 2012. For each booming tract and each counterfactual procedure, we regress the observed change in the number of residents from each residential tract who work in the treated workplace tract on the predicted change. An unbiased prediction procedure should yield a slope of one and an intercept of zero. Panel A of Figure 3 depicts the distribution of these coefficients for both procedures. For our covariates-based model, the slope coefficients are roughly centered on one (median of 0.99), and the intercept coefficients are roughly centered on zero (median of 0.01).

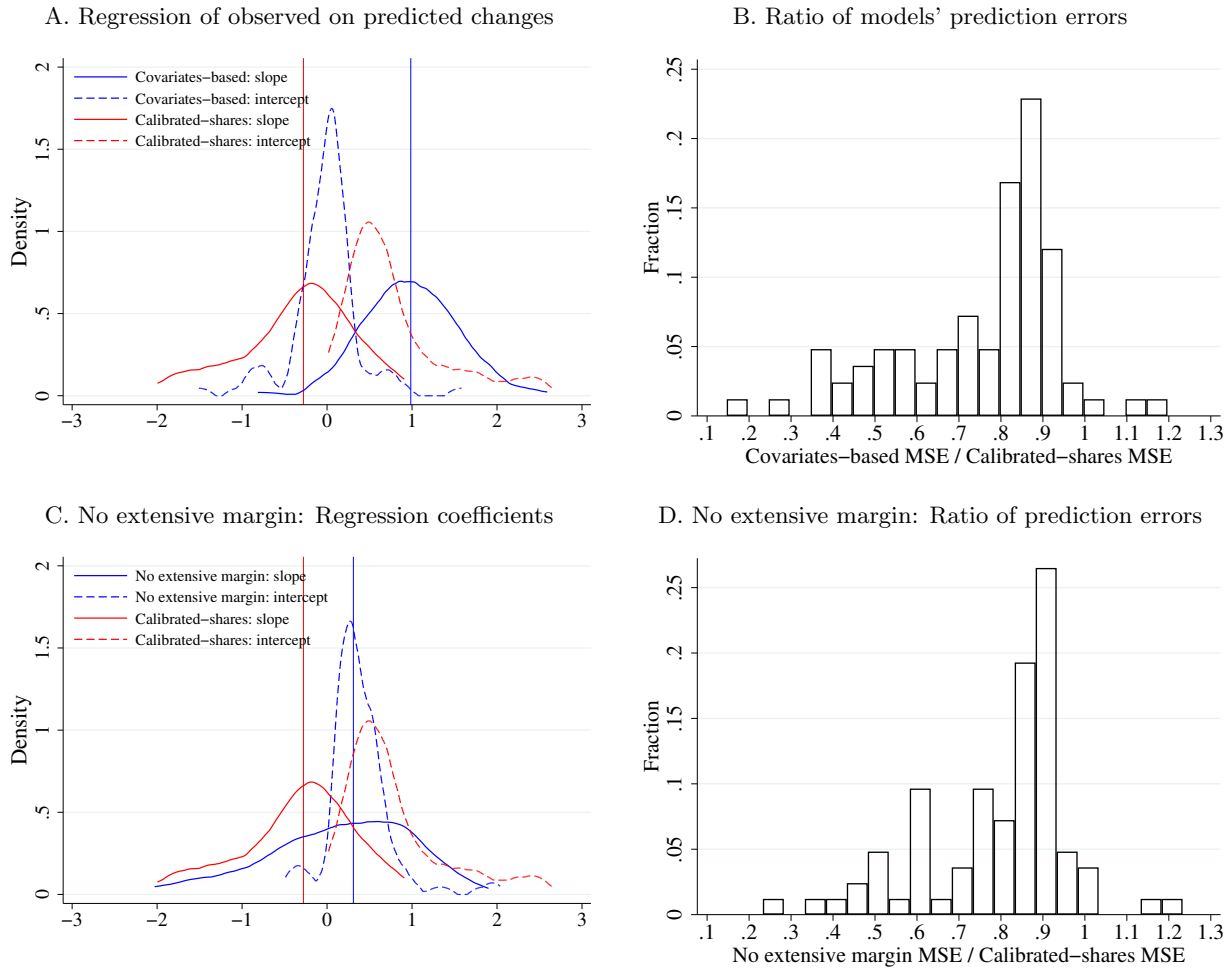
The calibrated-shares procedure does not perform as well. Across the 83 events, the median slope coefficient is -0.28, and the median intercept coefficient is 0.60. That is, the calibrated-shares procedure's predictions are negatively correlated with observed outcomes in more than half of the events. As a result, the covariates-based approach typically has a lower forecast error. Panel B of Figure 3 contrasts the two models' MSEs for each event. The covariates-based approach has a lower MSE than the calibrated-shares procedure in 80 of the 83 events.

Looking at in-sample fit for the 2010 commuting flows would suggest a very different contrast between the two approaches: the χ^2 test statistic for in-sample fit for the calibrated-shares procedure is literally zero, while the estimated covariates-based model would be rejected by a χ^2 test (see Appendix B.6). This warns against evaluating model specifications by in-sample fit. The covariates-based specification, while not perfectly describing the observed outcomes in 2010, is

[24] Figure B.1 depicts the employment changes for these two tracts.

[25] Because the 83 employment booms simultaneously occurred from 2010 to 2012, we compute the vector of 83 productivity increases that matches the observed changes in employment in these 83 workplace tracts. The results obtained when computing 83 separate productivity increases, which neglects the simultaneity, are virtually identical to those in Figure 3.

Figure 3: Comparison of models' predictive performance across 83 events



NOTES: "Calibrated-shares" refers to using the observed shares in the exact hat algebra defined by equations (5)–(7). "Covariates-based" in Panels A and B refers to the parameterization described in Section 3.2. "No extensive margin" in Panels C and D retains only the intensive-margin predictions of the covariates-based model. This case uses the covariates-based model's predictions for residence-workplace pairs with $\ell_{kn} > 0$ in the 2010 data and predicts zero change for pairs with $\ell_{kn} = 0$ in the 2010 data. In Panel A, 3 slope and 1 intercept coefficients are not depicted for the calibrated-shares procedure. In Panel C, 3 slope and 1 intercept coefficients are not depicted for the calibrated-shares procedure. Vertical lines depict the medians of the slope coefficient distributions.

much more informative about how commuting flows change with local employment booms. The calibrated-shares procedure predicts changes in commuting flows very poorly.

The covariate-based approach's superior predictive power is not due solely to the presence of zeros in the commuting matrix. To illustrate this, we produce predictions using a hybrid approach that uses the covariates-based model's predictions for residence-workplace pairs that have non-zero commuters in the 2010 data and predicts zero change for pairs where the baseline flow is zero (as in the calibrated-shares procedure). This hybrid approach, which retains only the intensive-margin predictions of the covariates-based model, still substantially outperforms the calibrated-

shares procedure (Figure 3C and Figure 3D).

3.5.2 Temporal aggregation and geographic aggregation

One may try to smooth out idiosyncratic elements of the data before fitting the model. For example, one might pool multiple years of data. In many empirical settings, particularly historical contexts, consecutive years of data are not available. In the case of LODES, annual data are available since 2002. We average commuter and wage observations for 2008–2010 before applying each approach.

Pooling three years of data yields a modest improvement for the calibrated-shares procedure (see Figure B.2). Its predictions are now positively correlated with observed outcomes for 69% of the events. Nonetheless, the covariates-based approach typically forecasts the changes in commuter counts much better. The covariates-based approach's slope coefficients are closer to one (median of 0.98 vs. 0.16), and its intercept coefficients are closer to zero (0.03 vs. 0.35). The covariates-based approach applied to the pooled data has a lower MSE than the pooled calibrated-shares procedure in 74 of the 83 events.

Using larger spatial units is another potential way to address the overfitting problem. We can aggregate the 2,160 residential tracts into 195 Neighborhood Tabulation Areas (NTAs) defined by the New York City Department of City Planning. We consider both aggregating the tract-level predictions up to NTA-level predictions and defining locations in the model to be NTAs.

When we aggregate the predicted changes in number of commuters to the booming workplace tracts from residential tracts up to residential NTAs, the covariates-based model still outperforms the calibrated-shares procedure (see Figure B.2). When aggregating predictions, the covariates-based model has a lower MSE than the calibrated-shares procedure in 55 of the 83 events.

Estimating the model using NTAs reduces the number of locations by an order of magnitude and thus the number of location pairs by two orders of magnitude. 35 NTAs had 2010–2012 employment booms. As described in Appendix B.3, the covariates-based approach and the calibrated-shares procedure perform similarly well when they are applied to NTAs. The covariates-based model has a lower MSE than the calibrated-shares procedure in 18 of the 35 events. In line with our Monte Carlo results, the calibrated-shares procedure is viable when locations are defined as NTAs such that there are more than 50 individuals per pair of locations.

Using larger spatial units avoids the overfitting problem that arises when the calibrated-shares procedure is applied to granular settings. But doing so leaves one unequipped to address the many research questions that concern small spatial units such as neighborhoods. Research does not need to avoid granular settings: the covariates-based approach performs well when applied to small spatial units.

3.5.3 Exact hat algebra using shares from an approximated matrix

Temporal and geographic aggregation are simply two of many statistical tools that researchers could use to reduce idiosyncratic noise in granular settings. We now consider a noise-reduction strategy that can be applied to a single year of data from a granular setting (without geographic

aggregation). The idea is to use the shares from an approximation of the commuting matrix in the exact hat algebra, which will reduce overfitting to the extent that the approximation reduces idiosyncratic noise.

In particular, we use a rank-restricted singular value decomposition (SVD) of the observed commuting matrix. The truncated SVD is widely used both to remove noise and compress data (Eldén, 2007, Ch. 6). Low-rank models are pervasive in data science and effective because matrices from many different data domains – movie preferences, text documents, medical records – are approximately low rank (Udell and Townsend, 2019). In economics, Conlon, Mortimer, and Sarkis (2023) explore using a low-rank approximation of the substitution matrix in demand estimation.

The commuting matrix $\mathbf{L} = \{\ell_{kn}\}$ of dimension $K \times N$ can be decomposed into three separate matrices $\mathbf{U}$, $\mathbf{\Sigma}$, and $\mathbf{V}$ such that $\mathbf{L} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$. This is an SVD of the commuting matrix where $\mathbf{U}$ is a $K \times K$ orthonormal matrix (the left singular vectors matrix), $\mathbf{\Sigma}$ is a non-negative $K \times N$ diagonal matrix with diagonal elements representing the singular values sorted in descending order, and $\mathbf{V}$ is an $N \times N$ orthonormal matrix (the right singular vectors matrix). The rank-$\mathfrak{r}$-restricted SVD retains only the $\mathfrak{r}$ largest singular values and sets the remaining singular values to zero. We replace negative elements of this rank-restricted matrix by zeros and rescale all values so that the total number of individuals in the resulting approximation is equal to that in the observed commuting matrix.[26]

This approximation requires the researcher to choose the rank $\mathfrak{r}$. The rank-1 approximation would underfit the data because it is merely the product of a k component and an n component, lacking a bilateral kn component. Using a full-rank matrix would mean using the observed commuting matrix. The statistics literature suggests computationally cheap rules of thumb for selecting the rank, such as visually identifying a point in a "scree plot" in which adjacent singular values differ little or selecting all SVD layers that explain more than a given share of the total variation (Cohen, 2021, p.496). Alternatively, one could also use a Monte Carlo analysis to pick the optimal rank. We discuss these approaches in Appendix B.4 and use them in our application to New York City. More importantly, it is computationally trivial to vary the rank in this approximation, making it easy to conduct sensitivity analyses.

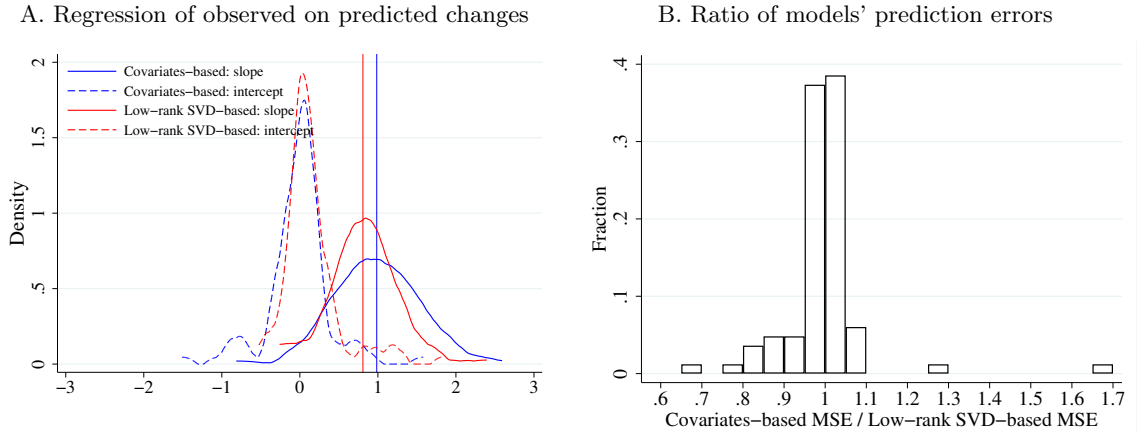
When studying neighborhood employment booms in New York City, using such an approximation of the commuting matrix yields substantially improved predictive performance over calibrating the model to the observed shares. We use an approximation derived from a rank-16 SVD, selected by a 0.5% threshold for the share of singular values. A Monte Carlo simulation using the covariates-based model as the data-generating process would have selected a rank of 18 (see Appendix B.4.2). Comparing the observed and rank-approximated commuting matrices visualized in Appendix Figure B.4 suggests that these approximations capture much of the relevant observed variation.[27]

[26] This replacement means the approximated matrix may not be rank $\mathfrak{r}$. A non-negative rank-$\mathfrak{r}$ approximation, which is costlier to compute, delivers very similar counterfactual predictions. See Appendix B.4.3.

[27] For example, the rank-16 SVD predicts 815 commuters for the Columbia University outlier, quite close to the observed 827 commuters and much larger than the covariates-based model's prediction of 70. One feature of the observed matrix that these low-rank approximations do not capture is the large counts on the diagonal of the matrix: the proclivity for individuals to reside and work in the same tract. Appendix Figure B.4 shows that both an

Plugging this approximation of the shares into equations (5)–(7) yields counterfactual predictions that are much closer to the predictions from the covariates-based model estimated on the full data. As shown in Figure 4, the median regression slope is 0.81, and the mean squared error is similar to that of the covariates-based approach. The regression slope coefficient from the rank-16 SVD is closer to 1.0 in 46 of the 83 events. This performance is not very sensitive to the precise choice of rank used in the approximation. Appendix Table B.1 reports the slope and MSE for a range of ranks. The MSE for the median event is similar for ranks 5 through 20.

Figure 4: Predictive performance across 83 events using low-rank approximation



NOTES: This figure depicts the predictive performance of using an approximation of the commuting matrix in exact hat algebra. The approximation is derived from a rank-16 singular value decomposition of the observed commuting matrix, as described in Appendix B.4. The covariates-based model is the same as in Figure 3. Panel A depicts the distributions of slope and intercept coefficients, and the vertical lines denote the median slope coefficients.

Using a noise-reducing approximation of the matrix of spatial outcomes in exact hat algebra may be fruitful in other contexts. The approximation guards against overfitting the spatial linkages between pairs of locations even in the absence of a high-quality covariate that predicts bilateral flows. Moreover, the computational costs of our suggested approximation are very low, as any modern linear algebra library allows one to compute the SVD of million-element matrices very quickly.

3.5.4 Interactive fixed effects

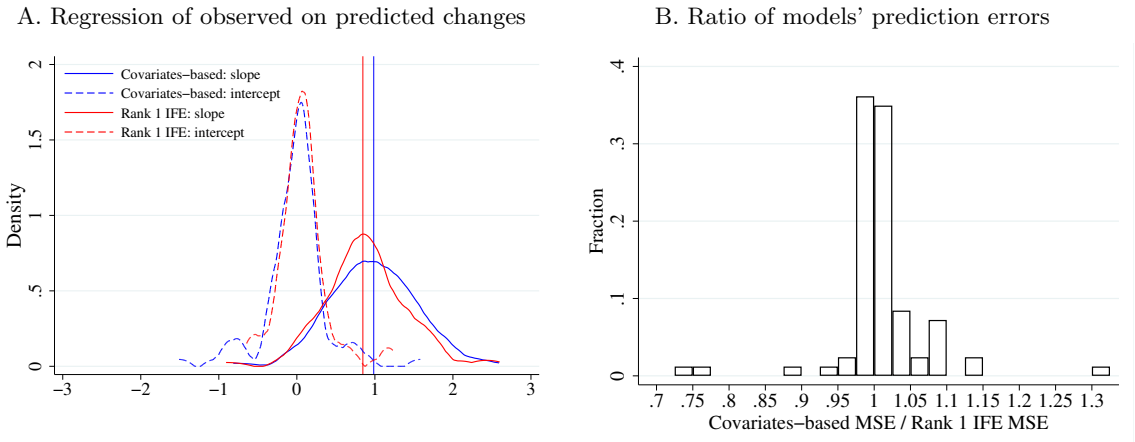
One can leverage the explanatory power of observed covariates while allowing for more flexibility by extending the covariates-based approach to have an explicitly modeled unobserved component of commuting costs. As detailed in Appendix B.5, we estimate an interactive-fixed-effects (IFE) specification that assumes the disutility component of commuting costs has an R-dimensional factor structure, $\lambda_{kn} = \exp(\psi_k'\gamma_n)$ (Bai, 2009; Chen, Fernández-Val, and Weidner, 2021). For intermediate

approximation derived from a rank-200 SVD and the covariates-based model do capture this pattern.

values of R, this specification is more flexible than the covariates-based model ($R = 0$) without going to the extreme of one parameter per pair of locations (the full-rank case). As with the covariates-based model, one can use the equilibrium shares of the estimated IFE model with equations (5)–(7) to compute counterfactual outcomes.

A rank-1 IFE specification offers a modest improvement over the covariates-based model in the event studies. As an encompassing specification, the IFE necessarily outperforms the covariates-based model (CBM) in terms of in-sample fit (see pseudo-R^2 in Appendix Table B.4 and χ^2 test in Appendix B.6). The event studies, summarized in Figure 5, examine its out-of-sample predictive performance. In Panel A, the IFE slope coefficient for the median event is 0.85, which is farther from 1.0 than the CBM. While not centered on 1.0, its distribution of slope coefficients is narrower such that the IFE slope coefficient is closer to 1.0 in 42 of the 83 events. The IFE's forecast error is smaller than that of the CBM in the typical event: Panel B shows that its MSE is lower in 47 of the 83 events. Thus, the more flexible interactive-fixed-effects specification offers a modest improvement in predictive performance. Rank-2 and rank-3 IFE specifications perform very similarly (see Appendix Figure B.6).

Figure 5: Predictive performance of the rank-1 IFE specification across 83 events



NOTES: This figure summarizes the predictive performance of the covariates-based model and the rank-1 IFE specification. Panel A depicts the distributions of slope coefficients, and the vertical lines denote the median slope coefficients.

3.5.5 Sensitivity analysis

One difference between the approaches described above is in their use of estimated or observed wages. The covariates-based model and interactive-fixed-effects specification estimate wages using workplace fixed effects. By contrast, observed wages are used in the calibrated-shares procedure and in the exact hat algebra with the SVD-approximated shares. We have verified that using alternative wage values—such as equal wages across locations—does little to alter the predictions

of the calibrated-shares procedure in the event studies.

We find similar event-study results when employing different assumptions about preferences and production. Thus far, we have assumed idiosyncratic preferences are spatially independent, as in prior research. Making residential locations in the same NTA closer substitutes in a nested-logit demand system has little effect on the predictions produced by either method (see Appendix B.7). Similarly, the changes in commuters predicted by both methods are not sensitive to the elasticity of substitution between goods σ (see Appendix B.8) nor to introducing local increasing returns at workplaces (see Appendix B.9).

3.6 Counterfactuals in continuum models: Takeaways

The predictive performance of quantitative spatial models applied to granular empirical settings is highly sensitive to the baseline equilibrium shares used in the exact hat algebra. The incumbent methods—the covariates-based approach and calibrated-shares procedure—produce sharply contrasting predictions for bilateral commuting flows because they parameterize spatial linkages very differently. Using shares from a covariates-based model predicts changes in commuting flows remarkably well. Using observed shares to calibrate millions of parameters suffers from an overfitting problem. Aggregating data across a few years does not remedy this problem. Using shares produced from a low-rank approximation of the commuting matrix alleviates the overfitting problem and produces predictions comparable to the covariates-based approach. An interactive-fixed-effects specification that extends the covariates-based approach offers modest improvements in predictive performance at greater computational cost.

When are the counterfactual predictions from the calibrated-shares procedure likely to suffer an overfitting problem? Generally, this depends on the magnitude of errors in the baseline shares and the sensitivity of counterfactual outcomes of interest to these errors. To examine the calibrated-shares procedure's finite-sample performance in a specific context, we recommend using a covariates-based model as the data-generating process in Monte Carlo simulations, as in Section 3.4.[28] The inputs necessary for such simulations are the same as those used in the calibrated-shares procedure, assuming the researcher has used a covariate to estimate the commuting elasticity. Verifying finite-sample performance is relevant beyond spatial economics, as calibrating model parameters to match observed market shares is also common in fields such as international trade and industrial organization. Whether perfectly matching the market shares is a feature or a bug depends on the context. In settings with noisy observed shares, using shares produced by data-smoothing procedures, such as low-rank approximations, or using more parsimonious models is likely to deliver better counterfactual predictions.

[28] Our code to simulate quantitative urban models is at https://github.com/jdingel/DingelTintelnotSEGS.

4 A spatial model with a finite number of individuals

Thus far, we studied the continuum model assuming that the observed data were a finite sample drawn from that model. The overfitting problem that troubles the calibrated-shares procedure in granular settings stems from substantial idiosyncratic noise embodied in the observed baseline shares. These idiosyncrasies matter because there are few decision makers per spatial link, even when we observe, as researchers often do, the universe of decision makers. In this section, we explore a related implication: when baseline outcomes depend on individual idiosyncrasies, so do counterfactual outcomes. In such settings, the patterns of substitution elicited by changes in economic primitives are shaped by individuals' idiosyncratic preferences. Unfortunately, the continuum model discussed in Sections 2 and 3 does not capture this. Contrary to individual idiosyncrasies affecting outcomes in granular settings, the premise of the continuum model is that they are integrated away.

We introduce a spatial model with a finite number of individuals in which equilibrium outcomes depend on individual idiosyncrasies. Its outcomes differ from the continuum model in two ways. First, because these models are not linear, the equilibrium prices in a continuum model need not equal the average equilibrium prices from a model with finite individuals. Second, counterfactual changes in quantities and prices depend on which individuals change their decisions in response to the exogenous shock, which depends on their vectors of unobserved idiosyncratic preferences. There is therefore uncertainty about realized counterfactual outcomes, as the distribution of idiosyncrasies induces a distribution of counterfactual changes.

In a model with a finite number of people, individuals' decisions affect wages and rents, raising two issues. First, do individuals internalize the effects of their own choices on local labor supplies and land demands? Second, are individuals able to enumerate the prices induced by every possible combination of others' choices? For tractability, we assume that individuals have common point-mass beliefs about wages and rents. Therefore, individuals act as price takers, as in Gabaix (2011). In our application, we assume that, given the model parameters, individuals have beliefs about wages and rents that are the equilibrium prices of the conventional continuum model. We label these beliefs "continuum-case rational expectations" because they would be rational if there were a continuum of individuals.[29] This model with a finite number of individuals is intentionally similar to the conventional continuum model: they coincide as the number of individuals becomes arbitrarily large and can be estimated similarly.

[29] Similar simplifications to agents' beliefs have been made in industrial organization (e.g., oblivious equilibrium in Weintraub, Benkard, and Van Roy 2008) and macroeconomics (e.g., Krusell and Smith 1998). Specifically, in an oblivious equilibrium, a finite number of firms optimize assuming that the industry state always equals its long-run expected value, even though the industry state is not constant given a finite number of firms and idiosyncratic shocks. In the Krusell and Smith (1998) model, heterogeneous agents condition their choices on moments from the distribution of state variables (rather than tracking the entire distribution).

4.1 Setup

Our model with a finite number of individuals features the same utility function, technology, and commuting costs as in Section 2, so the set of primitives $\Upsilon \equiv \{L, \{A_n\}, \{T_k\}, \{\bar{\delta}_{kn}\}, \{\lambda_{kn}\}, \alpha, \epsilon, \sigma\}$ is the same.[30] It differs in two important respects. First, the aggregate labor endowment L is embodied in I individuals each supplying $\frac{L}{I}$ units of labor. Second, we introduce a timing assumption: individuals first choose their residence-workplace pairs, and then all markets clear given individuals' locations. This timing would not matter if there were a continuum of individuals.

We denote the set of all idiosyncratic preferences ν_{kn}^i by $\boldsymbol{\nu}^I$. As before, these are i.i.d. draws from the standard Gumbel distribution. The superscript I indicates that the set grows with the number of individuals.

We make the following assumptions about information and expectations. All workers know the economic primitives Υ and have (common) expectations about equilibrium prices r_k and w_n. The price expectations are point-mass beliefs, such that each individual assigns 100% probability to a vector of wages $\{\tilde{w}_n\}$ and a vector of land prices $\{\tilde{r}_k\}$. Worker i knows her idiosyncratic preferences $\{\nu_{kn}^i\}$.

Decisions are made and markets clear in the following order. Based on beliefs $\{\tilde{w}_n\}$ and $\{\tilde{r}_k\}$, each worker chooses the residential location and the work location that maximize expected utility

$$\tilde{U}_{kn}^i = \underbrace{\epsilon \ln\left(\tilde{w}_n / \left(\tilde{P}^{1-\alpha} \tilde{r}_k^{\alpha} \delta_{kn}\right)\right)}_{\equiv \tilde{U}_{kn}} + \nu_{kn}^i, \tag{9}$$

where $\tilde{P} = \left[\sum_n \left(\tilde{w}_n / A_n\right)^{1-\sigma}\right]^{1/(1-\sigma)}$. After these decisions are made, individuals cannot relocate.[31] As detailed below, realized equilibrium land prices r_k and wages w_n are those that clear goods, labor, and land markets given individuals' residential and workplace locations.

Assuming point-mass beliefs about prices considerably simplifies the analysis. Otherwise, individuals would need to compute expectations using (equilibrium) probabilities for an enormous set of feasible allocations. For example, an economy with only 10 individuals and 16 residence-workplace pairs has more than 3 million possible allocations.[32] For empirically relevant magnitudes, this is infeasible. While our approach permits arbitrary point-mass beliefs about wages and rents, in our analysis below we assume that the expected prices are the equilibrium wages and rents of the continuum model with the same economic primitives Υ.

[30] Model extensions that introduce trade costs, residential amenities, local increasing returns, and production employing land are presented in Appendix C.11.

[31] The assumption that individuals make irreversible decisions is common in static spatial models. For example, in the open-city model of Ahlfeldt et al. (2015), individuals choose to live in Berlin based on expected utility, which is equal to the reservation level of utility in the wider economy. Individuals who choose Berlin and realize utility below the city-wide average cannot leave. Redding and Rossi-Hansberg (2017), Heblich, Redding, and Sturm (2020), Owens, Rossi-Hansberg, and Sarte (2020), and Brinkman and Lin (2022) make the same irreversibility assumption. In dynamic location choice models (e.g., Diamond, McQuade, and Qian 2019), moving costs cause agents to not necessarily change locations in response to shocks.

[32] With I individuals and N^2 residence-workplace pairs, the set of possible allocations (the support of the multinomial distribution) contains $\binom{I+N^2-1}{N^2-1} = \frac{(I+N^2-1)!}{(N^2-1)!I!}$ elements. For $I = 10$ and $N = 4$, this is about 3.27×10^6.

4.2 Equilibrium

We distinguish between a *trade equilibrium*, which clears goods and land markets taking individuals' locations as given, and a *commuting equilibrium with finitely many individuals*, in which individuals choose locations based on beliefs about the trade equilibrium that will result.

Definition 4.1. Trade equilibrium. Given the labor allocation $\{\ell_{kn}\}$ and economic primitives Υ, a trade equilibrium is a set of wages $\{w_n\}$ and land prices $\{r_k\}$ satisfying equations (3) and (4).

Note that these market-clearing conditions are the same as in the continuum model.[33] Given point-mass beliefs $\{\tilde{w}_n\}$ and $\{\tilde{r}_k\}$ and idiosyncratic preferences $\boldsymbol{\nu}^I$, individuals choose the residence-workplace pair that maximizes their utility in equation (9).

Definition 4.2. Commuting equilibrium with finitely many individuals. Given a number of individuals I, economic primitives Υ, idiosyncratic residence-workplace preferences $\boldsymbol{\nu}^I$, and a set of point-mass beliefs $(\{\tilde{w}_n\}, \{\tilde{r}_k\})$, a commuting equilibrium with finitely many individuals is a labor allocation $\{\ell_{kn}\}$, wages $\{w_n\}$, and land prices $\{r_k\}$ such that

- $\ell_{kn} = \frac{L}{I} \sum_{i=1}^{I} \mathbf{1}\{\tilde{U}_{kn} + \nu^i_{kn} > \tilde{U}_{k'n'} + \nu^i_{k'n'} \ \forall (k', n') \neq (k, n)\}$; and
- wages $\{w_n\}$ and land prices $\{r_k\}$ are a trade equilibrium given the labor allocation $\{\ell_{kn}\}$.

We now define a set of price beliefs that are the equilibrium prices of the continuum model from Section 2 with the same economic primitives.[34]

Definition 4.3. Rational expectations for the continuum case. Given economic primitives Υ, $\{\tilde{w}_n\}$ and $\{\tilde{r}_k\}$ are "continuum-case rational expectations" if $\{w_n\} = \{\tilde{w}_n\}$ and $\{r_k\} = \{\tilde{r}_k\}$ constitute an equilibrium of the continuum model defined in Section 2.2 for the same economic primitives Υ.

Distinguishing between the number of individuals I and aggregate labor L allows us to study locational decisions as $I \to \infty$ without changing aggregate labor supply. As Appendix C.9 shows, the labor allocation in the $I \to \infty$ limit is the same as in the continuum model in equation (2). Note that given idiosyncratic preferences and point-mass beliefs, the commuting equilibrium with finitely many individuals is unique.

4.3 Estimation of economic primitives

One can estimate a covariates-based specification of this model in the same way we estimated its continuum-model counterpart in Section 3.2. We take $\alpha, \sigma, \bar{\delta}_{kn}$ as given, impose $\lambda_{kn} = 1 \ \forall kn$, and

[33] The trade equilibrium is unique. Given the labor allocation, there is a unique set of relative wages satisfying equation (3), as shown in Appendix C.8. Equation (4) can be rewritten as $r_k = \frac{\alpha}{T_k} \sum_n \frac{\ell_{kn}}{\bar{\delta}_{kn}} w_n$, so there is a unique set of land prices associated with that wage vector.

[34] The wage and rent beliefs in Definition 4.3 are unique if the parametric condition for uniqueness in the continuum model in Section 2.2 is satisfied.

assume that ν^i_{kn} are i.i.d draws from a standard Gumbel distribution.[35] The resulting log likelihood function is

$$\mathcal{L} \equiv \sum_{k,n} \ell_{kn} \ln \left[\mathbb{P}(\tilde{U}^i_{kn} > \tilde{U}^i_{k'n'} \ \forall k'n' \neq kn) \right] = \sum_{k,n} \ell_{kn} \ln \left[\frac{\tilde{w}^\epsilon_n \left(\tilde{r}^\alpha_k \bar{\delta}_{kn} \right)^{-\epsilon}}{\sum_{k',n'} \tilde{w}^\epsilon_{n'} \left(\tilde{r}^\alpha_{k'} \bar{\delta}_{k'n'} \right)^{-\epsilon}} \right]. \tag{10}$$

This is equation (8) with rent and wage beliefs instead of equilibrium rents and wages. The maximum-likelihood estimate of the commuting elasticity ϵ is identical to that in the covariates-based continuum model (in column 1 of Table 1). The remaining economic primitives $\{T_k\}$ and $\{A_n\}$ can be obtained from the estimated rent and wage beliefs: the residence and workplace fixed effects are proportional to $\tilde{r}_k^{-\alpha\epsilon}$ and $\tilde{w}_n^\epsilon$, respectively. If $\{\tilde{r}_k\}$ and $\{\tilde{w}_n\}$ are continuum-case rational expectations, plugging these estimated beliefs into equation (2) yields the continuum-case labor allocation. Given α, σ, $\{\tilde{r}_k\}$, $\{\tilde{w}_n\}$, $\{\delta_{kn}\}$, and that labor allocation, equations (3) and (4) can be solved to obtain $\{T_k\}$ and $\{A_n\}$. These estimated land and productivity parameters are identical to those of the covariates-based continuum model.

4.4 Price dispersion and ex post regret

The realized equilibrium rents and wages vary with individual idiosyncrasies, so they generically differ from the point-mass beliefs about rents and wages that govern individuals' choices of residences and workplaces. How often would individuals choose a different residence-workplace pair at the realized equilibrium prices if they were able? In our New York City setting, ex post regret is rare: 96% of individual have idiosyncratic preferences such that they would not want to switch. See Appendix E for details of these results. Given the return to making the model tractable and computationally feasible, we judge this magnitude to be modest.

4.5 Contrasts with the continuum model

With a finite number of individuals, individual idiosyncrasies affect equilibrium outcomes. Given economic primitives Υ, this model produces a distribution of outcomes associated with the distribution of idiosyncratic preference shocks $\boldsymbol{\nu}^I$, whereas the continuum model delivers deterministic outcomes. Such a contrast has been emphasized by studies of how aggregate outcomes depend on firm-level productivity shocks (e.g., Gabaix 2011; di Giovanni, Levchenko, and Mejean 2014; Eaton, Kortum, and Sotelo 2013; Gaubert and Itskhoki 2021).

Given the same primitives Υ, the expected equilibrium labor allocation of this model equals that of the continuum model. This is because the multinomial distribution of residence-workplace outcomes $\{\ell_{kn}\}$ defined by the probabilities in equation (10) has expected values equal to the quantities in equation (2). By contrast, the model's equilibrium prices solve a system of non-linear

[35] The assumption that $\lambda_{kn} = 1$ can be relaxed by allowing the unobserved disutility component of commuting costs to have a factor structure, $\lambda_{kn} = \exp(\psi'_k \gamma_n)$, as in Section 3.5.4 and Appendix B.5. If individuals have continuum-case rational expectations, the economic primitives of this specification of the model with a finite number of individuals match those of the interactive-fixed-effects specification of the continuum model.

equations, so their expected values need not equal the continuum-case prices. For our baseline parameter values, these differences are small (Appendix E.1).

We focus on the implications of these idiosyncrasies for counterfactual outcomes. We compute the change in outcomes induced by a change in economic primitives from Υ to Υ' given idiosyncratic preferences $\boldsymbol{\nu}^I$.[36] In special cases, one can analytically characterize the uncertainty stemming from idiosyncrasies. In Appendix C.10, we describe a two-workplace economy. Let p denote the share of individuals expected to switch workplaces in response to a productivity change in the continuum model. In the model with a finite number of individuals, the variance of the share of individuals who switch workplaces is $\frac{p(1-p)}{I}$. In the general case, simulating realizations of $\boldsymbol{\nu}^I$ yields a distribution of counterfactual changes. The dispersion in counterfactual changes represents uncertainty stemming from individual idiosyncrasies. Although the analytical result describes a special case, it will turn out to approximate the uncertainty about the share of individuals switching workplaces in our Amazon HQ2 application well.

5 Application to Amazon HQ2 in Long Island City

We now examine the predictions of quantitative spatial models for the economic consequences of Amazon's aborted second headquarters (HQ2) in Long Island City, a controversial and widely discussed proposal. In November 2018, the company announced it would hire more than 25,000 employees in 4 million square feet of office space.[37] However, Amazon scrapped the project in February 2019 after facing a fierce backlash from local politicians and community members concerned about corporate subsidies and gentrification.

This large counterfactual change to a single workplace allows us to illustrate how choices of methods affect predictions. Section 5.1 shows how the geographic incidence of a productivity shock on local rents is governed by the parameterization of spatial linkages. Section 5.2 uses the model with a finite number of individuals to quantify the uncertainty about these counterfactual outcomes. Individual idiosyncrasies make neighborhood-level counterfactual changes considerably uncertain, even when examining a productivity shock shifting 1% of employment to a single tract.

5.1 Geographic incidence of Amazon HQ2

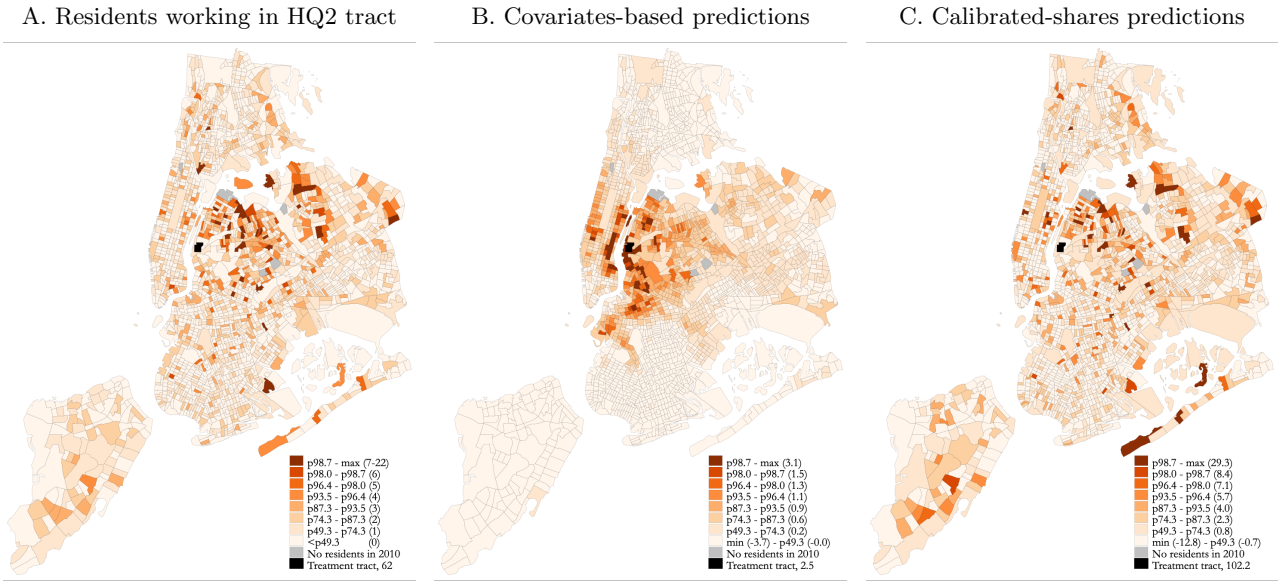
We analyze the consequences of a productivity increase in the Long Island City tract that would cause employment to rise by 25,000 workers from its 2010 level. Figure 6A depicts the distribution of residences among workers employed in that tract in 2010. To contrast the predictions of the

[36] This is akin to the computation of counterfactual changes holding a realization of firms' idiosyncratic unit costs fixed in Eaton, Kortum, and Sotelo (2013, Section 6.1).

[37] Amazon had requested proposals for HQ2 locations from cities and states across North America. In November 2018, it announced two winners, saying that it would hire more than 25,000 employees each in Long Island City in New York and Arlington in Virginia (Amazon, 2018). More details can be found in the Memorandum of Understanding between New York State, New York City and Amazon. Berkes and Gaetani (2023) introduce a model that features productivity spillovers and skill heterogeneity to analyze Amazon HQ2. Following Berkes and Gaetani (2023), the "treated" tract in our counterfactual scenario is 36081000700.

continuum model when using the covariates-based model and the calibrated-shares procedure, we find the productivity increases $\hat{A}_{n^*}^{\text{CBM}}$ and $\hat{A}_{n^*}^{\text{CSP}}$ that raise employment by 25,000 in the Amazon workplace tract n^*, holding all other economic primitives fixed.[38] Appendix F.1 reports additional results from using alternative values of the labor demand elasticity σ, from using a nested-logit demand system, and from the SVD-based and IFE specifications of the continuum model.

Figure 6: Amazon HQ2 counterfactual change in residents



NOTES: Panel A depicts the number of residents in each tract who work in the treated Long Island City tract in the 2010 LODES data. The legend of Panel A reports the percentiles corresponding to the integer number of residents, excluding the treatment tract. Panels B and C depict changes in the number of residents predicted by the covariates-based model and calibrated-shares procedure, respectively. The legend percentile cutoffs in Panels B and C correspond to those in Panel A. The covariates-based predictions for residents describe both the continuum model and the model with finite individuals because the change in the former equals the expected change in the latter.

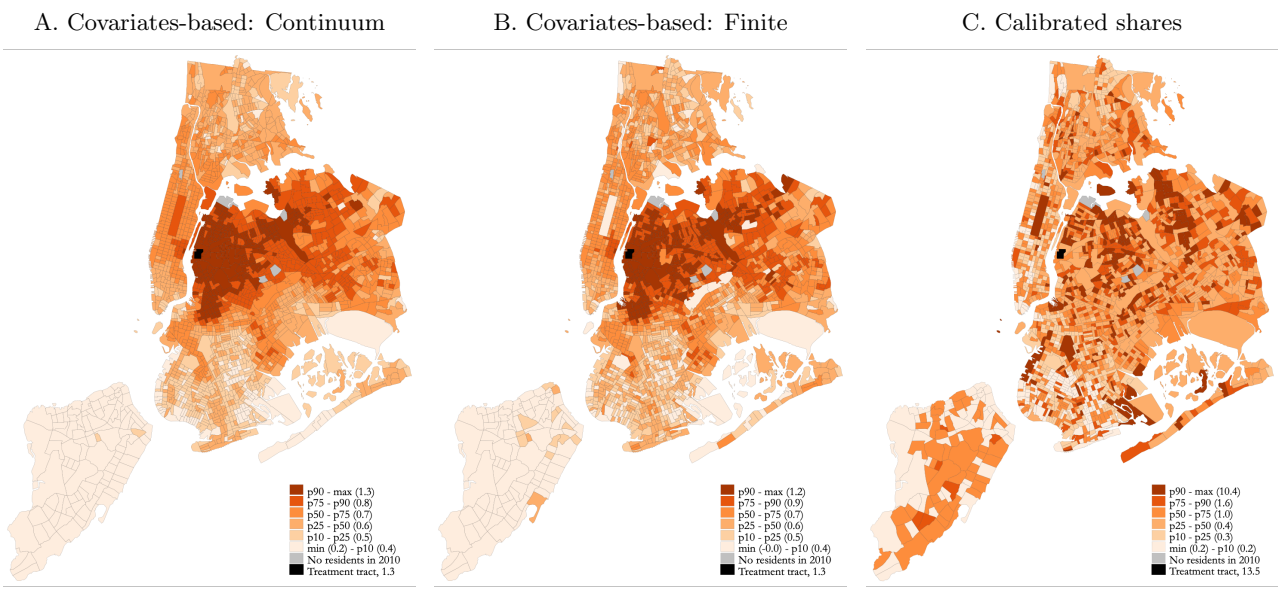
The counterfactual changes in residential populations are sensitive to the method used: while the covariates-based approach predicts small changes in all tracts, the calibrated-shares procedure predicts more varied and less spatially correlated changes. The covariates-based model's predicted changes in the number of residents in each tract, shown in Figure 6B, are modest in size and closely related to transit times to Long Island City. The calibrated-shares procedure's predicted changes, shown in Figure 6C, are very large for some residential tracts. The calibrated-shares procedure predicts increases in residents in neighborhoods scattered across the city, but little increase in residents in adjacent neighborhoods with short commutes to Long Island City. The calibrated-shares procedure's predictions are closely tied to the initial numbers of residents working in the treated tract: the spatial patterns in Figure 6A and Figure 6C are remarkably similar.

The incidence of Amazon HQ2 on residential rents predicted by the two methods differs consider-

[38] That is, we find the value of $\hat{A}_{n^*}^{\text{CBM}}$ such that the solution to equations (5)–(7) yields $\hat{\ell}_{kn}^{\text{CBM}}$ so that $\sum_k (\hat{\ell}_{kn^*}^{\text{CBM}} - 1)\ell_{kn^*}^{\text{CBM}} = 25,000$ and similarly for the CSP case.

ably. The greater variation in changes in the number of residents predicted by the calibrated-shares procedure is accompanied by greater variation in the predicted changes in real rents, $\hat{r}_k/\hat{P}$. The calibrated-shares procedure predicts rent increases of 1.6% to 10.4% for the top decile of tracts (Figure 7C). It suggests that in 12 census tracts, real rents would increase by at least 5%. By contrast, the covariates-based model predicts more modest increases of 0.8% to 1.3% for the top decile of affected tracts (Figures 7A and 7B).[39] The geographic incidence differs notably. Figure 7 shows that the calibrated-shares procedure predicts much larger rent increases in neighborhoods far from Amazon HQ2, such as Staten Island in the southwest corner of the map.[40] By contrast, the political opposition to Amazon HQ2 based on concerns about gentrification largely concerned housing prices in nearby constituencies (Goodman, 2019).

Figure 7: Predicted changes in rents



NOTES: These maps depict percent changes in rents, $(\hat{r}_k/\hat{P} - 1) \times 100$. 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 rent change using the calibrated-shares procedure.

Unlike the changes in residents and rents, the changes in workers, wages, and welfare predicted by the two methods are quite similar. Appendix Figures F.1 and F.2 show the changes in workers and wages. These predictions are similar because the geographic pattern of employment changes in response to a productivity shock is closely tied to the initial number of workers employed in each tract, which the two methods agree upon. Similar to the special case presented in Appendix C.4 and summarized in Section 3.3, residential responses to a workplace productivity shock largely depend on the baseline share of a single residence-workplace pair, which differs between the two methods,

[39] The covariates-based continuum model and model with a finite number of individuals predict very similar rent changes (compare Panels A and B).

[40] Appendix Figure F.3 plots rent changes against distance to the treated tract for both methods.

whereas employment responses depend on the workplace's share of total employment. The citywide welfare consequence of the productivity shock, derived in Appendix C.7, is also very similar across methods. This occurs because the welfare effect of this single-tract employment boom is largely invariant to spatial linkages: computing counterfactual changes using a parameterization with no spatial linkages, $\delta_{kn} = 1\ \forall kn$, also yields a very similar welfare consequence.[41]

5.2 Uncertainty about counterfactual predictions induced by idiosyncrasies

Counterfactual predictions come with various forms of uncertainty. For example, to capture uncertainty about parameter values, researchers can bootstrap the counterfactual predictions of a continuum model. Beyond parameter uncertainty, when the number of individuals is finite, there is a distribution of equilibrium outcomes associated with the distribution of individuals' idiosyncratic preferences. To capture the uncertainty stemming from individual idiosyncrasies that accompanies a change in economic primitives, we simulate the counterfactual change from Υ to Υ' for 100 realizations of idiosyncratic preferences $\boldsymbol{\nu}^I$ and construct a 90% confidence interval by reporting the 5^{th} and 95^{th} percentiles of the distributions of changes.

The predicted tract-level changes in residents and rents are small relative to the idiosyncratic variation. Figure 8A plots the 90% confidence interval for each tract's predicted change in residents against its mean predicted change. Every tract's 90% confidence interval for the change in residents has a positive upper bound and a negative lower bound. In this sense, none of the predicted changes in residents are distinguishable from zero. Figure 8C depicts the results for rents. For about 15% of residential tracts, the predicted changes in real rents are distinguishable from zero.

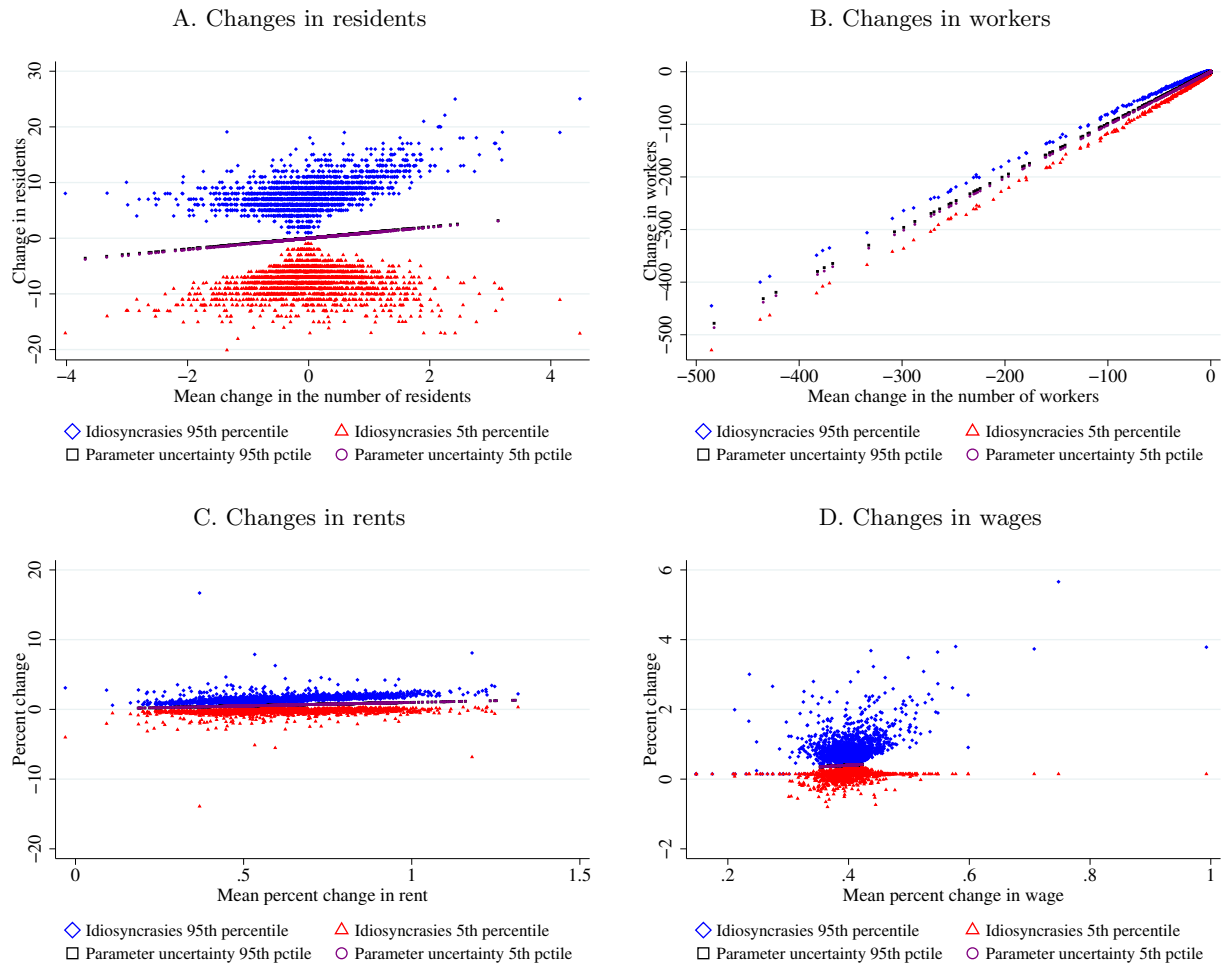
Since the predicted changes in the number of workers for some tracts are larger, many of them are meaningfully large relative to the idiosyncratic variation. As the number of workers varies across tracts much more than the number of residents, so too do the widths of the confidence intervals. Figure 8B depicts these confidence intervals for the changes in the number of workers. For hundreds of workplace tracts, the 90% confidence interval includes both negative and positive values.[42] For tracts predicted to lose many workers, the predicted change is sufficiently large relative to the uncertainty that even the 95^{th} percentile of changes is negative. It out turns that a binomial distribution with I draws and p equal to the expected change in the share of employment (the analytical result for uncertainty for the two-workplace case mentioned in Section 4.5), is a good approximation of the 90% confidence intervals depicted in Figure 8B.[43] For the vast majority of

[41] This finding does not imply that the two methods will always yield similar welfare conclusions. For counterfactual scenarios where spatial linkages matter for citywide welfare consequences, they may differ. For example, the citywide welfare benefit of faster commuting between two tracts that had zero commuters in the 2010 data would be zero under the calibrated-shares procedure but positive when using the baseline shares from the covariates-based parameterization.

[42] While workplaces are generically substitutes for one another, the Amazon expansion alters the pattern of residential rents, which can cause some individuals to choose a new residence-workplace pair that includes a non-Amazon workplace. An (idiosyncratic) increase in labor supply can even reduce real wages in such workplaces, as shown in Figure 8D.

[43] For example, for $I = 2,488,905$ and the workplace with an expected decrease of 300 employees ($p \times I = 300$), the binomial distribution's 90% confidence interval for the employment decrease, $pI \pm 1.64\sqrt{p(1-p)I} \approx [272, 328]$,

Figure 8: Variation in counterfactual changes: Idiosyncrasies vs. parameter uncertainty



NOTES: These plots depict variation in counterfactual changes induced by individual idiosyncrasies in the model with a finite number of individuals alongside variation attributable to parameter uncertainty about $\{A_n\}$, $\{T_k\}$, and ϵ in the covariates-based continuum model. We summarize the former by plotting the 5^{th} and 95^{th} percentiles of predicted changes in quantities and prices across 100 simulations of the model with a finite number of individuals. In Panels A, B, C, and D, the horizontal axes correspond to the values depicted in Figures 6B, F.1B, 7B, and F.2B, respectively. Panels B and D exclude the Amazon HQ2 workplace tract. In Panel A, the 90% confidence interval for the change in residents includes zero for all tracts. In Panel B, one outlier with an employment decline of 1,483 is not depicted, and the 90% confidence interval for the change in employment includes zero for 413 of the 2142 non-Amazon workplaces. The 90% confidence interval for the change in employment for the Amazon workplace tract is 24,777 to 25,349. In Panel C, real rent changes are $(\hat{r}_k/\hat{P}-1)\times 100$, and 322 out of 2160 origin tracts have a positive change in rents at the 5^{th} percentile. In Panel D, the real wage changes are $(\hat{w}_n/\hat{P}-1)\times 100$, and 1950 out of 2143 destination tracts have a positive change in wages at the 5^{th} percentile. We quantify parameter uncertainty by producing 100 counterfactual predictions by sampling with replacement 2,488,905 individuals from the observed commuting matrix 100 times. For each of these 100 samples, we estimate the covariates-based continuum model and compute its counterfactual predictions. The plots depict the 5^{th} and 95^{th} percentiles of predicted changes in quantities and prices across these 100 bootstrap samples. The variation attributable to parameter uncertainty is considerably smaller than the variation attributable to individual idiosyncrasies: the bootstrap 5^{th} and 95^{th} percentiles are almost indistinguishable.

aligns closely with that in Figure 8B.

workplace tracts, the uncertainty over real wages is sufficiently small that the 5th-percentile change for the real wage is greater than zero (see Figure 8D).[44]

The uncertainty stemming from individual idiosyncrasies in the model with finitely many individuals is larger than the variation in counterfactual predictions associated with parameter uncertainty in the continuum model in this setting. To compute the latter, we sample with replacement 2.5 million individuals from the observed commuting matrix 100 times. For each of these 100 samples, we estimate the covariates-based continuum model and compute its counterfactual predictions. Figure 8 shows that the variation in counterfactual changes induced by parameter uncertainty is much smaller than the variation induced by different realizations of idiosyncrasies.

We also investigate the uncertainty stemming from time-varying fundamentals by re-doing the counterfactual analysis using 2008 data (instead of 2010 data). When using the covariates-based parameterization, the counterfactual predictions are not very sensitive to the choice of baseline year: all the real rent changes and more than 95% of the real wage changes produced by the 2008 estimated model are inside the 90% confidence intervals of the predictions from the 2010 estimated model depicted in Figure 8.[45]

As we move from tract-level outcomes to examining larger geographic units, the uncertainty about counterfactual outcomes induced by individual idiosyncrasies diminishes but remains substantial. If we aggregate tract-level changes into 20 bins based on distance to the Amazon HQ2 tract, the mean predicted changes in residents and rents decline with distance to the Amazon HQ2 tract. For all but the first and twentieth ventiles, the 90% confidence interval for the change in residents includes zero (see Figure F.7). For every ventile, the 90% confidence interval for the rent change includes the citywide change. The uncertainty about the change in real aggregate output stemming from individual idiosyncrasies is negligible.[46] When we apply the model with a finite number of individuals to 195 Neighborhood Tabulation Areas rather than tracts, the 90% confidence interval for the change in residents includes zero for 193 of the 195 NTAs. The confidence intervals for the predicted change in real rent include the citywide change for 132 NTAs (see Appendix Figure F.9).

The uncertainty about counterfactual tract-level changes in residents and rents is similar when residential locations are more substitutable and when labor demand is more elastic. Uncertainty is similar when we employ a nested-logit specification in which residential locations in the same NTA are closer substitutes (see Appendix Figure F.10). When labor demand is perfectly elastic ($\sigma = \infty$), wages are invariant to individual idiosyncrasies. Uncertainty about rent changes is similar to the uncertainty in the baseline case ($\sigma = 4$), though the mean rent increase is smaller so more tracts have confidence intervals that include zero (see Appendix Figure F.11).

The predicted consequences of the most closely followed corporate headquarters decision in recent memory for local economic outcomes are subject to considerable uncertainty stemming from

[44] If labor demand were perfectly elastic $\sigma = \infty$, equilibrium wages would be independent of residual labor supply and would exhibit no uncertainty: $w_n \propto A_n$.

[45] By contrast, the calibrated-shares procedure's predictions are much more sensitive to the use of 2008 versus 2010 data. The correlation between the real rent changes shown in Figure 7C and those obtained using the 2008 data is only 0.23 (0.73 for real wages).

[46] The 90% confidence interval for the increase in aggregate output is 0.41% to 0.43%.

individual idiosyncrasies. This notable uncertainty would be missed by the conventional continuum approach. Given the magnitude of uncertainty stemming from individual idiosyncrasies that we document in this case, we believe this uncertainty is likely relevant in many empirical applications that consider the geographic distribution of economic outcomes.

When should a researcher quantify the uncertainty about counterfactual predictions attributable to individual idiosyncrasies? Our results suggest the geographic scale of the predicted changes is key: individual idiosyncrasies induce little uncertainty about aggregate outcomes. When studying spatially fine outcomes, the answer depends on the empirical setting, such as the size and nature of the counterfactual shock and the expected number of individuals choosing a pair of locations. To assess whether they can rule out uncertainty attributable to individual idiosyncrasies, researchers can repeatedly draw finite labor allocations from the continuum model at counterfactual parameters Υ' to compute a distribution of differences between finite-sample counterfactual outcomes and the expected outcome at baseline parameters Υ. This distribution of differences is more dispersed than the distribution of counterfactual changes from the model with a finite number of individuals that fixes the vector of idiosyncratic preferences $\boldsymbol{\nu}^I$.[47] If the dispersion of the former is small, then the uncertainty stemming from individual idiosyncrasies is small. For example, if the distribution of continuum-model differences does not include zero, then the confidence interval from the model with finite individuals excludes zero. When uncertainty attributable to individual idiosyncrasies cannot be ruled out, computing and reporting this uncertainty will produce more informative predictions about local economic outcomes.

6 Conclusion

Economists increasingly use spatially fine data and quantitative spatial models to compute counterfactual general-equilibrium outcomes. The smaller the number of individuals behind each economic outcome reported in these granular settings, the less compelling the conventional modeling assumption that there is a continuum of individuals. These applied general-equilibrium models need to reliably predict economic changes if policymakers are to use them to inform their decisions (Kehoe, 2005; Bryan, Glaeser, and Tsivanidis, 2020; Adão, Arkolakis, and Esposito, 2022).

We document that some conventional methods perform poorly when applied to granular settings. Calibrating pair-specific parameters when there are many pairs of locations relative to the number of decision makers fits the model to the idiosyncratic component of individual decisions. We recommend that researchers use Monte Carlo simulations to verify the finite-sample behavior of their procedures for producing counterfactual predictions in their empirical setting. We show how to use low-rank matrices to capture unobserved spatial linkages without overfitting when computing

[47] This distribution of differences is more dispersed than the distribution of counterfactual changes because an idiosyncratically high outcome at counterfactual parameters Υ' would also be higher at baseline parameters Υ given the same idiosyncratic preferences $\boldsymbol{\nu}^I$. This can be seen by contrasting Figure 9 in Dingel and Tintelnot (2021), which depicted such differences, to Figure 8 above. Sampling from the continuum model is computationally cheaper than computing counterfactual outcomes holding idiosyncratic preferences fixed in the model with a finite number of individuals.

counterfactual outcomes using exact hat algebra.

Counterfactual outcomes for particular locations vary with individual idiosyncrasies. We introduce a model with a finite number of individuals in which equilibrium outcomes depend in part on the idiosyncratic component of individuals' choices. This estimated model yields a distribution of equilibrium outcomes when computing counterfactual scenarios. Evaluating the counterfactual consequences of Amazon's proposed second headquarters in Long Island City, we find that there is considerable uncertainty about most of the predicted changes in local outcomes. Since all empirical settings feature a finite number of individuals, the continuum assumption has been made in the interest of modeling convenience, not realism. The model with a finite number of individuals is tractable and features a covariates-based specification that can be estimated using the same data as its continuum counterpart. Thus, it can be applied in the same settings in which economists have thus far assumed a continuum.

References

Adão, Rodrigo, Arnaud Costinot, and Dave Donaldson. 2024. "Putting Quantitative Models to the Test: An Application to the U.S.-China Trade War*." *The Quarterly Journal of Economics*, 140(2): 1471–1524.

Adão, Rodrigo, Costas Arkolakis, and Federico Esposito. 2022. "General Equilibrium Effects in Space: Theory and Measurement." Working paper.

Ahlfeldt, Gabriel M., Stephen J. Redding, Daniel M. Sturm, and Nikolaus Wolf. 2015. "The Economics of Density: Evidence From the Berlin Wall." *Econometrica*, 83(6): 2127–2189.

Allen, Treb, and Costas Arkolakis. 2014. "Trade and the Topography of the Spatial Economy." *The Quarterly Journal of Economics*, 129(3): 1085–1140.

Allen, Treb, Costas Arkolakis, and Xiangliang Li. 2016. "Optimal City Structure." Working paper.

Allen, Treb, Costas Arkolakis, and Xiangliang Li. 2023. "On the Equilibrium Properties of Network Models with Heterogeneous Agents." NBER Working Paper 27837.

Amazon. 2018. "Amazon selects New York City and Northern Virginia for new headquarters." press release.

Armenter, Roc, and Miklós Koren. 2014. "A Balls-and-Bins Model of Trade." *American Economic Review*, 104(7): 2127–51.

Arzaghi, Mohammad, and J. Vernon Henderson. 2008. "Networking off Madison Avenue." *Review of Economic Studies*, 75(4): 1011–1038.

Bai, Jushan. 2009. "Panel data models with interactive fixed effects." *Econometrica*, 77(4): 1229–1279.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. "High-Dimensional Methods and Inference on Structural and Treatment Effects." *Journal of Economic Perspectives*, 28(2): 29–50.

Berkes, Enrico, and Ruben Gaetani. 2023. "Income Segregation and the Rise of the Knowledge Economy." *American Economic Journal: Applied Economics*, 15(2): 69–102.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa. 2019. "A Distributional Framework for Matched Employer Employee Data." *Econometrica*, 87(3): 699–739.

Brinkman, Jeffrey, and Jeffrey Lin. 2022. "Freeway Revolts! The Quality of Life Effects of Highways." *The Review of Economics and Statistics*, 1–45.

Bryan, Gharad, Edward Glaeser, and Nick Tsivanidis. 2020. "Cities in the Developing World." *Annual Review of Economics*, 12: 273–297.

Carvalho, Vasco M., and Basile Grassi. 2019. "Large Firm Dynamics and the Business Cycle." *American Economic Review*, 109(4): 1375–1425.

Chen, Mingli, Iván Fernández-Val, and Martin Weidner. 2021. "Nonlinear factor models for network and panel data." *Journal of Econometrics*, 220(2): 296–324.

Cohen, Mike X. 2021. *Linear Algebra: Theory, Intuition, Code.* sincXpress.

Conlon, Christopher, Julie Holland Mortimer, and Paul Sarkis. 2023. "Estimating Preferences and Substitution Patterns from Second-Choice Data Alone." Working paper.

Correia, Sergio, Paulo Guimarães, and Tom Zylkin. 2020. "Fast Poisson estimation with high-dimensional fixed effects." *The Stata Journal*, 20(1): 95–115.

Costinot, Arnaud, and Andrés Rodríguez-Clare. 2014. "Trade Theory with Numbers: Quantifying the Consequences of Globalization." Vol. 4 of *Handbook of International Economics*, Chapter 4, 197–261. Elsevier.

Couture, Victor, Jonathan I. Dingel, Allison E. Green, Jessie Handbury, and Kevin R. Williams. 2022. "JUE Insight: Measuring movement and social contact with smartphone data: a real-time application to COVID-19." *Journal of Urban Economics*, 127.

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.

Davis, Morris A., and Francois Ortalo-Magne. 2011. "Household Expenditures, Wages, Rents." *Review of Economic Dynamics*, 14(2): 248–261.

Dekle, Robert, Jonathan Eaton, and Samuel Kortum. 2008. "Global Rebalancing with Gravity: Measuring the Burden of Adjustment." *IMF Staff Papers*, 55(3): 511–540.

di Giovanni, Julian, Andrei A. Levchenko, and Isabelle Mejean. 2014. "Firms, Destinations, and Aggregate Fluctuations." *Econometrica*, 82(4): 1303–1340.

Diamond, Rebecca, Tim McQuade, and Franklin Qian. 2019. "The Effects of Rent Control Expansion on Tenants, Landlords, and Inequality: Evidence from San Francisco." *American Economic Review*, 109(9): 3365–94.

Dingel, Jonathan, Antonio Miscio, and Donald Davis. 2021. "Cities, Lights, and Skills in Developing Economies." *Journal of Urban Economics*, 125(103174).

Dingel, Jonathan I., and Felix Tintelnot. 2021. "Spatial Economics for Granular Settings." Working paper.

Donaldson, Dave, and Adam Storeygard. 2016. "The View from Above: Applications of Satellite Data in Economics." *Journal of Economic Perspectives*, 30(4): 171–98.

Eaton, Jonathan, Samuel Kortum, and Sebastian Sotelo. 2013. "International Trade: Linking Micro and Macro." In *Advances in Economics and Econometrics: Tenth World Congress*. Vol. 3. Cambridge University Press.

Eldén, Lars. 2007. *Matrix Methods in Data Mining and Pattern Recognition.* Society for Industrial and Applied Mathematics.

Ellison, Glenn, and Edward L Glaeser. 1997. "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach." *Journal of Political Economy*, 105(5): 889–927.

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.

Gabaix, Xavier. 2011. "The Granular Origins of Aggregate Fluctuations." *Econometrica*, 79(3): 733–772.

Galichon, Alfred. 2022. "On the representation of the nested logit model." *Econometric Theory*, 38(2): 370–380.

Gandhi, Amit, Zhentong Lu, and Xiaoxia Shi. 2023. "Estimating demand for differentiated products with zeroes in market share data." *Quantitative Economics*, 14(2): 381–418.

Gaubert, Cecile, and Oleg Itskhoki. 2021. "Granular Comparative Advantage." *Journal of Political Economy*, 129(3).

Goodman, J. David. 2019. "Why Amazon Is Caught in an Unexpected Brawl in New York." *New York Times*, 12 February.

Graham, Matthew R., Mark J. Kutzbach, and Brian McKenzie. 2014. "Design Comparison of LODES and ACS Commuting Data Products." Center for Economic Studies, U.S. Census Bureau Working Papers 14-38.

Guimarães, Paulo, Octávio Figueiredo, and Douglas Woodward. 2003. "A Tractable Approach to the Firm Location Decision Problem." *The Review of Economics and Statistics*, 85(1): 201–204.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The elements of statistical learning: data mining, inference, and prediction.* Vol. 2, Springer.

Hastie, Trevor, Robert Tibshirani, and Martin Wainwright. 2015. *Statistical Learning with Sparsity: The Lasso and Generalizations.* CRC Press.

Heblich, Stephan, Stephen J Redding, and Daniel M Sturm. 2020. "The Making of the Modern Metropolis: Evidence from London." *The Quarterly Journal of Economics*, 135(4): 2059–2133.

Hodrick, Robert J., and Edward C. Prescott. 1997. "Postwar U.S. Business Cycles: An Empirical Investigation." *Journal of Money, Credit and Banking*, 29(1): 1–16.

Holmes, Thomas J., and Holger Sieg. 2015. "Structural Estimation in Urban Economics." *Handbook of Regional and Urban Economics*, Vol. 5, 69–114. Elsevier.

Hortaçsu, Ali, Olivia R. Natan, Hayden Parsley, Timothy Schwieg, and Kevin R. Williams. 2023. "Demand Estimation with Infrequent Purchases and Small Market Sizes." *Quantitative Economics*, 14(4): 1251–1294.

Hsiao, Allan. 2022. "Coordination and Commitment in International Climate Action: Evidence from Palm Oil." Working paper.

Hsieh, Cho-Jui, and Inderjit S Dhillon. 2011. "Fast coordinate descent methods with variable selection for non-negative matrix factorization." Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1064–1072.

Kalouptsidi, Myrto. 2014. "Time to Build and Fluctuations in Bulk Shipping." *American Economic Review*, 104(2): 564–608.

Kehoe, Timothy J. 2005. "An Evaluation of the Performance of Applied General Equilibrium Models on the Impact of NAFTA." In *Frontiers in Applied General Equilibrium Modeling: In Honor of Herbert Scarf*. 341–377. Cambridge University Press.

Kehoe, Timothy J., Pau S. Pujolàs, and Jack Rossbach. 2017. "Quantitative Trade Models: Developments and Challenges." *Annual Review of Economics*, 9(1): 295–325.

Krebs, Oliver, and Michael P. Pflüger. 2019. "On the Road (Again): Commuting and Local Employment Elasticities in Germany." IZA Discussion Papers 12257.

Kreindler, Gabriel, and Yuhei Miyauchi. 2023. "Measuring Commuting and Economic Activity inside Cities with Cell Phone Records." *The Review of Economics and Statistics*, , (WP2020-006): 1–48.

Krusell, Per, and Anthony A. Smith. 1998. "Income and Wealth Heterogeneity in the Macroeconomy." *Journal of Political Economy*, 106(5): 867–896.

McFadden, Daniel L. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics*. Ed. P. Zarembka, 105–142. Academic Press.

McFadden, Daniel L. 1978. "Modelling the Choice of Residential Location." In *Spatial Interaction Theory and Planning Models*. North Holland.

Mogstad, Magne, Joseph P Romano, Azeem M Shaikh, and Daniel Wilhelm. 2024. "Inference for Ranks with Applications to Mobility across Neighbourhoods and Academic Achievement across Countries." *The Review of Economic Studies*, 91(1): 476–518.

Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg. 2018. "Commuting, Migration, and Local Employment Elasticities." *American Economic Review*, 108(12): 3855–90.

Owens, Raymond, III, Esteban Rossi-Hansberg, and Pierre-Daniel Sarte. 2020. "Rethinking Detroit." *American Economic Journal: Economic Policy*, 12(2): 258–305.

Panigrahi, Piyush. 2022. "Endogenous Spatial Production Networks: Quantitative Implications for Trade and Productivity." Working paper.

Perez-Cervantes, Fernando. 2016. "Insurance Against Local Productivity Shocks: Evidence from Commuters in Mexico." Banco de México Working Papers 2016-19.

Proost, Stef, and Jacques-François Thisse. 2019. "What Can Be Learned from Spatial Economics?" *Journal of Economic Literature*, 57(3): 575–643.

Quan, Thomas W., and Kevin R. Williams. 2018. "Product variety, across-market demand heterogeneity, and the value of online retail." *The RAND Journal of Economics*, 49(4): 877–913.

Redding, Stephen J., and Esteban Rossi-Hansberg. 2017. "Quantitative Spatial Economics." *Annual Review of Economics*, 9(1): 21–58.

Ridout, Martin S. 2009. "Generating random numbers from a distribution specified by its Laplace transform." *Statistics and Computing*, 19: 439–450.

Rosenthal, Stuart S., and William C. Strange. 2020. "How Close Is Close? The Spatial Reach of Agglomeration Economies." *Journal of Economic Perspectives*, 34(3): 27–49.

Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Raymond Owens. 2010. "Housing Externalities." *Journal of Political Economy*, 118(3): 485–535.

Rutherford, Thomas F. 1995. "Constant Elasticity of Substitution Functions: Some Hints and Useful Formulae." Notes prepared for GAMS General Equilibrium Workshop held December, 1995 in Boulder Colorado.

Sanders, Bas. 2024. "Counterfactual Sensitivity in Quantitative Trade and Spatial Models."

Schoefer, Benjamin, and Oren Ziv. 2022. "Productivity, Place, and Plants." *The Review of Economics and Statistics*, 1–46.

Severen, Christopher. 2021. "Commuting, Labor, and Housing Market Effects of Mass Transportation: Welfare and Identification." *The Review of Economics and Statistics*, 1–99.

Silva, J. M. C. Santos, and Silvana Tenreyro. 2006. "The Log of Gravity." *The Review of Economics and Statistics*, 88(4): 641–658.

Sotelo, Sebastian. 2019. "Practical Aspects of Implementing the Multinomial PML Estimator." Working paper.

Train, Kenneth. 2009. *Discrete Choice Methods with Simulation.* Cambridge University Press.

Tsivanidis, Nick. 2023. "Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogota's TransMilenio." Working paper.

Udell, Madeleine, and Alex Townsend. 2019. "Why Are Big Data Matrices Approximately Low Rank?" *SIAM Journal on Mathematics of Data Science*, 1(1): 144–160.

Waddell, Sonya Ravindranath, and Pierre Daniel Sarte. 2016. "From Stylized to Quantitative Spatial Models of Cities." *Economic Quarterly*, 169–196.

Weintraub, Gabriel Y., C. Lanier Benkard, and Benjamin Van Roy. 2008. "Markov Perfect Industry Dynamics With Many Firms." *Econometrica*, 76(6): 1375–1411.

Zárate, Román David. 2023. "Spatial Misallocation, Informality and Transit Improvements: Evidence from Mexico City." Working paper.

Appendix

A Monte Carlo simulations

We conduct a Monte Carlo simulation to examine the finite-data performance of the covariates-based approach and the calibrated-shares procedure. In these simulations, the data-generating process is our estimated covariates-based model of New York City in 2010. We impose a counterfactual 9% increase in productivity ($\hat{A}_n = 1.09$) for the "treated" tract containing 200 Fifth Avenue, which generates an increase in employment matching the observed change from 2010 to 2012.[48] There are no changes in the productivity of other workplace tracts ($\hat{A}_n = 1$), no changes in land endowments ($\hat{T}_k = 1\ \forall k$), and no changes in commuting costs ($\hat{\bar{\delta}}_{kn} = 1\ \forall k, n$). In the limiting case, as $I \to \infty$, the calibrated-shares procedure would perfectly describe the changes in commuting flows associated with this productivity increase. Thus, any predictive failure when drawing finite-sample realizations from this data-generating process are due to problems stemming from the finite number of individuals.

A.1 Monte Carlo exercise: Continuum changes

In the continuum model with parameter values estimated in Section 3.2, the "true" change in outcomes is given by combining the relative changes in exogenous parameters and the baseline shares using equations (5)–(7).

We compute predicted changes using the covariates-based model and the calibrated-shares procedures in 100 simulations of finite-sample data. In each of the 100 simulations, we implement the following steps:

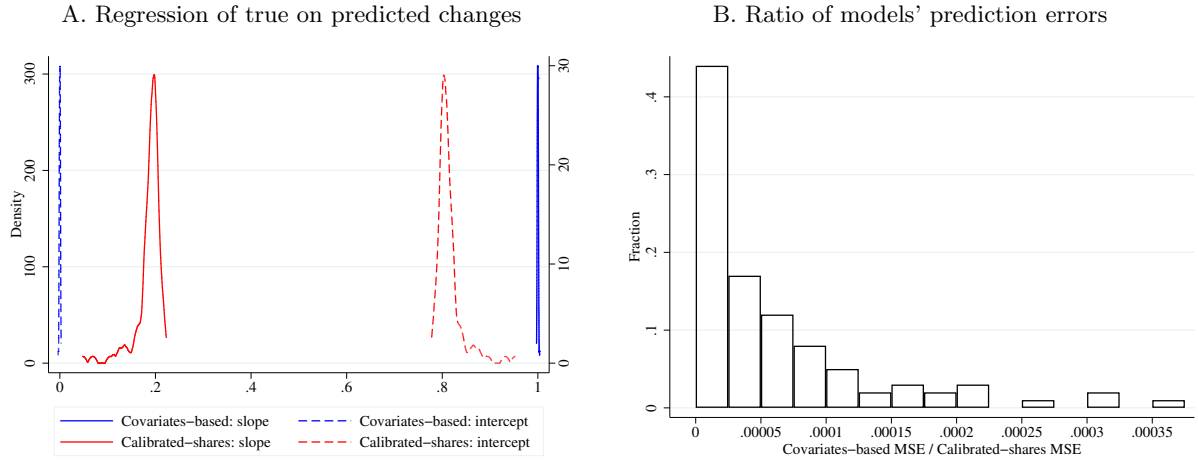
1. To generate "observed" baseline data, draw a finite-sample realization of the labor allocation from the parameterized continuum model.
2. Estimate ϵ and other parameters of the covariates-based model using the "observed" realization. Compute baseline shares ℓ_{kn} and y_{kn} for the covariates-based model using these parameter estimates. Compute baseline shares ℓ_{kn} and $y_{kn} = \ell_{kn} w_n / \bar{\delta}_{kn}$ for the calibrated-shares procedure using the finite-sample realization of ℓ_{kn} drawn in the previous step and the "true" values of w_n and $\bar{\delta}_{kn}$ used in the data-generating process.
3. Using equations (5)–(7), compute the increase in productivity required to match the true change in employment for the "treated" tract for both the covariates-based approach and the calibrated-shares procedure.
4. Compute the predicted change in commuter counts ($\ell_{kn}(A') - \ell_{kn}(A)$) and rents ($\hat{r}_k / \hat{P} - 1$) using both the covariates-based approach and the calibrated-shares procedure.

[48] In 2011, Tiffany & Co. moved its corporate headquarters to 260,000 square feet of office space at 200 Fifth Avenue.

5. Regress the true change in commuters destined for the treated tract on the changes predicted by the covariates-based approach. Regress the true change in commuters destined for the treated tract on the changes predicted by the calibrated-shares procedure. Regress the true change in real rents on the changes predicted by the covariates-based approach and calibrated-shares procedure, respectively. Compute the mean squared error of each set of predictions.

The results of these 100 simulations are reported in Figure 2, Panel A of Table 2, Figure A.1, and Table A.1. Figure A.1 shows that the calibrated-shares procedure predicts changes in rents poorly with the number of individuals in the data. Table A.1 shows that the slope coefficient becomes close to one when the number of individuals is two or three orders of magnitude larger.

Figure A.1: Monte Carlo: Calibrated-shares procedure predict rents poorly



NOTES: This figure describes predicted rent changes from the same simulations that produced the predicted commuter changes in Figure 2 and Panel A of Table 2. This figure depicts the regression coefficients and mean squared errors from 100 simulations in which $I = 2,488,905$, the data-generating process is the estimated covariates-based model of New York City in 2010, and the counterfactual change is an 9% increase in productivity in one workplace tract. The predicted change in real rents is compared to the continuum-model change in real rents.

Table A.1: Calibrated-shares procedure's finite-sample performance predicting rent changes

I	2.5	5	12.5	25	50	125	250	2560
Calibrated-shares: slope	0.187	0.312	0.535	0.695	0.822	0.918	0.959	0.996
Calibrated-shares: intercept	0.814	0.689	0.465	0.306	0.178	0.082	0.041	0.004
Calibrated-shares: MSE	444.852	219.325	85.426	42.944	21.375	8.708	4.310	0.420

NOTES: This table describes predicted rent changes from the same simulations that produced the predicted commuter changes in Panel A of Table 2.

A.2 Monte Carlo exercise: Finite-sample changes

In any empirical application, the researcher does not observe changes in outcomes for a continuum of individuals. The observed changes in outcomes come from finite-sample draws from the pre- and

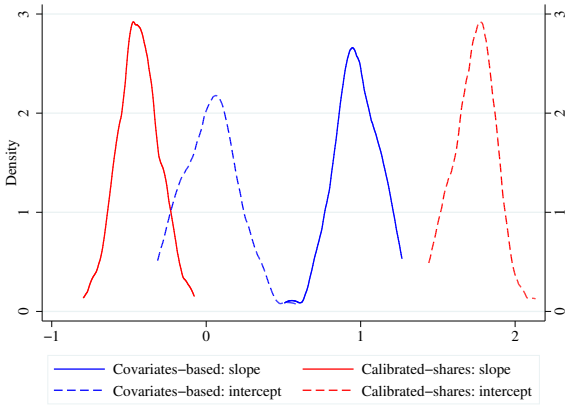
post-shock data-generating processes. We simulate this setting by drawing two labor-allocation realizations for each simulation, the first from the estimated model of New York City in 2010 and the second from that model with 9% higher productivity for the treated workplace tract. For each simulation, we execute six steps. The first four are identical to those described in the previous section, Appendix A.1. The last two are

5. Generate "observed" post-shock data by drawing a finite-sample realization of the labor allocation from the continuum model at the counterfactual productivity vector, which has 9% higher productivity for the treated workplace tract. Compute "observed" changes in commuter counts by subtracting the step-1 baseline realization of commuter counts from these realized post-shock commuter counts.
6. Regress the "observed" changes in commuter counts destined for the treated workplace tract on the changes predicted by the calibrated-shares procedure. Regress the "observed" changes in commuter counts destined for the treated workplace tract on the changes predicted by the covariates-based approach.

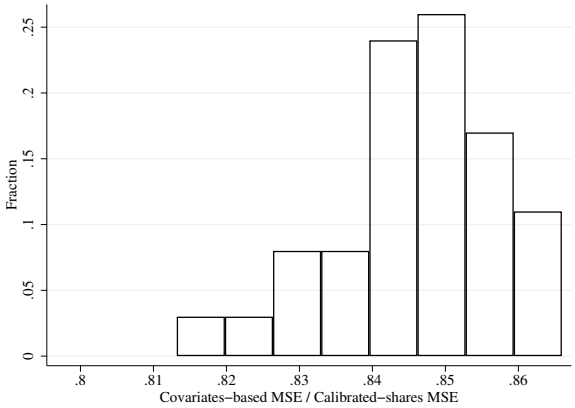
The results of these simulations are reported in Panel B of Table 2 and Figure A.2. The additional finite-sample noise raises the MSEs for both methods relative to the simulations described in Section A.1.

Figure A.2: Calibrated-shares procedure overfits in finite-sample simulations

A. Regression of finite-sample changes on predicted changes



B. Ratio of models' prediction errors



NOTES: This figure depicts the regression coefficients and MSEs from 100 simulations in which $I = 2,488,905$, the data-generating process is a finite-sample realization from the estimated covariates-based model of New York City in 2010, and a finite-sample realization from that model with a 9% increase in productivity in one workplace tract. "Covariates-based" means applying the parameterization described in Section 3.2 to the simulated data. "Calibrated-shares" means using the observed (simulated) shares in the exact hat algebra defined by equations (5)–(7).

A.3 Monte Carlo simulations with unobserved disutility, $\lambda_{kn} \neq 1$

Next, we consider a data-generating process in which commuting costs have both an observed and unobserved component in order to examine the relative predictive performance of the covariates-

based model and calibrated-shares procedure in granular settings. The commuting costs are now determined by

$$\ln \delta_{kn} = \ln \bar{\delta}_{kn} + \ln \lambda_{kn}$$
$$\ln \lambda_{kn} \overset{\text{iid}}{\sim} \mathcal{N}\left(0, \Lambda^2 \times \text{Var}(\ln \bar{\delta}_{kn})\right), \quad \Lambda \in \mathbb{R}^+.$$

The parameter Λ scales the standard deviation of the unobserved component of commuting costs, stated relative to the standard deviation of the observed component. The simulations shown in Table 2 are the case in which $\Lambda = 0$, so that $\lambda_{kn} = 1 \ \forall k, n$. When $\Lambda > 0$, the covariates-based model omits an unobserved component of commuting costs that the calibrated-shares procedure may, in principle, capture. Since $\ln \bar{\delta}_{kn}$ and $\ln \lambda_{kn}$ are orthogonal by assumption, the covariates-based model's predictions will be unbiased, but its forecast errors will increase with the magnitude of unobserved variation Λ. For $\Lambda = 0.1$, 0.25, 0.5, and 1.0, we compute 100 simulations using 100 realizations of the λ_{kn} matrix.

Tables A.2 and A.3 present the results of these simulations for continuum-model changes and finite-sample changes, respectively. As established in the $\Lambda = 0$ case, which appears in Table 2, the calibrated-shares procedure is highly data demanding. For all values of Λ reported in Tables A.2 and A.3, the calibrated-shares procedure's predictions become unbiased (slope near one) only as the number of individuals becomes orders of magnitude larger than the true value. Since the calibrated-shares procedure can capture the unobserved component of commuting costs, its mean forecast error becomes smaller than that of the covariates-based model when the unobserved variation is sufficiently large. For the number of individuals in the empirical setting ($I = 2,488,905$), this occurs only when the unobserved component of commuting costs varies at least a quarter as much as the observed component ($\Lambda \geq 0.25$) when we evaluate continuum-model changes and more than half as much as the observed component ($\Lambda > 0.5$) when we evaluate finite-sample changes.

A.4 Monte Carlo simulations: Finite-sample changes with fixed ν^I

This appendix revisits the Monte Carlo investigation of Section 3.4 using a data-generating process in which the economic primitives change from Υ to Υ' but the set of idiosyncratic preferences ν^I is fixed. This leverages the model introduced in Section 4 as the data-generating process. Contrasted with the finite-sample outcomes with iid-ν idiosyncrasies, these finite-sample simulations are substantially closer to the continuum-change outcomes.

Figure A.3 is the fixed-ν analogue of Figure 2, depicting the distributions of the regression coefficients and relative MSE for the covariates-based model and calibrated-shares procedure. The covariates-based model outperforms the calibrated-shares procedure. The regression coefficient results are very similar to Figure 2. The covariates-based model remains an unbiased predictor, while the calibrated-shares procedure has a median slope of 0.79. The ratio of mean squared errors is closer to one in these simulations because the noisy finite-sample outcomes mean that both approaches have larger forecast errors than in the continuum-outcomes case. The MSE ratio in the median simulation is about 84%.

Table A.2: Monte Carlo simulations with $\lambda_{kn} \neq 1$: Regressand is continuum change

		Slope (mean)		MSE (mean)	
Λ	I	Covariates-based	Calibrated-shares	Covariates-based	Calibrated-shares
0	2.5	0.9984	0.7793	0.0014	0.2280
0	5	0.9992	0.8759	0.0007	0.1133
0	12.5	0.9997	0.9483	0.0003	0.0451
0	25	0.9998	0.9725	0.0001	0.0227
0	50	0.9999	0.9868	0.0001	0.0113
0	125	1.0000	0.9951	0.0000	0.0045
0	250	1.0000	0.9973	0.0000	0.0023
0	2560	1.0000	0.9997	0.0000	0.0002
0.1	2.5	0.9968	0.7869	0.0380	0.2269
0.1	5	0.9977	0.8810	0.0373	0.1129
0.1	12.5	0.9983	0.9501	0.0370	0.0453
0.1	25	0.9984	0.9731	0.0368	0.0226
0.1	50	0.9985	0.9865	0.0368	0.0113
0.1	125	0.9986	0.9944	0.0367	0.0045
0.1	250	0.9986	0.9971	0.0367	0.0023
0.1	2560	0.9986	0.9996	0.0367	0.0002
0.25	2.5	0.9957	0.8199	0.2405	0.2262
0.25	5	0.9961	0.9025	0.2399	0.1129
0.25	12.5	0.9969	0.9581	0.2395	0.0454
0.25	25	0.9971	0.9787	0.2393	0.0226
0.25	50	0.9971	0.9889	0.2393	0.0113
0.25	125	0.9972	0.9954	0.2393	0.0045
0.25	250	0.9972	0.9978	0.2393	0.0023
0.25	2560	0.9972	0.9997	0.2392	0.0002
0.5	2.5	0.9954	0.8923	1.1084	0.2271
0.5	5	0.9962	0.9451	1.1076	0.1130
0.5	12.5	0.9969	0.9782	1.1071	0.0448
0.5	25	0.9971	0.9877	1.1071	0.0223
0.5	50	0.9972	0.9944	1.1070	0.0113
0.5	125	0.9973	0.9980	1.1069	0.0045
0.5	250	0.9973	0.9986	1.1069	0.0022
0.5	2560	0.9973	0.9999	1.1069	0.0002
1	2.5	1.0057	0.9691	7.0808	0.2186
1	5	1.0063	0.9847	7.0806	0.1090
1	12.5	1.0065	0.9946	7.0809	0.0435
1	25	1.0069	0.9958	7.0802	0.0219
1	50	1.0068	0.9985	7.0805	0.0109
1	125	1.0070	0.9992	7.0803	0.0044
1	250	1.0071	0.9994	7.0803	0.0022
1	2560	1.0070	0.9998	7.0803	0.0002

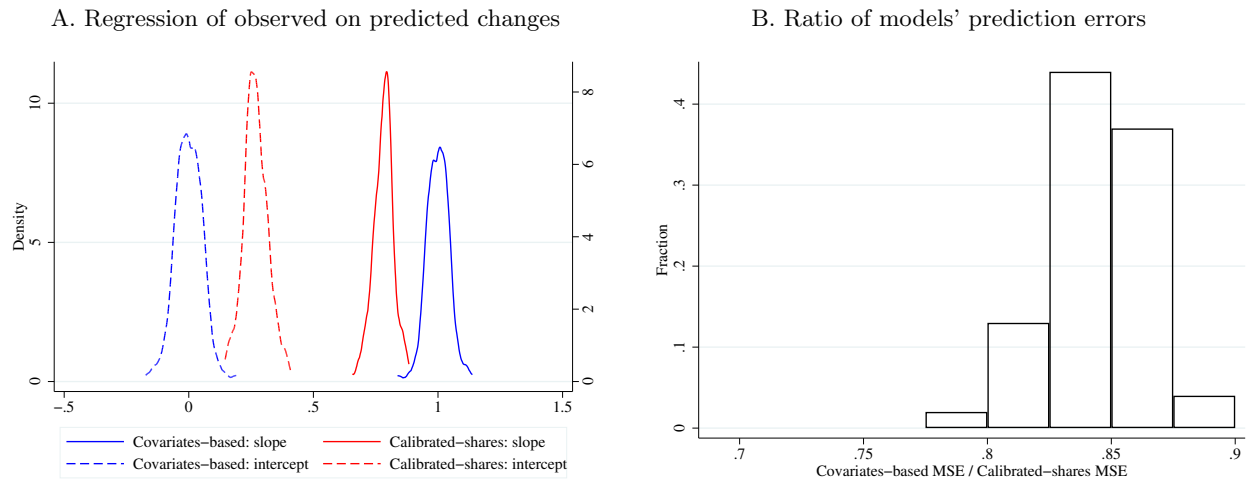
NOTES: This table reports the mean outcomes across 100 simulated event studies for both the covariates-based model and the calibrated-shares procedure. The data-generating process is the covariates-based model estimated on data for New York City in 2010 augmented by an unobserved component of commuting costs of magnitude Λ (see text). The regressand is the continuum change in commuters. The value of I is stated in millions of individuals. The value labeled "2.5" million is in fact $I = 2,488,905$. The $\Lambda = 0$ results also appear in Panel A of Table 2.

Table A.3: Monte Carlo simulations with $\lambda_{kn} \neq 1$: Regressand is finite-sample change

		Slope (mean)		MSE (mean)	
Λ	I	Covariates-based	Calibrated-shares	Covariates-based	Calibrated-shares
0	2.5	0.9744	-0.4372	14.4853	17.1252
0	5	0.9981	0.2281	7.1769	8.5177
0	12.5	1.0060	0.6695	2.8879	3.4181
0	25	0.9939	0.8216	1.4331	1.6968
0	50	1.0035	0.9169	0.7146	0.8463
0	125	1.0034	0.9687	0.2872	0.3402
0	250	1.0007	0.9832	0.1432	0.1696
0	2560	0.9997	0.9979	0.0140	0.0166
0.1	2.5	0.9938	-0.3548	14.5123	17.1386
0.1	5	0.9901	0.2324	7.2340	8.5081
0.1	12.5	1.0004	0.6782	2.9092	3.4009
0.1	25	0.9994	0.8340	1.4768	1.7056
0.1	50	0.9978	0.9148	0.7557	0.8522
0.1	125	0.9932	0.9606	0.3234	0.3398
0.1	250	0.9977	0.9813	0.1795	0.1704
0.1	2560	0.9983	0.9979	0.0505	0.0164
0.25	2.5	0.9709	-0.1494	14.6285	17.0442
0.25	5	1.0028	0.3911	7.4322	8.5047
0.25	12.5	0.9930	0.7333	3.1229	3.4121
0.25	25	0.9940	0.8639	1.6771	1.6987
0.25	50	0.9915	0.9270	0.9570	0.8495
0.25	125	0.9947	0.9693	0.5244	0.3405
0.25	250	0.9987	0.9875	0.3820	0.1687
0.25	2560	0.9968	0.9977	0.2524	0.0166
0.5	2.5	0.9973	0.3116	15.6180	17.1638
0.5	5	1.0082	0.6617	8.4153	8.5620
0.5	12.5	1.0039	0.8589	4.0031	3.4027
0.5	25	0.9925	0.9212	2.5287	1.6891
0.5	50	0.9990	0.9616	1.8189	0.8509
0.5	125	0.9991	0.9883	1.3973	0.3351
0.5	250	0.9950	0.9911	1.2505	0.1701
0.5	2560	0.9968	0.9989	1.1195	0.0166
1	2.5	0.9945	0.7774	20.9828	17.1174
1	5	1.0069	0.9059	14.2812	8.4447
1	12.5	1.0080	0.9710	10.3390	3.3886
1	25	1.0083	0.9688	8.4068	1.7038
1	50	1.0082	0.9929	7.8493	0.8545
1	125	1.0054	0.9971	7.4572	0.3435
1	250	1.0059	0.9976	7.1958	0.1702
1	2560	1.0075	0.9996	7.0915	0.0163

NOTES: This table reports the mean outcomes across 100 simulated event studies for both the covariates-based model and the calibrated-shares procedure. The data-generating process is the covariates-based model estimated on data for New York City in 2010 augmented by an unobserved component of commuting costs of magnitude Λ (see text). The regressand is the simulation-specific change in commuters from a realization drawn from the model using pre-shock parameter values to a realization drawn from the model using post-shock parameter values. The value of I is stated in millions of individuals. The value labeled "2.5" million is in fact $I = 2,488,905$. The $\Lambda = 0$ results also appear in Panel B of Table 2.

Figure A.3: Calibrated-shares procedure performs poorly in fixed-ν^I simulations



NOTES: This figure depicts the regression coefficients and MSEs from 100 simulations in which $I = 2,488,905$, the data-generating process is the estimated covariates-based model of New York City in 2010 with fixed idiosyncratic preferences ν^I, and the counterfactual change is a 9% increase in productivity in one workplace tract.

Fixing the idiosyncratic preferences reduces the excess churn in simulations with iid ν draws: across 100 simulations, the standard deviation of the total employment increase in the "treated tract" is about 185 in the iid-ν case and 51 in the fixed-ν case. Eliminating excess churn reduces the forecast error for both approaches, but their relative predictive performance is similar to the iid case. Across 100 simulations, the covariates-based model's average MSEs with iid and fixed idiosyncratic shocks are 14.49 and 1.23, respectively. The calibrated-shares procedure's average MSEs are 17.13 and 1.45, respectively.

B Event studies

This appendix presents additional details regarding the event studies examined in Section 3.5.

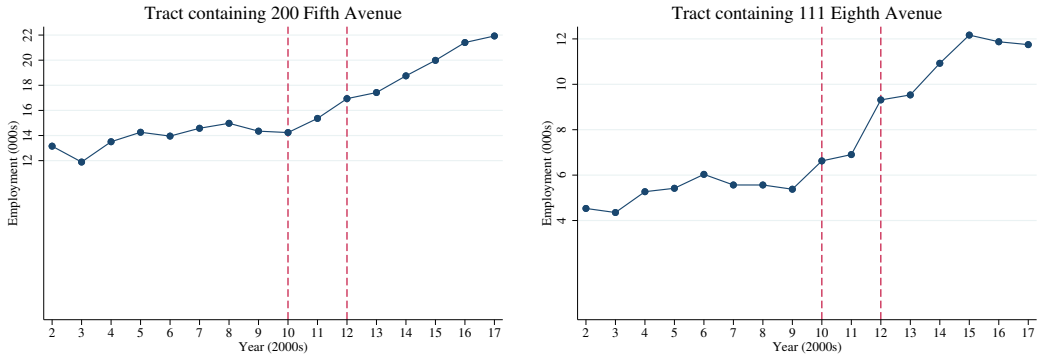
B.1 Two employment booms

Figure B.1 depicts the time series of total employment in two New York City census tracts that contain 200 Fifth Avenue and 111 Eighth Avenue. Between 2010 and 2012, Tiffany & Co. and Google both substantially increased their employment in these two tracts. These tracts are two of the 83 examined in the event studies in Section 3.5.

B.2 Event studies using temporal and geographic aggregation

Figure B.2 present the results of pooling multiple years of data and aggregating the predicted changes in number of commuters to the booming workplace tracts from residential tracts up to residential NTAs.

Figure B.1: Employment in two of the event-study tracts



NOTES: This figure depicts the number of primary jobs held by New York City residents in tracts 36061005800 and 36061008300 in the LODES data.

B.3 Estimating the model using larger spatial units

One may try to avoid the calibrated-shares procedure's overfitting problem by analyzing larger spatial units. Studying New York City's 195 Neighborhood Tabulation Areas (NTAs) instead of tracts reduces the number of locations by an order of magnitude and the number of location pairs by two orders. To do so, we aggregate tract-to-tract commuter counts up to NTA-to-NTA flows, compute NTA-to-NTA commuting costs as the average of tract-to-tract commuting costs $\bar{\delta}_{kn}$, and compute NTA-level wages using employment-weighted averages of tract-level wages. We use both the covariates-based model and the calibrated-shares procedure to predict changes in NTA-to-NTA commuting flows for the 35 NTAs that had 2010-2012 employment growth of at least 12.5% and at least 2,000 employees in 2010.

When analyzing these substantially larger spatial units, the covariates-based approach and the calibrated-shares procedure perform similarly. Figure B.3 shows that the covariates-based model's median slope coefficient is closer to one. The covariates-based model has a lower MSE than the calibrated-shares procedure in 18 of the 35 events.

B.4 Noise reduction via low-rank approximation

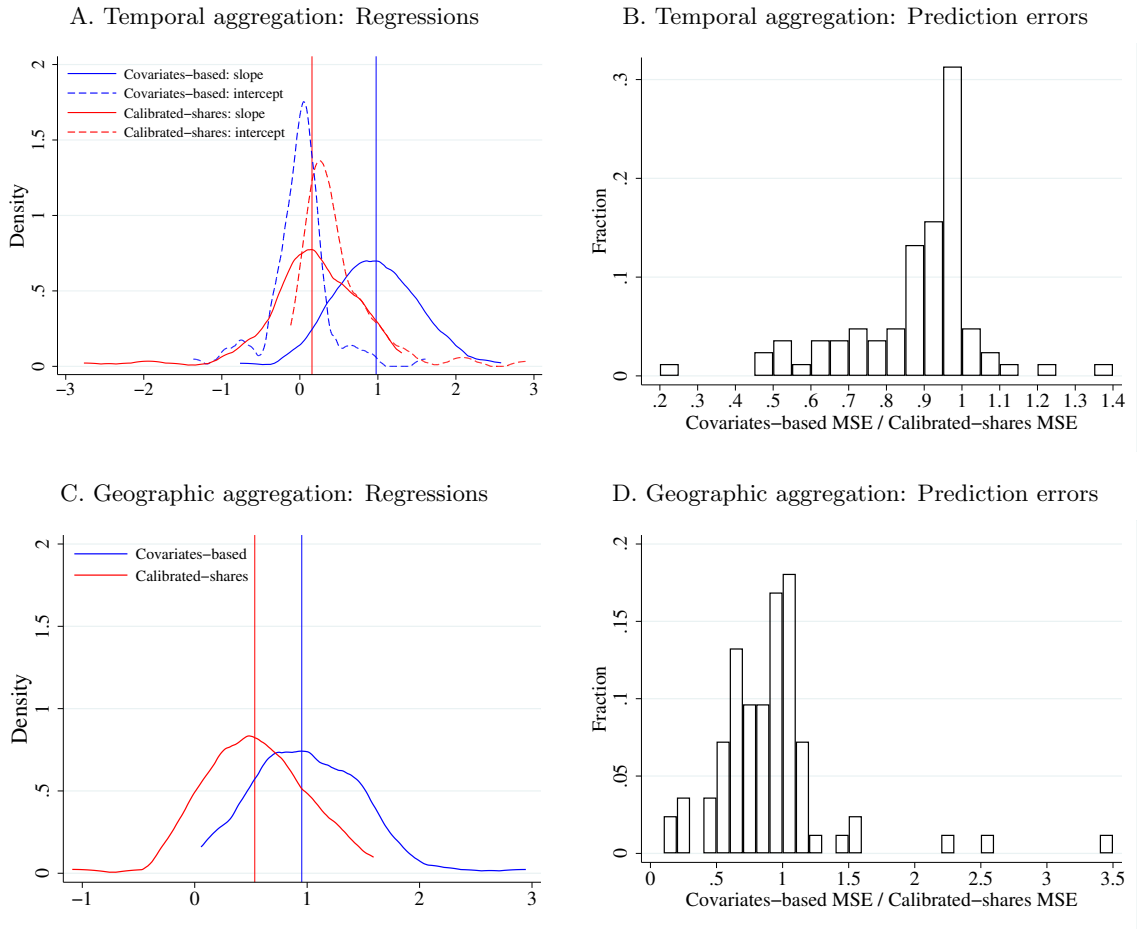
B.4.1 The approximation procedure

One method for guarding against idiosyncratic noise in the observed commuting matrix is to use values of ℓ_{kn} derived from a low-rank approximation of the matrix rather than the observed values when plugging ℓ_{kn} and $y_{kn} = \ell_{kn} w_n / \bar{\delta}_{kn}$ into equations (5)–(7). That is, for some particular rank $\mathfrak{r}$, replace the commuting matrix $\mathbf{L} = [\ell_{kn}]$ with

$$\tilde{\mathbf{L}} = \arg \min_{\mathbf{M} \in \mathbf{R}^{K \times N}} \|\mathbf{L} - \mathbf{M}\|_F^2, \quad \text{subject to} \quad \text{rank}(\mathbf{M}) \leq \mathfrak{r}, \tag{B.1}$$

where $\|\cdot\|_F$ is the Frobenius matrix norm. The solution to this problem can be computed from the singular value decomposition (SVD) of the original matrix $\mathbf{L}$. In particular, the optimal approximation $\tilde{\mathbf{L}}$ is given by the following algorithm (Hastie, Tibshirani, and Wainwright, 2015):

Figure B.2: Models' predictive performance with temporal and geographic aggregation



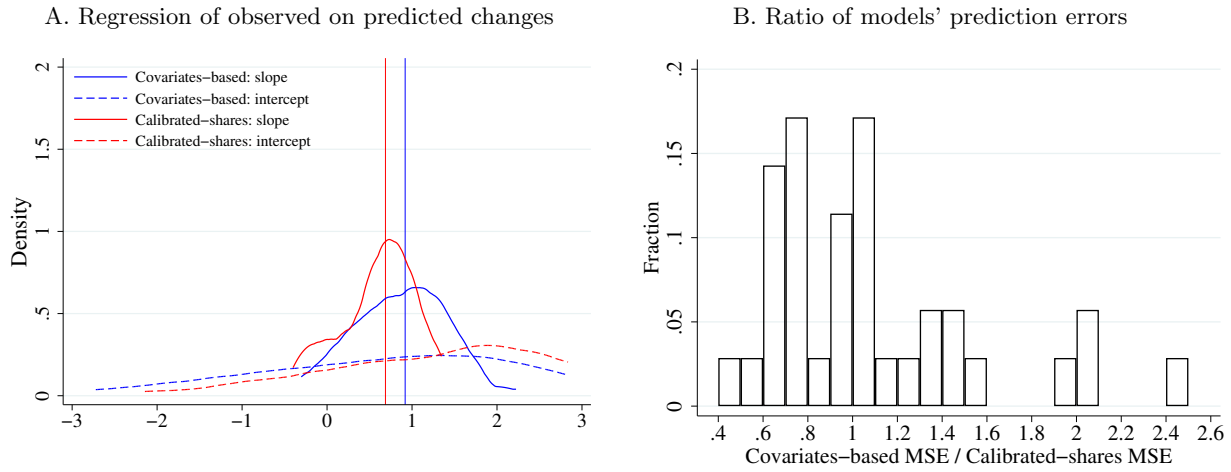
NOTES: This figure contrasts the covariates-based model and calibrated-shares procedure's predictions for 83 tract-level employment booms when employing temporal aggregation or geographic aggregation. Panels A and B depict results when using baseline data for 2008-2010 instead of only 2010. Panels C and D depict predictive performance after aggregating the tract-level predictions underlying Panels A and B of Figure 3 up to 195 Neighborhood Tabulation Areas. We sum the predicted change in commuters to the booming workplace tract across all residential tracts in an NTA. In panel A,
In Panel C, only the slope coefficients are depicted. Vertical lines depict the medians of the slope coefficient distributions.

- Compute the SVD of $\mathbf{L} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, where $\mathbf{U}$ is a $K \times K$ orthonormal matrix, $\mathbf{\Sigma}$ is a non-negative $K \times N$ diagonal matrix with diagonal elements representing the singular values sorted in descending order, and $\mathbf{V}^T$ is an $N \times N$ orthonormal matrix.

- Keep only the first (largest) $\mathfrak{r}$ singular values in Σ, to obtain $\tilde{\mathbf{L}} := U \begin{pmatrix} \Sigma_{\mathfrak{r}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} V^T$, where $\Sigma_{\mathfrak{r}}$ is the upper left $\mathfrak{r} \times \mathfrak{r}$ block in Σ.

However, this approximation is imperfect for our purposes on two accounts: first, while the original matrix $\mathbf{L}$ is non-negative (as it consists of counts), there is no guarantee that the rank-$\mathfrak{r}$ approximation is also non-negative. We therefore set all negative entries in $\tilde{\mathbf{L}}$ to 0. Additionally, $\tilde{\mathbf{L}}$ must be rescaled so that the sum across its entries matches the original; that is, we obtain $\tilde{\mathbf{L}}^{\text{SVD}}_{\mathfrak{r}}$

Figure B.3: Models' predictive performance across 35 NTA-level employment booms



NOTES: This figure depicts predictive performance when fitting both approaches using data aggregated up to the 195 Neighborhood Tabulation Areas defined by the New York City Department of City Planning. In panel A, 10 intercept coefficients are not depicted for the covariates-based model. 5 intercept coefficients are not depicted for the calibrated-shares procedure. Panel A depicts the distributions of slope coefficients, and the vertical lines denote the median slope coefficients.

by replacing negative entries in the rank-$\mathfrak{r}$ SVD approximation with zeros, then rescaling such that $\sum_{k,n}\left(\tilde{\mathbf{L}}_{\mathfrak{r}}^{\text{SVD}}\right)_{kn} = \sum_{k,n}\ell_{kn}$.

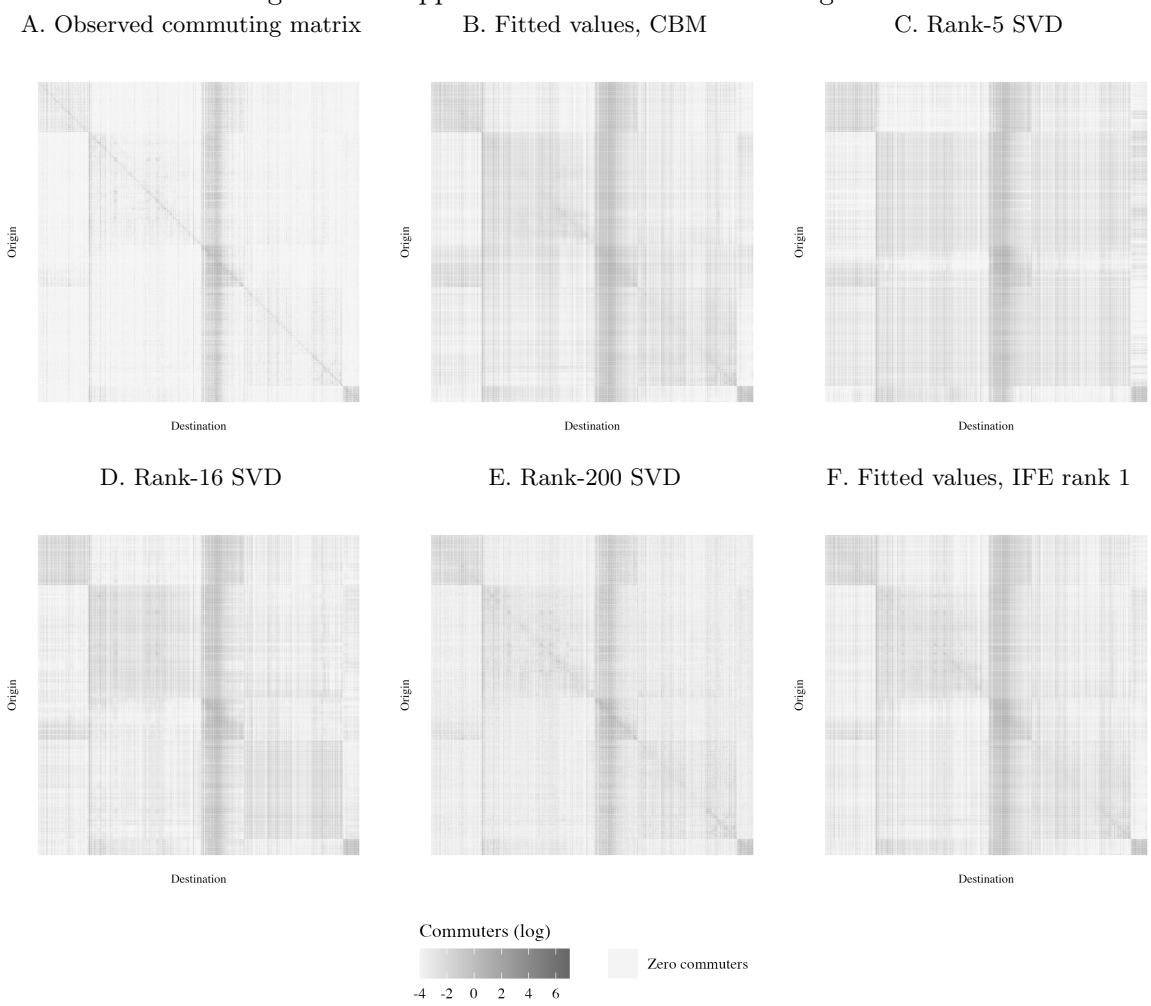
Figure B.4 visualizes the approximated matrices derived from three choices of rank. alongside the observed 2010 data and the fitted values from the covariates-based and rank-1 interactive-fixed-effects specifications. (The IFE procedure is described in detail in Appendix B.5.) The SVD-approximated commuting matrices preserve many features of the observed commuting matrix even under fairly stringent rank restrictions. In particular, workplaces with a large number of commuters are preserved, as are clusters of tracts with comparatively high or low commuting shares. Some features, however, are discarded by the approximation. Most notably, the higher values on the diagonal of the matrix (which corresponds with commuters who choose to live and work in the same tract) are missing from low-rank approximations (e.g., rank 5 and rank 16) but are contained in higher-rank approximations (e.g., rank 200). This is because the diagonal of a matrix is a fundamentally "high-rank" phenomenon and cannot in general be replicated by a low-rank approximation.

Table B.3 summarizes the sparsity generated by approximations derived from various ranks, including both the SVD-based procedure described above and the non-negative factorization procedure described in Section B.4.3.

B.4.2 Selecting the rank

When applying any rank-restricted algorithm, the rank of the approximation is a key choice. Since we are using low-rank approximation to reduce idiosyncratic noise, the choice of the rank parameter $\mathfrak{r}$ balances the trade-off between the signal and noise in observed commuting flows. We

Figure B.4: Approximations of the commuting matrix



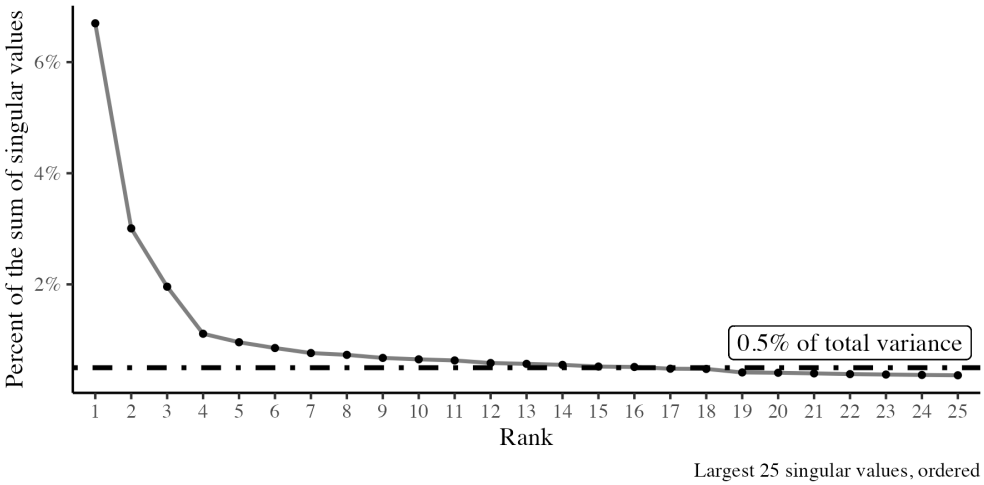
NOTES: This figure depicts the observed 2010 commuting matrix, the fitted covariates-based model (CBM) of ℓ_{kn}, approximations derived from rank-5, rank-16, and rank-200 singular value decompositions, and a rank-1 interactive-fixed-effects specification. Rows and columns are ordered by FIPS codes, so that all but eight adjacent entries are in the same county.

present multiple procedures from which a researcher can select to choose the rank used; some are computationally and conceptually simple (drawing from the established norms in machine learning), while a Monte Carlo simulation is more complex and tailored to a particular context. We describe each procedure in turn and present results, including their suggested rank.

Selecting the rank by simple methods After computing the SVD of the commuting matrix (which takes only seconds), we obtain an ordered list of singular values $\sigma_1, \sigma_2, \ldots, \sigma_N$ with $\sigma_i \geq \sigma_j$ for $i < j$. These are depicted in a scree plot in Figure B.5, which normalizes the singular values by dividing each by the sum of all singular values (i.e., $\frac{\sigma_{i'}}{\sum_i \sigma_i}$) and restricts attention to the 25 largest singular values. This normalization allows us to interpret each singular value as representing a fraction of the matrix's total variation (Cohen, 2021, p.499).

Figure B.5 allows us to evaluate two distinct rank criteria. The first is an informal procedure

Figure B.5: Scree plot (ordered singular values) for NYC 2010 commuting matrix



NOTES: This figure shows the magnitude of successive singular values for the NYC 2010 commuting matrix. The dashed line indicates the threshold of 0.5% of the sum of the singular values.

known as the "elbow" method, which is a method of visual inspection that attempts to identify the singular values that are substantially larger than the rest. The second, the variance criterion, is numerically founded but nonetheless requires a cutoff chosen by the researcher (Cohen, 2021, p.496). For the variance criterion, we compare the magnitude of each singular value with 0.5% of the sum of singular values; that is, we identify the set $\{\sigma_i \,|\, \sigma_i \geq 0.005 \sum_j \sigma_j\}$. This cutoff is shown with a dashed horizontal line in the scree plot. The two methods offer substantially different answers in our case, with the elbow method suggesting a lower rank of 4 and the 0.5% variance criterion suggesting a rank of 16. We select the more generous of these values and use rank 16 as our preferred specification in the main text.

Selecting the rank using Monte Carlo simulations Another method for selecting the rank is dependent on the research setting permitting Monte Carlo simulations. One can select the rank based on ranks' performance across Monte Carlo simulations (using regression coefficients and MSEs summarizing commuting-flow predictions). While substantially more computationally complex than the simple methods above, this procedure yields results that are more closely tied to the particular structure of the empirical setting.

Starting from the steps outlined in Appendix A.1 across 100 Monte Carlo draws from the continuum model, we follow steps 1 through 5 with two exceptions. First, we consider only changes in commuter flows, not rents; and second, we substitute the SVD-approximated commuting counts $\tilde{\ell}_{kn}^{\text{SVD}}$ for the "observed" counts, ℓ_{kn}, which are drawn from the data-generating process. Note that this requires us to approximate the commuting matrix separately for each of the Monte Carlo draws. As in Appendix A.1, we estimate the commuting elasticity ϵ for each of the Monte Carlo draws. The inputs into this commuting elasticity estimation are the "observed" values as in the covariates-based model, not the SVD-approximated commuting matrix. Finally, we raise productivity in the same tract (containing 200 Fifth Avenue) as in the Appendix A.1 simulations.

Table B.1 reports the performance across ranks in these Monte Carlo simulations. Rank 18 performs best in these simulations based on MSEs and is nearly best based on regression coefficients. Rank 16, which is suggested by the variance criteria shown in Figure B.5, performs very similarly.

The lower panel of Table B.1 presents the event study performance (for the same 83 employment booms as in Section 3.5.3) for approximations derived from a wide variety of ranks. The rank-16 results are shown in Figure 4.

B.4.3 Non-negative matrix factorization

While the previously outlined SVD procedure is one means of implementing noise reduction, truncating the commuting shares to be non-negative means that the resulting approxmiation is not necessarily low-rank. Another approach is to use an algorithm that imposes the non-negativity constraint. That is, we find a solution to the problem

$$\tilde{\mathbf{L}}^{\text{NNMF}} = \arg \min_{\mathbf{M} \in \mathbf{R}^{K \times N}} \|\mathbf{L} - \mathbf{M}\|_F^2 \quad \text{subject to} \quad \text{rank}(\mathbf{M}) \leq \mathfrak{r}, \text{ and } [\mathbf{M}]_{kn} \geq 0. \tag{B.2}$$

Solving this problem is more difficult and, unlike in the unconstrained case in which the solution is uniquely determined by SVD, there may be multiple (local) optima. We use the Greedy Coordinate Descent algorithm from Hsieh and Dhillon (2011) with a fixed initialization as implemented in the Julia package `NMF`. This procedure produces low-rank approximated matrices that deliver very similar performance as the computationally simpler SVD procedure described above. The results for event studies are shown in Table B.2.

B.4.4 Results

Table B.1 reports the event study and Monte Carlo performance of the SVD-approximated matrices. Table B.2 reports the event study performance of the non-negative matrix factorization (NNMF) approximations across ranks. As the non-negative matrix factorization appears to provide no meaningful improvement over the simpler SVD approximation procedure, we do not present any further analysis via Monte Carlo simulations for that procedure. Finally, Table B.3 reports the proportion of the approximated commuting matrix's entries that are zero across ranks for both approximation methods. While we used $\bar{\delta}_{kn}$ to calculate $y_{kn} = w_n \ell_{kn} / \bar{\delta}_{kn}$ in line with the rest of the paper, we have verified that the SVD approximation predictions are nearly identical when using $y_{kn} = w_n \ell_{kn}$.

B.5 Interactive fixed effects

B.5.1 Modeling unobserved commuting costs

This section extends the covariates-based model to incorporate unobserved disutility components of commuting costs. We assume that λ_{kn} has a low-rank factor structure. Specifically, we parameterize

Table B.1: Summary of SVD Monte Carlo and event study results by rank

	Rank																			
	1	2	3	4	5	6	8	10	12	14	15	16	18	20	50	100	500	1000	1500	full rank
Monte Carlo performance																				
Slope	1.03	1.05	1.04	1.04	1.02	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.99	.91	.78	.78	.78	.78	.78
Int.	-.041	-.059	-.056	-.049	-.021	-.012	.001	-.002	-.002	-.001	-.001	.001	.003	.014	.116	.268	.271	.271	.272	.272
MSE	.1579	.0461	.0460	.0460	.0372	.0357	.0333	.0322	.0323	.0320	.0318	.0315	.0311	.0311	.0930	.2241	.2279	.2280	.2280	.2280
Event study performance																				
Slope	.73	.70	.71	.80	.83	.86	.85	.83	.83	.82	.82	.81	.80	.79	.62	.32	-.43	-.47	-.47	-.46
Int.	.06	.14	.14	.09	.08	.08	.10	.13	.13	.14	.14	.15	.16	.17	.30	.51	.80	.82	.82	.82
MSE	10.53	10.38	10.37	10.29	10.27	10.26	10.27	10.30	10.32	10.32	10.38	10.40	10.45	10.48	10.96	11.71	12.94	13.23	13.35	13.39

NOTES: This table reports the predictive performance, measured by the average values of the regression slope and intercept coefficients and MSE, using SVD approximations of various ranks across two cases: a) 100 Monte Carlo simulations, with each drawn from the parameterized continuum model with a 9% productivity shock to the tract containing 200 Fifth Avenue, where the performance is evaluated relative to the continuum model's predictions; and b) the 83 event studies examined in Section 3.5.

Table B.2: Summary of NNMF event study results by rank

	Rank														
	1	2	3	4	5	6	8	10	12	14	16	18	20	15	50
Event study performance															
Slope	.73	.70	.71	.78	.78	.81	.85	.87	.87	.87	.87	.86	.83	.81	.71
Int.	.06	.14	.14	.11	.12	.10	.10	.10	.10	.10	.10	.11	.13	.15	.25
MSE	10.53	10.38	10.37	10.28	10.27	10.29	10.30	10.24	10.25	10.25	10.26	10.31	10.34	10.42	10.88

NOTES: This table reports the predictive performance, as measured by the average values of the regression slope and intercept coefficients and MSE, across the 83 event studies examined in Section 3.5 using various ranks of NNMF low-rank approximated commuting matrices.

Table B.3: Percentage (%) of approximated commuting matrix entries equal to zero

	Rank																			
	1	2	3	4	5	6	8	10	12	14	15	16	18	20	50	100	500	1000	1500	2143
SVD	0	1.9	2	6	11	12	15	15	16	17	17	17	17	18	22	26	37	41	42	85
NNMF	0	.4	.1	3	4.1	3.9	3.6	1.9	1.6	1.5	1.4	1.4	1	1.1	.5					

NOTES: This table reports the proportion of entries in the approximated commuting matrices, across ranks, that are equal to zero. These values are shown for both the SVD and the NNMF procedures, though we omit high-rank approximations via NNMF because of computational burden. The SVD procedure results in many more zeros, particularly for larger ranks, than the NNMF, largely due to the generation of negative values by SVD prior to our truncating them to zero.

the overall commuting costs δ_{kn} as

$$\delta_{kn} = \bar{\delta}_{kn} \times \underbrace{\exp(\psi_k' \gamma_n)}_{=\lambda_{kn}^{\text{IFE}}},$$

where ψ_k and γ_n are $R \times 1$ vectors that give rise to an R-dimensional factor structure. The log likelihood function then becomes

$$\ln \mathcal{L}^{\text{IFE}} = \sum_{k,n} \ell_{kn} \ln \left[\frac{w_n^{\epsilon} \left(r_k^{\alpha} \bar{\delta}_{kn} \exp(\psi_k' \gamma_n) \right)^{-\epsilon}}{\sum_{k',n'} w_{n'}^{\epsilon} \left(r_{k'}^{\alpha} \bar{\delta}_{k'n'} \exp(\psi_{k'}' \gamma_{n'}) \right)^{-\epsilon}} \right]. \tag{B.3}$$

In practice, to implement the estimation of the parameters, we follow Chen, Fernández-Val, and Weidner (2021), who estimate interactive fixed effects in a Poisson model. Here, multiple practical difficulties arise. Most critically, the likelihood is no longer convex; in particular, there is the possibility of multiple local optima. To guard against this, we use multiple randomized starting points and select the parameter estimates that produce the highest likelihood. Another difficulty is that the computational cost of this procedure rises rapidly with the factor dimension R. We consider $R \leq 3$. Note that in all specifications, additive fixed effects are already included to obtain estimates for w_n^ϵ and r_k^α; that is, R gives the dimension of the "purely" interactive fixed effects.[49]

In the main text, we show the estimates from the model with $R = 1$ for two reasons. First, we take advantage of the relative computational ease of the $R = 1$ case to estimate from a large number of starting points and find a consistent convergence pattern that suggests our estimate is the true optimum. Second, the eigenvalue ratio test suggested by Chen, Fernández-Val, and Weidner (2021) indicates that $R = 1$ is the optimal factor structure among the set of options $\{1, 2\}$. Possible choices of R span the range from the covariates-based specification (when $R = 0$) to the calibrated-shares procedure (the full-rank case).[50] The event study results for rank $R = \{2, 3\}$ are very similar to those for rank $R = 1$ and are shown below.

B.5.2 Estimation results

Table B.4 presents the estimates from maximizing the log likelihood function in equation (B.3) using the algorithm outlined by Chen, Fernández-Val, and Weidner (2021). We show results for $R = 0, 1, 2, 3$, where $R = 0$ is identical to the covariates-based model and the OLS column is included for the sake of comparison. The magnitude of the commuting elasticity ϵ diminishes with the rank of the factor structure. As expected, the model fit metrics increase with the factor structure rank, as higher ranks correspond to more free parameters to attain better (in-sample) fit.

B.5.3 Event study results for rank $R = \{2, 3\}$

In Figure B.6, we display the event study performance for IFE factor structure ranks 2 and 3. The extreme similarity in performance across the ranks suggests that rank 1 is sufficient to capture most of the expected gains from using IFE over the CBM.

[49] This differs from the notation in Chen, Fernández-Val, and Weidner (2021), who include the additive fixed effects in the factor dimension R. Our purely interactive factor dimension is, in their notation, R_2.

[50] Consider the case of a 2×2 commuting matrix A. This matrix can be decomposed into two 2×2 matrices (e.g., AI). However, there are many such decompositions, illustrating a failure of identification when fitting the commuting matrix with full-rank interactive fixed effects. Notice also that simply limiting the dimensions of the decomposition is not sufficient. In the 2×2 case, the only matrices A that can be written as the product of a 2×1 with a 1×2 vector are those with rank 1. As our commuting matrix is full rank, this suggests why a very high rank (and accompanying identification issues) would be needed to reproduce the calibrated-shares procedure using only high-rank interactive fixed effects.

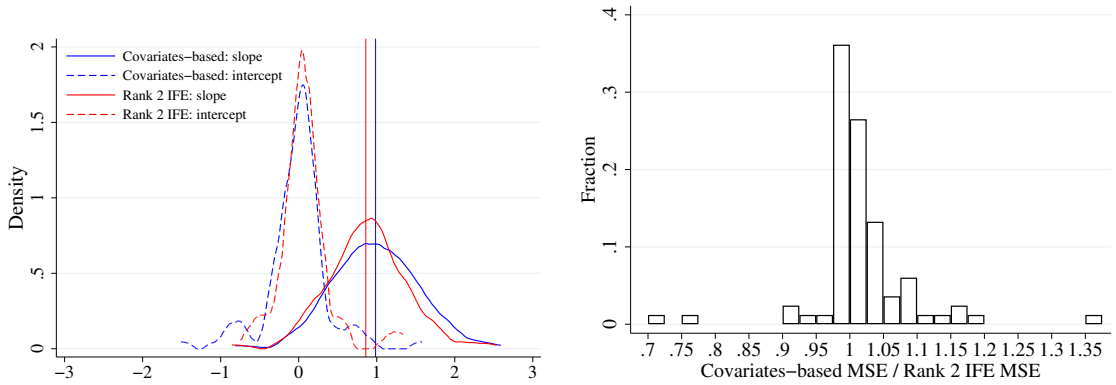
Table B.4: Commuting elasticity estimates from IFE specification

	OLS	MLE			
		$R = 0$	$R = 1$	$R = 2$	$R = 3$
Commuting cost	-2.307	-7.986	-7.177	-6.654	-6.36
(pseudo-)R^2	0.561	0.662	0.685	0.694	0.701
Location pairs	690,673	4,628,880			
Commuters	2,488,905	2,488,905			

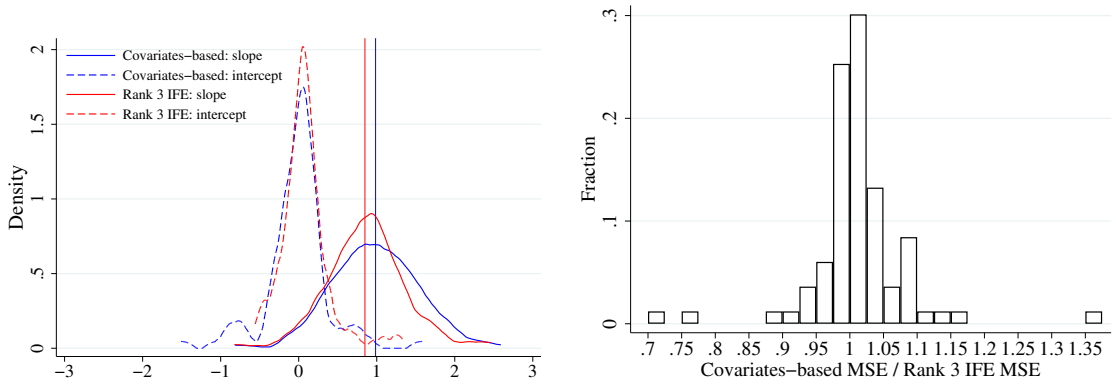
NOTES: All specifications include residence fixed effects and workplace fixed effects. The columns under "MLE" present the results from the maximum likelihood estimation described in the text, with R denoting the interactive fixed effects factor structure rank. The "OLS" column presents the results of estimating the log version of equation (2) by ordinary least squares, omitting observations where $\ell_{kn} = 0$. The model-fit statistic is pseudo-R^2 for MLE and R^2 for OLS.

Figure B.6: Event study performance: varying factor structure rank, IFE

A. Rank 2



B. Rank 3



NOTES: This figure summarizes the predictive performance of the covariates-based model and the ranks 2 and 3 IFE specifications, analogous to Figure 5. Panel A depicts the distributions of slope coefficients, and the vertical lines denote the median slope coefficients.

B.6 Tests of in-sample fit

In this appendix section, we assess the in-sample fit of different model parameterizations. The χ^2 test contrasts the expected and observed frequencies of outcomes and compares this difference to a critical value from the test-statistic distribution. Since in our application the expected frequencies of outcomes are small, the test statistic may not follow the asymptotic χ^2 distribution. We therefore construct the test statistic's distribution using a parametric bootstrap.

The test statistic is $\chi^2 = \sum_{kn} \frac{(\ell_{kn}^{\text{obs}} - \mathbb{E}[\ell_{kn}])^2}{\mathbb{E}[\ell_{kn}]}$, where ℓ_{kn}^{obs} is the observed number of individuals who reside in k and work in n in the data and $\mathbb{E}[\ell_{kn}]$ is the expected number of commuters (given by ℓ_{kn} in equation (2)). We construct the model-implied distribution of this test statistic from a parametric bootstrap in the following way:

1. We calculate the test statistic using

$$\chi^{2\text{ simulated}} = \sum_{kn} \frac{(\ell_{kn}^{\text{sim}} - \mathbb{E}[\ell_{kn}])^2}{\mathbb{E}[\ell_{kn}]}, \quad \ell_{kn}^{\text{sim}} \overset{\text{iid}}{\sim} \text{Multinomial}(I, \{p_{kn}\})$$

   where $p_{kn} \equiv \Pr(U_{kn}^i > U_{k'n'}^i \ \forall (k', n') \neq (k, n))$ is the model-implied probability of choosing kn defined in equation (8) and $I = 2,488,905$ is the total number of commuters.

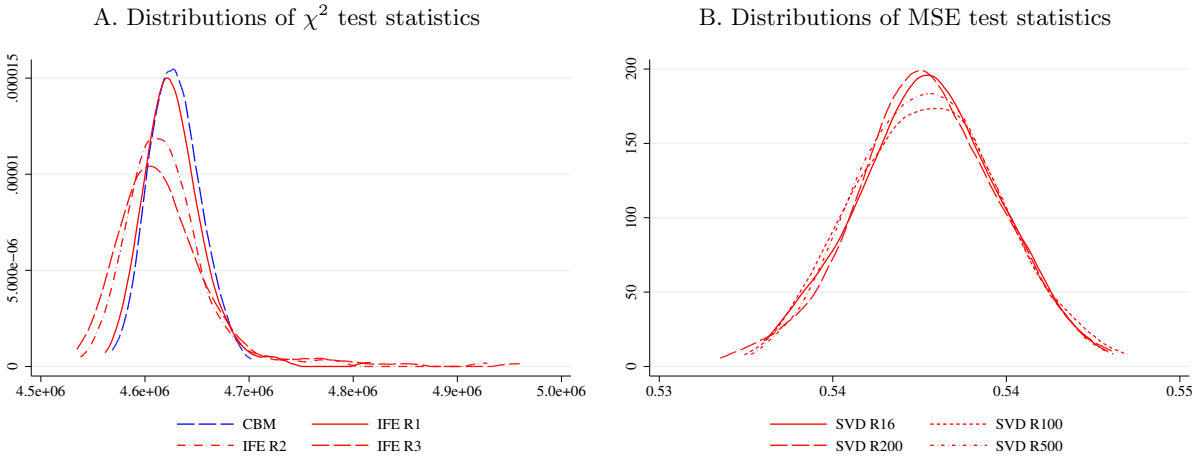
2. We repeat the first step 1,000 times to obtain the parametric-bootstrap-based distribution $\hat{f}\left(\chi^{2\text{ simulated}}\right)$.

The χ^2 test statistic for the covariates-based model is extremely large (calculated as 8.445e299, though numerical error may be sizable). This very large value of the test statistic is driven by outlier observations for which the model's expected number of commuters is extremely small (i.e., nearly equal to zero). By comparison, the test statistics for the interactive-fixed-effects specifications are much smaller: 1.006e7, 1.268e7, and 8.962e6 for ranks $R = 1$, 2, and 3, respectively. Nonetheless, all these test statistics exceed the 99$^{\text{th}}$ percentiles of the distributions of the χ^2 test statistics shown in Figure B.7A.

By contrast, the calibrated-shares procedure has perfect in-sample fit by construction and a χ^2 test statistic of zero. This fit is too good to be true: the test statistic of zero for the observed data is lower than all simulated values in the distribution generated by the parametric bootstrap.

We also assess the in-sample fit produced by the rank-restricted singular value decomposition (SVD) introduced in Section 3.5.3. One issue for this specification is that the rank-restricted SVD assigns zeros to some residence-workplace pairs that have positive counts in the data. A zero-probability event occurring rejects the parameterized model. We calculate the mean squared error (MSE) as an alternative test statistic that sidesteps the divide-by-zero issue. Specifically, we calculate MSE $= \frac{1}{K \times N} \sum_{kn} (\ell_{kn}^{\text{obs}} - \mathbb{E}[\ell_{kn}])^2$, where K is number of residence tracts and N is the number of workplace tracts. The parameterized bootstrapped distributions of MSE test statistics are shown in Figure B.7B. The approximated commuting matrix derived from a rank-110 SVD would lie within the 99-percent confidence interval, whereas higher-rank approximations exhibit an in-sample fit that is too good to be true.

Figure B.7: Distributions of χ^2 and MSE test statistics



NOTES: The test statistic in the left panel is $\chi^2 = \sum_{kn} \frac{(\ell_{kn}^{\text{obs}} - \mathbb{E}[\ell_{kn}])^2}{\mathbb{E}[\ell_{kn}]}$, where ℓ_{kn}^{obs} is the number of commuters from k to n and $\mathbb{E}[\ell_{kn}]$ is the expected value of commuters from k to n (given by ℓ_{kn} in the continuum model's equation (2)). We depict the distributions between their 0.5th and 99.5th percentiles. The test statistics using the commuter counts in the 2010 LODES data are 8.445e299, 1.006e7, 1.268e7, and 8.962e6 for the CBM, rank-1 IFE, rank-2 IFE, and rank-3 IFE, respectively. The test statistic in the right panel is MSE $= \frac{1}{K \times N} \sum_{kn} (\ell_{kn}^{\text{obs}} - \mathbb{E}[\ell_{kn}])^2$, where ℓ_{kn}^{obs} is the number of commuters from k to n, $\mathbb{E}[\ell_{kn}]$ is the expected value of commuters, K is number of residence tracts, and N is the number of workplace tracts. We depict the distributions between their 0.5th and 99.5th percentiles. The test statistics using the commuter counts in the 2010 LODES data are 1.467, 0.5725, 0.3581, and 0.1799 for the rank-16, rank-100, rank-200, and rank-500 singular value decompositions, respectively.

B.7 Spatially correlated idiosyncratic preferences

This appendix section relaxes the assumption that idiosyncratic preferences are independent across granular spatial units. In particular, preferences for residence-workplace pairs have a nested-logit structure, such that individual i's indirect utility of residing in k and working in n is

$$\mathring{U}_{kn}^i = \underbrace{\mathring{\epsilon} \ln \left(\frac{w_n}{r_k^{\alpha} P^{1-\alpha} \delta_{kn}} \right)}_{\equiv \mathring{U}_{kn}} + \mathring{\nu}_{kn}^i, \tag{B.4}$$

where $\mathring{U}_{kn}$ denotes the mean utility of choice kn and the parameter $\mathring{\epsilon}$ governs the importance of mean utility relative to idiosyncratic preferences in the nested-logit specification. We allow the idiosyncratic preferences $\mathring{\nu}_{kn}^i$ to be correlated across residences within Neighborhood Tabulation Areas (NTAs). We assume that the vector of idiosyncratic preferences has a cumulative distribution $F(\mathring{\nu}^i) = \exp\left\{ -\sum_z \left(\sum_{(k,n) \in B_z} e^{-\mathring{\nu}_{kn}^i / \zeta} \right)^{\zeta} \right\}$, where B_z denotes the set of residence-workplace kn pairs with the residence k in NTA z, and ζ governs the correlation of idiosyncratic preferences among alternatives in the set B_z. Whereas draws of idiosyncratic preferences between residence-workplace pairs with residences in distinct NTAs remain uncorrelated, the within-residence-NTA correlation of idiosyncratic preferences is $1 - \zeta^2$. Given equation (B.4) and this distribution of

idiosyncratic preferences, the fraction of people residing in k and working in n is

$$\mathbb{P}_{kn} = \underbrace{\frac{w_n^{\mathring{\epsilon}/\zeta}(r_k^{\alpha}\delta_{kn})^{-\mathring{\epsilon}/\zeta}}{\sum_{(k',n')\in B_z} w_{n'}^{\mathring{\epsilon}/\zeta}(r_{k'}^{\alpha}\delta_{k'n'})^{-\mathring{\epsilon}/\zeta}}}_{=\mathbb{P}_{kn|B_z}} \times \underbrace{\frac{\left[\sum_{(k',n')\in B_z} w_{n'}^{\mathring{\epsilon}/\zeta}\left(r_{k'}^{\alpha}\delta_{k'n'}\right)^{-\mathring{\epsilon}/\zeta}\right]^{\zeta}}{\sum_{z'}\left[\sum_{(k',n')\in B_{z'}} w_{n'}^{\mathring{\epsilon}/\zeta}\left(r_{k'}^{\alpha}\delta_{k'n'}\right)^{-\mathring{\epsilon}/\zeta}\right]^{\zeta}}}_{=\mathbb{P}_{B_z}},$$

where $\mathbb{P}_{B_z}$ denotes the probability of residing in NTA z and $\mathbb{P}_{kn|B_z}$ denotes the probability of selecting residence-workplace pair kn among residents of NTA z.

Under the assumption of nested-logit preferences and no unobserved disutility ($\delta_{kn} = \bar{\delta}_{kn} \forall kn$), the counterfactual-baseline ratio of commuters is

$$\hat{\ell}_{kn} = \left(\frac{\hat{w}_n}{\hat{r}_k^{\alpha}\hat{\bar{\delta}}_{kn}}\right)^{\mathring{\epsilon}/\zeta} \times \frac{\left[\sum_{(k',n')\in B_z}\left(\frac{\hat{w}_{n'}}{\hat{r}_{k'}^{\alpha}\hat{\bar{\delta}}_{k'n'}}\right)^{\mathring{\epsilon}/\zeta} \times \mathbb{P}_{k'n'|B_z}\right]^{\zeta-1}}{\sum_{z'}\mathbb{P}_{B_{z'}}\left[\sum_{(k',n')\in B_{z'}}\left(\frac{\hat{w}_{n'}}{\hat{r}_{k'}^{\alpha}\hat{\bar{\delta}}_{k'n'}}\right)^{\mathring{\epsilon}/\zeta} \times \mathbb{P}_{k'n'|B_{z'}}\right]^{\zeta}} \text{ if } \ell_{kn} > 0.$$

This specification yields the following log likelihood function:

$$\ln\mathcal{L} \equiv \sum_{k,n}\ell_{kn}\ln\mathbb{P}_{kn} = \sum_{k,n}\ell_{kn}\ln\left[\frac{w_n^{\mathring{\epsilon}/\zeta}(r_k^{\alpha}\delta_{kn})^{-\mathring{\epsilon}/\zeta} \times \left[\sum_{(k',n')\in B_z} w_{n'}^{\mathring{\epsilon}/\zeta}\left(r_{k'}^{\alpha}\delta_{k'n'}\right)^{-\mathring{\epsilon}/\zeta}\right]^{\zeta-1}}{\sum_{z'}\left[\sum_{(k',n')\in B_{z'}} w_{n'}^{\mathring{\epsilon}/\zeta}\left(r_{k'}^{\alpha}\delta_{k'n'}\right)^{-\mathring{\epsilon}/\zeta}\right]^{\zeta}}\right].$$

We sequentially estimate the parameters of the nested-logit model (Train, 2009). In the first step, we maximize the log likelihood with respect to $\mathring{\epsilon}/\zeta$, residence fixed effects, workplace fixed effects, and nest-specific fixed effects. Given the fixed effects, the estimate of $\mathring{\epsilon}/\zeta$ will be equal to the estimate of ϵ in the logit case. Note that the residence fixed effects are unique only up to a nest-specific normalization. We normalize the residence fixed effects relative to $r_k^{-\alpha\mathring{\epsilon}/\zeta}$ for some residence k in each NTA z. We index the baseline residence within a NTA as $\bar{k}(z)$. Once we impose the normalization, we have residence fixed effects $\left(\frac{r_k}{r_{\bar{k}(z)}}\right)^{-\alpha\mathring{\epsilon}}$. We can therefore write

$$\mathbb{P}_{kn|B_z} = \frac{\left[\left(\frac{r_k}{r_{\bar{k}(z)}}\right)^{-\alpha\mathring{\epsilon}} w_n^{\mathring{\epsilon}}\bar{\delta}_{kn}^{-\mathring{\epsilon}}\right]^{1/\zeta}}{\sum_{(k',n')\in B_z}\left[\left(\frac{r_{k'}}{r_{\bar{k}(z)}}\right)^{-\alpha\mathring{\epsilon}} w_{n'}^{\mathring{\epsilon}}\bar{\delta}_{k'n'}^{-\mathring{\epsilon}}\right]^{1/\zeta}} = \frac{\left[\left(\frac{r_k}{r_{\bar{k}(z)}}\right)^{-\alpha\mathring{\epsilon}} w_n^{\mathring{\epsilon}}\bar{\delta}_{kn}^{-\mathring{\epsilon}}\right]^{1/\zeta}}{I_z}, \quad \text{and} \quad \mathbb{P}_{B_z} = \frac{r_{\bar{k}(z)}^{-\alpha\mathring{\epsilon}} I_z^{\zeta}}{\sum_{z'} r_{\bar{k}(z')}^{-\alpha\mathring{\epsilon}} I_{z'}^{\zeta}},$$

where I_z is the inclusive value of nest z. In the second step, we compute the inclusive values using the estimates from the first step. Having the estimates of the inclusive values, $\hat{I}_z$, we maximize the

following log likelihood function with respect to $r_{\bar{k}(z)}$:

$$\ln \tilde{\mathcal{L}} \equiv \sum_z \left[\sum_{(k,n)\in B_z} \ell_{kn} \right] \ln \mathbb{P}_{B_z} = \sum_z \left(\left[\sum_{(k,n)\in B_z} \ell_{kn} \right] \times \ln \left[\frac{r_{\bar{k}(z)}^{-\alpha\mathring{\epsilon}} \hat{I}_z^{\zeta}}{\sum_{z'} r_{\bar{k}(z')}^{-\alpha\mathring{\epsilon}} \hat{I}_{z'}^{\zeta}} \right] \right)$$

Note that we cannot simultaneously identify $r_{\bar{k}(z)}$ and parameter ζ because the number of parameters exceeds the number of alternatives. We assume that the value of parameter ζ is known, in order to estimate $r_{\bar{k}(z)}$ up to a proportionality coefficient. We compute the equilibrium prices and the remaining economic primitives $\{T_k\}$ and $\{A_n\}$ from the estimated fixed effects and the market-clearing conditions.

We estimate the model for two values of parameter ζ: 0.25 and 0.75. We estimate each stage of the model by using the Poisson pseudo maximum likelihood estimator described in Section 3.2.When ζ is 0.75, the elasticity parameter $\mathring{\epsilon}$ is 5.99 ($= 7.99 \times \zeta$). The implied productivities are unchanged, and the implied land endowments have a correlation of 0.99 with the land endowments in the baseline model. When ζ is 0.25, the elasticity parameter $\mathring{\epsilon}$ is 2.0. The implied productivities are unchanged, and the implied land endowments have a correlation of 0.61 with the land endowments in the baseline model.

As long as the nested-logit model and the logit-model agree on the unconditional choice probabilities (i.e., they imply the same $\mathbb{P}_{k'n'}$), we note that in the nested-logit model there is greater within-nest substitution in response to a change in commuting costs than in the simple logit model. This can be seen in the second of the following three elasticities, noting that $\mathbb{P}_{k'n'|B_z} > \mathbb{P}_{k'n'}$:

$$\frac{\partial \ln \mathbb{P}_{kn}}{\partial \ln \delta_{k'n'}} = \begin{cases} -\epsilon(1 - \mathbb{P}_{k'n'|B_z}) - \mathring{\epsilon}(\mathbb{P}_{k'n'|B_z} - \mathbb{P}_{k'n'}) & \text{if } k' = k,\ n' = n \text{ and } (k,n) \in B_z \\ \epsilon\left(\zeta \mathbb{P}_{k'n'} + (1-\zeta)\mathbb{P}_{k'n'|B_z}\right) & \text{if } k' \neq k,\ (k',n') \in B_z \text{ and } (k,n) \in B_z \\ \mathring{\epsilon}\mathbb{P}_{k'n'} & \text{if } (k',n') \notin B_z \text{ and } (k,n) \in B_z \end{cases}$$

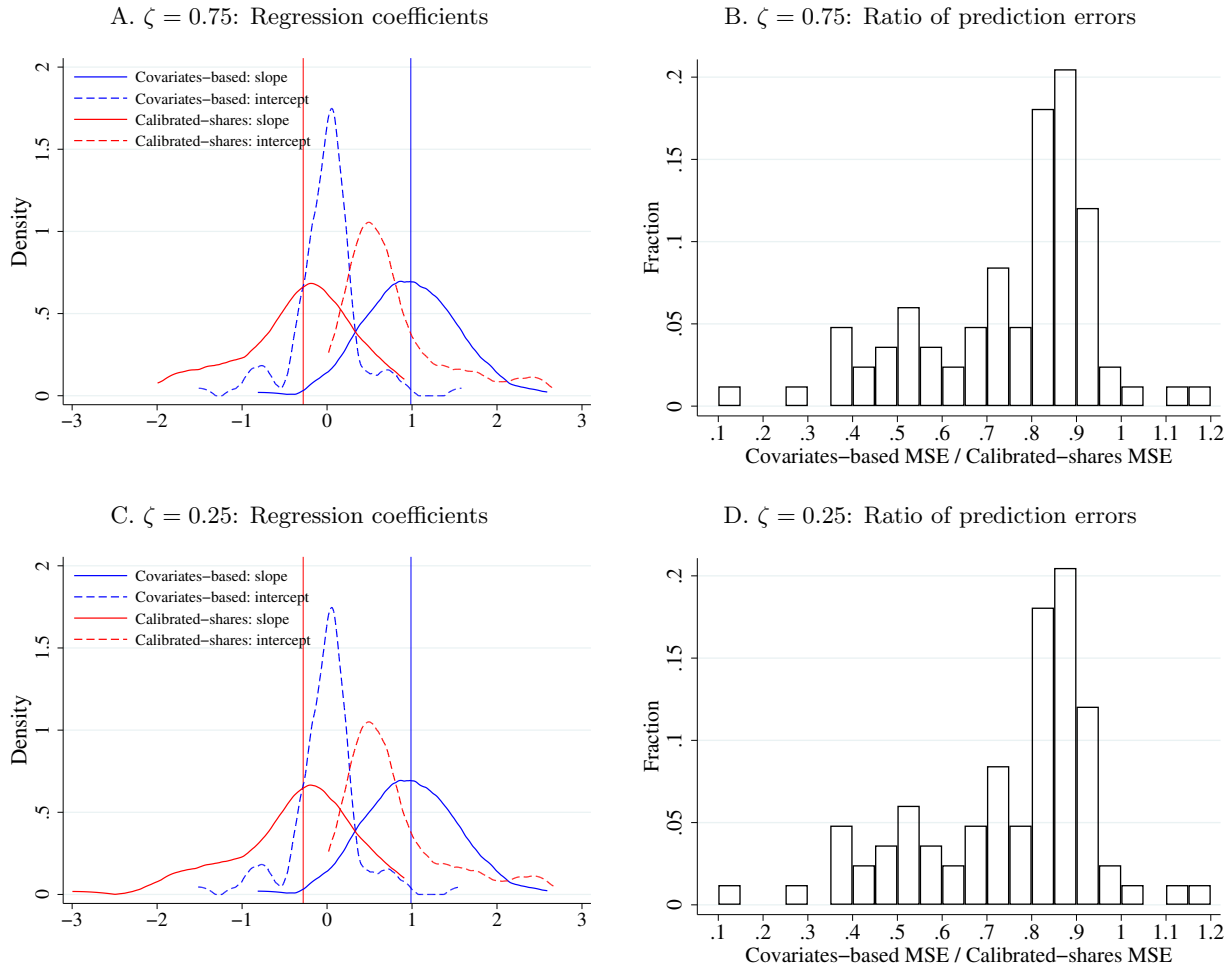
We use the model to compare the predictive performance of the covariates-based model and the calibrated-shares procedure in 83 employment booms that are detailed in Section 3.5. Figure B.8 contrasts the predictions of the two procedures for each value of the parameter ζ. As the figure shows, the results are robust to the alternative assumptions about preferences. Even when idiosyncratic preferences are highly correlated within residence NTAs, the covariates-based model considerably outperforms the calibrated-shares procedure.

B.8 Alternative labor demand elasticities

We compare the predictive performance of the covariates-based model and the calibrated-shares procedure for varying values of σ, the local labor demand elasticity. For both $\sigma = 1.1$ and $\sigma = \infty$, the covariates-based model has a lower mean squared error than the calibrated-shares procedure in 80 of the 83 events.[51] The covariates-based model outperforms the calibrated-shares procedure

[51] Note that the specification with $\sigma = 1.1$ does not satisfy the sufficient condition for uniqueness.

Figure B.8: Comparison of methods' predictive performance with nested-logit ν^i



NOTES: The pairs of panels in this figure depict results analogous to those in Figure 3 for the nested-logit model introduced in this appendix with nesting parameter $\zeta = 0.75$ (Panels A and B) and $\zeta = 0.25$ (Panels C and D). The results for $\zeta = 0.75$ and $\zeta = 0.25$ are not numerically identical, but they typically coincide to two decimal places and are visually indistinguishable.

regardless of the value of σ because their predictions are not very sensitive to σ. Varying σ does not meaningfully change the slope, intercept, or MSEs. As an example, Figure B.9 shows that the slope coefficients obtained when using $\sigma = 1.1$ or $\sigma = \infty$ are very similar to those for $\sigma = 4.0$ in the calibrated-shares procedure.

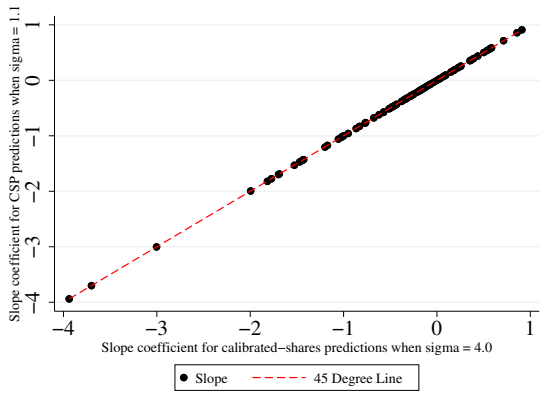
B.9 Local increasing returns to scale

This appendix section details the process for computing counterfactual predictions with local increasing returns and shows the results of such an exercise. In short, the predictions from both the covariates-based model and the calibrated-shares procedure are not sensitive to local increasing returns over the range of relevant values.

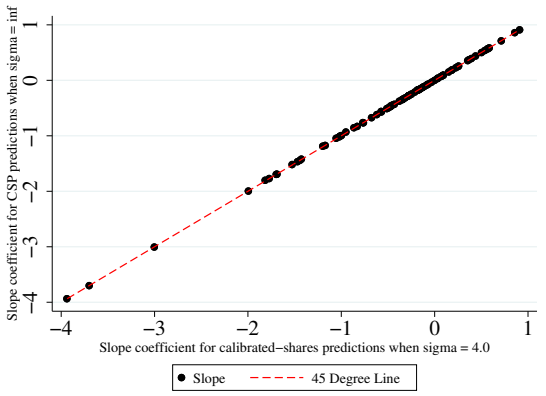
Per Appendix C.11.4, we assume that the increasing returns are external to the firm and that the

Figure B.9: Event-study slope coefficients by local labor demand elasticity σ

A. Slope coefficients for $\sigma = 1.1$ versus $\sigma = 4.0$



B. Slope coefficients for $\sigma = \infty$ versus $\sigma = 4.0$



NOTES: The slope coefficients come from regressions of the observed change in commuters on the predicted change in commuters for each of the 83 tract-level employment booms.

production function in location n is given by $q_n = \bar{A}_n L_n^{\eta+1}$. This implies the goods-market-clearing condition is

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

This can be rearranged to solve for w_n:

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

To express the counterfactual outcomes via exact hat algebra, take ratios to obtain:

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

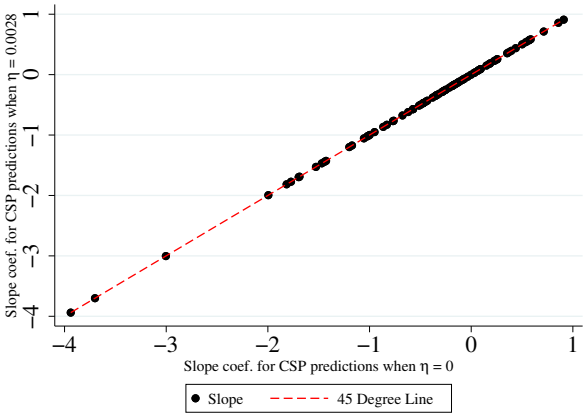
This expression for changes in wages generalizes equation (5) to the case of local increasing returns. Equations (6) and (7) are unchanged. The change in the CES price index is

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

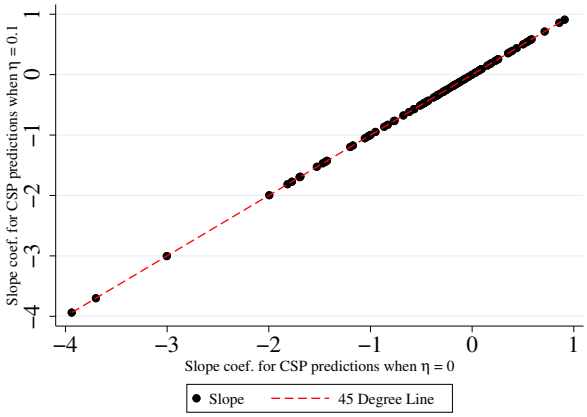
In the event studies of 83 local employment booms, the covariates-based model outperforms the calibrated-shares procedure regardless of the value of η; varying η does not meaningfully change the slope, intercept, or MSEs. The differences in predicted commuting flows typically differ only on the order of 10^{-2} relative to the baseline, even when using values of η larger than the sufficient condition for the equilibrium uniqueness. As an example, Figure B.10 shows that the slope coefficients obtained when using $\eta = .0028$ or $\eta = 0.1$ are very similar to those for $\eta = 0$ in the calibrated-

Figure B.10: Event-study slope coefficients by local increasing returns η

A. Slope coefficients for $\eta = 0.0028$ versus $\eta = 0$



B. Slope coefficients for $\eta = 0.1$ versus $\eta = 0$



NOTES: The slope coefficients come from regressions of the observed change in commuters on the predicted change in commuters for each of the 83 tract-level employment booms. The baseline case, $\eta = 0$, is the model without increasing returns.

shares procedure.

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, Arkolakis, and Li (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, Arkolakis, and Li (2023), which leave the conclusion of their Theorem 1 Part ii.b unchanged.

Following Remark 5 of Allen, Arkolakis, and Li (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 $(\mathbf{A})_{hh'} \equiv \left|\left(\mathbf{B\Gamma}^{-1}\right)_{hh'}\right|$:

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

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 $\mathbf{B\Gamma}^{-1}$. 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, Arkolakis, and Li (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, Arkolakis, and Li (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 $f_k(\lambda)$:

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

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

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

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, Arkolakis, and Li (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:

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

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):

$$\begin{aligned}
\hat{\ell}_{kn} &= \frac{\frac{w_n'^{\epsilon}(r_k'^{\alpha}\bar{\delta}'_{kn}\lambda'_{kn})^{-\epsilon}}{\sum_{k',n'} w_{n'}'^{\epsilon}(r_{k'}'^{\alpha}\bar{\delta}'_{k'n'}\lambda'_{k'n'})^{-\epsilon}}}{\frac{w_n^{\epsilon}(r_k^{\alpha}\bar{\delta}_{kn}\lambda_{kn})^{-\epsilon}}{\sum_{k',n'} w_{n'}^{\epsilon}(r_{k'}^{\alpha}\delta_{k'n'}\lambda_{k'n'})^{-\epsilon}}} \\
&= \frac{\hat{w}_n^{\epsilon}(\hat{r}_k^{\alpha}\hat{\bar{\delta}}_{kn}\hat{\lambda}_{kn})^{-\epsilon}}{\frac{\sum_{k',n'} w_{n'}'^{\epsilon}(r_{k'}'^{\alpha}\bar{\delta}'_{k'n'}\lambda'_{k'n'})^{-\epsilon}}{\sum_{k',n'} w_{n'}^{\epsilon}(r_{k'}^{\alpha}\bar{\delta}_{k'n'}\lambda_{k'n'})^{-\epsilon}}} \\
&= \frac{\hat{w}_n^{\epsilon}(\hat{r}_k^{\alpha}\hat{\bar{\delta}}_{kn}\hat{\lambda}_{kn})^{-\epsilon}}{\sum_{k',n'} \frac{w_{n'}'^{\epsilon}(r_{k'}'^{\alpha}\bar{\delta}'_{k'n'}\lambda'_{k'n'})^{-\epsilon}}{w_{n'}^{\epsilon}(r_{k'}^{\alpha}\bar{\delta}_{k'n'}\lambda_{k'n'})^{-\epsilon}} \frac{w_{n'}^{\epsilon}(r_{k'}^{\alpha}\bar{\delta}_{k'n'}\lambda_{k'n'})^{-\epsilon}}{\sum_{k',n'} w_{n'}^{\epsilon}(r_{k'}^{\alpha}\bar{\delta}_{k'n'}\lambda_{k'n'})^{-\epsilon}}} \\
&= \frac{\hat{w}_n^{\epsilon}(\hat{r}_k^{\alpha}\hat{\bar{\delta}}_{kn}\hat{\lambda}_{kn})^{-\epsilon}}{\sum_{k',n'} \hat{w}_{n'}^{\epsilon}(\hat{r}_{k'}^{\alpha}\hat{\bar{\delta}}_{k'n'}\hat{\lambda}_{k'n'})^{-\epsilon}\frac{\ell_{k'n'}}{L}}.
\end{aligned}$$

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 are do not vary across locations: $B_k = 1\ \forall k$. The probability of choosing residential-workplace pair kn is now

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

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, Rossi-Hansberg, and Sarte (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

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

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:

$$\hat{\ell}_{kn^*} = \frac{\hat{A}^{\epsilon}_{n^*}}{\hat{A}^{\epsilon}_{n^*}\frac{\sum_{k'}\ell_{k'n^*}}{L} + \frac{L-\sum_{k'}\ell_{k'n^*}}{L}} = \hat{L}_{n^*} \ \forall k$$
$$\Delta\ell_{kn^*} = \left(\hat{\ell}_{kn^*} - 1\right)\ell_{kn^*} = \left(\hat{L}_{n^*} - 1\right)\ell_{kn^*}.$$

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 $p_{kn} \equiv \mathbb{P}(U^i_{kn} > U^i_{k'n'} \ \forall (k', n') \neq (k, n))$ 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).[52] 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.[53] The two-way fixed-effects estimator is $\hat{p}_{kn} = s_k s_n$, where $s_k = \sum_n s_{kn}$ 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

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

[52] 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 $\mathbb{E}(s_{kn}) = p_{kn}$.

[53] 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 $\lambda_{kn}^{-\epsilon}$.

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

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

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 $p_{kn} \equiv \mathbb{P}(U^i_{kn} > U^i_{k'n'} \, \forall (k', n') \neq (k, n))$. 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.[54] 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

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

where the first term is the square of the bias of the estimator and the second term is its sample variance.[55] 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 54). 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

[54] 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}.

[55] 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}.

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 $\frac{1+2\mathfrak{d}_{kn}}{2p_kp_n+2\mathfrak{d}_{kn}} \pm \frac{\sqrt{1+4\mathfrak{d}_{kn}(1-p_kp_n)}}{2p_kp_n+2\mathfrak{d}_{kn}}$, and the inequality above is true when $\lambda_{kn}^{-\epsilon}$ is close enough to one to lie between these roots.[56] 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.[57]

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.

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

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

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

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

[57] 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.

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 $\frac{1}{I}\sum_{i=1}^{I} \mathbf{1}\left\{\tilde{U}^i_{kn}(\boldsymbol{\nu}^I) > \tilde{U}^i_{k'n'}(\boldsymbol{\nu}^I), \forall (k',n') \neq (k,n)\right\}$ converges to the mean of a binary random variable that is equal to one with probability $\mathbb{P}(\tilde{U}^i_{kn} > \tilde{U}^i_{k'n'}, \forall (k',n') \neq (k,n))$. Thus,

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

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 $p \equiv \mathbb{P}\left(\tilde{U}'_2 - \tilde{U}'_1 < \nu^i_1 - \nu^i_2 < \tilde{U}_2 - \tilde{U}_1\right) = F\left(\tilde{U}_2 - \tilde{U}_1\right) - F\left(\tilde{U}'_2 - \tilde{U}'_1\right)$ 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,

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

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 $\tilde{U}^i_{kn}$ 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

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

where L_n is labor and T^P_n 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 $P = \left(\sum_n c_n^{1-\sigma}\right)^{1/(1-\sigma)}$.

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

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

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

$$\sum_{n'} r_{n'} T^P_{n'} = \frac{\beta}{1-\beta} \sum_{n'} w_{n'} \sum_{k'} \frac{\ell_{k'n'}}{\bar{\delta}_{k'n'}} = \frac{\beta}{1-\beta} \sum_{n'} w_{n'} L_{n'}.$$

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.

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, Rossi-Hansberg, and Sarte (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, Kutzbach, and McKenzie, 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.

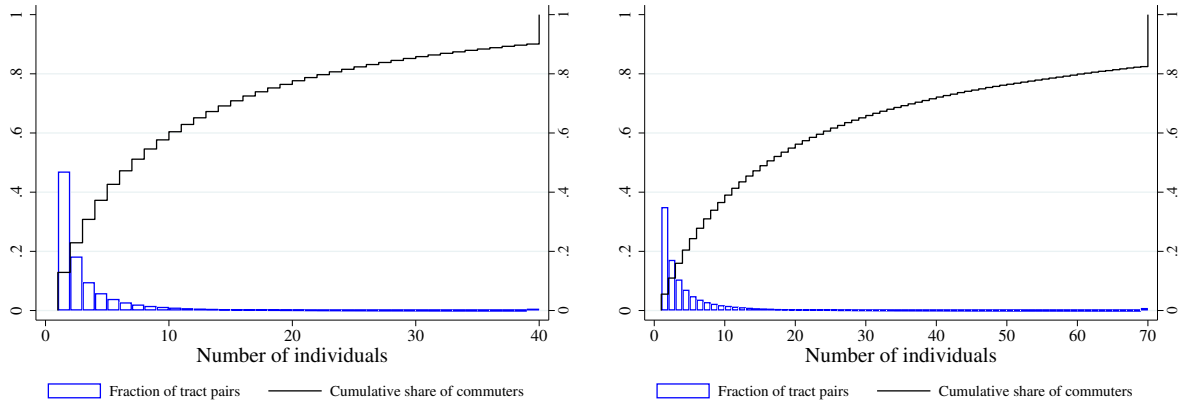
The spatial concentration of employment contributes to the sparsity of these commuting matrices. The median tract in Detroit has 465 employees working in it. Since Detroit has 1,166 residential tracts, at least 60% of locations must have zero residents commuting to this workplace.

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).[58] 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

[58] We follow Monte, Redding, and Rossi-Hansberg (2018) by restricting attention to county pairs that are less than 120 kilometers apart.

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 99$^{\text{th}}$ 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.

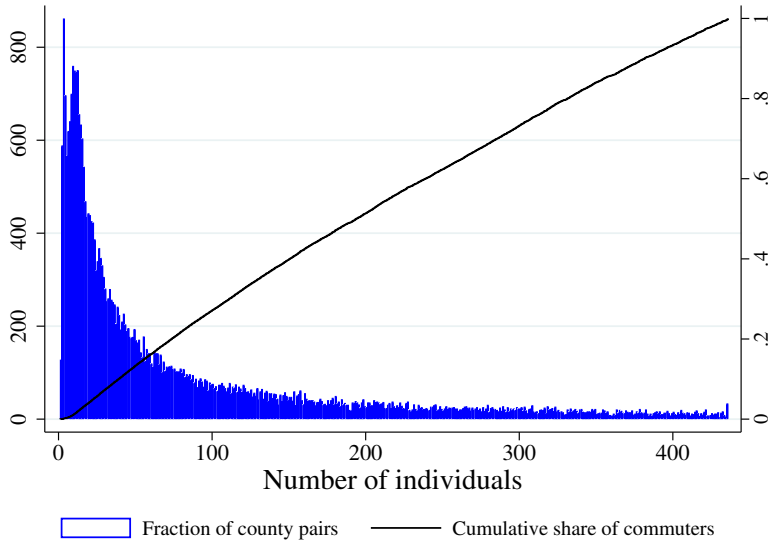
half of the county pairs report zero commuters. Among those reporting positive values, Figure D.2 shows that this distribution is skewed, so that many thousands of county pairs have small numbers of commuters. Unlike census tracts, however, these small flows constitute a small share of the total number 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, 2019). 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).[59] There are 7 million cross-municipality commuters between 131,620 pairs of municipalities. Thus, the average off-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.

[59]Dingel, Miscio, and Davis (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.

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

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.

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.[60] 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 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.

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 magni-

[60] 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.

Table D.3: The impersistence of commuting counts for pairs of US counties

2006–2010 \ 2011–2015	Initial Share(%)	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.

tudes 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.

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

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.

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.

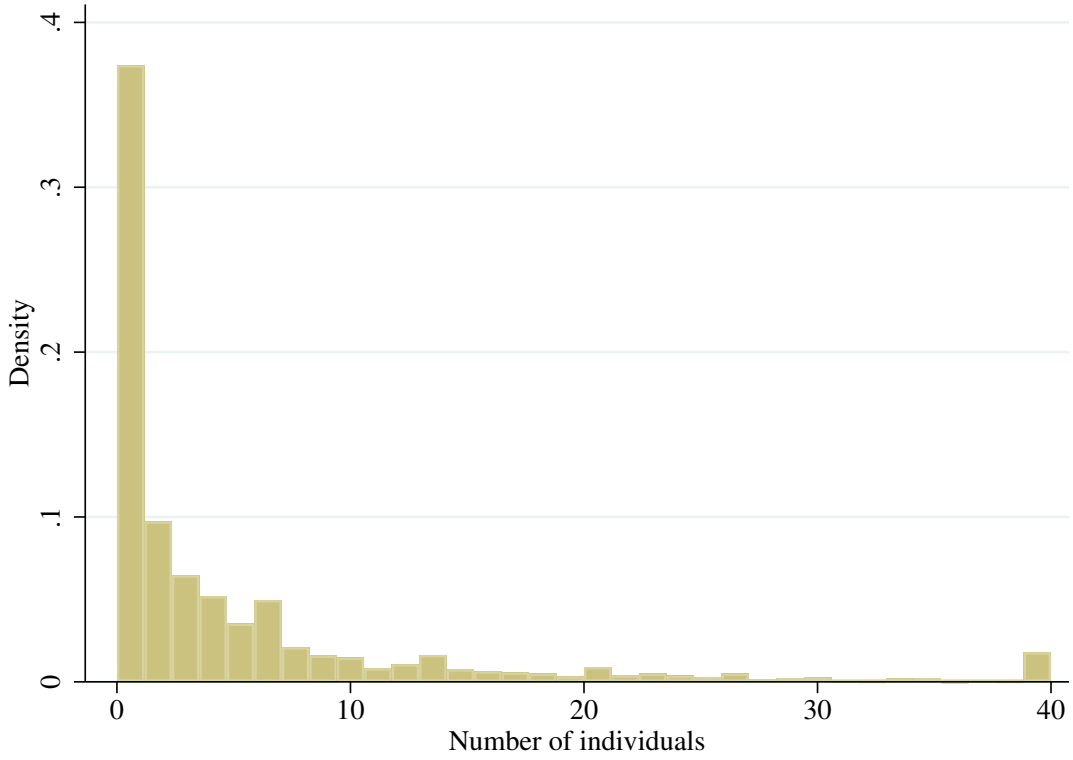
In many empirical settings, asymmetric zeros may simply reflect small flows rather than evidence of very asymmetric commuting costs. When most pairs of locations have zero commuters, and the modal pair with a positive number reflects the decision of only one commuter or one survey respondent, the difference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

Table D.5: Transition matrix for state migration flows between 2001 and 2002

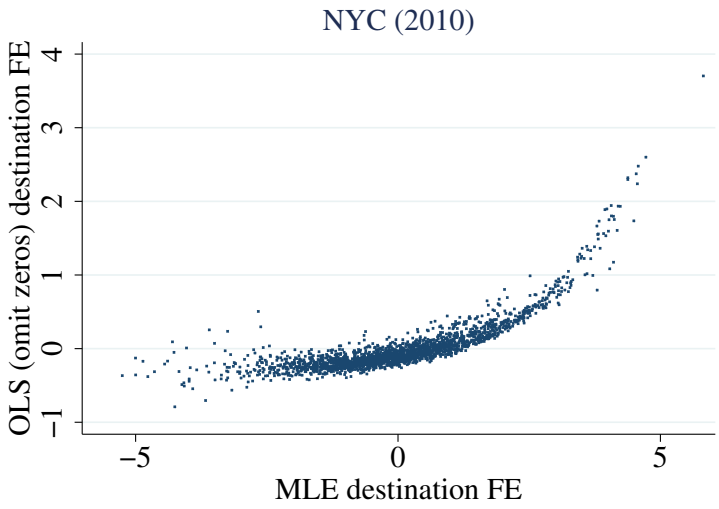
	2002						
2001	0	1	2	3	4	5	6+
0	0.59	0.16	0.13	0.05	0.03	0.01	0.02
1	0.40	0.17	0.17	0.10	0.06	0.05	0.06
2	0.32	0.16	0.21	0.09	0.07	0.07	0.09
3	0.23	0.16	0.21	0.09	0.08	0.09	0.14
4	0.26	0.13	0.15	0.10	0.06	0.06	0.25
5	0.13	0.09	0.12	0.14	0.14	0.06	0.31
6+	0.03	0.03	0.06	0.06	0.06	0.07	0.69

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 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.[61] 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

[61] 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.

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 and the block-level workplace area characteristics data for New York City are obtained from LODES version 7.4, published by the US Census Bureau.[62] Following Owens, Rossi-Hansberg, and Sarte (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`

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.

The workplace area characteristic data for NYC are downloaded from https://lehd.ces.census.gov/data/lodes/LODES7/ny/wac/, from `ny_wac_S000_JT01_2002.csv.gz` to `ny_wac_S000_JT01_2017.csv.gz` for primary jobs.

In these files, the relevant variables are

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

- `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, Rossi-Hansberg, and Sarte (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-tact 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, Rossi-Hansberg, and Sarte (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/CB0600C` 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.

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.[63]

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 $L_n = \sum_k \frac{\ell_{kn}}{\delta_{kn}}$. 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

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

The equilibrium wage in the continuum model is $w_n(\mathbb{E}[L_n])$, while the mean equilibrium wage in the model with a finite number of individuals is $\mathbb{E}[w_n(L_n)]$. 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

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

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.

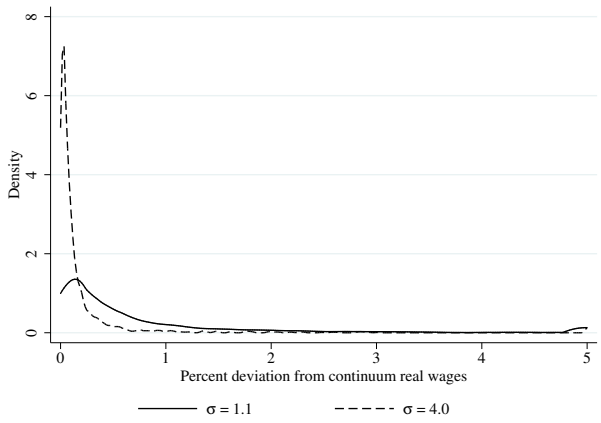
Using $f_n(L_n)$ to denote the probability mass function of supplied labor, the difference between the continuum model's wage and the mean wage in the model with a finite number of individuals is

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

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

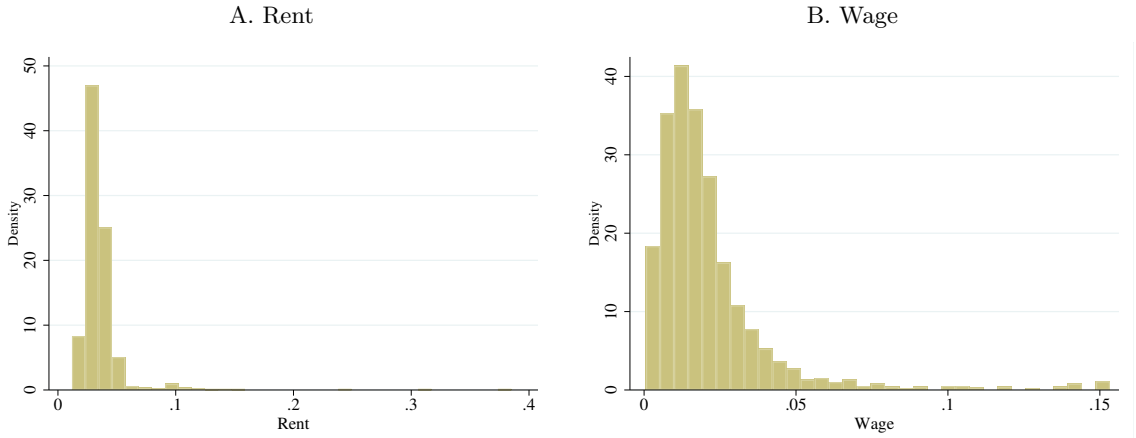


NOTES: This figure depicts the absolute percentage-point deviation of the mean realized wage from the continuum-case rational expectation of the wage across the 2,143 workplace tracts. We compute the mean wage using 100,000 simulations of the model with a finite number of individuals. This mean excludes 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

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

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 $\tilde{U}_{k,n}^{i}$. 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:

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

which implies that

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

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.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.1: Distribution of ex post regrets

s	Share with regret	Unconditional distribution					Conditional distribution	
		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: Price dispersion across simulation methods

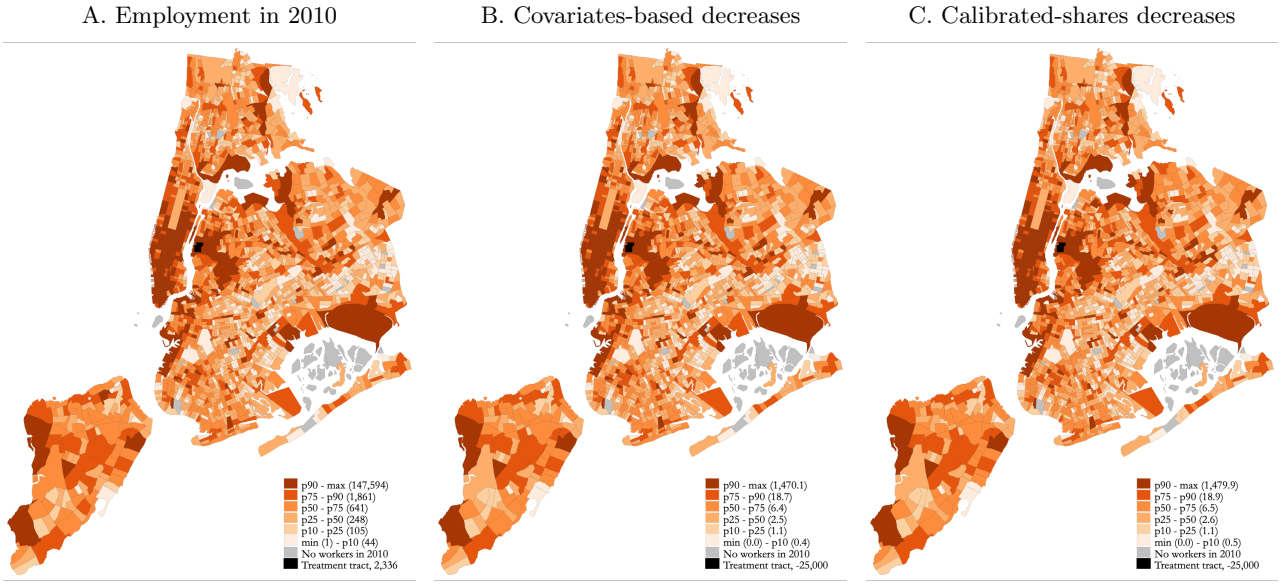
Simulation count	mean	p5	p10	p25	p50	75	p90	p95
Wage								
100,000	0.021	0.004	0.006	0.010	0.016	0.025	0.039	0.054
10	0.023	0.004	0.005	0.009	0.015	0.025	0.039	0.057
Rent								
100,000	0.035	0.021	0.024	0.027	0.032	0.038	0.045	0.051
10	0.034	0.016	0.019	0.024	0.031	0.038	0.049	0.058

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.

F Amazon HQ2 counterfactual

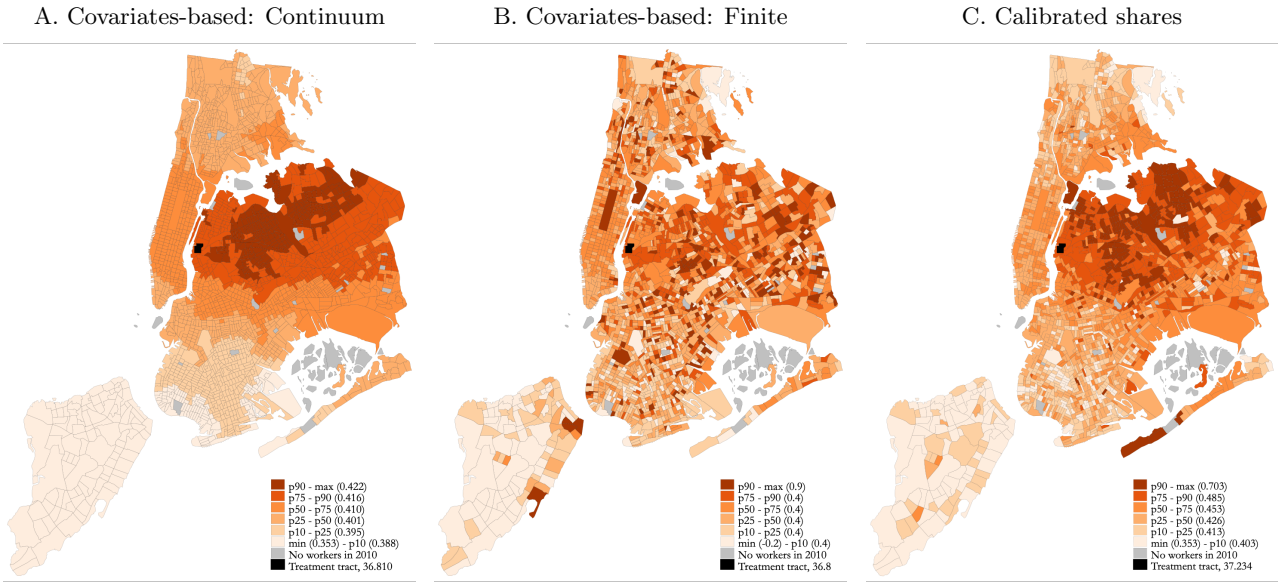
F.1 Geographic incidence of Amazon HQ2

Figure F.1: Amazon HQ2 counterfactual decreases in workers



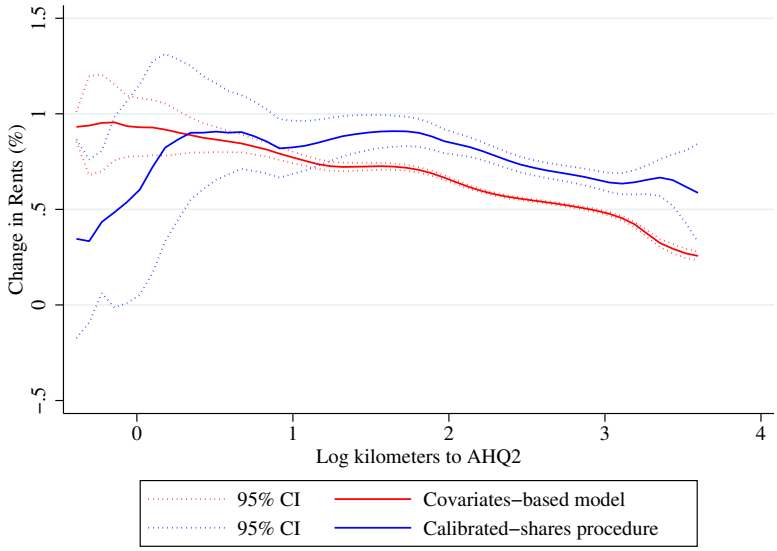
NOTES: Panel A depicts the number of workers employed in each tract in 2010. Panels B and C depict decreases in the number of workers predicted by the covariates-based model and calibrated-shares procedure, respectively. As total population is fixed, the number of workers decreases in all tracts except 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

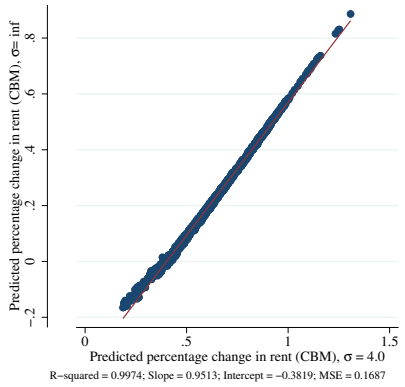


NOTES: This figure plots the changes in rents predicted by the covariates-based continuum model and the calibrated-shares procedure as functions of (log) distance to the Amazon HQ2 tract. The percent change in rents is $\left(\frac{\hat{r}_k}{\hat{P}}-1\right)\times 100$.

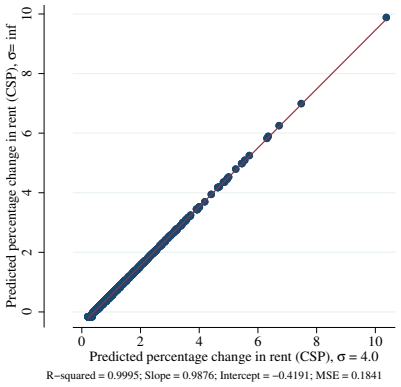
As alternatives to $\sigma = 4$, we consider the values $\sigma = 1.1$ and $\sigma = \infty$. The predicted changes in commuters to the Amazon HQ2 tract from both the covariates-based model and the calibrated-shares procedure are invariant to the labor demand elasticity σ: regressing predictions for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

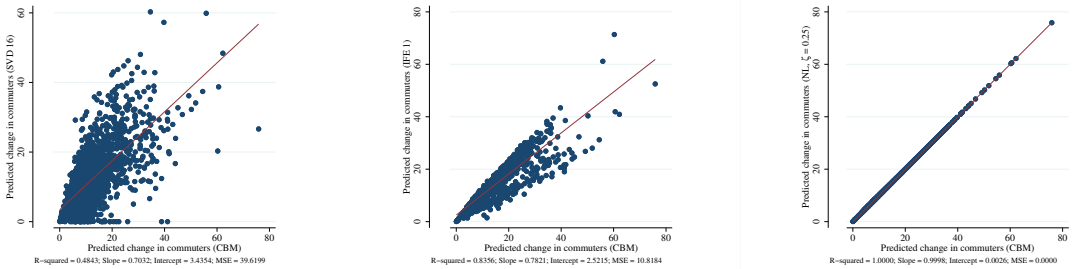


B. CSP comparison of real rents for $\sigma = 4.0$ and $\sigma = \infty$



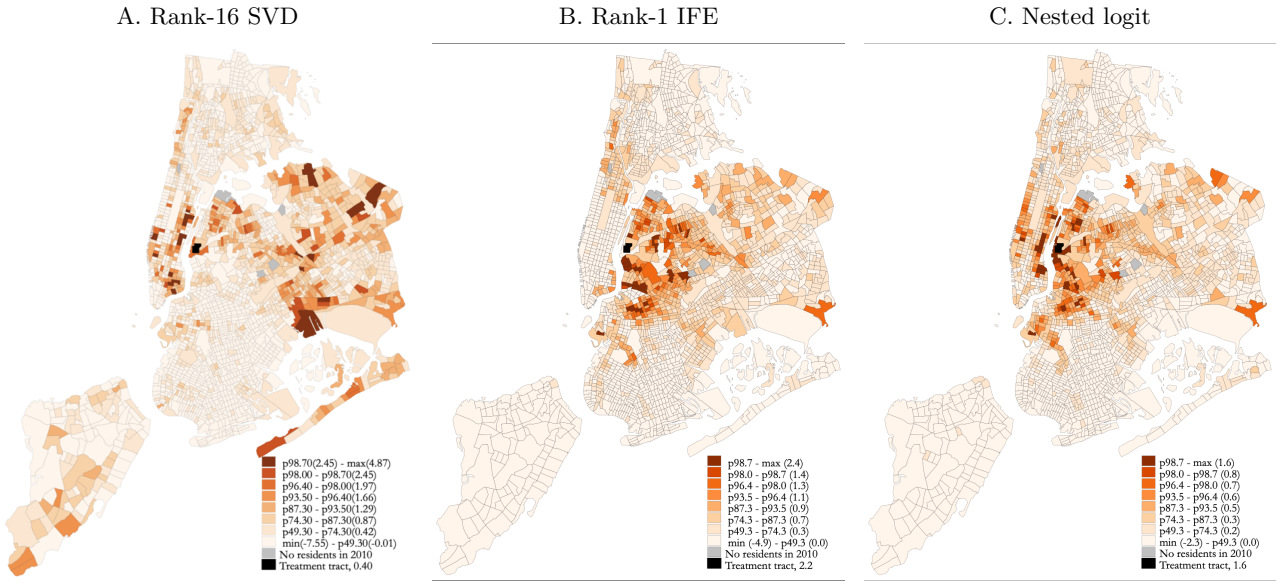
NOTES: These plots compare the percent change in real rents, $(\hat{r}/\hat{P} - 1) \times 100$, predicted by the covariates-based model (left panel) and calibrated-shares procedure (right panel) for the labor demand elasticities of $\sigma = 4.0$ and $\sigma = \infty$. In both panels, the predicted 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.

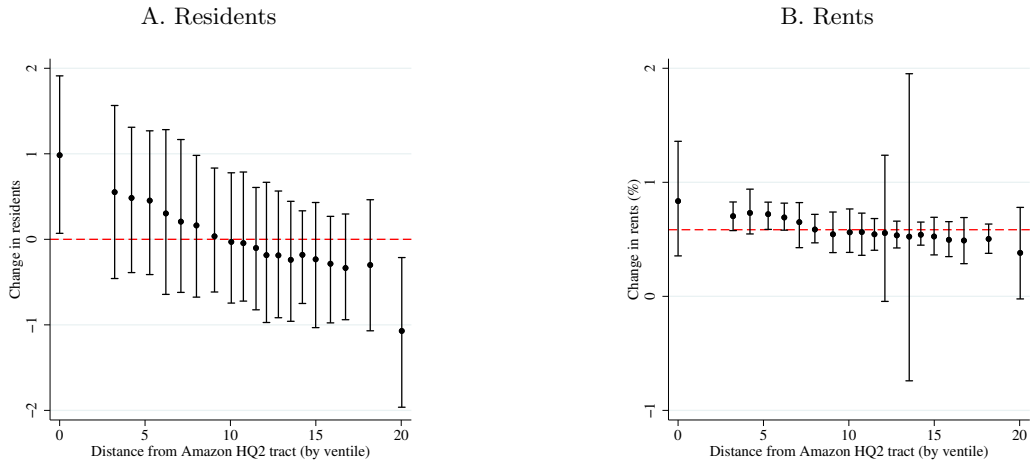
Figure F.6: Predicted changes in residents from SVD, IFE, and nested-logit specifications



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.

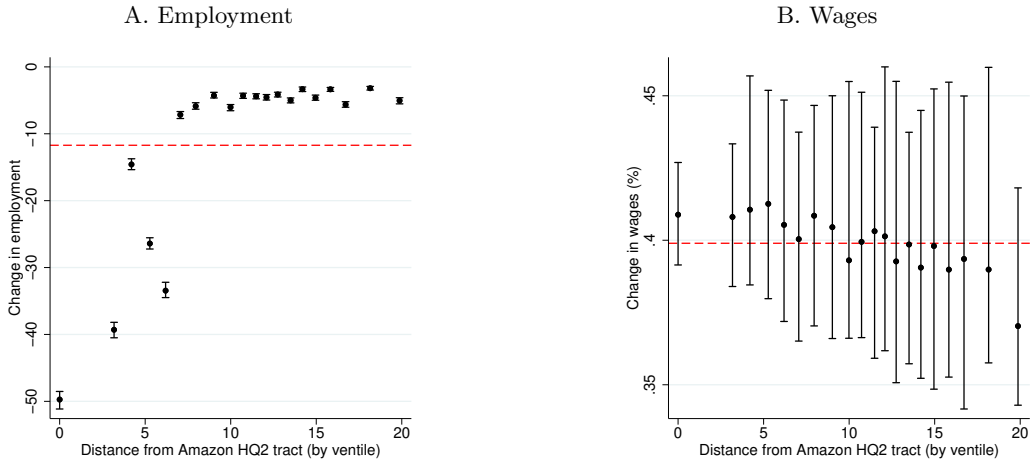
F.2 Uncertainty about Amazon HQ2 predictions due to idiosyncrasies

Figure F.7: Changes in residents and rents by distance ventile



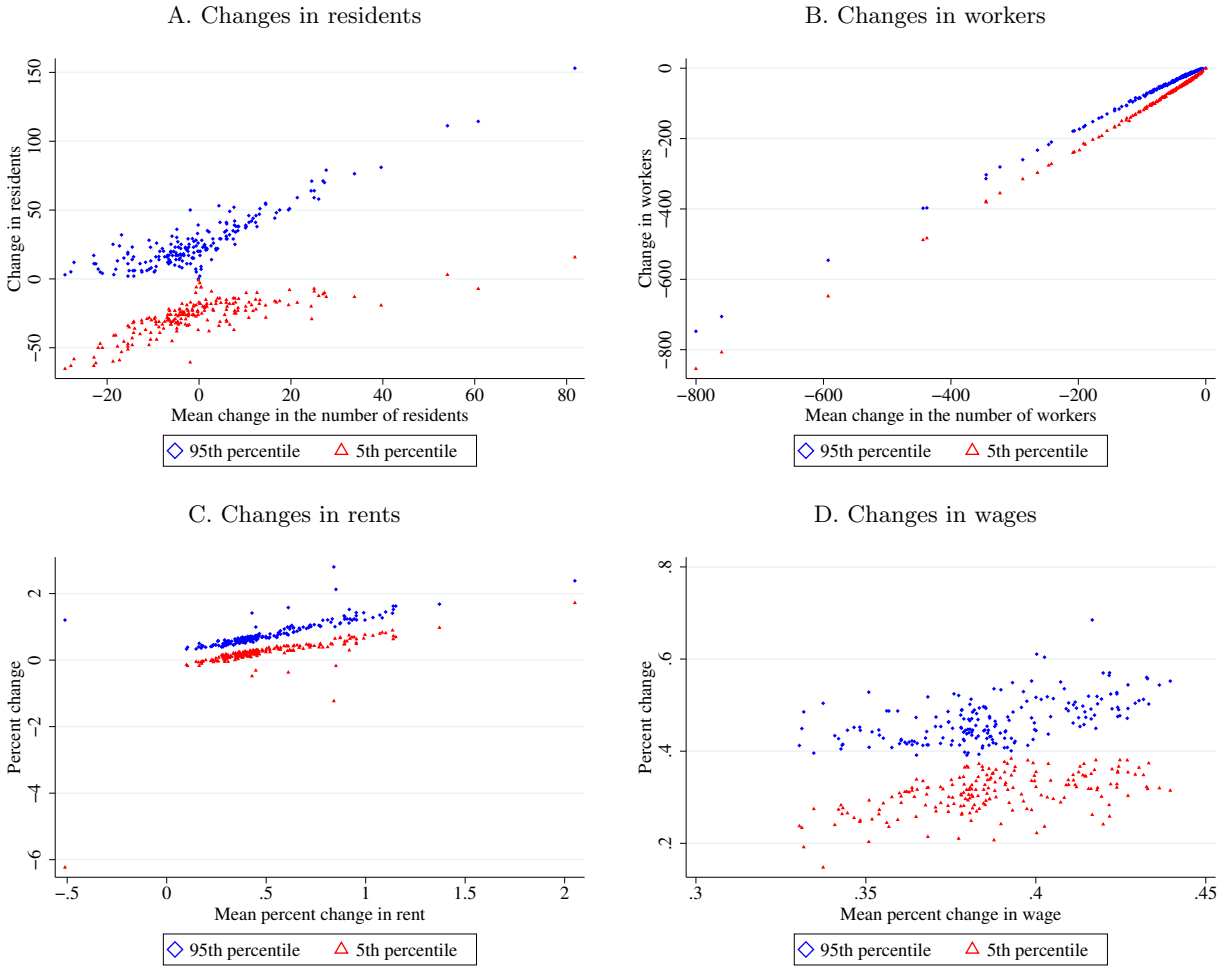
NOTES: This figure depicts the distribution of average tract-level outcomes by ventile of distance to the Amazon HQ2 tract. Each ventile contains about 108 residential tracts. Citywide average outcomes are shown with a dashed red line. Each point is the mean value of the ventile's outcomes across 100 simulations. The confidence intervals are constructed for each ventile from the 5^{th} and 95^{th} percentile outcomes of the 100 simulations. The percent change in real rents is defined as $\left(\frac{\hat{r}_k}{\hat{P}} - 1\right) \times 100$. 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 20^{th} 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



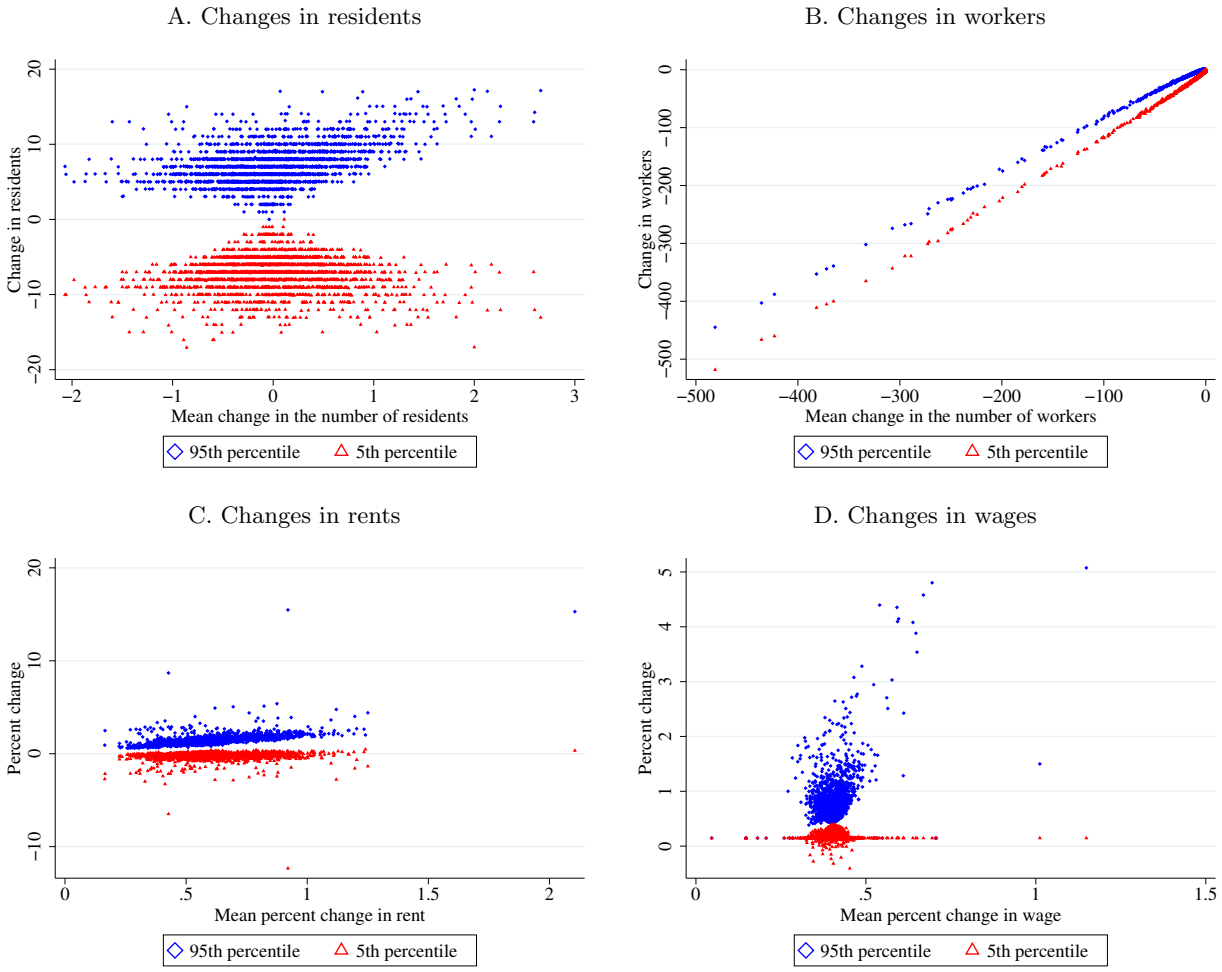
NOTES: This figure depicts the distribution of average tract-level outcomes by ventile of distance to the Amazon HQ2 tract. Each ventile contains about 107 workplace tracts. Citywide average outcomes are shown with a dashed red line. Each point is the mean value of the ventile's outcomes across 100 simulations. The confidence intervals are constructed for each ventile from the 5^{th} and 95^{th} 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)



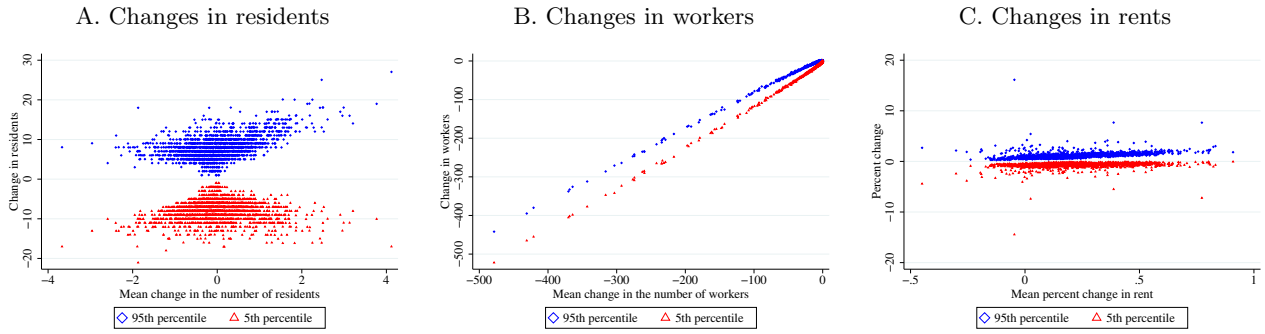
NOTES: These plots depict variation in counterfactual changes induced by individual idiosyncrasies in the NTA-level model with a finite number of individuals. The panel depicts the 5^{th} and 95^{th} percentiles of predicted changes in quantities and prices across 100 simulations of the NTA-level model 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^{th} and 95^{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 5^{th} and 95^{th} 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 505 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 24,777 to 25,349. In Panel C, real rent changes are $(\hat{r}_k/\hat{P} - 1) \times 100$, and 0 out of 2160 origin tracts have a positive change in rents at the 5^{th} percentile.