Spatial Economics for Granular Settings*

Jonathan I. Dingel
Chicago Booth, NBER, and CEPR

Felix Tintelnot
University of Chicago, NBER, and CEPR

29 September 2020

Abstract

We introduce a general-equilibrium model of a "granular" spatial economy populated by a finite number of people. Our quantitative model is designed for application to the growing body of fine spatial data used to study economic outcomes for regions, cities, and neighborhoods. Conventional approaches invoking the law of large numbers are ill-suited for such empirical settings. We evaluate quantitative spatial models' out-of-sample predictions using event studies of large office openings in New York City. Our granular framework improves upon the conventional continuum-of-individuals model, which perfectly fits the pre-event data but produces predictions uncorrelated with the observed changes in commuting flows. Applying our granular model to Amazon's proposed HQ2 in New York City reveals that the project's predicted consequences for most neighborhoods are small relative to the idiosyncratic component of individual decisions in this setting. We propose a convenient approximation for researchers to quantify the "granular uncertainty" accompanying their counterfactual predictions.

Keywords: commuting, granularity, gravity equation, quantitative spatial economics
JEL codes: C25, F16, R1, R13, R23

*We are grateful to Levi Crews and Mingjie Zhu for excellent research assistance. We thank Rodrigo Adão, Gabriel Ahlfeldt, Kirill Borusyak, Victor Couture, Teresa Fort, John Huizinga, Erik Hurst, Yuhei Miyauchi, Stephen Redding, Chris Severen, Daniel Sturm, and numerous conference and seminar participants for helpful feedback, and Thomas Holmes for his discussion at the NBER Summer Institute. 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. We gratefully acknowledge support from the National Science Foundation under grant number 2018609. jdingel@chicagobooth.edu and tintelnot@uchicago.edu

1 Introduction

Economists increasingly use quantitative spatial models to evaluate urban policies such as infrastructure investments and land-use planning decisions. The growing availability of economic data observed at increasingly finer spatial scales offers tremendous potential for new insights. However, if policymakers are to rely on these models to inform their decision-making, researchers must establish that they reliably capture relevant features of the data and perform well in making counterfactual predictions (Bryan, Glaeser, and Tsivanidis, 2019, p.31).

Quantitative spatial models characterize spatial linkages by the gravity equation, which says that the volume of interactions between any two locations increases with their sizes and decreases with the bilateral frictions between them (Redding and Rossi-Hansberg, 2017). The gravity equation, in turn, is derived under the assumption that the volume of interactions between any two locations is large enough to wash out any idiosyncratic choices of individuals. But when locations are defined at fine spatial scales, the interactions between them are "granular." That is, individual decision-makers are large relative to the economic outcomes being analyzed. In such settings, the idiosyncratic choices of individual decision-makers do not wash out. By assuming a continuum of individuals, the extant literature has neglected granularity's influence on its counterfactual predictions.

In this paper, we introduce a new framework for quantitative spatial analysis in granular settings. In doing so, we make four main contributions. First, we document that a number of empirical settings studied by economists feature commuting matrices that exhibit granularity. Second, we develop a quantitative spatial model that features a finite number of individuals and is suitable for counterfactual analysis. While incumbent models assume a continuum of individuals as a modeling convenience, our granular framework is just as tractable and matches the basic fact that no empirical setting has a continuum of individuals. Third, we assess the "out-of-sample" predictions of quantitative spatial models using event studies of neighborhood employment booms in New York City. Regressing observed changes in commuting flows on the models' predicted changes shows that our granular model predicts these changes better than the conventional procedure that fits the pre-event data perfectly. Fourth, we show that counterfactual predictions are accompanied by "granular uncertainty" due to the idiosyncratic components of equilibrium outcomes and propose a simple approximation of this dispersion. Applying this approach to an evaluation of Amazon's proposed HQ2 in New York City suggests that the predicted counterfactual changes for most census tracts are small compared to these outcomes' granular uncertainty.

Section 2 shows that granularity is common in empirical settings of interest. Focusing on the United States, we examine tract-to-tract and county-to-county commuting matrices used in prior research. These settings are sufficiently granular that the commuting matrices are sparse: half or more of geographic pairs—within reasonable commuting distance—have zero commuters. We document that the commuter counts vary considerably from one year to the next, which may be a symptom of granularity. This impersistence poses a challenge for calibration methods that exactly replicate the commuting flows observed in a given year. Furthermore, commuting matrices' zeros are often asymmetric. As a result, calibration procedures that rationalize zero-commuter

observations by infinite commuting costs impose severely asymmetric commuting costs despite the fact that daily commutes are round-trip journeys.

In Section 3, we propose a gravity-based general-equilibrium framework suitable for computing counterfactual outcomes in applications featuring granular commuting matrices. We follow the general-equilibrium approach employed in quantitative spatial models (e.g., Redding and Rossi-Hansberg 2017), but we assume that the number of individuals in the economy is finite. The key modeling challenge is that individual decisions affect wages and rents. We assume that individuals act as price takers and optimize given their beliefs about wages and rents. Employing standard discrete-choice methods (e.g., Train 2009), we model individuals choosing residence-workplace pairs on the basis of wage beliefs, rent beliefs, commuting costs, and an individual-specific idiosyncratic preference. We estimate the model assuming the price beliefs are "continuum-case rational expectations," which are the equilibrium prices in a model with the same parameters but a continuum of individuals. In our granular model, equilibrium outcomes depend in part on the idiosyncratic component of individuals' choices. Thus, where the continuum model would yield a single equilibrium allocation, our framework produces a *distribution* of economic outcomes instead. Notably, a single realization of the stochastic process often produces zeros for equilibrium outcomes that have strictly positive probabilities.

The continuum assumption has been made for reasons of convenience, not realism. Since our granular model is just as tractable as the continuum approach, it can be applied in the same empirical settings of interest. Section 4 shows that the model's parameters can be estimated using the same data on commuting flows typically employed in the estimation of continuum-of-individuals models. Moreover, since our granular framework coincides with the standard model in the limit, it can be used to quantify the role of granularity in economic outcomes and assess the (un)suitability of the continuum approximation.

In Section 5, we use our granular framework to analyze counterfactual scenarios. Our framework's predictions about counterfactual outcomes can differ considerably from the conventional calibrated-shares procedure that equates observed shares and underlying probabilities. Examining the "out-of-sample" performance of quantitative spatial models is an important step in establishing their value as an input into urban planning and other policymaking decisions.

First, we contrast these approaches' predictions for changes in commuting flows using observed outcomes after a local economic shock. Commuting flows are consistently reported in available data and the key spatial linkage between neighborhoods in many models of cities. We study events in which there is a large, discrete increase in employment in a single census tract, often arising from the arrival or expansion of a large employer such as Google or Tiffany & Co. We increase the productivity parameter for that tract to match the increase in total employment and compare the predicted changes in bilateral commuting flows to those observed in the data. When examining 78 tract-level employment booms in New York City, we find that our granular framework predicts the observed changes in commuting flows better than the conventional calibrated-shares procedure in 76 of these events. While predictive performance varies across events, regressing the observed

change in commuters on the granular model's prediction typically yields an estimated slope of one, whereas the estimated slope for the calibrated-shares procedure is often near zero or negative. In other words, the counterfactual predictions from the calibrated-shares procedure are *negatively* correlated with the observed changes in commuting in more than half of the 78 events.

At first glance, it may be surprising that our granular model predicts changes in commuter counts better than the procedure that calibrates parameters to perfectly fit the pre-event flows. For intuition, note that each observation in the pre-event cross section has three parts: a time-invariant component correlated with included covariates, a time-invariant component uncorrelated with those covariates, and a time-varying component. On the one hand, the advantage of the calibrated-shares procedure is its potential to capture real spatial linkages not predicted by covariates like transit times, such as the fact that a very large share of Columbia University employees reside in nearby university-owned residences. On the other hand, the calibrated-shares procedure matches every bit of variation in the data, including transitory individual idiosyncrasies. This risks overfitting the third component of the pre-event observations and failing to predict post-event outcomes. In our event studies, the overfitting problem dominates. A Monte Carlo simulation shows that granularity alone is sufficient to seriously impair the calibrated-shares procedure's predictive power when the model is otherwise correctly specified. We then explore the extent to which pooling several years of pre-event data can improve the models' predictions. While pooling three years of data improves the calibrated-shares procedure's predictions such that they are positively correlated with the observed changes in almost two-thirds of the events, the granular model still outperforms the calibrated-shares procedure in 65 of 78 events.

A second implication of granularity is granular uncertainty: given economic primitives, the number of individuals choosing a residence-workplace pair is a stochastic outcome due to their idiosyncratic preferences. Thus, counterfactual predictions involve both parameter uncertainty and granular uncertainty. We illustrate the importance of the latter in the context of evaluating the local economic effects of Amazon's proposed—but later abandoned—HQ2 in New York City. We find economically significant differences between the (mean) predictions of the granular model and the conventional calibrated-shares procedure, but we also find that most predicted tract-level changes are small relative to the accompanying granular uncertainty. We propose that researchers use a simple approximation based on the number of decision makers and the (change in) modeled probabilities to quantify the magnitude of granular uncertainty accompanying their counterfactual predictions. Our finding that granular uncertainty is large relative to the predicted neighborhood-level outcomes induced by a headline-grabbing headquarters decision suggests that granular uncertainty is likely important for counterfactual outcomes in many granular settings.

We contribute to the rapidly expanding literature on quantitative spatial economics.[1] A growing

[1] In gravity-based quantitative models, a location's access to all other locations' goods and factors, which is impeded by bilateral frictions, influences the magnitude and character of its economic activity. See Proost and Thisse (2019) for a recent survey of these models, Donaldson (2015) on market access in goods trade, Redding and Turner (2015) on transport costs for both goods and commuters, and Dingel (2017) on the income composition of goods market access.

body of research aims "to provide an empirically relevant quantitative model to perform general equilibrium counterfactual policy exercises" concerning topics such as transportation-infrastructure investment, neighborhood revitalization programs, and local business tax incentives (Redding and Rossi-Hansberg, 2017, p.23).[2] As anticipated by Holmes and Sieg (2015, p.106), "[a]s research in urban and regional applications takes advantage of new data sets at high levels of geographic resolution, it permits the study of interactions at narrow levels, where there may be relatively few decision-makers," but quantitative spatial models have largely abstracted from discreteness in the underlying economic environment.

Our event studies of new office openings contribute to a much smaller literature assessing the predictive power of quantitative spatial models using observed economic changes. Ahlfeldt et al. (2015) demonstrate their quantitative model's ability to capture how changes in floor prices associated with Berlin's division and reunification correlated with distance to the pre-war central business district. Monte, Redding, and Rossi-Hansberg (2018) verify a qualitative implication of their quantitative spatial model—larger employment responses to a local productivity increase in counties more open to commuting—using the million-dollar-plant openings examined by Greenstone, Hornbeck, and Moretti (2010). Adão, Arkolakis, and Esposito (2020) compare US commuting zones' actual and predicted changes in employment and wages in response to the "China shock" of Autor, Dorn, and Hanson (2013) for different calibrations of spatial links. Kreindler and Miyauchi (2020) use a gravity model and cell-phone–derived commuting flows to predict workplace wage levels in developing-economy cities that lack conventional data coverage. We explore the consequences of the widespread continuum assumption for the predictive power of quantitative spatial models in granular settings.

Finally, our framework is related to prior research about granular economies. A growing literature examines the importance of granularity for aggregate fluctuations in macroeconomics (Gabaix, 2011; Carvalho and Grassi, 2019) and international economics (di Giovanni and Levchenko, 2012; di Giovanni, Levchenko, and Mejean, 2014; Gaubert and Itskhoki, 2018; Daruich, Easterly, and Reshef, 2019). In addition to its relevance for aggregate fluctuations, firm-level granularity has been studied as one explanation for zeros in the international trade matrix (Eaton, Kortum, and Sotelo, 2013; Armenter and Koren, 2014).Mogstad et al. (2020) raise concerns about inference for ranks in granular settings. Our paper makes two contributions to these various strands of the literature. First, we provide a general-equilibrium modeling approach to granularity, namely continuum-case rational expectations, that is easy for practitioners to implement. Second, we demonstrate the importance of addressing granularity when making predictions about counterfactual outcomes and suggest a means of quantifying the granular uncertainty accompanying predicted changes.

[2]Heblich, Redding, and Sturm (2020); Severen (2019); Tsivanidis (2019); and Zárate (2019) study transportation-infrastructure investments. Owens, Rossi-Hansberg, and Sarte (2020) evaluate strategic visions to revitalize Detroit. Berkes and Gaetani (2020) compare the effects of siting Amazon's HQ2 in four different New York City neighborhoods.

2 Granular empirical settings

This section documents that granularity is common in empirical settings in which researchers and policymakers are interested in the incidence of local economic shocks. We primarily focus on the United States, examining tract-to-tract and county-to-county commuting matrices. After establishing that these settings are granular, we document two additional patterns that are likely symptoms of granularity: the commuter counts are impersistent over time, and zeros in the matrices are often asymmetric.

2.1 Data sources

We employ commuting matrices at two different spatial scales for the United States. First, we use data on commuting between census tracts taken from the Longitudinal Employer-Household Dynamics, Origin-Destination Employment Statistics (LODES). For illustrative purposes, we restrict our attention to pairs of tracts in the Detroit urban area and New York City.[3] Second, we use data on commuting between US counties taken from the American Community Survey (ACS), which surveys roughly 5% of the US population over a five-year period.[4]

An immediate implication of granularity is that some data providers perturb observations to protect confidentiality. For example, the published LODES tract-level workplace employment counts are noise-infused. Moreover, the published LODES commuting matrices report synthetically generated residence-workplace pairs.[5] 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.[6]

2.2 Tract-to-tract commuting matrices are granular

When working with fine spatial units like US census tracts, simply comparing the number of people and the number of residence-workplace pairs reveals that the empirical setting is granular. For example, the Detroit urban area has about 1.3 million pairs of census tracts and about 1.3 million commuters. Thus, the average cell in the commuting matrix contains one person. As 74% of tract pairs have zero commuters between them, the average positive cell contains about four people. As shown in the left panel of Figure 1, nearly half of the tract pairs with a positive number of

[3] 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. Davis et al. (2019) use the 2010 wave to compute the joint distribution of residences and workplaces for New York City. Following Owens, Rossi-Hansberg, and Sarte (2020), we use the primary-job counts in LODES.

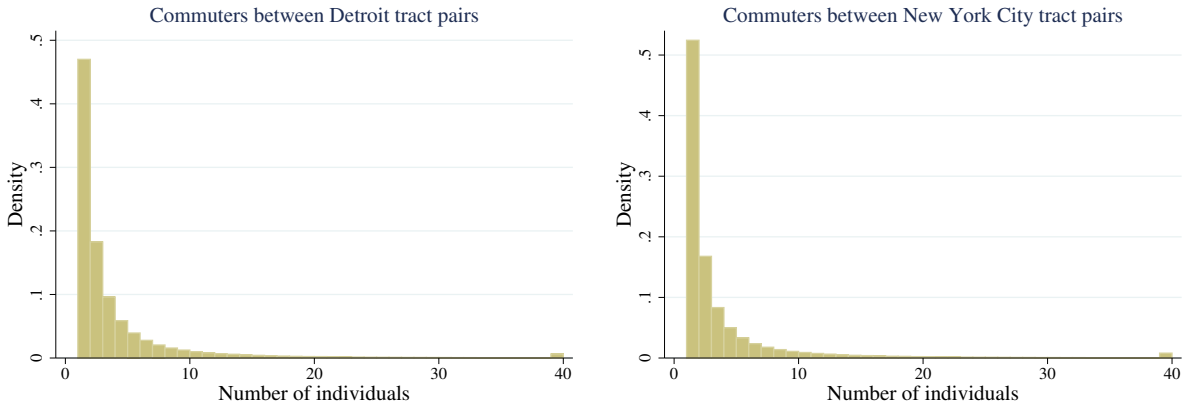
[4] Monte, Redding, and Rossi-Hansberg (2018) use the 2006–2010 wave of this survey to estimate a gravity model of commuting.

[5] 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."

[6] The econometric consequences of noise infusion and synthetic observations are specific to each use case. For example, see the discussion of these issues in Couture and Handbury (2019). Calibration approaches that perfectly match every (synthetic) observation are likely the research design most sensitive to these procedures.

commuters have only one commuter. 12.9% of Detroit commuters are the sole commuter in their cell of the commuting matrix. For such observations, a change in one individual's choice alters the commuting flow by 100%. 42.6% of Detroit commuters have five or fewer commuters in their cell of the commuting matrix. The right panel of Figure 1 shows similar patterns in New York City.[7]

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



NOTES: These histograms report the number of tract pairs in the Detroit urban area (left panel) and New York City (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. There are 1166 residential census tracts in Detroit. There are 2168 residential census tracts in New York City.

In fact, commuting matrices describing tract-to-tract flows for large US cities are inherently granular. Tracts are defined by the US Census Bureau such that the typical tract has 4,000 residents.[8] Thus, both the number of commuters and the number of tracts are roughly proportionate to the city's residential population, and the mean value of a cell in the tract-to-tract commuting matrix is inversely proportionate to that population.[9] In metropolitan areas with million of residents, the mean cell value is about one commuter. Since commuting flows take integer values, smaller mean values of cells make the commuting matrix necessarily granular and possibly even sparse.[10]

The spatial concentration of employment contributes to the sparsity of commuting matrices in settings with small spatial units. When these small spatial units are defined by street blocks or residential counts, the concentration of employment (relative to residences) implies that many

[7] About 2.5 million of New York City's 8 million residents are employed in New York City. Since New York City's commuting matrix has more than 4.6 million cells, its average cell contains slightly more than half a commuter, and this matrix is almost necessarily sparse. 40.7% of New York City commuters have five or fewer commuters in their cell of the commuting matrix.

[8] Geography Program glossary: "Census tracts generally have a population size between 1,200 and 8,000 people, with an optimum size of 4,000 people."

[9] To be more precise, in a city with R residents and I workers, there are about $N \approx R/4000$ tracts, and the mean value of each commuting cell in a tract-to-tract matrix is approximately $\frac{I}{N^2} = \frac{I}{R}\frac{16 \text{ million}}{R}$.

[10] The typical employed resident has only one place of employment. This imposes an integer constraint on commuting flows not relevant for trade flows or transaction flows, since firms can export to multiple destinations and shoppers can patronize multiple businesses.

workplace destinations have a smaller number of workers than the number of residential origins. The Detroit urban area illustrates the consequence. The median tract in the Detroit urban area 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.

Urban economists have studied other empirical settings in which the fine spatial resolution of the geographic units employed makes the commuting matrix necessarily granular. Roughly three-quarters of the cells in the Los Angeles metropolitan commuting matrix studied by Severen (2019) are empty.[11] Tsivanidis (2019) models the behavior of Bogota's 8 million residents commuting between almost 7 million pairs of tracts. Zárate (2019) examines Mexico City, which has 9 million people commuting between 13 million pairs of tracts. Ahlfeldt et al. (2015) model the behavior of about 3 million Berliners choosing among about 254 million pairs of city blocks.

2.3 The US county-to-county commuting matrix is granular

The commuting matrix for US counties is granular despite the large number of people relative to the number of county pairs. In the 2006–2010 ACS data, there are 136 million commuters (with commutes less than 120 kilometers).[12] 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. Figure 2 shows that this distribution is skewed, so that many thousands of county pairs have small numbers of commuters.

In practice, the granularity of this empirical setting 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 1. More than half of 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 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.[13]

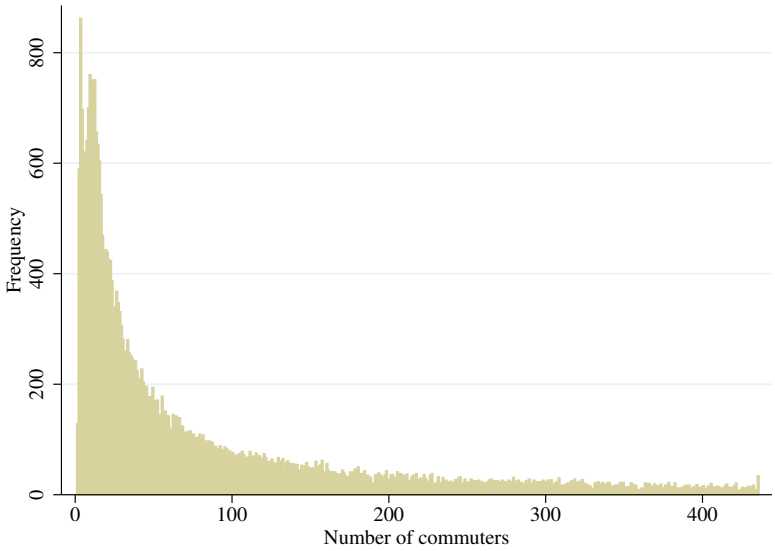
Commuting data from other countries also exhibit granularity. For example, one-quarter of Germany's county (Kreisfreie Städte and Landkreise) pairs within 120 kilometers have fewer than 10 commuters (Krebs and Pflüger, 2019). In Brazil's 2010 Censo Demográfico, which reports estimates based on a 10% sample of the population, there are 81 million commuters (with commutes

[11] Comparing Tables 2 and F4 of Severen (2019) shows that each column of the latter has about four times as many observations as its counterpart in the former.

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

[13] At face value, this would seem to imply that one could not reject the null hypothesis that there were no commuters for such county pairs. The unsuitability of the normal approximation for the binomial distribution in this case is another symptom of granularity.

Figure 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.

less than 60 kilometers).[14] 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.

[14]Dingel, Miscio, and Davis (2019) use these data to construct metropolitan areas based on commuting flows.

Table 1: Zeros in US county-to-county commuting matrix

Dataset	Zero Pairs	Positive Pairs	MOE >X (percent)
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 margin of error exceeds the (strictly positive) reported number of commuters.

2.4 Commuting counts are often impersistent

We now describe the first of two patterns that are likely symptoms of granularity. 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 granularity. 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 2, which presents the transition matrix for pairs of tracts in the Detroit urban area between the years 2013 and 2014, demonstrates considerable impersistence. 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 primarily is a symptom of the fact that in both years about three-quarters of observations are zero.[15] 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 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, as reported in Appendix A.

There is considerable churn even when using larger geographic units. Table 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 15% 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 symptom of granularity, these findings caution against procedures that infer structural parameters from the relative magnitudes of these counts. The difference between one commuter and two commuters (or one respondent and two respondents in a finite sample) is little evidence that the latter outcome was twice as probable. Similarly, procedures that rationalize observations with zero commuters

[15] 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 tract pairs in each period, nine-sixteenths of the pairs would be zero in both periods. That would not be evidence of persistence.

Table 2: The impersistence of commuting counts for pairs of tracts in Detroit and New York City

(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.

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.

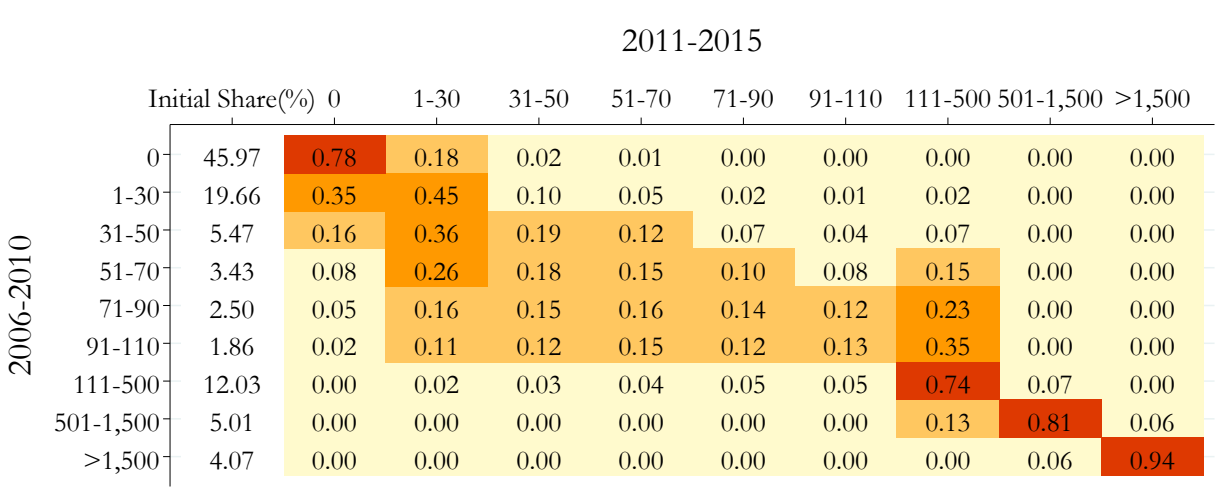
2.5 Zeros are often asymmetric

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

The fact that commuting matrices' zeros are often asymmetric poses a puzzle for calibration procedures that rationalize zero-commuter observations by infinite commuting costs. This interpretation of zeros implies severely asymmetric commuting costs, even though daily commutes are roundtrip 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 be a symptom of granularity 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.

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



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 2 shows no obvious bunching. We found similar impersistence when using alternative boundaries.

3 Model

This section introduces a model of granular residential and workplace location choices. We combine tools used to describe a finite number of individuals' decisions employed in the discrete-choice literature (e.g., Train 2009) with the general-equilibrium approach employed in quantitative spatial models (e.g., Redding and Rossi-Hansberg 2017). Extensions of the core model that introduce trade costs, residential amenities, local increasing returns, and production employing land are presented in Appendix B.

The key challenge in modeling a finite number of people is that individual decisions can affect wages and rents. This raises 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? In the interest of tractability, we assume that individuals have common point-mass beliefs about wages and rents. Therefore, individuals act as price takers, as in the granular economy of Gabaix (2011). In practice, we assume that these beliefs are "continuum-case rational expectations." In other words, given the model parameters, each individual forms their beliefs about wages and rents using the conventional continuum model, which is a limit case of our model as the number of individuals goes to infinity.

3.1 Setup

We consider a closed economy populated by I individuals. There is a set of locations indexed by k or n. Each location k is endowed with a quantity of land, T_k, that is owned by immobile

landlords who consume only goods.[16] Each location is endowed with the technology to produce a differentiated good (i.e., the Armington assumption). For simplicity, we abstract from trade costs.

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

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

where w_n denotes the wage in location n, Φ_k denotes the price index of location k, $\delta_{kn} \geq 1$ denotes the commuting cost between k and n, U_{kn} denotes the mean utility of choice kn, and ν_{kn}^i is the idiosyncratic utility draw of individual i for living in k and working in n. The parameter ϵ governs the importance of mean utility relative to the idiosyncratic utility draw. Given Cobb-Douglas preferences, the price index in location k is $\Phi_k = r_k^\alpha P^{1-\alpha}$, where r_k denotes the local land price, and P denotes the common CES price index for goods. We assume that ν_{kn}^i is drawn from a type 1 extreme value distribution.

Production of each location's differentiated good is linear in labor. The goods market is perfectly competitive.[17] Each individual is endowed with L/I units of labor, such that the aggregate labor endowment is L. Commuting is costly because time spent commuting is not spent working, so that individuals residing in k and working in n earn only w_n/δ_{kn} because they only spend $1/\delta_{kn}$ of their time working.[18] Workers produce in location n with a linear production technology: $q_n = A_n L_n$, where L_n is the quantity of labor supplied by workers working in n and A_n is that location's productivity. Given this production function and perfect competition, the price of location n's output is w_n/A_n for all consumers. Thus, the CES price index is $P = \left[\sum_n \left(w_n/A_n\right)^{1-\sigma}\right]^{1/(1-\sigma)}$.

We make the following assumptions about information and expectations. All workers know the primitives of the model (L, $\{A_n\}$, $\{T_k\}$, $\{\delta_{kn}\}$, α, ϵ, σ) and have (common) expectations about the equilibrium variables r_k and w_n. We assume that the price expectations are point-mass beliefs, such that each individual assigns 100% probability to a single vector of wages and a single vector of land prices. Denote these belief vectors by $\{\tilde{w}_n\}$ and $\{\tilde{r}_k\}$. 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 $\{\tilde{r}_k\}$,

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

[17] Given established isomorphisms between perfect-competition and monopolistic-competition models of economic geography (e.g., Allen and Arkolakis 2014), this assumption is not crucial to our results. See discussion in Appendix B.

[18] Alternatively, we could assume that δ_{kn} represents a utility shifter that does not affect working time. Modeling δ_{kn} as reduced working time facilitates the commuting-elasticity estimation discussed in Section 4.

each worker chooses the residential location and the work location that maximize expected utility

$$\tilde{U}^i_{kn} = \underbrace{\epsilon \ln\left(\frac{\tilde{w}_n}{\tilde{P}^{1-\alpha}\tilde{r}_k^{\alpha}\delta_{kn}}\right)}_{\equiv \tilde{U}_{kn}} + \nu^i_{kn}, \tag{2}$$

where $\tilde{P} = \left[\sum_n (\tilde{w}_n/A_n)^{1-\sigma}\right]^{1/(1-\sigma)}$. After these decisions are made, workers are immobile and cannot relocate.[19] 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.

The assumption that individuals have point-mass beliefs about wages, $\{\tilde{w}_n\}$, and rents, $\{\tilde{r}_k\}$, considerably simplifies the analysis. Otherwise, individuals would need to compute the prices associated with all possible residence-workplace allocations and then solve for the fixed-point probabilities that each of these occurs. The set of feasible allocations is large. For example, a granular economy with only 10 individuals and 16 workplace-residence pairs would have more than 3 million possible allocations.[20] For empirically relevant magnitudes, computing the number of possible allocations leads to such large numbers that they cause overflow problems in standard software. While our approach is general in the sense that it allows for any kind of point-mass beliefs about wages and rents, in our analysis below we assume that the expected prices are the equilibrium wages and rents in a model with the same economic primitives (L, $\{A_n\}$, $\{T_k\}$, $\{\delta_{kn}\}$, α, ϵ, σ) and a continuum of individuals ($I \to \infty$).

3.2 Equilibrium

For expositional clarity, we distinguish between a *trade equilibrium*, which clears markets taking individuals' locations as given, and a *granular commuting equilibrium*, in which a finite number of individuals choose their locations based on beliefs about the trade equilibrium that will result.

Goods market clearing equates each location's output to the quantity demanded. If ℓ_{kn}/δ_{kn} denotes the labor supply of individuals living in k and working in n, then output in location n is $A_n \sum_k \ell_{kn}/\delta_{kn}$. 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. The demand for each differentiated good stemming from CES preferences means that equating quantity supplied

[19] 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. Brinkman and Lin (2019), Heblich, Redding, and Sturm (2020), and Redding and Rossi-Hansberg (2017) make the same irreversibility assumption. Similarly, Tsivanidis (2019) assumes that residences and workplaces are chosen sequentially.

[20] 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.

and quantity demanded requires

$$A_n \sum_k \frac{\ell_{kn}}{\delta_{kn}} = \frac{(w_n/A_n)^{-\sigma}}{P^{1-\sigma}} \sum_{k',n'} \frac{\ell_{k'n'}}{\delta_{k'n'}} w_{n'} \qquad \forall n. \tag{3}$$

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:

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

We next define the *trade equilibrium*, which clears markets taking individuals' locations as given.

Definition 3.1. Trade equilibrium. Given the labor allocation $\{\ell_{kn}\}$ and economic primitives (L, $\{A_n\}$, $\{T_k\}$, $\{\delta_{kn}\}$, α, ϵ, σ), a trade equilibrium is a set of wages $\{w_n\}$ and land prices $\{r_k\}$ such that equations (3) and (4) are satisfied.

Remark. Given the labor allocation $\{\ell_{kn}\}$, there is a unique set of wages and land prices satisfying equations (3) and (4). Since $\sigma > 1$, there is a unique set of wages satisfying equation (3), by application of Theorem 1 of Allen, Arkolakis, and Takahashi (2020). Equation (4) can be rewritten as $r_k = \frac{\alpha}{T_k} \sum_n \frac{\ell_{kn}}{\delta_{kn}} w_n$, so there is a unique set of land prices associated with that wage vector.

Individuals' choices of location are made on the basis of their beliefs about the subsequent trade equilibrium. Given belief vectors $\{\tilde{w}_n\}$ and $\{\tilde{r}_k\}$ and the distribution of $\{\nu^i_{kn}\}$, equation (2) implies that the probability of any individual choosing kn as her residence-workplace pair is

$$\Pr(U^i_{kn} > U^i_{k'n'} \ \forall (k',n') \neq (k,n)) = \frac{\tilde{w}_n^{\epsilon} \left(\tilde{r}_k^{\alpha} \delta_{kn}\right)^{-\epsilon}}{\sum_{k',n'} \tilde{w}_{n'}^{\epsilon} \left(\tilde{r}_{k'}^{\alpha} \delta_{k'n'}\right)^{-\epsilon}}. \tag{5}$$

With these probabilities in hand, we define a granular commuting equilibrium as the labor allocation resulting from drawing I realizations of this process and clearing markets.

Definition 3.2. Granular commuting equilibrium. Given a number of individuals I, economic primitives (L, $\{A_n\}$, $\{T_k\}$, $\{\delta_{kn}\}$, α, ϵ, σ), and a set of point-mass beliefs ($\{\tilde{w}_n\}$, $\{\tilde{r}_k\}$), a granular commuting equilibrium is a labor allocation $\{\ell_{kn}\}$, wages $\{w_n\}$, and land prices $\{r_k\}$ such that

- $\ell_{kn} = \frac{L}{I} \sum_{i=1}^{I} \mathbf{1}\{\tilde{U}^i_{kn} > \tilde{U}^i_{k'n'} \ \forall (k',n') \neq (k,n)\}$ is the labor allocation resulting from I independent draws from the probability mass function in equation (5); and
- wages $\{w_n\}$ and land prices $\{r_k\}$ are a trade equilibrium, per Definition 3.1, given the labor allocation $\{\ell_{kn}\}$.

Distinguishing between the number of individuals I and the aggregate labor supply L allows us to apply the law of large numbers to locational decisions without changing aggregate labor supply.

As Appendix B.1 shows, since I is the number of draws from the probability mass function in equation (5), fixing L and taking $I \to \infty$ causes the labor allocation in the limiting case to be

$$\ell_{kn} = L \times \Pr(U^i_{kn} > U^i_{k'n'} \ \forall (k', n') \neq (k, n)) = L \times \frac{\tilde{w}_n^{\epsilon} \left(\tilde{r}_k^{\alpha} \delta_{kn}\right)^{-\epsilon}}{\sum_{k',n'} \tilde{w}_{n'}^{\epsilon} \left(\tilde{r}_{k'}^{\alpha} \delta_{k'n'}\right)^{-\epsilon}}. \tag{6}$$

Note that this limiting labor allocation still depends on beliefs. We now define the idea that, for this limiting case, rational expectations are point-mass beliefs that coincide with the resulting wages and prices.

Definition 3.3. Rational expectations for the continuum case. Given economic primitives (L, $\{A_n\}$, $\{T_k\}$, $\{\delta_{kn}\}$, α, ϵ, σ), $\{\tilde{w}_n\}$ and $\{\tilde{r}_k\}$ are "continuum-case rational expectations" if $\{\tilde{w}_n\}$ and $\{\tilde{r}_k\}$ constitute a trade equilibrium for the labor allocation $\{\ell_{kn}\}$ given in equation (6).

As $I \to \infty$, if individuals' point-mass beliefs are continuum-case rational expectations, then the model coincides with the standard continuum model of Redding and Rossi-Hansberg (2017). That is, combining equations (3) and (4) with equation (6) evaluated at $\tilde{w}_n = w_n$ and $\tilde{r}_k = r_k$ is the standard spatial-equilibrium model with a continuum of individuals.[21]

For finite I, a given set of economic primitives (L, $\{A_n\}$, $\{T_k\}$, $\{\delta_{kn}\}$, α, ϵ, σ) and point-mass beliefs ($\{\tilde{w}_n, \tilde{r}_k\}$) generate a *distribution* of equilibria associated with realizations of the process generating ν^i_{kn} and hence $\{\ell_{kn}\}$.[22]

Even though our implementation of the granular model uses the continuum model to generate expectations for wages and rents, our granular framework differs from the conventional model in fundamental ways. First, given the same economic primitives (L, $\{A_n\}$, $\{T_k\}$, $\{\delta_{kn}\}$, α, ϵ, σ), the granular model generates integer-valued outcomes, including zeros, unlike the continuum model. When commuting costs are a parsimonious function of finite transit times, the continuum model generically cannot generate such outcomes. Second, applying the continuum logic to granular settings leads to very different estimates of the economic primitives that suffer overfitting problems, as discussed in the following sections.

4 Estimation

In this section, we estimate the granular model using the tract-to-tract commuting data from New York City discussed in Section 2. We also discuss estimation of the continuum model that we will compare to the granular model.

[21] In the absence of agglomeration forces, this quantitative spatial model with a continuum of individuals has a unique equilibrium (see Ahlfeldt et al. (2015)). Therefore, the continuum-case rational expectations of wages and land rents are unique.

[22] Redding and Rossi-Hansberg (2017, p.38) anticipated this implication of granularity: "At smaller spatial scales (e.g., blocks within cities), one might expect such random idiosyncratic factors to be more important relative to the systematic deterministic components of a model (e.g., natural resource abundance) than at larger spatial scales (e.g., across regions or countries)."

We set the values of a few parameters prior to estimating those remaining. We assume each individual supplies one unit of labor, so that $L = I$ is equal to the number of employed individuals. Following the literature, we impose $\alpha = 0.24$, based on Davis and Ortalo-Magne (2011), and $\sigma = 4$, following Monte, Redding, and Rossi-Hansberg (2018).

For commuting costs, we distinguish between observed and unobserved components in order to encompass both the parameterization we employ when estimating our model and specifications employed in the existing literature.

$$\delta_{kn} = \underbrace{\bar{\delta}_{kn}}_{\text{observed}} \times \underbrace{\lambda_{kn}}_{\text{unobserved}}$$

We construct the observed part of the commuting costs by assuming that each worker has H hours that is spent either working or commuting. We compute $\bar{\delta}_{kn} = \frac{H}{H - t_{kn} - t_{nk}}$, where $H = 9$ and t_{kn} is the transit time from k to n according to Google Maps.[23] A wide variety of assumptions about the unobserved component λ_{kn} are possible. To highlight the contrast between our granular model and the incumbent literature, we focus on two extreme cases.[24] To estimate the granular model, we assume that there is no unobserved component, $\lambda_{kn} = 1 \; \forall k, n$. To parameterize the continuum model, we will assume that the unobserved component is appropriately mean-independent of $\bar{\delta}_{kn}$ and takes whatever values rationalize observed commuting counts.

Given observed values of commuting costs $\delta_{kn} = \bar{\delta}_{kn}$, the granular model's remaining parameters can be estimated by maximum likelihood. Using the probability mass function in equation (5) and denoting the number of individuals who chose the kn pair by ℓ_{kn}, the log likelihood function is

$$\ln \mathcal{L} \equiv \sum_k \sum_n \ell_{kn} \ln \left[\Pr(U^i_{kn} > U^i_{k'n'} \; \forall (k', n') \neq (k, n)) \right] = \sum_k \sum_n \ell_{kn} \ln \left[\frac{\tilde{w}^{\epsilon}_n \left(\tilde{r}^{\alpha}_k \delta_{kn} \right)^{-\epsilon}}{\sum_{k',n'} \tilde{w}^{\epsilon}_{n'} \left(\tilde{r}^{\alpha}_{k'} \delta_{k'n'} \right)^{-\epsilon}} \right].$$

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. As shown by Guimarães, Figueirdo, and Woodward (2003), the maximization of this likelihood function is numerically equivalent to a Poisson pseudo-maximum-likelihood estimator (PPMLE) that is available for a variety of software packages.[25]

The continuum model's commuting elasticity ϵ is estimated by combining equation (6) with a mean-independence assumption about the unobserved component λ_{kn}. Two distinct assumptions

[23] 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. Using $H = 8$ or $H = 10$ yields very little change in model fit relative to the $H = 9$ results reported in Table 4. For New York City, we use Google Maps public-transit times collected by Davis et al. (2019). We impute missing observations by predicting transit times using physical distance, imputing transit times for fewer than 4.1% of tract pairs in New York City.

[24] Among many intermediate cases, one might assume that λ_{kn} has a random-effects or an interactive-fixed-effects structure. Similarly, one could enrich $\bar{\delta}_{kn}$ by introducing more observed covariates or more flexible functions of transit times.

[25] See Sotelo (2019) for a discussion of the relationship between the multinomial and Poisson pseudo-maximum-likelihood estimators in the context of gravity models. In practice, we use the Stata package `ppmlhdfe` of Correia, Guimarães, and Zylkin (2020).

might be made. If one assumes $\mathbb{E}\left(\lambda_{kn}^{-\epsilon}|\tilde{r}_k, \tilde{w}_n, \bar{\delta}_{kn}\right) = 1$, the constant-elasticity function can be estimated using the Poisson pseudo-maximum-likelihood estimator, as shown in Silva and Tenreyro (2006). In this case, the granular and continuum models yield the same estimate of the commuting elasticity ϵ.[26] If one takes the logarithm of each side of equation (6), assumes $\mathbb{E}\left(\ln \lambda_{kn}|\tilde{r}_k, \tilde{w}_n, \bar{\delta}_{kn}\right) = 0$, 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. As emphasized 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 for the granular model is presented in column 1 of Table 4. 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 utility draws are modestly dispersed. Estimating the commuting elasticity using OLS yields a much lower elasticity, likely due to the well-understood selection bias associated with omitting three-quarters of observations from the estimation sample.[27] In what follows, we use the estimate reported in column 1 for both the granular and continuum models.

Table 4: Commuting elasticity estimates

	NYC (2010)	
	MLE	OLS
Commuting cost	-7.986	-2.307
	(0.307)	(0.0516)
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 "MLE" columns present the results from the maximum likelihood estimation described in the text. The "OLS" columns present the results of estimating the log version of equation (6) 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.

For the granular model, the remaining economic primitives $\{T_k\}$ and $\{A_n\}$ can be inferred from the maximum-likelihood estimates under the assumption of continuum-case rational expectations. Given values of the commuting elasticity ϵ and the land expenditure share α, individuals' beliefs about land prices and wages are transformations of the estimated fixed effects. In particular, in equation (5) the residence fixed effect is proportionate to $\tilde{r}_k^{-\alpha\epsilon}$ and the workplace fixed effect is proportionate to $\tilde{w}_n^{\epsilon}$. If $\{\tilde{r}_k\}$ and $\{\tilde{w}_n\}$ are continuum-case rational expectations, plugging these

[26] While both approaches arrive at the PPMLE, their economic assumptions differ. Note that the discrete-choice model of McFadden (1974) associated with equation (1) assumes that idiosyncratic tastes are identically distributed, while Silva and Tenreyro (2006) propose using the PPMLE to estimate constant-elasticity models in the presence of heteroskedastic λ_{kn} errors.

[27] This selection bias also has substantial implications for the estimated fixed effects, as discussed in Appendix C. Recent research in industrial organization draws attention to the problems associated with zero-valued market shares (Quan and Williams, 2018; Gandhi, Lu, and Shi, 2019; Dubé, Hortaçsu, and Joo, 2020).

estimated beliefs into equation (6) yields the continuum-case labor allocation. Given α, σ, $\{\tilde{r}_k\}$, $\{\tilde{w}_n\}$, $\{\delta_{kn}\}$ and that continuum-case labor allocation, equations (3) and (4) can be solved to yield values of $\{T_k\}$ and $\{A_n\}$.

For the continuum model, one must make further assumptions about the unobserved commuting cost λ_{kn} in order to infer the model's remaining parameters. The most common practice is to not restrict the distribution of λ_{kn} but assume that these parameters take values such that the continuum model exactly replicates the observed commuter counts. As we describe in the following section, the parameters $\{T_k\}$, $\{A_n\}$, and $\{\lambda_{kn}\}$ are not separately identified in this case, but one can compute counterfactual outcomes based on combinations of these parameters that are identified.

5 Counterfactual analysis and event studies

Researchers employing quantitative spatial models aim to provide a parsimonious general-equilibrium framework in order to predict counterfactual outcomes.[28] The value of these quantitative models for policy analysis therefore depends largely on their ability to predict the spatial distribution of economic consequences of a shock.[29] This section contrasts the predictions of our granular framework and the conventional procedure for both observed outcomes of past events and counterfactual outcomes of interest in New York City.

5.1 Counterfactual analysis in spatial models

Prior research using quantitative spatial models with a continuum of individuals has made either of two extreme assumptions about the unobserved component of commuting costs, λ_{kn}. Researchers who do not observe commuting flows between the geographic units they study, such as Ahlfeldt et al. (2015), have assumed that there is no such component, $\lambda_{kn} = 1 \; \forall k, n$. Researchers studying settings where bilateral commuting data are available have typically not employed this assumption, likely because the continuum model's commuting flows cannot match the observed data without a structural residual. Instead, the far more common approach to counterfactual analysis in quantitative spatial models is to choose values of λ_{kn} and other parameters such that the model's probabilities exactly replicate the observed quantities. This approach is known as "calibrated shares form" or "exact hat algebra" in the international trade literature (Rutherford, 1995; Dekle, Eaton, and Kortum, 2008; Costinot and Rodríguez-Clare, 2014). Recent studies employing this technique include Allen,

[28] Redding and Rossi-Hansberg (2017, p.23): "[T]his research does not aim to provide a fundamental explanation for the agglomeration of economic activity, but rather to provide an empirically relevant quantitative model to perform general equilibrium counterfactual policy exercises." Waddell and Sarte (2016, p.190): "The development of the new quantitative equilibrium models has initiated a more robust and realistic framework with which to model cities... it will also allow for more robust counterfactual policy exercises that can inform practitioners and policymakers regarding strategies for urban development."

[29] Bryan, Glaeser, and Tsivanidis (2019, p.31): "If quantitative models are to provide useful policy insights, their results have to be trusted. First, researchers must establish that their model captures relevant features of the data or (ideally) can replicate the real-world response to a policy change." More broadly, Kehoe (2003) argues that "it is the responsibility of modelers to demonstrate that their models are capable of predicting observed changes, at least ex post."

Arkolakis, and Li (2015); Perez-Cervantes (2016); Waddell and Sarte (2016); Monte, Redding, and Rossi-Hansberg (2018); Heblich, Redding, and Sturm (2020), Krebs and Pflüger (2019); Severen (2019); and Owens, Rossi-Hansberg, and Sarte (2020).

While not all the parameters of quantitative spatial models are identified by the calibrated-shares procedure, it allows one to compute the counterfactual equilibrium outcomes associated with a proportionate changes in these parameters. In Appendix D, we apply this calibrated-shares procedure to the continuum-of-individuals variant of the model in Section 3. Given a labor allocation $\{\ell_{kn}\}$, wages $\{w_n\}$, labor demand elasticity σ, and commuting elasticity ϵ, one can compute proportionate changes in wages, rents, and labor allocations for any combination of proportionate changes in productivity $\{\hat{A}_n\}$, land $\{\hat{T}_k\}$, and commuting costs $\{\hat{\delta}_{kn}\}$. Implicitly, the procedure calibrates initial combinations $\{A_n\}$, $\{T_k\}$, and $\{\delta_{kn}\}$ that rationalize the (observed) initial equilibrium values of $\{\ell_{kn}\}$ and wages $\{w_n\}$. A natural concern is that this procedure might overfit the observed data at the cost of predicting how the economy responds to changes.[30]

By contrast, we estimate the parameters of our granular model without perfectly equating model probabilities and observed outcomes, as described in Section 4. Since model probabilities and equilibrium outcomes are not identical in our approach, there is a distribution of counterfactual equilibrium outcomes associated with counterfactual parameter values. In particular, different values of the idiosyncratic preference shocks $\{\nu^i_{kn}\}$ generate different equilibrium allocations and prices for the same counterfactual values of economic primitives. In practice, we numerically simulate the model many times and present moments that summarize this distribution of counterfactual outcomes.

5.2 Event studies: Predicting observed commuting responses to a local shock

As a first means of contrasting the calibrated-shares procedure with our granular model, we examine each approach's predictions for large, discrete changes in employment in individual census tracts in New York City. We investigate changes in commuting patterns to workplace tracts that had substantial employment growth. Commuting flows are spatial links that govern quantitative spatial models' predictions about the incidence of local economic shocks on outcomes like land prices. In particular, we conduct an "event study" for each of the 78 workplace tracts in New York City that had a two-year increase in total employment from 2010 to 2012 of at least 500 employees and at least 15% from a 2010 level of at least 500 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

[30]Most notably, this procedure infers that unchosen outcomes are zero-probability events, thereby assigning infinite commuting costs to residence-workplace pairs with zero commuters. More broadly, equating relative probabilities with relative counts means that structural parameters are often inferred from "lumpy" comparisons of one versus two versus three commuters.

Avenue. These locations are two of the 78 workplace tracts we examine.[31]

5.2.1 Contrasting predictions for 78 tract-level employment booms

We estimate or calibrate the model based on 2010 data for New York City. We estimate our granular model using data on tract-to-tract commuting flows by maximum likelihood, as described in Section 4. The calibrated-shares procedure perfectly rationalizes observed commuting flows and wages, as described in Appendix D.[32] Next, we compute the increase in productivity required to match the observed 2010–2012 change in employment for the "treated" tract. Since the two procedures fit the 2010 data differently, this productivity increase need not be the same.[33] 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.

Figure 3 contrasts the two procedures' predictions for the changes in commuter flows associated with all 78 employment booms from 2010 to 2012. For each event and each 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 granular model, the slope coefficients are roughly centered on one (median of 0.95) and the intercept coefficients are roughly centered on zero (median of 0.02). The calibrated-shares procedure does not perform as well. Across the 78 events, the median slope coefficient is -0.24 and the median intercept coefficient is 0.57. That is, the calibrated-shares procedure's predictions are negatively correlated with observed outcomes in more than half the events. As a result, the granular model typically has a lower forecast error. Panel C of Figure 3 contrasts the two models' root mean squared errors (RMSE) for each event.[34] The granular model has a lower RMSE than the calibrated-shares procedure in 76 of the 78 events. While the calibrated-shares procedure necessarily has better in-sample fit, it performs poorly when predicting changes in commuting flows.

Panel B of Figure 3 shows that the predictions from the continuum-limit of the granular model are very similar to the mean of 1,000 simulations of the granular model. The granular model's superior predictive power is not due solely to the presence of zeros in the commuting matrix. To

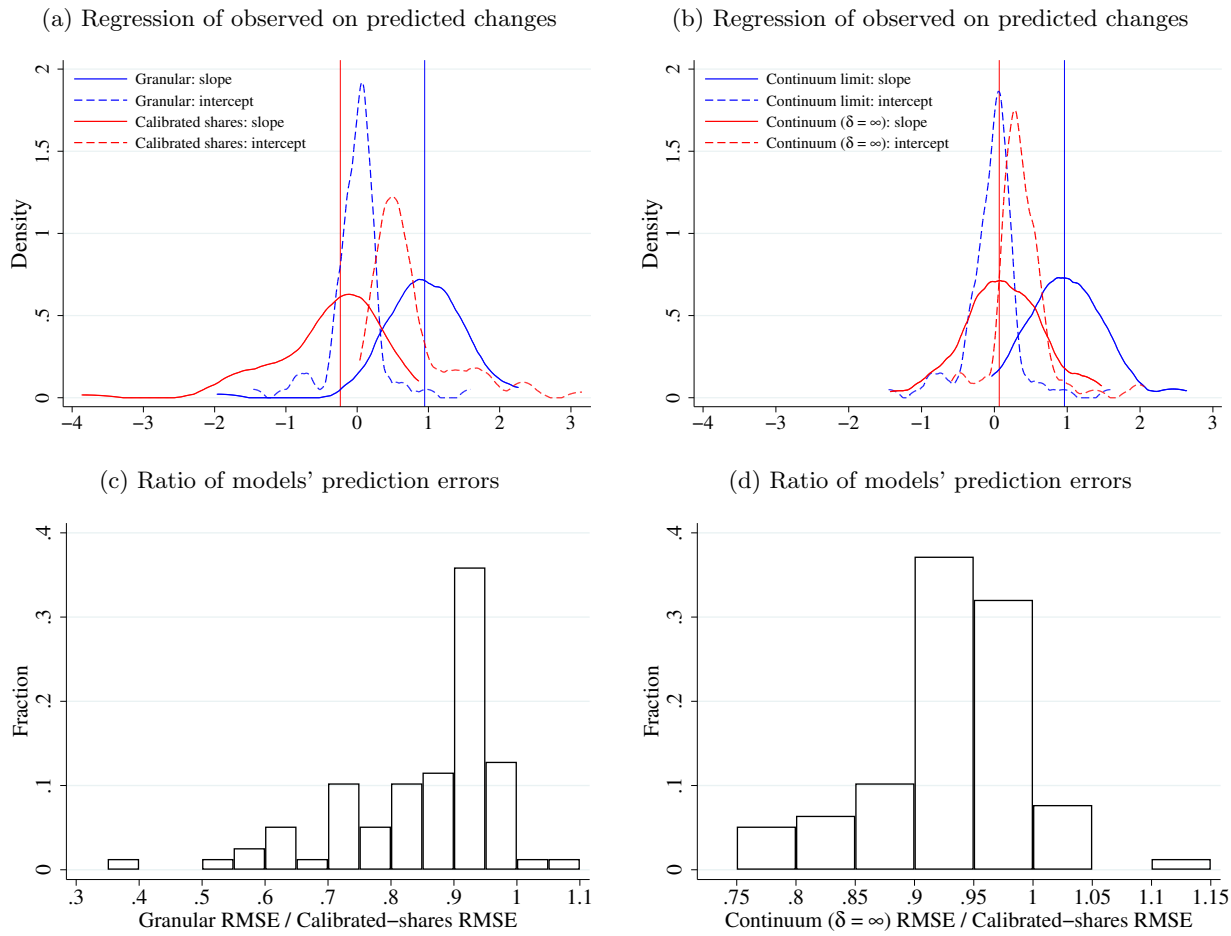
[31] Figure D.1 depicts the employment changes for these two tracts.

[32] The largest commuting flow in the 2010 LODES data for New York City is 827 commuters who reside between 110th and 114th Streets in Morningside Heights and work at adjacent Columbia University. The estimated granular model 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 advantage of the calibrated-shares procedure is its potential to capture such linkages.

[33] For the granular model, we compute the productivity increase using the limiting $I \to \infty$ case. This slightly reduces the computational burden of searching over potential productivity increases. We have verified that the average of 1,000 or more granular counterfactual simulations is very similar to the limiting case (under the same parameter values).

[34] Comparing mean absolute errors (MAE) yields the same conclusion. Characterizing forecast errors using RMSE or MAE is preferable to a measure such as R^2. As a squared correlation, R^2 would not distinguish between forecasts that are positively or negatively correlated with outcomes nor reveal a constant bias in a forecast. If expressed in terms of the sum of squared prediction errors relative to the variance of the observed changes, this "R^2" would be negative when the sum of squared prediction errors is larger, obviating the analogy with the in-sample fit of a linear regression.

Figure 3: Comparison of models' predictive performance across 78 events



NOTES: "Granular" in panels A and C refers to the mean of 1,000 simulations of the granular model. "Calibrated shares" in panels A, C, and D refers to the procedure defined in Appendix D. "Continuum limit" in panel B refers to the $I \to \infty$ limit of the granular model. "Continuum ($\delta = \infty$)" in panels B and D is a hybrid model in which $\delta_{kn} = \bar{\delta}_{kn}$ if $\ell_{kn} > 0$ in the 2010 data and $\delta_{kn} = \infty$ if $\ell_{kn} = 0$ in the 2010 data. In panel 3a, 2 outlying calibrated-shares observations not depicted. In panel 3b, 1 outlying continuum observation and 1 outlying continuum ($\delta = \infty$) observation not depicted. Vertical lines depict medians of slope coefficient distributions.

illustrate this, we compare the predictions to those of a granular model in which commuting costs are infinite for residence-workplace pairs that have zero commuters in the baseline 2010 data. The infinite-commuting-costs variant of the granular model reported in panel D of Figure 3 produces predictions that have lower root mean square errors than the predictions of the calibrated-shares procedure. Since the two procedures depicted in that panel handle zero-commuter flows identically, the granular model's superior predictive performance is due in part to better predicting commuter-count changes for residence-workplace pairs that have positive counts in the baseline 2010 data.

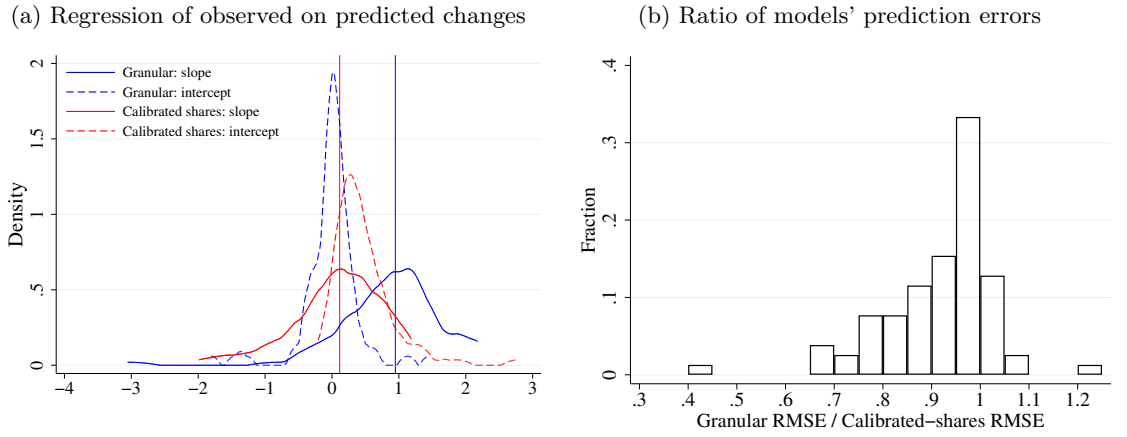
5.2.2 The role of granularity in the calibrated-shares procedure's predictive failure

Why does the calibrated-shares procedure that perfectly fits the pre-event data produce worse predictions than the parsimoniously parameterized granular model? A Monte Carlo exercise described in detail in Appendix D.3 suggests that this failure could be entirely due to granularity. In the simulations, the data-generating process is our estimated model of New York City in 2010, and we impose a counterfactual productivity increase in one tract. With a continuum of individuals, the calibrated-shares procedure would perfectly predict the associated changes in commuting flows. In practice, the calibrated-shares procedure has very limited predictive power when applied to a finite ($I \approx 2.5$ million) sample drawn from the data-generating process. Regressing simulated observed changes on the change predicted by the calibrated-shares procedure yields a slope of only 0.21 in the median simulation. The slope of the granular model is centered on one, as expected, and its RMSE is 91% of the calibrated-shares forecast error in the median simulation. Thus, granularity alone can severely limit the calibrated-shares procedure's predictive power.

5.2.3 Alternative prediction strategies

As an alternative to using a granular model, one might consider strategies to "smooth" granular data to be fed into the continuum model. For example, one might hope to eliminate idiosyncratic elements of the data by averaging across multiple years of data. In many empirical settings, particularly historical contexts, consecutive years of data are not available. In the case of the LODES data, annual data are available since 2002, and pooling may average out noise introduced by the confidentiality-protecting perturbations mentioned in Section 2.

Figure 4: Comparison of models' predictive performance across 78 events (pooled pre-event data)



NOTES: In panel A, 5 outlying calibrated-shares observations not depicted. Vertical lines depict medians of slope coefficient distributions.

To explore the gain from pooling multiple years of data, we average commuter and wage observations for 2008–2010 before applying our granular model and the calibrated-shares procedure. As depicted in Figure 4, pooling the data yields a modest improvement for the calibrated-shares

procedure. Its predictions are now positively correlated with observed outcomes for more than 60% of the events. Nonetheless, the granular model typically forecasts the changes in commuter counts much better. The granular model's slope coefficients are closer to one (median of 0.95 vs. 0.11) and its intercept coefficients are closer to zero (0.01 vs. 0.38). The granular model applied to the pooled data has a lower RMSE than the pooled calibrated-shares procedure in 65 of the 78 events.

The predictive power of the granular model could potentially be improved further. Recall that our implementation of the granular model uses only transit times to characterize commuting costs, leaving many unused degrees of freedom. Depending on the context, one could add borough-pair fixed effects, interactive fixed effects, or additional covariates such as the frequency of public-transit connections, vehicular transit times, and so forth. By contrast, the calibrated-shares procedure is the extreme case of introducing a structural error λ_{kn} for each pair of locations, thereby overfitting the observed data and failing to predict out of sample.

5.3 The effects of Amazon HQ2 on New York City neighborhoods

Having established that the granular model predicted commuting changes better than the conventional calibrated-shares procedure, we now examine these procedures' predictions for the economic consequences of Amazon's aborted second headquarters ("HQ2") in Long Island City, a controversial and widely discussed proposal. We focus on understanding the procedures' contrasting predictions and investigating the magnitude of granular uncertainty accompanying these predictions, rather than a full examination of the potential effects of Amazon's HQ2. We therefore employ the baseline model of Section 3, which abstracts from productivity spillovers or skill heterogeneity among workers.[35]

In September 2017, Amazon requested proposals for a second headquarters from cities and states across North America. 238 proposals were evaluated in a year-long competition. In November 2018, the company 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). The former was to have 4 million square feet office space, with anticipated expansion opportunities for another 4 million square feet and 15,000 more workers.[36] 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.

5.3.1 Predicting counterfactual outcomes

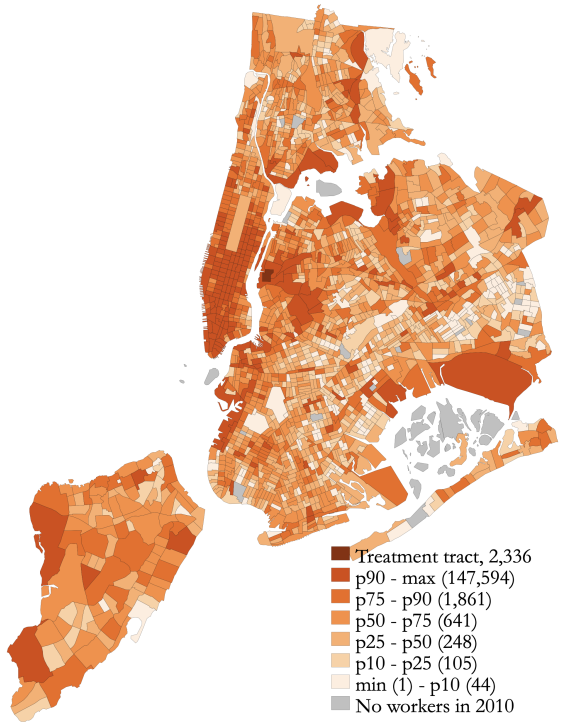
What general-equilibrium effects of Amazon HQ2 would quantitative spatial models predict? Using each of the two methods, we analyze the consequences of a productivity increase in the Long Island City tract that would cause employment to rise by 25,000 workers from the 2010 baseline levels.[37] Figure 5a depicts employment by tract in 2010, which is particularly concentrated in

[35] Berkes and Gaetani (2020) introduce a model that features these mechanisms and analyze Amazon HQ2.
[36] More details can be found in the Memorandum of Understanding between the State, City and Amazon.
[37] Following Berkes and Gaetani (2020), the "treated" tract is 36081000700.

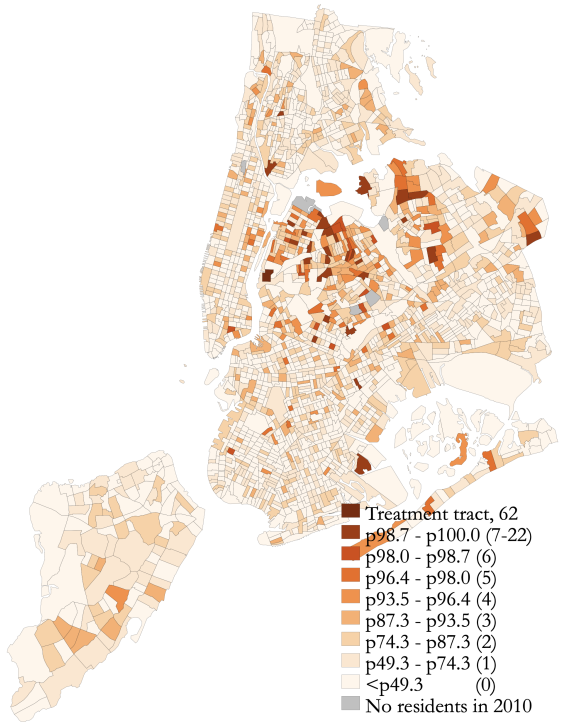
lower Manhattan and Queens. Figure 5b depicts the distribution of residences among those workers employed in the treated Long Island City tract in 2010. Notably, 49.3% of residential tracts have zero residents employed at that workplace tract. The calibrated-shares procedure rationalizes those zeros with infinite commuting costs.

Figure 5: New York City employment in 2010

(a) Number of workers



(b) Number of residents working at the treatment tract



NOTES: Panel 5a depicts the number of workers in each tract. Panel 5b depicts the number of residents in each tract who work in the treatment tract in the LODES 2010 data. That legend reports the percentiles associated with the integer number of residents represented by each color.

The predicted change in the number of residents in each tract due to the productivity increase in Long Island City is depicted in Figure 6a for the granular model and in Figure 6b for the calibrated shares approach. There is a striking contrast between the two methods' predicted residential changes. The granular model's predicted changes are modest in size and closely related to transit times to Long Island City. The calibrated-shares procedure's predicted changes are very large for some residential tracts and are closely tied to the initial numbers of residents working in the treated tract, as depicted in Figure 5b.[38] As a consequence, the according to the calibrated shares approach rents in the top decile of tracts would increase between 1.6 and 10.9 percentage points (Figure 7b). By comparison, the predicted rent increase for the top decile of affected tracts is quantitatively more modest between 0.8 and 1.3 percentage points (Figure 7a). Whereas the calibrated shares

[38]Figure D.3 in the Appendix depicts the close relationship between the initial number of residents working in the treated tract and the calibrated-shares procedure's predicted change in residents.

approach suggests that 49% of the workers in Long Island City are located in Queens, the granular model suggests only 33% of workers at the Amazon 2nd headquarter location would be located in Queens. The change in the number of workers by tract is given by Figures 6c and 6d. There are almost no differences in the predictions between the two methods and the predictions are closely tied to the initial number of workers by tract from Figure 5a. Relatedly, the real wage changes are relatively similar between the calibrated share and granular model, ranging between 0.35 and 0.67 percentage points for the non-treatment tracts (see Figures 7c and 7d).

5.3.2 Granular uncertainty

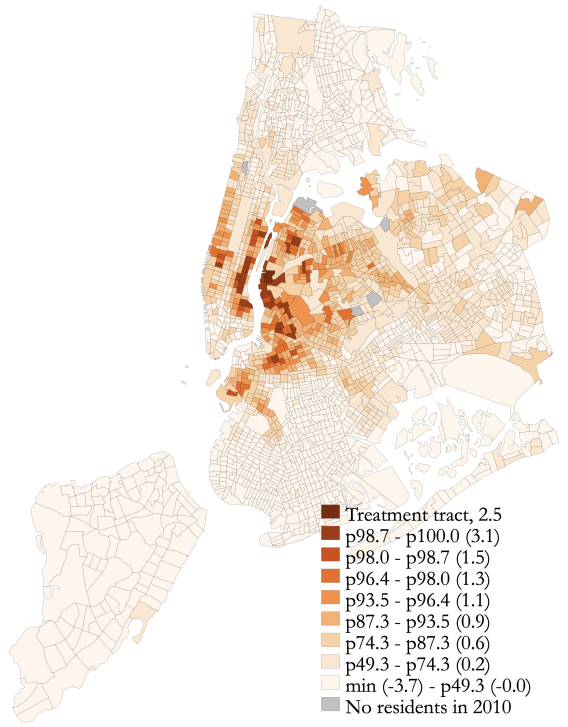
While the granular model's average prediction across many simulations is very similar to the prediction when evaluated at the granular model's continuum limit, one important implication of granularity is that there is a distribution of possible equilibrium outcomes. That is, counterfactual predictions are subject to granular uncertainty even if the model primitives (L, $\{A_n\}$, $\{T_k\}$, $\{\delta_{kn}\}$, α, ϵ, σ) are known with certainty. The magnitude of this granular uncertainty relative to the predicted counterfactual changes depends both on the empirical setting and the counterfactual shock examined. We consider the creation of 25,000 jobs in one tract to be a large shock, as Amazon's HQ2 choice was a high-profile corporate decision that elicited numerous proposals from local policymakers. Are the counterfactual changes in outcomes across different parts of New York City large relative to the granular uncertainty driven by the idiosyncratic components of individuals' location decisions?

To characterize the granular uncertainty around the granular model's average prediction for tract-level responses to the counterfactual arrival of Amazon HQ2 in Long Island City, we simulate the granular model's counterfactual outcomes 10,000 times. In particular, we simulate 10,000 labor allocations induced by the idiosyncratic preference parameters given the counterfactual economic primitives and solve for the accompanying 10,000 equilibria. From each of these counterfactual equilibrium outcomes, we subtract the mean outcome of the granular economy prior to the counterfactual change. We use the dispersion in the change in equilibrium outcomes across simulations to characterize the granular uncertainty that should accompany predictions about counterfactual changes.

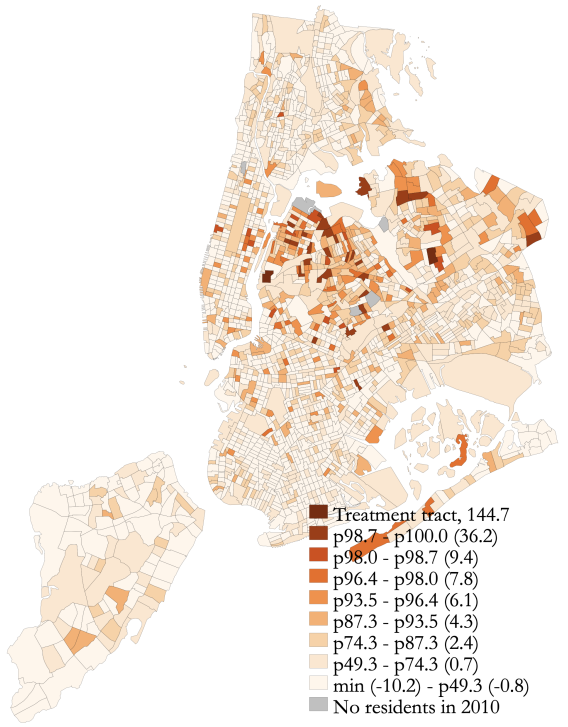
Before turning to the granular uncertainty revealed by these simulations, we discuss a simple approximation of granular uncertainty for the labor allocation that is cheap to compute and valid when the empirical setting is not "too granular." In the granular model, the number of individuals who choose residence-workplace pair kn, ℓ_{kn}, is a realization of I draws from a multinomial distribution over $K \times N$ outcomes with probabilities given by equation (5). Denote by p_k and p'_k the baseline and counterfactual probabilities of choosing any pair in which the residence is k. The standard deviation of $\sum_n \ell_{kn}$ is $\frac{L}{\sqrt{I}}\sqrt{p_k(1-p_k)}$, and thus the standard deviation of the change $\sum_n \Delta\ell_{kn}$ when $I = L$ is $\mathfrak{s} \equiv \sqrt{I}\sqrt{p_k(1-p_k) + p'_k(1-p'_k)}$. Thus, in the $I = L$ case, we can approximate the 90% confidence interval for the change in the number of residents in tract k by $I \times (p'_k - p_k) \pm 1.645\mathfrak{s}$. Using the normal distribution to approximate the distribution of outcomes

Figure 6: Predicted change in the number of residents or workers

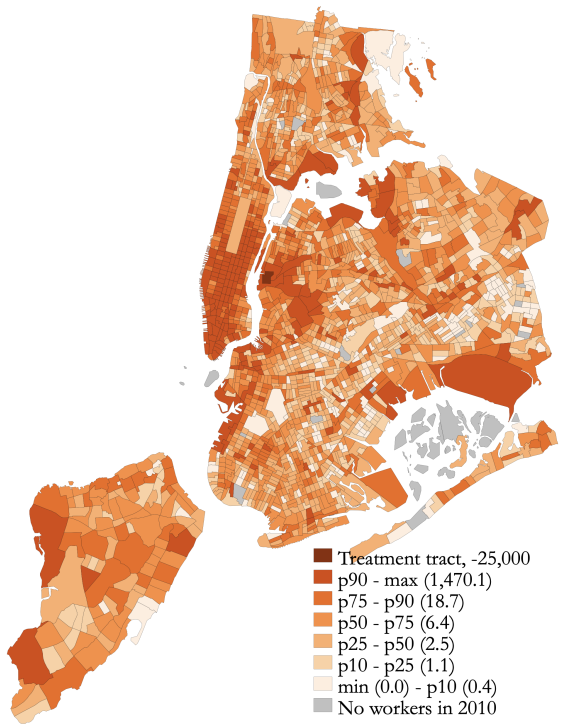
(a) Change in residents, granular model



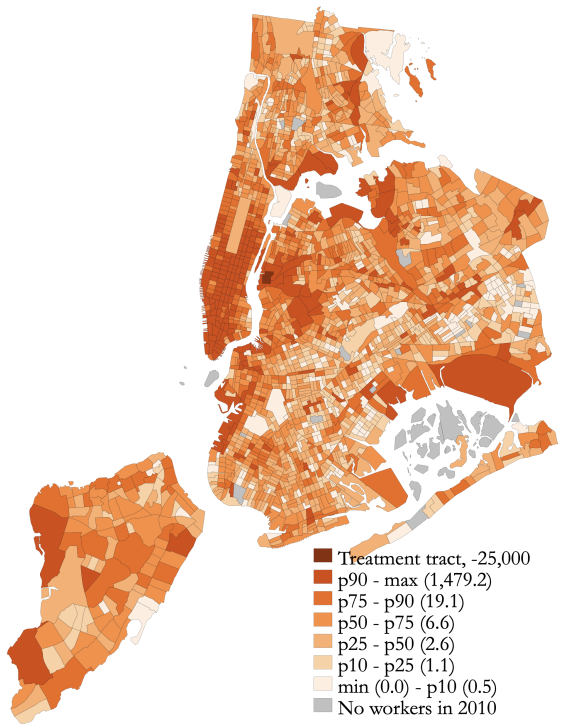
(b) Change in residents, calibrated-shares



(c) Decrease in workers, granular model



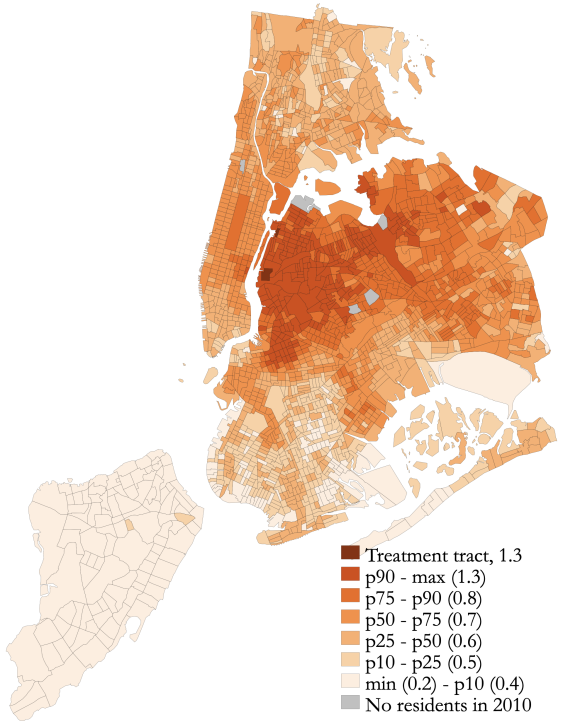
(d) Decrease in workers, calibrated-shares



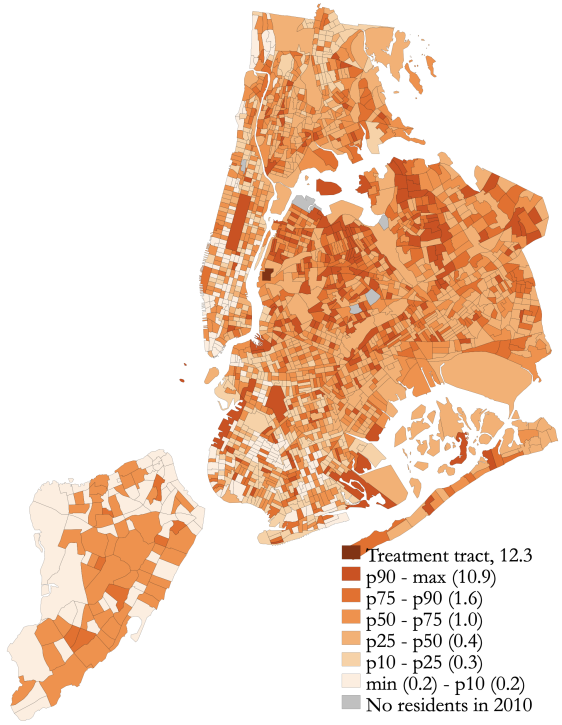
NOTES: Panels 6a and 6b depict (mean) changes in the number of residents predicted by the granular model and calibrated-shares procedure, respectively. Panels 6c and 6d depict (mean) decreases in the number of workers predicted by the granular model and calibrated-shares procedure, respectively. Note that the number of workers decreases in all tracts except the Amazon HQ2 location. The granular model's predicted mean changes are computed using the continuum limit ($I \to \infty$), since these average quantities are linear functions of the model probabilities. The legend percentile cutoffs in panels A and B correspond to those in Figure 5b; those in panels C and D to Figure 5a.

Figure 7: Predicted changes in prices

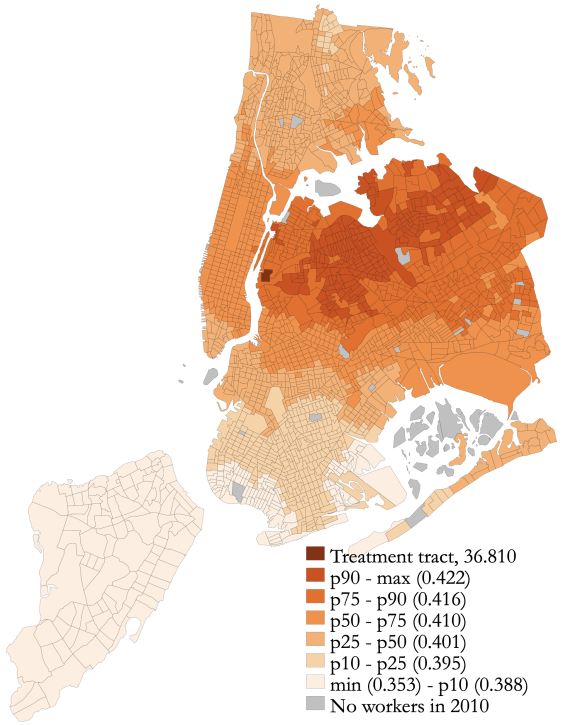
(a) Change in rent, granular model



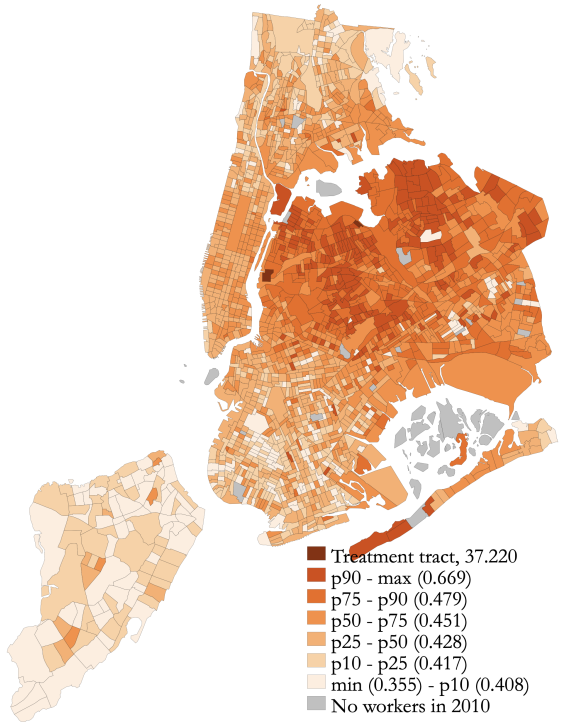
(b) Change in rent, calibrated-shares procedure



(c) Change in wage, granular model



(d) Change in wage, calibrated-shares procedure



NOTES: These maps depict percentage-point changes in prices. Panel 7a and 7b depict each tract's predicted change in rent (r_k/P) using the continuum limit $(I \to \infty)$ of the granular model and the calibrated-shares procedure, respectively. Panel 7c and 7d depict each tract's predicted change in wage (w_n/P) using each approach.

for tract k works well when $I \times p_k$ is sufficiently large. When the expected value of the outcome is small, the approximated confidence intervals for the number of people may include negative values. These obviously infeasible outcomes would be evidence that the empirical setting is so granular that the approximation is inappropriate, warranting further investigation of the granular uncertainty.

Using this approximation reveals that the predicted changes in residents produced by both the granular model and the calibrated-shares procedure are small relative to the granular uncertainty in this setting. As shown in Figure 6, the predicted changes overwhelmingly involve fewer than three residents in the granular model and fewer than ten residents in the calibrated-shares procedure. For the 2160 tracts in New York City with roughly equal number of residents, $\sqrt{p_k(1-p_k)+p'_k(1-p'_k)} \approx \sqrt{2 \times \frac{1}{2160} \times \frac{2159}{2160}}$, so the width of the approximated confidence interval for a tract with the average number of residents is about 158.[39] Thus, the granular uncertainty is roughly an order of magnitude greater than the single-digit predicted changes in the number of residents. As Figure 8's panel A shows, the distribution of changes in number of residents predicted by the granular model is virtually indistinguishable from a wide interval centered on zero for all tracts in New York City.

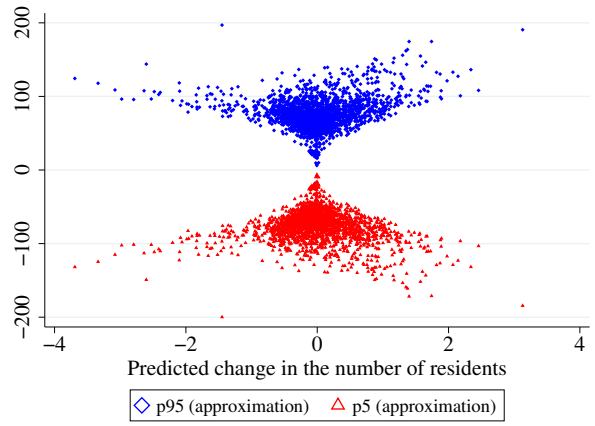
Since the granular model and calibrated-shares procedure predict large changes in the number of workers for some tracts, some of these predictions have the potential to be meaningfully large relative to the magnitude of granular uncertainty. Since the number of workers varies across tracts much more than the number of residents, the width of the approximated confidence intervals also varies much more. Figure 8's panel B depicts these approximated confidence intervals for the changes in the number of workers. Zero change lies below the 95^{th} percentile for all workplaces.

Next, we summarize the distributions of the predicted changes in residents and workers across 10,000 simulations of the granular model. While requiring more computations, this simulation-based approach relaxes the assumptions of the previous approximation. For example, the number of commuters choosing a residence-workplace pair cannot be negative. We find that the simulated confidence intervals are typically narrower than those employing the approximation based on the normal distribution. This makes little difference for changes in residents, as seen by comparing panels A and C of Figure 8, since the granular uncertainty is so large relative to predicted changes. In the case of changes in workers, however, the narrower confidence intervals produced by the simulations allow us to say that some predicted changes are large relative to granular uncertainty. For tracts predicted to lose more than 250 workers, the change is sufficiently large relative to granular uncertainty that even the 90^{th} percentile of changes is negative. For most workplaces, however, the approximated 90% confidence interval includes both negative and positive values. Thus the granular uncertainty involved in studying neighborhood-level outcomes is large relative to the predicted consequences of the most hotly followed corporate headquarters decision in recent memory. This suggests that quantitative spatial analyses of counterfactual outcomes must address the economic uncertainty that accompanies granular settings.

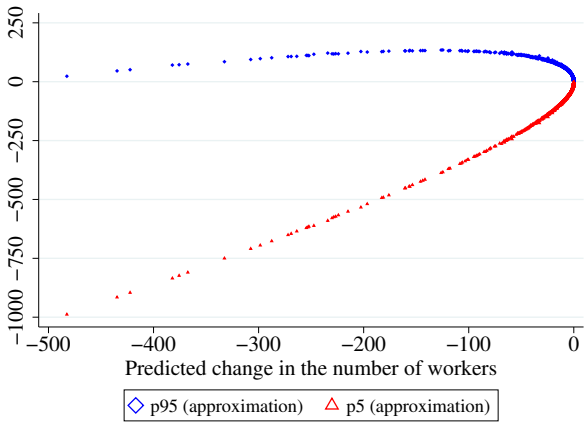
[39] $2 \times 1.645 \times \sqrt{I}\sqrt{2 \times \frac{1}{2160} \times \frac{2159}{2160}} \approx 3.29 \times \sqrt{I} \times \sqrt{\frac{2}{2160}} \approx 3.29 \times 1577.6 \times 0.03 \approx 158.$

Figure 8: Distribution of predicted change in the number of residents or workers

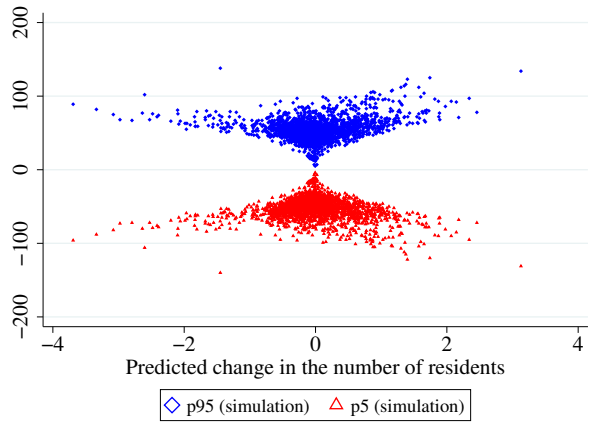
(a) Distribution of change in residents, approximation



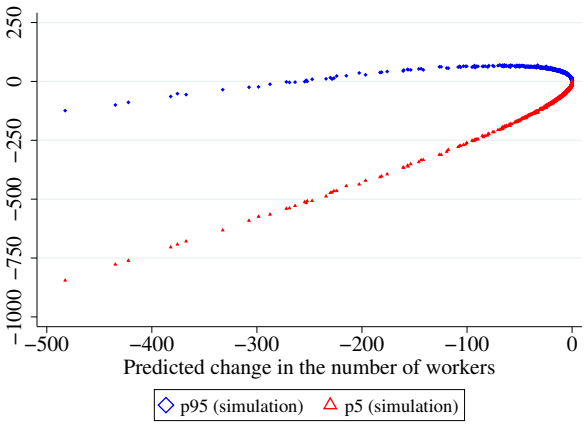
(b) Distribution of change in workers, approximation



(c) Distribution of change in residents, simulation



(d) Distribution of change in workers, simulation



NOTES: The plots depict the 5^{th} and 95^{th} percentiles of predicted change in the number of residents or workers under the granular model, by approximation and simulation. For the approximation method, we follow the steps mentioned in the main text. Panel 8a plots for the predicted change in the number of residents and panel 8b plots for the predicted change in the number of workers. For the simulation method, we simulate 10,000 equilibria of the counterfactual under the granular model, calculate values of the 5^{th} and 95^{th} percentiles, and subtract the continuum-limit of the granular model at baseline parameter values. Panel 8c plots for the predicted change in the number of residents and panel 8d plots for the predicted change in the number of workers. Treated tract 36081000700 and 1 outlying tract excluded in panel 8b and 8d.

6 Conclusion

Economists increasingly have access to spatially fine data and use quantitative spatial models to compute counterfactual general-equilibrium outcomes. The smaller the number of individuals behind each economic outcome reported in these data, the less compelling the conventional modeling assumption that there is a continuum of individuals. We document that empirical settings of interest exhibit granular commuting matrices and discuss the consequences of such granularity for economic modeling, estimation, and predicting counterfactual outcomes. In short, the conventional calibrated-shares procedure risks overfitting in-sample observations at the cost of failing to predict out-of-sample outcomes. We need to evaluate the performance of these applied general-equilibrium models in predicting observed changes if policymakers are going to rely on them to inform their decisions (Kehoe, 2003; Bryan, Glaeser, and Tsivanidis, 2019).

In this paper, we propose a granular model that can generate patterns of outcomes like those in the data, show how to estimate its parameters by maximum likelihood, and contrast its predictions about local economic shocks to those of the conventional continuum model. Assuming that commuting costs for millions of residence-workplace pairs are a parsimonious function of observed covariates rather than millions of pair-specific parameters guards against overfitting granular data. Studying 78 tract-level employment booms in New York City during 2010–2012, we find that our granular model predicts changes in commuting flows better than the calibrated-shares procedure.

Since all empirical settings feature a finite number of individuals, the continuum assumption has been made in the interest of modeling convenience, not realism. Our granular model is equally tractable and leverages the same data as the continuum approach, so it can be applied in the same settings in which economists have thus far assumed a continuum. Moreover, since our granular framework coincides with the standard model in the limit, it can be used to quantify the role of granularity in economic outcomes and assess the (un)suitability of the continuum approximation. Since we have provided evidence that a great variety of spatial settings exhibit granularity, this quantitative framework should find widespread application.

References

Adão, Rodrigo, Costas Arkolakis, and Federico Esposito. 2020. "General Equilibrium Indirect Effects in Space: Theory and Measurement."

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. 2015. "Optimal City Structure."

Allen, Treb, Costas Arkolakis, and Yuta Takahashi. 2020. "Universal Gravity." *Journal of Political Economy*, 128(2): 393–433.

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

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

Autor, David H., David Dorn, and Gordon H. Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review*, 103(6): 2121–68.

Berkes, Enrico, and Ruben Gaetani. 2020. "Income Segregation and Rise of the Knowledge Economy."

Brinkman, Jeffrey, and Jeffrey Lin. 2019. "Freeway Revolts!" Federal Reserve Bank of Philadelphia Working Papers 19-29.

Bryan, Gharad, Edward Glaeser, and Nick Tsivanidis. 2019. "Cities in the Developing World." National Bureau of Economic Research Working Paper 26390.

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

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, and Jessie Handbury. 2019. "Urban Revival in America."

Daruich, Diego, William Easterly, and Ariell Reshef. 2019. "The surprising instability of export specializations." *Journal of Development Economics*, 137: 36 – 65.

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, and Andrei A. Levchenko. 2012. "Country Size, International Trade, and Aggregate Fluctuations in Granular Economies." *Journal of Political Economy*, 120(6): 1083–1132.

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

Dingel, Jonathan I. 2017. "The Determinants of Quality Specialization." *Review of Economic Studies*, 84(4): 1551–1582.

Dingel, Jonathan I, Antonio Miscio, and Donald R Davis. 2019. "Cities, Lights, and Skills in Developing Economies." National Bureau of Economic Research Working Paper 25678.

Donaldson, Dave. 2015. "The Gains from Market Integration." *Annual Review of Economics*, 7(1): 619–647.

Dubé, Jean-Pierre H, Ali Hortaçsu, and Joonhwi Joo. 2020. "Random-Coefficients Logit Demand Estimation with Zero-Valued Market Shares." National Bureau of Economic Research Working Paper 26795.

Eaton, Jonathan, Samuel Kortum, and Sebastian Sotelo. 2013. "International Trade: Linking Micro and Macro." In *Advances in Economics and Econometrics: Tenth World Congress*. Vol. III, , ed. Eddie Dekel Daron Acemoglu, Manuel Arellano. Cambridge University Press.

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

Gandhi, Amit, Zhentong Lu, and Xiaoxia Shi. 2019. "Estimating Demand for Differentiated Products with Zeroes in Market Share Data."

Gaubert, Cecile, and Oleg Itskhoki. 2018. "Granular Comparative Advantage." National Bureau of Economic Research Working Paper 24807.

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.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. 2010. "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings." *Journal of Political Economy*, 118(3): 536–598.

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

Heblich, Stephan, Stephen J Redding, and Daniel M Sturm. 2020. "The Making of the Modern Metropolis: Evidence from London." *The Quarterly Journal of Economics*. qjaa014.

Holmes, Thomas J., and Holger Sieg. 2015. "Structural Estimation in Urban Economics." In *Handbook of Regional and Urban Economics*. Vol. 5 of *Handbook of Regional and Urban Economics*, , ed. Gilles Duranton, J. Vernon Henderson and William C. Strange, 69 – 114. Elsevier.

Kehoe, Timothy J. 2003. "An evaluation of the performance of applied general equilibrium models of the impact of NAFTA." *Staff Report*.

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

Kreindler, Gabriel, and Yuhei Miyauchi. 2020. "Measuring Commuting and Economic Activity inside Cities with Cell Phone Records."

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

McFadden, Daniel L. 1978. "Modelling the Choice of Residential Location." In *Spatial Interaction Theory and Planning Models*. , ed. A. Karlqvist, L. Lundqvist, F. Snickars and J. Weibull. North Holland.

Mogstad, Magne, Joseph P Romano, Azeem Shaikh, and Daniel Wilhelm. 2020. "Inference for Ranks with Applications to Mobility across Neighborhoods and Academic Achievement across Countries." National Bureau of Economic Research Working Paper 26883.

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.

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, and Matthew Turner. 2015. "Transportation Costs and the Spatial Organization of Economic Activity." In *Handbook of Regional and Urban Economics*. Vol. 5, , ed. Gilles Duranton, J. V. Henderson and William C. Strange, Chapter 20, 1339–1398. Elsevier.

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

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.

Severen, Christopher. 2019. "Commuting, Labor, and Housing Market Effects of Mass Transportation: Welfare and Identification."

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."

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

Tsivanidis, Nick. 2019. "Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogota's TransMilenio."

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

Zárate, Román David. 2019. "Factor Allocation, Informality and Transit Improvements: Evidence from Mexico City."

Appendix – For Online Publication

A Additional results for granular empirical settings

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 A.1 shows that, for tract pairs with fewer than ten commuters reported, a gravity-based estimate predicts the following year's value better than its current value does. The fitted values from a gravity model estimated using 2013 data have a higher R^2 for predicting observed 2014 values than the observed 2013 values in both Detroit and New York City. Using the observed 2013 values yields better predictions only for the tract pairs with the largest 2% of commuter counts.

Table A.1: Gravity-based estimates predict 2014 commuter counts 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 (5) as described in Section 4. The "gravity: distance" column uses $\ln \delta_{kn} = \ln \text{distance}_{kn}$ rather than commuting costs defined in Section **??**. The "2013 values" columns report the R^2 obtained by regressing the 2014 number of commuters on the 2013 number of commuters. The two "2013 values" columns differ because the regression sample is within observations where predictions from the gravity model available. Under the distance specification, observations where $k = n$ are dropped. For NYC, there are two observations where commute is infeasible.

B Theory

We first demonstrate the claim made in the main text that the continuum model is a limiting case of the granular model. The following subsections present extensions of the bare-bones model in Section 3 that introduce trade costs, the use of land in production, residential amenities, and local increasing returns. The model of perfect competition with local increasing returns should be isomorphic to a model of monopolistic competition with free entry, per the logic in Allen and Arkolakis (2014).

B.1 Continuum model as limiting case of the granular model

We derive the result that the granular model coincides with 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 (6) holds as $I \to \infty$. Conditional on the labor allocation $\{\ell_{kn}\}$, granularity plays no role in the trade equilibrium, which will coincide with that of a continuum model.

Definition 3.2 says that labor supplied by residents of k who work in n is

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

Note that $\mathbf{1}\left\{\tilde{U}_{kn}^{i} > \tilde{U}_{k'n'}^{i}, \forall (k',n') \neq (k,n)\right\}$ is a binary random variable that is equal to one with probability $\Pr(\tilde{U}_{kn}^{i} > \tilde{U}_{k'n'}^{i}, \forall (k',n') \neq (k,n))$. Thus, $\mathbb{E}\left(\mathbf{1}\left\{\tilde{U}_{kn}^{i} > \tilde{U}_{k'n'}^{i}, \forall (k',n') \neq (k,n)\right\}\right) = \Pr\left(\tilde{U}_{kn}^{i} > \tilde{U}_{k'n'}^{i}, \forall (k',n') \neq (k,n)\right)$. Since the idiosyncratic preference vectors are independent and identically distributed random variables, the law of large numbers implies that, using equation (5),

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

As a result, as $I \to \infty$, $\ell_{kn} \to L \times \frac{\tilde{w}_{n}^{\epsilon}(\tilde{r}_{k}^{\alpha}\delta_{kn})^{-\epsilon}}{\sum_{k',n'} \tilde{w}_{n'}^{\epsilon}(\tilde{r}_{k'}^{\alpha}\delta_{k'n'})^{-\epsilon}}$.

B.2 Trade costs

Relative to the model in Section 3, 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 $\Phi_k = r_k^{\alpha} P_k^{1-\alpha}$, where the local CES price index for goods is $P_k = \left[\sum_n (\tau_{nk} w_n / A_n)^{1-\sigma}\right]^{1/(1-\sigma)}$.

In this environment, the equivalent of equation (2) 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,

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

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:

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

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 3.1, with equations (B.1) and (4) serving as the relevant market-clearing conditions.

With trade costs, the expression for the probability of choosing residential-workplace pair kn analogous to equation (5) is

$$\Pr(U^i_{kn} > U^i_{k'n'} \ \forall (k',n') \neq (k,n)) = \frac{\tilde{w}^\epsilon_n \left(\tilde{r}^\alpha_k \tilde{P}^{1-\alpha}_k \delta_{kn}\right)^{-\epsilon}}{\sum_{k',n'} \tilde{w}^\epsilon_{n'} \left(\tilde{r}^\alpha_{k'} \tilde{P}^{1-\alpha}_{k'} \delta_{k'n'}\right)^{-\epsilon}}. \tag{B.2}$$

The definition of a granular commuting equilibrium is akin to Definition 3.2 with $\{\tau_{kn}\}$ added to the list of economic primitives and expression (B.2) serving as the relevant probability mass function.

B.3 Firms produce using land

Relative to the model in Section 3, we now assume that firms produce output using labor and land inputs via a Cobb-Douglas production function. Workers produce in location n with a Cobb-Douglas production technology:

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

where L_n is the labor supply of workers working in location n and ϕ_n is the share of land in location n that is used in production. As a result, the unit cost of production in location n is $c_n = \frac{1}{A_n} r^\beta_n w^{1-\beta}_n$. Thus, the CES price index is $P = \left[\sum_n c^{1-\sigma}_n\right]^{1/(1-\sigma)}$.

Since output in location n is $A_n(\phi_n T_n)^\beta \left(\sum_k \ell_{kn}\right)^{1-\beta}$, equating quantity supplied and quantity

demanded requires

$$A_n(\phi_n T_n)^{\beta}\left(\sum_k \ell_{kn}/\delta_{kn}\right)^{1-\beta} = \frac{(c_n)^{-\sigma}}{P^{1-\sigma}}\sum_{n'}\left(r_{n'}\phi_{n'}T_{n'} + w_{n'}\sum_{k'}\ell_{k'n'}/\delta_{k'n'}\right). \tag{B.3}$$

Equating the fixed land supply T_k to the sum of the quantities of land demanded by firms and by residents requires

$$T_k = \alpha\sum_n \frac{\ell_{kn}}{\delta_{kn}}\frac{w_n}{r_k} + \beta^{\frac{1}{1-\beta}}\left(\frac{w_k}{r_k}\right)\sum_{k'}\frac{\ell_{k'k}}{\delta_{k'k}}. \tag{B.4}$$

The definition of a trade equilibrium is akin to Definition 3.1, with equations (B.3) and (B.4) serving as the relevant market-clearing conditions. The definition of a granular commuting equilibrium is akin to Definition 3.2 with the Cobb-Douglas production parameter β added to the list of economic primitives.

B.4 Residential amenities

Relative to the model in Section 3, we now assume that each location is endowed with a residential amenity B_k. Individual i's utility from living in k and working in n is now

$$U^i_{kn} = \epsilon \ln\left(\frac{B_k w_n}{\Phi_k \delta_{kn}}\right) + \nu^i_{kn}.$$

With residential amenities, the expression for the probability of choosing residential-workplace pair kn analogous to equation (5) is

$$\Pr(U^i_{kn} > U^i_{k'n'}\ \forall (k',n') \neq (k,n)) = \frac{(B_k\tilde{w}_n)^{\epsilon}\left(\tilde{r}^{\alpha}_k\delta_{kn}\right)^{-\epsilon}}{\sum_{k',n'}(B_{k'}\tilde{w}_{n'})^{\epsilon}\left(\tilde{r}^{\alpha}_{k'}\delta_{k'n'}\right)^{-\epsilon}}. \tag{B.5}$$

The definition of a trade equilibrium in Definition 3.1 remains unchanged. The definition of a granular commuting equilibrium is akin to Definition 3.2 with the residential amenities $\{B_k\}$ added to the list of economic primitives and expression (B.5) serving as the relevant probability mass function.

B.5 Local increasing returns

Relative to the model in Section 3, 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 $\bar{A}_n\left(\sum_k \ell_{kn}/\delta_{kn}\right)^{1+\eta}$, equating quantity supplied and quantity

demanded requires

$$\bar{A}_n \left(\sum_k \ell_{kn}/\delta_{kn} \right)^{1+\eta} = \frac{{w_n}^{-\sigma} \bar{A}_n^{\sigma} \left(\sum_k \ell_{kn}/\delta_{kn} \right)^{\eta\sigma}}{P^{1-\sigma}} \sum_{k',n'} \frac{\ell_{k'n'}}{\delta_{k'n'}} w_{n'}. \tag{B.6}$$

The remaining results are unchanged after appropriately substituting in equation (B.6). Equation (4), which clears the land market, and expression (5) are unchanged by the introduction of local increasing returns. The definition of a trade equilibrium is akin to Definition 3.1, with equations (B.6) and (4) serving as the relevant market-clearing conditions. The definition of a granular commuting equilibrium is akin to Definition 3.2 with $\{\bar{A}_n\}$ and η added to the list of economic primitives.

C Estimation results

Much of the prior literature has used measures of distance rather than transit time. Table C.1 shows that the consequences of dropping observations with zero commuters reported in Section 4 are the same if one uses distance rather than transit time as the observable measure of bilateral commuting costs. Dropping the tract pairs with zero commuters produces estimates of the commuting elasticity that are substantially smaller.

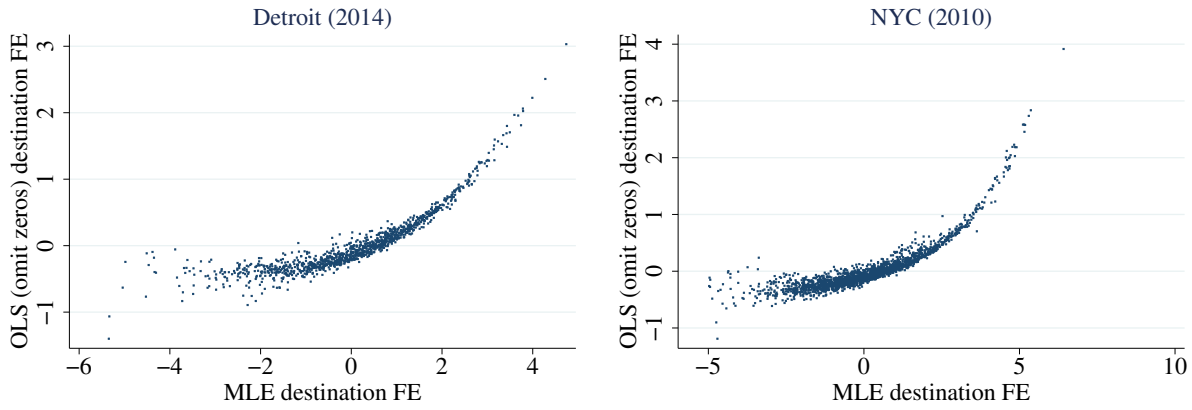
Table C.1: Commuting elasticity estimates using distance

	Detroit (2014)		NYC (2010)	
	MLE	OLS	MLE	OLS
Distance (log)	-1.098	-0.599	-0.998	-0.380
	(0.0135)	(0.0128)	(0.0205)	(0.00820)
Model fit (R^2 or pseudo-R^2)	0.609	0.539	0.673	0.588
Location pairs	1,352,565	357,301	4,626,742	688,626
Commuters	1,281,738	1,281,738	2,443,053	2,443,053

NOTES: All specifications include workplace fixed effects and residence fixed effects. The covariate δ_{kn} is now the geodesic distance between tracts k and n. The estimation sample is restricted to $k \neq n$ observations. Columns 1 and 3 presents the results from the maximum likelihood estimation described in Section 4. Columns 2 and 4 estimates the gravity equation in logs via OLS. The model-fit statistic is the pseudo-R^2 in columns 1 and 3 and R^2 in columns 2 and 4.

Figure C.1 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

Figure C.1: Destination fixed effects from tract-to-tract gravity regressions with distance covariate



NOTES: These plots depict the destination fixed effects estimated in Table C.1's column 1 (horizontal axis) and column 3 (vertical axis). The left panel plots these fixed effects for Detroit; the right panel plots them for New York City. See the notes to Table C.1 for details.

workplace locations with fewer workers.[40] Figure C.1 contrasts the OLS (vertical axis) and maximum likelihood (horizontal axis) estimates of the workplace fixed effects. The difference is stark: the range of the OLS estimates is half that of the maximum-likelihood estimates due to 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. This stark contrast is equally evident if one uses $\bar{\delta}_{kn}$ as the commuting-cost covariate instead of geodesic distance.

D Counterfactuals

D.1 Calibrated-shares procedure

This section describes a calibration procedure, known as "exact hat algebra" or "calibrated share form" in the international trade literature (Rutherford, 1995; Dekle, Eaton, and Kortum, 2008), often applied in quantitative spatial economics. This procedure makes a quantitative spatial model perfectly match the observed spatial distribution of economic outcomes and allows one to compute counterfactual equilibrium outcomes expressed as proportionate changes from the initial equilibrium.

The procedure relies on rewriting the equilibrium equations to describe ratios of counterfactual

[40] 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.

and initial equilibrium variables. To emulate the typical approach in the literature, we assume that commuting costs δ_{kn} are preference shifters rather than reduced labor hours. In this specification, the three equations that jointly define the equilibrium of the model in Section 3 when there is a continuum of individuals with continuum-case rational expectations, such that $\tilde{w}_n = w_n$ and $\tilde{r}_k = r_k$, are equation (6) and slight variants of equations (3) and (4):

$$\begin{aligned} A_n \sum_k \ell_{kn} &= \frac{(w_n/A_n)^{-\sigma}}{P^{1-\sigma}} \sum_{k',n'} \ell_{k'n'} w_{n'} \\ T_k &= \alpha \sum_n \ell_{kn} \frac{w_n}{r_k} \\ \ell_{kn} &= L \times \frac{w_n^{\epsilon} \left(r_k^{\alpha} \delta_{kn}\right)^{-\epsilon}}{\sum_{k',n'} w_{n'}^{\epsilon} \left(r_{k'}^{\alpha} \delta_{k'n'}\right)^{-\epsilon}}. \end{aligned}$$

The researcher knows the labor demand elasticity σ and commuting elasticity ϵ and observes the initial equilibrium labor allocation ℓ_{kn} and wages w_n. Denote counterfactual equilibrium values of wages, rents, and labor allocation by w'_n, r'_k, and ℓ'_{kn}. Denote the counterfactual-initial ratio of a variable x by $\hat{x} \equiv \frac{x'}{x}$. Assume that $\ell'_{kn} = 0$ if $\ell_{kn} = 0$, so that, in a slight abuse of notation, we can write $\hat{\ell}_{kn} = 1$ if $\ell_{kn} = 0$.

One can solve for the counterfactual equilibrium variables associated with combinations of counterfactual-initial ratios of productivities $\hat{A}_n$, land endowments $\hat{T}_k$, and commuting costs $\hat{\delta}_{kn}$. Tedious manipulation of the three equations above leads to the following three equations written in terms of elasticities, initial equilibrium values, and counterfactual-initial ratios:

$$\hat{w}_n = \hat{A}_n^{\frac{\sigma-1}{\sigma}} \left(\sum_n \left(\frac{\hat{w}_n}{\hat{A}_n} \right)^{1-\sigma} \frac{w_n \sum_k \ell_{kn}}{\sum_{k',n'} w_{n'} \ell_{k'n'}} \right)^{\frac{-1}{\sigma}} \left(\frac{\sum_k \hat{\ell}_{kn} \ell_{kn}}{\sum_k \ell_{kn}} \right)^{\frac{-1}{\sigma}} \left(\sum_{k',n'} \frac{\hat{\ell}_{k'n'} \ell_{k'n'} \hat{w}_{n'} w_{n'}}{\sum_{k',n'} \ell_{k'n'} w_{n'}} \right)^{\frac{1}{\sigma}} \tag{D.1}$$

$$\hat{r}_k = \hat{T}_k^{-1} \sum_n \frac{\hat{\ell}_{kn} \ell_{kn} \hat{w}_n w_n}{\sum_n \ell_{kn} w_n} \tag{D.2}$$

$$\hat{\ell}_{kn} = \begin{cases} 1, \text{ if } \ell_{kn} = 0 \\ \dfrac{\hat{w}_n^{\varepsilon} (\hat{r}_k^{\alpha} \hat{\delta}_{kn})^{-\varepsilon}}{\sum_{k',n'} \hat{w}_{n'}^{\varepsilon} \hat{r}_{k'}^{-\alpha\varepsilon} \hat{\delta}_{k'n'}^{-\varepsilon} \frac{\ell_{kn}}{L}}, \text{ if } \ell_{kn} > 0 \end{cases} \tag{D.3}$$

These three equations deliver $\hat{w}_n$, $\hat{r}_k$, and $\hat{\ell}_{kn}$ given the elasticities σ and ϵ, initial equilibrium values ℓ_{kn} and w_n, and counterfactual-initial ratios $\hat{A}_n$, $\hat{T}_k$, and $\hat{\delta}_{kn}$.

Note that this procedure implicitly rationalizes zero-commuters observations with infinity 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 $\delta'_{kn} \neq 0$ and $\delta_{kn} = \infty$.

Note that this procedure does not identify the initial values of the parameters A_n, T_k, and δ_{kn}. Given the elasticities σ and ϵ and the initial equilibrium values ℓ_{kn} and w_n, equations (3), (4), and (6) are insufficient to separately identify T_k and δ_{kn}. One would also need to observe land prices

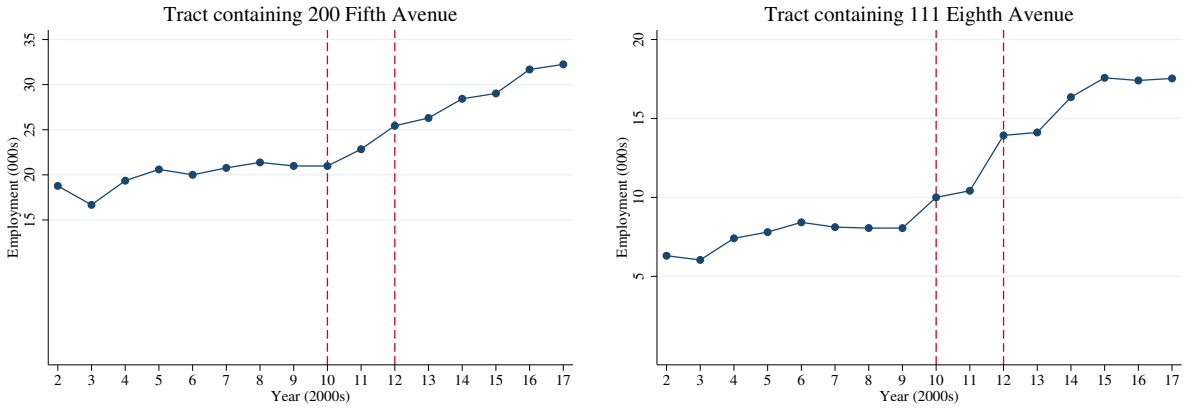
r_k to separate these parameters.

When implementing this procedure in Section 5, we compute tract-level workplace wages using LODES and ZIP Business Patterns data following Owens, Rossi-Hansberg, and Sarte (2020).

D.2 Event studies

Figure D.1 depicts the time series of total employment in two New York City census tracts that contain 200 Fifth Avenue and 111 Eighth Avenue. In Section 5, we examine tract-to-tract commuting flows within New York City for which these two tracts are workplaces (commuting destinations).

Figure D.1: Employment increases in the anchor-tenant event-study tracts



NOTES: This figure depicts the number of primary jobs in tracts 36061005800 and 36061008300 in the LODES data.

D.3 Monte Carlo exercise

A Monte Carlo exercise shows that granularity alone can substantially reduce the predictive power of the calibrated-shares procedure. In this simulation, the data-generating process is our estimated model of New York City in 2010. We impose a counterfactual 18% increase in productivity for the tract containing 200 Fifth Avenue. 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 granular ($I = 2,488,905$) realizations from this data-generating process are due to problems stemming from the "continuum of individuals" assumption.

Given our estimated model of New York City in 2010, we simulate 100 "event studies" as in Section 5. For each of the 100 events, we implement the following five steps:

1. To mimic "observed" data, draw one realization from the data-generating process at its 2010 parameter values and one realization at its counterfactual parameter values. Compute the difference in employment in the tract containing 200 Fifth Avenue.

2. Estimate the granular model and calibrate the continuum model to match the realization drawn from the data-generating process at its 2010 parameter values.

3. Compute the increase in productivity required to match the "observed" change in employment for the "treated" tract for both the calibrated-shares procedure and the granular model. These increases may differ from each other and the true 18% increase.

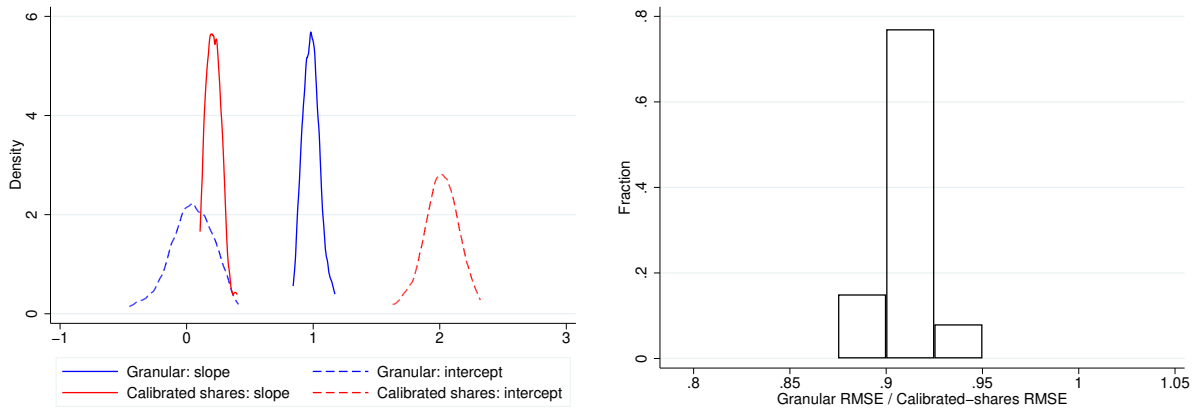
4. Compute the predicted change in commuter counts for the calibrated-shares procedure and 1,000 simulations of the granular model.

5. Regress the "observed" changes in commuter counts destined for the tract containing 200 Fifth Avenue on the calibrated-shares predicted changes. Regress the "observed" changes in commuter counts destined for the tract containing 200 Fifth Avenue on the mean of the predicted changes from 1,000 granular simulations.

We summarize the 100 simulations by plotting the coefficients and relative RMSEs in Figure D.2, which is analogous to Figure 3 in Section 5.

The Monte Carlo simulation shows that our estimation procedure recovers model parameters well enough when the granular model is the data-generating process to predict counterfactual changes in commuter counts. By contrast, the calibrated-shares procedure has limited predictive power when applied to granular data. Figure D.2 shows that regressing the "observed" changes in commuter counts on the values predicted by the calibrated-shares procedure yields a slope that is close to zero and an intercept that is far from it. 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, granularity alone can severely limit that procedure's predictive power.

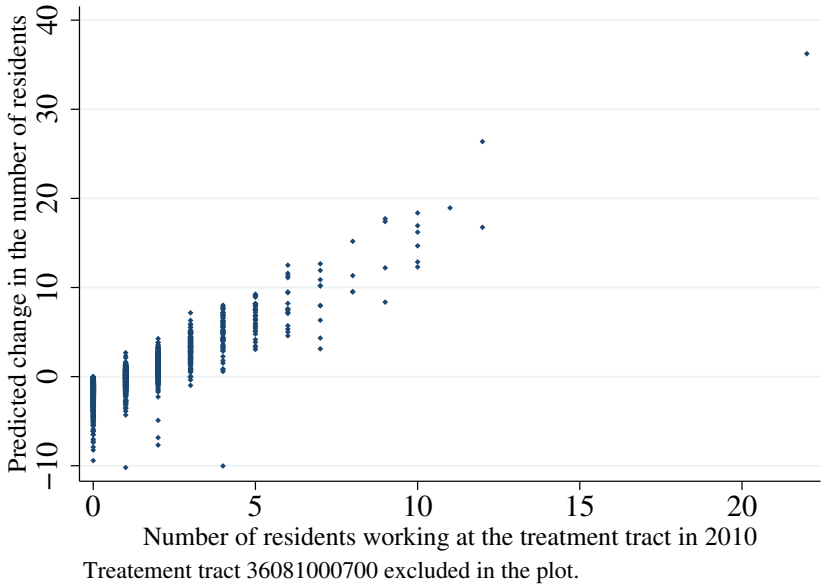
Figure D.2: Monte Carlo: Calibrated-shares procedure performs poorly with granular data



D.4 Event study: Amazon HQ2

Comparing Figures 5b and 6b reveals that the spatial pattern of changes in residents is closely related to the initial number of residents who work in the tract in which Amazon HQ2 would be located. Figure D.3 depicts the relationship between these two variables more explicitly in a scatterplot.

Figure D.3: Contrasting predicted change in residents and observed data



NOTES: This plot depicts the predicted change in the number of residents by calibrated-shares procedure against the number of residents working at the treatment tract in 2010.