Online Appendix: Not for Publication

Sample Construction

A1. Data Sources

Cross-sectional Data on Remote Work

Our first source of data to measure the prevalence of remote work in the U.S. economy and the earnings of remote workers are the decennial census between 1980 and 2000 and the American Community Survey thereafter. These surveys include a question asked of those ages 16 and over who were employed and at work in the previous week on the method of transportation usually used to get to work. We define as remote workers those who reported "work from home" on this question.[59] We measure the hourly earnings rate of workers by dividing the total pre-tax wage and salary income by the number of hours worked during the year.[60] This definition of remote work allows us to obtain a consistent measure of remote work and its compensation from 1980 to 2023 through repeated cross-sectional samples.[61] On average, every year, about 5 million workers are listed as remote workers. However, the decennial census and the ACS do not contain information on the number of hours worked at home.

Longitudinal Data on Remote Work

Our second source of data is the National Longitudinal Survey of Youth - 1979 Cohort (NLSY79), a nationally representative sample of 12,686 young men and women born between the years 1957 through 1964 in the United States. We focus on data reported from 1998 onward, a period through which the reporting of hours worked is consistent. Every two years, the survey reports the number of hours worked at home in a week. This allows us to observe variation in remote work on the intensive margin. We obtain workers' hourly rates directly from the survey. Thanks to the longitudinal design of the survey, we can observe workers' transitions into and out of remote work over time.

Foot Traffic Data

We use data from a private provider, SafeGraph, which produces foot traffic information at the business place level by collecting anonymized phone GPS tracking from third-party applications. Specifically, we use the Neighborhood Patterns

[59]In the decennial census, the question on means of transportation to work was asked in the "long form" questionnaire, which was administered to approximately 1 out of every 6 housing units in the United States.

[60]The income variable is "incwage". It includes wages, salaries, commissions, cash bonuses, tips, and other money income received from an employer. Payments-in-kind or reimbursements for business expenses are not included. The number of hours worked in the year is computed by multiplying the reported number of hours worked in a typical week by the number of weeks worked in the year.

[61]In 2023, no information on workers' industry of work is available, so only include this year for the descriptive evidence in Section II.

dataset, which contains information on foot traffic to census block groups every month. This dataset has several key advantages. First, it captures foot traffic at a fine geographical level.[62] This allows us to specifically focus on mobility to cities' CBDs. Second, it provides a count of individual visits by category of trip, including visits that are likely to be for work. We use this feature to define CBDs. Third, the visits to any block groups are disaggregated by block group of origin. This allows us to control for unobserved permanent characteristics of origin block groups by studying mobility to the CBD from each block group separately.

Housing Prices and Rents

We rely on house prices and rents aggregated at the zip code level by Zillow.[63] Our measure of house prices is the Zillow House Value Index (ZHVI), which controls for common house characteristics and adjusts for seasonality. We measure rents with the Zillow Observed Rental Index (ZORI), an index of house rents obtained by averaging asking rents after controlling for house quality indicators. Both series include prices and rents of single- and multi-family units. These datasets have the advantage of being reported monthly before and during the COVID-19 pandemic, allowing us to track the changes in housing prices and rent gradients in about 200 CBSAs. While the monthly Zillow data is well suited for within-city comparison across time, the relatively low number of observed zip codes in smaller CBSAs makes it is less amenable to comparisons across cities. To obtain more precise estimates of housing rent gradients for a larger number of CBSAs, we turn to data from the ACS at the census block group level.[64] We use the median rent paid in a block group during the 2015-2019 period. This, along with other block group-level controls defined below, allows us to estimate housing rent gradients before the COVID-19 pandemic more precisely and for almost 300 CBSAs.

Local Housing Market Characteristics

We collect zip code level variables to control for variation in housing market characteristics. From the Census Bureau, we obtain the median household income, the share of households in the top national income quintile, the median age, and the proportion of Black residents. There is significant variation in housing market characteristics at the CBSA level. We complement these measures with local geographic features collected by Lee and Lin (2018). We aggregate these census tract-level indicators at the zip code level.

[62]There are 220,684 census block groups in the United States.

[63]The datasets are publicly available from https://www.zillow.com/research/data/.

[64]The data can be obtained from IPUMS' National Historical Geographic Information System (NHGIS) portal Manson et al. (2022).

A2. Definitions of Data Samples

Census/ACS Sample

We combine the decennial census waves in 1980, 1990, and 2000 with the yearly waves of the American Community Survey between 2005 and 2023. We start from observations from all individuals recorded in each wave, weighted by their personal sampling weight. We only include workers who are employed, of age between 25 and 64, who do not work in agriculture or in the military, and who are not self-employed.[65] We restrict our sample to individuals who have not moved to a different metropolitan area since the previous year in order to allocate the labor delivery, and earnings of the individual to the appropriate labor market. To limit measurement error in earnings, we further restrict our sample to only include full-time workers, employed for the full year. We only include individuals born in the United States. Finally, we only include individuals who reside in one of the Core-based Statistical areas in the United States.

We define the main variables of interest as follows. An individual is categorized as a remote worker if they answer "work from home" to the question "What is your means of transportation to work?".[66] Our main measure of earnings is log hourly wage. We compute it by dividing yearly wage income by the estimated number of hours worked in the year. To obtain the total number of hours worked per year, we multiply the usual number of hours worked per week by the total number of weeks worked in the year.[67] For individuals located in small CBSAs (Micropolitan Statistical Areas), the CBSA is sometimes missing in IPUMS. We use a crosswalk from the PUMA (smallest geographical areas recorded in IPUMS) to CBSAs, by the intermediary of a crosswalk from PUMA to counties, and from counties to CBSAs.

Our final sample contains 15,872,072 observations, each corresponding to an individual in a given year. There are 637 distinct CBSAs covered by the sample. We report the number of observations, the share of remote workers, and the median hourly pay of commuters and remote workers per year in Table A1.

[65] Specifically, we exclude workers in 1-digit industries 1 and 14 from the 1990 Census industry classification ("Agriculture, Forestry, and Fisheries" and "Active Duty Military/NA/NC", respectively), workers in 1-digit occupations 4 and 7 from the 1990 Census occupation classification ("Agriculture, Forestry, and Fishing Occupations" and "Military Occupations/NA", respectively), and workers whose reported "class" or worker is self-employed (variable `classwkr` equal to 1).

[66] This is coded as `tranwork` = 80 in IPUMS.

[67] The yearly income variable is `incwage`. The number of weeks is coded into bins by the variables `wkswork*`, and the usual number of hours is recorded in `uhrswork`.

Table A1— Summary Statistics for the Census/ACS Sample

Year	Number of Observations	Share of Remote Workers (pct)	Median Hourly Pay of Commuting Workers (2021 $)	Median Hourly Pay of Remote Workers (2021 $)
1980	1,112,304	0.42	29.3	19.0
1990	1,788,782	0.64	26.8	22.5
2000	2,178,352	1.31	26.9	32.5
2005	517,591	1.51	26.7	35.9
2006	543,922	1.65	26.1	36.3
2007	549,974	1.80	26.4	36.2
2008	593,101	2.00	25.6	37.0
2009	569,593	2.16	25.7	35.8
2010	560,349	2.33	26.0	34.7
2011	540,518	2.45	25.6	35.7
2012	557,632	2.56	25.1	35.4
2013	575,307	2.72	24.9	35.0
2014	579,939	2.86	25.0	35.5
2015	592,265	3.02	24.9	36.5
2016	602,106	3.38	25.2	36.7
2017	618,611	3.61	25.8	37.6
2018	627,578	3.79	26.0	37.9
2019	637,308	4.29	26.4	37.0
2020	482,726	18.53	26.0	39.0
2021	601,567	21.98	25.7	41.1
2022	646,785	17.94	25.8	39.2
2023	395,762	16.74	27.2	39.3
All	15,872,072	5.70	25.8	38.5

Data from the Census/ACS. Hourly pay is adjusted for the Consumer Price Index and defined as the total annual earnings divided by the total number of hours worked in the year.

NLSY Sample

Our NLSY sample is constructed from the public use NLSY79 dataset.[68] We only include individuals from the main sample.[69] Besides this restriction, we include all individuals.

We define the main variables of interest as follows. When estimating the remote

[68]Note that there also exists the NLSY97, which follows individuals born between 1980 and 1984. Unfortunately, for the purposes of our analysis, the number of hours worked from home is not reported in this panel.

[69]That is, we exclude individuals from the supplemental and military samples that disproportionately survey individuals in the military and economically disadvantaged non-black, non-Hispanic individuals.

Table A2— Summary Statistics for the NLSY Sample

Year	Number of Observations	Average Number of Hours Worked at Home	Median Hourly Pay of Commuting Workers (2021 $)	Median Hourly Pay of Remote Workers (2021 $)
2002	30,541	1.34	21.2	28.2
2004	30,548	1.64	21.8	28.2
2006	30,545	1.61	21.5	31.0
2008	30,548	2.07	21.2	29.1
2010	30,549	2.14	21.2	29.4
2012	30,540	2.61	21.2	30.3
2014	30,548	2.96	21.2	31.2
2016	30,549	3.45	21.7	32.3
2018	30,546	3.52	21.6	31.9
2020	30,544	7.90	20.9	32.7
2022	30,542	7.27	20.4	32.0
All	336,000	3.08	21.3	30.8

Data from the NLSY79. Hourly pay is adjusted for the Consumer Price Index and is reported for the main job during the specified two-year period.

work premium, we use a continuous measure of remote work based on the number of hours worked at home per week. When a discrete definition is necessary, for example, when determining transitions into and from remote work, we use 24 hours of work at home per week as the threshold. Our main measure of earnings is hourly rates of pay, reported directly in the data. The geographical location is coded by the region of the United States. The four regions are Northeast, North-Central, South, and West.

Our final sample comprises, on average, 30,548 observations per year, representing observations from 5,656 individuals. We report the number of observations, average number of hours worked at home, and median hourly pay of commuters and remote workers per year in Table A2.

SafeGraph Sample

We measure visits to CBSAs' CBDs with the Neighborhood Patterns dataset, which contains monthly data between January 2019 and July 2022. Every month, the data is reported in a table that lists the number of phone devices that stopped for a visit inside a census block group by block group of origin. A visit is defined as the recording of the same device inside a polygon identified as a "place of interest" (i.e., a place of business) for more than a minute. SafeGraph also reports for each block group and each month the number of devices with a primary nighttime location inside the specified block group. We refer to this number as the number of residing devices in a block group.

Visits are categorized based on the behavior associated with the visit. Our main category of interest is the one associated with visits featuring a "work" behavior. "Work" visits are those recorded Monday through Friday between 7:30 am and 5:30 pm and that dwelled for at least six hours. Ideally, we would use "work" visits for all our analysis of visits to the CBD. However, this category only accounts for a small share of the total number of visits, leading the number of recorded "work" visits to be more volatile over time and less precisely measured.

For this reason, we use two different measures of visits in our analysis. First, we use the number of "work" visits from all block groups. We use the average number of monthly "work" visits in 2019 to define the location of the CBD. Second, we use the number of visits of any category by block group of origin. This is our preferred measure to study the variation over time of foot traffic patterns to the CBD.

Once the CBDs of each CBSA are defined according to the procedure defined in Section A.A3, we define mobility to the CBD at the origin block group of level. For each block group i, month t, in a CBSA j, we compute the number of visits $N_{ii't}$ originating from block group i that have as destination a block group that is inside a CBD of CBSA j, divided by the number N_{it}^{res} of sampled devices residing in block group i at t:

$$sh_{it}^{CBD} = \frac{\sum_{i' \in CBD_j} N_{ii't}}{N_{it}^{res}}, \tag{A1}$$

noting that we allow CBSAs to have multiple CBDs.

We exclude block groups with fewer than 300 residing devices in a month. Our sample contains data on 896 CBSAs, with 203,075 distinct census block groups. The median CBSA in the sample has 59 distinct block groups with available foot traffic data to the CBD.

ACS Housing Rents Sample

We assemble a dataset of housing rents and local market characteristics at the block group level from the 2015-2019 ACS. The main outcome variable in our sample of block group-level rents is the median rent paid in the block group.

Our controls include the percentage of dwellings by type of structure, number of bedrooms, and construction decade, all based on the 2015-2019 ACS data. Additionally, we include controls for block-group characteristics such as the percentage of Hispanic, Black, and Asian populations. We also include controls from Lee and Lin (2017) that are defined at the census tract level.[70] These include the average slope, the maximum temperature in July, the minimum temperature in January, annual precipitations, the distance to the closest port, the distance to

[70] We assign the same value of these geographic controls to all block groups inside the same tract.

Table A3— Summary Statistics for Occupation Groups

Code	Occupation	2019 Share of Total Employment	2019 Share That is Remote
43	Executive, Administrative, and Managerial	0.118	0.067
95	Management Related Occupations	0.050	0.088
124	Mathematical and Computer Scientists	0.041	0.107
156	Architecture and Engineering	0.025	0.041
196	Life, Physical, and Social Services	0.012	0.058
206	Community and Social Services	0.020	0.042
215	Legal	0.011	0.051
255	Education, Training, and Library	0.070	0.026
296	Art, Entertainment, and Media	0.016	0.083
354	Healthcare Practitioners	0.068	0.026
365	Healthcare Support	0.032	0.055
395	Protective Service	0.024	0.017
416	Food Preparation and Serving	0.046	0.012
425	Building Cleaning and Maintenance	0.026	0.017
465	Personal Care and Service	0.017	0.031
496	Sales and Related	0.085	0.060
593	Office and Administrative Support	0.123	0.041
613	Farming, Forestry, and Fishing	0.002	0.073
694	Construction and Extraction	0.049	0.015
762	Installation, Repair, and Maintenance	0.032	0.017
896	Production	0.066	0.013
983	Transportation and Material Moving	0.064	0.016
999	Unemployed not Classified	0.001	0.016

Occupation groups based on the occupation classification used in the 2000 decennial census, denoted `OCC` in IPUMS.

the closest river, the distance to the closest lake, the distance to the closest shore, and the surface area of the block group.

Finally, we compute the distance of each block group to the closest CBD of the CBSA, using the latitude and longitude of the centroid of the block group.

Occupations

We define 22 groups of occupations based on the occupation classification used in the 2000 decennial census, denoted `OCC` in IPUMS. From the 2000 decennial census until the latest ACS, the 2000 Census classification is available. For the decennial census waves in 1980 and 1990, the 1990 Census classification is available. We use a crosswalk between the 1990 and 2000 Census classifications to

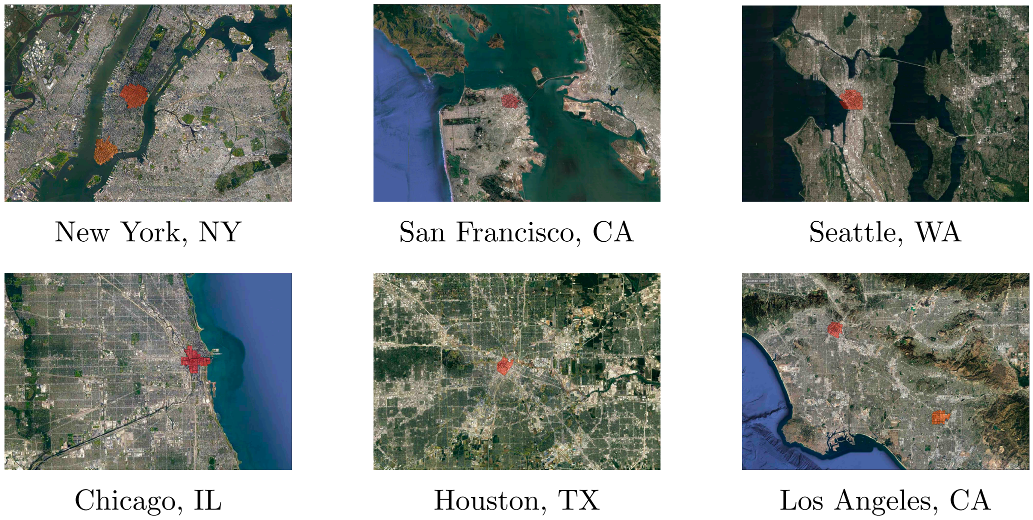


Figure A1. Central Business Districts of Selected Large Cities

obtain a consistent definition of our occupation groups.

The NLSY uses the 2000 Census occupation classification from 2002 onward. Prior to 2002, the NLSY uses the 1970 Census occupation classification. We use a crosswalk from the 1970 Census to the 1990 Census classifications, then from the 1990 Census to the 2000 Census classifications to obtain a consistent definition.

We report descriptive statistics about these occupation groups in Table A3.

A3. Defining Central Business Districts

For each CBSA j, chosen radius r, we define the geographic coordinates of the k^{th} CBD of j as

$$z^{(1)}_{cbd\ j}(r) = \arg\max_{z} \sum_{i \in cbsa_j} \mathbb{1}\{dist(z_i, z) \leq r\} \frac{N_{commute\ i2019}}{land\_area_i}, \tag{A2}$$

where i is a block group, z_i its coordinates (latitude and longitude), $dist(\cdot,\cdot)$ the geodesic distance, $N_{commute\ i2019}$ the total number of work-related visits recorded in block group i during the year 2019 from SafeGraph.

For $k \geq 2$, we impose that the center of the k^{th} CBD is not inside previous CBDs. We retain the k^{th} CBD if it attracts at least 80% as many commuters as the main CBD and if it does not overlap with CBDs 1 to $k-1$. We define the distance of each zip code to the CBD as the distance to the nearest CBD.

Supportive Evidence on City-Specific Patterns

B1. Change in Visits to the CBD for Selected Cities

In Figure B1, we report the time series of visits to the CBD of six selected cities. The estimates are the coefficients for monthly indicators in a regression at the block group-month level. The outcome variable is the number of visits to the CBD of the CBSA from the specified block group, divided by the number of residing devices in that block group, as defined in (A1), scaled by the value of this ratio in January 2020. That is, our outcome variable is $y_{it} = sh_{it}^{CBD}/sh_{it_0}^{CBD}$, where i is a block group, t is a month, and t_0 is set to January 2020. We then estimate the following regression for each CBSA:

$$y_{it} = \sum_{s=s_0}^{S} \beta_s \mathbb{1}\{t = s\} + \epsilon_{it},$$

where $s_0 = 1$ for October 2019, and $S = 34$ for July 2022. We restrict the sample to block groups that are located less than 50 kilometers away from the CBD, and cluster standard errors at the census tract level, which contains 3.4 block groups on average.

Figure B1 reports the estimates of β_s for the CBSAs of Los Angeles, CA, Chicago, IL, Boston, MA, San Diego, CA, Durham, NC, and Burlington, VT, along with the 95 percent confidence interval around each estimate. All six CBSAs experience a sudden drop in visits to the CBD in March-April 2020, corresponding to the period of nation-wide stay-at-home orders at the onset of the COVID-19 pandemic. The share of visits to the CBD drops to about 20 percent of their value just three months earlier. The patterns in the following months are similar for the three cities in the top panel. After some ebbs and flows likely explained by the various waves of COVID-19 infections, visits to the CBD in these CBSAs stabilize at, or below, 50 percent of January 2020 levels. In the three cities in the lower panel, however, the patterns of visits to the CBD is starkly different: after a small temporary upset in the winter of 2021, visits to the CBD continue to increase and return to pre-pandemic levels, or close to, by the end of the sample period.

B2. Office Trips for Selected Cities

In Section II.A, we measure trips to the CBD as the share of *all* trips originating from a census block group and ending in the CBD, averaged across all block groups. This measure, therefore, captures not only workers' trips to the office, but also non-work-related trips, such as visits to retail stores, restaurants, and other amenities.

In this section, we confirm that, for the CBSAs for which they are available, office trips to the CBD follow the same patterns as all other trips. We use an

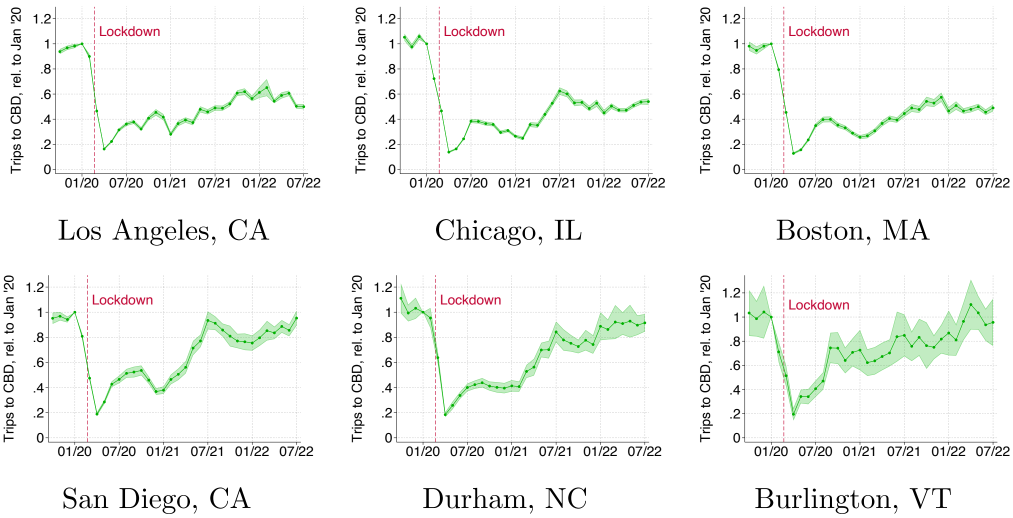


Figure B1. Change in Visits to the CBD for Selected Cities

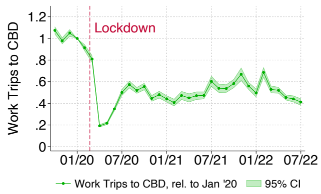
Note: Estimates of monthly indicators in a separate regression for each CBSA of visits to the CBD at the block group-month level. The outcome variable is the number of visits to the CBD from the specified block group, divided by the number of residing devices in that block group, as defined in (A1), scaled by the value of this ratio in January 2020. The shaded areas represent the contour delimited by the consecutive 95 percent confidence intervals of each monthly indicator. Standard errors are clustered at the census tract of origin.

alternative measure of trips to the CBD based only on the subset of trips that are labeled as *workplace* trips in the Neighborhood Patterns dataset. Workplace visits are defined by SafeGraph as those recorded Monday through Friday between 7:30 am and 5:30 pm and that dwelled for at least six hours. Workplace visits are recorded less consistently than all trips and are less often reported in smaller CBSAs. For the nine CBSAs analyzed in Figures 2 and B1, however, workplace visits are reported consistently. We report the monthly estimates of *office* trips to these CBSA's CBDs in Figure B2.

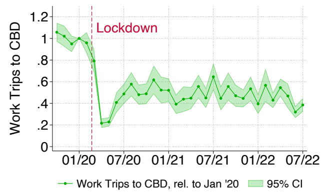
Figure B2 reveals the same patterns as those observed in Figures 2 and B1. For all cities, the first months of the pandemic saw a drastic drop in trips to around 20 percent of their initial value. The recovery, however, has been quite different. While Madison, San Diego, Durham, and Burlington essentially went back to their pre-pandemic levels by the end of our sample period, CBD office trips for New York, San Francisco, Los Angeles, Chicago, and Boston have stabilized at around 40 percent of their pre-pandemic levels.

B3. Evidence of Population Changes Across CBSAs

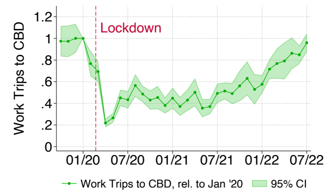
In this section, we motivate our choice of considering a closed-city model by documenting that, while there have been noticeable changes in CBSAs' popu-



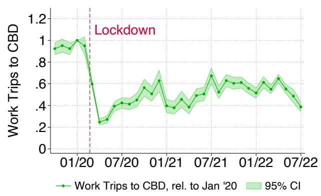
(a) New York, NY



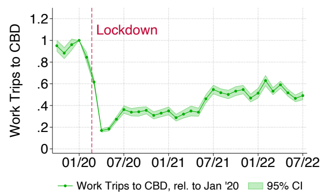
(b) San Francisco, CA



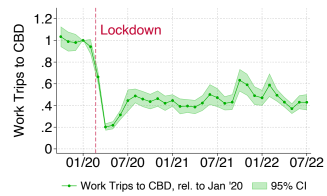
(c) Madison, WI



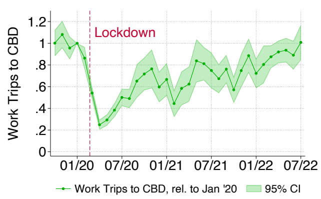
(d) Los Angeles, CA



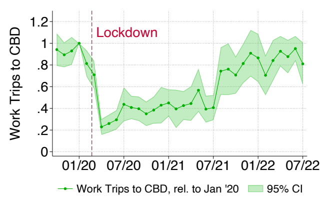
(e) Chicago, IL



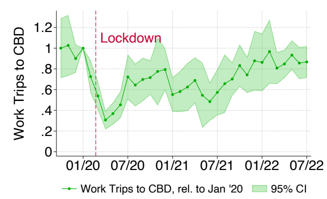
(f) Boston, MA



(g) San Diego, CA



(h) Durham, NC



(i) Burlington, VT

Figure B2. Office Trips to CBD relative to January 2020

Note: Estimates of monthly indicators in a separate regression for each CBSA of *workplace* visits to the CBD—those recorded Monday through Friday between 7:30 am and 5:30 pm that dwelled for at least six hours—at the block group-month level. The outcome variable is the number of visits to the CBD from the specified block group, divided by the number of residing devices in that block group, as defined in (A1), scaled by the value of this ratio in January 2020. The shaded areas represent the contour delimited by the consecutive 95 percent confidence intervals of each monthly indicator. Standard errors are clustered at the census tract of origin.

lations since 2019, those changes are small in comparison to the large shifts in remote work, trips to CBD, and house price gradients documented in Sections I and II.

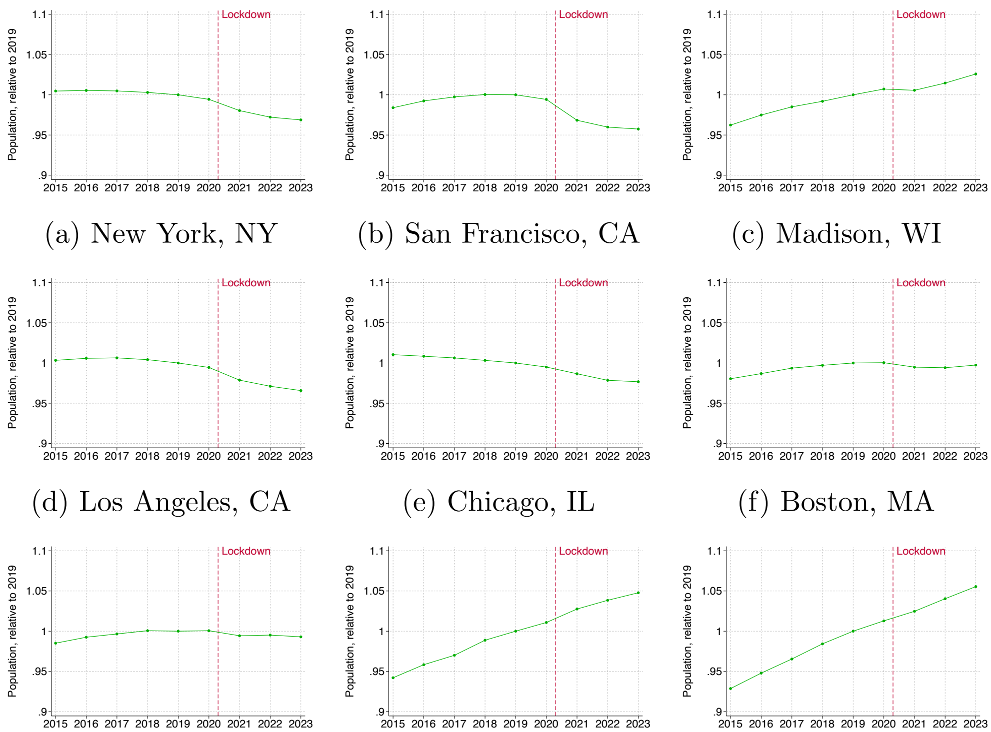
We look at estimates of CBSAs' population published by the US Census Bureau. We build a yearly time series from 2015 to 2023 by appending a file with estimates from 2015 to 2020, included, and one from 2020, included, to 2023. Some CBSAs' delineations changed in 2020, leading to two different population estimates for that same year, one for each delineation. To avoid jumps due to these delineation changes, we adjusted the estimates in 2020-2023 by the difference in the two estimates in 2020. In Figure B3, we start by documenting the changes in population for the nine cities used as examples throughout the paper. Several cities, particularly those that experienced large decreases in trips to CBD, have also seen their population decline between 2019 and 2023. New York City's population declined by 3.1 percent, while San Francisco's population declined by 4.3 percent, the strongest decline in that group. CBSAs that returned to pre-pandemic levels of trips to CBD, such as Madison, WI, Durham, NC, Burlington, VT, and San Diego, CA, appear to be following their pre-pandemic trends. Note that the decline in population in the largest cities is consistent with our findings that these cities offer lower levels of welfare in the low-commuting equilibrium.

In Figure B4a, we show that this deviation from trend of the CBSA population for large cities, but not for smaller cities, appears to be systematic. Large CBSAs with employment above 1.5 million grew on average by 3.7 percent between 2015 and 2019, but only by 1.9 percent between 2019 and 2023. Small CBSAs with employment below 150 thousand, however, grew almost at the same rate between 2015 and 2019 (0.4 percent) and between 2019 and 2023 (0.5 percent). Hence, on average, small CBSAs grew at the same rate in the four years preceding and following the COVID-19 pandemic, while large CBSAs grew by 1.8 percentage points less. In comparison, these large CBSAs saw a decrease in trips to CBD by at least 40 percent.

In Figure B4b, we find that, even though larger cities appear to have experienced larger deviation from their population trends since the pandemic, there is little, if any, correlation between the population change between 2019 and 2023 and the change in trips to CBD between January 2020 and July 2022. We view this as further evidence that the reallocation of workers across CBSAs is unlikely to be the essential driving force behind the dramatic changes in city structure we document in Sections I and II.

B4. Establishment Employment Concentration by City Size

Our theory considers a competitive production sector in which no single firm is large enough to affect the equilibrium selection through its own demand for in-person and remote workers. In practice, however, some CBSAs host large establishments that can represent a substantial share of that CBSA's employment. If smaller CBSAs are more likely to host such large establishments, they may be



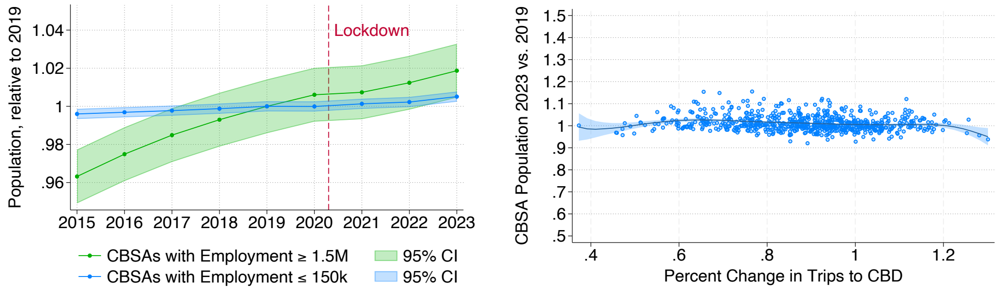
(a) New York, NY (b) San Francisco, CA (c) Madison, WI

(d) Los Angeles, CA (e) Chicago, IL (f) Boston, MA

(g) San Diego, CA (h) Durham, NC (i) Burlington, VT

Figure B3. Change in Total Population Relative to 2019

Note: Population estimates from "Annual Resident Population Estimates and Estimated Components of Resident Population Change for Metropolitan and Micropolitan Statistical Areas and Their Geographic Components for the United States," available at `https://www.census.gov/data/tables/time-series/demo/popest/2020s-total-metro-and-micro-statistical-areas.html`.



(a) Change in Population in Large vs. Small CBSAs

(b) Population Change vs. Change in CBD Trips

Figure B4. Change in Total Population Relative to 2019

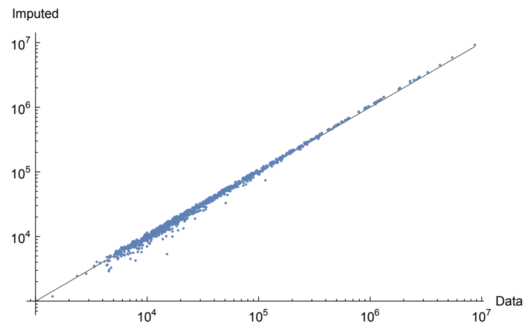
Note: Population estimates from "Annual Resident Population Estimates and Estimated Components of Resident Population Change for Metropolitan and Micropolitan Statistical Areas and Their Geographic Components for the United States," available at `https://www.census.gov/data/tables/time-series/demo/popest/2020s-total-metro-and-micro-statistical-areas.html`. Panel (a) represents the average CBSA population relative to 2019 for i) CBSAs with total employment above 1.5 million (25 largest CBSAs), ii) CBSAs with total employment below 150,000 (663 smallest CBSAs). The shaded areas represent the consecutive 95 percent confidence intervals of each yearly estimate. Panel (b) reports the change in CBSA population between 2019 and 2023 against the change in visits to the CBD in July 2022 relative to January 2020. Each marker represents a CBSA. The solid line represents a fitted kernel-weighted local polynomial smoothing, with the shaded area corresponding to the 95 percent confidence interval around the fitted value.

more likely to return to high levels of commuting if those establishments participate in bringing workers back to the office.

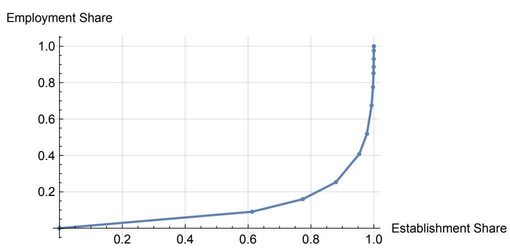
In this section, we show that, perhaps surprisingly, smaller MSAs (the subset of CBSAs for which we can perform the exercise) are *less* likely to host establishments that account for a large share of their employment. We interpret this finding as indicative that the differential rates of return to the office between large and small CBSAs are unlikely to be facilitated by the idiosyncratic decisions of the largest establishments.

We examine the empirical relationship between the MSA-level concentration of employment in large establishments and the total employment size of the MSA. To this aim, we estimate, for each MSA j, the share of employment accounted for by the fraction p of largest establishments, $\chi_j(p) \in (0,1)$, and relate it to the log employment in the MSA j, $\log(emp_j)$.

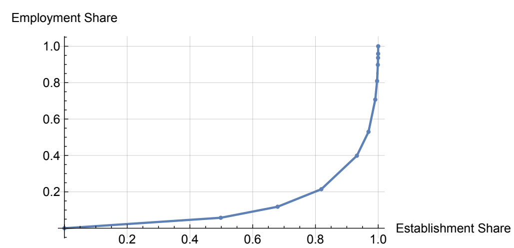
We start by constructing a Lorenz curve for the total MSA j's employment in 2019. The curve maps the share $e_j(s) \in (0,1)$ of MSA j's establishments smaller than s into the share of employment $\zeta_j(s) \in (0,1)$ accounted for by those establishments. The 2019 County Business Patterns' MSA file (CBP) contains, for each MSA j, the total employment in the MSA E_j, the total number of establishments N_j, and the number of establishments $n_j(r)$ in each of 11 different employment brackets $[t_l(r), t_u(r)]$, for $r = 1, ..., 11$, plus a residual category $t_l(12)$



(a) Fitted vs. Actual CBSA Employment



(b) Lorenz Curve, New York, NY



(c) Lorenz Curve, Madison, WI

Figure B5. Estimating employment by bracket size from CBSA County Business Patterns

Note: Panel (a) of this figure shows the relationship between actual CBSA employment on the x axis, and estimated employment after recovering the optimal weights ω, b, m from equation B4, on the y axis; the black line is the 45 degree line. Panels (b) and (c) show the implied Lorenz curves for the metro areas of New York, NY and Madison, WI.

with employment above 5,000. With these data, we seek to 1) estimate employment in each bracket with an appropriate weight between the lower and upper end, 2) estimate employment in the top bracket where we only know the lower end, 3) to minimize the sum of squared deviations between estimated and actual employment across CBSAs. We address each of these points as follow:

1) We approximate employment in brackets size $[t_l(r), t_u(r)]$, for $r \leq 11$, with

$$emp(r;\omega,b) = t_u(r)\left[\omega\cdot\varphi(t_u(r);b)\right] + t_l(r)\left[1-\omega\cdot\varphi(t_u(r);b)\right], \tag{B1}$$

where $\omega \in [0,1]$ is a weight common across all ranges, and $\varphi(t_u(r);b) \in [0,1]$ is a "bending" function that shifts the common weight differentially across ranges as a function of a parameter $b \in [0,1]$. The values of ω and b will be chosen optimally, as described below. Note that if $\omega = 0.5, \varphi(t_u(r);b) = 1$ for all r, the approximation is simply the average of the brackets' endpoints; since the establishment size distribution is not uniform, one might conjecture that the appropriate midpoint varies across the intervals. In particular, if the establishment size is Pareto, the appropriate midpoint is closer to the left end for lower employment ranges than it is for higher employment ranges. We then seek a function $\varphi(t_u(r);b)$ that assumes values between 0 and 1, is increasing in the endpoint $t_u(r)$, and admits "no bending" as a possibility, that is, $\varphi(t_u(r);b) = 1$ for all $t_u(r)$. One such function is

$$\varphi(x;b) = b + (1-b)\left(\frac{x-4}{4995}\right)^{1/2}, \tag{B2}$$

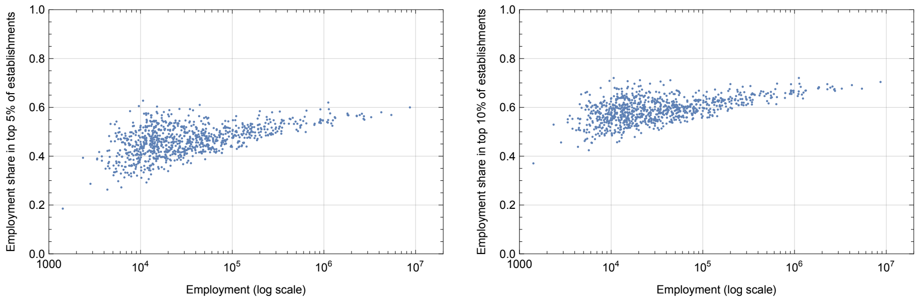
defined between the values of $x = 4$ (the upper end of the lowest employment bracket, 1 to 4) and $x = 4999$ (the upper end of the highest employment bracket, from 2,500 to 4,999 employees), and $b \in [0,1]$. This function approaches a flat line at 1, and the heterogeneity among midpoints becomes less extreme, as $b \to 1$.

2) With regard to the second constraint, we parameterize employment in the residual interval as $emp(12;m) = 5{,}000m$, with $m \geq 1$. The total imputed employment in a CBSA can then be written as:

$$\tilde{e}_j(\omega,b,m) = \sum_{r=1}^{11} emp_j(r;\omega,b)n_j(r) + emp_j(12;m)n_j(12). \tag{B3}$$

3) We then minimize

$$\rho(\omega,b,m) = \sum_j \left(\frac{\tilde{e}_j(\omega,b,m)}{E_j} - 1\right)^2, \tag{B4}$$



(a) Top 5%, all MSAs

(b) Top 10%, all MSAs

Figure B6. Employment Concentration in Largest Establishments by MSA Size

Note: Estimated share of employment accounted for by the top 5 and 10 percent establishments in size. Each dot represents a Metropolitan Statistical Area. The MSA employment is obtained from the 2019 County Business Patterns. The share of employment in the top establishments is obtained by constructing the Lorenz curve of employment share by establishment size for each MSA.

under the constraints that $\omega \in [0,1], b \in [0,1], m \geq 1$.

The best fit for our procedure is $b = 1, \omega = 0.4475, m = 1$, which rejects heterogeneous weights and suggests a common weight higher than 0.5 on the highest end of each bracket. Panel (a) of figure B5 reports a comparison between data and fitted CBSA employment, across CBSAs. This procedure delivers data on the number of establishments and an estimate of the total employment in each of the brackets, $emp_j(r; 0.447, 1)n_j(r)$, for each bracket and CBSA. We build the Lorenz curve for each MSA by computing the share of establishments up to a given size bin s, $e_j(s) = \sum_{r=1}^{s} n_j(r)/N_j$, and the cumulative employment share accounted for by these establishments, $\zeta_j(s) = \sum_{r=1}^{s} emp_j(r)/E_j$. For each MSA j, we then linearly interpolate the set of points $(e_j(s), \zeta_j(s))$ to obtain the function $\tilde{\zeta}_j(s)$, giving, for each $s \in (0,1)$, the share of employment accounted for by the $100e$ percent smallest establishments in that MSA. Finally, for a chosen share of top establishments p, we compute the share of employment accounted for by the $100p$ percent largest establishments in MSA j as $\chi_j(p) = 1 - \tilde{\zeta}_j(1-p)$. Two examples of the function $\chi_j(p)$ for New York and Madison, WI, are reported in panels (b) and (c) of Figure B5.

Figure B6 shows the relationship between city size and the share of employment accounted for by the largest 5% and 10% of establishments, respectively. We find that smaller cities tend to have employment less concentrated in a few large establishments. We interpret this finding as indicative that the return to the office in smaller cities is less likely to be triggered by the decision of a few large employers mandating a return to the office.

B5. Changes in CBSAs' Teleworkable Employment

Our theory considers workers in a representative occupation with in-person and remote productivities that are calibrated as the employment-weighted average of our occupation-specific productivity estimates. This assumption abstracts from the possibility for workers to change their occupations over time. However, shifts in cities' occupational composition could partly explain their diverging trajectories if, in response to the pandemic lockdowns, workers in large cities decided to transition in large numbers to occupations more amenable to remote work, while workers in smaller cities did not.

In this section, we show that this was not the case. First, while there is a noticeable transition toward occupations with higher teleworkable index, it is of much lower magnitude than the shift in trips to CBD. Second, that shift is slightly more pronounced in *smaller* CBSAs, indicating that the dramatic switch to remote work in large cities is not explained by a disproportional shift toward teleworkable occupations.

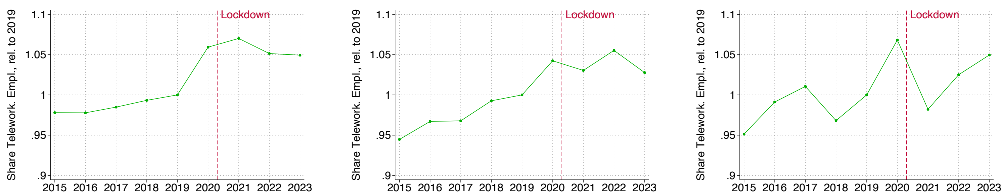
In Figure B7, we report the evolution of the share of teleworkable employment, computed as the employment-weighted average of the time-invariant measures of occupational teleworkable index by Dingel and Neiman (2020). In New York City, we observed a marked, but limited, increase in teleworkable employment. The share increased by 2.3 percent between 2015 and 2019 and by 5.0 percent between 2019 and 2023. In San Francisco, the share of teleworkable employment has increased on average at a rate comparable to pre-pandemic. Our measure of teleworkable employment is more volatile in Madison; on average, it also increased at a similar rate to pre-pandemic.

We investigate these patterns systematically across large and small CBSAs in Figure B8. On average, the share of teleworkable employment increased by 5 percent between 2019 and 2020 in small and large CBSAs. By 2023, the increase was 7.9 percent in small CBSAs, and 5.8 percent in large CBSAs. We view this finding as indicating that changes in CBSAs' occupational composition are unlikely to be a key driver of the heterogeneous responses described in Section II.

B6. Descriptive Evidence Using Census Employment

Individual Mobility

In Section II, we characterize CBSAs by their total non-farm employment in 2019, published by the BEA. While this variable provides a measure of city size for all CBSAs, it is distinct from the measure of city size that we use in the quantification of the model in several ways. First, it is based on counting the number of jobs in the CBSAs, rather than on the number of workers, as is the case in our Census data. Since some workers hold more than one job, the BEA measure is typically larger than our Census measure. Second, because the Census data is based on individual records, we can count the number of commuters and



(a) New York, NY (b) San Francisco, CA (c) Madison, WI

Figure B7. Change in Share of Teleworkable Employment for Selected Cities

Note: Estimates are constructed by computing CBSAs' yearly occupational composition from the ACS, and multiplying each occupational share by the teleworkable index from Dingel and Neiman (2020).

calibrate our model to workers who are more likely to face the trade-off between remote work and commuting that our theory speaks to. For example, the Census allows us to focus on workers between 25 and 64 years of age and to exclude self-employed workers. As a result, our Census-based measure of CBSA employment is 56 percent lower than the BEA measure on average. The correlation between the log of the two employment measures is 0.89.

Using the Census individual data to measure commuting decisions, we can compute total employment for 598 out of the 847 CBSAs represented in Figure 3a. In Figures B9a and B9b, we reproduce the analogous figures to Figures 3a and 3 with this smaller sample of CBSAs. In Figure B9a, we use lower thresholds of employment (1 million for "large" CBSAs, 100,000 for "small" CBSAs) to reflect the fact that our measure of employment from the Census is lower than the BEA measure on average. These thresholds of Census employment reproduce the partition of CBSAs in Figure 3a, with the first group containing the top 3 percent of CBSAs in terms of employment, the second containing the bottom 75 percent of CBSAs.

The patterns of trips to CBD over time are very similar to those described in Section II.A. Namely, CBSAs of all sizes experience a large decline of trips to CBD of 70 to 80 percent during the stay-at-home policies in 2020, with larger CBSAs only recovering to 60 percent of their pre-pandemic values, while smaller CBSAs return at least to 90 percent on average. Figure B9b confirms that the relationship between the size of the drop in trips to CBD and CBSA employment became stronger by July 2022 than in April 2020.

Distance Gradients in House Prices

Figure B10a shows that the divergence in the average distance elasticity of house prices relative to January 2020 between smaller and larger cities is visible also when we measure size with Census employment.

Panel (a) of Figure B10b confirms that the flattening of the housing price gradients occurs in similar magnitude for cities of all sizes in January 2021. Panel

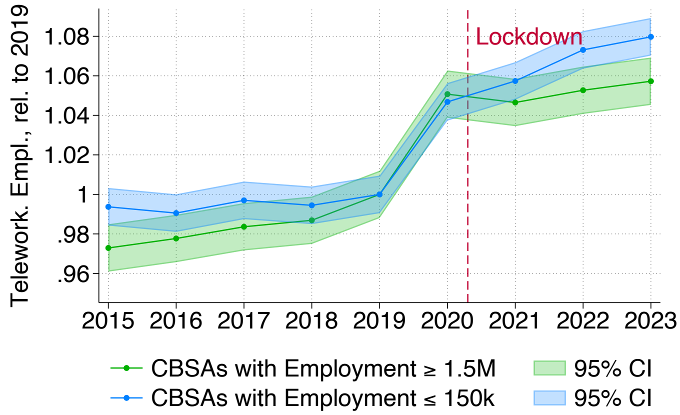


Figure B8. Change in Share of Teleworkable Employment in Large vs. Small CBSAs

Note: Plotted coefficients represent the average share of teleworkable employment in a CBSA relative to 2019 for i) CBSAs with total employment above 1.5 million (25 largest CBSAs), ii) CBSAs with total employment below 150,000 (663 smallest CBSAs). The shaded areas represent the consecutive 95 percent confidence intervals of each yearly estimate.

(b) confirms that even in the sample of cities ordered by Census employment, the drop in the gradient is more strongly related to city size by December 2022.

Quantification Exercises

C1. Estimation of the Transition Elasticity and Transition Costs

Predicting Transition Shares

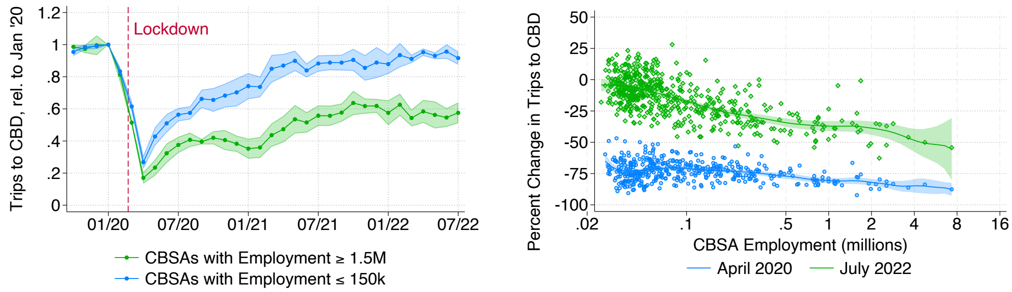
We estimate a probit model to control for individual characteristics in transition probabilities $\ell\ell'$:

$$Y_{i\ell\ell' t} = \Phi\left(\delta_t^{r(i)} + X_{it}\beta\right),$$

where $Y_{i\ell\ell' t}$ is an indicator of individual i transitioning from ℓ to ℓ', $r(i)$ is individual i's education-region group (among the eight possible groups), $X_{i,t}$ contains individual i's age, age squared, tenure at the job, tenure squared, occupation, industry. We define $\hat{\lambda}^r_{\ell\ell' t} = \Phi\left(\hat{\delta}^r_t + \bar{X}_{it}\hat{\beta}\right)$ the predicted transition shares evaluated at mean values of X. The dependent variable in the estimation regression is then defined as

$$y^r_{\ell\ell' t} \equiv \ln \frac{\hat{\lambda}^r_{\ell\ell' t}(\hat{\lambda}^r_{\ell'\ell' t+1})^{\beta}}{\hat{\lambda}^r_{\ell\ell,t}(\hat{\lambda}^r_{\ell\ell' t+1})^{\beta}}.$$

We estimate the log hourly rates for each education group-region-year after



(a) April 2020 Relative to January 2020 (b) July 2022 Relative to January 2020

Figure B9. Response of Trips to the CBD, Census Employment

Note: Panel (a) reports the estimates of the average volume of visits to the CBD from block groups located in i) CBSAs with total Census employment above 1 million (19 largest CBSAs), ii) CBSAs with total employment below 100,000 (450 smallest CBSAs) expressed as a share of their values in January 2020. Panel (b) reports the change in the average level of visits to each CBSA's CBD between January 2020 and April 2020, and January 2020 and July 2022, against our measure of employment from the Census. Each marker represents a CBSA. The solid line represents a fitted kernel-weighted local polynomial smoothing, with the shaded area corresponding to the 95 percent confidence interval around the fitted value.

controlling for individual characteristics

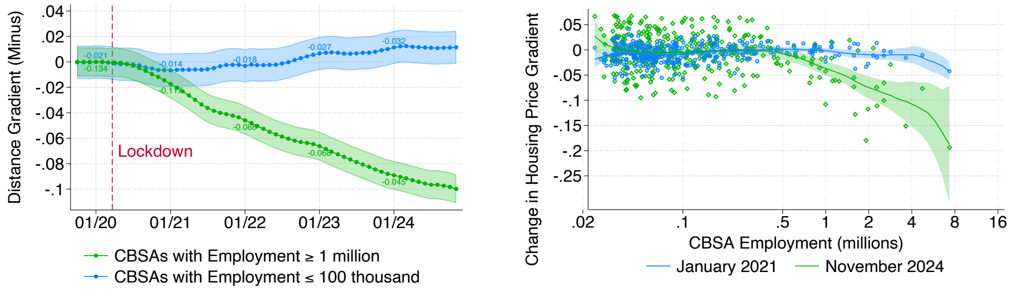
$$\ln w_{it} = X_{it}\beta + \zeta_i + \omega_{it},$$

where X_{it} contains individual i's age, age squared, tenure at job, tenure squared, occupation, industry, and region. We define $\ln w^r_{\ell' t+1}$ as the average value of $\hat{\omega}_{it+1}$ for individuals in education-region r. With 4 regions, 2 education groups, 9 periods (every two years from 2002 to 2018), 2 transition types, we obtain a maximum of $4 \times 2 \times 9 \times 2 = 144$ observations.

Results for the Baseline Specification

Table C1 reports the results of the estimation of (21) and the implied structural parameters.[71] In the first two columns, we report the OLS and IV estimates from the specification in (21), respectively. Consistent with the intuition that workers only respond to the part of the wage ratio that they observe, better captured by using the past wage ratio as an instrument, the IV estimates of the coefficient on the log hourly wage difference are higher than the OLS estimates. They provide estimates of the elasticity of transitions to wages s, and of the fixed transition cost that are statistically significant at the 1 percent level. Note that, in practice, the

[71] The total number of possible observations is the product of the number of groups r (8) with the number of distinct years (9 for every two years between 2000 and 2016), times two for each direction of transition (to remote, or to commute), i.e. 144 observations. In practice, however, there are only 48 observations with non-missing values of $y^r_{\ell\ell' t}$.



(a) Jan. 2021 Relative to Jan. 2020 (b) Dec. 2022 Relative to Jan. 2020

Figure B10. Response of House Price Gradients, Census Employment

Note: Panel (a) reports the difference between the average housing price gradients in January 2020 and each following month, using the same Census employment thresholds as in Figure B9a. A negative value for the difference in housing price gradients implies a lower premium for housing located close to the CBD. The value of the average housing price gradient in each group is reported next to the marker of the associated group every six months. The shaded areas represent the contour delimited by the consecutive 95 percent confidence intervals of each monthly indicator. Panel (b) reports the difference between the housing price gradients in January 2020 and January 2021, and January 2020 and November 2024, against our measure of employment from the Census. Each marker represents a CBSA. The solid line represents a fitted kernel-weighted local polynomial smoothing, with the shaded area corresponding to the 95 percent confidence interval around the fitted value.

number of commuters, L^r_{ct+1}, and total employment, L^r_{t+1}, are highly correlated, and so the coefficients on these two covariates are noisy.

In addition to the parameters s, F, and $\xi - \theta$, Equation (21) in principle also allows to recover the following parameters:

$$\begin{aligned} \alpha_0 - \ln \frac{\bar{B}}{\bar{\tau}} &= \frac{1}{\eta_0}\left(\frac{\eta_2}{\mu} + \left(\frac{\eta_4}{\mu} - \frac{\eta_5}{1-\mu}\right)\ln \pi\right), \\ \alpha_1 &= \frac{\eta_3}{\eta_0 \mu}, \\ \gamma &= \frac{2}{\eta_0}\left(\frac{\eta_4}{\mu} - \frac{\eta_5}{1-\mu}\right). \end{aligned} \tag{C1}$$

However, since η_2, η_3, η_4, and η_5 are imprecisely estimated, the implied values for these parameters, as well as for $\xi - \theta$, are uninformative. For example, the implied value of $\xi - \theta$ is -1.70, with a standard error of 19.9. This is why we set these parameters to zero for the rest of our analysis, except for $\bar{\tau}$, as described in Section IV.E.

Alternative Definition of Transitions by City Size

In this section, we conduct an alternative estimation of (21), with transition shares defined by subsets of CBSAs with similar population rather than by groups

Table C1— Estimates of the Gravity Equation for Mode Transitions and Structural Parameters

Covariate	OLS $y^r_{\ell\ell',t}$	IV $y^r_{\ell\ell',t}$	Parameter	OLS Estimates	IV Estimates
$\ln \frac{w^r_{\ell',t+1}}{w^r_{\ell,t+1}}$	1.954	3.088	s	0.472	0.298
	(0.411)	(0.998)		(0.099)	(0.096)
$d_{\ell\ell'}$	4.579	3.926	F	2.818	1.785
	(6.141)	(6.035)		(0.843)	(0.696)
$d_{\ell\ell'}(t+1)$	-0.078	-0.082	$\xi - \theta$	-5.672	-1.693
	(0.055)	(0.067)		(30.522)	(19.801)
$d_{\ell\ell'} L^r_{c,t+1}$	-2.902	-1.449			
	(14.519)	(14.796)			
$d_{\ell\ell'} L^r_{t+1}$	2.660	1.255			
	(14.301)	(14.571)			
cons	-0.469	-0.469			
	(0.103)	(0.097)			
N	48	48			
R^2	0.26	0.19			

Robust standard errors in parentheses. Structural parameters estimated with the Delta method.

of region-educational attainment.

While the region-education groups used in our main specification allow for differences in transition patterns across broad geographical areas, they pool observations from many different cities in the same region. These cities may offer different conditions to remote workers. Pooling observations from CBSAs with similar population recognizes that larger cities may offer different remote work conditions than smaller cities—at the expense of pooling observations from different regions.

In order to partition cities into subsets with similar population, we first obtain information on the CBSA and county of each individual from the restricted-use NLSY Geocode dataset.[72] When the CBSA information is missing, we infer it from the county FIPS code, if available. For each CBSA j, $j = 1, \ldots, J$, we compute the average population $\overline{pop}_j$ over the 2000-2018 period. Given a number of subsets K, we order CBSAs by increasing average population, such that the CBSAs indexed by j and $j+1$ satisfy $\overline{pop}_j \leq \overline{pop}_{j+1}$ for any $j = 1, \ldots, J-1$. We then define the threshold indices $(\tau_k)_{k=0,\ldots,K}$, such that $\tau_0 = 0$, $\tau_K = J$, and for $k = 1, \ldots, K-1$, τ_k is defined recursively as the smallest index such that all

[72]A description of the NLSY Geocode data is available at https://www.bls.gov/nls/request-restricted-data/nlsy-geocode-data.htm.

CBSAs indexed between τ_{k-1} and τ_k account for at least $1/K$ of the total average population across all CBSAs,

$$\tau_k = \min_{\tau=\tau_{k-1},\ldots,J} \sum_{j=\tau_{k-1}}^{\tau} \overline{pop}_j \geq \frac{1}{K}\sum_{j}^{J} \overline{pop}_j, \qquad k = 1,\ldots,K-1.$$

For each $k = 1,\ldots,K$, we refer to the subset $\{j|\tau_{k-1} < j \leq \tau_k\}$ of CBSAs as subset k. We then define transition shares by counting the number of transitions into and out of remote work of individuals located in CBSAs that are part of the same subset k. By construction, subset $k = 1$ contains many small CBSAs, while subset K contains few large ones.

In Table C2, we describe the results from estimating (21) with $K = 5$ subsets of cities, defining remote work as spending either $d = 1$ or $d = 2$ days per week at home. We compare the OLS estimator with the Poisson Pseudo Maximum Likelihood (PPML) estimator. Given the limited number of remote workers observed per subset k and year t, the transition shares $\lambda^k_{\ell\ell' t}$ are sometimes computed based on a small number of observations. We define a minimum threshold $\bar{n}_{remote}$ of remote workers, such that observations $y^k_{\ell\ell' t}$ are included in the estimation only if more than $\bar{n}_{remote}$ remote workers are observed in subset k, year t. In the case where we define remote work as $d = 1$ day per week at home, we set $\bar{n}_{remote} = 20$, leaving 62 observations; with $d = 2$, we set $\bar{n}_{remote} = 10$, leaving 66 observations.

The estimates of s and F from the OLS and PPML estimator, with $d = 1$, are reported in Columns (1) and (2). The estimates of s are similar, and not significantly different from our main estimate of 0.3. The estimates of F are also close to our main estimate of 1.78 The estimates obtained from defining remote work as $d = 2$ days or more per week are reported in Columns (3) and (4). The OLS and PPML yield slightly higher estimates of both s and F, although our main values of s and F are contained in the 95% confidence intervals of these estimates.

C2. Estimation of the Agglomeration Externalities

Derivation of the Estimation Equation

This appendix derives the estimating equation for the agglomeration elasticity from an explicit production function, rather than postulating a wage equation. Doing so clarifies the origin of each term in equation (25) and highlights how the presence of a city-wide production factor modifies the interpretation of coefficients.

We extend the baseline model to include a city-wide factor of production K_{jt} (e.g., infrastructure, transit quality, or other shared inputs) that augments the productivity of all workers equally, regardless of their labor delivery mode. This factor scales with city size and introduces an additional channel through which total population affects wages.

Table C2— Estimates of the Gravity Equation Based on City Size

	(1)	(2)	(3)	(4)
	Remote = 1 day		Remote = 2 days	
	OLS	PPML	OLS	PPML
$\ln \frac{w^r_{\ell' t+1}}{w^r_{\ell t+1}}$	1.439	1.384	1.071	0.886
	(0.380)	(0.389)	(0.482)	(0.460)
$d_{\ell\ell'}$	1.100	1.105	0.513	0.080
	(0.686)	(0.661)	(0.933)	(0.764)
$d_{\ell\ell'}(t+1)$	-0.051	-0.060	0.133	0.027
	(0.046)	(0.046)	(0.077)	(0.046)
$d_{\ell\ell'} L^r_{c,t+1}$	-0.152	-5.603	8.880	11.60
	(14.81)	(14.71)	(17.19)	(14.73)
$d_{\ell\ell'} L^r_{t+1}$	0.103	5.546	-8.899	-11.60
	(14.78)	(14.69)	(17.16)	(14.71)
cons	-0.339	-0.231	-0.406	-0.272
	(0.065)	(0.060)	(0.065)	(0.060)
Implied parameters				
s	0.640	0.666	0.860	1.041
	(0.169)	(0.187)	(0.387)	(0.540)
F	2.773	1.962	4.458	3.610
	(0.910)	(0.780)	(2.020)	(2.158)
N	62	62	66	66
R^2	0.14		0.20	

Robust standard errors in parentheses. Structural parameters estimated with the Delta method.

Production is linear in equipped labor. Let M_j denote the mass of firms in city j, each choosing the quantities of commuters ℓ_c and remote workers ℓ_m according to

$$y_{jt} = \left(A_{jt} \tilde{L}_{cjt}^{\delta_j} \ell_c + z_j^{\mu} \left(A_{jt} \tilde{L}_{cjt}^{\delta_j} \right)^{1-\mu} \ell_m \right) K_{jt}^{\nu_j}.$$

Efficiency units delivered to the CBD are defined as in (2):

$$\tilde{L}_{cjt} \equiv L_{cjt}^{\mu} L_{jt}^{1-\mu}, \qquad S_{cjt} \equiv \frac{L_{cjt}}{L_{jt}}.$$

Profit maximization implies that the wage of a worker with delivery mode $\ell \in$

$\{c, m\}$ satisfies

$$w_{\ell jt} = \begin{cases} w_{cjt} = \bar{w}\, K_{jt}^{\nu_j} A_{jt} \tilde{L}_{cjt}^{\delta_j} & \text{if } \ell = c, \\ w_{mjt} = \bar{w}\, K_{jt}^{\nu_j} z_j^{\mu} \left(A_{jt} \tilde{L}_{cjt}^{\delta_j} \right)^{1-\mu} & \text{if } \ell = m. \end{cases}$$

Relative to the baseline specification, wages now scale with K_{jt}. We assume the city-wide factor grows proportionally with city size,

$$K_{jt} = g_{jt} L_{jt}^{\vartheta_j},$$

and define $\bar{A}_{jt} \equiv A_{jt} g_{jt}$ and $\varsigma_j \equiv \nu_j \vartheta_j$ as the elasticity of productivity with respect to total employment. Substituting into the commuter wage equation yields

$$\log w_{cjt} = \text{const} + \log \bar{A}_{jt} + \mu\delta_j \log S_{cjt} + (\delta_j + \varsigma_j) \log L_{jt}.$$

Taking Δt-period differences ($\Delta x_t \equiv x_t - x_{t-\Delta t}$) gives

$$\Delta \log w_{cjt} = \Delta \log \bar{A}_{jt} + \mu\delta_j \Delta \log S_{cjt} + (\delta_j + \varsigma_j) \Delta \log L_{jt}.$$

Equation (25) thus follows directly from this structural setup. The parameter δ_j captures *agglomeration externalities*—the elasticity of commuter productivity with respect to the density of CBD work activity—while ς_j captures *capital-deepening or shared-input effects*: as total city size L_{jt} grows, firms and workers benefit from more abundant non-labor factors such as infrastructure, management systems, or supplier networks. Empirically, $(\delta_j + \varsigma_j)$ governs the sensitivity of wages to changes in total city size, whereas $\mu\delta_j$ determines the response to changes in the commuter share S_{cjt}.

Main Specification

The dependent variable in Equation (28) is the log change in the hourly wages of commuters in each CBSA and year. We build it using our ACS sample and a similar approach to the one taken in Section C.C5, but focusing on city-wide wages of commuters only instead of sector-specific wages of all workers.

Namely, we define the dependent variable as the residual commuter wages for a CBSA-year $\bar{w}_{jt}$ estimated from the regression of all commuters' hourly wage on individual characteristics and for each CBSA, and year t,

$$\log w_{it} = \sum_{j=1}^{J} \mathbb{1}_{cbsa_{it}=j} \left(\gamma_{jt} X_{it} + \bar{w}_{jt} \right) + u_{it}, \tag{C2}$$

where the vector of individual covariates X_{it} contains a constant and worker i's log of the levels and squares of their work experience and age, and indicators for each occupation, levels of educational attainment, part-year work status, marital status, race, number of children, and gender.

We specify δ_j as a linear combination of CBSA j's labor-value-added employment shares in each of the four industry groups. We use estimates of the labor-value-added shares γ_s, $s = 1, \ldots, 22$, of each of the 1997-NAICS 2-digit industries from Rossi-Hansberg, Sarte and Schwartzman (2021) and compute the employment shares $shemp_{sj0}$ of each industry s in each CBSA j in 1980 in our ACS sample. We then compute the labor-value-added employment share in each industry group as,

$$s_{gj0} \equiv \sum_{s \in g} \gamma_s shemp_{sj0}, \tag{C3}$$

where $s \in g$ indicates that industry s is part of the industry group g.

Table C3 reports the results. In Panel (A), for each industry group $g \geq 1$, β_g represents the agglomeration strength in sector g, relative to nontradable industries (RUC), $g = 0$. Focusing on the IV estimates in column (2), we find that all industry groups have a higher agglomeration strength than nontradables, with the highest value for manufacturing, followed by accommodation, trade, and transportation. In Panel (B), the industry-specific coefficients are combined with each CBSA j's labor-value-added adjusted employment shares by group, s_{gj0} to form estimates δ_j. Our preferred IV estimates in column (2) indicate a mean value of δ_j of 0.067 and a standard deviation of 0.022 across the 622 CBSAs in the sample. In Appendix C.C2, we perform robustness exercises for the specifications of Equation (28); in Appendix C.C2, we define agglomeration externalities using only industry estimates from Rossi-Hansberg, Sarte and Schwartzman (2021). We find similar qualitative results in both exercises.

Robustness of the Main Specification

We test the robustness of our main specification for the estimation of agglomeration externalities by leaving out each year from the sample alternatively. The results are summarized in Table C4. They indicate that while the estimated agglomeration parameters vary somewhat if we drop one year, the mean value remains within an interval of 0.06 and 0.09.

Alternative Estimates of Agglomeration Externalities

In this section, we construct alternative estimates of agglomeration externalities and test the robustness of our quantification. We build city-specific estimates relying on industry-specific estimates of agglomeration externalities obtained by Rossi-Hansberg, Sarte and Schwartzman (2021). They specify a production function featuring labor productivity that is influenced by external effects stemming from the local composition of labor. They recover the endogenous levels of local

Table C3— Estimates of Agglomeration Externalities by Industry Group and CBSA

(A) Agglomeration Externality by Industry Group	Parameter	(1) OLS	(2) IV
Health and education (HE)	β_1	0.967	0.613
		(0.165)	(0.099)
Professional and other services (PS)	β_2	2.015	0.512
		(0.586)	(0.522)
Manufacturing (M)	β_3	0.749	1.166
		(0.833)	(0.549)
Accommodation, trade, transport. (ATT)	β_4	0.897	0.786
		(2.259)	(2.364)
	β_0	-0.412	-0.149
		(0.228)	(0.255)
Observations		9,152	9,152
First Stage Within-R^2			0.479
(B) Agglomeration Externality by CBSA			
Mean		0.043	0.067
Standard deviation		0.045	0.022

Panel (A) reports estimates of industry-group agglomeration externalities from Equation (28), with an intercept and slope of $\Delta \log L_{jt}$ for each CBSA, and year fixed effects. The dependent variable is the log time difference of commuter wages in each CBSA-year according to Equation (C2) in Appendix C.C2. The instrumental variables are the predicted changes in the number of commuters, defined in (26). The regression is based on observations for 622 CBSAs and 15 years. Robust standard errors in parentheses. We confirm the strength of the first stage by reporting the R^2 of the regression of $\Delta \log S_{cjt}$ on IV_{jt} after residualizing each on an intercept and slope of $\Delta \log L_{jt}$ for each CBSA, and year fixed effects. Panel (B) reports the first two moments of the distribution of positive CBSA-specific estimates of agglomeration externalities, computed as $\delta_j = \beta_0 + \sum_{g=1}^{4} \beta_g s_{gj0}$.

Table C4— Robustness Exercises for the Estimation of Agglomeration Externalities

Year Excluded	$p_{50}(\hat{\delta})$	$sd(\hat{\delta})$
2007	0.071	0.022
2008	0.065	0.025
2009	0.082	0.025
2010	0.086	0.032
2011	0.081	0.032
2012	0.073	0.024
2013	0.060	0.026
2014	0.073	0.022
2015	0.090	0.021
2016	0.060	0.017
2017	0.053	0.021
2018	0.061	0.029
2019	0.089	0.017
2020	0.034	0.032
2021	0.097	0.051

The table report robustness exercises in the estimation of the agglomeration externality from Equation (28). In particular, each row repeats our estimation excluding one year at a time from the sample, and reports the mean and standard deviation of δ_j across 622 CBSAs.

productivity that rationalize observed data on wages, employment, and input-output linkages as an equilibrium. Given their measures of productivity, they estimate the degree of local externalities for four industry groups and two broad occupational categories.[73]

We construct our alternative measure of city-specific agglomeration externalities by computing the average of the four industry-specific externality parameters weighted by each city's employment shares in the four industries. We compute these industry shares using the County Business Patterns in 2019.[74] Specifically, we compute the city-specific averages of the externality parameters τ_L in Table 1 of Rossi-Hansberg, Sarte and Schwartzman (2021). The average value of the agglomeration externality across 624 CBSAs is 0.089, which is close to estimates obtained in different contexts by other studies.[75] Figure C1 compares our main

[73]The industry groups, based on the NAICS 1997 classification, are Health and education (HE), Professional and other services (PS), Manufacturing (M), Accommodation, trade, and transportation (ATT). The two occupational categories are Cognitive Non-Routine (CNR) and non Cognitive Non-Routine (non-CNR). They define CNR occupations as all those with SOC-2 classifications 11 to 29 and non-CNR occupations as all other occupations.

[74]We use the crosswalk provided by Rossi-Hansberg, Sarte and Schwartzman (2021) to map the four groups of industries defined by NAICS 1997 into the industries used by the County Business Patterns. See Appendix Table A.16 page 80 of Rossi-Hansberg, Sarte and Schwartzman (2019).

[75]Duranton and Puga (2023) estimate an overall agglomeration externality of 0.08 in the United States,

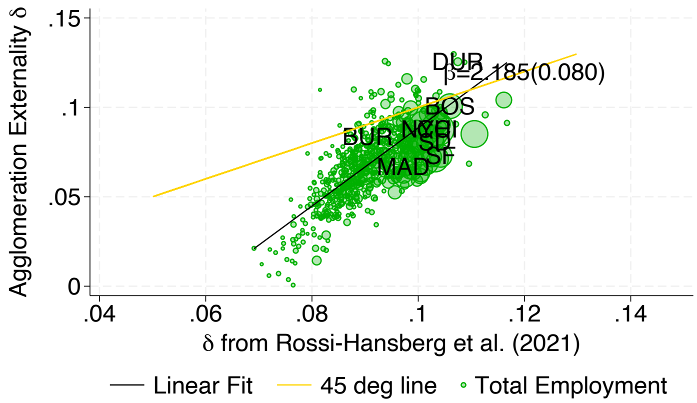


Figure C1. Agglomeration Externalities from Main Specification vs. Constructed from Rossi-Hansberg, Sarte and Schwartzman (2021)

Note: This figure reports the CBSA-specific values of δ_j implied by our main specification on the y axis and by the alternative method described in Section C.C2 on the x axis.

estimates to those from this alternative method, revealing a good correlation between the two methods. The R^2 is 0.23.

C3. Estimation of the Remote Work Premium

Estimates from the ACS

In addition to our preferred estimation of the remote work premium based on individual longitudinal data in the NLSY, we also estimate here the remote work premium based on data from the American Community Survey. Hence, rather than defining remote work based on the number of hours worked at home, we define remote work as a binary variable. Specifically, we define $d_{remote_{it}}$ as an indicator that equals one if an individual reports "work at home" as their means of transportation to work, and then predict $\hat{d} = 1$ to represent remote work.

We then estimate the national trend and an intercept for each of the 22 occupation groups, denoted by o. Specifically, we estimate the following equation:

$$\ln w_{it} = (\psi^0_{o(i)} + \psi^1 t) d_{remote_{it}} + X_{it}\beta_{o(i)} + \epsilon_{it},$$

where $\ln w_{it}$ is the log of the hourly wage for individual i at time t, $\psi^0_{o(i)}$ and ψ^1 represent the intercept and the national trend for occupation group o, respectively, and X_{it} is a vector of individual-level control variables, including age, age squared, tenure at the job, tenure squared, industry, marital status, race, education-year,

once adding the static and dynamic effects of city size on individual productivity.

number of children, metro area-year. We focus on full-time US-born workers in metropolitan areas observed between 2000 and 2018.

Finally, we estimate the remote work premium in city j as follows:

$$\psi_{jt} = \sum_o sh_{oj}^{2018}(\psi_o^0 + \psi^1 t)\hat{d},$$

where sh_{oj}^{2018} is the share of workers in city j working in occupation o in 2018. The results are reported in Table C5.

Remote Work Premium by Occupations' Teleworkable Index

Our theory predicts that cities' heterogeneous paths in the wake of the pandemic lockdowns can be explained without large changes, or any change at all, in the productivity of remote work—a feature that we view as appealing.

However, there may have been some changes in the productivity of remote work, and these changes may have been heterogeneous across cities. One possibility is that occupations that offer a high potential for telework may have experienced larger increases in productivity, leading CBSAs with higher employment shares in those occupations to experience higher average increases in remote work productivity.

We investigate this possibility by plotting a version of Figure 1b for two groups of occupations in the ACS: those above and below the median share of teleworkable index, according to Dingel and Neiman (2020).

The occupation groups above the median value of teleworkable index are "Executive, Administrative, and Managerial Occupation", "Management Related Occupations," "Mathematical and Computer Scientists," "Architecture and Engineering," "Life, Physical, and Social Services," "Community and Social Services," "Legal," "Education, Training, and Library," "Art, Design, Entertainment, Sports and Media," "Sales and Related," and "Office and Administrative Support."

Those below the median value of teleworkable index are "Healthcare Practitioners and Technical," "Healthcare Support," "Protective Services," "Food Preparation and Serving Related," "Building and Grounds Cleaning and Maintenance," "Personal Care and Service," "Farming, Forestry, and Fishing," "Construction and Extraction," "Installation, Repair, and Maintenance," "Production," "Transportation and Material Moving," and "Experienced Unemployed not Classified by Occupation."

We report the results in Figure C2. The two series show that, expectedly, the premium for remote work has been higher for occupations with high teleworkable index. However, both groups have experienced parallel upward trends since 2005, with no clear sign of differential growth, including after 2020. We view this as another indication that large heterogeneous changes in the productivity of remote work of specific occupations is unlikely to be the main driver of the large divergence in city structure since the COVID-19 pandemic.

Table C5— Remote Work Premium for Occupation Groups

	(1) NLSY	(2) ACS
Remote	-0.006 (0.001)	0.029 (0.000)
Remote × Year	0.001 (0.000)	0.006 (0.000)
Management Related Occupations × Remote	0.000 (0.001)	-0.058 (0.000)
Math. and Computer Scientists × Remote	0.002 (0.001)	-0.021 (0.000)
Architecture and Engineering × Remote	0.000 (0.001)	-0.009 (0.000)
Life, Physical, and Social Services × Remote	-0.002 (0.004)	0.013 (0.001)
Community and Social Services × Remote	-0.001 (0.003)	-0.190 (0.001)
Legal × Remote	0.002 (0.002)	-0.060 (0.001)
Education, Training, and Library × Remote	0.003 (0.002)	-0.064 (0.000)
Art, Design, Entertainment, Sports × Remote	-0.003 (0.003)	-0.080 (0.001)
Healthcare Practitioners × Remote	-0.001 (0.002)	-0.151 (0.000)
Healthcare Support × Remote	-0.002 (0.003)	-0.373 (0.001)
Protective Services × Remote	-0.000 (0.003)	-0.086 (0.001)
Food Preparation and Serving × Remote	-0.004 (0.003)	-0.153 (0.001)
Building Cleaning and Maintenance × Remote	-0.003 (0.003)	-0.198 (0.001)
Personal Care and Servic × Remote	-0.003 (0.002)	-0.325 (0.001)
Sales and Related × Remote	0.002 (0.001)	0.140 (0.000)
Office and Administrative Support × Remote	-0.002 (0.001)	-0.050 (0.000)
Farming, Forestry, and Fishing × Remote	-0.009 (0.004)	-0.038 (0.001)
Construction and Extraction × Remote	-0.003 (0.004)	-0.111 (0.001)
Installation, Repair, Maintenance × Remote	-0.002 (0.001)	-0.125 (0.001)
Production × Remote	-0.001 (0.003)	-0.069 (0.001)
Transportation, Material Moving × Remote	-0.009 (0.007)	-0.141 (0.001)
Indiv. Controls	Yes	Yes
Year FE	Yes	Yes
Occ. & Ind. FE	Yes	Yes
Indiv. FE	Yes	No
Observations	40,385	15,872,072

Robust standard errors in parentheses. Occupation groups based on the occupation classification used in the 2000 decennial census, denoted `OCC` in IPUMS.

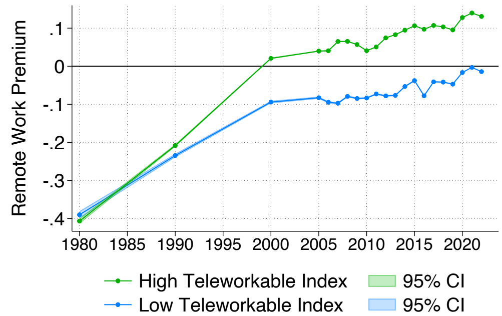


Figure C2. Remote Work Hourly Wage Premium – Low vs. High Teleworkable Index Occupations

Note: This figure reports the coefficients of yearly remote work indicators from the regression of log hourly wages, including the same controls as in Panel (b) of Figure 1b, for two subsamples: individuals working in occupation groups with above and below median teleworkable index, as defined by Dingel and Neiman (2020). The shaded areas represent the 95 confidence interval around the estimates.

C4. Estimation of the Transportation Costs Elasticity

Comparison to the Literature

As reported in Table 1, the mean transportation-cost elasticity across the 278 CBSAs in our sample is 0.026. This value is lower than some other estimates in the literature: for instance, Duranton and Puga (2023) obtain an elasticity of 0.077 from a pooled bid-rent regression covering a broad set of U.S. metropolitan areas. This appendix clarifies why our estimate is consistent with our model's structure and how the difference arises from the mapping between the empirical rent gradient and the underlying transportation-cost elasticity. Using our data but applying the Duranton and Puga (2023) empirical specification, we replicate their result, confirming that the discrepancy stems from model interpretation rather than data or estimation differences.

Duranton and Puga (2023) derive the spatial rent gradient from a linear commuting cost specification, where a resident located at distance x from the CBD incurs a cost $T_{it}(x) = \tau_{it} x^{\gamma}$, so that $d \ln T_{it}(x)/d \ln x = \gamma$. In their additive framework, commuting costs enter directly in dollars rather than in utility, and equilibrium requires rents $P_{it}(x)$ to relate to commuting costs according to

$$P_{it}(0) - P_{it}(x) = T_{it}(x).$$

Differentiating implies $d \ln(P_{it}(0) - P_{it}(x))/d \ln x = \gamma$: the rent gradient in levels exactly mirrors the elasticity of commuting costs with distance. Because housing consumption is fixed (each worker consumes one unit of land), the slope of rents directly identifies γ.

By contrast, our framework embeds the same functional form for $T_{it}(x) = \tau_{it} x^{\gamma}$ in a utility-based model with *endogenous housing consumption*. In this case, commuting costs affect utility multiplicatively, and the rent gradient reflects both commuting and housing-consumption elasticities. From our equation (12), we obtain, with the same notation as above,

$$\frac{d \ln P_{it}(x)}{d \ln x} = -\gamma \frac{(1-\mu_{\ell})}{(1-\alpha)},$$

and the general expression in equation (33) gives

$$\frac{d \ln P_{it}(x)}{d \ln x} = -\frac{\gamma}{1-\alpha} \left(1 - \mu + \mu \frac{L_{cit}}{L_{it}}\right).$$

Before the widespread adoption of remote work ($L_{cit} \simeq L_{it}$), this simplifies to

$$\frac{d \ln P_{it}(x)}{d \ln x} \simeq -\frac{\gamma}{1-\alpha}.$$

Since the share of non-housing consumption is $\alpha = 0.76$, rents must decline roughly four times faster than commuting costs to maintain spatial indifference. A measured rent gradient of -0.077 thus implies $\gamma \simeq 0.02$ in our model, whereas in Duranton and Puga (2023)'s additive model it would imply $\gamma = 0.077$.

To confirm this interpretation, Table C6 re-estimates the average rent gradient using the set of CBSAs, controls, and functional form closest to those in Duranton and Puga (2023). Both analyses use ACS data on median gross rents at the block-group level, though our period is 2015–2019 (theirs is 2008–2012). When using the log of median rent as the dependent variable—as in our baseline specification—we obtain a slope of -0.070. When we instead use the log of the difference in rents between the most expensive block group and each location—as in Duranton and Puga (2023)—we obtain 0.071, closely matching their 0.077. Hence, the lower value of γ in our results reflects not a weaker spatial relationship in the data but a different theoretical mapping from rent gradients to commuting costs. Models with fixed housing consumption identify the rent slope directly as γ, while in our model with adjustable housing, the rent gradient scales with $1/(1-\alpha)$, leading to proportionally lower implied values of γ for the same empirical gradient.

Transportation Costs Elasticity and City Size

In this section, we report the relation between the estimated distance elasticity of transportation costs, γ, and CBSA size, across the 622 cities in our sample. This relation is depicted in Figure C3, where we exclude a few outliers for visibility. We estimate that a doubling of city size is associated to a distance that is 0.001 smaller, relative to a sample mean of 0.026.

Table C6— Average Transportation Costs Elasticity Across CBSAs

	$\ln(P_{it}(x))$ (1)	$\ln(P_{it}(0) - P_{it}(x))$ (2)
Log distance to city center	-0.070	0.071
	(0.007)	(0.007)
CBSA indicators	Yes	Yes
Block-group and dwelling controls	Yes	Yes
R^2	0.289	0.214
Observations	90,421	90,421

This table reports OLS estimates of the average bid-rent gradient across the sample of CBSAs closest to the set of MSAs used in Duranton and Puga (2023). Units of observation are city block-groups and the dependent variable is i) log of the median housing rental price in the block-group in column (1); ii) the log of the difference in the median housing rental price between the most expensive block-group in the city and the block-group under consideration in column (2). Controls in both columns are percentages of Hispanic, Black, and Asian population, performance in standardized tests of the closest public school relative to the city average, waterfront indicator, and ruggedness. Dwelling controls are percentage dwellings in block-group by type of structure, number of bedrooms, and construction decade. The R2 is within city.

Allowing for Endogenous Housing Supply

In this subsection, we examine how allowing the housing supply—and thus population density—to respond to the rent level affects equilibrium outcomes. We show that, while housing elasticity modifies the location and magnitude of the kinks in the rent function, the relationship between the average rent–distance slope and the transportation-cost elasticity γ remains almost identical to that in the baseline model with fixed housing supply, especially when the commuter share is large, as in our estimation sample.

Assume that housing supply $h(d)$ at distance d from the CBD on each ray follows

$$h(d) = ar(d)^{\eta}, \qquad a > 0, \quad \eta \geq 0. \tag{C4}$$

Here η measures the elasticity of local housing density with respect to rent. When $\eta = 0$, housing supply is perfectly inelastic.

The slope of the rent function is determined by agents' joint optimization of consumption, housing, and residential location, taking rents as given. This slope is therefore unaffected by the elasticity of housing supply and remains given by (13), reproduced below:

$$r(d) = \begin{cases} \left(\bar{D}/d\right)^{\gamma/(1-\alpha)} \left(\bar{d}/\bar{D}\right)^{\mu\gamma/(1-\alpha)} r_a & \text{if } d < \bar{d} \\ \left(\bar{D}/d\right)^{(1-\mu)\gamma/(1-\alpha)} r_a & \text{if } d \geq \bar{d} \end{cases}.$$

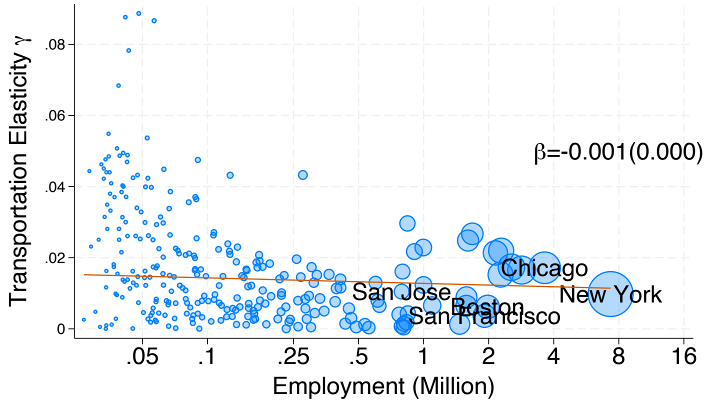


Figure C3. Transportation Costs Elasticity, γ, and CBSA size

Note: This figure reports the city-specific γ_j against the CBSA's BEA employment. Each γ_j is estimated as described in Section IV.D. The size of the circles represents CBSA's employment.

As a result, the average elasticity of rent with respect to distance, κ, remains

$$\kappa \equiv \frac{1}{\bar{D}} \int_{x=0}^{\bar{D}} \frac{\partial r(x)}{\partial x} \frac{x}{r(x)} dx = -\frac{\gamma}{1-\alpha} \left[(1-\mu) + \mu \frac{\bar{d}}{\bar{D}} \right].$$

Hence, the mapping between κ_j—which we estimate for each CBSA j in Section IV.D—and the transportation-cost elasticity γ_j differs only through the ratio $\bar{d}/\bar{D}$, whose expression changes when $\eta > 0$. As shown in Section III.D, when $\eta = 0$, $\bar{d} = (L_c/\pi)^{1/2}$ and $\bar{D} = (L/\pi)^{1/2}$, implying $\bar{d}/\bar{D} = (L_c/L)^{1/2}$.

To characterize $\bar{d}/\bar{D}$ under elastic supply, define $\phi \equiv \eta\gamma/(1-\alpha)$. Integrating housing units over the commuter disk and the entire city yields:

(C5)
$$L_c = \int_0^{\bar{d}} h(x) 2\pi x \, dx = \left(\bar{D}^{\,1-\mu} \bar{d}^{\mu}\right)^{\phi} r_a^{\eta} \, 2\pi a \, \frac{\bar{d}^{\,2-\phi}}{2-\phi} = \frac{2\pi a \, r_a^{\eta}}{2-\phi} \, \bar{D}^{\,\phi(1-\mu)} \bar{d}^{\,2-\phi(1-\mu)},$$

(C6)
$$L = L_c + \int_{\bar{d}}^{\bar{D}} h(x) 2\pi x \, dx = L_c + \frac{2\pi a \, r_a^{\eta}}{2-\phi(1-\mu)} \left(\bar{D}^{\,2} - \bar{D}^{\,\phi(1-\mu)} \bar{d}^{\,2-\phi(1-\mu)}\right).$$

Eliminating constants and powers of $\bar{D}$ gives the closed-form relationship

$$\left(\frac{\bar{d}}{\bar{D}}\right)^{2-\phi(1-\mu)} = \frac{(2-\phi)\, L_c}{(2-\phi(1-\mu))\, L - \phi\mu \, L_c}. \tag{C7}$$

Two implications follow immediately:
(i) Fixed housing is the limiting case. When $\eta = 0$ (so $\phi = 0$), (C7) reduces to

$$\left(\frac{\bar{d}}{\bar{D}}\right)^2 = \frac{L_c}{L} \quad \Rightarrow \quad \frac{\bar{d}}{\bar{D}} = \left(\frac{L_c}{L}\right)^{1/2},$$

which corresponds to the benchmark used in the main text.
(ii) With elastic housing, deviations are small for empirically plausible parameters. For benchmark parameter values $\gamma = 0.027$, $\alpha = 0.76$, $\mu = 0.6$, and a pre-pandemic commuter share $L_c/L = 0.96$, a high housing-supply elasticity $\eta = 3$—near the upper bound of empirical estimates for U.S. metropolitan areas—yields $\bar{d}/\bar{D} = 0.976$ from equation (C7), compared to $(L_c/L)^{1/2} = 0.980$. This difference corresponds to a deviation of only -0.4 percent. Even under extremely elastic housing supply, the resulting correction is thus quantitatively negligible. Across the empirically relevant range $0 \leq \eta \leq 3$, deviations remain below one-half of one percent, confirming that the mapping between the estimated rent gradient κ and the transportation-cost elasticity γ derived under inelastic housing supply remains an excellent approximation.

C5. Estimation of the Elasticity of Substitution Between Remote and In-Person Workers

Model with Multi-Sector CES Production and Inelastic Labor Supply

The model in Section III focuses on a representative production sector that implicitly assumes a linear production function in remote and in-person workers, such that the relative wage of remote and in-person workers is independent of their relative labor supply. These assumptions translate into a simple strategy for estimating the remote work premium based on the relative wages of remote and in-person workers.

In this Section, we provide evidence in support of these assumptions. We develop an alternative model that allows for multiple sectors and a CES production in remote and in-person workers. We derive estimating equations for the elasticity of substitution between remote and in-person workers and find very large values that do not reject the linear specification in the main model. For that reason, we maintain the linear production model as our main model.

Consider a disk-shaped city with an exogenous population of size L. There are two final goods in the economy: consumption and housing. The final consumption good Q, the price of which we normalize to 1, is produced according to a Cobb-Douglas function over intermediate sectoral goods,

$$\text{(C8)} \qquad Q = \prod_{s=1}^{S} q_s^{\beta_s}, \qquad \beta_s \in (0,1),\ s = 1, \ldots, S,, \qquad \sum_{s=1}^{S} \beta_s = 1.$$

Sectoral goods are produced in the CBD by profit-maximizing firms under constant returns to scale according to a CES production function combining in-person and remote workers,

$$q_s = A_s \left[A_{cs} L_{cs}^{\theta} + A_{ms} L_{ms}^{\theta} \right]^{1/\theta}, \qquad \theta \leq 1, \tag{C9}$$

where L_{cs} is the number of in-person workers, L_{ms} is the number of remote workers, and $\theta = (\rho - 1)/\rho$, where ρ is the elasticity of substitution between remote and in-person workers. The efficiency of in-person and remote workers in sector s depends on the total economic activity that takes place in the city center across all sectors, which we measure, as in Section III, by defining

$$\tilde{L}_c \equiv L_c^{\mu} L^{1-\mu}, \tag{C10}$$

a function of the current number of commuters $L_c = \sum_{s=1}^{S} L_{cs}$, the total city size $L = L_c + L_m$, and the fraction of days per week a remote worker spends at home, μ. We postulate that the factor shares for each type of labor are given by

$$A_{cs} \equiv \frac{\left(A_s \tilde{L}_c^{\delta} \right)^{\mu}}{z_s^{\mu} + \left(A_s \tilde{L}_c^{\delta} \right)^{\mu}}, \qquad A_{ms} \equiv \frac{z_s^{\mu}}{z_s^{\mu} + \left(A_s \tilde{L}_c^{\delta} \right)^{\mu}}. \tag{C11}$$

The productivity of a commuting worker depends on the in-office productivity in that sector $A_s > 0$, and on how many workers are on average in the CBD, $\tilde{L}_c$, with an elasticity $\delta > 0$. Similarly, the productivity of a remote worker depends on the remote work productivity in that sector z_s. Hence, the optimal ratio of remote workers increases in their productivity, the more so as they spend a higher share of the week at home. The wage for workers in mode ℓ in sector s resulting from profit maximization is given by

$$w_{\ell s} = p_s A_s^{\theta} A_{\ell s} L_{\ell s}^{\theta - 1} q_s^{1-\theta}, \tag{C12}$$

where p_s is the price of sectoral good s. The wage of commuters relative to remote workers in sector s is therefore given by

$$\frac{w_{ms}}{w_{cs}} = \left(\frac{z_s}{A_s \tilde{L}_c^{\delta}} \right)^{\mu} \left(\frac{L_{ms}}{L_{cs}} \right)^{\theta - 1}. \tag{C13}$$

Absentee landlords invest their effort to produce housing on their unit-size plot and receive $r(d)$ per unit of housing. Land has an alternative use that yields r_a.

Equation (C13) will be the basis of our estimation strategy of the elasticity of substitution between remote and in-person workers. We aim to estimate $\phi \equiv \theta - 1 = -1/\rho$.

Estimates of the Elasticity of Substitution Between Remote and In-Person

Consider equation (C13) for different cities $j = 1, \dots, J$ and different periods $t = 1, \dots, T$. Let δ_j be the agglomeration externality in city j. We assume that the technology z_{st} is constant across cities and let in-office productivity be the product of sector-year, city-year terms, and a residual $A_{sjt} = \bar{A}_{st}\bar{a}_{jt}\tilde{a}_{sjt}$. Taking logs and adding time and city indices, equation (C13) writes

$$\log \frac{w_{msjt}}{w_{csjt}} = D_{st} + G_{jt} + \phi \log \frac{L_{msjt}}{L_{csjt}} + \epsilon_{sjt}, \tag{C14}$$

where $D_{st} \equiv \mu \log \frac{z_{st}}{\bar{A}_{st}}$, $G_{jt} \equiv -\mu\bar{a}_{jt} - \mu\delta_j \log \tilde{L}_{cjt}$, $\phi \equiv \theta - 1$, and the error term $\epsilon_{sjt} = -\mu \log \tilde{a}_{sjt}$ captures unobserved productivity shocks specific to each sector-city-year.

A threat to identification in (C14) is the presence of contemporaneous unobserved innovations to the productivity of any sector s in a city j. These innovations are likely to affect both the relative wages and labor supply of remote workers. We estimate ϕ using past ratios of the relative labor supply, $L_{msjt'}/L_{csjt'}$, with $t' < t$, as instrument for the ratio in period t, under the exclusion restriction that those past ratios are orthogonal to the contemporaneous innovations in the sector-city component of productivity.

We estimate Equation (C14) with the ACS data, using all years between 1980 and 2022, 638 CBSAs, and 33 distinct industry groups. In an alternative specification, we replace industry groups with the 22 occupation groups defined in Section A.A2. We define the dependent variable as the residual remote work premium for a sector-CBSA-year ψ_{sjt} estimated from the regression of all workers' hourly wage on individual characteristics and indicators of being remote in each sector and CBSA, and year t,

$$\log w_{it} = \sum_{s=1}^{S} \mathbb{1}_{sec_{it}=s} \left(\gamma_{st} X_{it} + \sum_{j=1}^{J} \mathbb{1}_{cbsa_{it}=j} \left(\beta_{jst} + \psi_{jst} \mathbb{1}_{remote_{it}=1} \right) \right) + u_{it}, \tag{C15}$$

where the vector of individual covariates X_{it} contains a constant and worker i's log of the levels and squares of their work experience and age, and indicators for levels of educational attainment, part-year work status, marital status, race, number of children, and gender.

We report the estimates of ϕ from equation (C14) in Table C7. Recall that $\phi = -1/\rho$, so for any value of $\rho \in [0, \infty)$, $\phi \in (-\infty, 0)$, and values of ϕ close to zero correspond to very high values of the elasticity of substitution ρ.

The exclusion restriction in our estimation of Equation (C14) is that the levels of the innovations in city-sector productivity are independent over time. Next, we consider a different exclusion restriction and assume that the changes in in-

Table C7— Estimates of Substitution Parameter Between Remote and In-Person

	By Occupation				By Industry			
	Levels		Time Diff.		Levels		Time Diff.	
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7) OLS	(8) IV
$\log \frac{L_{m,sjt}}{L_{c,sjt}}$	0.005	-0.002			-0.001	-0.017		
	(0.002)	(0.015)			(0.002)	(0.014)		
$\Delta \log \frac{L_{m,sjt}}{L_{c,sjt}}$			0.003	0.006			-0.003	0.005
			(0.003)	(0.008)			(0.003)	(0.008)
CBSA-Y FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Group-Y FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	104796	44095	44095	25134	102642	41331	41331	23733

Estimates of $\phi = -1/\rho$, where ρ is the elasticity of substitution between remote and in-person workers. Columns (1)-(4) define worker groups as the 22 occupation groups in Section A.A2; Columns (5)-(8) define worker groups as the 33 two-digit 1990-industry codes in the ACS. Columns (1)-(2) and (5)-(6) estimate Equation (C14); Columns (3)-(4) and (7)-(8) estimate Equation (C16). The dependent variable is the estimated remote work hourly wage premium, in levels or in time-difference, from Equation (C15). The period of reference for the definition of instrumental variables is $t' = t - 3$, and time differences are defined with $\tau = 3$.

novations are independent over time. This allows for the estimation of ϕ in time differences. For any $\tau \in \mathbb{N}$, define $\Delta x_t = x_t - x_{t-\tau}$ for any variable x. Then, a consistent estimate of ϕ can be obtained using past values of the change in the relative supply of remote workers, $\Delta \log L_{msjt'}/L_{csjt'}$, with $t' < t$,

$$\Delta \log \frac{w_{msjt}}{w_{csjt}} = d_{st} + g_{jt} + \phi \Delta \log \frac{L_{msjt}}{L_{csjt}} + \varepsilon_{sjt}, \tag{C16}$$

where $d_{st} \equiv \Delta D_{st}$, $g_{jt} \equiv \Delta G_{jt}$, and $\varepsilon_{sjt} \equiv \Delta \epsilon_{sjt}$. We report the estimates of ϕ from equation (C16) in Table C7, where instruments are defined for a reference period $t' = t - 3$ and time differences are defined with $\tau = 3$. In all our specifications, whether dividing workers by occupation or industry groups, estimating the elasticity of substitution using levels or time differences, with OLS or instrumental variables, the estimates are consistently centered on zero, implying a near-infinite elasticity of substitution between remote and in-person workers.

From Table C7, we conclude that the association between the relative labor supply of remote workers and their compensation is consistent with a very large elasticity of substitution, and with setting $\phi = 0$, i.e., $\theta = 1$ in Equation (C9), implying that the production function is well represented by a linear function as in the paper.

Empirical Determinants of Multiplicity and Robustness

D1. Correlates Between Multiplicity, Trip Shortfall, and Model Parameters

In this subsection we ask what determines, empirically, whether a city is inside its cone. The several dimensions of heterogeneity that we have exploited to quantify the cities' parameters, like industrial and occupational composition and the rent gradients, all jointly contribute to this position. To organize our discussion, we leverage our theory, which indicates that the position of a city with respect to its cone is influenced by the city's size, agglomeration and transportation costs elasticities, and the relative productivity of remote work.

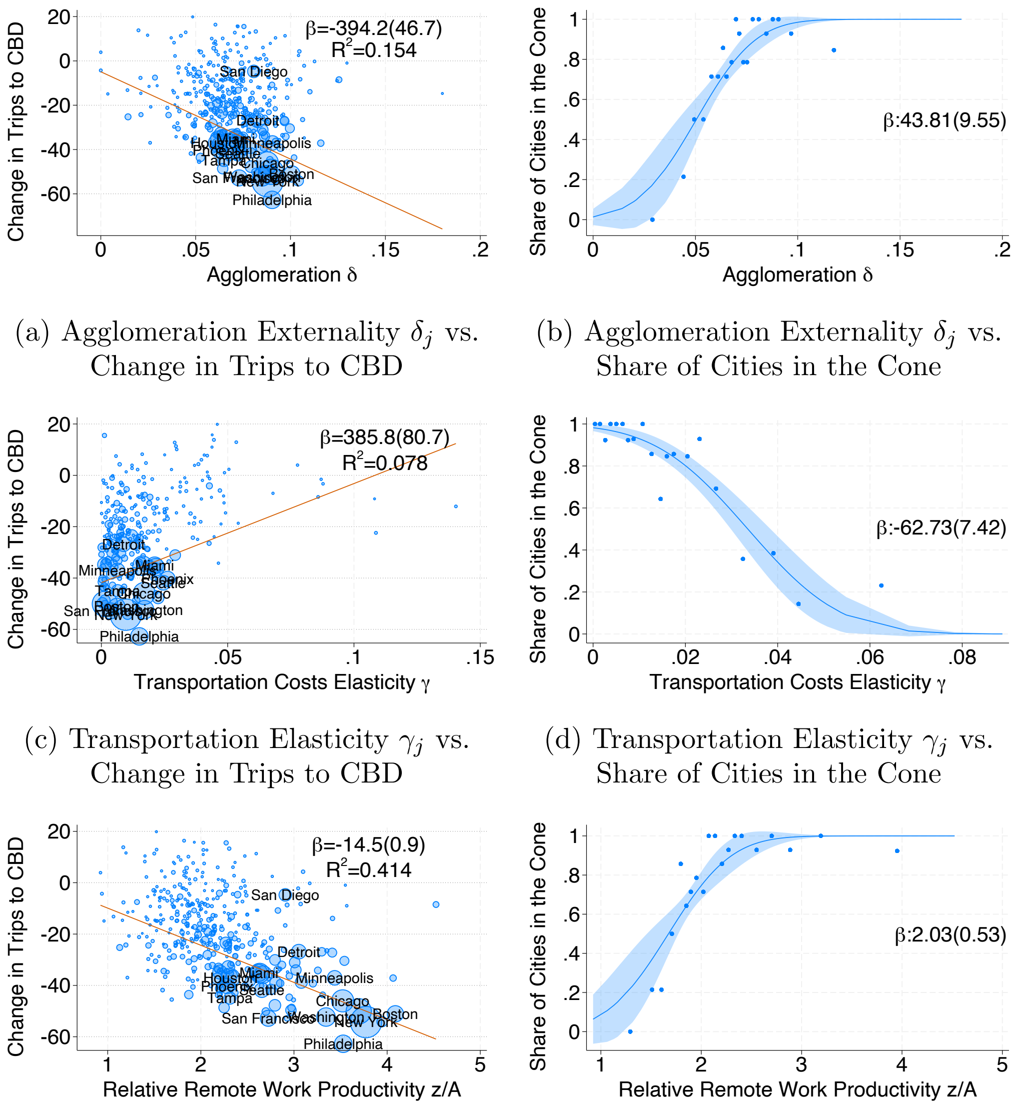
The left side of Figure D1 reports the relationship between the change in visits to the CBD and our estimates of agglomeration externalities, δ_j, transportation elasticities γ_j, and relative remote productivity z_j/A_j. Each bubble represents a city of population proportional to the bubble area. The right side of Figure D1 reports the share of cities that are predicted to be in their cone of multiplicity as a function of the same parameters. Blue dots indicate the fraction of cities in the cone in the calibrated model for 20 quantiles of the variable indicated on the horizontal axis.

The left side of the figure shows that cities that experienced larger declines in visits to the CBD are those that have a larger agglomeration externality, a smaller transportation costs elasticity, and a higher relative productivity of remote work. These relationships are quite strong, as indicated by the magnitude and significance of the population-weighted linear fit in the figures and values of R^2 between 0.08 and 0.41.

The right side of the figure shows that the relationship between the drop in visits to the CBD and the key agglomeration parameters is rooted in the workings of the model. The figures show that calibrated cities with higher agglomeration externalities δ_j, lower transportation elasticities γ_j, and higher relative productivity of remote work z_j/A_j are those that are the most likely to be inside the cone of multiplicity.[76] Therefore, our model predicts that these are the cities that are the most likely to remain in a low-commuting equilibrium after a temporary shock such as a lockdown.

In Table D1, we explore the relative ability of these factors to relate to CBD trips, and their contribution to whether each city is inside or outside its cone – the multivariate versions of the left and right columns of Figure D1. We note that these partial correlations should be interpreted with caution, since is it their joint combination that determines whether or not a city exhibits multiple stationary equilibria. In particular, in Column (1) we run a linear regression of the shortfall in trips to the CBD on standardized values of $L_j, z_j/A_j, \delta_j$, and γ_j. We find that the variation in agglomeration elasticity and the relative remote work productivity

[76]Panel (d) in this figure excludes some large outliers focusing on $\gamma_j < 0.1$. This choice is made for the visibility of the graph, and the estimates return unchanged messages without these restrictions.



(a) Agglomeration Externality δ_j vs. Change in Trips to CBD

(b) Agglomeration Externality δ_j vs. Share of Cities in the Cone

(c) Transportation Elasticity γ_j vs. Change in Trips to CBD

(d) Transportation Elasticity γ_j vs. Share of Cities in the Cone

(e) Rel. Remote Work Productivity z_j/A_j vs. Change in Trips to CBD

(f) Rel. Remote Work Productivity z_j/A_j vs. Share of Cities in the Cone

Figure D1. Visits to CBD and Share of Cities in the Cone vs. City Parameters

Table D1— Probit Regression of Cone Indicator on Model-Predicted CBSA Characteristics

	(1) Trip shortfall	(2) $\mathbb{1}_j^{\text{cone}}$
Employment	-2.37	2.12
	(1.74)	(0.57)
Relative Remote Productivity, z_j/A_j	-16.50	-6.56
	(4.07)	(1.30)
Agglomeration Externality, δ_j	12.66	8.69
	(3.24)	(1.43)
Transportation Cost Elasticity, γ_j	1.54	-18.60
	(0.91)	(2.49)
Observations	274	274
R^2	0.334	
Pseudo R^2		0.826

Column (1) of this table shows the results of a linear regression of the shortfall of trips in July 2022 relative to January 2020, as described in the main text, on $\mathbb{1}_j^{cone}$ on $L_j, z_j/A_j, \delta_j$, and γ_j. Column (2) shows the results of a probit regression on the same set of variables. All the regressors have been standardized. Robust standard errors in parentheses.

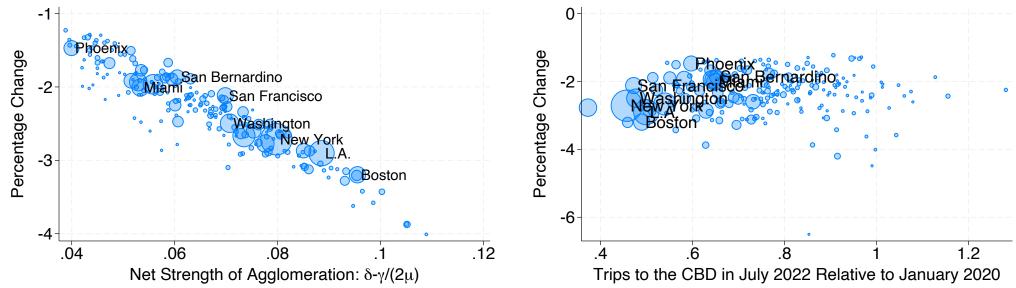
are the most strongly associated to the drop in trips to CBD. In column (2), we estimate a probit regression of the indicator of whether the city is inside its cone on standardized values of $L_j, z_j/A_j, \delta_j$, and γ_j. Empirically, we find a large role for the distance elasticity of transportation costs, γ_j, which is natural given the large variability we find in the data, followed by the agglomeration externality, δ_j, and relative productivity of remote work. The high pseudo-R^2 suggests that there is little residual variation to be explained by the interactions of these factors.

D2. Welfare Effects from Switching to Remote Equilibrium for Largest Cities

In Table D2, we report the values of our estimates for the change in visits to the CBD and the welfare of the low-commuting stationary equilibrium relative to the high-commuting stationary equilibrium. Our sample is the largest 45 U.S. cities for which we can compute the relevant parameters and the commuting trips to the CBD.

In the table, Column 2 reports the employment from the 2019 Census that we use in our calibrations, as described in Appendix B.B6. Column 3 reports the change in visits to the CBD for each city from January 2020 to July 2022, as described in Appendix B.B1.[77] Column 4 reports the welfare ratio $\mathcal{W}(L^{ss}_{c\,low})$ $/\mathcal{W}(L^{ss}_{c\,high})$, defined in Section VI. These values are only computed for the

[77]These values are a simple transformation of those plotted on the y axis of Panel (b) of Figure 3 (the complement to 100).



(a) With Net Agglomeration Strength (b) With CBD Trip Decline

Figure D2. Correlates of Welfare Change $\hat{\mathcal{W}}$

CBSAs that are in their cone of multiplicity according to our estimation. We indicate by "U" the CBSAs that are outside their cone, for which there exists no low-commuting stationary equilibrium.

Figure D2 shows correlates of $\hat{\mathcal{W}}$, the welfare in the low-commuting relative to the high-commuting stationary equilibrium. These scatterplots can only be computed for the set of cities in the cone of multiplicity. The left panel shows a scatterplot between $\hat{\mathcal{W}}$ and the net strength of agglomeration externalities, $\delta - \gamma/(2\mu)$. The right panel shows a scatterplot between $\hat{\mathcal{W}}$ and the share of trips to the CBD in July 2022, relative to January 2020. The marker size is proportional to total city employment.

D3. Robustness to Alternative Definitions of Switching

This section tests the robustness of our empirical results to an alternative definition of a city's return to the high-commuting equilibrium. In addition to the measure of switching behavior used in Section V, we define a more stringent indicator, $\mathbb{1}_j^{back(0,30)}$, which equals one if trips in July 2022 are at or above their pre-pandemic level, and zero if they declined by more than 30%. This alternative threshold captures only the most complete recoveries in commuting activity.

As shown in Table D3, the results remain broadly consistent with those in Table 2. The cone indicator $\mathbb{1}_j^{cone}$ continues to be strongly correlated with the probability that a city returns to its pre-pandemic level of CBD trips, with a pseudo R^2 of 0.28 when used as the only explanatory variable. The magnitude of the coefficient is similar across specifications, although it becomes smaller and less precisely estimated once total CBSA employment is included as a control. This confirms that the predictive power of the cone indicator is not driven by a particular definition of "return," but is robust across alternative thresholds.

Table D2— Change in Trips to CBD and Welfare Loss in Remote Equilibrium for a Group of 45 Cities

CBSA	2019 Census Employment	Trips to CBD Jul'22 / Jan'20	Welfare Change $\mathcal{W}(L^{ss}_{c,low})$ $/\mathcal{W}(L^{ss}_{c,high})$
New York-New Jersey-Long Island, NY-NJ-PA	7,334,144	0.457	0.973
Los Angeles-Long Beach-Santa Ana, CA	4,748,281	0.499	0.971
Chicago-Joliet-Naperville, IL-IN-WI	3,630,720	0.540	0.973
Dallas-Fort Worth-Arlington, TX	2,842,177	0.648	0.980
Washington-Arlington-Alexandria, DC-VA-MD-WV	2,548,552	0.477	0.975
Atlanta-Sandy Springs-Marietta, GA	2,284,940	0.642	0.981
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	2,270,070	0.373	0.972
Miami-Fort Lauderdale-Pompano Beach, FL	2,149,512	0.649	0.980
Boston-Cambridge-Quincy, MA-NH	1,979,049	0.490	0.968
San Francisco-Oakland-Fremont, CA	1,914,574	0.472	0.979
Phoenix-Mesa-Glendale, AZ	1,680,550	0.597	0.985
Seattle-Tacoma-Bellevue, WA	1,606,160	0.582	0.981
Detroit-Warren-Livonia, MI	1,578,962	0.732	0.974
Minneapolis-St. Paul-Bloomington, MN-WI	1,575,451	0.630	0.971
Riverside-San Bernardino-Ontario, CA	1,468,173	0.652	0.981
Tampa-St. Petersburg-Clearwater, FL	1,095,802	0.551	0.981
Charlotte-Gastonia-Rock Hill, NC-SC	999,016	0.766	0.982
Portland-Vancouver-Hillsboro, OR-WA	996,704	0.523	0.978
Austin-Round Rock-San Marcos, TX	906,522	0.639	0.983
Cincinnati-Middletown, OH-KY-IN	845,880	0.699	0.975
Pittsburgh, PA	841,832	0.689	0.975
Kansas City, MO-KS	832,707	0.655	0.977
Sacramento–Arden-Arcade–Roseville, CA	811,475	0.514	0.981
Las Vegas-Paradise, NV	807,521	0.504	0.971
Nashville-Davidson–Murfreesboro–Franklin, TN	798,725	0.660	0.974
San Jose-Sunnyvale-Santa Clara, CA	791,561	0.459	0.968
Columbus, OH	791,421	0.585	0.975
Indianapolis-Carmel, IN	771,329	0.615	0.977
Virginia Beach-Norfolk-Newport News, VA-NC	624,538	0.669	0.981
Providence-New Bedford-Fall River, RI-MA	612,059	0.696	0.967
Milwaukee-Waukesha-West Allis, WI	599,389	0.728	0.969
Raleigh-Cary, NC	558,030	0.719	0.983
Richmond, VA	534,987	0.564	0.985
Salt Lake City, UT	488,013	0.596	0.978
Oklahoma City, OK	461,898	0.809	0.981
Hartford-West Hartford-East Hartford, CT	451,318	0.680	0.973
New Orleans-Metairie-Kenner, LA	434,260	0.498	0.974
Buffalo-Niagara Falls, NY	412,839	0.609	0.972
Grand Rapids-Wyoming, MI	412,449	0.513	0.971
Rochester, NY	395,591	0.608	0.969
Omaha-Council Bluffs, NE-IA	389,758	0.777	0.982
Tulsa, OK	365,181	0.755	0.979
Worcester, MA	346,749	0.658	0.969
Madison, WI	276,371	0.959	U
Baton Rouge, LA	270,329	0.748	U

The letter "U" in Column 4 indicates that the CBSA is outside the multiplicity cone according to our estimates.

D4. *Robustness to Alternative Values of Transportation and Agglomeration Elasticities*

In this subsection, we investigate the sensitivity of our main results to alternative calibrations of the model that shift the values of all transportation elasticities

Table D3— Determinants of CBSAs' Changes in Trips to CBD: Observable Characteristics

	$\mathbb{1}_j^{\text{back}(5,20)}$			$\mathbb{1}_j^{\text{back}(0,30)}$		
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbb{1}_j^{\text{cone}}$	-1.26	-1.16	-1.10	-1.92	-1.38	-1.19
	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.08)
CBSA Employment			-6.59			-26.66
			(0.01)			(0.00)
Share of Teleworkable Employment		-5.60	-6.07		-2.06	5.86
		(0.08)	(0.04)		(0.76)	(0.31)
Share of Trips to CBD in 2019		-0.29	-0.59		-0.53	-2.95
		(0.31)	(0.03)		(0.46)	(0.00)
s_{j0} Health and education		5.87	8.50		0.09	18.17
		(0.47)	(0.31)		(0.99)	(0.23)
s_{j0} Professional services		-17.26	10.66		-39.13	-31.33
		(0.10)	(0.26)		(0.04)	(0.14)
s_{j0} Manufacturing		39.63	52.73		45.56	120.75
		(0.04)	(0.01)		(0.10)	(0.00)
s_{j0} Accomm., trade, transport.		4.75	14.07		-7.05	37.77
		(0.72)	(0.32)		(0.74)	(0.18)
Constant	0.37	2.31	-2.61	0.91	4.52	-12.46
	(0.07)	(0.36)	(0.30)	(0.00)	(0.19)	(0.03)
Observations	188	188	188	99	99	99
Pseudo R^2	0.137	0.392	0.494	0.280	0.578	0.831

Columns (1)–(3) report probit estimates of $\mathbb{1}_j^{back(5,20)}$ on $\mathbb{1}_j^{cone}$ and the listed controls. Columns (4)–(6) use the alternative indicator $\mathbb{1}_j^{back(0,30)}$. Robust p-values in parentheses.

up, or all the agglomeration elasticity down, by the same level. In particular, when raising a city's γ by a shift parameter s, we recalibrate the parameter $\bar{\tau}$ so that the city's initial commuting population matches its 2019 levels as described above, and recompute its position with respect to its cone. When reducing a city's δ by a shift parameter s, we recompute the values of A, z/A and $\bar{\tau}$ to match commuting population and remote work premium, as described above, and again recompute its position with respect to its cone. After shifting one of these two elasticities by the same amount for all cities, we re-estimate the relation between the in-cone indicator and trip shortfalls, rent flattening, and city size, as in the main text.

We report the results of these exercises in Tables D4 and D5. The first table contains robustness to increasing γ progressively by 0.01, up to 0.04: this exercise shifts the mean of γ up to 0.0667 in the most conservative case. The second table contains robustness to reducing δ progressively by 0.01, up to 0.04: this exercise shifts the mean of δ down to 0.0327 in the most conservative case.[78] In all cases, the in-cone indicator is significantly associated to our outcomes of interest; naturally, as γ grows, or δ shrinks, the fraction of cities in the cone reduces, and so does the strength of these associations.

D5. Robustness to Alternative Values of Amenity and Congestion Externalities

Our baseline calibration sets $\xi - \theta = 0$, consistent with the central values for residential amenity and congestion elasticities reported in the literature. Existing estimates point to $\xi \approx 0.15$ (Glaeser, Kolko and Saiz, 2001; Ahlfeldt et al., 2015; Redding, 2022) and to values of θ in the range $[0.13, 0.17]$ (Couture, Duranton and Turner, 2018; Akbar et al., 2023; Almagro et al., 2024), implying a small difference between the two parameters. In this Appendix, we assess the sensitivity of our main results to alternative values of $\xi - \theta$ within this empirically relevant range. Specifically, we let $\xi - \theta$ vary between -0.03 and $+0.03$. For each value, we recalibrate city-level fundamentals and recompute the cone indicator, its relationship with trip shortfalls, gradient flattening, and city size, and the resulting welfare effects.

Table D6 reports our first set of robustness results. For each value of $\xi - \theta$ (column 1), we estimate probit regressions of the cone indicator on three outcomes: the trip shortfall through July 2022 (columns 2–3), the flattening of the distance gradient through November 2024 (columns 4–5), and city size (columns 6–7). Across all specifications, the estimated slope remains statistically significant and close in magnitude to our preferred baseline, which is reproduced in the central row corresponding to $\xi - \theta = 0$. As expected, increasing $\xi - \theta$ raises the fraction of cities in the cone, reflecting stronger net agglomeration forces.

Table D7 extends Table 3 in the main text by reporting the distribution of welfare changes implied by alternative values of $\xi - \theta$. Relative to our baseline

[78] The number of cities in the sample drops slightly as we reduce the mean since the downward shift pushes some of them into a negative δ.

Table D4— Robustness with Respect to Transportation Costs Elasticities

		(1)	(2)	(3)	(4)	(5)
	Shift s	0	+0.01	+0.02	+0.03	+0.04
Statistics on γ	p10	0.002	0.012	0.022	0.032	0.042
	p50	0.014	0.024	0.034	0.044	0.054
	p90	0.044	0.054	0.064	0.074	0.084
	Mean	0.027	0.037	0.047	0.057	0.067
Trip Shortfall	β	-3.80	-3.67	-3.45	-2.98	-2.93
	s.e.	(0.61)	(0.59)	(0.57)	(0.55)	(0.58)
Flattening at 11/24	β	3.87	2.86	3.43	2.93	3.97
	s.e.	(1.41)	(1.33)	(1.32)	(1.37)	(1.64)
City Size	β	0.80	0.64	0.60	0.40	0.37
	s.e.	(0.14)	(0.13)	(0.10)	(0.08)	(0.07)
Multiplicity	Number of Cities	278	278	278	278	278
	Fraction in Cone	0.76	0.69	0.55	0.43	0.29

This table reports the coefficients (β), standard errors (s.e.) of the regressions of the in-cone indicator $\mathbb{1}_j^{cone}$ on the trip shortfall, the rent flattening, and total log employment in a city, for alternative calibrations of the distribution of γ with a mean shifted by s.

mean welfare loss of 2.3%, the mean loss ranges from 1.6% to 3.3% as the balance between agglomeration and congestion externalities shifts by six percentage points. When amenity agglomeration externalities become more important, the welfare loss from coordinating away from downtown correspondingly increases. Importantly, even when congestion externalities exceed amenity agglomeration externalities by three percentage points—so that a remote equilibrium reduces net congestion disutility—the implied welfare losses remain positive.

Table D5— Robustness with Respect to Agglomeration Externalities

		(1)	(2)	(3)	(4)	(5)
	Shift s	0	-0.01	-0.02	-0.03	-0.04
Statistics on δ	p10	0.047	0.038	0.029	0.019	0.010
	p50	0.071	0.061	0.051	0.041	0.031
	p90	0.092	0.082	0.072	0.062	0.052
	Mean	0.071	0.061	0.052	0.042	0.033
Trip Shortfall	β	-3.80	-3.96	-3.38	-2.94	-2.33
	s.e.	(0.61)	(0.60)	(0.58)	(0.56)	(0.63)
Flattening at 11/24	β	3.87	3.00	3.00	3.56	2.13
	s.e.	(1.41)	(1.33)	(1.33)	(1.53)	(1.75)
City Size	β	0.80	0.66	0.51	0.37	0.25
	s.e.	(0.14)	(0.13)	(0.10)	(0.07)	(0.07)
Multiplicity	Number of Cities	278	277	274	272	266
	Fraction in Cone	0.76	0.67	0.51	0.33	0.19

This table reports the coefficients (β), standard errors (s.e.) of the regressions of the in-cone indicator $\mathbb{1}_j^{cone}$ on the trip shortfall, the rent flattening, and total log employment in a city, for alternative calibrations of the distribution of δ with a mean shifted by s.

Table D6— Robustness of the Relationship Between Multiplicity and City Outcomes with Respect to Net Amenity Externalities $\xi - \theta$

	Trip Shortfall		Flattening 11/24		City Size		Number	Fraction
$\xi - \theta$	β	s.e.	β	s.e.	β	s.e.	of Cities	in Cone
-0.03	-3.00	0.56	3.56	1.52	0.38	0.07	278	0.32
-0.02	-3.42	0.57	3.00	1.33	0.51	0.09	278	0.50
-0.01	-4.00	0.60	2.94	1.33	0.66	0.13	278	0.67
0.00	-3.80	0.61	3.87	1.41	0.80	0.14	278	0.76
0.01	-3.01	0.64	3.09	1.51	0.70	0.17	278	0.87
0.02	-2.83	0.70	2.89	1.61	0.77	0.22	278	0.90
0.03	-2.28	0.71	4.23	1.66	0.81	0.28	278	0.92

This table reports the coefficients (β), standard errors (s.e.) of the regressions of the in-cone indicator $\mathbb{1}_j^{cone}$ on the trip shortfall, the rent flattening, and total log employment in a city, for alternative calibrations of the difference between amenity and congestion externalities of $\xi - \theta$.

Table D7— Robustness of the Welfare Effects with Respect to Net Amenity Externalities $\xi - \theta$

$\xi - \theta$	Min	p10	p25	p50	p75	p90	Max	Mean	N
-0.03	0.9	1.1	1.3	1.5	1.8	2.1	2.8	1.6	87
-0.02	0.9	1.1	1.4	1.7	2.1	2.3	3.2	1.7	136
-0.01	0.8	1.4	1.6	1.9	2.3	2.7	3.6	2.0	183
0.00	1.2	1.6	1.9	2.2	2.6	3.1	4.0	2.3	208
0.01	1.1	1.8	2.1	2.6	3.0	3.3	4.4	2.6	238
0.02	1.5	2.1	2.5	2.9	3.4	3.8	4.9	2.9	247
0.03	1.6	2.5	2.9	3.3	3.8	4.1	5.3	3.3	254

This table reports the distribution of Welfare losses across CBSAs from switching to the remote equilibrium for alternative calibrations of the difference between amenity and congestion externalities, $\xi - \theta$.

Theory

E1. Rewriting the Average Value of a Worker

We start by deriving a relationship that will be useful in the derivations: the expression of the option value $\Omega\left(\bar{\varepsilon}_t^{\ell}\right)$ of a given labor delivery mode ℓ as a function of the transition probability $\lambda_t^{\ell\ell}$.

By adding and subtracting the continuation value in the current labor delivery mode, the equation for U_ℓ can be rewritten as

$$\begin{aligned}
&U_\ell\left(\{\varepsilon_{ct}, \varepsilon_{mt}\}; L_{ct}, \omega^t\right)\\
&= u_\ell\left(L_{ct}; A_t, T_t, z_t\right) + \max_{\ell'}\left\{\beta V_{\ell'}\left(L_{ct+1}; \omega^{t+1}\right) - F_{\ell\ell' t} + \varepsilon_{\ell' t}\right\}\\
&= u_\ell\left(L_{ct}; A_t, T_t, z_t\right) + \beta V_\ell\left(L_{ct+1}; \omega^{t+1}\right) - F_{\ell\ell,t} +\\
&\max_{\ell'}\left\{\beta\left[V_{\ell'}\left(L_{ct+1}; \omega^{t+1}\right) - V_\ell\left(L_{ct+1}; \omega^{t+1}\right)\right] - \left(F_{\ell\ell' t} - F_{\ell\ell t}\right) + \varepsilon_{\ell' t}\right\}\\
&= u_\ell\left(L_{ct}; A_t, T_t, z_t\right) + \beta V_\ell\left(L_{ct+1}; \omega^{t+1}\right) - F_{\ell\ell t} + \max_{\ell'}\left\{\bar{\varepsilon}_t^{\ell\ell'} + \varepsilon_{\ell' t}\right\},
\end{aligned} \tag{E1}$$

where

$$\bar{\varepsilon}_t^{\ell\ell'}\left(L_{ct+1}\right) \equiv \beta\left[V_{\ell'}\left(L_{ct+1}; \omega^{t+1}\right) - V_\ell\left(L_{ct+1}; \omega^{t+1}\right)\right] - \left(F_{\ell\ell' t} - F_{\ell\ell t}\right) \tag{E2}$$

is the average value of switching from ℓ to ℓ'. Taking the expectation of U_ℓ with respect to ε_t,

$$\begin{aligned} V_\ell\left(L_{ct}; z_t\right) &= u_\ell\left(L_{ct}; A_t, T_t, z_t\right) + \beta V_\ell\left(L_{ct+1}; \omega^{t+1}\right) - F_{\ell\ell t} + E_t \max_{\ell'}\left\{\bar{\varepsilon}_t^{\ell\ell'} + \varepsilon_{\ell' t}\right\} \\ &\equiv u_\ell\left(L_{ct}; A_t, T_t, z_t\right) + \beta V_\ell\left(L_{ct+1}; \omega^{t+1}\right) - F_{\ell\ell t} + \Omega\left(\bar{\varepsilon}_t^{\ell}\right), \end{aligned} \tag{E3}$$

where $\bar{\varepsilon}_t^\ell \equiv \left\{\bar{\varepsilon}_t^{\ell c}, \bar{\varepsilon}_t^{\ell m}\right\}$. $\Omega\left(\bar{\varepsilon}_t^\ell\right)$ is the value of the option of being able to move out of ℓ if circumstances require, or the "option value" of ℓ.

Under our distributional assumption for ε_t we can write the share of workers in ℓ state transitioning to ℓ' as

$$\lambda_t^{\ell\ell'} = \frac{\exp\left(\bar{\varepsilon}_t^{\ell\ell'}\left(L_{ct+1}\right)/s\right)}{\exp\left(\bar{\varepsilon}_t^{\ell c}\left(L_{ct+1}\right)/s\right) + \exp\left(\bar{\varepsilon}_t^{\ell m}\left(L_{ct+1}\right)/s\right)}. \tag{E4}$$

We can use this expression to obtain the average expected switching values. In fact,

$$\frac{\lambda_t^{\ell\ell'}}{\lambda_t^{\ell\ell}} = \exp\left(\bar{\varepsilon}_t^{\ell\ell'}/s - \bar{\varepsilon}_t^{\ell\ell}/s\right) \Longrightarrow s\ln\frac{\lambda_t^{\ell\ell'}}{\lambda_t^{\ell\ell}} = \bar{\varepsilon}_t^{\ell\ell'} - \bar{\varepsilon}_t^{\ell\ell} \Longrightarrow \bar{\varepsilon}_t^{\ell\ell'} = s\ln\frac{\lambda_t^{\ell\ell'}}{\lambda_t^{\ell\ell}}. \tag{E5}$$

To obtain the option value of being in ℓ, recall that

$$\Omega\left(\bar{\varepsilon}_t^\ell\right) = s\log\left[\exp\left(\bar{\varepsilon}_t^{\ell c}/s\right) + \exp\left(\bar{\varepsilon}_t^{\ell m}/s\right)\right]. \tag{E6}$$

Since

$$\begin{aligned} \log\lambda_t^{\ell\ell} &= \bar{\varepsilon}_t^{\ell\ell}/s - \log\left[\exp\left(\bar{\varepsilon}_t^{\ell c}/s\right) + \exp\left(\bar{\varepsilon}_t^{\ell m}/s\right)\right] \\ &= -\log\left[\exp\left(\bar{\varepsilon}_t^{\ell c}/s\right) + \exp\left(\bar{\varepsilon}_t^{\ell m}/s\right)\right], \end{aligned} \tag{E7}$$

we have that

$$\Omega\left(\bar{\varepsilon}_t^\ell\right) = -s\ln\lambda_t^{\ell\ell}. \tag{E8}$$

E2. Existence of a Stationary Equilibrium

To prove that a stationary equilibrium always exists, we proceed in steps. First, Lemma E.1 shows that the characterization of a stationary equilibrium can be reduced to the zeros of one equation in one unknown, the share of commuters who remain commuters next period, λ^{cc}. Lemma E.2 establishes a useful property of one component in such equation. Finally, Proposition E.1 shows that this equation always has at least one solution.

In a stationary equilibrium, the system is characterized by 11 unknowns, u_c, u_m, L_c, V_c, V_m, λ^{cc}, λ^{mm}, $\bar{\varepsilon}^{cc}$, $\bar{\varepsilon}^{cm}$, $\bar{\varepsilon}^{mc}$, $\bar{\varepsilon}^{mm}$, and 11 equations: the flow utility values that account for the relation between $\bar{d}, \bar{D}$, and L_c, L:

$$u_c = \log r_a^{\alpha-1} \bar{\alpha} \frac{B(\tilde{L}_c)}{\tau(\tilde{L}_c)} \frac{w_c}{\left(\tilde{L}_c/\pi\right)^{\gamma/2}}, \tag{E9}$$

$$u_m = \log r_a^{\alpha-1} \bar{\alpha} T^{\mu} \left(\frac{B(\tilde{L}_c)}{\tau(\tilde{L}_c)}\right)^{1-\mu} \frac{w_m}{(L/\pi)^{\gamma(1-\mu)/2}}, \tag{E10}$$

the law of motion for L_c,

$$L_c = L_c \lambda^{cc} + (L - L_c)(1 - \lambda^{mm}), \tag{E11}$$

the average value functions by mode,

$$V_c = u_c + \beta V_c - s \ln \lambda^{cc} - F_{cc}, \tag{E12}$$

$$V_m = u_m + \beta V_m - s \ln \lambda^{mm} - F_{mm}, \tag{E13}$$

the transition probabilities,

$$\lambda^{cc} = \frac{\exp\left(\bar{\varepsilon}^{cc}/s\right)}{\exp\left(\bar{\varepsilon}^{cc}/s\right) + \exp\left(\bar{\varepsilon}^{cm}/s\right)}, \tag{E14}$$

$$\lambda^{mm} = \frac{\exp\left(\bar{\varepsilon}^{mm}/s\right)}{\exp\left(\bar{\varepsilon}^{mc}/s\right) + \exp\left(\bar{\varepsilon}^{mm}/s\right)}, \tag{E15}$$

and the expected values from switching

$$\bar{\varepsilon}^{cc} = 0, \tag{E16}$$

$$\bar{\varepsilon}^{mm} = 0, \tag{E17}$$

$$\bar{\varepsilon}^{cm} = \beta\left(V_m - V_c\right) - \left(F_{cm} - F_{cc}\right), \tag{E18}$$

$$\bar{\varepsilon}^{mc} = \beta\left(V_c - V_m\right) - \left(F_{mc} - F_{mm}\right). \tag{E19}$$

The following lemma shows that we can always reduce this system of equations to one equation in one unknown.

LEMMA E.1: *The characterization of the stationary equilibrium (E9)-(E19) can be reduced to one equation in one unknown,*

$$f\left(\lambda^{cc}\right) \equiv \frac{1}{1 + \exp\left(\bar{\varepsilon}^{cm}\left(\lambda^{cc}\right)/s\right)} - \lambda^{cc} = 0, \tag{E20}$$

where

$$\bar{\varepsilon}^{cm}\left(\lambda^{cc}\right) \equiv \frac{\beta}{1-\beta}\left[e_f\left(\lambda^{cc}\right) + e_o\left(\lambda^{cc}\right) - \left(F_{mm} - F_{cc}\right)\right] - F_{cm} + F_{cc}. \tag{E21}$$

is the value of switching from c to m; the function

$$e_o\left(\lambda^{cc}\right) \equiv s \ln \lambda^{cc} \frac{1-\phi\lambda^{cc}}{1-\lambda^{cc}}. \tag{E22}$$

is the difference in the option values of staying in m versus staying in c; and the function

$$e_f\left(\lambda^{cc}\right) \equiv u_m\left(L_c\left(\lambda^{cc}\right)\right) - u_c\left(L_c\left(\lambda^{cc}\right)\right). \tag{E23}$$

is the difference in the flow value of utilities, with

$$L_c\left(\lambda^{cc}\right) = \frac{\left(1-\phi\right)\lambda^{cc}}{\left(1-\lambda^{cc}\right)\left(1-\phi\lambda^{cc}\right)+\left(1-\phi\right)\lambda^{cc}} L \equiv \sigma\left(\lambda^{cc}\right) L, \tag{E24}$$

$$\phi \equiv \left[1-\exp\left(\frac{F_{cc}+F_{mm}-F_{cm}-F_{mc}}{s}\right)\right] \leq 1. \tag{E25}$$

PROOF:

Rearrange the value of switching from c to m (eq. (E18)) as $\beta\left(V_c - V_m\right) = -\bar{\varepsilon}^{cm} - F_{cm} + F_{cc}$. *Substitute this expression into eq. (E19) for* $\bar{\varepsilon}^{mc}$ *to obtain*

$$\begin{aligned} \bar{\varepsilon}^{mc} &= \beta\left(V_c - V_m\right) - F_{mc} + F_{mm} \\ &= -\bar{\varepsilon}^{cm} - F_{cm} + F_{cc} - F_{mc} + F_{mm}. \end{aligned} \tag{E26}$$

Rearranging the average value functions (E12)-(E13), one can write

$$\begin{aligned} V_c &= \left(u_c - s\ln\lambda^{cc} - F_{cc}\right)/\left(1-\beta\right), \\ V_m &= \left(u_m - s\ln\lambda^{mm} - F_{mm}\right)/\left(1-\beta\right). \end{aligned}$$

and use these expressions into (E18) to obtain

$$\begin{aligned} \bar{\varepsilon}^{cm} &= \beta\left(V_m - V_c\right) - F_{cm} + F_{cc} \\ &= \frac{\beta}{1-\beta}\left[u_m - s\ln\lambda^{mm} - F_{mm} - \left(u_c - s\ln\lambda^{cc} - F_{cc}\right)\right] - F_{cm} + F_{cc} \\ &= \frac{\beta}{1-\beta}\left[u_m - u_c + s\ln\lambda^{cc}/\lambda^{mm} - F_{mm} + F_{cc}\right] - F_{cm} + F_{cc}. \end{aligned} \tag{E27}$$

We aim to express the right-hand side of this expression only as a function of λ_{cc}. *To do that, note that eq. (E14) and (E16) together imply*

$$\lambda^{cc} = \frac{1}{1+\exp\left(\bar{\varepsilon}^{cm}/s\right)} \Longrightarrow \exp\left(-\bar{\varepsilon}^{cm}/s\right) = \lambda^{cc}/\left(1-\lambda^{cc}\right). \tag{E28}$$

Use (E17) and (E26) into eq. (E15) for λ^{mm},

$$\lambda^{mm} = \frac{1}{\exp\left(\left(-\bar{\varepsilon}^{cm} - F_{cm} + F_{cc} - F_{mc} + F_{mm}\right)/s\right) + 1}. \tag{E29}$$

We can then use eq. (E28) in (E29) to obtain,

$$\begin{aligned}
\lambda^{mm} &= \frac{1}{\exp\left(-\bar{\varepsilon}^{cm}/s\right)\exp\left(\left(-F_{cm} + F_{cc} - F_{mc} + F_{mm}\right)/s\right) + 1} \\
&= \frac{1}{\lambda^{cc}/\left(1-\lambda^{cc}\right)\exp\left(\left(-F_{cm} + F_{cc} - F_{mc} + F_{mm}\right)/s\right) + 1} \\
&= \frac{\left(1-\lambda^{cc}\right)}{\lambda^{cc}\exp\left(\left(-F_{cm} + F_{cc} - F_{mc} + F_{mm}\right)/s\right) + \left(1-\lambda^{cc}\right)} \\
&= \frac{\left(1-\lambda^{cc}\right)}{1-\lambda^{cc}\left[1-\exp\left(\left(-F_{cm} + F_{cc} - F_{mc} + F_{mm}\right)/s\right)\right]}.
\end{aligned}$$

This implies

$$\begin{aligned}
\lambda^{mm} &= \frac{1-\lambda^{cc}}{1-\phi\lambda^{cc}}, \\
\textit{with } \phi &\equiv \left[1-\exp\left(\frac{F_{cc}+F_{mm}-F_{cm}-F_{mc}}{s}\right)\right] \leq 1.
\end{aligned} \tag{E30}$$

Use (E30) to write the stationary equilibrium value of L_c *(from eq. (E11)) as*

$$\begin{aligned}
&L_c = L_c\lambda^{cc} + \left(L-L_c\right)\left(1-\lambda^{mm}\right), \\
&L_c - L_c\lambda^{cc} + L_c\left(1-\lambda^{mm}\right) = L\left(1-\lambda^{mm}\right), \\
&L_c\left(1-\lambda^{cc} + \frac{\left(1-\phi\right)\lambda^{cc}}{1-\phi\lambda^{cc}}\right) = L\frac{\left(1-\phi\right)\lambda^{cc}}{1-\phi\lambda^{cc}}, \\
&L_c\left(\left(1-\lambda^{cc}\right)\left(1-\phi\lambda^{cc}\right) + \left(1-\phi\right)\lambda^{cc}\right) = L\left(1-\phi\right)\lambda^{cc}, \\
&L_c = \frac{\left(1-\phi\right)\lambda^{cc}}{\left(1-\lambda^{cc}\right)\left(1-\phi\lambda^{cc}\right) + \left(1-\phi\right)\lambda^{cc}}L \equiv \sigma\left(\lambda^{cc}\right)L. \qquad \text{(E31)}
\end{aligned}$$

Making use of (E30) and (E31), we can then rewrite (E27) as

$$\begin{aligned}
\bar{\varepsilon}^{cm} &= \frac{\beta}{1-\beta}\left[u_m\left(L_c\left(\lambda^{cc}\right)\right) - u_c\left(L_c\left(\lambda^{cc}\right)\right) + s\ln\lambda^{cc}/\lambda^{mm} - F_{mm} + F_{cc}\right] \\
&\quad -F_{cm} + F_{cc} \\
&= \frac{\beta}{1-\beta}\left[u_m\left(L_c\left(\lambda^{cc}\right)\right) - u_c\left(L_c\left(\lambda^{cc}\right)\right) + s\ln\lambda^{cc}\frac{1-\phi\lambda^{cc}}{1-\lambda^{cc}} - F_{mm} + F_{cc}\right] \\
&\quad -F_{cm} + F_{cc}. \qquad \text{(E32)}
\end{aligned}$$

LEMMA E.2: *The function* $e_o(\lambda)$ *in eq. (E22) increases from -∞ to +∞ as*

$\lambda \in (0,1)$.

PROOF:

The derivative of $e_o(\lambda) = s \ln\left(\lambda \frac{1-\phi\lambda}{1-\lambda}\right)$ *with respect to* λ *is*

$$\begin{aligned} s\left[\frac{1}{\lambda} - \frac{\phi}{1-\phi\lambda} + \frac{1}{1-\lambda}\right] &= s\left[\frac{1}{\lambda} + \frac{-\phi(1-\lambda) + (1-\phi\lambda)}{(1-\phi\lambda)(1-\lambda)}\right] \\ &= s\left[\frac{1}{\lambda} + \frac{1-\phi}{(1-\phi\lambda)(1-\lambda)}\right] \\ &= s\left[\frac{(1-\phi\lambda)(1-\lambda) + (1-\phi)\lambda}{(1-\phi\lambda)(1-\lambda)\lambda}\right] \\ &= s\frac{1-2\phi\lambda+\phi\lambda^2}{(1-\phi\lambda)(1-\lambda)\lambda}. \end{aligned} \tag{E33}$$

This function is always positive because the numerator is a convex parabola in λ*, and the minimum is reached at* $\lambda = 2\phi/(2\phi) = 1$*, where the numerator is equal to* $1 - 2\phi + \phi = 1 - \phi > 0$*. Note that* $\lim_{\lambda\to 0} \lambda \frac{1-\phi\lambda}{1-\lambda} = 0$*, and* $\lim_{\lambda\to 1} \lambda \frac{1-\phi\lambda}{1-\lambda} = +\infty$*, so that the function* $\lim_{\lambda\to 0} e_o(\lambda) = -\infty$ *and* $\lim_{\lambda\to 1} e_o(\lambda) = +\infty$*.*

Recall that $e_o(\lambda^{cc})$ is the option value of a remote worker minus the option value of a commuting worker. Lemma E.2 says that: 1) as the equilibrium number of commuters grows, the relative option value of being a remote worker grows; 2) as we approach a purely remote equilibrium ($\lambda^{cc} \to 0$), switching from c to m implies a loss in option values that approaches $+\infty$; as we approach a purely commuting equilibrium ($\lambda^{cc} \to 1$), switching from c to m implies a gain in the option values that approaches $+\infty$. In this sense, $e_o(\lambda)$ acts as a dynamic congestion force, always increasing the value of becoming a remote worker when there are more commuters.

PROPOSITION E.1: *There exists at least one stationary equilibrium* $L_{c,ss}$*.*

PROOF:

Consider the equation $f(\lambda) = 0$ *characterizing stationary equilibria from (E20).* $f(\lambda)$ *is continuous over* $\lambda \in (0,1)$*. We aim at showing that* $\lim_{\lambda\to 0} f(\lambda) > 0$ *and* $\lim_{\lambda\to 1} f(\lambda) < 0$*, so that* $f(\lambda) = 0$ *(and there exists at least one stationary equilibrium) for some* $\lambda \in (0,1)$ *by the intermediate value theorem.*

From (E21), consider the behavior of

$$\bar{\varepsilon}^{cm}(\lambda^{cc}) \equiv \frac{\beta}{1-\beta}\left[e_f(\lambda^{cc}) + e_o(\lambda^{cc}) - (F_{mm} - F_{cc})\right] - F_{cm} + F_{cc}.$$

The difference in the option values $e_o(\lambda^{cc})$ *behaves as* $\lim_{\lambda\longrightarrow 0} e_o(\lambda^{cc}) = -\infty$ *and* $\lim_{\lambda\longrightarrow 1} e_o(\lambda^{cc}) = +\infty$ *from Lemma E.2.*

Using the functional forms in Assumption 1 in the indirect utility functions

(E9) and (E10), we can write

$$
\begin{aligned}
e_f\left(\lambda^{cc}\right) &\equiv u_m\left(\lambda^{cc}\right)-u_c\left(\lambda^{cc}\right) \\
&= \mu\left(\log T-\log \frac{\bar{B}}{\bar{\tau}}-(\delta+\xi-\theta-\gamma/2)\log L-\frac{\gamma}{2}\log \pi+\log \frac{Z}{A}\right) \\
&- \mu^2(\delta+\xi-\theta-\gamma/(2\mu))\log \sigma(\lambda^{cc}). \qquad \text{(E34)}
\end{aligned}
$$

As $\lambda \to 1$, this difference approaches a finite number. When $\lambda \to 0$, this difference approaches $-\infty$ if $\delta+\xi<\theta+\gamma/(2\mu)$, and $+\infty$ if $\delta+\xi>\theta+\gamma/(2\mu)$,

These observations imply that $\lim_{\lambda^{cc}\longrightarrow 1}\bar{\varepsilon}^{cm}\left(\lambda^{cc}\right)=+\infty$. Moreover, Assumption 2 guarantees that $\lim_{\lambda^{cc}\to 0}\bar{\varepsilon}^{cm}\left(\lambda^{cc}\right)=-\infty$. It follows that $\lim_{\lambda\longrightarrow 0} f\left(\lambda\right)=1$ and $\lim_{\lambda\longrightarrow 1} f\left(\lambda\right)=-1$. Hence, eq. (E20) always has at least one solution.

E3. Sufficient Conditions for a Unique Stationary Equilibrium

In this subsection, we show sufficient conditions for the stationary equilibrium to be unique and how this stationary equilibrium changes with some of the model's parameters. The proof of this result in Proposition E.2 is based on establishing conditions for the monotonicity of $f\left(\lambda\right)$ in eq. (E20). Lemma E.3 below preliminarily establishes the monotonicity of L_c in the share of commuters that remain commuters, λ^{cc}.

LEMMA E.3: *The function $\sigma\left(\lambda\right)$ in eq. (E24) is monotonically increasing from 0 to 1 when $\lambda\in[0,1]$.*
PROOF:

Computing the derivative,

$$
\begin{aligned}
&\text{(E35)}\quad \sigma'\left(\lambda\right) \\
&= (1-\phi)\frac{(1-\lambda)(1-\phi\lambda)+(1-\phi)\lambda-\lambda(1-\phi)+\lambda(1-\phi\lambda)+\phi\lambda(1-\lambda)}{\left[(1-\lambda)(1-\phi\lambda)+(1-\phi)\lambda\right]^2} \\
&= \frac{(1-\phi)\left(1-\lambda^2\phi\right)}{\left[(1-\lambda)(1-\phi\lambda)+(1-\phi)\lambda\right]^2}.
\end{aligned}
$$

which is always positive.

PROPOSITION E.2: *Assume $\delta+\xi<\theta+\gamma/(2\mu)$. Then there is a unique stationary equilibrium value of the commuting population $L_{c,ss}$. This stationary equilibrium value monotonically increases with A and $\bar{B}$, and monotonically falls with Z, T and $\bar{\tau}$.*
PROOF:

Recall eq. (E34), stating that

$$\begin{aligned} e_f(\lambda^{cc}) &\equiv u_m(\lambda^{cc}) - u_c(\lambda^{cc}) \\ &= \mu\left(\log T - \log\frac{\bar{B}}{\bar{\tau}} - (\delta+\xi-\theta-\gamma/2)\log L - \frac{\gamma}{2}\log\pi + \log\frac{Z}{A}\right) \\ &\quad -\mu^2(\delta+\xi-\theta-\gamma/(2\mu))\log\sigma(\lambda^{cc}). \end{aligned} \tag{E36}$$

When $\delta+\xi<\theta+\gamma/(2\mu)$, this difference is increasing in λ^{cc} from Lemma E.3 and it follows that $\bar{\varepsilon}^{cm}(\lambda^{cc})$ in eq. (E21) is monotonically increasing in λ^{cc}, mapping the interval $(0,1)$ into $(-\infty,+\infty)$. Hence, the function $f(\lambda^{cc})$ in (E20) is monotonically decreasing in λ^{cc} and maps the interval $(0,1)$ into itself. Equation (E20) has then a unique solution for λ^{cc}, which maps into a unique value for $L_c=\sigma(\lambda_{cc})L$.

Note also that A, Z, T, $\bar{B}$ and $\bar{\tau}$ only affect the value of $\bar{\varepsilon}^{cm}(\lambda^{cc})$ in (E21) through u_m-u_c; in particular, $\bar{\varepsilon}^{cm}\left(\lambda^{cc};A,\bar{B},Z,T,\tau\right)$ grows with Z, T, and $\bar{\tau}$ while falling with A and $\bar{B}$ for any given λ^{cc}. Hence, $f(\lambda^{cc})$ in (E20) shifts uniformly down with Z, T, $\bar{\tau}$ and up with A and $\bar{B}$ in its domain $\lambda^{cc}\in(0,1)$. This implies that the solution to (E20) and $L_c=\sigma(\lambda_{cc})L$ also moves in the same direction with these parameters.

E4. Necessary and Sufficient Conditions for Multiple Stationary Equilibria

In Proposition E.1 we have shown that there is always at least one stationary equilibrium because $f(\lambda)$ starts positive, ends negative, and is continuous. In Proposition E.3, we show that for any level of (sufficiently strong) agglomeration forces, there is a range of values z/A where the economy exhibits multiple stationary equilibria. Proposition E.3 makes use of the following four lemmas. Lemma E.4 characterizes necessary and sufficient conditions for multiplicity in terms of the function $f(\lambda)$ in eq. (E20): multiple stationary equilibria exist if $f'(\lambda)>0$ at a stationary equilibrium, which can be restated into a condition on $\partial\bar{\varepsilon}^{cm}(\lambda)/\partial\lambda$, the slope of the expected value of switching from c to m with respect to λ. Lemma E.5 considers such slope and it shows that, at a candidate $\hat{\lambda}$, one can always make $\partial\bar{\varepsilon}^{cm}(\lambda)/\partial\lambda$ arbitrarily negative by choosing an externality term δ above a threshold δ^*. This fact is useful because the slope of $\bar{\varepsilon}^{cm}$ can be used to control the slope of $f(\lambda)$. Lemma E.6 characterizes some properties of this threshold δ^*. Lemma E.7 shows that for any level of δ, we can always find a ratio of the exogenous remote work to in-office productivity, z/A, that makes an arbitrary $\hat{\lambda}$ a stationary equilibrium. Proposition E.3 uses these lemmas to show that for a high enough δ, we can always find a z/A that makes some $\hat{\lambda}$ a stationary equilibrium where $f'(\hat{\lambda})>0$.

LEMMA E.4: *Consider an economy Ω. The economy has multiple stationary* *equilibria if and only if there exist one $\hat{\lambda}$ such that*

$$f(\hat{\lambda}) = 0, \tag{E37}$$

$$\left.\frac{\partial f(\lambda)}{\partial \lambda}\right|_{\lambda=\hat{\lambda}} > 0. \tag{E38}$$

Moreover, if such $\hat{\lambda}$ exists, then it holds true that:

$$\left.\frac{\partial \bar{\varepsilon}^{cm}(\lambda)}{\partial \lambda}\right|_{\lambda=\hat{\lambda}} < -\frac{s}{\hat{\lambda}\left(1-\hat{\lambda}\right)}. \tag{E39}$$

PROOF:

From Lemma E.1, a stationary equilibrium is characterized by $f(\hat{\lambda}) = 0$. Suppose

$$\left.\frac{\partial f(\lambda)}{\partial \lambda}\right|_{\lambda=\hat{\lambda}} > 0. \tag{E40}$$

Since $f(\lambda)$ always starts at $+1$ and ends at -1, $f(\lambda)$ will cross zero at least two more times, one to the left and one to the right of $\hat{\lambda}$. In other words, the equation $f(\lambda)$ has more than one solution, that is, the economy exhibits multiple stationary equilibria.

Note that, using the definition of f in eq. (E20),

$$\begin{aligned}
\frac{\partial f(\lambda)}{\partial \lambda} &= -\frac{\exp\left(\bar{\varepsilon}^{cm}(\lambda)/s\right)}{\left(1+\exp\left(\bar{\varepsilon}^{cm}(\lambda)/s\right)\right)^2}\frac{1}{s}\frac{\partial \bar{\varepsilon}^{cm}(\lambda)}{\partial \lambda} - 1 > 0, \\
\frac{\partial \bar{\varepsilon}^{cm}(\lambda)}{\partial \lambda} &< -s\frac{\left(1+\exp\left(\bar{\varepsilon}^{cm}(\lambda)/s\right)\right)^2}{\exp\left(\bar{\varepsilon}^{cm}(\lambda)/s\right)}. \qquad \text{(E41)}
\end{aligned}$$

Additionally, using the same definition, at a stationary equilibrium it holds true that

$$f\left(\hat{\lambda}\right) \equiv \frac{1}{1+\exp\left(\bar{\varepsilon}^{cm}\left(\hat{\lambda}\right)/s\right)} - \hat{\lambda} = 0 \implies \tag{E42}$$

$$1+\exp\left(\bar{\varepsilon}^{cm}\left(\hat{\lambda}\right)/s\right) = 1/\hat{\lambda} \tag{E43}$$

$$\exp\left(\bar{\varepsilon}^{cm}\left(\hat{\lambda}\right)/s\right) = \left(1-\hat{\lambda}\right)/\hat{\lambda}. \tag{E44}$$

Hence, using equations (E43) and (E44) in eq. (E41), we have

(E45) $$\left.\frac{\partial \bar{\varepsilon}^{cm}(\lambda)}{\partial \lambda}\right|_{\lambda=\hat{\lambda}} < -\frac{s}{\hat{\lambda}\left(1-\hat{\lambda}\right)}.$$

LEMMA E.5: $\forall \hat{\lambda} \in (0,1), K>0, \exists\ \delta^*\left(\hat{\lambda}, K\right) > 0$:

$$\left.\frac{\partial \bar{\varepsilon}^{cm}(\lambda, \delta)}{\partial \lambda}\right|_{\lambda=\hat{\lambda}} < -K \iff \delta + \xi > \delta^*\left(\hat{\lambda}, K\right).$$

PROOF:

Consider the expression for $\bar{\varepsilon}^{cm}(\lambda)$ from (E21):

$$\bar{\varepsilon}^{cm}(\lambda) \equiv \frac{\beta}{1-\beta}\left[e_f(\lambda) + e_o(\lambda) - (F_{mm} - F_{cc})\right] - F_{cm} + F_{cc}.$$

The first derivative of this function with respect to λ is

$$\frac{\partial \bar{\varepsilon}^{cm}(\lambda)}{\partial \lambda} = \frac{\beta}{1-\beta}\left[e_f'(\lambda) + e_o'(\lambda)\right].$$

We know that $e_o'(\lambda) > 0$. From eq. (E34), we can easily compute e_f' so that our condition becomes

$$\left.\frac{\partial \bar{\varepsilon}^{cm}(\lambda)}{\partial \lambda}\right|_{\lambda=\hat{\lambda}} < -K \iff \left.e_f'(\lambda)\right|_{\lambda=\hat{\lambda}} < -\left.e_o'(\lambda)\right|_{\lambda=\hat{\lambda}} - \frac{1-\beta}{\beta}K \iff$$

(E46) $$\delta + \xi > \frac{1}{\mu^2}\left[\frac{\sigma(\lambda)}{\sigma'(\lambda)}\left(e_o'(\lambda) + \frac{1-\beta}{\beta}K\right)\right]_{\lambda=\hat{\lambda}} + \frac{\gamma}{2\mu} + \theta \equiv \delta^*\left(\hat{\lambda}, K\right).$$

LEMMA E.6: *Define*

(E47) $$\hat{\delta}(\lambda) \equiv \frac{1}{\mu^2}\left[\frac{\sigma(\lambda)}{\sigma'(\lambda)}\left(e_o'(\lambda) + \frac{1-\beta}{\beta}\frac{s}{\lambda(1-\lambda)}\right)\right] + \frac{\gamma}{2\mu} + \theta$$

as the threshold function $\delta^*(\lambda, K)$ for $K = \frac{s}{\lambda(1-\lambda)}$ in eq. (E46). Then, $\hat{\delta}(\lambda)$ reaches a minimum value of $\eta_{min} > \gamma/(2\mu) + \theta$. If $\phi \in (0,1)$, then $\hat{\delta}(\lambda)$ is convex in λ.

PROOF:

The function $\hat{\delta}(\lambda)$ is defined for $\hat{\lambda} \in (0,1)$, is continuous, and differentiable. Since the term in the square brackets is always positive, the function is also always positive. To further characterize $\hat{\delta}(\lambda)$, we study its limiting properties. To compute $\lim_{\hat{\delta}\to 0}\hat{\delta}(\hat{\lambda})$, we substitute the expressions for $\sigma(\lambda)$, $e_o'(\lambda)$ and $\sigma'(\lambda)$

from equations (E31), (E33), and (E36) and obtain

$$\lim_{\hat{\lambda}\to 0} \hat{\delta}(\hat{\lambda}) = \frac{s}{\mu^2\beta} + \frac{\gamma}{2\mu} + \theta \equiv \hat{\delta}_0 > 0.$$

Further, $\lim_{\hat{\delta}\to 1} \hat{\delta}(\hat{\lambda}) = +\infty$ since, from (E31) and (E36), $\lim_{\hat{\lambda}\to 1} \sigma\left(\hat{\lambda}\right) = 1$ and $\lim_{\hat{\lambda}\to 1} \sigma'\left(\hat{\lambda}\right) = 1$; and from Lemma E.2, $\lim_{\hat{\lambda}\to 1} e_o'\left(\hat{\lambda}\right) = +\infty$.
The first derivative of $\hat{\delta}(\lambda)$, after some algebra, is given by:

$$\begin{aligned} \hat{\delta}'(\lambda) = & - \frac{s}{\mu^2}\frac{4\phi(1-2\phi\lambda+\phi\lambda^2)}{(1-\phi\lambda^2)(1-\phi\lambda)} \\ & + \frac{s}{\mu^2}\frac{(1-2\phi\lambda+\phi\lambda^2)^2(1-3\phi^2\lambda^2-3\phi\lambda^2+4\phi^2\lambda^3+\phi)}{(1-\phi\lambda^2)^2(1-\phi\lambda)^2(1-\lambda)^2} \\ & + \frac{s}{\mu^2}\frac{1-\beta}{\beta}\frac{1+2\phi(2\lambda-2\lambda^2-1)-\phi^2\lambda^2(2-4\lambda+\lambda^2)}{(1-\phi\lambda^2)^2(1-\lambda)^2}. \end{aligned} \tag{E48}$$

Moreover,

$$\lim_{\hat{\lambda}\to 0} \hat{\delta}'(\lambda) < 0 \iff \phi > \frac{1}{2+\beta}. \tag{E49}$$

Hence, the function begins by sloping down if ϕ is large enough. Assume that $\hat{\delta}(\lambda)$ is convex, which we will prove right below. In this case, the (unique) minimum is achieved for a $\eta_{min} : \gamma/(2\mu) + \theta < \eta_{min} < \hat{\delta}_0$. If the function begins by sloping up, then $\eta_{min} \equiv \hat{\delta}_0$.

It remains to be proven that $\hat{\delta}(\lambda)$ is convex. To this end, it is thus sufficient to prove the convexity of the simplified function $\hat{\delta}(\lambda)_{simple}$, which takes out the constant term $\frac{\gamma}{\mu} + \theta$ and the positive multiplicative term $\frac{s}{\mu^2}$:

$$\begin{aligned} \hat{\delta}(\lambda)_{simple} &= \frac{\lambda\left(1-2\phi\lambda+\phi\lambda^2\right)}{1-\phi\lambda^2}\left[\frac{1-2\phi\lambda+\phi\lambda^2}{(1-\phi\lambda)(1-\lambda)\lambda} + \frac{1-\beta}{\beta}\frac{1}{\lambda(1-\lambda)}\right] \\ &= \frac{(1-2\phi\lambda+\phi\lambda^2)}{(1-\phi\lambda^2)(1-\lambda)}\left[\frac{1-2\phi\lambda+\phi\lambda^2}{(1-\phi\lambda)} + \frac{1-\beta}{\beta}\right] \\ &= \underbrace{\frac{(1-2\phi\lambda+\phi\lambda^2)^2}{(1-\phi\lambda^2)(1-\lambda)(1-\phi\lambda)}}_{\text{first term } \Xi_1} + \frac{1-\beta}{\beta}\underbrace{\frac{(1-2\phi\lambda+\phi\lambda^2)}{(1-\phi\lambda^2)(1-\lambda)}}_{\text{second term } \Xi_2}. \end{aligned}$$

We argue that each of the terms Ξ_1 and Ξ_2 is convex, which will prove the convexity of their sum.

The second derivative of Ξ_1 is:

$$\frac{d^2\Xi_1}{d\lambda^2} = \frac{2}{(1-\lambda)^3} + \phi^2\left(\frac{2}{(1-\phi\lambda)^3} - \frac{8x(\phi\lambda^2+3)}{(1-\phi\lambda^2)^3}\right).$$

We prove by contradiction that $\frac{d^2\Xi_1}{d\lambda^2}$ cannot be negative. Recall that the range of values for ϕ, λ is $\phi \in (0,1), \lambda \in (0,1)$. The first term of the expression $\frac{2}{(1-\lambda)^3}$, is always positive. If $\frac{d^2\Xi_1}{d\lambda^2}$ is negative at any value of λ, then the term $\left(\frac{2}{(1-\phi\lambda)^3} - \frac{8x(\phi\lambda^2+3)}{(1-\phi\lambda^2)^3}\right)$ must be negative. If this is so, then:

$$\begin{aligned}
\frac{d^2\Xi_1}{d\lambda^2} &= \frac{2}{(1-\lambda)^3} + \phi^2\left(\frac{2}{(1-\phi\lambda)^3} - \frac{8x(\phi\lambda^2+3)}{(1-\phi\lambda^2)^3}\right)\\
&\geq \frac{2}{(1-\lambda)^3} + \left(\frac{2}{(1-\phi\lambda)^3} - \frac{8x(\phi\lambda^2+3)}{(1-\phi\lambda^2)^3}\right)\\
&\geq \frac{2}{(1-\lambda)^3} + \left(\frac{2}{(1-\lambda)^3} - \frac{8x(\phi\lambda^2+3)}{(1-\phi\lambda^2)^3}\right)\\
&\geq \frac{2}{(1-\lambda)^3} + \left(\frac{2}{(1-\lambda)^3} - \frac{8x(\lambda^2+3)}{(1-\lambda^2)^3}\right)\\
&= \frac{4}{(1+\lambda)^3}\\
&\geq 0.
\end{aligned}$$

The fourth line follows because $\frac{d}{da} = -\frac{8x(\phi\lambda^2+3)}{(1-\phi\lambda^2)^3} = -\frac{16\lambda^3(5+\phi\lambda^2)}{(-1+\phi\lambda^2)^4} \leq 0$, so the term is most negative at $\phi = 1$. We have shown that $\frac{d^2\Xi_1}{d\lambda^2}$ is weakly positive. Thus, Ξ_1 is convex.

We now move on to the second term term. The second derivative of Ξ_2 is:

$$\frac{d^2\Xi_2}{d\lambda^2} = \frac{2(1-\phi\lambda^2)^3 - 4\phi^2\lambda(\phi\lambda^2+3)(1-\lambda)^3}{(1-\lambda)^3(1-\phi\lambda^2)^3}.$$

The denominator is obviously positive for $\phi \in (0,1), \lambda \in (0,1)$. We want to prove that the numerator is positive. Let us observe so the numerator, which we denote as $N_2(\lambda)$:

$$N_2(\lambda) = 2(1-\phi\lambda^2)^3 - 4\phi^2\lambda(\phi\lambda^2+3)(1-\lambda)^3.$$

By inspection, $N_2(\lambda)$ is decreasing in ϕ for $\phi \in (0,1)$. As such we only need to prove that $N_2(\lambda)$ is positive for $\phi = 1$ to show that $N_2(\lambda)$ is always positive.

$$\begin{aligned} N_2(\lambda) &= 2(1-\phi\lambda^2)^3 - 4\phi^2\lambda(\phi\lambda^2+3)(1-\lambda)^3 \\ &\geq 2(1-\lambda^2)^3 - 4\lambda(\lambda^2+3)(1-\lambda)^3 \\ &= 2(1-\lambda)^6 \\ &\geq 0 \text{ for } \lambda \in (0,1). \end{aligned}$$

We have shown that both the numerator and denominator of the second derivative of Ξ_2 are positive. As such, Ξ_2 is convex.

We have shown that Ξ_1 and Ξ_2 are convex, their sum $\hat{\delta}(\lambda)_{simple}$ is also convex, which in turn implies that our original function $\hat{\delta}(\lambda)$ is convex.

Lemma E.6 essentially states what is the minimum level of agglomeration forces that overcomes dynamic congestion forces if at a candidate λ we want $f'(\lambda) > 0$, a necessary condition for such λ to be an unstable equilibrium in a multiple equilibrium configuration.

LEMMA E.7: *For each $\hat{\lambda}$ and δ, there exists a ratio z/A such that $f\left(\hat{\lambda}; \delta, z/A\right) = 0$. This ratio is described by the function*

$$\begin{aligned} \rho(\hat{\lambda}) \quad &= \quad \exp\left\{\frac{1}{\mu}\left[\frac{1-\beta}{\beta}(F_{cm}-F_{cc}) + F_{mm} - F_{cc}\right]\right\} \times \pi^{\gamma/2}\frac{\bar{B}}{\bar{\tau}T} \\ &\times \quad \frac{\sigma(\hat{\lambda})^{\mu(\xi+\delta-\theta)-\gamma/2}}{(1-\phi\hat{\lambda})^{s/\mu}} \times \left(\frac{1-\hat{\lambda}}{\hat{\lambda}}\right)^{s/(\beta\mu)} \times L^{\xi+\delta-\theta-\gamma/2}. \end{aligned} \tag{E50}$$

PROOF:

Using the definition of $\bar{\varepsilon}^{cm}\left(\hat{\lambda}\right)$ from eq. (E21) in the condition for a stationary equilibrium from (E20),

$$\begin{aligned} \hat{\lambda} \quad &= \quad \frac{1}{1+\exp\left(\bar{\varepsilon}^{cm}\left(\hat{\lambda}\right)/s\right)} \implies s\ln\frac{1-\hat{\lambda}}{\hat{\lambda}} = \bar{\varepsilon}^{cm}\left(\hat{\lambda}\right), \\ s\ln\frac{1-\hat{\lambda}}{\hat{\lambda}} \quad &= \quad \frac{\beta}{1-\beta}\left[e_f\left(\hat{\lambda}\right) + e_o\left(\hat{\lambda}\right) - (F_{mm}-F_{cc})\right] - F_{cm} + F_{cc}. \end{aligned}$$

Using $e_f \equiv u_m - u_c$ and the result in eq. (E34), after some manipulation, we obtain

$$\begin{aligned} \frac{Z}{A} \quad &= \quad \exp\left\{\frac{1}{\mu}\left[\frac{1-\beta}{\beta}(F_{cm}-F_{cc}) + F_{mm} - F_{cc}\right]\right\} \times \pi^{\gamma/2}\frac{\bar{B}}{\bar{\tau}T} \\ &\times \quad \frac{\sigma(\hat{\lambda})^{\mu(\xi+\delta-\theta)-\gamma/2}}{(1-\phi\hat{\lambda})^{s/\mu}} \times \left(\frac{1-\hat{\lambda}}{\hat{\lambda}}\right)^{s/(\beta\mu)} \times L^{\xi+\delta-\theta-\gamma/2} \equiv \rho(\hat{\lambda}). \end{aligned} \tag{E51}$$

L_c is a function of $\hat{\lambda}$ and is bounded between 0 and L. One can verify using eq. (E22) that the sum in the curly bracket $-e_o\left(\hat{\lambda}\right) + \frac{1-\beta}{\beta} s \ln \frac{1-\hat{\lambda}}{\hat{\lambda}}$ goes to $+\infty$ for $\hat{\lambda} \to 0$ and to $-\infty$ for $\hat{\lambda} \to 1$. Hence, the image of $\hat{z}$ spans $(0, +\infty)$ as $\hat{\lambda}$ varies in $(0, 1)$.

PROPOSITION E.3: *Assume $F_{cc} = F_{mm} = 0$. Then, there exist a finite threshold $\eta_{min} > \theta + \gamma/(2\mu)$, and a set $\mathcal{Z} \subset \mathbb{R}_{++}$, such that:*

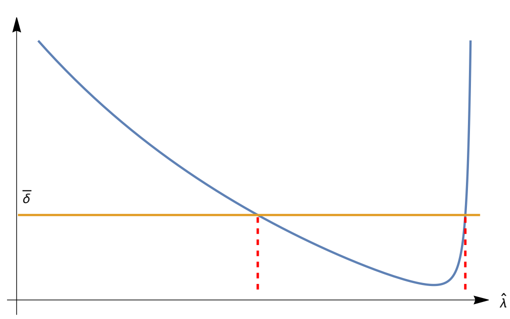
i) for $\delta + \xi > \eta_{min}$, $\mathcal{Z}$ is a non-empty interval (Z_{min}, Z_{max}) and there are multiple stationary equilibria if $z/A \in \mathcal{Z}$; further, Z_{min} and Z_{max} grow with L and $\bar{B}$ and fall with T and $\bar{\tau}$.

ii) there is a unique stationary equilibrium in all other cases.

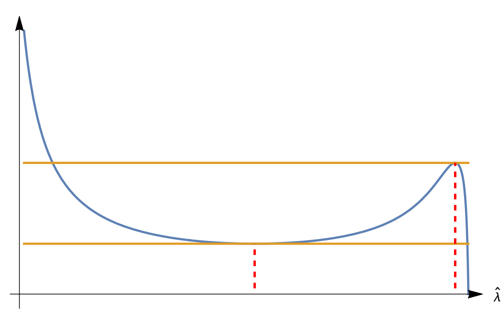
PROOF:

From Lemma E.4, we know that equations (E37) and (E39) characterize an economy with multiple stationary equilibria. We seek necessary and sufficient conditions for those equations to hold.

Consider the equation $\hat{\delta}(\hat{\lambda})$ in Lemma E.6. From Lemmas E.4 and E.5, $f'(\hat{\lambda}) > 0$ if and only if we pick a $\bar{\delta} : \bar{\delta} + \xi \geq \hat{\delta}(\hat{\lambda})$. Note that $\hat{\delta}(\hat{\lambda})$ is defined for $\hat{\lambda} \in (0, 1)$, is continuous, and differentiable. Also, from Lemma E.6, it is convex in $\hat{\lambda}$ and it reaches a global minimum value of $\hat{\delta}_{min}$. An example of this function is plotted in Panel (a) of Figure E1.



(a) The function $\hat{\delta}(\hat{\lambda})$



(b) The function $\rho(\hat{\lambda})$

Figure E1. Identifying the Range of Parameters for Multiple Stationary Equilibria

Our proof proceeds with a taxonomic discussion of the implications of different values of $\bar{\delta}$. This discussion can be organized around features of the set

$$\Lambda(\bar{\delta}) \equiv \{\hat{\lambda} : \hat{\delta}\left(\hat{\lambda}\right) \leq \bar{\delta} + \xi\}. \tag{E52}$$

Note that by construction, for each $\hat{\lambda} \in \Lambda(\bar{\delta})$, eq. (E39) is satisfied if $f(\hat{\lambda}, \bar{\delta}, z/A) = 0$.

Consider part *i)* of the proposition, and fix a $\bar{\delta} : \bar{\delta} + \xi > \hat{\delta}_{min} > \gamma/(2\mu) + \theta$. Then, the set $\Lambda(\bar{\delta})$ is non-empty; moreover, since $\hat{\delta}$ is convex from Lemma E.6, the set $\Lambda(\bar{\delta})$ is an interval of the form $\Lambda(\bar{\delta}) = (\lambda_{min}(\bar{\delta}), \lambda_{max}(\bar{\delta}))$. Consider then the set

$$\mathcal{Z}(\bar{\delta}) = \{z/A : z/A = \rho\left(\hat{\lambda}\right) \text{ for some } \hat{\lambda} \in \Lambda(\bar{\delta})\}. \tag{E53}$$

The set $\mathcal{Z}$ is the image of the set Λ in z/A (see Panel (b) of Figure E1) By construction, for any z/A in this set, there is some $\hat{\lambda}$ for which $\hat{\lambda}, z/A$ is a stationary equilibrium, that is, eq. (E37) holds. Since at the same $\hat{\lambda}$ eq. (E39) also holds, the economy exhibits multiple stationary equilibria. Note also that since $\rho(\lambda)$ is continuous and Λ is an interval, $\mathcal{Z}(\bar{\delta})$ is also an interval and can be written as

$$\begin{aligned} \mathcal{Z}(\bar{\delta}) &= \left(Z_{min}(\bar{\delta}), Z_{max}(\bar{\delta})\right), \\ Z_{min}(\bar{\delta}) &= min_{\lambda \in \Lambda(\bar{\delta})}\ \rho(\lambda), \\ Z_{max}(\bar{\delta}) &= max_{\lambda \in \Lambda(\bar{\delta})}\ \rho(\lambda). \end{aligned}$$

To show how $Z_{min}(\bar{\delta}), Z_{max}(\bar{\delta})$ vary with L, we consider the expression for $\rho(\lambda)$ in eq. (E51). If λ^- is the minimizer when the total city size is L, then λ^- is also the minimizer when the total city size is kL, for $k > 0$; additionally, $\rho(\lambda, kL) = k^{\bar{\delta}+\xi-\theta-\gamma/2}\rho(\lambda, L)$. Since $\bar{\delta} + \xi - \theta - \gamma/2 > 0$ is satisfied whenever $\mathcal{Z}$ is non-empty, then Z_{min} grows with L. A similar argument shows that Z_{max} also grows with L. Further, similar arguments can be made to show that these boundaries grow with $\bar{B}$ and fall with $\bar{\tau}$ and T.

To prove part *ii)*, note that if $\bar{\delta}+\xi > \hat{\delta}_{min}$ but $z/A \notin \mathcal{Z}(\bar{\delta})$, then by construction none of the $\hat{\lambda} \in \Lambda(\bar{\delta})$ are a stationary equilibrium where (E39) holds; however, since a stationary equilibrium always exists, the equilibrium must be unique. Similarly, if $\bar{\delta} + \xi < \hat{\delta}_{min}$, then the set $\Lambda(\bar{\delta})$ in eq. (E52) is empty; hence, there is no $\hat{\lambda}$ for which eq. (E37) holds at a candidate stationary equilibrium. Again, since a stationary equilibrium always exists, it must be unique.

Proposition III.3 then follows from the fact that L_c is monotone increasing in λ^{cc} by Lemma E.3.

E5. Relative Transition Shares

LEMMA E.8: *Suppose $F_{cc} = F_{mm} = 0$ and that $F_{cm} = F_{mc} = F$. Then*

$$y_{\ell\ell' t} \equiv \ln \frac{\lambda_{\ell\ell' t}\lambda^{\beta}_{\ell'\ell' t+1}}{\lambda_{\ell\ell t}\lambda^{\beta}_{\ell\ell' t+1}} = \frac{\beta}{s}\left(u_{\ell' t+1} - u_{\ell t+1}\right) - \frac{(1-\beta)F}{s}. \tag{E54}$$

PROOF:

Combining (18) and (17), the share of workers in state ℓ transitioning to ℓ' is given by

$$\lambda_{\ell\ell' t} = \frac{\exp\left(\beta V_{\ell' t+1} - F_{\ell\ell',t}\right)^{1/s}}{\exp\left(\beta V_{ct+1} - F_{\ell c,t}\right)^{1/s} + \exp\left(\beta V_{mt+1} - F_{\ell mt}\right)^{1/s}}, \tag{E55}$$

where we denote $V_{\ell t+1} = V_\ell\left(L_{ct+1}; \omega^{t+1}\right)$. We first obtain an expression of $\lambda_{\ell\ell' t}$ as a function of $V_{\ell t}$, $V_{\ell' t+1}$ and transition costs only by re-writing (E3) with time indices only,

$$V_{\ell t} = u_{\ell t} + \beta V_{\ell t+1} - F_{\ell\ell t} + \Omega_{\ell t}, \tag{E56}$$

and re-writing (E8) after replacing the transition share by its expression in (E55):

$$\Omega_{\ell t} = -s \ln\left(\frac{\exp\left(\beta V_{\ell t+1} - F_{\ell\ell t}\right)^{1/s}}{\exp\left(\beta V_{ct+1} - F_{\ell ct}\right)^{1/s} + \exp\left(\beta V_{mt+1} - F_{\ell mt}\right)^{1/s}}\right). \tag{E57}$$

Substituting for $\Omega_{\ell t}$ in (E56), we get the following expression for the value of labor delivery mode ℓ:

$$V_{\ell t} = u_{\ell t} + s \ln\left(\exp\left(\beta V_{ct+1} - F_{\ell ct}\right)^{1/s} + \exp\left(\beta V_{mt+1} - F_{\ell mt}\right)^{1/s}\right), \tag{E58}$$

or equivalently,

$$\exp\left(\beta V_{ct+1} - F_{\ell ct}\right)^{1/s} + \exp\left(\beta V_{mt+1} - F_{\ell mt}\right)^{1/s} = \exp\left(V_{\ell t} - u_{\ell t}\right)^{1/s}. \tag{E59}$$

Recognizing the left hand side as the denominator in (E55), we obtain the following expressions for the share of workers in state ℓ transitioning to ℓ' at t, in state ℓ transitioning to ℓ at t, in state ℓ' transitioning to ℓ' at $t+1$ and in state ℓ transitioning to ℓ' at $t+1$, respectively, as:

$$\lambda_{\ell\ell' t} = \exp\left(\beta V_{\ell' t+1} - F_{\ell\ell' t} + u_{\ell t} - V_{\ell t}\right)^{1/s}. \tag{E60}$$

$$\lambda_{\ell\ell t} = \exp\left(\beta V_{\ell t+1} - F_{\ell\ell t} + u_{\ell t} - V_{\ell t}\right)^{1/s}, \tag{E61}$$

$$\lambda_{\ell'\ell' t+1} = \exp\left(\beta V_{\ell' t+2} - F_{\ell'\ell' t+1} + u_{\ell' t+1} - V_{\ell' t+1}\right)^{1/s}, \tag{E62}$$

$$\lambda_{\ell\ell' t+1} = \exp\left(\beta V_{\ell' t+2} - F_{\ell\ell' t+1} + u_{\ell t+1} - V_{\ell t+1}\right)^{1/s}. \tag{E63}$$

We get the following expression for the ratio of the discounted probabilities:
(E64)

$$\frac{\lambda_{\ell\ell' t}\left(\lambda_{\ell'\ell' t+1}\right)^{\beta}}{\lambda_{\ell\ell t}\left(\lambda_{\ell\ell' t+1}\right)^{\beta}} = \exp\left(\beta\left(u_{\ell' t+1} - u_{\ell t+1}\right) + F_{\ell\ell t} - F_{\ell\ell' t} + \beta\left(F_{\ell\ell' t+1} - F_{\ell'\ell' t+1}\right)\right)^{1/s}.$$

Using the assumption imposed in Proposition III.3 that $F_{\ell\ell t} = 0$ for all t and $\ell \in \{m, c\}$, and that $F_{lm} = F_{ml} = F$, the logarithm of this ratio can be written as

$$\ln\frac{\lambda_{\ell\ell' t}\left(\lambda_{\ell'\ell' t+1}\right)^{\beta}}{\lambda_{\ell\ell t}\left(\lambda_{\ell\ell' t+1}\right)^{\beta}} = \frac{\beta}{s}\left(u_{\ell' t+1} - u_{\ell t+1}\right) - \frac{1-\beta}{s}F. \tag{E65}$$

E6. Testing for Multiple Equilibria

It is in principle possible for models like ours to exhibit multiple equilibria, not only multiple stationary equilibria. For example, a city could start in an initial condition with low commuting, but individual expectations of a return to high commuting might put the system on a self-fulfilling path. In this subsection, we describe how we test – and find no evidence of – this possibility.

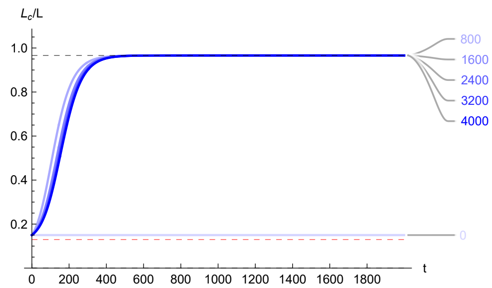
In particular, let us examine first a case where, starting from an initial condition that we expect to converge toward the high-commuting equilibrium, the system might actually converge toward the low commuting one. For cities with multiple stationary equilibria, we have always found 2 stable and one unstable stationary equilibrium. We therefore start by setting an initial condition for L_{c0} above the unstable equilibrium. The top panels of Figure E2 set $L_{c0}/L = 0.15$ for the metro area of New York, which is above the unstable stationary equilibrium (dashed red line).

In Panel (a), we set as initial guesses for the time path of the value functions V_c and V_m their respective value in the high-commuting stationary equilibrium. We require, as a convergence criterion, both that the sup-norm relative distance between successive guesses of the value functions is less than a given threshold and that the terminal value of these value functions is within the same relative tolerance from the high-commuting stationary equilibrium value. We set this threshold to 10^{-6}. The panel shows successive time paths of L_c/L at the indicated iteration number. In this exercise, our value function iteration algorithm converges to a solution in a little more than 4,000 iterations.

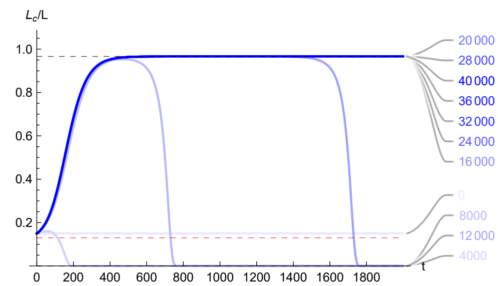
Starting from the same $L_{c0}/L = 0.15$, Panel (b) sets as initial guesses for the time path of the value functions V_c and V_m their respective value in the low-commuting stationary equilibrium. We impose the same convergence criterion, requiring this time that the terminal value of the value functions converge to the low-commuting stationary equilibrium. In this case, our value function iteration algorithm does not converge within 40,000 iterations. Panel (b) shows that along the iteration process, the guesses attempt to reach the low commuting equilibrium, but they move progressively away from it.

The bottom panels of Figure E2 report analogous exercises where we instead start from an initial condition below the unstable equilibrium, namely $L_{c0}/L = 0.10$. In Panel (c), we again attempt to converge toward the high-commuting equilibrium, with no success. In Panel (d), we successfully converge towards the low-commuting equilibrium in about 3,600 iterations.

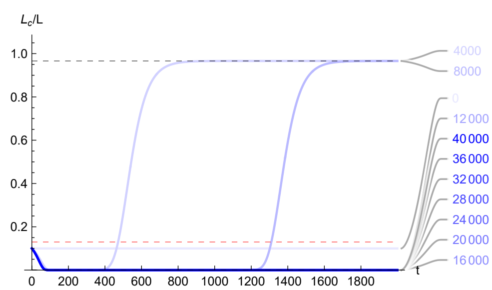
In unreported results, we have also experimented with other initial conditions farther away from either side of the unstable equilibrium, and other cities of varying sizes (like San Francisco, CA, Kansas City, MO-KS, Milwaukee, WI, Salt Lake City, UT, and Providence, RI), always reaching the same conclusion.



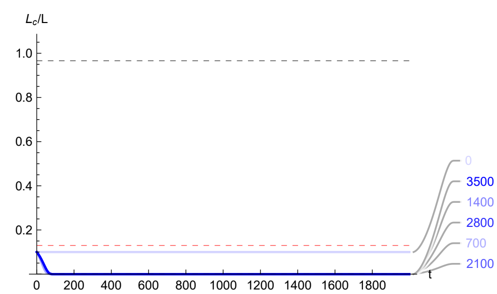
(a) $L_{c0}/L = 0.15$, shooting toward high-commuting equilibrium



(b) $L_{c0}/L = 0.15$, shooting toward low-commuting equilibrium



(c) $L_{c0}/L = 0.10$, shooting toward high-commuting equilibrium



(d) $L_{c0}/L = 0.10$, shooting toward low-commuting equilibrium

Figure E2. Tests for multiple equilibria