Highways, Commuting and Trade: Unpacking Suburban Growth

Daniel Velásquez*

University of Michigan

Preliminary and Incomplete - October 2024

Newest version here

Abstract: I study the effect of U.S. Interstate Highway System on suburbanization. Highways reduced the cost of moving workers (via commuting) and goods (via trade). I leverage variation in highway construction dates and driving time reductions to show that new highways induce both commuting and trade shocks. I find that, on average, a rise of one standard deviation in market access rises population and employment by about 10% after 40 years. However, a rise of one standard deviation in commuting access rises population and employment by about 1% after controlling for market access changes. I develop a quantitative model of trade, commuting, migration, and scale externalities, and map it to my reduced-form estimates through indirect inference. The model is designed to capture the average reduced-form effects from the data. However, because the introduction of highways induced heterogeneous changes in market and commuting access, the model allows me to characterize the contributions of each factor separately and explore their interactions. Using the calibrated model, simulations show highways account for 15% of suburban growth and 33% of the decline in urban cores. Trade costs reductions, and not only commute costs reductions, shaped the rise of the suburbs and the decline of urban cores.

*Department of Economics, University of Michigan (email: danielvc@umich.edu). I am very grateful to Dominick Bartelme, Agostina Brinatti, Sun Kyoung Lee, Andrei Levchenko, Sebastian Sotelo, Stephen Terry, and Dean Yang for comments and suggestions. The standard disclaimer applies.

1 Introduction

Suburbanization has become a defining feature of the U.S. since 1950, with suburban counties emerging as major population and employment centers, while urban cores have lost residents and jobs. This spatial reorganization has far-reaching consequences—contributing to environmental degradation, reduced density, and rising fiscal and maintenance burdens for cities.[1] Yet the factors driving this shift remain debated,[2] and unpacking its causes is crucial for informing future policy. While much of the debate has focused on how highways have reduced commuting costs, their role in lowering trade costs may play a key part in driving suburbanization.

I revisit the impact of the Interstate Highway System, and offer a complementary view by showing that trade cost reductions—and not only commute costs—caused suburban growth and urban decline. I highlight how city networks shape the effects of highways. For example, reduced driving times on I-95 between Baltimore and New York provide far greater value to firms shipping goods than those on I-70 connecting Baltimore with Pittsburgh due to New York's sheer size. Moreover, the distinction between commuting and trade costs is relevant because the long distance between Baltimore and New York makes commuting impractical. Thus, any influence of New York on Baltimore's spatial distribution of economic activity comes through trade. As trade costs fall, migration and job reallocation follow, particularly toward suburban areas with direct highway access. In addition, new residents may work and consume locally, while also enhancing amenities and productivity through scale externalities. This nationwide perspective clarifies how improvements in trade accessibility in one place can ripple across commuting, trade, and migration networks, reinforcing suburbanization patterns throughout the U.S.

Against this backdrop, this paper addresses two main questions: To what extent did the construction of the Interstate Highway System contribute to the rise of suburbs and the decline of urban cores in the U.S.? And how much of this effect can be attributed to the reduction of trade costs? To this end, I proceed in two steps. First, I show empirically that highways substantially reduced trade costs, not only commuting costs. Second, I quantify the role highways in the rise of the suburbs and the decline of urban cores due to trade cost reductions, alongside the fall in commuting costs.

I start by constructing a new data set with dated and georeferenced placements of highways by mile. These data allows me to compute bilateral driving times between counties for each decade. Using

[1] See Brueckner (2000); Glaeser and Kahn (2004); Ewing et al. (2003).

[2] See Cullen and Levitt (1999) on crime; Boustan (2010) on racial tensions and white flight; Reber (2005) on schooling; Baum-Snow and Hartley (2020) on amenity value, see Baum-Snow (2007) on highway construction.

these data, I leverage variation in highway construction dates and driving time reductions to estimate how changes in access to employment centers (i.e., commuting access) and in access to markets for shipping goods (i.e, trade access) affect both population and employment growth in U.S. counties between 1950 and 2020. I find that a one standard deviation increase in trade access raises population and employment by 9.6% and 11.8%, while a one standard deviation increase in commuting access results in only a 0.03% and 1.5% rise. In other words, on average, trade cost reductions have a greater role than commute costs reductions in driving both population and employment growth at the county level. This result holds across a battery of robustness tests, including concerns about the endogeneity of highway placement.

In the second step, I develop a quantitative spatial model of trade, commuting, migration, and production and amenity externalities. The model includes both tradable and non-tradable sectors, along with a floorspace construction sector for added quantitative realism. I use the model to separately understand the contributions of trade and commute costs on suburbanization, while also exploring their interaction.

To guide the quantitative analysis, I derive a set of analytical results from a stripped-down version of the model's equilibrium. Suppose there is a set of commercial locations producing goods and another set of residential locations where workers live. Because of heterogeneity in trade and commute costs, each location sends workers to a particular subset of commercial areas while importing goods from another subset.

In this environment, I explore the migration consequences of trade cost reductions. Reductions in trade costs trigger migration responses as workers seek lower prices. Labor demand and supply raises based on existing trade and commute linkages. For instance, new residents consume goods from certain areas, which raises labor demand there. At the same time, they commute to other areas, increasing labor supply. Knowing the locations of these shifts in labor demand and supply is crucial, as they determine to what extent migration responses amplify the effect of initial reductions in trade costs. Then, I outline the migration effects of commute cost cuts. Lowering commute costs has qualitatively similar effects on population and employment growth as reducing trade costs, but it yields quantitatively different results. In this case, migration arises from improved commuting conditions rather than lower prices, resulting in different increases in labor supply and demand across space compared to those driven by trade cost reductions. Finally, I determine under what conditions drops in commute and trade costs cause suburban growth and decline of urban cores. Whether a decrease in trade or commute costs between residential and commercial areas drives suburban growth depends on (i) whether suburbs are among the most affected areas by these costs reductions, (ii) whether migration responses are stronger in suburban locations relative to other areas, and (iii) whether areas experiencing a greater influx of new residents consume from and commute to suburban areas.

I use an indirect inference approach to estimate amenity externalities and calibrate the remaining parameters. The comparative statics show how amenity externalities shape the total impact of trade and commute access, and when trade access effects are amplified more than commute access. First, traditional models assume constant scale elasticities, where externalities grow with population, but endogenous amenities intensify migration by making growing locations more attractive. Second, whether amenity externalities amplify trade access effects more than commute access depends on the magnitude of migration responses and their correlation with trade and commuting linkages. For example, migration triggered by trade cost reductions can lead to labor supply shifts if migrants settle and work in areas that produce goods. In contrast, commute cost reductions may direct migration toward locations reliant on imports, limiting local labor demand shifts. Thus, in this example, stronger amenity externalities amplify the effects of trade cost reductions more than those of commute cost reductions.

Assuming amenity externalities increase proportionally with population growth (i.e. constant scale elasticity) can be quite restrictive. Under this assumption, it is possible to only rationalize the effect of trade access and not of commute access, or vice-versa. To generalize this assumption, and for parsimony, I follow Bartelme, Li, and Velasquez (2024). I assume that amenity externalities are only active within a middle range of population sizes.[3] This assumption gives the model greater flexibility to match the average effects of commute and trade access.[4] Consider a reduction in trade costs. The correlation between triggered migration flows and existing trade and commuting linkages varies across the county-size distribution. This dependence with county size is not necessarily the same produced by commute cost reductions. Thus, the threshold at which externalities become inactive determines the average effect of trade and commute access. In favor of this assumption, I find that the data and

[3] An interpretation is that congestion forces become too large in the largest locations, while the smallest locations do not have enough people to sustain the production of endogenous amenities. TFP externalities, which as of this version of the paper are calibrated from the literature, are assumed to remain constant, meaning that large cities retain their productivity advantages, as well-established in the literature.

[4] Note that the average effect of commute or trade access is a weighted average of its impact on each county.

the indirect inference procedure strongly suggest the presence of heterogeneous amenity externalities. The estimated threshold at which these externalities become inactive is around 200,000. Additionally, we can confidently reject a threshold close to infinity, which would be consistent with a constant scale elasticity framework.

Once the model is calibrated to the U.S economy, I use it to quantify the extent to which highways contribute to suburban growth and urban core decline. I also use it to examine how the interaction between trade and commuting influences these outcomes. I simulate a counterfactual scenario without the Interstate Highway System, holding all other shocks between 1960 and 2020 constant. I evaluate three cases: highways reduce only commute costs, only trade costs, and both simultaneously. Comparing the observed data to the counterfactual scenario where highways reduce both trade and commute costs shows that highways account for 15% of suburban growth and 33% of the decline in urban cores. When highways reduce only commute costs, the model shows that the fall in commute costs is responsible of 88% of the effect of highways on suburban growth, but cannot explain the decline of urban cores. Rather, suburban growth comes at the expense of rural areas. In contrast, trade cost reductions explain the remaining 12% of the effect of highways on suburban growth, and 100% of the effect of highways on the decline of urban cores.

In this research, I contribute to four strands of the literature. First, I advance the debate on the causes of suburbanization.[5] I emphasize that highways not only reduce commute costs (Baum-Snow, 2007; 2020; Baum-Snow et al., 2017) but also reduce trade costs (Michaels, 2008; Faber, 2014; Duranton and Turner, 2014), further reinforcing suburbanization patterns and contributing to urban core decline. Furthermore, I highlight the role of city networks in shaping trade access effects, and therefore, suburbanization. Finally, I provide a novel quantification of highway effects that includes not only cores and suburbs but also smaller cities and rural areas.

Second, I contribute to the empirical and theoretical literature evaluating the impact of new transport infrastructure by examining the interplay between trade costs and commute costs.[6] While transport infrastructure is typically associated with reductions in either trade costs or commute costs, I analyze both dimensions together. By using the new data set of bilateral driving times I constructed,

[5]See Cullen and Levitt (1999) on crime; Boustan (2010) on racial tensions and white flight; Reber (2005) on schooling; Baum-Snow and Hartley (2020) on amenity value, see Baum-Snow (2007) on highway construction.

[6]Trade: Michaels, 2008; Duranton et al., 2014; Faber, 2014; Allen and Arkolakis, 2014, 2023; Donaldson, 2018; Monte et al., 2018; Sotelo, 2020; Frye, 2024. Commuting: Baum-Snow, 2007, 2020; Duranton and Turner, 2012; Monte et al., 2018; Heblich et al., 2020; Baum-Snow et al., 2020; Severen, 2021; Zarate, 2021; Brinkman and Lin, 2022; Tsivanidis, 2023; Weiwu, 2023, 2024; Velasquez, 2024.

I empirically demonstrate that, once I account for changes in trade access due to highway construction, the effect of commute access on population and employment growth becomes negligible, on average. In contrast, trade access produces population and employment growth. By utilizing these average responses of trade and commute access, I calibrate a general equilibrium model of trade, commuting, migration, and externalities. The model is designed to capture the average reduced-form effects from the data. However, because the introduction of highways induced heterogeneous changes in market and commuting access, the model allows me to characterize the contributions of each factor separately and explore their interactions.

Third, I contribute to economic geography models of trade and migration. I consider how migration amplifies initial impacts of trade and commute access changes. My work builds on Monte et al. (2018), who also explore the relevance of commuting patterns. I provide direct empirical evidence on the impacts of trade and commute access changes on population and employment growth. I also demonstrate how externalities improve the match between empirical findings and model predictions regarding the effects of trade and commute access. Finally, I analyze how the interaction between trade, commuting, and migration shapes suburbanization in the U.S., with migration reinforcing the effects of trade and commute access.

Fourth, there is a long tradition of research on scale externalities, commonly known as agglomeration economies.[7] Some focus on amenity externalities.[8] I provide new estimates of dynamic amenity externalities and evidence of heterogeneous scale elasticities using an indirect inference approach. Donaldson and Allen (2023) explore path dependence, showing how dynamic amenity and production externalities can create multiple steady states. Bartelme, Li, and Velasquez (2024) extend this framework by incorporating heterogeneous scale elasticities. They examine how a temporary shock can have long-term effects on middle-sized locations when agglomeration economies are sufficiently strong. My contribution lies in analyzing the long-run impacts of a permanent shock, focusing on how heterogeneous scale externalities, paired with commute and trade linkages, shape long-term outcomes across urban cores, suburbs, medium metros and rural areas.

[7] See Krugman (1991), Ciccone and Hall (1996), Glaeser and Maré (2001), Lucas and Rossi-Hansberg (2002), Duranton and Puga (2004), Combes, Duranton, and Gobillon (2008), Allen and Arkolakis (2014, 2023) , de la Roca and Puga (2017), Kline and Moretti (2014), Ahlfeldt, Redding, Sturm, and Wolf (2015).

[8] Bayer, P., Ferreira, F. and McMillan, R. (2007), Diamond (2016), Leonardi, and Moretti (2023), Almagro and Domínguez-Iino (2024), De la Roca, Parkhomenko, and Velasquez (2024)

2 Background, Data and Stylized Facts

2.1 Background and Data

Background. Though conceived in 1947 with a 37,324-mile plan—designed primarily for national defense, not urban development—the Interstate Highway system did not truly materialize until the Federal-Aid Highway Act of 1956 spurred construction of 42,800 miles of interconnected freeways in many construction waves. By 1990, virtually all of the original plan was executed, transforming the nation's road network from limited-access miles to a vast, integrated grid. Crucially, many highways were constructed following a "ray" pattern, connecting outlying areas directly to central city cores (Baum-Snow, 2007). This reshaped the structure of U.S. cities by reducing commute times between suburbs and urban cores. Similarly, highway segments connected many cities, effectively reducing trade costs between U.S. markets (Duranton et al., 2014).

Data. I compile several datasets. To study population characteristics at the county level, I use the Decennial Census and the American Community Survey Data (1950-2020).[9] To document the distribution of jobs, I rely on the County Business Patterns Database (1950-2020) by Eckert et al., (2021). To map out commuting flows over space and time, I rely on the Journey to Work Database (1970-2020).

To construct bilateral driving times between counties, I utilize the 2005 Highway Performance Monitoring System (HPMS), the PR-511 data set,[10] and various historical and planned transportation data. I merge PR-511 files with HPMS road data to georeference highway construction dates. Using the Dijkstra algorithm and a speed of 70 miles per hour, I calculate bilateral commuting times between county centroids for each decade from 1960 to 2000.[11] I also compute driving times assuming highways followed the exact paths of the 1947 Interstate plan, the 1920 Pershing Map, the 1528-1850 historical routes of exploration, and railroads around 1898 (Baumsnow, 2007; Duranton and Turner 2012; Michaels et al, 2019; Frye, 2023; Brinkman and Lin, 2024).[12]

After interpolating county-level data across decades and excluding Puerto Rico, Alaska, and other

[9] To have a consistent definition of census counties, I interpolate them using Eckert et al. (2020) crosswalks. Elsewhere, I also rely on Census-tract level data. I interpolate census-tracts using crosswalks from Lee and Lin (2018) and the Longitudinal Tract Data Base.

[10] The PR-511 data set identifies the year of construction of each highway segment (Baum-Snow, 2007; Baum-Snow, 2020; Weiwu, 2024).

[11] Centroids were determined using 2000s population density data from WorldPop.

[12] Procedures are detailed in the Data Appendix.

islands, the final dataset includes 3,092 counties for each decade from 1950 to 2020.

County classification. I use the 1990 NCHS Urban-Rural Classification Scheme for counties. This classification divides counties in large metropolitan areas (1 million+ population) into two groups: "core" metro counties (those containing all or part of the largest central city) and "fringe" metro counties (suburbs).[13] It also includes medium metros (population 250,000–999,999), small metros (50,000–249,999), and nonmetropolitan counties with cities (those with cities of 10,000 or more residents) and without cities (do not contain any part of a city of 10,000 or more residents).

This classification is ideal to study the spatial distribution of economic activity. First, the NCHS scheme captures not just population size but also functional relationships—such as the distinction between core and fringe counties in large metros—making it ideal for studying commuting patterns and suburbanization trends. Second, by categorizing both metropolitan and non-metropolitan areas, it allows for comparative analysis across a full continuum of urbanization, making it easier to assess how suburban areas differ from cores, medium and small metros, and rural areas. Moreover, every county in the U.S. is assigned a classification.

Figure 1 illustrates how the NCHS classification captures the spatial structure of metropolitan and non-metropolitan areas, distinguishing between core, suburban, and smaller metro regions. The figure shows the New York and Hartford metropolitan areas, highlighting how the scheme differentiates New York's core counties (e.g., New York, Kings, and Bronx) from suburban counties like Bergen and Rockland. Similarly, we can differentiate from Hartford's core and fringe areas.

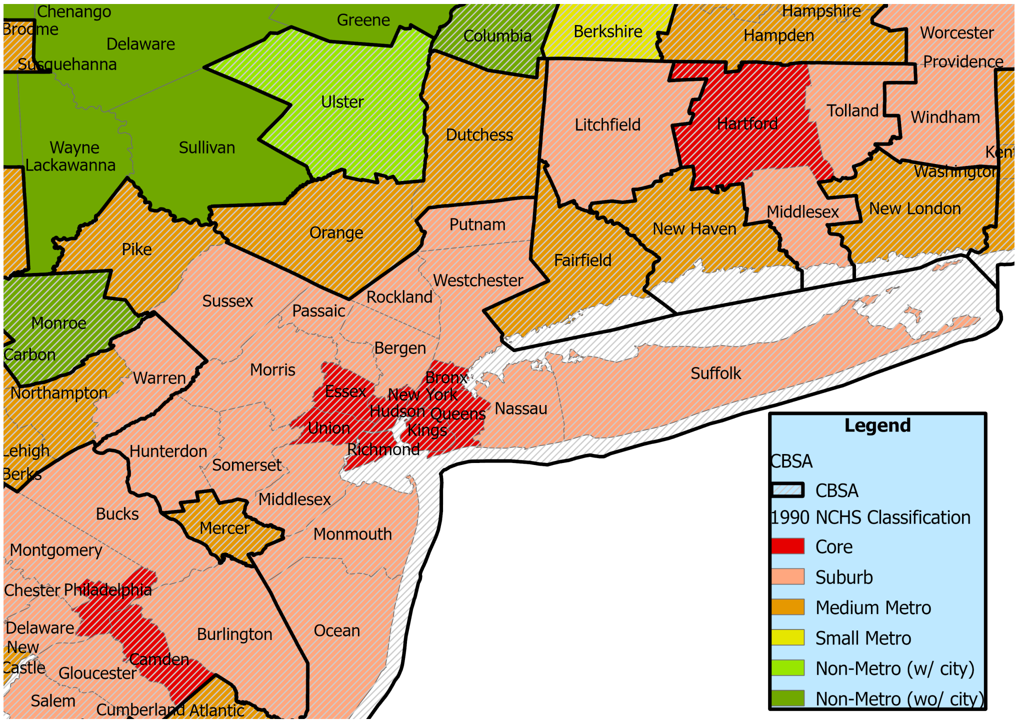
2.2 Stylized Facts

In this section, I highlight two key facts. First, suburbs have emerged as both employment and population centers at the expense of core and rural counties respectively. This paper shows that this was due to reductions in commute costs, and crucially, the fall in trade costs. Second, commuting to suburbs has grown, showing that suburbs now absorb both residents and jobs from surrounding areas. A key insight of this paper is that lower trade costs decentralize jobs to suburbs. This effect intensifies when workers can commute easily, enabling suburbs to scale up their workforce as demand for their goods rises. Therefore, it is essential to study trade, commuting, and migration together to

[13] This refers to counties that either: (i) are in metropolitan areas of 1 million or more but are not classified as large central metro, or (ii) are in areas with less than 1 million population but are adjacent to a large central metro county in a neighboring metropolitan area.

fully understand suburbanization in the U.S.

Figure 1: New York and Hartford



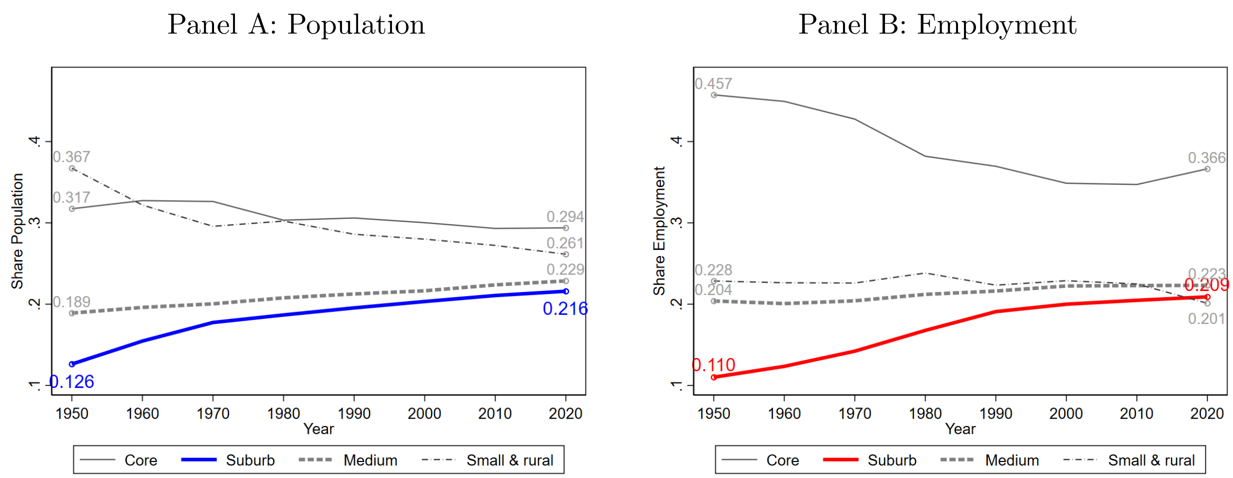
Fact 1: Suburbs emerged as population and employment centers at the expense of rural and core counties, respectively.

Suburbs have emerged as major population and employment centers. Initially accounting for about one-tenth, suburbs now represent roughly one-fifth of the total population and employment. However, while population losses were primarily concentrated in rural areas, employment declines were more pronounced in core counties.

Figure 2 shows the nationwide share of population (Panel A) and employment (Panel B) from 1950 to 2020 across different urban classifications as defined by the NCHS classification. I focus on four groups: core, suburbs, medium metros and rural (which also include small metros). In Panel A, population in suburbs rose from 12.6% to 21.6%. This shift was accompanied by a modest decrease in core counties (31.7% to 29.4%), a significant decline in rural counties (from 36.7% to 26.1%), and a modest rise in medium metros (18.9% to 22.9%). In contrast, in Panel B, the rise of employment in suburbs (11.0% to 20.9%) coincided with a sizable decline in core counties (45.7% to 36.6%), a modest decline in rural areas (22.8% to 20.1%) and rise in medium metros (20.4% to 22.3%).

This paper studies how reductions in trade costs in addition to the fall in commute costs due to highways explain Fact 1.

Figure 2: Population and employment share out of the nationwide total



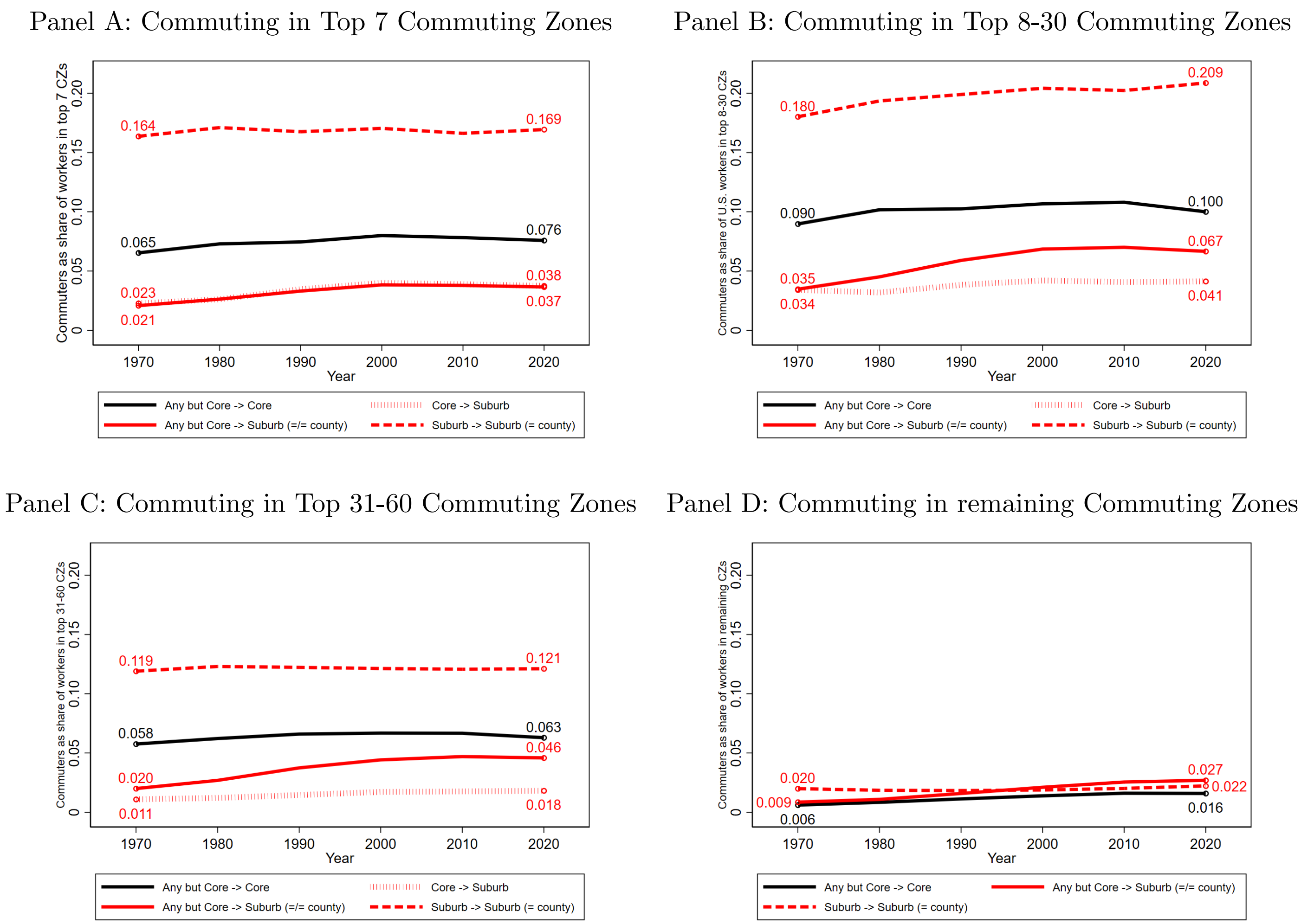
The y-axis shows the nationwide share of population or employment by classification according to the 1990 NCHS Urban-Rural Classification Scheme for counties. Small metros and nonmetropolitan counties were pooled together. For example, $Share_{k,t} = \frac{\sum_{i \in k} Pop_{i,t}}{\sum_i^S Pop_{i,t}}$, where S denotes total number of counties, i indexes counties, and k indexes categories (e.g. "Fringe"). Population data comes from the Decennial Census and the American Community Survey Data. Employment data comes from the County Business Patterns Database. To make sure counties definition over time are consistent, I interpolate county level data using Eckert et al. (2020) crosswalks.

Fact 2: Suburban counties have become increasingly important as commuting destinations, especially in mid-sized commuting zones.

Suburban counties have become increasingly important as commuting destinations. While suburb-to-core commutes have grown, commutes into suburban areas have increased even more in some areas. This increase in suburban commuting is most evident in mid-sized commuting zones, compared to both the smallest and largest zones.

Figure 3 contains four panels, each representing commuting patterns within a specific group of commuting zones. The sample is divided into four groups: the top 7 zones, zones ranked 8-30, 31-60, and the remaining zones, with each group accounting for roughly 25% of the U.S. population. Each panel shows the share of workers commuting between different origins and destinations within these zones (e.g., core to suburb).

Figure 3: Share of workers by type of commute out of total number of workers in size bin



This figure is divided in four panels or size bins. Each account for about 25% of the population. Panel A depicts the top 7 commuting zones. Panel B, the top 8-30 commuting zones. Panel C, the top 31-60 commuting zones. Panel D, the remaining commuting zones. The y-axis shows the share of commuters by 'type' of commute according to origin and destination out of the total number of workers in a given size bin, i.e., $Share_{k,bin,t} = \frac{\sum_{i\in(k,bin)} Workers_{i,t}}{\sum_{i\in(bin)} Workers_{i,t}}$, where i indexes counties, k types of commute, and bin size bin. The graph depicts these 4 types of commutes: noncore-to-core, noncore-to-suburb, core-to-suburb, suburb-to-suburb (within the same county). These shares do not sum one because there are omitted categories. Commuting flow data comes from the Journey To Work Database (1970-2020). To make sure counties definition over time are consistent, I interpolate commuting flows using Eckert et al. (2020) crosswalks.

The rise of suburb-to-suburb commuting over suburb-to-core commuting in mid-sized commuting zones highlights the interaction between trade trade and commuting access in driving suburban employment growth. In large cities, specialized industries (e.g., finance in New York) concentrate in core areas, attracting workers from various locations via commuting. Jobs in these industries are less likely to relocate to suburban areas, even when trade costs decrease due to strong agglomeration forces. In contrast, mid-sized commuting zones have a higher share of manufacturing employment.

Thus, employment is more prone to decentralization in response to trade cost reductions. This is because lower trade costs make suburban locations more attractive for manufacturing firms that rely on shipping goods to and from other markets. Commuting access also plays a role but operates differently—suburbs in mid-sized zones draw workers mainly from nearby non-core areas rather than pulling them toward the urban core. Thus commuting enables suburbs to scale up their workforce as demand for their goods rises in response to trade cost reductions.

Therefore, in the quantitative section of the paper, I set up a model with a realistic geography at the county level commute, trade and migration linkages.

3 Reduced form: The Effect of Commuting and Trade Access

I highlighted the post-1950 shift of population and employment to suburbs, with population decline in rural areas and employment losses in large central counties. I also emphasized the need to account for both commuting and trade, as suburban employment growth is linked across areas through trade.

In this section, I provide direct evidence of the commuting and trade access effects of highways by leveraging the fact that highways reduced both commute times and trade costs, but did so in an heterogeneous way across locations and dates. The impact of a new highway segment depends on the connections it creates: it can link a county to major employment hubs, increasing commuting access, or to large consumer markets where firms can ship their goods, enhancing trade access.

3.1 Measurement of commuting and trade access

Commuting access. A standard model of commuting as in Ahlfeldt et al. (2015) and Tsivanidis (2023) features the following relationship between changes in a location's commuting accessibility, and changes in commute times:

$$\Delta \log \Phi_{i,t}^{W} \quad \propto -\sum_j \pi_{ij}^0 \cdot \Delta\tau_{i,j,t} \tag{1}$$

where subscripts are: origin i, destination j, and decade t. Δ indicates a decade-by-decade change. The commuting access measure, $\Delta \log \Phi_{i,t}^{W}$, indicates that a location i's commuting accessibility to jobs will increase to the extent that commuting times are reduced along the pairs (i, j) that are used by a greater share of workers, as measured by the commuting flows, π_{ij}^0.

Trade access. A standard model of trade would predict that changes in goods' market access relate to changes in driving times in the following way:

$$\Delta \log \Psi_d^T \propto -\sum_o \lambda_{od} \Delta \log \tau_{o,d,t} \tag{2}$$

where λ_{od} are purchases by d of goods produced in o. Unfortunately, there is no data on trade flows λ_{od} for the earliest periods of my sample. Thus, I follow Donaldson and Hornbeck (2016) and approximate trade shares by noting that $\Psi_d \approx \sum_{o=1}^{N} (\varsigma_{od})^{-\varepsilon} L_o$, where $\varsigma_{od} = \tau_{od}^{\varkappa}$ are the trade costs between o and d. Suppose $\varsigma_{od} = \varsigma_{do}$. $\varkappa$ is an elasticity mapping bilateral driving times to trade costs. Thus, it can be shown that

$$\Delta \log \Psi_d^T \propto -\sum_o \underbrace{\equiv \frac{L_o (\varsigma_{od})^{-\varepsilon}}{\Psi_d^T}}_{\tilde{\lambda}_{od}} \Delta \log \tau_{o,d,t} \tag{3}$$

where $-\varkappa\varepsilon$ is the trade elasticity, which can be calibrated to -2 (Boehm et al., 2023), -1.5 (Monte et al., 2018), or -1.4 (Duranton et al., 2014). The trade access $\Delta \log \Psi_d^T$ says that when driving times between counties are reduced, trade access will increase as large markets become more accessible.

Comparing trade and commuting access. Unlike the measure of commuting access, trade access assigns positive weights to distant locations—while virtually no one commutes from San Francisco to New York, goods produced in San Francisco are still consumed in New York. By measuring commuting and trade access this way, we can make a wide range of comparisons due to the variation in both measures across counties and dates. Counties can experience an increase in commuting access without a corresponding rise in trade access, or vice versa. Some counties receive both shocks, while others experience (almost) no change. Furthermore, even if a new highway segment is built in a different county, it can still increase access, either from a commuting or trade perspective to the extent that the road network is further developed.

Figure 4 illustrates two examples. Gray lines show highway segments built by 1960, and blue lines show those built by 1970. Red counties represent large population centers (top 5% in 1950). Panel A focuses on the Atlanta metro area, highlighting Coweta County (southwest) and Jackson County (northeast). Both counties improved their commuting access to central Atlanta, but Jackson also gained access to a key trade route. As a result, both counties saw commuting access increases

(94th percentile), but Jackson's trade access rose more (83rd percentile vs. 54th for Coweta). Panel B zooms out to the Greenville-Anderson metro area, where counties saw trade access improvements (80th percentile or higher) even without new local highway segments. This is thanks to network completion elsewhere, indicated by the black arrows.

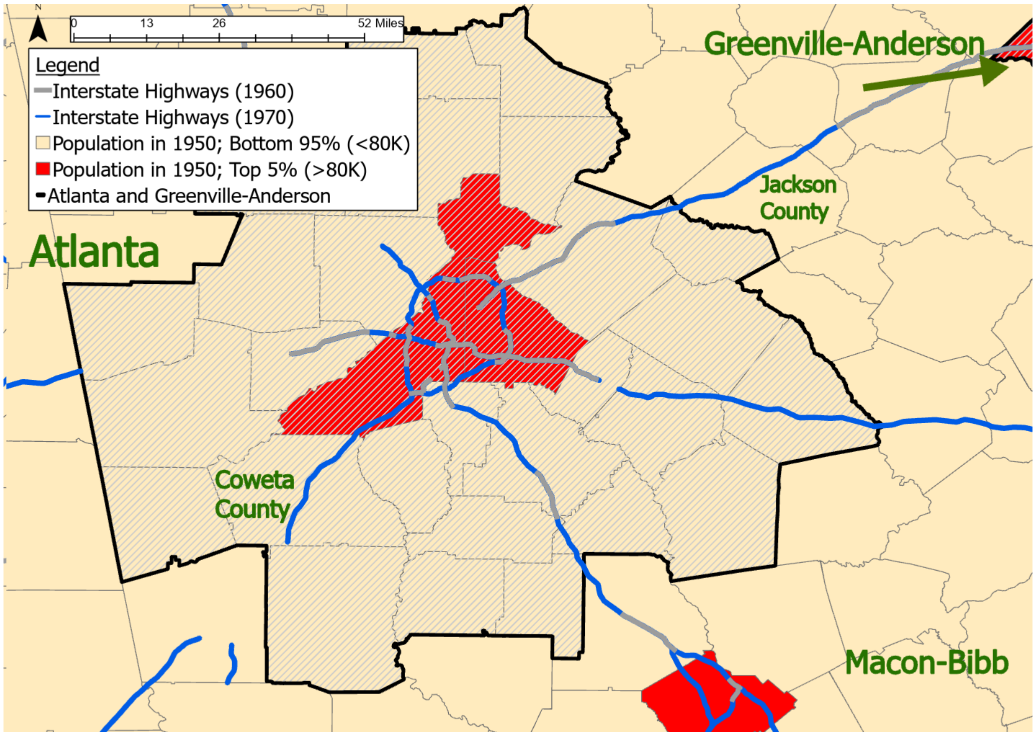
Another example, Figure 5, highlights the distinct spatial patterns of changes in commuting (Panel A) and trade access (Panel B) from 1960 to 1970. Commuting access increases are more localized, with many regions showing access growth in at least one county. In contrast, trade access spans a wider area. This difference is underscored by a relatively low correlation of 23.3% between the two measures.[14]

Limitations, solutions, and advantages. Consider a simple Two Way Fixed Effect (TWFE) regression relating outcomes such as population and employment to trade and commute access as defined earlier, conditioning on fixed effects and controls. This approach has a few limitations. First, GIS-based driving time measurements might be inaccurate due to varying traffic speeds, unlike the assumed constant speed of 70 mph. There could also be measurement error when merging PR-511 data with the HPMS roads shapefile due to the lack of segment identifiers. If measurement error is classic, OLS would be biased towards zero. Second, highway placement is influenced by political factors, which might lead to biased results if highways are strategically placed in areas that are growing less (OLS would provide a biased toward zero estimate) or in areas that are growing more (OLS would over estimate the true impact). Third, commuting access data starts from 1970, when a portion of the highway network was already built. Fourth, the impact of highways could be persistent over time. As many authors have pointed out, a simple TWFE regression is prone to bias from heterogeneous impacts over time (Callaway and Sant'Anna, 2021; Sun and Abraham, 2021; Dube et al., 2023; Callaway et al., 2024). This is because, if a unit was treated in the past and the effect of the treatment is persistent, then this unit is 'contaminated' and should not be used as a comparison unit for units that are treated later in the sample. Finally, one could argue that within county commuting is still a relevant margin to study.

[14] This is the correlation between changes in trade and commuting access over the period 1970-1960. The average decade-by-decade correlation over the period 1960-2000 is 32.8%.

Figure 4: Illustrative examples

Panel A: Atlanta, Georgia



Panel B: Atlanta and Greenville-Anderson Metro areas

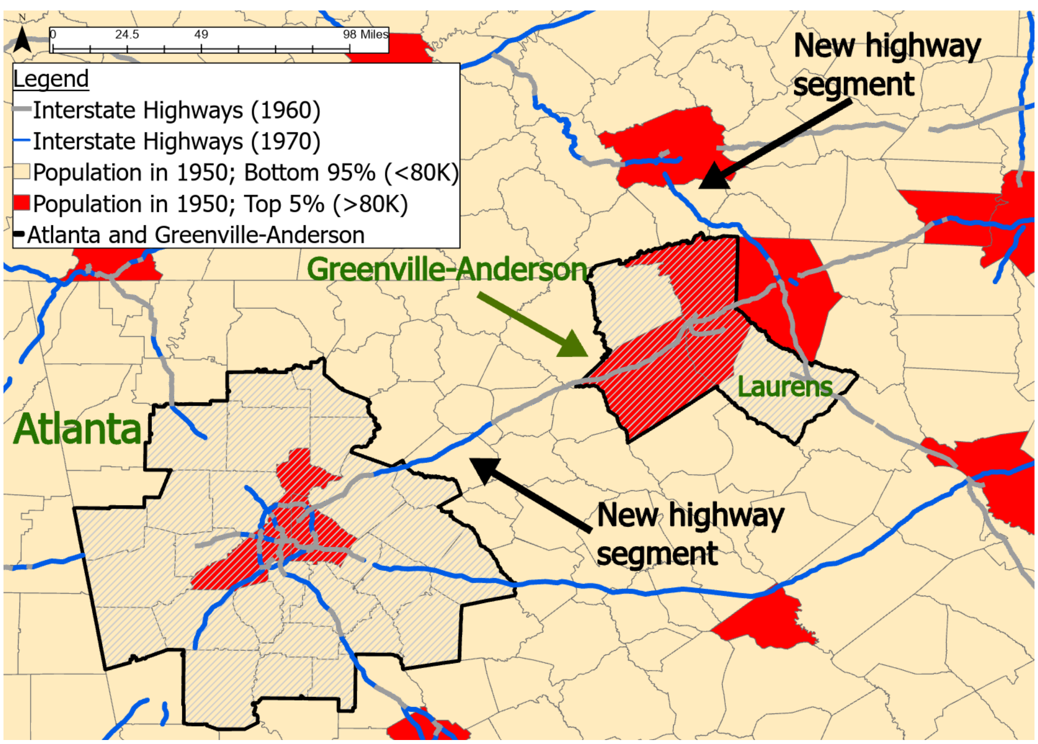
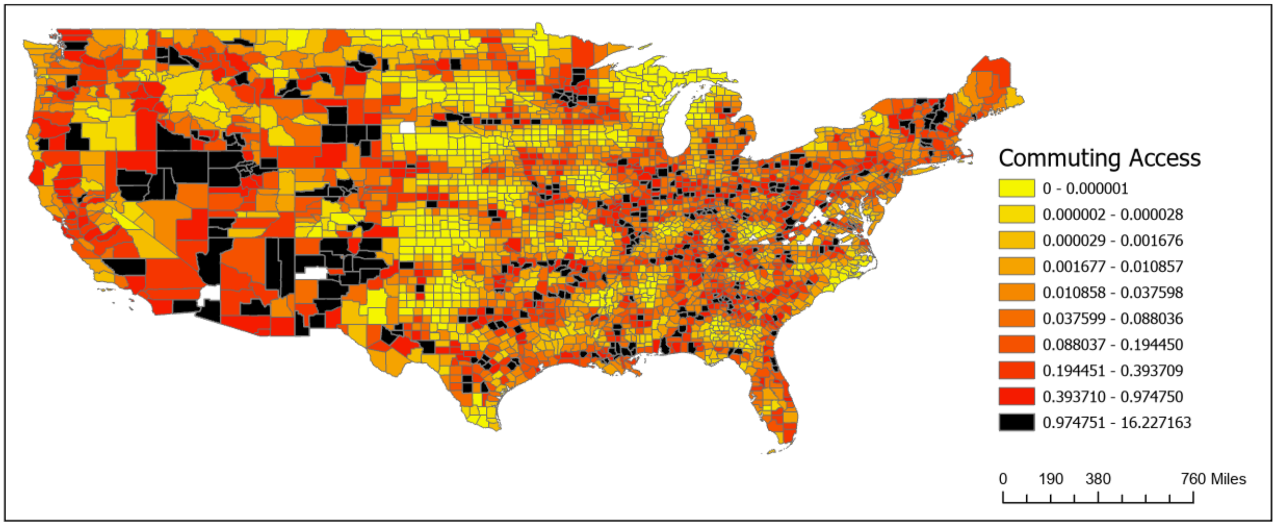
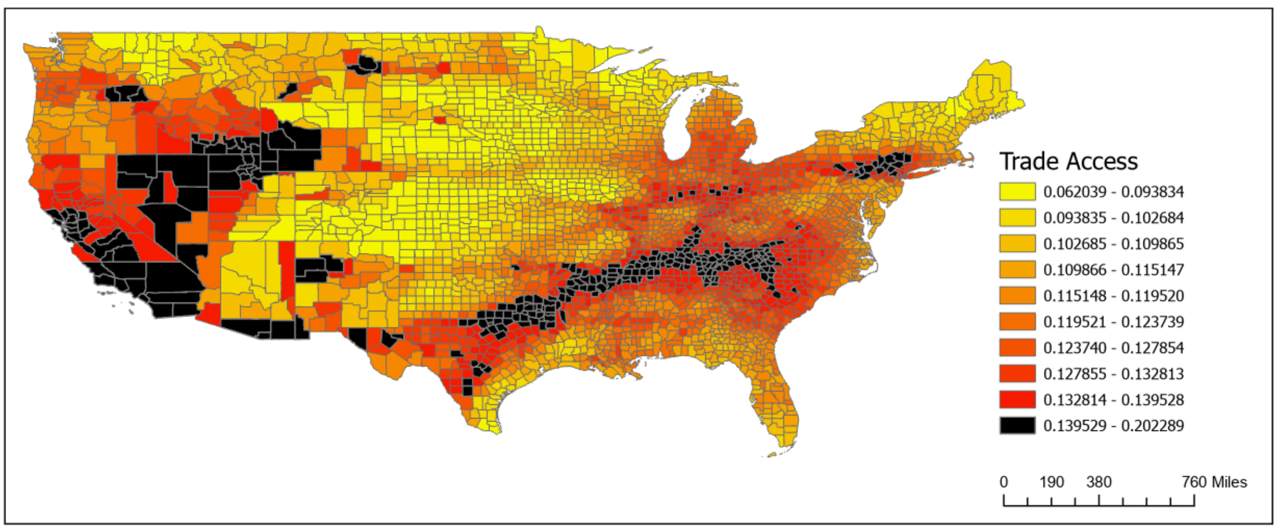


Figure 5: Changes in commute and trade access between 1970 and 1960

Panel A: Commuting Access



Panel B: Trade Access



To address the first two issues, I use an instrumental variables estimation with four instruments previously used in the literature: the 1947 Interstate plan, the 1920 Pershing Map, the 1528-1850 historical routes of exploration, and railroads around 1898 (Baumsnow, 2007; Duranton and Turner 2012; Michaels et al, 2019; Frye, 2023; Brinkman and Lin, 2024). I further discuss these instruments later in the text. To tackle the third issue, I conduct a dynamic commuting gravity estimation, showing that commuting flows adjust slowly to changes in commute times, mitigating concerns about using 1970 data instead of 1960.[15] Additionally, following Borusyak, Hull, and Jaravel (2019), exogenous

[15]Estimating a dynamic commuting gravity equation and examining its consequences could be a standalone paper.

independent shocks ensure that shift-share estimators remain consistent even when shares are endogenous. This is because these shocks are uncorrelated with bias from the shares. With a large set of shocks, the bias averages out. In this setting, changes in driving times, especially after using the instruments, can be interpreted as a set of exogenous shocks. For the fourth issue, I apply the local projection-based differences-in-differences (LP-DiD) method from Dube et al. (2023), which uses local projections to estimate dynamic effects and avoid bias by dropping 'unclean' observations. Details are expanded upon in the next section. For the last issue, I argue that the objective of this paper is to study suburbanization at the county level and not within counties. Figure 2 already shows that suburbanization at this scale was a major phenomenon. I find it unlikely that commuting within counties is able to explain suburbanization between counties. Nonetheless, to have something to say regarding within-county commuting, later in the paper I rely on census-tract level outcomes to study the impact of commute and trade access on population, while conditioning on census-tract fixed effects.

While I acknowledge these limitations and address them the best way possible, there is also some advantages with respect to alternative measures. First, to study the effect of highways one alternative would be to use dummies or perhaps count the number of miles of highways per county. The problem with this idea is that it would be hard to disentangle the trade effect from the commute effect. By relying on the commute and trade access measures I can use a richer variation in the data (i.e. the location of a county within the road network), and even test for interaction effects. Second, I could use highways dummies and a measure of distance to the closest central business district to try to pin down the commute access effect. This is because one usually expects the commute access effect to be greater on the fringe than on the central business district. However, the commuting access measure I utilize already handles this idea directly by using commuting flows as shares. My approach also do not rely on assumptions of what is a central business district and how far it is from a given county. Third, I could use long differences instead of decade-by-decade differences, thus avoiding the bias from persistent effects without relying on LP-DiD estimates. I argue that by relying on shorter differences (i.e., decade by decade), and estimating dynamic impulse responses as it is explained in the next section, I can leverage finer variation giving me a better chance of differentiating trade access effects from commute access effects.

However, these estimates help calibrate my static spatial model. I am currently working on a separate paper focused on these topics.

3.2 Empirical strategy

LP-DiD. A simple two-way fixed effects (TWFE) regression relating commute and trade access measures to outcomes would likely be biased if the effects of access are persistent over time. To address this, I apply the LP-DiD approach (Dube et al., 2023), which leverages panel data to estimate dynamic impulse responses. The LP-DiD approach utilizes local projections to estimate these dynamic effects and ensures unbiased estimates through the so-called "clean-control" condition. This condition avoids bias by excluding "unclean" observations, i.e., those that were treated in an earlier period.

One challenge in this setting is that all locations are treated to varying degrees, given the continuous nature of access measures. To address this, I restrict the sample for each wave of highway expansion by defining a treated group of locations that experience an increase in access above a specific threshold. For each treatment wave, I set the threshold at the 50th percentile of decade-by-decade changes in access measures.[16] More precisely, for each highway expansion wave s (e.g., 1970, 1980, 1990, or 2000), I classify locations into a treated group of counties j that experienced a significant increase in access measures and a control group of counties j that are either treated in the future or never treated.

To estimate the impact of highway construction on population and employment growth, I use a series of two-period panel regressions centered around the construction event. For each highway wave s, the baseline period is $s-10$, and the post-treatment period is $s+h$, where h is the horizon of interest (e.g., $h=40$). Let $q \in \{s-10, s+h\}$. The regression specification is:

$$\underbrace{\ln y_{j,s,q}}_{\text{Census/CBPD}} = \rho_h \underbrace{\log \Phi^{W,0}_{j,s,q}}_{\text{commuting}} + \varphi_h \underbrace{\log \Psi^{T,0}_{j,s,q}}_{\text{trade}} + \underbrace{\gamma_{s,q,h} + \gamma_{j,s,h}}_{\text{FE}} + \underbrace{\varpi'_h X_{j,s,q}}_{\text{controls}} + \tilde{\mu}_{j,s,h} \tag{4}$$

In the regression model, the terms $\log \Phi^{W,0}_{j,s,q}$ and $\log \Psi^{T,0}_{j,s,q}$ represent the commuting and trade market access, respectively, for county j during wave s, either before the construction of the highway, or just after the construction of the highway. Specifically, the commuting access term $\log \Phi^{W,0}_{j,s,q}$ is defined as:

$$\log \Phi^{W,0}_{j,s,q} = \begin{cases} \log \Phi^{W}_{j,s-10} & \text{if } q = s-10, \\ \log \Phi^{W}_{j,s+0} & \text{if } q = s+h. \end{cases}$$

[16] I conduct sensitivity analyses to confirm the robustness of the results to different thresholds.

Similarly, the trade access term $\log \Psi^{T,0}_{j,s,q}$ is defined as:

$$\log \Psi^{T,0}_{j,s,q} = \begin{cases} \log \Psi^{T}_{j,s-10} & \text{if } q = s - 10, \\ \log \Psi^{T}_{j,s+0} & \text{if } q = s + h. \end{cases}$$

These definitions ensure that the market access measures align with the baseline period $(s-10)$ and the post-treatment horizon $(s+h)$. The key idea is to compute the first change in market access due to highway expansion and use it to estimate the impulse response. By using these definitions in the two-period panel regressions, I measure how the initial change in market access $(\log \Phi^{W}_{j,s+0} - \log \Phi^{W}_{j,s-10})$ influences the outcomes such as population and employment growth. Put it differently, this setup captures the cumulative dynamic effects of highways over time, allowing me to study how improvements in access propagate over multiple decades. Finally, I demean $\log \Phi$ and $\log \Psi$, and re-scale with standard deviation of their change between 1970 and 1960. In all regressions I include the following set of controls interacted with wave s and horizon h fixed effects: population in 1950, change in log population between 1960 and 1950, income in 1950, share of the population with college education in 1950, employment in 1950, change in log employment between 1960 and 1950, ratio of tradable to non-tradable employment, urban dummies in 1950, and state fixed effects.

Finally, when trade and commute access both improve simultaneously, the interaction effect can be negative, meaning the two channels substitute for each other. For instance, if trade access improves, making it easier for firms to move goods across regions, but commute access also increases, workers may prefer to commute to jobs in other regions instead of working locally. This weakens the local employment benefits that improved trade access would otherwise provide. To capture such interactions, in some specifications I add the interaction term:

$$\ln y_{j,s,q} = \rho_h \log \Phi^{W,0}_{j,s,q} + \varphi_h \log \Psi^{T,0}_{j,s,q} + \zeta_h \underbrace{\left(\log \Phi^{W,0}_{j,s,q} \cdot \log \Psi^{T,0}_{j,s,q}\right)}_{\text{interaction}} + \gamma_{s,q,h} + \gamma_{j,s,h} + \varpi_h' X_{j,s,q} + \tilde{\mu}_{j,s,q}. \quad (5)$$

If the interaction coefficient ζ_h is negative, it suggests that the improvement in commute access reduces the marginal benefit of trade access. I then predict the average commute access effect as $\rho_h + \zeta_h \log \bar{\Psi}^{T,0}_{j,s,q}$, and the average trade access effect as $\varphi_h + \zeta_h \log \bar{\Phi}^{W,0}_{j,s,q}$.

2SLS LP-DiD. Highway placement may not be random, which requires the use of instruments to address potential endogeneity. For the instruments to be valid, they must satisfy both relevance and exogeneity conditions. I employ four commonly used instruments from the literature.

First, I discuss the exogeneity condition. The first instrument is the 1947 Plan, developed by the Bureau of Public Roads to promote intercity trade and national defense (Baum-Snow, 2007, 2019; Duranton and Turner, 2012). Since the focus was on national defense and intercity trade, the prevailing argument is that this plan may not be correlated with a city's internal fundamentals. However, it could still be related to fundamentals at the city level. To further address this concern, I also use the 1921 Pershing Plan, created under General John J. Pershing to prioritize military needs rather than trade (Michaels et al., 2019; Frye, 2023). Additionally, I utilize historical routes as instruments based on the premise that they are unlikely to correlate with city fundamentals from 1950 to 2020 while still predicting highway location. Specifically, I follow Duranton and Turner (2012) by using exploration routes from the 16th to 19th centuries (shapefiles from Brinkman and Lin, 2024) and historical railroads as of 1898 (shapefiles from Atack, 2015).

Second, regarding the relevance condition, Appendix XXX compares each of these instruments with highway placement by the year 2000. Visual inspection confirms that these instruments effectively predict the location of highways. While these instruments predict highway placement, they must also influence the timing of highway construction without being directly related to the outcome of interest. By interacting these instruments with time trends, I capture the temporal dynamics of highway placement. The underlying assumption is that regions with lower commute times tended to have highways built first (or vice-versa). Specifically, I compute GIS driving times assuming speeds of 70 mph along these routes and calculate commuting and trade access measures as follows:

$$\tilde{\Psi}_d^k \propto -\sum_o \tilde{\lambda}_{od}^0 \tau_{od,t}^k \quad \text{and} \quad \tilde{\Phi}_{i,t}^k \propto -\sum_j \pi_{ij}^0 \tau_{ij,t}^k \tag{6}$$

where k indexes each of these four instruments. I interact each of these variables with time trends and use them as instruments for commuting access, trade access, and their interaction.

3.3 Main results

This section establishes that the average effect of commute access on population and employment disappears once I control for trade access.

OLS results. I first present the results of increasing commuting access without controlling for trade access. Then, I add the controls for trade. Finally, I add the interaction of trade and commuting. I repeat this exercise with the variable for trade access. I use log population and log employment as outcomes.

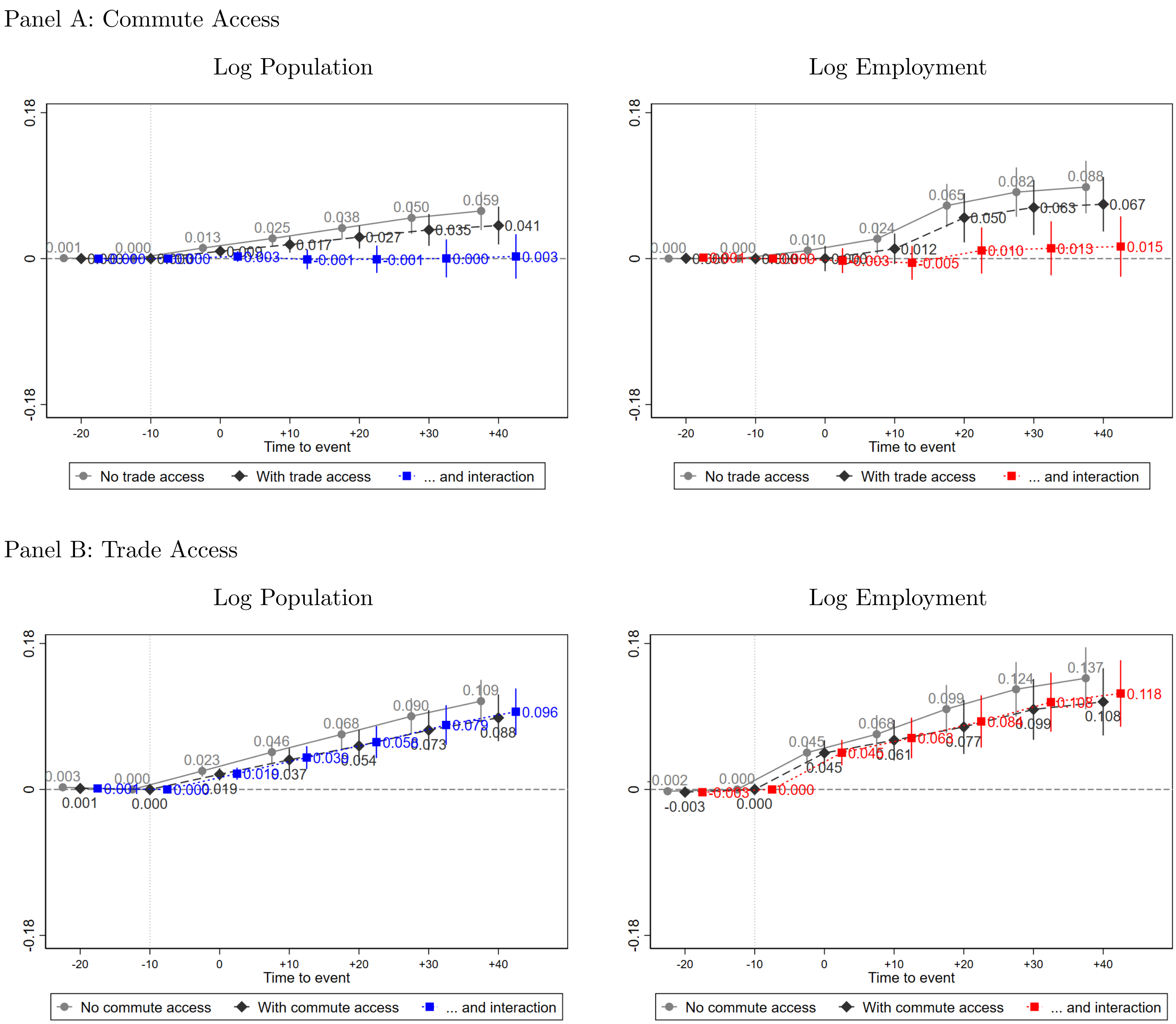
Table A.1 in the Appendix displays the coefficients for the estimation when $h = 40$, illustrating the effects observed after 40 years, specifically for log population. Column (1) shows that a one standard deviation increase in commute access results in a 5.9% increase in log population. In Column (2), we see that a one standard deviation increase in trade access leads to a 10.9% increase in population. However, when both terms are included together in the model, each coefficient decreases in magnitude. Finally, the interaction term reveals a negative coefficient, suggesting that the simultaneous increases in trade and commute access may dampen the individual effects of each. We observe similar trends for log employment in Table A.2. In the subsequent analysis, I will calculate the average impact of increasing commute access.

Panel A of Figure 6 shows that increasing commute access by one standard deviation increases population by 5.9% and employment by 10.8% after 40 years. Adding the trade access term reduces these coefficients to 4.1% in the case of population and 6.7% in the case of employment. Adding the interaction term and computing average effects leads to a commute effect of 0.03% and 1.5% for population and employment. Both are statistically indistinguishable from zero.

I turn to study the effect of increasing trade access. Panel B of Figure 6 shows that increasing trade access by one standard deviation increases population and employment by 10.9% and 13.7% after 40 years. Controlling for commute access reduces these coefficients to 8.8% and 10.8% respectively. Controlling for the interaction between trade and commute access, and computing the average impact of increasing trade access leads to an impact of 9.6% on population and 11.8% on employment. Figure A.2 shows that the impact is greater for employment than for population.

All in all, OLS regressions show that, on average, highways affect population and employment at the county level primarily through trade access, not commuting access. The effects are greater for employment than for population. Since highway placement may not be random, even when pre-trends are parallel, in the next section, I use a 2SLS estimator to confirm that trade access mediates the effect of highways.

Figure 6: OLS estimates of the effect of commuting and trade access

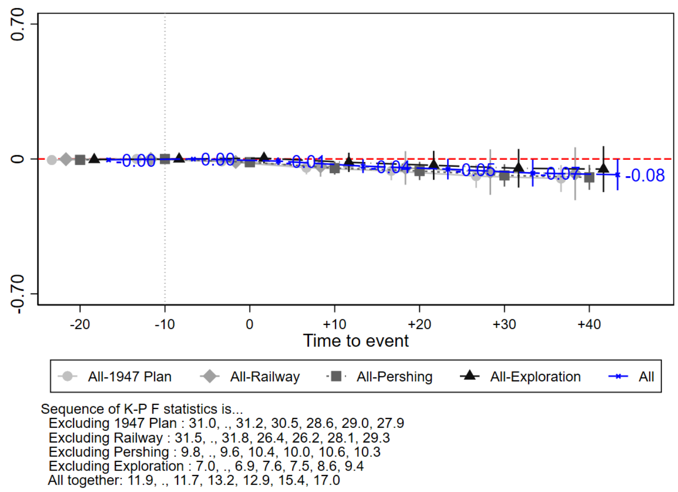


2SLS results. Figure 7 summarizes the main findings. Panel A shows the average effect of increasing commute access (when controlling for trade access and the interaction). Panel B does so for trade access. In each panel there are two outcomes, population and employment. Each figure depicts in blue and red the results from using all instruments at the same time. The figure also depicts in the background the results from excluding a particular instrument (e.g., the 1947 Plan) while keeping the remaining instruments in the estimation (e.g., "All-1947 Plan"). Kleibergen-Paap F-statistics are reported below each picture. First stage is strong in most specifications, except when excluding exploration routes from the 16th-19th centuries.

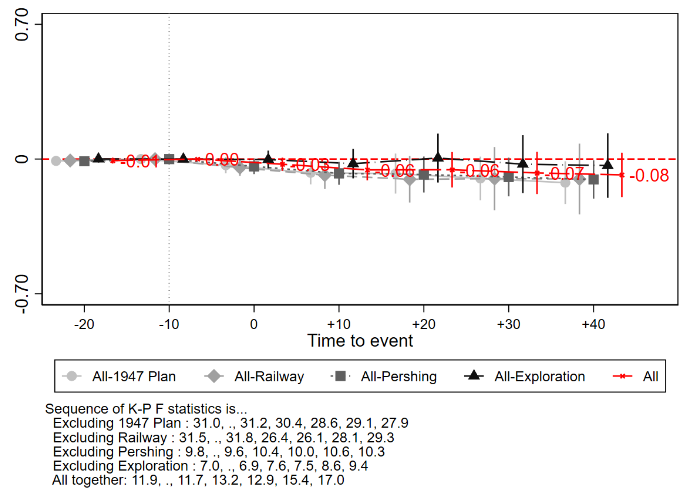
Figure 7: 2SLS estimates of the effect of commuting and trade access

Panel A: Commute Access

Log Population

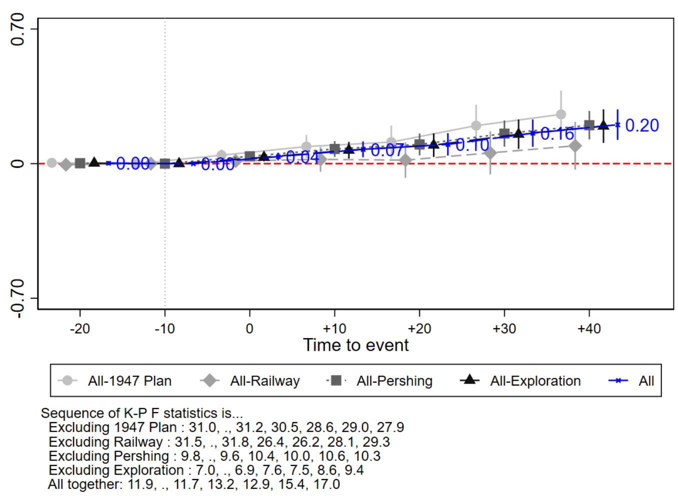


Log Employment

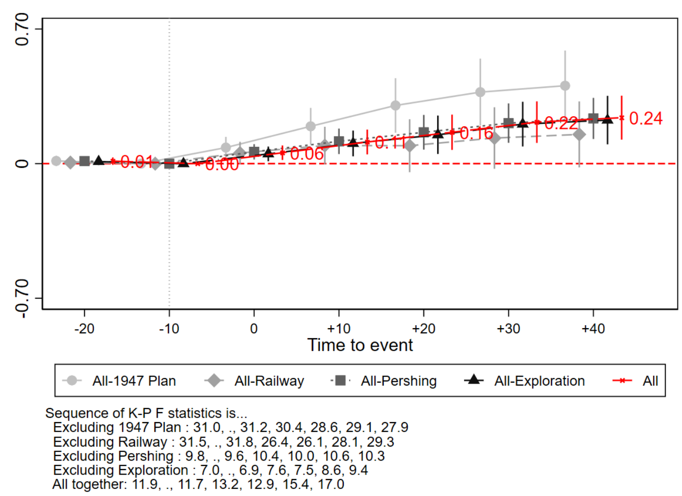


Panel B: Trade Access

Log Population



Log Employment



There are three key takeaways. First, 2SLS regressions confirm that, on average, highways impact population and employment through trade access, not commuting access. After 40 years, the effect of commute access is negative and insignificant, while trade access has a positive and significant impact.[17] Second, these results hold across different sets of instruments. Third, the 2SLS results reveal that OLS underestimates the effect (OLS is lower than 2SLS) of trade access and overestimates the effect of commute access (OLS is very close to zero, whereas 2SLS is negative but not significant).

3.4 Other results

In Appendix A.5, I present other results that are relevant to this context. First, I analyze how commuting and trade access affect peripheral counties compared to central ones, across commuting zones of different sizes. Second, in previous sections, I introduced the interaction between commute and trade access and treated it like a control, while also computing the average effect of either commute or trade access. In this appendix, I explore the sign and magnitude of this interaction term with greater detail. Third, in my baseline estimates, identification comes from a comparison of access measures between counties, even when they belong to different commuting zones but within the same state. In this appendix, I introduce commuting zones fixed effects interacted with year-by-wave fixed effects to make comparisons within commuting zones. Fourth, I study the trade and commute access effects within counties, that is, at the census tract level. Fifth, employing a dynamic gravity equation accounting for persistent effects, I show that commuting flows between counties do respond to declines in commute times.

4 Theory

I develop a Quantitative Spatial Model of the U.S., where locations are connected through trade, commuting, and migration. The model features a discrete set of locations, indexed by $i, j, o, d, , n \in \mathcal{S}$, and periods $t \in \mathcal{T}$.

Firms are static and workers myopic. Each location is endowed with land, which, combined with labor, produces floorspace. Floorspace and labor are then used to produce either a tradable or non-tradable good, with the tradable good also requiring intermediate inputs. Tradable goods are aggregated into a tradable goods bundles for consumption and for intermediate use. Workers live for two periods: in the first, they are born and decide where to migrate, and in the second, they consume the final good, demand floorspace, supply labor, and have children in the location they migrated to.

[17] Figure A.3 shows that the impact is greater for employment than for population, albeit sometimes the difference is not statistically significant.

The model introduces dynamic externalities in both the tradable sector and residential amenities, similar to the approaches in Allen and Donaldson (2022) and Bartelme, Li, and Velasquez (2024).

4.1 Preferences, commuting and migration

Preferences. There is a continuum of individuals indexed by ω, each living for two periods. In the first period ("childhood"), they consume what their parents consume. In the second period ("adulthood"), they choose where to live and work, supply labor inelastically, consume a final good and housing, and give birth to ζ_t children. The utility of an individual ω born in n, living in i, and commuting to j is:

$$U_{n,i,j,t}(\omega) = \nu_i(\omega) \cdot \xi_{n,i,t} \cdot u_{i,t} \cdot \frac{\epsilon_j(\omega)}{d_{i,j,t}} \cdot C_{i,t}^{\alpha} H_{i,t}^{1-\alpha} \tag{7}$$

where $\nu_i(\omega)$ and $\epsilon_j(\omega)$ are idiosyncratic preferences for living in i and commuting to j. $d_{i,j,t} = \exp(\kappa\tau_{i,j,t})$ is the commute iceberg cost between origin i and destination j, and $\xi_{n,i,t}$ is the migration iceberg cost between origin n and destination i, both of which depend on driving times. $u_{i,t}$ are amenities, $C_{i,t}$ is final good consumption, and $H_{i,t}$ is residential floorspace consumption. Households maximize this utility subject to the budget:

$$y_{n,i,j,t}(\omega) \equiv w_{j,t} = P_{i,t}C_{i,t} + R_{i,t}H_{i,t} \tag{8}$$

where $w_{j,t}$ is the wage at j, $P_{i,t}$ is the price of the final good, $R_{i,t}$ is the rental price of residential floorspace.

Commuting. Conditional on their place of residence, individuals draw location-specific preferences $\epsilon_j(\omega)$ from a Frechet distribution: $P(\epsilon_j \leq \epsilon) = \exp\left(-K\epsilon^{-\theta}\right)$, where θ is the shape parameter. They seek the location with the highest utility, which translates to the highest net income: $y_{i,j,t}(\omega) = \frac{w_j\epsilon_j(\omega)}{d_{i,j,t}}$. In other words, they look for the location that provides the highest income after accounting for commuting costs and commuting preferences. Dispersion in commuting preferences affects sensitivity to commute costs. When all locations are homogeneous from a preference perspective, then a small reduction in commute costs triggers a big reaction. Put it differently, a higher θ (i.e. lower dispersion) means higher sensitivity. On average, the probability of choosing location j given residence i is:

$$\pi_{ij|i,t} \quad = \quad \frac{\left(\frac{w_{j,t}}{d_{i,j,t}}\right)^{\theta}}{\Phi_{i,t}} \tag{9}$$

where commuting acess is $\Phi_{i,t} = \sum_{l=1}^{S}\left(\frac{w_{l,t}}{d_{i,l,t}}\right)^{\theta}$.

The expected income conditional on residence i is:

$$\mathcal{E}[y|i,n] \quad = \quad \sum_{j} \pi_{ij|i,t} w_j \equiv \bar{y}_{i,t} \tag{10}$$

The expected utility, given residence i and birthplace n, is:

$$\mathcal{E}\left[U_{n,i,j,t}(\omega) \mid i,n\right] \quad \propto \quad \nu_i(\omega) \cdot \underbrace{\frac{\xi_{n,i,t} u_{i,t}}{P_{i,t}^{\alpha} R_{i,t}^{1-\alpha}} \cdot \Phi_{i,t}^{1/\theta}}_{\bar{U}_{n,i,t}} \equiv U_{n,i,t}^{\mathcal{E}} \tag{11}$$

Migration. Assuming the idiosyncratic term $\nu_i(\omega)$ follows a Frechet distribution with shape parameter η, workers' utility is also distributed Frechet: $P(U_{n,i,t}^{\mathcal{E}} \leq x) = \exp\left[-\left(U_{n,i,t}^{\mathcal{E}}\right)^{\eta} x^{-\eta}\right]$. Workers choose the location offering the highest utility in the period they reach adulthood. The probability of choosing location i is $\pi_{i|n,t}^{R} = P\left(U_{n,i,t}^{\mathcal{E}} \geq \max_{l \neq i} U_{n,l,t}^{\mathcal{E}}\right)$. Thus, the number of individuals moving from location n to i is:

$$L_{n,i,t} \quad = \quad \zeta_t \frac{\bar{U}_{n,i,t}^{\eta}}{\sum_{i \in N} \bar{U}_{n,i,t}^{\eta}} L_{n,t-1} \tag{12}$$

This equation says migration flows increase towards locations with higher indirect utility $\bar{U}_{n,i,t}$ relative to outside options $\sum_{i \in N} \bar{U}_{n,i,t}^{\eta}$, and from origins with more residents $L_{n,t-1}$. From these migration flows I can compute total population per location-time period. Once I know population per location, I can compute expenditure on final goods (i.e. tradable and non-tradable goods), and residential floorspace.

This completes the discussion of the household side of the model. Next, we turn to the production of tradable, non-tradable goods, and final goods, as well as the costs associated with shipping tradable goods across locations.

4.2 Production of tradable, non-tradable, and final goods

Tradable and non-tradable goods. This model is an application of the Eaton-Kortum framework. Each location has two sectors: tradable (T) and non-tradable (NT), each producing a continuum of goods indexed by $\eth$. The efficiency of producing good $\eth$ in each sector-location is a realization of a random variable a_o, drawn from a Frechet distribution $F_o(a) = \exp(-A_o a^{-\varepsilon})$, where ε is the shape parameter (governing comparative advantage, with lower ε implying more heterogeneity and stronger comparative advantage), and A_o represents absolute advantage. The cost of purchasing a good from location o in location d in sector T is given by the random variable $p_{od} = c_o\varsigma_{od}/a_o$, where c_o is the unit cost in o, and ς_{od} are the trade costs between o and d. In the non-tradeable sector we have $\varsigma_{od}^{NT} \to \infty$ for all $o \neq d$. In the tradeable sector we have that $\varsigma_{od}^{T} \geq 1$ for all o, d. For both sectors, ς_{oo} is normalized to one.

Firms in each sector use structures $\tilde{H}_o$ and labor $\tilde{L}_o$. In the tradable sector, they also use intermediate inputs, all aggregated with an elasticity of substitution σ. In principle, σ can be below one, which would imply that labor, floorspace, and intermediate inputs are gross complements. The cost bundle in sector $s \in \{T, NT\}$ is:

$$c_{o,t}^{s} = \left(\varrho_L^s w_{o,t}^{1-\sigma} + \varrho_H^s \tilde{R}_{o,t}^{1-\sigma} + (1 - \varrho_L^s - \varrho_H^s) P_{o,t}^{T\,1-\sigma}\right)^{\frac{1}{1-\sigma}} \tag{13}$$

where P_o^T is the CES price index intermediate goods, which is the same price index of tradable goods, and $\varrho_L^{NT} + \varrho_H^{NT} = 1$.

Final goods. The final good is a CES aggregate of the tradable and non-tradable goods with elasticity σ_D. The corresponding price indexes are:

$$P_{i,t}^{NT} = \left[\int_0^1 \left(p_{i,t}^{NT}(\eth)\right)^{1-\sigma_D} d\eth\right]^{\frac{1}{1-\sigma_D}} \quad \text{and} \quad P_{i,t}^{T} = \left[\int_0^1 \left(p_{i,t}^{T}(\eth)\right)^{1-\sigma_D} d\eth\right]^{\frac{1}{1-\sigma_D}} \tag{14}$$

$$P_{i,t} = \left[b\left(P_{i,t}^{T}\right)^{1-\sigma_D} + (1-b)\left(P_{i,t}^{NT}\right)^{1-\sigma_D}\right]^{\frac{1}{1-\sigma_D}} \tag{15}$$

From quantitative point of view, including the non-tradable sector in the model is important because it grew considerably between 1950-2020 as a result of structural change. From a qualitative point of view, it is important because the non-tradable sector interacts closely with trade and commuting. If tradable and non-tradable goods are complements, a decrease in trade costs may shift the household expenditure share from tradable to non-tradable goods, causing a reallocation of employment betweens sectors. Moreover, a reduction in commuting costs (i.e. a labor supply shock to a first order) primarily affect prices in the labor-intensive sector, decreasing household expenditure share in that sector and increasing it elsewhere. In both cases, complementarity amplifies the initial shock, leading to a larger overall employment increase.

4.3 Trade

The likelihood that county o supplies a particular good to county d is the probability λ_{od} that o's price turns out to be the lowest. Recall that due perfect competition, $p^s_{od} = \frac{c^s_o \varsigma^s_{od}}{a^s_o}$ for both sectors. Thus, the lowest price realization in county d is $p_d = \min\{p_{od} : o = 1, \dots, N\}$. Given a_o follows a Frechet distribution $P(a_o \leq a) = \exp(-A_o a^{-\varepsilon})$, we have that the probability that county o provides a good at the lowest price to county d is:

$$\lambda_{od|d} = \frac{A_o \left(c_o \varsigma_{od}\right)^{-\varepsilon}}{\Psi_d}$$

where $\Psi_d = \sum_{o=1}^{N} A_o \left(c_o \varsigma_{od}\right)^{-\varepsilon}$ summarizes states of technology, input costs, and geographic barriers, and their influence on prices in each county d. Notice that for the non-tradable sector $\Psi^{NT}_d = A^{NT}_d c^{NT-\varepsilon}_d$.

Due to the Frechet assumption, goods purchased by destination counties have a price distribution that does not depend on the source. A source county with better technology, lower input costs, or fewer barriers will sell a wider variety of goods, until the price distribution of the goods sold by the source county matches the overall price distribution in the destination. Since county's d average expenditure per good does not vary by source, the fraction of goods that county d buys from county o is also the fraction of its expenditure on goods from county o:

$$\frac{X^s_{od}}{X^s_d} = \lambda^s_{od} = \frac{A^s_o \left(c^s_o \varsigma^s_{od}\right)^{-\varepsilon}}{\Psi^s_d} \tag{16}$$

Finally, to derive the price indexes, we plug the distribution of realized prices into the price indexes of equations 14.[18] Assuming $\sigma_D < 1 + \varepsilon$, the price indexes become:

$$P_d^{NT} = cons \cdot \left(\Psi_d^{NT}\right)^{-1/\varepsilon} \quad \text{and} \quad P_d^{T} = cons \cdot \left(\Psi_d^{T}\right)^{-1/\varepsilon} \tag{17}$$

with $const = \left[\Gamma\left(\frac{\varepsilon+1-\sigma_D}{\varepsilon}\right)\right]$.

4.4 Floorspace and land

Each location is endowed with T_d units of land. The production of floorspace in a location o relies on labor and land, which are aggregated into a CES production function with elasticity of substitution ν.

If labor and land are substitutes, differences in land prices can affect the production mix. For instance, in suburban areas where land prices are typically lower due to its abundance, firms tend to use more land than other inputs. This explains why factories, warehouses, and homes are generally larger in suburban and rural areas compared to central locations.

Solving the problem of the firm yields the following cost bundle for floorspace:

$$c_o^{\text{Cons}} = \left(\beta w_o^{1-\nu} + (1-\beta) Q_o^{1-\nu}\right)^{\frac{1}{1-\nu}}, \tag{18}$$

where β represents labor-specific productivity and $1-\beta$ represents land-specific productivity, normalized to sum to 1. Due perfect competition, the price index for floorspace R_o is:

$$R_o = \frac{\left(\beta w_o^{1-\nu} + (1-\beta) Q_o^{1-\nu}\right)^{\frac{1}{1-\nu}}}{Z_o}, \tag{19}$$

where Z_o denotes the total factor productivity of the construction sector.

Total floorspace H is allocated between residential and commercial uses. This allocation is determined by the total demand from residential uses (from households) and commercial uses (the sum of demand from the tradable and non-tradable sectors).

Finally, revenue from floorspace production, $H_o R_o$, is distributed between labor and land payments. Specifically, the share of revenue allocated to labor payments is $\beta \frac{w_o^{1-\nu}}{c_o^{1-\nu}}$, and the share allocated to land payments is $(1-\beta)\frac{Q_o^{1-\nu}}{c_o^{1-\nu}}$. Developers, who are responsible for producing floorspace, own the land and

[18] $G_d(p) = 1 - \exp\left(-\Psi_d p^{\varepsilon}\right)$.

use their land payments to consume the final good.

4.5 Scale externalities

The model incorporates externalities in two ways: through total factor productivity (TFP) and amenities. For TFP, I use a *constant scale elasticity* formulation (Allen, and Donaldson, 2023), while for amenities, I adopt a *heterogeneous scale elasticity* (Bartelme, Li, Velasquez, 2024). The respective specifications are:

$$
\begin{aligned}
A_{j,t}^{T} &= \bar{A}_{j,t}^{T} \cdot \left(\tilde{L}_{j,t-1}^{T}\right)^{\delta}, \\
u_{i,t} &= \bar{u}_{i,t} \cdot \begin{cases} \gamma_0^{\phi} & \text{if } L_{i,t-1} < \gamma_0 \\ (L_{i,t-1})^{\phi} & \text{if } \gamma_0 \leq L_{i,t-1} \leq \gamma_1 \\ \gamma_1^{\phi} & \text{if } L_{i,t-1} > \gamma_1 \end{cases}
\end{aligned} \tag{20}
$$

The key difference between these formulations lies in how externalities respond to changes in scale. The constant scale elasticity used for TFP assumes proportional productivity gains with increases in the labor force. In contrast, the heterogeneous scale elasticity for amenities allows the strength of externalities to vary across regions. Large locations experience congestion effects that limit the benefits of additional people, while small locations lack the critical mass needed to generate endogenous amenities. Mid-sized regions—such as suburbs—achieve a more favorable balance.

One may wonder why I use heterogeneous externalities instead of a constant scale elasticity formulation. Furthermore, one might question the choice of assuming a heterogeneous scale elasticity only for amenities. The first reason is to better match the reduced-form evidence, as a constant scale elasticity fails to capture the observed effects of trade and commute access increases. Second, while using heterogeneous scale elasticities for both TFP and amenities is theoretically possible, it would substantially increase the parameter space. In practice, I employ an indirect inference approach to align the effects of commuting and trade access in the data with those in the model. Expanding the model to estimate six parameters would make this process intractable. Third, in practice, I find that amenity externalities play a larger role in explaining the observed effects than TFP externalities, making it more practical to allow for heterogeneity where it matters most.

Section 4.7 clarifies the mapping between trade and commute access effects with externalities.

Trade access primarily shifts labor demand by increasing consumption from producing regions. Lower goods prices trigger migration, raising labor supply. In contrast, commute access directly increases labor supply but has a more uncertain effect on labor demand, as it depends on changes in both income and migration flows. They key idea is that although both shocks are amplified by amenity externalities, their effects differ across regions. Allowing stronger amenity externalities in mid-sized locations gives the model additional flexibility to amplify trade and commute shocks differently across areas with varying population sizes.

Finally, I assume dynamic externalities in the model rather than static. This assumption allows for multiple steady states while maintaining a unique equilibrium path, providing a natural mechanism for equilibrium selection within the model framework.

4.6 Market clearing and equilibrium

Market clearing. The model reduces to market clearing conditions for employment and floorspace, where I have omitted the t subscript:

$$\underbrace{\tilde{L}_d w_d}_{\text{Commuting}} = \underbrace{\varrho_L^T \left(\frac{w_d}{c_d^T}\right)^{1-\sigma} X_d^T}_{\text{Tradable Sector}} + \underbrace{\varrho_L^{NT} \left(\frac{w_d}{c_d^{NT}}\right)^{1-\sigma} X_d^{NT}}_{\text{Non-Tradable Sector}} + \underbrace{\beta \left(\frac{w_o}{c_o^{Cons}}\right)^{1-\nu} X_d^{Cons}}_{\text{Construction Sector}} \tag{21}$$

$$\underbrace{\frac{Q_d T_d}{(1-\beta)\left(\frac{Q_d}{c_d^{Cons}}\right)^{1-\nu}}}_{=H_d R_d} = \underbrace{\varrho_H^T \left(\frac{\tilde{R}_d}{c_d^T}\right)^{1-\sigma} X_d^T}_{\text{Tradable Sector}} + \underbrace{\varrho_H^{NT} \left(\frac{\tilde{R}_d}{c_d^{NT}}\right)^{1-\sigma} X_d^{NT}}_{\text{Non-Tradable Sector}} + \underbrace{(1-\alpha)\bar{y}_d L_d}_{\text{By Workers}} \tag{22}$$

The left-hand side of the employment equation represents labor supply, which depend on commuting costs. The right hand side is total labor demand which depends on trade linkages. Moreover, the left-hand side of the floorspace equation, which is the supply of floorspace, already imposes land market clearing.

To close the model, I use the fact that exports from county i are equal to total income of the tradable sector $X_i^T = \sum_d X_{id}^T = \sum_d \lambda_{id|d} X_d^T$, with balanced trade ensuring exports equal imports: $X_i^T = P_i^T C_i^{W,T} + P_i^T C_i^{H,T} + (1 - \varrho_L^T - \varrho_H^T)\left(\frac{P_i^T}{c_i^T}\right)^{1-\sigma} X_i^T$, with $w_i \tilde{L}_i^T = \varrho_L^T \left(\frac{w_i^T}{c_i^T}\right)^{1-\sigma} X_i^T$. From these, I obtain an expression for X_d^T which I plug into the employment and floorspace market clearing conditions:

$$X_d^T = \sum_o \frac{A_d^T \left(c_d^T \varsigma_{do}^T\right)^{-\varepsilon}}{\Psi_o^T} \left(\underbrace{P_o^T C_o^{W,T}}_{\text{By Workers}} + \underbrace{P_o^T C_o^{H,T}}_{\text{By Landlords}} + \underbrace{\frac{\left(1-\varrho_L^T-\varrho_H^T\right)}{\varrho_L^T} \left(\frac{P_o^T}{w_o^T}\right)^{1-\sigma} w_o \tilde{L}_o^T}_{\text{By Tradable Firms}} \right) \tag{23}$$

where $P_d^T C_d^{W,T} = b\left(\frac{P_d^T}{P_d}\right)^{1-\sigma} \alpha \bar{y}_d L_d$ is the consumption by workers, and $P_d^T C_d^{H,T} = b\left(\frac{P_d^T}{P_d}\right)^{1-\sigma} Q_d T_d$ is the consumption by landowners. The system of equations for employment and floorspace solves for wages (w_d) and land prices (Q_d).[19]

Equilibrium. We are ready to define the equilibrium of the model.

Definition 1. Given geography of bilateral costs $\left\{\varsigma_{od,t}^{NT}, \varsigma_{od,t}^{T}, d_{od,t}, \xi_{n,i,t}\right\}_{o,d,t}$, set of characteristics and initial conditions $\left\{A_{d,t}^T, A_{d,t}^{NT}, u_{d,t}, T_d, Z_{d,t}, \bar{L}_0\right\}_{d,t}$, and parameters $\left\{\varrho_L^T, \varrho_H^T, \varrho_L^{NT}, \rho, \beta, \alpha, \theta, \eta, \varepsilon, \sigma, \sigma_D, \nu, \gamma, \phi, \zeta_t\right\}$, the equilibrium is a path of prices, $\{w_{d,t}, Q_{d,t}\}$, and allocations, $\left\{H_{d,t}, \tilde{L}_{d,t}\right\}_{d,t}$, such that there is floorspace market clearing, labor market clearing, total population adds up, $\sum_i L_{it} = \zeta_t \sum_i L_{it-1}$.

4.7 Comparative statics in a simplified environment

To build intuition about the mechanics behind the effects of increasing commute and trade access and their mapping to the model's parameters, I simplify the model and conduct comparative statics. Specifically, I assume a production process that uses labor as the sole input—excluding floorspace and intermediate inputs. I also consider a tradable sector in each location. Furthermore, I assume there are no price responses to access changes, and the model is not necessarily in a steady state. Finally, I assume for simplicity that amenity externalities are contemporaneous. In other words, the model simplifies to a framework with one sector, one input, but with commuting, trade, myopic migration, and contemporaneous externalities. Given these assumptions, the model collapses into the following labor market clearing condition

$$\underbrace{\tilde{L}_d w_d}_{\text{Labor Supply}} = \underbrace{\sum_o \left[\frac{A_d \left(c_d \varsigma_{do}\right)^{-\varepsilon}}{\Psi_o} \left(\bar{y}_o L_o\right)\right]}_{\text{Labor Demand}} \tag{24}$$

[19] We can substitute $R_d = \frac{\left(\beta w_d^{1-\nu} + (1-\beta) Q_d^{1-\nu}\right)^{\frac{1}{1-\nu}}}{Z_d}$ into the system to express R_d in terms of Q_d.

where:

$$\tilde{L}_d = \sum_i \pi_{ij|i} L_i \quad L_i = \sum_n \zeta \frac{\bar{U}_{ni}^{\eta}}{\sum_{i \in N} \bar{U}_{ni}^{\eta}} L_{n,-1} \quad \bar{U}_{ni} = \frac{\xi_{ni} u_i}{P_i} \cdot \bar{y}_i$$

$$\bar{y}_i = \sum_j \pi_{ij|i} w_j \qquad \pi_{ij|i} = \frac{\left(\frac{w_j}{d_{ij}}\right)^{\theta}}{\Phi_i} \qquad \Phi_i = \sum_{l=1}^{S} \left(\frac{w_j}{d_{ij}}\right)^{\theta}$$

$$\Psi_i = A_j \left(c_j \varsigma_{ji}\right)^{-\varepsilon} \qquad P_i = cons \cdot \left(\Psi_i\right)^{-1/\varepsilon} \qquad c_j = \frac{w_j}{A_j}$$

Commute Access. *Reductions in commute costs boost labor supply directly through commuting and indirectly through migration, but their impact on labor demand remains uncertain due to potentially competing income and migration effects.*

When commute costs from i to j decrease, more workers commute to j, directly raising its labor supply. This increased job accessibility makes i a more desirable place to live, prompting migration into i. As new residents settle in i, they commute to j for work, increasing the labor supply to j. Some of the new residents will also consume from j, thereby rising labor demand in j. However, the change in commute flows also affects the average income of residents in i, increasing or decreasing labor demand. This is because the average income in i will increase if j is a high-wage location, or will decrease if j is a low-wage location. As commute costs to j is reduced, workers in the margin will be willing to reduce commute costs even by sacrificing earnings if needed. So, in sum, labor supply in j will rise but the effect on j's labor demand is uncertain.

The initial commute response is regulated by the *commute elasticity* (θ). The extent of the migration-driven shifts depends on two parameters: the *migration elasticity* (η), which measures how sensitive migration is to changes in income and job access, and the *amenity externality* ($\phi \mathcal{I}_{(i=mid)}$), which amplifies migration in locations with scale externalities in amenity production. Together, these parameters shape how reductions in commute costs propagate through migration and labor markets across regions.

To see this, I take the derivative of labor supply and labor demand with respect to log commute costs $\log d_{ij}$:

$$\underbrace{-\frac{\partial}{\partial \log d_{ij}} \log \tilde{L}_j}_{\Delta \text{ in labor supply}} = \frac{\pi_{ij|i} L_i}{\tilde{L}_j} \left(\underbrace{\left(1 - \pi_{ij|i}\right)\theta}_{\Delta \text{ in commute flows}} + \underbrace{\frac{\sum_n \frac{L_{ni}}{L_i} \left(1 - \frac{\bar{U}_{ni}^{\eta}}{\sum_{i \in N} \bar{U}_{nl}^{\eta}}\right)}{1 - \phi\eta\mathcal{I}_{(i=mid)} \sum_n \frac{L_{ni}}{L_i} \left(1 - \frac{\bar{U}_{ni}^{\eta}}{\sum_{i \in N} \bar{U}_{n,l}^{\eta}}\right)} \pi_{ij|i}\eta}_{\Delta \text{ in migration flows}} \right)$$

$$\underbrace{-\frac{\partial}{\partial \log d_{ij}} \log \tilde{L}_j}_{\Delta \text{ in labor demand}} = \frac{\lambda_{ji|i} \bar{y}_i L_i}{w_j \tilde{L}_j} \left(\underbrace{\theta \frac{\pi_{ij|i} w_j}{\bar{y}_i} \left(1 - \frac{\bar{y}_i}{w_j}\right)}_{\Delta \text{ in average income}} + \underbrace{\frac{\sum_n \frac{L_{ni}}{L_i} \left(1 - \frac{\bar{U}_{ni}^{\eta}}{\sum_{i \in N} \bar{U}_{n,l}^{\eta}}\right)}{1 - \phi\eta\mathcal{I}_{(i=mid)} \sum_n \frac{L_{ni}}{L_i} \left(1 - \frac{\bar{U}_{ni}^{\eta}}{\sum_{i \in N} \bar{U}_{nl}^{\eta}}\right)} \pi_{ij|i}\eta}_{\Delta \text{ in migration flows}} \right)$$

The first equation shows that reducing commute costs between origin i and destination j impacts labor supply to j in two ways: through changes in commute flows and through migration flows. The term $\frac{\pi_{ij|i} L_i}{\tilde{L}_j}(1 - \pi_{ij|i})\theta$ captures how much labor supply to j increases, which depends on the extent to which i contributes to j's total employment. The component $(1 - \pi_{ij|i})\theta$ of this term measures the elasticity of commute flows, indicating how sensitive commuting is to reductions in commute costs.

Changes in commute flows also influence job accessibility in i, which triggers a migration response. The effect on job accessibility is given by $\frac{\pi_{ij|i} L_i}{\tilde{L}_j}\pi_{ij|i}$, meaning it is larger when the share of people commuting to j from i is high. Higher job accessibility in i triggers migration into i, especially when job accessibility contributes more to utility, as measured by $\pi_{ij|i}\eta$. This migration response is amplified when amenity externalities are strong, reflected by the denominator $1 - \phi\eta\mathcal{I}_{(i=mid)} \sum_n \frac{L_{ni}}{L_i} \left(1 - \frac{\bar{U}_{ni}^{\eta}}{\sum_{i \in N} \bar{U}_{n,l}^{\eta}}\right)$. The term $\sum_n \frac{L_{ni}}{L_i} \left(1 - \frac{\bar{U}_{ni}^{\eta}}{\sum_{i \in N} \bar{U}_{n,l}^{\eta}}\right)$ in both the numerator and denominator simply captures how all locations l becomes less desirable as utility in i rises.

The second equation follows a similar logic for labor demand, although the final result is uncertain. Lower commute costs influence labor demand through average income in i and migration into i, both of which indirectly affect labor demand in j. The effect on average income ($\theta \frac{\pi_{ij|i} w_j}{\bar{y}_i} \left(1 - \frac{\bar{y}_i}{w_j}\right)$) is positive if j is a high-wage location ($w_j > \bar{y}_i$), and negative if j is a low-wage location ($w_j < \bar{y}_i$). The effect on migration flows to i due to lower prices is always positive. The strength of these effects depends on how important i is as a source of revenue for j, measured by $\frac{\lambda_{ji|i} \bar{y}_i L_i}{w_j \tilde{L}_j}$.

To sum up, reducing commute costs directly raises labor supply through increased commuting flows while also triggering migration, which further shifts both labor supply and demand. The extent of these migration-driven adjustments depends crucially on heterogeneous amenity externalities ($\phi\mathcal{I}_{(i=mid)}$), which amplify migration effects, particularly in mid-sized locations.

Trade Access. *Reductions in trade costs affects labor markets through direct increases in labor demand and subsequent migration effects, which in turn increases both labor supply and labor demand.*

When trade costs between a producing region j and a consuming region i, fall, consumption by i from goods produced in j increases, raising labor demand in j. At the same time, the lower prices make i more attractive as a place to live, leading to migration into i. This increase in population subsequently raises labor supply to j through commuting, while also stimulating labor demand as new residents will consume goods produced in j. The effect on trade flows is regulated by the trade elasticity ε. As with commute access, the magnitude of these migration-driven shifts is shaped by the *migration elasticity* (η) and the presence of *amenity externalities* ($\phi(\mathcal{I}_{(i=mid)})$), which amplify migration in locations with scale externalities.

More precisely, I take the derivative of labor supply and labor demand with respect to log trade costs, $\log \varsigma_{ji}$:

$$\underbrace{-\frac{\partial}{\partial \log \varsigma_{ji}} \log \tilde{L}_j}_{\Delta \text{ in labor supply}} = \frac{\pi_{ij|i} L_i}{\tilde{L}_j} \left(\underbrace{\frac{\sum_n \frac{L_{ni}}{L_{it}} \left(1 - \frac{\bar{U}_{ni}^{\eta}}{\sum_{l' \in N} \bar{U}_{n,l'}^{\eta}}\right)}{1 - \phi\eta \mathcal{I}_{(i=mid)} \sum_n \frac{L_{ni}}{L_{it}} \left(1 - \frac{\bar{U}_{ni}^{\eta}}{\sum_{l' \in N} \bar{U}_{n,l'}^{\eta}}\right)} \lambda_{ji|i}\eta}_{\Delta \text{ in migration flows}} \right)$$

$$\underbrace{-\frac{\partial}{\partial \log \varsigma_{ji}} \log \tilde{L}_j}_{\Delta \text{ in labor demand}} = \frac{\lambda_{ji|i} \bar{y}_i L_i}{w_j \tilde{L}_j} \left(\underbrace{\left(1 - \lambda_{ji|i}\right)\varepsilon}_{\Delta \text{ in trade flows}} + \underbrace{\frac{\sum_n \frac{L_{ni}}{L_{it}} \left(1 - \frac{\bar{U}_{ni}^{\eta}}{\sum_{l' \in N} \bar{U}_{n,l'}^{\eta}}\right)}{1 - \phi\eta \mathcal{I}_{(i=mid)} \sum_n \frac{L_{ni}}{L_{it}} \left(1 - \frac{\bar{U}_{ni}^{\eta}}{\sum_{l' \in N} \bar{U}_{n,l'}^{\eta}}\right)} \lambda_{ji|i}\eta}_{\Delta \text{ in migration flows}} \right)$$

The first equation indicates that reductions in trade costs affect labor supply to j through migration. The component $\frac{\pi_{ij|i} L_i}{\tilde{L}_j}$ shows that the effect is larger when origin i accounts for a significant share of employment in j. Similarly as before, the migration effect is governed by the term $\frac{\sum_n \frac{L_{ni}}{L_i}\left(1-\frac{\bar{U}_{ni}^{\eta}}{\sum_{l' \in N} \bar{U}_{nl'}^{\eta}}\right)}{1-\phi\eta\mathcal{I}_{(i=mid)} \sum_n \frac{L_{ni}}{L_i}\left(1-\frac{\bar{U}_{ni}^{\eta}}{\sum_{l' \in N} \bar{U}_{nl'}^{\eta}}\right)} \lambda_{ji|i}\eta$, which captures how changes in prices affect migration flows to i.

The second equation highlights the effect of trade access on labor demand. The term $\frac{\lambda_{ji|i} \bar{y}_i L_i}{w_j \tilde{L}_j}$ reflects the importance of i as a source of revenue for j. To the extent that this term is bigger,then the effect on trade flows $\left(1 - \lambda_{ji|i}\right)\varepsilon$ will translate into a larger shift of the labor demand in j. The migration response in i further amplifies labor demand in j. Larger amenity externalities in middle-sized locations, indicated by $\mathcal{I}_{(i=mid)}$, make these regions particularly responsive to reductions in trade costs.

Together, these equations show that the impact of trade access flows through multiple channels: it not only directly increases labor demand by raising trade flows but also indirectly shifts the labor supply and demand through migration responses, which are regulated by the amenity externality $\phi\mathcal{I}_{(i=mid)}$.

Rationalizing reduced form-impacts and heterogeneous scale elasticities. The case for using heterogeneous scale elasticities, rather than a constant one, rests on how these elasticities amplify the effects of trade access and commute access differently across locations of different population size. Recall that the data shows a positive average effect for trade access, but commute access effects are close to zero.

The key idea is that changes in migration flows magnify the relative effect of trade access to commute access in proportion to the ratio $\lambda_{ji|i}/\pi_{ij|i}$ as shown by equations above. If this ratio is low on average, migration would over-amplify commute access effects relative to trade access effects. If this ratio is high, migration would attenuate commute access effects relative to trade access effects. Applying a constant-scale elasticity for amenity externalities would further reinforce these patterns.

With heterogeneous scale elasticities, we can fine-tune these effects. For example, if the ratio $\lambda_{ji|i}/\pi_{ij|i}$ is particularly low in large counties, commute access effects will appear too large in these locations under constant elasticities. By turning off amenity externalities in large counties, we limit the magnitude of commute access effects in these counties, driving down the average impact we calculate for all regions. Conversely, if the ratio $\lambda_{ji|i}/\pi_{ij|i}$ is high in mid-sized locations, trade access effects will likely dominate over commute access effects in these areas. Since amenity externalities remain active in mid-sized counties, this boosts the relevance of trade access effects, helping us align the model's predictions with observed patterns in the data.

This section employs a stylized example. In real life we have a set of commute and trade costs changes that happen simultaneously. In any case, this example allows us to understand how heterogeneous amenity externalities can aid the model in rationalizing the observed reduced-form effects of trade and commute access.

5 Bridging Theory and Data

In this section, I describe the process of matching the model to real data, estimating key parameters, and calibrating the remaining ones.

5.1 Model inversion

Based on the knowledge on some parameters, and availability of some data, I can invert the model and recover all the model's unobservables, which are required to solve the model. A key feature of the model is that we can invert for unobservables, and then decompose them into endogenous (i.e. due externalities) and exogenous unobservables. The following theorems formalize these ideas.

THEOREM 5.1: *Given parameter* $\{\theta\}$*, data* $\{\pi_{od}, w_d\}$*, and bilateral commute costs* $\{d_{od}\}$*, there exists a unique set of commuting preferences* $\{K_{od}\}$ *that rationalize bilateral commute flows.*

Proof. See Appendix A.1 □

THEOREM 5.2: *Given parameters* $\{\nu, \beta\}$ *and data* $\{\tilde{L}_d, w_d, H_d, R_d\}$*, there is a unique set of construction TFP,* $\{Z_d\}_d$*, and land supply,* $\{T_d\}_d$*, that rationalize the data as an equilibrium of the model.*

Proof. See Appendix A.2 □

THEOREM 5.3: *If* $\alpha \bar{y}_{|d} L_d + (1-\beta)\left(\frac{Q_d}{c_d^{Cons}}\right)^{1-\nu} R_d H_d \geq \tilde{L}_d^{NT} w_d, \forall d$*, given parameters* $\{\sigma, \sigma_D, \varepsilon, \theta, \nu, \varrho^T, \varrho_L^{NT}, b, \alpha, \beta\}$*, data* $\{L_d, \tilde{L}_d^s, w_d, H_d, R_d,\}$*, and bilateral costs* $\{\varsigma_{od}^{NT}, \varsigma_{od}^T, d_{od}\}$*, there is a unique set of TFP in the tradable and non-tradable sectors* $\{A_d^T, A_d^{NT}\}_d$ *that rationalize the data as an equilibrium of the model.*[20]

Proof. See Appendix A.3 □

THEOREM 5.4: *If* $\alpha \bar{y}_{|d} L_d + (1-\beta)\left(\frac{Q_d}{c_d^{Cons}}\right)^{1-\nu} R_d H_d \geq \tilde{L}_d^{NT} w_d, \forall d$*, given parameters* $\{\sigma, \sigma_D, \varepsilon, \theta, \nu, \varrho^T, \varrho_L^{NT}, b, \alpha, \beta, \eta\}$*, data* $\{L_d, L_{d,-1} \tilde{L}_d^s, w_d, H_d, R_d,\}$*, and bilateral costs* $\{\varsigma_{od,t}^{NT}, \varsigma_{od}^T, d_{od}, \xi_{od}\}$*, there is a unique set of amenities* $\{u_d\}$ *that rationalize the data as an equilibrium of the model.*

Proof. See Appendix A.4 □

[20] $L_{d,-1}$ indicates population in the previous period.

The first part of Theorem 5.1 uses data on wages, population, and employment, to find commuting preferences that satisfy labor market clearing, given commuting equations. Note that population and employment data may come from different sources. As long the nationwide number of workers and population add up, there will be no problem. If one brings, perhaps from a different source, data on the share of workers that live in the same location where they work, we can also match these shares, provided the total workers already implied by this share does not surpass the number of workers we observe in the employment data.[21]

Theorem 5.2 uses labor and labor market clearing in the construction sector plus perfect competition, along with data on floorspace, wages, and sectoral employment, to infer land supply and construction productivity.

Theorem 5.3 leverages labor market clearing in both the tradable and non-tradable sectors, alongside perfect competition, to recover unobserved sectoral productivity using data on floorspace units and prices, wages, and sectoral employment. A sufficient condition is that the maximum income available for spending within a location is at least as large as the wage bill in the non-tradable sector, which we observe from the data. In models without commuting, this condition is trivially satisfied.[22] Once this condition holds, we can recover the total factor productivity (TFP) in the non-tradable sector by ensuring the income spent on workers in the non-tradable sector matches the wage bill implied by the data.

Finally, Theorem 5.4 uses all previous theorems, and hence requires the same data and parameters, plus migration elasticities η and bilateral migration frictions $\xi_{ni,t}$. It relies on the migration equations derived from the model to rationalize the time series of population across counties, $L_{i,t}$.

5.2 Calibration

This section clarifies which parameter values I estimate from the data, and which I calibrate from existing literature.

Externalities. Scale externalities for TFP, δ, are calibrated following updated estimates in the literature (Bartelme et al., 2024). I choose a value of $\delta = 0.1$, which is close to the median across

[21]If one observes counts of workers commuting from origins to destinations, and do not rely on total population and employment data that come from other sources, one can always find a set of origin-destination commuting preferences that rationalize the observed flows as an equilibrium.

[22]For instance, in a model without commuting or a floorspace market, the condition simplifies to $\bar{y}_d L_d \geq \tilde{L}_d^{NT} w_d$, where income and population equal wages and total employment ($\tilde{L}_d^{NT} + \tilde{L}_d^{T}$).

sectors. I estimate amenity externalities following an indirect inference approach, which is explained in detail in the following section.

Commuting parameters. I assume an exponential representation of commute costs following papers in the literature: $d_{ij} = \exp(\kappa\tau_{ij})$. Appendix A.7 estimates the relationship between commute times and commute flows in the short- and long-run. For the model, I assume an average semi-elasticity of $\kappa\theta$=0.0375 which is the average of all the coefficients shown in Appendix A.7.[23] Then, I use estimates of θ from the literature to calibrate θ and κ. In particular, Monte et al. (2018) estimates suggest a value of $\theta = 3.3$. From, there I get a value of $\kappa = 0.0375/3.3 = 0.0114$.

Migration parameters. I calibrate migration costs as done typically in the literature and assume: $\xi_{ij} = \tau_{ij}^{-\varkappa_m}$. For the sake of computing migration costs, I fix 1960 driving times to represent county distances. This assumes highways influence migration only indirectly through prices, commuting access, and amenities, but not directly through reductions in driving times. From 2000–2010 migration data, I estimate a gravity equation and find $\varkappa_m\eta = 1.25$ (see Appendix XXX for details). Following the literature, I set $\eta = 4$. Population growth ζ_t is calibrated directly from the data.

Trade parameters. I calibrate trade costs as $\varsigma_{ji} = B_{s(i),s(j)}\tau_{ji}^{\varkappa_T}$. Using Monte et al. (2018), I set the elasticity of driving times to trade flows as $\varkappa_T\varepsilon = 1.5$, with $\varepsilon = 4$ from Broda and Weinstein (2006). The term $B_{s(i),s(j)}$ representing home bias equals one for intra-state trade and exceeds one for inter-state trade. I calibrate $H_{s(i),s(j)}$ to match 1992 data on average intra-state expenditure from the Freight Analysis Framework.

Floorspace production parameters. Based on Albouy et al., (2018), the share of land in floorspace production, $1 - \beta$, is about 0.4. The elasticity of substitution between land and labor, drawn from Epple et al., (2010), Ahlfeldt and McMillen (2018, 2020), and Combes et al., (2021), ranges from 0.75 to 1.4. I select the value of $\nu = 1.20$ to reflect increased substitution toward land as its price falls.[24]

[23] This paper presents a static model of commuting, despite the dynamic gravity equation indicating differences between short- and long-run elasticities. I am currently developing a paper that explores the implications of dynamic commuting. Since commuting is static in this paper, I simply assume an average elasticity across time. If I instead a calibrate the elasticity in this paper using the value after 40 years ($\theta\kappa = 0.060$), every decade's commuting will be excessively responsive to commuting times.

[24] 1.20 is the middle value that allows for substitution (i.e. above one) and that is within the range of estimates (below 1.4). Note that estimates in the literature concern land and capital, not labor as in this paper. Future versions may include intermediate inputs as a proxy of capital.

Tradable and non-tradable production parameters. For the production of tradables, intensities for labor, floorspace, and intermediate sum to one ($\sum_k \varrho_k^T = 1$). For non-tradables, the only factors are floorspace and labor, and their intensities sum to one ($\sum_k \varrho_k^{NT} = 1$). I assume labor and land intensities of $\varrho_L^T = \varrho_H^T = 0.25$ for tradables, with intermediates at $1 - \varrho_L^T - \varrho_H^T = 0.50$, aligning with input-output data and Valentinyi and Herrendorf (2008).[25] For non-tradables, I assume labor and land intensities of $\varrho_L^{NT} = 2/3$ and $\varrho_L^T = 1/3$, following Valentinyi and Herrendorf.[26] Regarding, the elasticity of substitution between floorspace and labor, I assume a value of $\sigma = 0.3$, the midpoint value in Behrens, et al., (2022) range of 0.2-0.4.

Household consumption parameters. From the Bureau of Labor Statistics Expenditure Surveys, the share of household consumption allocated to land is approximately $1 - \alpha = 1/3$. The intensity of tradables in the final goods bundle, denoted by ρ, is directly derived from the data and calibrated to ensure market clearing in the tradable goods market each decade. I assume an elasticity of substitution between tradables and non-tradables of $\sigma_D = 0.5$, roughly consistent with findings in the literature (e.g. 0.44 in Stockmand and Tesar, 1995; 0.74 in Mendoza, 1991). Finally, intermediate share is chosen to ensure Walras given normalization enforced in the model.

Theorems and unobservables. With these parameters (except for externalities), I apply Theorems 5.1-5.4 to back out unobservables for each county-by-decade. This allows me to match county-level data by decade on population, employment, bilateral commuting, wages, floorspace prices, and floorspace supply. I then use the recovered amenities to perform the indirect inference approach outlined in the next section.

5.3 Indirect inference

The goal of this section is to estimate the amenity externality function, which includes three parameters: the slope of the externality, ϕ, and two location thresholds, γ_0 and γ_1. I utilize an indirect inference approach and map the externality parameters to the reduced-form evidence I reported previously. The primary advantage of employing an indirect inference approach is that it allows us to

[25] See Table 7, Valentinyi and Herrendorf (2008), with re-normalized values excluding non-tradable intermediates, and capital relabeled as floorspace.

[26] See Table 7, Valentinyi and Herrendorf (2008), with re-normalized values excluding intermediates, and capital relabeled as floorspace.

leverage the general equilibrium responses generated by the model and compare them to the observed data. In essence, we use these responses to inform our estimates of the parameters.

To avoid estimating three parameters, I calibrate γ_0 using the average rural population of 3,000, meaning externalities are inactive below this threshold. Using Theorems 5.1-5.4, I recover unobservables for each decade, fix them to 1960 levels, and solve the model under two scenarios: one with no highways and one with highways from the Pershing Plan, reducing commute and trade costs. The underlying assumption is that highway assignment according to the Pershing Plan is as good as random. I re-estimate the reduced-form regressions from Section 3.3 using the heterogeneous scale elasticity formulation for amenities. By comparing the model's moments to the data, I search for the amenity elasticity ϕ and the upper threshold γ_1 that maximizes the match betwen the model's reduced-form regressions and the data's reduced-form evidence. I use the IV estimates as target moments. That is, I target four moments: the trade and commute access effect on population and employment. In other words, I minimize the quadratic loss function:

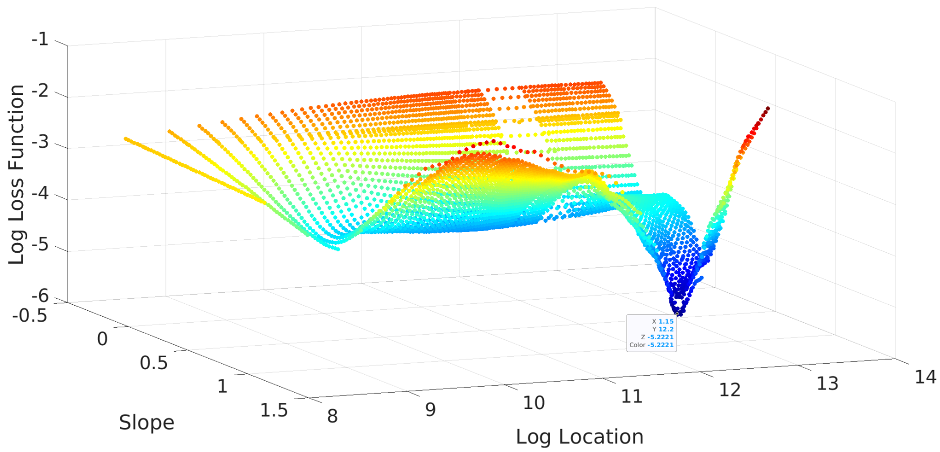
$$Loss = [\beta^{data} - \beta^{model}] I [\beta^{data} - \beta^{model}]' \tag{25}$$

where I is the identity matrix. Panel A of Figure 8 presents the resulting loss function. As the location parameter approaches infinity, the loss function increases, indicating a poorer fit between the model and the data. This serves as evidence against the constant scale formulation. Similarly, as the slope parameter approaches zero, the loss function increases for every location parameter, providing evidence against the no externality formulation.

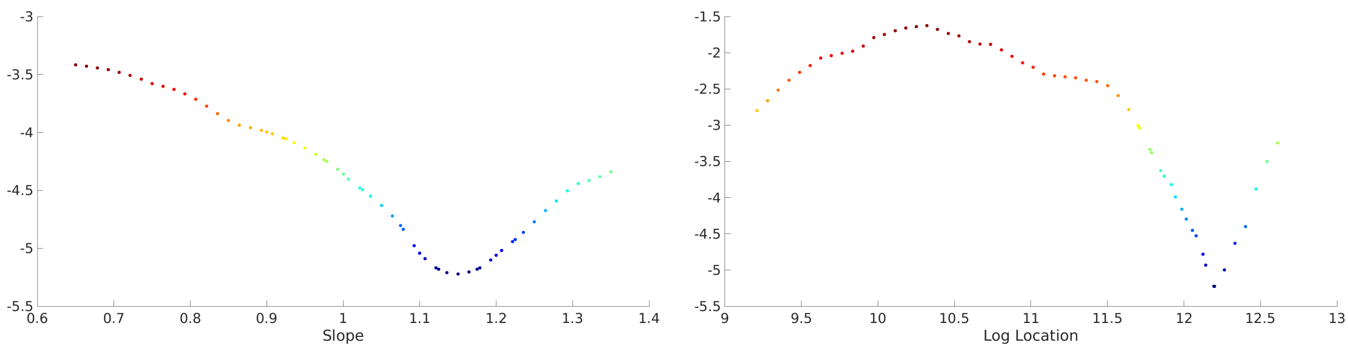
I estimate a slope parameter of $\phi = 1.15$ and a location threshold of $\gamma_1 = 198,790$. Panel B of Figure 8 shows sharp identification, especially for the location parameter. The estimated effects are: trade access increases population by 0.20951 (data = 0.20) and employment by 0.17462 (data = 0.24), while commute access decreases population by -0.10729 (data = -0.08) and employment by -0.096923 (data = -0.08). The model closely matches population changes but underestimates trade access effects on employment.

Figure 8: Loss Function

Panel A: Log Loss vs Slope and Log Location



Panel B: Fixing one parameter to its estimated value while varying the remaining one



Regarding the magnitude of these parameter estimates, some comments are in order. While the slope is larger than typical values in the literature, it remains reasonable for two reasons. First, in contrast to constant scale elasticities—where large locations experience unbounded improvements as externalities never diminish—heterogeneous scale elasticities limit the amplification of externalities to the range between γ_0 and γ_1. This ensures that beyond γ_1, growth stabilizes, yielding a unique steady state in partial equilibrium for large locations. Second, some studies assume negative amenity externalities to capture congestion effects associated with fixed land. While my model includes variable floorspace, land remains fixed, which introduces an endogenous congestion force. My calibration already accounts for these features, allowing the estimated amenity externalities to be higher, and even positive, as seen here. Moreover, my calibration also considers imperfect myopic migration given the Frechet assumption for migration preferences, where the shape parameter acts as another dispersion force.

With regards the location parameters, their value highlights how congestion offsets amenity externalities in the largest counties, and the lack of critical mass in the smallest counties. Approximately 1,800 out of 3,100 counties lie within the interval $\gamma_0 = 3,000$ to $\gamma_1 = 198,790$, with the 35 largest counties falling outside the upper threshold. This suggests that in these large counties, congestion forces are strong enough to offset any positive amenity externalities, which is a plausible finding. Importantly, this does not imply that living in large cities offers no advantages. My calibration assumes TFP externalities follow a constant scale formulation, meaning productivity gains from agglomeration in large cities remain intact.

6 Counterfactuals: Unpacking Suburban Growth

In this section, I use the model calibrated to the U.S. economy, matching county-level data on population, employment, commuting flows, wages, floorspace prices, and floorspace supply from 1950 to 2020. I first simulate an economy without highways and compare it to the simulated economy with highways (matching observed data). This allows me to decompose the overall impact of highways. Furthermore, I isolate the commute effect by introducing highways but allowing only reductions in commuting costs. Similarly, the trade effect isolates the impact of highways by allowing only trade cost reductions. The difference between the total effect and the sum of the trade and commute effects is what I refer to as the interaction effect.

Aggregate effects. Table 1 shows my main findings. The table is divided into two panels: Panel A reports changes in population share by NCHS classification, while Panel B shows changes in employment share. Each row presents changes in percentage points. The first row in each panel shows the observed change in shares between 1960 and 2020 (Data Δ). The second row displays the total effect of highways by comparing the simulated economy with and without highways. The next two rows decompose this total effect into the commute effect (only reduced commuting costs) and the trade effect (only reduced trade costs). Finally, the last row shows the interaction effect, which captures the additional impact when both commuting and trade cost reductions are introduced simultaneously. Each value represents the change in share for the respective category (Core, Suburbs, Medium, and Small/Rural areas).

Highways account for 0.94 percentage points of the 6.14 percentage point increase in population share attributed to suburbs (15.3%), while they explain -0.88 percentage points of the -3.33 percentage point decline in core areas (26.6%). In the data, medium cities grew by 3.27 percentage points, while small and rural areas shrank by 6.10 percentage points. In contrast, the model predicts medium cities shrinking by 0.60 percentage points and small and rural places growing by 0.53 percentage points.

Table 1: The effect of Interstate Highways on Suburbanization

Panel A: Population share by NCHS classification				
	Core	**Suburbs**	**Medium**	**Small/Rural**
Data Δ (2020-1960)	-0.0333	0.0614	0.0327	-0.0610
Total Effect (hwy - no hwy)	-0.0088	0.0094	-0.0060	0.0053
Commute Effect	0.0013	0.0083	0.0036	-0.0131
Trade Effect	-0.0100	-0.0035	-0.0125	0.0260
Interaction Effect	-0.0001	0.0047	0.0030	-0.0075
Panel B: Employment share by NCHS classification				
	Core	**Suburbs**	**Medium**	**Small/Rural**
Data Δ (2020-1960)	-0.0157	0.0504	0.0319	-0.0669
Total Effect (hwy - no hwy)	-0.0027	0.0030	-0.0041	0.0040
Commute Effect	0.0033	0.0054	0.0019	-0.0103
Trade Effect	-0.0074	-0.0047	-0.0086	0.0206
Interaction Effect	0.0015	0.0022	0.0026	-0.0063

I unpack the growth driven by highways across three terms: commute-only, trade-only, and interaction effects. The reduction in commute costs contributed to the relative decline of small and rural areas compared to medium cities, suburbs, and core counties, with suburbs experiencing the most growth. Regarding trade access, the reduction in trade costs shifted populations away from core and medium cities toward small and rural areas. While suburbs also contracted, they did so to a lesser extent than cores, leading to suburbanization: the decline of cores and the rise of suburbs.

The interaction effect shows that simultaneous reductions in trade and commute costs further enhanced suburban growth, particularly at the expense of rural areas. Notably, the interaction effect accounts for 50% (0.0047/0.0094) of the total increase in suburbs, underscoring the significance of trade in the suburbanization process, which is characterized by the decline of cores and the rise of suburbs.

For employment, as shown in Panel B, highways similarly explain about 17.2% of the decline in core areas and 6% of the rise in suburbs. Qualitatvely the commute, trade and interaction effects show similar patterns.

Understanding aggregate and reduced-form results. While the reduced-form analysis shows null impacts for commute access and positive effects for trade access, counterfactual simulations reveal that both commute and trade costs contribute to suburbanization. This apparent discrepancy arises because reduced-form estimates capture the *average effects* of increasing each of access measures, while counterfactuals allow us to examine the individual impacts of reducing trade and commute costs *on each county*. Since, the population share in the suburbs for a given period is defined as $\text{Suburbs Share} = \sum_{i \in \text{Suburb}} L_{it} / \sum_{i \in \text{All}} L_{it}$, the effect on suburbanization for a specific access variable can be expressed as:

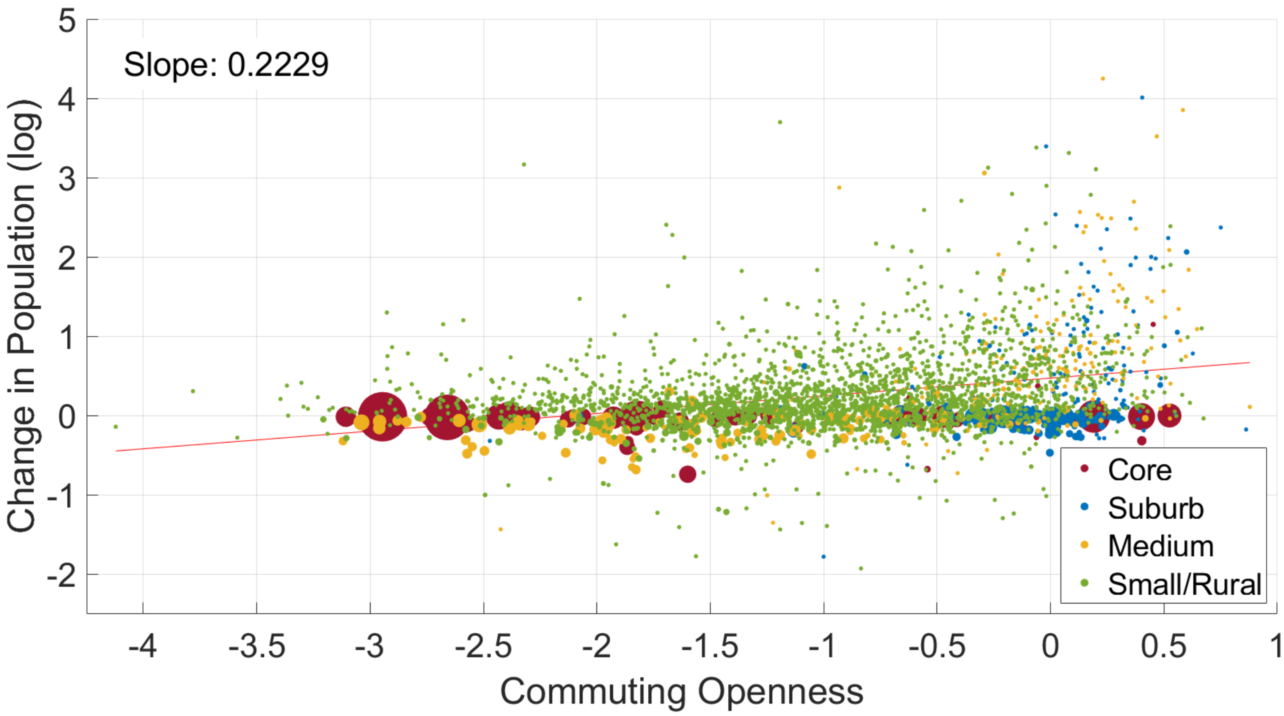
$$\begin{aligned} \Delta \text{Suburbs Share} &= \frac{\sum_{i \in \text{Suburb}} \Delta L_{it}}{\sum_{i \in \text{All}} L_{it}} - \frac{\sum_{i \in \text{Suburb}} L_{it}}{\sum_{i \in \text{All}} L_{it}} \times \frac{\sum_{i \in \text{All}} \Delta L_{it}}{\sum_{i \in \text{All}} L_{it}} \\ & \frac{\sum_{i \in \text{Suburb}} \text{Effect}_i \Delta \text{Access}_i}{\sum_{i \in \text{All}} L_{it}} - \frac{\sum_{i \in \text{Suburb}} L_{it}}{\sum_{i \in \text{All}} L_{it}} \times \frac{\sum_{i \in \text{All}} \text{Effect}_i \Delta \text{Access}_i}{\sum_{i \in \text{All}} L_{it}} \end{aligned} \tag{26}$$

In this context, I have estimated the average effect $\mathbf{E}[\text{Effect}_i]$ in the reduced form, which I used to calibrate the model through the indirect inference approach. In other words, I use the average trade and commute access effects as target moments. However, by performing the counterfactual exercise, we implicitly construct the objects Effect_i for each county, and from there compute the aggregate effects. The intuition of how the model implicitly constructs Effect_i is in Section 4.7.

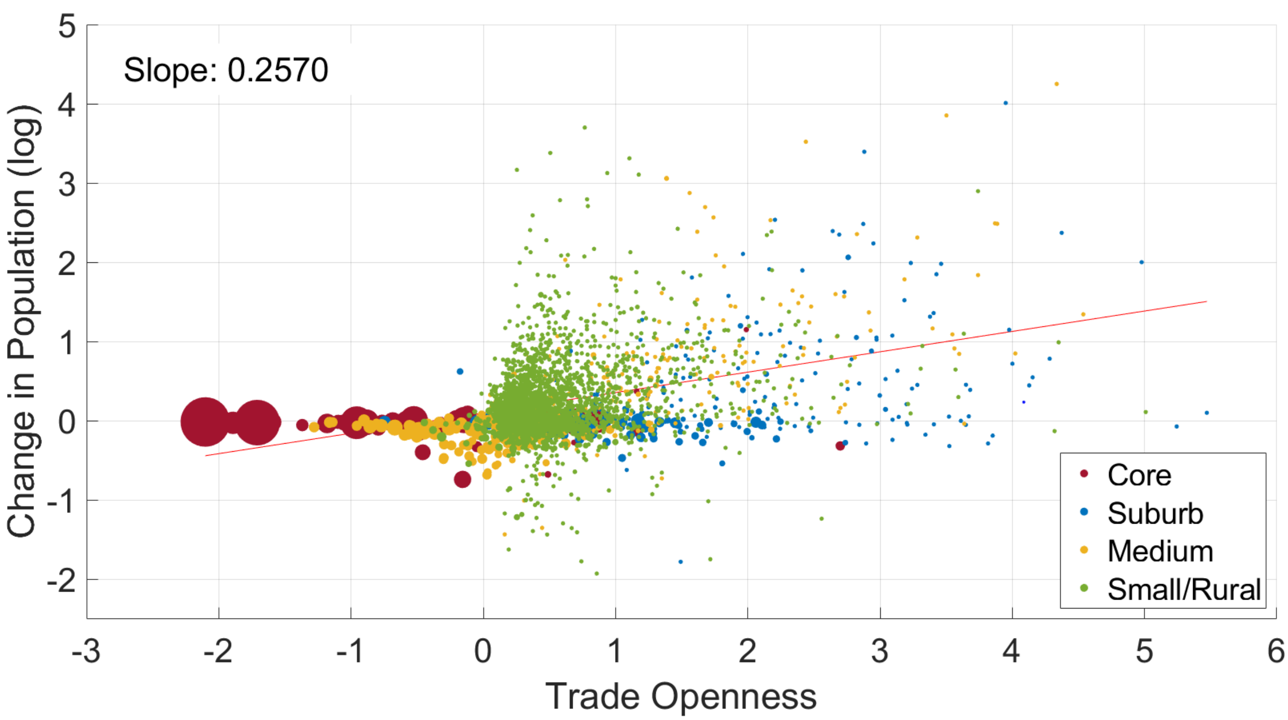
Heterogeneity across counties. My results reveal substantial heterogeneity in population growth across counties, measured as the log difference between a world with highways and one without. Figure 9 illustrates this: Panel A plots population growth against commuting openness, defined as the log of worker inflows plus outflows relative to total employment. Panel B plots population growth against trade openness, measured as the log of 'imports' from other counties plus 'exports' to other counties relative to total sales. Dots are color-coded by NCHS classification (core, suburb, medium, or small/rural).

Figure 9: Relationship between growth in population and in job

Panel A: Initial commuting openness and population growth



Panel B: Initial trade openness and population growth

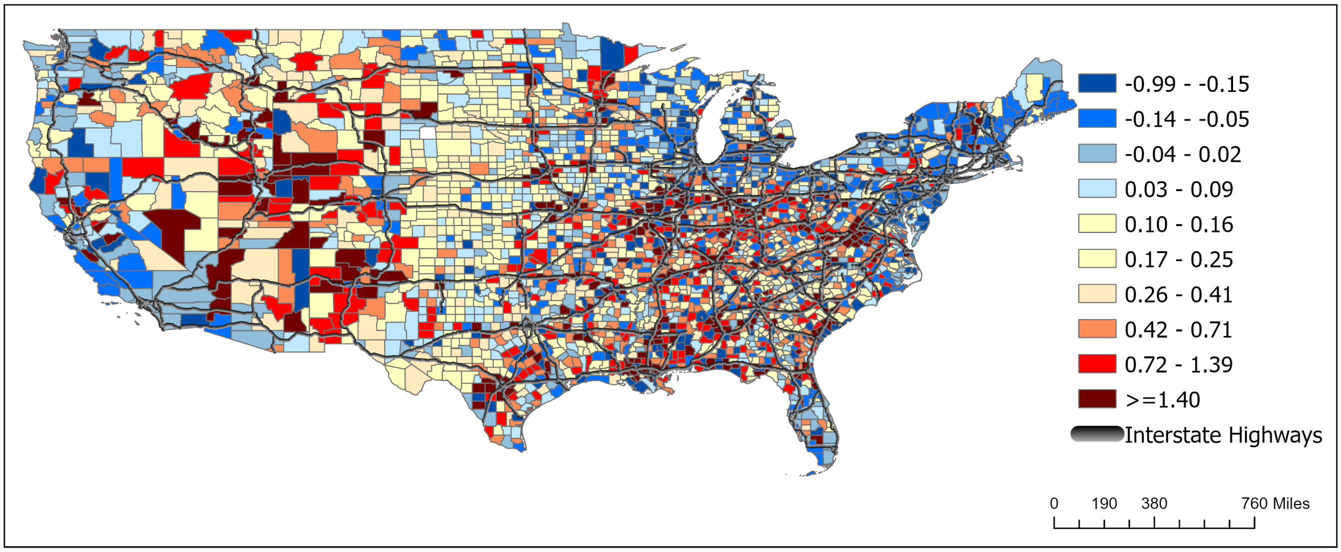


This figure allows us to better understand the aggregate results in Table 1. First, counties more open to commuting or trade benefit more from reductions in these costs, consistent with the comparative statics derived in section 4.7. This is confirmed by the positive slopes shown in both panels. Second, suburbs exhibit particularly high commuting openness, with a steeper regression line for suburban counties (not plotted) compared to the average regression line (plotted), suggesting that reductions in commute costs drive stronger population growth in these areas. Third, for trade, openness is higher in suburbs and small/rural areas, while medium and core metros show less initial openness. This pattern aligns with Table 1, where trade cost reductions led to decreases in the share of the population accounted by these areas.

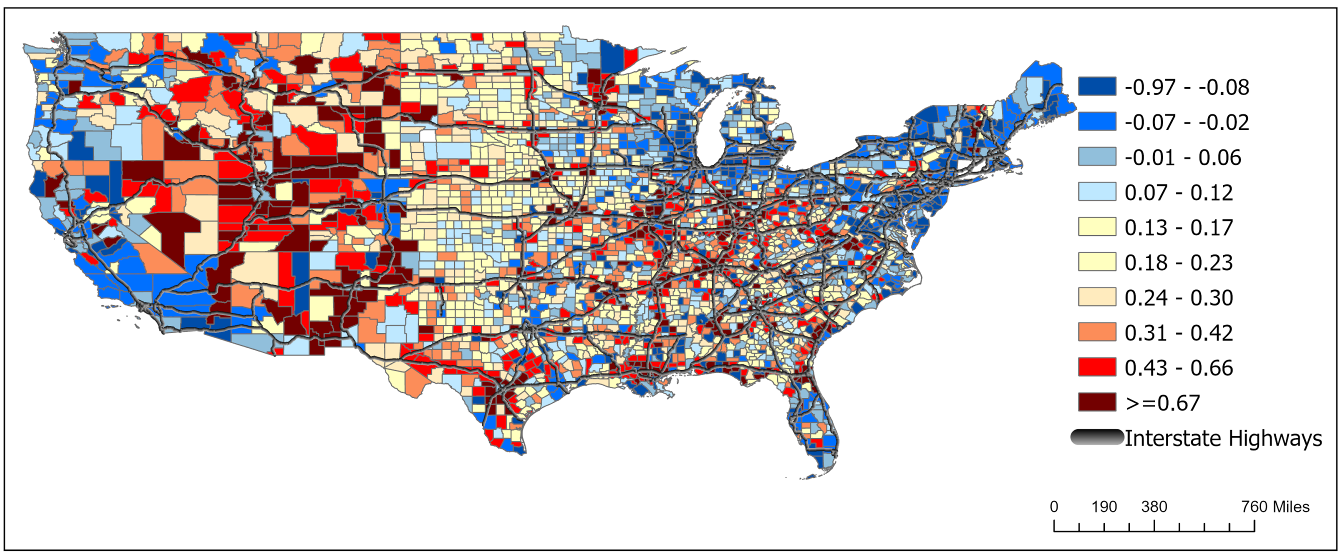
Figure 10 further highlights the heterogeneity in population and employment growth across U.S. counties. First, the maps reveal that growth disproportionately occurred inland, away from the coasts. Since many of the largest counties are located along the coasts, this suggests a trend toward decentralization. Second, while population and employment growth are correlated, they do not align perfectly. For example, the employment decline was more pronounced in coastal areas, particularly in the southwest (e.g., Los Angeles), than population decline. Conversely, the central west region experienced stronger employment growth relative to population growth.

Figure 10: Population and employment growth across counties

Panel A: Population growth (highways - no highways)



Panel B: Employment growth (highways - no highways)



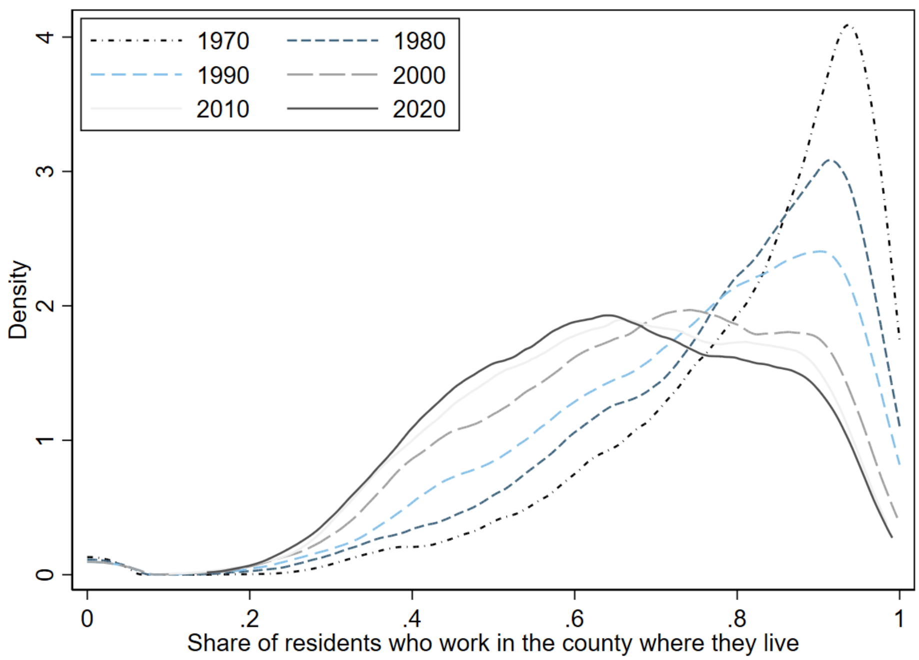
7 Conclusions

TBD

A Appendix

A.1 Figures

Figure A.1: Distribution across counties of the share of people working on the same county they live

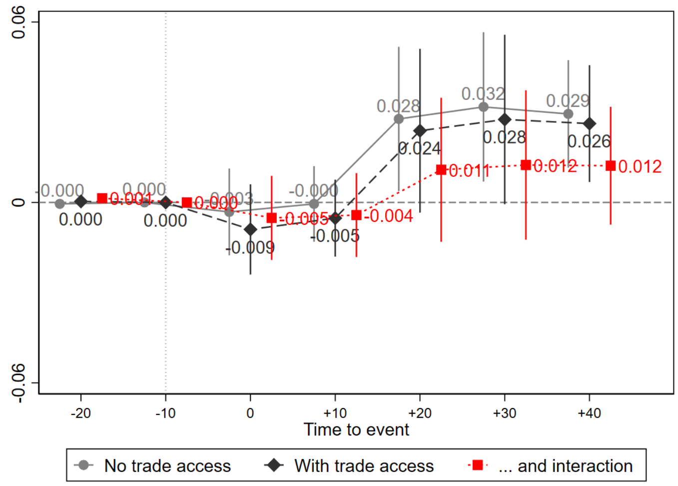


To make sure counties definition over time are consistent, I interpolate commuting flows using Eckert et al. (2020) crosswalks. Commuting flow data comes from the Journey To Work Database (1970-2020).

Figure A.2: OLS estimates of the effect of commuting and trade access

Panel A: Commute Access

Log Employment - Log Population



Panel B: Trade Access

Log Employment - Log Population

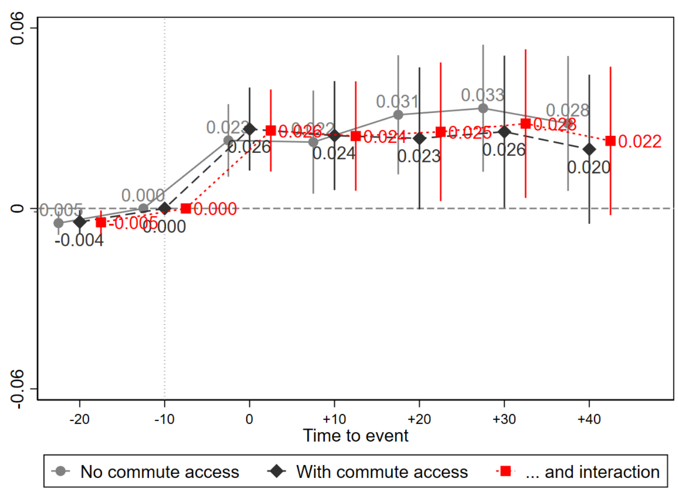
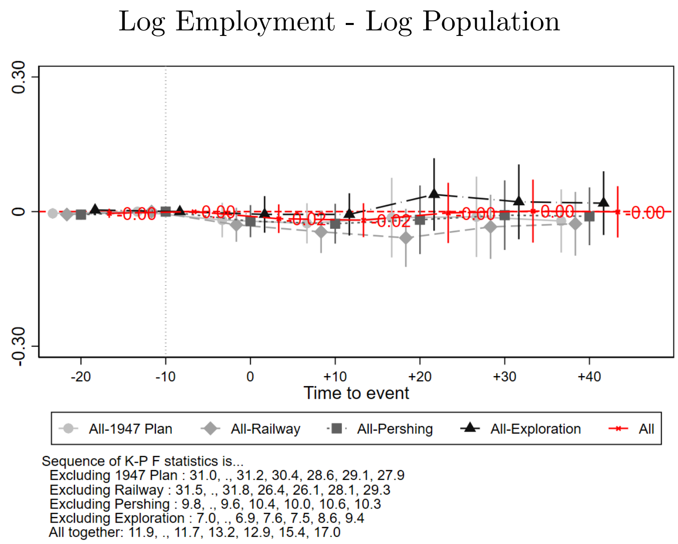


Figure A.3: 2SLS estimates of the effect of commuting and trade access

Panel A: Commute Access



Panel B: Trade Access

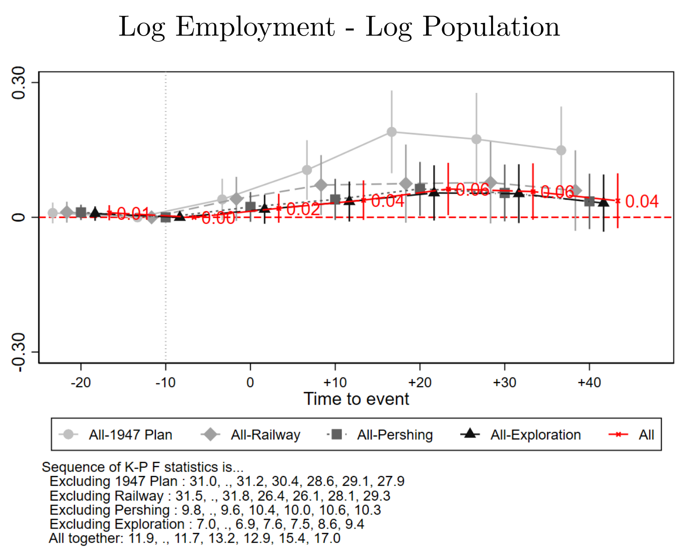
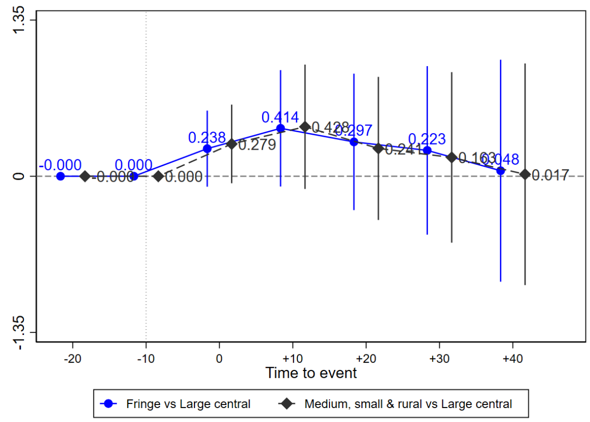


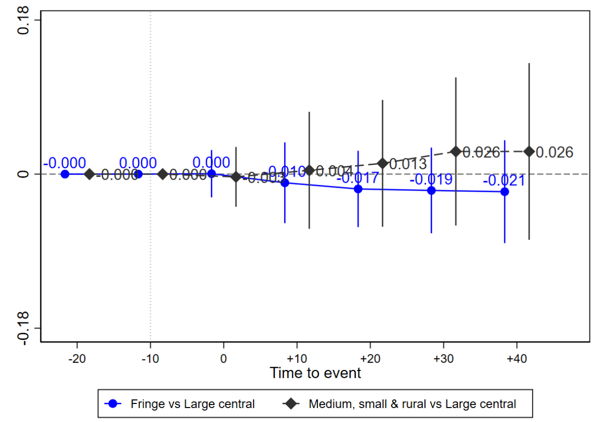
Figure A.4: OLS estimates of the effect of commute and trade access on log population by commuting zone and county NCHS classification

Panel A: Top 7 commuting zones

Commute

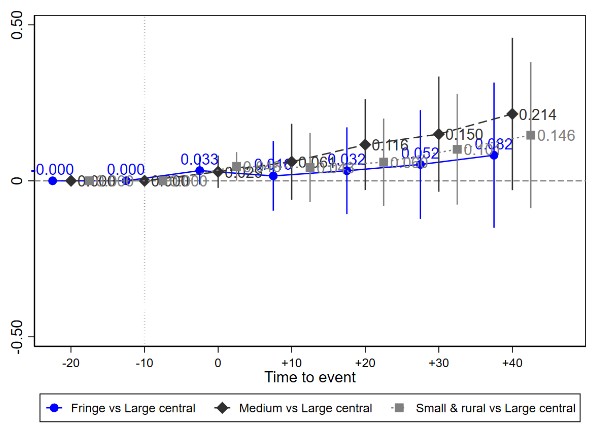


Trade

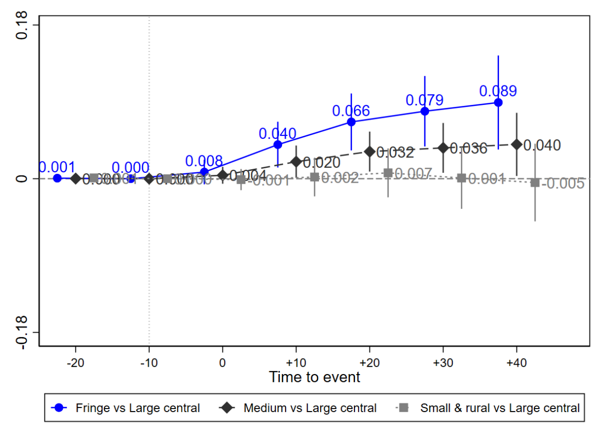


Panel B: Top 8-60 commuting zones

Commute

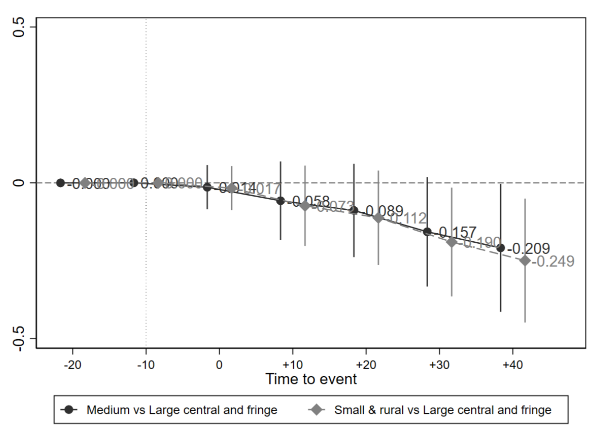


Trade



Panel C: Remaining commuting zones

Commute



Trade

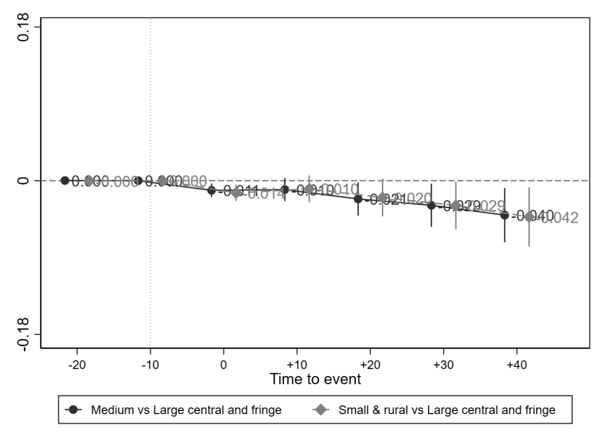
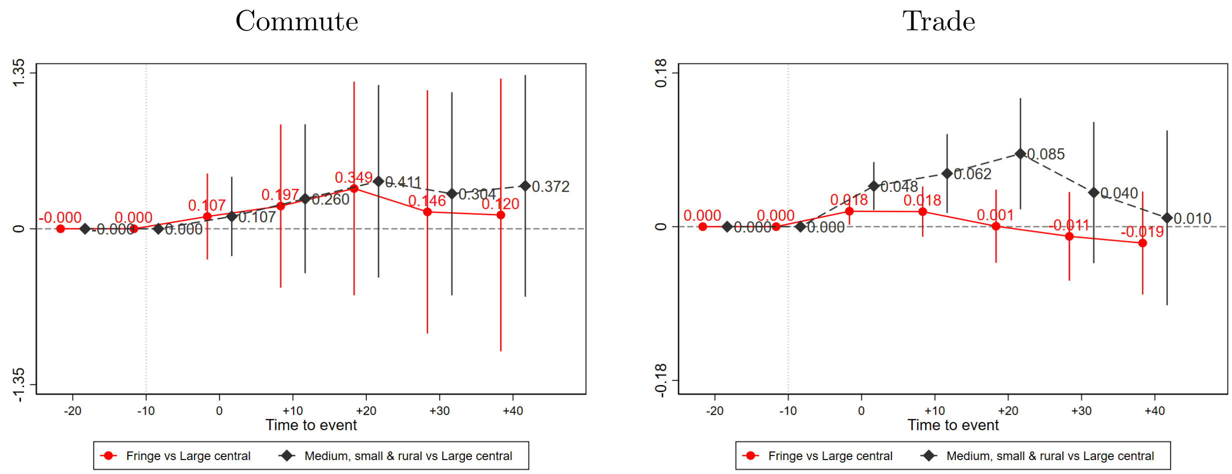
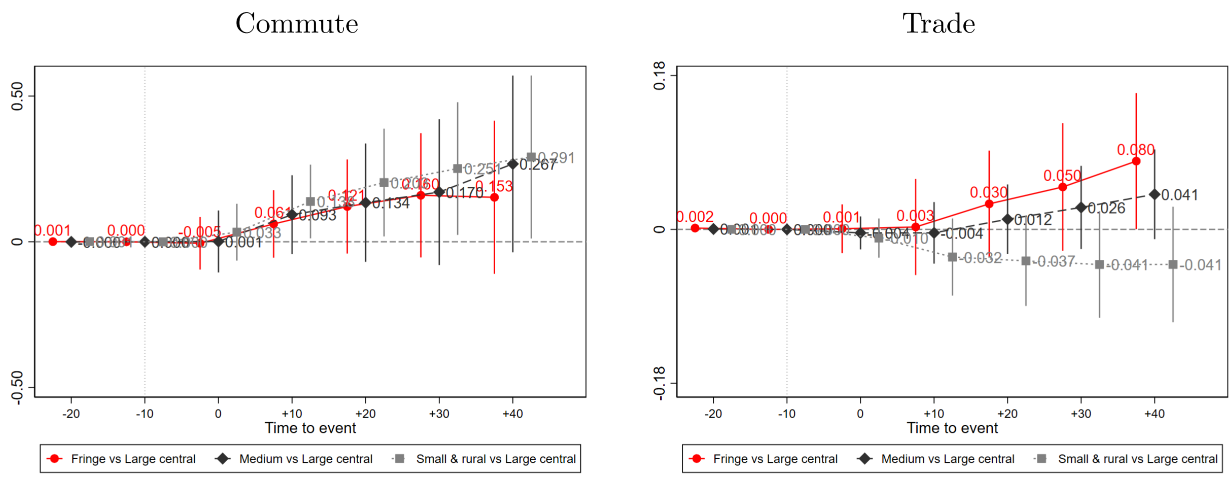


Figure A.5: OLS estimates of the effect of commute and trade access on log employment by commuting zone and county NCHS classification

Panel A: Top 7 commuting zones



Panel B: Top 8-60 commuting zones



Panel C: Remaining commuting zones

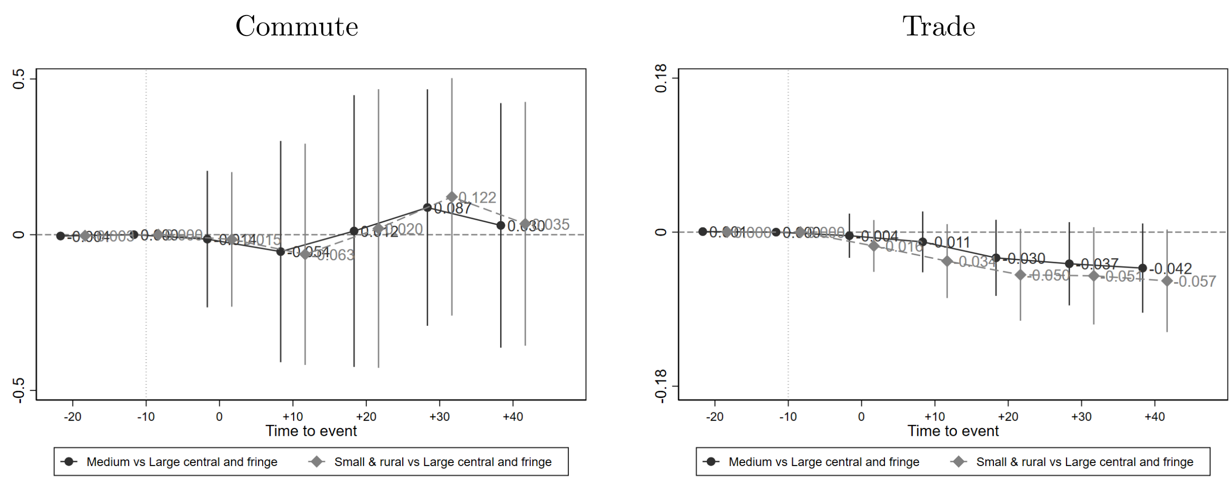
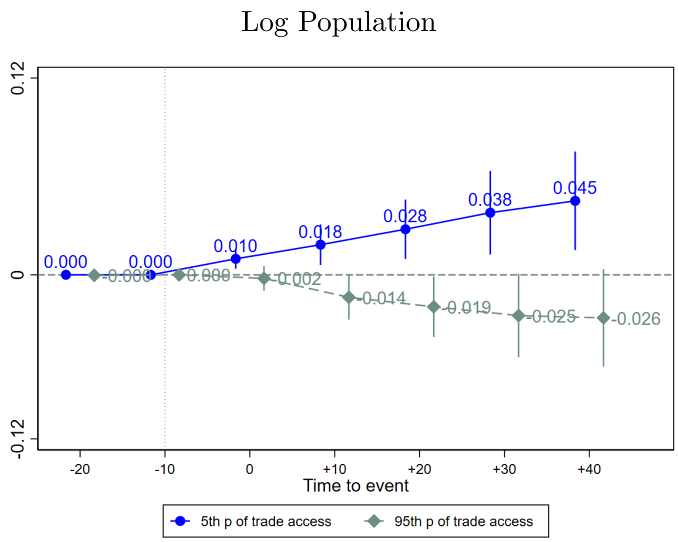
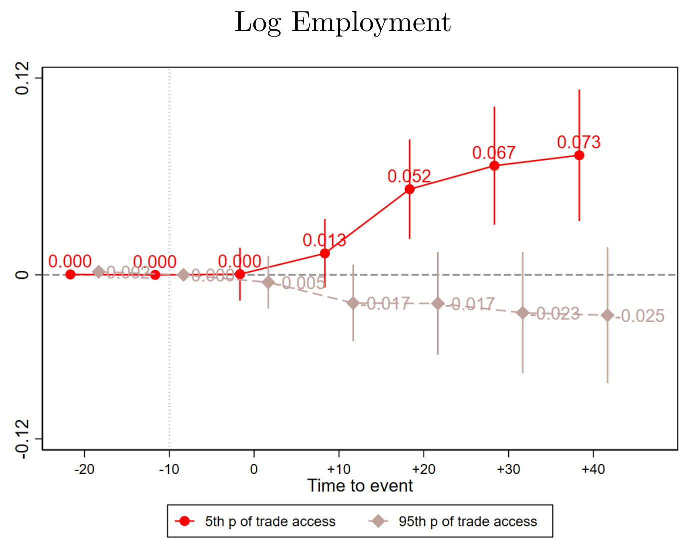


Figure A.6: OLS estimates of the effect of commuting and trade access and their interaction

Panel A: Commute Access





Panel B: Trade Access

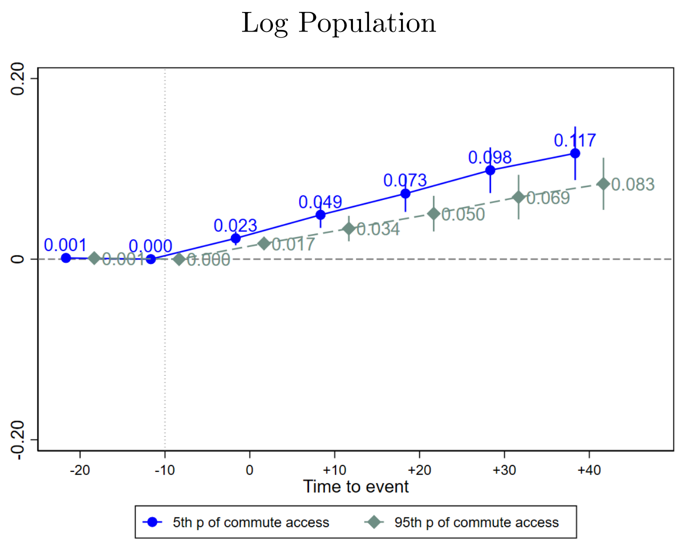
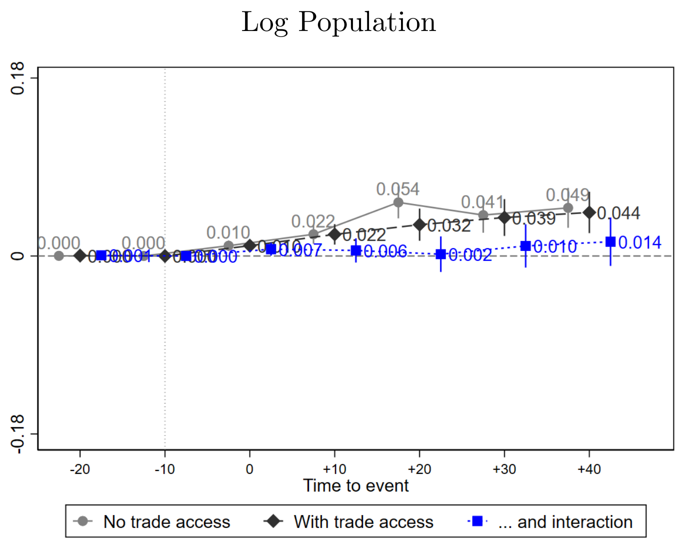
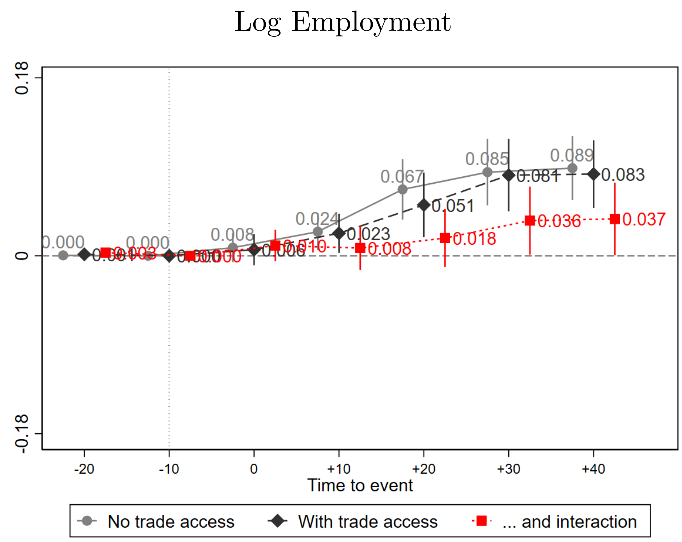


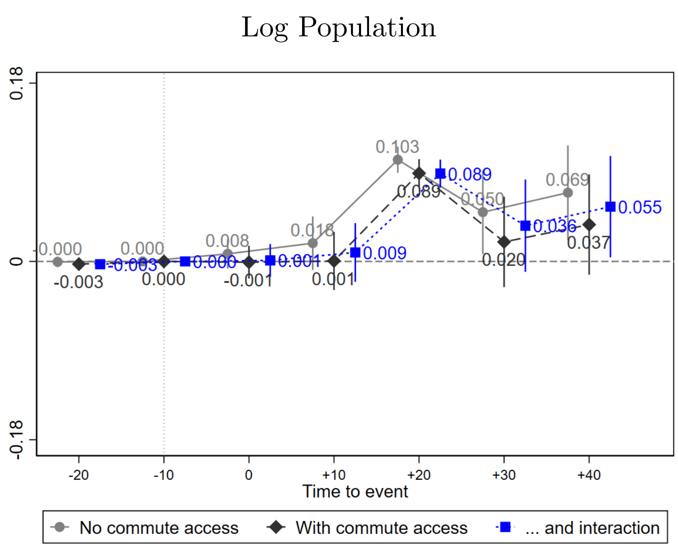
Figure A.7: OLS estimates of the effect of commuting and trade access conditioning on CZ times year-by-wave FE

Panel A: Commute Access





Panel B: Trade Access



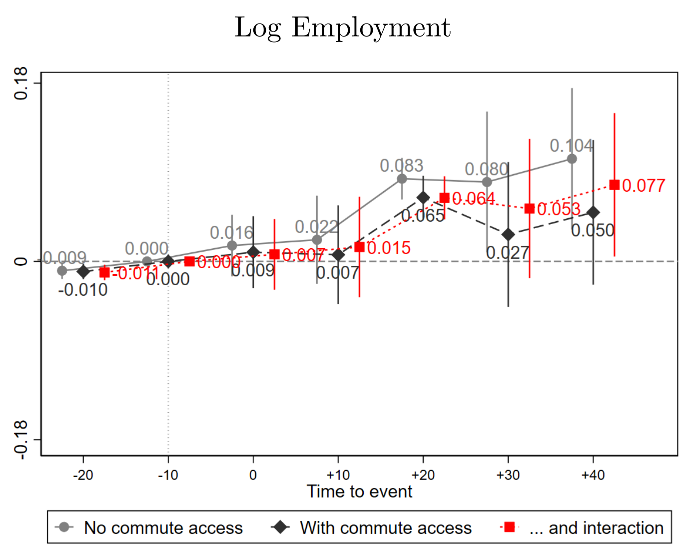
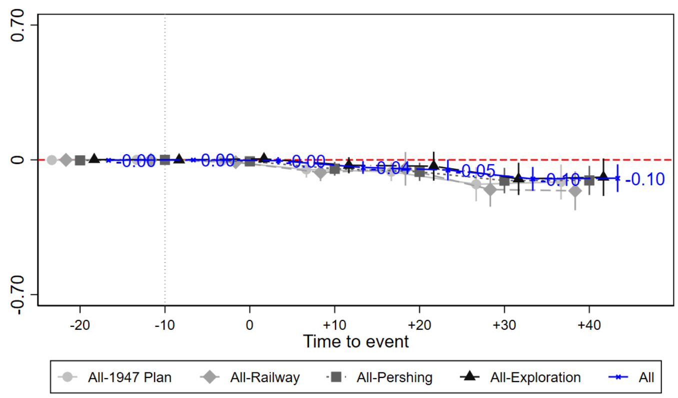


Figure A.8: 2SLS estimates of the effect of commuting and trade access conditioning on CZ times year-by-wave FE

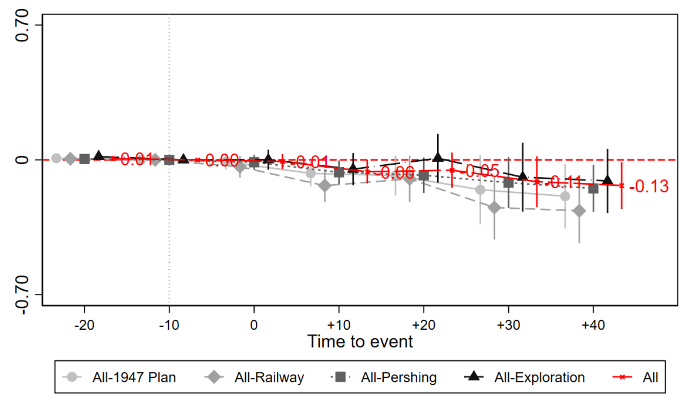
Panel A: Commute Access

Log Population



Sequence of K-P F statistics is...
Excluding 1947 Plan : 13.4, ., 13.2, 11.3, 28.6, 9.7, 9.8
Excluding Railway : 19.9, ., 18.5, 15.7, 26.1, 13.8, 14.3
Excluding Pershing : 12.1, ., 11.0, 11.1, 10.0, 9.8, 9.2
Excluding Exploration : 5.7, ., 5.5, 6.7, 7.5, 6.3, 6.6
All together: 21.3, ., 20.0, 18.9, 12.9, 16.2, 16.0

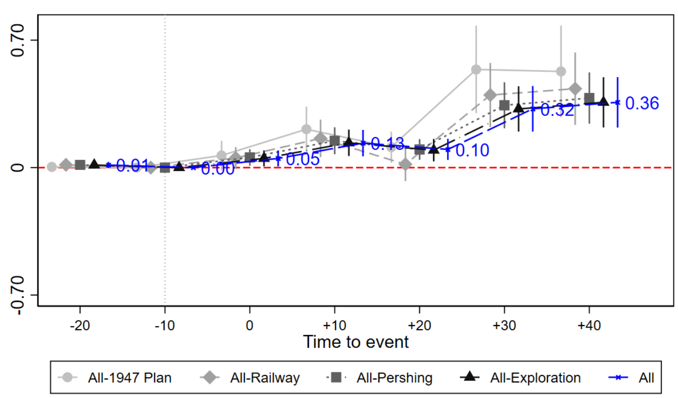
Log Employment



Sequence of K-P F statistics is...
Excluding 1947 Plan : 13.9, ., 13.8, 11.9, 27.1, 10.2, 10.3
Excluding Railway : 17.3, ., 17.0, 14.7, 25.9, 13.0, 12.9
Excluding Pershing : 10.9, ., 10.6, 10.4, 10.8, 9.3, 8.7
Excluding Exploration : 5.1, ., 5.2, 6.3, 7.8, 6.0, 6.2
All together: 19.2, ., 19.0, 17.8, 12.9, 15.5, 15.0

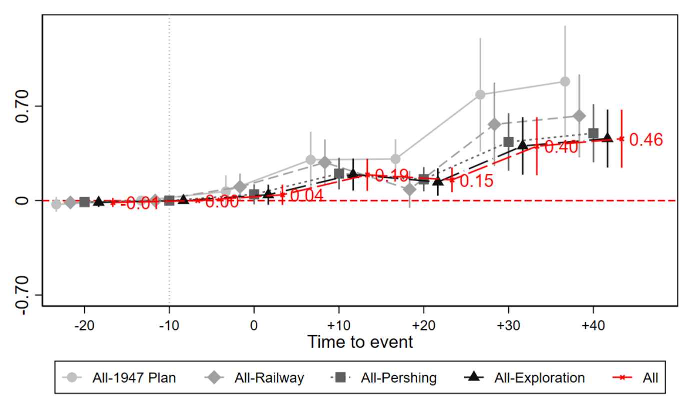
Panel B: Trade Access

Log Population



Sequence of K-P F statistics is...
Excluding 1947 Plan : 13.4, ., 13.2, 11.3, 28.6, 9.7, 9.8
Excluding Railway : 19.9, ., 18.5, 15.7, 26.1, 13.8, 14.3
Excluding Pershing : 12.1, ., 11.0, 11.1, 10.0, 9.8, 9.2
Excluding Exploration : 5.7, ., 5.5, 6.7, 7.5, 6.3, 6.6
All together: 21.3, ., 20.0, 18.9, 12.9, 16.2, 16.0

Log Employment

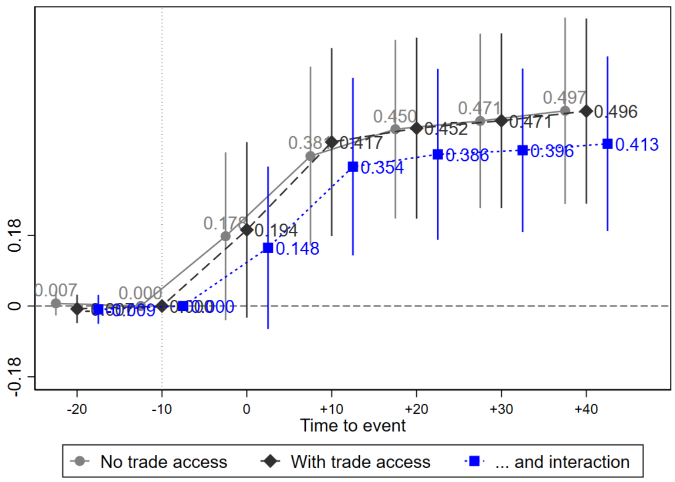


Sequence of K-P F statistics is...
Excluding 1947 Plan : 13.9, ., 13.8, 11.9, 27.1, 10.2, 10.3
Excluding Railway : 17.3, ., 17.0, 14.7, 25.9, 13.0, 12.9
Excluding Pershing : 10.9, ., 10.6, 10.4, 10.8, 9.3, 8.7
Excluding Exploration : 5.1, ., 5.2, 6.3, 7.8, 6.0, 6.2
All together: 19.2, ., 19.0, 17.8, 12.9, 15.5, 15.0

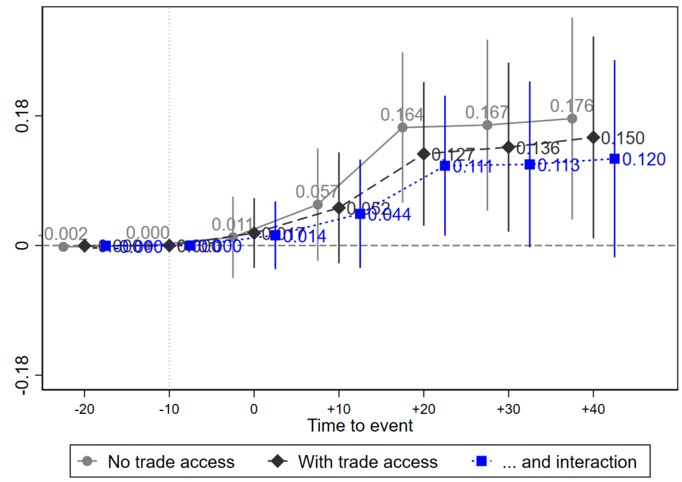
Figure A.9: OLS estimates of the effect of commuting and trade access on census-tracts' population

Panel A: Commute Access

Log Population (unweighted)

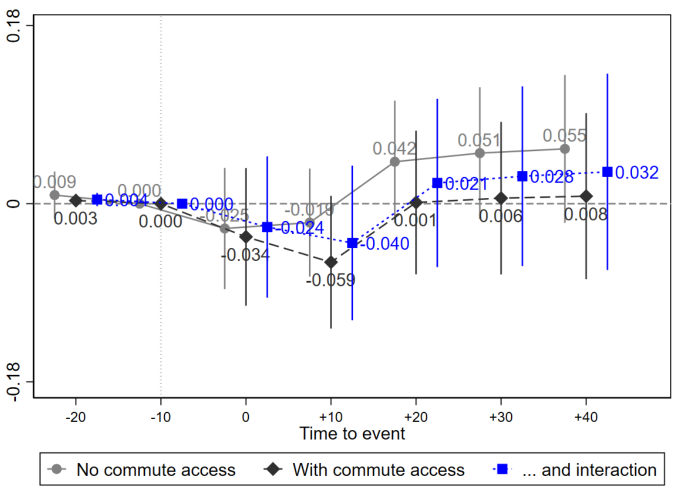


Log Population (weighted)

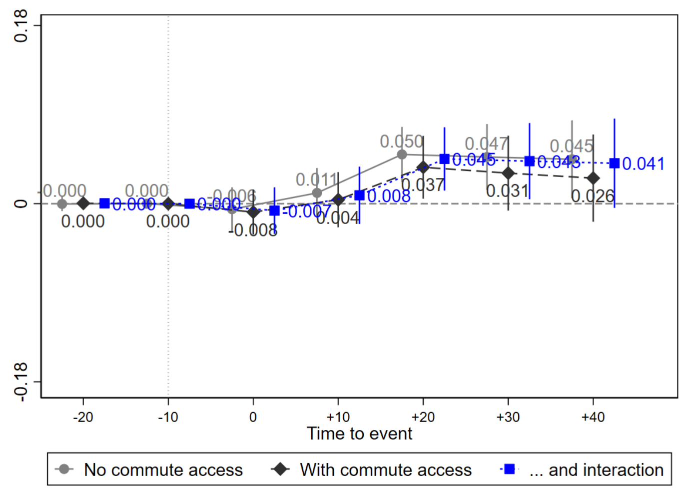


Panel A: Trade Access

Log Population (unweighted)



Log Population (weighted)



A.2 Tables

Table A.1: Commute and Trade Access Effects on Log Population after 40 years

	(1)	(2)	(3)	(4)
	Log Population	Log Population	Log Population	Log Population
Commute Access (C)	0.0588***		0.0408***	0.00235
	(0.0121)		(0.0119)	(0.0140)
Trade Access (G)		0.109***	0.0884***	0.0958***
		(0.0134)	(0.0147)	(0.0146)
Interaction Term (I)				-0.00192***
				(0.000330)
Observations	13336	13258	12248	12248

Notes: Standard errors clustered at the county-level in parentheses. * p¡0.10, ** p¡0.05, *** p¡0.01

Table A.2: Commute and Trade Access Effects on Log Employment after 40 years

	(1)	(2)	(3)	(4)
	Log Employment	Log Employment	Log Employment	Log Employment
Commute Access (C)	0.0882***		0.0671***	0.0145
	(0.0165)		(0.0172)	(0.0190)
Trade Access (G)		0.137***	0.108***	0.118***
		(0.0195)	(0.0211)	(0.0209)
Interaction Term (I)				-0.00263***
				(0.000474)
Observations	13336	13258	12248	12248

Notes: Standard errors clustered at the county-level in parentheses. * p¡0.10, ** p¡0.05, *** p¡0.01

A.3 Solution algorithm

A.4 Data Appendix

A.5 Other results

Heterogeneity across commuting zones

Facts 1 and 2 in Section 2.2 established that suburbs, especially in mid-sized commuting zones, emerged as key employment hubs. Section 3.3 established that the average effect of trade access is positive, whereas the average effect of commute access is statistically not distinguishable from zero. This section explores the intersection of these two ideas. That is, it analyzes how commuting and trade access affect peripheral counties compared to central ones, across commuting zones of different sizes.

How? I restrict the sample by commuting zone size. Then, I re-run regressions, interacting commute and trade access with indicators for large central areas (i.e. core areas), suburbs, medium metros, and small metros pooled with rural areas (following the NCHS classification). In most cases, I choose large central counties as base category. Thus, coefficients are interpreted in relative terms.

Figure A.4 presents population results. Panel A focuses on the top 7 commuting zones (LA, NY, Chicago, Houston, SF, NJ, Boston).[27] The findings reveal that, after 40 years, commuting and trade access effects on smaller areas, relative to large central counties, are statistically insignificant. Employment results, shown in Figure A.5, follow a similar pattern.[28]

Panel B examines the top 8-60 commuting zones. The key finding is that trade access drives population and employment decentralization, while commuting access plays a smaller role. First, regarding trade access, Figure A.4 shows that it significantly affects population in fringe counties compared to large central counties. The same is true for medium metros relative to large central counties, though to a lesser extent. Figure A.5 show similar patterns for employment. Second, regarding commute access, it also has positive relative impacts but these are statistically insignificant[29]. Overall, these findings support Facts 1 and 2: increased trade access has made suburbs and intermediate cities important employment hubs.

Panel C focuses on the remaining commuting zones. For this panel, I group counties classified as large central and fringe into a single category, again because of small sample considerations. I use this category as base category.[30] For population, both commute and trade access show negative coefficients across all areas, meaning that rises in commute and trade access have *larger* impacts on fringe counties relative to medium metros (and fringe counties relative to smaller metros and rural areas). For employment, only trade effects are statistically significant. In other words, both commute

[27] I group medium, small, and rural counties to ensure a larger sample. Within the biggest commuting zones there are just a few counties classified as either medium, small or rural.

[28] However, in the short-run (0-10 years), only commute access has relative impacts for population, while only trade access has relative impacts for employment.

[29] Although they are statistically significant for employment as per Figure A.5.

[30] This is because within the smallest commuting zones there are just a few counties classified as either large central or fringe.

and trade access increase population density in the smallest commuting zones, though trade access only does so for employment.

In summary, suburbs have become major employment hubs due to trade access. In large commuting zones, decentralization produced by commute and trade effects is not persistent. However in mid-sized commuting zones decentralization is persistent. In the smallest commuting zones, commute access concentrates population and employment towards higher-density areas, while trade access only does so for employment.[31]

Interaction. In previous sections, I introduced the interaction between commute and trade access and treated it like a control, while also computing the average effect of either commute or trade access. In this section, I explore the sign and magnitude of this interaction term with greater detail.

Theoretically, when trade and commute access both improve simultaneously, the interaction effect can be negative, meaning the two channels substitute for each other. For instance, if trade access improves, making it easier for firms to move goods across regions, but commute access also increases, workers may prefer to commute to jobs in other regions instead of working locally. This weakens the local employment benefits that improved trade access would otherwise provide.

Figure A.6 shows the commute and trade effect for population and employment. I compute the effect of increasing commute access assuming that trade access increases at the 5th percentile and 95th percentile to illustrate the interaction's magnitude. The reverse is done for trade access. Since commute access effects decrease with trade access, there is a negative interaction. When trade access is low enough, commute access becomes positive and statistically significant, even when the average impact of commute access is zero in Figure 6. In contrast, the impact of trade access remains positive, even with high commute access changes.

Controlling for commuting zones fixed effects interacted with year-by-wave fixed effects. In the baseline, I control for state fixed effects interacted with year-by-wave fixed effects. Identification comes from a comparison of access measures between counties, even when they belong to different commuting zones but within the same state.

An interesting comparison would be between counties with highways leading to different connections, but within the same commuting zone. For example, from the point of view of a city like Baltimore

[31] 2SLS results are available in Appendix, but conclusions are similar despite weak first-stage results in some subsamples.

or Washington DC, finding more economic activity around a highway leading to Manhattan, NY, versus one leading to Charlottesville, VA would highlight the trade access effects of highways. To do a comparison similar to this example, I introduce commuting zones fixed effects interacted with year-by-wave fixed effects (and lose the state fixed effects). The commuting zone fixed effects interacted with year-by-wave fixed effects will be perfectly colinear with the access measures of commuting zones comprised by a single county. Thus, identification comes from comparing access measures between two or more counties within the same commuting zone.

OLS results are available in Appendix A.7. Overall results are similar to the baseline, but with a few nuances. First, commuting effects are positive and statistically significant for employment: a one standard deviation increase in commuting access rises employment by 3.7% (in Figure 6 the effect was of 1.5% but statistically, zero). For population, the effect is still indistinguishable from zero. Second, trade effects are almost halved: increasing trade access rises population by 5.5% (instead of 9.6%) and employment by 7.7% (instead of 11.8%).

While OLS estimates give commuting access a bigger role than in the baseline, the opposite is true for 2SLS estimates. Estimating the effect of trade and commute access by 2SLS, while also controlling for commuting zone times year-by-wave fixed effects strengthens the impact of trade (see Appendix A.8). Trade effects nearly double, with 40-year coefficients for population and employment increasing to 0.36 and 0.46, respectively (rather than 0.20 and 0.24). Commuting effects remain negative but they are now statistically significant.

Census tracts. This paper examines whether gains in commuting and trade access from highways explain the suburbanization of population and employment at the county level. To explore suburbanization within counties, that is, at the census tract level, I construct a panel of census tracts from 1950 to 2020. Only population data is available and not employment.[32] I rerun the regression using census-tract population as the outcome and introduce census tract fixed effects to trace out how population changes in response to a county-wide increase in trade and commute access.[33] The analysis is done both with and without weighting by tract population in 1950. If trade access only affects population when weighted, it suggests that trade access increases population in the most populous tracts.

[32]To ensure consistent census tract definitions, I use crosswalks from Lee and Lin (2018) and the Longitudinal Tract Data Base. The sample includes only census tracts in MSAs with a positive population in 1950. If a tract was not part of an MSA in 1950 but appears later, its 1950 data is recorded as missing.

[33]I also lose the state-by-year-by-wave fixed effects to avoid over saturation of the regression with fixed effects.

OLS results (Appendix A.9) show that without weights, commute access increases population, while trade access has no effect after 40 years. However, when population weights are applied, commute access has a smaller impact than without weights. In turn, trade access shows a statistically significant positive effect. This suggest that trade access effects are concentrated on the most populous tracts,[34] whereas commute access effects are so in the least populous tracts.[35]

In summary, trade access decentralizes population between counties but concentrates it within counties. Moreover, commuting access has no effect on suburbanization between counties but decentralizes population within counties.

Does commuting between counties respond to changes in commute times? There may be concerns that studying changes in commute access at the county level is unfructuous, under the prior that most commuting occurs within counties and changes in driving times between counties have little impact. First, Figure A.1 already suggests otherwise. It shows that many residents work outside their home county, and that this number has been increasing over time. Second, I exploit changes in driving times and commute flows to study the relationship directly.

In this section, I estimate a dynamic gravity equation for commuting flows using data from 1970 to 2020, leveraging changes in commute times caused by staggered highway construction. Results in Appendix A.7 show that a one-minute increase in commute time reduces the probability of commuting by 1.8% in the short run, but this effect grows to 6.0% after 40 years. These findings align with prior research on commuting semi-elasticities but offer a novel contribution by separately identifying short- and long-run elasticities.

A.6 Theorems and propositions

THEOREM A.1: *Given parameter $\{\theta\}$, data $\{\tilde{L}_d, L_o, w_d\}$, and bilateral commute costs $\{d_{od}\}$, there exists a unique set of commuting preferences $\{K_d\}$ that rationalize total employment per location as an equilibrium. Additionally, if data on $\pi_{dd|d}$ is available and $\pi_{dd|d} L_d \leq \tilde{L}_d$, there is a unique set of preferences over non-origin destinations $\{K_{d \neq o}\}$ and origin-origin pairs $\{\tilde{K}_{oo}\}$, such that both commuting shares (over origin-origin pairs), $\pi_{dd|d}$, and employment data, $\tilde{L}_d$, are rationalized.*

[34] Or, at the very least, trade effects are somewhat constant across tracts of different sizes since coefficients are similar when the regression is weighted or unweighted.

[35] I do not report 2SLS estimates because instruments in this subsample are weak: the Kleibergen-Paap F-statistics are around 2.

Proof. For the first part, notice that commuting preferences over destinations are homogeneous of degree zero in the following system of equations:

$$\sum_{o} \pi_{od|o} L_o = \underbrace{\tilde{L}_d}_{data} \tag{27}$$

where

$$\pi_{ij|i} \quad = \quad \frac{\left(K_j \frac{w_j}{d_{i,j}}\right)^{\theta}}{\Phi_i} \tag{28}$$

with commuting access as $\Phi_i = \sum_{l=1}^{S} \left(K_j \frac{w_l}{d_{i,l}}\right)^{\theta}$.

For the second part, we can divide the procedure in two steps. First, we match data on terms away from the diagonal of commuting flows. Second, we match data on elements in the diagonal.

For the first step, we use market clearing:

$$\sum_{o \neq d} \pi_{od|o} L_o = \underbrace{\tilde{L}_d - \pi_{dd|d} L_d}_{data} \tag{29}$$

where

$$\pi_{ij|i} \quad = \quad \frac{\left(K_j \frac{w_j}{d_{i,j}}\right)^{\theta}}{\Phi_i}, \forall i \neq j \tag{30}$$

Here, we require that $\pi_{ii|i} L_i \leq \tilde{L}_i$ so that the right hand side is positive.

For the second step, we use the data on $\pi_{dd|d}$ plus the equations for commuting:

$$\pi_{dd|d} = \frac{\left(\tilde{K}_{dd} w_d\right)^{\theta}}{\Phi_d} \tag{31}$$

□

THEOREM A.2: *Given parameters $\{\nu, \beta\}$ and data $\{\tilde{L}_d, w_d, H_d, R_d\}$, there is a unique set of construction TFP $\{Z_d\}$ and land supply $\{T_d\}$ that rationalize the data as an equilibrium of the model.*

Proof. From labor market clearing in the construction sector, we infer unit costs as:

$$\tilde{L}_d^C w_d = \underbrace{\beta \left(\frac{w_d}{c_d^{Cons}}\right)^{1-\nu} R_d H_d}_{\text{by construction}} \quad \Rightarrow \quad c_d^{Cons} = \beta^{\frac{1}{1-\nu}} \left(\frac{\tilde{L}_d^C w_d^{\nu}}{R_d H_d}\right)^{-\frac{1}{1-\nu}}$$

We then recover productivities in the construction sector:

$$R_d = \frac{c_d^{Cons}}{Z_d}$$

Using the cost function definition $c_o^{Cons} = \left(\beta w_o^{1-\nu} + (1-\beta) Q_o^{1-\nu}\right)^{1/(1-\nu)}$, we can infer land prices Q_o:

$$(1-\beta) Q_d^{1-\nu} = \left(\frac{R_d H_d}{\tilde{L}_d^C} - w_d\right) \beta w_d^{-\nu}$$

Finally, from land market clearing:

$$(1-\beta) \left(\frac{Q_o}{c_o^{Cons}}\right)^{1-\nu} R_o H_o = Q_o T_o$$

we can recover land supply T_o. □

THEOREM A.3: *If $\alpha \bar{y}_{|d,t} L_{d,t} + (1-\beta) \left(\frac{Q_d}{c_d^{Cons}}\right)^{1-\nu} R_{d,t} H_{d,t} \geq \tilde{L}_{d,t}^{NT} w_{d,t}, \forall d$, given parameters $\{\sigma, \sigma_D, \varepsilon, \theta, \nu, \varrho^T, \varrho_L^{NT}, b, \alpha, \beta\}$, data $\{L_{d,t}, \tilde{L}_{d,t}^s, w_{d,t}, H_{d,t}, R_{d,t}, \}$, and bilateral costs $\{\varsigma_{od,t}^{NT}, \varsigma_{od,t}^{T}, d_{od,t}\}$, there is a unique set of TFP in the tradable and non-tradable sectors $\{A_{dt}^T, A_{dt}^{NT}\}$ that rationalize the data as an equilibrium of the model.*

Proof. From previous theorems we have inferred prices and supply in the land market, plus commute flows and therefore average income. These will be inputs later in the proof.

We start by computing compute unit costs and the price index in the non-tradable sector:

$$c_o^{NT} = \left(\varrho^{NT} w_o^{1-\sigma} + \left(1-\varrho^{NT}\right) \tilde{R}_o^{1-\sigma}\right)^{1/(1-\sigma)} \quad \text{and} \quad P_o^{NT} = cons \cdot \left(A_d^{NT} c_o^{NT-\varepsilon}\right)^{-1/\varepsilon} \tag{32}$$

with $const = \left[\Gamma\left(\frac{\varepsilon+1-\sigma_D}{\varepsilon}\right)\right]$.

We use labor market clearing the in the non-tradable and tradable sectors to search for the set of

TFPs that solves the following system of equations

$$\underbrace{\tilde{L}_d^T w_d}_{\text{Data}} = \underbrace{\varrho_L^T \left(\frac{w_d}{c_d^T}\right)^{1-\sigma} X_d^T}_{\text{Tradable Sector}} \tag{33}$$

$$\underbrace{\tilde{L}_d^{NT} w_d}_{\text{Data}} = \underbrace{\varrho_L^{NT} \left(\frac{w_d}{c_d^{NT}}\right)^{1-\sigma} X_d^{NT}}_{\text{Non-Tradable Sector}} \tag{34}$$

where:

$$X_d^T = \sum_o \frac{A_d^T \left(c_d^T \varsigma_{do}^T\right)^{-\varepsilon}}{\Psi_o^T} \left(\underbrace{P_o^T C_o^{W,T}}_{\text{By Workers}} + \underbrace{P_o^T C_o^{H,T}}_{\text{By Landlords}} + \underbrace{\frac{(1-\varrho_L^T-\varrho_H^T)}{\varrho_L^T} \left(\frac{P_o^T}{w_o^T}\right)^{1-\sigma} w_o \tilde{L}_o^T}_{\text{By Tradable Firms}} \right) \tag{35}$$

$$X_d^{NT} = \underbrace{(1-b)\left(\frac{P_d^{NT}}{P_d}\right)^{1-\sigma} \alpha \bar{y}_d L_d}_{\text{By Workers}} + \underbrace{(1-b)\left(\frac{P_d^{NT}}{P_d}\right)^{1-\sigma} Q_d T_d}_{\text{By Landlords}} \tag{36}$$

$$P_d^T = cons \cdot \left(\sum_{o=1}^{N} A_o^T \left(c_o^T \varsigma_{od}^T\right)^{-\varepsilon}\right)^{-1/\varepsilon} \tag{37}$$

and $P_d^T C_d^{W,T} = b\left(\frac{P_d^T}{P_d}\right)^{1-\sigma} \alpha \bar{y}_d L_d$ is the consumption by workers, and $P_d^T C_d^{H,T} = b\left(\frac{P_d^T}{P_d}\right)^{1-\sigma} Q_d T_d$ is the consumption by landowners. Every element in these equations, except for A_{NT} and A_T is either known directly from the data (e.g. $L_d^s w_d$), inferred from a previous theorem (e.g. $Q_d T_d$), or inferred in this theorem (e.g. P_d^T).

In fact, one can prove that the system is homogenous of degree 0, provided that parameters ϱ_x^T scale appropriately to a given normalization. Multiply TFP in the non-tradable and tradable sectors by a constant c. Notice that this term cancels out everywhere except in the demand of intermediate inputs by tradable firms, where we are left with:

$$c^{-(1-\sigma)/\epsilon} \frac{(1-\varrho_L^T-\varrho_H^T)}{\varrho_L^T} \left(\frac{P_o^T}{w_o^T}\right)^{1-\sigma} w_o \tilde{L}_o^T \tag{38}$$

Hence, we need to set the new $\frac{(1-\tilde{\varrho}_L^T-\tilde{\varrho}_H^T)}{\tilde{\varrho}_L^T}$ to equal to $(1/c^{-(1-\sigma)/\epsilon})\frac{(1-\varrho_L^T-\varrho_H^T)}{\varrho_L^T}$, so that the c term cancels out here as well.

Finally, notice that total income (by workers and landlords) in the right hand side of the non-tradables' labor market clearing equation is either observed directly in the data, or inferred from the data in a previous step. Then, the model can only match the wage bill in the non-tradable sector if the share of income spent on the non-tradable sector is large enough. This condition is trivially satisfied in a model without commuting but with a non-tradable sector.

$\square$

THEOREM A.4: *If* $\alpha\bar{y}_{|d,t}L_{d,t} + (1-\beta)\left(\frac{Q_d}{c_d^{Cons}}\right)^{1-\nu} R_{d,t}H_{d,t} \geq \tilde{L}_{d,t}^{NT} w_{d,t}, \forall d$, *given parameters* $\{\sigma, \sigma_D, \varepsilon, \theta, \nu, \varrho^T, \varrho_L^{NT}, b, \alpha, \beta, \zeta_t\}$, *data* $\{L_{d,t}, \tilde{L}_{d,t}^s, w_{d,t}, H_{d,t}, R_{d,t},\}$, *and bilateral costs* $\{\varsigma_{od,t}^{NT}, \varsigma_{od,t}^{T}, d_{od,t}, \xi_{od,t}\}$, *there is a unique set of amenities* $\{u_d\}$ *that rationalize the data as an equilibrium of the model.*

Proof. From previous theorems we recover $P_{i,t}$, and $\bar{y}_{i,t}$. From migration equations we have the following system of equations which, in every period, is homogeneous of degree one in amenities $u_{i,t}$:

$$L_{i,t} = \sum_n \zeta_t \frac{\bar{U}_{n,i,t}^{\eta}}{\sum_{i\in N}\bar{U}_{n,i,t}^{\eta}} L_{n,t-1} \tag{39}$$

$$\bar{U}_{n,i,t} = \frac{\xi_{n,i,t}u_{i,t}}{P_{i,t}^{\alpha}R_{i,t}^{1-\alpha}} \cdot \bar{y}_{t,|i} \tag{40}$$

$\square$

A.7 Dynamic Commuting Gravity Equation

In this section, I estimate a dynamic gravity equation using commute flows data and commute time changes from 1970-2020. Due to the staggered nature of highway construction, I exploit variation in changes in commute times produced at different dates to estimate a dynamic gravity equation of commuting. I follow Dube et al., (2023) and their local projection-based differences-in-differences (LP-DiD) approach. The local projection approach is a statistical technique introduced in Jorda (2005). The basic idea is exploit panel data to estimate dynamic impulse responses. The LP-DiD approach utilize local projections to estimate dynamic effects in addition to the so-called 'clean-control' condition. The clean control condition avoids bias by dropping 'unclean' observations.

What are the units of observation in this context? Commuting is a very local phenomenon. For example, even if a trip from San Francisco to New York would take 20 hours instead of 40 hours, still probably nobody would do this commute. In other words, flows between locations that are extremely distant between each other are obviously insensitive to changes in driving times. This way, I define my units of observation as all origin-destination pairs (i, j) that were at a distance of less than 5 hours in 1960. Results are robust to changes around the 5-hours threshold.

How do I build 'treatment' and 'control' groups? In this setting, treatment comes in many waves as highways were constructed in 1970, 1980, 1990, and 2000. For each treatment wave happening on date s, I define a 'treatment' and a 'control' group in the following way. A pair (i, j) is said to be 'treated' if commute times between locations i and j were reduced for the first time at date s. A pair (i, j) is said to be a 'control' pair if it is going to be treated in the future, but has not been treated yet; or, if it was never treated at all. In this way, for treatment wave s, I drop all pairs (i, j) that were treated for the first time before wave s. By this point, we already have a mapping between treatment wave s, and origin-destination pairs (i, j) that belong to either the treatment or the control group of treatment wave s. Finally, for each treatment wave s, I keep in the sample all observations at periods that are around a treatment window: $t \in \big(s - H, s + H\big)$.

I stack treatment and control groups for each treatment wave s and estimate the following regressions, clustering standard errors at county-pair level:

$$\Delta_h \log \pi_{i,j,t} = \theta_h \Delta_0 \tau_{i,j,t} + \underbrace{\phi_{i,t,h} + \phi_{j,t,h}}_{\text{FE}} + \epsilon_{i,j,t,h} \tag{41}$$

$$\Delta_h \tau_{i,j,t} = \alpha_h \Delta_0 \tau_{i,j,t} + \underbrace{\sigma_{i,t,h} + \sigma_{j,t,h}}_{\text{FE}} + \varepsilon_{i,j,t,h} \tag{42}$$

where the parameter of interest is $\tilde{\theta}_h = \theta_h / \alpha_h$. This is equivalent to using the first change in commute times $\Delta_0 \tau_{i,j,t}$ as an instrument for changes in commute times happening at any time after the first treatment: $\Delta_h \tau_{i,j,t}$.[36] The advantage in gravity settings is that it allows the researcher to control for for origin-decade fixed effects and destination-decade fixed effects. They account for idiosyncratic changes in the amenity value of origin locations, and for idiosyncratic changes in productivity of destination locations.

In Figure A.10, I show the results of estimating equation (41), which is akin to a reduced form regression. This figure reveals that when commute times rises by one minute between location i and location j, the probability of commuting to j decreases by 1.9%, approximately. However, this impact rises over time in absolute terms. After 40 years, the probability of commuting to j decreases by 6.0%. This semi-elasticity of -0.060 is statistically different at a 5% level than the -0.018 estimated for the short-run effect. Hence, changes in commute times have long-lived and growing effects on commuting

[36] I estimate equation (41) by PPML, and equation (42) by OLS

flows as cities and structures are built. This is a novel fact. Importantly, I do not detect differential pre-trends in commute flows.

The values of these coefficients are consistent with previous estimates in the literature. For instance, depending on the specification, Tsivanidis (2023) finds an impact between -0.025 and -0.053 over a span of 20 years in a sample of urban planning zones within Bogotá, Colombia (an urban planning zone is a somewhat smaller area than a typical municipality). Ahlfeldt et al. (2015), find a semi-elasticity of -0.070 in a cross-section of origin-destination districts in Berlin. Zarate (2022) finds a semi-elasticity between -0.028 and -0.042 in a cross-section of origin-destination municipalities within Mexico city. Velasquez (2024) estimates a semi-elasticity between -0.047 and -0.067 in a cross-section of origin-destination municipalities within Lima, Perú.

The main takeaways from this section are that (i) changes in commute times between counties produce changes in commute flows, (ii) the effects are greater the longer the time horizon. Also, (iii) while previous estimates of the semi-elasticity between commute times and commute flows are in the same ballpark, in most studies, it is unclear whether estimates of the commuting elasticity is a long-run elasticity, an elasticity over a fixed time horizon, or a mix of short- and long-run elasticities. My study provides estimates of the short- and the long-run elasticities, as well as their path over time.

Figure A.10: Dynamic Gravity Equation of Commuting