Commuting, Labor, and Housing Market Effects of Mass Transportation: Welfare and Identification*

Christopher Severen†
Federal Reserve Bank of Philadelphia

June 2018

Abstract

Using a panel covering all bilateral commuting, I estimate a quantitative spatial general equilibrium model to quantify the welfare effect of Los Angeles Metro Rail and isolate the commuting benefit from other channels. The Metro increases commuting between connected locations by 7%-13% relative to control pairs selected using historical proposed routes; other margins are not affected. A novel strategy interacts local innovations with intraurban geography to identify local housing and labor elasticities; estimates indicate limited mobility. The Metro generates a welfare gain of $246 million annually by 2000, one-third the annualized cost. More recent data show additional, dynamic commuting growth.

Keywords: subway, commuting, gravity, economic geography, local labor supply

JEL Codes: J61, L91, R13, R40

*Many have provided valuable advice during the development of this paper; I would especially like to thank Elliot Anenberg, Valerie Bostwick, Jeff Brinkman, Olivier Deschênes, Jonathan Dingel, Gilles Duranton, Jessie Handbury, Stephan Heblich, Kilian Heilmann, Nicolai Kuminoff, Jeff Lin, Kyle Meng, Paulina Oliva, and Wei You. First draft was December 10, 2015.

†Address: Research Department, Federal Reserve Bank of Philadelphia, Ten Independence Mall, Philadelphia, PA 19106. Telephone: (215) 574-4113. Email: chris.severen@gmail.com or Chris.Severen@phil.frb.org.

1 Introduction

High commuting costs limit consumer choice and inhibit mobility within cities. Governments consider a broad range of interventions to mitigate the costs of distance and congestion. Subway and light rail systems are increasingly popular: Since the 1980s, Atlanta, Austin, Dallas, Denver, Houston, Los Angeles, Miami, Phoenix, Portland, and Seattle have all built systems. But rail transit is expensive, costing from three to ten times as much per mile for light rail as roads (and potentially much more for subways).[1] Do the benefits of urban rail transit projects outweigh their costs, particularly in car-oriented cities?

I study the effects of Los Angeles Metro Rail on welfare, commuting, and the spatial distribution of people and prices. Los Angeles is a large, automobile-dominated city that built a relatively extensive rail transit network in ten years from scratch, making LA Metro Rail particularly relevant for other cities considering rail transit. I describe a quantitative economic geography model of a city that is composed of a collection of labor and housing markets (census tracts) connected by commuting to provide a framework for analyzing the impacts of LA Metro Rail. Central to this model is a gravity-like equation of commuting flows. In order to estimate this equation, I assemble a unique dataset that includes all census tract-to-census tract commuting flows and times in 1990 and 2000 in the greater LA region. I provide the first causal estimates of transit's effect on bilateral commuting using three complementary strategies. The first two strategies exploit historical maps of streetcar and proposed subway routes; I select plausible control tract pairs based on proximity to these paths. The third approach defines control pairs by adjacency to treated pairs and provides a lower bound on the effect size. The effect of LA Metro Rail is substantial: commuting increases by 7%-13% between connected tract pairs by 2000.

To determine whether transit shifts other margins and quantify welfare effects, I require local labor and housing market elasticities. Census tract-scaled elasticities that condition on geography and have not been rigorously estimated; I develop a new strategy and assemble unique data to identify these parameters. My approach begins with a local implementation of the Bartik (1991) instrument that exploits local (census tract) variation in labor demand within the city. I employ workplace wage data to estimate the local labor supply elasticity. Interacting tract-specific labor demand shocks with the spatial configuration of the city generates additional instruments that identify all remaining labor and housing market elasticities. The elasticity of local (tract-level) labor supply governs the homogeneity of household preferences over locations, and thus how responsive agents are to changes in prices, amenities, and commuting costs. This parameter is essential to translate treatment effects to utility. It has a low value, indicating agents are heterogeneous in their preferred locations and relatively unwilling to move in response to changes in local characteristcs. I also find that tract-level housing supply is inelastic. Using estimated elasticities and observed data, I can distinguish the commuting effect of transit from other channels and ra-

1. Light rail in the US typically costs \$50-\$150 million per mile.

tionalize various findings about urban rail transit in a single framework. The commuting effect dominates; impacts from non-commuting channels (e.g., amenities) appear minimal.

An advantage of my empirical setting is that I observe bilateral, tract-level residence-workplace commuting flows and wages by workplace. In other settings, these data are not observed and are instead recovered by combining other data with modeling assumptions that explain labor supply and commuting primarily in terms of travel time (e.g., Ahlfeldt et al. 2015; Allen, Arkolakis, and Li 2015). Because I observe this data, I can an implement an urban economic geography model without these assumptions, as well as determine how reasonable such assumptions are. Observed wage explains about 2% of the variation in the modeled, adjusted wage term. Travel time and distance explain at most 20% of time-invariant variation in commuting flows. Together, these results suggest that significant and persistent idiosyncrasies play an important role in determining who works where. Travel-time based accessibility (e.g., market access or price index) terms, when used within cities, smooth away much unmodeled variation and may attribute too much to changes in travel times (e.g., Heblich, Redding, and Sturm 2018; Tsivanidis 2018).

Transit increases the attractiveness of commuting between connected tracts, leading to substantial welfare gains through increased commuting capacity. Preferred estimates show that by 2000, LA Metro Rail generates about $246 million in surplus for commuters annually, or roughly $13.30 for every ticketed ride. However, these benefits amount to about one-third of the annualized cost of construction and net operating expenses. I also examine additional changes in commuting after a further fifteen years. Tracts first connected before 2000 see an additional increase in commuting of 5%-7% by 2015, and tracts connected after 2000 experience a 4%-9% increase in commuting by 2015. Taking these additional gains into account, benefits approach costs, but only break even under the most generous assumptions. This analysis suggests that rail transit is unlikely to be cost effective over its first two or three decades as measured by its primary output, commuting.[2]

The primary effect of transit is to change the utility cost of travel. Quantifying these effects is challenging. Following early applications of discrete choice modeling (e.g., McFadden 1974), one literature estimates the benefits of transit in a modal choice framework. These models require knowledge of what travel costs are (travel time, reliability, availability of complementary activities, such reading or working), and accurate measures of their value. However, empirical estimates vary widely.[3] A hedonic literature notes that commuting benefits of transit are capitalized into housing and land prices (Baum-Snow and Kahn 2000; McMillen and McDonald 2004). How-

2. One important caveat is that while I calculate the commuting effects over a twenty-five year window, I can only examine other channels between 1990 and 2000. Unmeasured benefits include increased mobility for non-commuters (those unemployed or not in the labor force) and environmental factors. Like other transit agencies, the LA County Metropolitan Transportation Authority has historically received more than two-thirds of its capital expenditure budget from the federal government. Thus, rail could be viewed as cost effective from a local public finance perspective.

3. Summarizing many studies, Small and Verhoef (2007) notes value of time estimating varying from 35%-84% of wage rate for commuting, and from 20%-90% of wage rate for personal travel.

ever, this approach primarily considers homeowner benefits, potentially excluding other portions of the commuting public. Hedonics cannot easily isolate commuting benefit from other channels, such as residential amenities (Chen and Whalley 2012; Kahn 2007) or disamenities (Bowes and Ihlanfeldt 2001). A common identification threat concerns demand spillovers; this could lead to violations of the stable unit treatment value assumption (SUTVA).[4] Metropolitan-level analysis suggests that public transit expansion may enhance aggregate productivity and employment growth (Chatman and Noland 2014; Duranton and Turner 2012), but has at most a small effect on population growth (Gonzalez-Navarro and Turner 2016) and does little to reduce city-level commute times (Duranton and Turner 2011). However, aggregate analysis does not capture the role of local factors nor the impact of transit on urban form.

The particular research setting is of great interest: Los Angeles is a car-oriented city that installed a relatively large rail network over a ten-year period, resulting in forty-six stations by 2000. The experience of Los Angeles is more informative for most cities considering rail-based mass transit than studies from older, denser cities (e.g., Gibbons and Machin 2005). It is an active line of inquiry whether new mass transit infrastructure in less dense cities provides appreciable benefits, particularly given the newer role of cities as centers of consumption in addition to production (Baum-Snow, Kahn, and Voith 2005; Billings 2011; Glaeser, Kolko, and Saiz 2001). Interest in understanding the economic consequences of the LA Metro has indeed been high, and there is a budding line of research on the topic. I contribute by studying a new outcome, commuting, using complementary approaches that exploit historical maps and spatial adjacencies, and by developing credible estimates of welfare impacts.[5]

2 A model of urban location choice

Transportation improvements change how agents interact with and across space. I describe a simple, quantitative model of a city with internal commuting to rationalize interactions between housing and labor markets across space.[6] The model generates a log-linear system of simultaneous equations that describe commuting as well as housing and labor markets. The city consists of a collection of N locations, operationalized as census tracts, that each contain a labor market

4. Gibbons and Machin (2005) show that transit can displace housing demand elsewhere; Donaldson and Hornbeck (2016) discuss the importance of modeling general equilibrium when evaluating transportation infrastructure.

5. The use of historical and proposed routes is a mainstay of regional economics (e.g., Baum-Snow 2007), but these tools have been seldom applied within cities (for an exception, see Heilmann 2017). The adjacency approach applies the intuition of Dube, Lester, and Reich (2010) to bilateral flow data.

6. The model is similar to that of Ahlfeldt et al. (2015), with four notable differences: (i) origin-destination pairs are subject to idiosyncratic preferences, (ii) local housing supply can be exogenously shifted, (iii) land use is exogenously determined between housing and productive uses, and (iv) agglomeration and consumption externalities are excluded from the primary model. The first two are useful generalizations that rationalize variation in observed commuting flows. The latter two match the empirical setting (there is little scope for land use adjustment and externalities are essentially time invariant and captured by tract fixed effects), simplify exposition, and have little impact. I detail how these changes alter identification in Section 4; results are robust to their incorporation.

and a housing market. There is no restriction on where agents live and work conditional on being within the city, so agents choose the place of residence-place of work pair that maximizes utility conditional on commuting costs. The model links local, observable equilibrium outcomes to unobservable economic fundamentals.

Joint market household decision: Labor supply and housing demand

Atomistic households make location and consumption decisions. For the location decision, households choose a tract of work and a tract of residence. Conditional on choosing to live in location i, households face per unit housing costs Q_i and receive amenity $\tilde{B}_i$. Conditional on choosing place of work j, households inelastically provide one unit of labor in exchange for wage W_j. Given the joint location choice and prices, households make decisions over consumption of housing and a composite good. Specifically, household o chooses location pair ij, consumption $\mathcal{C}$, and housing H to maximize the following Cobb-Douglas utility function:

$$\max_{\mathcal{C},H,\{ij\}} U_{ijo} = \max_{\mathcal{C},H,\{ij\}} \frac{\nu_{ijo}\tilde{B}_i}{\delta_{ij}} \left(\frac{\mathcal{C}}{\zeta}\right)^{\zeta} \left(\frac{H}{1-\zeta}\right)^{1-\zeta} \quad \text{s.t.} \quad \mathcal{C} + Q_i H = W_j$$

where ν_{ijo} is household o's idiosyncratic preference for location pair ij. The cost of commuting between i and j is captured by $\delta_{ij} = e^{\kappa\tau_{ij}}$, where τ_{ij} is the observed travel time. The share of household expenditures on housing is $1-\zeta$. This generates the following indirect utility function conditional on location pair choice ij:

$$v_{o|ij} = \frac{\nu_{ijo}\tilde{B}_i W_j Q_i^{\zeta-1}}{\delta_{ij}}$$

Given this specification, optimal housing consumption for household o conditional on location pair ij is given by $H_{ijo} = (1-\zeta)W_j/Q_i$.[7]

To map indirect utility to choice probabilities, assume $\nu_{ij}(o)$ is distributed Fréchet with scale parameter $\tilde{\Lambda}_{ij} = T_i E_j D_{ij}$ and shape parameter $\epsilon > 0$. The cdf of ν is thus:

$$F_{ij}(\nu) = e^{T_i E_j D_{ij} \nu^{-\epsilon}}$$

The scale parameter captures mean idiosyncratic preference for location pair ij: T_i captures the mean utility of residing in i, E_j the mean non-wage utility of working in j, and D_{ij} an unobserved pair-specific shift in the utility of a particular commute. The shape parameter governs the degree of homogeneity in preferences: for high ϵ, agents view location pairs homogeneously, while for

7. Any indirect utility function with a multiplicatively separable idiosyncratic component could be employed. For example, the sorting literature uses a nested CES parameterization (Epple and Sieg 1999). Davis and Ortalo-Magné (2011) show that expenditure shares on housing are relatively constant through time in cities in the United States, supporting the Cobb-Douglas assumption. I retain this assumption to maintain comparability with the existing literature.

low ϵ, their valuations are quite heterogeneous. With this distributional assumption, utility maximization yields a simple proportional formula for commuting flows. The share of the population that chooses residential location i and place of work j is:

$$\pi_{ij} = \frac{\tilde{\Lambda}_{ij}\left(\delta_{ij}Q_i^{1-\zeta}\right)^{-\epsilon}(\tilde{B}_iW_j)^{\epsilon}}{\sum_r\sum_s\tilde{\Lambda}_{rs}\left(\delta_{rs}Q_r^{1-\zeta}\right)^{-\epsilon}(\tilde{B}_rW_s)^{\epsilon}} \tag{1}$$

To relate commuting shares to observable commuting flows, multiply π_{ij} by the population of the market as a whole ($\bar{N}$), so that $N_{ij} = \pi_{ij}\bar{N}$.

The city can be viewed either as existing in autarky or being nested in a large, open economy. This assumption makes little difference outside of welfare calculations (due to homothetic preferences). In an open economy, no spatial arbitrage requires that the average welfare from moving to the city equal the reservation utility of living anywhere else. The expected value of moving to the city is:

$$\mathbb{E}[U_{ijo}] = \Gamma\left(\frac{\epsilon-1}{\epsilon}\right)\cdot\left[\sum_r\sum_s\tilde{\Lambda}_{rs}\left(\delta_{rs}Q_r^{1-\zeta}\right)^{-\epsilon}(\tilde{B}_rW_s)^{\epsilon}\right]^{1/\epsilon} \tag{2}$$

where $\Gamma(\cdot)$ is the gamma function and the aggregate population $\bar{N}$ is implicitly defined. Free mobility thus requires $\mathbb{E}[U_{ijo}] = \bar{U}$, and aggregate population changes to maintain $\bar{U}$.

Production: Labor demand

A continuum of measure zero firms produces a globally tradable commodity in each location j under perfect competition.[8] Firms select competitively available labor N^Y and land L^Y inputs to maximize profits under constant returns to scale.[9] Production is multiplicatively separable in local productivity A_j and a technology that is identical across j:

$$Y = A_jF(N_j^Y, L_j^Y) \tag{3}$$

Because of the atomistic size of firms, land use decisions are made in accordance with profit maximization despite the locally fixed available quantity of land.[10] Perfect competition in labor markets implies that firms pay workers the marginal product of labor: $W_j = A_jF_N(N_j^Y, L_j^Y)$. I assume

8. The primary focus of this study is the flow of people rather than the flow of goods, so I assume that goods are uniformly available and globally traded. This differs from studies of regional trade that are interested in trade in goods (Allen and Arkolakis 2016; Donaldson 2018; Faber 2014; Monte, Redding, and Rossi-Hansberg 2015).

9. A number of recent papers that compare metropolitan outcomes permit heterogeneity in workforce productivity, generally by education level (Diamond 2016; Adão 2016). Other work has focused on the interaction of skill and the distribution of economic activity within and across cities (Davis and Dingel 2014). I abstract away from this in an effort to focus on very local effects — in the empirical application, what is one metropolitan statistical area in other papers is here more than 2,500 unique locations.

10. Individual firms make unconstrained input decisions, but aggregate land use is predetermined, as is standard in many urban models, e.g., Glaeser et al. (2008).

Cobb-Douglas production technology: $F(N^Y, L^Y) = (N^Y)^\alpha (L^Y)^{1-\alpha}$.[11] The key parameter is $\alpha - 1$; the implicit assumption of constant returns to scale plays little role. Inverse labor demand is given by:

$$W_j = \alpha A_j \left(\frac{L_j^Y}{N_j^Y} \right)^{1-\alpha} \tag{4}$$

Housing supply

Housing is produced by measure zero builders using land for housing L^H and material inputs M. A local, multiplicatively separable housing productivity term $\tilde{C}_i$ captures local cost drivers such as geography (e.g., terrain) and regulation. Materials are readily available in all locations at the same cost, but aggregate local land supply for housing is predetermined.[12] Convexity in land pricing serves as a congestive force, driving up prices in desirable locations until agents look elsewhere. As is standard in the literature, I specify housing production to take Cobb-Douglas form: $H = (L^H)^\phi M^{1-\phi} \tilde{C}_i$.[13] Developers sell housing in location i in a competitive market at per unit price Q_i to maximize profit: $Q_i H - P_i^L L^H - P^M M$. The exogenous price of construction materials P^M is common to all locations, and the local price of land P^L responds to housing demand (discussed below).

Because detailed data on housing production is not available, I utilize the zero profit condition to develop an empirical formula for housing costs. The first order condition of developer profit with respect to construction materials gives:

$$Q_i = \frac{P^M}{(1-\phi)\tilde{C}_i} \left(\frac{M}{L^H} \right)^\phi \tag{5}$$

Substituting this into into the developer's profit function and enforcing the zero profit condition implied by perfect competition gives construction material demand: $M^* = \frac{1-\phi}{\phi} \frac{L^H P_j^L}{P^M}$. Enforcing zero profits gives $Q_i = (P_i^L L^H + P^M M)/((L_i^H)^\phi M_i^{1-\phi} \tilde{C}_i)$. Substituting in M^* gives the cost function:

$$Q_i = \left(\frac{1-\phi}{\phi} \right)^{1-\phi} \frac{(P_i^L)^\phi (P^M)^{1-\phi}}{\phi \tilde{C}_i} \tag{6}$$

11. Ottaviano and Peri (2012) argue that Cobb-Douglas is reasonable when estimating labor demand curves in the US because the long run expenditure share of income on labor is roughly constant.

12. This simplifies the model while maintaining fidelity to the setting. Strong zoning and the medium time frame of this study may not match the temporal patterns required for land use change; many studies of land use or housing supply examine only long-run changes (e.g., Saiz (2010) uses a thirty-year window). Including land use measures does not greatly change identification or results. There is little evidence of differential land use near transit.

13. Ahlfeldt et al. (2015) show that the Cobb-Douglas specification of floor space production using land and materials as inputs is a reasonable approximation for modern Berlin. Combes, Duranton, and Gobillon (2012) find the same using French data, as do Epple, Gordon, and Sieg (2010) in Allegheny County, PA.

The cost of housing can thus be interpreted as an average between the cost of land and the cost of construction materials, and is deflated by housing productivity.

The price of land responds to changes in demand and land availability: I parameterize it as a function of local housing density $P_i^L = (H_i/L_i^H)^{\tilde{\psi}}$, where the parameter $\tilde{\psi} > 0$ captures local price elasticity of land with respect to density.[14] This parameter provides a congestive force to the model. Taking the expression for land price, combining with Equation (6), and compressing notation relates housing supply, price, and land availability:

$$Q_i = C_i \left(\frac{H_i}{L_i^H} \right)^{\psi} \tag{7}$$

where $\psi = \tilde{\psi}\phi$, and $C_i = (1-\phi)^{\phi}(P^M)^{1-\phi}/\tilde{C}_i\phi^{\phi+1}$ captures the local inverse efficiency in housing production. As housing productivity $\tilde{C}_i$ increases, C_i falls, so increases in housing productivity (decreases in C_i) increase the quantity of housing supplied at any price.

Equilibrium characterization

In equilibrium, labor and housing markets clear in all locations. Labor market clearing requires that local labor demand equal supply:

$$N_i^Y = \sum_r \bar{N}\pi_{ri} \tag{8}$$

Commuting shares (1) determine employment in any location. With frictional commuting, workers benefit from residing near work locations. Given the assumptions on household preferences, housing demand is a constant fraction of the ratio of wage to housing price. Aggregate housing demand in i is the sum of wage-rent ratios weighted by commuting flows—this takes into account heterogeneity in income stemming from variation in place of work. Housing market clearing requires that the local housing supply equal demand:

$$H_i = (1-\zeta) \sum_s \bar{N}\pi_{is} \frac{W_s}{Q_i} \tag{9}$$

Given model parameters $\{\alpha, \epsilon, \zeta, \psi, \kappa\}$, reservation utility $\bar{U}$, vectors of land availability by use $\{\mathbf{L^Y}, \mathbf{L^H}\}$, vectors of residential fundamentals $\{\tilde{\mathbf{B}}, \mathbf{C}, \mathbf{T}\}$, vectors of place of work fundamentals $\{\mathbf{A}, \mathbf{E}\}$, and matrices of residential-place of work pair fundamentals $\{\mathbf{D}, \boldsymbol{\tau}\}$, the general equilibrium of the model can be referenced by the price vectors $\{\mathbf{W}, \mathbf{Q}\}$, the commuting vector $\boldsymbol{\pi}$, and the scalar population measure $\bar{N}$. The elements of the equilibrium are determined by the following equations: commuting shares (1), profit maximization and competition in production (4), com-

14. I discuss an alternate way to close the model in the Appendix. Because I have data on land use, rather than floor space or use, I frame the model and analysis in terms of land.

petition in housing production and land use (7), labor market clearing (8), and housing market clearing (9), as well as no spatial arbitrage (2) in an open economy.

Proposition 1 (Existence and uniqueness)**.** *Consider the equilibrium defined by equations (1), (4), (7), (8), and (9):*

i) At least one equilibrium exists across residential locations with strictly positive quantities of residential land and work locations with strictly positive quantities of land used in production.

ii) There is at most one equilibrium if

$$\frac{2\epsilon(\epsilon+1)(1-\alpha)(1-\zeta)}{1+\epsilon(1-\alpha)} - 1 \leq \frac{1}{\psi} \tag{10}$$

Proof. See the Appendix. □

Existence makes use of the assumption that land use is predetermined and requires that positive residential land translates to a positive measure of residents and that positive land in production translates to a positive measure of workers. However, existence does not require positive commuting flows between all locations. The presence of zero commuting flows is a common characteristic of commuting data. The uniqueness condition requires that the elasticity of housing supply ($1/\psi$) be larger than a function of preference homogeneity and other parameters. The left-hand term is increasing in ϵ: The more homogeneous preferences are, the more elastic housing supply must be to ensure a single equilibrium.[15]

Recovering funadmentals

The model may have multiple equilibria, though this is unlikely given the low value of ϵ (representing high heterogeneity in idiosyncratic location preference) that I estimate. Regardless, for a given set of parameters, there is a unique mapping from the observed data to local fundamentals. Model parameters are estimated using these fundamentals and the observed values of the endogenous variables in combination with instruments to define moment conditions. Because these equations can be formulated as a system of three separate, linear equations, standard identification assumptions for generalized method of moments (GMM) estimators hold. $\tilde{B}_i$ and T_i enter isomorphically; let $B_i = T_i\tilde{B}_i^\epsilon$ and $\Lambda_{ij} = B_i E_j D_{ij}$.[16] Local fundamentals $\mathbf{A}$, $\mathbf{C}$, and $\mathbf{\Lambda}$ can be expressed as unique functions of data and parameters:

15. In this setting, agglomeration does not alter the equilibrium uniqueness condition; see the Appendix.

16. This mapping marks significant divergence from the framework in Ahlfeldt et al. (2015), where local fundamentals consist of a composite workplace term that combines $\mathbf{A}$ and $\mathbf{E}$, a residential term that combines $\mathbf{B}$ and $\mathbf{T}$, and omits any location or pair specific variation in housing supply $\mathbf{C}$ or commute utility $\mathbf{D}$. Note that the components of $\mathbf{\Lambda}$ are not uniquely identified from the data; I use statistical arguments to separate $\mathbf{B}$, $\mathbf{E}$, and $\mathbf{D}$.

Proposition 2 (Residual uniqueness). *Given parameters* $\{\alpha, \epsilon, \zeta, \psi, \kappa\}$, *observed data* $\{\mathbf{W}, \mathbf{Q}, \boldsymbol{\pi}, \bar{N}\}$, *and commuting times* $\boldsymbol{\tau}$, *then there exists a unique set of fundamentals* $\{\mathbf{A}, \mathbf{C}, \boldsymbol{\Lambda}\}$ *that are consistent with the data being an equilibrium of the model.*

Proof. See the Appendix. □

3 Data

I draw from a variety of sources to develop a decadal panel of intraurban outcomes from 1990 and 2000.[17] The sample covers Los Angeles County and four adjacent counties (Orange, Riverside, San Bernardino, and Ventura) and captures most commuting into and out of Los Angeles County (where all transit installations have occurred). Furthermore, this five-county area is economically distinct from other conurbations; it is reasonable to assume that it captures most relevant local interactions. While there is a rich amount of data available, there are some difficulties in obtaining consistent data over the sample period. I briefly discuss data sources and geographic normalization below; additional details can be found in the Appendix. Summary information, shown in Table 1, reveals that nearly 75% of commuters drove to work alone 1990.

Geographic data normalization. The standard unit of observation in this paper is a census tract or tract pair using 1990 Census geography. Tract definitions change over time, and data products that provide consistent geographies do not include many of the primary variables of interest in this study.[18] I normalize to 1990 geography because it involves the least amount of data manipulation and minimizes rounding issues. For data from 2000, I collect data at the block group or tract level. I overlay the 2000 geographies on 1990 tract definitions to spatially assign them to 1990 tracts. For block groups that map into multiple tracts, I weight block groups to tracts proportionally by area. I perform a similar process for origin-destination pairs.

Commuting flow data. Spatially disaggregate commuting data come from the 1990 Census Transportation Planning Package (CTPP). The 1990 CTPP reports tract-to-tract commuting flows for all tracts in the Los Angeles area. The 2000 CTPP gives tract-to-tract commuting flows for all census tracts in the U.S. I normalize geographies to construct a panel of tract-to-tract commuting flows for all tracts within the five county region surrounding Los Angeles. Reporting standards have changed over time; when I combine CTPP data across years, I apply consistent rounding rules.[19]

17. While I provide some additional results using more recent data, my primary analysis focuses on 1990 to 2000 because changes in reporting and data availability. I discuss these issues in more detail in the Appendix.

18. Commuting flows and wage at place of work are available only in original geographies.

19. The geographic normalization results some pairs with fractional commuting flows in 2000. To overcome this issue, I treat a flow of less than 1 as 0 in the empirics. I experiment with alternative cutoffs, as discussed in the Appendix.

Place of residence and place of work data. I draw aggregate data on residential census tracts and block groups from the National Historic Geographic Information System (NHGIS). I also use Geolytics' Neighborhood Change Database (NCDB) to validate identifying assumptions. The CTPP contains *census tract of work* wage data unavailable elsewhere, and employment by industry (in 18 aggregate Standard Industrial Classification (SIC) codes). More recent CTPP products do not include this variable, which is the primary data limitation that restricts my primary analysis to the 1990-2000 period.

Transit data and treatment; other data sources. I obtain shapefiles with location data on transit stations and lines from the Los Angeles County Metropolitan Transportation Authority and combine this with published information on the timing of station and line openings. To construct labor demand shocks, I draw from IPUMS microdata on all workers outside of California from the 1990 and 2000 Censuses. I obtain a panel of spatial land use data from the Southern California Association of Governments (SCAG). Finally, I use historical data on proposed transit systems and rail lines to develop plausible counterfactuals (described in Section 6).

4 Identification and estimation

Local labor and housing market elasticities provide the mapping between local fundamentals (and interventions that shift them) and observed prices and quantities. Consistent estimates of the elasticities are required to use observable data to learn about changes to local fundamentals and to simulate counterfactual scenarios. I develop an identification strategy that uses panel variation in wages at place of work, housing prices, and commuting flows, permitting the incorporation of tract and tract-pair fixed effects to flexibly control for unobserved, time-invariant characteristics that confound identification. This is important as persistent, difficult-to-capture characteristics can play an anchoring role in cities (Lee and Lin 2018).

All components of the model are expressed in the commuting flow (1), wage setting (4), and housing price (7) equations. Local fundamentals are potentially functions of covariates ($A = A(X)$ and so on) such as transit proximity. Log-linearizing multiplicatively separable terms delivers a tractable linear system:

$$w_j = g_0 + (\alpha - 1)n_j^Y + \ln(A_j) \tag{11}$$

$$n_{ij} = g_1 + \epsilon w_j - \epsilon(1-\zeta)q_i - \epsilon\kappa\tau_{ijt} + \ln(B_i E_j D_{ij}) \tag{12}$$

$$q_i = g_2 + \psi h_i + \ln(C_i) \tag{13}$$

where $n_j^Y = \ln\left(\bar{N}\sum_r \pi_{rj}/L_j^Y\right)$ is log employment density, $h_i = \ln\left((1-\zeta)\bar{N}\sum_s \pi_{is}W_s/Q_iL_i^H\right)$

is log housing density, and g capture remaining constants.[20] Equation (12) highlights where my approach diverges from common practice in the economic geography literature. Instead of relying on market access terms to summarize spatial relationships, I model these connections.

This system can be re-expressed to more clearly represent the connections between supply and demand in multiple markets, as well as better exposit the identification strategy. First, I separate the unobservables into time varying and time invariant components, so that $\ln(A_{jt}) = \bar{a}_j + a_{jt}$, etc. This leads to the following system (omitting constants):

$$\text{Labor demand in } i: \quad w_{jt} = \tilde{\alpha} n_{jt}^{Y} + \bar{a}_j + a_{jt} \tag{14}$$

$$\text{Labor supply to } i: \quad \omega_{jt} = \epsilon w_{jt} + \bar{e}_j + e_{jt} \tag{15}$$

$$\text{Commuting between } i \text{ and } j: \quad n_{ijt} = \omega_{jt} + \theta_{it} - \epsilon\kappa\tau_{ijt} + \bar{d}_{ij} + d_{ijt} \tag{16}$$

$$\text{Housing demand in } i \quad \theta_{it} = \tilde{\zeta} q_{it} + \bar{b}_i + b_{it} \tag{17}$$

$$\text{Housing supply in } i: \quad q_{it} = \psi h_{it} + \bar{c}_i + c_{et} \tag{18}$$

where $\tilde{\alpha} = \alpha - 1$, $\tilde{\zeta} = -\epsilon(1-\zeta)$. The system resembles standard linear supply and demand models, but for many interconnected housing and labor markets. OLS estimates of supply and demand elasticities will be biased due to simultaneous determination of prices and quantities.

4.1 A general approach to identifying local elasticities

I develop a local implementation of a shift-share instrument that takes advantage of available data and geography. I leverage plausibly exogenous panel variation in tract-level labor demand, interacting local labor demand shocks with the distance between tracts to create exogenous variation in local economic conditions. This approach requires little reliance on model structure, expresses identifying assumptions in terms of transparent moment conditions, and exploits the panel nature of the data to control for local, time invariant characteristics. The moment conditions identify all four elasticities, though I focus primarily on ϵ and ψ, as these two embed information about the local economic environment and cannot be estimated from microdata.[21] While the model described in Section 2 of the paper abstracts away from many potential confounding factors, I discuss the strengths and weaknesses of the proposed identification strategy to such model variants.

20. That is, $g_0 = \ln(\alpha)$, $g_1 = \ln(\bar{N}) - \ln\left(\sum_r \sum_s \Lambda_{rs}\left(e^{\kappa\tau_{rs}} Q_r^{1-\zeta}\right)^{-\epsilon} W_s^{\epsilon}\right)$, and $g_2 = 0$. Linearizing equations (1), (4), and (7) has empirical consequences: Commuting flows of zero are excluded. This is relatively standard when modeling gravity; I follow this approach because most tract-pairs connected by transit have non-zero flows. Other approaches are available, e.g., Silva and Tenreyro (2006).

21. In contrast, α and ζ could be estimated from microdata. Nonetheless, they are theoretically identified and estimates are reasonable.

Summary

Identification requires a demand or supply shock that shifts one of Equations (11) to (13) but is excludable from the others. I construct tract-level labor demand shocks from changes in national wage and employment levels and ex ante local employment shares by industry. After controlling for year and census tract fixed effects, the remaining variation consists of changes in wages and employment determined from ex ante, local industrial composition. These shocks are relevant if they are correlated with changes in local productivity (Δa_{jt}) and excludable if they are uncorrelated with changes in the other local fundamentals. Under these assumptions, the labor demand shock traces out the labor supply curve. Housing demand in nearby locations shifts in response. Because these downstream housing demand effects will be stronger nearer the workplace origination of the shock, I take a linear combination of labor demand shocks with weights determined by a spatial decay function to map the labor demand shocks to a residential tract. This derived housing demand instrument traces out the housing supply curve.

To identify housing demand, I require an instrument that shifts housing supply. For agents who work in j and live in i, a labor demand shock for agents who work elsewhere (in j') but live in i shifts effective housing supply in i. That is, a labor demand shock for workers $n_{ij'}$ with $j' \neq j$ translates into a housing supply shock to workers n_{ij} (i.e., as long as housing supply is not perfectly elastic). I again use spatial decay weights to determine an appropriate relationship between labor demand shocks and the housing market. Finally, labor demand shocks in one location alter wages and induce workers to shift employment location. In the absence of spillovers in labor demand, the labor demand shocks in one location shift labor supply in nearby locations (conditional on the local shock), tracing out labor demand.

Detailed description of moment condition construction

Let $R_t^{q,Nat}$ be average national wage or total national employment in industry q in year t, $N_{j,0}^q$ be the number of workers in each two-digit SIC industry q in the initial year (1990) in tract j, and $N_{i,0} = \sum_q N_{j,0}^q$ the ex-ante total employment in tract i. The labor demand shock is formed by interacting changes in wages or employment with ex ante local employment shares and summing across industries:

$$\Delta z_{jt}^{LD,R} = \sum_q \frac{R_t^{q,Nat} - R_0^{q,Nat}}{R_0^{q,Nat}} \cdot \frac{N_{j,0}^q}{N_{j,0}}$$

It is important to note that this demand shock embeds information on ex ante industry shares. When used as an instrument, an implicit assumption is that changes in non-productivity latent variables (e.g., amenities) are uncorrelated with prior industry structure. To ensure that local innovations in productivity do not drive national changes, I exclude all workers in California.

The demand shock instruments the change in wage to identify the slope of labor supply. Identification requires that the labor demand shock is uncorrelated with change in labor supply. Place of residence-by-year fixed effects control for changes in residential amenities that may be correlated with labor demand shocks (see Equation 12). The corresponding moment condition is:

$$\mathbb{E}[\Delta z_{jt}^{LD,R} \times (\Delta e_{jt} + \Delta d_{ijt})] = 0,\ \forall i, j \tag{M-1}$$

This moment condition identifies ϵ when the labor demand shock is uncorrelated with changes to workplace amenities and changes in the mean utility associated with commutes to that location. In fact, this can be weakened further by estimating place of work-by-year fixed effects to use as the dependent variable measuring labor supply conditional on commuting. I discuss and interpret this and the following identification assumptions later in this section.

A labor demand shock in one location shifts demand for housing in locations where workers might live (both by changing wage and by changing employment), and thus can be used to instrument changes in housing quantity to identify the slope of the housing supply curve. However, this requires mapping the labor demand shock to residential locations. I describe a housing shock to residential location i of the form $\Delta z_{it}^{HD,X} = \mathbf{z}_t^{LD,X} \cdot \boldsymbol{\vartheta}_i$ where the weights ϑ are predetermined functions of distance. I define the weight of interaction between any two places i and j as a parametric decay function of the predicted travel time between the tract centroids of the two locations, according to posted speed limits. Specifically,

$$\Delta z_{it}^{HD,R}(\rho) = \sum_s \frac{e^{-\rho\delta_{is}} \Delta z_{st}^{LD,X}}{\sum_s e^{-\rho\delta_{is}}}$$

where δ_{js} is the travel time between j and s, and ρ the spatial decay parameter. Intuitively, positive labor demand shocks near a location have, on average, a larger influence on housing demand than similar shocks farther away.[22] The resulting inverse-distance weighted labor demand shock identifies ψ, the inverse price elasticity of housing supply:

$$\mathbb{E}[\Delta z_{it}^{HD,R}(\rho) \times \Delta c_{it}] = 0,\ \forall i \tag{M-2}$$

Although both elements of M-2 relate to tract i, the housing demand shock draws on labor demand shocks from any j; I consider later how selecting particular subsets of these j may be useful.

Residents of a single location commute to many different locations for work. Consider the residents of a location i. Workers who live in i and work in j are sensitive to the housing demands of workers who work in j' that live in i; they share the local housing market. A labor demand

22. There is a similarity of this approach to the spatial weighting matrix used in spatial econometrics. A key difference is that I use a weighting matrix to develop instruments, rather than require that I correctly define the parametric relationship between locations.

shock to workers ij' can change the effective housing supply to workers ij. Thus labor demand shocks for ij' workers can be used to instrument changes in housing prices for ij workers and identify the slope of housing demand. To develop an average measure of the shocks for $ij', j' \neq j$, I employ inverse weighting as before, but excluding own tract j:

$$\Delta z^{HS,R}_{i(-j)t}(\rho) = \sum_{s \neq j} \frac{e^{-\rho\delta_{is}} \Delta z^{LD,X}_{st}}{\sum_{s \neq j} e^{-\rho\delta_{is}}}$$

Using the place of work-by-year fixed effects in Equation (12) to control for changes in workplace amenities, the following moment condition identifies $\epsilon(1-\zeta)$:

$$\mathbb{E}[\Delta z^{HS,R}_{i(-j)t}(\rho) \times (\Delta b_{it} + \Delta d_{ijt})] = 0, \; \forall\, i, j' \neq j \tag{M-3}$$

This instrument varies for every commuting pair. It is generally difficult to recover estimates of housing demand without microdata due to difficulties in quantifying housing services. Nonetheless, because tract pairs generate more variation than do individual tracts, this approach (along with an estimate of ϵ) recovers reasonable estimates of the household expenditure share on housing $1-\zeta$.

Finally, workers employed at place j observe the labor demand shock to $j' \neq j$, and may respond by leaving j for j'. This suggests that a labor demand shock at j' can be used to instrument changes in employment at j, functioning as a labor supply in j and identifying labor demand. But this is reflected through residential location, rather than through location at place of work. Consider residents in i: A positive shock to j' entices more workers from i the closer j' is to i, rather than the closer j' is to j. The following weighting uses this intuition and interacts with distance twice:

$$\Delta z^{LS,R}_{jt}(\rho) = \sum_r \left(\frac{e^{-\rho\delta_{rj}}}{\sum_r e^{-\rho\delta_{rj}}} \sum_{s \neq j} \frac{e^{-\rho\delta_{sr}} \Delta z^{LD,X}_{st}}{\sum_{s \neq j} e^{-\rho\delta_{sr}}} \right)$$

The own tract labor demand shock is excluded in order to remove mechanical correlation with local changes in productivity. The corresponding moment condition is:

$$\mathbb{E}[\Delta z^{LS,R}_{jt}(\rho) \times \Delta a_{jt})] = 0, \; \forall j \tag{M-4}$$

This gives an estimate of the share of production income that goes to non-labor expenses, $\alpha - 1$. The moment condition above provides an alternative way to estimate this parameter that is conceptually similar to competing characteristics instrument of Berry, Levinsohn, and Pakes (1995).

Because the instruments described above are all weighted averages of the labor demand shock, the identifying assumptions can be reframed in terms of more transparent moment conditions defined in terms of the labor demand shock (note that A-1 is identical to M-1):

$$\mathbb{E}[\Delta z_{jt}^{LD,R} \times (\Delta e_{jt} + \Delta d_{ijt})] = 0,\ \forall\, ij \tag{A-1}$$

$$\mathbb{E}[\Delta z_{jt}^{LD,R} \times \Delta c_{it}] = 0,\ \forall\, ij \tag{A-2}$$

$$\mathbb{E}[\Delta z_{j't}^{LD,R} \times (\Delta b_{it} + \Delta d_{ijt})] = 0,\ \forall\, ij' \neq ij \tag{A-3}$$

$$\mathbb{E}[\Delta z_{j't}^{LD,R} \times \Delta a_{jt}] = 0,\ \forall\, j' \neq j \tag{A-4}$$

Proposition 3. *Assume A1, A2, A3, and A4 are true, $\rho > 0$, $\mathbb{E}[\Delta z_{jt}^{LD,R} \times \Delta w_{jt}] \neq 0$, housing demand is downward sloping, and labor and housing supply are upward sloping. Then M1, M2, M3, and M4 are satisfied and the model is identified.*

Proof. Assumptions A-1 to A-4 are derived from M-1 to M-4 using the definitions of the instruments. The requirement that $\rho > 0$ ensures variation in the labor demand shock across space. The requirements are standard regularity conditions for identification in a system of simultaneous equations. □

Furthermore, the presence of data on wages at place of work and commuting flows in combination with Equation (16) suggests high-dimensional fixed effects may be useful to control for unobserved confounders. Assumptions A-1 and A-3 can be weakened to exploit this:

$$\mathbb{E}[\Delta z_{jt}^{LD,R} \times \Delta e_{jt}] = 0,\ \forall\, j \tag{A-1a}$$

$$\mathbb{E}[\Delta z_{j't}^{LD,R} \times \Delta b_{it}] = 0,\ \forall\, i \tag{A-3a}$$

Discussion and comparison to existing approaches

Assumptions A-1 and A-1a require that the local labor demand shocks be uncorrelated with changes in the local work amenity at j. They are weaker than either standard identification assumptions using shift-share labor demand shocks or those common the economic geography literature. When the labor demand shock is used as an instrument for observed employment or wage to trace out labor supply, identification requires the shocks be orthogonal to any non-wage determinants of labor supply (residential amenities, commuting costs, workplace amenities). In my notation, such a condition is $\mathbb{E}[\Delta z_{jt}^{LD,R} \Delta f_j(\mathbf{B}, \boldsymbol{\delta}, \mathbf{D}, \mathbf{E})] = 0$, where Δf_j averages over locations. In contrast, Assumptions A-1 and A-1a clarify the spatial requirements for identification and are robust to local correlation between improvements in residential amenities and the productivity shocks.

The economic geography literature typically identifies this key parameter from cross-sectional variation related to commuting. For example, Ahlfeldt et al. (2015) and Allen, Arkolakis, and Li (2015) condition on the time use parameter (κ) estimated from auxiliary models, and require that unobserved, origin-destination specific amenties are orthogonal to travel time. This is problematic if people of an unobservable type consistently select particular residential and workplace locations, and labor supply or commuting times are correlated with these types. Indeed, I later show that time-invariant pair-fixed effects explain 60% of the variation in commute flows over time, most of which cannot be explained by travel time or distance. Ahlfeldt et al. (2015) also require that there be no variation in workplace amenities or non-wage match quality ($\mathbb{E}[\ln(E_j)^2] = 0$). I later show this assumption is improbable. Monte, Redding, and Rossi-Hansberg (2015) specify production within a trade framework and recover productivity from cross-sectional trade flows and an assumed elasticity of substitution σ. They assume the productivity is orthogonal to workplace and origin-destition specific amenties to recover the labor supply elasticity.[23] Such indirect strategies are necessary because wage at place of work is typically unobserved. By comparison, Assumptions A-1 and A-1a utilize workplace wage data and the census of commuting flows, as well as rely on panel variation to control potentially confounding, persistent characteristics $\bar{e}_j$.

Assumption A-2 requires labor demand shocks be uncorrelated with changes in (inverse) housing productivity, Δc_{it}, which measures how efficiently developers provide housing density. If the labor demand shocks are correlated with these housing productivity innovations, ψ is not identified. One potential concern with Assumption A-2 is through labor reallocation: If labor demand shocks alter the pool of workers available for construction, there could be cause for concern.[24] Alternatively, if the productivity shocks lead changes in local regulatory conditions that make residential density easier (or harder) for developers to provide, the moment condition will not hold. To address these concerns, Assumption A-2 controls time-invariant confounding factors in $\bar{c}_i$; given the slow, frictional nature of land use change, this is significant. Further, I can modify Assumption A-2 to the own-tract labor demand shock:

$$\mathbb{E}[\Delta z_{jt}^{LD,R} \times \Delta c_{it}] = 0, \; \forall\, i \neq j \tag{A-2a}$$

This requires productivity shocks in a location be uncorrelated with innovations in nearby housing productivity. In fact, this could be further depending on concerns about specific types of endogeneity: If changes in local zoning in response to local productivity shocks were a concern, this condition could include only productivity shocks in other zoning districts (or cities). Labor demand shocks have been used to estimate the aggregate housing supply elasticity (Diamond 2016; Saiz 2010), though the use of spatially heterogeneous labor demand shocks within cities to identify

23. More precisely, Ahlfeldt et al. (2015) first estimate $\epsilon\kappa$, and then use the assumption that $\mathbb{E}[\ln(E_j)^2] = 0$ to match the variation in modeled wages to highly aggregated observable wages. Tsivanidis (2018) estimates κ from models of commuting, and runs additional models instrumenting τ with plausibly exogenous variation in transit connections to limit endogeneity concerns. Like Ahlfeldt et al. (2015), Tsivanidis (2018) also assumes $\mathbb{E}[\ln(E_j)^2] = 0$, but it does not impact identification of ϵ in his framework. Allen, Arkolakis, and Li (2015) estimate $\epsilon\kappa$, but then put more structure around time use to parameterize κ and identify ϵ. Monte, Redding, and Rossi-Hansberg (2015) assume $\mathbb{E}[\ln(A_j(\sigma)) \times \ln(E_j D_{ij})|\sigma, \kappa] = 0$. I use my notation when discussing other papers.

24. The model assumes that housing (the local good) does not require labor for production. Because housing stock is largely fixed, that seems reasonable for this relatively short panel (in terms of building life). For the inclusion of workers in local production, see Albouy (2016).

a local elasticity is new.

Assumptions A-3 and A-4 are less central, as the parameters they identify ($\tilde{\alpha}$ and $\tilde{\zeta}$) can be estimated from microdata. Nonetheless, they can be viewed as providing an additional test of identification. Assumption A-3 requires that labor demand shocks in one location do not change amenities and commute utility in *other* locations, and controls for persistent residential and commuting amenities, $\bar{b}_i$ and $\bar{d}_{ij}$. That is, any innovations to amenities in j or the unobserved commute utility between i and j must be uncorrelated with national innovations in labor demand and the share of industry in 1990 in a third location $j' \neq j$. A potential cause for concern is that high, positive labor demand shock in location j' could lead to more civic investment in public amenities in nearby locations i due to fiscal averaging within cities. Assumption A-3a is similar, but weaker in that the unobserved commuting shock does not enter. Assumption A-4 requires that labor demand shocks in one location be uncorrelated with nearby changes in productivity. Though tract fixed effects $\bar{a}_j$ control for most spatial correlation in industrial location, this identification assumption may not strictly hold. Nonetheless, estimates appear reasonable.

Identification with Agglomeration and Endogenous Land Use

The model presented in Section 2 abstracts away from agglomerative forces and endogenous land use determination. I summarize here how allowing these elements alters identification; details are in Appendix C. Identification of ϵ and ψ is still possible when agglomeration influences productivity and residential amenities, though $\tilde{\alpha}$ and $\tilde{\zeta}$ can no longer be identified without ex ante knowledge of the parameters that govern these forces. Agglomerative forces only confound demand elasticities; given the exogenous demand shocks, supply elasticities are still identified. Agglomeration tends to be highly path dependent, therefore fixed effects $\bar{a}_j$ and $\bar{b}_i$ control for most of these forces (Davis and Weinstein 2008). Furthermore, Ahlfeldt et al. (2015) show that these forces mostly dissipate within a few (five) minutes of travel time, limiting their role to confound.

Land use is relatively fixed and is primarily captured by the tract fixed effects $\bar{a}_i$ and $\bar{c}_i$. However, I observe measures of land use (zoning in 1990 and 2001) for housing and for production that allow me to test model robustness. The estimating equations for labor demand and housing supply are phrased in terms of employment and housing density, respectively. Assumptions A-2 and A-4 require productivity shocks to be orthogonal to changes in a location's ability to provide employment density and housing density. In other words, because these conditions are framed in terms of observable density by use, endogenous land use determination does not pose a direct threat to identification.[25]

25. Endogenizing land use is not trivial, however. It requires an additional market clearing condition, changes properties of the model equilibrium, and alters counterfactual simulations. Such changes make a small numerical differences due to the constancy of land use in this context. Note that if land use were unobserved, identification would become more challenging as land use becomes a latent term.

4.2 Estimating the effects of interventions

Estimates of $\{\tilde{\alpha}, \epsilon, \tilde{\zeta}, \psi\}$ permit recovery of local economic fundamentals by removing the simultaneous components of the supply, demand, and commuting equations. These economic fundamentals represent economic characteristics of a place that exist outside of a market equilibrium. In combination with market forces, these fundamentals determine the equilibrium prices and distribution of people. The fundamentals contain information about local productivity, housing supply, and transportation networks, and can be used to study how policy interventions shift supply and demand. This strategy removes concerns about (i) simultaneity and (ii) spatial interference through market forces, making it particularly useful when seeking to understand how a policy separately impacts supply and demand, or when a policy might exhibit price spillovers.

Consider a local intervention, T. In general, the intervention could impact any local fundamental. The following econometric framework permits estimating the effect of the intervention on local fundamentals:

$$n_{ijt} = \omega_{jt} + \theta_{it} - \epsilon\kappa\tau_{ijt} + \varsigma_{ij}^{D} + \lambda^{D} T_{ijt} + \varepsilon_{ijt}^{D} \tag{19}$$

$$\hat{\mathbf{Y}}_{it} = \boldsymbol{\lambda} T_{it} + \boldsymbol{\varsigma}_i + \boldsymbol{\varepsilon}_{it} \tag{20}$$

where $\boldsymbol{\lambda} = \{\lambda^A, \lambda^B, \lambda^C, \lambda^E\}$ and λ^D are the effects to be estimated. $\hat{\mathbf{Y}} = \{\widehat{\mathbf{a}}, \widehat{\mathbf{b}}, \widehat{\mathbf{c}}, \widehat{\mathbf{e}}\}$ contains the four non-commuting fundamentals. Standard econometric techniques (difference-in-difference, fixed effects, instrumental variables) can then be performed on the above system. While the full sample should be used to estimate the structural elasticities, the effects of interventions can be estimated using a restricted sample if needed to overcome selection bias. Economic theory may dictate additional restrictions on the system, i.e., some λ may be zero.[26] In Section 6, I discuss identifying and estimating treatment effects $\boldsymbol{\lambda}$.

5 Structural elasticities in Los Angeles: Results

Estimation of the structural elasticities can proceed either using Equations (11) to (13), or estimating tract-by-year fixed effects (ω_{jt} and θ_{it}) and using Equations (14) to (18). The advantage of the latter set of equations is that, for consistently estimated fixed effects, the identifying assumptions are weaker (Conditions A-1a and A-3a instead of A-1 and A-3). The cost is finite sample measurement error in the fixed effects and the potential loss of useful variation when using instruments that can vary by tract pair. I experiment with both approaches. When using the fixed effects approach, I estimate the place of work-by-year and place of residence-by-year fixed effects year by year from the commuting flow data; I discuss this in detail in the Appendix.

26. Interestingly, when some λ can be assumed equal to zero, and others are non-zero, treatment can be used as an instrument to identify some or all of the structural elasticities.

I first estimate the shape parameter ϵ, which corresponds to homogeneity in location preference; this parameter also represents a labor supply elasticity that conditions on commuting and residential geography. Because the instrument does not vary by tract pair, there is little benefit of using Equation (12) over (15). I proceed estimating ϵ from Equation 15, which uses place of work-by-year fixed effects as the dependent variable. Table 2 shows results using both the wage and employment variants of the instrumental variable. The first stage is sufficiently strong across all specifications, although it appears that using the wage instrument alone is preferable.[27] The value of ϵ varies between 0.50 and 1.03 depending on the specification. I take $\epsilon = 0.544$ from Column (1) as the preferred estimate, as it exploits variation only from the wage instrument and uses the largest sample. The low value of ϵ implies workers are quite heterogeneous in their location preferences.[28]

The estimates in Table 2 are substantially lower than values commonly found in economic geography models of commuting. Ahlfeldt et al. (2015) require $\mathbb{E}[\ln(E_j)^2] = 0$ to identify ϵ, implying a cross-sectional regression of $\hat{\omega}$ on log-wage at place of work should have an R^2 close to 1. I find an R^2 of 0.02. This heterogeneity is persistent: time-invariant fixed effects account for 83% of panel variation in $\hat{\omega}$, whereas changes in wage explain just 2%. While the large residual variation of $\ln(E_{jt})$ points to persistent sorting, differences in estimates of ϵ also relate to time: elasticities implicitly reflect a period during which adjustments can be made. To ensure counterfactuals are internally valid, it is important that elasticities correspond to the time horizon under study.[29] In fact, Allen, Arkolakis, and Li (2015) add more structure around this parameter and recover a similar value (0.71). The estimates here are also closer to more standard estimates of labor supply elasticities (e.g., Suárez Serrato and Zidar 2014; Falch 2010). This low value has important implications for studies of urban structure, as preference heterogeneity limits the locational responsivity of agents to changing local conditions.

The remaining structural parameters are identified using instruments constructed from the labor demand shocks and a spatial decay parameter, $\rho > 0$, that governs how the labor demand shocks propagate across space. I experiment with different values of ρ and settle on a range of $\ln(\rho) \in [-10, -2]$.[30] The labor demand shocks should propagate through the economy following the same decay as commuting, as these shocks will affect nearby markets only to the extent that workers are willing to commute to and from those markets. This implies $\rho = \epsilon\kappa$, and suggests using values of $\ln(\rho) \in [-7.5, -4.5]$.[31] Results are shown in Tables 3 through 5 for $\ln(\rho) = -7.5$,

27. Testing Hansen's J statistic does not reject the null that both instruments are uncorrelated with the error.

28. Technically, $\epsilon \in (0, 1]$ invalidates the use of Equation (2) for welfare calculation. In Section 7 and the Appendix, I show that this is a nonissue: Equation (1) is isomorphic to a multinomial logit expression for consumer welfare.

29. If Δt were very small, there would be no adjustment. Indeed, estimates of ϵ and κ from cross-sectional data likely capture a long-run relationship ($\Delta t \to \infty$). More elastic values correspond to longer adjustment horizons.

30. The value of ρ impacts efficiency but not identification for $\rho \in (0, \infty)$. As $\rho \to 0$, the spatial correlation of the shocks increases. In the limit, there is no variation in the instrument, and the system is not identified. On the other hand, as $\rho \to \infty$, shocks do not influence activity elsewhere (autarky), and the system is not identified.

31. Corresponding to estimates of $\epsilon\kappa$ that vary between -0.0005 and -0.011, depending on whether time-invariant

while Figure 1 provides a graphical representation over a range of values.

Instrumental variables estimates of the inverse housing supply elasticity from Equation (13) appear in Table 3. These estimates imply housing supply elasticities of about 0.68 when no adjustment is made for income-driven variation in quantity (Columns 1 and 2), and about 0.85 when income can influence housing quantity (Column 4). Including available residential land decreases these estimates.[32] Column 2 excludes the own tract labor demand shock when aggregating the instrument; this permits local housing productivity to covary with the local labor demand shock. The estimate coefficient is almost identical to Column 1. Panel A of Figure 1 shows how estimates of ψ from Column 4 of Table 3 respond to changes in ρ. Over the preferred range of decay values, point estimates are fairly constant. All results suggest that local, tract-level housing provision is inelastic in the Los Angeles region from 1990 to 2000. Saiz (2010) finds the median long-run inverse housing supply elasticity among major U.S. metropolitan areas to be about 1.75; his estimate for the Los Angeles area is 0.63. My estimates similarly point to limited medium-run scope for adjustment in local housing stock. This matches anecdotal and empirical evidence on the highly regulated California housing market (Quigley and Raphael 2005).

It is usually difficult to estimate household expenditure shares (which map to the elasticity of housing demand) or labor demand elasticities in urban models that use aggregated data. However, estimates in Tables 4 and 5 broadly concur with values derived from other sources, lending additional credibility to the identification strategy as a whole. Table 4 gives estimates of Equation (12) using $\Delta z^{HS,X}_{i(-j)t}$ as instruments for housing prices to determine $\epsilon(1-\zeta)$, the elasticity of housing demand. As when estimating the housing supply elasticity, I primarily use the employment variant of the instrument—it more strongly captures the transmitted effects of the labor demand shocks. The own tract can be excluded from the regression to limit concerns about the labor demand shock driving confounding changes in amenities. Results are generally insignificant and vary between -0.42 and -0.60 when the own tract is included, and between -0.61 and -0.83 when the own tract is excluded. Panel B of Figure 1 corresponds to the specification in Column 1. Point estimates imply a housing expenditures share of between 40% and 100% of income, higher than microdata suggest but not unreasonable at the lower for high cost areas.[33]

Finally, I estimate the inverse elasticity of labor demand ($\alpha - 1$) using demand shocks to nearby census tracts as an instrument. Results, shown in Table 5 and Panel C of Figure 1, vary between -0.20 and -0.39, implying labor's share of income is roughly 0.7. Column 2 of Table 5 includes the own-tract demand shock, $\Delta z^{LD,R}_{jt}$, as a control (recall that the instrument is $\Delta z^{LS,R}_{jt}(\rho)$). This

pair fixed effects are included; see the Appendix.

32. If the assumption of Cobb-Douglas housing production is correct, then the coefficient on land should be negative and equivalent in magnitude to the housing level coefficient, ψ. The coefficients are not statistically different in absolute value from each other.

33. If ζ were given, this could be used as an overidentification test for Assumption A-1. IPUMS microdata indicate that the median household expenditure share on *renting* is about 0.26 for this time period, though there are a number of differences in calculating income and housing costs that could explain the mean difference.

permits limited spatial correlation (to the extent the observed labor demand shocks are spatially correlated), and implies a slightly higher labor share of income. Column C includes the log measure of land zoned for productive uses.[34] Because the instrument uses spatial decay equivalent to ρ^2 instead of ρ, in Figure 1 the instrument becomes weaker at larger values of ρ.

The ability to generate reasonable estimates of α and ζ provides confidence in this interconnected approach to identification. Estimation of these parameters is more demanding than ϵ and ψ, both in terms of the stringency of the moment conditions and in the amount of exogenous variation needed to avoid weak instrument problems. Overall, these results suggest that interacting locally defined labor demand shocks with spatial structure can be used to create broad, omni-purpose tools for identifying local price elasticities.

6 The effects of the LA Metro on local fundamentals

I use Equations 19 and 20 to study the effects of the LA Metro Rail on local economic fundamentals. These fundamentals are informed by the structural parameter estimates in the previous section, as are the counterfactual simulations in the next section. I define two definitions of treatment, T: (i) tracts that either receive a transit station or whose centroid is within 500 meters of a transit station by the end of 1999, and (ii) tracts for which any part of the tract is within 500 meters of a transit station by the end of 1999.[35] The first, narrower definition encompasses 53 tracts, while the second encompasses 145 tracts. A map of stations is shown in Figure 2. Table 1 reports summary statistics on transit station location and ridership.

6.1 Difference-in-differences using historical plans, routes, and adjacencies

I use historical data to restrict the sample to treated census tracts and two subsets of untreated tracts that could have plausibly received treatment but did not. These serve as control groups for difference-in-difference analysis of the effects of the LA Metro Rail on bilateral commuting flows and other urban outcomes. First, I draw from *Comprehensive Rapid Transit Plan for the City and County of Los Angeles* by Kelker, De Leuw and Company in 1925.[36] The firm was tasked to create a feasible rail transit network to accommodate Los Angeles' booming population in the 1920s. The plan was eventually defeated for two primary reasons: (i) the transit system would have been

34. The model implies that the coefficients on land and employment should be equivalent in absolute value. They are statistically indistinguishable. I also estimate an alternative specification using agglomerative spillovers fixing a decay rate at 0.3167, as estimated in Ahlfeldt et al. (2015). Results over a grid of decay parameters values are typically similar. With agglomeration, estimates $\alpha - 1$ hover around -0.41. The strength of the agglomerative spillover is roughly $\mu = 0.05$ and only marginally significant with larger ρ.

35. The transportation literature typically draws circles of 400 to 800 meters around transit stations. The use of tract geography here does not allow such precision, but the two definitions are meant to roughly correspond to these distances. Additional results with alternative definitions are available upon request.

36. This map is available through the LACMTA library and is online at https://www.metro.net/about/library/archives/visions-studies/mass-rapid-transit-concept-maps/.

operated by private business, an arrangement that invited skepticism during this pre-Depression era of increased inequality; and (ii) elevated portions of the system were opposed by business interests (Fogelson 1967). I also utilize this document to derive the locations of all Pacific Electric Railroad (PER) lines installed in 1925. The PER was a system of at-grade standard gauge track and its associated transit service that served Los Angeles during the first half of the 20th century. LA Metro Rail has used its rights of way. Control tracts are mapped with treated tracts in Figure 2; the original maps from Kelker, De Leuw and Company (1925) are shown in Figure 3.

Control tracts are those that did not receive treatment and are within 500 meters of: (i) the Kelker, De Leuw and Co. proposal, "1925 Plan Sample"; or (ii) the PER's lines, "PER Sample". Identification of the effects of transit on local fundamentals requires a parallel trend assumption: Changes in local fundamentals experienced by control tracts are the same, on average, as those that received a transit connection would have experienced had transit not been built. This assumption is supported by several pieces of evidence. First, land assembly for transportation infrastructure is costly and difficult to coordinate. The control groups generally do not include places where land assembly for transit would have been prohibitively expensive. Second, the historical evolution of treated and untreated control tracts is similar, as proximity to historical transit generated path dependent changes in urban form (Brooks and Lutz 2016). Third, some control tracts are incidentally untreated; they are are along next to a transit line are not close to a station (it is suboptimal to locate transit stations too close to each other, Crampton 2000). Finally, the westward expansion of the Red and Purple lines was delayed by an unexpected geologic shock—the explosion of a clothing store due to natural methane seepage. The affected tracts are in both control groups, and the system is now expanding into this dense corridor.

Two validity checks are available. The first compares the exogenous labor demand shocks experienced by tracts that become treated with the control group. Similarity indicates that locations that received transit did not systematically undergo increased growth due to aggregate industrial trends. Table 6 reports results from the following model:

$$\Delta z_{jt}^{LD,R} = \beta \Delta T_{jt} + \varsigma_t + \Delta \varepsilon_{it} \tag{21}$$

Tracts are very similar under the narrow definition of treatment (Panel A). Under the broad definition of treatment (Panel B), treated tracts experience somewhat more negative shocks, though these are very small in magnitude. Differences disappear in specifications that include subcounty-by-year fixed effects to control for subregional trends.

Another test determines if parallel trends hold prior to treatment (pre-trends). Unfortunately, I cannot directly perform this test on commuting flows or local fundamentals – commuting data begins in 1990. I instead test for pre-trends in other, more readily available covariates from an alternative data source, Geolytics Neighborhood Change Database (NCDB). I use tract characteristics for 1970, 1980, and 1990.[37] Results presented in Table 7 report the relative changes in characteristics of treated tracts to controls by 1980 and 1990, and the p-value of the test of whether the effects in both years are jointly zero. All specifications include subcounty-by-year fixed effects. The results provide evidence of both similarity and divergence between treated and control locations. Treated locations show similar evolutions in the share of the population that is black, and in household income, housing units, and housing values. The covariates describing housing map to fundamentals, suggesting pre-trends are parallel. However, control tracts are also becoming more Hispanic, foreign-born, poorer, and less educated.

Because it is not clear how to map these demographic changes to fundamentals, I probe robustness in two ways. Throughout the main results, I allow for differential trends on baseline characteristics (1990 levels). I also utilize the spatial correlation in characteristics of census tracts to develop a third approach to estimating treatment effects using tract adjacencies and high-dimensional fixed effects in the manner of Dube, Lester, and Reich (2010). Because tracts (or tract pairs) are being compared only with adjacent tracts (or doubly adjacent tract pairs), the parallel trends assumption may be more credible. Estimates from this adjacency approach are prone to attenuation if the spatial scale of the effect is mismeasured, and so represent *a lower bound on the true effect size.* The Appendix describes this approach in more detail.

6.2 Effects of LA Metro Rail on commuting

Tables 8 through 10 present the first causal estimates of the effect of transit on bilateral commuting flows from variants of Equation (19):

$$n_{ijt} = \omega_{jt} + \theta_{it} + \varsigma_{ij} + \mathbf{s}_{s_i s_j t} + \lambda^D T_{ijt} + x'_{ijt}\beta + \varepsilon^D_{ijt} \tag{22}$$

where $T_{ijt} = T_{it}T_{jt}$ indicates whether an origin-destination tract pair is connected by LA Metro Rail. All specifications include residence-by-year, place of work-by-year, and tract pair fixed effects. The origin- and destination-by-year fixed effects control for tract level, time-varying characteristics that could influence treatment, like average income or employment. Some specifications include subcounty of residence-by-subcounty of work-by-year fixed effects, $\mathbf{s}_{s_i s_j t}$, and time-pair varying covariates, such as highway proximity and commute time.[38]

37. Because I retain the NCDB's 2010 tract geography, the results are not directly comparable to the main data. The test still serves as a proxy. Because there are more census tracts in 2010 than 1990, pre-trends tests include a larger sample.

38. I-105 expanded during this time. I include measures of highway proximity to control for confounding effects (Brinkman and Lin 2017).

Effects from 1990-2000

Table 8 indicates that the first stage of LA Metro Rail development led to an average increase of 11%-14% in commuting between tracts nearest transit stations (Panel A). Results are significant across specifications, and robust to the inclusion of controls and subcounty pair-by-year fixed effects. Under the broad definition of treatment (Panel B), the effect varies for 4%-8% and is generally significant. Highway controls have a small effect on coefficient estimates; including travel time does not alter results. The effect under the broad definition of treatment is smaller than with the narrow definition (commuters are more likely to utilize nearer stations). Fixed effects limit confounding factors; only pair-specific shocks to net-of-time commuting utility are potentially confounding. Adjacency estimates (reported in the Appendix) suggest a lower bound on the effect size of 8-9% for the narrow definition of treatment. Given the estimates of ϵ derived in the prior sections, this suggests a willingness to pay for transit of 13%-26% of income. At the lower end, this roughly accords with difference between average car ownership costs and the costs of using transit.

These results represent an alternative approach to evaluating the benefits of transportation infrastructure. Relative to the hedonic literature, the effect is solely attributable to commuting. This approach also diverges from the economic geography literature, which typically assumes the only pair-specific element that changes flows are travel times. While travel times influence commuting, there are many other determinants. To illustrate, the top panel of Figure 4 graphs the raw log share of commuters by travel time, and the bottom panel demeans the commuting flows by place of residence and place of work fixed effects. The standard approach fits a line to the data in Panel B to recover travel time disutility, and jettisons remaining variation (such an approach unavoidable if all commuting is not observed, e.g., Ahlfeldt et al. 2015). Yet there is far more variation conditional on travel time than between variation between the expected mean at minimum and maximum travel time. Between 75-80% of cross-sectional variation in commuting remains after controlling for origin and destination fixed and travel time.[39] Nearly 60% of this residual variation is persistent. This suggests that elements other than travel time may be more important than travel time in determining commuting flows, such as non-time commuting characteristics (such as reliability, Bates et al. 2001; Small and Verhoef 2007) or persistence in matching heterogeneous worker types and neighborhoods (LeRoy and Sonstelie 1983; Gabe and Abel 2011).

Some specifications condition on travel time, though of course travel time may respond to transit. While data suppression in the 2000 CTPP limits my ability to model carefully modal choice, I can tests for effects of transit on average commute time.[40] Table 9 shows the results of a

39. Monte, Redding, and Rossi-Hansberg (2015) find an R^2 of about 0.80 in their county-county commuting regression, while I find an R^2 of 0.20-0.25 in the analogous tract-tract commuting regression. The difference in explanatory power is likely due to spatial scale: distance is a powerful determinant of commuting at larger scales (county centroids are typically 20-30 miles apart), but less so within cities.

40. With unsuppressed data, the model could nest modal choice decisions: the structural residuals $\ln(D_{ijt})$ can be

regression model similar to Equation (22) but with travel time as the outcome. Transit has no effect on commute times, providing geographically disaggregate evidence that supports the finding in Duranton and Turner (2011) that transit does not appreciably reduce congestion. This result also suggests that back-of-the-envelope calculations in Anderson (2014) to infer the long-run benefits of transit from short-term service disruptions are too generous.[41]

Effects from 2002-2015

I use data from the 2002 and 2015 LEHD Origin-Destination Employment Statistics (LODES) to look for additional effects of transit on commuting flows in more recent years.[42] Because LA Metro Rail expanded during this period, I estimate a variant of Equation (22) on the LODES panel with two different effects: (i) *New Transit* is the effect of new stations (built after 2002) on bilateral commuting flows, while (ii) *Existing Transit* is the additional increase in commuting experienced by the stations built earlier (between 1990 and 2002). The results, shown in Table 10, indicate that new stations increase commuting by 9%-10% by 2015. While substantial, it is smaller than the effect of stations built from 1990-2000, likely because stations built after 2002 connect more suburban, less transit-friendly locations. Previously connected tracts continue to experience commuting growth, of another 7%-8% by 2015. This is evidence that either (i) aggregate commuting flows take decades to adjust to new transit modes (i.e., *habituation*), or (ii) that there are increasing returns in transit network size. I return to this distinction in Section 7.

6.3 Other margins of the LA Metro

While there is strong evidence that transit increases commuting between connected locations, there is little evidence it affects other margins: Transit does not consistently shift local fundamentals. Estimates of the effect of transit on productivity, residential amenities, housing productivity, and workplace amenities are shown in Table 11. Under the narrow definition of treatment, transit has no non-commuting effect after conditioning on regional trends and changes in highway structure. The zeros estimated for productivity are relatively large and negative, while point estimates on the other margins are small and insignificant. These results indicate that transit is not generating large-scale non-commuting benefits or costs within the immediate proximity of stations.[43]

thought of as the inclusive or log-sum value from lower nest mode choice logit model. Otherwise, fitting a modal choice model requires knowledge (or GIS predictions) of travel times associated with unchosen modes (as in Tsivanidis 2018). I rely on reported travel times, which are more realistic given the difficulty recreating historic travel conditions.

41. Anderson (2014) uses an unexpected, temporary labor strike that disrupted LA Metro Rail in 2003 to credibly identify the short-run effects on increased congestion.

42. The 2006/10 CTPP is difficult to compare with previous CTPP data because of changes in reporting standards, aggregation of data across years, and additional changes in geographies. Notably, it does not report comparable workplace wage.

43. Results using adjacencies, other elasticities, and the broad definition of treatment are in the Appendix. Estimates using the broad definition of treatment are generally smaller than with the narrow definition. As with the narrow definition, estimates are insignificant for productivity, residential amenity, and housing productivity fundamentals.

Surprisingly, there is no apparent effect of LA Metro Rail on (non-commuting) residential amenities in any specification. The implication is that hedonic estimates of the effect of transit reflect a commuting benefit rather than other related neighborhood amenities. [44] It is important to note that these results apply to LA Metro between 1990 and 2000; I cannot extend the non-commuting analysis to more recent years. The network was limited in size and connectivity at this time. As the transit network has expanded, it has become more valuable in terms of transportation connectivity. Responses that depend on scale, or are slower to respond (zoning laws can take decades to evolve), could potentially manifest in recent years. LA Metro Rail's primary effect between 1990-2000 is to expand commuter connectivity in Los Angeles: the city can accommodate more people with transit.

6.4 Robustness checks: sorting and land use

Transit users may differ from those who do not use transit (Glaeser, Kahn, and Rappaport 2008; LeRoy and Sonstelie 1983). If so, transit could induce equilibrium sorting. While the data limit the explicit addition of heterogeneity to the model, I find no evidence of differential trends in median household income between treated and control census tracts. Results are shown in Table 12, Panel A; point estimates are small and insignificant. Figure 5 shows the relationship between transit and rail usage by income centiles in 1990 and 2000, and reveals that there is little relationship between income and rail usage. This does not rule out changes in the distributions of income within census tracts, but shows that sorting of workers is not a first-order, medium-run impact.

The model assumes predetermined land use. While identification of the structural elasticities and transit effects are robust to this, counterfactual simulation may or may not be. I use SCAG zoning maps to test this channel and find little evidence of association between land use change and treatment. As Table 12 (Panel B) indicates, the change in residential land use is very small and statistically insignificant in richer fixed effects specifications. This is unsurprising given strict zoning and the relatively fixed nature of land use in urban settings due to the slow depreciation of buildings, and accords with the finding in Schuetz, Giuliano, and Shin (2016) that zoning hinders transit-oriented development near rail stations.

The effect on workplace amenity appears to be negative, though an effect on this margin is unexpected under the broad definition of treatment, when it was insignificant in Table 11. While this might represent an effect of transit (e.g. reduced parking near, but not immediately adjacent, to stations), more likely is a labor match story unrelated to transit.

44. See Ahlfeldt (2011), Ahlfeldt and Wendland (2009), Baum-Snow and Kahn (2000), Gibbons and Machin (2005), and McMillen and McDonald (2004). The results in Table 11 are consistent with Schuetz (2015), who shows that retail (a consumption amenity) does not increase near new rail transit stations in California between 1992 and 2009. This addresses a concern that transit serves to increase access to consumption amenities rather than just to jobs; this channel is not modeled. However, it does not appear to be a significant factor.

7 Welfare calculations and additional quantitative exercises

To estimate counterfactuals, I employ the succinct hat notation of Dekle, Eaton, and Kortum (2008), letting $\hat{X}_{it} = X'_{it}/X_{it}$ represent the relative change of an observed or estimated variable X under the counterfactual X'. This approach avoids using levels of fundamentals. Results are easily interpretable and given as a ratio to the observed price or population level. Furthermore, after solving the model in terms of updated equilibrium prices and populations, estimation of the counterfactual proceeds easily via an iterative algorithm that quickly finds a fixed point representing a counterfactual equilibrium. I lay out the algorithm in the Appendix.

7.1 LA Metro Rail and welfare

I estimate counterfactual values of $\hat{\mathbf{W}}, \hat{\boldsymbol{\pi}}, \hat{\mathbf{Q}}$ (and sometimes $\hat{\bar{N}}$) relative to observed data in 2000 under various combinations of the estimated structural elasticities $\{\alpha, \epsilon, \zeta, \psi\}$. Using estimates of the fundamental effects of transit from the preceding section, I define alternative scenarios by adjusting fundamentals so $\hat{X}_i = 1 - \lambda^X T_i$, for $X \in \{A, B, C, D, E\}$. The assumption of an open or closed city plays an important role. In a closed city, total population does not adjust. This means that there are real utility gains; these gains are equalized across the city through general equilibrium movements in prices. The model delivers a simple expression for welfare changes as a function of changes in local fundamentals and prices—a hat-notation variant of Equation (2):

$$\%\Delta \text{ Welfare } \approx \ln \hat{\bar{U}} = \frac{1}{\epsilon} \ln \left(\frac{\hat{B}_i \hat{E}_j \hat{D}_{ij} \hat{W}_j^{*\epsilon} \hat{Q}_i^{*-\epsilon(1-\zeta)}}{\hat{\pi}_{ij}^*} \right) \tag{23}$$

for any pair ij, where $\hat{X}^*$ indicates the equilibrium value of X in the counterfactual under autarky (that is, fixing $\hat{\bar{N}} = 1$). There is one technical issue: the expectation underlying this term is infinite for $\epsilon \leq 1$, even though Equation (23) can be calculated for any value of $\epsilon \neq 0$. Instead of enforcing this arbitrary parameter restriction, I show (in the Appendix) that Equation (23) can be isopmorphically expressed in a multinomial logit framework, with ϵ naturally taking the roll of marginal utility of income (Train 2009). This expands permissible values of ϵ; Equation (23) (and similar expressions) can be used for $\epsilon > 0$, not just $\epsilon > 1$. Because utility is homogeneous of degree one in wage, a proportional change in utility translates to an equivalent proportional change in wage. To convert this to levels, I multiply the proportional change in utility by the average annual wage ($31,563) and aggregate population of workers (6.73 million) in 2000. Annualized costs are $641 to $796 million.[45]

45. System cost for lines and stations completed by 1999 is $8.7 billion (2016 dollars). The annual operating subsidy for the rail portion of LA Metro's operations is about $162 million. I annualize the capital cost by assuming it is paid for in a loan with 30 year horizon and a 6% discount rate, giving an annualized capital cost of $635 million. Total annual costs are thus $796 million. Instead, assuming a 5% discount rate and 50 year horizon results in total annual costs of $641 million. Full details are in the Appendix.

Instead, if the city is open, aggregate population adjusts so that the expected utility in the city is equivalent to $\bar{U}$. Thus aggregate welfare for incumbent residents is unchanged: No spatial arbitrage means that the expected utility of city residence is $\bar{U}$ both before and after the change in fundamentals, so I instead report changes in total population. This statistic captures the change in the population the city can accommodate under transit with no change to utility.

Welfare effects by 2000

Table 13 reports the changes in aggregate welfare and population due to LA Metro in percentage and dollar terms. Columns (1) - (6) use $\lambda^D = 0.135$ and simulate the counterfactual removing this effect from tracts that receive treatment under the narrow definition. Column (1) uses preferred estimates and indicates an annual benefit of $246.4 million, a 0.12% increase of aggregate welfare. Under an open economy, the employed population of the Los Angeles region is larger by 0.27% with transit, or roughly 18,000 people. Columns (2) - (6) experiment with other elasticity values. These values are relatively unaffected by the housing supply elasticity (which varies between 0.62 and 0.85) or housing's share of income (varies between 0.30 and 0.76), but are highly influenced by ϵ. This is because ϵ maps changes in $\Lambda_{ij} = B_i E_j D_{ij}$ to utility according to $\hat{\Lambda}^{1/\epsilon}$, see Equation (23). For low values of ϵ, households are heterogeneous and stick to their chosen location pair, so large observed changes in commuting flows correspond to large effects on utility. Under high ϵ, households move in response to small changes in prices and fundamentals, and large changes in commuting flows do not correspond to large utility effects.

Columns (7) and (8) report results under the broad definition of treatment and the corresponding treatment effect, $\lambda^D = 0.058$. The specifications are more generous in that the set of tracts connected by transit are much larger. Because this increases the connected number of origin-destination pairs, welfare results are almost twice as large as when under the narrow definition despite the smaller value of λ^D. Column (7) pairs the broad specification with a low value of ϵ and produces the most generous non-dynamic results I report, indicating an annual benefit of $435 million. Under an open economy, this specification implies the employed population of Los Angeles is 0.34% larger with transit than it would otherwise be, about 23,000 people. The divergence between the closed and open equilibrium results under the narrow and broad definitions of treatment come from increased general equilibrium congestion in housing markets under the broad definition.

To show how the benefits of the subway change with network size, I perform a series of counterfactual exercises that simulate growing the network one station at a time. I grow the network in chronological order of station openings out from downtown Los Angeles. Results are shown in Figure 6. Each dot represents the cumulative benefit of a network including that station and all previously built stations; the color of the dot corresponds to that station's line. Panel A plots the results as a function of residential population, while Panel B shows the same results as a function of workplace population. Because employment hubs in downtown Los Angeles and near LAX receive treatment, the scale of Panel B is much greater (employment is more concentrated than housing). Figure 6 reveals significant differences in both the implied take up and benefits of individual stations. Some stations are near a great many jobs, but provide links for few residents. Panel A suggests linear returns to connecting residential population, while Panel B suggests increasing returns to connecting employment centers.

A general conclusion across all specifications is that the benefits of transit do not exceed costs by the year 2000. There are a number of aspects that this cost-benefit comparison misses: the benefits of LA Metro Rail to non workers and utility from non commuting trips. Further, if congestion has decreased as a result of transit, there may be additional benefits associated with increased traffic efficiency—though the reduced form results reported in Table 9 imply that this effect is not present. Decreased air pollution may provide an additional benefit; a generous estimate using parameters from Gendron-Carrier et al. (2018) and Los Angeles' mid-1990s birthrate suggests an additional gain of about $180 million annually.

Scale effects and dynamic response, through 2015 and beyond

The most significant omission of the 1990-2000 analysis is that early lines form the initial core of what has since grown to be a larger network. I use the dynamic results in Table 10 to calculate welfare impacts over an even longer run. To preserve the same cost base, I only consider the longer run benefits of stations built before 2000 for cost-benefit comparisons (the *Existing Transit* treatment). These tracts experience an additional commuting increase of 7.2% by 2015. This could be due to some combination of (i) slow habituation or (ii) returns to network size. Only habituation effects should be used to extend welfare calculations of the pre-2000 system, so I assign the entire *Existing Transit* effect to habituation to provide a generous bound.[46] Under these assumptions, the welfare benefit is $379 million annually, or an increase of 28,000 workers in an open city. Performing the same exercise using the broad definition of transit (and $\lambda^D = 0.105$) leads to a welfare benefit of $788 million annually, and is the only simulation in which the benefits of the system built by 2000 equal or exceed the costs. Taken together, this evidence suggests that the benefits of rail are unlikely to greatly surpass the costs in the first two to three decades after construction.

Panels C to F of Figure 6 extend the station-by-station analysis to all currently built and future stations and incorporating the effects from Table 10. New stations built after 2000 increase commuting between connected locations by 9.3% under the narrow definition of treatment, and 4.4% under the broad definition. Figure 6 interprets the *Existing Transit* effect as increasing returns to network size, and increases the commuting effect of existing stations by 7.2%/35 for each station built (thirty-five additional stations where constructed by 2015; details and related results are in the Appendix). These results assume that there are no other, non-commuting effects of transit

46. I run the same counterfactual as before, but with $\lambda^D = 0.207$.

(which I cannot test after 2000 because of data limitations). Since 2000, there are increasing returns to connecting residential population, but decreasing returns in workplace connections; this is the opposite pattern as before 2000. This highlights the importance of understanding not only where people live and work, but where the people who live in one location actually work; travel time alone is a poor measure of such idiosyncracies.

8 Conclusion

This paper develops and estimates an equilibrium model of a city wherein costly commuting connects housing and labor markets, and uses this model to estimate the welfare impacts of Los Angeles Metro Rail. The model is sufficiently parsimonious to permit transparent identification and estimation of all parameters, yet better reflects the observed spatial distribution of economic activity than commonly used market access approaches. The elasticity of labor supply plays a key role governing homogeneity in location preference. A small value indicates agents are relatively unwilling to relocate and are not very responsive to changes in local conditions or policies. Conversely, it also implies that observed responses to transit correspond to large utility gains. Estimates of the remaining elasticities are in line with previous studies, supporting the view that Southern California has a constrained housing supply.

I provide new insights into how transit influences city structure by isolating the commuting benefit of transit from other margins. LA Metro Rail increases commuting between census tracts near stations by 13.5% in the first decade after construction, relative to control groups selected by proposed and historical transit locations. There is little evidence of effects through other channels. Conservative welfare estimates point to a range of positive annual benefits of the system from $131 million to $247 million by 2000. These welfare benefits are smaller than the costs of LA Metro's light rail and subway lines, though more generous estimates using broader definitions of treatment come closer. I also provide evidence of dynamic effects due to increasing returns or habituation; if these effects are because of slow adjustment, the benefits of LA Metro Rail's network as of 1999 may just exceed costs. While these welfare estimates leave out some other benefits of transit (such as benefits for non-workers), the results here warrant a note of caution to cities expecting rail investment to lead to large increases in worker welfare within ten to twenty-five years.

References

Adão, Rodrigo. 2016. "Worker Heterogeneity, Wage Inequality, and International Trade: Theory and Evidence from Brazil."

Ahlfeldt, Gabriel. 2011. "If Alonso was right: modeling accessibility and explaining the residential land gradient." *Journal of Regional Science* 51 (2): 318–338.

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

Ahlfeldt, Gabriel M, and Nicolai Wendland. 2009. "Looming stations: Valuing transport innovations in historical context." *Economics Letters* 105 (1): 97–99.

Albouy, David. 2016. "What are cities worth? Land rents, local productivity, and the total value of amenities." *Review of Economics and Statistics* 98 (3): 477–487.

Allen, Treb, and Costas Arkolakis. 2016. "The Welfare Effects of Transportation Infrastructure Improvements."

Allen, Treb, Costas Arkolakis, and Xiangliang Li. 2014. "On the existence and uniqueness of trade equilibria."

———. 2015. "Optimal City Structure."

Anderson, Michael L. 2014. "Subways, strikes, and slowdowns: The impacts of public transit on traffic congestion." *American Economic Review* 104 (9): 2763–2796.

Bartik, Timothy J. 1991. *Who Benefits from State and Local Economic Development Policies?* Books from Upjohn Press. W.E. Upjohn Institute for Employment Research.

Bates, John, John Polak, Peter Jones, and Andrew Cook. 2001. "The valuation of reliability for personal travel." *Transportation Research Part E: Logistics and Transportation Review* 37 (2): 191–229.

Baum-Snow, Nathaniel. 2007. "Did highways cause suburbanization?" *Quarterly Journal of Economics* 122 (2): 775–805.

Baum-Snow, Nathaniel, and Matthew E Kahn. 2000. "The effects of new public projects to expand urban rail transit." *Journal of Public Economics* 77 (2): 241–263.

Baum-Snow, Nathaniel, Matthew E Kahn, and Richard Voith. 2005. "Effects of urban rail transit expansions: Evidence from sixteen cities, 1970-2000." *Brookings-Wharton Papers on Urban Affairs:* 147–206.

Bayer, Patrick, Nathaniel Keohane, and Christopher Timmins. 2009. "Migration and hedonic valuation: The case of air quality." *Journal of Environmental Economics and Management* 58 (1): 1–14.

Bayer, Patrick, and Christopher Timmins. 2005. "On the equilibrium properties of locational sorting models." *Journal of Urban Economics* 57 (3): 462–477.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile prices in market equilibrium." *Econometrica* 63 (4): 841–890.

Billings, Stephen B. 2011. "Estimating the value of a new transit option." *Regional Science and Urban Economics* 41 (6): 525–536.

Bowes, David R, and Keith R Ihlanfeldt. 2001. "Identifying the impacts of rail transit stations on residential property values." *Journal of Urban Economics* 50 (1): 1–25.

Brinkman, Jeffrey, and Jeffrey Lin. 2017. "Freeway Revolts!"

Brooks, Leah, and Byron Lutz. 2016. "Vestiges of Transit: Urban Persistence at a Micro Scale."

Chatman, Daniel G, and Robert B Noland. 2014. "Transit service, physical agglomeration and productivity in US metropolitan areas." *Urban Studies* 51 (5): 917–937.

Chen, Yihsu, and Alexander Whalley. 2012. "Green infrastructure: The effects of urban rail transit on air quality." *American Economic Journal: Economic Policy* 4 (1): 58–97.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2012. "The costs of agglomeration: House and land prices in French cities."

Crampton, Graham R. 2000. "Urban economic structure and the optimal rail system." *Urban Studies* 37 (3): 623–632.

Davis, Donald R, and Jonathan I Dingel. 2014. "The comparative advantage of cities." *NBER Working Paper* 20602.

Davis, Donald R, and David E Weinstein. 2008. "A search for multiple equilibria in urban industrial structure." *Journal of Regional Science* 48 (1): 29–65.

Davis, Morris A, and François Ortalo-Magné. 2011. "Household expenditures, wages, rents." *Review of Economic Dynamics* 14 (2): 248–261.

Dekle, Robert, Jonathan Eaton, and Samuel Kortum. 2008. *Global rebalancing with gravity: measuring the burden of adjustment.* Technical report. National Bureau of Economic Research.

Diamond, Rebecca. 2016. "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000." *American Economic Review:* 479–524.

Donaldson, Dave. 2018. "Railroads of the Raj: Estimating the impact of transportation infrastructure." *American Economic Review* 108 (4-5): 899–934.

Donaldson, Dave, and Richard Hornbeck. 2016. "Railroads and American economic growth: A "market access" approach." *Quarterly Journal of Economics* 131 (2): 799–858.

Dube, Arindrajit, T William Lester, and Michael Reich. 2010. "Minimum wage effects across state borders: Estimates using contiguous counties." *Review of Economics and Statistics* 92 (4): 945–964.

Duranton, Gilles, and Matthew A Turner. 2011. "The fundamental law of road congestion: Evidence from US cities." *American Economic Review* 101 (6): 2616–2652.

———. 2012. "Urban growth and transportation." *Review of Economic Studies* 79 (4): 1407–1440.

Epple, Dennis, Brett Gordon, and Holger Sieg. 2010. "A new approach to estimating the production function for housing." *American Economic Review* 100 (3): 905–924.

Epple, Dennis, and Holger Sieg. 1999. "Estimating Equilibrium Models of Local Jurisdictions." *Journal of Political Economy* 107 (4): 645–681.

Faber, Benjamin. 2014. "Trade integration, market size, and industrialization: evidence from China's National Trunk Highway System." *Review of Economic Studies* 81 (3): 1046–1070.

Falch, Torberg. 2010. "The elasticity of labor supply at the establishment level." *Journal of Labor Economics* 28 (2): 237–266.

Fogelson, R.M. 1967. *The Fragmented Metropolis: Los Angeles, 1850-1930.* University of California Press. ISBN: 9780520082304.

Gabe, Todd M, and Jaison R Abel. 2011. "Specialized knowledge and the geographic concentration of occupations." *Journal of Economic Geography* 12 (2): 435–453.

Gendron-Carrier, Nicolas, Marco Gonzalez-Navarro, Stefano Polloni, and Matthew A Turner. 2018. "Subways and urban air pollution." *NBER Working Paper* 24183.

Gibbons, Stephen, and Stephen Machin. 2005. "Valuing rail access using transport innovations." *Journal of Urban Economics* 57 (1): 148–169.

Glaeser, Edward L, et al. 2008. *Cities, agglomeration, and spatial equilibrium.* Oxford University Press.

Glaeser, Edward L, Matthew E Kahn, and Jordan Rappaport. 2008. "Why do the poor live in cities? The role of public transportation." *Journal of Urban Economics* 63 (1): 1–24.

Glaeser, Edward L, Jed Kolko, and Albert Saiz. 2001. "Consumer city." *Journal of Economic Geography* 1 (1): 27–50.

Gonzalez-Navarro, Marco, and Matthew A Turner. 2016. "Subways and urban growth: Evidence from Earth."

Heblich, Stephan, Stephen J Redding, and Daniel M Sturm. 2018. "The Making of the Modern Metropolis: Evidence from London."

Heilmann, Kilian. 2017. "Transit Access and Neighborhood Segregation: A Study of the Dallas Light Rail System."

Kahn, Matthew E. 2007. "Gentrification Trends in New Transit-Oriented Communities: Evidence from 14 Cities That Expanded and Built Rail Transit Systems." *Real Estate Economics* 35 (2): 155–182.

Lee, Sanghoon, and Jeffrey Lin. 2018. "Natural Amenities, Neighbourhood Dynamics, and Persistence in the Spatial Distribution of Income." *Review of Economic Studies* 85 (1): 663–694. doi:10.1093/restud/rdx018.

LeRoy, Stephen F, and Jon Sonstelie. 1983. "Paradise lost and regained: Transportation innovation, income, and residential location." *Journal of Urban Economics* 13 (1): 67–89.

Logan, John R, Zengwang Xu, and Brian J Stults. 2014. "Interpolating US decennial census tract data from as early as 1970 to 2010: A longitudinal tract database." *Professional Geographer* 66 (3): 412–420.

McFadden, Daniel. 1974. "The measurement of urban travel demand." *Journal of public economics* 3 (4): 303–328.

McMillen, Daniel P, and John McDonald. 2004. "Reaction of house prices to a new rapid transit line: Chicago's midway line, 1983–1999." *Real Estate Economics* 32 (3): 463–486.

Monte, Ferdinando, Stephen J Redding, and Esteban Rossi-Hansberg. 2015. "Commuting, migration and local employment elasticities." *NBER Working Paper* 21706.

Ottaviano, Gianmarco IP, and Giovanni Peri. 2012. "Rethinking the effect of immigration on wages." *Journal of the European Economic Association* 10 (1): 152–197.

Quigley, John M, and Steven Raphael. 2005. "Regulation and the high cost of housing in California." *American Economic Review* 95 (2): 323–328.

Saiz, Albert. 2010. "The geographic determinants of housing supply." *Quarterly Journal of Economics* 125 (3): 1253–1296.

Schuetz, Jenny. 2015. "Do rail transit stations encourage neighbourhood retail activity?" *Urban Studies* 52 (114): 2699–2723.

Schuetz, Jenny, Genevieve Giuliano, and Eun Jin Shin. 2016. "Does Zoning Help or Hinder Transit Oriented (Re)Development?"

Silva, JMC Santos, and Silvana Tenreyro. 2006. "The log of gravity." *Review of Economics and Statistics* 88 (4): 641–658.

Small, Kenneth, and Erik Verhoef. 2007. *The Economics of Urban Transportation.* New York, NY: Routledge.

Suárez Serrato, Juan Carlos, and Owen Zidar. 2014. "Who benefits from state corporate tax cuts? A local labor markets approach with heterogeneous firms." *NBER Working Paper* 20289.

Train, Kenneth E. 2009. *Discrete choice methods with simulation.* Cambridge University Press.

Tsivanidis, Nick. 2018. "The Aggregate and Distributional Effects of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio."

Figure 1: Estimated structural elasticities (except labor supply) for different values of ρ, the spatial decay of labor demand shock

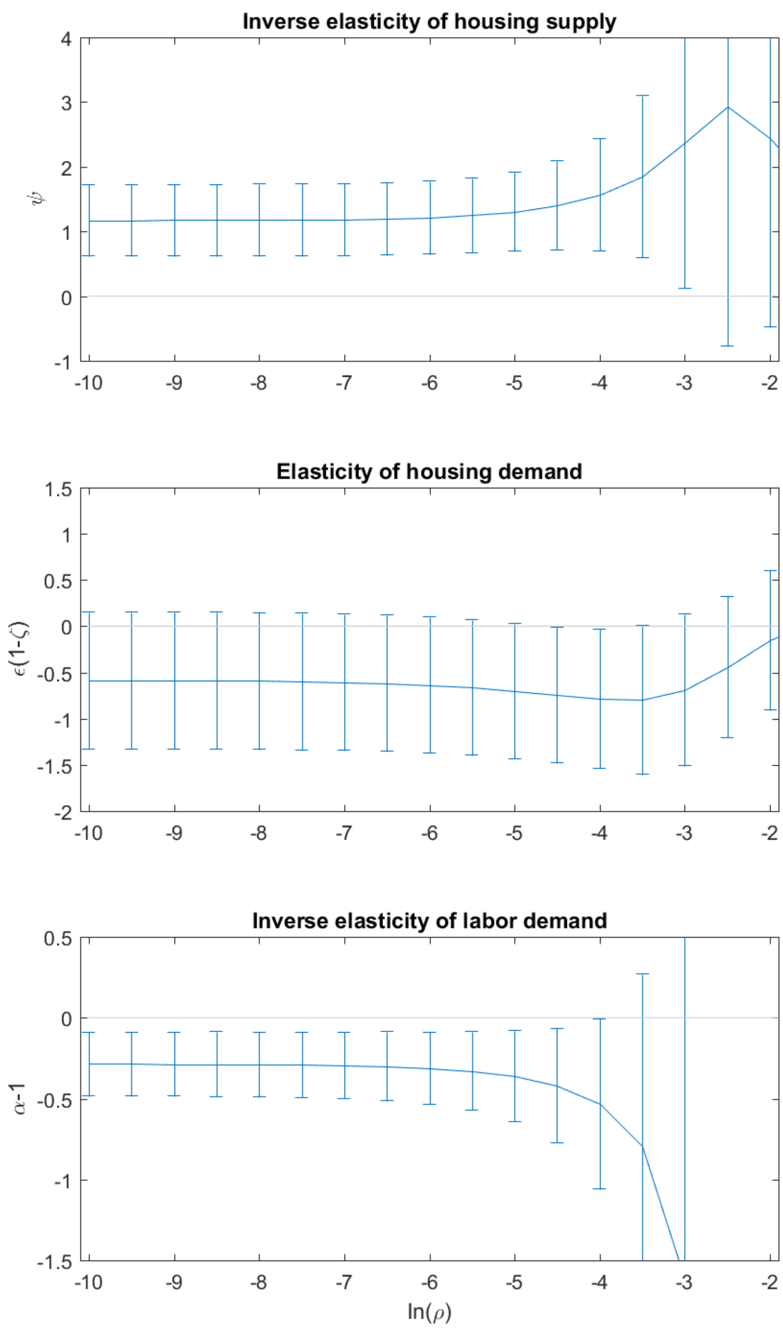


Figure 2: Map of LA Metro lines and stations and the 1925 Plan

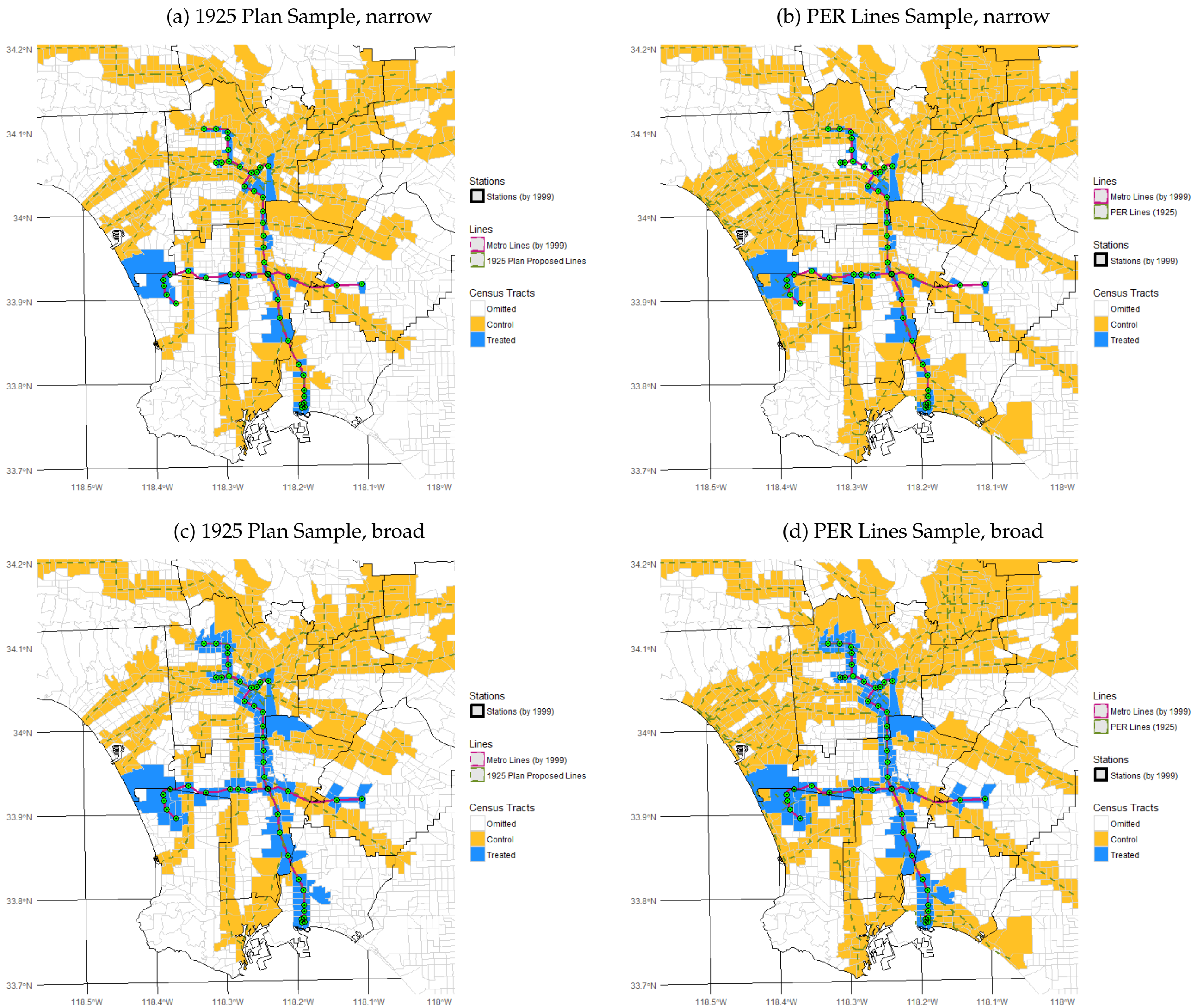


Figure 3: Map of Proposed LA Metro Lines and PER Lines in Kelker, De Leuw and Co. (1925)

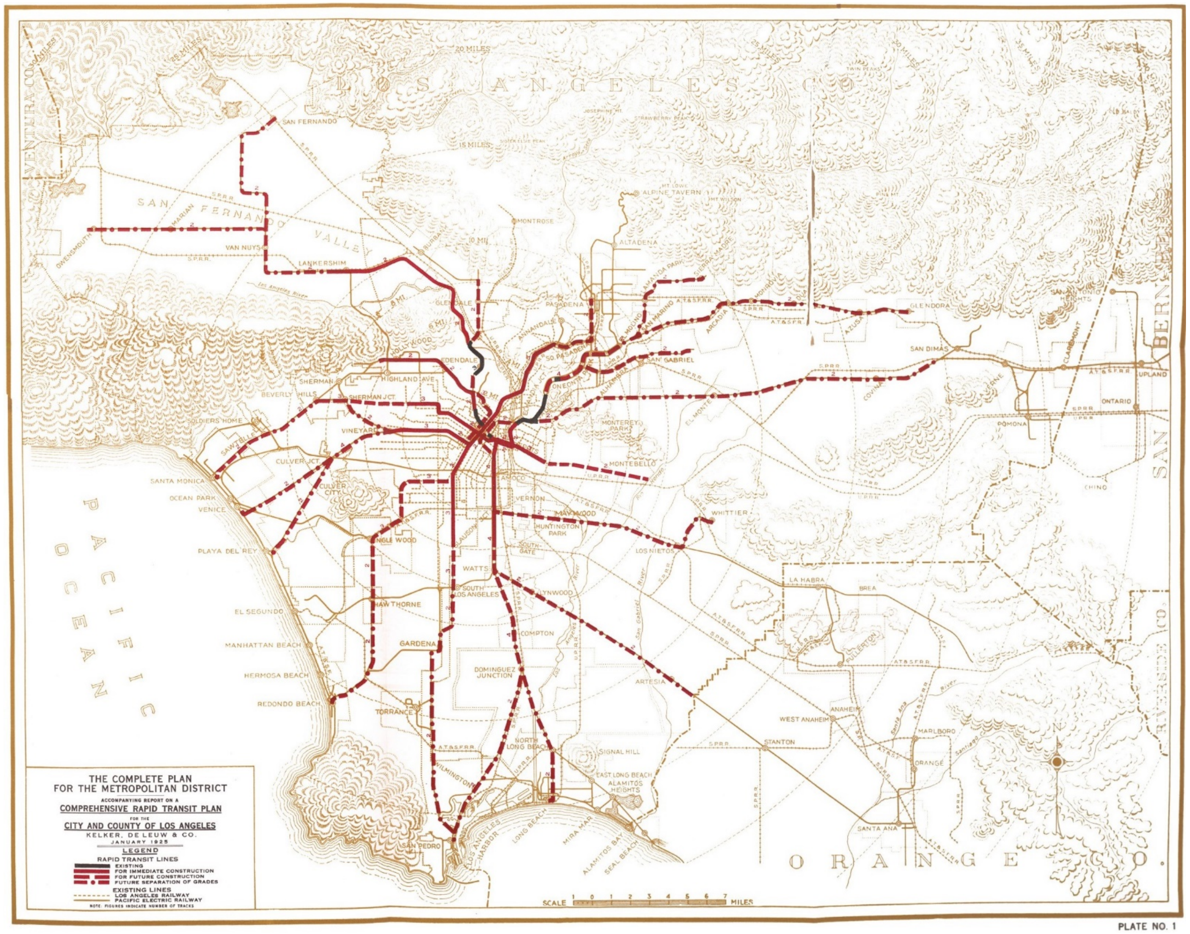


PLATE NO. 1

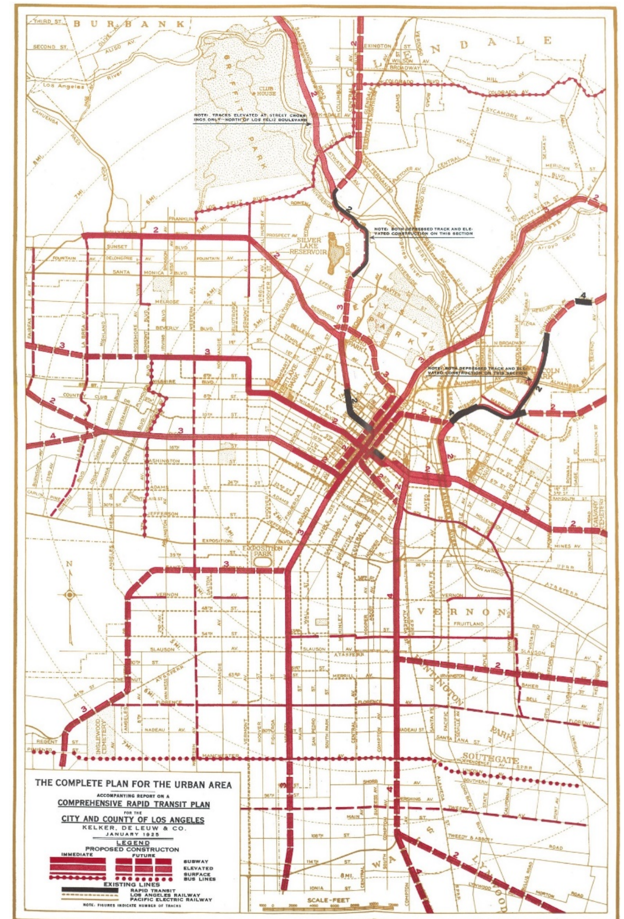


PLATE NO. 2

Figure 4: Heterogeneity in commute share by travel time, 1990. There is more variation conditioning on commute time than across commute times.

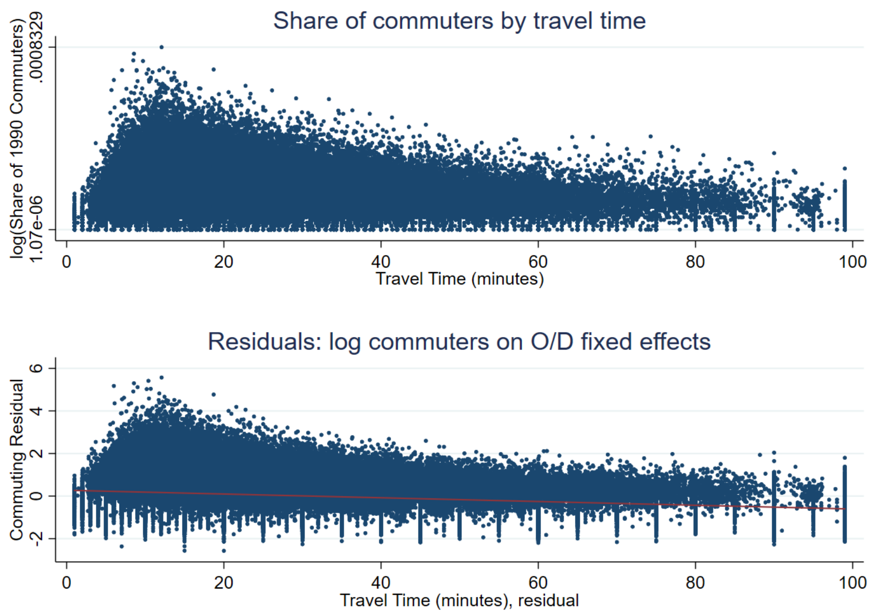


Figure 5: Take-up of LA Metro Rail for commuting does not vary by income, but overall take-up of transit (including bus) does.

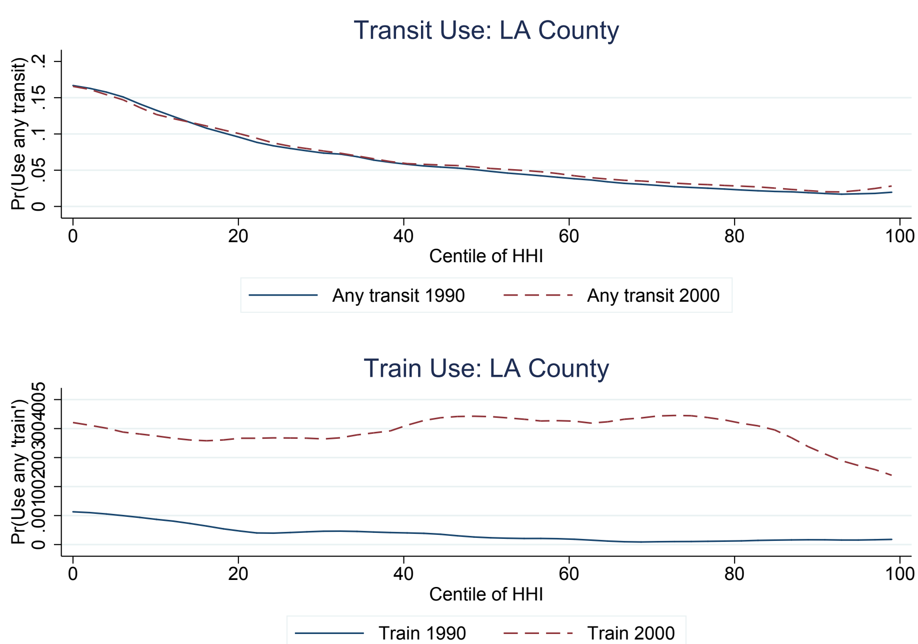


Figure 6: Cumulative benefits of network growth, dots are stations, colors represent subway line.

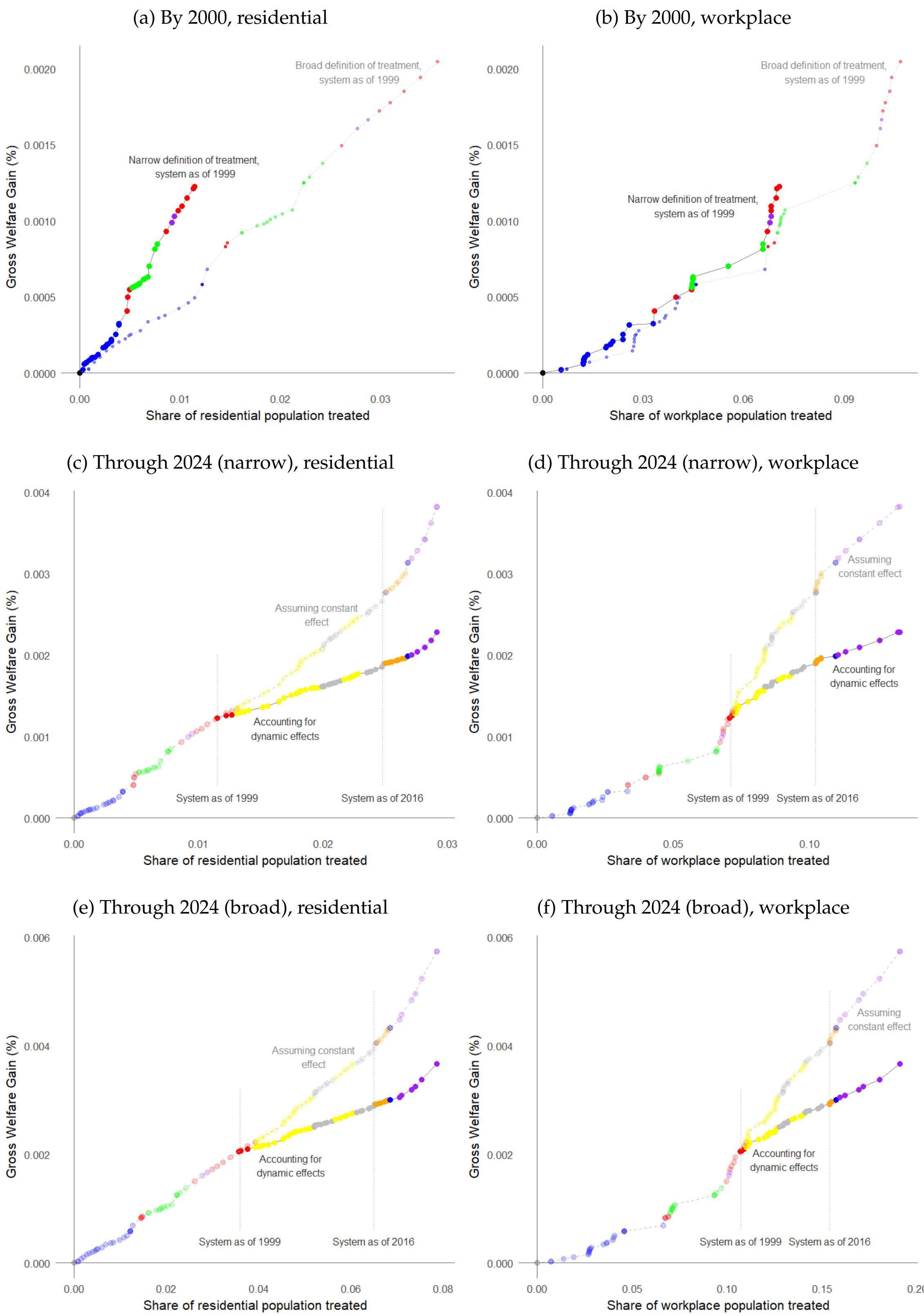


Table 1: Descriptive statistics on transportation in Los Angeles and station placement

	LA County		Full Sample	
	Centroid $< 500m$ (1)	Any $< 500m$ (2)	Centroid $< 500m$ (3)	Any $< 500m$ (4)
A. Pre-treatment tract characteristics (1990)				
% workers at POW tract that receive treatment	11.3%	19.5%	7.2%	12.3%
% workers at RES tract that receive treatment	2.7%	8.1%	1.6%	4.8%
% workers that receive transit connection RES-POW	0.6%	2.9%	0.4%	1.7%
% workers commuting via: Drive alone	71.8%		74.5%	
% workers commuting via: Carpool	15.8%		15.8%	
% workers commuting via: Bus	6.9%		4.6%	
B. Commuting characteristics				
No. commuters per tract pair (1990)	3.40 [7.51]			
No. commuters per tract pair (2000)	3.39 [8.28]			
Commute time (minutes, 1990)	26.3 [16.8]			
Commute time (minutes, 2000)	28.0 [18.3]			

Data from Census micro records (from IPUMS) and 1990 CTPP. LA County restricts analysis only to workers both living and residing in Los Angeles county, while the full sample includes all five counties in the main sample. Brackets indicate standard deviation. Commute times are weighted by flows.

Table 2: IV estimates of labor supply elasticity (ϵ)

	Mean utilty of place of work			
	$\hat{\omega}_{it}$ (1)	$\hat{\omega}_{it}$ (2)	$\hat{\omega}_{it}$ (3)	$\hat{\omega}_{it}$ (4)
$\ln$(Wage)	0.544**	0.501	0.774**	1.030*
	(0.210)	(0.412)	(0.214)	(0.407)
F-stat	19.87	17.48	17.14	13.39
F-stat (KP)	17.67	15.28	15.17	11.52
Wage instrument	X	X	X	X
Empl. instrument			X	X
ω estimated	Yr-by-yr	Panel	Yr-by-yr	Panel
N	4,822	4,708	4,822	4,708

Fixed effects IV estimates of regression of ω_{it} on w_{it}. All estimates include a tract fixed effect. KP refers to the Kleinbergen-Papp F statistic. Place of work-by-year fixed effects ω_{it} estimated separately (year by year) in columns (1) and (3), and together (in a single regression) in columns (2) and (4). Standard errors clustered by tract in parentheses: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 3: IV estimates of inverse housing supply elasticity (ψ)

	ln(House Value)				
	q_{it} (1)	q_{it} (2)	q_{it} (3)	q_{it} (4)	q_{it} (5)
ln(Population)	1.476**	1.465**	1.973**		
	(0.391)	(0.398)	(0.606)		
ln(Hous. Consump.)				1.181**	1.626**
				(0.284)	(0.451)
ln(Res. Land)			-1.183		-1.335
			(0.748)		(0.821)
Housing Supply Elasticity ($1/\psi$)	0.678	0.683	0.507	0.847	0.615
F-stat	26.95	27.58	16.89	23.55	14.02
F-stat (KP)	21.87	20.50	15.78	19.27	13.84
Empl. instrument	All	Not i	All	All	All
N	4,560	4,556	4,554	4,510	4,504

Fixed effects IV estimates of regression of median house value on population, housing consumption, and residential land, using $\ln(\kappa) = -7.5$ and employment IV. Reported estimates are of the inverse housing supply elasticity ψ, and F-statistics from the first stage (standard and Kleinbergen-Papp). All estimates include a tract fixed effect. Column (2) excludes own tract during instrument construction. Standard errors clustered by tract in parentheses: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 4: IV estimates of housing demand elasticity ($-\epsilon(1-\zeta)$)

	ln(Flow)				
	n_{ijt} (1)	n_{ijt} (2)	n_{ijt} (3)	n_{ijt} (4)	n_{ijt} (5)
ln(House Value)	-0.597	-0.593	-0.830*	-0.415	-0.613^{+}
	(0.377)	(0.377)	(0.381)	(0.317)	(0.318)
Travel Time		-0.0005**			
		(0.0001)			
F-stat	342.0	342.7	337.4	248.7	244.9
F-stat (KP)	229.3	229.6	227.3	213.3	211.2
POW-x-year FE	X	X	X	X	X
Empl. instrument	X	X	X	X	X
Wage instrument				X	X
Sample	All	All	Not ii	All	Not ii
N	287,748	287,748	282,902	287,748	282,902

Fixed effects IV estimates of regression of flows on median housing values, using $\ln(\kappa) = -7.5$. Reported estimates are of the housing demand elasticity $-\epsilon(1-\zeta)$, and F-statistics from the first stage (standard and Kleinbergen-Papp). All estimates include tract-of-work-by-year and tract-pair fixed effects. Columns (3) and (5) exclude the own tract from the regression. Standard errors clustered by tract pair in parentheses: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 5: IV estimates of inverse labor demand elasticity ($\alpha - 1$) using wage IV

	ln(Wage)		
	w_{jt} (1)	w_{jt} (2)	w_{jt} (3)
ln(Employment)	-0.291**	-0.202**	-0.393^{+}
	(0.103)	(0.072)	(0.238)
ln(Prod. Land)			0.357^{+}
			(0.200)
F-stat	5.24	6.39	7.25
F-stat (KP)	3.30	2.77	5.33
Own shock as control		Y	Y
N	4,884	4,884	4,862

Fixed effects IV estimates of regression of wage on employment, using $\ln(\kappa) = -7.5$ and wage instrument. Reported estimates are of the inverse labor demand elasticity ($\alpha - 1$), and F-statistics from the first stage (standard and Kleinbergen-Papp). All estimates include tract fixed effects. Standard errors clustered by tract in parentheses: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 6: Control group validity: labor demand shocks and treatment status

			1925 Plan Sample				PER Lines Sample			
	$z_{jt}^{LD,e}$ (1)	$z_{jt}^{LD,w}$ (2)	$z_{jt}^{LD,e}$ (3)	$z_{jt}^{LD,e}$ (4)	$z_{jt}^{LD,w}$ (5)	$z_{jt}^{LD,w}$ (6)	$z_{jt}^{LD,e}$ (7)	$z_{jt}^{LD,e}$ (8)	$z_{jt}^{LD,w}$ (9)	$z_{jt}^{LD,w}$ (10)
A. Tract centroid within 500 meters of station										
1[Transit]	-0.004	-0.002	-0.006	0.000	-0.003	0.000	-0.006	0.002	-0.002	0.001
	(0.006)	(0.003)	(0.006)	(0.007)	(0.003)	(0.003)	(0.006)	(0.007)	(0.003)	(0.003)
N	5,074	5,074	1,422	1,422	1,422	1,422	1,884	1,880	1,884	1,880
B. Any part of tract within 500 meters of station										
1[Transit]	-0.002	-0.003	-0.004	0.005	-0.004*	-0.000	-0.005	0.005	-0.004*	0.000
	(0.004)	(0.002)	(0.004)	(0.005)	(0.002)	(0.002)	(0.004)	(0.005)	(0.002)	(0.002)
N	5,074	5,074	1,458	1,458	1,458	1,458	1,924	1,920	1,924	1,920
Tract FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Subcty-x-yr FE	-	-	-	Y	-	Y	-	Y	-	Y

Each column of each panel presents the results of a different regression of the labor demand shock (measured in wage or employment) on treatment status, for twenty total. Regressions include year and tract fixed effects, and some include subcounty-by-year fixed effects. Standard errors clustered by tract in parentheses: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 7: Covariate balance prior to treatment, narrow definition of treatment

	% Black (1)	% Hisp. (2)	% Foreign Born (3)	ln emp (4)	% Transit (5)	% HS Grads (6)	% Coll. Grads (7)	Pov. Rate (8)	ln HHI (9)	ln hous. units (10)	ln hous. value (11)	ln ave. rent (12)	% Moved <5 yrs (13)
						A. 1925 Plan Sample							
1980×1[Transit]	0.008	0.021	0.036*	0.083^{+}	0.014^{+}	-0.027**	-0.006	0.017*	-0.002	0.007	-0.022	0.393	-0.028*
	(0.020)	(0.018)	(0.014)	(0.043)	(0.008)	(0.010)	(0.005)	(0.008)	(0.020)	(0.028)	(0.039)	(0.283)	(0.013)
1990×1[Transit]	0.018	0.046^{+}	0.042*	0.054	0.027*	-0.042*	-0.023**	0.020^{+}	-0.036	-0.017	-0.033	0.400	-0.017
	(0.027)	(0.025)	(0.017)	(0.049)	(0.012)	(0.017)	(0.008)	(0.011)	(0.031)	(0.039)	(0.047)	(0.280)	(0.012)
Joint test	[0.745]	[0.133]	[0.032]	[0.147]	[0.062]	[0.022]	[0.016]	[0.083]	[0.424]	[0.727]	[0.766]	[0.305]	[0.097]
N	2856	2856	2856	2856	2856	2856	2856	2856	2851	2856	2768	2231	2856
						B. PER Sample							
1980×1[Transit]	0.009	0.027	0.039**	0.087*	0.017*	-0.028**	-0.009^{+}	0.017*	0.003	0.007	-0.040	0.226	-0.028*
	(0.019)	(0.018)	(0.014)	(0.043)	(0.008)	(0.010)	(0.005)	(0.008)	(0.021)	(0.028)	(0.038)	(0.290)	(0.013)
1990×1[Transit]	0.009	0.065**	0.055**	0.071	0.031**	-0.049**	-0.028**	0.022*	-0.030	-0.019	-0.059	0.242	-0.022^{+}
	(0.026)	(0.024)	(0.017)	(0.050)	(0.011)	(0.017)	(0.008)	(0.011)	(0.031)	(0.038)	(0.047)	(0.287)	(0.012)
Joint test	[0.894]	[0.009]	[0.005]	[0.135]	[0.021]	[0.011]	[0.003]	[0.068]	[0.435]	[0.679]	[0.405]	[0.454]	[0.074]
N	3654	3654	3654	3654	3654	3654	3654	3654	3647	3653	3562	2792	3654

Each column of each panel presents the results of a different regression, for twenty-six total. Regressors include the narrow definition of treatment interacted with indicators for years prior to treatment. All regressions include time, tract, and subcounty-by-year fixed effects. P-values for joint tests given in brackets. Standard errors clustered by tract in parentheses: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 8: Effect of transit on commuting flows (λ^D) from 1990-2000, historical control groups, flows≥1

		1925 Plan Sample				PER Lines Sample			
	n_{ijt}	n_{ijt}	n_{ijt}	n_{ijt}	n_{ijt}	n_{ijt}	n_{ijt}	n_{ijt}	n_{ijt}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A. Narrow treatment definition: tract centroid within 500 meters of station									
1[Transit]	0.102*	0.138*	0.138*	0.126*	0.135*	0.124*	0.124*	0.112*	0.120*
	(0.049)	(0.055)	(0.055)	(0.055)	(0.055)	(0.053)	(0.053)	(0.053)	(0.053)
B. Broad treatment definition: any part of tract within 500 meters of station									
1[Transit]	0.044$^+$	0.065*	0.065*	0.047	0.058$^+$	0.076**	0.076**	0.059*	0.064*
	(0.025)	(0.031)	(0.031)	(0.032)	(0.033)	(0.029)	(0.029)	(0.029)	(0.030)
Tract Pair FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
POW-X-Yr FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
RES-X-Yr FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Travel Time	-	-	Y	Y	Y	-	Y	Y	Y
Highway Control	-	-	-	Y	Y	-	-	Y	Y
Sbcty-X-Sbcty-X-Yr FE	-	-	-	-	Y	-	-	-	Y
N	291,000	49,996	49,996	49,996	49,992	73,520	73,520	73,520	73,498

High-dimensional fixed effects estimates of λ^D. All estimates include tract of work-by-year, tract of residence-by-year, and tract pair fixed effects. Travel time is measured in minutes. Standard errors clustered by tract pair in parentheses: $^+ p < 0.10$, $^* p < 0.05$, $^{**} p < 0.01$

Table 9: Effect of transit on commute time, historical comparison

		Commute time, 1925 Plan Sample				Commute time, PER Sample			
	τ_{ijt}	τ_{ijt}	τ_{ijt}	τ_{ijt}	τ_{ijt}	τ_{ijt}	τ_{ijt}	τ_{ijt}	τ_{ijt}
	All	All	All	All	Drivers	All	All	All	Drivers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A. Tract centroid within 500 meters of station									
1[Transit]	-1.50	-0.93	-0.88	-0.66	-1.44	-0.74	-0.72	-0.44	-0.82
	(1.08)	(1.20)	(1.20)	(1.21)	(2.05)	(1.17)	(1.17)	(1.17)	(1.81)
B. Any part of tract within 500 meters of station									
1[Transit]	-0.83	-0.25	-0.19	-0.00	1.17	-0.14	-0.10	0.19	1.11
	(0.57)	(0.73)	(0.74)	(0.76)	(1.23)	(0.67)	(0.68)	(0.69)	(1.05)
Tract Pair FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
POW-X-Yr FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
RES-X-Yr FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Highway Control	-	-	Y	Y	Y	-	Y	Y	Y
Sbcty-X-Sbcty-X-Yr FE	-	-	-	Y	Y	-	-	Y	Y
N	311,340	53,468	53,468	53,468	15,090	78,424	78,424	78,400	23,562

High-dimensional fixed effects estimates of the effect of transit on commute times. All estimates include tract of work-by-year, tract of residence-by-year, and tract pair fixed effects. Travel time is measured in minutes, and is for all commuters or drivers only as indicated. Standard errors clustered by tract pair in parentheses: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 10: Dynamic effect of transit on commuting flows (λ^D) from 2002-2015, historical control groups, flows$\geq$1, narrow

			Commuting Flows, 1925 Plan Sample				Commuting Flows, PER Sample			
	n_{ijt}	n_{ijt}	n_{ijt}	n_{ijt}	n_{ijt}	n_{ijt}	n_{ijt}	n_{ijt}	n_{ijt}	n_{ijt}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(9)
A. Tract centroid within 500 meters of station										
1[New Transit]	0.102**	0.106**	0.093**	0.101**	0.085**	0.093**	0.081**	0.089**	0.080**	0.087**
	(0.023)	(0.023)	(0.025)	(0.025)	(0.025)	(0.025)	(0.024)	(0.024)	(0.024)	(0.024)
1[Existing Transit]		0.082**		0.080**		0.072**		0.077**		0.079**
		(0.024)		(0.026)		(0.026)		(0.025)		(0.025)
B. Any part of tract within 500 meters of station										
1[New Transit]	0.044**	0.050**	0.038**	0.052**	0.032*	0.044**	0.034**	0.045**	0.031*	0.044**
	(0.012)	(0.012)	(0.014)	(0.014)	(0.014)	(0.014)	(0.013)	(0.013)	(0.013)	(0.014)
1[Existing Transit]		0.059**		0.056**		0.047**		0.055**		0.059**
		(0.011)		(0.013)		(0.014)		(0.012)		(0.013)
Tract Pair FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
POW-X-Yr FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
RES-X-Yr FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Sbcty-X-Sbcty-X-Yr FE	-	-	-	-	Y	Y	-	-	Y	Y
N	1,993,198	1,993,198	237,438	237,438	237,424	237,424	351,688	351,688	351,658	351,658

High-dimensional fixed effects estimates of λ^D. All estimates include tract of work-by-year, tract of residence-by-year, and tract pair fixed effects. Travel time is measured in minutes. Standard errors clustered by tract pair in parentheses: $^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 11: Transit and non-commuting fundamentals, narrow definition of transit

		1925 Plan Sample				PER Sample			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				A. Effect on productivity λ^A with $\alpha - 1 = -0.291$					
1[Transit]	-0.127**	-0.068^{+}	-0.043	-0.053	-0.044	-0.086*	-0.056	-0.069^{+}	-0.053
	(0.034)	(0.035)	(0.038)	(0.037)	(0.037)	(0.035)	(0.036)	(0.037)	(0.036)
N	4,884	1,400	1,400	1,400	1,400	1,846	1,844	1,844	1,842
				B. Effect on residential amenity λ^B with $\epsilon(1-\zeta) = 0.597$					
1[Transit]	-0.002	0.006	-0.014	-0.022	-0.019	0.009	-0.004	-0.008	-0.007
	(0.027)	(0.027)	(0.030)	(0.033)	(0.033)	(0.027)	(0.029)	(0.031)	(0.031)
N	4,526	1,292	1,292	1,292	1,292	1,690	1,688	1,688	1,688
				C. Effect on inv. housing efficiency λ^C with $\psi = 1.181$					
1[Transit]	0.080*	0.035	-0.014	-0.008	0.011	0.050	-0.014	-0.014	-0.001
	(0.040)	(0.041)	(0.043)	(0.048)	(0.045)	(0.040)	(0.042)	(0.045)	(0.043)
N	4,494	1,274	1,274	1,274	1,274	1,660	1,658	1,658	1,658
				D. Effect on workplace amenity λ^E with $\epsilon = 0.544$					
1[Transit]	0.032	0.001	-0.007	-0.010	-0.013	0.012	-0.002	-0.003	-0.008
	(0.024)	(0.025)	(0.026)	(0.026)	(0.026)	(0.024)	(0.026)	(0.025)	(0.025)
N	4,822	1,400	1,400	1,400	1,400	1,844	1,844	1,844	1,842
Tract FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Subcty-by-Yr FE	-	-	Y	Y	Y	-	Y	Y	Y
Highway Control	-	-	-	Y	Y	-	-	Y	Y
Other Controls	-	-	-	-	Y	-	-	-	Y

Results from thirty-six regressions of transit proximity on estimated local fundamentals. Other controls are 1990 levels of log household income, percentage black, percentage with high school degrees, and percentage of local employment in manufacturing. Standard errors clustered by tract in parentheses: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 12: Transit, income change, and land use change, narrow

		1925 Plan Sample				PER Sample			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	A. Change in log household income								
1[Transit]	-0.003	0.012	0.012	0.010	0.009	0.003	0.006	0.007	0.009
	(0.019)	(0.019)	(0.020)	(0.021)	(0.021)	(0.019)	(0.020)	(0.020)	(0.021)
N	4,956	1,394	1,394	1,394	1,394	1,820	1,818	1,818	1,818
	B. Change in log residential land								
1[Transit]	-0.013**	0.001	0.001	0.000	0.000	0.000	0.001	0.001	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
N	4,950	1,396	1,396	1,396	1,396	1,858	1,854	1,854	1,852
Tract FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Subcty-by-Yr FE	-	-	Y	Y	Y	-	Y	Y	Y
Highway Control	-	-	-	Y	Y	-	-	Y	Y
Other Controls	-	-	-	-	Y	-	-	-	Y

Results from eighteen regressions of transit proximity on log household income. Other controls are 1990 levels of percentage black, percentage with high school degrees, and percentage of local employment in manufacturing. Standard errors clustered by tract in parentheses: $^{+}\ p < 0.10$, $^{*}\ p < 0.05$, $^{**}\ p < 0.01$

Table 13: Welfare estimates in 2000 (in $2016)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Parameters								
α	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680
ϵ	0.544	1.030	1.030	0.544	0.544	0.544	0.544	1.030
ζ	0.700	0.700	0.597	0.700	0.237	0.700	0.700	0.700
ψ	1.180	1.180	1.180	1.626	1.180	1.180	1.180	1.180
Change in fundamentals	Narrow	Narrow	Narrow	Narrow	Narrow	Narrow	Broad	Broad
λ_D	0.135	0.135	0.135	0.135	0.135	0.135	0.058	0.058
Closed Economy								
Annual Δ in welfare	0.116%	0.061%	0.062%	0.116%	0.116%	0.116%	0.205%	0.109%
(in millions of $2016)	246.4	130.5	130.7	246.4	247.2	247.1	435.3	230.5
Open Economy								
Population Δ	0.269%	0.143%	0.131%	0.260%	0.194%	0.219%	0.341%	0.181%
Annual cost (6%, 30yy)	-$797 mil.							
Annual cost (5%, 50yr)	-$641 mil.							

See text for description.

For Online Publication

Appendices and Supplemental Results

to accompany

Commuting, Labor, and Housing Market Effects of Mass Transportation: Welfare and Identification

by

Christopher Severen

June 2018

Appendix A: Discussion of data

In this Appendix section, I discuss all the sources of data that this project draws from and details relevant to sample construction. I pay particular attention to normalization. I also compare the CTPP and LEHD LODES data sources, and explain why they are not suitable to be used together.

A1. Sources

- Census Transportation Planning Project (CTPP)
  - 1990 Urban Part II: Place of Work, Census Tract
  - 1990 Urban Part III: Journey-to-Work, Census Tract
  - 2000 Part 2
  - 2000 Part 3
  - 2006/10 Part 3 (not used in current draft)
- National Historical Geographic Information System (NHGIS)
  - Shapefiles, Block Group and Census Tract, 1990, 2000, and 2010
  - Census, Block Group and Census Tract aggregates, 1990 and 2000
- Longitudinal Employer-Household Dynamics (LEHD) Origen-Destination Employment Statistics (LODES)
  - Aggregated to tract-to-tract flows, 2002 and 2015, using constant 2010 geographies
- Geolytics Neighborhood Change Database (NCDB)
  - Census aggregates in constant 2010 geographies from 1970-2010
- Los Angeles County Metropolitan Transportation Authority (LACMTA)
  - Shapefiles on LA Metro stations and lines
  - Opening dates for stations and lines
  - Ridership data
  - Kelker, De Leuw and Company (1925). I georeference this map in ArcGIS, and then processe it in R to provide geographic data to delineate the 1925 Plan and PER Line samples.
- IPUMS USA
  - Microdata on employment, wage, and industry by MSA for all non-CA residents, 1980-2000.
  - Microdata on transit in the 1990 and 2000 Censuses for LA area residents.
- Southern California Association of Governments
  - Land use and zoning maps: 1990, 1993, 2001, 2005.
- National Highway Planning Network
  - Shapefiles for the Century Freeway (I-105)

A2. Data construction details

See data map on following pages. Where there have been significant and arbitrary data processing decisions, I denote this by P#. See Figure A3 to reference data processing.

Geographic normalization

Through my primary analysis (all results from 1990 and 2000, excluding the check on pre-trends), the unit of observation is the census tract according to 1990 Census geographies. The Transportation Analysis Zones used in Southern California in the 1990 CTPP are equivalent to census tracts from the 1990 Census that have been subdivided by municipal boundaries if they overlay multiple jurisdictions. I merge TAZs in 1990 that cross municipal boundaries and assign them to the corresponding census tract. Data from the 2000 CTPP and 2000 Census are both in 2000 geographies. I therefore overlay shapefiles delineating 2000 geographies on 1990 census tracts to develop a crosswalk that translates 2000 data into 1990 geographies.[A.1] Where possible, I use 2000 block group data and shapefiles to refine the crosswalk. More precisely, to create the crosswalk, I intersect the 2000 census tracts and census block group files with 1990 census tracts, and then clean to provide a set of weights to be used in converting 2000 data to the 1990 geographies. Note that the intersection method varies according to whether summation or averaging is desired. If summing, weights are the portion of a 2000 geography that overlays the 1990 census tract. If averaging, weights are the portion of the 1990 census tract that is covered by a 2000 geography. In all cases, I excluded intersected values that cover less than 0.5% of the targeted area to reduce noise (P1).[A.2]

To normalize 2000 flows and travel times to 1990 geographies, the crosswalk is merged twice into the data, once by origin and once by destination (using the Stata command joinby to ensure all combinations were made). I then collapse this data by 1990 origin-destination pairs, taking the raw sum areal weights as the 1990 flow counts and using the areal weights to determine travel times. Many travel times are not disclosed in the 2000 data, and are treated as missing and are ignored. The 2000 CTPP data do not report actual counts, instead rounding to the nearest 5 (except for 1-7, which is labeled 4). In order to treat 1990 and 2000 data similarly, I develop two approaches that are conservative, though they throw away potentially useful variation. Both are similar, but differ in how they treat small numbers. In approach (P2a), I divide flows by 5, and round to the nearest digit. In approach (P2b), I change any flow values between 1 and 4 inclusive to be 4, and divide by 5 and round to the nearest digit. Small digits are different in the two years: in 1990, digits <4 have actual meaning, whereas in 2000 digits <4 can only have been created through the areal weighting process. Both approaches accommodate these differences in a different way, and offer different truncation points (2.5 for approach (a), and 1 for approach (b)). Approach (b) is my preferred specification. For all flows-by-mode, I follow approach (b), as not doing so would result in significant left-truncation. I also drop all pairs with a value of 0 in both 1990 and 2000 for approach (b) (P4). A small number of locations failed to merge. The flows in these amounted to 0.4% of the population.

I exclude census tracts from the eastern edges of San Bernardino and Riverside counties on the

A.1. This is essentially the reverse process of the Longitudinal Tract Data Base in Logan, Xu, and Stults (2014); I bring current data to 1990 geographies because merging tracts induces less error than (perhaps incorrectly) splitting tracts.

A.2. There are constant small realignments of census blocks (which aggregate to tracts) to account for roads, construction, lot mergers, etc. I choose the 0.5% threshold because it is unlikely that this represented a substantive change in the census tract, but rather just a minor border adjustment.

Channel Islands.

Labor demand shock construction

I construct wage and employment variants of the Bartik (1991) labor demand shock using Census microdata from 1990 and 2000. I exclude all workers in California. To create measures of national changes in labor demand, I calculate the change in wage or employment by two digit SIC industry from 1990 to 2000. I then interact this with the 1990 employment share by industry at each census tract of work to create a local measure of (plausibly exogenous) change in labor demand. While it would be preferable to use 1980 employment share by industry at tract of work, I have not been able to locate such data.

I then follow the approach described in Section 4 and interact the labor demand shock with the distance between tracts to model how the shock dissipates into adjacent markets. Because each tract may be joined to a different number of tracts, I weight by distance and exclude tracts that experience zero commuting flows (P3).

Data trimming

The various processes above produce relatively standardized data that accords reasonably well with ad hoc probes of quality. However, there are instances of extreme values that becomes influential observations during estimation. I experimented with a number of approaches to deal with this: (i) doing nothing, (iia) winsorizing in levels, (iib) trimming in levels, (iiia) winsorizing in changes, and (iiib) trimming in changes, where all winsorizing and trimming takes places at the 1st and 99th centiles. I ultimately settled on (iiib) trimming in changes, because it reduces the number of influential observations and removes observations with implausible-seeming characteristics from the data. I also remove observations with top-coded data where applicable. If a variable was top-coded differently in different years, I standardized the top code to the most conservative year.

Construction of treatment and control groups

The Dorothy Peyton Gray Transportation Library of LACMTA hosts historical data on proposed transit plans for the Los Angeles area, including the Kelker, De Leuw and Company (1925) plan. I obtain high-resolution digital copies of Plates 1 and 2 of this document and georeference them in ArcGIS using immutable landmarks and political boundaries. I then trace the proposed lines and the existing PER lines from this map, and convert these traces into shapefiles.

To define treatment status, I spatially join shapefiles on actual LA Metro Rail stations from LACMTA to both census tract centroids and boundaries. I define treatment in two ways:

- A narrower definition that requires that either (i) the distance from a tract boundary to a station be exactly 0, or (ii) the distance from a tract centroid to a station be less than 500 meters. Condition (i) implies that the stations lies within the census tract.

- A broader definition that requires just that the distance from a tract boundary to a station be less than 500 meters.

All treated tracts are included in all estimates. To develop a set of control tracts, I spatially join the shapefiles descended from the Kelker, De Leuw and Company (1925) document to the census tract shapefiles, and keep all tracts that have boundaries within 500 meters of the tracks. This assigns non-treated tracts to a control group for three different reasons: (i) they lie along spurs of proposed track that were never built, (ii) they are near a built track but distant from a station, (iii) they lie slightly farther away from stations than nearby treated tracts. Previous iterations of this paper have used alternative definitions of these control groups, but the use of a 500 meter boundary seems to provide the closest comparison. I perform this separately for 1990 tract geographies (for the main specifications) and 2010 tract geographies (for use with the NCDB and LEHD LODES).

A3. Fixed effects calculations

Residential and workplace fixed effects

Equation (16) ideally can be estimated using a high-dimensional fixed effects estimator on the following specification:

$$n_{ijt} = \omega_{jt} + \theta_{it} - \epsilon\kappa\tau_{ijt} + \varsigma_{ij}^{D} + \ln(D_{ijt})$$

However, some tract pairs do not have positive flows in both periods (recall that n reports log-flows). This results in these pairs being dropped from the estimating equation. The fixed effects ω_{jt} and θ_{it} have economic content, and are proportional to market access terms found in standard trade models. Furthermore, I use the fixed effects directly to identify structural elasticities. Omitting pairs with zero observations in one year may lead to inconsistent estimates of the fixed effect values. Instead, I estimate the equations separately for each year:

$$\begin{aligned} n_{ij0} &= \omega_{j0} + \theta_{i0} - \epsilon\kappa\tau_{ij0} + \ln(D_{ij0}) \\ n_{ij1} &= \omega_{j1} + \theta_{i1} - \epsilon\kappa\tau_{ij1} + \ln(D_{ij1}) \end{aligned}$$

The disadvantage of this is approach is that it does not allow for pair-specific time invariant fixed effects. However, travel time τ explains much of those fixed effects. Results are relatively robust to both approaches.

Permanence of unobserved determinants of commuting flows

Roughly half of the variation in commuting flows is due to unobserved, time-invariant characteristics. This details the analysis that leads to this conclusion. First, I run gravity models of commuting in the cross section with and without travel time:

$$n_{ij} = \omega_j + \theta_i - \epsilon\kappa\tau_{ij} + \ln(D_{ij})$$

The inclusion of travel time increases the R^2 from 0.20 to 0.26 in 1990, and 0.17 to 0.21 in 2000. This suggests that while travel time plays an important role, other factors are important.

Second, I run the panel gravity model with and without pair fixed effects:

$$n_{ijt} = \omega_{jt} + \theta_{it} + \varsigma_{ij}^{D} + \ln(D_{ij})$$

The model without pair fixed effects has an R^2 of 0.18 (or 0.21 if restricted to observations that have two non-zero commuting flows), but jumps to 0.81 with pair fixed effects. This regression excludes travel time. Regressing the pair fixed effects on travel time or network-based measures of travel cost:

$$\widehat{\varsigma_{ij}^{D}} = \epsilon\kappa\tau_{ij} + \ln(D_{ij})$$

lead to R^2 values of between 0.07 and 0.19.

Putting this together, pair fixed effects explain about 60% of the variation in the panel, conditioning on origin- and destination-by-year fixed effects. Only about 20% of these fixed effects are explained by commuting time, meaning 80% are not. Putting these together, we see that about 48% of commuting flows are explained by time-invariant, non-distance characteristics.

A4. CTPP vs. LODES

I draw data primarily from the CTPP. There are a number of advantages and a few disadvantages of the CTPP over another popular source of data, the Longitudinal Employer-Household Dynamic (LEHD) Origin-Destination Employment Statistics (LODES). The benefits of CTPP data:

1. In CTPP data, place of work is determined from household responses to a particular set of census questions. The response indicates where an individual worked in the week prior to the census, which may or may not correspond to a fixed establishment. LODES data come from federal tax records, and so identify people as working at the address on a firm's tax statement. Thus for firms with several establishments, there may be clustering at the mailing location that is not indicative of actual workplace. This is particularly true for large, multi-establishment firms.

2. The CTPP included median and mean wage at place of work prior in the 1990 and 2000 enumerations. LODES provides only a few large bins. Accurate measures of local wage at place of work are key to this analysis, and a novel contribution to the urban trade literature.

3. CTPP data include reported travel times. Thus, these estimates take into account congestion and other item unobservable to route planning GIS systems that may induce measurement error.

4. CTPP location data is accurately reported, while there is some geographic randomization (within block group) in LODES data to preserve confidentiality.

5. The CTPP data go back to 1990, while LODES does not begin until 2002. Thus, with CTPP I can fully capture commuting in 'pre' and 'post' periods.

Benefits of LODES data:

1. LODES data provide annual measures of commuting between locations since 2002, and the geocoding of workplace mailing address has a higher match rate than in the CTPP.

2. The CTPP has rather odd rounding rules that induce more measurement error in low commute-flow tract pairs. LODES has no such rounding rules (though there is geographic jittering).

3. LODES is calculated with consistent geography over time, while the CTPP is estimated using whatever geographies are decided upon by state census and transportation entities. This means that CTPP data must undergo geographic normalization, while LODES data does not.

There are two further disadvantages to the CTPP data: (i) not all fields from the 1990 and 2000 CTPP are reported in the 2006/10 CTPP. Important for this paper is the lack of wage at place of work data in 2006/10. (ii) Industry coding changed between the 1990 and 2000 census reports.

I have tried combining data sources to provide a more complete panel of commuting flows across time. There are a number of issues with this approach, namely concern that measurement error in flows drowns out meaningful variation in observed commuting flow changes over time. In fact, this seems to be the case when combining the 1990 CTPP with 2002 LODES data, or the 1990 and 2000 CTPP data with more recent LODES data. Further, the lack of wage at place of work data in LODES is a severe disadvantage. While I have experimented with alternative (fixed effects) methods to estimate wage at place of work, measurement error swamps meaningful measurement.

A5. Welfare and Cost Benefit Analysis

This section details the costs of the subway built by 2000. I do not track costs since 2000, as the calculation becomes much less clear with more recent data. To compare the costs and benefits of transportation interventions, I require annualized estimates of costs to compare with the annualized welfare benefits calculated in the text. Costs consist of two components: (i) the annualized cost of capital investment in rail, railcars, stations, and similar expenses, and (ii) net operating expenses (operating costs less revenues). Spreadsheet available by request.

Annualized Capital Expenditure

Cost information is from a consolidation of capital expenditures on lines built before 2000 from fiscal budgets.[A.3] After adjusting all costs to 2015 dollars, the total capital expenditure for the rail, rolling stock, and stations built prior to 2000 is $8.7 billion. To annualized this, I assume annual payments are made on this principal balance over a 30-year horizon with 6% interest rate (the interest rate used for some internal calculations by LA Metro). This gives an annualized capital cost of $634.6 million. This does not include other financing charges, the cost of planning, or some other expenses.

Operating Subsidies

Like most transit systems in the United States, LA Metro has incomplete farebox recovery, meaning that it subsidizes a portion of every ride. For rail in 2001, the farebox recovery ratio was about 20%. To estimate the welfare effects, I need to look at the *net* subsidy: operating costs less fare revenue. I have not found operating expenses from 1999 or 2000, but I do have expenses from 2001 and 2002. I use these numbers as a proxy. Rail (light and heavy) operations total $202.4 million in 2015 dollars, and rail fare revenue is $40.2 million. The net subsidy is $162.2 million per year.

Dynamic Welfare Results notes

Treatment effects are assigned in the following manner for the dynamic welfare effect graphs (Figure **??**) for narrow (broad) treatment effect:

A.3. Found here: `http://demographia.com/db-rubin-la-transit.pdf`.

- Stations built between 1990 and 2000:
  - Receive initial treatment effect of 0.135 (0.058)
  - After 2000, receive additional 0.072/35 (0.047/35) for the each of the next 35 stations build (until 2015)
  - After 2015, receive no additional benefits
- Stations built between 2000 and 2015:
  - Receive initial treatment effect of 0.093 (0.044)
  - After 2015, receive additional 0.072/35 (0.047/35) for the each of the next 35 stations build (until beyond sample)
- Stations built after 2015:
  - Receive initial treatment effect of 0.093 (0.044)

Appendix B: Proofs

Proposition 1

To establish Proposition 1i (existence), I utilize a fixed point argument and homogeneity. To establish Proposition 1ii, I make use of Theorem 1ii from Allen, Arkolakis, and Li (2014) (AAL) and the Perron-Frobenius Theorem.

Existence in a closed economy: Land use is assumed to be predetermined. Denote the set of location pairs with positive land use for housing and production as $\mathcal{C} = \{ij : L_i^H > 0 \text{ and } L_j^Y > 0\}$, and the cardinality of $\mathcal{C}$ as $N_{\mathcal{C}}$. Assume that $L_i^H > 0 \Leftrightarrow \sum_s \pi_{is} > 0$ and $L_j^Y > 0 \Leftrightarrow \sum_r \pi_{rj} > 0$. The model can be entirely expressed in terms of the aggregate population $\bar{N}$, the data on land use, local fundamentals, travel costs, and commuting shares $\{L_i^H, L_j^Y, A_j, \tilde{B}_i, C_i, D_{ij}, E_j, T_i, \delta_{ij}, \pi_{ij}\}_{\forall ij \in \mathcal{C}}$. Note that the commuting shares and aggregate population are endogenous, all else is given.

The commuting share from ij can be written as an implicit function of the vector of all commuting shares, population, exogenous variables, and models parameters: Define $\mathcal{T}_{ij}(\boldsymbol{\pi}; \bar{N})$:

$$\mathcal{T}_{ij}(\boldsymbol{\pi}; \bar{N}) = \frac{\frac{\Lambda_{ij}}{\delta_{ij}^{\epsilon}} \cdot \frac{\breve{A}_j^{\epsilon}}{\left(\bar{N} \sum_r \pi_{rj}\right)^{\epsilon(1-\alpha)}} \cdot \left(\bar{N} \breve{C}_i \cdot \sum_s \frac{\pi_{is} \breve{A}_s}{\left(\bar{N} \sum_r \pi_{rs}\right)^{1-\alpha}}\right)^{\frac{-\epsilon\psi(1-\zeta)}{1+\psi}}}{\sum_r \sum_s \frac{\Lambda_{rs}}{\delta_{rs}^{\epsilon}} \cdot \frac{\breve{A}_s^{\epsilon}}{\left(\bar{N} \sum_{r'} \pi_{r's}\right)^{\epsilon(1-\alpha)}} \cdot \left(\bar{N} \breve{C}_r \cdot \sum_{s'} \frac{\pi_{rs'} \breve{A}_{s'}}{\left(\bar{N} \sum_{r'} \pi_{r's'}\right)^{1-\alpha}}\right)^{\frac{-\epsilon\psi(1-\zeta)}{1+\psi}}}$$

with $\breve{A}_j = \alpha A_j {L_j^Y}^{1-\alpha}$ and $\breve{C}_i = (1-\zeta) C_i^{1/\psi} {L_i^H}^{-1}$. An equilibrium of the model is the vector $\boldsymbol{\pi}$ and aggregate population $\bar{N}$ such that $\boldsymbol{\pi}$ is a fixed point of $\mathcal{T}_{ij}(\boldsymbol{\pi}; \bar{N})$ and the no spatial arbitrage condition is satisfied. First, note that $\mathcal{T}_{ij}(\boldsymbol{\pi}; \bar{N})$ is homogeneous of degree zero in $\bar{N}$, so $\mathcal{T}_{ij}(\boldsymbol{\pi}; \bar{N}) = \mathcal{T}_{ij}(\boldsymbol{\pi})$ and the existence of commuting shares is independent of aggregate population.

Consider $\mathcal{T}_{ij}(\boldsymbol{\pi})$. By assumption, for all $ij \in \mathcal{C}$, we have $L_i^H > 0$, $L_j^Y > 0$, and $\sum_r \pi_{rj} > 0$ and $\sum_s \pi_{is} > 0$. This implies that $\pi_{ij} \geq 0$, and $\pi_{ij} \leq 1$ because π represent shares. Stacking equations, equilibrium commuting shares are a fixed point $\mathcal{T}(\boldsymbol{\pi}^{FP}) = \boldsymbol{\pi}^{FP}$. The function $\mathcal{T} : [0,1]^{N_{\mathcal{C}}} \to [0,1]^{N_{\mathcal{C}}}$ is continuous and maps a compact, convex set into itself. Therefore, by the Brouwer fixed point theorem, an equilibrium vector $\boldsymbol{\pi}^{FP}$ exists. In a closed economy, aggregate population is fixed, so this establishes existence.

Existence in an open economy: In an open economy, existence of equilibrium follows from *Existence in a closed economy*, but also the no spatial arbitrage that requires expected utility to be equalized to $\bar{U}$ in equilibrium. Denote element ij of $\boldsymbol{\pi}^{FP}$ be π_{ij}. Rewriting the no spatial arbitrage condition:

$$\bar{N} = \left(\frac{\bar{U}}{\Gamma\left(\frac{\epsilon-1}{\epsilon}\right) \cdot \left(\sum_r \sum_s \frac{\Lambda_{rs}}{\delta_{rs}^{\epsilon}} \cdot \frac{\breve{A}_s^{\epsilon}}{\left(\bar{N} \sum_{r'} \pi_{r's}\right)^{\epsilon(1-\alpha)}} \cdot \left(\bar{N} \breve{C}_r \cdot \sum_{s'} \frac{\pi_{rs'} \breve{A}_{s'}}{\left(\bar{N} \sum_{r'} \pi_{r's'}\right)^{1-\alpha}}\right)^{\frac{-\epsilon\psi(1-\zeta)}{1+\psi}} \right)^{1/\epsilon}} \right)^{\frac{1}{1-\alpha\left(1+\frac{\psi(1-\zeta)}{1+\psi}\right)}}$$

Given $\boldsymbol{\pi}^{FP}$, existence requires that the preceding equation give a real, finite value of $\bar{N}$. This is the case so long as $\epsilon > 1$ and $\alpha \neq \frac{1+\psi}{1+\psi(2-\zeta)}$.

Uniqueness: Rearranging the system in Equations (1), (4), (8), (9), and (7) into a more convenient form gives:

$$\begin{aligned}
W_j^{\frac{1+\epsilon(1-\alpha)}{1-\alpha}} \Omega_j &= \bar{N}^{-1} K_{0j} \sum_s W_s^{\epsilon} \Omega_s \\
\Omega_j &= \sum_r K_{1rj} Q_r^{-\epsilon(1-\zeta)} \\
Q_i^{-\epsilon(1-\zeta)-\frac{1+\psi}{\psi}} \Phi_i &= \bar{N}^{-1} K_{2i} \sum_s W_s^{\epsilon} \Omega_s \\
\Phi_i &= \sum_s K_{1is} W_s^{\epsilon+1}
\end{aligned}$$

where $K_{0j} = \breve{A}_j^{1/(1-\alpha)}$, $K_{1ij} = \Lambda_{ij}\delta_{ij}^{-\epsilon}$, and $K_{2i} = \breve{C}_i^{-1/\psi^2}$ are functions of predetermined parameters.

This transforms the model into the form of Equation 1 in AAL. Let $\mathbb{G}$ represent the matrix of exponents on the left hand side of the above system in the order (W, Ω, Q, Φ), and let $\mathbb{B}$ be the corresponding exponents on the right hand side:

$$\mathbb{G} = \begin{pmatrix} \frac{1+\epsilon(1-\alpha)}{1-\alpha} & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -\epsilon(1-\zeta) - \frac{1+\psi}{\psi} & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbb{B} = \begin{pmatrix} \epsilon & 1 & 0 & 0 \\ 0 & 0 & -\epsilon(1-\zeta) & 0 \\ \epsilon & 1 & 0 & 0 \\ \epsilon+1 & 0 & 0 & 0 \end{pmatrix}$$

Note that $\mathbb{G}$ is invertible. To address uniqueness, define $\mathbb{A} = \mathbb{B}\mathbb{G}^{-1}$ and $\mathbb{A}^+$ to be the element-wise absolute value of $\mathbb{A}$. That is,

$$\mathbb{A}^+ = \begin{pmatrix} \frac{\epsilon[(1-\alpha)-\mu]}{1+\epsilon(1-\alpha)} & \frac{1}{1+\epsilon(1-\alpha)} & 0 & 0 \\ 0 & 0 & \frac{\epsilon(1-\zeta)}{\epsilon(1-\zeta)+\frac{1+\psi}{\psi}} & \frac{\epsilon(1-\zeta)}{\epsilon(1-\zeta)+\frac{1+\psi}{\psi}} \\ \frac{\epsilon[(1-\alpha)-\mu]}{1+\epsilon(1-\alpha)} & \frac{1}{1+\epsilon(1-\alpha)} & 0 & 0 \\ \frac{(\epsilon+1)[(1-\alpha)-\mu]}{1+\epsilon(1-\alpha)} & \frac{(\epsilon+1)(1-\alpha)}{1+\epsilon(1-\alpha)} & 0 & 0 \end{pmatrix}$$

Theorem 1ii in AAL establishes that there is a unique equilibrium to the model if the spectral radius (largest eigenvalue) of $\mathbb{A}^+$ is less than or equal to one. Thus, uniqueness is established when $\rho(\mathbb{A}^+) \leq 1$.

Because $\mathbb{A}^+$ corresponds to a strongly connected graph and is nonnegative, it is irreducible. The Perron-Frobenius Theorem states that a nonnegative, irreducible matrix has a positive spectral radius with corresponding strictly positive eigenvector. So finding a condition under which $\rho(\mathbb{A}^+) \leq 1$ is identical to determining conditions under which $\mathbb{A}^+\mathbf{x} \leq \mathbf{x}$ for $\mathbf{x} \gg 0$. Solving the implied system of inequalities gives condition (10).[A.4]

A.4. To ensure the algebra is correct, I have numerically verified $\rho(\mathbb{A}^+) \leq 1$ iff Equation (10) holds.

Proposition 2

Existence: A_i is uniquely determined from:[A.5]

$$A_i = \frac{W_i}{\alpha} \left(\frac{\sum_r \bar{N} \pi_{ri}}{L_i^Y} \right)^{1-\alpha}$$

and C_i is uniquely determined from:

$$C_i = Q_i^{1+\psi} \left(\frac{L_i^H}{\sum_s \bar{N} \pi_{is} W_s} \right)^{\psi}$$

Define an excess demand function:

$$\mathcal{D}_{ij}(\mathbf{\Lambda}) = \pi_{ij} - \frac{\Lambda_{ij} W_j^{\epsilon} \left(\delta_{ij} Q_i^{1-\zeta} \right)^{-\epsilon}}{\sum_r \sum_s \Lambda_{rs} W_s^{\epsilon} \left(\delta_{rs} Q_r^{1-\zeta} \right)^{-\epsilon}} = 0$$

Note that $\mathcal{D}$ is continuous and homogeneous of degree zero. Homogeneity implies that $\mathbf{\Lambda}$ can be rescaled and restricted to the unit simplex: $\{\mathbf{\Lambda} : \sum_r \sum_s \Lambda_{rs} = 1\}$. This means that $\mathcal{D} : [0,1]^{N^2} \to [0,1]^{N^2}$. So $\mathcal{D}$ is a continuous function from a compact, convex set into itself; the Brouwer fixed point theorem guarantees existence.

Uniqueness: To establish uniqueness, note that by homogeneity of degree zero, we have $\sum_r \sum_s \mathcal{D}_{rs}(\mathbf{\Lambda}) = 0$. Define $M_{ij} = W_j^{\epsilon} \left(\delta_{ij} Q_i^{1-\zeta} \right)^{-\epsilon}$. The Jacobian of $\mathcal{D}$ has diagonal elements:

$$-\frac{M_{ij} \cdot \left(\left(\sum_r \sum_s \Lambda_{rs} M_{rs} \right) - \Lambda_{ij} M_{ij} \right)}{\left(\sum_r \sum_s \Lambda_{rs} M_{rs} \right)^2} < 0$$

and off-diagonal elements

$$\frac{\Lambda_{ij} M_{ij} M_{\{ij\}'}}{\left(\sum_r \sum_s \Lambda_{rs} M_{rs} \right)^2} > 0$$

where $\{ij\}'$ refers to an origin destination pair such that $i' \neq i$ and/or $j' \neq j$. Thus the aggregate excess demand function exhibits gross substitution, and equilibrium is unique.[A.6]

Proposition 3

Under the assumption that $e^{-\kappa \delta_{ij}}$ are fixed terms, $\Delta z_{it}^{HD,X}$, $\Delta z_{(-i)jt}^{HS,X}$, and $\Delta z_{it}^{LS,X}$ are linear combinations of $\Delta z_{jt}^{LD,X}$. Therefore, A2 maps directly into M2, and A4 maps directly into M4. Condition M3 imples $\mathbb{E}[\Delta z_{j't}^{LD,X} \Delta \ln(B_{it} D_{ijt})] = 0, \; \forall \, i, j' \neq j$, which holds given M1.

A.5. Uniqueness of A holds under agglomeration, the other terms are unaffected.

A.6. See Proposition 17.F.3 in Mas-Colell, Whinston, and Green. An alternative approach could be to use weak diagonal dominance of this positive matrix (following Bayer and Timmins (2005) but for weaker conditions).

Appendix C: Identification under Alternative Assumptions

In this section, I discuss identification under more general settings that those in the simplest version of the model presented in the paper. The two modification I consider are: (i) endogenous land use determination (no zoning), and (ii) the presence of agglomeration and other forces. As discussed in the paper, it is unlikely that either of these plays a significant role in the environment presented in the paper. Nonetheless, it is illustrative to work through these variations.

I first display the identification assumptions from the main text. Though these will continue to be necessary, they will not be sufficient.

$$\mathbb{E}[\Delta z_{jt}^{LD,R} \times \Delta \ln(E_{jt} D_{ijt})] = 0, \; \forall \, ij \tag{A-1}$$

$$\mathbb{E}[\Delta z_{jt}^{LD,R} \times \Delta \ln(C_{it})] = 0, \; \forall \, ij \tag{A-2}$$

$$\mathbb{E}[\Delta z_{j't}^{LD,R} \times \Delta \ln(B_{it} D_{ijt})] = 0, \; \forall \, ij' \neq ij \tag{A-3}$$

$$\mathbb{E}[\Delta z_{j't}^{LD,R} \times \Delta \ln(A_{jt})] = 0, \; \forall \, j' \neq j \tag{A-4}$$

and their simplified versions that accomodate the presence of rich fixed effects:

$$\mathbb{E}[\Delta z_{jt}^{LD,R} \times \Delta \ln(E_{jt})] = 0, \; \forall \, j \tag{A-1a}$$

$$\mathbb{E}[\Delta z_{jt}^{LD,R} \times \Delta \ln(C_{it})] = 0, \; \forall \, i \neq j \tag{A-2a}$$

$$\mathbb{E}[\Delta z_{j't}^{LD,R} \times \Delta \ln(B_{it})] = 0, \; \forall \, i \tag{A-3a}$$

Below, I describe additional conditions for identification, and their plausibility, under various changes to model form and data availability.

C1. Agglomeration in Productivity and Residential Amenity

To describe how the presence of agglomerative forces change identification assumption, define residential and productive spillovers as in Ahlfeldt et al. (2015):

$$\text{Productive agglomeration (A-augmenting):} \qquad \Upsilon_{jt} = \Upsilon\left(\sum_s k_{\Upsilon,js}\left(\frac{N_{st}^Y}{L_{st}^Y}\right)\right)$$

$$\text{Residential agglomeration (B-augmenting):} \qquad \Psi_{it} = \Psi\left(\sum_r k_{\Psi,ir}\left(\frac{N_{rt}^H}{L_{rt}^H}\right)\right)$$

where k here represent distance kernals and $N_i^H = \bar{N} \sum_s \pi_{is}$ is residential (employed) population.

If the parameters for the spillovers are known (of both the effects and the distance functions), then it is not necessary to develop new identification assumptions. Instead, the following substitutions can be made:

$$w_{jt} - \ln(\Upsilon_{jt}) \text{ for } w_{jt} \text{ in the labor demand equation}$$
$$\theta_{it} - \ln(\Psi_{it}) \text{ for } \theta_{it} \text{ in the housing demand equation}$$

Note that these equations reveal why the presence of these forces has little effect in this setting: they are mostly captured by the fixed effects $\bar{a}_j$ and $\bar{b}_i$.

If the spillovers are omitted from the model, additional moment conditions are required. Moment conditions presented in Assumptions A-1, A-1a, A-2, and A-2a do not change. Recall that those assumptions identify the key parameters of interest. Moment conditions corresponding to A-3, A-3a, and A-4 are tightened:

$$\mathbb{E}[\Delta z_{j't}^{LD,R} \times \Delta \ln(B_{it}\Psi_{it}D_{ijt})] = 0, \; \forall \; ij' \neq ij$$
$$\mathbb{E}[\Delta z_{j't}^{LD,R} \times \Delta \ln(B_{it}\Psi_{it})] = 0, \; \forall \; i$$
$$\mathbb{E}[\Delta z_{j't}^{LD,R} \times \Delta \ln(A_{jt}\Upsilon_{jt})] = 0, \; \forall \; j' \neq j$$

For these to hold, two additional assumptions are required in addition to Assumptions A-3 (or A-3a) and A-4:

$$\mathbb{E}[\Delta z_{j't}^{LD,R} \times \Delta \ln(\Psi_{it})] = 0, \; \forall \; i \tag{S-3}$$
$$\mathbb{E}[\Delta z_{j't}^{LD,R} \times \Delta \ln(\Upsilon_{jt})] = 0, \; \forall \; j' \neq j \tag{S-4}$$

If these conditions hold in addition to Assumptions A, the model is identified.

However, recall that instrument relevant requires $\mathbb{E}[\Delta z_{jt}^{LD,R} \times \Delta \ln(A_{jt})] \neq 0$. Both Ψ and Υ depend on nearby density, so to the extent location j' is near i or j, productivity shocks influence density and Assumptions S-3 and S-4 are unlikely to hold in a strict sense. However, they may hold approximately: There is significant autocorrelation in the population mass in locations from decade to decade. While this makes separately identifying agglomeration force difficult, in the context of the model presented here, this stickiness aids identification because much of $\Delta\Psi$ and $\Delta\Upsilon$ are captured by time-invariant tract fixed effects.

C2. Endogenous Land Use

If land use is observed (as here) and the amount of land used in housing and production is determined by market forces, no additional assumptions need be made for identification. This is not true for the theoretical model or counterfactual simulations; both would need to be modified with an additional market clearing condition to account for the additional degree of freedom.

One minor change in interpretation of parameter values must be made if land use is endogenous. The assumption of congestion in the relationship between land price and residential density can no longer be supported: $P_i^L \neq (H_i/L_i^H)^{\tilde{\psi}}$. This is because of the price of land also depends on the demand for land for production (and so congestion occurs through displacing employment instead of density costs). ψ has no role in this alternate model. However, because total output (housing) is observable, we can modify the model to derive an estimating equation very similar to that in the main paper.

Consider the developer's problem. Zero profits implies $Q_iH_i = P_i^L L^H + P^M M$, and the first order conditions deliver an expression for M under profit maximization. This results in the expression:

$$Q_iH_i = \frac{1}{\phi}P_i^L L^H$$

which just requires that a constant fraction of developer income be spent on land. Solving this for

P_i^L and substituting into Equation (6) and solving for Q_i delivers the equilibrium expression:

$$Q_i = \left(\frac{H_i}{L_i^H}\right)^{\frac{\phi}{1-\phi}} \mathfrak{C}_i$$

where $\mathfrak{C}_i = \frac{1-\phi}{\phi^2} P_M \tilde{C}_i^{1/(\phi-1)}$ contains the same elements as C_i. In fact, the estimating equation based on the above expression is isomorphic to that in the main text. Here, however, we identify $\frac{\phi}{1-\phi}$ instead of ψ. Note that under this interpretation, ϕ (the share of land in construction costs) is between 0.54 and 0.66, according to the estimates in Table 3. This is higher that a relatively standard value of 0.25 from Combes, Duranton, and Gobillon (2012), Epple, Gordon, and Sieg (2010), and Ahlfeldt et al. (2015). However, in Southern California land value anecdotally makes up high share of transacted real estate value. Alternatively, this could be seen as evidence in favor in immutable zoning.

As a quick aside, to complete the theoretical model, it is necessary to specify a land market clearing condition. I assume that the total land in a tract available for any use is fixed at $\bar{L}_i$; market clearing then requires $L_i^H + L_i^Y = \bar{L}_i$.[A.7] This condition can be rewritten (using Equation 4):

$$H_i \left(\frac{\mathfrak{C}_i}{Q_i}\right)^{\frac{1-\phi}{\phi}} + N_i^Y \left(\frac{W_i}{\alpha A_i}\right)^{\frac{1}{1-\alpha}} = \bar{L}_i$$

This equation, in conjunction with the model in the main text, is sufficient to pin down land use.[A.8]

C3. Agglomeration and Endogenous Land Use

Because endogenous land use did not alter identification, identification with both agglomeration and endogenous land use does requires the same assumptions as for the case with agglomeration: Assumptions S-3 and S-4 in addition to Assumptions A.

A.7. Note that this implies $\Delta L_{it}^Y = -\Delta L_{it}^H$.

A.8. Note that we can also rewrite this market clearing condition as an analytic expression of the observable prices, quantities, parameters, and the unobservable price of land:

$$\frac{\phi Q_i H_i}{P_i^L} + N_i^Y \left(\frac{(1-\alpha) W_i N_i^Y}{\alpha P_i^L}\right)^{\frac{1}{\alpha}} = \bar{L}_i$$

The price and land can be be calculated from this expression.

Appendix D: Checking robustness through adjacencies

I implement the adjacency strategy of Dube, Lester, and Reich (2010) to non-flow regressions by assigning each treated tract and the set of adjacent, but untreated tracts to a group g. Each group is permitted to have its own time trend. Thus, this approach treats each treated location as providing its own experiment and control group. Specifically, I estimate versions of the following model:

$$y_{igt} = \lambda T_{it} + \varsigma_i + \varsigma_{gt} + \varepsilon_{igt}$$

where y_{igt} is the outcome of observation i in group g in year t, where y_{igt} and ε_{igt} are repeated across groups as many times as necessary. Standard errors are clustered across i, and thus allow for both correlation across time and the correlation across groups induced by using some i in multiple g. Results generally support those using standard difference-in-differences estimates in the main paper, with the exception of the workplace amenity effect. See discussion in the main paper.

Porting the Dube, Lester, and Reich (2010) method to flow data is less straightforward and potentially problematic, as high dimension fixed effects already absorb much variation that would be captured by groups. It is not apparent how to define groups, either. For each treated pair ij, I form a group of controls consisting of all pairs of tracts that are adjacent to both treated tracts and for which one tract is not treated. Thus, for each ij there is a group g of $\#_i \times \#_j$ tract pairs, where $\#$ represents the number of controls plus the one treated tract. I estimate the following equations:

$$y_{ijgt} = \lambda T_{ijt} + \varsigma_{ij} + \varsigma_{gt} + \varepsilon_{ijgt}$$
$$y_{ijgt} = \lambda T_{ijt} + \varsigma_{ij} + \varsigma_{it} + \varsigma_{jt} + \varsigma_{gt} + \varepsilon_{ijgt}$$

The first maintains the spirit of a difference-in-difference estimator, but adds adjacency groups. The second model is in the spirit of a gravity model, with time-varying origin and destination fixed effects. Estimates under the narrow definition of treatment point to 0.08-0.09, though estimates from the second model are imprecise. Estimates under the broad definition of treatment are between 0.00-0.07, though imprecise. The lack of precision is due to little residual variation after all the fixed effects.

Appendix E: Counterfactual Estimation

First, note that the following hold:

$$\hat{W}_i = \hat{A}_i \hat{\bar{N}}^{\alpha-1} \left(\frac{\sum_r \pi_{ri}\hat{\pi}_{ri}}{\sum_r \pi_{ri}} \right)^{\alpha-1}$$
$$\hat{Q}_i = \hat{C}_i^{1/(1+\psi)} \left(\frac{\hat{\bar{N}} \sum_s \pi_{is}\hat{\pi}_{is} W_s \hat{W}_s}{\sum_s \pi_{is} W_s} \right)^{\psi/(1+\psi)}$$
$$\hat{\pi}_{ij} = \frac{\hat{B}_i \hat{D}_{ij} \hat{W}_j^{\epsilon} \hat{Q}_i^{-\epsilon(1-\zeta)}}{\sum_r \sum_s \pi_{rs} \hat{B}_r \hat{D}_{rs} \hat{W}_s^{\epsilon} \hat{Q}_r^{-\epsilon(1-\zeta)}}$$

where $\hat{\bar{N}} = 1$ in a closed economy. In the case of the open economy, aggregate population can adjust, ensuring no arbitrage between the city and outside locations. To account for this, define:

$$\hat{\bar{N}} = \left(\sum_r \sum_s \pi_{rs} \hat{B}_r \hat{D}_{rs} \left(\hat{A}_s \left(\frac{\sum_{r'} \pi_{r's}\hat{\pi}_{r's}}{\sum_{r'} \pi_{r's}} \right)^{\alpha-1} \right)^{\epsilon} \times \left(\hat{C}_r \cdot \left(\frac{\sum_{s'} \pi_{rs'}\hat{\pi}_{rs'} W_{s'} \hat{A}_{s'} \left(\frac{\sum_{r'} \pi_{r's}\hat{\pi}_{r's}}{\sum_{r'} \pi_{r's}} \right)^{\alpha-1}}{\sum_{s'} \pi_{rs'} W_{s'}} \right)^{\psi} \right)^{\frac{-\epsilon(1-\zeta)}{1+\psi}} \right)^{\frac{1+\psi}{\epsilon[(1+\psi)-\alpha(1+\zeta\psi)]}}$$

Simulating counterfactuals

I use the following algorithm to simulate counterfactuals:

1. Make an initial guess of wages and housing prices: $\{\hat{W}_i^{(0)}\}$, $\{\hat{Q}_i^{(0)}\}$. It is useful to set these equal to 1.
2. Estimate $\{\hat{\pi}_{ij}^{(0)}\}$ using $\{\hat{W}_i^{(0)}\}$, $\{\hat{Q}_i^{(0)}\}$, and $\{\pi_{ij}\}$.
3. Estimate $\hat{\bar{N}}^{(0)}$ using $\{\hat{W}_i^{(0)}\}$, $\{W_i\}$, $\{\hat{Q}_i^{(0)}\}$, $\{\hat{\pi}_{ij}^{(0)}\}$, and $\{\pi_{ij}\}$
4. Main Loop:
    (a) Define $\{\hat{Q}_i^{(temp)}\}$ using $\hat{\bar{N}}^{(t-1)}$, $\{\hat{W}_i^{(t-1)}\}$, $\{W_i\}$, $\{\hat{\pi}_{ij}^{(t-1)}\}$, and $\{\pi_{ij}\}$
    (b) Define $\{\hat{W}_i^{(temp)}\}$ using $\hat{\bar{N}}^{(t-1)}$, $\{\hat{\pi}_{ij}^{(t-1)}\}$, and $\{\pi_{ij}\}$
    (c) Define $\{\hat{\pi}_{ij}^{(temp)}\}$ using $\{\hat{W}_i^{(t)}\}$, $\{\hat{Q}_i^{(t)}\}$, and $\{\pi_{ij}\}$.
    (d) Define $\hat{\bar{N}}^{(temp)}$ using $\{\hat{W}_i^{(t)}\}$, $\{W_i\}$, $\{\hat{Q}_i^{(t)}\}$, $\{\hat{\pi}_{ij}^{(t)}\}$, and $\{\pi_{ij}\}$
    (e) Update $\hat{X}^{(t)} = \xi \hat{X}^{(temp)} + (1-\xi)\hat{X}^{(t-1)}$ for $\hat{X} \in \{\hat{Q}, \hat{W}, \hat{\pi}, \hat{\bar{N}}\}$, where ξ is a weight that disciplines updating.

(f) Estimate movement as:

$$\Delta = \sum_r |\hat{W}_r^{(t)} - \hat{W}_r^{(t-1)}| + \sum_r |\hat{Q}_r^{(t)} - \hat{Q}_r^{(t-1)}| + \frac{1}{N} \sum_r \sum_s |\hat{\pi}_{rs}^{(t)} - \hat{\pi}_{rs}^{(t-1)}| + |\hat{\hat{N}}^{(t)} - \hat{\hat{N}}^{(t-1)}|$$

(g) Stop when movement is below convergence criterion

Appendix F: Welfare under $\epsilon \leq 1$ (Frechet is Multinomial Logit)

First, I show that the expression in Equation (23) has an equivalent log-sum representation.[A.9] Begin by dividing counterfactual and factual expected utilities (from Equation 2):

$$\hat{\bar{U}} = \frac{\mathbb{E}[U'_{ijo}]}{\mathbb{E}[U_{rso}]} = \frac{\Gamma\left(\frac{\epsilon-1}{\epsilon}\right) \cdot \left(\sum_{\{ij\}} \tilde{\Lambda}'_{ij} \left(\delta'_{ij} Q_i'^{1-\zeta}\right)^{-\epsilon} (\tilde{B}'_i W'_j)^{\epsilon}\right)^{1/\epsilon}}{\Gamma\left(\frac{\epsilon-1}{\epsilon}\right) \cdot \left(\sum_{\{rs\}} \tilde{\Lambda}_{rs} \left(\delta_{rs} Q_r^{1-\zeta}\right)^{-\epsilon} (\tilde{B}_r W_s)^{\epsilon}\right)^{1/\epsilon}}$$

$$= \left(\frac{\sum_{\{ij\}} \tilde{\Lambda}'_{ij} \left(\delta'_{ij} Q_i'^{1-\zeta}\right)^{-\epsilon} (\tilde{B}'_i W'_j)^{\epsilon}}{\sum_{\{rs\}} \tilde{\Lambda}_{rs} \left(\delta_{rs} Q_r^{1-\zeta}\right)^{-\epsilon} (\tilde{B}_r W_s)^{\epsilon}}\right)^{1/\epsilon} \tag{F-1}$$

where $\{ij\} = \{rs\}$ track summation sets. Substituting in Equation (1) for some particular ij into the above twice (once for π'_{ij} and once for π_{ij}) and taking logs gives Equation (23).

From Train (2009), the change in consumer welfare due to changes of the characteristics of the elements in the choice set are:

$$\mathbb{E}[\bar{\mathcal{W}}'] - \mathbb{E}[\bar{\mathcal{W}}] = \frac{1}{\mu} \ln\left(\frac{\sum_{k \in K_1} e^{V'_k}}{\sum_{k \in K_0} e^{V_k}}\right) \tag{F-2}$$

where here μ is the marginal utility of income. Let:

$$\begin{aligned} V'_k &= \ln\left(\tilde{\Lambda}'_{ij} \left(\delta'_{ij} Q_i'^{1-\zeta}\right)^{-\epsilon} (\tilde{B}'_i W'_j)^{\epsilon}\right) \\ V_k &= \ln\left(\tilde{\Lambda}_{rs} \left(\delta_{rs} {Q_r}^{1-\zeta}\right)^{-\epsilon} (\tilde{B}_r W_s)^{\epsilon}\right) \\ \mu &= \epsilon \\ K_0 = K_1 &= \{ij\} = \{rs\} \end{aligned}$$

Taking logs of Equation F-1 then delivers Equation F-2. Note that $\mu = \epsilon$ is natural as ϵ already captures the utility effect of wage dollars. Thus the Frechet framework is identical to a multinomial logit framework where the utility from choice ij is:

$$\mathcal{U}_{ijo} = \ln\left(\tilde{\Lambda}'_{ij} \left(\delta'_{ij} Q_i'^{1-\zeta}\right)^{-\epsilon} (\tilde{B}'_i W'_j)^{\epsilon}\right) + \varepsilon_{ijo}$$

for ϵ_{ijo} distributed iid extreme value. In fact, this is very precisely (up to interpretation of amenity terms and trade costs) the specification often used in the discrete location choice literature (e.g.

A.9. Thanks to Wei You for noting that (23) and a log-sum expression are interchangeable:

$$\hat{\bar{U}} = \left(\frac{\hat{\tilde{\Lambda}}_{ij} (\hat{\tilde{B}}_i \hat{W}_j)^{\epsilon} \hat{Q}_i^{-\epsilon(1-\zeta)}}{\hat{\pi}_{ij}}\right)^{1/\epsilon} = \left(\sum_{\{ij\}} \pi_{ij} \hat{\tilde{\Lambda}}_{ij} \left(\hat{\delta}_{ij} \hat{Q}_i^{1-\zeta}\right)^{-\epsilon} (\hat{\tilde{B}}_i \hat{W}_j)^{\epsilon}\right)^{1/\epsilon}$$

Bayer, Keohane, and Timmins 2009). To map interpretation of the change in consumer welfare between the two frameworks, note:

$$\mathbb{E}[\bar{\mathcal{W}}'] - \mathbb{E}[\bar{\mathcal{W}}] = \ln \mathbb{E}[U'_{ijo}] - \ln \mathbb{E}[U_{ijo}] = \ln \hat{\hat{U}} \approx \%\Delta \text{ Welfare}$$

That is, welfare change is naturally expressed in relative terms (rather than monetary terms) when used with Frechet framework. Equation F-2 only requires $\epsilon > 0$, and so Equation (23) can be used for welfare evaluation when $\epsilon \in (0, 1]$ and well as $\epsilon > 1$.

Appendix Figures and Tables

Figure A1: Glossary of variables and parameters

Parameters	Interpretation
ϵ	Homogeneity of location preferences (and wage elasticity of labor supply)
ζ	Household expenditure share on non-housing goods
$\tilde{\zeta} = -\epsilon(1-\zeta)$	Price elasticity of housing demand
α	Share of (production) income spent on labor
$\tilde{\alpha} = \alpha - 1$	Inverse wage elasticity of labor demand
ϕ	Share of housing income spent on land
$\tilde{\psi}$	Congestive cost of housing
$\psi = \tilde{\psi}\phi$	Inverse price elasticity of housing supply
κ	Semi-elasticity of commuting with respect to travel time
ρ	Spatial decay for instrumental variable construction
λ^x	Treatment effect for outcome x
Variables	**Interpretation**
A	Workplace productivity
$B = T\tilde{B}^{\epsilon}$	Gross residential amenity
$\tilde{B}$	Simple residential amenity
T	Mean residential utility
C	Inverse housing efficiency
$\tilde{C}$	Housing productivity
D	Mean utility commute (net of time)
E	Workplace amenity (net of wage)
$\mathcal{C}$	Consumption
H	Housing quantity
W	Wage
Q	Housing price
$\delta = e^{\kappa\tau}$	Commuting friction
τ	Travel time
π	Commuting share
$\bar{N}$	Aggregate population
N^Y	Employment at place of work
L^Y	Land used for production
M	Housing materials
L^H	Land used for housing
P^M	Price of housing materials
$P^L = (H/L^H)^{\psi}$	Price of land

Figure A2: Timeline of transportation in Los Angeles

1925	Comprehensive Rapid Transit Plan for the County of Los Angeles, Kelker, De Leuw and Co. develop at the request of local governments
1951	Los Angeles Metropolitan Transit Authority (LAMTA) formed
1961	Pacific Electric (Red Cars) end of service
1963	Los Angeles Railway (Yellow Cars) end of service
1964	Southern California Rapid Transit District (SCRTD) formed from LAMTA
3/24/1985	Ross Dress for Less methane explosion in Wilshire-Fairfax
1985	Construction begins on LA Metro Rail
11/20/1985	Department of Transportation and Related Agencies Appropriation Act (1986) includes language prohibiting funding of tunnels for transit along Wilshire corridor due to concerns about methane (HR 3244)
7/14/1990	Blue line opens
2/15/1991	Metro Center station opens
1993	Los Angeles County Metropolitan Transportation Authority forms from SCRTD
1/30/1993	Red line opens, connects system to Union Station
10/14/1993	Century Freeway (I-105) opens
8/12/1995	Green line opens in median of Century Freeway
7/13/1996	Red line expands to Wilshire/Vermont
6/12/1999	Red line expands to Hollywood/Vine
6/24/2000	Red line expands to North Hollywood
7/26/2003	Gold line opens
2006	Purple line renamed from Red line branch
9/20/2006	HR 3244 amended to remove prohibitions on funding of tunnels for transit along Wilshire corridor
11/15/2009	Gold line expands in East LA
4–6/2012	Expo line opens
3/5/2016	Gold line expands to Azusa
5/20/2016	Expo line expands to Santa Monica

Figure A3: Data Map

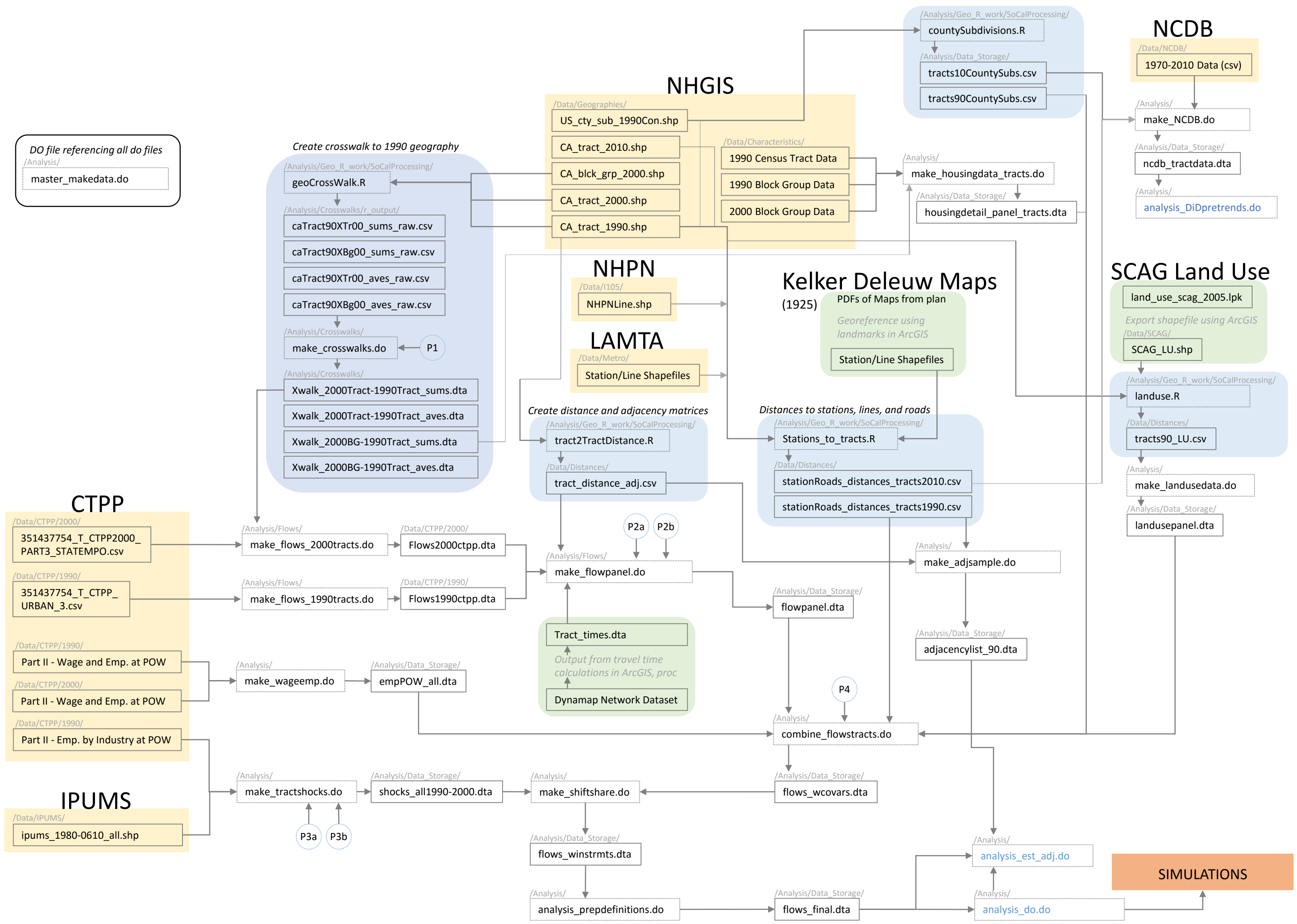


Figure A4: Ridership, 1990-2000

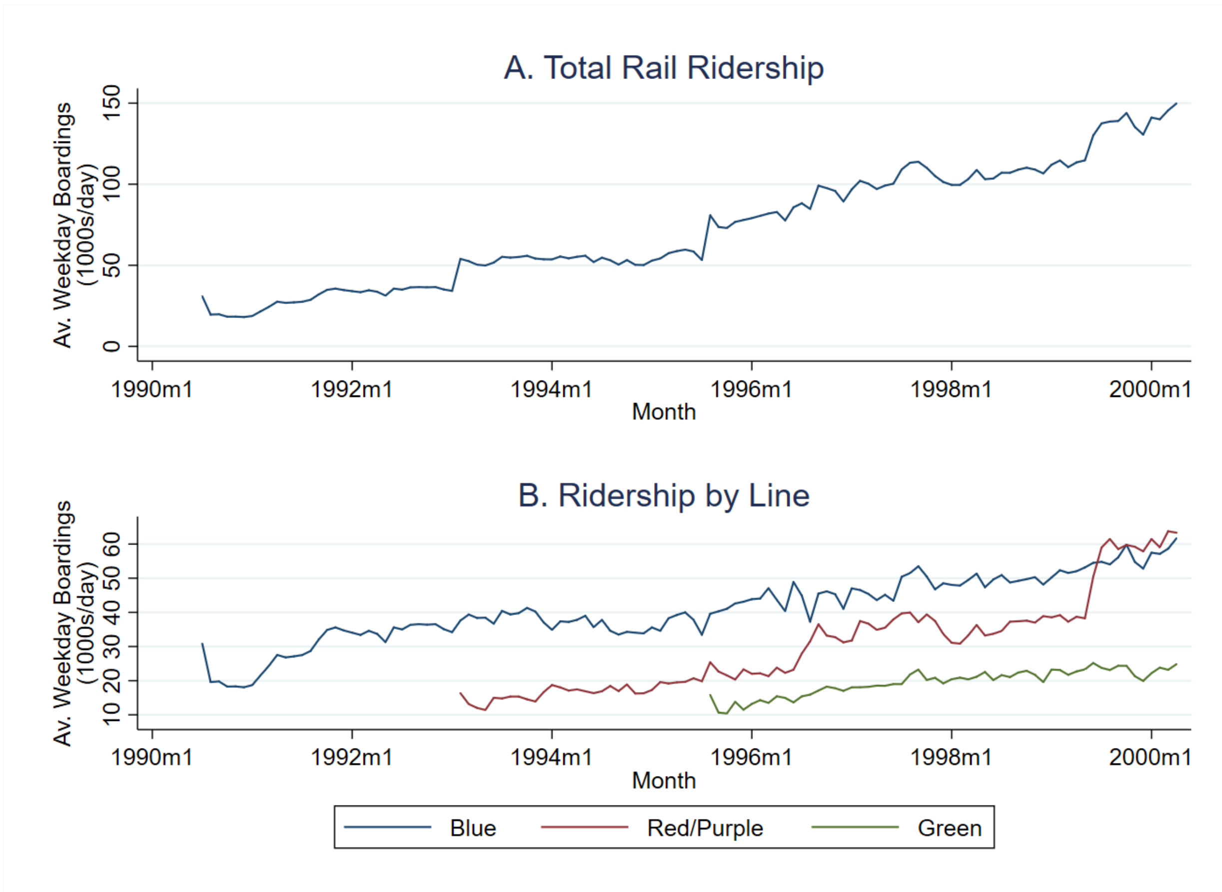


Figure A5: Ridership, 1990-2014

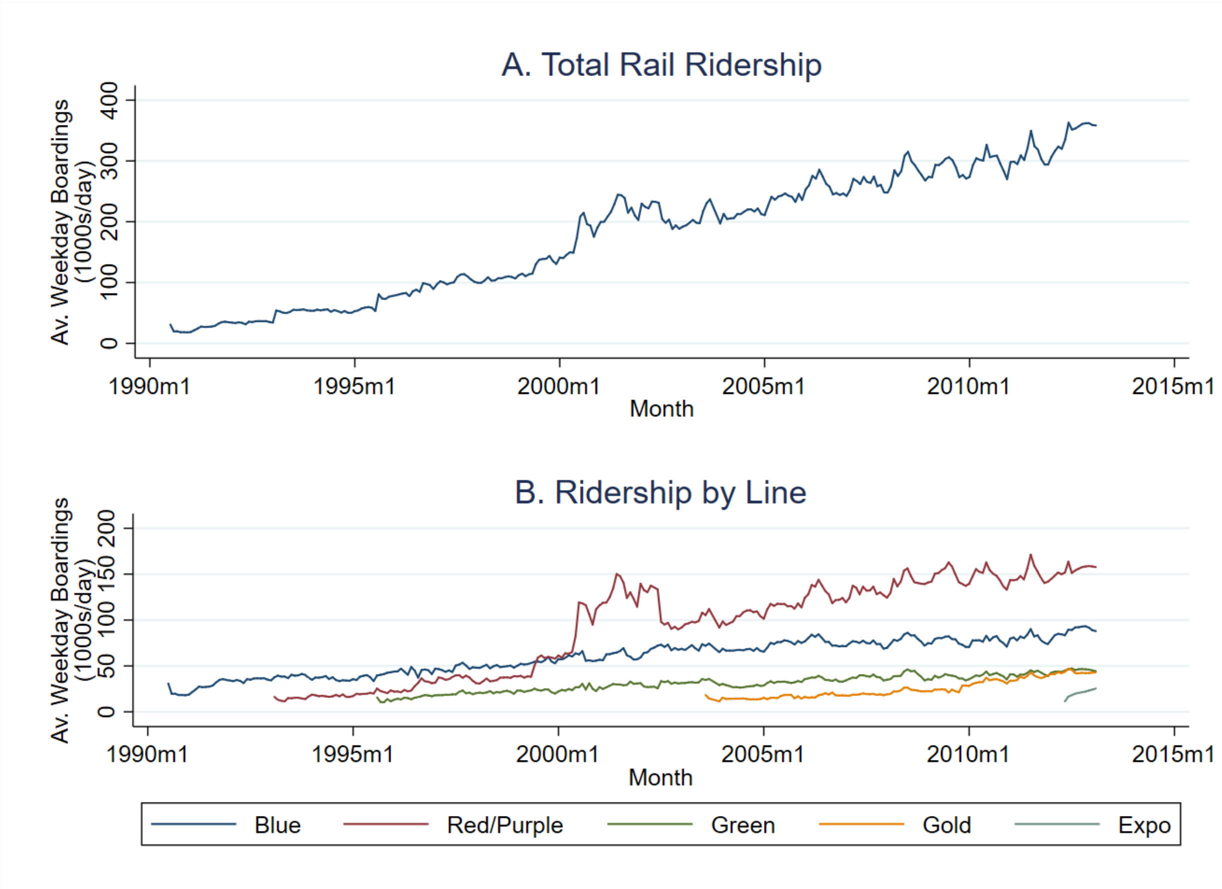
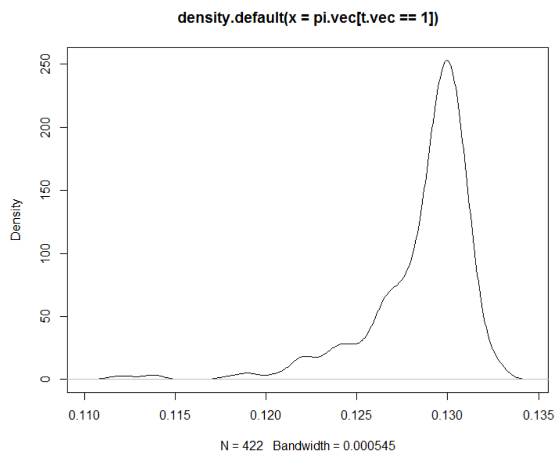
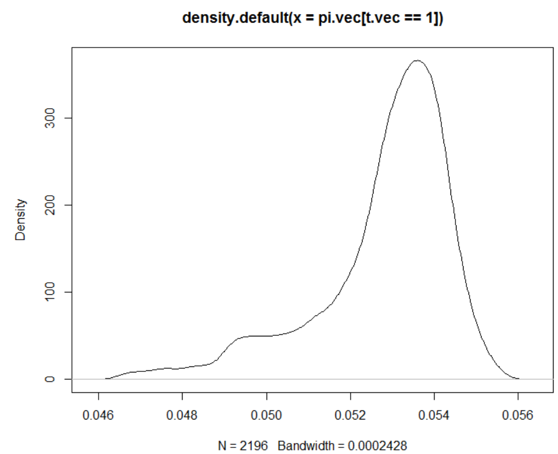


Figure A6: Distribution of counterfactual commuting effects (corresponding to Columns 1 and 7 of Figure 15)

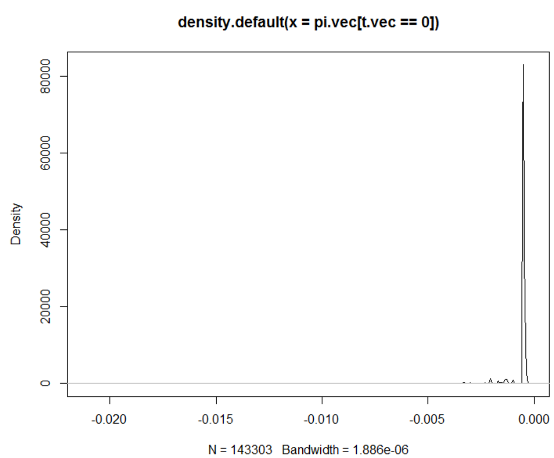
(a) Treated tract pairs, narrow



(b) Treated tract pairs, broad



(c) Untreated tract pairs, narrow



(d) Untreated tract pairs, broad

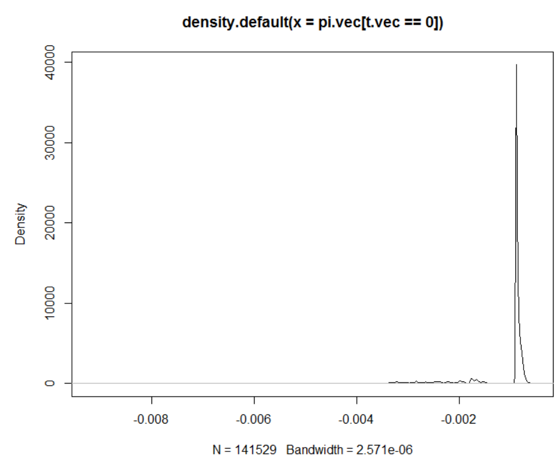


Table A1: IV estimates of labor supply elasticity (ϵ), untrimmed

	Mean utilty of place of work			
	$\hat{\omega}_{it}$	$\hat{\omega}_{it}$	$\hat{\omega}_{it}$	$\hat{\omega}_{it}$
	(1)	(2)	(3)	(4)
$\ln(\text{Wage})$	0.596^{*}	0.984^{+}	0.942^{**}	1.500^{**}
	(0.238)	(0.515)	(0.304)	(0.573)
F-stat	19.82	18.74	13.45	11.48
F-stat (KP)	17.56	16.36	11.40	9.88
Wage instrument	X	X	X	X
Empl. instrument			X	X
ω estimated	Yr-by-yr	Panel	Yr-by-yr	Panel
N	4,916	4,802	4,916	4,802

Fixed effects IV estimates of regression of ω_{it} on w_{it}. All estimates include a tract fixed effect. KP refers to the Kleinbergen-Papp F statistic. Sample is not trimmed to exclude highest and lowest centiles of changes in wage. Place of work-by-year fixed effects ω_{it} estimated separately (year by year) in columns (1) and (3), and together in columns (2) and (4). Standard errors clustered by tract in parentheses: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table A2: Covariate balance prior to treatment, broad

	% Black (1)	% Hisp. (2)	% Foreign Born (3)	ln emp (4)	% Transit (5)	% HS Grads (6)	% Coll. Grads (7)	Pov. Rate (8)	ln HHI (9)	ln hous. units (10)	ln hous. value (11)	ln ave. rent (12)	% Moved <5 yrs (13)
							A.1925 Plan Sample						
1980×1[Transit]	-0.009	0.021	0.036**	0.051*	0.007	-0.019**	-0.000	0.014*	-0.039	-0.012	-0.027	0.497*	-0.042**
	(0.014)	(0.013)	(0.010)	(0.025)	(0.005)	(0.007)	(0.004)	(0.006)	(0.028)	(0.018)	(0.027)	(0.242)	(0.009)
1990×1[Transit]	-0.011	0.054**	0.053**	0.035	0.017*	-0.038**	-0.018**	0.021**	-0.041^{+}	-0.026	-0.060^{+}	0.507*	-0.023**
	(0.019)	(0.018)	(0.012)	(0.029)	(0.007)	(0.011)	(0.007)	(0.007)	(0.021)	(0.024)	(0.032)	(0.237)	(0.008)
Joint test	[0.784]	[0.002]	[0.000]	[0.124]	[0.043]	[0.003]	[0.008]	[0.007]	[0.103]	[0.567]	[0.162]	[0.049]	[0.000]
N	2856	2856	2856	2856	2856	2856	2856	2856	2851	2856	2768	2231	2856
							B. PER Sample						
1980×1[Transit]	-0.007	0.029*	0.040**	0.054^{+}	0.011*	-0.020**	-0.004	0.014*	-0.030	-0.011	-0.054*	0.258	-0.040**
	(0.012)	(0.012)	(0.009)	(0.028)	(0.005)	(0.007)	(0.004)	(0.006)	(0.026)	(0.018)	(0.024)	(0.259)	(0.008)
1990×1[Transit]	-0.021	0.079**	0.071**	0.054^{+}	0.022**	-0.046**	-0.025**	0.023**	-0.034	-0.027	-0.093**	0.280	-0.029**
	(0.018)	(0.017)	(0.012)	(0.030)	(0.006)	(0.011)	(0.007)	(0.007)	(0.022)	(0.024)	(0.029)	(0.254)	(0.008)
Joint test	[0.299]	[0.000]	[0.000]	[0.116]	[0.002]	[0.000]	[0.001]	[0.003]	[0.278]	[0.507]	[0.004]	[0.140]	[0.000]
N	3654	3654	3654	3654	3654	3654	3654	3654	3647	3653	3562	2792	3654

Each column of each panel presents the results of a different regression, for twenty-six total. Regressors include the broad definition of treatment interacted with indicators for years prior to treatment. All regressions include time, tract, and subcounty-by-year fixed effects. P-values for joint tests given in brackets. Standard errors clustered by tract in parentheses: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table A3: Effect of transit on commuting flows (λ^D), historical comparison, flows≥2.5

		1925 Plan Sample				PER sample			
	n_{ijt}	n_{ijt}	n_{ijt}	n_{ijt}	n_{ijt}	n_{ijt}	n_{ijt}	n_{ijt}	n_{ijt}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A. Tract centroid within 500 meters of station									
1[Transit]	0.100*	0.135*	0.134*	0.124*	0.133*	0.121*	0.121*	0.110*	0.118*
	(0.049)	(0.055)	(0.055)	(0.055)	(0.055)	(0.053)	(0.053)	(0.053)	(0.053)
B. Any part of tract within 500 meters of station									
1[Transit]	0.038	0.059$^+$	0.059$^+$	0.043	0.053	0.070*	0.070*	0.055$^+$	0.060$^+$
	(0.025)	(0.032)	(0.032)	(0.032)	(0.033)	(0.029)	(0.029)	(0.030)	(0.031)
Tract Pair FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
POW-X-Yr FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
RES-X-Yr FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Travel Time	-	-	Y	Y	Y	-	Y	Y	Y
Highway Control	-	-	-	Y	Y	-	-	Y	Y
Subcty-X-Subcty-X-Yr FE	-	-	-	-	Y	-	-	-	Y
N	286,752	49,322	49,322	49,322	49,318	72,542	72,542	72,542	72,520

High-dimensional fixed effects estimates of λ^D. All estimates include tract of work-by-year, tract of residence-by-year, and tract pair fixed effects. Travel time is measured in minutes. Standard errors clustered by tract pair in parentheses: $^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table A4: Transit and non-commuting fundamentals, broad definition of transit

	(1)	1925 Plan Sample (2)	(3)	(4)	(5)	PER Sample (6)	(7)	(8)	(9)
				A. Effect on productivity λ^A with $\alpha - 1 = -0.291$					
1[Transit]	-0.086**	-0.026	0.009	-0.003	0.010	-0.043+	-0.003	-0.013	0.009
	(0.023)	(0.024)	(0.027)	(0.028)	(0.028)	(0.023)	(0.026)	(0.028)	(0.027)
N	4,884	1,434	1,434	1,434	1,434	1,886	1,884	1,884	1,882
				B. Effect on residential amenity λ^B with $\epsilon(1-\zeta) = 0.597$					
1[Transit]	0.004	0.015	-0.009	-0.013	-0.010	0.016	0.005	0.001	0.003
	(0.019)	(0.019)	(0.022)	(0.024)	(0.026)	(0.019)	(0.021)	(0.023)	(0.023)
N	4,526	1,324	1,324	1,324	1,324	1,716	1,714	1,714	1,714
				C. Effect on inv. housing efficiency λ^C with $\psi = 1.181$					
1[Transit]	0.086**	0.042+	-0.019	-0.017	0.008	0.060*	-0.010	-0.008	0.008
	(0.023)	(0.025)	(0.027)	(0.032)	(0.032)	(0.024)	(0.026)	(0.030)	(0.029)
N	4,494	1,306	1,306	1,306	1,306	1,686	1,684	1,684	1,684
				D. Effect on workplace amenity λ^E with $\epsilon = 0.544$					
1[Transit]	0.009	-0.025	-0.046*	-0.055**	-0.062**	-0.015	-0.038*	-0.046*	-0.054**
	(0.014)	(0.015)	(0.018)	(0.018)	(0.019)	(0.015)	(0.018)	(0.018)	(0.018)
N	4,822	1,436	1,436	1,436	1,436	1,884	1,884	1,884	1,882
Tract FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Subcty-by-Yr FE	-	-	Y	Y	Y	-	Y	Y	Y
Highway Control	-	-	-	Y	Y	-	-	Y	Y
Other Controls	-	-	-	-	Y	-	-	-	Y

Results from thirty-six regressions of transit proximity on estimated local fundamentals. Other controls are 1990 levels of log household income, percentage black, percentage with high school degrees, and percentage of local employment in manufacturing. Standard errors clustered by tract in parentheses: $^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table A5: Transit and housing supply (λ^C), $\psi = 1.465$, trimmed

		Hous. Supply Shifter, 1925 Plan Sample				Hous. Supply Shifter, PER Sample			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
			A. Tract centroid within 500 meters of station						
1[Transit]	0.096*	0.057	0.005	0.006	0.016	0.068^{+}	0.002	0.002	0.011
	(0.038)	(0.039)	(0.041)	(0.044)	(0.042)	(0.039)	(0.040)	(0.043)	(0.042)
N	4,542	1,294	1,294	1,294	1,294	1,690	1,688	1,688	1,688
			B. Any part of tract within 500 meters of station						
1[Transit]	0.087**	0.052*	-0.013	-0.010	0.005	0.063*	-0.011	-0.008	0.004
	(0.025)	(0.026)	(0.027)	(0.032)	(0.031)	(0.025)	(0.027)	(0.030)	(0.029)
N	4,542	1,326	1,326	1,326	1,326	1,716	1,714	1,714	1,714
Tract FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Subcty-by-Yr FE	-	-	Y	Y	Y	-	Y	Y	Y
Highway Control	-	-	-	Y	Y	-	-	Y	Y
Other Controls	-	-	-	-	Y	-	-	-	Y

Results from eighteen regressions of transit proximity on regressand $q_{it} - \hat{\psi} h_{it}$ with $\hat{\psi} = 1.89$. Other controls are 1990 levels of log household income, percentage black, percentage with high school degrees, and percentage of local employment in manufacturing. Standard errors clustered by tract in parentheses: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table A6: Transit and housing demand (amenities, λ^B), $\epsilon(1-\zeta) = 0.597$, trimmed

		Amenity Level, 1925 Plan Sample				Amenity Level, PER Sample			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
			A. Tract centroid within 500 meters of station						
1[Transit]	0.046	0.052	0.022	0.019	0.020	0.049	0.028	0.024	0.024
	(0.031)	(0.032)	(0.034)	(0.036)	(0.034)	(0.032)	(0.034)	(0.035)	(0.034)
N	4,528	1,294	1,294	1,294	1,294	1,692	1,690	1,690	1,690
			B. Any part of tract within 500 meters of station						
1[Transit]	0.046*	0.058**	0.027	0.031	0.037	0.052*	0.033	0.033	0.035
	(0.021)	(0.022)	(0.024)	(0.027)	(0.026)	(0.021)	(0.023)	(0.025)	(0.025)
N	4,528	1,326	1,326	1,326	1,326	1,718	1,716	1,716	1,716
Tract FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Subcty-by-Yr FE	-	-	Y	Y	Y	-	Y	Y	Y
Highway Control	-	-	-	Y	Y	-	-	Y	Y
Other Controls	-	-	-	-	Y	-	-	-	Y

Results from eighteen regressions of transit proximity on regressand $\hat{\theta}_{it} + \epsilon(1-\zeta)q_{it}$ with $\epsilon(1-\zeta) = 0.46$. Other controls are 1990 levels of log household income, percentage black, percentage with high school degrees, and percentage of local employment in manufacturing. Standard errors clustered by tract in parentheses: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table A7: Transit and housing demand (amenities, λ^B), $\epsilon(1-\zeta) = 0.544 * (1 - 0.65) = 0.19$

	(1)	Amenity Level, 1925 Plan Sample (2)	(3)	(4)	(5)	Amenity Level, PER Sample (6)	(7)	(8)	(9)
A. Tract centroid within 500 meters of station, FEs estimate year-by-year									
1[Transit]	-0.037^{+}	-0.030	-0.031	-0.033	-0.028	-0.030	-0.025	-0.024	-0.019
	(0.021)	(0.022)	(0.024)	(0.026)	(0.027)	(0.022)	(0.023)	(0.024)	(0.025)
N	4,526	1,292	1,292	1,292	1,292	1,690	1,688	1,688	1,688
B. Any part of tract within 500 meters of station, FEs estimate year-by-year									
1[Transit]	-0.023	-0.016	-0.019	-0.020	-0.012	-0.017	-0.011	-0.012	-0.004
	(0.015)	(0.016)	(0.018)	(0.020)	(0.022)	(0.015)	(0.017)	(0.019)	(0.020)
N	4,526	1,324	1,324	1,324	1,324	1,716	1,714	1,714	1,714
C. Tract centroid within 500 meters of station, FEs estimated in panel									
1[Transit]	0.011	0.016	0.005	0.007	0.011	0.011	0.007	0.008	0.011
	(0.026)	(0.027)	(0.029)	(0.030)	(0.030)	(0.026)	(0.028)	(0.029)	(0.029)
N	4,528	1,294	1,294	1,294	1,294	1,692	1,690	1,690	1,690
D. Any part of tract within 500 meters of station, FEs estimated in panel									
1[Transit]	0.020	0.027	0.018	0.026	0.035	0.020	0.018	0.021	0.029
	(0.017)	(0.018)	(0.020)	(0.023)	(0.023)	(0.017)	(0.019)	(0.021)	(0.022)
N	4,528	1,326	1,326	1,326	1,326	1,718	1,716	1,716	1,716
Tract FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Subcty-by-Yr FE	-	-	Y	Y	Y	-	Y	Y	Y
Highway Control	-	-	-	Y	Y	-	-	Y	Y
Other Controls	-	-	-	-	Y	-	-	-	Y

Results from thirty-six regressions of transit proximity on regressand $\hat{\theta}_{it} + \epsilon(1-\zeta)q_{it}$ with $\epsilon(1-\zeta) = 0.46$. Other controls are 1990 levels of log household income, percentage black, percentage with high school degrees, and percentage of local employment in manufacturing. Standard errors clustered by tract in parentheses: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table A8: Transit and workplace utility (λ^E), $\epsilon = 1.030$, trimmed

		Workplace Utility, 1925 Plan Sample				Workplace Utility, PER Sample			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				A. Tract centroid within 500 meters of station					
1[Transit]	0.072*	0.026	0.012	0.010	0.006	0.043	0.020	0.022	0.014
	(0.034)	(0.035)	(0.037)	(0.037)	(0.037)	(0.035)	(0.037)	(0.036)	(0.036)
N	4,822	1,400	1,400	1,400	1,400	1,844	1,844	1,844	1,842
				B. Any part of tract within 500 meters of station					
1[Transit]	0.030	-0.021	-0.051^{+}	-0.061*	-0.070*	-0.003	-0.040	-0.047^{+}	-0.061*
	(0.022)	(0.023)	(0.027)	(0.027)	(0.028)	(0.023)	(0.026)	(0.027)	(0.027)
N	4,822	1,436	1,436	1,436	1,436	1,884	1,884	1,884	1,882
Tract FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Subcty-by-Yr FE	-	-	Y	Y	Y	-	Y	Y	Y
Highway Control	-	-	-	Y	Y	-	-	Y	Y
Other Controls	-	-	-	-	Y	-	-	-	Y

Results from eighteen regressions of transit proximity on regressand $\hat{\omega}_{it} - \epsilon w_{it}$ with $\epsilon = 1.50$. Other controls are 1990 levels of log household income, percentage black, percentage with high school degrees, and percentage of local employment in manufacturing. Standard errors clustered by tract in parentheses: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table A9: Transit, income change, and land use change, broad

		1925 Plan Sample				PER Sample			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				A. Change in log household income					
1[Transit]	-0.015	0.001	0.004	0.005	0.007	-0.009	-0.005	-0.004	-0.001
	(0.012)	(0.013)	(0.014)	(0.016)	(0.016)	(0.013)	(0.014)	(0.015)	(0.015)
N	4,956	1,430	1,430	1,430	1,430	1,878	1,876	1,876	1,876
				B. Change in log residential land					
1[Transit]	-0.013**	0.001	0.002	0.002	0.001	0.001	0.002	0.002	0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
N	4,950	1,432	1,432	1,432	1,432	1,898	1,894	1,894	1,892
Tract FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Subcty-by-Yr FE	-	-	Y	Y	Y	-	Y	Y	Y
Highway Control	-	-	-	Y	Y	-	-	Y	Y
Other Controls	-	-	-	-	Y	-	-	-	Y

Results from eighteen regressions of transit proximity on log household income. Other controls are 1990 levels of percentage black, percentage with high school degrees, and percentage of local employment in manufacturing. Standard errors clustered by tract in parentheses: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Additional Results – Adjacencies

Table A10: Control group validity, adjacency sample: labor demand shocks and treatment

	Narrow treatment		Broader treatment	
	$z_{jt}^{LD,e}$ (1)	$z_{jt}^{LD,w}$ (2)	$z_{jt}^{LD,e}$ (3)	$z_{jt}^{LD,w}$ (4)
1[Transit]	0.011^{+}	0.000	0.006	0.001
	(0.007)	(0.003)	(0.004)	(0.002)
Tract FE	Y	Y	Y	Y
Group-x-year FE	Y	Y	Y	Y
N	2,194	2,194	2,194	2,194

Each column presents the results of a different regression of the labor demand shock (measured in wage or employment) on treatment status. Regressions include year, tract, and group fixed effects. Standard errors clustered by tract in parentheses: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table A11: Effect of transit on commuting flows (λ^D), adjacency sample, flows≥1

	Narrow treatment			Broader treatment		
	n_{ijt} (1)	n_{ijt} (2)	n_{ijt} (3)	n_{ijt} (4)	n_{ijt} (5)	n_{ijt} (6)
A. No Origin- or Destination-by-Year Fixed Effects						
1[Transit]	0.088^{+}	0.087^{+}	0.080	0.002	0.002	-0.004
	(0.052)	(0.052)	(0.052)	(0.027)	(0.027)	(0.028)
B. With Origin- and Destination-by-Year Fixed Effects						
1[Transit]	0.099	0.099	0.092	0.069	0.069	0.060
	(0.074)	(0.074)	(0.074)	(0.049)	(0.049)	(0.050)
Tract Pair FE	Y	Y	Y	Y	Y	Y
Group-x-year FE	Y	Y	Y	Y	Y	Y
Travel Time	-	Y	Y	-	Y	Y
Highway Control	-	-	Y	-	-	Y
N	217,517	217,517	217,517	217,517	217,517	217,517

High dimensioned fixed effects estimates of λ^D. All estimates include tract pair and group-by-year fixed effects. Travel time is measured in minutes. Standard errors clustered by tract pair in parentheses: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table A12: Transit and housing supply (λ^C), $\psi = 1.181$, adjacency sample, trimmed

	Narrow treatment			Broader treatment		
	n_{ijt} (1)	n_{ijt} (2)	n_{ijt} (3)	n_{ijt} (4)	n_{ijt} (5)	n_{ijt} (6)
1[Transit]	-0.046	-0.026	-0.043	-0.036	-0.023	-0.020
	(0.056)	(0.059)	(0.052)	(0.027)	(0.029)	(0.028)
Tract FE	Y	Y	Y	Y	Y	Y
Group-x-year FE	Y	Y	Y	Y	Y	Y
Highway Control	-	Y	Y	-	Y	Y
Other Controls	-	-	Y	-	-	Y
N	1,994	1,994	1,994	1,994	1,994	1,994

Regressions include year, tract, and group fixed effects. Other controls are 1990 levels of log household income, percentage black, percentage with high school degrees, and percentage of local employment in manufacturing. Standard errors clustered by tract in parentheses: $^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table A13: Transit and housing demand (amenities, λ^B), $\epsilon(1-\zeta) = 0.597$, adjacency sample, trimmed

	Narrow treatment			Broader treatment		
	n_{ijt} (1)	n_{ijt} (2)	n_{ijt} (3)	n_{ijt} (4)	n_{ijt} (5)	n_{ijt} (6)
1[Transit]	-0.023	-0.023	-0.019	-0.003	-0.003	-0.004
	(0.033)	(0.035)	(0.034)	(0.020)	(0.022)	(0.021)
Tract FE	Y	Y	Y	Y	Y	Y
Group-x-year FE	Y	Y	Y	Y	Y	Y
Highway Control	-	Y	Y	-	Y	Y
Other Controls	-	-	Y	-	-	Y
N	1,916	1,916	1,916	1,916	1,916	1,916

Regressions include year, tract, and group fixed effects. Tract of residence-by-year fixed effects used in dependent variable are estimated year-by-year. Other controls are 1990 levels of log household income, percentage black, percentage with high school degrees, and percentage of local employment in manufacturing. Standard errors clustered by tract in parentheses: $^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table A14: Transit and labor demand (productivity, λ^A), $\alpha - 1 = -0.291$, adjacency sample, trimmed

	Narrow treatment			Broader treatment		
	n_{ijt}	n_{ijt}	n_{ijt}	n_{ijt}	n_{ijt}	n_{ijt}
	(1)	(2)	(3)	(4)	(5)	(6)
1[Transit]	-0.110*	-0.061	-0.076	-0.022	0.009	0.006
	(0.047)	(0.051)	(0.048)	(0.030)	(0.033)	(0.030)
Tract FE	Y	Y	Y	Y	Y	Y
Group-x-year FE	Y	Y	Y	Y	Y	Y
Highway Control	-	Y	Y	-	Y	Y
Other Controls	-	-	Y	-	-	Y
N	2,168	2,168	2,168	2,168	2,168	2,168

Regressions include year, tract, and group fixed effects. Other controls are 1990 levels of log household income, percentage black, percentage with high school degrees, and percentage of local employment in manufacturing. Standard errors clustered by tract in parentheses: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table A15: Transit and workplace utility (λ^E), $\epsilon = 0.544$, adjacency sample, trimmed

	Narrow treatment			Broader treatment		
	n_{ijt}	n_{ijt}	n_{ijt}	n_{ijt}	n_{ijt}	n_{ijt}
	(1)	(2)	(3)	(4)	(5)	(6)
1[Transit]	-0.022	-0.056^{+}	-0.056^{+}	-0.066**	-0.090**	-0.089**
	(0.031)	(0.032)	(0.032)	(0.020)	(0.021)	(0.021)
Tract FE	Y	Y	Y	Y	Y	Y
Group-x-year FE	Y	Y	Y	Y	Y	Y
Highway Control	-	Y	Y	-	Y	Y
Other Controls	-	-	Y	-	-	Y
N	2,178	2,178	2,178	2,178	2,178	2,178

Regressions include year, tract, and group fixed effects. Other controls are 1990 levels of log household income, percentage black, percentage with high school degrees, and percentage of local employment in manufacturing. Standard errors clustered by tract in parentheses: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$