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

Christopher Severen†
Federal Reserve Bank of Philadelphia

August 2021

Abstract

I study Los Angeles Metro Rail's effects using panel data on bilateral commuting flows, a quantitative spatial model, and historically motivated quasi-experimental research designs. The model separates transit's commuting effects from local productivity or amenity effects, and spatial shift-share instruments identify inelastic labor and housing supply. Metro Rail connections increase commuting by 16% but do not have large effects on local productivity or amenities. Metro Rail generates $94 million in annual benefits by 2000, or 12–25% of annualized costs. Accounting for reduced congestion and slow transit adoption adds, at most, another $200 million in annual benefits.

Keywords: subway, commuting, gravity, economic geography, local labor supply
JEL Codes: J61, L91, R13, R31, 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, Xavier d'Haultfœuille, Stephan Heblich, Kilian Heilmann, Nicolai Kuminoff, Jeff Lin, Kyle Meng, Yuhei Miyauchi, Paulina Oliva, and Wei You, as well as several anonymous referees.

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

1 Introduction

High commuting costs limit consumer choice and mobility within cities. Governments invest large sums in urban rail transit infrastructure to mitigate the costs of distance and congestion. What are the benefits of these investments, and what margins of urban economic geography do they shift?

I study the effects of Los Angeles Metro Rail on commuting, non-commuting margins, and welfare. I assemble data on census tract-to-census tract commuting flows in 1990 and 2000 and develop a new approach to measure the effects of transit on commuting using these flows. Identification of the commuting effect exploits bilateral and panel aspects of the data. To study the non-commuting effects of infrastructure and quantify welfare impacts, I describe a quantitative spatial general equilibrium model and develop a new strategy to identify its key parameters.

A panel gravity equation, which can be embedded in the model, estimates the commuting effect of transit on bilateral flows. Instead of comparing single locations, identification hinges on selecting *pairs* of locations that satisfy treatment ignorability. This means comparing changes in flows between pairs of locations that both receive treatment to changes in flows between pairs of locations in which just one, or neither, receives treatment. I leverage unanticipated shocks to route construction, historical streetcar routes, and proposed subway lines to limit selection concerns.[1] The commuting effect is substantial: flows increase 11%–16% between connected tract pairs that both contain stations by 2000. Tract pairs both slightly more distant from new stations see increases of 8%–14%; there is no effect farther away. I also find evidence of moderate medium-run congestion improvements by 2000 and further commuting growth of 9%–12% between previously connected tracts by 2015. These results suggest that latent demand may not fully displace reduced congestion in the medium run, in contrast to the Fundamental Law of Congestion (Downs 1962).

The model distinguishes the commuting effect of transit from non-commuting effects. Non-commuting effects impact all residences or workplaces near a station, rather than just commuters using transit-served routes. Conditional on model parameters, data on local housing and labor prices and quantities map to tract-specific and time-varying non-commuting fundamentals (e.g., productivity, amenities). I then estimate the effect of transit on changes in these fundamentals with a differences-in-differences strategy, which controls for confounding time-invariant factors (e.g., proximity to the coast). The commuting effect dominates; impacts from non-commuting channels appear minimal.

The model quantifies welfare and accounts for general equilibrium adjustments. Panel data on average wage and industry mix *at place of work* identify key model parameters.[2] Foremost is the local extensive-margin elasticity of labor supply to a tract, which governs how responsive agents are to changes in prices, amenities, and commuting costs. It is also essential for translating treat-

1. See Redding and Turner (2015) for a review of this challenge and common solutions.

2. Workplace wage is often unobserved. I use panel workplace wage data to test a key identifying assumption in Ahlfeldt et al. (2015) meant to overcome unobserved workplace wage and find that is unlikely to hold precisely.

ment effects to welfare. Estimates indicate a low value, implying agents heterogeneous in their preferred locations and relatively unwilling to move in response to changes in local characteristics. I also interact local labor demand shocks with the spatial configuration of the city to estimate a tract-scaled housing supply elasticity.

By 2000, LA Metro Rail generates a baseline \$94 million (in 2016 dollars) in annual surplus (12%–25% of the annualized cost) through commuting channels.[3] A bootstrap 95% confidence interval of this benefit is 2%–64% of costs. Upper bounds on additional annual benefits are \$132 million from reduced congestion and \$76 million from slow adjustment in commuting behavior. Back-of-the-envelope calculations of air pollution benefits may add \$100 million annually. Though substantial, the commuting benefit alone of LA Metro Rail does not clearly exceed its cost over its first two decades, highlighting the difficulty in shifting commuting behavior in restrictively zoned, car-oriented cities. Had zoning allowed 10% greater residential density near transit, the benefits of transit would have doubled.

This approach bridges the hedonic method of valuing transportation infrastructure (e.g., McMillen and McDonald 2004) with a generalized form of the travel time-based method typical of quantitative urban models (e.g., Ahlfeldt et al. 2015; Monte, Redding, and Rossi-Hansberg 2018; Tsivanidis 2018). The hedonic method gives a real estate-mediated measure of the overall effect of transit due to both commuting and non-commuting margins, but cannot disentangle these margins nor account for general equilibrium effects.[4] In contrast, quantitative urban models often only allow transit to shift commuter market access.

Panel commuting flow data offer a substantial improvement over common practice in the literature, which relies on cross-sectional flow data to parameterize gravity models and infer market access. The literature often interprets changes in travel times as perfectly determining changes in commuting flows. Instead, I directly use commuting flows, which reflect travel times but also capture any other hard-to-measure, route-specific characteristics of commute choice (e.g., pleasantness or reliability). The panel aspect permits the use of pair-specific fixed effects, which control for the time-invariant components of these hard-to-measure characteristics. Moreover, commuting flow data obviate the need to estimate historical travel times from modern routing engines.

Los Angeles is a populous, car-oriented region that built an extensive rail network within a decade, so its experience may be more informative for many cities considering rail-based mass transit than evidence from older, denser cities. It is an active line of inquiry whether new mass transit infrastructure in less dense cities provides appreciable benefits, particularly given the newer

3. There are three important caveats. First, while I calculate the commuting effects over a 25-year window, I examine other channels only from 1990 to 2000 due to data limitations. Second, the welfare analysis assumes that agents have homothetic preferences. Finally, transit may benefit non-commuting travel or have environmental effects on cities as a whole. I discuss these in Section 8.

4. Equilibrium effects (e.g., price spillovers) violate the stable unit treatment value assumption and can invalidate hedonic analysis. Donaldson and Hornbeck (2016) highlight the importance of equilibrium adjustments when evaluating transportation infrastructure.

role of cities as centers of consumption (Baum-Snow, Kahn, and Voith 2005). There is a budding line of research examining the consequences of LA Metro Rail.[5]

I describe the setting and data in Section 2. Section 3 discusses identification and estimation of the commuting effect. Section 4 develops and characterizes the model. Section 5 addresses the second empirical challenge: estimating the model's elasticities. I describe estimating the non-commuting effects of transit in Section 6. Section 7 reports baseline welfare analysis, and Section 8 discusses extensions and discusses broader conclusions.

2 Setting and Data: Transit in Los Angeles

Automobiles have long been a common feature of mobility in Los Angeles. Angelenos adopted automobiles in large numbers during rapid urban growth in the 1920s, leading to early complaints of crowded streets and attempts to relieve traffic delays (Fogelson 1967). Increasing congestion in the 1960s and 1970s led to several failed referendums to expand rail transit. By 1980, the situation reached a political tipping point and voters passed Proposition A—a sales tax increase to fund rail transit. The proposed plan combined heavy rail (subway) and light rail operations in an interconnected urban rail transit system. Construction on the system began in 1985 and the first light rail line (the Blue Line) partially opened in mid-1990 (construction delays meant it did not reach its urban termini until early 1991). The subway line faced routing difficulties and first opened in 1993. It was expanded in 1996 and 1999 and is now run as two lines: the Red and Purple Lines. Another light rail line initially meant to connect to the international airport (the Green Line) opened in 1995 largely in the median of a new freeway, but without reaching the airport.[6]

To study the effects of rail transit in Los Angeles, I develop a panel of tract-level variables in 1990 and 2000 that covers Los Angeles County and four adjacent counties (Orange, Riverside, San Bernardino, and Ventura). This five-county area is economically distinct from nearby conurbations and captures most relevant local interactions. I briefly discuss primary data sources and processing below; additional details can be found in the Appendix.

Geo-normalization. The standard unit of observation is a census tract or tract pair using 1990 Census geography. I normalize to 1990 geography to minimize rounding errors. I areally weight more recent geographies when crosswalking to 1990 tracts.

Commuting flow data. Tract-to-tract commuting flow data are primarily from the 1990 and 2000 Census Transportation Planning Packages (CTPP). I normalize origin-destination pairs to 1990 geography and apply consistent rounding and suppression rules to create a consistent panel. I also use commuting flow data covering 2002 and 2015 from the Longitudinal Employer-Household Dynamics Origin Destination Employment Statistics (LODES), normalized to 2010 geographies.

5. See, for example, Redfearn (2009), Schuetz (2015), and Schuetz, Giuliano, and Shin (2018).

6. LA Metro Rail's line names recently changed. I use the older designations for continuity. The system continues to grow, with 6 lines, 93 stations, and 106 miles of rail as of 2016.

Due to methodological differences, CTPP and LODES data are not combined.

Place of residence and place of work data. Data on residential census tracts and block groups are from the National Historic Geographic Information System (NHGIS) and Geolytics' Neighborhood Change Database (NCDB). The CTPP contains *tract of work* wage data unavailable elsewhere and employment by industry in 18 aggregate Standard Industrial Classification (SIC) codes. I trim the data to exclude implausible changes between 1990 and 2000 levels.

Transit data and treatment; other data sources. I combine geodata on Metro Rail transit stations and lines from the Los Angeles County Metropolitan Transportation Authority (LACMTA) with published information on the timing of openings. To construct labor demand shocks, I use IPUMS microdata on all workers outside of California from the 1990 and 2000 Censuses. Panel land use data are from the Southern California Association of Governments (SCAG).

3 Commuting Effects of LA Metro Rail

The number of commuters from residential tract i to workplace tract j at time t, denoted N_{ijt}, depends on residential tract characteristics, θ_{it}, workplace tract characteristics, ω_{jt}, and trip characteristics τ_{ijt}. Let T denote some function of proximity to transit. Commuting is:

$$N_{ijt} = N_{ijt}\Big(\theta_{it}(T_{it}), \omega_{jt}(T_{jt}), \tau_{ijt}(T_{it}, T_{jt})\Big). \tag{1}$$

The commuting effect of transit isolates transit's effect connecting i and j: $\frac{\partial N}{\partial \tau}\frac{\partial \tau}{\partial T}$. Transit can generally shift residential or workplace characteristics as well. Comparing flows by T alone does not differentiate commuting effects from other margins.

Panel bilateral commuting flow data can separate these margins. Bilateral data allow flexibly controlling for residential and workplace characteristics and shocks, while temporal variation allows controlling for time-invariant pair-specific characteristics. Let T_{ijt} denote proximity to transit at both i and j. I estimate:

$$\ln(N_{ijt}) = \omega_{jt} + \theta_{it} + T'_{ijt}\lambda^D + x'_{ijt}\beta + \bar{d}_{ij} + \iota_{s_i s_j t} + d_{ijt}, \tag{2}$$

where $\bar{d}_{ij}$ are pair fixed effects and d_{ijt} captures unobserved variation in commuting between locations over time.[7] Residential and workplace tract-by-year fixed effects, θ_{it} and ω_{jt}, capture non-commuting effects of transit.

Equation (2) is a panel gravity equation, wherein the pair fixed effects subsume distance.[8]

7. While the structural model interprets d_{ijt} as an unobserved utility shifter, Dingel and Tintelnot (2020) highlight how granularity can produce measurement error.

8. $N_{ijt} = 0$ for some observations, so $\ln(N_{ijt})$ is undefined. I estimate high-dimension fixed-effects Poisson PML models to show robustness (Larch et al. 2019). Results are broadly consistent, as only tract pairs that are zero in one period and non-zero in the other period can lead to differences (always-zero pairs separate out).

These fixed effects also capture unobserved determinants of commuting flows—such as bus service, ease of parking, pleasantness, or even workplace-residential matching—to the extent that such features are time invariant. Some specifications use subcounty-by-subcounty-by-year fixed effects ($\iota_{s_i s_j t}$) to capture regional shifts in commuting patterns and allow flexible trends in regional integration. In x_{ijt}, I include time-varying measures of proximity to the Century Freeway, which opened in the mid-1990s.

Treatment, T_{ijt}, reflects three mutually exclusive, binary definitions of proximity of *both a residential and a workplace tract* to LA Metro Rail stations open before the end of 1999[9]:

i) *O & D contain station*: Both tracts either contain a transit station or have their centroid within 500 meters of a transit station.

ii) *O & D <250m from station*: Both tracts have some part within 250 meters of a transit station, but i) is not true.

iii) *O & D <500m from station*: Both tracts have some part within 500 meters of a transit station, but neither i) nor ii) is true.

3.1 Identification

Equation (2) estimates a dyadic difference-in-differences (DD) design supplemented with origin- and destination-by-year fixed effects. Interpreting λ^D as the average causal effect of LA Metro Rail on commuting flows requires assuming parallel counterfactual trends: In the absence of treatment, commuting between treated and control tract pairs would have evolved similarly on average, *conditional on separable changes to residential and workplace locations*. This substantially relaxes standard DD assumptions. Time-varying origin and destination fixed effects largely control for confounding factors (e.g., school quality, zoning).

While nonrandom siting of transit threatens the parallel trends assumption, origin and destination fixed effects remove concerns about where stations are placed. I limit pair-specific selection concerns with two approaches. First, I use historical data giving the locations of a proposed subway network and former streetcar lines, which embeds shocks to route placement and thus selects pairs of tracts that share common historical trends and could have both plausibly been treated by 2000. The second approach considers only tract pairs that are both near existing or not-then-built transit stations, comparing pairs along the same subway line with pairs that are not.

History & Shocks designs. Kelker, De Leuw & Co. (1925) designed a rail transit network to accommodate Los Angeles' booming population in the 1920s (the plan was shelved partly because of skepticism of rail companies and a preference for tunneled lines).[10] The plans also show former

9. For a hypothetical square tract of median area (1.38km^2) with a uniformly distrbuted population, average distance to a station for each proximity is i) 444m–888m, ii) 689m–888m, and iii) 901m–1219m; see Appendix A.4.

10. The relevant maps from Kelker, De Leuw & Co. (1925) are shown in Appendix Figure H1.

Pacific Electric Railroad (PER) lines, an at-grade passenger rail system. I define two samples as the union of tract pairs near LA Metro Rail by 2000 and pairs that lie near: (i) the Kelker, De Leuw & Co. (1925) subway proposal, "1925 Subway Plan"; or (ii) PER lines, "PER Sample". The 1925 Subway Plan itself has two variants: an "Immediate" plan meant for immediate construction, and a plan meant to accommodate buildout ("All"). These research designs are mapped in Figure 1.

The validity of these groups as controls is supported by several lines of reasoning and evidence. First, many control pairs contain one "end" (either the origin or destination) that is treated, though the other end is not. Such control pairs compare changes in the number of commuters residing in i who work in j (which receives a transit linkage) to the number who work in j' (which does not). Similarly, workers in j who reside in i are compared with those who reside in i'. These comparisons control for many potential unobserved motives for changing commuting behavior.

Second, control tracts selected by this approach are near historical transit corridors, as are treated locations. Brooks and Lutz (2019) show that locations near former streetcar routes in Los Angeles are generally on similar land use trajectories. This promotes viability of transit siting, and land use itself can have an impact on travel behavior (Duranton and Turner 2018).

Third, there was significant variation in the timing of route construction due to reasons likely orthogonal to transit demand. Notably, a geologic shock limited westward expansion of the Red/Purple Line (its original routing was along Wilshire Boulevard, one of the densest corridors in Los Angeles). Methane seeping into a nearby clothing store exploded in March 1985. Even though the explosion was not related to subway construction, legislation soon restricted tunneling along Wilshire. This corridor is served by transit in both the 1925 Subway and PER samples, and in almost every plan since. Similarly, the Green Line's route was built partially within an under-construction highway to minimize construction costs. Its westward end was meant to serve Los Angeles International Airport (LAX), but the Federal Aviation Administration's concerns about electromagnetic interference detoured this alignment to the south. Construction is underway on connections between the system and the Wilshire corridor and LAX, so these shocks can be seen as generating plausibly exogenous variation to the timing of treatment.

Routes also reflect other non-transportation concerns, such as satisfying political pressures, ensuring political support for transit revenue allocations, and spurring political favor from competing oversight agencies.[11] To illustrate, politicians demanded that heavy rail serve the San Fernando Valley, despite the cost and difficulty of doing so. It was also deemed necessary to connect Long Beach to ensure access to its portion of state gas tax revenue. At one point, a particularly serpentine route was dubbed the "wounded knee" because it touched so many local political jurisdictions. In sum, Elkind (2014, p. 50)'s statement that "politics, outside circumstances, and the geography of power … played an outsized role in influencing where the new rail lines would go"

11. The political setting is presented in Elkind (2014), who notes that "plotting a subway through the politically decentralized landscape of Los Angeles meant ceding control to numerous fiefdoms of federal, state, and local politicians" (p. 70).

indicates the presence of factors unrelated to travel demand in the planning process.

Finally, I provide econometric evidence. While the data do not permit testing pre-trends in *tract-pair* flows, I examine pre-trends in *tract*-level housing and labor market characteristics using NCDB data from 1970 to 1990.[12] Among economic variables later captured in the model (columns 1–4 of Appendix Table H2), there are no significant differences in pre-trends in employed population or household income in any sample; there is marginal evidence of pre-trends in housing values and number of households. However, the Immediate 1925 Subway Plan specification shows no evidence of differential pre-trends in model-relevant characteristics, and so is preferred.

I also investigate differential trends in neighborhood and travel characteristics (columns 5–10 of Appendix Table H2). In some samples, residents of treated tracts were becoming less college educated and more impoverished and were moving less often. Evidence on differences in travel pre-trends is mixed: in two samples, the share of household with no cars was decreasing prior to treatment. In none of the samples was the commuting share by auto differentially changing, but in all samples the relative transit commuting share transit was increasing slightly. The Immediate 1925 Subway Plan sample shows the least evidence of pre-trends across all variables.

Recall that the fixed effects in Equation (2) render it robust to any tract-level pre-trends, whether observable or not. While tract-level pre-trends do not impede identification of the commuting effect, they may confound the non-commuting analysis in Section 6.

Same Line designs. These designs compare tract pairs on the same line to tract pairs that lie along different lines. This design expects that tract pairs along the same line are "more treated" than those that are not. Parallel trends are violated if planners targeted directly connecting (by the same line) locations that would have seen larger increases in commuting anyway relative to other treated—but indirectly connected—locations. I consider two variants. The first includes locations treated by 2000 and locations not yet treated in 2000 but that are treated by 2015. In this 'Ever Treated' variant, treated pairs both lie near stations along the same line and control pairs both lie near stations open by 2015 but that are not along on the same line. In the second, more stringent variant (Treated by 2000), I restrict control pairs to those that lie near a station open by 2000.

There are two caveats regarding these designs. First, they identify the commuting effect of transit only if it is infinitely costly to switch from one line to another (i.e., a very high transfer penalty). Under a finite (but non-zero) transfer penalty, these designs reflect the marginal effect of being directly connected relative to being indirectly connected. Second, these designs rely on fewer observations and are less precise.

12. NCDB is, by default, normalized to 2010 geographies. I use the same treatment rules, but this results in higher observation counts due to denser tracts in 2010 than 1990.

3.2 Commuting flow estimates

Panel A of Table 1 reports estimates of λ^D on commuting flows between 1990 and 2000. Columns 1–3 show results from the full sample, successively adding in measures of treatment proximity, subcounty pair-by-year fixed effects, and controls. Columns 4–6 reflect the History & Shocks designs. Columns 7–8 report results from the Same Line designs. Clustered standard errors are robust to correlation within tract pairs, residential tracts, and workplace tracts.

LA Metro Rail increased commuting 10%–22% between tracts nearest transit stations by 2000. Estimates are significant across all specifications. The preferred specification (column 6) indicates an increase of 14.9 log points (16%). Slightly fewer transit-proximate tract pairs see an effect of 8%–14%, with a preferred estimate of 12.8 log points (14%). Tract pairs at a farther distance from stations see no significant effect.[13]

Several features of the results are reassuring. First, estimates are ordered by proximity. Together these form a specification test: effects should concentrate near stations with little no effect farther away. Second, as the control group becomes more targeted (and sample size decreases), point estimates become larger (ascending from left to right). This ordering implies that control tract pairs experience progressively less commuting growth than in the full sample. As these designs increasingly select connections between older, more mature parts of the city, this is reasonable. Finally, Same Line estimates are a bit larger than the History & Shocks estimates, but also less precise. This suggests that proximity to directly connected stations is more important than simply being near a transit station.[14]

Panel B of Table 1 extends the analysis to determine if there are additional effects from 2002–2015 using LODES data. Because LA Metro Rail expanded during this period, I estimate a variant of Equation (2) with two different effects: *Existing connections* for the additional increase between tracts connected by stations built earlier (between 1990 and 2002), and *New connections* for stations built after 2002. *Existing connections* experience additional commuting growth of 8%–12% by 2015 if both tracts were previously connected and 5%–10% if the tracts were a bit farther from stations. *New connections* increase commuting by 10%–14% between tracts that both contain new stations by 2015 and up to 9% for those slightly more distant from stations. While substantial, these effects are likely smaller than for tracts connected between 1990 and 2000 because stations built between 2002 and 2015 are generally more suburban.

13. See footnote 9 for interpretation of these distance bins and robustness in Appendix Figure H6. Similar estimates result from Poisson PML specifications; see Appendix Table H4.

14. Appendix Table H5 directly tests this. Though noisy, effects are always for locations along the same line. The data allow further exploration in this vein. Interacted proximity bins for origin and destination tracts indicate a greater effect of proximity at the destination than the origin, suggesting commuters respond more to closer workplace-to-station proximity than to closer residence-to-station proximity (Appendix Table H6). Treatment does not generally alter the extensive margin of connection, but its effect may increase slightly over time (Appendix Table H3).

3.3 Commuting time (congestion) estimates

A common motivation for rail transit is to relieve automobile congestion. Anderson (2014) finds that a 2003 labor strike that disrupted LA Metro Rail service increased nearby automobile congestion by 47%. However, that strike lasted 35 days. It is unclear how to map that short-run response to the long run. Duranton and Turner (2011) find no aggregate evidence that transit decreases automobile travel. This notion, called the "Fundamental Law of Congestion" (Downs 1962), suggests that congestion improvements (and downstream benefits like air pollution) may be transitory.

I combine reported CTPP travel times with inferred travel routes to test for decreased congestion due to transit.[15] I map the fastest driving route between all location pairs in my sample and calculate the share of each route ($\ell_{ij\in k}/\ell_{ij}$) that falls within five mutually exclusive distance buffers k (omitting the share of the route more than 4km distant). I then estimate:

$$\ln(\tau_{ijt}) = \sum_k \lambda_{\tau,k} \frac{\ell_{ij\in k}}{\ell_{ij}} 1_{[t=2000]} + \omega_{jt} + \theta_{it} + \varsigma_{ij} + \varepsilon_{ijt}, \tag{3}$$

where τ_{ijt} is the average reported travel time from i to j in year t.

Table 2 shows results for two measures of travel time: log travel time across all modes and log travel time for private cars. Results for all-mode travel times are negative but mostly insignificant; they may reflect mode switching to transit. Car-only results, however, show clear evidence of a reduction in travel time.[16] Routes that are entirely within 250m meters of lines by 2000 see a 15.0 log point (14%) reduction in travel time. For routes lying entirely 250m–500m from lines by 2000, travel times decrease by 18.9 log points (17%). Though not statistically different, the larger effect slightly farther from the rail lines may indicate slower travel due to at-grade crossings. Portions of routes that lie farther than 500m from a rail line see no change in travel times.

The estimates in columns 3–4 of Table 2 are roughly one-third those in Anderson (2014). This suggests substantial but incomplete attenuation of congestion benefits over a period of 5–10 years, indicating that congestion benefits of transit may not be entirely transitory. Any downstream benefits, like improved air quality, may similarly last over longer time horizons.

4 A Model of Urban Location Choice

I turn to a quantitative urban model of residential and workplace choice to recover the non-commuting effects of transit and translate the effects of transit to welfare. The model links local, observable equilibrium outcomes to local, unobservable economic fundamentals (e.g., productiv-

15. The setting is comparable to Anderson (2014), though he uses a different research design and focuses primarily on observed highway travel speeds. See Brinkman (2016) and Allen and Arkolakis (2019) for endogenous congestion.

16. Sample sizes are smaller than in Table 1 due to disclosure restrictions, which is also why the sample for private car travel time is yet smaller. I also drop pairs for which the implied trip speed is greater than 80mph.

ity, amenities). The model includes a collection of N locations in a city, operationalized as census tracts, that each contain a labor market and a housing market.[17]

Joint market household decision: Labor supply and housing demand

Atomistic households make consumption decisions and choose a tract of work and a tract of residence. Conditional on residential location i, households face per-unit housing costs Q_i and receive amenity $\tilde{B}_i$. Conditional on place of work j, households inelastically provide one unit of labor in exchange for wage W_j. Given locations 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 Cobb-Douglas utility:

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

where ν_{ijo} is household o's idiosyncratic preference for location pair ij. The commuting cost between i and j is captured by $\delta_{ij} \geq 1$. The share of household expenditures on housing is $1-\zeta$. Indirect utility conditional on location pair ij is:

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

Housing consumption for o conditional on ij is $H_{ijo} = (1-\zeta)W_j/Q_i$.

I assume ν_{ijo} 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 low ϵ, their valuations are heterogeneous. The share of the population that chooses residential location i and place of work j under the Fréchet assumption is:

$$\pi_{ij} = \frac{\tilde{\Lambda}_{ij}\left(\delta_{ij} Q_i^{1-\zeta}\right)^{-\epsilon} (\tilde{B}_i W_j)^{\epsilon}}{\sum_r \sum_s \tilde{\Lambda}_{rs}\left(\delta_{rs} Q_r^{1-\zeta}\right)^{-\epsilon} (\tilde{B}_r W_s)^{\epsilon}}. \tag{4}$$

17. The model is similar to Ahlfeldt et al. (2015), with five differences: (i) origin-destination pairs can differ in mean utility, which permits deriving Equation (2) from the model; (ii) a local housing efficiency parameter captures differences in local regulations and per unit housing costs; (iii) land use between housing and production is exogenously determined; (iv) endogenous externalities are omitted (though I discuss model variants with endogenous agglomeration for welfare calculations in the Appendix); and (v) the model can be written as a system of three equations log-linear in data and fundamentals.

Observable commuting flows are $N_{ij} = \pi_{ij}\bar{N}$, where $\bar{N}$ is total population.

The city can be viewed as either 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 elsewhere, so:

$$\mathbb{E}[U_{ijo}] = \Gamma\left(\frac{\epsilon-1}{\epsilon}\right) \cdot \left[\sum_r \sum_s \tilde{\Lambda}_{rs} \left(\delta_{rs} Q_r^{1-\zeta}\right)^{-\epsilon} (\tilde{B}_r W_s)^{\epsilon}\right]^{1/\epsilon}, \tag{5}$$

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

Production: Labor demand

Measure-zero firms produce a globally tradable commodity in each location j under perfect competition. Firms select labor N^Y and land L^Y inputs to maximize profits under constant returns to scale. Production is multiplicatively separable in local productivity A_j and a technology that is identical across j: $Y = A_j F(N_j^Y, L_j^Y)$. Because of the atomistic size of firms, land use decisions are made in accordance with profit maximization despite the locally fixed quantity of land. Perfect competition in labor markets implies that firms pay workers the marginal product of labor: $W_j = A_j F_N(N_j^Y, L_j^Y)$. I assume Cobb-Douglas production technology: $F(N^Y, L^Y) = (N^Y)^{\alpha}(L^Y)^{1-\alpha}$. Inverse labor demand is given by:

$$W_j = \alpha A_j \left(L_j^Y / N_j^Y\right)^{1-\alpha}. \tag{6}$$

Housing supply

Measure-zero builders construct housing using land L^H and material inputs M. A local, multiplicatively separable housing productivity term $\tilde{C}_i$ captures cost drivers such as geography and regulation. Materials are readily available in all locations at the same cost, but local land supply for housing is predetermined.[18] Convexity in land pricing serves as a congestive force, driving up prices in desirable locations until agents look elsewhere. I specify Cobb-Douglas housing production: $H = (L^H)^{\phi} M^{1-\phi} \tilde{C}_i$. Developers sell housing in location i in a competitive market at unit price Q_i to maximize profit: $Q_i H - P_i^L L^H - P^M M$. The price of construction materials P^M is exogenous and common to all locations.

Because detailed data on housing production are not available, I utilize a zero-profit condition to develop an empirical formula for housing costs. The first-order condition for developer profits

18. This simplifies the model while maintaining fidelity to the setting. Strong zoning and the medium-run time frame of this study may not match the temporal patterns required for land use change.

with respect to construction materials gives:

$$Q_i = \frac{P^M}{(1-\phi)\tilde{C}_i}\left(\frac{M}{L^H}\right)^{\phi}. \tag{7}$$

Substituting this into the developer's profit function and enforcing zero-profit conditions gives construction material demand, $M^* = \frac{1-\phi}{\phi}\frac{L^H P_j^L}{P^M}$, as well as $Q_i = (P_i^L L^H + P^M M)/((L_i^H)^{\phi} M^{1-\phi}\tilde{C}_i)$. Substituting in M^* gives the cost function: $Q_i = C_i (P_i^L)^{\phi}$, where $C_i = (P^M)^{1-\phi}/(1-\phi)^{1-\phi}\phi^{\phi}\tilde{C}_i$ captures the inverse efficiency in housing production.

The price of land, P_i^L, 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. This parameter provides a congestive force to the model. Combining the expression for land price with Equation (7) relates housing supply, price, and land availability:

$$Q_i = C_i \left(H_i/L_i^H\right)^{\psi}, \tag{8}$$

where $\psi = \tilde{\psi}\phi$. 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 demand equal supply locally:

$$N_i^Y = \sum_r \bar{N}\pi_{ri}. \tag{9}$$

Given Cobb-Douglas 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, reflecting heterogeneity in income stemming from variation in wage (i.e., place of work). Housing market clearing requires that the local housing supply equal demand:

$$H_i = (1-\zeta)\sum_s \bar{N}\pi_{is}\frac{W_s}{Q_i}. \tag{10}$$

Given model parameters $\{\alpha, \epsilon, \zeta, \psi\}$, 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 commuting fundamentals and costs $\{\mathbf{D}, \boldsymbol{\delta}\}$, an equilibrium is referenced by price vectors $\{\mathbf{W}, \mathbf{Q}\}$, commuting vector $\boldsymbol{\pi}$, and scalar population measure $\bar{N}$. Existence and uniqueness conditions are (proofs are in Appendix B):

Proposition 1. *Consider the equilibrium defined by Equations (4), (6), (8), (9), (10):*

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

ii) There is at most one equilibrium if

$$\frac{2\epsilon(\epsilon+1)(1-\alpha)(1-\zeta)}{1+\epsilon(1-\alpha)} - 1 \leq \frac{1}{\psi}. \tag{11}$$

Inversion

Though the model may have multiple equilibria, 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. $\tilde{B}_i$ and T_i enter isomorphically; let $B_i = T_i \tilde{B}_i^{\epsilon}$ and $\Lambda_{ij} = B_i E_j D_{ij} \delta_{ij}^{-\epsilon}$.[19] Local fundamentals $\mathbf{A}$, $\mathbf{C}$, and $\boldsymbol{\Lambda}$ can be expressed as unique functions of data and parameters:

Proposition 2. *Given parameters* $\{\alpha, \epsilon, \zeta, \psi\}$ *and observed data* $\{\mathbf{W}, \mathbf{Q}, \boldsymbol{\pi}, \bar{N}\}$*, a unique set of fundamentals* $\{\mathbf{A}, \mathbf{C}, \boldsymbol{\Lambda}\}$ *exists and is consistent with the data being an equilibrium of the model.*

5 Model Identification and Estimation

Local labor and housing market elasticities provide a 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 use of fixed effects to control for unobserved, time-invariant characteristics that confound identification (for example, location near a port, on a pleasant hillside, or in town with stringent land use regulations).[20]

All components of the model are expressed in the commuting flow (4), wage setting (6), and housing price (8) equations. Denote log values in lowercase letters. Adding time subscripts, including tract and tract-pair fixed effects (e.g., $\ln(A_{jt}) = \bar{a}_j + a_{jt}$), and letting n^Y_{jt} be log employment density, h_{it} be log housing density, the g be constants, $\tilde{\alpha} = \alpha - 1$, and $\tilde{\zeta} = -\epsilon(1-\zeta)$, these equations

19. This mapping diverges from Ahlfeldt et al. (2015). Note that the components of $\boldsymbol{\Lambda}$ are not uniquely identified from the data; I use statistical arguments to separate $\mathbf{B}$, $\mathbf{E}$, and $\mathbf{D}\boldsymbol{\delta}^{-\epsilon}$.

20. Persistent, difficult-to-measure amenities play an anchoring role in cities (Lee and Lin 2018). Strong land use regulation likely locks in such anchors in Southern California (Severen and Plantinga 2018).

deliver a tractable system log-linear in data and fundamentals (see Appendix B):

$$\begin{aligned}
\text{Labor demand:} \quad & w_{jt} = g_{0t} + \tilde{\alpha} n_{jt}^{Y} + \bar{a}_j + a_{jt} && (12)\\
\text{Commuting:} \quad & n_{ijt} = g_{1t} + \underbrace{\epsilon w_{jt} + \bar{e}_j + e_{jt}}_{\substack{=\omega_{jt},\\ \text{Labor supply}}} + \underbrace{\tilde{\zeta} q_{it} + \bar{b}_i + b_{it}}_{\substack{=\theta_{it},\\ \text{Housing demand}}} + \tilde{d}_{ijt} && (13)\\
\text{Housing supply:} \quad & q_{it} = g_{2t} + \psi h_{it} + \bar{c}_i + c_{it} && (14)
\end{aligned}$$

Local fundamentals are potentially functions of covariates ($\bar{a}_j + a_{jt} = a(x_{jt})$ and so on), such as transit proximity. Equation (13) interprets Equation (2) structurally, with fixed effects $\omega_{jt} = \epsilon w_{jt} + \bar{e}_j + e_{jt}$ and $\theta_{it} = \tilde{\zeta} q_{it} + \bar{b}_i + b_{it}$, and utility-equivalent commuting costs such that $\tilde{d}_{ijt} = T_{ijt}'\lambda^D + x_{ijt}'\beta + \bar{d}_{ij} + \iota_{s_i s_j t} + d_{ijt}$.

5.1 A general approach to identifying local elasticities

I develop a local shift-share instrument to overcome simultaneity in Equations (12)–(14). I leverage tract-level variation in shocks to labor demand, interacting these with the distance between tracts to create variation in local economic shocks. I focus on identification of ϵ (the elasticity of labor supply) and ψ (the inverse elasticity of housing supply), as these embed information about the local economic environment and cannot be calibrated from microdata.[21]

Identification requires a demand or supply shock that shifts one of Equations (12)-(14) 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. Effective variation comes from changes in wages and employment determined by ex ante local industrial composition. These shocks are relevant if correlated with changes in local productivity (Δa_{jt}) and excludable if uncorrelated with changes in other local fundamentals. Under these assumptions, labor demand shocks trace out the labor supply curve. Housing demand nearby shifts in response. Because this downstream response will be stronger nearer the workplace origination of the shock, I take a linear combination of labor demand shocks with weights determined by spatial decay and commuting to map labor demand shocks to a residential tract and trace out the housing supply curve.

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 sums interactions

21. I also develop moment conditions that can identify all housing and labor demand elasticities in Appendix D, which exploit further interactions of city geography and local economic shocks.

of changes in wages or employment with ex ante employment shares across industries:

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

Because the demand shock embeds information on ex ante industry shares, an implicit identification assumption is that changes in non-productivity latent variables (e.g., amenities) are uncorrelated with prior industry structure. To ensure that local factors do not drive national changes, I exclude all workers in California.

5.2 The labor supply elasticity (Fréchet shape parameter)

The shape parameter ϵ governs the homogeneity of location preference and is an extensive-margin labor supply elasticity that conditions on commuting and residential geography. I estimate ϵ using a two-step strategy. In the first step, I recover $\hat{\omega}_{jt}$ from Equation (13) estimated by either a linear model or Poisson PML. In the second step, $\hat{\omega}_{jt}$ is the dependent variable in the wage equation: $\Delta\hat{\omega}_{jt} = \epsilon\Delta w_{jt} + \Delta e_{jt}$. Instrumenting wage with the labor demand shock z_{jt} identifies the labor supply elasticity if:

$$\mathbb{E}[\Delta z_{jt} \times \Delta e_{jt}] = 0, \ \forall j. \tag{M-1}$$

Moment condition M-1 requires that the labor demand shock in a tract j be uncorrelated with unobservable changes in factors that shift labor supply to that tract, such as workplace-specific amenities or accessibility to tract j.[22] Place of residence-by-year fixed effects (θ_{it}) control for any changes in residential amenities that may be correlated with labor demand shocks.

Table 3 shows results using three different methods of recovering $\hat{\omega}_{it}$ instrumenting with the wage variant of the labor demand shock. Columns 1–2 estimate ω_{it} jointly using both years of data in a log-linear panel. Columns 3–4 use a separate Poisson PML model for each year to estimate ω_{it}, conditioning on bilateral travel costs. Columns 5–6 jointly use both years of data, but with a Poisson PML estimator and pair fixed effects. As such, columns 3–4 include all tract pairs with zero flows, columns 5–6 drop tract pairs that have zeros flows in both time periods, and columns 1–2 omit tract pairs with zero flows in any years (see footnote 8).

The first stage is significant, has the right sign, and is of a reasonable magnitude across all specifications.[23] Using the log-linear estimator of $\hat{\omega}$, estimates of ϵ are about 1. Unlike the commuting analysis, zeros matter because any individual tract may have many incoming zeros. Estimates of ϵ based on $\hat{\omega}$ from Poisson PML models vary between between 2.18 and 2.90. The inclusion of

22. Any differences in Δd_{ijt} that apply to tract j as a whole confound $\hat{\omega}_{jt}$ and potentially contaminate e_{jt}.

23. The first stage captures the transmission of national wage shocks to local wages, so a value near 1 is expected. A robust bootstrap 95% confidence interval is [0.562, 8.856] with a median value of 2.157; see Appendix E. Appendix Tables H7 and H8 provide specification tests from Goldsmith-Pinkham, Sorkin, and Swift (2020) that correspond to columns 5–6 of Table 3.

subcounty-by-year fixed effects in columns 4 and 6 identifies the effect more narrowly from sub-regional variation. I take $\epsilon = 2.18$ from Column 6 as the preferred estimate. The generally low value of ϵ implies that workers are quite heterogeneous in location preferences and is roughly in line with extensive-margin labor supply elasticities found in studies of labor markets: Falch (2010) estimates labor supply elasticities between 1.0 and 1.9, Suárez Serrato and Zidar (2016) find values between 0.75 and 4.2, and Albouy and Stuart (2020) recover 1.98.

M-1 is a local, rather than citywide, moment condition, and is weaker than standard applications of shift-share labor supply instruments. When a citywide labor demand shock is used to trace out labor supply, identification requires the shock be orthogonal to non-wage determinants of labor supply (i.e., $\mathbb{E}[\Delta \bar{z}_t \cdot (\Delta \bar{b}_t + \Delta \bar{d}_t + \Delta \bar{e}_t)] = 0$, where $\bar{\cdot}$ averages over locations within a city). This suggests potential pitfalls in standard applications of shift share instruments. First, changes in residential amenities, commuting costs, and workplace amenities must be uncorrelated with labor demand shocks *regardless of where in the city they occur*. Second, changes in amenities cannot be correlated with the labor demand shock, whereas in my design, even unobserved and time-varying residential amenities are not confounding. Finally, in the standard design, changes in commuting costs cannot be correlated with labor demand shocks locally or elsewhere within the city.

The urban economic geography literature often identifies ϵ from a combination of modeling assumptions and cross-sectional variation in travel time. Because I observe workplace wage, I can test a prominent assumption. Ahlfeldt et al. (2015) select ϵ to rescale the variance of model-implied wages.[24] This implicitly disallows positive correlation between w and e, and requires either (i) $\mathbb{E}[w_j e_j]/\mathbb{V}[e_j] = -1/2\epsilon$ or (ii) $\mathbb{V}[e_j] = 0$.[25] Panel C of Table 3 presents tests of (i), where estimates of ϵ from Panel A are enforced to calculate $e_j = \hat{\omega}_j - \hat{\epsilon} w_j$. Most specifications show a negative correlation between w_j and e_j, as required if $\mathbb{E}[w_j e_j]/\mathbb{V}[e_j] = -1/2\epsilon$. The first row of p-values treats ϵ as a constant, the second as a random variable (see Appendix D for details). If ϵ is fixed, most specifications reject the null hypothesis that $\mathbb{E}[w_j e_j]/\mathbb{V}[e_j] = -1/2\epsilon$. However, if uncertainty in ϵ is taken into account, only Columns 3 and 4 reject the null.

Supplementary results in Appendix D indicate that wages explain only a relatively small amount of the variation in workplace fixed effects ω. Similarly, Kreindler and Miyauchi (2020)—who also observe workplace wage—find only a modest cross-sectional relationship. Together, these results highlight the importance of carefully considering the relationship between workplace

24. Other papers take varied, ad hoc approaches to unobserved workplace wages. For example, Monte, Redding, and Rossi-Hansberg (2018) assume an elasticity of substitution σ, specify a trade-in-goods model to recover productivity from cross-sectional trade flows, then assume recovered productivity is orthogonal to workplace and origin-destination specific amenities $\mathbb{E}[a_j(\sigma) \times (e_j + d_{ij})] = 0, \forall i, j$, implicitly requiring correct model specification.

25. To see this, Ahlfeldt et al. (2015) require $\mathbb{V}(\mathring{\omega}_j) = \mathbb{V}(w_{\bar{j}})$, where $\mathring{\omega} = \mathring{w}_j + \frac{1}{\epsilon} e_j$, $\mathring{w}_j$ are model-implied wages, and $\mathbb{V}(w_{\bar{j}})$ is the variance of average wage across twelve districts of Berlin. Rearranging, noting that e_j is mean zero, and substituting observed wage w_j for $\mathring{w}_j$ and $w_{\bar{j}}$ yields $2\epsilon\mathbb{E}[w_j e_j] + \mathbb{V}[e_j] = 0$. Consider three cases: (i) if w_j and e_j are negatively correlated, $\mathbb{E}[w_j e_j]/\mathbb{V}[e_j] = -1/2\epsilon$; (ii) $\mathbb{E}[w_j e_j] \Leftrightarrow \mathbb{V}[e_j] = 0$; and (iii) positive correlation w_j and e_j imply a negative variance of e.

wages and workplace amenities. While the variance assumption used to identify ϵ in Ahlfeldt et al. (2015) is unlikely to hold precisely or be universally reasonable, it may not be unreasonable where w_j and e_j are expected to have moderate negative correlation.

5.3 The (inverse) housing supply elasticity

A labor demand shock in one location shifts demand for housing in locations where workers might live, and thus can be used to identify the slope of the housing supply curve. To map labor demand shocks to residential housing shocks, I use linear combinations of shocks $\Delta z_{it}^{HD} = \mathbf{z}_t \cdot \boldsymbol{\vartheta}_i$, with weights ϑ that decay in travel time between locations that ever have positive commuting:

$$\Delta z_{it}^{HD}(\rho) = \sum_s \frac{e^{-\rho\vartheta_{is}} 1_{\check{n}_{is}>0} \Delta z_{st}}{\sum_s e^{-\rho\vartheta_{is}} 1_{\check{n}_{is}>0}},$$

where ϑ_{js} is the travel time between j and s, ρ is spatial decay, and $\check{n}_{is}$ denotes the maximum flow value from i to s in any year. With $\rho > 0$, labor demand shocks nearer a residential location i are more important than labor demand shocks farther from i. The resulting inverse-travel time weighted labor demand shock can be used to instrument housing density and identify ψ, the inverse price elasticity of housing supply, under the following condition:

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

Although both elements of M-2 relate to tract i, the housing demand shock draws on labor demand shocks from any j (including i).

M-2 requires labor demand shocks be uncorrelated with changes in (inverse) housing productivity, Δc_{it}, which reflects the efficiency of housing provision. One potential concern with Assumption A-2 is whether local zoning responds to local labor demand shocks. An alternative version drops the most local component of the labor demand shock (from i itself):

$$\Delta z_{it}^{HD,a}(\rho) = \sum_{s \neq i} \frac{e^{-\rho\vartheta_{is}} 1_{\check{n}_{is}>0} \Delta z_{st}}{\sum_{s \neq i} e^{-\rho\vartheta_{is}} 1_{\check{n}_{is}>0}}.$$

This identifies housing supply if $\mathbb{E}[\Delta z_{it}^{HD,a}(\rho) \times \Delta c_{it}] = 0, \ \forall i$, which is more easily parsed as a function of the labor demand shock itself:

$$k_{ij} \mathbb{E}[\Delta z_{jt} \times \Delta c_{it}] = 0, \ \forall \ i \neq j, \tag{M-2a}$$

where $k_{ij} = e^{-\rho\vartheta_{ij}} 1_{\check{n}_{ij}>0}$ is a weight. Condition M-2a requires that innovations in housing efficiency be uncorrelated with nearby productivity shocks.

Implementing these moment conditions requires choosing the spatial decay parameter, $\rho > 0$, governing how labor demand shocks propagate across space. I experiment with different values in $\ln(\rho) \in [-10, -2]$. I report results for $\ln(\rho) = -5.5$, estimated in differences using the employment instrument. Table 4 estimates the inverse housing supply elasticity in Equation 14 under M-2 (odd columns) and M-2a (even columns). First stage estimates (panel D) are significant and negative, indicating that a positive productivity shock decreases nearby residential density. Productivity shocks increase per worker demand for residential floorspace, which—given restrictive zoning in this area—either manifests as larger (likely single-family) homes or drives residential mobility to less dense housing options.

Estimates imply supply elasticities of 0.45–0.46 without income-driven adjustment in quantity (columns 1 and 2), and 0.57–0.62 when income can influence housing quantity (columns 3–6). Columns 3 and 4 provide a specification test for Equation (8).[26] Even columns exclude the own-tract labor demand shock when aggregating the instrument (under M-2a); this permits local housing productivity to covary with the local labor demand shock. Estimates are similar. Results suggest that local, tract-level housing provision is inelastic from 1990 to 2000. While Saiz (2010) finds the population-weighted housing supply elasticity in large U.S. cities is 1.75, the estimate for the Los Angeles area is 0.63, which is very close to the results in columns 3–6 of Table 4.

6 Non-commuting Effects of Transit

Given parameters $\{\epsilon, \psi, \alpha, \zeta\}$ and data on workplace wages, residential housing prices, and commuting, the model delivers straightforward expressions to recover local economic fundamentals—they are the model residuals $\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{E}\}$.[27] These economic fundamentals represent economic characteristics of a place that exist outside of a market equilibrium. In combination with market forces, fundamentals determine equilibrium prices and the distribution of people.

Because these residuals embed information on local fundamentals, they can be used to study the effects of policy. Consider a local intervention, T. In general, the intervention could impact any local fundamental. I model the effect of T on fundamentals with panel data as

$$\hat{\mathbf{Y}}_{it} = \boldsymbol{\lambda} T_{it} + \varsigma_i + \boldsymbol{\varepsilon}_{it}, \tag{15}$$

where $\boldsymbol{\lambda} = \{\lambda^A, \lambda^B, \lambda^C, \lambda^E\}$ are the effects to be estimated and $\hat{\mathbf{Y}} = \{\widehat{\mathbf{a}}, \widehat{\mathbf{b}}, \widehat{\mathbf{c}}, \widehat{\mathbf{e}}\}$ are the logged (non-commuting) fundamentals. Standard research designs can be used to identify $\boldsymbol{\lambda}$ (the full sample should be used to estimate the structural elasticities, but intervention effects may use a

26. The coefficients in columns 3–4 in Table 4 should be equal in magnitude and opposite in sign. They are not statistically different in absolute value (Panel C). A robust bootstrap 95% confidence interval is [1.008, 3.672] with a median value of 1.608 for the results in column 6; see Appendix E for details.

27. I assume $\zeta = 0.65$, implying that the household expenditure share on housing is $1-\zeta = 0.35$ and that the elasticity of housing demand is $-\epsilon(1-\zeta) = -0.76$. I assume that labor's share in production is $\alpha = 0.68$.

different sample). I use a continuous measure of tract location relative to transit stations:

$$T_{it} = \text{Proximity}_i^{\bar{d}} = \frac{\max\{0, \bar{d} - \min_k\{\text{dist}_i(\text{MetroStation}_k)\}\}}{\bar{d}} \in [0, 1],$$

where k indexes stations and $\bar{d}$ is some maximum distance (either 500m or 1km). This normalizes proximity to be one when a tract contains a station and zero if a tract is more than $\bar{d}$ from a station. I reuse the History & Shocks designs to identify λ^A and λ^B.[28]

Data limitations prohibit pre-trend analysis in fundamentals. However, recall the pre-trend comparisons in Section 3 (Appendix Table H2). The Immediate 1925 Plan sample does not exhibit differential pre-trends in model-relevant variables. However, because there are some pre-trends in other variables, I include the 1990 levels of sociodemographic variables (income, education, and manufacturing employment) to allow for differential trends according to initial conditions.

I find little evidence of large non-commuting effects of LA Metro Rail. Table 5 reports the estimated effect of changes in transit proximity between 1990 and 2000 on tract-level productivity (panels I and II) and residential amenities (panels III and IV). I assume exogenous fundamentals in panels I and III, but also show results from a model extension with endogenous agglomeration effects in panels II and IV (see Appendix D). Results are similar across different values of $\bar{d}$, different research designs, and whether or not endogenous agglomeration effects are accounted for. The one exception is that there is some evidence of a positive residential amenity effect in the PER sample. However, it is countered by smaller and insignificant estimates in other samples.

These results are perhaps surprising, as these margins have been the subject of considerable research. Duranton and Turner (2012) find evidence that transit increases city-level productivity and employment growth, while Kahn (2007) and Billings (2011) show some gentrifying effects of transit and that transit can anchor local development. Table 5 indicates that Los Angeles does not mirror such experience (at least by 2000) and is consistent with Schuetz (2015), who does not find that new rail transit stations generally increase consumption amenities in California.

There are two important caveats to the results in Table 5. These results apply only 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, responses that depend on scale—or are slow—could manifest in more recent years. Second, the estimates in Table 5, while generally insignificant, are not precisely estimated zeros. For example, the productivity estimates are all between about 0.03 and 0.05 log points. Their insignificance does not preclude small positive effects.

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. However, Figure 2 shows that rail usage among commuters is constant across most of the income distribution. Fur-

28. I assume that $\lambda^C = \lambda^E = 0$ as it is unlikely that transit itself can shift either of these margins. Appendix Table H10 tests these assumptions.

thermore, median household income in treated tracts does not appear to diverge after treatment (see Appendix Table H11). The lack of sorting response may be related to strict zoning and land use regulation (e.g., Quigley and Raphael 2005).[29]

7 Welfare Effects by 2000

I use the model, estimated and selected parameters, and treatment status to estimate counterfactuals and calculate welfare changes. I employ hat notation (Dekle, Eaton, and Kortum 2008), letting $\hat{X}_{it} = X'_{it}/X_{it}$, where $X'it$ is a counterfactual value. An iterative algorithm recovers counterfactual endogenous vectors $\{\hat{\mathbf{W}}, \hat{\mathbf{Q}}\}_{\forall i}$, $\{\hat{\boldsymbol{\pi}}\}_{\forall ij \in \mathcal{C}^+}$ (where $\mathcal{C}^+$ is the set of ij pairs with positive flows) relative to their observed values in 2000 (see Appendix C for details). Alternative scenarios are defined by adjusting fundamentals so that $\hat{X}_{i(j)} = \exp(-\lambda^X T_{i(j)})$, for $X \in \{A, B, C, D, E\}$. In the scenarios I consider below, I maintain the assumption that $\lambda^C = \lambda^E = 0$ and enforce the insignificance of other variables.[30]

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 (5):

$$\%\Delta \text{ Welfare } \approx \ln \hat{\bar{U}} = \frac{1}{\epsilon} \ln \left(\frac{\hat{B}_i \hat{E}_j \hat{D}_{ij} \hat{W}_j^{*\epsilon} \hat{Q}_i^{*-\epsilon(1-\zeta)}}{\hat{\pi}_{ij}^*} \right) \tag{16}$$

for each ij, where $\hat{X}^*$ indicates the equilibrium value of X in the counterfactual under autarky (that is, fixing $\hat{\bar{N}} = 1$). Because utility is homogeneous of degree one in wage, a proportional change in utility is equivalent to a 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. Instead, if the city is open, its total population $\hat{\bar{N}}$ also adjusts so that the expected utility in the city remains $\bar{U}$. Thus aggregate welfare for incumbent residents is unchanged. Because no spatial arbitrage means $\bar{U}$ in an open city is unchanged in response to changes in fundamentals, I instead report changes in total population.

Annualized costs combine two elements: (i) operating subsidies and (ii) annualized capital expenditures. The annual operating subsidy for the rail portion of LA Metro's operations for FY 2001–2002 is about \$162 million (2016 dollars). Total system cost for lines and stations completed by 1999 is \$8.7 billion (2016 dollars). I provide several annualizations of capital expenses. LA

29. I find little evidence of zoning changes near new LA Metro Rail stations; see Appendix Table H11.

30. That is, $\lambda^A = \lambda^B = 0$ from Table 5, and the third element of λ^D corresponding to proximity iii) is also 0 (from Table 1). Appendix Table H13 experiments with other λ^A and λ^B and reduced land use regulation.

Metro's borrowing terms at the time were about 6%, so the annual payment for a 30-year loan is roughly $635 million. However, subways last for a long time, so it may be appropriate to use a lower social discount rate. With a discount rate of 2.5% over an infinite horizon, capital expenditures are $218 million per year. Combining with the operating subsidy yields an annualized cost between $380 million and $797 million per year.

Table 6 reports the changes in aggregate welfare and population due to LA Metro Rail. Panels A and B report baseline parameter values and estimates, respectively (column 1), as well as bootstrap values (columns 2–3). The bootstrap procedure preserves the correlation structure across $\{\epsilon, \psi, \lambda^{D'}\}$; see Appendix E for details. LA Metro Rail by 2000 generates an annual baseline benefit of $93.6 million in 2016 USD (a 0.044% increase in welfare). In an open economy, the employed population of the Los Angeles region is 0.088% higher with LA Metro Rail. The 95% bootstrap confidence interval is [$11.9 million, $380.5 million] (an increase of welfare between 0.006% and 0.179%). The baseline commuting benefit of LA Metro Rail by 2000 is about 16% of an annualized cost of $597 million; the baseline benefit most likely lies between 2% and 64% of this annualized cost. Under the lowest assumption of the annualized costs, the baseline benefit is 25% of cost (confidence interval from 3% to 100%) by 2000. Under the highest assumption, the baseline benefit is 12% of cost (confidence interval from 1% to 48%) by 2000.

A general conclusion across baseline specifications is that the commuting benefit of rail transit in Los Angeles does not exceed its cost by 2000. Regardless of the discount rate, baseline benefits are a bit more than half of the operating subsidy of $162 million. However, the baseline commuting benefits do not cover the capital expenses at standard discount rates.

8 Extensions and Discussion

The baseline estimate of the benefit of LA Metro Rail in Section 7 solely reflects non-congestion commuting benefits of roughly $100 million per year. I consider several additional margins of benefits that the baseline analysis excludes, as there is significant uncertainty about their magnitude or persistence. Accounting for these margins may increase the benefit of LA Metro Rail by up to $300 million annually, roughly in line with annualized costs. I then discuss headwinds that may have limited the upside benefits of LA Metro Rail.

Longer-run Commuting Effects. Continued commuting growth from 2002–2015 between previously connected stations indicates that (i) aggregate commuting flows take decades to adjust to new transit modes (i.e., *habituation*), and/or (ii) there are increasing returns to transit network size. Only when assuming that this additional growth is due to habituation can we combine the additional benefits using the same cost basis as in Section 7.[31] Calculating welfare combining the effects of Panel A and the *existing connection* effects of Panel B in Table 1 yields $169.2 million an-

31. If the benefit is due to network effects, the cost basis should be adjusted to reflect network expansion costs.

nually by 2015. The increase of $75.6 million over the baseline is an upper bound estimate of the habituation effect. Unfortunately, data limitations prevent testing for longer-run effects on non-commuting margins.

Congestion. As discussed in Section 3.3, the Fundamental Law of Congestion suggests no long-run effect of transit capacity on road congestion. However, Table 2 shows that measurable congestion benefits persist several years after transit lines open. Implicitly assuming these congestion benefits last in perpetuity, I combine them with a two-step gravity-based estimate of the elasticity of commuting with respect to travel time of -0.239 (see Appendix F). The benefit of LA Metro Rail accounting for decreased congestion from transit is $225.9 million annually by 2000 (an increase of $132.3 million from baseline).

Air Pollution. Another benefit may be decreased air pollution from decreased congestion. Gendron-Carrier et al. (2021) show that subways lead to a mild decrease in air pollution in highly polluted cities. Applying their estimates to Los Angeles County, LA Metro Rail by 2000 would have led to reductions in air pollution corresponding to 50.4 fewer infant deaths annually. As the medium-run results travel-time savings in Table 2 are roughly one-third of those in Anderson (2014), I take one-third of the potential avoided infant deaths as a baseline long-run estimate of 16.8 fewer infant deaths. Assuming a standard value of statistical life of $6 million, this generates $101 million annually in additional benefits.[32]

Agglomeration. Incorporating agglomeration changes welfare little, because the relative effects of LA Metro Rail in any one location are not large. There are two margins to consider: At the metropolitan level, suppose a simple agglomerative force increases productivity by 5% everywhere for each doubling of city population. Welfare is homogeneous of degree one in wage, and wage is proportional to productivity, so productivity in an open city increases by $\ln(1 + 0.05 \times 0.00088) \approx 0.0044\%$, about $9 million annually. The other margin is local agglomeration. Including these forces as implemented in Ahlfeldt et al. (2015) slightly decreases the welfare generated by LA Metro Rail (to $91.8 million), indicating that LA Metro Rail slightly decentralizes population.

8.1 Discussion: Headwinds

Some particular characteristics of LA Metro Rail and Los Angeles warrant note when interpreting these results to draw broader conclusions. Of note are targeting and disperse commuting patterns, zoning, costs, and federal funding.

LA Metro Rail does not connect the residences and workplaces of many commuters. Only 1%–3% of the 1990 population of Los Angeles County both lived and worked in tract pairs near rail stations by 2000 (Appendix Table H1). Figure 3 plots the likelihood of becoming treated by ex

32. This calculation is imprecise and only meant to give a sense of magnitude. An alternative calculation using results from Deryugina et al. (2019) indicates $13 million–$14 million annually in benefits due to decreased mortality and hospital expenditures among local Medicare recipients.

ante (1990) flows for pairs ij with $i \neq j$ in the Immediate 1925 Plan sample. While the positive relationship indicates that transit was sometimes placed where it could have a larger effect, many high-flow pairs are not connected. Linking denser corridors (e.g., Wilshire Boulevard) would have generated greater gains. Regardless, Los Angeles has a polycentric distribution of jobs and residences (Redfearn 2007) that is less amenable to transit adoption. Indeed, residing close to a transit station increases the likelihood of using LA Metro Rail by only 0.8 percentage points without conditioning on workplace on a base of zero (Appendix Table H12).[33]

Land use regulations inhibit the ability of locations receiving stations to adjust building stock (Bunten and Rolheiser 2020; Schuetz, Giuliano, and Shin 2018). Essentially no land was converted to residential use near LA Metro Rail before 2000 (Appendix Table H11). Moreover, Proposition U, which passed in 1986, halved allowable density throughout much of Los Angeles just before LA Metro Rail opened. Such legislation combined with political constraints meant the "coordinated land use and rail planning ... died a gory death" (Elkind 2014, p. 71). Restrictive zoning likely slowed adjustment to and adoption of LA Metro Rail; longer-run analysis may well find larger effects. Indeed, if land use regulations had eased to permit 10% higher residential density in census tracts that contained transit stations, LA Metro Rail's effect would have been 44%–140% larger without much (or any) additional expense (see Appendix Table H13).

While I take costs as given, lowering the capital and operating costs of transit would aid cost-benefit calculations. There is a growing body of evidence that infrastructure is more costly in the United States than elsewhere in the world and that these costs have been increasing over time (Brooks and Liscow 2019; Levy 2016; Mehrotra, Turner, and Uribe 2019). While there is not yet a clear consensus on solutions to these differences, if the costs of LA Metro Rail were, for example, half of their observed levels, estimated benefits would meet or exceed costs.

Finally, capital expenditures on transit are largely funded with federal dollars in the U.S., with states and localities making up the difference. From 1996 and 1999, 26%–45% of LACMTA's capital expenditures were funded with local dollars. It is therefore possible for a benefit-cost calculation considering only local costs to be positive, which may be the relevant decision margin for local decision makers.

There are also margins to which this paper does not speak. City-wide effects are difficult to measure with this approach. Nor can I directly speak to benefits resulting in better transit provision for non-commuting trips (though this margin could show up as a residential amenity, which I do not find). This framework does not capture the benefits for non-workers. Such effects are particularly important for equity concerns and are unfortunately understudied. Relatedly, the

33. Tsivanidis (2018) finds that Bus Rapid Transit (BRT) in Bogotá increases welfare roughly 40 times more than the baseline effect of LA Metro Rail. Comparing commuter behavior across the two cities suggests that the difference is due to relatively low adoption of rapid transit in Los Angeles. Before Bogotá's BRT was built, 73% of commuters took the bus; afterwards, the BRT had 2.2 million trips/day and 36% of commuters used it. In contrast, in Los Angeles before Metro Rail, 5%–7% of commuters took the bus; by 2000, just 0.4% of commuters used the subway, with at most 150,000 trips/day.

assumption of homothetic preferences may limit my ability to measure more sizable utility gains for populations with greater transportation cost sensitivity.

9 Conclusion

I develop a method for evaluating the benefits of transit from commuting flow data and estimate the impact of Los Angeles Metro Rail. A parsimonious model permits transparent identification and estimation, but can still isolate commuting effects from non-commuting channels (e.g., amenities). LA Metro Rail increases commuting between the census tracts nearest to stations by 16% in the first decade after construction, relative to control groups selected by proposed and historical transit routes. There is some evidence that Metro Rail reduces congestion in the medium run.

The elasticity of labor supply plays a key role in the model and governs homogeneity in location preference; its small estimated value indicates agents are relatively unwilling to relocate and are not very responsive to changes in local conditions. Conversely, this implies that observed responses to transit correspond to significant utility gains.

Baseline welfare estimates show positive annual benefits of LA Metro Rail to be $94 million by 2000. These welfare benefits are smaller than the operational and capital costs of LA Metro's light rail and subway lines. When combined with additional congestion effects and back-of-the-envelope calculations on the benefits of reduced air pollution, total benefits may be up to $300 million per year more. While the estimates omit some margins (such as benefits for non-workers), results warrant a note of caution to cities—and particularly polycentric, automobile-oriented cities—expecting rail investment to dramatically alter their commuting environment within 10 to 25 years.

References

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.

Albouy, David, and Bryan A Stuart. 2020. "Urban population and amenities: The neoclassical model of location." *International Economic Review* 61 (1): 127–158.

Allen, Treb, and Costas Arkolakis. 2019. "The welfare effects of transportation infrastructure improvements." *NBER Working Paper,* no. w25487.

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

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.

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

Brinkman, Jeffrey C. 2016. "Congestion, agglomeration, and the structure of cities." *Journal of Urban Economics* 94:13–31.

Brooks, Leah, and Zachary D Liscow. 2019. "Infrastructure costs."

Brooks, Leah, and Byron Lutz. 2019. "Vestiges of transit: Urban persistence at a microscale." *Review of Economics and Statistics* 101 (3): 385–399.

Bunten, Devin Michelle, and Lyndsey Rolheiser. 2020. "People or parking?" *Habitat International* 106:102289.

Dekle, Robert, Jonathan Eaton, and Samuel Kortum. 2008. "Global rebalancing with gravity: Measuring the burden of adjustment." *NBER Working Paper,* no. w13846.

Deryugina, Tatyana, Garth Heutel, Nolan H Miller, David Molitor, and Julian Reif. 2019. "The mortality and medical costs of air pollution: Evidence from changes in wind direction." *American Economic Review* 109 (12): 4178–4219.

Dingel, Jonathan I, and Felix Tintelnot. 2020. "Spatial economics for granular settings." *NBER Working Paper,* no. w27287.

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

Downs, Anthony. 1962. "The law of peak-hour expressway congestion." *Traffic Quarterly* 16 (3).

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.

———. 2018. "Urban form and driving: Evidence from US cities." *Journal of Urban Economics* 108:170–191.

Elkind, Ethan N. 2014. *Railtown: The fight for the Los Angeles metro rail and the future of the city.* Oakland, CA: University of California Press.

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.* Oakland, California: University of California Press.

Gendron-Carrier, Nicolas, Marco Gonzalez-Navarro, Stefano Polloni, and Matthew A Turner. 2021. "Subways and urban air pollution." *American Economic Journal: Applied Economics* Forthcoming.

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.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift. 2020. "Bartik instruments: What, when, why, and how." *American Economic Review* 110 (8): 2586–2624.

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.

Kelker, De Leuw & Co. 1925. *Report and recommendations on a comprehensive rapid transit plan for the City and County of Los Angeles.* Technical report. Chicago.

Kreindler, Gabriel, and Yuhei Miyauchi. 2020. "Measuring commuting and economic activity inside cities with cell phone records." *NBER Working Paper,* no. w28516.

Larch, Mario, Joschka Wanner, Yoto V Yotov, and Thomas Zylkin. 2019. "Currency unions andtTrade: A PPML re-assessment with high-dimensional fixed effects." *Oxford Bulletin of Economics and Statistics* 81 (3): 487–510.

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.

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.

Levy, Alon. 2016. *Why Costs Matter.* https://pedestrianobservations.com/2016/01/31/why-costs-matter/. Accessed: 2021-02-14.

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.

Mehrotra, Neil, Matthew A Turner, and Juan Pablo Uribe. 2019. "Does the US have an infrastructure cost problem? Evidence from the Interstate Highway System."

Monte, Ferdinando, Stephen J Redding, and Esteban Rossi-Hansberg. 2018. "Commuting, migration, and local employment elasticities." *American Economic Review* 108 (12): 3855–3890.

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

Redding, Stephen J, and Matthew A Turner. 2015. “Transportation costs and the spatial organization of economic activity.” In *Handbook of Regional and Urban Economics,* 5th ed., edited by Gilles Duranton, J Vernon Henderson, and William C Strange, 1339–1398. Elsevier.

Redfearn, Christian L. 2007. “The topography of metropolitan employment: Identifying centers of employment in a polycentric urban area.” *Journal of Urban Economics* 61 (3): 519–541.

———. 2009. “How informative are average effects? Hedonic regression and amenity capitalization in complex urban housing markets.” *Regional Science and Urban Economics* 39 (3): 297–306.

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 (14): 2699–2723.

Schuetz, Jenny, Genevieve Giuliano, and Eun Jin Shin. 2018. “Does zoning help or hinder transit-oriented (re)development?” *Urban Studies* 55 (8): 1672–1689.

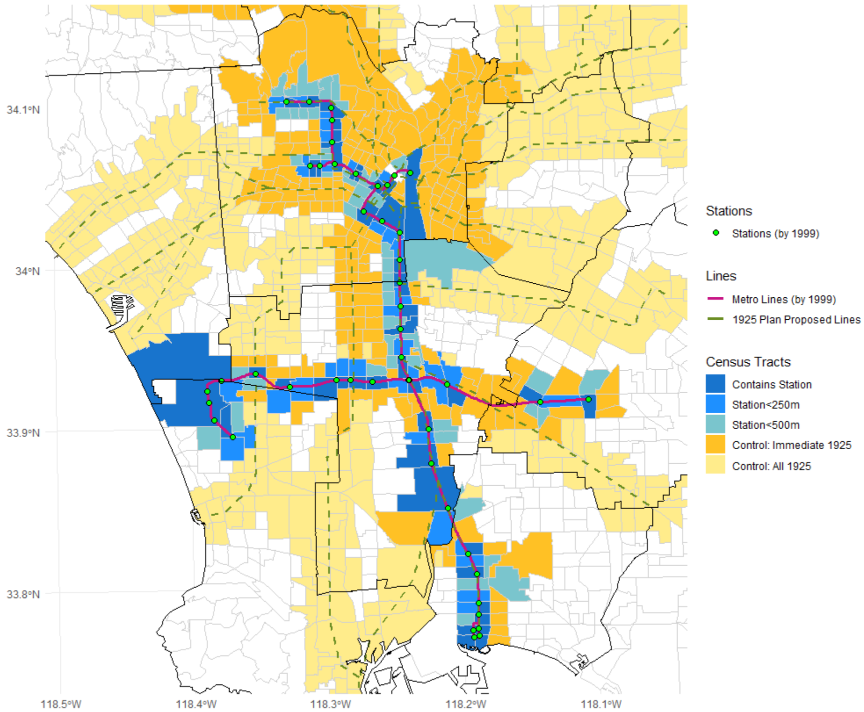
Severen, Christopher, and Andrew J Plantinga. 2018. “Land-use regulations, property values, and rents: Decomposing the effects of the California Coastal Act.” *Journal of Urban Economics* 107:65–78.

Suárez Serrato, Juan Carlos, and Owen Zidar. 2016. “Who benefits from state corporate tax cuts? A local labor markets approach with heterogeneous firms.” *American Economic Review* 106 (9): 2582–2624.

Tsivanidis, Nick. 2018. “The aggregate and distributional effects of urban transit infrastructure: Evidence from Bogotá’s TransMilenio.”

Figure 1: Map of LA Metro lines, stations, and the 1925 Plan and PER Lines

(a) 1925 Plan Sample



(b) PER Lines Sample

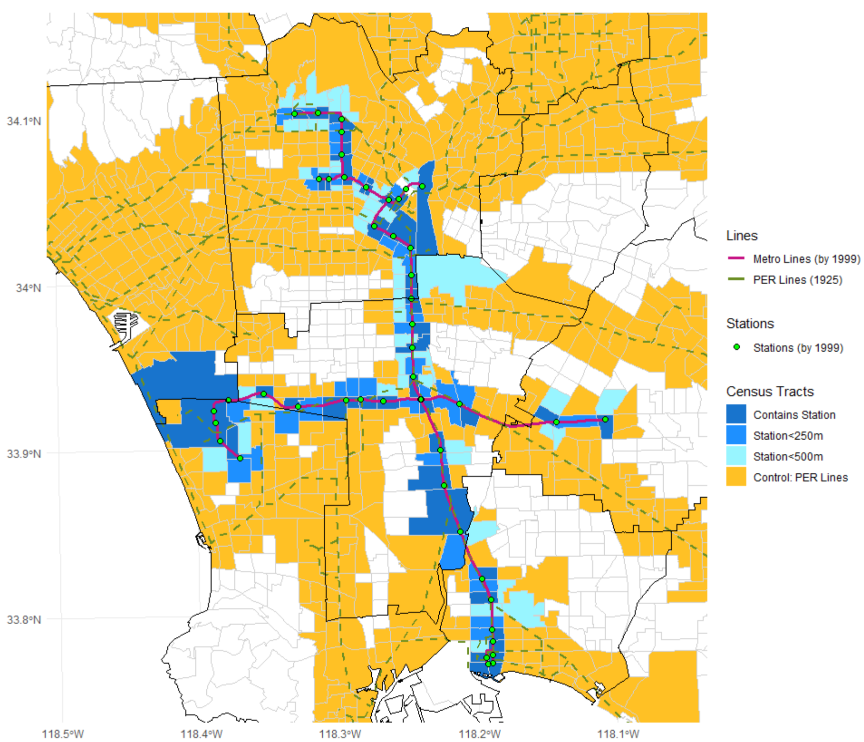


Figure 2: Take-up of LA Metro Rail by commuters does not vary by income, but overall take-up of transit (including bus) does.

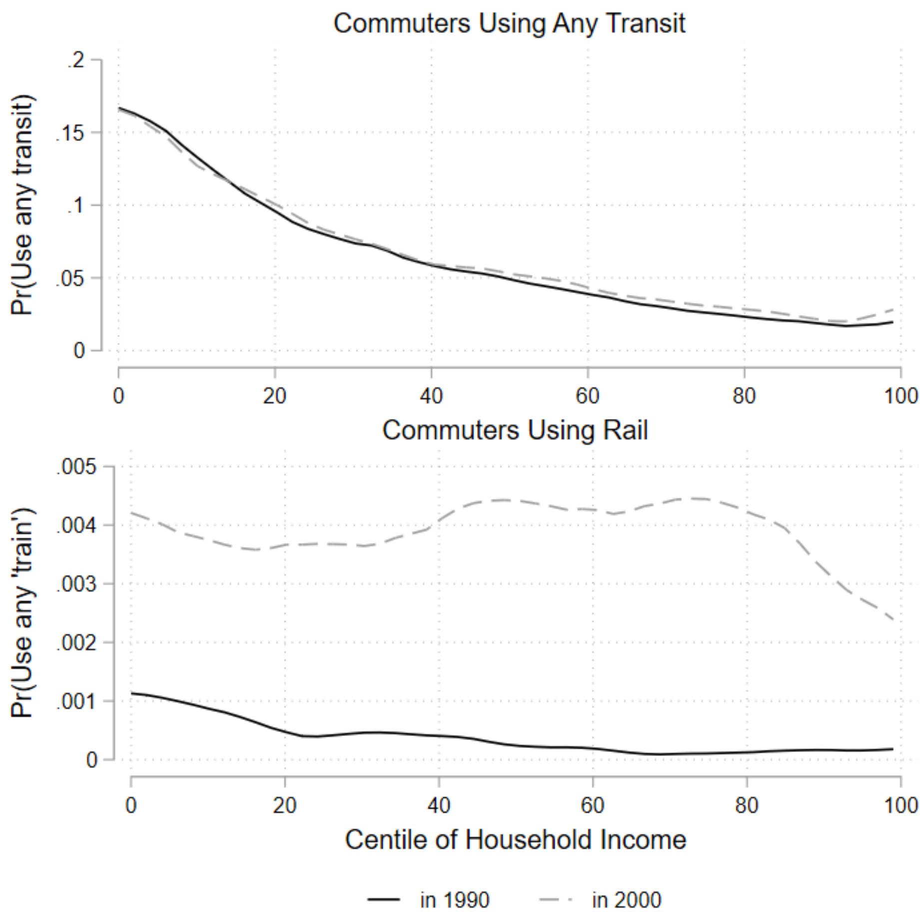


Figure 3: Tract pairs with higher ex ante commuting flows are a bit more likely to receive LA Metro Rail by 2000, but many high commuting pairs do not.

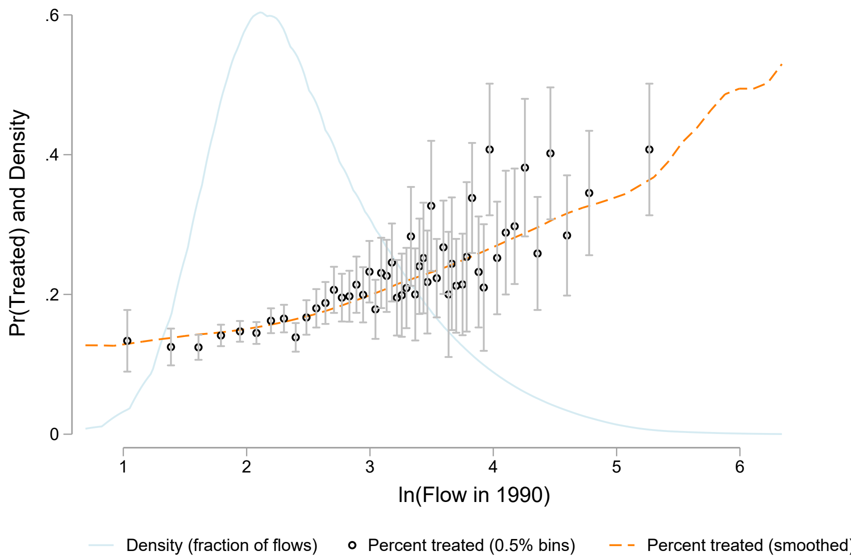


Table 1: Effect of transit on commuting flows by 2000 (log-linear)

	Full Sample			History & Shocks			Same Line	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Commuting effects 1990–2000 (CTPP data)								
O & D contain station	0.100***	0.111***	0.100***	0.111**	0.145***	0.149**	0.157*	0.201**
	(0.038)	(0.038)	(0.038)	(0.044)	(0.045)	(0.061)	(0.089)	(0.093)
O & D <250m from station		0.076*	0.056	0.090*	0.104**	0.128**	0.051	0.061
		(0.046)	(0.047)	(0.051)	(0.051)	(0.065)	(0.075)	(0.079)
O & D <500m from station		0.001	-0.013	0.020	0.014	0.012	-0.032	0.070
		(0.037)	(0.036)	(0.040)	(0.042)	(0.053)	(0.067)	(0.067)
N	291532	291532	291110	99480	74408	19222	8280	4496
B. Commuting effects 2002–2015 (LODES data)								
New connections								
O & D contain station	0.108***	0.110***	0.094***	0.105***	0.112***	0.133***		
	(0.034)	(0.034)	(0.032)	(0.033)	(0.031)	(0.036)		
O & D <250m from station	0.038	0.038	0.022	0.035	0.046**	0.083***		
	(0.024)	(0.024)	(0.024)	(0.024)	(0.023)	(0.027)		
O & D <500m from station	0.032	0.031	0.017	0.035*	0.025	0.045*		
	(0.022)	(0.023)	(0.021)	(0.020)	(0.020)	(0.025)		
Existing connections								
O & D contain station		0.091**	0.084**	0.099***	0.098***	0.112***		
		(0.038)	(0.036)	(0.034)	(0.032)	(0.030)		
O & D <250m from station		0.049*	0.048*	0.059*	0.061*	0.091***		
		(0.027)	(0.029)	(0.032)	(0.035)	(0.029)		
O & D <500m from station	0.056**	0.048*	0.043*	0.041	0.028	0.032		
	(0.022)	(0.025)	(0.025)	(0.025)	(0.025)	(0.029)		
N	1993198	1993198	1992702	514082	385278	105794		
Control Group	All	All	All	PER	Full '25 Plan	Immed. '25 Plan	Ever Treated	Treated by 2000
Standard Three-Way FEs	Y	Y	Y	Y	Y	Y	Y	Y
Subcounty Pair-×-Year FEs	-	-	Y	Y	Y	Y	Y	Y
Highway Controls	-	-	Y	Y	Y	Y	Y	Y

High-dimensional fixed effects estimates of λ^D with log-linear estimator; standard three-way fixed effects are tract of work-by-year, tract of residence-by-year, and tract pair. Outcome is log commuting flow. Treatment variables are mutually exclusive. Column titles refer to design: tracts pairs on any lines are treated in Columns (1)-(6), while only tract pairs on the same line are treated in Columns (7) and (8). Panel A uses CTPP data and is normalized to 1990 geographies. Panel B uses LODES data normalized to 2010 geographies. Standard errors clustered by tract pair, tract of residence, and tract of work in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Does transit decrease congestion?

	$\ln(\text{Time}_{ijt}^{\text{All}})$		$\ln(\text{Time}_{ijt}^{\text{Car}})$	
	(1)	(2)	(3)	(4)
Share of route <250m from transit line	-0.079*	-0.020	-0.144*	-0.150*
	(0.043)	(0.042)	(0.084)	(0.085)
Share of route 250m-500m from transit line	-0.051	-0.023	-0.197**	-0.189**
	(0.062)	(0.063)	(0.089)	(0.093)
Share of route 500m-1km from transit line	-0.036	-0.016	0.011	-0.024
	(0.048)	(0.047)	(0.076)	(0.075)
Share of route 1km-2km from transit line	-0.039	-0.022	-0.052	-0.070
	(0.028)	(0.028)	(0.057)	(0.054)
Share of route 2km-4km from transit line	-0.006	0.012	0.013	0.002
	(0.019)	(0.019)	(0.032)	(0.035)
N	286392	286142	89614	89432
Control Group	All	All	All	All
Standard Three-Way FEs	Y	Y	Y	Y
Subcounty Pair-$\times$-Year FEs	-	Y	-	Y
Transit Controls	Y	Y	Y	Y
Highway Controls	Y	Y	Y	Y

High-dimensional fixed effects estimates of the changes in the share of a route near transit on log travel time; standard three-way fixed effects are tract of work-by-year, tract of residence-by-year, and tract pair. Average reported travel time from the CTPP reflects all modes in Columns (1)-(2) and only automobiles in Columns (3)-(4). Times with implied speeds greater than 80mph are excluded. Treatment variables are mutually exclusive. Highway controls are shares of a route within 250m and 1km of the Century Freeway, and transit controls are the treatment variables in Table 1. Standard errors clustered by tract pair, tract of residence, and tract of work in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: IV estimates of labor supply elasticity (ϵ)

	$\Delta\hat{\omega}_{jt}$					
	(1)	(2)	(3)	(4)	(5)	(6)
A. IV Estimates of ϵ						
$\Delta \ln(W_{jt})$	1.284*	0.807	2.904**	2.373**	2.512**	2.180*
	(0.698)	(0.693)	(1.221)	(1.206)	(1.094)	(1.171)
B. First Stage						
Δz_{jt}	1.150***	0.918***	1.165***	0.939***	1.165***	0.939***
	(0.332)	(0.333)	(0.331)	(0.332)	(0.331)	(0.332)
F-stat (CD)	48.2	27.9	50.8	29.8	50.8	29.8
F-stat (KP)	12.0	7.6	12.4	8.0	12.4	8.0
N	2432	2426	2533	2525	2533	2525
C. Ahlfeldt et al. (2015) Specification Test ($H_0 : \mathbb{E}[w_j e_j]/\mathbb{V}[e_j] = -1/2\epsilon$)						
$\mathbb{E}[w_j e_j]/\mathbb{V}[e_j]$	-0.313	-0.219	-0.015	0.010	-0.178	-0.165
p-value of H_0, ϵ fixed	[0.018]	[0.000]	[0.000]	[0.000]	[0.137]	[0.000]
p-value of H_0, $\hat{\epsilon}$ random var.	[0.729]	[0.455]	[0.031]	[0.037]	[0.819]	[0.605]
$\hat{\omega}$ estimated:	Linear Panel		PPML Yr-by-Yr		PPML Panel	
Subcounty-$\times$-Year FEs	-	Y	-	Y	-	Y

Panel instrument variable (IV) estimates of regression of $\hat{\omega}_{jt}$ on w_{it} estimated in differences using wage instrument. CD and KP refer to the Cragg-Donald and Kleibergen-Paap tests, respectively. Weighted by 1990 workplace employment. Columns 1-2 use a log-linear panel specification to estimate ω_{jt}; columns 3-6 use PPML estimation (estimated year-by-year in columns 3-4 using distance as a gravity term and as a panel in columns 5-6 with ij fixed effects). Place of work-by-year fixed effects ($\hat{\omega}_{jt}$) estimated in the panel in columns 1 and 3 with ij fixed effects. Panel C presents specification tests of Ahlfeldt et al. (2015), see Section 5.2. Sample size reflects count of differenced observations. Robust standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: IV estimates of inverse housing supply elasticity (ψ)

	$\Delta q_{it} = \Delta \ln(Q_{it})$					
	(1)	(2)	(3)	(4)	(5)	(6)
A. IV Estimates of ψ						
$\Delta \ln(\text{Density}_{it})$	2.325**	2.168***				
	(0.939)	(0.807)				
$\Delta \ln(\text{Hous. Cons.}_{it})$			1.762***	1.644***		
			(0.533)	(0.476)		
$\Delta \ln(\text{Res. Land}_{it})$			-2.082**	-1.927**		
			(0.970)	(0.945)		
$\Delta \ln(\text{Hous. Density}_{it})$					1.687***	1.602***
					(0.508)	(0.453)
B. Housing Supply Elasticity						
$1/\psi$	0.430**	0.461***	0.568***	0.608***	0.593***	0.624***
	(0.174)	(0.172)	(0.172)	(0.176)	(0.178)	(0.177)
C. Specification Test ($H_0 : \psi_{\text{Hous. Cons.}}/\psi_{\text{Res. Land}} = -1$)						
$\psi_{\text{Hous. Cons.}}/\psi_{\text{Res. Land}}$			-0.846	-0.853	-1	-1
p-value of H_0			[0.590]	[0.616]		
D. First Stage						
$\Delta z_{it}^{HD(,a)}(\rho)$	-1.397***	-1.468***	-2.398***	-2.303***	-1.945***	-2.003***
	(0.525)	(0.504)	(0.711)	(0.721)	(0.588)	(0.571)
F-stat (CD)	5.5	6.1	9.6	10.7	10.9	11.5
F-stat (KP)	7.1	8.5	11.3	12.4	10.9	12.3
N	2232	2232	2175	2175	2175	2175
Empl. instrument	All	Not i	All	Not i	All	Not i

Panel instrument variable (IV) estimates of regression of median house value on population, housing consumption, and residential land, using $\ln(\rho) = -5.5$ and estimated in differences using the employment instrument. Housing prices excluded if either year is top-coded. Weighted by 1990 number of homeowners. Columns 2, 4, and 6 exclude own tract during instrument construction. CD and KP refer to the Cragg-Donald and Kleibergen-Paap tests, respectively. Sample size reflects count of differenced observations. Robust standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Transit and non-commuting fundamentals (other effects of transit)

	$\Delta\hat{\mathbf{Y}}_{it}$ (Productivity and Amenities)							
	$\bar{d}$ = 500m				$\bar{d}$ = 1km			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Effect on productivity								
I.) λ^A estimated using $\Delta\hat{A}$								
Proximity$_i \times t$	0.033	0.029	0.016	0.020	0.041	0.040	0.026	0.034
	(0.039)	(0.039)	(0.039)	(0.043)	(0.037)	(0.038)	(0.037)	(0.043)
N	2509	1167	934	394	2509	1167	934	394
II.) λ^A estimated using $\Delta\hat{\mathcal{A}} = \Delta\hat{A} - \mu\Delta\ln(\Upsilon)$								
Proximity$_i \times t$	0.028	0.029	0.017	0.021	0.036	0.041	0.026	0.036
	(0.038)	(0.039)	(0.038)	(0.042)	(0.036)	(0.037)	(0.037)	(0.042)
N	2469	1167	934	394	2469	1167	934	394
Effect on residential amenity level								
III.) λ^B estimated using $\Delta\hat{B}$								
Proximity$_i \times t$	0.053	0.070**	0.047	0.008	0.038	0.057*	0.027	-0.027
	(0.033)	(0.034)	(0.034)	(0.036)	(0.029)	(0.032)	(0.031)	(0.035)
N	2160	994	815	343	2160	994	815	343
IV.) λ^B estimated using $\Delta\hat{\mathcal{B}} = \Delta\hat{B} - \eta\Delta\ln(\Omega)$								
Proximity$_i \times t$	0.049	0.066**	0.043	0.007	0.034	0.054*	0.023	-0.027
	(0.031)	(0.032)	(0.032)	(0.034)	(0.028)	(0.030)	(0.029)	(0.032)
N	2153	993	814	343	2153	993	814	343
Control Group	All	PER	Full '25 Plan	Immed. '25 Plan	All	PER	Full '25 Plan	Immed. '25 Plan

Results from thirty-two regressions of transit proximity on local fundamentals. Here, the distance effect of agglomeration decays at the values in Ahlfeldt et al. (2015). All regressions include tract fixed effects, subcounty-by-year fixed effects, and controls. Controls include changes in highway proximity and 1990 levels of log household income, share of residents with at least a high school degree, and manufacturing employment. Sample size reflects number of differenced tracts. Robust standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Welfare estimates in 2000 (in $2016)

	Preferred Estimate (1)	Bootstrapped Results: Median (2)	Bootstrapped Results: 95% CI (3)
A. Parameter Values			
ϵ	2.180	2.157	[0.562, 8.856]
ψ	1.602	1.608	[1.008, 3.672]
λ^D, O & D contain station	0.149	0.136	[0.016, 0.257]
λ^D, O & D <250m from station	0.128	0.117	[0.022, 0.211]
B. Bootstrapped Welfare Change (Primary Model)			
Closed Economy			
Annual Δ in welfare	0.044%	0.040%	[0.006%, 0.179%]
(in $2016)	**$93.6 mil.**	$85.1 mil.	**[$11.9 mil., $380.5 mil.]**
Open Economy			
Population Δ	0.088%	0.081%	[0.008%, 0.348%]
C. Extensions ($2016)	Total		Addition to $93.6 mil.
Congestion†	$225.9 mil.		$132.3 mil.
Network Effects through 2015†	$169.2 mil.		$75.6 mil.
Agglomeration	$91.9 mil.		-$1.7 mil.
Air Pollution Mortality	$194.4 mil.		$100.8 mil
D. Annualized Cost ($2016)			
Capital Costs			
At 6% over 30 years	-$635 mil.		
At 5% over 50 years	-$479 mil.		
At 5% in perpetuity	-$435 mil.		
At 2.5% in perpetuity	-$218 mil.		
Operational subsidy in 2002	-$162 mil.		
Total Cost	**-$380 to -$797 mil.**		

Op. subsidy refers to the annual operation subsidy. Other parameters are $\zeta = 0.65$, $\alpha = 0.68$, $\epsilon\kappa = -0.239$, and λ^D as reported in column 6 of Table 2 (with the coefficient corresponding to distance iii) set to 0). Bootstrap results reflect 400 wild bootstrap draws. See text and appendices for details. †indicates an upper bound on the effect.

For Online Publication

Appendices and Supplemental Results

to accompany

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

by

Christopher Severen

July 2021

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.

A.1 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
- 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) Origin-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 1970-2000
- Los Angeles County Metropolitan Transportation Authority (LACMTA)
  - Shapefiles of LA Metro stations and lines
  - Opening dates for stations and lines
  - Ridership data
  - Kelker, De Leuw & Company (1925). I georeference this map in ArcGIS, and then process 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)
- Dynamap Road Network (Tele Atlas)
  - Calculate travel times using ArcGIS
- HERE API
  - Calculate travel times using HERE API, interfaced via HereR package in R
- Graphhopper routing engine and OpenStreetMap data
  - Develop shapefiles for fastest driving routes between tracts, and intersect with buffers around LA Metro Rail lines to determine route exposure transit.

A.2 Data construction details

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 these 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

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.

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), as well as a small number of flows that failed to merge.

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 accord reasonably well with ad hoc probes of quality. However, there are instances of extreme values that become 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. While I select (i) for all results reported in the main paper, the extended identification results for the labor demand elasticities use data cleaned according to (iiib), likely 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; this matters most for the estimation of ψ and recovery of residential amenities $\mathbf{B}$. If a variable was top-coded differently in different years, I standardize the top code to the most conservative year.

Travel time, route construction, and exposure to transit

Travel times were originally calculated using Dynamap data and ArcGIS, however, I no longer have access to that data. I also have calculated travel times using HERE's API `developer.here.com` to ensure that those numbers are reasonable (they are). In general, I use the Dynamap-generated travel times because it was based on an older model of the street network, and thus is more likely to match the driving environment under study (in 1990 and 2000) than HERE's travel times, which reflect the current street network.

To determine which driving routes were most exposed to LA Metro Rail, I use a local instance of Graphhopper, an open source routing engine. I provide it data from OpenStreetMap on all of Southern California, and feed it all combinations of pairs of origins and destinations. This returns a shapefile for each route. I intersect these routes with buffers of various distances around LA Metro Rail lines, and assign to each buffer bin the share of the route that lies within that bin.

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 & 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.[A.3] 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:

i) *O & D contain station*: Both tracts either contain a transit station or have their centroid within 500 meters of a transit station.

ii) *O & D <250m from station*: Both tracts have some part within 250 meters of a transit station, but i) is not true.

iii) *O & D <500m from station*: Both tracts have some part within 500 meters of a transit station, but neither i) nor ii) is true.

Only stations open before the end of 1999 are considered. Appendix Figure H6 uses a single measure treatment variable that determines treatment based on various combinations of maximum distances from centroid and perimeter to station. As expected, under very stringent definitions, the estimated effect is large and positive. As distances increase, the estimated effect drops to zero.

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 & 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 tracks that were never built, (ii) they are near a built track but distant from a station, or (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).

A.3 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:

A.3. Maps are available through the LACMTA library and online at https://www.metro.net/about/library/archives/visions-studies/mass-rapid-transit-concept-maps/.

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 items unobservable to route planning GIS systems that may induce measurement error.

4. CTPP location data are 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 do 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.

A.4 Distance calculations on an idealized geography

To help interpret and contextualize the estimated coefficients on the commuting flow treatment bins in Section 4, I describe below functions of the distances implied by the treatment bins *on an idealized geography*. Real-world census tracts show a wide variety of shapes and vary in density in ways that make such calculations on actual geography challenging. The approach below abstracts away from from both random census tract shaping and variation in density within census tracts.

Suppose all locations on a map are covered by square tracts with an identical area A; we will use the median tract area 1.38km^2 for the calculations below. This corresponds to a square with edges of length $\ell = \sqrt{A} = 1.17\text{km}$. Consider a station at the origin. We can calculate the average distance to a square with vertices (starting in the SW corner of the square and proceeding clockwise) at (a, b), $(a, b + \ell)$, $(a + \ell, b + \ell)$, $(a + \ell, b)$ by using the following expectation:

$$\int_b^{b+\ell} \int_a^{a+\ell} \frac{\sqrt{x^2 + y^2}}{A} dxdy$$

Note that this expectation is valid only under a uniform density in the square. It also uses straight-line distance: In the presence of buildings that obstruct straight-line walking paths, this measure will therefore understate the true distance. Denote by d_ϵ some tiny distance, positive distance to be used as a limit for the calculations below (i.e., it will equal 0m for calculation).

Table A1 maps the treatment bins presented in Section 4 to several descriptions of minimum, maximum, and average distance. The simple measure is *Range*, which is just the minimum and maximum distances from a station to a point in a tract that falls under definition i), ii), or iii). Under definition i), the minimum *Range* is just 0m and the range assumes a station located at a vertex of the tract. The minimum *Range* for definitions ii) and iii) assumes the station is nearest to the midpoint of one side of the tract, at a perpendicular distance of d_ϵ and $250m$ respectively. The maximum *Range* assumes the station is located diagonally away from a vertex of the tract, with a perpendicular distance of $250m$ and $500m$

The *Station at centroid* measure assumes that a station is located at the centroid of the tract. The expectation above is therefore evaluated at the given values of a and b in the table; this corresponds to the average distance from the origin to any point in a square of area A that is centered at the origin. If the station is a centroid, however, no tracts can fall under definition ii) or iii), because the station is $\ell/2 > 500m$ from the nearest edge.

The *Minimum Average Distance* and *Maximum Average Distance* measures instead calculate the average distance from the station if it were as close as possible or as far as possible, respectively, according to the expectation criteria above. For i), that means putting the station in the centroid of the tract or at a vertex of the tract. For ii), it means placing the station just outside the midpoint of an edge of the tract or 250m diagonally away from the vertex of the tract. For iii), it means placing the station 250m perpendicularly away from the midpoint of an edge of the tract or 500m diagonally away from the vertex of the tract. Numerical values of *Station at centroid*, *Minimum Average Distance*, and *Maximum Average Distance* were calculated using Wolfram Alpha's online integral evaluator (note that analytic solutions exist).

Table A1: Distances for different treatment bins under uniform square geography of median area

	Range	Station at centroid	Min. Ave. Dist.	Max. Ave. Dist.
Distances under i)	[0m, 1655m]	444m $a = -\frac{\ell}{2}$ $b = -\frac{\ell}{2}$	444m $a = -\frac{\ell}{2}$ $b = -\frac{\ell}{2}$	888m $a = -d_\epsilon$ $b = -d_\epsilon$
Distances under ii)	(0m, 2008m)	- $a = .$ $b = .$	689m $a = d_\epsilon$ $b = -\frac{\ell}{2}$	888m $a = d_\epsilon$ $b = d_\epsilon$
Distances under iii)	[250m, 2363m)	- $a = .$ $b = .$	901m $a = 250m + d_\epsilon$ $b = -\frac{\ell}{2}$	1219m $a = 250m + d_\epsilon$ $b = 250m + d_\epsilon$

B Proofs and Algebra

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 $\mathcal{T}_{ij}(\boldsymbol{\pi}; \bar{N})$:

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

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 $\boldsymbol{\pi}^{FP}$ as π_{ij}. Rewriting the no spatial arbitrage condition:

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

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: Consider now the set of places with the positive land use for either housing or production, denoted $\mathcal{J}$ (a theorem referenced below requires that the set of possible housing locations be the same as the set of possible production locations). Rearranging the system in Equations (4), (6), (8), (9), and (10) into a more convenient form gives:

$$
\begin{aligned}
W_j^{\frac{1+\epsilon(1-\alpha)}{1-\alpha}} \Psi_j &= \bar{N}^{-1} K_{0j} \sum_{s \in \mathcal{J}} W_s^{\epsilon} \Psi_s \\
\Psi_j &= \sum_{r \in \mathcal{J}} K_{1rj} Q_r^{-\epsilon(1-\zeta)} \\
Q_i^{-\epsilon(1-\zeta)-\frac{1+\psi}{\psi}} \Phi_i &= \bar{N}^{-1} K_{2i} \sum_{s \in \mathcal{J}} W_s^{\epsilon} \Psi_s \\
\Phi_i &= \sum_{r \in \mathcal{J}} K_{1is} W_s^{\epsilon+1}
\end{aligned}
$$

where $K_{0j} = \breve{A}_j^{1/(1-\alpha)}$, $K_{1ij} = \Lambda_{ij}$, 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, Ψ, 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 (11).[B.1]

Proposition 2

Existence: A_i is uniquely determined from:

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

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:

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

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 $M_{ij} = W_j^{\epsilon}\left(Q_i^{1-\zeta}\right)^{-\epsilon}$. 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.[B.2]

B.1. To ensure the algebra is correct, I have numerically verified $\rho(\mathbb{A}^+) \leq 1$ iff Equation (11) holds.

B.2. See Proposition 17.F.3 in Mas-Colell, Whinston, and Green, *Microeconomic Theory* (Oxford University Press, 1995). An alternative approach could be to use weak diagonal dominance of this positive matrix (following Bayer and Timmins (2005) but for weaker conditions).

Rewriting the Model as a Three Linear Equation System

Taking logs of Equations (4), (6), and (8) gives the following cross-sectional system (where lower-case letters represent the log counterparts of level variables):

$$w_j = g_0 + (\alpha - 1)n_j^Y + \ln(A_j) \tag{B-1}$$
$$n_{ij} = g_1 + \epsilon w_j - \epsilon(1-\zeta)q_i - \epsilon\delta_{ij} + \ln(B_i E_j D_{ij}) \tag{B-2}$$
$$q_i = g_2 + \psi h_i + \ln(C_i) \tag{B-3}$$

where $n_j^Y = \ln\left(\bar{N}\sum_r \pi_{rj}/L_j^Y\right)$ is log employment density and $h_i = \ln\left((1-\zeta)\bar{N}\sum_s \pi_{is}W_s/Q_i L_i^H\right)$ is log housing density. The g capture remaining constants: $g_0 = \ln(\alpha)$, $g_1 = \ln(\bar{N}) - \ln\left(\sum_r\sum_s \Lambda_{rs}\left(\delta_{ij}Q_r^{1-\zeta}\right)^{-\epsilon} W_s^\epsilon\right)$, and $g_2 = 0$. Local fundamentals are potentially functions of covariates ($A = A(X)$ and so on) such as transit proximity.

This system can be re-expressed to more clearly represent the supply and demand linkages and better exposit the identification strategy. First, separate the unobservables into time-varying and time-invariant components, so that $\ln(A_{jt}) = \bar{a}_j + a_{jt}$, etc. Under the assumption that land use and travel times are constant, this means making the structural assumptions:

$$\ln(A_{jt} L_{jt}^{Y\,1-\alpha}) = \bar{a}_j + a_{jt} \tag{B-4}$$
$$-\epsilon\delta_{ijt} + \ln(B_{it}E_{jt}D_{ijt}) = \bar{b}_i + b_{it} + \bar{e}_j + e_{jt} + \tilde{d}_{ijt} \tag{B-5}$$
$$\ln(C_{it} L_{it}^{H\,-\psi}) = \bar{c}_i + c_{it} \tag{B-6}$$

where $\tilde{d}_{ijt}$ captures statistical measures of travel costs:[B.3]

$$\tilde{d}_{ijt} = T_{ijt}'\lambda^D + x_{ijt}'\beta + \bar{d}_{ij} + \iota_{s_i s_j t} + d_{ijt}.$$

Altogether, substituting into Equations (B-1) to (B-3) gives:

Labor demand in j:	$w_{jt} = g_{0t} + \tilde{\alpha} n_{jt}^Y + \bar{a}_j + a_{jt}$	(B-7)
Labor supply to j:	$\omega_{jt} = \epsilon w_{jt} + \bar{e}_j + e_{jt}$	(B-8)
Commuting between i and j:	$n_{ijt} = g_{1t} + \omega_{jt} + \theta_{it} + \tilde{d}_{ijt}$	(B-9)
Housing demand in i	$\theta_{it} = \tilde{\zeta} q_{it} + \bar{b}_i + b_{it}$	(B-10)
Housing supply in i:	$q_{it} = g_{2t} + \psi h_{it} + \bar{c}_i + c_{et}$	(B-11)

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.

B.3. That is, $-\epsilon\delta_{ijt} + \ln(D_{ijt}) = T_{ijt}'\lambda^D + x_{ijt}'\beta + \bar{d}_{ij} + \iota_{s_i s_j t} + d_{ijt}$.

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

First, I show that the expression in Equation (16) has an equivalent log-sum representation. Begin by dividing counterfactual and factual expected utilities (from Equation 5):

$$\hat{\bar{U}} = \frac{\mathbb{E}[U'_{ijo}]}{\mathbb{E}[U_{rso}]} = \frac{\Gamma\left(\frac{\epsilon-1}{\epsilon}\right) \cdot \left(\sum_{\{ij\}} \tilde{\Lambda}'_{ij} \left(\delta'_{ij} Q_i'^{1-\zeta}\right)^{-\epsilon} (\tilde{B}'_i W'_j)^{\epsilon}\right)^{1/\epsilon}}{\Gamma\left(\frac{\epsilon-1}{\epsilon}\right) \cdot \left(\sum_{\{rs\}} \tilde{\Lambda}_{rs} \left(\delta_{rs} Q_r^{1-\zeta}\right)^{-\epsilon} (\tilde{B}_r W_s)^{\epsilon}\right)^{1/\epsilon}} = \left(\frac{\sum_{\{ij\}} \tilde{\Lambda}'_{ij} \left(\delta'_{ij} Q_i'^{1-\zeta}\right)^{-\epsilon} (\tilde{B}'_i W'_j)^{\epsilon}}{\sum_{\{rs\}} \tilde{\Lambda}_{rs} \left(\delta_{rs} Q_r^{1-\zeta}\right)^{-\epsilon} (\tilde{B}_r W_s)^{\epsilon}}\right)^{1/\epsilon} \tag{B-12}$$

where $\{ij\} = \{rs\}$ track summation sets. Substituting in Equation (4) for some particular ij into the above twice (once for π'_{ij} and once for π_{ij}) and taking logs gives Equation (16).

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

$$\mathbb{E}[\bar{\mathcal{W}}'] - \mathbb{E}[\bar{\mathcal{W}}] = \frac{1}{\mu} \ln \left(\frac{\sum_{k \in K_1} e^{V'_k}}{\sum_{k \in K_0} e^{V_k}}\right) \tag{B-13}$$

where here μ is the marginal utility of income.[B.4] Let:

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

Taking logs of Equation (B-12) then delivers Equation (B-13). 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., Bayer, Keohane, and Timmins 2009). To map interpretation of the change in consumer welfare between the two frameworks, note:

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

B.4. Thanks to Wei You for noting that (16) and a log-sum expression are interchangeable:

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

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 (16) can be used for welfare evaluation when $\epsilon \in (0, 1]$ and well as $\epsilon > 1$.

C Cost-benefit Calculations

This section details the costs of the subway built by 2000, and additional discussion and details on back-of-the-envelope calculations of the air pollution benefit of transit.

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, rail cars, stations, and similar expenses, and (ii) net operating expenses (operating costs less revenues), and are simply summed:

$$\text{Total Annual Cost} = \text{Operating Subsidy} + \text{Annualized Capital Expenditure}$$

C.1 Annualized Capital Expenditure

Cost information is from a consolidation of capital expenditures on lines built before 2000 from fiscal budgets.[C.1] 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.

However, LA Metro's internal cost of borrowing may not be a suitable social discount rate, and the 30-year horizon may be too short. I provide several alternative definitions: (i) 5% interest over a 50-year horizon, (ii) 5% over an infinite horizon, and (iii) 2.5% over an infinite horizon. For (i) and (ii), the 5% rate is roughly equal to a low-yielding municipal bonds in 2000. For (iii), the 2.5% rate is low, roughly equal to the recent cost of borrowing, and is meant to represent a policy maker that highly values future generation or is uncertain about future discount rates (see Weitzman 1998). Once built, subways typically remain in operation for the long run (perhaps forever).

C.2 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 use the *net* subsidy: operating costs less fare revenue. Operating expenses from 1999 or 2000 are unavailable, so I use operating expenses from 2001 and 2002 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.

C.3 Air Pollution Benefits

Notation in this section is independent of anywhere else in the paper.

There is substantial uncertainty about the precise mapping of non-point sources of pollution to health damages. For example, Anderson (2020) discusses meta-analysis that suggests that many transportation-based pollutants generated on highways (including PM2.5) decay to background

C.1. Source: http://demographia.com/db-rubin-la-transit.pdf.

levels over a short distance. On the other hand, both Chen and Whalley (2012) and Gendron-Carrier et al. (2021) link citywide air pollution levels to subway openings, reflecting either further dispersion horizons or reduced congestion spread out over a dense network of streets. A literature in transportation science also seeks to understand the dispersion of roadway-generated pollutants, often using atmospheric dispersion models.

I provide two methods to roughly calculate potential the air pollution benefits of LA Metro Rail. Any approach must take a stance on whether the effects are transient or expected to persist because the Fundamental Law of Congestion argument discussed in the paper. Both methods below take the view that there is some medium-run persistence of reduced congestion due to transit.

Note, however, that the primary benefits of both methods are in terms of reduced mortality for at-risk (very young and old) populations, because I have been unable to find research that provides all the links required size the effect of the morbidity channel. Additional research in the morbidity effects of transportation-induced reductions in air pollution and health is needed, especially for the prime-age population.

Infant Mortality and Subway Openings

Subway openings in polluted cities reduce PM10 by 3.2 μg/m^3 (Gendron-Carrier et al. 2021), who also review the literature and show a roughly linear relationship of 10 infant deaths per 100,000 births per μg/m^3 of PM10. Given the number of births in Los Angeles County in 2000, this corresponds to a short-run effect of 50.4 fewer infant deaths annually. In Section 3.3, I find a long-run effect of LA Metro Rail on congestion to be about one-third of the short-run effect estimated in Anderson (2014), so I discount this effect by two-thirds. Finally, I use a standard value of statistical life of 6,000,000 to translate avoided mortality to a dollar measure.
Formula & Details:

$$(A \times B \times C) \times \delta \times V_L$$

- A: Subway opening reduces PM10 by 3.2 μg/m^3 (Gendron-Carrier et al. 2021)
- B: 10 infant deaths per 100,000 birth per μg/m^3 of PM10 (Gendron-Carrier et al. 2021)
- C: Number of births in Los Angeles County in 2000: 157,508 (US Natality Files)
- δ: Long-run discounting: 1/3 (based on congestion results in Section 3.3)
- V_L: Statistical present value of life at birth: $6,000,000

Medicare-Eligible Mortality and Subway Ridership

By 2000, LA Metro Rail had roughly 150,000 weekday boardings, or about 1.1% of Los Angeles County's total daily trips (assuming 3.8 trips per household per day from Gendron-Carrier et al. (2021), 2.62 people per household, and 9,500,000 residents in Los Angeles County). Assume that all these trips would have otherwise taken place by automobile (rather than bus) and that there is no induced demand. Given that about 55% percent of Los Angeles' PM2.5 levels are generated from roadway traffic, and that average PM2.5 was 20 μg/m^3 in 2000, this corresponds to a 0.12 μg/m^3 reduction in PM2.5. From Deryugina et al. (2019), a 1 μg/m^3 increase in PM2.5 destroys

2.99 life-years per million Medicare recipients per day (measured over a three-day period). Scaling that up by the number of days in a year and a value per life-year saved of $100,000, and then finally by the local Medicare-eligible population (approximate by those aged 65 or older), the annual benefit is about $13 million for older Los Angeles County residents.

Formula & Details:

$$\frac{L_W}{\frac{D}{H} \times P_{LA}} \times R_{LA} \times M \times Y_L \times \frac{365 \text{ days}}{\text{year}} \times V_{LY} \times (p_{65+} \times P_{LA})$$

D: Household trips per day in 2009: 3.8 (Gendron-Carrier et al. 2021).

H: People per household in 2000 in US: 2.62.

L_W: Weekday LA Metro Rail trips in 2000: 150,000

P_{LA}: Population of Los Angeles County in 2000: 9,500,000

R_{LA}: Fraction of overall PM2.5 emission from roadways in LA: 55% (HEI Panel on the Health Effects of Traffic-Related Air Pollution 2010).

M: Los Angeles air pollution (PM2.5) in 2000: 20 μg/m^3 (*Current Air Quality and Trends in the South Coast Air Quality Management District* 2000)

Y_L: Life-years saved per μg/m^3 reduction in PM2.5 per million medicare recipients per day: 2.99 (Deryugina et al. 2019).

V_{LY}: Value per life-year saved: $100,000 (Deryugina et al. 2019).

p_{65+}: Share of population 65 and older in Los Angeles in 2000: 10%

D Model Extensions and Alternative Identification

D.1 Additional Identification Methods and Results

This section extends the approach of interacting labor demand shocks with geography to identify the remaining (housing and labor demand) elasticities. I also discuss two modifications to the standard identification framework: (i) endogenous land use determination (no zoning), and (ii) the presence of agglomeration and other forces.

Residents of one location commute to many different locations for work. Workers who live in i and work in j are sensitive to the housing demands of workers who work in j' but also live in i. A labor demand 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 exclude own tract j:

$$\Delta z^{HS}_{i(-j)t}(\rho) = \sum_{s \neq j} \frac{e^{-\rho\vartheta_{is}} 1_{\check{n}_{is}>0} \Delta z_{st}}{\sum_{s \neq j} e^{-\rho\vartheta_{is}} 1_{\check{n}_{is}>0}}$$

Note that place of work-by-year fixed effects (ω_{jt}) control for changes in workplace amenities. The following moment condition identifies $\tilde{\zeta} = \epsilon(1-\zeta)$:

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

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 express more variation than individual tracts, this approach can identify the housing demand elasticity.

Finally, workers employed at 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 shock 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:

$$\Delta z^{LS}_{jt}(\rho) = \sum_{r} \left(\frac{e^{-\rho\vartheta_{rj}} 1_{\check{n}_{rj}>0}}{\sum_r e^{-\rho\vartheta_{rj}} 1_{\check{n}_{rj}>0}} \sum_{s \neq j} \frac{e^{-\rho\vartheta_{sr}} 1_{\check{n}_{is}>0} \Delta z_{st}}{\sum_{s \neq j} e^{-\rho\vartheta_{sr}} 1_{\check{n}_{is}>0}} \right)$$

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

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

This identifies the labor demand elasticity, $\tilde{\alpha} = \alpha - 1$, and provides an alternative way to estimate this parameter that is conceptually similar to the 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 made more transparent. The following reframe M-1 through M-4 in terms of a labor demand shock (note A-1 is identical to M-1):

$$\mathbb{E}[\Delta z_{jt} \times (\Delta e_{jt} + \Delta d_{ijt})] = 0, \ \forall \, ij \tag{A-1}$$
$$\mathbb{E}[\Delta z_{jt} \times \Delta c_{it}] = 0, \ \forall \, ij \tag{A-2}$$
$$\mathbb{E}[\Delta z_{j't} \times (\Delta b_{it} + \Delta d_{ijt})] = 0, \ \forall \, ij' \neq ij \tag{A-3}$$
$$\mathbb{E}[\Delta z_{j't} \times \Delta a_{jt}] = 0, \ \forall \, j' \neq j \tag{A-4}$$

Proposition 3. *Assume A-1, A-2, A-3, and A-4 are true,* $\rho > 0$*,* $\mathbb{E}[\Delta z_{jt} \times \Delta w_{jt}] \neq 0$*, housing demand is downward sloping, and labor and housing supply are upward sloping. Then M-1, M-2, M-3, and M-4 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. □

It is difficult to estimate household expenditure shares or labor demand elasticities in urban models that use aggregated data (e.g., Diamond 2016). Table D1 uses the employment variant of $\Delta z^{HS}_{i(-j)t}$ to instrument for housing prices to determine $\epsilon(1-\zeta)$, the elasticity of housing demand. The own tract can be excluded from the regression to limit concerns about the labor demand shock driving confounding changes in amenities. Results are significant and vary between -1.00 and -0.78. With $\epsilon = 2.18$, these imply a housing expenditure share between 36% and 45% of income, somewhat higher than microdata suggest but not unreasonable for high-cost areas.

Table D1: IV estimates of housing demand elasticity ($-\epsilon(1-\zeta)$)

	$\Delta n_{ijt} = \Delta \ln(N_{ijt})$		
	(1)	(2)	(3)
A. IV Estimates of $-\epsilon(1-\zeta)$			
$\Delta \ln(\text{House Value})$	-0.782**	-0.778**	-1.008***
	(0.378)	(0.377)	(0.382)
B. First Stage			
$\Delta z^{HS}_{i(-j)t}(\rho)$	1.139***	1.140***	1.135***
	(0.074)	(0.074)	(0.074)
F-stat (CD)	363.2	364.0	356.5
F-stat (KP)	239.7	240.0	236.5
N	143593	143593	141188
Sample	All	All	not ii
Travel Time	-	Y	-

Panel instrument variable (IV) estimates of regression of flows on median housing values, using $\ln(\rho) = -5.5$. Estimated in differences using employment instrument. CD and KP refer to the Cragg-Donald and Kleibergen-Paap tests, respectively. Variables are trimmed to exclude extreme values (see text). All estimates include tract-of-work-by-year and tract-pair fixed effects. Standard errors clustered by tract in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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 D2 vary substantially. Column 2 of Table D2 includes the own-tract demand shock, Δz_{jt}, as a control (recall that the instrument is $\Delta z^{LS}_{jt}(\rho)$). This permits limited spatial correlation (to the extent the observed labor demand shocks are spatially correlated), and implies a slightly higher labor share of income. The measure may not be unreasonable, particularly given the power of the first stage results (the magnitude is less interpretable because of repeated aggregation across origins and destinations). Column 3 includes the log measure of land zoned for productive uses, but this is measured poorly in the data.[D.1] Similarly, Column 4 indicates too large estimates.

D.1. Unlike residential land, it is difficult to classify different types of land used in production. For example, it is unclear whether to add land used for storage. Further, the data show some unusual changes across waves.

Table D2: IV estimates of inverse labor demand elasticity $(\alpha - 1)$

	$\Delta w_{jt} = \Delta \ln(W_{jt})$			
	(1)	(2)	(3)	(4)
A. IV Estimates of $(\alpha - 1)$				
$\Delta \ln(\text{Employment})$	-0.679*	-0.346*	-1.389	
	(0.382)	(0.179)	(1.595)	
$\Delta \ln(\text{Prod. Land})$			1.806	
			(2.050)	
$\Delta \ln(\text{Emp. Density})$				-0.999
				(0.888)
B. Specification Test $(H_0 : \psi_{\text{Employment}}/\psi_{\text{Prod. Land}} = -1)$				
$\psi_{\text{Employment}}/\psi_{\text{Prod. Land}}$			-0.769	
$[\cdot] = \Pr(H_0)$			[0.004]	
C. First Stage				
$\Delta z^{LS}_{jt}(\rho)$	-15.205**	-22.181***	-6.625	-8.878
	(7.367)	(7.747)	(7.242)	(7.245)
F-stat (CD)	5.4	10.6	1.0	1.9
F-stat (KP)	4.3	8.2	0.8	1.5
N	2442	2442	2385	2385
Own shock as control	-	Y	Y	Y

Panel instrument variable (IV) estimates of regression of employment, employment density and land in production, using $\ln(\rho) = -5.5$. CD and KP refer to the Cragg-Donald and Kleibergen-Paap tests, respectively. Variables are trimmed to exclude extreme values (see text). Columns 2-4 include the own tract labor demand shock as a control. Standard errors clustered by tract in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Reasonable estimates of $\tilde{\alpha}$ and $\tilde{\zeta}$ provide confidence in this interconnected approach to identification and serve as an informal test of overidentification. 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, omnipurpose tools for identifying local price elasticities.

D.2 Extension with externalities

I also consider a model extension with production and residential externalities. I model these as in Ahlfeldt et al. (2015), who define production and residential externalities as an inverse-distance weighted sum of employment and residential population, respectively. This recasts endogenous measures of productivity and amenities in terms of an exogenous fundamental, the observable distribution of population, and four new parameters. Specifically,

$$A_j = \mathcal{A}_j \Upsilon_j^{\mu}, \qquad \Upsilon_j = \sum_n \frac{e^{\kappa_A d_{jn}} \sum_r \pi_{rn}}{L_j^Y}$$
$$B_i = \mathcal{B}_i \Omega_i^{\eta}, \qquad \Omega_i = \sum_n \frac{e^{\kappa_B d_{ni}} \sum_s \pi_{ns}}{L_i^H}$$

where I use parameters from Ahlfeldt et al. (2015) as needed.

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:

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

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:

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

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

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

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 Ω 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 $\Delta\Omega$ and $\Delta\Upsilon$ are captured by time-invariant tract fixed effects.

D.3 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 imply $Q_i H_i = P_i^L L^H + P^M M$, and the first-order conditions deliver an expression for M under profit maximization. This results in the expression:

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

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. This is higher than 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 $L_i^H + L_i^Y = \bar{L}_i$.[D.2] 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.[D.3]

D.2. Note that this implies $\Delta L_{it}^Y = -\Delta L_{it}^H$.

D.3. 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.

D.4 Agglomeration and Endogenous Land Use

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

D.5 Specification Tests Using Workplace Wage

I provide a test of the identifying assumption of ϵ in Ahlfeldt et al. (2015) in Panel C of Table 3 of the main paper. Using ρ in this subsection to denote $\mathbb{E}[w_j e_j]/\mathbb{V}[e_j]$, I estimate by OLS:

$$w_j = \rho \hat{e}_j + u_j \tag{1}$$

using year 2000 data and $\hat{e}_j = \hat{\omega}_j - \hat{\epsilon} w_j$, where $\hat{\omega}$ and $\hat{\epsilon}$ correspond to the estimation method and results presented in Panels A and B of Table 3. Identification in Ahlfeldt et al. (2015) requires a weakly negative correlation, which is partially supported by my results.

To test whether $\mathbb{E}[w_j e_j]/\mathbb{V}[e_j] = -1/2\epsilon$, I offer two approaches. The first ignores uncertainty around estimates ϵ, recognizing that my IV estimates are somewhat imprecise and that other approaches may give much more precise parameter estimates. The p-value is simply a test of the linear restriction $\rho = -1/2\epsilon$. That the null is rejected in most columns indicates that $\mathbb{E}[w_j e_j]/\mathbb{V}[e_j] = -1/2\epsilon$ is unlikely to hold precisely.

The second approaches estimates a system by GMM. First, I estimate $\hat{e}_j = \hat{\omega}_j - \hat{\epsilon} w_j$ enforcing the IV point estimates of ϵ in Panel A of Table 3. Then I jointly estimate

$$\Delta \hat{\omega}_{jt} = \epsilon \Delta w_{jt} + \Delta e_{jt} \tag{2}$$

$$w_j = \rho \hat{e}_j + u_j \tag{3}$$

via system GMM, where I assume:

$$\mathbb{E}[\Delta z_{jt} \times \Delta e_{jt}] = 0 \tag{4}$$

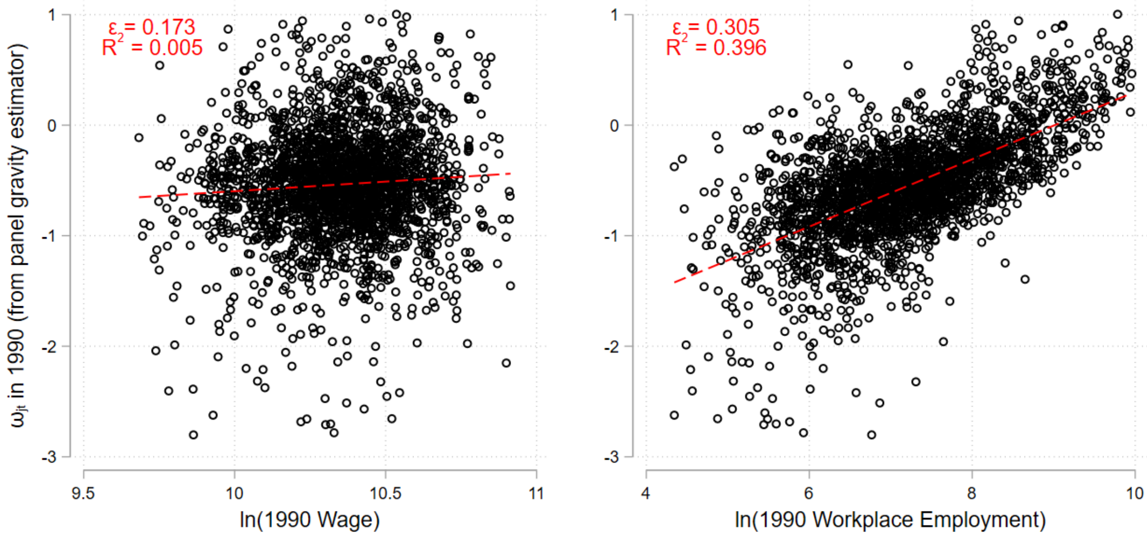
$$\mathbb{E}[\hat{e}_j \times u_j] = 0 \tag{5}$$

and test the non-linear restriction that: $\rho = -1/2\epsilon$. This test has less power, and accordingly fails to reject the null in four out of the six columns presented in Table 3. Interestingly, the null is rejected in precisely the cases where correlation between w_j and e_j is weakly negative or even positive, according with the statistical argument in Section 5 of the main paper.

I also provide a comparison of ω and w in order to maintain comparability with Kreindler and Miyauchi (2020). Panel A of Figure D1 plots $\hat{\omega}_j$ as recored by the destination fixed effects (and which is equal to $\epsilon w_j + e_j$) against w_j using data from Greater Los Angeles in 1990. If $\mathbb{E}[w_j e_j] = 0$, then wages and labor supply are not simultaneously determined, and then the slope in Panel A of Figure D1 (0.17) is equal to ϵ. If $\mathbb{E}[e_j^2] = 0$, this figure would reveal a one-to-one mapping between ω_j and w_j. Instead, ω_j is more closely related to workplace employment levels (Panel B of Figure D1), highlighting the severity of simultaneity when wage is unobserved.

Importantly, note the difference in scale between the variation in ω and in w, and the relatively low R-squared. This suggests that, non-wage workplace amenities play a first-order role in determining the distribution of workplace population.

Figure D1: Does $\omega = \epsilon w$, and if not, what is it capturing?



E Counterfactual Estimation & Bootstrap Procedure

E.1 Counterfactual Estimation

First, note that the following hold:

$$\hat{W}_i = \hat{A}_i \hat{\bar{N}}^{\alpha-1} \left(\frac{\sum_r \pi_{ri} \hat{\pi}_{ri}}{\sum_r \pi_{ri}} \right)^{\alpha-1} \tag{E1}$$

$$\hat{Q}_i = \hat{C}_i^{1/(1+\psi)} \left(\frac{\hat{\bar{N}} \sum_s \pi_{is} \hat{\pi}_{is} W_s \hat{W}_s}{\sum_s \pi_{is} W_s} \right)^{\psi/(1+\psi)} \tag{E2}$$

$$\hat{\pi}_{ij} = \frac{\hat{B}_i \hat{E}_j \hat{D}_{ij} \hat{W}_j^{\epsilon} \hat{Q}_i^{-\epsilon(1-\zeta)}}{\sum_r \sum_s \pi_{rs} \hat{B}_r \hat{E}_s \hat{D}_{rs} \hat{W}_s^{\epsilon} \hat{Q}_r^{-\epsilon(1-\zeta)}} \tag{E3}$$

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

An equilibrium is a fixed point in wages, housing prices, and commuting flows (in a closed economy). I use the algorithm below with an adaptive updating weight to find a new equilibrium after a shock. I first simulate the closed economy counterfactual, then use that solution as the initial for the open economy counterfactual, if required:

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. Set the initial updating weight $\xi^{(0)} \in (0, 1)$ (typically I use 0.8).
2. Calculate $\{\hat{\pi}_{ij}^{(0)}\}$ using $\{\hat{W}_i^{(0)}\}$, $\{\hat{Q}_i^{(0)}\}$, and $\{\pi_{ij}\}$.
3. Main Loop:
   (a) Calculate $\{\hat{Q}_i^{(temp)}\}$ using $\{\hat{W}_i^{(t-1)}\}$, $\{W_i\}$, $\{\hat{\pi}_{ij}^{(t-1)}\}$, and $\{\pi_{ij}\}$.
   (b) Calculate $\{\hat{W}_i^{(temp)}\}$ using $\{\hat{\pi}_{ij}^{(t-1)}\}$, and $\{\pi_{ij}\}$.
   (c) Calculate $\{\hat{\pi}_{ij}^{(temp)}\}$ using $\{\hat{W}_i^{(t)}\}$, $\{\hat{Q}_i^{(t)}\}$, and $\{\pi_{ij}\}$.

(d) Update $\hat{X}^{(t)} = \xi^{(t)}\hat{X}^{(temp)} + (1-\xi^{(t)})\hat{X}^{(t-1)}$ for $\hat{X} \in \{\hat{Q}, \hat{W}, \hat{\pi}\}$, where ξ is a weight that disciplines updating.

(e) Calculate movement as (with $\mathcal{N}$ the number of pairwise observations):

$$\Delta^{(t)} = \frac{1}{\mathcal{N}} \sum_r \sum_s |\hat{\pi}_{rs}^{(t)} - \hat{\pi}_{rs}^{(t-1)}|.$$

(f) If $\Delta^{(t)} \geq \Delta^{(t-1)}$, set a new $\xi^{(t+1)} < \xi^{(t)}$.

(g) Stop when movement is below convergence criterion.

4. Initial guess for $\hat{\bar{N}}^{(0)}$ using $\{\hat{W}_i^{(temp)}\}$, $\{W_i\}$, $\{\hat{Q}_i^{(temp)}\}$, $\{\hat{\pi}_{ij}^{(temp)}\}$, and $\{\pi_{ij}\}$, resetting $\xi^{(0)}$.

5. Main Loop:

(a) Calculate $\{\hat{Q}_i^{(temp)}\}$ using $\hat{\bar{N}}^{(t-1)}$, $\{\hat{W}_i^{(t-1)}\}$, $\{W_i\}$, $\{\hat{\pi}_{ij}^{(t-1)}\}$, and $\{\pi_{ij}\}$.

(b) Calculate $\{\hat{W}_i^{(temp)}\}$ using $\hat{\bar{N}}^{(t-1)}$, $\{\hat{\pi}_{ij}^{(t-1)}\}$, and $\{\pi_{ij}\}$.

(c) Calculate $\{\hat{\pi}_{ij}^{(temp)}\}$ using $\{\hat{W}_i^{(t)}\}$, $\{\hat{Q}_i^{(t)}\}$, and $\{\pi_{ij}\}$.

(d) Calculate $\hat{\bar{N}}^{(temp)}$ using $\{\hat{W}_i^{(t)}\}$, $\{W_i\}$, $\{\hat{Q}_i^{(t)}\}$, $\{\hat{\pi}_{ij}^{(t)}\}$, and $\{\pi_{ij}\}$.

(e) Update $\hat{X}^{(t)} = \xi^{(t)}\hat{X}^{(temp)} + (1-\xi^{(t)})\hat{X}^{(t-1)}$ for $\hat{X} \in \{\hat{Q}, \hat{W}, \hat{\pi}, \hat{\bar{N}}\}$, where ξ is a weight that disciplines updating.

(f) Calculate movement as:

$$\Delta = \frac{1}{\mathcal{N}} \sum_r \sum_s |\hat{\pi}_{rs}^{(t)} - \hat{\pi}_{rs}^{(t-1)}|.$$

(g) If $\Delta^{(t)} \geq \Delta^{(t-1)}$, set a new $\xi^{(t+1)} < \xi^{(t)}$.

(h) Stop when movement is below convergence criterion.

With Agglomeration

When modeling agglomeration, I use the following system:

$$\hat{\Upsilon}_i = \frac{\sum_n e^{\delta_A d_{in}} \sum_r \pi_{rn}\hat{\pi}_{rn}}{\sum_n e^{\delta_A d_{in}} \sum_r \pi_{rn}}$$

$$\hat{\Omega}_i = \frac{\sum_n e^{\delta_B d_{ni}} \sum_s \pi_{ns}\hat{\pi}_{ns}}{\sum_n e^{\delta_B d_{ni}} \sum_s \pi_{ns}}$$

$$\hat{W}_i = \hat{A}_i \hat{\Upsilon}_i^{\mu} \hat{\bar{N}}^{\alpha-1} \left(\frac{\sum_r \pi_{ri}\hat{\pi}_{ri}}{\sum_r \pi_{ri}} \right)^{\alpha-1}$$

$$\hat{Q}_i = \hat{C}_i^{1/(1+\psi)} \left(\frac{\hat{\bar{N}} \sum_s \pi_{is}\hat{\pi}_{is} W_s \hat{W}_s}{\sum_s \pi_{is} W_s} \right)^{\psi/(1+\psi)}$$

$$\hat{\pi}_{ij} = \frac{\hat{B}_i \hat{E}_j \hat{D}_{ij} \hat{W}_j^{\epsilon} \hat{Q}_i^{-\epsilon(1-\zeta)} \hat{\Omega}_i^{\eta}}{\sum_r \sum_s \pi_{rs} \hat{B}_r \hat{E}_s \hat{D}_{rs} \hat{W}_s^{\epsilon} \hat{Q}_r^{-\epsilon(1-\zeta)} \hat{\Omega}_r^{\eta}}$$

E.2 Additional Counterfactual Exercises

Appendix Table H13 presents a variety of additional model results. I describe them here. The table contains three columns, reflecting partial equilibrium, general equilibrium, and general equilibrium with spillovers. The general equilibrium with and without spillovers are as presented in the main text and in these appendices.

The partial equilibrium results are calculated in a different manner. Rather than being the result of a fixed point algorithm, these are akin to only initializing the first round of fixed point process. That is, for Equations (E1)–(E2), let wages and housing prices reflect just the changes in fundamentals:

$$\hat{A} \rightarrow \hat{W}$$
$$\hat{C} \rightarrow \hat{Q}$$

Next, feed the changes in $\hat{B}$, $\hat{D}$, and $\hat{E}$ into the following variant of Equation (E3), which ignores the price changes:

$$\hat{\pi}_{ij} = \frac{\hat{B}_i \hat{E}_j \hat{D}_{ij}}{\sum_r \sum_s \pi_{rs} \hat{B}_r \hat{E}_s \hat{D}_{rs}}$$

Then, feed the updated prices and $\hat{\pi}$ vector into the welfare formula:

$$\ln \hat{\bar{U}} = \frac{1}{\epsilon} \ln \left(\frac{\hat{B}_i \hat{E}_j \hat{D}_{ij} \hat{W}_j^{\epsilon} \hat{Q}_i^{-\epsilon(1-\zeta)}}{\hat{\pi}_{ij}} \right)$$

Thus, this partial equilibrium effect accounts for mobility induced changes to B, D, and E, and changes to utility induced by changes in wage or housing prices (only because of changes in A and C), but does not reflect mobility induced by changes in prices or changes in prices induced by mobility.

The first panel of Appendix Table H13 corresponds to the results presented in the main text. The second panel assumes a $\hat{C}$ just in tracts that contain a transit stations such that the partial equilibrium effect results a 10% higher residential population in those tracts (allowing $\hat{C} \rightarrow \hat{Q} \rightarrow \hat{\pi}$). The third panel assumes that $\lambda^B = 0.05$ using the 500m measure of treatment proximity. The fourth panel assumes that $\lambda^A = 0.04$ using the 500m measure of treatment proximity.

Combining historic and future counterfactuals

All of the main counterfactual simulations presented in Table 6 except the dynamic 2015 effects are *historic* in nature: by how much would the region be worse off if LA Metro Rail were removed, relative to observed outcomes in 2000. The dynamic 2015 effect takes that as given and instead asks additionally about the *future*: by how much is the region better off assuming commuting growth through 2015, relative to the observed data in 2000 (which includes the transit system as built). The bottom three panels of Appendix Table H13 all assume that the additional effects are *future* effects, taking place after 2000.

E.3 Bootstrapping Procedure

To my knowledge, no off-the-shelf bootstrapping procedure works well for developing joint bootstrap estimates of the parameters of the model presented in Equations (12)–(14). A central challenge is the need to preserve the correlation structure of parameters estimated across these equations. This, in turn, is complicated by the use of both IV to estimate some parameters, and the possible presence of dyadic correlation structures in the estimation of other parameters. This multiple-cluster case suffers from a potential degeneracy issue that renders standard bootstrapping approaches for dyadic data, such as the pigeonhole bootstrap (Owen 2007), as overly conservative (Davezies, D'Haultfœuille, and Guyonvarch 2019; Menzel 2020). Fortunately, recent research offers bootstrapping procedures that produce more accurate confidence intervals for IV and for dyadic data under some additional assumptions (Davidson and MacKinnon 2010; Davezies, D'Haultfœuille, and Guyonvarch 2019; Menzel 2020). For an overview of inference under multiple clusters and networks more generally, see Graham (2020).

I therefore describe a hybrid wild bootstrap procedure that combines recent results on bootstrapping under multi-way error structures by Menzel (2020) and the wild restricted efficient residual bootstrap for IV by Davidson and MacKinnon (2010). There are two key features of this approach. Foremost and *essentially*, the same sets of bootstrapping weights are used across equations, so this procedure captures the correlation of estimates across equations. This is vital for correct model inference. Second, the approach used for each equation is relatively efficient.[E.1]

A central challenge is bootstrapping the dyadic estimating equation in a way that reflects the three-way cluster error structure implemented in the paper. Experimentation revealed that both the pigeonhole bootstrap (Owen 2007) and the related method presented in Davezies, D'Haultfœuille, and Guyonvarch (2019) are much more conservative than the three-way cluster. There is likely a good reason: the bootstrap can fail in degenerate settings, such as if errors are approximately iid (Davezies, D'Haultfœuille, and Guyonvarch 2019; Graham 2020; Menzel 2020). Such degeneracy is quite likely conditional on rich sets of fixed effects (Menzel 2020).

Modern techniques rely on the correspondence of dyadic data to exchangeable arrays (Davezies, D'Haultfœuille, and Guyonvarch 2019). Menzel (2020) suggests a solution in the degenerate case.[E.2] In the particular case of an additive regression model with homoskedastic errors (which is not unlikely conditional on fixed effects), a double differencing approach has excellent convergence properties, but has a non-standard limiting distribution. However, he develops a bootstrap procedure that approximates the limiting distribution. This is a form of wild bootstrap that uses a convolution of two Gamma-distributed random variables to form bootstrap residuals. Each weight corresponds to 'one dimension' of the dyadic data, and so can naturally be used as the wild bootstrap weight for non-dyadic estimating equations. Those are IV estimators, so I turn to the wild restricted efficient residual bootstrap by Davidson and MacKinnon (2010), but use the same Gamma-distributed weights as used for the dyadic data.

In order fit my research design into the double-differencing framework of Menzel (2020), I first take differences of Equation (2); this implicitly controls for autocorrelation with origin-destination pair (see Davezies, D'Haultfœuille, and Guyonvarch 2019, for a similar first step). Let $\Delta\ddot{x}_{ijt}$ denote the double-differenced value of Δx_{ijt} with respect to i and j, where $\Delta x_{ijt} = x_{ij,t=1} - x_{ij,t=0}$. That

E.1. Note that the procedure developed here allows for correlation within $\{i, j\}$ pairs over time, across pairs $\{i, j\}$ and $\{i, j'\}$ for $j' \neq j$, and across pairs $\{i, j\}$ and $\{i', j\}$ for $i' \neq i$, but does not allow correlation between $\{i, j\}$ and $\{j', i\}$, even for $j = j'$. How to handle this case is an active area of research; see Graham (2020) for more discussion.

E.2. Menzel (2020) also offers a more sophisticated adaptive bootstrap procedure.

is,

$$\Delta\ddot{x}_{ijt} = \Delta x_{ijt} - \bar{\Delta x}_{iJt} - \bar{\Delta x}_{Ijt} + \bar{\Delta x}_{IJt}$$

where $\bar{\Delta x}_{iJt} = J^{-1}\sum_j \Delta x_{ijt}$, $\bar{\Delta x}_{Ijt} = I^{-1}\sum_i \Delta x_{ijt}$, and $\bar{\Delta x}_{IJt} = (I \times J)^{-1}\sum_i\sum_j \Delta x_{ijt}$ are one- and two-way averages.

The estimating equations for the bootstrap are:

$$\Delta\ddot{n}_{ijt} = \Delta\ddot{T}_{ijt}\lambda + \Delta\ddot{d}_{ijt} \tag{D1}$$

$$\Delta\Omega_{jt} = \epsilon\Delta w_{jt} + \Delta e_{jt} \tag{D2.IV}$$

$$\Delta w_{jt} = \pi_\epsilon \Delta z_{jt} + \Delta u_{\epsilon,jt} \tag{D2.1st}$$

$$\Delta q_{it} = \psi\Delta h_{it} + \Delta c_{it} \tag{D3.IV}$$

$$\Delta h_{it} = \pi_\psi \Delta z_{it}^{HD,a}(\rho) + \Delta u_{\psi,it} \tag{D3.1st}$$

where I've omitted writing covariates or subcounty and subcounty-pair fixed effects for parsimony (but include them in estimation), let $\lambda = \lambda^D$, and assumed that Ω_{jt} is fixed. The bootstrapping procedure below applies the same set of wild bootstrap residual weights to create estimates using the method of Menzel (2020) for Equation (D1) and Davidson and MacKinnon (2010) for the two sets of IV equations in (D2) and (D3).

Note that Equations (D2) and (D3) are precisely identical to those presented in the paper. Equation (D1) is estimated in a slightly different way than that in the paper, in that it does not use iterative demeaning (instead employing the double difference). Therefore, point estimates are slightly different. These are presented in Table E1. While I provide results from the full and PER samples for comparison, I rely on the Immediate 1925 Plan Sample for the bootstrap procedure. Because the coefficient on "O & D <500m from station" is close to zero and insignificant, I do not use it in welfare simulations (instead setting it equal to 0).

Table E1: Effect of Transit on Commuting Flows by 2000 (time differenced then double differenced)

	(1)	(2)	(3)
O & D contain station	0.088**	0.098**	0.134**
	(0.039)	(0.045)	(0.058)
O & D <250m from station	0.048	0.079	0.115*
	(0.047)	(0.051)	(0.061)
O & D <500m from station	-0.022	0.009	0.004
	(0.036)	(0.039)	(0.049)
N	145555	49740	9611
Control Group	All	PER	Immed. '25 Plan
Standard Three-Way FEs	Y	Y	Y
Subcounty Pair-×-Year FEs	Y	Y	Y
Highway Controls	Y	Y	Y

Standard errors clustered by tract pair, tract of residence, and tract of work in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Bootstrap Routine

The bootstrap implemented in the paper takes as fixed ζ and α, assigns λ^D under distance iii) (O&D<500m from station) equal to zero, assigns $\lambda^A = \lambda^B = 0$ consistent with empirical results, and maintains the assumption that $\lambda^C = \lambda^E = 0$. The bootstrapped input parameters are ϵ, ψ, and $\lambda^{D'}$ under distance i) and ii). I use $\mathcal{B} = 400$ wild bootstrap draws.

1. Separately estimate Equations D1, D2, and D3 as above, using IV for Equations D2 and D3. Recover parameter estimates $\tilde{\lambda}$, $\tilde{\epsilon}$, and $\tilde{\psi}$; as well as residual estimates $\Delta\hat{\ddot{d}}_{ijt}$.

2. Estimate the equations below using OLS while enforcing the identified coefficients $\tilde{\epsilon}$ and $\tilde{\psi}$, then recover the residuals $\Delta\hat{e}_{jt}$ and $\Delta\hat{c}_{it}$:

$$\Delta\Omega_{jt} - \tilde{\epsilon}\Delta w_{jt} = \text{covariates} + \Delta e_{jt}$$
$$\Delta q_{it} - \tilde{\psi}\Delta h_{it} = \text{covariates} + \Delta c_{it}$$

3. Because the second stage residuals in Equations D2.1st and D3.1st do not reflect correlation with $\Delta\hat{e}_{jt}$ and $\Delta\hat{c}_{it}$, using them to create the bootstrap sample is inefficient. Instead, separately estimate:

$$\Delta w_{jt} = \pi_\epsilon \Delta z_{jt} + \varpi_\epsilon \Delta\hat{e}_{jt} + \Delta o_{\epsilon,jt}$$
$$\Delta h_{it} = \pi_\psi \Delta z_{it}^{HD,a}(\rho) + \varpi_\psi \Delta\hat{c}_{it} + \Delta o_{\psi,it}$$

   and then save the estimates $\hat{\pi}_\epsilon$ and $\hat{\pi}_\psi$, then generate first stage residuals incorporating the predicted contribution of the second stage correlation to the estimated residuals as follows:

$$\Delta\hat{u}_{\epsilon,jt} \equiv \hat{\varpi}_\epsilon \Delta\hat{e}_{jt} + \Delta\hat{o}_{\epsilon,jt}$$
$$\Delta\hat{u}_{\psi,it} \equiv \hat{\varpi}_\psi \Delta\hat{c}_{it} + \Delta\hat{o}_{\psi,it}$$

4. Draw $\tilde{\mathcal{B}} = \mathcal{B} * 1.025$ sets of $2\mathcal{N}$ iid random variables from a Gamma distribution with shape parameter 4 and scale parameter 1/2, and subtract 2 from these values (such that these variables are mean 0, variance 1, and skewness 1). For each set of random variables $b \in 1, \ldots, \tilde{\mathcal{B}}$, index elements in the first half by $i \in 1, \ldots, \mathcal{N}$ and in the second half by $j \in 1, \ldots, \mathcal{N}$, so that, e.g., ω_{b_i} is the bootstrap weight for correspond to location i in bootstrap sample b.

5. Generate bootstrap samples and recover bootstrap estimates. That is, for each b:

   (a) Create bootstrap residual values as follows:

$$\Delta\hat{\ddot{d}}^*_{ijt} \equiv \omega_{b_i} \cdot \omega_{b_j} \cdot \Delta\hat{\ddot{d}}_{ijt}$$
$$\Delta\hat{e}^*_{jt} \equiv k^*_\epsilon \cdot \omega_{b_j} \cdot \Delta\hat{e}_{jt}$$
$$\Delta\hat{u}^*_{\epsilon,jt} \equiv l^*_\epsilon \cdot \omega_{b_j} \cdot \Delta\hat{u}_{\epsilon,jt}$$
$$\Delta\hat{c}^*_{it} \equiv k^*_\psi \cdot \omega_{b_i} \cdot \Delta\hat{c}_{it}$$
$$\Delta\hat{u}^*_{\psi,jt} \equiv k^*_\psi \cdot \omega_{b_i} \cdot \Delta\hat{u}_{\psi,jt}$$

where $k^* = (\mathcal{N}/(\mathcal{N}-k))^{1/2}$ and $l^* = (\mathcal{N}/(\mathcal{N}-l))^{1/2}$ are scale adjustments wherein the denominators are the residual degrees of freedom for the uninstrumented and instrumented equations, respectively.

(b) Create the bootstrap sample as follows, where Equations SMP.3 and SMP.5 are respectively defined before SMP.2 and SMP.4:

$$\Delta\hat{\ddot{n}}^*_{ijt} \equiv \Delta\ddot{T}_{ijt}\tilde{\lambda} + \Delta\hat{\ddot{d}}^*_{ijt} \tag{SMP.1}$$

$$\Delta\hat{\Omega}^*_{jt} \equiv \tilde{\epsilon}\Delta\hat{w}^*_{jt} + \Delta\hat{e}^*_{jt} \tag{SMP.2}$$

$$\Delta\hat{w}^*_{jt} \equiv \hat{\pi}_\epsilon\Delta z_{jt} + \Delta\hat{u}^*_{\epsilon,jt} \tag{SMP.3}$$

$$\Delta\hat{q}^*_{it} \equiv \tilde{\psi}\Delta\hat{h}^*_{it} + \Delta\hat{c}^*_{it} \tag{SMP.4}$$

$$\Delta\hat{h}^*_{it} \equiv \hat{\pi}_\psi\Delta z^{HD,a}_{it}(\rho) + \Delta\hat{u}^*_{\psi,it} \tag{SMP.5}$$

(c) Separately estimate Equations BS1, BS2, and BS3 as below, using IV for equation BS2 and BS3. Save bootstrap estimates $\{\lambda^*, \epsilon^*, \psi^*\}_b$.

$$\Delta\hat{\ddot{n}}^*_{ijt} = \Delta\ddot{T}_{ijt}\lambda^* + \Delta\ddot{d}^*_{ijt} \tag{BS1}$$

$$\Delta\hat{\Omega}^*_{jt} = \epsilon^*\Delta\hat{w}^*_{jt} + \Delta e^*_{jt} \tag{BS2.IV}$$

$$\Delta\hat{w}^*_{jt} = \pi^*_\epsilon\Delta z_{jt} + \Delta u^*_{\epsilon,jt} \tag{BS2.1st}$$

$$\Delta\hat{q}^*_{it} = \psi^*\Delta\hat{h}^*_{it} + \Delta c^*_{it} \tag{BS3.IV}$$

$$\Delta\hat{h}^*_{it} = \pi^*_\psi\Delta z^{HD,a}_{it}(\rho) + \Delta u^*_{\psi,it} \tag{BS3.1st}$$

6. Remove any sets b of bootstrap estimates for which $\epsilon^* < 0$ or $\psi^* < 0$. Randomly sample from the remaining sets of bootstrap estimates to ensure $\mathcal{B}$ replicates.

7. For each set of bootstrap estimates, simulate the model using $\{\lambda^*, \epsilon^*, \psi^*\}_b$ to create a bootstrapped welfare estimate $\mathcal{W}^*_b$

8. Report the 95% central CI discarding the tails of $\mathcal{W}^*_b, \forall b$.

F Gravity and Commuting Costs

First, a note on measures of travel time/cost:

- τ^{GIS}: Network travel times are calculated from route-querying software. These are *available only in a cross-section*, but are available for every origin-destination pair.
- τ^{Obs}: Observed travel times come from the CTPP. They are *panel data*, but available only between pairs with positive commuting that satisfy disclosure requirements.

I calculate gravity comparing both of these measures and several techniques. The availability of panel data creates a challenge for gravity models using cross-sectional measures of travel time: There is no time variation, and so time-invariant pair fixed effects absorb all variation in travel times, and κ cannot be identified.

To illustrate, Table F1 reports estimates of $\epsilon\kappa$ from the following models, with and without pair fixed effects

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

Without pair fixed effects, the elasticity of commuting with respect to travel time is -0.41 (in column 1), substantially smaller in magnitude than the -1 sometimes used in trade. In comparison, including pair fixed effects generates an estimate of 0.07, indicating that small increases in travel time may actually increase commuting. This captures the tricky issue with using panel data for measuring gravity: For tract pairs that see increases in commuting, congestion may increase, causing increases in travel time (and reverse causality). However it is striking that the sign switches.

Observed travel times in the census are averages of recalled times across all commuters, and may be subject to measurement error. How bad might this measurement error be? For comparison, column (3) reports the results from:

$$n_{ijt} = \omega_{jt} + \theta_{it} - \epsilon\kappa\tau_{ij}^{\text{GIS}} + \lambda^{D\prime}T_{ijt} + \ln(D_{ijt})$$

using the GIS-calculated measure of travel time. It is about 50% larger in magnitude than column 1. This indicates that there is measurement error, but also that observed travel times likely contain substantial signal.

An ideal approach would control for time-invariant determinants of commuting between pairs, but still allow recovering the elasticity of commuting. I propose a two-step approach, first recovering estimates of ς_{ij}^{D} from above then regressing them on a measure of time:

$$\hat{\varsigma}_{ijt}^{D} = -\epsilon\kappa\tau_{ijt}^{\text{Obs}} + u_{ijt}$$
$$\hat{\varsigma}_{ij}^{D} = -\epsilon\kappa\tau_{ij}^{\text{GIS}} + u_{ij}$$

where $\hat{\varsigma}_{ij0}^{D} = \hat{\varsigma}_{ij1}^{D}$ in the first model. Note that, unlike for the ω and θ fixed effects, it is unreasonable to assume asymptotic arguments as the pair fixed effects are estimated from a short panel. This means there may be large measurement error $\hat{\varsigma}_{ij}^{D}$, although expanding the time dimension of the panel likely improves this margin.

Columns 4–7 show the results of the second stage of this two-step process. In columns 4–5, $\hat{\varsigma}_{ij}^{D}$ are estimates from a log-linear model, while in columns 6–7, $\hat{\varsigma}_{ij}^{D}$ are estimates from a PPML model. Note that the coefficients on observed and GIS-calculated measures of travel time are roughly similar now. The log-linear first step results are also a bit smaller than the PPML first step results.

Table F1: Measuring Gravity in the Panel

	n_{ijt}			$\hat{\varsigma}_{ij}^{D}$ (Two-Step Estimator)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\ln(\tau_{ijt}^{\text{Obs}})$	-0.414***	0.073***		-0.239***		-0.391***	
	(0.010)	(0.005)		(0.012)		(0.011)	
$\ln(\tau_{ij}^{\text{GIS}})$			-0.600***		-0.257***		-0.368***
			(0.013)		(0.014)		(0.012)
N	717073	276128	771999	282757	143353	726261	628129
Origin- & Destination-by-Year FEs	Y	Y	Y	na	na	na	na
Pair FEs	-	Y	-	na	na	na	na
First Step Estimated by:	na	na	na	Lin.	Lin.	PPML	PPML

Estimates of marginal disutility of travel time; outcome is log commuting flow in columns 1–3, and pair FEs derived from gravity in columns 4–7. Standard errors clustered by tract pair (columns 1, 2, 3, 4, 6), tract of residence (all columns), and tract of work (all columns) in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

G Appendix References

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

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

Anatolyev, Stanislav. 2013. "Instrumental variables estimation and inference in the presence of many exogenous regressors." *The Econometrics Journal* 16 (1): 27–72.

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

———. 2020. "As the wind blows: The effects of long-term exposure to air pollution on mortality." *Journal of the European Economic Association* 18 (4): 1886–1927.

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

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.

Chao, John C, Jerry A Hausman, Whitney K Newey, Norman R Swanson, and Tiemen Woutersen. 2014. "Testing overidentifying restrictions with many instruments and heteroskedasticity." *Journal of Econometrics* 178:15–21.

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.

Chernozhukov, Victor, and Christian Hansen. 2008. "The reduced form: A simple approach to inference with weak instruments." *Economics Letters* 100 (1): 68–71.

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

Current Air Quality and Trends in the South Coast Air Quality Management District. 2000. Technical report. Diamond Bar, CA: South Coast Air Quality Management District.

Davezies, Laurent, Xavier D'Haultfœuille, and Yannick Guyonvarch. 2019. "Empirical process results for exchangeable arrays." *arXiv preprint arXiv:1906.11293.*

Davidson, Russell, and James G MacKinnon. 2010. "Wild bootstrap tests for IV regression." *Journal of Business & Economic Statistics* 28 (1): 128–144.

Deryugina, Tatyana, Garth Heutel, Nolan H Miller, David Molitor, and Julian Reif. 2019. "The mortality and medical costs of air pollution: Evidence from changes in wind direction." *American Economic Review* 109 (12): 4178–4219.

Diamond, Rebecca. 2016. "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000." *American Economic Review* 106 (3): 479–524.

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.

Gendron-Carrier, Nicolas, Marco Gonzalez-Navarro, Stefano Polloni, and Matthew A Turner. 2021. "Subways and urban air pollution." *American Economic Journal: Applied Economics* Forthcoming.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift. 2020. "Bartik instruments: What, when, why, and how." *American Economic Review* 110 (8): 2586–2624.

Graham, Bryan S. 2020. "Dyadic regression." *The Econometric Analysis of Network Data:* 23–40.

Hausman, Jerry A, Whitney K Newey, Tiemen Woutersen, John C Chao, and Norman R Swanson. 2012. "Instrumental variable estimation with heteroskedasticity and many instruments." *Quantitative Economics* 3 (2): 211–255.

HEI Panel on the Health Effects of Traffic-Related Air Pollution. 2010. *Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects. HEI Special Report 17.* Technical report. Boston, MA: Health Effects Institute.

Kolesár, Michal, Raj Chetty, John Friedman, Edward Glaeser, and Guido W Imbens. 2015. "Identification and inference with many invalid instruments." *Journal of Business & Economic Statistics* 33 (4): 474–484.

Kreindler, Gabriel, and Yuhei Miyauchi. 2020. "Measuring commuting and economic activity inside cities with cell phone records." *NBER Working Paper,* no. w28516.

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

Menzel, Konrad. 2020. *Bootstrap with cluster-dependence in two or more dimensions.*

Owen, Art B. 2007. "The pigeonhole bootstrap." *Annals of Applied Statistics* 1 (2): 386–411.

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

Weitzman, Martin L. 1998. "Why the far-distant future should be discounted at its lowest possible rate." *Journal of Environmental Economics and Management* 36 (3): 201–208.

H Additional and Supplementary Results

Summary of Additional Results

- Figure H1: Map of Proposed LA Metro Lines and PER Lines in Kelker, De Leuw & Co. (1925)
- Figure H2: LA Metro Rail Ridership, 1990-2000
- Figure H3: LA Metro Rail Ridership, 1990-2014
- Figure H4: Glossary of variables and parameters
- Figure H5: Timeline of transportation in Los Angeles
- Figure H6: Robustness to different distance bin assumptions
- Table H1: Descriptive statistics on transportation in Los Angeles and station placement
- Table H2: Pre-trends in tract-level characteristics, 1970-1990
- Table H3: Treatment is not related to changes in zero flows or years opened
- Table H4: Effect of transit on commuting flows by 2000 (PPML with HDFEs)
- Table H5: Effects are larger for tract pairs on the same line
- Table H6: Interactions of residential and workplace station proximity
- Table H7: Bartik tests from Goldsmith-Pinkham, Sorkin, and Swift (2020), no subcounty fixed effects
- Table H8: Bartik tests from Goldsmith-Pinkham, Sorkin, and Swift (2020), with subcounty fixed effects
- Table H9: Transit and non-commuting fundamentals with half spatial decay
- Table H10: Transit and non-mutable fundamentals
- Table H11: Transit, income change, and land use change (robustness)
- Table H12: Effect of LA Metro Rail on Residential Commute Share using Rail Transit
- Table H13: Welfare effects of model extensions (% change in welfare)

Appendix Figures and Tables

Figure H1: Map of Proposed LA Metro Lines and PER Lines in Kelker, De Leuw & Co. (1925)

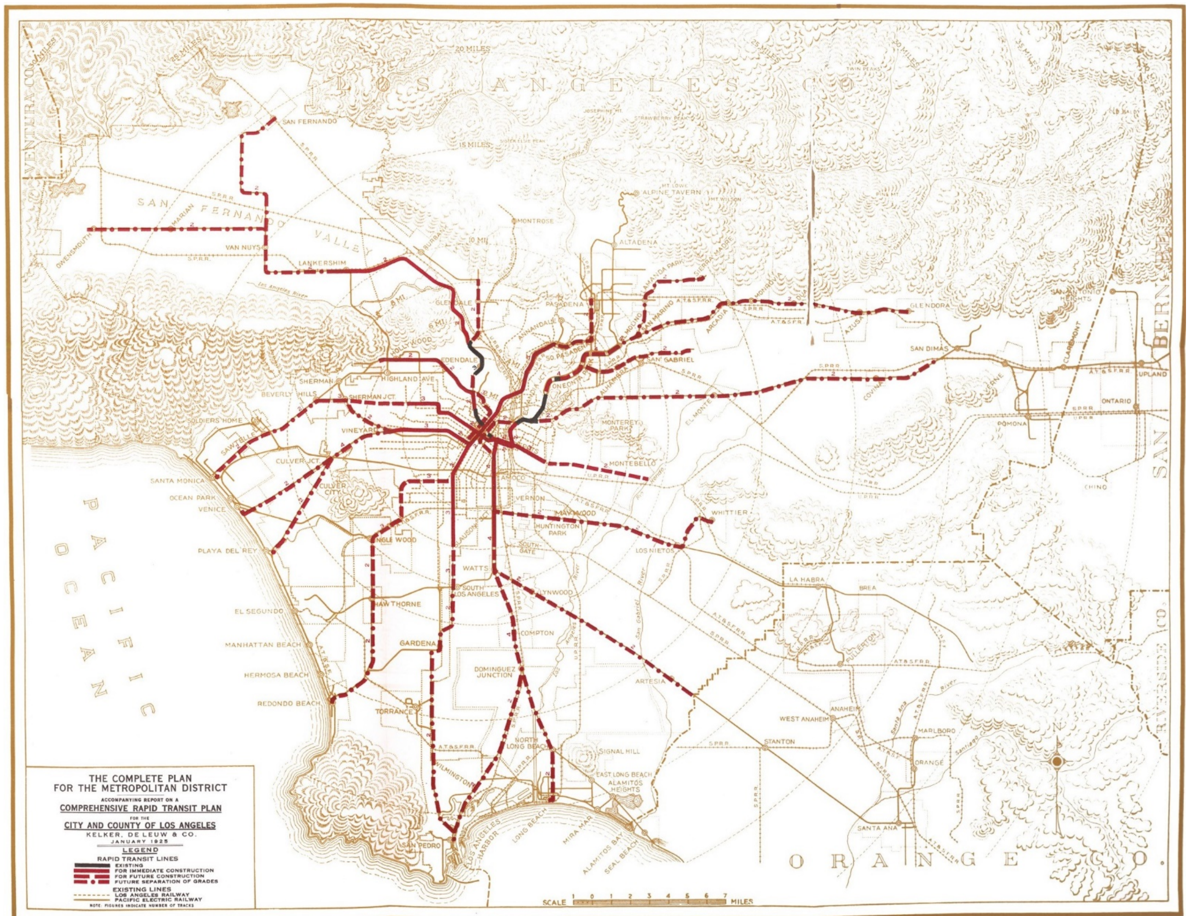


PLATE NO. 1

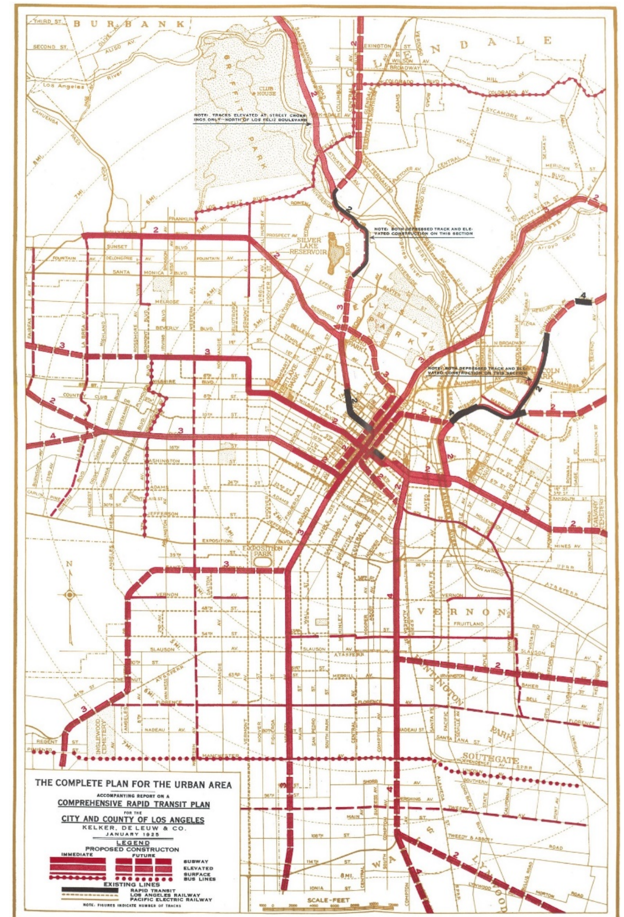


PLATE NO. 2

Figure H2: LA Metro Rail Ridership, 1990-2000

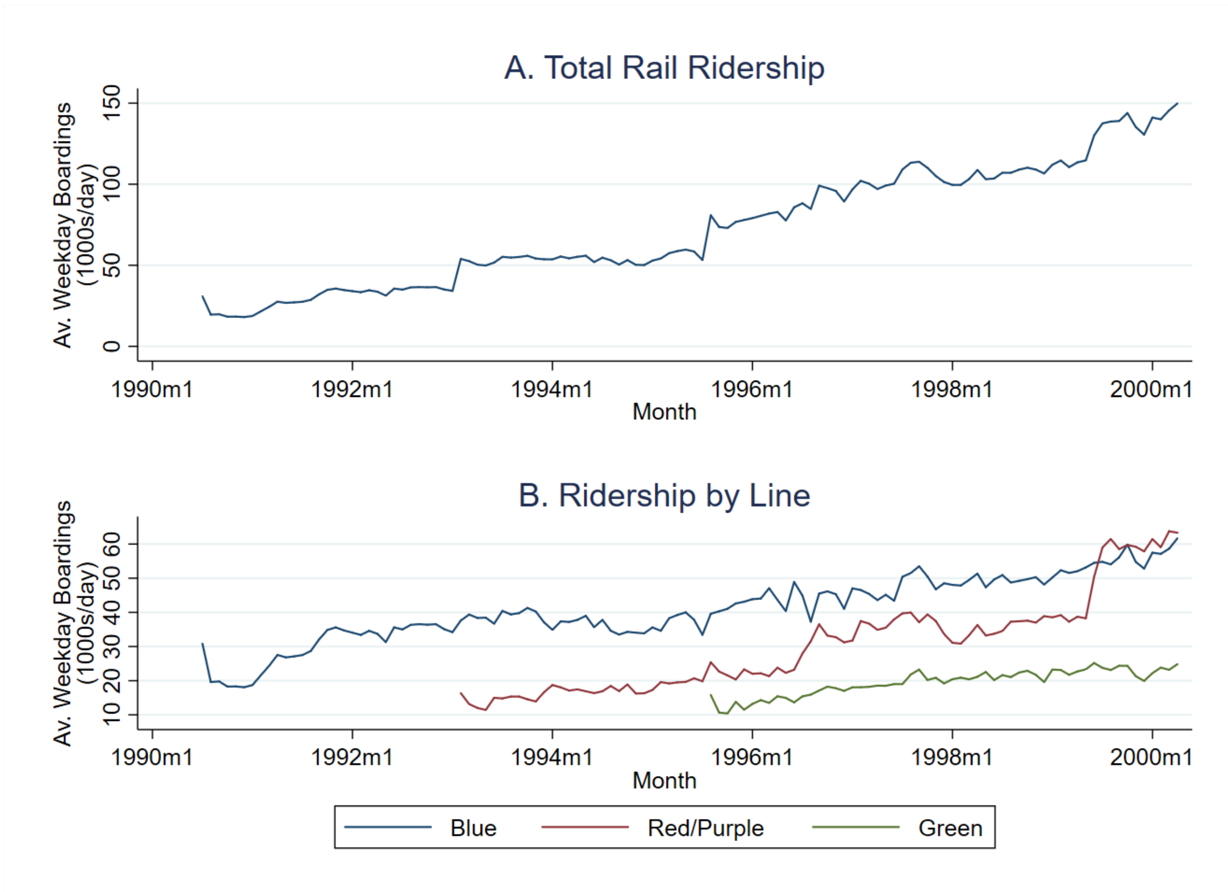


Figure H3: LA Metro Rail Ridership, 1990-2014

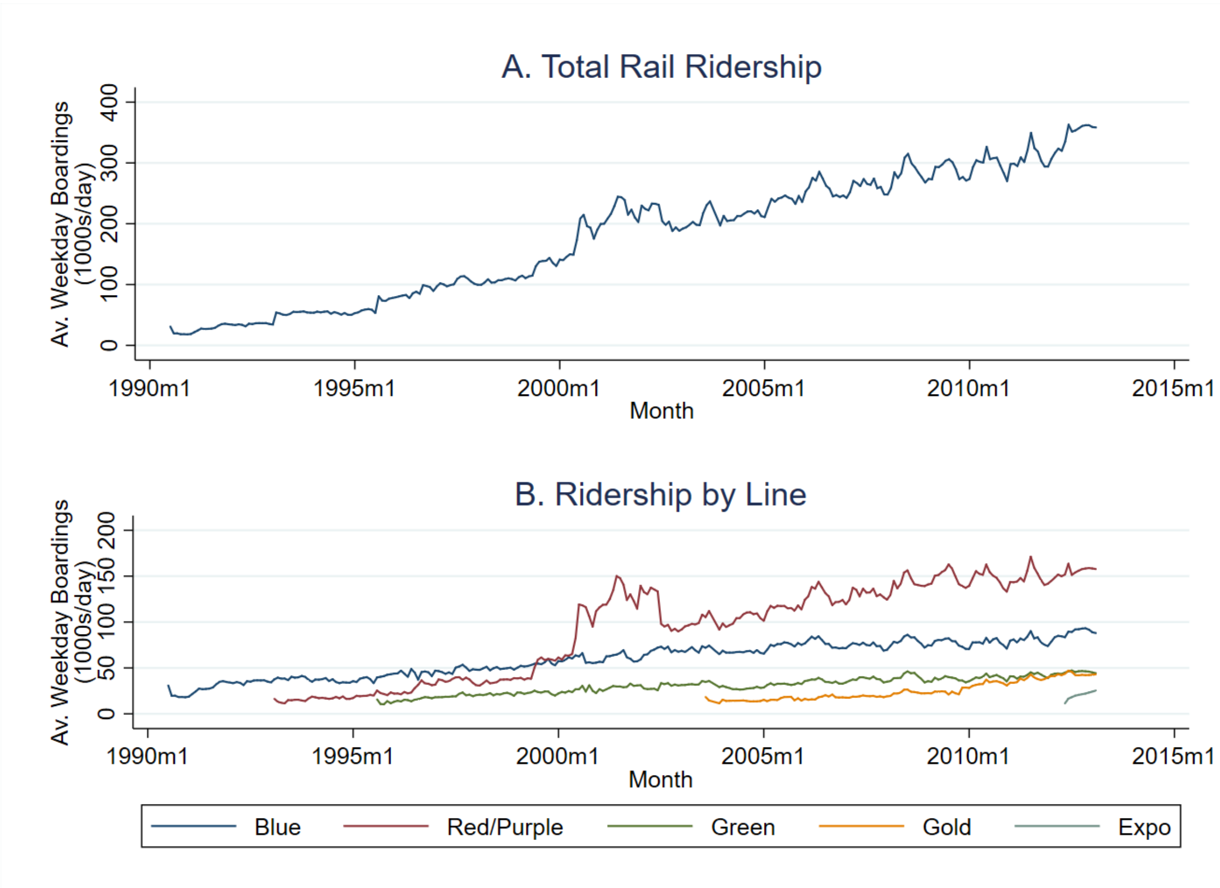


Figure H4: Glossary of variables and parameters

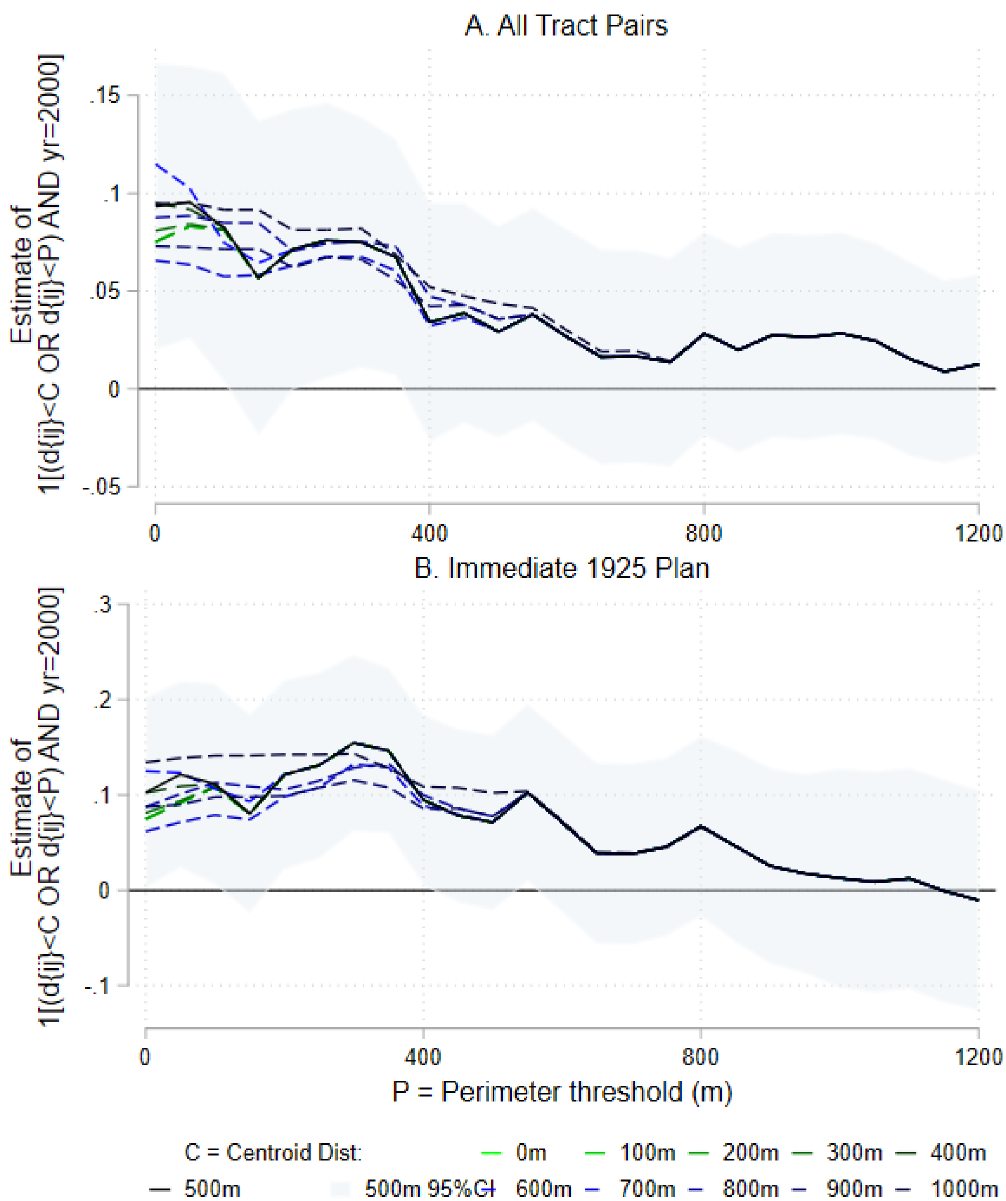
Parameters	Interpretation
ϵ	Homogeneity of location preferences (and wage elasticity of labor supply)
ζ	Household expenditure share on non-housing goods
$\tilde{\zeta} = -\epsilon(1-\zeta)$	Price elasticity of housing demand
α	Share of (production) income spent on labor
$\tilde{\alpha} = \alpha - 1$	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 H5: Timeline of transportation in Los Angeles

1925	Comprehensive Rapid Transit Plan for the County of Los Angeles, Kelker, De Leuw & Co. developed 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 H6: Robustness to different distance bin assumptions



Effect of a single treatment under various definitions of i) as estimated by Equation (2). P represents the distance from a station to the exterior perimeter of the tract if the tract does not contain a station and is zero if the tract contains a station; in the main text it is zero. C represents the second criterion of i) in the main text, the limiting centroid distance; it is 500m in the main text.

Table H1: Descriptive statistics on transportation in Los Angeles and station placement

	LA County		Full Sample	
	(1)	(2)	(3)	(4)
A. Pre-treatment mode choice characteristics (1990)				
% 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				
Commute time (minutes, 1990)	26.3 [16.8]			
Commute time (minutes, 2000)	28.0 [18.3]			
C. % of pre-treatment population that becomes treated	Centroid $< 500m$	Any $< 500m$	Centroid $< 500m$	Any $< 500m$
% workers at POW tract that receive treatment	11.3%	19.4%	7.2%	12.3%
% workers at RES tract that receive treatment	2.6%	8.1%	1.6%	4.8%
% workers that receive transit connection RES-POW	0.6%	2.9%	0.4%	1.7%

Data from Census micro records (from IPUMS) and 1990 CTPP. LA County restricts analysis 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 H2: Pre-trends in tract-level characteristics, 1970-1990

	Model-relevant				Other characteristics			Travel characteristics		
	ln Res. Emp. (1)	ln #HHs (2)	ln HHI (3)	ln House Value (4)	% Coll. Grads (5)	Pov. Rate (6)	% Moved <5yrs (7)	%HHs No Car (8)	%Com. Use Auto (9)	%Com. Use Transit (10)
All Tracts										
Proximity$_i^{500m} \times t$	0.025	-0.032*	-0.015	-0.023	-0.015***	0.015***	-0.016***	-0.015***	0.004	0.012***
	(0.020)	(0.016)	(0.012)	(0.014)	(0.003)	(0.004)	(0.005)	(0.006)	(0.005)	(0.004)
N	11643	11632	11556	11383	11650	11651	11651	7774	11644	11644
PER Sample										
Proximity$_i^{500m} \times t$	0.003	-0.036**	-0.015	-0.042***	-0.015***	0.012***	-0.014**	-0.014**	0.004	0.012***
	(0.022)	(0.017)	(0.014)	(0.016)	(0.004)	(0.005)	(0.006)	(0.006)	(0.005)	(0.004)
N	3696	3695	3689	3591	3696	3696	3696	2464	3696	3696
Full 1925 Plan Sample										
Proximity$_i^{500m} \times t$	-0.003	-0.032*	-0.018	-0.023	-0.011***	0.010**	-0.012**	-0.010	0.006	0.008*
	(0.021)	(0.017)	(0.013)	(0.016)	(0.004)	(0.005)	(0.006)	(0.006)	(0.005)	(0.005)
N	2886	2886	2881	2788	2886	2886	2886	1924	2886	2886
Immediate 1925 Plan Sample										
Proximity$_i^{500m} \times t$	0.023	-0.010	-0.015	-0.004	-0.007*	0.007	-0.007	-0.010	0.006	0.011**
	(0.021)	(0.019)	(0.015)	(0.019)	(0.004)	(0.006)	(0.006)	(0.007)	(0.006)	(0.005)
N	1236	1236	1236	1158	1236	1236	1236	824	1236	1236
Subcounty-×-Year FEs	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Estimates show pre-trends from 1970-1990 for tracts treated between 1990-1999, except Column (8), which covers only 1980–1990. HH is households, and HHI is household income. Data are from the Neighborhood Change Database and reflect 2010 geographies. All regressions include tract and subcounty-by-year fixed effects. Standard errors clustered by tract in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table H3: Treatment is not related to changes in zero flows or years opened

	$1_{N_{ijt}>0}$			$\ln(N_{ijt})$				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
O & D contain station	0.015	0.008	0.005	0.185***	0.150**	0.098**	0.167**	0.142**
	(0.034)	(0.031)	(0.038)	(0.061)	(0.058)	(0.040)	(0.082)	(0.059)
O & D <250m from station	0.005	0.000	-0.026	0.148**	0.105*	0.058	0.145*	0.128**
	(0.027)	(0.026)	(0.033)	(0.061)	(0.058)	(0.046)	(0.077)	(0.065)
O & D <500m from station	0.052**	0.035	0.019	0.073	0.036	-0.009	0.030	0.013
	(0.024)	(0.022)	(0.027)	(0.058)	(0.053)	(0.035)	(0.071)	(0.052)
Years Open				0.073	0.036		0.030	
				(0.058)	(0.053)		(0.071)	
Years Open × O & D contain station						0.004		0.027*
						(0.013)		(0.015)
Years Open × O & D <250m from station						-0.010		-0.011
						(0.015)		(0.018)
Years Open × O & D <500m from station						-0.014		-0.012
						(0.010)		(0.013)
N	1263082	1262478	69614	291532	291110	291110	19222	19222
Control Group	All	All	Immed. '25 Plan	All	All	All	Immed. '25 Plan	Immed. '25 Plan
Standard Three-Way FEs	Y	Y	Y	Y	Y	Y	Y	Y
Subcounty Pair-×-Year FEs	-	Y	Y	-	Y	Y	Y	Y
Highway Controls	-	Y	Y	-	Y	Y	Y	Y

High-dimensional fixed effects estimates of transit on an indicator for positive flows (Columns 1-3) or log commuting flow (Columns 4-8); standard three-way fixed effects are tract of work-by-year, tract of residence-by-year, and tract pair. Sample consists of all non-missing/singular tract pairs. Years opened is relative to the newest station nearest either the origin or destination, centered on 5 years (the mean value). Standard errors clustered by tract pair, tract of residence, and tract of work in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table H4: Effect of transit on commuting flows by 2000 (PPML)

	Full Sample			History & Shocks			Same Line	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
O & D contain station	0.135**	0.164**	0.125*	0.107	0.131*	0.163**	0.207*	0.152
	(0.069)	(0.074)	(0.065)	(0.068)	(0.068)	(0.081)	(0.123)	(0.125)
O & D <250m from station		0.135**	0.101	0.129*	0.122*	0.120	0.210**	0.158
		(0.067)	(0.062)	(0.067)	(0.068)	(0.079)	(0.093)	(0.103)
O & D <500m from station		0.129**	0.077	0.084*	0.071	0.063	0.096	0.101
		(0.058)	(0.049)	(0.050)	(0.051)	(0.057)	(0.070)	(0.077)
N	1261978	1261978	1259440	407832	310924	69596	26270	12172
Control Group	All	All	All	PER	Full '25 Plan	Immed. '25 Plan	Ever Treated	Treated by 2000
Standard Three-Way FEs	Y	Y	Y	Y	Y	Y	Y	Y
Subcounty Pair-×-Year FEs	-	-	Y	Y	Y	Y	Y	Y
Highway Controls	-	-	Y	Y	Y	Y	Y	Y

High-dimensional fixed effects estimates of λ^D with PPML estimator; standard three-way fixed effects are tract of work-by-year, tract of residence-by-year, and tract pair. Outcome is commuting flow. Treatment variables are mutually exclusive. Column titles define treatment: tracts pairs on any lines are treated in Columns (1)-(6), while only tract pairs on the same line are treated in Columns (7) and (8). Standard errors clustered by tract pair, tract of residence, and tract of work in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table H5: Effects are larger for tract pairs on same line

	(1)	(2)	(3)	(4)	(5)	(6)
O & D contain station						
Same line	0.151***	0.163***	0.144***	0.153***	0.192***	0.205***
	(0.053)	(0.053)	(0.051)	(0.055)	(0.058)	(0.071)
Not same line	0.026	0.035	0.040	0.056	0.087	0.075
	(0.076)	(0.077)	(0.076)	(0.084)	(0.082)	(0.088)
O & D <250m from station						
Same line		0.094	0.064	0.096	0.115*	0.145**
		(0.058)	(0.060)	(0.062)	(0.062)	(0.071)
Not same line		0.046	0.049	0.087	0.095	0.105
		(0.059)	(0.059)	(0.062)	(0.064)	(0.079)
O & D <500m from station						
Same line		0.032	0.015	0.048	0.046	0.041
		(0.044)	(0.042)	(0.046)	(0.048)	(0.060)
Not same line		-0.069	-0.072	-0.037	-0.051	-0.048
		(0.046)	(0.044)	(0.046)	(0.046)	(0.056)
N	291532	291532	291110	99480	74408	19222
Control Group	All	All	All	PER	Full '25 Plan	Immed. '25 Plan
Standard Three-Way FEs	Y	Y	Y	Y	Y	Y
Subcounty Pair-×-Year FEs	-	-	Y	Y	Y	Y
Highway Controls	-	-	Y	Y	Y	Y

High-dimensional fixed effects estimates of λ^D with log-linear estimator; standard three-way fixed effects are tract of work-by-year, tract of residence-by-year, and tract pair. Outcome is log commuting flow. Treatment variables are mutually exclusive. Standard errors clustered by tract pair, tract of residence, and tract of work in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table H6: Interactions of residential and workplace station proximity

	(1)			(2)		
	D contains station	D<250m from station	D<500m from station	D contains station	D<250m from station	D<500m from station
O contains station	0.094**	0.146*	-0.144*	0.108*	0.171*	-0.197**
	(0.038)	(0.077)	(0.076)	(0.063)	(0.100)	(0.094)
O<250m from station	0.021	0.017	0.043	0.096	0.118	0.051
	(0.059)	(0.074)	(0.103)	(0.075)	(0.110)	(0.111)
O<500m from station	0.009	-0.012	0.034	0.087	0.083	0.056
	(0.049)	(0.060)	(0.047)	(0.063)	(0.084)	(0.069)
N		291110			19222	
Control Group		All			Immediate '25 Plan	
Standard Three-Way FEs		Y			Y	
Subcounty Pair-×-Year FEs		Y			Y	
Highway Control		Y			Y	

High-dimensional fixed effects estimates of λ^D with log-linear estimator; standard three-way fixed effects are tract of work-by-year, tract of residence-by-year, and tract pair. Outcome is log commuting flow. Treatment variables. Standard errors clustered by tract pair, tract of residence, and tract of work in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table H7: Bartik tests from Goldsmith-Pinkham, Sorkin, and Swift (2020), no subcounty fixed effects

Panel A: Negative and positive weights

	Sum	Mean	Share
Negative	-0.245	-0.035	0.164
Positive	1.245	0.113	0.836

Panel B: Correlations of Industry Aggregates

	α_k	g_k	β_k	F_k	Var(z_k)
α_k	1				
g_k	0.043	1			
β_k	0.132	0.463	1		
F_k	0.148	-0.178	-0.477	1	
Var(z_k)	0.490	-0.112	-0.047	-0.025	1

Panel C: Top 5 Rotemberg weight industries

	$\hat{\alpha}_k$	g_k	$\hat{\beta}_k$	95 % CI	Ind Share
Manufacturing (durable)	0.395	-0.058	3.526	N/A	13.135
Transportation	0.226	-0.110	0.558	(-1.6,2.0)	4.141
FIRE	0.193	0.106	2.248	(-0.4,12.4)	7.805
Personal Services	0.124	0.178	5.181	(2.4,22.6)	3.565
Wholesale Trade	0.112	-0.058	-0.897	(-2.5,0.2)	4.919

Panel D: Estimates of β_k for positive and negative weights

	α-weighted sum	Share of overall β	Mean
Negative	-0.315	-0.126	1.730
Positive	2.827	1.126	1.985

Panel E: Alternative estimates and overidentification

	Bartik	TSLS	LIML	MBTSLS	HFUL
$\Delta \ln(W_{jt})$	2.512	1.457	2.911	1.571	2.470
	(1.094)	(0.662)	(1.835)	(0.705)	(5.065)
Over ID Test Stat.		34.2	22.2		23.9
p-value		[0.01]	[0.14]		[0.09]

This table reports statistics about Rotemberg weights and alternate IV estimators as suggested in Goldsmith-Pinkham, Sorkin, and Swift (2020). The results correspond to Table 3 column 5. In all cases, statistics reflect normalized growth rates. Panel A reports the share and sum of negative Rotemberg weights. Panel B reports correlations between the weights (α_k), the national component of growth (g_k), the just-identified coefficient estimates (β_k), the first-stage F-statistic of the industry share (F_k), and the variation in the industry shares across locations (VAR(z_k)). Panel C reports variation in the weights across years. Panel D reports the top five industries according to the Rotemberg weights. The 95% CI uses the weak-instrument robust CI from Chernozhukov and Hansen (2008) over a range from -25 to 25 and is N/A if it exceeds that range, and Ind Share is the industry share (multiplied by 100). Panel E reports a variety of alternative estimates. TSLS uses each industry share separately as instruments. LIML reports estimates using the limited information maximum likelihood estimator with the same set of instruments. MBTSLS uses Anatolyev (2013) and Kolesár et al. (2015) with the same set of instruments. HFUL uses the HFUL estimator (Hausman et al. 2012) with the same set of instruments, and the J-statistic from Chao et al. (2014). p-values are in brackets.

Table H8: Bartik tests from Goldsmith-Pinkham, Sorkin, and Swift (2020), with subcounty fixed effects

Panel A: Negative and positive weights					
	Sum	Mean	Share		
Negative	-0.290	-0.048	0.184		
Positive	1.290	0.108	0.816		
Panel B: Correlations of Industry Aggregates					
	α_k	g_k	β_k	F_k	Var(z_k)
α_k	1				
g_k	-0.051	1			
β_k	-0.009	0.460	1		
F_k	0.180	-0.347	-0.468	1	
Var(z_k)	0.524	-0.112	-0.218	0.092	1
Panel C: Top 5 Rotemberg weight industries					
	$\hat{\alpha}_k$	g_k	$\hat{\beta}_k$	95 % CI	Ind Share
Manufacturing (durable)	0.457	-0.058	2.980	N/A	13.149
Transportation	0.262	-0.110	-0.305	(-1.80,0.80)	4.143
FIRE	0.144	0.106	3.398	N/A	7.814
Health	0.137	0.050	-1.110	(-6.80,1.40)	7.047
Wholesale Trade	0.100	-0.058	-2.690	(-7.80,-1.10)	4.924
Panel D: Estimates of β_k for positive and negative weights					
	α-weighted sum	Share of overall β	Mean		
Negative	-0.178	-0.082	0.735		
Positive	2.358	1.082	3.066		
Panel E: Alternative estimates and overidentification					
	Bartik	TSLS	LIML	MBTSLS	HFUL
$\Delta\ln(W_{jt})$	2.180	0.450	25.760	0.464	.
	(1.171)	(0.542)	(1363.467)	(0.813)	()
Over ID Test Stat.		75.9	0.8		.
p-value		[0.00]	[1.00]		[.]

This table reports statistics about Rotemberg weights and alternate IV estimators as suggested in Goldsmith-Pinkham, Sorkin, and Swift (2020). The results correspond to Table 3 column 6. In all cases, statistics reflect normalized growth rates. Panel A reports the share and sum of negative Rotemberg weights. Panel B reports correlations between the weights (α_k), the national component of growth (g_k), the just-identified coefficient estimates (β_k), the first-stage F-statistic of the industry share (F_k), and the variation in the industry shares across locations (VAR(z_k)). Panel C reports variation in the weights across years. Panel D reports the top five industries according to the Rotemberg weights. The 95% CI uses the weak-instrument robust CI from Chernozhukov and Hansen (2008) over a range from -25 to 25 and is N/A if it exceeds that range, and Ind Share is the industry share (multiplied by 100). Panel E reports a variety of alternative estimates. TSLS uses each industry share separately as instruments. LIML reports estimates using the limited information maximum likelihood estimator with the same set of instruments. MBTSLS uses Anatolyev (2013) and Kolesár et al. (2015) with the same set of instruments. HFUL uses the HFUL estimator (Hausman et al. 2012) with the same set of instruments, and the J-statistic from Chao et al. (2014). p-values are in brackets.

Table H9: Transit and non-commuting fundamentals with half spatial decay

	$\bar{d}$ = 500m				$\bar{d}$ = 1km			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Effect on productivity								
I.) λ^A *estimated using* $\Delta\hat{\mathcal{A}} = \Delta\hat{A} - \mu\Delta\ln(\Upsilon^{Far})$								
Proximity$_i \times t$	0.031	0.030	0.017	0.021	0.040	0.041	0.027	0.035
	(0.039)	(0.039)	(0.038)	(0.043)	(0.037)	(0.038)	(0.037)	(0.043)
N	2469	1167	934	394	2469	1167	934	394
Effect on residential amenity level								
IV.) λ^B *estimated using* $\Delta\hat{\mathcal{B}} = \Delta\hat{B} - \eta\Delta\ln(\Omega^{Far})$								
Proximity$_i \times t$	0.050	0.068**	0.045	0.007	0.035	0.056*	0.025	-0.028
	(0.032)	(0.033)	(0.033)	(0.035)	(0.029)	(0.031)	(0.030)	(0.034)
N	2149	994	815	343	2149	994	815	343
Control Group	All	PER	Full '25 Plan	Immed. '25 Plan	All	PER	Full '25 Plan	Immed. '25 Plan

Results from sixteen regressions of transit proximity on local productivity after removing agglomeration. Here, the distance effect of agglomeration decays at half the values in Ahlfeldt et al. (2015). All regressions include tract fixed effects, subcounty-by-year fixed effects, and controls. Controls include changes in highway proximity and 1990 levels of log household income, share of residents with at least a high school degree, and manufacturing employment. Sample size reflects number of differenced tracts. Standard errors clustered by tract in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table H10: Transit and non-mutable fundamentals

	$\Delta\hat{\mathbf{Y}}_{it}$							
	$\bar{d}$ = 500m				$\bar{d}$ = 1km			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Effect on inverse housing supply efficiency $\Delta\hat{C}$**,** λ^C								
Proximity$_i \times t$	0.006	0.007	-0.024	-0.041	0.007	0.012	-0.031	-0.053
	(0.049)	(0.051)	(0.051)	(0.058)	(0.044)	(0.047)	(0.048)	(0.057)
N	2172	996	818	348	2172	996	818	348
B. Effect on workplace amenity $\Delta\hat{E}$**,** λ^E								
Proximity$_i \times t$	-0.016	-0.083	-0.064	-0.115	-0.004	-0.087	-0.062	-0.138*
	(0.070)	(0.072)	(0.075)	(0.084)	(0.066)	(0.069)	(0.072)	(0.084)
N	2516	1168	935	395	2516	1168	935	395
Control Group	All	PER	Full '25 Plan	Immed. '25 Plan	All	PER	Full '25 Plan	Immed. '25 Plan

Results from sixteen regressions of transit proximity on local fundamentals. All regressions include tract fixed effects, subcounty-by-year fixed effects, and controls. Controls include changes in highway proximity and 1990 levels of log household income, share of residents with at least a high school degree, and manufacturing employment. Sample size reflects number of differenced tracts. Standard errors clustered by tract in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table H11: Transit, income change, and land use change (robustness)

	$\Delta\hat{\mathbf{Y}}_{it}$							
	$\bar{d}$ = 500m				$\bar{d}$ = 1km			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Change in residential land								
Proximity$_i \times t$	0.006***	0.002	0.001	0.001	0.007***	0.003	0.002	0.001
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
N	2468	1152	920	385	2468	1152	920	385
B. Change in household income								
Proximity$_i \times t$	-0.016	-0.006	-0.006	-0.019	-0.025	-0.014	-0.015	-0.034*
	(0.017)	(0.017)	(0.017)	(0.018)	(0.016)	(0.017)	(0.017)	(0.019)
N	2476	1142	915	380	2476	1142	915	380
Control Group	All	PER	Full '25 Plan	Immed. '25 Plan	All	PER	Full '25 Plan	Immed. '25 Plan

Results from sixteen regressions of transit proximity on residential land use and household income. All regressions include tract fixed effects, subcounty-by-year fixed effects, and controls. Controls include changes in highway proximity and 1990 levels of log household income, share of residents with at least a high school degree, and manufacturing employment. Sample size reflects number of differenced tracts. Standard errors clustered by tract in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table H12: Effect of LA Metro Rail on residential commute share using rail transit

	Residential Commute Share using Rail Transit							
	$\bar{d}$ = 500m				$\bar{d}$ = 1km			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Proximity$_i \times t$	0.0084***	0.0081***	0.0081***	0.0079***	0.0078***	0.0077***	0.0077***	0.0076***
	(0.0015)	(0.0014)	(0.0014)	(0.0015)	(0.0012)	(0.0012)	(0.0013)	(0.001)
N	2262	1037	848	371	2262	1037	848	371
Control Group	All	PER	Full '25 Plan	Immed. '25 Plan	All	PER	Full '25 Plan	Immed. '25 Plan

Results from eight regressions of transit proximity on subway/light rail commute share. All regressions include tract fixed effects, subcounty-by-year fixed effects, and controls. Controls include changes in highway proximity and 1990 levels of log household income, share of residents with at least a high school degree, and manufacturing employment. Sample size reflects number of differenced tracts. Standard errors clustered by tract in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table H13: Welfare effects of model extensions (% change in welfare)

	Partial Equilibrium (1)	General Equilibrium (2)	General Equilibrium + Spillovers (3)
Baseline: Commuting Effect through 2000	0.04410	0.04404	0.04324
+ Dynamic Effect (through 2015)	0.08032	0.07965	0.07838
+ Congestion Effect	0.10671	0.10632	0.10431
+ Dynamic & Congestion Effects	0.14295	0.14196	0.13946
Baseline & Reduced Land Use Regulation	0.12234	0.10574	0.10371
+ Dynamic Effect (through 2015)	0.15859	0.14266	0.14040
+ Congestion Effect	0.18500	0.16806	0.16481
+ Dynamic & Congestion Effects	0.22127	0.20500	0.20153
Baseline & 5% Amenity	0.10757	0.10688	0.10478
+ Dynamic Effect (through 2015)	0.14544	0.14355	0.14118
+ Congestion Effect	0.17022	0.16920	0.16588
+ Dynamic & Congestion Effects	0.20811	0.20589	0.20230
Baseline & 4% Productivity	0.36608	0.28847	0.28735
+ Dynamic Effect (through 2015)	0.40241	0.32548	0.32400
+ Congestion Effect	0.42888	0.35090	0.34857
+ Dynamic & Congestion Effects	0.46524	0.38794	0.38524

This table gives the percentage change in welfare for various model scenarios (e.g., 0.04409 is a 0.044% change in welfare). Columns 1–3 show partial equilibrium results, general equilibrium results the ignore endogenous agglomeration, and general equilibrium results that account for endogenous agglomeration. Partial equilibrium reflects only changes in items that feed into the utility function, but no feedbacks (but effects to A and C do have price effects). The first panel includes the results in the main paper, whereas the other panels present experiments presuming other effects become present: Reduced Land Use Regulation assumes $\lambda^C = -0.3297$ just in tracts containing transit stations, such that residential density increases in those locations by 10%; 5% Amenity assumes $\lambda^B = 0.05$ with the 500m proximity measure of treatment; and 4% Productivity assumes $\lambda^A = 0.04$ with the 500m proximity measure of treatment.