Political Preferences and Transport Infrastructure: Evidence from California's High-Speed Rail*

Pablo Fajgelbaum†, Cecile Gaubert‡, Nicole Gorton, Eduardo Morales§, and Edouard Schaal*

†UCLA and NBER
‡Berkeley and NBER
§Princeton and NBER
*CREI, ICREA, UPF and BSE

July 2024

Abstract

We study how political preferences shaped California's High-Speed Rail (CHSR), a large transportation project approved by referendum in 2008. Voters' support responded significantly to the projected economic gains in their tract of residence, as measured by a quantitative model of high-speed rail matched to CHSR plans. Given this response, a revealed-preference approach comparing the proposed network with alternative designs identifies strong planner's preferences for political support. The optimal politically-blind design would have placed the stations nearer to California's dense metro areas, where it was harder to sway votes, thus increasing the projected economic gains.

*We thank Andrii Parkhomenko, Jonathan Dingel, and Giacomo Ponzetto for their conference discussions, and seminar participants at ASSA Meetings, Berkeley, Chicago, CREI, Chicago-Princeton Spatial Economics Conference, Harvard-MIT, Michigan, NBER Transportation, Penn State, St. Louis Fed, SED, Timbergen, Toronto, UC Irvine, UCLA ITS, University of Washington, UTDT, West Coast Spatial Workshop, and Wisconsin for helpful feedback. We thank the NBER and the US DOT for financial support. Cecile Gaubert acknowledges support from NSF CAREER grant #1941917. Edouard Schaal acknowledges financial support from the Spanish Ministry of Economy and Competitiveness, through the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2015-0563) and the European Research Council Starting Grant 804095. This study was made possible with funding received by the University of California Institute of Transportation Studies from the State of California for the California Resilient and Innovative Mobility Initiative (RIMI). This paper is disseminated under the sponsorship of the State of California in the interest of information exchange and does not necessarily reflect the official views or policies of the State of California.

1 Introduction

The efficiency of transportation systems constitutes a central question in spatial economics. Using quantitative spatial frameworks, recent research has studied the optimality of transport networks, finding that inefficiencies are pervasive and that observed transport systems could have been designed in alternative welfare-improving ways. Recent studies include Fajgelbaum and Schaal (2020) and Allen and Arkolakis (2022) for highways, Brancaccio et al. (2024) for ports, Kreindler et al. (2023) and Almagro et al. (2024) for bus systems, and Frechette et al. (2019), Buchholz (2022), and Brancaccio et al. (2023) for taxis or bulk shipping.

Why are transport networks inefficient? In this paper, we study the role of households' and policymakers' preferences in shaping the projects that are implemented. In the process of designing transport networks, policymakers may take into account the popular approval elicited by these projects, as well as distributional impacts among constituencies. Investments driven by these motives –for example, targeting areas to maximize political support– will generally differ from investments driven by aggregate welfare considerations alone. Quantifying these differences requires a methodology to estimate the weight that these motivations play in the planner's preferences, as well as gauging whether, and by how much, public support responds to the economic impacts of transport investment projects.

We make progress on these questions in the context of California's High-Speed Rail (CHSR), one of the most expensive transport projects attempted in U.S. history. Proposition 1A, on the ballot in the 2008 general election, asked Californians whether they approved of initiating funding for the CHSR (it passed with 52.6% in favor). The features of the project that were known when voting –through business plans, environmental reports, and the ballot's text– resulted from years-long planning by authorities who anticipated putting the CHSR up to a public vote. We estimate Californian politicians' and voters' preferences, and study their role in shaping the proposed CHSR. To that end, we use a novel framework of optimal high-speed rail network design combined with voting data and CHSR planning data.[1]

We first develop and estimate a spatial model that incorporates specificities of high-speed rail passenger travel. Then, using census-tract data, we estimate the relationship between favorable votes in the 2008 referendum and the expected local economic impacts of the CHSR predicted by the model. To address endogeneity concerns, we build instruments based on random station placements along alternative CHSR designs entertained early in the planning process. We then embed the spatial model and the estimated voters' responses into the problem of a politically-minded planner who decides where to locate stations. We estimate bounds on the policymakers' preference parameters by comparing the actual CHSR design with alternatives that were not selected. Finally, we solve for the optimal station placements under alternative policymakers' preferences.

[1] Since it was approved by referendum in 2008, the CHSR has been mired in a myriad of financial, legal, and implementation troubles; doubts linger on whether it will ever be operational. We do incorporate the possibility of failure in the expectations of voters and policymakers, as we explain in detail below. Our focus is *not* on understanding why the project faced so much trouble, but on studying what its initial design reveals about the preferences of voters and policymakers for this infrastructure project, and on how these preferences shaped its design.

We summarize three main takeaways of our analysis. First, we find voters responded significantly to the model-implied projected economic impact of the CHSR in their tract of residence. This result shows that economic voting may be a significant driver of policy preferences over transport infrastructure. It also gives credibility to our estimated model as a predictor of the spatial distribution of economic impacts of the CHSR. Second, the CHSR design implies strong planner preferences for votes: deviations that would have increased aggregate welfare while reducing votes were not implemented, thus identifying a positive lower bound. Third, these preferences for votes partly shaped the network: the optimal politically-blind CHSR design would have concentrated stations closer to urban areas where it was harder to sway votes. In so doing, the projected gross economic benefits of the CHSR would have increased. We thus conclude that attaining popular approval was an important driver of the CHSR design.

We now describe each step of our analysis. First, we develop and estimate a quantitative spatial model of high-speed passenger travel. The goal of the model is to obtain the distribution across census tracts of the relative real-income impacts of the CHSR, to then use voting data from the 2008 referendum to estimate the responsiveness of votes to these relative impacts. Compared to canonical urban frameworks centered on commuting, such as Ahlfeldt et al. (2015) and Monte et al. (2018), our model includes three distinct features. First, while facilitating commuting into urban centers is an important role of high-speed rail systems (Zheng and Kahn, 2013), long-distance rail connections also confer benefits to infrequent business or leisure travelers. We incorporate these additional travel purposes and rely on the California Household Travel Survey to quantify the parameters determining their importance.[2] Second, as CHSR usage would depend on access to competing travel modes, we include a choice over transport modes (between car, air, and public transit) for each origin-destination pair and travel purpose. Third, we incorporate that travel decisions depend both on travel time (as in standard frameworks) and on monetary trip costs, an a priori relevant feature given the low ticket prices announced by CHSR planners.[3]

To estimate the key model parameters, we rely on gravity equations for commuting, leisure, and business travel across California's census tracts. The resulting estimates reveal how travel time and cost differences by route are valued by Californians, as well as how preferences over different travel modes vary across regions, demographic groups, and travel purpose. Using the estimated model, we then simulate travelers' choices over destinations and mode of transport in the hypothetical scenario in which the CHSR becomes available, and obtain as a result the spatial distribution of the potential real income effects of the CHSR across 7866 census tracts.[4]

[2] We model long-distance business and leisure travel decisions as including an intensive margin, the number of trips, which we observe in the data. We model business travel as an input in production, with firms (rather than travelers) deciding the destination of business trips as a function of characteristics that make certain destinations more profitable for business connections. Through this channel, the high-speed rail network may affect TFP by better matching business travelers to business hubs. Bernard et al. (2019) and Dong et al. (2020) provide evidence that the high-speed rail in Japan and China, respectively, raised business or research productivity by facilitating face-to-face interactions.

[3] The prices projected in the 2022 update to the CHSR business plan are considerably higher than the 2008 forecasts. We account for both of these forecasts by considering alternative scenarios when quantifying our model.

[4] Our setting is sparse, with about 15 million workers for about 64 million origin-destination pairs of census tracts;

Our model predicts that larger urban centers such as San Francisco or Los Angeles, and areas closer to the location of the planned CHSR railway stations, would on average benefit relatively more from the implementation of the CHSR; in contrast, more sparsely populated areas like Central-Valley have lower potential gains. These model-based predictions are qualitatively consistent with existing empirical assessments of passenger transport systems.[5] Moreover, as shown in the next step, the voting patterns in the 2008 referendum responded to these model-implied real income gains from the CHSR, providing validation of the model predictions.

We compute these model-implied real-income gains of the CHSR using official CHSR business plans as available in two different years: 2008 (when the vote took place) and 2022. The 2008 information turned out to be unrealistic: the construction of the CHSR progressed at a much slower pace and costs were higher than forecasted. In contrast, the 2022 CHSR business plan acknowledges risk, costs, and timeline increases. Whether we use the 2008 or the 2022 scenarios affects the level of economic impacts across tracts; for example, in our baseline model, we compute a slightly negative population-weighted net impact using 2008 cost projections, but a much stronger negative net impact under the 2022 scenario.[6] However, the ranking of winners and losers does not change: the economic impacts when using 2008 and 2022 forecasts is highly correlated across census tracts. We use this distribution to estimate if there was a response of votes to expected real income impacts in the next step.

In the second step, we estimate the elasticity of favorable votes in the 2008 referendum to the model-implied expected economic gains from the CHSR. Identifying this response is key for our goal of determining whether planning authorities took into account public support when deciding the design of the CHSR. By following a structural approach to measure the private economic benefits of a policy, we depart from the prior literature that tries to determine the role of those benefits in voters' choices. These previous studies project votes on reduced-form variables that plausibly capture individual exposure to a policy.[7] Our reliance on a structural model is essential to recover a structural elasticity and to characterize counterfactual optimal networks.

Our structural approach is not exempt from identification challenges, crucially the possibility

moreover, the long distance travel information that we use comes from travel surveys collected on limited samples. Consequently, to ensure that our estimates of model parameters are not affected by an incidental parameters problem, we follow the suggestion in Dingel and Tintelnot (2020) and project all travel costs and destination-specific effects on vectors of observed characteristics.

[5]Recent studies of the the spatial impacts of public transit systems include Tsivanidis (2019) for rapid buses in Bogota, Severen (2021) and Tyndall (2021) for the light rail in Los Angeles and other US cities, Gupta et al. (2022) for the NYC subway, Zárate (2022) for subway lines in Mexico City, Khanna et al. (2023) for cable-car in Medellin, and Borusyak and Hull (2023) for the Chinese rail. Broadly speaking, these studies find positive impacts of proximity to transit connections on commuting flows, formal employment, wages, or land prices. Koster et al. (2022) finds employment losses for smaller areas connected to the Japanese high-speed rail.

[6]We obtain smaller gross benefits (before capital costs) than the official numbers calculated by CHSR authorities using different methods than ours.

[7]E.g., Van Patten and Méndez (2022), studying a referendum in Costa Rica on whether to sign a FTA with the U.S., project favorable votes on voters' exposure to U.S. trade. Following Deacon and Shapiro (1975), many studies use ballots to draw inference about demand for private versus collective goods. Kahn and Matsusaka (1997) and Holian et al. (2013) correlate votes in ballot initiatives, including the CHSR Proposition 1A, with proxies for economic exposure such as industry of voters. Alesina and Giuliano (2011) review research that uses survey data to measure preferences for private vs. public value in income taxation.

that the model-implied economic benefits of the CHSR are correlated with other variables, such as preferences over public goods or political ideology, that also impacted voters' support for the CHSR.[8] We take several steps to deal with these identification challenges. First, we control for a host of tract-level covariates including county fixed effects, demographic characteristics, party affiliation, votes in related referenda, and distance to stations. Second, we instrument for the economic impact of the CHSR using alternative CHSR designs with random placement of stations across feasible routes, according to transport engineers, that were announced in years prior to the 2008 vote. Third, we conduct the estimation under different model specifications in terms of economic mechanisms and voters' expectations on CHSR costs at the time of voting.

Regardless of the model variant and the identification strategy, we find that voters are responsive to the model-based expected economic impact of the CHSR. Moreover, once we instrument, this elasticity is robust to the set of controls. Depending on whether 2008 or 2022 cost predictions are used, an extra 0.03-0.10 percentage points in local expected economic gains swayed one percentage point of local votes. This high responsiveness of votes to projected economic impacts implies that policymakers who value public support may have shifted the supply of infrastructure towards certain areas based on the marginal impact on public support, at the expense of where it may have been socially more desirable.

In the third and final step, we estimate the preferences of a social planner designing the CHSR and then compute counterfactual optimal designs under alternative preferences. We model a policy-maker choosing the distribution of stations along the technologically feasible routes linking Northern and Southern California, including the proposed CHSR and its main alternative along the Interstate 5 highway. We assume that the observed CHSR maximized a weighted sum of two components: a sum of tract-specific real income impacts of the CHSR (with tract-specific Pareto weights as function of demographics), and the total voter support for the project.[9]

To estimate the planner's preferences, we follow a revealed-preference approach in the spirit of Goldberg and Maggi (1999) in international trade and Bourguignon and Spadaro (2012) in public finance. Our approach more specifically relates to Adão et al. (2023)'s analysis of U.S. tariffs. Like them, we use a fully-specified quantitative model to construct perturbations of the planner's objective function in response to counterfactual policies, and then estimate the planner's preferences that rationalize the observed policy as maximizing the value of the planner's objective function.

A challenge in our context is that closed-form solutions for optimal policies, which are typically used in the previous literature, are unavailable. Furthermore, marginal perturbations to stations' locations have little identifying power. Our approach therefore uses discrete deviations from the observed station placement –for example, by shifting a CHSR station from its designated location to the next-largest urban area without a proposed station. In doing so, we derive revealed-preference moment inequalities following Pakes (2010) and Pakes et al. (2015). These deviations set bounds

[8]Voting on subjective considerations, referred to as expressive voting in the political sciences literature (Hillman, 2010), helps to rationalize phenomena such as high voting turnout (Brennan and Hamlin, 1998).

[9]Burgess et al. (2015) find empirical evidence that a president's ethnicity or birthplace explains patterns of road investments in Kenya.

on the planner's preferences; for instance, deviations that increase aggregate votes but reduce aggregate income define an upper bound on how much the planner likes the former relative to the latter. We use the moment-inequality inference procedure in Andrews and Soares (2010) to compute confidence sets for the planner's preference parameters.

Our results show strong planner preferences for votes: for an additional percentage point of favorable votes, the planner trades off up to 0.89% aggregate income gains using 2008 cost projections (or 0.14% using 2022 cost projections). The estimates also imply some preference for areas with a larger share of college graduates. The planner is, thus, far from utilitarian.

Finally, we compute optimal networks under counterfactual planner preferences. The optimal CHSR design for an *apolitical* planner differs substantially from the proposed plan, with many stations located closer to the main metropolitan areas. The reason is these metro areas have low voting elasticities: they would have supported the CHSR regardless of private economic gains. Thus, in the absence of electoral motives, policymakers would have placed stations closer to these locations. Doing so would have increased by 15%-25% the projected gross economic benefits of the network, depending on what cost projections are used. Nearly all of this difference between the proposed CHSR and what would have been selected by a utilitarian planner comes from the preferences for votes, with the estimated heterogeneity in preferences over demographic groups playing a minor role.

The paper proceeds as follows. Section 2 gives some background on the CHSR. Section 3 lays out our quantitative model of the CHSR's economic impacts. Section 4 presents the estimation of the model parameters and the distribution of CHSR's local economic impacts. Section 5 estimates the effect of these local economic impacts on votes. Section 6 embeds the model and voter responses into a planners' problem to estimate its preferences and implement counterfactual optimal designs. Section 6 concludes. A full description of the model as well as details on data sources and implementation appears in the Online Appendix.

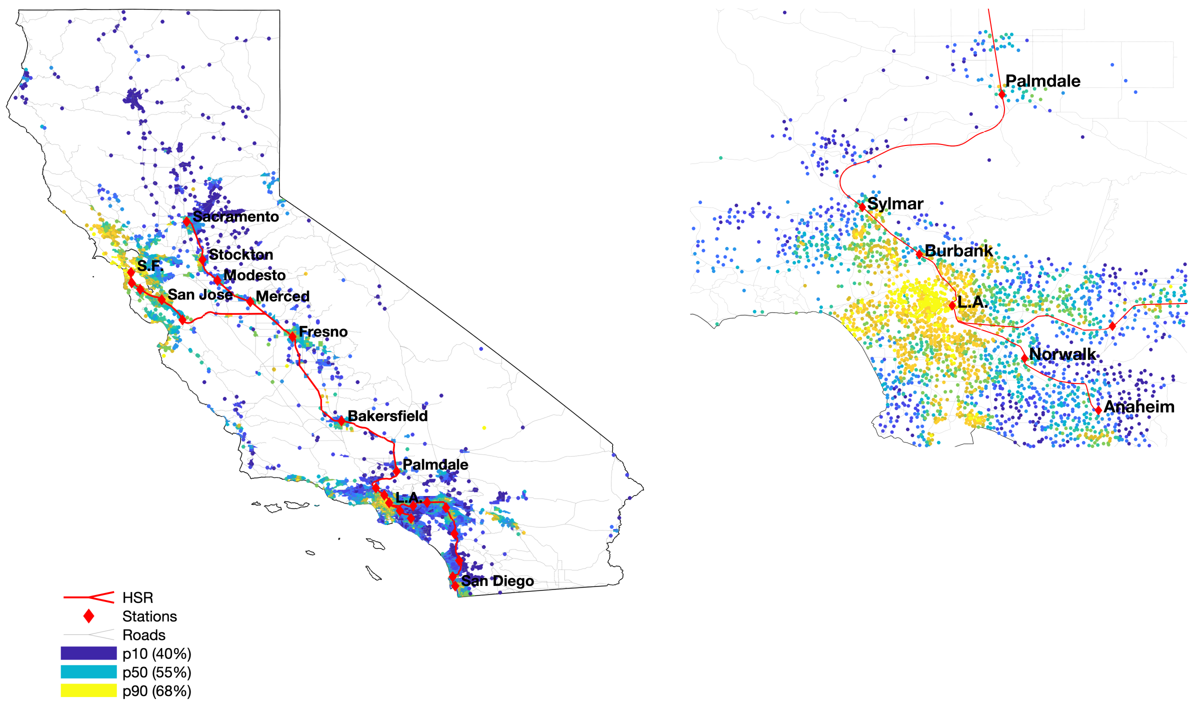
2 Background

In 1996, the California state legislature established the California High Speed Rail Authority (CAHSRA) to explore the creation of a high-speed rail network that would connect the main urban centers in northern California to those in southern California. In August 2008, the California legislature approved that Proposition 1A would appear on the ballot in the November 2008 general election. The proposition asked California voters to approve the issuance of nearly $10 billion in bonds to initiate funding for the CHSR, to be complemented by federal funding and private investors.

The CHSR project described in Proposition 1A had to satisfy several criteria. First, it had to connect San Francisco to Los Angeles and Anaheim, and include stations in Sacramento, the San Francisco Bay Area, the Central Valley, the Inland Empire, Orange County, and San Diego. Second, it had to travel at 200 miles per hour or faster, making the trip from San Francisco to

Figure 1: CHSR Route and Proposition 1A Votes

(a) California (b) LA County



Los Angeles Union Station in at most two hours and 40 minutes. Third, the maximum number of stations in the entire network was 24. Fourth, it had to be completed by 2033.

Proposition 1A was approved by 52.6% of votes. Participation amounted to 94% of voters who cast a vote for president. Figure 1 shows the share of positive votes on Proposition 1A in each census tract. Each point is the population centroid of a tract; in denser areas the entire tract is colored. Bright yellow areas were more supportive, while dark blue areas were less supportive. Broadly speaking, support was stronger in urban centers (Los Angeles, San Diego, San Francisco, San Jose, Fresno, and Sacramento) and declining in distance to the railway line. Counties of the greater San Francisco bay area (e.g., Marin and Sonoma) show clusters of strong support, while in Los Angeles the support is more concentrated in central areas.

Construction began in 2015, suffering many technical, legal, and financing troubles since then. Construction is now focused on the Central Valley segment, a 180 miles-long stretch (out of about 800 miles). The estimated total costs for Phase I, connecting San Francisco to LA, have doubled to around $100 billion in 2022 (California High Speed Rail Authority, 2022).

Voters in 2008 may have expected that the description of the CHSR project in Proposition 1A was too optimistic. We incorporate this possibility into our analysis below, by considering voters' expectations that align either with the projections in the 2008 plan or with updated projections released in 2022, which acknowledge the difficulties since the vote.

3 Framework

This section gives an overview of the theoretical framework. We outline our assumptions on voters' preferences as well as the model of the CHSR's economic impacts. We defer the discussion of the planner's problem to Section 6.

3.1 Utility and Voting

Consider a resident ω of a location i. Her utility $u_\omega(s)$ depends on whether the CHSR is approved to be built ($s = Y$ for Yes) or not ($s = N$ for No). We assume:

$$u_\omega(s) = \mathbb{E}\left[\ln W(i,s) \mid \mathcal{I}_i\right] + \ln a(i,s) + \varepsilon^u_\omega(s). \tag{1}$$

The first component, $\mathbb{E}\left[\ln W(i,s) \mid \mathcal{I}_i\right]$, is the expected real income of residents of location i in state s. The expectation is taken over the distribution of future shocks to fundamental economic characteristics in all locations, conditional on the current information set $\mathcal{I}_i$ of residents of location i. We describe the term $\ln W(i,s)$ below in Section 3.2. The second component, $\ln a(i,s)$, captures other determinants of preferences for the CHSR that are common across all residents of location i, such as political affiliation or environmental preferences. Finally, $\varepsilon^u_\omega(s)$ captures an individual-specific component preferences.

Individuals vote for the policy option s that delivers the highest utility. Aggregating over individuals, our setup corresponds to a probabilistic voting model, with the fraction of positive votes for the CHSR in location i defined as:

$$v(i) = \Pr\left[u_\omega(Y) > u_\omega(N)\right]. \tag{2}$$

We assume the idiosyncratic shocks $\varepsilon^u_\omega(s)$ are iid across residents and type-I extreme-value distributed with shape parameter θ_V:

$$\Pr\left(\varepsilon^u_\omega(s) < x\right) = e^{-e^{-\theta_V x}}. \tag{3}$$

Throughout the paper, for a generic variable $X(i,s)$, we let

$$\hat{X}(i) \equiv \frac{X(i,Y)}{X(i,N)} \tag{4}$$

denote the ratio of variable $X(i,s)$ between an equilibrium where the vote is approved and where it is not. Then, the fraction of voters in location i that support the CHSR takes the standard logit form:

$$v(i) = \frac{e^{\theta_V\left(\mathbb{E}\left[\ln \hat{W}(i)|\mathcal{I}_i\right] + \ln \hat{a}(i)\right)}}{1 + e^{\theta_V\left(\mathbb{E}\left[\ln \hat{W}(i)|\mathcal{I}_i\right] + \ln \hat{a}(i)\right)}}. \tag{5}$$

In this expression, $\hat{W}(i)$ is the real income differences in location i depending on whether the CHSR vote is approved, while $\hat{a}(i)$ measures other determinants of preferences for the project.

3.2 Quantitative Model of High-Speed Rail

We measure the economic impacts of the CHSR using as a basis a commuting model à la Ahlfeldt et al. (2015) augmented to allow for: long-distance leisure and business trips, a mode-of-travel decision with origin-specific preferences over travel modes, a monetary cost of travel, and uncertainty over project completion when calculating welfare. We provide here an overview of key forces and measurement equations, and we refer to Appendix A for a full model description. As in the previous section, $\hat{X}$ means the ratio between the value of variable i when the CHSR vote is approved and when it is not.

If the CHSR is approved, households expect to start paying right away a yearly tax t to fund the capital costs of the full CHSR project. Voters are however uncertain over the project completion: they expect the CHSR to be operational no sooner than T years after the vote, with a yearly probability of completion equal to p afterwards. Hence, if approved, the annualized net (log-) real-income impact of the CHSR is:

$$\ln \hat{W}(i) = \underbrace{(1-R)\ln(1-t)}_{\text{upfront tax}} + \underbrace{R \ln \hat{V}(i)}_{\text{net gain if completed}}, \qquad (6)$$

where $R \equiv (1+r)^{-T} \frac{p}{r+p}$ is an effective discount rate that incorporates the time discount and non-completion risk, and where $\hat{V}(i)$ captures net real income impacts of the CHSR conditional on becoming operational. The first term in (6) is an annualized expected upfront tax to be paid even before the CHSR is operational and the second term is a gain if the CHSR is completed (net of taxes that continue to be paid).

To evaluate $\hat{V}(i)$, we assume residents of a location i choose a location where to work, and also make infrequent leisure and business long-distance trips.[10] Travelers for any of these purposes (commuting, business, or leisure) perceive different destinations as imperfect substitutes. Travelers also make a choice of transport mode on each origin-destination pair, perceiving different travel modes as imperfect substitutes. Commuters choose between car, public transit, and walking/biking; and long-distance travelers choose between car, public transit, and air.

Across origins, residents may vary in their preferences for each travel mode. E.g., everything else equal, some demographic groups may have a stronger preference for traveling by car than via public transit. Data on mode usage by census tract allows us to estimate these tract-specific preferences for modes of transport. However, the lack of data on CHSR usage implies that preferences for it must be assimilated to preferences for one of the available modes. We assume that, if available, the CHSR would be perceived as a perfect substitute for public transit in the case of commuters; and as a perfect substitute for air travel in the case of business and leisure travelers. Of course, someone initially traveling by car may switch travel mode and use the CHSR when available.

[10] We let the development of the CHSR impact commuting and travel choices, but assume it does not impact residential choices. In Section 5, we show there is a significant response of voting decisions to own-tract economic outcomes, consistent with voters believing in 2008 that the economic impact of the CHSR in their location of residence will affect them in the future. An extension where workers face a constant per-period probability of migration would preserve a log-linear relationship between $\hat{W}(i)$ and $\hat{V}(i)$ as in (6), although with different structural parameters in that relationship.

To account for the possibility that workers may internalize to different extents the impact that the construction of the CHSR will have on equilibrium prices, we consider two model variants: a simple *baseline model* with fixed wages and land prices, and a more sophisticated *general equilibrium model*.

Baseline Model In our baseline model, the economic impacts of the CHSR come exclusively from changes in the time and monetary cost of traveling, and from the tax required to finance the infrastructure. Given the microfoundation in Appendix A, the annual real-income change for residents of tract i if the CHSR becomes operational, $\hat{V}(i)$, can be written as follows:

$$\hat{V}(i) = \hat{\Omega}_C(i)\,\hat{\Omega}_L(i)\,. \tag{7}$$

The component $\hat{\Omega}_C(i)$ equals the change in labor income net of commuting cost. More specifically, it captures commuters' expected time savings and changes in travel costs, and it can be written as:

$$\hat{\Omega}_C(i) \equiv \left(\sum_{j\in\mathcal{J}}\sum_{m\in\mathcal{M}_\mathcal{C}} \lambda_C(i,j,m)\left(\frac{\hat{I}(i,j,m)}{\hat{\tau}(i,j,m)^{\rho}}\right)^{\theta_C}\right)^{\frac{1}{\theta_C}}. \tag{8}$$

In this expression, $\hat{I}(i,j,m)$ is the change in disposable income (net of commuting costs and taxes) for a commuter from residence tract i to workplace tract j using transport mode m. As shown in appendix equation (A.34),

$$\hat{I}(i,j,m) = -\chi(i,j,m)\,\hat{p}_C(i,j,m) + (1+\chi(i,j,m))\,(1-t)\,\hat{y}(i,j)\,,$$

where $\chi(i,j,m)$ is the share of commuting costs in disposable income for someone traveling from i to j through mode m before the CHSR is operational, $\hat{p}_C(i,j_C,m_C)$ is the change in the monetary cost of this commuting route, t is the tax levied to finance the CHSR's capital costs, and $\hat{y}(i,j)$ is the change in pre-tax income. In this baseline model, $\hat{y}(i,j)=1$. Thus, in this baseline model, a commuter's disposable income is impacted by the CHSR through only two channels: the train ticket price and the tax levied to finance the CHSR's capital costs. In the general-equilibrium model, described below, $\hat{y}(i,j)$ includes changes in wages and in land rents.

The term $\hat{\tau}(i,j,m)$ in (8) is the change in travel time from i to j through mode m, converted into a dollar-equivalent value by the elasticity ρ. The elasticity θ_C captures the extent to which residents substitute across commuting destinations or travel modes when their relative appeals change. Finally, for each (i,j,m), the corresponding changes in time and monetary costs are weighted by the share $\lambda_C(i,j,m)$, the fraction of tract-i residents that commute to j using mode m in the equilibrium without the CHSR. These shares capture heterogeneous tract-specific preferences over transport modes and commuting destinations.

The second component of (7), $\hat{\Omega}_L(i)$, captures the economic impact of the CHSR on leisure travel. This term is also a weighted average of time and cost changes:

$$\hat{\Omega}_L(i) \equiv \left(\sum_{j\in\mathcal{J}}\sum_{m\in\mathcal{M}_L} \lambda_L(i,j,m)\left(\frac{1}{\hat{p}_L(i,j,m)\,\hat{\tau}(i,j,m)^{\rho}}\right)^{\mu_L\theta_L}\right)^{\frac{1}{\theta_L}}. \tag{9}$$

The term $\lambda_L(i, j, m)$ is the fraction of leisure travelers from i to j using mode m absent the CHSR, $\hat{p}_L(i, j, m)$ is the change in the ticket cost of travel, $\hat{\tau}(i, j, m)^\rho$ is the monetary-equivalent change in the time cost of travel, and θ_L captures how substitutable destinations and transport modes when traveling for leisure. Compared to equation (8), the expression in equation (9) accounts for the fact that not all workers travel for leisure through the share of leisure travel in total expenditure, μ_L.[11]

Equations (7) to (9) completely determine the real-income effects of the CHSR in our baseline model. In short, these reduced-form equations capture weighted averages of time- and cost- changes from the CHSR, with tract-specific weights that capture the likelihood that the CHSR will be adopted given the observed travel patterns of each tract, and with elasticities that capture the value of time and the rate at which residents substitute travel modes and destinations.

General Equilibrium Model Our general-equilibrium model (the "GE" model) nests the baseline model. In this case, the CHSR further impacts local amenities, productivities, land rents, and wages. The annual real-income change $\hat{V}(i)$ is now given by the appendix equation (A.42).

Total factor productivity can now change through two mechanisms. First, through standard spillover effects: as workers change their workplace location in response to the CHSR, worker density may change and, as result, location-specific productivities $A(i)$ change. Second, through business trips, which now enter as a shifter of total factor productivity (see Appendix A.3 for details). As a result, total factor productivity in tradable goods changes as follows:

$$\hat{\Omega}_B(i) = \hat{A}(i)\left(\sum_{j\in\mathcal{J}}\sum_{m\in\mathcal{M}_B}\lambda_B(i,j,m)\left(\frac{\hat{A}(j)}{\hat{p}_B(i,j,m)\,\hat{\tau}(i,j,m)^\rho}\right)^{\mu_B\theta_B}\right)^{\frac{1}{\theta_B}}. \tag{10}$$

Total factor productivity of a firm in i depends on productivity $A(i)$ at that location and on productivity $A(j)$ at the destination of business trips. It also depends negatively on the time and monetary cost of business trips. The term $\lambda_B(i, j, m)$ is the fraction of business travelers from i going to j using mode m absent the CHSR. In turn, local productivities impact both wages and land rents through standard market clearing. Wages impact income of workers, while land rents in tract i are capitalized into housing values and, in this way, affect the income of that tract's residents who are homeowners.

4 Distribution of CHSR's Local Economic Impacts

In this section, we explain how we estimate the parameters of the economic model described in Section 3.2. We then describe the local income effects of the CHSR predicted by the model.

[11]There are two key differences between (8) and (9). First, the change in the monetary cost of commuting travel $\hat{p}_C$ enters as a negative additive shifter in disposable income $\hat{I}(i, j, m)$ in (8), while the monetary cost of leisure travel $\hat{p}_L$ enters multiplicatively in (9). Second, the role of the monetary cost of travel is modulated by the share of spending in leisure travel, μ_L, in (9). These differences reflect the different ways in which commuting and leisure travel enter in preferences. Spending on commuting is non-homothetic: travelers spend an amount of money commuting during a fixed amount of days through the year, with their remaining income divided between consumption, housing, and leisure trips. Spending across these items is then determined according to a homothetic function, with weight μ_L on leisure trips.

4.1 Data

We conduct the analysis at the level of census tracts. Our sample covers 7,866 census tracts housing 98.5% of the statewide population. We rely on information on commuting flows from the 2006-2010 American Community Survey (ACS) (U.S. Census Bureau, 2010b) and on leisure and business trips from the California Household Travel Survey (CAHTS) (California Department of Transportation, 2012) conducted between 2010 and 2012. The CAHTS records trips longer than 50 miles over an 8-week period. To compute travel time across various transport modes, we rely on Google Maps for car and bus transit, and on official rail and air time schedules. The monetary cost of car travel is computed combining information on trip length with estimates of average driving costs per mile, while for bus, rail, and air we use information from the American Public Transportation Association (American Public Transit Association, 2010), the Bureau of Transportation Statistics (Bureau of Transportation Statistics, 2008b), and various rail operators. We construct time and monetary costs of traveling by the CHSR using information from the 2008 and 2022 CHSR business plan (California High Speed Rail Authority, 2008, 2022). We provide additional details on the data used in our analysis in Appendix B.

In addition to the estimates discussed in Section 4.2, our counterfactual predictions also rely on tract-level information on the number of residents, labor income, land-rent income, the share of floor space used for housing, and share of local landowners (see sources in Appendix B). When computing the counterfactuals with general-equilibrium effects, we borrow spillover elasticities from the literature as detailed in Appendix B. The gravity estimates and the counterfactuals are implemented at the census tract level using data from circa 2019.[12]

4.2 Gravity Estimates

The local real income effects of the high-speed rail, as determined by (7) to (10), depend on the fraction of commuters, leisure travelers, and business travelers by origin-destination and mode ($\lambda_k(i, j, m)$ for $k = C, L, B$); the substitution elasticities θ_C, θ_L, and θ_B; and the parameters ρ, μ_L, and μ_B. These parameters determine the preferences of the residents of each census tract for traveling to each other tract by different transport modes and, as a result, inform the extent to which these residents would modify their trip destinations or mode of transport were the CHSR to become available.

We summarize here our strategy to estimate these parameters and the resulting estimates, and provide additional details in Appendix D.

Commuting There is a large literature estimating preferences for destinations in commuting decisions; e.g., Monte et al. (2018), Tsivanidis (2019), and Heblich et al. (2020). The specification

[12] The fact that the model-implied local real-income impacts of the CHSR use data from 2019 implies that they rely on information that was not available to voters on occasion of the 2008 vote on Proposition 1A. Section 5 discusses how we account for this in our estimation of the weight that voters' expected real-income from the CHSR had on their preferences for the CHSR.

of the worker's commuting decision in our model, as detailed in Appendix A, enriches those baseline specifications in three dimensions. First, we allow for preferences over modes of transport that are heterogeneous across locations of residence. Second, we account for destination- and transport-mode-specific monetary costs. Third, we allow labor income in a destination to depend on the worker's place of residence, accounting in this way for heterogeneity in worker skill composition across origin tracts. These three elements impact the likelihood that residents of a particular census tract would use the CHSR for commuting, were it to become available.

To estimate the parameter vector (θ_C, ρ) and the commuting share $\lambda_C(i,j,m)$ for every origin, destination, and transport mode, we rely on the following relationship:[13]

$$\lambda_C(i,j,m) = \frac{\left(\frac{I(i,j,m)}{D_C(i,m)}\right)^{\theta_C} \tau(i,j,m)^{-\theta_C \rho}}{\sum_{j' \in \mathcal{J}} \sum_{m' \in \mathcal{M}_C} \left(\frac{I(i,j',m')}{D_C(i,m')}\right)^{\theta_C} \tau(i,j',m')^{-\theta_C \rho}}, \tag{11}$$

where $I(i,j,m)$ is the disposable income of someone commuting from i to j by transport mode m, $\tau(i,j,m)$ denotes the travel time between locations i and j by mode m, and $D_C(i,m)$ is a preference shifter for transport mode m specific to commuters residing in location i. We consider three feasible modes of transport for commuting: $\mathcal{M}_C = \{\text{private vehicle, public transport, walk/bike}\}$.

We perform the estimation in two steps. In the first step, we estimate θ_C and ρ using a Generalized Method of Moments (GMM) estimator that exploits variation in the choice of destination conditional on origin and transport mode. The estimates of θ_C and ρ are robust to assumptions on the taste shifters $D_C(i,m)$, which vary exclusively by origin and mode. Our estimates are $\hat{\theta_C} = 3.35$ (with robust standard error equal to 0.10) and $\hat{\rho} = 0.21$ (robust s.e. equal to 0.006). This last parameter captures the percentage increase in wages in a destination that would leave workers indifferent if commuting time were to increase by one percent. The estimate of ρ indicates, for example, that workers would require a 10% higher wage to accept a job with 50% longer commute, an approximately 25-minute increase for the average commuter.

In the second step, we estimate $D_C(i,m)$ for all origin census tracts and modes of transport. We model $D_C(i,m)$ as a function of observed origin-specific demographic covariates, $X_C(i)$, with mode-specific coefficients $\Psi_C(m)$. We estimate these coefficients using again a GMM estimator, relying on observed variation across origins with different demographics $X_C(i)$ in their use of different transport modes.

We present these estimates in Appendix Table A.1. The preferences over transport mode differ across census tracts; for example, tracts with larger shares of younger and more educated workers have a weaker preference for commuting by car, and tracts with larger shares of nonwhite residents have a stronger preference for public transport. These values of $D_C(i,m)$ impact the usage of modes of transport in addition to the impact that commuting times and monetary costs may have.

Combining $\left(\hat{\theta}_C, \hat{\rho}\right)$ and the estimates of $D_C(i,m)$ with the expression in (11), we generate model-predicted commuting shares $\lambda_C(i,j,m)$ for all origin-destination pairs and transport modes. We then use these shares in (8) to quantify part of $\hat{W}(i)$.[14]

[13] This expression and (A) correspond to an equilibrium without the CHSR as defined in Appendix A.

[14] Dingel and Tintelnot (2020) recommend using model-implied predicted shares rather than directly observed

Business and Leisure Travel The model-predicted share of business or leisure trips with origin in census tract i and destination in a census tract j that use a mode of transit m is:

$$\tilde{\lambda}_k(i,j,m) = \frac{\left(\frac{Z_k(i,j)}{D_k(i,m)}\right)^{\mu_k\theta_k} \tau(i,j,m)^{-\rho\mu_k\theta_k} p_k(i,j,m)^{-\mu_k\theta_k-1}}{\sum_{j\in\mathcal{J}}\sum_{m\in\mathcal{M}_k}\left(\frac{Z_k(i,j)}{D_k(i,m)}\right)^{\mu_k\theta_k} \tau(i,j,m)^{\rho\mu_k\theta_k} p_k(i,j,m)^{-\mu_k\theta_k-1}} \quad (12)$$

for $k = L$ (leisure) or $k = B$ (business). The shifter $Z_k(i,j)$ captures the leisure or business appeal of destination tract j among travelers from census tract i; $D_k(i,m)$ is a preference shifter for using transport mode m among travelers for purpose k from i; $\tau(i,j,m)$ denotes the travel time between locations i and j by mode m; and $p_k(i,j,m)$ is the monetary cost per round trip. For both leisure and business travel, we consider three feasible modes of transport: $\mathcal{M}_B = \mathcal{M}_L = \{\text{private vehicle, public transport, airplane}\}$. Our procedure to estimate the parameters in (12) implements a two-step GMM estimator similar to that used to estimate the parameters in (11).

In the first step, we condition on the estimate of ρ and estimate θ_k and μ_k for $k = L, B$. For leisure travel, our estimate of the coefficient on log travel time is $\mu_L\theta_L\rho = 0.46$ (robust s.e. equal to 0.14). The estimate of the analogous coefficient in the case of business travel is $\mu_B\theta_B\rho = 0.59$ (robust s.e. equal to 0.22).[15] Our estimates of the commuting equation described above imply a coefficient on log travel time of $\theta_B\rho = 0.71$, reflecting a higher disutility of time spent traveling in the case of commuting trips than in the case of leisure and business trips.

As the expression in (12) illustrates, μ_k and θ_k are not separately identified. We thus calibrate μ_k using external data sources. For leisure travel, we set $\mu_L = 0.05$, consistently with BLS information on U.S. households' annual share of spending on travel, including transportation, food away from home, and lodging.[16] For business travel, μ_B equals the share of the firm's value added spent on its employees' business travel. We set $\mu_B = 0.015$ following industry reports.[17] Given these calibrated values of μ_L and μ_B and our estimate of ρ, the coefficients on log travel time imply estimates of θ_L and θ_B equal to 43.8 (robust s.e. equal to 13.4) and 185.6 (robust s.e. equal to 68.2), respectively. Thus, travelers perceive business and leisure destinations as highly substitutable.

In the second step, we estimate $Z_k(i,j)$ for all origin and destination census tracts by modeling these terms as a function of observed origin- and destination-specific covariates and a vector of coefficients on those covariates. Using the estimates of θ_k and ρ, the calibrated value of μ_k, the estimates of $Z_k(i,j)$ for all origin and destination census tracts, and the expression in (12), we generate model-predicted shares of long-distance trips $\tilde{\lambda}_k(i,j,m)$, for $k = L, B$ for all origin-destination pairs and transport modes.

shares to compute the implications of quantitative spatial models, in particular in sparse settings like ours.

[15] This standard error uses information on monetary costs to measure $p_k(i,j,m)$. As described in Appendix C, these monetary costs are by construction highly correlated with travel time, generating collinearity. If we exclude $p_k(i,j,m)$ from our estimation, we obtain similar coefficients on log-travel time ($\mu_B\theta_B\rho$=0.55) but a much smaller robust standard error (equal to 0.04).

[16] The US Travel Association reports leisure travel spending in the ballpark of 800 billion USD for 2019 (U.S. Travel Association 2020 Answer Sheet), which as a share of that year's US private consumption expenditure of 14,400 billion (FRED) yields a similar share of 5.6%.

[17] The US Travel Association and the Global Business Travel Association both report US business travel spending in the ballpark of 340 billion USD in 2019, which corresponds to about 1.5% of US GDP in that year.

4.3 Alternative CHSR Scenarios

When computing the counterfactual real-income impact of the CHSR, we consider alternative scenarios in terms of the possible timeline and costs of the CHSR and in terms of economic forces included in the model, summarized in Table A.2.

First, we consider a "2008 Business Plan" scenario which uses the information on the project published at the time of the vote (California High Speed Rail Authority, 2008). According to this plan, the full project would be operational by 2030, tickets would be set at 50% of the typical airfare, and the present value of the total capital cost of implementing Phase-I of the project (linking San Francisco to Los Angeles) would be $33 billion. Second, we consider a "2022 Business Plan" scenario, which uses the information from the 2022 updated business plan (California High Speed Rail Authority, 2022). According to this alternative plan, the full project would be completed in about double the number of years than announced in the 2008 plan, ticket prices are also doubled, and the present value of the capital costs for Phase I increase to at least $77 billion; furthermore, the probability that the forecasts of this alternative plan are satisfied is set to only 65%.[18]

For each of these two scenarios, we compute counterfactual real-income impacts using both the baseline and the general-equilibrium models described in Section 3.2, using the system (A.32)-(A.41) in Appendix A.8.

4.4 Impact of CHSR on Travel Time and Costs

An important input to determining the counterfactual impact of the CHSR is the change in travel times and monetary costs caused by the introduction of the CHSR. Appendix C details the construction of the travel times and monetary costs in the settings with and without the CHSR. In short, we calibrate a multi-modal transport network for California (including travel by car, air, bus or rail, and bike) to match observed travel times from the data. We then allow travelers to use the CHSR, assuming that it is used only when it is beneficial in utility terms, given the time gain and ticket price, as described in Appendix A.6.

The 2008 and 2022 CHSR business plans promised sizable potential travel time reductions. To gauge the size of the shock, consider potential CHSR use among "directly impacted" travelers: those originally using public transit (either bus or rail) when commuting, or those originally using public transit or air travel for long-distance trips. Of course, as the estimated model indicates, the CHSR may also draw travelers by altering their mode of transport (e.g., away from using the car) or the destination of their commuting, leisure, or business trips.

As illustrated in Table 1, 1.4% of all commuters, 2.3% of all long-distance leisure travelers, and 9.7% of all long-distance business travelers travel (in the observed equilibrium without the CHSR) on routes where the CHSR would be utility improving according to the model. Among these directly

[18]As Phase-I was about 60% of the full project, we set the cost of the total project by adjusting the capital costs proportionally to the length of the full network. The 2022 plan expects that, by 2030, only the Central-Valley segment (180 miles) would have been completed; to obtain an expected completion date for the whole project, we extrapolate proportionally from this initial deadline assuming a constant time until completion per mile.

Table 1: The CHSR Shock

	% Initial Travelers Directly Better Off	Time Gain		Cost Change 2008 Business Plan		Cost Change 2022 Business Plan	
	Pub. Trans. or Air	median	75 ptile	med	75p	med	75p
Commute	1.4%	34' (38%)	53' (48%)	-21%	-9%	3%	17%
Leisure	2.3%	21' (11%)	47' (32%)	-60%	-50%	-28%	-13%
Business	9.7%	12' (5%)	26' (12%)	-63%	-57%	-31%	-21%

Note: The first column shows the fraction of all travelers within each travel purpose who, before the CHSR becomes available, travels on routes where the CHSR is utility-improving (considering both time savings and monetary costs) and therefore used when available, assuming that the CHSR may only directly replace public transit or air. The remaining columns show moments from the traveler-weighted distribution of time and cost changes across the origin-destination-modes within each travel purpose where the CHSR is used when available.

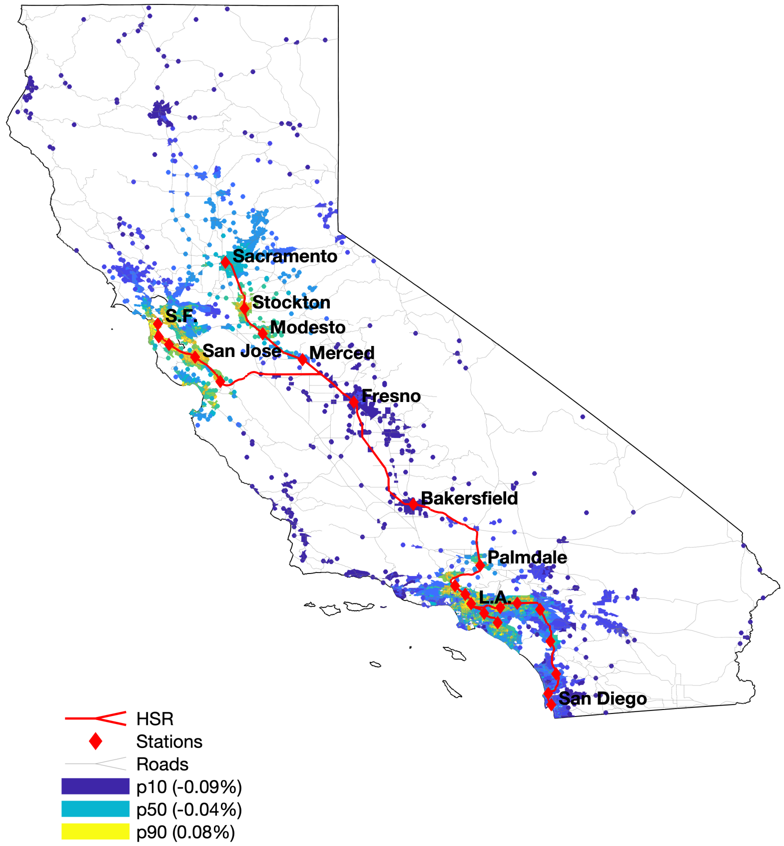
impacted travelers, the reductions in travel time are substantial. For example, the median time gain for a commuter via public transit is 38%. The monetary gains disappear for commuters with the higher projected ticket prices updated in the 2022 Business Plan, but these updated forecasts still predict large time and pecuniary gains for leisure and business travelers. The difference in the predicted monetary gains between commuters and long-distance travelers comes from the former using the relatively cheap public transit option and the latter using the more expensive air travel option.

4.5 Distribution of the CHSR Impacts across Census Tracts

We now compute the distribution across census tracts of the expected real-income effects of the CHSR, $\hat{W}(i)$, as defined in (6). Figure 2 plots real-income effects in the baseline economic model and the predictions of the 2008 Business Plan, and Figure 3 zooms in on the tracts in the San Francisco Bay Area and the Los Angeles county. Bright yellow tracts gain the most, and dark blue tracts lose the most. The effects are heterogeneous, with a small share of winning tracts. The top 10% of tracts experience real-income gains between 0.1% and 1.1% per year, while 70.3% of all tracts lose. When using instead the 2022 Business Plan, our baseline model predicts that 99.5% of all tracts lose. Table 2 describes the distribution of $\hat{W}(i)$ across model versions.

The maps show an intuitive gradient of gains as a function of distance to the stations. However, distance is not the only determinant of gains from CHSR adoption; for example, travel patterns and preferences over (or ease of access to) different travel modes also matter, and they enter in our estimates through the gravity estimation. Appendix Table A.3 illustrates these determinants by showing regressions of the tract-level real-income changes on several covariates. The regressions confirm that tracts closer to stations gain more. However, tracts with a higher percentage of commuters by public transit or of long-distance travelers by air in the initial equilibrium also gain

Figure 2: Spatial Distribution of Real-Income Gains from the CHSR



more, as CHSR stations tend to be close to standard rail stations and airports.[19] In addition, tracts located in Los Angeles county gain more than what the regression covariates would predict. Central Valley locations like Fresno or Bakersfield gain less and the San Francisco Bay Area gains more, but only if general-equilibrium effects are not accounted for. These location-specific results illustrate the importance of accounting for the specific travel patterns of different regions of California.

The distribution in space of the real-income gains of the CHSR is very similar across all model variants, with a correlation across census tracts that is greater of 90% across any two variants. I.e., the ranking of relative winners and losers is approximately the same, regardless of whether we incorporate general-equilibrium forces in the model, and regardless of whether we model the CHSR according to the 2008 or the 2022 projected costs, travel speed, and ticket prices. However, the level of the implied gains varies significantly depending on whether we incorporate in our analysis the predictions of the 2008 Business Plan or those of the 2022 Plan. In our baseline model (without GE effects), we obtain a population-weighted loss of -0.01% using the projections in the 2008 Plan, and a much larger aggregate loss of -0.31% using the projections in the 2022 Plan, as shown in the last column of Table 2.[20]

[19] Appendix Figure A.1 shows the spatial distribution of CHSR stations alongside rail stations (which belong to the public transport network) and airports.

[20] Using different methods than in our analysis, initial 2008 estimates by the High-Speed Rail Authority (California

Figure 3: Spatial Distribution of Real-Income Gains from the CHSR (SF and LA)

(a) SF Bay Area (b) LA County

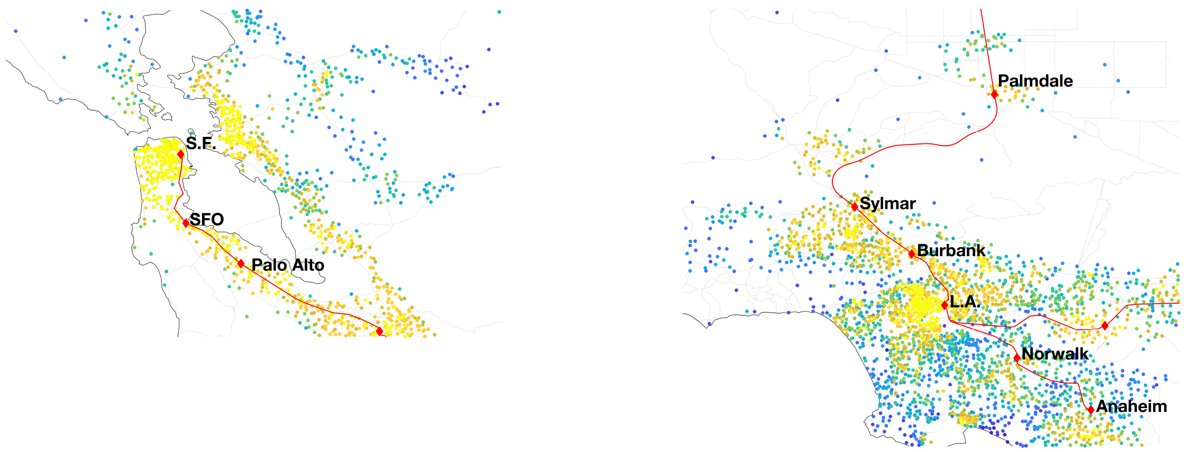


Table 2: Distributional Impacts from the CHSR

	Moments of $\hat{W}(i)$			Pop-Weighted Avg	
	p10	p50	p90	Gross	Net
2008 Business Plan	-0.09%	-0.04%	0.08%	0.13%	-0.01%
2008 Business Plan, with GE	0.01%	0.10%	0.27%	0.25%	0.13%
2022 Business Plan	-0.34%	-0.32%	-0.27%	0.04%	-0.31%
2022 Business Plan, with GE	-0.32%	-0.29%	-0.27%	0.05%	-0.28%

Note: The first three columns reports moments from the annual real-income gain $\hat{W}(i)$ defined in (6). The last two columns are population-weighted average of $\hat{W}(i)$ across census tracts. The next to last column reports the weighted average before capital costs (setting the lump-sum tax t to zero in our calculations). The last column includes these costs.

Appendix table A.4 further decomposes these aggregate effects into those stemming from the upfront tax and from the commuting, leisure travel, and business trips components defined in conditions (6)-(10). We find a substantial role for leisure and business trips in driving the aggregate impacts, with the gross benefits from leisure and business travel being at least as large as those from commuting across all cases.

High Speed Rail Authority, 2008) estimated net present-discounted gains of $97bn in 2008 USD. In our baseline, the analog number is $-2.4bn. The project update to the 2022 CHSR business plan (California High Speed Rail Authority, 2023) includes a benefit-cost analysis such that, if only "high-speed rail user benefits" are included (corresponding to the forces we include in the baseline model), the phase-I of the CHSR (from San Francisco to Los Angeles and Anaheim) leads to a present-discounted *loss* of $-15bn in 2021 dollars. Our 2022 scenario without GE, implemented on the full CHSR (instead of just phase-I), yields a loss of $-136.5bn in 2021 USD. The 2023 CHSR update also reports gains of $26bn for only phase I when "wider economic benefits for worker and firms" are further included in addition to rail user benefits. Our GE estimate for the 2022 case yields a loss of $-123.1bn for the full network.

5 Effect of Local Economic Impacts on Votes

Armed with the estimated model of the real-income gains of the CHSR, we now estimate the parameter θ_V in (5), which determines the impact of expected real-income in shaping voters' preferences for the CHSR.

5.1 Estimation Strategy

When estimating voters' response to expected real income in the 2008 referendum, two potential identification issues arise. First, we may measure voters' expected real-income gain from the CHSR with error. Second, there may be other determinants of voters' CHSR preferences that are correlated with their expected real-income gain from the CHSR. While the first is an instance of measurement error in the covariate of interest, the second is an example of omitted variables correlated with that covariate. We discuss here these identification concerns and the strategies we follow to address them.

Measurement Error We do not observe voters' expectations of the CHSR's real income impact, $\mathbb{E}\left[\ln \hat{W}(i) \mid \mathcal{I}_i\right]$ in (5). Instead, we identify θ_V as the coefficient on the model-implied ex-post real income impact of the CHSR, which we thus use as proxy for voters' expectations; i.e., our proxy is $\ln \hat{W}(i)$, as defined in (6) and computed using the estimates described in Section 4.5. Because this proxy is constructed using information on realized fundamentals (such as productivities) from 2019, we denote it in this section by $\hat{W}_{19}(i)$. Without loss of generality, differences between voters' expectations and this proxy may be decomposed into two terms: (a) voters' expectational error, denoted by $\epsilon_{W,1}(i)$; and, (b) model misspecification, i.e., any mismatch between the true ex-post CHSR impact and that predicted by our model, denoted by $\epsilon_{W,2}(i)$. Thus, we can write:

$$\mathbb{E}\left[\ln \hat{W}(i) \mid \mathcal{I}_i\right] = \ln \hat{W}_{19}(i) - \epsilon_{W,1}(i) - \epsilon_{W,2}(i). \quad (13)$$

If voters' expectations are rational, the error $\epsilon_{W,1}(i)$ is analogous to classical measurement error, biasing OLS estimates of θ_V towards zero. Any IV estimator that uses as instrument a variable that belongs to voters' information set at the time of the vote avoids this source of bias (Dickstein and Morales, 2018). We present below estimates that use as such instrument the model-implied real-income impact of the CHSR constructed using fundamentals from 2008, denoted by $\ln \hat{W}_{08}(i)$. If voters' expectations are rational and the 2008 fundamentals belong to their information sets, this instrument will be mean independent of the expectational error $\epsilon_{W,1}(i)$. If, furthermore, our model were to correctly capture the ex post economic impact of the CHSR, then the resulting IV estimator of θ_V would be unaffected by measurement error.

Importantly, rationality of expectations is sufficient but not necessary for our IV estimator of θ_V to avoid measurement error bias. As discussed below, when estimating θ_V, we condition on a set of controls, one of them being the share of registered Democrats in each census tract. Thus, our IV estimator of θ_V is not affected by voters' expectational errors even if voters have biased beliefs about the real income impact of the CHSR, as long as the bias in expectations aligns with their party affiliation.

Characterizing the bias due to the model misspecification error $\epsilon_{W,2}(i)$ is more complicated, as it depends on which particular model aspect is inaccurate. We address this potential bias by exploring how robust our estimates of θ_V are to the alternative economic models described in Table A.2. As a reminder, these models differ in the projected CHSR capital costs (they either use the projections in the 2008 or the 2022 Business Plan) and in the economic forces incorporated in the model (they either abstract from or incorporate general-equilibrium effects).

Omitted Variables The determinants of voters' preferences entering through the component $\hat{a}(i)$ in (5) may be correlated with voters' expectations of the CHSR's real income impact. To limit the risk that our estimates of θ_V suffer from omitted variable bias, we introduce proxies for subjective considerations that may play a role in voters' preferences for the CHSR. Formally, we assume that:

$$\ln \hat{a}(i) \equiv \sum_{k=1}^{K} \tilde{\beta}_k X_k(i) + \epsilon_a(i). \tag{14}$$

We introduce three sets of covariates $X_k(i)$. The first set proxies for voters' ideology: it includes the tract-specific share of registered Democrats and the tract-specific vote shares in favor of two propositions, Prop. 10 and Prop. 1B, on clean energy and transportation projects, respectively.[21] We interpret the vote share in support of Prop. 10 as a proxy for voters' environmental concerns, and the vote share in support of Prop. 1B as a proxy for voters' willingness to back transportation infrastructure spending in general. The second set of covariates measures demographic characteristics: the tract-specific share of residents who are nonwhite, college-educated, or under 30. Finally, the third set of covariates measures the time it would take voters in each tract to reach the closest CHSR station. This distance measure accounts for any correlation between the location of the CHSR stations and voters' CHSR preferences that is not captured by the two other sets of covariates.[22] In addition to these covariates, we control in all specifications for county fixed effects. Consequently, the identification of θ_V is based on variation across census tracts within counties.

We conceptualize all these political, demographic, and proximity covariates, as well as the county fixed-effects, as controls that help us obtain consistent estimates of θ_V. The interpretation of the coefficients on these covariates is not relevant for our purpose; those coefficients may capture the impact of multiple treatments and, thus, may have multiple valid interpretations. For example, the share of registered Democrats in a tract may capture both differences in CHSR preferences

[21] Prop. 10 ("Bonds for Alternative Fuels Initiative"), on the ballot at the same time as the CHSR, asked voters to authorize the government of CA to issue \$5 billion in bonds for alternative fuel projects (it failed with 40.6% of votes in favor). Prop. 1B ("Transportation Bond Measure"), on the ballot in 2006, asked voters to authorize the government of CA to issue \$19.9 billion in bonds for transportation projects (it passed with 61.4% of votes in favor). Source: *ballotpedia.org*.

[22] Note that $\ln \hat{W}(i)$ accounts for preferences for using different means of transport (car, public transit, airplane, or biking), as described in Section 4.2. Thus, the measure of distance to the projected CHSR stations included among the set of controls captures any effect on CHSR preferences of being close to CHSR stations other than the potential usage of the CHSR were it to become available, which enters through $\ln \hat{W}(i)$.

driven by political ideology and differences in beliefs about the future real income impact of the CHSR.

Instrument #1: Random Stations The covariates $X_k(i)$ described above, together with $\ln \hat{W}_{19}(i)$, account for a large share of the variation in vote shares across locations.[23] All determinants of voters' CHSR preferences not controlled for by these covariates are accounted for by the term $\epsilon_a(i)$ in (14). If these omitted variables are correlated with the model-implied CHSR income impact, $\hat{W}_{19}(i)$, the OLS estimate of θ_V will be biased. If they are also correlated with the model-implied impact when fundamentals are set to their 2008 values, the TSLS estimate of θ_V that uses $\ln \hat{W}_{08}(i)$ as an instrument will also be biased.

Thus, we also present TSLS estimates of θ_V that rely on an alternative instrument that we build as follows: in a first step, we simulate one hundred *counterfactual* CHSR networks by randomizing the location of 24 stations along the projected CHSR railway line; in a second step, for each of these counterfactual networks and all census tracts, we compute the associated model-implied real-income change using 2008 fundamentals; finally, in a third step, we compute the average of these real income changes across the one hundred simulated networks. More specifically, denoting the instrument by $\hat{W}_{08}^{IV}(i)$, we build it as:

$$\hat{W}_{08}^{IV}(i) = \frac{1}{100}\sum_{n=1}^{100} \ln\left(\hat{W}_{08}^{cf}(i,n)\right), \tag{15}$$

where $\hat{W}_{08}^{cf}(i,n)$ is the model-predicted change in welfare in location i from the counterfactual CHSR design n built using 2008 fundamentals.

Instrument #2: Random Stations and Random Paths We construct a second instrument where, instead of randomizing the location of stations along the projected CHSR railway line, we use information on three alternative railway lines that were considered in the early stages of the CHSR design process (US DOT, 2005); see Appendix Figure A.2. Relevant for the validity of this second instrument, these three alternative CHSR routes were selected as candidates to link Los Angeles to San Francisco on the basis of technical feasibility and cost savings.[24] Thus, political considerations may have determined the final choice between the three alternative routes (and the location of the 24 stations), but they were not behind the decision to limit the set of possible routes to these three alternative railway lines. Hence, our second instrument uses the formula described in (15), where each counterfactual CHSR design n is determined by first drawing randomly one of these three potential routes, and then randomly locating the 24 stations along that route.

As these two instruments incorporate information on neither the actual location of the 24 projected stations (in the case of the first instrument) nor on the actually projected CHSR railway

[23] The R-squared of the OLS estimated regression is above 0.9 in the specification with the largest set of covariates.

[24] There are two possible routes from Northern to Southern California: along the coast or through the center of the state (along the I-5 highway or via the Central Valley). The topology of the coastal area makes it more difficult for trains to reach top speed and is costlier in terms of building. Along the Interstate 5, the terrain is flatter and construction is cheaper but the stations are farther away from population centers. The third option, via the Central Valley, ranks in between these alternatives in terms of both proximity to population and expected costs.

line (in the case of the second instrument), they may be valid even if the location of the CHSR stations (or the railway line, in the case of the second instrument) had been determined with the goal of favoring tracts with systematically larger or smaller values of the unobserved term $\epsilon_a(i)$. Moreover, as the instruments only use information on 2008 fundamentals, they will be mean independent of the expectational error $\epsilon_W(i)$ under the conditions discussed above.

5.2 Estimates of Voter Preferences

Combining (14) with (13) and (5), we obtain our estimating equation:

$$\ln\left(\frac{v(i)}{1-v(i)}\right) = \theta_V \ln \hat{W}(i) + \sum_{k=1}^{K} \beta_k X_k(i) + \epsilon(i), \tag{16}$$

where $\beta_k \equiv \theta_V \tilde{\beta}_k$ and where $\epsilon(i) \equiv \theta_V(\epsilon_a(i) - \epsilon_{W,1}(i) - \epsilon_{W,2}(i))$ includes unobserved components of voters' preferences, measurement error, and model misspecification error. In Table 3, we report OLS and TSLS estimates of θ_V and the β_k's. In this table, we compute the real-income impact of the CHSR using the economic model without general-equilibrium effects and the cost projections from the 2008 Business Plan (we discuss alternative specifications below). In the first four columns, we present OLS estimates for regression specifications that progressively account for a larger set of regressors. In column (1), we include no covariate other than the model-implied welfare impact of the CHSR. In column (2), we add proxies for political ideology; in column (3), we add variables that capture the demographic composition of each census tract's residents; and, in column (4), we add as a control the log of the shortest time distance from each tract to its closest CHSR station.

The OLS estimated value of θ_V equals 64.40 (with robust s.e. equal to 3.77) in the specification (4) with the largest set of controls. This estimate of θ_V is robust to whether we control for voters' demographic characteristics or by the distance to the closest HSR station. The estimates of the coefficients on the various controls reveal that a larger support for the CHSR is predicted by the following census tract characteristics: a larger share of registered Democrats; a larger support for Prop. 10 (support for alternative fuel vehicles); a larger support for Prop. 1b (support for transportation projects); a larger share of residents who are white, college-educated, or under 30 years of age; and proximity to a CHSR station.

Columns (5) to (7) report IV estimates. Column (5) uses as instrument the 2008-based real income measure, column (6) uses the instrument that additionally randomizes the location of the stations alone, and column (7) uses the instrument that additionally randomizes the location of both the railway line. For the three instruments we obtain large first-stage F-statistics.

The fact that the IV estimate of θ_V in column (5) that uses as instrument the 2008-based real income measure is larger than the OLS estimate (72.45 vs. 64.40) is consistent with the latter being downward biased due to voters' expectational errors. The fact that the IV estimates of θ_V computed using both the random-station and the random-path instruments in columns (6) and (7) are smaller than that reported in column (5) reveals that the CHSR design favored census tracts whose residents were, for unobserved reasons, more predisposed to support the proposed CHSR.

Table 3: Estimates of Voting Equation

Inst. Var.:	None - OLS				$\ln(\hat{W}_{08})$	Random Station	Random Path
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$log(\hat{W}_{19})$	159.73[a]	72.35[a]	65.37[a]	64.40[a]	72.45[a]	66.97[a]	63.68[a]
	(6.11)	(4.46)	(3.58)	(3.77)	(4.49)	(4.54)	(5.29)
Log-odds Dem. Sh.		0.30[a]	0.38[a]	0.38[a]	0.38[a]	0.38[a]	0.38[a]
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Environ.: Prop. 10		1.13[a]	2.41[a]	2.41[a]	2.40[a]	2.41[a]	2.41[a]
		(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Transp.: Prop. 1b		1.54[a]	0.81[a]	0.81[a]	0.80[a]	0.80[a]	0.81[a]
		(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Sh. non-White			-0.17[a]	-0.17[a]	-0.17[a]	-0.17[a]	-0.17[a]
			(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Sh. College			0.75[a]	0.74[a]	0.74[a]	0.74[a]	0.74[a]
			(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Sh. Under 30			0.19[a]	0.19[a]	0.19[a]	0.19[a]	0.18[a]
			(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Log. Dist. Station				-0.00	-0.00	-0.00	-0.00
				(0.00)	(0.00)	(0.00)	(0.00)
F-stat					1783	966	850
Num. Obs.	7861	7861	7861	7861	7861	7861	7861

Note: [a] denotes 1% significance level. Robust standard errors in parenthesis. All specifications control for county fixed effects.

Results across Model Variants Appendix Table A.5 presents analogous estimates to those in columns (6) and (7) in Table 3, but using the alternative model variants described in Table A.2. The estimates of θ_V are larger under the costs and timeline projections from the 2022 Business Plan than under the projections in the 2008 Plan. Intuitively, this happens because there is a high correlation across tracts (greater than 91.2%) in the economic impacts predicted by the two model variants, but the variance of these tract-specific predictions is smaller when using 2022 information, resulting in a larger θ_V.[25] Conversely, the estimates of θ_V are smaller when the CHSR real-income impacts are computed using the model that incorporates general-equilibrium impacts; that model yields a distribution of $\ln \hat{W}(i)$ that is very correlated with that predicted by the model without general equilibrium effects, but more dispersed, resulting in a smaller estimate of θ_V.

Robustness Our estimates are robust to a range of alternative weighting schemes and sample selection criteria. Appendix Table A.6 presents estimates analogous to those in columns (6) and (7) in Table 3 but for specifications that weight the results in each tract by either the total number of votes (in columns (3) and (4)) or by the participation rate (in columns (5) and (6)). When weighting by the total number of votes, we obtain IV estimates of θ_V that are larger than the

[25] The belief p that the CHSR will be successfully completed enters the calculation of $\ln \hat{W}(i)$ by multiplying the per-period economic effects $\ln \hat{V}(i)$, as shown in equation (6). Hence, a lower p, reflecting the more pessimistic beliefs included in the 2022 Business Plan, leads to lower variance of $\hat{W}(i)$.

baseline (unweighted) estimates. Conversely, weighting the results by the participation rate yields estimates very similar to those of the unweighted estimator. Columns (7) and (8) in Appendix Table A.6 show that the estimates of θ_V become somewhat larger when we redo our estimation dropping the nearly 3,000 census tracts that are less than 5km away from the railway line.

5.3 Implications

The estimates imply we cannot reject the joint hypothesis that voters in the 2008 referendum cared about the expected real income impact of the CHSR, and that this variable is correlated with our model predictions. Thus, they provide support for our estimated model as a predictor of the spatial distribution of the real-income effects of the CHSR.

The estimates further illustrate that economic voting was a driver of policy preferences over the CHSR. Economic voting is a relevant driver of votes in electoral politics (Lewis-Beck and Stegmaier, 2000). However, when it comes to preferences over transport policies, case studies and survey evidence often point to low public support for projects with seemingly positive net economic impacts.[26] In contrast, we show that changes in the expected economic impact of the CHSR had a significant effect on voters' preferences.

Our estimates also determine the additional real income needed to sway votes. Consider an alternative CHSR design that would have increased expected economic gains in location i by $d\hat{W}(i)$ percentage points, compared to $\hat{W}(i)$ in the actual CHSR design. This change would have affected the favorable votes in location i by

$$dv(i) = \frac{\theta_V v(i)(1 - v(i))}{\hat{W}(i)} d\hat{W}(i) \tag{17}$$

percentage points. Table 4 uses this formula to compute the real income change needed to sway an extra percentage point of votes in favor of the CHSR. Across model variants, 0.03 to 0.10 extra percentage points in economic gains swayed one percentage point of votes in the median tract. To put this number in perspective, the last column of Table 4 reports the standard deviation of income

Table 4: Effect of Local Economic Impacts on the Vote

Model Variant	θ_V	Income Gain to Sway 1 pp (median tract)	Standard Deviation of Income Gain
2008 Business Plan	63.7	0.06%	0.12%
2008 Business Plan, with GE	43.2	0.10%	0.14%
2022 Business Plan	159.3	0.03%	0.05%
2022 Business Plan, with GE	96.7	0.04%	0.05%

Note: The θ_V corresponds to the IV estimate using the random routes instrument shown in Table A.5 for each model variant. The second column reports the 50th percentile of the distribution across tracts of the inverse of $dv(i)/d\hat{W}(i)$ in (17), which measures the extra gains $\hat{W}(i)$ required to sway 1% of votes around the equilibrium with the CHSR. The last column reports the standard deviation of the income gain $\hat{W}(i)$ across tracts.

[26] An example is road pricing; see Verhoef et al. (1997), Schade and Schlag (2003), and Noordegraaf et al. (2014).

gains across tracts corresponding to each model variant. These numbers imply that, across model variants, one standard deviation in income gains swayed between 1.2 and 2.0 percentage points of local votes.

Given the economic significance of these magnitudes, we conjecture that policymakers could have decided to allocate stations based on their impact on public support, at the expense of locations where stations were socially more desirable. We quantify this trade-off in the next section.

6 Policymakers' Preferences and Optimal Designs

We consider the problem of a planner who decides where to locate the CHSR stations. This problem captures, in reduced form, a complex process leading up to the referendum involving politicians and technical experts. Rather than modeling this process, we adopt an "as if" representation in which a hypothetical planner holds preferences over the welfare of different residents and over votes. We estimate the preferences of such social planner, identifying as a result the distributional biases emerging from this process.[27]

To set up the planner problem, one could follow an approach such as Fajgelbaum and Schaal (2020) or Kreindler et al. (2023), where the planner designs the entire network with complete flexibility. In our case, this would amount to deciding both the location of the railway lines and of the stations. However, doing so would raise methodological challenges due to dimensionality. It would also force us to model and quantify the potentially heterogeneous costs of building rail lines at different locations.

We can, instead, leverage that only a few alternative routes were identified as being feasible on the basis of cost and engineering considerations. More specifically, the CHSR was originally meant to connect Los Angeles and San Francisco, further branching out to San Diego and Sacramento in later phases. As discussed in Section 5.1, three possible routes linking Northern and Southern California were originally identified (US DOT, 2005): one ran along the coast, and the other two ran through the center of the state, either along the I-5 highway or through the Central Valley (see Appendix Figure A.2). The coastal route was subject to much higher costs, leaving the I-5 and the Central Valley routes as the two main options for policymakers to choose from; conditional on a route, policymakers had to decide the location of stations.

In the upcoming sections, we use observed distribution of stations along the actual CHSR route to estimate planners' preferences, and then compute counterfactual optimal distributions under alternative preferences along both the actual CHSR route and the main alternative along the I-5. Section 6.1 sets up the planner's problem, Section 6.2 discusses how we estimate the planner's preferences, and Section 6.3 solves for optimal station allocations with counterfactual planner's preferences.

[27] Baldwin (1987) points out that different political economy models map to a planner's problem with group-specific weights. Grossman and Helpman (2001) model lobby contribution games that map to a planner's problem with sector-specific Pareto weights; more recently, Adão et al. (2023) estimate the Pareto weights across industries and skill groups of a hypothetical planner setting US tariffs.

6.1 Planner's Problem

Formally, the planner chooses the geographic coordinates $\mathbf{d} = \{d_1, ..., d_{24}\}$ of the 24 CHSR stations, where each d_i belongs to a set $\mathscr{D}$ of feasible coordinates. Each design $\mathbf{d}$ maps to a distribution of travel times and travel costs, and, through those variables, to a utility level of each agent ω if the design is approved ($u_\omega(Y;\mathbf{d})$ defined in (1)) and to a share of favorable votes ($v(i;\mathbf{d})$ defined in (5)). The planner's valuation for design $\mathbf{d}$ given a realization of fundamentals is:

$$\mathcal{W}(\mathbf{d}) \equiv \sum_{i=1}^{I} \lambda_U(i) N_R(i) \mathbb{E}_\omega [u_\omega(Y;\mathbf{d}) - u_\omega(N)] + \lambda_V \sum_{i=1}^{J} N_R(i) v(i;\mathbf{d}). \tag{18}$$

The first term in (18) is a weighted sum across tracts i of the average (over agents ω living in i) expected utility gains from the CHSR being approved. This sum uses location-specific per-capita weights $\lambda_U(i)$ that capture the planner's preferences for each of the $N_R(i)$ residents of tract i. Thus, this term captures distributional impacts among constituencies. We define the $\lambda_U(i)$ as functions of observed covariates $Z_k(i)$ (including a constant, $Z_0(i) \equiv 1$):

$$\lambda_U(i) \equiv \sum_{k=0}^{K} b_k Z_k(i). \tag{19}$$

The second term in (18) is the total number of votes in favor of the CHSR; when $\lambda_V > 0$, the planner attaches a positive weight to favorable votes. This term captures that policymakers may take into account the popular approval elicited by a transport infrastructure project.

The design $\mathbf{d}^0$ that maximizes the planner's objective function in (18) is:

$$\mathbf{d}^0 = \arg\max_{\mathbf{d}\in\mathscr{D}^{24}} \mathbb{E}[\mathcal{W}(\mathbf{d})], \tag{20}$$

where the expectation is taken over future realization of the economic fundamentals, unknown at the time of designing the CHSR. Using the definition of u_ω in (1), and assuming voters' information about fundamentals at the time of voting is as rich as the planner's, the optimal design is

$$\mathbf{d}^0 = \arg\max_{\mathbf{d}\in\mathscr{D}^{24}} \mathbb{E}\left[\sum_{i=1}^{J} \lambda_U(i) N_R(i) \ln \hat{W}(i;\mathbf{d}) + \lambda_V \sum_{i=1}^{J} N_R(i) v(i;\mathbf{d})\right]. \tag{21}$$

Similar to voters, we assume the planner has rational expectations. Hence, for any design $\mathbf{d}^n$ different from the optimal design $\mathbf{d}^0$:

$$\mathbb{E}\left[\mathcal{W}(\mathbf{d}^n) - \mathcal{W}\left(\mathbf{d}^0\right)\right] \approx \sum_{i=1}^{J} \left(\lambda_U(i) + \lambda_V \frac{\partial v(i)}{\partial \ln \hat{W}(i)}\right) N_R(i) \Delta \ln \hat{W}(i, \mathbf{d}^n) - \epsilon(\mathbf{d}^n) \leq 0, \tag{22}$$

where

$$\Delta \ln \hat{W}(i, \mathbf{d}^n) \equiv \ln\left(\frac{\hat{W}(i;\mathbf{d}^n)}{\hat{W}(i;\mathbf{d}^0)}\right) = \ln\left(\frac{W(i, Y, \mathbf{d}^n)}{W(i, Y, \mathbf{d}^0)}\right) \tag{23}$$

is the log difference between the real income generated by the design $\mathbf{d}^n$ and that generated by the optimal design $\mathbf{d}^0$, and where $\epsilon(\mathbf{d}^n)$ is the planner's expectational error when evaluating the impact of the high-speed rail design.

6.2 Estimation of Planner's Preferences

We assume that the observed CHSR design corresponds to the optimal choice of the planner. Thus, according to (22), any deviation from the observed design must yield weakly negative returns. Condition (22) demonstrates how the trade-off between votes and real income in the planner's design helps us identify the planner's preferences. The planner has incentives to favor locations with higher Pareto weight $\lambda_U(i)$ or higher returns in terms of votes, as captured by $\partial v(i) / \partial \ln \hat{W}(i)$ defined in (17). Unchosen CHSR designs that increase the real income of a location i at the expense of some other location i' reveal an upper bound for $\lambda_U(i)$ relative to $\lambda_U(i')$. Similarly, unchosen CHSR designs that increase (reduce) aggregate (λ_U-weighted) real-income while reducing (increasing) aggregate votes reveal a lower (upper) bound for λ_V relative to the average of the $\lambda_U(i)$'s.

A possible approach to estimate the parameters entering (18) would be to exploit differences in the planner's payoff function in reaction to small perturbations to stations' locations, much like Goldberg and Maggi (1999) and Adão et al. (2023) do for tariffs. In that case, we could derive moments from the first-order condition in (22) holding as an equality rather than as an inequality. However, in our empirical setting, this approach has little identification power due to all covariates of interest changing smoothly in space. Moreover, closed-form solutions for optimal policies, which are relied upon for implementation in the previous papers on tariffs, are unavailable in our case. Instead, we use a moment inequality estimator that exploits discrete (rather than marginal) deviations from the planner's observed station placement, in which case we can construct moments from the first-order condition in (22) holding as a weak inequality. Specifically, we derive our moment inequalities following the revealed-preference approach introduced in Pakes (2010) and Pakes et al. (2015). We provide implementation details in Appendix F.

The main shortcoming of our estimator is the absence of unobserved (to the researcher) determinants of the planner's payoff function (e.g., cost shocks) that may vary across locations. Allowing for such unobserved determinants in our setting would only be feasible if we could determine ex ante (i.e., for all parameter values) whether the decisions to place stations in any two locations are complements or substitutes (see, e.g., Jia (2008), Arkolakis et al. (2023), and Castro-Vincenzi et al. (2024)). Unfortunately, in our setting, those decisions may be either complements or substitutes: depending on the placement of other stations, the planner's gains from an extra station may decrease or increase as stations are added. Therefore, to implement our moment inequalities, we must model the unobserved terms as expectational errors of the planner.

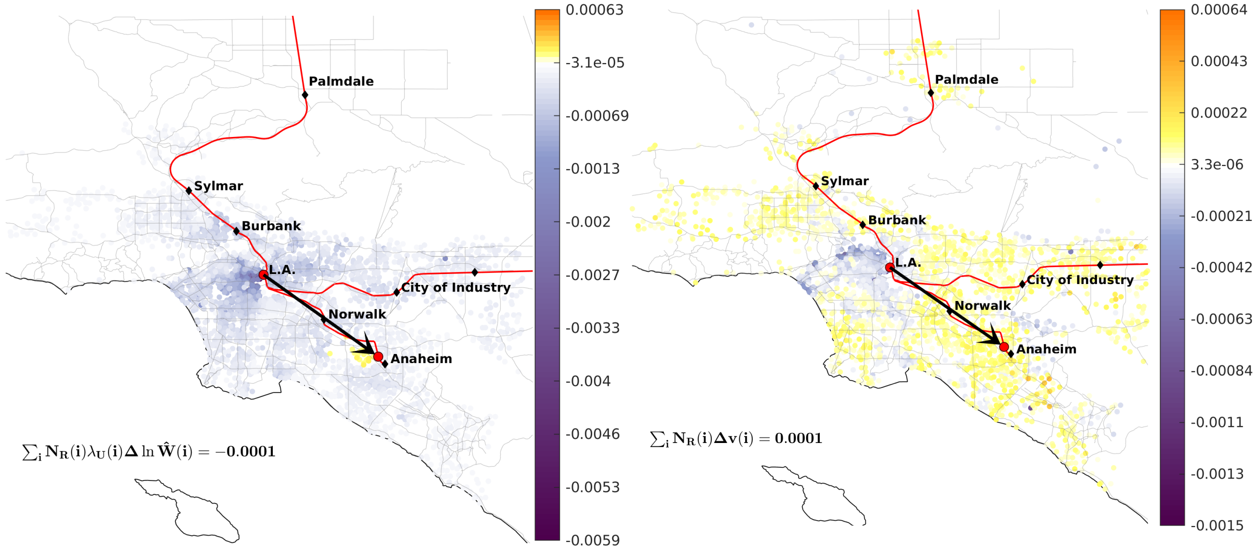
Construction of Perturbations and Moments To build the inequalities in (22), we first construct hundreds of perturbations to the distribution of stations along the proposed CHSR design. In each perturbation, one station is shifted to another location without a proposed station. The set of alternative locations is chosen to identify upper and lower bounds on the parameters: we use the peaks and troughs of the covariates $Z_k(i)$ and of the voting elasticity $\partial v(i) / \partial \ln \hat{W}(i)$ along the CHSR rail line.[28] We index each one-station deviation by n and let $\mathbf{d}^n$ be the associated design.

[28]Figure A.8 in Appendix F shows such peaks and throughs for a covariate $Z_k(i)$ that equals population density.

Figure 4: Example of Perturbation Identifying λ

Note:

(a) Welfare Change (b) Voting Semi-elasticity



the figures show the $\Delta \ln \hat{W}(i)$ for a specific perturbation using 2008 cost predictions, and the voting semi-elasticity at the initial equilibrium corresponding to our preferred estimate under those cost predictions.

Figure 4 shows one such perturbation, where the station in downtown LA is shifted towards Anaheim. The map on the left shows the real-income differences with the actual design, $\Delta \ln \hat{W}(i)$ in (23). The map on the right shows the voting semi-elasticity, $\partial \text{v}(i)/\partial \ln \hat{W}(i)$. Downtown LA is a low-voting-gradient area as it strongly supports the CHSR, so increasing their residents' real income barely changes votes in that location. Conversely, the areas closer to the new location of the station have higher voting elasticities with respect to real income. Thus, by shifting the station's location, this perturbation redistributes real income from low-voting-gradient areas of LA to higher-voting-gradient places in Orange County. As this perturbation implies a reduction in the real-income component of the planner's objective and an increase in aggregate votes, it helps identify an upper bound on the preference for votes λ_V relative to an average of the $\lambda_U(i)$.

For estimation, we build moment inequalities using mutually exclusive subsets of perturbations. Specifically, letting $\mathcal{N}$ be the set of all perturbations, each moment $e = 1, .., E$ is defined as the average value $m(\mathcal{N}_e)$ of the perturbations to the planner's objective, $\Delta \mathcal{W}(\mathbf{d}^n)$ defined in (22), across a subgroup $\mathcal{N}_e \subset \mathcal{N}$:

$$m(\mathcal{N}_e) \equiv \sum_{n \in \mathcal{N}_e} \Delta \mathcal{W}(\mathbf{d}^n) \leq 0 \tag{24}$$

for $e = 1, .., E$. As long as the subsets $\mathcal{N}_e$ are defined as a function of variables that belong to the information set of the planner, the assumption of rational expectations implies that the expectational errors $\epsilon(\mathbf{d}^n)$ in (22) are averaged out in $m(\mathcal{N}_e)$.[29] Then, using (19), we can rewrite

[29] I.e., $\sum_n \epsilon(\mathbf{d}^n) 1\{n \in N_e\} \to 0$ as $\sum_n 1\{n \in N_e\} \to \infty$. This result depends on two assumptions. First, for any perturbation n, the unobserved term $\epsilon(\mathbf{d}^n)$ is unknown to the planner at the time at which the optimal CHSR

the moment inequality e in (24), as

$$m(\mathcal{N}_e) \equiv \sum_{n\in\mathcal{N}_e}\left[\sum_{k=0}^{K} b_k Z_k(i) + \lambda_V\theta_V v(i)(1-v(i))\right] N_R(i)\,\Delta\ln\hat{W}(i,\mathbf{d}^n) \leq 0 \text{ for } e=1,..,E. \quad (25)$$

The inequalities in (25) provide E conditions that depend on observable covariates ($Z_k(i)$, $v(i)(1-v(i))$, $N_R(i)$), the model-generated real-income gains in each perturbation $\hat{W}(i,\mathbf{d}^n)$, and the previously estimated θ_V. Using these E conditions, we recover the $K+2$ parameters $b_0,..,b_K$ and λ_V. We group perturbations that help identify upper and lower bounds for our various parameters (see Appendix F.1 for details). Because the conditions hold as inequalities, we compute confidence sets for the parameters $\{b_k,\lambda_V\}$ following Andrews and Soares (2010). This procedure identifies an admissible set of parameter values such that, for a given confidence level, we are unable to reject the hypothesis that the data was generated by a parameter vector within the set.

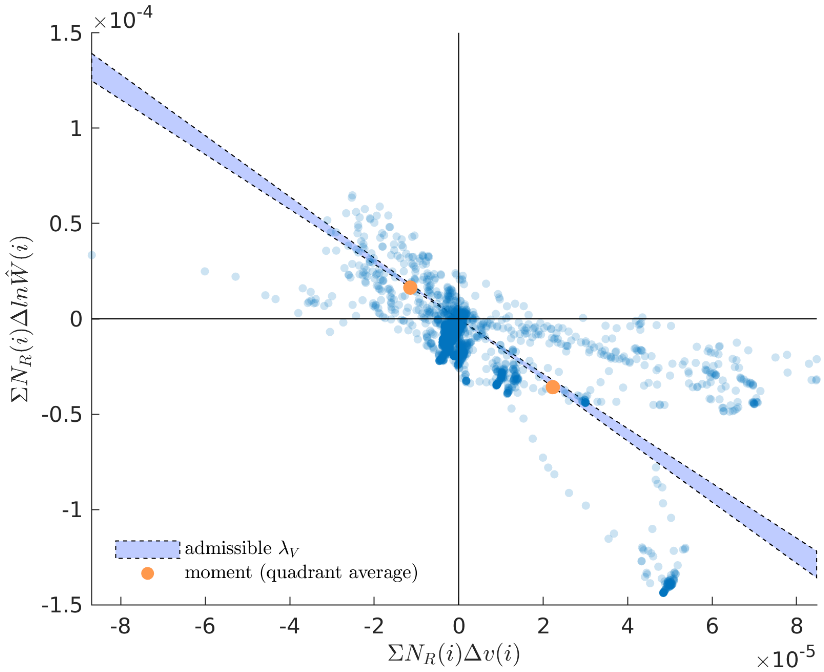
Pareto Frontier Between Votes and Welfare To understand how we form the moment inequalities and how they identify our parameters, we discuss here a restricted case in which the planner has no preferences over the distribution of real income in space (the planner's objective function has $b_k=0$ for $k=1,..,K$). Instead, it holds preferences over a utilitarian component of real income, $\sum_i N_R(i)\ln\hat{W}(i;\mathbf{d})$, and over aggregate votes, $\sum_i N_R(i)\,v(i;\mathbf{d})$, with weights b_0 and λ_V, respectively. Because only λ_V/b_0 is identified, we normalize $b_0\equiv 1$. Figure 5 plots all the perturbations, where the x-axis represents the (demeaned) total change in votes and the y-axis represents the change in the utilitarian component of the planner's welfare. The actual CHSR proposal corresponds to the point $(0,0)$.

As the only unknown parameter is λ_V, we form four moments $m(\mathcal{N}_e)$ by grouping the perturbations according to how they fall in the four quadrants displayed in the figure. The lower-left quadrant groups perturbations that yield both lower utilitarian welfare and lower votes than the actual CHSR proposal, trivially satisfying our inequality conditions. In turn, perturbations in the upper-right quadrant imply greater welfare and higher votes, and can only be rationalized by the planner's expectational errors.

In contrast, the perturbations in the upper-left and lower-right quadrants suggest a trade-off between welfare and votes. The orange dots represent the two moments $m(\mathcal{N}_e)$, that average across all perturbation within the corresponding quadrant. The values for λ_V (relative to b_0) consistent with these two moment inequalities are those that satisfy the condition $y+\lambda_V x\leq 0$, with (x,y) the coordinates of each of the two moments depicted in the orange dots. Visually, this set corresponds to drawing a line through the origin with (negative) slopes x^{UL}/y^{UL} and x^{LR}/y^{LR} (where UL and LR denote the upper-left and lower-right quadrants, respectively) and taking the intersection of the areas between these lines, as represented by the purple area on the graph. The set is $0\leq -x^{UL}/y^{UL}\leq\lambda\leq -x^{LR}/y^{LR}$.

design was chosen; this is guaranteed by the assumption that $\epsilon(\mathbf{d}^n)$ exclusively incorporates expectational errors of the planner. Second, across the perturbations n included in the same subset $\mathcal{N}_e$, the correlation across the different unobserved terms $\epsilon(\mathbf{d}^n)$ is sufficiently low such that the Law of Large Numbers applies.

Figure 5: Utilitarian Welfare vs. Votes



As the figure illustrates, the perturbations in the upper-left quadrant identify a strictly positive lower bound on λ_V, while those in the lower-right quadrant identify an upper bound. Our estimation procedure extends this logic to multiple covariates. Appendix F gives additional details on how the perturbations and moments are constructed in the general case in which the unknown parameter vector includes b_k for $k = 1, \ldots, K$, and on how we implement the procedure in Andrews and Soares (2010) to compute confidence sets.

Parameter Estimates Table 5 shows estimates of the b's and λ_V under the baseline model using both the projections contained in the 2008 and the 2022 business plans. We present results for specifications that allow the planner's Pareto weights to depend on population density, the share of college-educated residents, or the share of non-white residents. In the estimation, the parameter space is not restricted and, as a result, our confidence sets could include both positive and negative values for each parameter b_k or λ_V.

To simplify the interpretation, we standardize every covariate $Z_k(i)$ and normalize to 1 the population-weighted mean of the Pareto weights $\lambda_U(i)$. Each parameter b_k can thus be interpreted as capturing the impact of a one-standard deviation increase in $Z_k(i)$ relative to the average Pareto weight.[30] For each specification and parameter, we report the 95% confidence set projected on the corresponding parameter. Column (1) of Table 5 corresponds to Figure 5. Columns (2) to (4) allow for one b_k at a time, while column (5) reports an estimation with all the b_k's.

[30] We also demean the variable $\theta_V v(i)(1 - v(i))$. As a result, the constant we estimate is $b_0 + \lambda_V \mathbb{E}[\theta_V v(i)(1 - v(i))]$ and captures the overall utilitarian motive of the planner, including what is inherited from the voting block.

Table 5: Planner's Preferences Estimates

(a) 2008 Business Plan

Observable	Confidence Sets				
	(1)	(2)	(3)	(4)	(5)
Density		[-0.06, 1.10]			[-0.07, 0.25]
Share college			[0.42, 0.53]		[0.06, 0.44]
Share non-white				[-0.35, 0.53]	[-0.30, 0.20]
Votes	[-0.03, 1.60]	[0.39, 2.34]	[0.55, 1.13]	[0.24, 1.66]	[0.64, 1.14]
Constant	[1.00, 1.00]	[0.98, 1.00]	[0.99, 0.99]	[0.98, 1.01]	[0.99, 1.01]

(b) 2022 Business Plan

Observable	Confidence Sets				
	(1)	(2)	(3)	(4)	(5)
Density		[-0.06, 1.10]			[-0.06, 0.21]
Share college			[0.30, 0.53]		[0.06, 0.44]
Share non-white				[-0.35, 0.53]	[-0.30, 0.18]
Votes	[0.05, 0.26]	[0.06, 0.37]	[0.08, 0.14]	[0.05, 0.27]	[0.10, 0.18]
Constant	[1.00, 1.00]	[0.98, 1.00]	[0.99, 1.00]	[0.98, 1.01]	[0.99, 1.01]

Notes: The brackets indicate the min and the max of the 95% confidence set for each covariate.

Across all but one specification, our estimates indicate the planner has a strictly positive preference for votes. I.e., except for the case without any covariate in Column (1) of panel (a), the confidence set corresponding for the preferences over votes does not include 0. In our preferred specification in column (5), the coefficient on votes is centered at 0.89 using 2008 cost forecasts and at 0.14 using 2022 forecasts. These numbers mean that the planner is willing to give up 0.14% to 0.89% additional real income gains for an extra percentage point of favorable votes.

The confidence sets for the parameters entering the planner's Pareto weights also indicate that the planner assigns a larger weight to welfare in census tracts with a larger share of college-educated residents. In turn, we cannot rule out that the planner's weights are invariant to the census tract's population density and share of non-white residents (since the $\lambda_U(i)$ are defined as per-capita Pareto weights, the planner still assigns a larger weight to more populated tracts, even if the per-capita Pareto weight does not vary with population density). In sum, the planner is far from the utilitarian benchmark. Appendix Figure A.4 shows the distribution of total Pareto weights $\lambda_U(i) + \lambda_V \theta_V v(i)(1 - v(i))$ across tracts i implied by the centroid of our confidence set; these estimates imply a large variance in the Pareto weights.

6.3 Optimal Station Placement with Apolitical Preferences

To illustrate the importance of political and distributional considerations for the design of transportation policy, we compute the optimal station placement under alternative planners' preferences. The optimal station placement problem is non-convex due to substitutabilities and complementarities across stations, as well as due to the sigmoidal function –neither convex nor concave– defining the voting probabilities, $v(i; \mathbf{d})$ in (5). To optimize, we use a procedure that combines the perturbations from the estimation stage, simulated annealing, and a continuous optimizer. Appendix F.2 provides details.[31]

Optimal Stations Along the Actual CHSR Route We compare the proposed CHSR design to that preferred by an *apolitical* planner whose objective function does not depend on votes (i.e., $\lambda_V = 0$), thereby eliminating the planner's incentives to assign higher real income to higher voting-elasticity locations. We set the values of the remaining parameters entering the planner's Pareto weight to the centroid of the confidence sets from column (5) in Table 5.[32] We restrict the planner to place all stations along actual CHSR route, and explore below the setting in which the planner places the stations along the alternative rail line along the I-5 highway.

Figure 6 shows in red circles the stations that maximize the objective of the apolitical planner, in black diamonds the stations in the proposed CHSR design, and a color scale from blue to red to indicate the welfare changes in different locations. While not immediately apparent from the figure, the optimal design of the apolitical planner is quite different from the proposed one: the political motives shift stations away from high-density areas of Los Angeles and San Francisco. These locations have strong preferences in favor of the CHSR; as a result, the elasticity of the favorable vote share with respect to the real income of their residents is low. Hence, in the absence of political motives, several stations that, according to the original proposal, were located in suburban areas of San Francisco (such as San Jose), Los Angeles (Palmdale, Sylmar, Burbank) and San Diego (University City) are reallocated towards their corresponding metropolitan areas.

The relocation of stations, and their associated income impacts, are shown in Column (1) of Appendix Table A.7. In line with our previous explanation, the largest welfare gains from eliminating political preferences arise from stations that move closer to dense urban areas, also causing the largest vote losses.

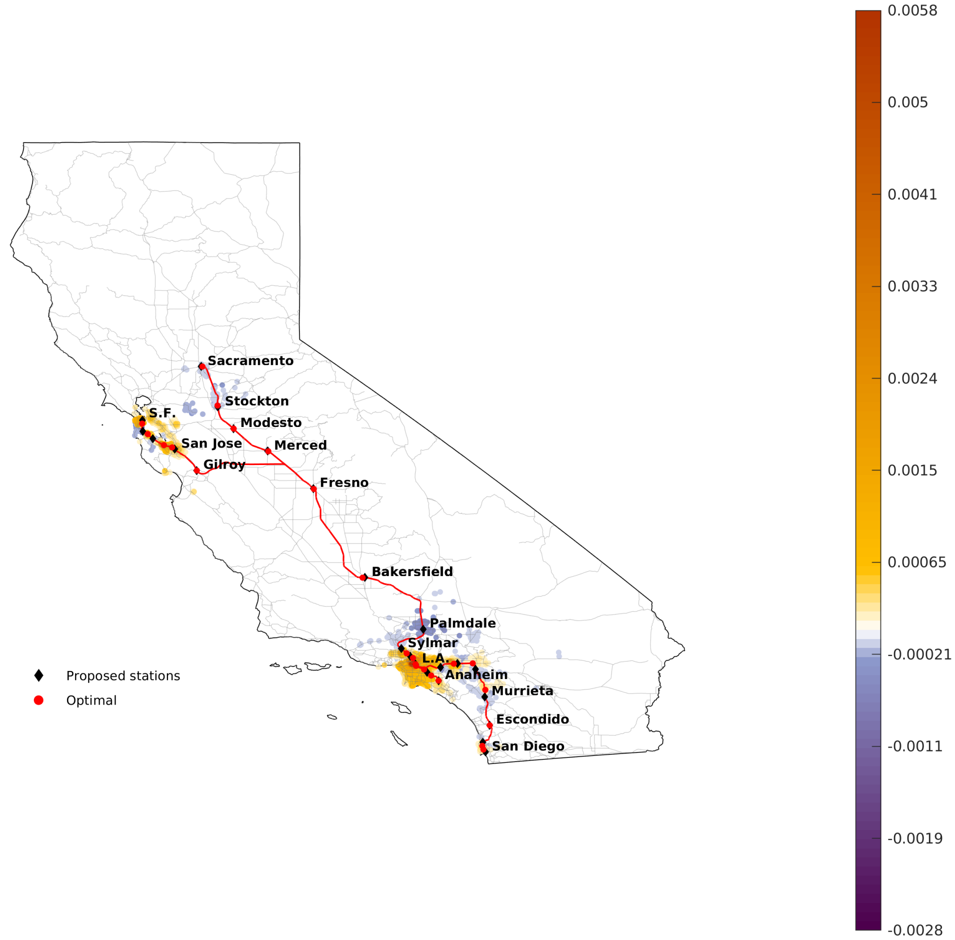
To gauge the distributional impacts of political preferences, Table 6 displays the cross-sectional impact of the counterfactual CHSR design using the 2008 cost predictions. We find that locations with low voting elasticity experience the largest gains. This implies a redistribution of welfare towards high population density urban areas, which also feature a higher share of non-white residents. Appendix Figure A.6 zooms in on the Los Angeles region, displaying the real income changes (left panel) and the voting elasticity (right panel). The relocation of stations towards more central locations redistributes welfare towards the city center and away from the suburbs, in line with a general redistribution pattern towards high density areas that have low voting elasticity.

As expected, aggregate real income increases and votes decrease when the planner does not care about votes. Compared to the impacts of original CHSR in Table 2, the gross population-weighted gains (without accounting for capital costs, which are kept constant when computing the optimal

[31] We verify that our procedure yields the best outcome when each station is individually shifted over a 10 km range from the optimum. Appendix Figure A.5 shows this check.

[32] Using the 2008 estimates from Panel (a) of Table 5, this implies: b = [0.07, 0.20, -0.07]. Appendix Figure A.3 displays the topology of the confidence set from column (5) by showing pairwise projections for each combination of the parameters b_k and λ.

Figure 6: Apolitical Planner, Optimal Stations along CHSR Route and Impacts vs. Proposed Plan



Notes: The maps shows the change in the real income of each location, $\Delta \ln \hat{W}\left(i, \mathbf{d}^{*}_{CHSR}\right)$ defined in (23), when comparing the optimal design d^{*}_{CHSR} of an apolitical planner along the CHSR route and the proposed CHSR. The outcomes correspond to using the baseline model under the 2008 business plan.

station placements) grow considerably, by 15% under 2008 costs, or by 25% under 2022 costs. The aggregate share of favorable votes declines only slightly (by 0.02 and 0.01 percentage points under 2008 and 2022 costs, respectively), although as mentioned we observe considerable variance in the changes in votes across locations.

The optimal station distribution of a utilitarian planner ($\lambda_U(i) = 1$ and $\lambda_V = 0$) does not differ significantly from the apolitical one, except that some stations are shifted from high to lower college-share areas, as illustrated for the case of San Diego in Appendix Figure A.7. The main beneficiaries from eliminating the planner's preference for high college-share locations around University City are areas in downtown San Diego, whose population exhibits a mix of high and less educated residents. Compared to the proposed design, the aggregate welfare gains associated with the optimal placement of stations of the utilitarian planner are very similar to those obtained by the apolitical planner, suggesting that the key factor behind the lower aggregate welfare gains associated to the proposed CHSR is the weight that the actual planner put on votes.

Table 6: Apolitical planner, Welfare Gains across Covariates

ΔReal Income by Quartile	1st	2nd	3rd	4th
Density	0.00%	0.00%	0.02%	0.05%
Share college	0.02%	0.02%	0.01%	0.02%
Share non-white	0.01%	0.01%	0.02%	0.04%
Voting elasticity	0.04%	0.02%	0.01%	0.01%

Notes: 2008 Business Plan estimates of real income gains of moving from proposed plan to apolitical optimum. Real income gains are population-weighted and computed within each quartile of the corresponding variable in each row.

Optimal Stations Along the I-5 Highway Route The alternative I-5 route differs from the proposed one in that it provides a more direct connection between San Francisco and Los Angeles. It would have also served fewer destinations, by removing the Sacramento, Anaheim, and San Diego segments from the full network.

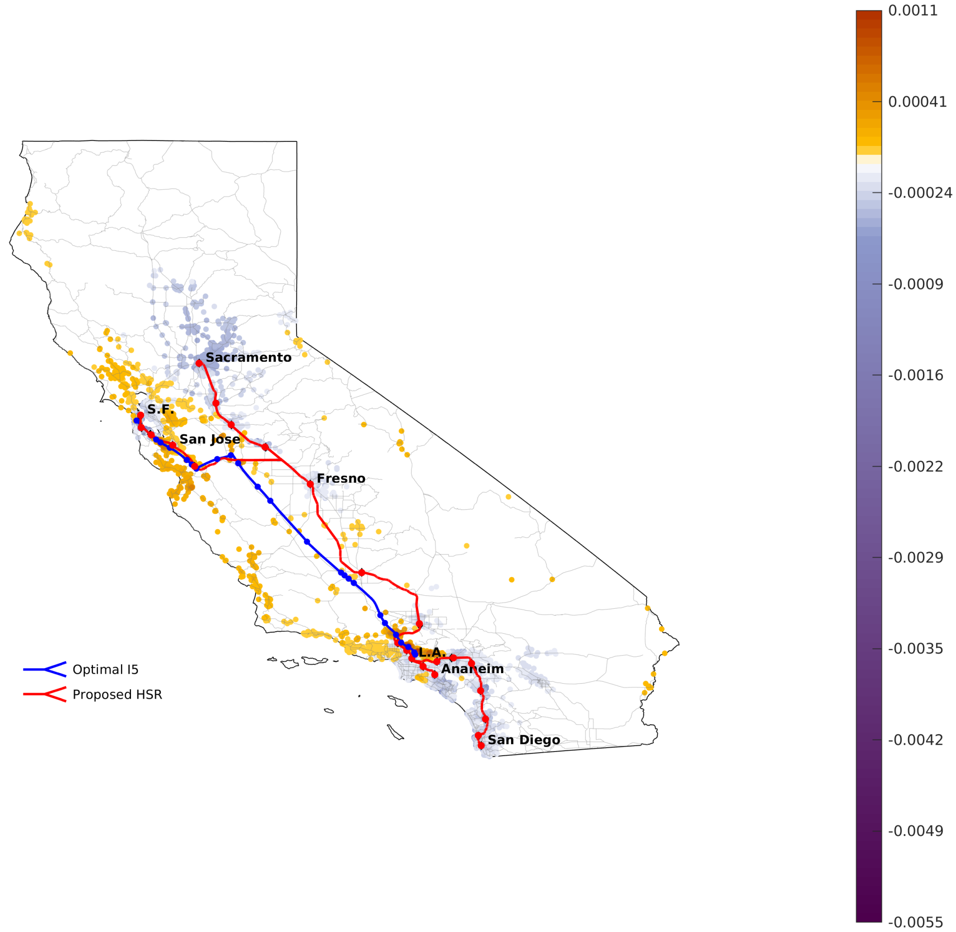
A possible reason why the I-5 route was not chosen is preference for votes. To determine whether this could have been a key factor, we proceed in two steps. First, as we ignore the capital cost of the I-5 route, we estimate the value that makes the actual policymakers, under our estimated preferences, indifferent between optimally locating stations along the I-5 or along the proposed CHSR route. This procedure yields a cost of 44.0 billion USD, or 12% less than the capital costs assuming 2008 forecasts. Given costs greater than this value, our model rationalizes why the planner chose the actual CHSR over the I-5 route.

Then, given these costs costs, we consider whether an apolitical planner would have preferred to optimally place stations along the alternative I-5 route or along the actual CHSR. Figure 7 displays the I-5 route with optimal station placement of the apolitical planner, and shows the welfare changes relative to the proposed plan. Due to its smaller geographical coverage, the I-5 route results in welfare losses for many locations along the Sacramento and San Diego branches. However, various locations in the greater San Francisco and Los Angeles areas benefit from a faster connection between the two large metropolitan areas and from a greater density of stations.

Aggregating across tracts and weighting by population, we find a slight loss for the apolitical planner compared to the proposed CHSR. Thus, a planner with no preferences for votes would have also chosen the actual CHSR route, through the Central Valley, over the I5 route. We conclude that the route along the I-5 highway was not discarded due to the type of political preferences that we estimate.[33]

[33] An alternative explanation is that the costs of the I5 route were in fact lower than our previous estimate, and that the route was discarded due to forces outside our model, such as a regional favoritism for the Central Valley beyond the preferences for demographic groups or votes that we have estimated.

Figure 7: Apolitical Planner, Optimal Stations along I5 Route and Impacts versus Proposed Plan



Notes: The maps shows the change in the real income of each location, $\Delta \ln \hat{W}\left(i, \mathbf{d}_{I5}^{*}\right)$ defined in (23), when comparing the optimal design d_{I5}^{*} of an apolitical planner along the I5 route and the proposed CHSR. The outcomes correspond to using the baseline model under the 2008 business plan.

7 Conclusion

We studied the role of policymakers' and households' preferences in shaping transportation infrastructure projects. We use the California High-Speed Rail as the basis of our study, leveraging the fact that we observe vote shares in favor of this project across California's census tracts. The estimates reveal that voters did respond to the expected real-income impacts of the CHSR. So, in this context, economic voting was a significant driver of policy preferences over transport infrastructure.

We posit that the proposed CHSR design represents the optimal choice of a planner, whose preferences we estimate using moment inequalities. Our estimator infers the planner's preferences by comparing the value of the planner's objective function under the proposed plan to that under alternative designs. We find that the planer selected a design that increased the total number of favorable votes in the 2008 referendum at the expense of alternative designs that would have resulted in a smaller electoral support but larger aggregate welfare. We thus conclude that attaining popular approval was an important driver of the proposed design of the California High-Speed Rail.

As a result of these preferences for gaining votes, stations were placed further away from the main metropolitan areas than it would have been optimal from an aggregate real income maximization perspective alone. Our economic model and moment inequality methodology may be generally applied to infer the considerations taken into account by policymakers in charge of the design of large and complex transport networks.

References

Adão, R., A. Costinot, D. Donaldson, and J. A. Sturm (2023): "Why is Trade Not Free? A Revealed Preference Approach," Tech. rep., National Bureau of Economic Research.

Ahlfeldt, G. M., S. J. Redding, D. M. Sturm, and N. Wolf (2015): "The economics of density: Evidence from the Berlin Wall," *Econometrica*, 83, 2127–2189.

Alesina, A. and P. Giuliano (2011): "Preferences for redistribution," in *Handbook of social economics*, Elsevier, vol. 1, 93–131.

Allen, T. and C. Arkolakis (2022): "The welfare effects of transportation infrastructure improvements," *The Review of Economic Studies*, 89, 2911–2957.

Almagro, M., F. Barbieri, J. C. Castillo, N. G. Hickok, and T. Salz (2024): "Optimal Urban Transportation Policy: Evidence from Chicago," Tech. rep., National Bureau of Economic Research.

American Public Transit Association (2010): "Public Transportation Fare Database," Tech. rep., https://www.apta.com/research-technical-resources/transit-statistics/fare-database/.

Andrews, D. W. and G. Soares (2010): "Inference for parameters defined by moment inequalities using generalized moment selection," *Econometrica*, 78, 119–157.

Arkolakis, C., F. Eckert, and R. Shi (2023): "Combinatorial Discrete Choice: A Quantitative Model of Multinational Location Decisions," Tech. rep., National Bureau of Economic Research.

Baldwin, R. (1987): "Politically realistic objective functions and trade policy PROFs and tariffs," *Economics Letters*, 24, 287–290.

Bernard, A. B., A. Moxnes, and Y. U. Saito (2019): "Production networks, geography, and firm performance," *Journal of Political Economy*, 127, 639–688.

Borusyak, K. and P. Hull (2023): "Nonrandom exposure to exogenous shocks," *Econometrica*, 91, 2155–2185.

Bourguignon, F. and A. Spadaro (2012): "Tax–benefit revealed social preferences," *The Journal of Economic Inequality*, 10, 75–108.

Brancaccio, G., M. Kalouptsidi, and T. Papageorgiou (2024): "Investment in Infrastructure and Trade: The Case of Ports," .

Brancaccio, G., M. Kalouptsidi, T. Papageorgiou, and N. Rosaia (2023): "Search frictions and efficiency in decentralized transport markets," *The Quarterly Journal of Economics*, 138, 2451–2503.

Brennan, G. and A. Hamlin (1998): "Expressive voting and electoral equilibrium," *Public choice*, 95, 149–175.

Buchholz, N. (2022): "Spatial equilibrium, search frictions, and dynamic efficiency in the taxi industry," *The Review of Economic Studies*, 89, 556–591.

Bureau of Transportation Statistics (2008a): "Airline Origin and Destination Survey (OND)," Tech. rep., U.S. Department of Transportation, https://www.transtats.bts.gov/Tables.asp?QO_VQ=EFI&QO_anzr=Nv4yv0r%FDb4vtv0%FDn0q%FDQr56v0n6v10%FDf748rB%FD%FLQOEO%FM&QO_fu146_anzr=b4vtv0%FDn0q%FDQr56v0n6v10%FDf748rB.

——— (2008b): "TranStats," Tech. rep., U.S. Department of Transportation, https://www.transtats.bts.gov/.

Burgess, R., R. Jedwab, E. Miguel, A. Morjaria, and G. Padró i Miquel (2015): "The value of democracy: evidence from road building in Kenya," *American Economic Review*, 105, 1817–51.

California Department of Transportation (2012): "2010-2012 California Household Travel Survey," Tech. rep., https://www.nrel.gov/transportation/secure-transportation-data/tsdc-california-travel-survey.html.

California High Speed Rail Authority (2008): "Business Plan 2008," *https://hsr.ca.gov/about/high-speed-rail-business-plans/2008-business-plan/*.

——— (2022): "Business Plan 2022," *https://hsr.ca.gov/about/high-speed-rail-business-plans/2022-business-plan/*.

——— (2023): "2023 Project Update Report," *https://hsr.ca.gov/about/project-update-reports/2023-project-update-report/*.

Castro-Vincenzi, J., E. Menaguale, E. Morales, and A. Sabal (2024): "Market Entry and Plant Location in Multiproduct Firms," Tech. rep., mimeo.

Deacon, R. and P. Shapiro (1975): "Private preference for collective goods revealed through voting on referenda," *The American Economic Review*, 65, 943–955.

Dickstein, M. J. and E. Morales (2018): "What do exporters know?" *The Quarterly Journal of Economics*, 133, 1753–1801.

Dingel, J. I. and F. Tintelnot (2020): "Spatial economics for granular settings," Tech. rep., National Bureau of Economic Research.

Dong, X., S. Zheng, and M. E. Kahn (2020): "The role of transportation speed in facilitating high skilled teamwork across cities," *Journal of Urban Economics*, 115, 103212.

Fajgelbaum, P. D. and E. Schaal (2020): "Optimal transport networks in spatial equilibrium," *Econometrica*, 88, 1411–1452.

Frechette, G. R., A. Lizzeri, and T. Salz (2019): "Frictions in a competitive, regulated market: Evidence from taxis," *American Economic Review*, 109, 2954–2992.

Goldberg, P. K. and G. Maggi (1999): "Protection for Sale: An Empirical Investigation," *American Economic Review*, 89, 1135–1155.

Grossman, G. M. and E. Helpman (2001): *Special interest politics*, MIT press.

Gupta, A., S. Van Nieuwerburgh, and C. Kontokosta (2022): "Take the Q train: Value capture of public infrastructure projects," *Journal of Urban Economics*, 129, 103422.

Heblich, S., S. J. Redding, and D. M. Sturm (2020): "The making of the modern metropolis: evidence from London," *The Quarterly Journal of Economics*, 135, 2059–2133.

Hillman, A. L. (2010): "Expressive behavior in economics and politics," *European Journal of political economy*, 26, 403–418.

Holian, M. J., M. E. Kahn, et al. (2013): "California voting and suburbanization patterns: implications for transit policy." Tech. rep., California. Dept. of Transportation.

Jia, P. (2008): "What happens when Wal-Mart comes to town: An empirical analysis of the discount retailing industry," *Econometrica*, 76, 1263–1316.

Kahn, M. E. and J. G. Matsusaka (1997): "Demand for environmental goods: Evidence from voting patterns on California initiatives," *The Journal of Law and Economics*, 40, 137–174.

Khanna, G., A. Nyshadham, D. Ramos-Menchelli, J. A. Tamayo, and A. Tiew (2023): *Spatial mobility, economic opportunity, and crime*, Harvard Business School.

Koster, H. R., T. Tabuchi, and J.-F. Thisse (2022): "To be connected or not to be connected? The role of long-haul economies," *Journal of Economic Geography*, 22, 711–753.

Kreindler, G., A. Gaduh, T. Graff, R. Hanna, and B. A. Olken (2023): "Optimal Public Transportation Networks: Evidence from the World's Largest Bus Rapid Transit System in Jakarta," Tech. rep., National Bureau of Economic Research.

Lewis-Beck, M. S. and M. Stegmaier (2000): "Economic determinants of electoral outcomes," *Annual review of political science*, 3, 183–219.

Monte, F., S. J. Redding, and E. Rossi-Hansberg (2018): "Commuting, migration, and local employment elasticities," *American Economic Review*, 108, 3855–90.

Noordegraaf, D. V., J. A. Annema, and B. van Wee (2014): "Policy implementation lessons from six road pricing cases," *Transportation Research Part A: Policy and Practice*, 59, 172–191.

Pakes, A. (2010): "Alternative models for moment inequalities," *Econometrica*, 78, 1783–1822.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2015): "Moment inequalities and their application," *Econometrica*, 83, 315–334.

Schade, J. and B. Schlag (2003): "Acceptability of urban transport pricing strategies," *Transportation Research Part F: Traffic Psychology and Behaviour*, 6, 45–61.

Severen, C. (2021): "Commuting, labor, and housing market effects of mass transportation: Welfare and identification," *Review of Economics and Statistics*, 1–99.

Tsivanidis, N. (2019): "Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio," .

Tyndall, J. (2021): "The local labour market effects of light rail transit," *Journal of Urban Economics*, 124, 103350.

University of California, Berkeley (2008): "Statewide Database," Tech. rep., https://statewidedatabase.org/.

U.S. Census (2008a): "Longitudinal Employer-Household Dynamics," Tech. rep., https://lehd.ces.census.gov/data/.

——— (2008b): "TIGER/Line Shapefiles, California, 2008," Tech. rep., https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.2008.html#list-tab-790442341.

U.S. Census Bureau (2010a): "2006-2010 5-Year American Community Survey," Tech. rep., American Community Survey, https://data.census.gov/.

——— (2010b): "2006-2010 5-Year American Community Survey Commuting Flows," Tech. rep., Census Transportation Planning Products, https://transportation.org/ctpp/datasets/.

US DOT (2005): "Final Program Environmental Impact Report/Environmental Impact Statement (EIR/EIS) for the proposed California High-Speed Train System," Tech. rep., US Department of Transportation and CA High Speed Rail Authority.

Valentinyi, A. and B. Herrendorf (2008): "Measuring factor income shares at the sectoral level," *Review of Economic Dynamics*, 11, 820–835.

Van Patten, D. and E. Méndez (2022): "Voting on a Trade Agreement: Firm Networks and Attitudes Toward Openness," Tech. rep., National Bureau of Economic Research.

Verhoef, E. T., P. Nijkamp, and P. Rietveld (1997): "The social feasibility of road pricing: a case study for the Randstad area," *Journal of Transport Economics and Policy*, 255–276.

Zárate, R. D. (2022): "Spatial misallocation, informality, and transit improvements: Evidence from mexico city," .

Zheng, S. and M. E. Kahn (2013): "China's bullet trains facilitate market integration and mitigate the cost of megacity growth," *Proceedings of the national academy of sciences*, 110, E1248–E1253.

Online Appendix to "Political Preferences and the Spatial Distribution of Infrastructure: Evidence from California's High-Speed Rail"

Pablo Fajgelbaum, Cecile Gaubert, Nicole Gorton, Eduardo Morales, Edouard Schaal

A Full Description of the Economic Model

We model an economy with a set $\mathcal{J}$ of tracts, each tract i with a fixed resident population $N_R(i)$ and connected to other tracts by various transport modes. Residents consume a traded good, floor space, and leisure trips. They choose where to commute to work, where to take leisure trips, how many such trips to make, and what transport mode to use for each travel purpose and origin-destination pair. The discrete choices of destination and travel mode are governed by idiosyncratic shocks to residents' preferences. Across tracts, residents are heterogeneous in average preferences over transport modes and leisure destination, efficiency units of labor, and ownership of the local floor space.

Traded good firms produce using labor, floor space, and business trips with constant returns to scale. They choose where to send workers to business trips, how many such trips to make, what transport mode to use, and through what route in the transport network. The discrete choices of business travel destination and travel mode are governed by idiosyncratic shocks to firms' productivity. Across tracts, firms are heterogeneous in terms of average preferences over transport modes and business destination, and productivity. In addition, tracts are heterogeneous in terms of the level of endogenous amenities enjoyed by their residents and the stock of floor space.

All the transport modes operate constant-returns technologies using the tradeable good as input, with ticket prices covering the price of each trip. The CHSR network is constructed with a fixed investment financed from income taxes. The rollout of the CHSR endows the economy with an option of making faster or cheaper trips along some routes within specific modes compared to the status quo.

In the presentation of the model, variables that are indexed by s may change either endogenously or exogenously based on whether the CHSR proposition passes $(s = Y)$ or not $(s = N)$.

A.1 Preferences

When the CHSR status is s, the utility v_ω of an individual ω living in tract i who travels to j_C for commuting and to j_L for leisure, by transport modes m_C and m_L respectively, is:

$$v_\omega(i, j_C, m_C, j_L, m_L, s) = \max_{C, H_C, T_L} B(i, s) \frac{C^{1-\mu_L-\mu_H} H_C^{\mu_H}}{d_C(i, j_C, m_C, s)} \left(\frac{q_L(i, j_L) B(j_L, s)}{d_L(i, j_L, m_L, s)} T_L \right)^{\mu_L} \varepsilon_\omega^C(j_C, m_C) \varepsilon_\omega^L(j_L, m_L). \tag{A.1}$$

subject to the budget constraint:

$$C + r\,(i,s)\,H_C + p_L\,(i,j_L,m_L,s)\,T_L = I\,(i,j_C,m_C,s) \tag{A.2}$$

Expression A.1 indicates that consumers derive utility from the amenities of their place of residence $B\,(i,s)$ and from the consumption of tradeable commodities C, housing H_C, and leisure trips T_L, with Cobb-Douglas shares μ_H and μ_L. The amenity term $B\,(i,s)$ may respond endogenously to the local density of economic activity as detailed below. These workers face disutility $d_C\,(i,j_C,m_C,s)$ from daily commuting travel. The utility that they derive from leisure trips depends negatively on time travelled $d_L\,(i,j_L,m_L,s)$ and positively on the quality of the destination visited, equal to a composite of an exogenous origin-destination component $q_L\,(i,j_L)$ (capturing, for example, that residents of some locations may on average be more likely to have relatives in some other specific location) and the destination-specific amenity $B\,(j_L,s)$. The last two terms of (A.1), $\varepsilon^C_\omega\,(j_C,m_C)$ and $\varepsilon^L_\omega\,(j_L,m_L)$, are idiosyncratic preference shocks for commuting and leisure travel to each destination by each travel mode.

The utility cost of travel is a power function of travel time $\tau_k\,(i,j,m,s)$ and depends on travel mode for both commuting ($k = C$) and leisure ($k = L$):

$$d_k\,(i,j,m,s) = D_k\,(i,m)\,\tau_k\,(i,j,m,s)^\rho \;\text{ for } k = C, L, \tag{A.3}$$

where ρ is the elasticity of travel disutility to travel time and $D_k\,(i,m)$ is a location-specific preference for traveling through transport mode m. This term captures that workers in different tracts may have different tastes for different modes of travel, such as a preference for using cars over public transit.

We turn now to describing the budget constraint A.2. In the expenditure side on the left, the price per unit of tradeable commodities (C) is normalized to 1 and the cost per unit of floor space for housing (H) is $r\,(i,s)$. The monetary cost per round-trip leisure travel from i to j using travel mode m in state s is $p_L\,(i,j,m,s)$. The right-hand side of the budget constraint (A.2) is the disposable income, defined as gross income $y\,(i,j_C,s)$ net of taxes $t\,(s)$ and annual commuting costs:

$$I\,(i,j_C,m_C,s) \equiv (1 - t\,(s))\,y\,(i,j_C,s) - p_C\,(i,j_C,m_C,s)\,T_C. \tag{A.4}$$

The tax rate $t\,(s)$ equals t if the CHSR is approved ($s = Y$) and 0 otherwise. Gross income comes from two sources, labor and home ownership:

$$y\,(i,j_C,s) \equiv e\,(i)\,w\,(j_C,s) + \eta\,(i)\,r\,(i,s)\,. \tag{A.5}$$

The returns to labor equal the efficiency units per resident of tract i, $e\,(i)$, times the wage per efficiency unit at destination, $w\,(j_C,s)$. So, within an origin tract, commuters to different destinations earn different wages based on $w\,(j_C,s)$; and, across origin tracts, commuters to the same destination earn different wages based on $e\,(i)$. The last term in (A.5) is the return to home-ownership, where $\eta\,(i)$ is the locally owned floor space per resident. An increase in land rents $r\,(i,s)$ reduces the real income of tract-i residents through the cost of housing, but it increases it through the returns to land as a function of $\eta\,(i)$.

Finally, the round-trip monetary commuting cost of traveling from i to j through means m in state s is $p_C(i,j,m,s)$. The annual commuting cost multiplies this per-trip cost by the number of working days through the year, T_C. Unlike for leisure, where the number of trips T_L is endogenously chosen, the number of commuting trips T_C is fixed by the number of working days. The resulting demand system is quasi-homothetic, with homothetic demand over C, H, and T_L after spending $p_C(i,j_C,m_C,s)\,T_C$ annually on commuting.

A.2 Indirect Utility and Welfare

Maximizing out the solutions for consumption C, housing H and number of leisure trips T_L, the solution to (A.1) gives indirect utility conditional on the origin, destinations, travel modes, and idiosyncratic preference shocks for destination:

$$v_\omega(i,j_C,m_C,j_L,m_L,s)=\frac{B(i,s)}{r(i,s)^{\mu_H}}\left(\frac{I(i,j_C,m_C,s)}{d_C(i,j_C,m_C,s)}\varepsilon_\omega^C(j_C,m_C)\right)\left(\frac{q_L(i,j_L)\,B(j_L,s)}{p_L(i,j_L,m_L,s)\,d_L(i,j_L,m_L,s)}\right)^{\mu_L}\varepsilon_\omega^L(j_L,m_L) \tag{A.6}$$

Each resident ω makes discrete choices of destination and transport mode for both commuting and leisure to maximizes indirect utility. These choices are represented by the quadruplet $\{j_C,j_L,m_C,m_L\}$. Destinations are chosen from the set of tracts $\mathcal{J}$ while the set of transport modes available for travel purpose $k=L,C$ is $\mathcal{M}_k$. We assume the idiosyncratic preference shocks for commuting and leisure travel $\varepsilon_\omega^C(j_C,m_C)$ and $\varepsilon_\omega^L(j_L,m_L)$ to be IID Type-I extreme value distributed:

$$\Pr\left(\varepsilon_\omega^k(j_k,m_k)<x\right)=e^{-e^{-\theta_k x}}\text{ for }k=C,L, \tag{A.7}$$

where θ_k maps to the (inverse) of the dispersion of shocks across travel modes and destinations for travel purpose $k=C,L$.

The average yearly real income of tract-i residents is defined as the expected value of indirect utility across the realizations of the ε_ω^C and ε_ω^L preference shocks, that is:

$$V(i,s)=\mathbb{E}_\omega\left[\max_{(j_C,j_L,m_C,m_L)\in\mathcal{J}^2\times\mathcal{M}_C\times\mathcal{M}_L}v_\omega(i,j_C,m_C,j_L,m_L,s)\right]. \tag{A.8}$$

Using standard properties of the extreme-value distributions for the shocks ε_ω^C and ε_ω^L, we can write this expression as:[1]

$$V(i,s)\propto\frac{B(i,s)}{r(i,s)^{\mu_H}}\frac{\Omega_C(i,s)}{\Omega_L(i,s)^{\mu_L}}, \tag{A.9}$$

with a constant of proportionality that is independent from i, and where $\Omega_C(i,s)$ captures average

[1] As an intermediate step, we exploit that the idiosyncratic preference shocks are independent, and so are the travel choices:

$$V(i,s)=\frac{B(i,s)}{r(i,s)^{\mu_H}}\mathbb{E}_\omega\left[\max_{j_C,m_C}\left(\frac{I(i,j_C,m_C,s)}{d_C(i,j_C,m_C,s)}\varepsilon_\omega^C(j_C,m_C)\right)\right]\mathbb{E}_\omega\left[\max_{j_L m_L}\left(\frac{q_L(i,j_L)\,B(j_L,s)}{p_L(i,j_L,m_L,s)\,d_L(i,j_L,m_L,s)}\right)^{\mu_L}\varepsilon_\omega^L(j_L,m_L)\right].$$

income net of commuting costs of residents of i,

$$\Omega_C(i,s) = \left(\sum_{j \in \mathcal{J}} \sum_{m \in \mathcal{M}_C} \left(\frac{I(i,j,m,s)}{d_C(i,j,m,s)} \right)^{\theta_C} \right)^{\frac{1}{\theta_C}}, \tag{A.10}$$

and where $\Omega_L(i,s)$ is akin to a quality-adjusted price index for leisure trips for residents of i, net of travel costs:

$$\Omega_L(i,s) \equiv \left(\sum_{j \in \mathcal{J}} \sum_{m \in \mathcal{M}_C} \left(\frac{p_L(i,j,m,s)\, d_L(i,j,m,s)}{q_L(i,j)\, B(j,s)} \right)^{-\mu_L \theta_L} \right)^{-\frac{1}{\mu_L \theta_L}}. \tag{A.11}$$

A.3 Tradeable Sector Firms

In the tradeable sector, we assume an exogenous measure of firms in each tract.[2] Tracts differ in their productivity $A(j,s)$. Each firm uses floor space H_Y, labor N_Y, and business trips T_B as inputs. A firm ω sending workers on T_B business trips to destination j_B using transport mode m_B produces output according to the Cobb-Douglas production function:

$$Y_\omega(j, H_Y, N_Y, R_B, j_B, m_B, s) = A(j,s)\, H_Y^{\mu_{H_Y}} N_Y^{1-\mu_{H_Y}-\mu_B} \left(\frac{q_B(j,j_B)\, A(j_B,s)}{d_B(j,j_B,m_B,s)} T_B \right)^{\mu_B} \varepsilon_\omega^B(j_B, m_B). \tag{A.12}$$

Business trips are factor-productivity enhancing, capturing in reduced form that they promote new supplier or customers. Specifically, the returns to business trips depend on the productivity of the destination $A(j_B,s)$, on an exogenous origin-destination productivity match $q_B(j,j_B)$ (capturing that firms in some locations may on average be more likely to find business partners in some specific locations), and negatively on a function of time traveled, captured by the function $d_B(j,j_B,m_B,s)$ defined as in (A.3). Finally, the return to business trips also depend on an idiosyncratic productivity shock $\varepsilon_\omega^B(j_B,m_B)$ for the destination and travel mode for these business trips. We assume them to be IID Type-I extreme value distributed:

$$\Pr\left(\varepsilon_\omega^B(j_B,m_B) < x\right) = e^{-e^{-\theta_B x}}, \tag{A.13}$$

where θ_B is the (inverse) of the dispersion of shocks across travel modes and destinations for business travel.

We assume that firms hire labor and floor space before observing the realizations of the idiosyncratic business opportunity shocks. Then, they choose the business trip destination (from the set of locations $\mathcal{J}$), the transport mode (from the set of available modes $\mathcal{M}_B$), and the number of trips T_B. Hence, a firm in j solves the problem:

$$\Pi = \max_{H_Y, N_Y} \mathbb{E}\left[\max_{(T_B, j_B, m) \in \left(\mathbb{R}^+ \times \mathcal{J} \times \mathcal{M}_B\right)} Y_\omega(j, H_Y, N_Y, R_B, j_B, m, s) - p_B(j, j_B, m, s)\, T_B \right] - w(j,s)\, N_Y - r(j,s)\, H_Y, \tag{A.14}$$

where $p_B(j,j_B,m_B,s)$ is the monetary cost per roundtrip business trip. Because conditional on floor space and labor there are decreasing returns to the number of trips, we can solve for the number of trips T_B, plug them back into the term within the expectation, and then integrate

[2] We normalize the measure of firms to 1 in every tract. This is without loss of generality, given the heterogeneity in productivity and amenities already allowed across tracts.

over realization of idiosyncratic business shocks using standard properties of the extreme value distribution defined in (A.13).

After these steps, we obtain a closed-form solution for the expected output net of business costs (the term within brackets in (A.14)). Specifically, the firm problem over floor space and labor can be re-written as follows:

$$\Pi = \max_{H_Y, N_Y} \left(\Omega_B (j,s) \, H_Y^{\mu_{H_Y}} N_Y^{1-\mu_{H_Y}-\mu_B} \right)^{\frac{1}{1-\mu_B}} - r(j,s) \, H_Y - w(j,s) \, N_Y. \tag{A.15}$$

where $\Omega_B(j,s)$ is an endogenous TFP term that depends on both the TFP of the location $A(j,s)$ and the distribution of business travel opportunities,

$$\Omega_B(j,s) \equiv \kappa_B A(j,s) \left(\sum_{j_B \in \mathcal{J}} \sum_{m \in \mathcal{M}_B} \left(\frac{q_B(j, j_B) \, A(j_B, s)}{p_B(j, j_B, m_B, s) \, d_B(j, j_B, m_B, s)} \right)^{\theta_B \mu_B} \right)^{\frac{1}{\theta_B}}, \tag{A.16}$$

where we have denoted $\kappa_B \equiv \mu_B^{\mu_B} (1-\mu_B)^{1-\mu_B}$.

A.4 Travel Choices

The travel decisions of workers and firms imply equations for shares and numbers of trips taken to a given destination. We use these equations to estimate key parameters of the model. Specifically, using standard properties of the extreme-value distributions for the shocks ε_ω^C and ε_ω^L, the solution to (A.8) gives the fraction of residents from i that commute to j using transport mode m,

$$\lambda_C(i, j_C, m_C, s) = \frac{\left(\frac{I(i, j_C, m_C, s)}{d_C(i, j_C, m_C, s)} \right)^{\theta_C}}{\sum_{j \in \mathcal{J}} \sum_{m \in \mathcal{M}_C} \left(\frac{I(i,j,m,s)}{d_C(i,j,m,s)} \right)^{\theta_C}}, \tag{A.17}$$

as well as the fraction of residents from i that travel for leisure to j through transport mode m:

$$\lambda_L(i, j_L, m_L, s) = \frac{\left(\frac{q_L(i, j_L) B(j_L, s)}{p_L(i, j_L, m_L, s) d_L(i, j_L, m_L, s)} \right)^{\mu_L \theta_L}}{\sum_{j \in \mathcal{J}} \sum_{m \in \mathcal{M}_L} \left(\frac{q_L(i,j) B(j,s)}{p_L(i,j,m,s) d_L(i,j,m,s)} \right)^{\mu_L \theta_L}}. \tag{A.18}$$

Similarly, from the solution to the firm's problem in A.14 and using standard properties of the extreme value shocks ε_ω^B, the fraction of firms from j sending workers on business trips to j_B takes the same functional form as (A.18):

$$\lambda_B(i, j_B, m_B, s) = \frac{\left(\frac{q_B(i, j_B) A(j_B, s)}{p_B(i, j_B, m_B, s) d_B(i, j_B, m_B, s)} \right)^{\mu_B \theta_B}}{\sum_{j \in \mathcal{J}} \sum_{m \in \mathcal{M}_B} \left(\frac{q_B(i,j) A(j,s)}{p_B(i,j,m,s) d_B(i,j,m,s)} \right)^{\mu_B \theta_B}}. \tag{A.19}$$

The last two expressions measure the shares of travelers (or firms) from location i making leisure (or business trips) to a destination. In the data, we observe the number of trips to each destination by travel purpose. In the model, the number of leisure trips from i to j_L through means m_L depends both on the share of travelers and on the intensity of travel. From the consumer problem (A.1), leisure trips are a constant share $\mu_L(i)$ of disposable income among location-i residents. Adding up this optimal choice across residents of i, we obtain that the total number of leisure trips from i

to j_L using mode m_L is:

$$T_L(i, j_L, m_L, s) = \lambda_L(j, j_L, m_B, s)\,\mu_L(i)\,\frac{N_R(i)\,\overline{I}(i)}{p_L(i, j_L, m_L, s)}, \tag{A.20}$$

where $\overline{I}(i)$ is the average disposable income among location i's residents, itself a function of where they commute for work:

$$\overline{I}(i) \equiv \sum_{j\in\mathcal{J}} \sum_{m\in\mathcal{M}_C} \lambda_C(i, j, m, s)\, I(i, j, m, s)\,. \tag{A.21}$$

Similarly, from the solution for T_B from (A.14), the total number of business trips from i to j_B through mode m_B is:

$$T_B(i, j_B, m_B, s) = \lambda_B(i, j_B, m_B, s)\,\frac{\mu_B}{1-\mu_B}\,\frac{Y(i, s)}{p_B(i, j_B, m_B, s)}. \tag{A.22}$$

Using (A.20) and (A.22) we obtain the gravity equation (12) that we bring to the data.

A.5 Spillovers

Firm productivity and residential amenities may respond endogenously to the level of local activity. We use similar functional forms as Ahlfeldt et al. (2015) and assume that spillovers respond to the density of workers in the location and in the surroundings:

$$A(j, s) = Z_A(j)\left(\sum_{k\in\mathcal{J}} e^{-\rho_A^{\text{spillover}}\tau^{min}(j,k,s)}\,\frac{\tilde{N}_Y(k, s)}{H(k)}\right)^{\gamma_A^{\text{spillover}}} \tag{A.23}$$

$$B(i, s) = Z_B(i)\left(\sum_{k\in\mathcal{J}} e^{-\rho_B^{\text{spillover}}\tau^{min}(j,k,s)}\,\frac{\tilde{N}_Y(k, s)}{H(k)}\right)^{\gamma_B^{\text{spillover}}} \tag{A.24}$$

where

$$\tilde{N}_Y(j, s) = \sum_i \lambda_C(i, j, s)\, N_R(i) \tag{A.25}$$

is the number of workers employed in j and $H(j)$ is the available floor space, so that $\frac{\tilde{N}_Y(j,s)}{H(k)}$ worker density in j and $\tau^{min}(j, k, s)$ is the fastest travel time across all modes over a given route. In Ahlfeldt et al. (2015), the congestion at residence (denominated B here) depends on how many people live around an area, while the agglomeration at destination (denominated A here) depends on how many people work around an area. Since we assume a fixed number of residents, in our case both spillovers are a function of the endogenous number of workers in the surrounding areas. Similarly, in Ahlfeldt et al. (2015), the surrounding density is discounted by the ρ elasticities times travel time. Since we have multiple travel mode, in our case we use the fastest travel time across all travel modes.

A.6 High-Speed Rail Use

The travel time $\tau_k(i, j, m, s)$ and the roundtrip monetary cost $p_k(i, j, m, s)$ introduced so far for each travel purpose $k = C, L, B$ correspond to the route chosen in the transport network to travel from origin i to destination j through mode m for purpose k in state s.

Without the CHSR (i.e. when $s = N$) we use the travel time corresponding to the fastest route by origin and destination for a given mode, and assign the monetary cost of this route. When the CHSR is available (i.e. when $s = Y$), the CHSR is treated as a perfect substitute to traveling via public transit (for commuters or long-distance travelers) or by air (for long-distance travelers). That is, the CHSR is used within mode m for each pair (i, j) if and only if the CHSR-based time-cost pair (τ, p) leads to a higher indirect utility than the pre-CHSR one.

Specifically, given the indirect utility (A.6), the commuting time and monetary cost when the CHSR is available, $\{\tau_C(\cdot, Y), p_C(\cdot, Y)\}$, is the solution to:

$$\max_{\{\tau_C(\cdot,N),p_C(\cdot,N)\},\{\tau^{CHSR}(\cdot),p^{CHSR}(\cdot)\}} \frac{(1-t)\,y(i,j,Y) - pT_C}{\tau^{\rho}}. \tag{A.26}$$

Similarly, as implied by the indirect utility (A.6) and by the definition of business productivity (A.16), the time and monetary cost for a leisure ($k = L$) or business ($k = B$) traveler when the CHSR is available is the $\{\tau_k(\cdot, Y), p_k(\cdot, Y)\}$ that solves

$$\min_{\{\tau_C(\cdot,N),p_C(\cdot,N)\},\{\tau^{CHSR}(\cdot),p^{CHSR}(\cdot)\}} p\tau^{\rho}. \tag{A.27}$$

A.7 Equilibrium Conditions

Equilibrium in the labor market of tract i dictates that the demand for efficiency units of labor in a location equals the supply of efficiency units to that location, $N(i, s)$:

$$N_Y(i,s) = \underbrace{\sum_{j\in\mathcal{J}} \lambda_C(j,i,s)\, e(j)\, N_R(j)}_{\equiv N(i,s)} \tag{A.28}$$

where $\lambda_C(j, i, s)$ fraction of commuters from j to i through any mode:

$$\lambda_C(j,i,s) \equiv \sum_{m\in\mathcal{M}_C} \lambda_C(j,i,m,s)\,. \tag{A.29}$$

Next, using the solution for consumer demand for floor space from (A.1) and for firm's demand for floor space and labor from (A.14), the equilibrium in the housing markets is:

$$N_R(i)\,\frac{\mu_H(i)\,\overline{I}(i)}{r(i,s)} + N_Y(i,s)\,\frac{w(i,s)}{r(i,s)}\,\frac{\mu_{H_Y}(i)}{1-\mu_{H_Y}(i)-\mu_B(i)} = H(i)\,, \tag{A.30}$$

where the first term in the left-hand side is the demand for floor space coming from residents of i, the second term is demand coming from firms located in i, and $H(i)$ is the supply of floor space in i. Finally, since tradeable firms operate subject to constant returns, the zero-profit conditions resulting from (A.14) dictates:

$$w(j,s)^{1-\mu_B-\mu_{H_Y}}\, r(j,s)^{\mu_{H_Y}} = \kappa_\Omega \Omega_B(j) \tag{A.31}$$

for some constant κ_Ω that is a function of μ_B and μ_{H_Y}.

An equilibrium consists of distributions of land prices $r(j, s)$, wages $w(j, s)$, and supplies of labor into tradeables $N_Y(i, s)$, such that:

i) the land market clearing condition (A.30) holds for all tracts;

ii) the labor market clearing condition (A.28) holds for all tracts i; and

iii) the zero-profit condition (A.31) holds for all tracts j.

Note that the system of equations defined by (A.30)-(A.31) include as unknowns the endogenous productivity term $\Omega(j)$, the agglomeration and amenity spillover functions $A(j,s)$ and $B(i,s)$, and the average income $\overline{I}(i)$. Using (A.16), (A.23), (A.24), and (A.21), all these endogenous variables can be expressed as functions of the endogenous variables $\{r(j,s), w(j,s), N_Y(i,s)\}$ which define the equilibrium.

A.8 System for Counterfactual Analysis

In this section we derive the system that we implement when running counterfactuals. For this, we now move to express the equilibrium value of every endogenous outcome in a scenario where $s = Y$ relative to its value in an equilibrium where $s = N$. We let

$$\hat{X}(\cdot) \equiv \frac{X(\cdot, Y)}{X(\cdot, N)}$$

be the ratio of variable X between its equilibrium value when $s = Y$ (so that the CHSR will be built with some probability) and when $s = N$ (the CHSR is not built).

CHSR Shock Starting from an initial equilibrium, the previous system of equilibrium conditions is impacted by potentially different travel times and monetary travel costs. Specifically, the shock to the system is given by time changes,

$$\hat{\tau}_k(i,j,m)$$

and by monetary travel cost changes,

$$\hat{p}_k(i,j,m)$$

for each travel purpose $k = C, L, B$ (commuting, leisure, or business travel). The pre- and post-CHSR levels of these variables are defined in (A.26) and (A.27). On route-mode combinations (i,j,m) where travelers do not choose CHSR, these shocks are $\hat{\tau}_k(i,j,m) = \hat{p}_k(i,j,m) = 1$. On route-mode combinations where CHSR is preferred to the pre-existing mode then either $\hat{\tau}_k(i,j,m) < 1$, $\hat{p}_k(i,j,m) < 1$, or both. To construct these shocks, we use the pre- and post-CHSR travel times and costs following the discussion in Section A.6. When $s = Y$, then disposable income also changes with the tax rate in a common away across locations:

$$\widehat{1-t} = 1 - t.$$

Equilibrium System in Relative Changes The equilibrium response to $\{\hat{\tau}(i,j,m), \hat{p}_k(i,j,m), 1-t\}$ consists in changes in land rents $\hat{r}(i)$, wages $\hat{w}(i)$, and labor supplies $\hat{N}_Y(i)$ such that:

i) The land market clears, i.e. (A.30) holds in the counterfactual equilibrium, which implies:

$$\hat{r}(i) = \frac{H_C(i,N)}{H(i)}\hat{\overline{I}}(i) + \left(1 - \frac{H_C(i,N)}{H(i)}\right)\hat{w}(i)\hat{N}_Y(i), \tag{A.32}$$

where $H_C \equiv N_R(i) \frac{\mu_H(i)\overline{I}(i)}{r(i,s)}$ is the aggregate housing demand in i and $\hat{\overline{I}}(i)$ is the change in average income of residents of i defined in (A.21),

$$\hat{\overline{I}}(i) = \sum_{j \in \mathcal{J}} \sum_{m \in \mathcal{M}_C} \frac{\lambda_C(i,j,m,N)}{\overline{I}(i,N)} \hat{\lambda}^C(i,j,m)\,\hat{I}(i,j,m), \tag{A.33}$$

where the change in disposable income next of taxes and commuting costs for commuters from i to j using mode m is

$$\hat{I}(i,j,m) = (1+\chi(i,j,m))(1-t)\,\hat{y}(i,j) - \chi(i,j,m)\,\hat{p}_C(i,j,m,N), \tag{A.34}$$

where $\chi(i,j,m) \equiv \frac{T_C p_C(i,j,m,N)}{I(i,j,m,N)}$ is the share of commuting costs in disposable income in the pre-CHSR scenario, and where the change in pre-tax income is

$$\hat{y}(i,j) = \left(1 - \frac{e(i)\,w(j,N)}{y(i,j,N)}\right)\hat{\overline{I}}(i) + \frac{e(i)\,w(j,N)}{y(i,j_C,N)}\hat{w}(j), \tag{A.35}$$

and, from (A.17), $\hat{\lambda}_C(i,j,m)$ is given by:

$$\hat{\lambda}_C(i,j_C,m_C) = \frac{\left(\frac{\hat{I}(i,j_C,m_C)}{\hat{d_C}(i,j_C,m_C)}\right)^{\theta_C}}{\sum_{j \in \mathcal{J}} \sum_{m \in \mathcal{M}_C} \lambda_C(i,j,m,N)\left(\frac{\hat{I}(i,j,m)}{\hat{d_C}(i,j,m)}\right)^{\theta_C}} \tag{A.36}$$

ii) the labor market clears, i.e. (A.28) holds in the counterfactual equilibrium, which implies

$$\hat{N}_Y(i) = \sum_{j \in \mathcal{J}} \left(\frac{\lambda_C(j,i,N)\,e(j)\,N_R(j)}{N(i,s)}\right)\hat{\lambda}_C(j,i) \tag{A.37}$$

where, in the supply side in the right-hand side of (A.37), $\hat{\lambda}_C(j,i)$ is given by

$$\hat{\lambda}_C(j,i) = \sum_{m \in \mathcal{M}_C} \left(\frac{\lambda_C(j,i,m,N)}{\lambda_C(j,i,N)}\right)\hat{\lambda}_C(j,i,m). \tag{A.38}$$

iii) the zero-profit condition (A.31) holds in a counterfactual scenario, i.e.

$$\hat{w}(j)^{1-\mu_B-\mu_{H_Y}}\,\hat{r}(j)^{\mu_{H_Y}} = \hat{\Omega}_B(j), \tag{A.39}$$

where, from (A.16), $\hat{\Omega}_B(i)$ is given by (10) in the text.

From (A.23), the agglomeration component of TFP in $\hat{\Omega}_B$ changes according to:

$$\hat{A}(j) = \left(\sum_{k \in \mathcal{J}} \frac{\tilde{N_Y}(k,N)/H(k)}{\sum_{k' \in \mathcal{J}} e^{-\rho_A^{\text{spillover}}\tau^{min}(j,k',N)}\tilde{N_Y}(k',N)/H(k')} e^{-\rho_A^{\text{spillover}}\tau^{min}(j,k,B)}\hat{\tilde{N}}_Y(k)\right)^{\gamma_A^{\text{spillover}}} \tag{A.40}$$

where from (A.25) the change in the number of workers employed in j is:

$$\hat{\tilde{N}}_Y(j) = \sum_i \left(\frac{\lambda_C(i,j,N)\,N_R(i)}{\tilde{N_Y}(j,s)}\right)\hat{\lambda_C}(i,j). \tag{A.41}$$

for $\hat{\lambda_C}(i,j)$ defined in (A.38).

A.9 Welfare Changes

From (7) to (9) we obtain the following expressions for the welfare change:

$$\hat{V}(i) = \left(\frac{\hat{B}(i)}{\hat{r}(i)^{\mu_H}}\right)\hat{\Omega}_C(i)\,\hat{\Omega}_L(i)\,. \tag{A.42}$$

The commuting component $\hat{\Omega}_C(i)$ changes according to (8), with $\hat{I}(i,j,m)$ in that expression given by (A.34). The leisure changes according to

$$\hat{\Omega}_L(i) \equiv \left(\sum_{j\in\mathcal{J}}\sum_{m\in\mathcal{M}_L}\lambda_L(i,j,m)\left(\frac{\hat{B}(j)}{\hat{p}_L(i,j,m)\,\hat{\tau}(i,j,m)^{\rho}}\right)^{\mu_L\theta_L}\right)^{\frac{1}{\theta_L}}. \tag{A.43}$$

In these expressions, endogenous amenities component $\hat{B}(i)$ satisfies a similar equation to $\hat{A}(j)$ in (A.40):

$$\hat{B}(j) = \left(\sum_{k\in\mathcal{J}}\frac{\tilde{N_Y}(k,N)/H(k)}{\sum_{k'\in\mathcal{J}}e^{-\rho_B^{\text{spillover}}\tau^{min}(j,k',N)}\tilde{N_Y}(k',N)/H(k')}e^{-\rho_B^{\text{spillover}}\tau^{min}(j,k,B)}\hat{\tilde{N}}_Y(k)\right)^{\gamma_B^{\text{spillover}}}. \tag{A.44}$$

Conditions (7) and (9) in the text follow from (A.42) and (A.43) when we set $\hat{B}(i) = \hat{r}(i) = 1$.

B Data Sources

Geographic Units The analysis is conducted at the tract level. Our sample comprises 7866 out 8057 census tracts in California's mainland that are populated and have positive employment and no missing data (U.S. Census (2008b)and U.S. Census Bureau (2010a)). These tracts account for 98.5% of the state's population and 97.6% of all tracts.

Voting We obtain data on the number of favorable and negative votes by precinct for Proposition 1A and for other ballots in 2006 and 2008 from the UC Berkeley's Statewide Database (University of California, Berkeley (2008)). We also use their crosswalk to construct a tract-level dataset of votes.

Commuting Data on commuting flows are taken from the American Community Survey (ACS) (U.S. Census Bureau, 2010b). The American Community Survey reports tract-to-tract data on commuting by transport mode as a part of the Census Transportation Planning Products.[3] To measure commuting flows, respondents answer the question: "At what location did this person work LAST WEEK?". We construct the flows excluding work-from-home workers, corresponding to less than 5% of statewide workers in this period. In addition, to measure the mode of travel, respondents are asked: "How did this person usually get to work LAST WEEK?". We classify car, truck, or van as the "car" mode. We classify the bus, subway, commuter rail, light rail, or ferry as

[3] The LEHD also reports commuting flows by origin and destination based on administrative data linking employees home locations with their employer's location. As we do not observe the frequency at which these trips are taken, these origin-destination flows may not reflect regular commuting.

the "public transit" mode. Finally, we classify the remainder, which includes biking and walking as the "walking or biking" mode.

Leisure and Business trips Leisure and business trips are compiled from the California Household Travel Survey (CAHTS) conducted between 2010 and 2012. The CAHTS records trips longer than 50 miles taken over a 8-week survey period. 18,008 households and 68,193 trips appear in the dataset. The data include information on the origin, destination, and residence census tract of each trip, the number of people on each trip, the travel mode, and the purpose of the trip. We classify each trip into a leisure trip if the purpose includes entertainment, vacation, shopping or visiting friends and family. The top leisure destinations are Disneyland, Yosemite, Mission Beach (San Diego), Downtown San Francisco, and Downtown San Diego. We classify each trip into a business trip of the purpose includes business meetings, conventions, or seminars. The top business destinations are the State Capitol in Downtown Sacramento, Downtown Los Angeles, Downtown San Francisco, and Downtown San Diego. Taken together, leisure and business trips account for 84% of all trips in the survey. The remaining trips include combined business and pleasure trips, medical trips, school-related activities, and trips for which the purpose is not stated.

Wages We use data on wages by workplace Census tract and residence Census tract from the 2008 and 2019 samples of the Longitudinal Employer-Household Dynamics (LEHD) Origin-Destination Employment Statistics published by the U.S. Census (U.S. Census (2008a)). The LEHD reports the number of workers in each workplace-residence Census tract pair who have monthly earnings below $1,250, between $1,250 and $3,333, and above $3,333. To construct an average wage along each route, we first measure average earnings within each of these three bins within California using the individual level American Community Survey samples in 2008 and 2019. We use these bins and our estimates of average earnings within each bin to compute average earnings among workers within each workplace-residence tract pair.[4]

Population and Demographics We define the number of working-age residents $N_R(i)$ entering in the model quantifications using the distribution of commuters originating in each census tract from each ACS with a re-scaling to match the working-age population of California according to the BLS.

We measure share of non-white residents, occupational composition, the share of residents with a college degree, and demographic covariates by census tract from the 2006-2010 and 2015-2019 American Community Survey five-year estimates.

Construction of Additional Variables Using the previous sources we construct additional variables needed for the implementation of the model. We construct disposable income $y(i, j_C, N)$ in (A.5) using its definition as the sum of labor income and locally owned land rents. We construct

[4] We use the 2012-2016 sample because the 2016-2020 sample is not yet available, and also encompasses the start of the COVID-19 pandemic, which saw large changes in commuting patterns.

the land rent component of income as $\eta(i) r(i,N) \equiv \frac{\eta_R(i) r(i) H_C(i,N)}{N_R(i)}$, where $\eta_R(i)$ is the share of homes that is owner-occupied from ACS and $r(i) H_C(i,N)$ are residential home values from ACS transformed to annualized rent-equivalent values. We measure the share of land used for residential purposes, $H_C(i,N)/H(i)$ entering in (A.32), using Zillow's ZTRAX data . We construct the disposable income $I(i, j_C, m_C, N)$ defined in (A.4) using $y(i, j_C, N)$, the round-trip commuting costs described in the next section, and $T_C = 250$ commuting trips throughout the year.

Additional Parameter Calibration We calibrate the remaining parameters using estimates in the literature. We assume that the share of firm expenditure on floor space is $\mu_{H,Y} = 0.20$, in line with Valentinyi and Herrendorf (2008). We use estimates from Ahlfeldt et al. (2015) to calibrate the productivity and amenity spillovers from equations A.23 and A.24. Specifically, using their preferred estimate from Table V, column 3, we assume:$\gamma_A^{\text{spillover}} = 0.071$, $\gamma_B^{\text{spillover}} = 0.155$ and $\rho_A^{\text{spillover}} = 0.361$, $\rho_B^{\text{spillover}} = 0.759$.

C Transport Network Details

Transport Network We construct a transport network to calibrate times and costs from each census tract origin to each census tract destination, for each possible transport mode: car, public transit (combining bus and rail stations), air travel, biking, and CHSR. The centroid is defined as the geographic centroid of the most populous Census block within each tract. We construct the road network based on the 2010 primary and secondary road shapefile for California obtained from the U.S. Census, which we transform into a graph. We connect the tract centroids to the road network by creating links from the centroids to closest node on the road network. The resulting road network has 72790 edges and 71957 nodes, of which 7866 nodes correspond to the tract centroids in our analysis.

To this network, we add a rail network with 185 train stations obtained from California's Department of Transportation for 2013, which is the closest available year to 2008.[5] We construct the air network including the 10 largest airports in California: LAX, SFO, SAN, OAK, SJC, SNA, SMF, ONT, BUR, LGB. We include the 24 unique air routes operating among these airports according to the Bureau of Transportation 2008 airline ticket dataset (Bureau of Transportation Statistics, 2008a).[6] For the CHSR, we obtain a shapefile of the planned route and stations at the time of the vote in November of 2008 from the University of California at Davis.[7] This shapefile includes the planned 24 stations plus two potential stations (Irvine and Tulare). We exclude these two additional stations, as a total of 24 stations is consistent with the description of the network in the original CHSR bill passed by the California legislature before the 2008 vote.

The resulting multi-modal transport networks includes the road network expanded with tract centroids, rail stations, airports, and the CHSR. We create artificial edges that connect the rail

[5] Available at: https://geodata.lib.utexas.edu/catalog/stanford-xd213bw5660.
[6] Available at: https://www.transtats.bts.gov/DatabaseInfo.asp?QO_VQ=EFI&Yv0x=D.
[7] Available at: https://databasin.org/datasets/7a9f1867f2e24a1e97ab10419a73b25a/.

stations, airports, and CHSR stations to their closest node on the original road network. Figure A.1 displays the roads, rail stations, airports, and CHSR line with stations used throughout the paper.

Travel Times We calibrate speeds by private car and public bus by assigning a travel time to each edge of the road network to match travel times by car only and by public transit only from Google Maps on a random sample of 10,000 origin and destination tracts. We assign travel times by multiplying the arc-length kilometer distance of each edge by its average speed, using one among 5 possible speed categories for each edge depending on its features (primary urban, primary rural, secondary urban, secondary rural, and artificial centroid-node link). We add a time constant to every trip that is independent from distance and captures waiting times.

All fastest routes are computed using the fast marching method. The fastest route via car is computed on the calibrated road network. The fastest route via public transit is defined as the fastest between traveling via public bus or via rail for each origin-destination pair. The fastest route via rail assumes that travelers use a car to the stations nearest to the origin and destination tracts, excluding origin-destination pairs where driving times to rail stations are greater than 2 hours one-way or the nearest station to the origin and destination tracts are the same. We use station-to-station rail times available in the websites of rail systems in California (ACE Rail, Amtrak, BART, CalTrain, Coaster, and Metrolink) and wait time of 17 minutes at the origin rail station.

Fastest routes on bike on the road network are calculated assuming an average speed of 20 km per hour in urban environments.

The fastest travel time by air is computed assuming road speeds to and from airports, allowing travelers to use any airport regardless of distance to origin and destination tract, using flight times from Google Maps and assuming a wait time of 90 minutes at the origin airport.

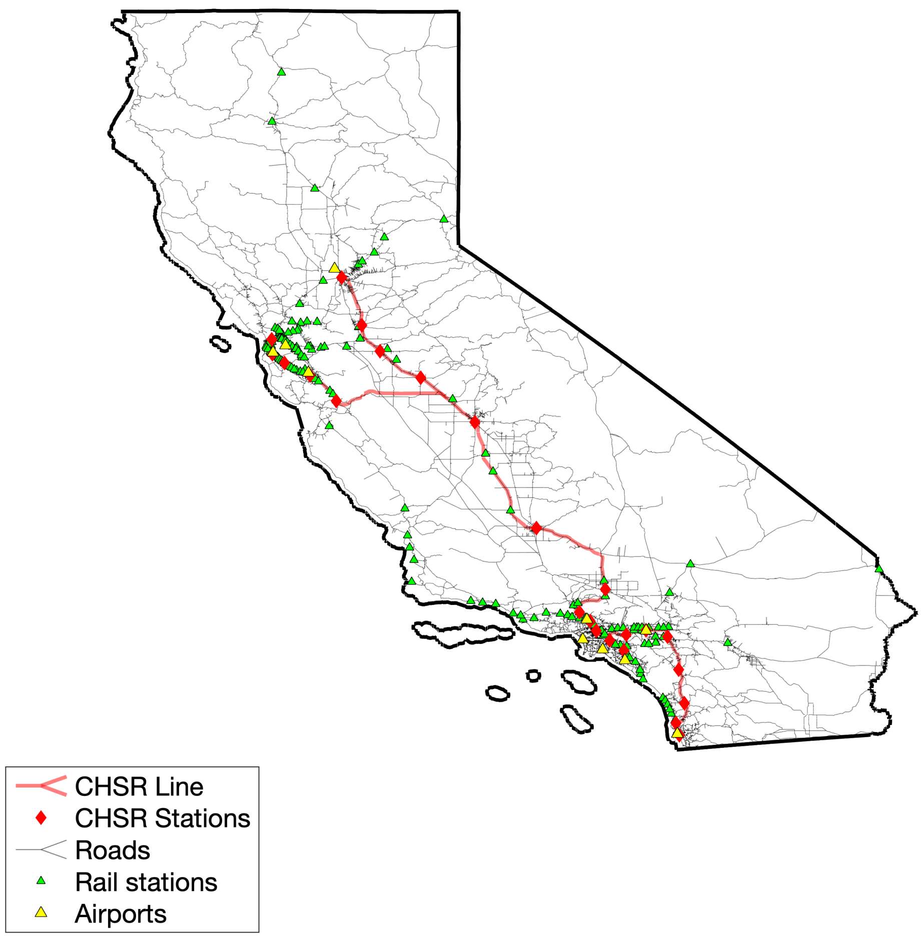
Paths and times via CHSR are defined similarly to via rail, with travelers using the stations closest (by car) to the origin and destination tracts. We use planned speeds between contiguous stations of the CHSR network from the 2008 Speed Rail Authority's Business Plan (California High Speed Rail Authority, 2008). Specifically, from the 2008 business plan we assign a speed of either 125, 175, or 220 miles per hour. The resulting pairwise travel times between all stations closely match those reported between major stations. In Section 6.3, when optimizing over stations along the alternative I5 design, we set the speed on each point of the I5 route equal to the speed on its closest point along the original CHSR design.

Travel Costs For car, the cost of travel from each origin to each destination on a given route is computed based the per-mile cost of fuel assuming a cost of $3.50 per gallon[8] and fuel efficiency of 21 miles per gallon[9]. The cost of traveling via bus equals the average one-way adult bus ticket price in the county where the origin census tract is located, according to the American Public Transportation Association (American Public Transit Association, 2010), and complemented with

[8]Source: 2008 Los Angeles Almanac, available at http://www.laalmanac.com/energy/en12.php.
[9]Source: US Department of Energy, available at https://www.fueleconomy.gov/feg/byclass/Midsize_Cars2008.shtml/.

data from each county's website when necessary (average bus fare across all counties is $2.15). A fixed and variable per-mile cost of traveling via rail is estimated for the Amtrak Capital Corridor in Northern California and applied to the entire rail network. This estimation yields a rail fare with a fixed cost of $2.92 and an extra cost of $0.20 per mile. The cost of air travel on every route is set to the average one-way ticket cost of $151 across routes according to the Bureau of Transportation Statistics.[10] To calculate the ticket price of the CHSR, we use as basis either the $55 ticket price from LA to San Francisco from the 2008 Business Plan (California High Speed Rail Authority, 2008) or the update of this number to $110 in the 2022 plan (California High Speed Rail Authority, 2022). We project these costs to the full HSR network using the fixed cost estimated for the Amtrak Capitol Corridor.

Figure A.1: Transport Network



Note: the figure shows the transport network used throughout the paper, including roads, rail stations, airports, and the CHSR.

[10] Available at https://www.transtats.bts.gov/DatabaseInfo.asp?QO_VQ=EFI&Yv0x=D .

D Implementation of the Gravity Equations

D.1 Commuting

We discuss the details of how we estimate the parameters of the gravity equation for commuting in 11. First, we measure disposable income $I(i,j,m)$, defined in A.4, as the difference between labor income and commuting costs. Consistent with the model, the labor income term is the product of an origin-specific component $e(i)$, which accounts for the possibility that workers that reside in different locations have different human capital, and a destination-specific component $w(j)$, which accounts for productivity differences across workplaces. We estimate these origin- and destination-specific components using the following estimating equation: $w^{data}(i,j) = \exp(\tilde{e}(i) + \tilde{w}(j)) + \varepsilon(i,j)$, where $w^{data}(i,j)$ denotes the observed average wage of workers who reside in i and work in j, $\tilde{e}(i) \equiv \ln(e(i))$, $\tilde{w}(j) \equiv \ln w(j)$, and $\varepsilon(i,j)$ accounts for factors affecting observed average wages that cannot be accounted for by an origin-specific and a destination-specific fixed effect. We assume that $\varepsilon(i,j)$ does not impact workers' commuting decisions and is mean-independent of the origin- and destination-specific components; e.g., it captures either measurement error in wages or wage shocks unexpected at the time of making commuting decisions. We take this approach because we do not observe $w^{data}(i,j)$ for every (i,j) pair.

Second, we consider as feasible commuting choices any pair of origin and destination census tracts in CA and transport mode such that the travel time is either less than 4 hours (when using either car or public transport) or less than 2 hours (when biking). As a result, we use about 33 million origin-destination pairs between which it is feasible to commute by car, 21 million pairs for public transit, and 5 million pairs for bike. For all of the 7,866 potential origin locations considered in our analysis, there is at least one destination that may be reached by car and at least one destination may reached by public transport.

Finally, because our information on commuting comes from a finite sample of residents, we allow for the possibility that the observed commuting shares $\lambda_C^{obs}(i,j,m)$ differ from the true ones, $\lambda_C(i,j,m)$, by a term $error_C(i,j,m)$ that captures sampling error: $\lambda_C^{obs}(i,j,m) = \lambda_C(i,j,m) + error_C(i,j,m)$.

We perform the estimation in two steps. In the first step, to estimate θ_C and ρ_C, we use variation in the choice of destination conditional on origin and transport mode. We use the two moment conditions

$$\mathbb{E}\left[\left(\frac{\lambda_C^{obs}(i,j,m)}{\sum_{j'}\lambda_C^{obs}(i,j',m)} - \frac{\lambda_C(i,j,m)}{\sum_{j'}\lambda_C(i,j',m)}\right)X(i,j,m)\right] = 0, \tag{A.45}$$

where $X(i,j,m) = (\ln(I(i,j,m)), \ln\tau(i,j,m))'$. We build sample analogues of these moment conditions by averaging across origins i, destinations j, and modes of transport m.

In the second step, to estimate the preferences $D_C(i,m)$ that residents of a census tract i have for a particular transport mode m, we model the origin- and mode-specific term $D_C(i,m)$ as a function of observed origin-specific covariates $X_C(i)$ with mode-specific coefficients $\Psi_C(m)$:

$$D_C(i,m)^{-\theta_C} \equiv \exp(\Psi_C(m)X_C(i)). \tag{A.46}$$

The vector $X_C(i)$ includes a constant, the share of residents who own a car, the share of residents who are under 30, the share of college-educated residents, the share of nonwhite residents, the log median income, and the log population density. To estimate these parameters, we use variation across origins in the share of commuters that use each transport mode m. Specifically, we use the mode m-specific moment conditions

$$\mathbb{E}\left[\left(\sum_j \lambda_C^{\text{obs}}(i,j,m) - \sum_j \lambda_C(i,j,m)\right) X_C(i)\right] = 0. \tag{A.47}$$

We build a sample analogue of these mode-specific moment conditions by taking an average across all origin census tracts i.

D.2 Business and Leisure

We describe the details of our procedure to estimate the gravity equation for long-distance leisure and business trips in (12). The size of the sample containing information on leisure and business trips is much smaller than that containing information on commuting trips and, in particular, contains no information for the residents of certain census tracts, making it impossible to estimate as free parameters the origin-destination and origin-mode unobserved effects $Z_k(i,j)$ and $D_k(i,m)$ in (12). To sidestep this data limitation, we write $Z_k(i,j)$ and $D_k(i,m)$ as a function of observable characteristics. Specifically, we let:

$$\left(\frac{Z_k(i,j)}{D_k(i,m)}\right)^{\mu_k\theta_k} \equiv \exp\left(\gamma_k(m) + \Psi_k X_k(j)\right) \tag{A.48}$$

for $k = L, B$, where $\gamma_k(m)$ is a mode and travel purpose-specific shifter, Ψ_k is a vector of purpose-specific parameters, and $X_k(j)$ is a vector of observed characteristics. In our empirical specification, $X_L(j)$ includes proxies for the amenity value of a destination: the log distance between j and the closest beach, a dummy variable for whether j is in a national park, the share of workers in j employed in the hospitality sector, and the log total population. The vector $X_B(j)$ includes the share of workers in management roles in the destination tract and its log total population.

We measure the share of business and leisure trips $\tilde{\lambda}_k(i,j,m)$, travel time $\tau(i,j,m)$, and travel costs $p_k(i,j,m)$ as indicated in Section 4.1. Given the limited size of our sample on business and leisure trips, we account for sampling error in our measure of $\tilde{\lambda}_k(i,j,m)$, which we denote as $\tilde{\lambda}_k^{obs}(i,j,m)$.

The separate identification of θ_k and ρ arises from the response of $\tilde{\lambda}_k(i,j,m)$ to the travel time variable $\tau(i,j,m)$ and the monetary cost term $p_k(i,j,m)$. While we follow standard procedures to measure travel times and, thus, are reasonably confident of its accuracy, our measure of the monetary cost of traveling between any two census tracts likely suffers from substantial measurement error. Consequently, we treat the term $p_k(i,j,m)$ merely as a control and its associated coefficient as a nuisance parameter, and assume that ρ equals the corresponding parameter entering the commuting equation.

We estimate the remaining parameters following a two-step estimation approach similar to that described in Section D.1 for commuting. In the first step, we identify θ_k and β_k through the following moment condition:

$$\mathbb{E}\left[\left(\frac{\tilde{\lambda}_k^{obs}(i,j,m)}{\sum_{j'\in\mathcal{J}_k}\tilde{\lambda}_k^{obs}(i,j',m)}-\frac{\tilde{\lambda}_k(i,j,m)}{\sum_{j'\in\mathcal{J}_k}\tilde{\lambda}_k(i,j',m)}\right)X_k(i,j,m)\right]=0, \tag{A.49}$$

for $k=L,B$, where $X_k(i,j,m)=(X_k(j),\ln\tau(i,j,m),\ln p_k(i,j,m))'$.

For each origin tract i, the moment above sums over j' in the denominators of the left hand side of (A.49). In the case of trips performed by airplane, we further exclude from the choice set those pairs of origin and destination tracts for which the travel time by airplane is larger than by car. The sample analogue of the moments in (A.49) averages over approximately 24 million pairs of origin and destination tracts among which it is feasible to travel by airplane, and over approximately 52 million pairs of tracts that may be reached by car or public transport.

In a second step, after normalizing the mode-specific shifter $\gamma_k(m)$ to equal zero for $m=$ airplane, we identify $\gamma_k(m)$ for $m=$ private vehicle and $m=$ public transport using variation across origins in the share of travelers that use each transport mode m, through the following m-specific moment conditions

$$\mathbb{E}\left[\sum_j\tilde{\lambda}_k^{\text{obs}}(i,j,m)-\sum_j\tilde{\lambda}_k(i,j,m)\right]=0. \tag{A.50}$$

We build a sample analogue of these mode-specific moment conditions by taking an average across all origin census tracts i.

E Additional Tables and Figures

Table A.1: Commuting Equation Estimates, Second Step

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Public Transport								
Constant	-1.05^{a}	-0.00	0.17	-0.93^{a}	-2.69^{a}	-4.47^{a}	-8.00^{a}	-15.69^{a}
	(0.05)	(0.13)	(0.29)	(0.12)	(0.10)	(1.56)	(0.34)	(1.09)
Sh. Car Owners		-1.18^{a}						0.49^{b}
		(0.16)						(0.22)
Sh. Under 30			-2.68^{a}					-2.57^{a}
			(0.59)					(0.33)
Sh. College-educated				-0.24				-1.26^{a}
				(0.25)				(0.21)
Sh. Nonwhite					3.78^{a}			2.96^{a}
					(0.24)			(0.18)
Log Median Inc.						0.31^{b}		0.91^{a}
						(0.14)		(0.09)
Log Pop. Density							0.94^{a}	0.61^{a}
							(0.05)	(0.03)
Private Vehicle								
Constant	0.95^{a}	-3.29^{a}	2.52^{a}	1.36^{a}	0.39^{a}	-5.30^{a}	-0.38^{c}	-15.74^{a}
	(0.04)	(0.10)	(0.25)	(0.09)	(0.08)	(1.20)	(0.22)	(0.82)
Sh. Car Owners		5.07^{a}						5.84^{a}
		(0.12)						(0.17)
Sh. Under 30			-3.50^{a}					-1.53^{a}
			(0.50)					(0.24)
Sh. College-educated				-0.94^{a}				-0.84^{a}
				(0.19)				(0.16)
Sh. Nonwhite					1.29^{a}			1.91^{a}
					(0.21)			(0.14)
Log Median Inc.						0.56^{a}		0.85^{a}
						(0.11)		(0.07)
Log Pop. Density							0.18^{a}	0.35^{a}
							(0.03)	(0.02)
Num. Obs.	23593	23593	23593	23593	23593	23593	23593	23593

Note: ^{a} denotes 1% significance; ^{b} denotes 5% significance; and ^{c} denotes 10% significance. Robust standard errors are displayed in parenthesis. All specifications are conditional on the estimates $\hat{\theta}_C = 3.35$ and $\hat{\rho}_C = 0.21$. The mode of transport excluded from the specification is "walk/bike."

Table A.2: Model Variants

	Indirect impacts	Costs and Expectations	
	(land, wages, spillovers)	Capital and Maintenance Costs	Completion Probability and Timeline
2008 Business Plan	no	2008 Forecasts	
2008 Business Plan, with GE	yes		
2022 Business Plan	no	2022 Forecasts	$p = 0.65$, $T = 33$
2022 Business Plan, with GE	yes		

Note: this table summarizes the differences across counterfactual scenarios. In the third column, T is the minimum number of years after the vote until the full CHSR project is operational, and p is the probability that it will become operational in each year after T if it has not done so before.

Table A.3: Determinants of Heterogenous Gains from CHSR

	2008	2008 GE	2022	2022 GE
	(1)	(2)	(3)	(4)
	b	b	b	b
Distance to nearest CHSR station	-0.00024***	-0.00039***	-0.00007***	-0.00011***
% public transit in commuting	0.01223***	0.01272***	0.00488***	0.00520***
% air in long distance travel	0.00852***	0.02653***	0.00139***	0.00416***
LA fixed effect	0.00027***	0.00042***	0.00012***	0.00010***
SF Bay Area fixed effect	0.00005**	-0.00016***	0.00009***	-0.00000
Central Valley fixed Effect	-0.00011***	0.00015***	-0.00005***	0.00002**
R2	0.864	0.809	0.859	0.774
N	7,866	7,866	7,866	7,866

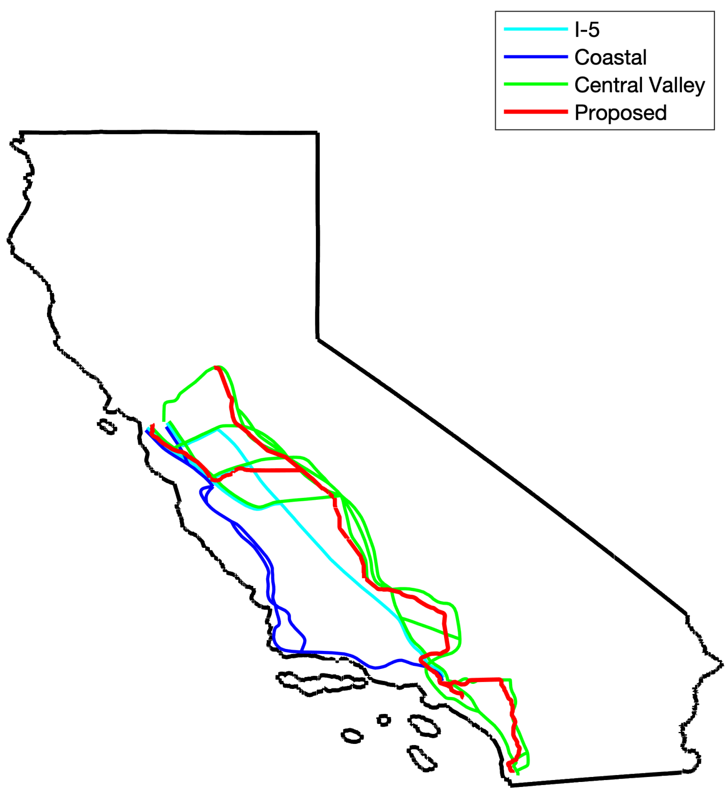
Note: The table shows, for each model variant, an OLS regression of the tract-level real-income change, $\hat{W}(i)$, on covariates. “Distance to nearest CHSR station” is the tract centroid’s log-distance to the nearest CHSR station; “% public transit in commute” is fraction of residents who, in the observed equilibrium without the CHSR, commute via public transit; “% air in long distance” is fraction of residents who, in the observed equilibrium without the CHSR, make long distance trips via air (defined as an average of that fraction for business and leisure travelers); and the LA, SF Bay Area and CV fixed effects correspond to dummies for whether the census tract is within LA county, one of the SF Bay Area counties, or one of the 18 counties in California’s Central Valley, respectively.

Table A.4: Gain by Type of Travel

	2008 Business Plan		2022 Business Plan	
		+GE		+GE
Upfront Tax	-0.09%	-0.09%	-0.28%	-0.28%
Commute (net of tax)	0.04%	0.02%	-0.04%	-0.03%
Leisure	0.05%	0.05%	0.01%	0.01%
Business	-	0.06%	-	0.01%
Total	-0.01%	0.13%	-0.31%	-0.28%

Note: This table splits the aggregate welfare change (defined as a population-weighted average of the $\hat{W}(i)$ in (6)) into the upfront tax component $(1-R)\ln(1-t)$ paid prior to the CHSR being operational (in the first row) and the annualized net gain conditional on the CHSR being operational, $R\ln\hat{V}(i)$. This net gain is further decomposed, from (7), into its commuting component $\hat{\Omega}_C(i)$ in (8) and the leisure component $\hat{\Omega}_L(i)$ in (9). The gains from commuters are computed as coming from their time and cost savings from HSR, net of taxes paid while the CHSR is operational, holding gross income constant. The "+GE" case also includes amenity spillovers and wage changes in this term.

Figure A.2: Potential CHSR Routes (1996)



Note: the figure shows a digitization of the three planned routes reprinted in page 113 of part 1 of the 2005 CHSR Environmental Impact Report (US DOT, 2005), available at https://hsr.ca.gov/wp-content/uploads/docs/programs/eir-eis/State_Wide_EIR_EIS_Volume_1_Part_1_of_3.pdf. These three routes (Central Valley, I5, and Coastal) are used to construct the instrument #2 in Section 5. The route in red is the actual planned CHSR as of the 2008 vote.

Table A.5: Estimates of Voting Equation, Alternative Models, All Covariates

Model:	2008 Plan		2008 Plan, GE		2022 Plan		2022 Plan, GE	
Inst. Var.:	Random Station (1)	Random Path (2)	Random Station (3)	Random Path (4)	Random Station (5)	Random Path (6)	Random Station (7)	Random Path (8)
$log(\hat{W}_{19})$	66.97^a	63.68^a	50.23^a	43.16^a	177.58^a	159.28^a	123.45^a	96.66^a
	(4.54)	(5.29)	(3.64)	(4.30)	(11.59)	(13.11)	(9.26)	(10.53)
Log-odds Dem. Sh.	0.38^a	0.38^a	0.38^a	0.38^a	0.38^a	0.38^a	0.38^a	0.38^a
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Environ.: Prop. 10	2.41^a	2.41^a	2.41^a	2.43^a	2.40^a	2.41^a	2.39^a	2.42^a
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Transp.: Prop. 1b	0.80^a	0.81^a	0.86^a	0.86^a	0.82^a	0.82^a	0.90^a	0.88^a
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Sh. non-White	-0.17^a	-0.17^a	-0.15^a	-0.15^a	-0.16^a	-0.16^a	-0.15^a	-0.15^a
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Sh. College	0.74^a	0.74^a	0.75^a	0.75^a	0.74^a	0.74^a	0.76^a	0.76^a
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Sh. Under 30	0.19^a	0.18^a	0.18^a	0.18^a	0.19^a	0.19^a	0.17^a	0.17^a
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Log. Dist. Station	-0.00	-0.00	-0.00	-0.01^b	-0.00	-0.01^c	-0.01^b	-0.01^a
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
F-stat	966	850	1093	1108	1111	810	1229	974
Num. Obs.	7861	7861	7861	7861	7861	7861	7861	7861

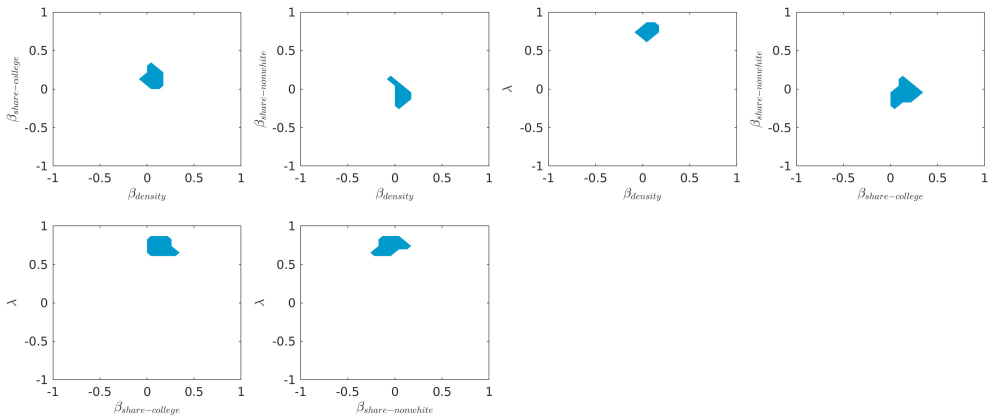
Note: ^{a} denotes 1% significance level. Robust standard errors in parenthesis. All specifications control for county fixed effects. Columns (1) and (2) present baseline estimates. Columns (3) and (4) present results for the model that incorporates general equilibrium effects. Columns (5) and (6) present results for the "pessimistic" model, which assumes a 0.5 probability that the CHSR is completed in 24 years. Columns (7) and (8) present results for a version of the model that allows the CHSR to be a perfect substitute to traveling by car.

Table A.6: Estimates of Voting Equation, Alternative Weighting and Sample Selection Criteria

	Baseline		Weighting - Num. Votes		Weighting - Participation		Selection - $\geq$ 5km line	
Inst. Var.:	Random Station (1)	Random Path (2)	Random Station (3)	Random Path (4)	Random Station (5)	Random Path (6)	Random Station (7)	Random Path (8)
$log(\hat{W}_{19})$	66.97^a	63.68^a	90.50^a	83.21^a	65.97^a	62.33^a	89.14^a	86.47^a
	(4.54)	(5.29)	(6.16)	(7.19)	(10.54)	(8.31)	(7.62)	(9.20)
Log-odds Dem. Sh.	0.38^a	0.38^a	0.41^a	0.41^a	0.42^a	0.42^a	0.40^a	0.40^a
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Environ.: Prop. 10	2.41^a	2.41^a	2.34^a	2.35^a	2.23^a	2.24^a	2.50^a	2.50^a
	(0.05)	(0.05)	(0.05)	(0.06)	(0.17)	(0.17)	(0.06)	(0.06)
Transp.: Prop. 1b	0.80^a	0.81^a	0.87^a	0.86^a	0.90^a	0.90^a	0.82^a	0.82^a
	(0.04)	(0.04)	(0.05)	(0.05)	(0.09)	(0.09)	(0.05)	(0.05)
Sh. non-White	-0.17^a	-0.17^a	-0.23^a	-0.23^a	-0.21^a	-0.21^a	-0.31^a	-0.30^a
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Sh. College	0.74^a	0.74^a	0.74^a	0.74^a	0.71^a	0.72^a	0.72^a	0.72^a
	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.03)	(0.02)	(0.02)
Sh. Under 30	0.19^a	0.18^a	0.23^a	0.23^a	0.23	0.23	0.26^a	0.26^a
	(0.03)	(0.03)	(0.03)	(0.03)	(0.14)	(0.14)	(0.03)	(0.03)
Log. Dist. Station	-0.00	-0.00	0.00	0.00	-0.01	-0.01	-0.00	-0.01
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
F-stat	966	850	1332	979	1162	1244	665	531
Num. Obs.	7861	7861	7861	7861	7861	7861	7861	7861

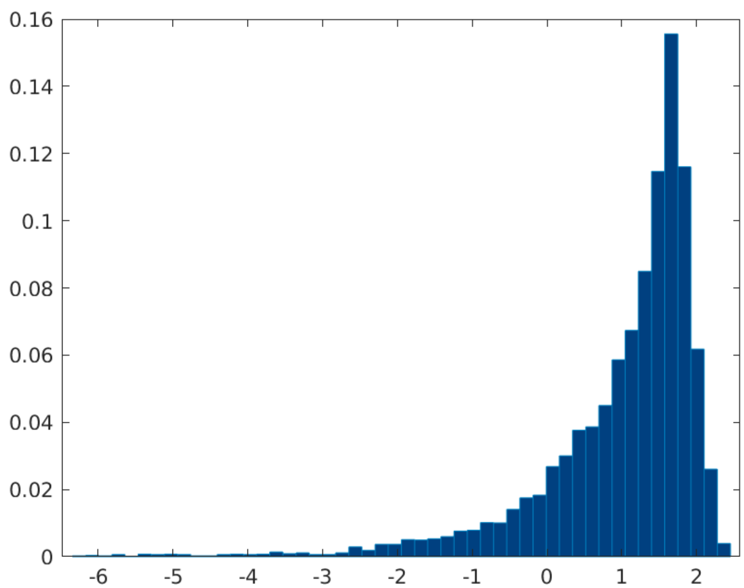
Note: ^{a} denotes 1% significance level. Robust standard errors in parenthesis. All specifications control for county fixed effects. Columns (1) and (2) present baseline estimates. Columns (3) and (4) present results where each census tract is weighted by the number of votes in the HSR referendum. Columns (5) and (6) present results where each census tract is weighted by the participation rate in the HSR referendum. Columns (7) and (8) present results where we exclude census tracts that are less than 5 km away from the railway line.

Figure A.3: 95% Confidence Set for Planner's Parameter Estimates



Notes: 2008 Business Plan estimates. Pairwise projections of the 95% confidence set from column (6) in Table 5 in spherical normalization (see Appendix F.1). In blue are denoted all the points $\left(b_k^*, b_j^*\right)$ such that there exists a vector $\left(\vec{b}, \lambda\right)$ with $b_k = b_k^*$ and $b_l = b_j^*$ which belongs to the confidence set.

Figure A.4: Distribution of Planning Weights



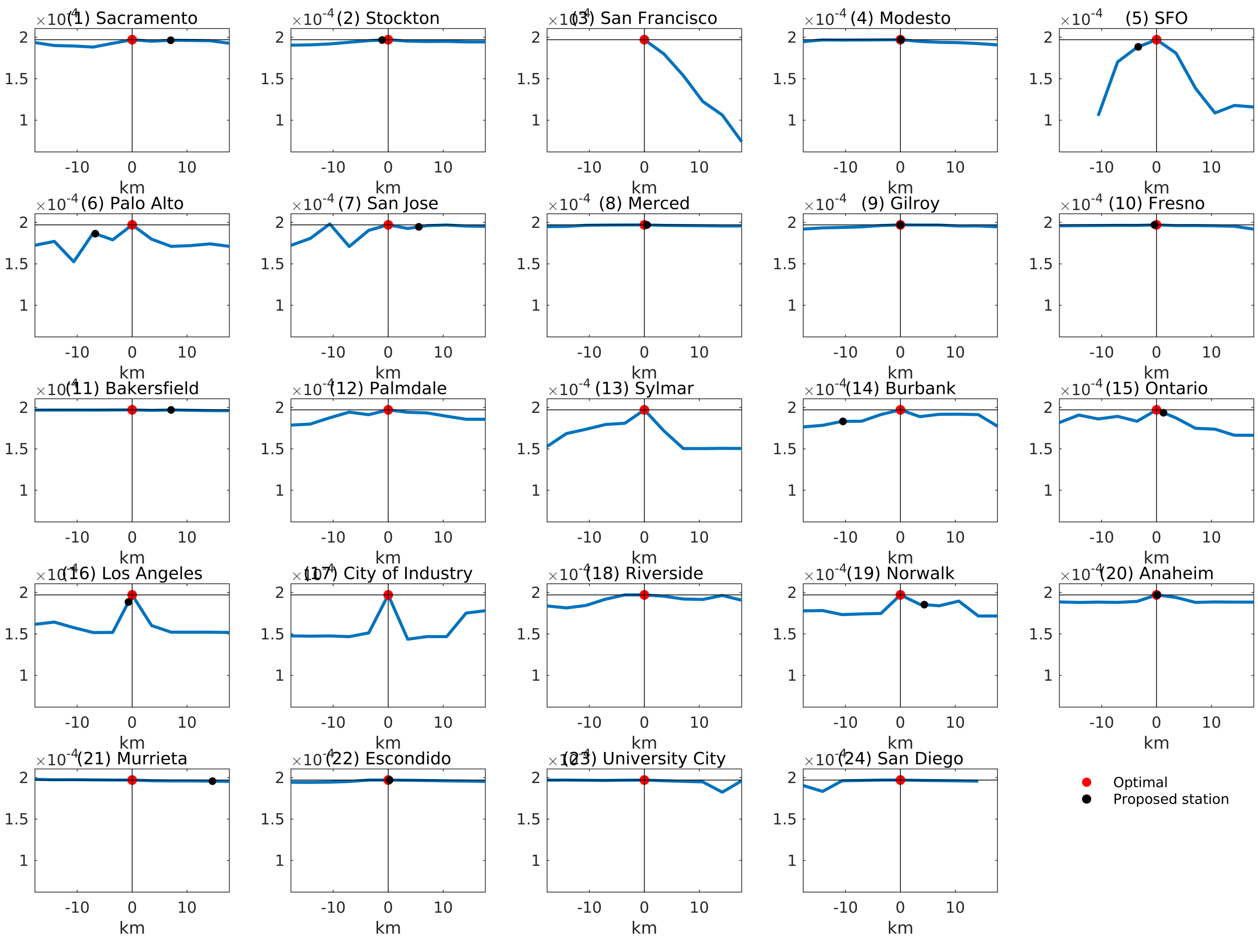
Notes: 2008 Business Plan estimates. Histogram of the effective planning weights, defined as $\lambda_U(i) + \lambda_V \theta_V v(i)(1 - v(i))$, from specification (5) in Table 5. The parameters are normalized so that the population-weighted mean of the Pareto weights is 1.

Table A.7: Stations in the Apolitical Planner vs. Proposed Plan (2008 Business Plan)

	(1)	(2)	(3)	(4)	(5)	(6)
Station	Distance (km)	ΔWelfare (%)	$\boldsymbol{\rho}_{\text{density}}$	$\boldsymbol{\rho}_{\text{share college}}$	$\boldsymbol{\rho}_{\text{share non-white}}$	$\boldsymbol{\rho}_{\text{voting semi-elasticity}}$
Sacramento	-1.61	-0.0001	0.046	0.019	0.11	0.029
Stockton	2.31	-0.0001	0.029	0.077	-0.0076	-0.01
San Francisco	5.63	0.0044	0.46	0.17	0.11	-0.59
Modesto	-0.13	0	-0.066	-0.12	-0.054	0.07
SFO	7.85	0.0007	0.09	0.052	0.0027	-0.12
Palo Alto	18.24	0.0007	0.046	0.068	0.066	-0.061
San Jose	-4.66	0	0.068	0.1	-0.016	-0.086
Merced	-1.01	0	-0.0039	-0.0074	0.0083	-0.004
Gilroy	0.01	0	0.0088	-0.044	-0.031	-0.0084
Fresno	1.13	0	-0.066	-0.086	-0.0092	0.041
Bakersfield	-3.39	0	-0.017	-0.0085	-0.033	0.041
Palmdale	46.9	0.0013	0.15	-0.0014	0.052	-0.026
Sylmar	35.86	0.0051	0.46	-0.12	0.28	-0.19
Burbank	5	0.0013	0.28	-0.03	0.14	-0.089
Ontario	-4.97	0.0004	0.049	-0.014	0.053	-0.013
Los Angeles	0.19	0.0004	0.4	-0.063	0.16	-0.19
City of Industry	-23.5	0.0033	0.35	-0.15	0.23	-0.12
Riverside	-10.8	0.0006	0.035	-0.044	0.05	0.018
Norwalk	-8.09	0.0023	0.15	-0.048	0.15	0.022
Anaheim	-0.03	0	-0.29	0.068	-0.2	0.022
Murrieta	-11.64	0.0002	0.017	-0.07	0.031	0.0017
Escondido	-0.04	0	-0.00089	0.035	-0.098	0.072
University City	5.93	0.0002	0.015	0.0051	-0.064	0.02
San Diego	-5.09	0.0003	0.043	0.082	-0.062	0.028

Notes: 2008 Business Plan estimates. Column (1) reports the distance in km between the optimal location and the proposed station. A positive number indicates a movement towards San Diego, negative towards San Francisco. Column (2) reports the change in aggregate welfare in basis points (%) given the Pareto weights $\lambda_U(i)$ between the optimal station placement minus the welfare corresponding to moving *only* the corresponding station back to its original location in the proposed plan. In columns (3)-(7), ρ_X show the correlation between $\Delta \log \hat{W}(i)$ and the corresponding covariate X in that one-deviation counterfactual from the proposed plan.

Figure A.5: Apolitical Planner, Robustness of Optimal Design

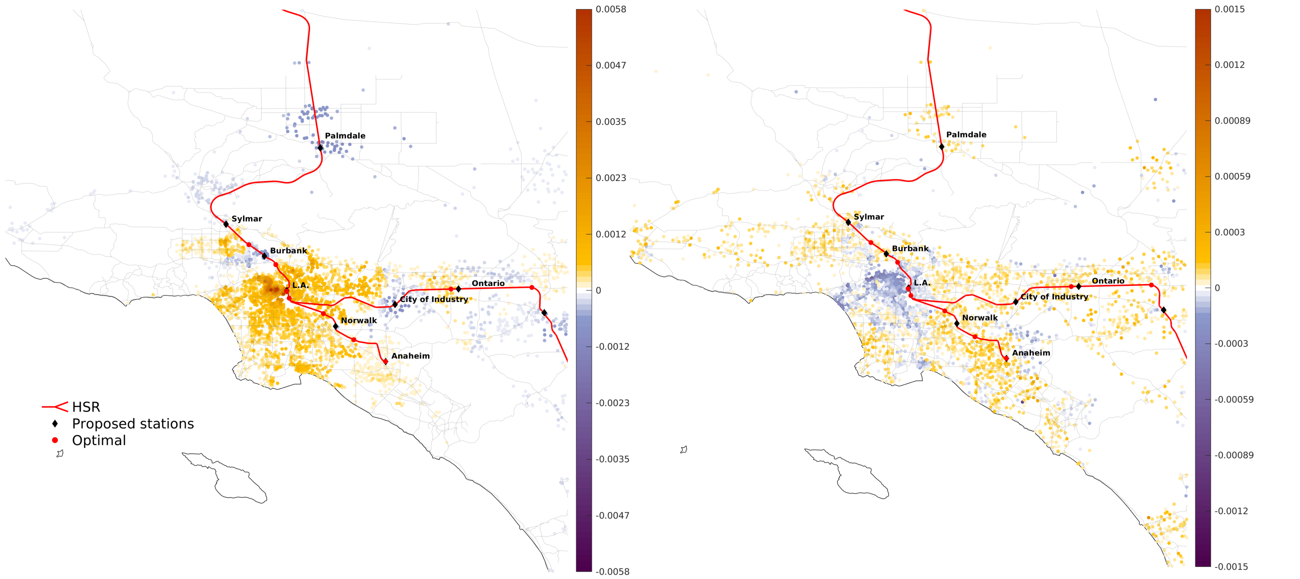


Notes: 2008 Business Plan estimates. Each panel displays how aggregate welfare is affected when moving the indicated individual station by about ±10km around the optimal location. Optimal stations are indicated in red, the proposed ones in black (sometimes outside).

Figure A.6: Apolitical Planner, Welfare vs. Votes in L.A.

Notes:

(a) Welfare Change (b) Voting Semi-elasticity

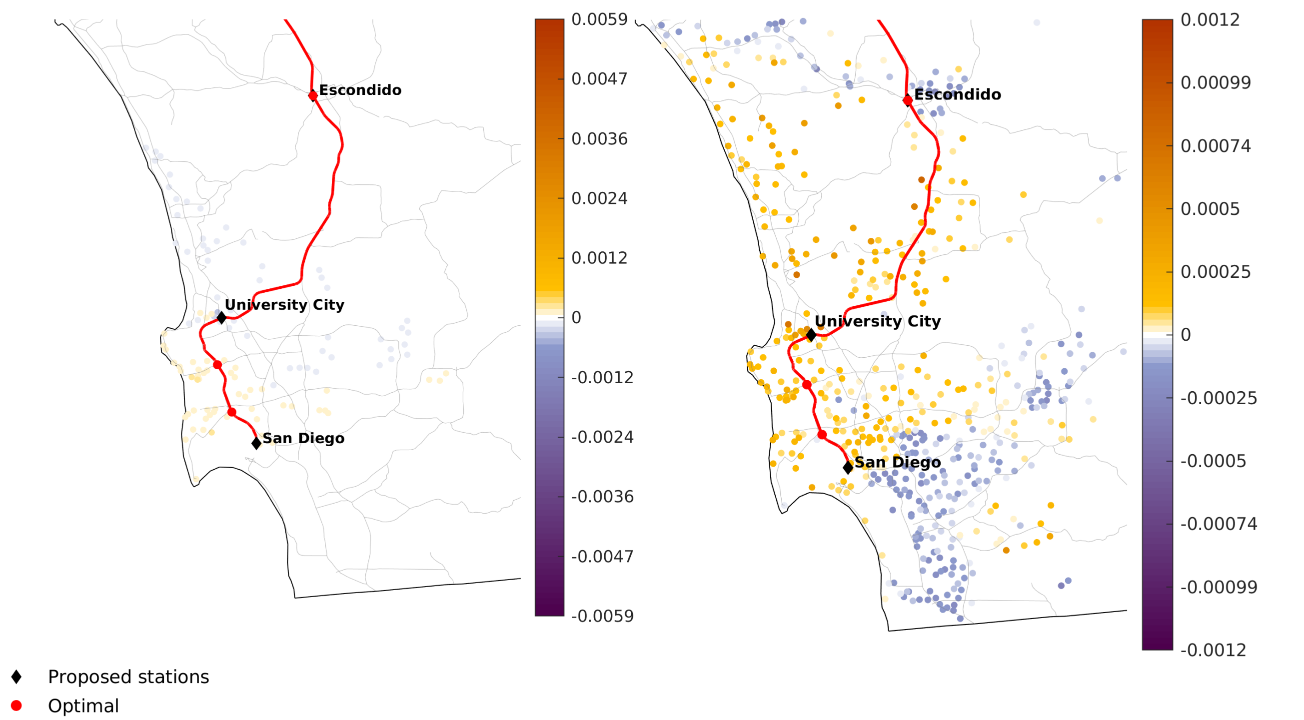


2008 Business Plan estimates.

Figure A.7: Utilitarian Planner, Welfare vs. Share College in San Diego

Notes:

(a) Welfare Change (b) Share College



2008 Business Plan estimates.

F Appendix to Section 6 (Planner Preferences)

This section contains additional details on the implementation of the planner's preference estimation and optimal station location problem.

F.1 Planner's Preferences Estimation

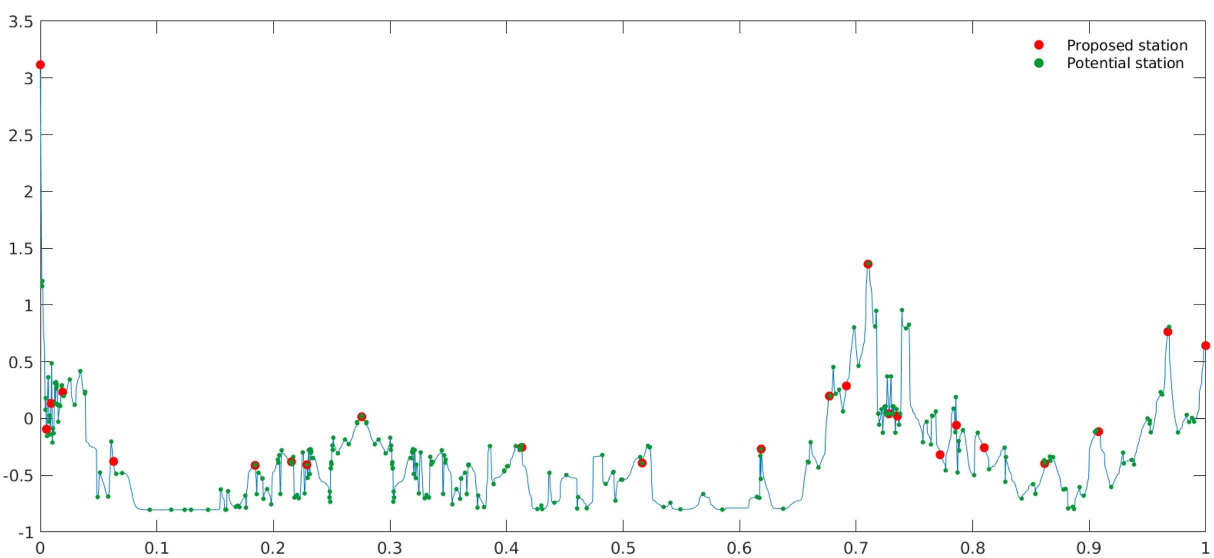
Design of the perturbations As indicated in subsection 6.2, our estimation relies on a moment inequality estimator based on a set of perturbations $n \in \mathcal{N}$ of the CHSR design that yield inequalities of the form

$$\sum_{i=1}^{J}\left[b_0 + \sum_{k=1}^{K} b_k Z_k\left(i\right) + \lambda_V \nabla v\left(i\right)\right] N_R\left(i\right) \Delta \ln \hat{W}\left(i, \mathbf{d}^n\right) - \epsilon\left(\mathbf{d}^n\right) \leq 0, \tag{A.51}$$

where $\nabla v\left(i\right) \equiv \theta_V v\left(i\right)\left(1 - v\left(i\right)\right)$ is the semi-elasticity of votes to real income. We generate these perturbations with the aim of identifying upper and lower bounds on the parameters b and λ_V. For each covariate $Z \in \mathcal{Z} = \{Z_1, \ldots, Z_K, \nabla v\}$, we construct a set of potential locations $\mathcal{L}\left(Z\right)$ corresponding to peaks and troughs of Z along the proposed CHSR lines. We order these locations as a function of their distance to San Francisco along the CHSR outline. To illustrate this procedure, Figure A.8 displays the variation in population density along the proposed CHSR outline highlighting its peaks and troughs as potential station locations (green) in comparison to the proposed ones (red).

For each covariate Z, we then define a set of perturbations $\mathcal{N}\left(Z\right)$ by moving each station

Figure A.8: Potential Station Locations based on Population Density



Notes: Smoothed population density $\hat{n}\left(i\right) = \sum_{i'=1}^{J} n\left(i'\right) e^{-\rho dist\left(i,i'\right)}$ where $\rho = 100$ and dist(i, i') is the arc-degree distance between Census tracts i and i'. The x-axis is an indicator of the location between San Francisco ($x = 0$) and San Diego ($x = 1$) as a fraction of the entire CHSR length, with Los Angeles corresponding to $x \simeq 0.7$. Potential locations for stations are identified as local peaks and troughs and indicated in green.

$s = 1 \ldots 24$ individually by a certain number of steps $k = -N_{steps} \ldots -1$ (towards San Francisco) and $k = 1 \ldots N_{steps}$ (towards San Diego) among the set of potential locations. In our baseline specification, we choose $N_{steps} = 2$, which, for a number of 24 stations, generates $24 \times 4 = 96$ perturbations per covariate.

Our final set of perturbations $\mathcal{N} = \cup_{Z \in \mathcal{Z}} \mathcal{N}(Z)$ is the union of these covariate-specific perturbations. In our largest specification (using as covariates population density, the share of college-educated residents, the share of non-white residents and votes) we obtain $4 \times 96 = 384$ perturbations to identify 5 parameters ($b_0, ..., b_K$ and λ_V).

Moment Conditions To lighten notation, we rewrite the welfare inequality (A.51) for perturbation n (i.e, design $\mathbf{d}^n$) as

$$\Delta \mathcal{W}(\mathbf{d}^n; \gamma) \equiv \sum_{k=1}^{K+2} \gamma_k X_k(n) - \epsilon(\mathbf{d}^n) \leq 0,$$

where $\gamma = (b_0, b_1, \ldots, b_K, \lambda_V)$ is the vector of parameters we want to estimate and

$$\mathbf{X}(n) \equiv \begin{pmatrix} \sum_i N_R(i) \Delta \ln \hat{W}(i, \mathbf{d}^n) \\ \sum_i N_R(i) \Delta \ln \hat{W}(i, \mathbf{d}^n) Z_1^T(i) \\ \vdots \\ \sum_i N_R(i) \Delta \ln \hat{W}(i, \mathbf{d}^n) Z_K^T(i) \\ \sum_i N_R(i) \Delta \ln \hat{W}(i, \mathbf{d}^n) \nabla v(i) \end{pmatrix}.$$

To obtain upper and lower bounds on each parameter γ_k, we create a set of moments indexed by $e = 1, \ldots, E$ with $E = 2 \times (K+2)$. Each moment is associated to a particular sign of the component X_k evaluated in 2008 at the time when the CHSR was designed. More precisely, for $e = 1, \ldots, E$, we define the moment

$$\hat{m}_e(\gamma) = \frac{\sum_{n \in \mathcal{N}_e} \Delta \mathcal{W}(\mathbf{d}^n; \gamma)}{|\mathcal{N}|}$$

and associated standard deviation

$$\hat{\sigma}_e(\gamma) = \left[\frac{\sum_{n \in \mathcal{N}_e} (\Delta \mathcal{W}(\mathbf{d}^n; \gamma) - \hat{m}_e(\gamma))^2}{|\mathcal{N}|} \right]^{\frac{1}{2}},$$

where $\mathcal{N}_e = \{n \in \mathcal{N} \mid X_e^{2008}(n) \geq 0\}$ for $e = 1, \ldots, K+2$ and $\mathcal{N}_e = \left\{n \in \mathcal{N} \mid X_{e-(K+2)}^{2008}(n) \leq 0\right\}$ for $e = K+3, \ldots, E$. Since the subsets $\mathcal{N}_e \subset \mathcal{N}$ are created using information from 2008, we can use the moment condition that $E\left[\epsilon(\mathbf{d}^n) | \mathcal{I}^{2008}\right] = 0$. Following the Modified Method of Moments from Andrews and Soares (2010), we construct the statistics

$$\hat{Q}(\gamma) = \sum_{e}^{E} \left(\max \left(\sqrt{|\mathcal{N}|} \frac{\hat{m}_e(\gamma)}{\hat{\sigma}_e(\gamma)}, 0 \right) \right)^2.$$

We construct a 95% confidence set over parameter γ as

$$\Gamma = \left\{ \gamma \mid \hat{Q}(\gamma) \leq cv_{95}(\gamma) \right\},$$

where the critical value $cv_{95}(\gamma)$ is computed by bootstrap.

Normalization The estimation procedure runs into the problem that the planner weights γ are not uniquely determined since for any positive real number $z > 0$, $z \times \gamma$ give the exact same planner preferences. We use two different types of normalizations:

- To guarantee having a bounded confidence set, we use a spherical normalization during the estimation stage, requiring that $\sum_k \left(\gamma_k^{sphere}\right)^2 = 1$.
- When reporting the result and to make the parameters more directly interpretable, we normalize γ so that the population-weighted average of $\lambda_U(i)$ is 1 across locations. Specifically, the means that for any γ^{sphere}, we define

$$\gamma_k^{norm} \equiv \frac{\gamma_k^{sphere}}{z\left(\gamma^{sphere}\right)}$$

  where

$$z\left(\gamma^{sphere}\right) = \sum_i N_R(i) \left(\gamma_0 + \sum_{k=1}^{K} \gamma_k^{sphere} Z_k(i)\right).$$

  Since the covariates Z_k are normalized to have mean 0 and standard deviation 1, γ_k^{norm} can be interpreted as is the relative impact on the per-capita Pareto weight of having covariate Z_k one standard deviation above its mean.

F.2 Planner Optimization

Procedure The optimal station location problem is highly non-convex optimization due to the presence of a sigmoidal functional form of the voting block (neither concave nor convex) and competing complementarities (convex) and substitutabilities (concave) arising in the placement of stations. The optimization problem requires the use of non-convex techniques. The optimization is done in three sequential steps.

First, we use the information in the precomputed perturbation set $\mathcal{N}$ used for estimation. For each station s, we identify the perturbation $n \in \mathcal{N}$ with the highest welfare among those that shifted the location of station s, and evaluate a new CHSR design in which station s is set to that best location. If total welfare increases, we accept this location; if not, we keep the initial location and move on to the next station.

Second, we use a simulated annealing method. A randomly selected station is shifted among the potential locations that were considered when constructing the perturbations. The new location is accepted with a probability that depends on the welfare obtained (1 if welfare increases, and a positive probability even if welfare decreases).

Third, we implement a continuous optimizer. This final step attempts to refine the local optimum obtained by the previous two steps by using a continuous optimizer (specifically, we use simplex or interior-point algorithms). We parametrize the CHSR outline with a cubic spline over its coordinates, allowing us to evaluate welfare over a continuous set of station locations.

These efforts to globally explore the candidate set of station locations do not guarantee identification of the global optimum. Nonetheless, we verify that our final result is a good candidate by checking that it yields the highest possible welfare when each station is moved individually within a range of 10 km from the proposed optimum along the CHSR outline, as shown in Figure A.5 for the case of the apolitical planner using 2008 cost predictions.

Dealing with Expectations Errors The planner's objective function is defined as the expectation of future welfare, evaluated using information available in year 2008. We cannot compute the expectation term in practice. To deal with this issue, we adopt the same strategy that we used in the estimation: for each potential design $\mathbf{d}$, we evaluate the planner's objective using time T data (when the CHSR was projected to be in service) and introduce a forecast error term $\epsilon(\mathbf{d};\gamma)$ for each given set of planner weights $\gamma = (b;\lambda)$, defined as

$$\begin{aligned}\epsilon(\mathbf{d};\gamma) = & \sum_i \lambda_U(i;b)\, N_R(i)\, \Delta \ln \hat{W}(i,\mathbf{d}) + \lambda_V \sum_i N_R(i)\, v(i;\mathbf{d}) \\ & - \mathbb{E}\left[\sum_i \lambda_U(i;b)\, N_R(i)\, \Delta \ln \hat{W}(i,\mathbf{d}) + \lambda_V \sum_i N_R(i)\, v(i;\mathbf{d}) \mid \mathcal{I}^{2008}\right].\end{aligned}$$

When computing optimal station locations for a counterfactual set of planner weight $\gamma = (b, \lambda_V)$, $\epsilon(\mathbf{d};\gamma)$ is unknown. When optimizing, we assume that the uncertainty driving the forecast error is independent of the planner's preferences and set $\epsilon(\mathbf{d};\gamma) \equiv \epsilon(\mathbf{d};\hat{\gamma})$ where $\hat{\gamma}$ is the estimated planner's weights. To evaluate $\epsilon(\mathbf{d};\hat{\gamma})$, we use the property that design $\mathbf{d}$ is not optimal under the estimated preferences $\hat{\gamma}$, i.e., that

$$\sum_i \lambda_U\left(i;\hat{b}\right) N_R(i)\, \Delta \ln \hat{W}(i,\mathbf{d}) + \hat{\lambda}_V \sum_i N(i)\, v(i;\mathbf{d}) - \epsilon(\mathbf{d};\hat{\gamma}) \leq 0.$$

To avoid penalizing alternative CHSR designs, we adopt the lowest possible value for $\epsilon(\mathbf{d};\hat{\gamma})$ and set it to

$$\epsilon(\mathbf{d};\hat{\gamma}) \equiv \max\left(0, \sum_i \lambda_U\left(i;\hat{b}\right) N_R(i)\, \Delta \ln \hat{W}(i,\mathbf{d}) + \hat{\lambda}_V \sum_i N_R(i)\, v(i;\mathbf{d})\right).$$