Political Preferences and the Spatial Distribution of 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

April 2024

Abstract

We study how political preferences shaped California's High-Speed Rail (CHSR), a large transportation project approved by referendum in 2008. Across census tracts, support for the project responded significantly to the projected economic gains at the time of voting, as measured by a quantitative model of high-speed rail matched to CHSR plans. Instrumenting with random station placements on feasible routes, we estimate that 0.1%-0.2% projected economic gains swayed 1% of votes at the median tract. Given this elasticity, a revealed-preference approach comparing the CHSR with counterfactual designs identifies strong policymakers' preferences for political support. A politically-blind planner would have placed the stations nearer to California's dense metro areas, doubling the projected economic gains.

*We thank Andrii Parkhomenko and Jonathan Dingel 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, 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 infrastructure investments constitutes a central question in spatial economics. Using quantitative spatial frameworks with a transport sector, recent research has studied the optimality of transport networks. Recent studies in this vein include Fajgelbaum and Schaal (2020) and Allen and Arkolakis (2022) for highways, Kreindler et al. (2023) and Almagro et al. (2024) for bus systems, Brancaccio et al. (2023) for bulk shipping, and Salz (2022) and Buchholz (2022) for taxis. This body of research finds that inefficiencies are pervasive: according to these parameterized models, observed transport systems could have been designed in welfare improving ways.

Why are transport networks seemingly inefficient? In this paper we study the role of policymakers' and households' preferences in shaping the projects that are implemented. We posit that 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 forces requires a methodology to reveal the planners' political motivations from observed transport networks. A key step towards this goal is 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), an electric high-speed rail originally conceived to connect San Francisco and Los Angeles that is among the most expensive transport projects ever attempted in U.S. history. In the 2008 general election, Proposition 1A asked Californians whether they approved of initiating funding for the CHSR. The features of the project that were known at the time of voting –manifested through official business plans, environmental reports, and the ballot's text– resulted from years-long planning by state authorities who anticipated and eventually approved putting the CHSR up to a public vote.

We tease out Californian politicians' and voters' policy preferences and study their role in shaping the proposed CHSR. To that end, we use a novel structural framework of high-speed rail network design combined with voting data and CHSR planning data.[1] We first develop a spatial model of passenger travel with features that capture specificities of high-speed rail, and quantify it using the CHSR business plans. Then we estimate, across census tracts, the response of favorable votes in the 2008 referendum to the expected local economic impacts of the CHSR as measured by the model. To address endogeneity concerns, we build instruments based on random station placements along feasible CHSR designs entertained by planners. Finally, we embed both the spatial model and the voting choice into the problem of a politically-minded planner who decides where to locate stations. We estimate the policymakers' preferences for popular approval and redistribution,

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

by comparing the actual CHSR design with non-proposed alternatives to obtain parameter bounds. We finally solve for optimal station distributions under alternative policymakers' preferences.

We summarize three main takeaways. First, we find that voters responded significantly to the projected economic impact of the CHSR in their tract. This result supports our estimated model as a predictor of the spatial distribution of economic impacts of the CHSR. It also shows that economic voting is a significant driver of policy preferences over transport infrastructure. Second, the CHSR design implies strong planner preferences for gaining votes; intuitively, this means that deviations that would have increased aggregate welfare but reduced votes were not implemented. Third, in the absence of these preferences for votes, the optimal CHSR design would have concentrated stations closer to urban areas, where it was harder to sway votes, and this would have doubled the projected gross economic benefits of the CHSR. We conclude that attaining popular approval was an important driver of transport network design.

We now describe each step. To start, we develop and estimate a quantitative spatial model of high-speed passenger travel. The goal of the model is to obtain a distribution of *relative* real-income impacts of the CHSR across census tracts forecasted at the time of voting, to then estimate the responsiveness of votes to these impacts. Compared to canonical urban frameworks centered on commuting, such as Ahlfeldt et al. (2015) and Monte et al. (2018), our framework 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 for their quantification.[2] Second, as potential CHSR usage depends on access to competing travel modes, we model a transport-mode decision (between car, air, and public transit) for each origin-destination and travel purpose by incorporating a standard approach in the tradition of McFadden et al. (1973). Third, travel decisions account for both time (as in the frameworks we build upon) and monetary trip costs, an a priori relevant feature given the low ticket prices announced by CHSR planners.[3]

We quantify the welfare-relevant elasticities of the model by estimating gravity equations for commuting, leisure, and business travel across California's census tracts. These 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. We are then able to simulate travelers' decisions to use the CHSR (were it to become available) and the spatial distribution of its potential real income effects across 7866 census tracts with heterogeneous access to California's multi-modal transport network, labor and business productivity, and leisure

[2]We model long-distance business and leisure travel decisions as including an intensive margin, the number of trips, that we also observe in the data. We model business travel as an input in production, with firms (rather than travelers) deciding the destination of business trips based on characteristics that make destinations more profitable for business connections. Through this channel, the transport network affects 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 use both of these forecasts in the alternative scenarios of our quantification.

travel amenities.[4]

The results show that larger urban centers such as San Francisco or Los Angeles and areas closer to the railway stations would on average benefit relatively more from the CHSR in terms of the time-cost reductions and expected wage and land-rent increases of its resident voters; in contrast, more sparsely populated areas like Central-Valley have lower potential gains. These model-based predictions cannot be contrasted with actual impacts because the CHSR has not been built; however, they are qualitatively consistent with existing empirical assessments of commuter systems.[5] Moreover, the voter regressions discussed below provide a validation of the model predictions, as they robustly show more support for the CHSR in tracts where the model predicts stronger expected real-income impacts.

These predictions use official CHSR information available in 2008, which turned out to be unrealistic: the construction of the CHSR progressed at a much slower pace, costs bloated, and doubts linger on whether the project will ever come to fruition. Therefore, we also implement the analysis using ex-ante more pessimistic (and ex-post more realistic) forecasts that incorporate information from the 2022 CHSR business plan, which acknowledges risk, costs, and timeline increases. We emphasize that these assumptions affect the level and the variance of expected economic impacts across tracts, but not the *relative* impacts: the spatial correlation of the gains using 2008 vs. 2022 forecasts is 99%. It is this distribution that matters to estimate if there was a response of votes to expected income in the next step.

As a second step, we estimate the elasticity of favorable votes to the expected economic gains. Identifying this response is crucial for our goal of estimating whether planning authorities take into account political support when deciding where to locate transport infrastructure. We undertake this estimation in a discrete-choice framework where voters derive utility from both economic outcomes and other considerations, such as preferences for public-good aspects of the policy or political identity (which likely matters for transport projects such as the CHSR that are championed by one political party).[6] For our ultimate purpose of analyzing network design, the key outcome of this estimation is the impact of expected economic gains on the votes; we do not need to take a stance on the nature of the remaining considerations behind the vote.

Our analysis of the importance of private economic determinants in the voting decision follows a novel strategy. We project favorable votes on the policy impacts predicted by our fully specified

[4]Our setting is sparse, with about 15 million workers for about 64 million origin-destination pairs of census tracts; moreover, the long distance travel information that we use comes from travel surveys. As suggested by Dingel and Tintelnot (2020), we implement Poisson Pseudo-Maximum likelihood estimation of travel elasticities.

[5]Recent studies of the the spatial impacts of public transit on employment, wages, or land prices 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 subway in New York, and Borusyak and Hull (2023) for the Chinese high-speed rail. Broadly speaking, these studies find positive impacts of proximity to transit connections on commuting flows, employment, wages, or land prices. Koster et al. (2022) finds employment losses for smaller areas connected to the Japanese high-speed rail.

[6]Voting on subjective considerations is referred to as expressive voting in the political sciences literature (Hillman, 2010), and it helps to rationalize phenomena such as high voting turnout (Brennan and Hamlin, 1998). Grossman and Helpman (2021) introduce an identity component alongside material outcomes in the voting decision over trade policies.

model; in contrast, most studies project votes on reduced-form variables that plausibly capture individual exposure to a policy.[7] This reliance on a structural model is essential to recover a key structural elasticity and to characterize counterfactual optimal networks.

Our estimation deals with important identification challenges, such as the correlation of economic benefits of the CHSR with unobserved components of beliefs. First, we instrument for the economic impact of the CHSR using alternative CHSR designs with random placement of stations across feasible routes identified by planners years prior to the vote. Second, we construct these instruments using pre-determined variables in 2008 to account for the measurement error inherent to using realizations in lieu of unobserved expectations (Dickstein and Morales, 2018). Third, we control for a host of tract-level covariates including county fixed effects, demographic characteristics, Democratic party affiliation, and votes in related referenda favoring green energies or transportation initiatives. Finally, we also conduct the estimation under different model specifications in terms of economic mechanisms and cost assumptions.

Regardless of the model assumptions and identification strategies, we find that voters were significantly responsive to the model-based expected economic impact of the CHSR. Moreover, once we instrument, this elasticity is highly stable regardless of what sets of controls are included. Depending on whether 2008 or 2022 cost predictions are used, an extra 0.1-0.2 percentage points in local expected economic gains swayed one percent of local votes (the standard deviation of votes was 0.1 percentage points).[8] This responsiveness implies that policymakers who value public support may want to shift the supply of infrastructure towards areas where its marginal impact on public support is greater, at the expense of where it may be socially more desirable. We quantify this trade-off in the next step.

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.[9] We model a policymaker choosing the distribution of stations along the main technologically feasible routes linking Northern and Southern California that were identified by transport engineers several years prior to the 2008 vote (US DOT, 2005). We assume that the observed distribution of stations maximized an objective function with two components: aggregate real income across tracts (with tract-specific Pareto weights as function of demographics) and total votes in favor of the project.

To estimate these 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.

[7]E.g., Van Patten and Méndez (2022), studying a referendum in Costa Rica on whether to sign a free-trade agreement with the U.S., project favorable votes on voters' industry or firm exposure to U.S. trade. Following Deacon and Shapiro (1975), many studies have used 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 determinants of political ideology and proxies for economic geography. In some of these studies, individual economic returns are captured by dummies for the industry of voters. Alesina and Giuliano (2011) review research that uses survey data to measure preferences for private vs. public value in income taxation.

[8]The elasticities of votes to expected economic gains are mechanically different depending on whether we use 2008 or 2022 business plans because, under the latter, the completion probability is lower, leading to estimating more elastic vote responses to projected economic gains.

[9]Burgess et al. (2015) and Alder and Kondo (2020) find empirical evidence that a president's ethnicity or birthplace explains patterns of road investments in Kenya and China, respectively.

The logic is to parametrize the planner's preferences such that its optimization is consistent with an observed policy. 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 in response to counterfactual policies, and then estimate planner's preferences such that those deviations cannot increase the value of the planner's objective.

A key challenge in our context is that closed-form solutions for optimal policies, which are typically used in the previous literature, are unavailable. Furthermore, small perturbations to stations' locations have little identifying power. Hence, we use discrete (rather than marginal) deviations from the planner's 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 on the planner's preferences: for instance, deviations that increase aggregate votes but reduce aggregate income define an upper bound on how much planners could have liked 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, as well as some preference for areas with a larger share of college graduates. Moreover, these estimates imply a large heterogeneity in tract-specific weights. The planner is, thus, far from the utilitarian benchmark.

Finally, we implement optimal networks under counterfactual preferences that shut down preferences for votes or for redistribution. The optimal designs of the CHSR for an *apolitical* planner differs substantially from the proposed plan by increasing the proximity of stations to the main metropolitan areas. The reason is that metro areas have low voting elasticities (they strongly support the project already). Hence, an absence of electoral motives would have incentivized policymakers to place stations closer to high-density locations with low expected political gains, where the marginal voter is far from the mode of political support. This reallocation doubles the projected economic gains of the CHSR. Finally, we find that the planner's preference for votes drives almost all of the difference between the observed network and a purely utilitarian optimal design, where the planner cares neither about votes or redistribution.

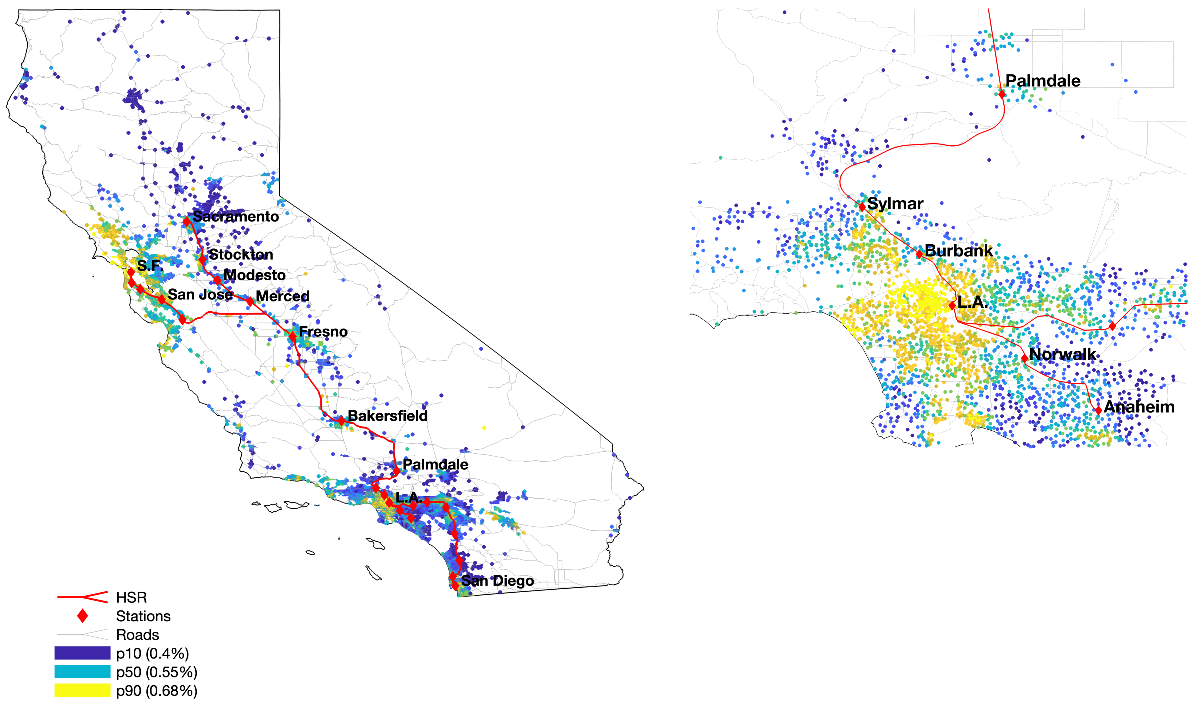
The paper proceeds as follows. Section 2 gives some background on the CHSR. Section 3 lays out our quantitative framework. Section 4 presents the data, the quantification of the model and the estimation of voters' preferences. Section 5 estimates the planner's preferences and presents model-based counterfactuals. 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

Figure 1: CHSR Route and Proposition 1A Votes

(a) California (b) LA County



legislature approved that Proposition 1A would appear on the ballot in the November 2008 general election, asking California voters to approve the issuance of nearly \$10 billion in bonds to initiate funding for the CHSR. The project's Phase-I (linking Los Angeles and San Francisco) was budgeted at \$33 billion, with complementary funding expected from federal funding and private investors.

The CHSR project voted upon in Proposition 1A was required to meet several criteria. First, it would connect San Francisco to Los Angeles and Anaheim, and would also include Sacramento, the San Francisco Bay Area, the Central Valley, Los Angeles, the Inland Empire, Orange County, and San Diego. Second, it would travel at at least 200 miles per hour, making the trip from San Francisco to Los Angeles Union Station in at most two hours and 40 minutes. Third, there would be no more than 24 stations across the entire network. Finally, the anticipated completion date as of 2008 was 2033.

Proposition 1A was approved by 52.6% of votes. Participation was high, equal to 94% of voters who cast a vote for president in the same election. 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, we observe stronger support in urban centers (Los Angeles, San Diego, San Francisco, San Jose, Fresno, and Sacramento) and declining support as we move away from 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 has ran behind schedule and costs have bloated. Construction is now focused on the Central Valley segment, a 180 miles-long stretch (out of the CHSR approximately 800 miles) that is expected to be completed by the mid 2030s. The estimated total costs for Phase I of the project have roughly doubled from around \$50 billion in 2008 (California High Speed Rail Authority, 2008) to around \$100 in 2022 (California High Speed Rail Authority, 2022).

The divergence between 2008 promises and ex-post outcomes may have been incorporated in the forecasts of voters. Therefore, we implement our analysis in two alternative baseline scenarios that correspond to either the ex-post exceedingly optimistic projections from 2008 or the more realistic updates from 2022.

3 Framework

This section gives an overview of the theoretical framework. We present 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), as follows:

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

The first component of utility, $\mathbb{E}\left[\ln W(i,s) \mid \mathcal{I}_i\right]$, is the average expected real income of residents of i for CHSR state s. The expectation is taken over future shocks to fundamental economic characteristics of different locations i, conditional on the current information $\mathcal{I}_i$ of residents of i. We detail the economic forces included in the term $\ln W(i,s)$ next, in Section 3.2. The second component, $\ln a(i,s)$, captures other systematic component of preferences across residents of i, such as a more favorable ideological bias to the CHSR project linked to political affiliation or environmental preferences. Finally, the shock $\varepsilon^u_\omega(s)$ captures idiosyncratic (mean-zero) variation in utility across residents of tract i stemming from either economic or non-economic considerations.

Individuals vote for the policy option s that delivers the highest utility (1). Aggregating over individuals, our setup corresponds to a probabilistic voting model (Deacon and Shapiro, 1975), 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 that the idiosyncratic shocks $\varepsilon^u_\omega(s)$ are Type-I extreme-value distributed across residents, with shape parameter θ_V:

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

As a result, the fraction of voters in i that support building the rail takes the standard logit form:

$$v(i) = \frac{e^{\theta_V\left(\mathbb{E}\left[\ln \hat{W}(i)|\mathcal{I}_i\right]+\ln \hat{a}(i)\right)}}{1+e^{\theta_V\left(\mathbb{E}\left[\ln \hat{W}(i)|\mathcal{I}_i\right]+\ln \hat{a}(i)\right)}}. \tag{4}$$

In this expression,

$$\hat{W}(i) \equiv \frac{W(i,Y)}{W(i,N)} \tag{5}$$

measures real income differences between a world with and without the planned infrastructure project, while $\hat{a}(i) \equiv a(i,Y)/a(i,N)$ measures other determinants of preferences for the project.

A central step in our analysis is to estimate the response of votes to the economic impact of the CHSR, θ_V in (4). We do so in Section 5. A difficulty is that we do not directly observe voters' expectations $\mathbb{E}\left[\ln \hat{W}(i) \mid \mathcal{I}_i\right]$ in (4). To make progress, we use a proxy for these expectations based on the distribution of CHSR impacts $\hat{W}(i)$ across locations predicted by a quantitative spatial model of the high-speed rail in California.[10] We give an overview of this model next.

3.2 Quantitative Model of High-Speed Rail

We summarize here the key forces included in our spatial model, as well as the equations for $\hat{W}(i)$. The heart of the model is a commuting model à la Ahlfeldt et al. (2015) that we augment with the specificities of our setting. Specifically, we allow for: long-distance leisure and business trips, a mode of travel decision with location-specific preferences over travel modes, a monetary cost of travel, and an element of uncertainty over project completion to calculate welfare.

The mechanics of this class of models are well established in the literature (see Redding and Rossi-Hansberg (2017) for a review). Therefore, to save space, we provide here a broad model overview, highlighting its key features and showing the intuitive model-based expressions we use for quantification. We refer the interested reader to the full self-contained description of the model and its microfoundations in Appendix E.

Project Uncertainty If the CHSR is approved ($s = Y$), we assume that households expect to start paying a yearly tax t to fund the CHSR right away (before the CHSR is built and operational). We also assume uncertainty over the project completion: voters expect the CHSR to be operational no sooner than T years after the vote, with a yearly probability of completion equal to p after that. Hence, the annualized (log-) real-income impact of the CHSR, if approved, is a weighted average of $\ln(1-t)$ and of the change in the annual real-income impact (net of taxes) from the CHSR being operational, $\ln \hat{V}(i)$:

$$\ln \hat{W}(i) = (1-R)\ln(1-t) + R\ln \hat{V}(i), \tag{6}$$

where $R \equiv (1+r)^{-T}\frac{p}{r+p}$ is an effective discount rate that incorporates the time discount, expected time until completion, and non-completion risk.

[10] In Section 5 we also discuss our strategies to deal with the potential biases introduced by replacing the voters' expectation by a model-based computation based on realized economic fundamentals.

Real-Income Impacts of the CHSR when Operational To evaluate $\hat{V}(i)$, we assume that residents of location i choose a location where to work, similar to Ahlfeldt et al. (2015).[11] In addition, they also make infrequent leisure and business long-distance trips. In each case (commuting or long-distance trips), travelers make a discrete choice of transport mode. Specifically: car, public transit, air, or walking/biking are imperfect substitutes, and residents of different locations vary in their preferences for travel modes so as to match the heterogeneity in mode use revealed by the data.

When available, the CHSR is a perfect substitute for public transit (for commuters), and for air (for business or leisure travelers). Of course, someone initially traveling by car may also choose to switch travel mode to use the CHSR when available.

In the analysis, we consider two model variants for the economic impacts perceived by voters: a simple *baseline model* and a more sophisticated *general equilibrium model*.

Measurement in the Baseline Model In our baseline model, the economic impacts of the CHSR come simply from the direct time and trip cost effects of the train, as well as from the tax burden to finance the infrastructure. Let $\hat{X}$ denote the ratio between value of variable X if the CHSR is constructed and its value if is not. Given the model microfoundations in Appendix E, the net annual real-income change conditional on the CHSR being operational, $\hat{V}(i)$, for residents of tract i can be decomposed formally as follows:

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

The first component $\hat{\Omega}_C(i)$ is the change in labor income net of commuting cost coming from the commuting component of the model. It captures the expected time savings and changes in travel costs of commuters, based on their ex-ante commuting patterns:

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

In this expression, $\hat{I}(i,j,m)$ is the change in disposable income for a commuter from residence tract i to workplace tract j using transport mode m. This disposable income is impacted by the CHSR for two reasons: the train ticket price and the CHSR tax levied to finance its capital costs.[12] Second, $\hat{\tau}(i,j,m)$ is the change in travel time from i to j through mode m, converted into a dollar-equivalent

[11] In the framework, the development of the CHSR impacts commuting and travel choices but not residential choice. The estimates of Section 5 show a significant response of voting decisions to own-tract economic outcomes, which means that voters must have assigned as significant chance to the local economic impacts –rather than the impacts of a typical migration destination– affecting them in the future. We could accommodate residential mobility by assuming that every year workers can change location at some fixed rate; then $\ln \hat{W}(i)$ would still be linear in $\ln \hat{V}(i)$ as in (6), but the coefficient in front of $\ln \hat{V}(i)$ would be smaller than in the baseline model. In this extension, the coefficient θ_V estimated (4) could still be inferred from the one we obtain in the current implementation, as they become a scaled version of each other.

[12] Specifically, with direct effects only, the change in disposable income is, defined in (A.39), is $\hat{I}(i,j,m) = (1+\chi^{pre}(i,j,m))(1-t) - \chi^{pre}(i,j,m)\,\hat{p}_C(i,j,m)$, where $\chi^{pre}(i,j,m)$ 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, and t is the tax to finance the CHSR.

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, these changes in time and costs for each (i, j, m) are weighted by the shares $\lambda_C(i, j, m)$, the fraction of tract-i residents that commute for work to j using mode m in an observed equilibrium without the CHSR. These shares capture heterogeneous preferences over modes across census tracts.

The second component of (7), $\hat{\Omega}_L(i)$, captures the economic impacts of the CHSR on leisure travel. Given the model microfoundations, this term is again a weighted average of time and cost changes on each route:

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

In this expression, $\lambda_L(i, j, m)$ is the fraction of leisure travelers from i choosing destination j using mode m absent the CHSR, $\hat{p}_L(i, j, m)$ is the change in the monetary cost of travel, $\hat{\tau}(i, j, m)^\rho$ is time savings due to the CHSR in dollar-equivalent terms, and θ_L captures the substitution across destinations and modes for leisure travel. Compared to the commuter term, this expression further includes the share of leisure travel in total expenditure, μ_L.[13]

Measurement with General Equilibrium Effects We also consider a full general equilibrium model variant (the "GE" model). Through the lens of this model, the CHSR is valued both because of the previous direct time and cost effects, and because of its indirect general equilibrium impacts on local amenities and productivities, land rents, and wages.

The impacts on productivity and amenity are channeled through standard spillover effects: for instance, as workers reallocate, location-specific productivity $A(j)$ respond to local change in the density of workers. In addition, a specificity of our model is that tradeable goods are produced with productivity-enhancing business trips (Bernard et al., 2019) in addition to labor and land. The productivity benefits of these trips depend positively on the productivity at destination, and negatively on time and travel cost. Formally, the development of the CHSR improves firms' productivity as follows:

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

This structure is again a very similar to (9), where now $\lambda_B(i, j, m)$ is the fraction of business travelers from i going to destination j using mode m absent the CHSR. These changes in productivity impact wages. In addition, land rents capitalize local productivity enhancements as they increase

[13]When comparing (8) and (9), one can notice two asymmetries: in (8), the monetary cost enters as a negative additive shifter in disposable income $\hat{I}(i, j, m)$, while in (9) it enters multiplicatively; and the latter is shaped by the intensity of leisure spending μ_L. These differences arise due to the non-homothetic nature of spending on commuting: assuming a fixed number of commuting days in the year, travelers spend a fixed amount of money in commuting and the remaining income is divided into consumption, housing, and leisure trips. Travelers decide how many trips to make to their preferred destination with homothetic preferences over leisure trips, consumption, and housing, with spending shares possibly varying across tracts.

demand for productive space. A share of the residents of each tract is homeowner and benefits from this capitalization.

In the baseline model we do not need any additional equations beyond (7) to (9) to measure the real-income effects of the CHSR. The GE model generates additional changes in amenities, land rents, and disposable income (through wages), as shown by their generalization in the appendix equations A.47 and (A.48).

4 Distribution of CHSR's Local Economic Impacts

In this section, we describe the estimation of the parameters of the economic model of the CHSR. We then present the location-specific real income effects of the CHSR predicted by the quantified model.

4.1 Data

We start with an overview of the data used in our analysis, deferring to Appendix A for details. We conduct the analysis at the level of census tracts. Our sample covers 7,866 census tracts housing 98.5% of the statewide population. To estimate the model, 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 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 cost 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 times and costs of traveling by CHSR using information from the 2008 CHSR business plan.

4.2 Gravity Estimates

To characterize the impact of the high-speed rail using (7), (8) and (9), we need estimates of the fraction of commuters, leisure travelers, and business travelers by origin-destination and mode $(\lambda_k\,(i,j,m)$ for $k=C,L,B)$; of the substitution elasticities θ_k; and of the parameters ρ, μ_L, and μ_B. Next, we discuss the procedure we follow to estimate these travel shares and parameters. The resulting estimates determine the preferences of the residents of each census tract for traveling to each other census tract by different transport modes and, as a result, inform the extent to which these residents would modify their trip destinations and mode of transport, were the CHSR to become available. We summarize here our main estimation strategy and results, and defer to Appendix C for the details of our estimation.

Commuting There is a large literature estimating preferences for destinations in commuting decisions. Our specification of the workers' commuting decisions detailed in Appendix E enriches the baseline model in the spatial economics literature (e.g., Monte et al. (2018), Tsivanidis (2019) and Heblich et al., 2020) in several dimensions. We allow labor income in a destination to depend on worker's place of residence, in order to control for heterogeneity worker composition across origin tracts. We also account for monetary costs by transport modes, and we estimate preferences for modes of transport that are heterogeneous across locations according to several demographic characteristics. All these elements substantially impact the likelihood that residents of a particular census tract uses the CHSR, were it to become available.

Specifically, we base the estimation of the parameter vector (θ_C, ρ) on the following model-implied relationship for the share of residents from i that commute to j using transport mode m:

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

In this expression, $I(i,j,m)$ is disposable income (annual labor income net of monetary costs of 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 systematic component of the preferences of commuters residing in location i for transport mode m. We consider three feasible modes of transport for commuting: $\mathcal{M}_C = \{\text{private vehicle, public transport, walk/bike}\}$.[14]

We perform the estimation in two steps. We first estimate θ_C and ρ_C using a Generalized Method of Moments (GMM) estimator that exploits variation in the choice of destination that differ in travel time and costs conditional on origin and transport mode, $\lambda_k(j \mid i,m)$. This estimation step yields estimates of θ_C and ρ_C that are independent from the taste shifters $D_C(i,m)$, which vary by origin and mode. We obtain an estimate of $\theta_C = 2.97$ (with robust standard error equal to 0.14) and an estimate of $\rho_C = 0.75$ (robust s.e. 0.04). 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 percentage point. Both estimates are consistent with the literature. For example, Severen (2019) computes an estimate of θ_C equal to 2.2 using tract-level data for Los Angeles, and Monte et al. (2018) uses county-to-county commuting data covering all of the US, and obtain estimates $\theta_C = 3.3$ and $\rho_C = 1.34$.

We then estimate the preferences $D_C(i,m)$ that residents of a census tract i have for transport mode m. We model $D_C(i,m)$ as an exponential function of observed origin-specific demographic covariates $X_C(i)$ with mode-specific coefficients $\Psi_C(m)$. To identify these coefficients, we use again a GMM estimator and rely 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. We find that, everything else constant, residents prefer commuting by car relative to public transport,

[14] All expressions in this section correspond to the pre-CHSR full model equilibrium, as defined in Appendix E. To save notation, we omit the index s in every variable; e.g., the expression in (11) corresponds to $\lambda_C(i, j_C, m_C, s)$ for $s = N$, as defined in (A.22).

and by public transport relative to biking or walking. However, these preferences differ substantially across census tracts, with younger and more educated census tracts having a weaker preference for commuting by car, and with more nonwhite tracts strongly preferring public transport.

Importantly, these preferences shift usage of modes of transport across census tracts on top of any differences that may arise from commuting times and monetary costs. Using these estimates and expression in (11), we generate model-predicted commuting shares $\lambda_C(i,j,m)$, and use these shares in (8) to quantify $\hat{W}(i)$.[15]

Business and Leisure Travel According to our model, the share of business or leisure trips with origin in a census tract i and destination in j that use a mode of transit m is:

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

for $k=L$ (leisure) or $k=B$ (business). The shifter $Z_k(i,j)$ captures the leisure or business appeal of traveling to j; $D_k(i,m)$ is a preference shifter for using 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: $\mathcal{M}_B = \mathcal{M}_L = \{$airplane, private vehicle, public transport$\}$. Our procedure to estimate the parameters in (12) implements a two-step GMM estimator similar to that used to estimate the gravity equation from commuters in (11).

For leisure travel we estimate a coefficient on log travel time $\mu_L\theta_L\rho = 1.20$ (robust s.e. 0.28), and for business travel we estimate $\mu_B\theta_B\rho = 1.65$ -1.65 (robust s.e. 0.82). This time elasticity was -2.24 in the case of commuting, reflecting a higher disutility of time spent traveling in the case of commuting trips than in the case of leisure and business trips. However, as the expression in (12) illustrates, μ_k and θ_k are not separately identified. Thus, we 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 revenue spent on its employees' business travel. We set $\mu_B = 0.015$ following industry reports.[17] In combination with the calibrated values of μ_L and μ_B and our estimate of ρ, the estimated coefficients on log travel time imply $\theta_L = 31.99$ (with robust standard error equal to 7.35) and $\theta_B = 146.98$ (with robust s.e. equal to 73.38). Thus, travelers perceive business and leisure destinations as highly substitutable.

Using these estimates and the expression in (12), we generate model-predicted shares of long-distance trips $\tilde{\lambda}_k(i,j,m)$, for $k=L,B$. In doing so, we compute trip shares for origin and destina-

[15] Dingel and Tintelnot (2020) recommend using predicted shares rather than the raw observed shares to simulate impacts in quantitative spatial models, in particular in sparse settings like ours.

[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.

tion tracts that are less than 50 miles away, which are not included in the CHTS survey data. We use these predicted shares in (9) and (10) in order to quantify $\hat{W}(i)$.

4.3 Alternative CHSR Scenarios

We implement counterfactuals for several scenarios, summarized in Appendix Table A.2.

The "2008 Business Plan" scenario corresponds to the highly optimistic announcements of the 2008 CHSR business plan (California High Speed Rail Authority, 2008). The plan assumed the full project to be operational by 2030, with present value of capital costs for Phase-I at \$33b (as Phase-I was about 60% of the full project, we proportionally adjust the total cost to \$50b). Ticket prices were set at 50% of the corresponding airfare and said to finance maintenance costs. The "2022 Business Plan" scenario corresponds to the 2022 CHSR Update Report (California High Speed Rail Authority, 2022). Capital costs for the full project are updated to \$130b,[18] ticket prices are doubled, and the completion timeline is 50% longer than the 2008 announcement.[19] The report also estimates a 65% probability of delivering the project as planned.

For each of these cost assumptions, we compute a counterfactual that only accounts for the direct effects of the CHSR on travel time and ticket prices (as discussed in the previous section) and one that further accounts for general equilibrium (labeled "with GE") which further allows for impacts through land prices, wages, and spillovers due to employment agglomeration.

We compute counterfactuals using the system (A.37)-(A.46) in Appendix E.8, implemented at the census tract level. In addition to the gravity variables in a pre-CHSR equilibrium that were computed in the previous subsection, we also use for model quantification 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 from the sources detailed in Appendix A. When computing the "GE" version of counterfactuals, we use spillover elasticities from the literature as also detailed in Appendix A.

For each variable, we choose the latest year for which data is available before 2020; doing so also avoids noise from Covid-19 shocks. This choice means that the real-income impacts are computed using ex-post realized fundamentals from circa 2019. Section 5 discusses our procedure to address the bias introduced by using these realized fundamentals, rather than unobserved expected fundamentals as indicated by the model, to estimate the voting equation.

A final crucial component of the counterfactuals is the change in travel time and ticket prices caused by the CHSR. Appendix B details the construction of the times and costs before and after 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 introduce the CHSR, which is endogenously used only when it is beneficial to do so given time gain and ticket price, as described in Appendix E.6.

[18]In the report, capital costs are in between \$77bn and \$113bn for phase-I of CSHR. We adjust proportionally to the size of the full network and take an intermediate point between these estimates.

[19]The plan sets a timeline of 2030 for a small Central-Valley segment (180 miles), so we proportionally adjust the completion forecast of the full network by 50% relative to the 2008 baseline for the full network.

4.4 Time and Cost Shocks

A few statistics are useful to put the potential time gains from the CHSR in perspective. The average commute in California is 31 minutes one-way, with 7% of trips lasting more than 60 minutes. For trips above 50 miles, the California Household Travel Survey (CAHTS), conducted between 2010 and 2012, shows that a quarter of Californians undertook such trips every year. Out of 64.4 million annual leisure trips, the mean trip is 143 minutes long; and out of 7.1 million annual business trips, the median is 123 minutes long. Car is the most common travel mode, used by 90% of commuters, with both public transit and walking or biking accounting each for approximately 5% of commuting trips. Travel by car accounted for 96% of leisure trips (with the remaining evenly split between air and public transit), while 88% of business trips were done by car and 10% via air.

Table 1: The CHSR Shock

	% Initial Travelers Directly Better Off	Time Gain		Cost Change			
				2008 Business Plan		2022 Business Plan	
	Pub. Trans. or Air	median	75 ptile	med	75p	med	75p
Commute	1.0%	28' (35%)	47' (45%)	-18%	-1%	9%	29%
Leisure	0.4%	25' (15%)	50' (36%)	-55%	-30%	-24%	-7%
Business	4.6%	14' (6%)	29' (14%)	-62%	-56%	-31%	-22%

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 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 origin-destination-modes within each travel purpose, conditioning on origin-destination-modes where the CHSR is used when available.

Against this backdrop, both the 2008 and 2022 business plans promised sizable potential gains of the CHSR. Table 1 shows how many travelers are directly exposed to the CHSR shock, and by how much. The first column shows the fraction of all travelers who, before the CHSR is available, travel on routes for which the CHSR would be used if available. The remaining columns report moments from the distributions of time and cost changes for these direct winners. The direct beneficiaries save a substantial amount of time; e.g., the median gain for a commuter via public transit is 35 % of commute time. These direct winners also save a substantial amount in travel costs at the most highly optimistic 2008 projection for ticket prices. In the 2022 Business Plan scenario with higher projections of ticket prices, there are still large pecuniary gains among leisure and business travelers who directly benefit. However, for commuters, their time savings are partially offset by higher commuting costs. This difference comes from the fact that commuters do not use air travel, and therefore only gain time when switching from the relatively cheap public transit to the CHSR.

4.5 Distribution of the CHSR Impacts across Census Tracts

We now compute the distribution of expected real-income effects of the CHSR across census tracts, $\hat{W}(i)$ defined in (6). Figure 2 plots real-income effects across tracts in the baseline scenario with 2008 predictions and Figure 3 zooms into Los Angeles and the San Francisco Bay Area. Bright yellow tracts gain the most, and dark blue tracts lose the most. The effects are highly heterogeneous across space. The top 10% of tracts experience real-income gains between 0.5% and 4.1% per year, while 38.7% of all tracts lose. In the baseline scenario with 2022 business plan information, 92.0% of all tracts lose. Among winners, the tract that gains the most gains 1.7% per year. Appendix Table A.3 summarizes moments of the distribution of $\hat{W}(i)$ across census tracks in each of the 4 cases of the model.

To gauge the drivers of this heterogeneity, Table 2 shows regressions of these tract-level real-income changes on tract-specific characteristics in the pre-CHSR equilibrium: each tract's distance to the nearest CHSR station, the percentage of car and public-transit travelers, the average commute time, and fixed effects for belonging to Los Angeles, the San Francisco Bay Area or a Central Valley county. As we could expect, tracts that are closer to stations gain more. By this logic, all the largest cities in California should gain. However, tracts with lower car usage, more prone to using public transportation, or with longer commute times gain more, because these tracts are more likely to adopt the CHSR in some outbound trip. Furthermore, tracts located in Los Angeles county gain over and above what these characteristics would predict, suggesting that, due to their specific travel patterns, they are more likely to adopt the CHSR. In contrast, Central Valley locations like Fresno or Bakersfield have low gains, both because of the direct effect of covariates like high car usage, and because of their specific travel patterns.

Importantly, the heterogeneity across space is consistent across the different cost scenarios, with correlation across tracts of at least 95.9% between any two scenarios. As we shall see in Section 5, despite the different aggregate effects across model variants shown in Table A.3, the robust patterns of spatial heterogeneity across model variants are always consistent with a strong elasticity of votes to real income. Furthermore, as we will see in the Section 6, the significant gap between the gains in L.A. and the Central Valley would have been even larger in the absence of political incentives for the planner.

In terms of aggregate welfare impacts the model implies aggregate gains of 0.16% under the overly optimistic 2008 CHSR promises, and losses of -0.21% based on 2022 CHSR information.[20]

[20] There are many differences in method, underlying data, and details of implementation between our analysis and existing cost-benefit analyses of the CHSR. Still, we can broadly compare the ballpark of our numbers against existing estimates that have considered related forces. Initial 2008 estimates by the High-Speed Rail Authority (California High Speed Rail Authority, 2008) estimated net present-discounted gains of \$97bn in 2008 USD. In our baseline, the analog number is \$69.9bn. The 2022 CHSR project update (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 case without indirect effects), 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 can be compared to this estimate, although it implements the full CHSR (instead of just phase-I) by scaling up costs in proportion to the additional miles of the total network relative to phase-I. This scenario yields a loss of \$-91.6 bn in 2021 USD. The 2023 CHSR update also reports gains of \$26bn for the Phase I when "wider economic benefits for worker and

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

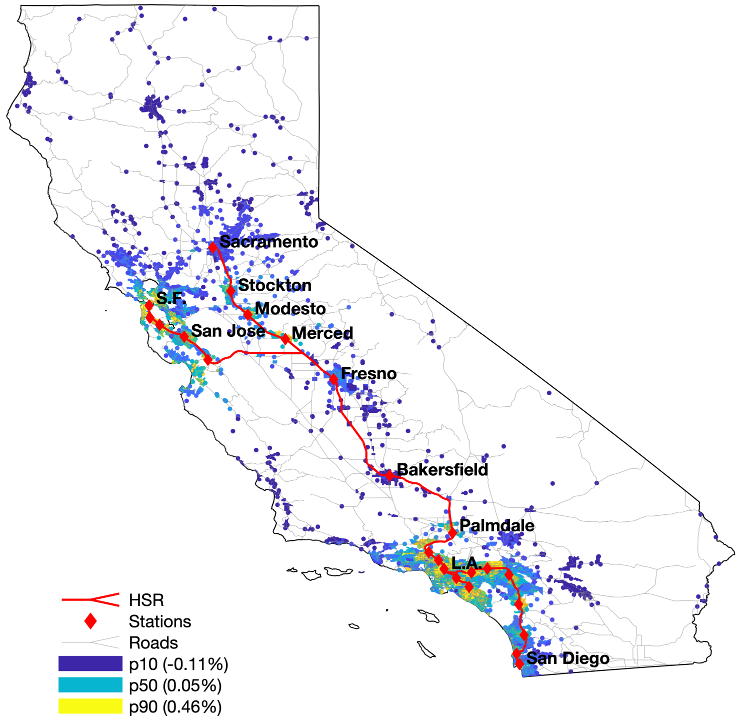


Table 2: Heterogenous Gains from CHSR

	2008	2008	2008 GE	2022	2022 GE
dist. to station	-0.00087***	-0.00100***	-0.00115***	-0.00047***	-0.00058***
% public transit	0.03817***	0.03399***	0.03489***	0.01668***	0.01714***
% car	0.00466***	0.00106**	0.00101*	0.00050**	0.00046*
commute time	0.00003	0.00118***	0.00135***	0.00052***	0.00063***
LA fixed effect		0.00109***	0.00096***	0.00057***	0.00052***
SF bay area f.e.		0.00032***	0.00053***	0.00030***	0.00037***
Central Valley f.e.		-0.00111***	-0.00071***	-0.00051***	-0.00047***
R2	0.684	0.714	0.656	0.717	0.666
N	7,866	7,866	7,866	7,866	7,866

Note: The table shows the result from running, for each model variant, an OLS regression of the tract-level real-income change, $\hat{W}(i)$, on observable tract characteristics. "dist. to station" is the tract centroid's distance to the nearest CHSR station, "% public transit" and % car" are the fraction of residents who initially commutes via public transit or via car, and "commute time" is the average commuting time by road. The LA,SF 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 of the Central Valley, respectively.

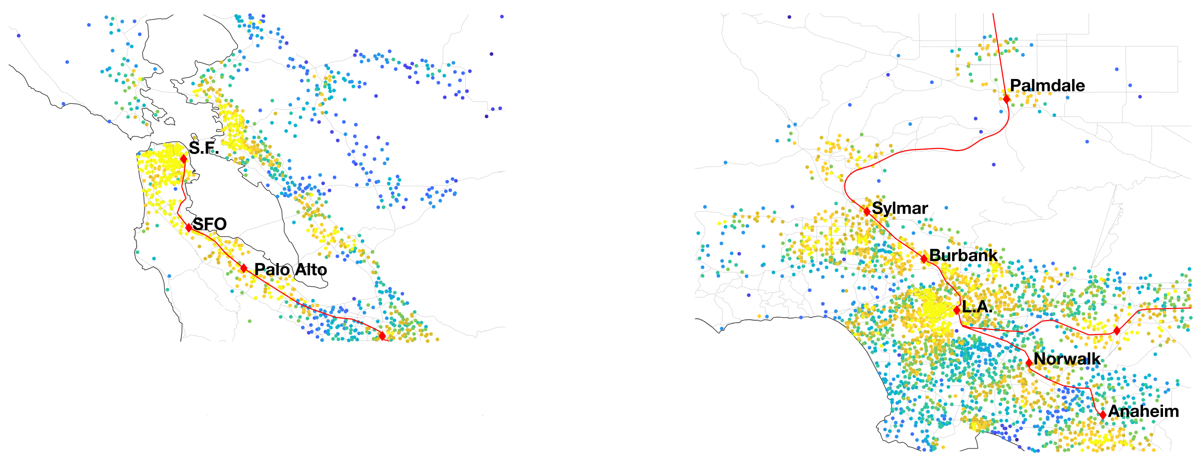
In Section 6 we compare these predictions with those from the optimal station distribution decided by a planner without political preferences.

Appendix table A.4 further decomposes the aggregate effects into those stemming from com-

firms" are further included in addition to rail user benefits. These can be compared with our estimates including indirect effect for the 2022 case. These still correspond to a loss, although lower: $-76.2 bn for the full network.

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

(a) SF Bay Area (b) LA County



Note: The maps shows percentiles of tract-level real income effects $\hat{W}(i)$ in the baseline scenario.

muting, leisure travel, and business trips. Without general-equilibrium effects, commuting account for almost all of the aggregate effect –even though long-distance leisure trips save more time than the average commute trip, aggregate commuting time dwarfs leisure travel time. The magnification of gains in the case with general-equilibrium impacts are due almost exclusively to the productivity-enhancing effects of business trips.

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 (4), 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' expectations of the 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 issue is an instance of measurement error in the covariate of interest, the second one is an example of omitted variables that are correlated with that covariate. If not properly addressed, both issues may bias our estimates of θ_V. We discuss in detail these identification concerns, as well as the strategies we follow to address them.

Measurement Error First, we discuss how we address potential measurement error in our covariate of interest. 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 (4). Thus, we identify θ_V as the coefficient on a proxy of those expectations. Specifically, we use the model-implied ex post real income impact of the CHSR as such proxy. That is, we use $\hat{W}(i)$, defined in (6) and computed in Section 4.5. Because this variable is constructed using realized fundamentals 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 for the CHSR impact predicted by our model, denoted by $\epsilon_{W,1}(i)$; and, (b), 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:

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

If voters' expectations are rational, their expectational error $\epsilon_{W,1}(i)$ is analogous to classical measurement error in the covariate of interest, biasing OLS estimates of θ_V towards zero. Addressing this issue requires an instrument that belongs to the voters' information set when they voted (Dickstein and Morales, 2018). We present below estimates that use as such instrument the model-implied ex post real income impact of the CHSR constructed using fundamentals from 2008, denoted by $\hat{W}_{08}(i)$. If voters' expectations are rational and the 2008 fundamentals belong to voters' information sets, this instrument is 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 estimator of θ_V to address measurement error bias. As we discuss 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 perfectly aligns with their party affiliation.

Characterizing the bias due to the potential model misspecification error $\epsilon_{W,2}(i)$ is more complicated, as it depends on its precise source; that is, it depends on which particular model aspect is inaccurate. To address this potential bias, we explore how robust our estimates of θ_V are to the alternative economic models described in Table A.2. These models span polar cases of beliefs about costs (using 2008 or 2022 estimates) and economic forces (either abstracting from or incorporating indirect general-equilibrium impacts).

Omitted Variables Second, we discuss how we address any potential correlation between our covariate of interest and other determinants of voters' preferences. These other determinants enter through the component $\hat{a}(i)$ in (4). Our first approach to limit the risk that our estimates of θ_V are affected by omitted variable bias is to introduce proxies for subjective considerations that may play an important role in determining 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 different sets of covariates $X_k(i)$ in (16). The first set bears a direct relationship to voters' political ideology: it includes the tract-specific share of registered Democrats and the tract-specific shares of votes cast in favor of two propositions, Prop. 10 and Prop. 1B, on different energy and transportation projects, respectively.[21] For each census tract, 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 of the residents of each census tract: the shares of residents who are nonwhite, college-educated, or under 30 years of age. Finally, the third set of covariates includes the time it would take voters in each census 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 three sets of 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.

Importantly, 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 the structural parameter θ_V. The interpretation of the coefficients on these covariates is not of immediate interest for our purposes; those coefficients may be capturing the impact of multiple treatments that are correlated with those controls. For example, the coefficient on the tract-specific share of registered Democrats may capture not only differences in preferences for the CHSR as a function of political ideology but, as we discussed above, also differences in expectations for the future real income impact of the CHSR. For our purposes, this is irrelevant, as we do not study the impact of changes in voters' party preferences.

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

[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] We highlight that the model to construct $\ln \hat{W}(i)$ already captures each census tract's preferences for using different means of transport (car, public transit, airplane, or biking), as described in our gravity estimation. These preferences drive the potential usage of the CHSR were it to become available, entering through $\ln \hat{W}(i)$.

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

$\ln \hat{W}_{08}(i)$ as an instrument will also be biased.

We consequently also present estimates of θ_V that rely on an instrument constructed as follows: in a first step, we simulate one hundred *counterfactual* CHSR networks by randomizing the location of the 24 stations along the effectively 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 stations along the actual CHSR proposal, 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.1. Crucially for the validity of this second instrument, these three alternative CHSR routes were selected on the basis of technical feasibility and cost savings alone.[24] In other terms, while political considerations may have determined the final choice between the three alternative routes (and the location of the 24 stations), political considerations were not behind the decision to limit the set of possible routes to these alternative 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 the drawn route. Our main specification adopts this instrument.

As these two instruments incorporate information on neither the actual location of the 24 projected stations (in the case of the first instrument) nor the actually projected CHSR railway line (in the case of the second instrument), they may be valid even if the location of the CHSR stations (or 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 (4), we obtain our estimating

[24] There are two possible routes from Northern to Southern California: along the coast or through the center of the state (along 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 I-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.

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)
$log(\hat{W}_{19})$	41.91^a	18.29^a	15.83^a	15.13^a	17.96^a	20.90^a	22.05^a
	(1.82)	(1.30)	(1.05)	(1.09)	(1.30)	(1.76)	(1.83)
Log-odds Dem. Sh.		0.30^a	0.39^a	0.38^a	0.38^a	0.39^a	0.39^a
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Environ.: Prop. 10		1.15^a	2.44^a	2.44^a	2.42^a	2.40^a	2.40^a
		(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Transp.: Prop. 1b		1.55^a	0.82^a	0.82^a	0.81^a	0.81^a	0.80^a
		(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Sh. non-White			-0.16^a	-0.17^a	-0.17^a	-0.18^a	-0.18^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	0.73^a
			(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Sh. Under 30			0.18^a	0.17^a	0.17^a	0.17^a	0.17^a
			(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Log. Dist. Station				-0.01^a	-0.01^c	-0.00	-0.00
				(0.00)	(0.00)	(0.00)	(0.00)
F-stat					1197	672	466
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.

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

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. We use voting data for Prop. 1A as well as other elections from UC Berkeley's Statewide Database (University of California, Berkeley, 2008). The sources for the remaining covariates are discussed in Section 4.1.

In Table 3, we report OLS and IV estimates of θ_V and the β_k's where the expected economic impact of the CHSR is built using the economic model with only direct CHSR effects and construction costs set to their 2008 announced levels. 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 the proxies for political ideology; in column (3), we add variables that capture the demographic composition of the residents of each census tract; and, in column (4), we add as a control the log of the shortest time distance from each tract to its closest station.

The estimated value of θ_V equals 15.13 (with robust standard error equal to 1.09) in the specification 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.

The remaining columns (5)-(7) report the IV estimates, with the same set of controls. Column (5) uses as instrument the 2008-based real income measure, column (6) uses the instrument that randomizes the location of the stations alone, and column (7) uses the instrument that randomizes the location of both the railway line and the stations. For these three instruments, we obtain large first-stage F-statistics.

The estimate of θ_V increases progressively as we move from column (4) to (7). The fact that the IV estimate of θ_V that uses as instrument the 2008-based real income measure is larger than the OLS estimate (17.96 vs. 15.13) 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 are larger than that reported in column (5) reveals that the CHSR design favored census tracts whose residents were, for unobserved reasons, less predisposed to support the proposed CHSR.

Results across Model Variants Appendix Table A.5 replicates the estimates from column (6) and (7) in Table 3 across the model variants described in Table A.2.

The estimates of θ_V are larger when we compute the real-income gains from the CHSR using the 2022 information than when we use the 2008 business plan to inform the CHSR cost and completion probabilities. To understand these estimates, one should bear in mind there is a high correlation across tracts (greater than 95.9%) in the economic impacts $\hat{W}(i)$ predicted by the two model variants; however, the variance of these tract-specific predictions is much smaller in the model that uses 2022 information.[25] This leads to a higher θ_V estimate.

The estimates of θ_V are smaller when incorporating general-equilibrium impacts. Given completion beliefs p as determined by a given business plan, the model that incorporates general-equilibrium effects 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. This extra dispersion results in a smaller estimate of θ_V.

Robustness Our results are also qualitatively very similar across a range of empirical specifications, including different weighting schemes, sample selection, and covariate selection. Appendix Table A.6 replicates the estimates from columns (6) and (7) in Table 3 but for specifications that weight the results in each census tract by either the total number of votes (in columns (3) and (4)) or by the participation rate (in columns (5) and (6)). When implementing either of these

[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 based on the 2022 business plan, leads to a higher θ_V.

two weighting schemes, we obtain IV estimates of θ_V that are slightly larger than the baseline (unweighted) estimates reported in Table 3. Columns (7) and (8) in Appendix Table A.6 show that the estimates of θ_V are minimally affected 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

We discuss key implications from our findings. First, the estimates and identification assumptions discussed above imply that we cannot reject the joint hypothesis that agents cared about the expected real income impact of the CHSR and that this variable is correlated with our model predictions. These estimates thus provide support for our estimated model as a predictor of the spatial distribution of economic impacts of the CHSR.

Second, the result also implies that economic voting is a driver of policy preferences over transport infrastructure. Economic voting is arguably a strong driver of votes in electoral politics (Lewis-Beck and Stegmaier, 2000); however this not necessarily the case regarding innovations in transport systems, with case studies and survey evidence suggesting low public acceptability for transport policies with seemingly positive net economic impacts.[26] In contrast, using data on actual votes, we show that changes in the expected economic impact of the CHSR does have a significant effect on voters' preferences.

Third, and crucial for our ultimate goal of recovering policymakers' preferences, our estimates imply that policymakers who value public support for infrastructure projects may want to shift the supply of infrastructure towards areas where its marginal impact on public support is greater, at the expense of where it may be socially more desirable. To get a sense of this trade-off, which we quantify in the next section, consider a change in the CHSR design that increases economic gains in a given location i by $d \ln \hat{W}(i)$ compared to the original CHSR design, at the expense of lower gains in other locations. This change impacts votes in tract i according to the following semi elasticity:

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

That is, the vote response to a given real income increase is lower in tracts with more extreme vote shares. 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, an extra 0.1-0.2 percentage points in expected economic gains swayed one percent of local votes in the median tract (to put this number in context, the standard deviation of favorable votes across census tracts was 0.1 percentage points). If the planner values votes, her decisions will be distorted: CHSR designs will be tweaked to favor locations where this elasticity is greater, rather than where it strictly maximizes aggregate real income. We evaluate this tradeoff in the next section.

[26] The most salient example is road pricing. See for example Verhoef et al. (1997), Schade and Schlag (2003), and Noordegraaf et al. (2014).

Table 4: Effect of Economic Preferences on the Vote

Model Variant	θ_V	Gain to Sway 1 pp (median tract)
2008 Business Plan	22.0	0.2%
2008 Business Plan, with GE	18.1	0.2%
2022 Business Plan	44.6	0.1%
2022 Business Plan, with GE	35.8	0.1%

Note: The θ_V reported corresponds to IV estimate using the random routes instrument shown in Table 3. The second column reports the 50th percentile of the distribution across tracts of the inverse of $\frac{dv(i)}{d\ln \hat{W}(i)}$ in (17). This inverse measures the real-income gains required to sway 1% of votes around the equilibrium with the CHSR.

6 Policymakers' Preferences and Optimal Designs

We embed the spatial model and the voting choice into the problem of a politically-minded planner who decides where to locate stations. Our goal is to estimate the preferences of a social planner designing the CHSR, and then compute counterfactual optimal designs under alternative preferences that shut down the political component of preferences.

The planning problem captures, in reduced form, a complex process –involving politicians, technical experts, and the corporate sector– leading up to the referendum.[27] Rather than modeling this process, we propose an "as if" representation, as those of a hypothetical planner with preferences over residents of different locations and aggregate votes. Our estimation identifies the distributional biases emerging from this complex process.[28] Even if voting over an infrastructure project is somewhat unusual, policymakers' concerns with popular approval doubtlessly are pervasive: authorities may keep an eye on this approval over an infrastructure project even when an actual vote does not take place.

Section 6.1 sets up the planners' problem, Section 6.2 estimates the planner's preferences, and Section 6.3 solve for the optimal station distributions of planners without political preferences.

6.1 Planner's Problem

To set up the planner problem, one could use an approach in the spirit of Fajgelbaum and Schaal (2020) for roads or Kreindler et al. (2023) for buses, where an entire network is flexibly designed.[29] In our case, this would amount to designing both the railway lines the locations of the

[27] See https://hsr.ca.gov/about/board-of- directors/board-members for the list of CHSR board members and https://hsr.ca.gov/about/board-of-directors/ for the committee rules. Committee members are chosen by elected officials and serve a four-year term.

[28] In the context of models of tariff formation, Baldwin (1987) points out that different political economy models can map to a planner's problem with group-specific weights; Grossman and Helpman (1994) model a political game with lobby contributions that can be represented by a planner's problem with sector-specific Pareto weights; and Adão et al. (2023) estimate the Pareto weights across industries and skill groups of a hypothetical planner setting US tariffs.

[29] Unlike in Fajgelbaum and Schaal (2020), our optimization problem is subject to non-convexities due to the mix of substitutabilities and complementarities among possible station locations.

station along these lines. Doing so would raise methodological challenges due to dimensionality, and would also raise the need to model the heterogeneous costs of building rail lines over space. However, in reality, only few routes were identified as being potentially feasible by planners based on cost and engineering considerations. In what follows, we therefore restrict the planner's problem to choose the location of stations along those feasible routes.

Specifically, since its inception the CHSR was meant to connect Los Angeles and San Francisco, further branching to San Diego and to Sacramento in later phases. As we have discussed in Section 5.1 when introducing our main instrument, three possible routes linking Northern and Southern California were identified several years prior to the 2008 vote (US DOT, 2005): one route ran along the coast, and the other two ran through the center of the state, either along I-5 highway or through the Central Valley (see Appendix Figure A.1). The coastal route was discarded due to much higher costs, leaving the I-5 and the Central Valley routes as the main technologically feasible options according to transport engineers. Therefore, the main decisions of the political authorities were: which of these two routes to use, and where to place the stations along the chosen route. When solving for optimal station placements under counterfactual planner preferences, we consider both routes.

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. Hence, each $\mathbf{d}$ maps to an utility level of each agent ω if the vote is approved ($u_\omega(Y; \mathbf{d})$ defined in (1)), and to a share of favorable votes $v(i; \mathbf{d})$ defined in (4). 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 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 location i. We define these weights as a function of observed covariates $Z_k(i)$ in each tract (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 California residents that vote in favor of the CHSR; when $\lambda_V > 0$, the planner attaches a positive weight to favorable votes.

Conditional on the project being approved, the design of the CHSR internalizes the non-economic component of preferences $\hat{a}(i)$ through the votes. We note that $\hat{a}(i)$ is independent from $\mathbf{d}$; i.e., the design impacts preferences only through the forces included in our economic framework. To restrict the dimensionality of the parameters to be estimated, we do not model location-specific costs of placing a station; while this heterogeneity could be more explicitly modeled, it is here indirectly captured in the estimation by the coefficients b_k.

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

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

where the expectation is taken over future realization of unknown fundamentals at the time of designing the CHSR. Using the definition of u_ω in (1), and assuming that voters' information set about fundamentals at the time of voting is at least as rich as the planner's, the optimal design $\mathbf{d}^0$ that maximizes the planner's objective function in (6.1) is:

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

We now consider optimality conditions. Similar to voters, we assume the planner has rational expectations. Hence, for any alternative design $\mathbf{d}^n$ different from the optimally chosen CHSR design $\mathbf{d}^0$, the objective function would have changed according to:

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

where $\frac{\partial v(i)}{\partial \ln \hat{W}(i)}$ was defined in (17), and where

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

is the log difference in real income between the counterfactual design $\mathbf{d}^n$ and the actual CHSR design $\mathbf{d}^0$, and where $\epsilon\left(\mathbf{d}^n\right)$ is the planner's expectational error when evaluating the impact of the high-speed rail design. Because we assume that the observed CHSR corresponds to the optimal choice of the planner, any deviation such as (22) must yield weakly negative returns.[30]

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 $\lambda_U\left(i\right)$ or higher returns in terms of votes, as captured by $\partial v\left(i\right)/\partial \ln \hat{W}\left(i\right)$ 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 $\lambda_U\left(i\right)$ relative to $\lambda_U\left(i'\right)$. 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 $\lambda_U\left(i\right)$'s.

6.2 Estimation of Planner's Preferences

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 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.

[30] Note that the realization of the planner's welfare given observed fundamentals, $\mathcal{W}\left(\mathbf{d}^n\right) - \mathcal{W}\left(\mathbf{d}^0\right)$, could be positive. I.e., there can be feasible designs that would have been preferred by the planner, had the planner known the fundamentals. Those cases are rationalized by sufficiently large expectational errors $\epsilon\left(\mathbf{d}^n\right)$.

Moreover, closed-form solutions for optimal policies, which they can exploit for implementation, are unavailable in our case when optimizing over geographic coordinates.

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 estimation is only feasible in binary-choice settings when one can determine ex ante whether any two binary choice decisions of a decision maker are complements or substitutes (see, e.g., Jia (2008) and Arkolakis et al. (2023)). Unfortunately, while our setup would belong to this class of binary-choice problems if we restricted the choice set $\mathscr{D}$ to be discrete, the binary decisions in our objective function 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 we add stations. Therefore, to implement our moment inequality approach, we must model all unobserved terms as expectational errors of the planner (see Jeon and Rysman (2024) for a similar approach).

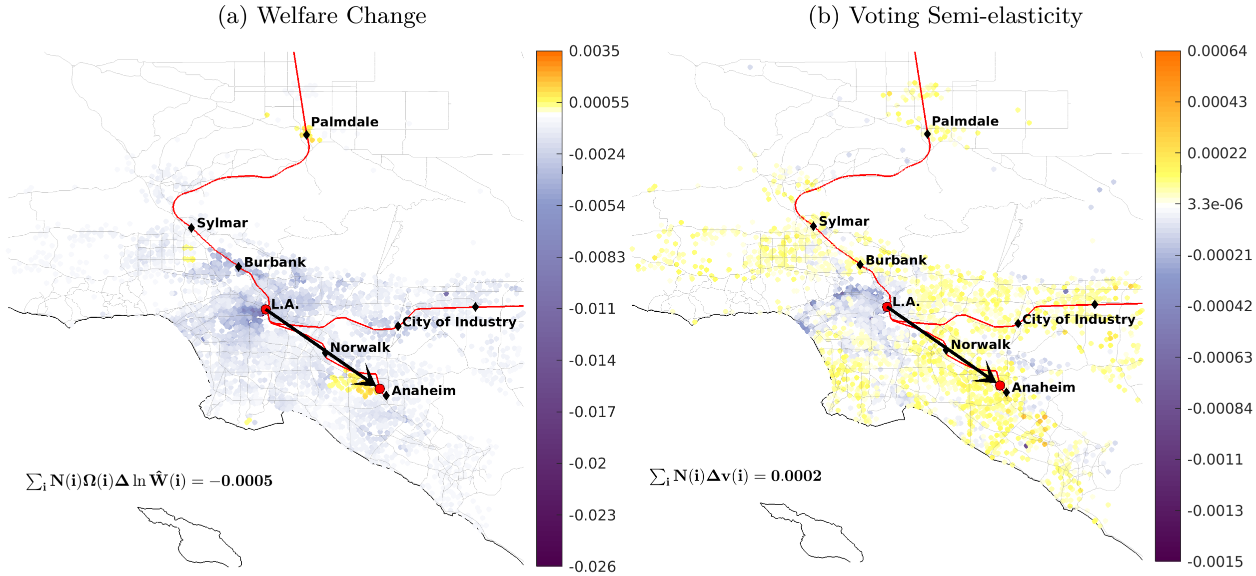
Construction of Perturbations and Moments To build the inequalities described in (22), we construct hundreds of perturbations to the distribution of stations along the proposed CHSR line. In each perturbation, a single station is moved to another location without a proposed station. The set of potential locations is chosen to identify upper and lower bounds on the parameters; specifically, we use 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 line.[31] We index each one-station deviation by n and let $\mathbf{d}^n$ be the associated design. In this way, we obtain N perturbations $\Delta\mathcal{W}(\mathbf{d}^n)$ as defined in (22).

Figure 4 shows one such perturbation, where the Los Angeles station is shifted towards Anaheim. The map on the left shows the real-income differences with the actual design, $\Delta \ln \hat{W}(i)$ defined in (23). The map on the right shows the voting semi-elasticity, $\partial \mathrm{v}(i)/\partial \ln \hat{W}(i)$. Downtown LA is a low-voting gradient area because it strongly supports the CHSR, so increasing real income barely changes votes. This perturbation redistributes real income from densely-populated, low-voting gradient areas of L.A to higher-voting gradient places in Orange County (closer to median voters). As this perturbation implies a reduction in the real-income component of the planner's objective and an increase in aggregate votes but was not chosen, it helps identify an upper bound on the preference for votes λ_V relative to an average of the $\lambda_U(i)$.

For the 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),

[31] Figure A.2 in the Appendix shows such peaks and throughs for $Z_k(i)$ equal to population density.

Figure 4: Example of Perturbation Identifying λ



Note: 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.

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 $m(\mathcal{N}_e)$.[32] Then, using (19), we can rewrite the moment inequality e in (24), as

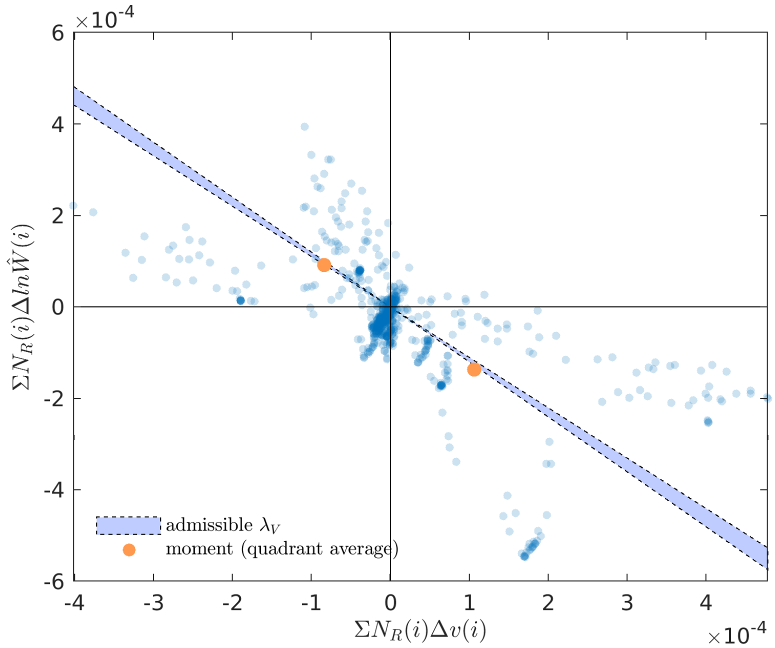
$$m(\mathcal{N}_e) \equiv \sum_{n \in \mathcal{N}_e} \left[\sum_{k=0}^{K} b_k Z_k(i) + \lambda_V \theta_V v(i)(1 - v(i)) \right] N_R(i) \Delta \ln \hat{W}(i, \mathbf{d}^n) \leq 0 \tag{25}$$

for $e = 1,..,E$.

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

[32] 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 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



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's objective function has $b_k = 0$ for all demographics $k = 1...K$ except for the constant term $b_0 \neq 0$ (so, no distributional preferences). In this case, the planner 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 ($\sum_i N_R(i) \ln \hat{W}(i; \mathbf{d}^n)$). The actual CHSR proposal corresponds to the $(0, 0)$ point.

In this case where we only need to identify the parameter 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 are uninformative because they 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 large expectational errors.

In contrast, the perturbations in the upper-left and lower-right quadrants are informative about the parameters because they suggest a trade-off between welfare and votes. The orange dots are two of the moments $m(\mathcal{N}_e)$ defined in (25), which average across all perturbation within each quadrant. The admissible values for λ_V (relative to b_0) are such that $y + \lambda_V x \leq 0$, where (x, y) are the coordinates of each of the two moments depicted in the orange dots. Visually, this set

Table 5: Planner's Preferences Estimates (2008 Business Plan)

Observable	Pareto weight parameters b and λ				
	(1)	(2)	(3)	(4)	(5)
Density		[0.00, 1.13]			[0.00, 0.14]
Share college			[0.23, 0.50]		[0.12, 0.50]
Share non-white				[-0.33, 0.50]	[-0.23, 0.14]
Votes	[2.60, 11.07]	[4.24, 21.26]	[5.52, 8.95]	[3.45, 18.23]	[6.79, 9.14]
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.

corresponds to drawing a line through the origin with (negative) slopes x^{UL}/y^{UL} and x^{LR}/y^{LR} (where UL and LR means the upper-left and lower-right quadrant) and taking the intersection of the areas between these lines, as represented by the purple area on the graph.[33]

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 determine admissible sets.

Parameter Estimates Table 5 shows estimates of the b's and λ_V under the 2008 Business Plan across specifications that allow for covariates that include: population density, the share of college-educated residents, and the share of non-white residents. In the estimation, the support of the parameters is not restricted, and each parameter b_k or λ_V can take any positive or negative value.

To simplify the interpretation, we standardize every covariate $Z_k(i)$ and impose that the population-weighted mean of the Pareto weights $\lambda_U(i)$ equals 1. Each parameter b_k can thus be interpreted as capturing the impact of a one-standard deviation increase on the covariate relative to the average Pareto weight.[34] For each specification and parameter, we report the minimum and the maximum of the admissible values corresponding to a 95% confidence set. Column (1) of Table 5 corresponds to the case discussed in 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.

The estimation reveals a strictly positive preference for votes across the various specifications. Population density and the share of non-white residents do not appear significantly different from 0, as 0 belongs to the 95% confidence set. Since the $\lambda_U(i)$ are defined as per-capita Pareto weights, b_{density} multiplies total population squared, so a value of this parameter equal to zero still implies a strong utilitarian motive, entering through the constant b_0. Our estimates in columns (2) and (5)

[33] I.e., the set is $0 \leq -x^{UL}/y^{UL} \leq \lambda \leq -x^{LR}/y^{LR}$.

[34] 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.

suggest a positive preference for census tracts with a larger share of college-educated residents. In what follows, we use column (5) as our baseline specification.[35] Table A.7 in the Appendix repeats the estimation using the 2022 cost predictions –the values of the b_k's are unchanged, while λ_V is rescaled due to the different magnitude of welfare gains and our different estimate for θ_V.

Appendix Figure A.4 shows the distribution of total Pareto weights $\lambda_U(i)+\lambda_V\theta_V v(i)(1-v(i))$ across tract i implied by the centroid of our confidence set. The estimates imply a large variance in the Pareto weights, so that the planner is far from the utilitarian benchmark.

6.3 Optimal Station Placement with Apolitical Preferences

To demonstrate the importance of political and distributional considerations for the design of transportation policy, we conduct counterfactual exercises that show the optimal station placement for a planner with different preferences.[36] To compute the counterfactual optimal designs, we use the centroid of our 95% the confidence set. Since population density and the share of non-white residents are insignificant, we use the centroid of the confidence set conditional on these two parameters being zero; this implies: $b = [0.00, 0.39, 0.00]$ for [density, share college, share non-white] and $\lambda = 8.27$.

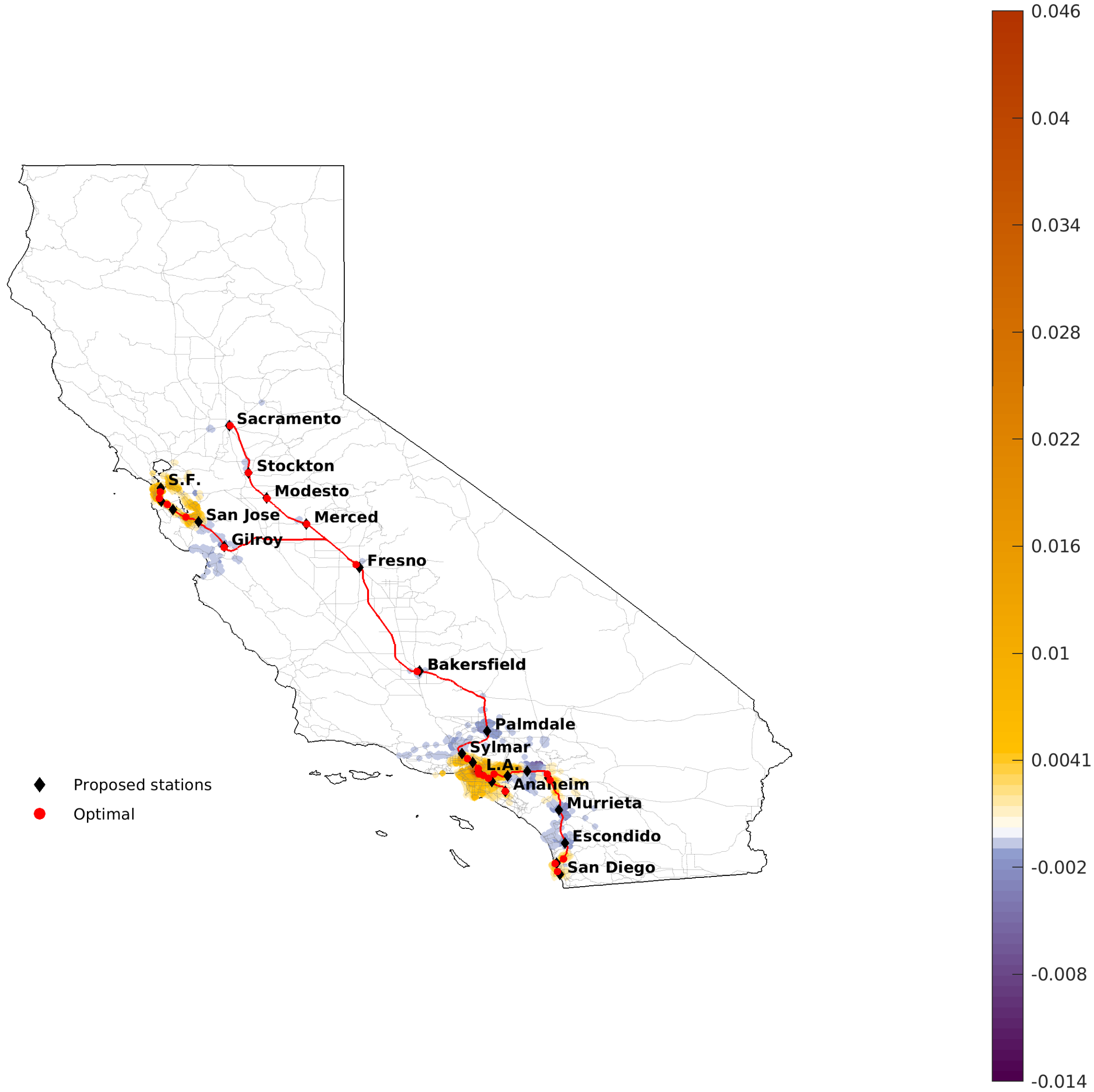
Optimal Stations Along the actual CHSR Design We compare the actual proposed CHSR design to that preferred by an *apolitical* planner whose objective function does not depend on votes (i.e., $\lambda_V = 0$), while the remaining parameters remain at the centroid of the confidence set, thereby eliminating the planner's incentives to assign higher real income to locations in higher voting-elasticity areas. In this first exercise, we restrict the planner choice set to the actual CHSR design, and below we allow for the alternative design along the I-5 highway.

Figure 6 shows in red the optimal location of stations corresponding to the apolitical planner, and in black the actual CHSR plan, along with the spatial distribution of real income changes. The optimal counterfactual design is quite different from the proposed plan. Broadly speaking, political motives shift stations away from high-density areas of L.A. and San Francisco. These locations receive the majority of commuting flows from nearby areas, yet are already strongly biased in favor of the CHSR (meaning that they have a low voting elasticity due to a strong preference for the high-speed rail already). Hence, in the absence of political motives, several originally proposed stations in suburban areas of San Francisco (such as San Jose), Los Angeles (Palmdale, Sylmar, Burbank) and San Diego (Escondido) reallocate towards their corresponding metropolitan areas.

[35] Figure A.3 in Appendix F displays the topology of the confidence set from column (5) by showing pairwise projections for each combination of the parameters b_k and λ.

[36] The optimal station placement problem is non-convex due to substitutabilities and complementarities across stations, as well as due to the sigmoidal function that defines the voting probabilities. For implementation, we use a three-step optimization procedure that combines the perturbations from the estimation stage, simulated annealing, and a continuous optimizer. Our procedure does not guarantee a global optimum, but we verify that it yields the best outcome when each station is individually shifted over 10 km range from the proposed optimum. Appendix Figure A.5 shows the planners' objective function when each station is reallocated and Appendix F.2 provides additional details.

Figure 6: Apolitical Planner, Optimal Station Placement



Notes: The maps shows the change in welfare $\Delta \log \hat{W}(i)$ in each location according to color scale on the right.

Table 6 shows the difference in aggregate real income (defined as the population-weighted average of $\ln \hat{W}(i)$) and votes between the optimal apolitical CHSR design and the actual CHSR proposal, under both the 2008 and the 2022 Business Plans. As expected, aggregate real income increases and the aggregate vote declines when the planner does not care about votes. The magnitudes are substantial: with 2008 costs, the planner foregoes 0.16% percentage points of aggregate real income to obtain 0.15% more votes, out of a baseline gains equal to 0.16% from the proposed CHSR in Table A.3. Using the 2022 cost projections, the aggregate loss of -0.21% becomes a loss of -0.12%. In sum, a politically-blind planner would have a designed a network with twice the expected economic gains (using 2008 costs) or about 40% smaller losses (using 2022 costs).

Table A.8 in the Appendix displays the cross-sectional impact of the counterfactual CHSR

Table 6: Apolitical planner, Statistics

	2008 Business Plan	2022 Business Plan
Aggregate Welfare vs. Baseline CHSR	0.16%	0.08%
[min max] across tracts	[-1.42%, 4.54%]	[-0.67%, 2.68%]
Aggregate Vote vs. Baseline CHSR	-0.15%	-0.08%
[min max] across tracts	[-12.89%, 1.35%]	[-7.62%, 0.46%]

Notes: The table shows differences in outcomes between the optimal counterfactual design $\mathbf{d^a}$ of an apolitical planner with $\lambda_V = 0$ and the baseline CHSR design. Aggregate welfare is the population-weighted average of real income differences with respect to the baseline CHSR, $\sum_{i=1}^{J} N_R(i) \ln \hat{W}(i; \mathbf{d^a})$. The rows [min max] show the support of $\ln \hat{W}(i; \mathbf{d^a})$ across locations.

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.[37] Appendix Figure A.6 zooms 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.

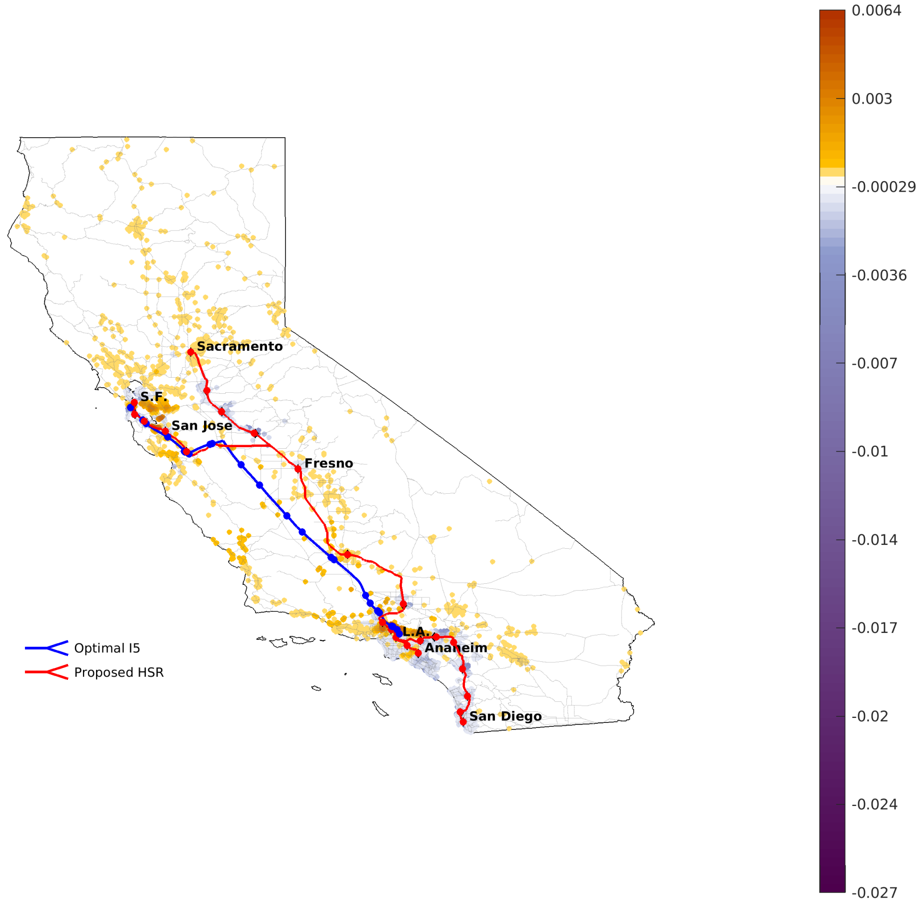
The optimal station distribution of a utilitarian planner ($\lambda_U(i) = 1$ and $\lambda_V = 0$) does not differ significantly from the apolitical one, except for some stations due to a redistribution from high to low college share areas, as is illustrated in the case of San Diego in Appendix Figure A.7. The main beneficiaries from eliminating the planner's preference for high college-share locations are areas in downtown San Diego with a mix of high and less educated residents. In spite of this, our results suggests overall aggregate welfare gains of 0.16% in the utilitarian case (2008 Business Plan), suggesting that the preference for votes drives most of the difference between the observed network and the optimal design, regardless of whether we consider the apolitical or utilitarian planner preferences.

Optimal Stations along the Alternative I-5 Design As we have discussed, the route running along the I-5 was a second candidate for the CHSR. The I-5 route differed from the proposed plan in that it provided a faster connection between San Francisco and Los Angeles through a more direct path along the I-5 highway. It also served less destinations by removing the Sacramento, Anaheim, and San Diego segments. A key argument in favor of this route was also its lower cost –the overall length of the I-5 route was 603.57km as opposed to 1242.02km for the proposed plan.

A natural question is whether this route was not chosen due political preferences. To answer

[37] Additional details are provided in Table A.9 in Appendix F, which shows how welfare changes between the optimal plan and one in which each station is individually moved back to its location in the original CHSR proposal. In line with our previous explanation, the largest welfare gains arise from stations that move closer to dense urban areas, also causing the largest vote losses (such as San Francisco, San Jose, Palmdale, Burbank, Sylmar)

Figure 7: Welfare Change between Optimal I5 and Proposed CHSR Plan



Notes: 2008 Business Plan estimates. The maps shows the change in welfare $\Delta \log \hat{W}(i)$ (I5 minus baseline) in each location according to color scale on the right under the estimated policymakers' preferences.

this question, we solve for the optimal distribution of the 24 CHSR stations along this route using the apolitical policymakers' preferences (i.e., setting $\lambda_V = 0$ while leaving the remaining parameters at the centroid of the confidence set). We reduce the building costs and taxes in proportion to the length of the I-5 route relative to the proposed plan. We then check if, contrary to the actual planner with political preferences who chose the Central Valley design, an apolitical planner would have chosen the I5 route.[38]

Figure 7 displays the I-5 route with the optimal station placement of an apolitical planner, and shows the welfare changes relative to the proposed plan using 2008 projections. The lower total costs of this route compared to the actual proposal results in welfare gains across the map, including even in areas served by the CHSR where the gains were modest in the proposed plan (for instance in the Central Valley). The main losers from this alternative design are areas South and

[38]We also solve for the optimal station distribution along the I5 using the estimated planner's utility with preferences for both redistribution and for votes. In this case, the planner is practically indifferent between the proposed CHSR and this optimal design on the I5. This result is not by construction, as the I5 route was not used to construct any perturbation for the estimation.

East of L.A., and locations along the San Diego and Sacramento corridors which no longer have access to the line.

Aggregating winners and losers, we find a total loss for the apolitical planner (-0.0804%) compared to the actual CHSR. This result suggests that the proposed CHSR plan indeed dominated the I-5 routes, and that the I-5 was not discarded due to the type of political preferences we have estimated.

7 Conclusion

We study 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 the spatial distribution of votes in favor or against this project across California's census tracts. The empirical analysis reveals that voters did respond to the expected real-income impacts of the CHSR. This result supports the model as a predictor of economic impacts of the CHSR, and it shows that economic voting is a significant driver of policy preferences over transport infrastructure.

We posit that the observed CHSR design represents the optimal choice of a planner, whose preferences we estimate using moment inequalities. The estimation compares the impact of the design chosen by the planner to that of unchosen designs, which reveals the planner preferences. We find strong planner's references for winning votes, which means that alternative designs that were not chosen would have increased aggregate welfare and reduced votes. A counterfactual optimal design by an apolitical planner increases the proximity of stations towards main metropolitan areas, which tend to strongly support the project in any case. As a result, the gross benefits from the project double compared to the actual design that incorporates political preferences. We conclude that attaining popular approval was an important driver of transport network design.

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.

Alder, S. and I. Kondo (2020): "Political Distortions and Infrastructure Networks in China: A Quantitative Spatial Equilibrium Analysis," .

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, 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/.*

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.

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 (1994): "Protection for Sale," *The American Economic Review*, 84, 833–850.

——— (2021): "Identity politics and trade policy," *The Review of Economic Studies*, 88, 1101–1126.

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.

Jeon, J. and M. Rysman (2024): "International Trade in Data: Investment and Usage of the Subsea Internet Cable Network," .

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.

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.

McFadden, D. et al. (1973): "Conditional logit analysis of qualitative choice behavior," .

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.

Redding, S. J. and E. A. Rossi-Hansberg (2017): "Quantitative spatial economics," *Annual Review of Economics*, 9.

Salz, T. (2022): "Intermediation and competition in search markets: An empirical case study," *Journal of Political Economy*, 130, 310–345.

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. (2019): "Commuting, labor, and housing market effects of mass transportation: Welfare and identification," *The Review of Economics and Statistics*, 1–99.

——— (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.

——— (2010b): "2006-2010 5-Year American Community Survey Commuting Flows," Tech. rep., Census Transportation Planning Products, http://data5.ctpp.transportation.org/ctpp1216/Browse/browsetables.aspx.

US DOT (2005): "Final Program Environmental Impact Report/Environmental Impact Statement (EIR/EIS) for the proposed California High-Speed Train System," *https://hsr.ca.gov/programs/environmental-planning/program-eir-eis-documents-for-the-statewide-high-speed-rail-system-tier-1/final-program-environmental-impact-report-environmental-impact-statement-eir-eis-for-the-proposed-california-high-speed-train-system-2005/*.

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.

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 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 University of California at Berkeley 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.[1] 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 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

[1] 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.

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.[2]

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.10) using its definition as the sum of labor income and locally owned land rents. We construct 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.37), using Zillow's ZTRAX data . We construct the disposable income $I(i, j_C, m_C, N)$ defined in (A.9) 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 $\mu_{H,Y} = 0.20$,

[2] 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.

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.28 and A.29. 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 $\rho_A^{\text{spillover}} = 0.361$, $\rho_B^{\text{spillover}} = 0.759$.

B 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 correspond to the centroids of our analysis. To this network we add a rail network with 192 train stations obtained from California's Department of Transportation for 2013, which is the closest available year to 2008.[3] 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)).[4] Finally, 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.[5] The 2008 CHSR map includes 24 stations and two potential stations (Irvine and Tulare). We exclude these two planned 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 transport networks includes the road network expanded with tract centroids, rail stations, airports, and CHSR stations. Similar to centroids, we create artificial edges that connect the rail stations, airports, and CHSR stations to their closest node on the original road network.

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 of 5 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. We closely match the Google times

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

for car trips, whereas we tend to somewhat under-estimate travel times for long trips via public bus.

The fastest route via car is computed on the calibrated road network (all fastest routes are computed using the fast marching method). The fastest route via public transit is defined as the fastest between traveling entirely via public bus or via a combination of car and rail for each origin-destination pair. The fastest route via rail assumes that travelers use the train stations nearest to the origin and destination tracts.[6] To construct rail times we use station-to-station rail times available in the websites of rail systems in California (ACE Rail, Amtrak, BART, CalTrain, Coaster, and Metrolink), centroid-to-rail times computed via car, 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 station 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 the alternative I5 design of the CHSR, the traveling speed is set in a way that mimics the proposed plan, as shown in Figure A.8.

Travel Costs by Mode 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[7] and fuel efficiency of 21 miles per gallon[8]. 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 data from each county's website when necessary (average bus fare across all counties equals \$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 \$2.9 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.[9] 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,

[6] We exclude origin-destination pairs where driving times to the rail station are greater than 2 hours or where the distance between stations is greater than between the original tracts.

[7] Source: 2008 Los Angeles Almanac, available at http://www.laalmanac.com/energy/en12.php.

[8] Source: US Department of Energy, available at https://www.fueleconomy.gov/feg/byclass/Midsize_Cars2008.shtml/.

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

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 same ratio of variable to fixed costs as a function of distance estimated for the Amtrak Capitol Corridor.

C Implementation of the Gravity Equations

C.1 Commuting

We discuss here in detail how we estimate the parameters of the gravity equation for commuting in 11. First, we allow for three possible travel modes: $\mathcal{M}_C = \{\text{walking or biking, car, public transport}\}$. Second, we assume the labor income term entering $I(i,j,m)$ 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 all other 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 measurement error in wages as well as wage shocks unexpected to workers when making their commuting decisions. We take this approach because we do not observe $w^{data}(i,j)$ for every (i,j) pair.

Third, we consider as feasible commuting choices any pair of origin and destination census tracts and transport mode such that both census tracts are in CA and the travel time is either less than 4 hours (when using either car or public transport) or less than 2 hours (when biking or walking). 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 or walking. 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. First, 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

$$\mathbb{E}\left[\left(\frac{\lambda_C^{obs}(i,j,m)}{\sum_{j'}\lambda_C^{obs}(i,j',m)} - \frac{\lambda_C(i,j,m)}{\sum_{j'}\lambda_C(i,j',m)}\right)X(i,j,m)\right] = 0, \tag{A.1}$$

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.

Second, 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 unknown mode-specific coefficients $\Psi_C(m)$:

$$D_C(i,m)^{-\theta_C} \equiv \exp\left(\Psi_C(m) X_C(i)\right).$$

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

$$\mathbb{E}\left[\left(\sum_j \lambda_C^{\text{obs}}(i,j,m) - \sum_j \lambda_C(i,j,m)\right) X_C(i)\right] = 0. \tag{A.2}$$

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

C.2 Business and Leisure

We describe the details 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-model unobserved effects $Z_k(i,j)$ and $D_k(i,m)$. 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:

$$\left(\frac{Z_k(i,j)}{D_k(i,m)}\right)^{\mu_k\theta_k} \equiv \exp\left(\gamma_k(m) + \Psi_k X_k(j)\right), \tag{A.3}$$

for $k = L, B$, where $\gamma_k(m)$ is a mode and purpose-specific parameter, 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; i.e., 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 C.1. In the first step, we identify θ_k and β_k through the following moment condition:

$$\mathbb{E}\left[\left(\lambda_k^{obs}\left(i,j \mid m\right)-\lambda_k\left(i,j \mid m\right)\right)X_k\left(i,j,m\right)\right]=0, \tag{A.4}$$

for $k=L,B$, where $X_k(i,j,m)=\left(X_k(j),\ln\tau\left(i,j,m\right),\ln p_k(i,j,m)\right)'$, and $\lambda_k^{obs}\left(i,j|m\right)$ and $\lambda_k\left(i,j|m\right)$ denote the observed and model-implied shares of trips of type k from location i to j conditional on the transport mode m.

As discussed in Section 4.1, we only observe data for leisure and business trips whose origin and destination tracts are at least 50 miles away; we account for this data selection and build sample analogues of the moment conditions in (A.4) by taking an average across all pairs of origin i and destination j that are at least 50 miles apart. Consequently, for each origin tract i, the denominator of $\lambda_k\left(i,j|m\right)$ sums only over this restricted choice set. 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.4) 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 model-specific shifter $\gamma_k\left(m\right)$ to equal zero for $m=$ airplane, we identify $\gamma_k\left(m\right)$ for $m=$ private vehicle and $m=$ public transport using the following two m-specific moment conditions

$$\mathbb{E}\left[\sum_j \tilde{\lambda}_k^{\text{obs}}\left(i,j,m\right)-\sum_j \tilde{\lambda}_k\left(i,j,m\right)\right]=0. \tag{A.5}$$

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

D Tables and Figures

Table A.1: Commuting Equation Estimates, Second Step

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Public Transport								
Constant	1.73^a	3.70^a	2.39^a	1.85^a	0.46^a	0.98	-5.32^a	-8.22^a
	(0.05)	(0.15)	(0.31)	(0.12)	(0.10)	(1.55)	(0.34)	(1.05)
Sh. Car Owners		-2.31^a						-0.47^b
		(0.18)						(0.21)
Sh. Under 30			-1.41^b					-2.91^a
			(0.64)					(0.32)
Sh. College-educated				-0.26				-1.46^a
				(0.25)				(0.23)
Sh. Nonwhite					2.96^a			1.46^a
					(0.24)			(0.18)
Log Median Inc.						0.07		0.59^a
						(0.14)		(0.09)
Log Pop. Density							0.95^a	0.70^a
							(0.05)	(0.03)
Private Vehicle								
Constant	2.64^a	-0.71^a	3.64^a	3.14^a	2.55^a	0.41	1.24^a	-7.98^a
	(0.04)	(0.13)	(0.27)	(0.09)	(0.07)	(1.17)	(0.23)	(0.84)
Sh. Car Owners		4.00^a						4.76^a
		(0.15)						(0.19)
Sh. Under 30			-2.20^a					-1.59^a
			(0.56)					(0.25)
Sh. College-educated				-1.17^a				-1.07^a
				(0.19)				(0.19)
Sh. Nonwhite					0.21			0.20
					(0.20)			(0.14)
Log Median Inc.						0.20^c		0.43^a
						(0.11)		(0.08)
Log Pop. Density							0.19^a	0.41^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 $\hat{\theta}_C = 2.97$ and $\hat{\rho}_C = 0.75$. We normalize the parameter vector $\Psi_D(m)$ for m = walking or biking, and, given this normalization, we estimate the parameter vectors corresponding to m = private vehicle and m = public transport.

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. In the "Only Gross Benefits" cases, we compute economic impacts assuming every cost (capital and maintenance through ticket prices) equal to zero.

Table A.3: Distributional Impacts from the CHSR

	$\hat{W}(i)$			
	p10	p50	p90	Agg
2008 Business Plan	-0.11%	0.05%	0.46%	0.16%
2008 Business Plan, with GE	-0.05%	0.13%	0.58%	0.24%
2022 Business Plan	-0.34%	-0.26%	-0.06%	-0.21%
2022 Business Plan, with GE	-0.34%	-0.26%	-0.06%	-0.17%

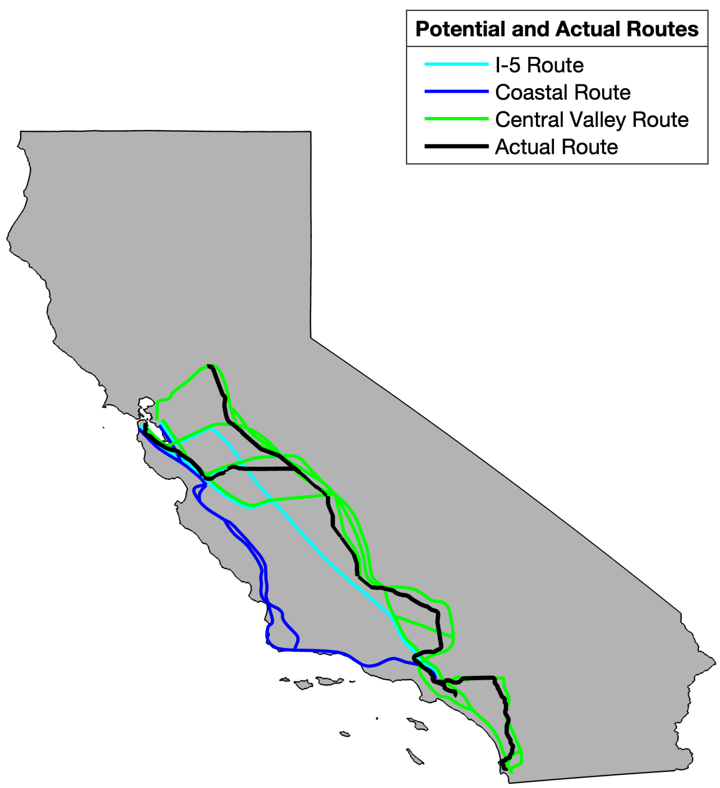
Note: The table reports moments from the per-year gain $\hat{W}(i)$ defined in 6. The aggregate in the last column is defined as the working-age population weighted average of $\hat{W}(i)$ across census tracts.

Table A.4: Gain by Type of Travel

	2008 Business Plan		2022 Business Plan	
		+GE		+GE
Commute	0.15%	0.18%	-0.20%	-0.08%
Leisure	0.01%	0.01%	-0.01%	-0.00%
Business	-	0.09%	-	-0.08%
Total	0.16%	0.24%	-0.21%	-0.17%

Note: This table splits the aggregate welfare change in (6) into the commuting component $\hat{\Omega}_C(i)$ in (8) and the leisure component $\hat{\Omega}_L(i)$ in (9). By construction of (7), these effects are log-additive in the case without GE effects. In the GE cases, the decomposition is not exact as elements are not log-additive. The gains coming from commuters are computed as coming from their time and cost savings from HSR, holding income constant, plus spillovers in amenities for residents. The gains coming from business travelers are computed as the welfare gains coming from endogenous wage changes, due both to time and cost saving in business traveling from the CHSR and productivity spillovers.

Figure A.1: Potential CHSR Routes (1996)



Note: this figure shows our 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. Each route includes multiple branches that could be used within a route. The route in black is the actual planned CHSR as of the 2008 vote.

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

Model:	No GE		GE		Pessimistic - No GE		Pessimistic - GE	
Inst. Var.:	Random Station	Random Path	Random Station	Random Path	Random Station	Random Path	Random Station	Random Path
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$log(\hat{W}_{19})$	20.90^a	22.05^a	18.64^a	18.14^a	43.32^a	44.62^a	37.47^a	35.76^a
	(1.76)	(1.83)	(1.58)	(1.61)	(3.60)	(3.66)	(3.15)	(3.15)
Log-odds Dem. Sh.	0.39^a	0.39^a	0.39^a	0.39^a	0.39^a	0.39^a	0.39^a	0.39^a
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Environ.: Prop. 10	2.40^a	2.40^a	2.39^a	2.40^a	2.40^a	2.39^a	2.39^a	2.39^a
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Transp.: Prop. 1b	0.81^a	0.80^a	0.84^a	0.84^a	0.81^a	0.81^a	0.85^a	0.85^a
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Sh. non-White	-0.18^a	-0.18^a	-0.17^a	-0.17^a	-0.18^a	-0.18^a	-0.17^a	-0.17^a
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Sh. College	0.74^a	0.73^a	0.74^a	0.74^a	0.73^a	0.73^a	0.74^a	0.74^a
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Sh. Under 30	0.17^a	0.17^a	0.17^a	0.17^a	0.18^a	0.18^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	-0.00	-0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
F-stat	672	466	719	516	683	488	743	543
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	Random Path	Random Station	Random Path	Random Station	Random Path	Random Station	Random Path
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$log(\hat{W}_{19})$	20.90^a	22.05^a	27.81^a	27.98^a	24.09^a	24.74^a	20.31^a	19.96^a
	(1.76)	(1.83)	(2.47)	(2.76)	(4.02)	(3.79)	(2.46)	(2.71)
Log-odds Dem. Sh.	0.39^a	0.39^a	0.41^a	0.41^a	0.42^a	0.43^a	0.41^a	0.41^a
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Environ.: Prop. 10	2.40^a	2.40^a	2.33^a	2.33^a	2.19^a	2.19^a	2.50^a	2.50^a
	(0.05)	(0.05)	(0.06)	(0.06)	(0.17)	(0.17)	(0.06)	(0.06)
Transp.: Prop. 1b	0.81^a	0.80^a	0.86^a	0.86^a	0.89^a	0.89^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	-0.18^a	-0.18^a	-0.25^a	-0.25^a	-0.23^a	-0.23^a	-0.30^a	-0.30^a
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Sh. College	0.74^a	0.73^a	0.73^a	0.73^a	0.70^a	0.70^a	0.73^a	0.73^a
	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.03)	(0.02)	(0.02)
Sh. Under 30	0.17^a	0.17^a	0.22^a	0.22^a	0.22	0.22	0.25^a	0.25^a
	(0.03)	(0.03)	(0.03)	(0.03)	(0.15)	(0.15)	(0.03)	(0.03)
Log. Dist. Station	-0.00	-0.00	0.01	0.01	-0.01	-0.01	-0.01^b	-0.01^b
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
F-stat	672	466	806	721	398	518	335	326
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.

E 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 ($s = N$).

E.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:

$$v_\omega(i, j_C, m_C, j_L, m_L, s) = \max_{C, H_C, T_L} B(i, s) \frac{C^{1-\mu_L-\mu_H} H_C^{\mu_H}}{d_C(i, j_C, m_C, s)} \left(\frac{q_L(i, j_L) B(j_L, s)}{d_L(i, j_L, m_L, s)} T_L \right)^{\mu_L} \varepsilon_\omega^C(j_C, m_C) \varepsilon_\omega^L(j_L, m_L). \tag{A.6}$$

subject to the budget constraint:

$$C + r(i, s) H_C + p_L(i, j_L, m_L, s) T_L = I(i, j_C, m_C, s) \tag{A.7}$$

Expression A.6 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.6), $\varepsilon_\omega^C(j_C, m_C)$ and $\varepsilon_\omega^L(j_L, m_L)$, 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 ($k = L$):

$$d_k(i, j, m, s) = D_k(i, m)\, \tau_k(i, j, m, s)^\rho \text{ for } k = C, L, \tag{A.8}$$

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.7. 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.7) is the disposable income, defined as gross income $y(i, j_C, s)$ net of taxes $t(s)$ and annual commuting costs:

$$I(i, j_C, m_C, s) \equiv (1 - t(s))\, y(i, j_C, s) - p_C(i, j_C, m_C, s)\, T_C. \tag{A.9}$$

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:

$$y(i, j_C, s) \equiv e(i)\, w(j_C, s) + \eta(i)\, r(i, s). \tag{A.10}$$

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.10) 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.

E.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.6) gives indirect utility conditional on the origin, destinations, travel modes, and idiosyncratic preference shocks for destination:

$$v_\omega\left(i, j_C, m_C, j_L, m_L, s\right)=\frac{B(i, s)}{r(i, s)^{\mu_H(i)}}\left(\frac{I\left(i, j_C, m_C, s\right)}{d_C\left(i, j_C, m_C, s\right)} \varepsilon_\omega^C\left(j_C, m_C\right)\right)\left(\frac{q_L\left(i, j_L\right) B\left(j_L, s\right)}{p_L\left(i, j_L, m_L, s\right) d_L\left(i, j_L, m_L, s\right)}\right)^{\mu_L(i)} \varepsilon_\omega^L\left(j_L, m_L\right) \tag{A.11}$$

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 for travel purpose $k = L, C$ is $\mathcal{M}_k$. We assume the idiosyncratic preference shocks for commuting and leisure travel $\varepsilon_\omega^C\left(j_C, m_C\right)$ and $\varepsilon_\omega^L\left(j_L, m_L\right)$ to be IID Type-I extreme value distributed:

$$\Pr\left(\varepsilon_\omega^k\left(j_k, m_k\right)<x\right)=e^{-e^{-\theta_k x}} \text{ for } k=C, L, \tag{A.12}$$

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:

$$V(i, s)=\mathbb{E}_\omega\left[\max_{\left(j_C, j_L, m_C, m_L\right) \in \mathcal{J}^2 \times \mathcal{M}_C \times \mathcal{M}_L} v_\omega\left(i, j_C, m_C, j_L, m_L, s\right)\right]. \tag{A.13}$$

Using standard properties of the extreme-value distributions for the shocks ε_ω^C and ε_ω^L, we can write this expression as:[10]

$$V(i, s)=\frac{B(i, s)}{r(i, s)^{\mu_H(i)}} \frac{\Omega_C(i, s)}{\Omega_L(i, s)^{\mu_L(i)}} \tag{A.14}$$

where $\Omega_C(i, s)$ captures average income net of commuting costs of residents of i,

$$\Omega_C(i, s)=\left(\sum_{j \in \mathcal{J}} \sum_{m \in \mathcal{M}_C}\left(\frac{I(i, j, m, s)}{d_C(i, j, m, s)}\right)^{\theta_C}\right)^{\frac{1}{\theta_C}}, \tag{A.15}$$

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:

$$\Omega_L(i, s) \equiv\left(\sum_{j \in \mathcal{J}} \sum_{m \in \mathcal{M}_C}\left(\frac{p_L(i, j, m, s) d_L(i, j, m, s)}{q_L(i, j) B(j, s)}\right)^{-\mu_L \theta_L}\right)^{-\frac{1}{\mu_L \theta_L}}. \tag{A.16}$$

E.3 Tradeable Sector Firms

In the tradeable sector, we assume a measure 1 of firms in each tract. 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

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

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

output according to the Cobb-Douglas production function:

$$Y_\omega(j, H_Y, N_Y, R_B, j_B, m_B, s) = A(j,s)\, H_Y^{\mu_{H_Y}}\, N_Y^{1-\mu_{H_Y}-\mu_B} \left(\frac{q_B(j, j_B)\, A(j_B, s)}{d_B(j, j_B, m_B, s)} R_B\right)^{\mu_B} \varepsilon_\omega^B(j_B, m_B). \tag{A.17}$$

Business trips are productivity-enhancing, capturing for example that they promote new supplier or customer relationships. 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 time traveled (captured by $d_B(j, j_B, m_B, s)$). 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:

$$\Pr\left(\varepsilon_\omega^B(j_B, m_B) < x\right) = e^{-e^{-\theta_B x}}, \tag{A.18}$$

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:

$$\Pi = \max_{H_Y, N_Y} \mathbb{E}\left[\max_{(T_B, j_B, m) \in \left(\mathbb{R}^+ \times \mathcal{J} \times \mathcal{M}_B\right)} Y_\omega(j, H_Y, N_Y, R_B, j_B, m, s) - p_B(j, j_B, m, s)\, T_B\right] - w(j,s)\, N_Y - r(j,s)\, H_Y, \tag{A.19}$$

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 over realization of idiosyncratic business shocks using standard properties of the extreme value distribution defined in (A.18).

After these steps, we obtain a closed-form solution for the expected output net of business costs (the term within brackets in (A.19)). Specifically, the firm problem over floor space and labor can be re-written as follows:

$$\Pi = \max_{H_Y, N_Y} \left(\Omega_B(j,s)\, H_Y^{\mu_{H_Y}}\, N_Y^{1-\mu_{H_Y}-\mu_B}\right)^{\frac{1}{1-\mu_B}} - r(j,s)\, H_Y - w(j,s)\, N_Y. \tag{A.20}$$

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,

$$\Omega_B(j,s) \equiv \kappa_B A(j,s) \left(\sum_{j_B \in \mathcal{J}} \sum_{m \in \mathcal{M}_B} \left(\frac{q_B(j, j_B)\, A(j_B, s)}{p_B(j, j_B, m_B, s)\, d_B(j, j_B, m_B, s)}\right)^{\theta_B \mu_B}\right)^{\frac{1}{\theta_B}}, \tag{A.21}$$

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

E.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.13) gives the fraction of residents from i that commute to j using transport mode m,

$$\lambda_C\left(i, j_C, m_C, s\right)=\frac{\left(\frac{I(i, j_C, m_C, s)}{d_C(i, j_C, m_C, s)}\right)^{\theta_C}}{\sum_{j \in \mathcal{J}} \sum_{m \in \mathcal{M}_C}\left(\frac{I(i, j, m, s)}{d_C(i, j, m, s)}\right)^{\theta_C}}, \tag{A.22}$$

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

$$\lambda_L\left(i, j_L, m_L, s\right)=\frac{\left(\frac{q_L(i, j_L) B(j_L, s)}{p_L(i, j_L, m_L, s) d_L(i, j_L, m_L, s)}\right)^{\mu_L \theta_L}}{\sum_{j \in \mathcal{J}} \sum_{m \in \mathcal{M}_L}\left(\frac{q_L(i, j) B(j, s)}{p_L(i, j, m, s) d_L(i, j, m, s)}\right)^{\mu_L \theta_L}}. \tag{A.23}$$

Similarly, from the solution to the firm's problem in A.19 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.23):

$$\lambda_B\left(i, j_B, m_B, s\right)=\frac{\left(\frac{q_B(i, j_B) A(j_B, s)}{p_B(i, j_B, m_B, s) d_B(i, j_B, m_B, s)}\right)^{\mu_B \theta_B}}{\sum_{j \in \mathcal{J}} \sum_{m \in \mathcal{M}_B}\left(\frac{q_B(i, j) A(j, s)}{p_B(i, j, m, s) d_B(i, j, m, s)}\right)^{\mu_B \theta_B}}. \tag{A.24}$$

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.6), 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:

$$T_L\left(i, j_L, m_L, s\right)=\lambda_L\left(j, j_L, m_B, s\right) \mu_L(i) \frac{N_R(i) \bar{I}(i)}{p_L\left(i, j_L, m_L, s\right)}, \tag{A.25}$$

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

$$\bar{I}(i) \equiv \sum_{j \in \mathcal{J}} \sum_{m \in \mathcal{M}_C} \lambda_C(i, j, m, s) I(i, j, m, s). \tag{A.26}$$

Similarly, from the solution for T_B from (A.19), the total number of business trips from i to j_B through mode m_B is:

$$T_B\left(i, j_B, m_B, s\right)=\lambda_B\left(i, j_B, m_B, s\right) \frac{\mu_B}{1-\mu_B} \frac{Y(i, s)}{p_B\left(i, j_B, m_B, s\right)}. \tag{A.27}$$

Using (A.25) and (A.27) we obtain the gravity equation (12) that we bring to the data.

E.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:

$$A(j,s) = Z_A(j) \left(\sum_{k \in \mathcal{J}} e^{-\rho_A^{\text{spillover}} \tau^{min}(j,k,s)} \frac{\tilde{N}_Y(k,s)}{H(k)} \right)^{\gamma_A^{\text{spillover}}} \tag{A.28}$$

$$B(i,s) = Z_B(i) \left(\sum_{k \in \mathcal{J}} e^{-\rho_B^{\text{spillover}} \tau^{min}(j,k,s)} \frac{\tilde{N}_Y(k,s)}{H(k)} \right)^{\gamma_B^{\text{spillover}}} \tag{A.29}$$

where

$$\tilde{N}_Y(j,s) = \sum_i \lambda_C(i,j,s) N_R(i) \tag{A.30}$$

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.

E.6 Route Choice and 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 $k = C, L, B$ are the time and the cost corresponding to the actual 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 between a given origin and destination for a given mode and assign a monetary cost to 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 by air (for business and leisure travelers). That is, the CHSR is used within mode m for each pair (i,j) if the CHSR-based time-cost pair (τ, p) leads to a higher indirect utility than the pre-CHSR one. Given the indirect utility (A.11), this trade-off is summarized by:

$$(\tau_C(\cdot,Y), p_C(\cdot,Y)) = \arg \max_{(\tau,p) \in \{(\tau_C(\cdot,N), p_C(\cdot,N)), (\tau^{CHSR}(\cdot), p^{CHSR}(\cdot))\}} \frac{(1-t)\, y(i,j,Y) - pT_C}{\tau^\rho}. \tag{A.31}$$

Similarly, as implied by the indirect utility (A.11) and by the definition of business productivity (A.21), a leisure ($k = L$) or business ($k = B$) traveler chooses a post-CHSR time and cost pair by solving:

$$\left(\tau_k\left(\cdot, Y\right), p_k\left(\cdot, Y\right)\right) = \arg \min_{(\tau,p)\in\left\{\left(\tau_C(\cdot,N), p_C(\cdot,N)\right), \left(\tau^{CHSR}(\cdot), p^{CHSR}(\cdot)\right)\right\}} p\tau^{\rho} \tag{A.32}$$

for $k = L, B$. When $s = Y$ (i.e., when the CHSR is available), the times and costs $\tau_k\left(i, j, m, Y\right)$ and $p_k\left(i, j, m, Y\right)$ on a given origin-destination may be different for commuters and for leisure or business travel, because the monetary costs from using the CHSR enter asymmetrically in indirect utility depending on the travel purpose.

E.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, $N\left(i, s\right)$:

$$N_Y\left(i, s\right) = \underbrace{\sum_{j\in\mathcal{J}} \lambda_C\left(j, i, s\right) e\left(j\right) N_R\left(j\right)}_{\equiv N(i,s)} \tag{A.33}$$

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

$$\lambda_C\left(j, i, s\right) \equiv \sum_{m\in\mathcal{M}_C} \lambda_C\left(j, i, m, s\right). \tag{A.34}$$

Next, using the solution for consumer demand for floor space from (A.6) and for firm's demand for floor space and labor from (A.19), the equilibrium in the housing markets is:

$$N_R\left(i\right) \frac{\mu_H\left(i\right) \overline{I}\left(i\right)}{r\left(i, s\right)} + N_Y\left(i, s\right) \frac{w\left(i, s\right)}{r\left(i, s\right)} \frac{\mu_{H_Y}\left(i\right)}{1 - \mu_{H_Y}\left(i\right) - \mu_B\left(i\right)} = H\left(i\right), \tag{A.35}$$

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 $H\left(i\right)$ is the supply of floor space in i. Finally, since tradeable firms operate subject to constant returns, the zero-profit conditions resulting from (A.19) dictates:

$$w\left(j, s\right)^{1-\mu_B-\mu_{H_Y}} r\left(j, s\right)^{\mu_{H_Y}} = \kappa_\Omega \Omega_B\left(j\right) \tag{A.36}$$

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

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

i) the land market clearing condition (A.35) holds for all tracts;

ii) the labor market clearing condition (A.33) holds for all tracts i; and

iii) the zero-profit condition (A.36) holds for all tracts j.

Note that the system of equations defined by (A.35)-(A.36) include as unknowns the endogenous productivity term $\Omega\left(j\right)$, the agglomeration and amenity spillover functions $A\left(j, s\right)$ and $B\left(i, s\right)$, and the average income $\overline{I}\left(i\right)$. Using (A.21), (A.28), (A.29), and (A.26), 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.

E.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.31) and (A.32). 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 E.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.35) holds in the counterfactual equilibrium, which implies:

$$\hat{r}(i) = \frac{H_C(i,N)}{H(i)}\hat{\bar{I}}(i) + \left(1 - \frac{H_C(i,N)}{H(i)}\right)\hat{w}(i)\hat{N}_Y(i), \tag{A.37}$$

where $H_C \equiv N_R(i)\frac{\mu_H(i)\bar{I}(i)}{r(i,s)}$ is the aggregate housing demand in i and $\hat{\bar{I}}(i)$ is the change in average income of residents of i defined in (A.26),

$$\hat{\bar{I}}(i) = \sum_{j\in\mathcal{J}}\sum_{m\in\mathcal{M}_C}\frac{\lambda_C(i,j,m,N)}{\bar{I}(i,N)}\hat{\lambda}^C(i,j,m)\hat{I}(i,j,m), \tag{A.38}$$

where the change in disposable income next of taxes and commuting costs for commuters from i to

j using mode m is

$$\hat{I}(i,j,m) = \frac{y(i,j,N)}{I(i,j,m,N)}(1-t)\,\hat{y}(i,j) - \frac{T_C p_C(i,j,m,N)}{I(i,j,m,N)}\hat{p}_C(i,j,m,N), \tag{A.39}$$

the change in pre-tax income is

$$\hat{y}(i,j) = \left(1 - \frac{e(i)\,w(j,N)}{y(i,j,N)}\right)\hat{\bar{I}}(i) + \frac{e(i)\,w(j,N)}{y(i,j_C,N)}\hat{w}(j), \tag{A.40}$$

and, from (A.22), $\hat{\lambda}_C(i,j,m)$ is given by:

$$\hat{\lambda}_C(i,j_C,m_C) = \frac{\left(\frac{\hat{I}(i,j_C,m_C)}{\hat{d_C}(i,j_C,m_C)}\right)^{\theta_C}}{\sum_{j\in\mathcal{J}}\sum_{m\in\mathcal{M}_C}\lambda_C(i,j,m,N)\left(\frac{\hat{I}(i,j,m)}{\hat{d_C}(i,j,m)}\right)^{\theta_C}} \tag{A.41}$$

ii) the labor market clears, i.e. (A.33) holds in the counterfactual equilibrium, which implies

$$\hat{N}_Y(i) = \sum_{j\in\mathcal{J}}\left(\frac{\lambda_C(j,i,N)\,e(j)\,N_R(j)}{N(i,s)}\right)\hat{\lambda}_C(j,i) \tag{A.42}$$

where, in the supply side in the right-hand side of (A.42), $\hat{\lambda}_C(j,i)$ is given by

$$\hat{\lambda}_C(j,i) = \sum_{m\in\mathcal{M}_C}\left(\frac{\lambda_C(j,i,m,N)}{\lambda_C(j,i,N)}\right)\hat{\lambda}_C(j,i,m). \tag{A.43}$$

iii) the zero-profit condition (A.36) holds in a counterfactual scenario, i.e.

$$\hat{w}(j)^{1-\mu_B-\mu_{H_Y}}\,\hat{r}(j)^{\mu_{H_Y}} = \hat{\Omega}_B(j), \tag{A.44}$$

where, from (A.21), $\hat{\Omega}_B(i)$ is given by (10) in the text.

From (A.28), the agglomeration component of TFP in $\hat{\Omega}_B$ changes according to:

$$\hat{A}(j) = \left(\sum_{k\in\mathcal{J}}\frac{\tilde{N}_Y(k,N)/H(k)}{\sum_{k'\in\mathcal{J}}e^{-\rho_A^{\text{spillover}}\tau^{min}(j,k',N)}\tilde{N}_Y(k',N)/H(k')}e^{-\rho_A^{\text{spillover}}\tau^{min}(j,k,B)}\hat{\tilde{N}}_Y(k)\right)^{\gamma_A^{\text{spillover}}} \tag{A.45}$$

where from (A.30) the change in the number of workers employed in j is:

$$\hat{\tilde{N}}_Y(j) = \sum_i\left(\frac{\lambda_C(i,j,N)\,N_R(i)}{\tilde{N}_Y(j,s)}\right)\hat{\lambda_C}(i,j). \tag{A.46}$$

for $\hat{\lambda_C}(i,j)$ defined in (A.43).

E.9 Welfare Changes

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

$$\hat{V}(i) = \left(\frac{\hat{B}(i)}{\hat{r}(i)^{\mu_H}}\right)\hat{\Omega}_C(i)\,\hat{\Omega}_L(i). \tag{A.47}$$

The commuting component $\hat{\Omega}_C(i)$ changes according to (8), with $\hat{I}(i,j,m)$ in that expression given by (A.39). The leisure changes according to

$$\hat{\Omega}_L(i) \equiv \left(\sum_{j\in\mathcal{J}}\sum_{m\in\mathcal{M}_L}\lambda_L(i,j,m)\left(\frac{\hat{B}(j)}{\hat{p}_L(i,j,m)\,\hat{\tau}(i,j,m)^{\rho}}\right)^{\mu_L\theta_L}\right)^{\frac{1}{\theta_L}}. \tag{A.48}$$

In these expressions, endogenous amenities component $\hat{B}(i)$ satisfies a similar equation to $\hat{A}(j)$ in (A.45):

$$\hat{B}(j) = \left(\sum_{k\in\mathcal{J}}\frac{\tilde{N_Y}(k,N)/H(k)}{\sum_{k'\in\mathcal{J}}e^{-\rho_B^{\text{spillover}}\tau^{min}(j,k',N)}\tilde{N_Y}(k',N)/H(k')}e^{-\rho_B^{\text{spillover}}\tau^{min}(j,k,B)}\hat{\tilde{N}}_Y(k)\right)^{\gamma_B^{\text{spillover}}}. \tag{A.49}$$

Conditions (7) and (9) in the text follow from (A.47) and (A.48) when we set $\hat{B}(i) = \hat{r}(i) = 1$.

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

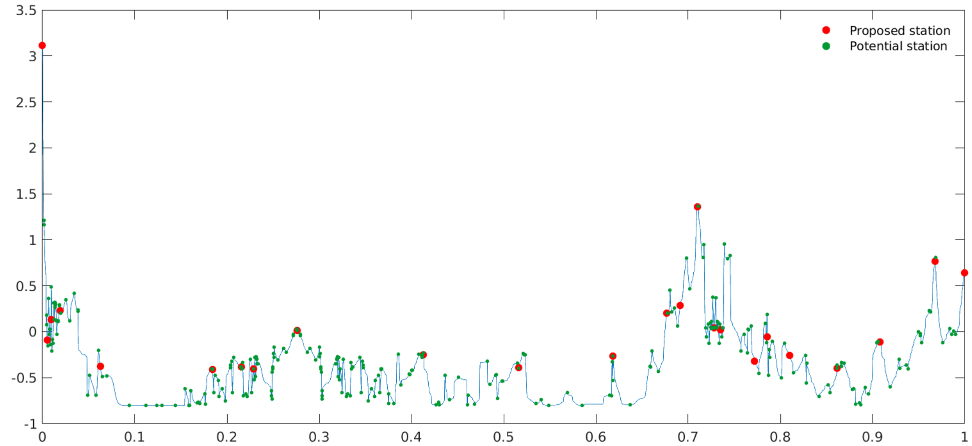
$$\sum_{i=1}^{J}\left[b_0 + \sum_{k=1}^{K} b_k Z_k(i) + \lambda_V \nabla v(i)\right] N_R(i)\,\Delta\ln\hat{W}(i,\mathbf{d}^n) - \epsilon(\mathbf{d}^n) \le 0, \tag{A.50}$$

where $\nabla v(i) \equiv \theta_V v(i)(1-v(i))$ 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 $\mathcal{L}(Z)$ 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.2 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 $\mathcal{N}(Z)$ by moving each station $s = 1\ldots 24$ 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-

Figure A.2: Potential Station Locations based on Population Density



Notes: Smoothed population density $\hat{n}(i) = \sum_{i'=1}^{J} n(i') e^{-\rho dist(i,i')}$ where $\rho = 100$ and $\text{dist}(i,i')$ 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.

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).

Moment Conditions To lighten notation, we rewrite the welfare inequality (A.50) 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 $\gamma = (b_0, b_1, \ldots, b_K, \lambda)$ 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 $\mathcal{N}_e = \left\{n \in \mathcal{N} \mid X_e^{2008}(n) \geq 0\right\}$ 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 $E\left[\epsilon\left(\mathbf{d}^n\right)|\mathcal{I}^{2008}\right] = 0$. 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 a highly non-concave optimization problem due to the presence of a sigmoidal 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:

1. *Perturbations.* We first use the information contained in the precomputed perturbation set $\mathcal{N}$. For each station s, we identify which perturbation $n \in \mathcal{N}$ yields 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.

2. *Simulated annealing.* To escape potential local optima and increase our chances of finding the global optimum, we use a type of simulated annealing method. A station is selected at random and is moved randomly among the potential locations 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).
3. *Continuous optimizer.* The final step of the optimization attempts to refine the local optimum obtained by the previous two steps by using a continuous optimizer (e.g. simplex or interior-point algorithms). We parametrize the CHSR outline with a cubic spline, allowing us to evaluate welfare over a continuous set of station locations.

Our extensive efforts to globally explore the set of station locations do not guarantee identification of the global optimum. Nonetheless, we verify that our final result is a promising candidate by showing 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 (see Figure A.5).

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

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

Table A.7: Planner's Preferences Estimates (2022 Business Plan)

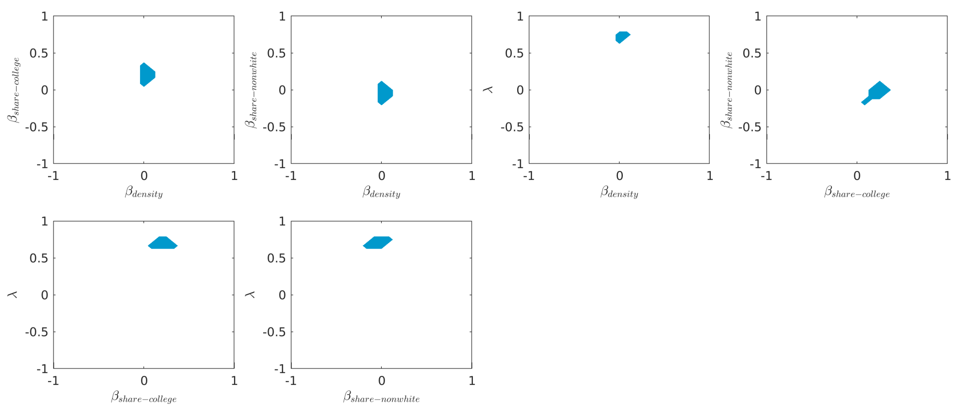
Observable	Pareto weight parameters b and λ				
	(1)	(2)	(3)	(4)	(5)
Density		[0.00, 1.13]			[0.00, 0.14]
Share college			[0.50, 0.50]		[0.36, 0.50]
Share non-white				[-0.33, 0.50]	[-0.14, 0.00]
Votes	[0.51, 2.19]	[0.84, 4.20]	[1.44, 1.44]	[0.70, 3.60]	[1.39, 1.81]
Constant	[1.00, 1.00]	[0.98, 1.00]	[0.99, 0.99]	[0.98, 1.01]	[0.99, 1.00]

Table A.8: Apolitical planner, Welfare Gains across Covariates

ΔReal Income by Quartile	1st	2nd	3rd	4th
Density	0.01%	0.05%	0.14%	0.43%
Share college	0.18%	0.16%	0.12%	0.17%
Share non-white	0.04%	0.07%	0.18%	0.32%
Voting elasticity	0.37%	0.14%	0.09%	0.05%

Notes: 2008 Business Plan estimates of real income gains of moving from proposed plan to apolitical optimum. Real income gains are population-weighted and provided in basis points. The rows [min max] show the support across locations.

Figure A.3: 95% Confidence Set for 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.

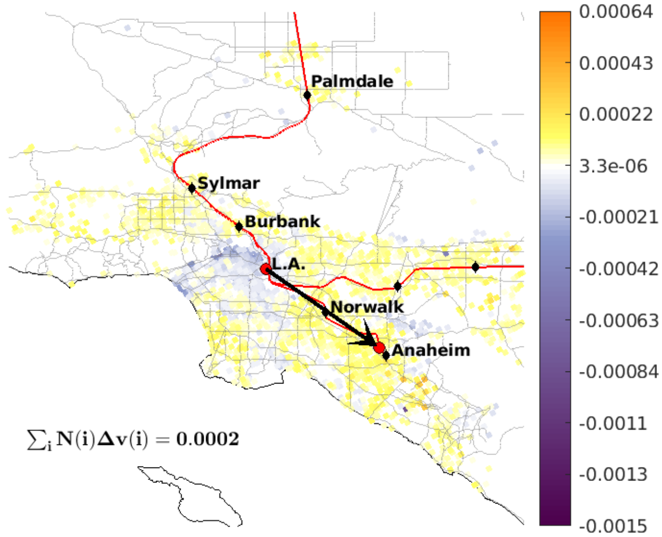
F.3 Additional Tables and Figures for the Planning Problem

Table A.9: Stations in the Apolitical Planner vs. Proposed 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	1.12	0	0.0025	-0.043	0.064	0.0041
Stockton	-0.36	0	-0.024	-0.031	-0.035	0.02
San Francisco	5.62	0.0311	0.38	0.15	0.13	-0.57
Modesto	-0.72	0	-0.03	-0.056	-0.0055	0.025
SFO	-4.47	0.0093	0.13	0.053	0.1	-0.21
Palo Alto	-9.81	0.0083	0.16	0.04	0.066	-0.24
San Jose	-15.99	0.0324	0.29	0.22	0.2	-0.41
Merced	-0.23	0	-0.009	-0.02	0.0083	0.0023
Gilroy	1.29	0.0003	0.071	0.026	0.013	-0.11
Fresno	-5.82	0.0001	-0.0067	-0.032	0.0057	0.0006
Bakersfield	-3.29	0	-0.023	-0.025	-0.014	0.032
Palmdale	55.31	0.0092	0.25	0.018	0.073	-0.11
Sylmar	9.94	0.0167	0.27	-0.13	0.2	-0.046
Burbank	15.67	0.0151	0.41	-0.062	0.2	-0.23
Ontario	-53.45	0.0016	0.21	-0.055	0.12	-0.12
Los Angeles	3.03	0.0152	0.44	-0.082	0.26	-0.22
City of Industry	-16.63	0.0069	0.11	-0.081	0.12	-0.0095
Riverside	-5.9	0.0074	-0.04	-0.14	0.057	0.088
Norwalk	6.33	0.0089	0.2	-0.083	0.19	0.0035
Anaheim	-0.01	0	-0.16	0.06	-0.13	-0.0032
Murrieta	-44.37	0.0027	-0.024	-0.1	0.057	0.031
Escondido	22.9	0.0043	0.0097	0.1	0.01	0.031
University City	2.23	0.0015	0.039	0.063	-0.05	0.013
San Diego	-4.67	0.0039	0.047	0.1	-0.07	0.033

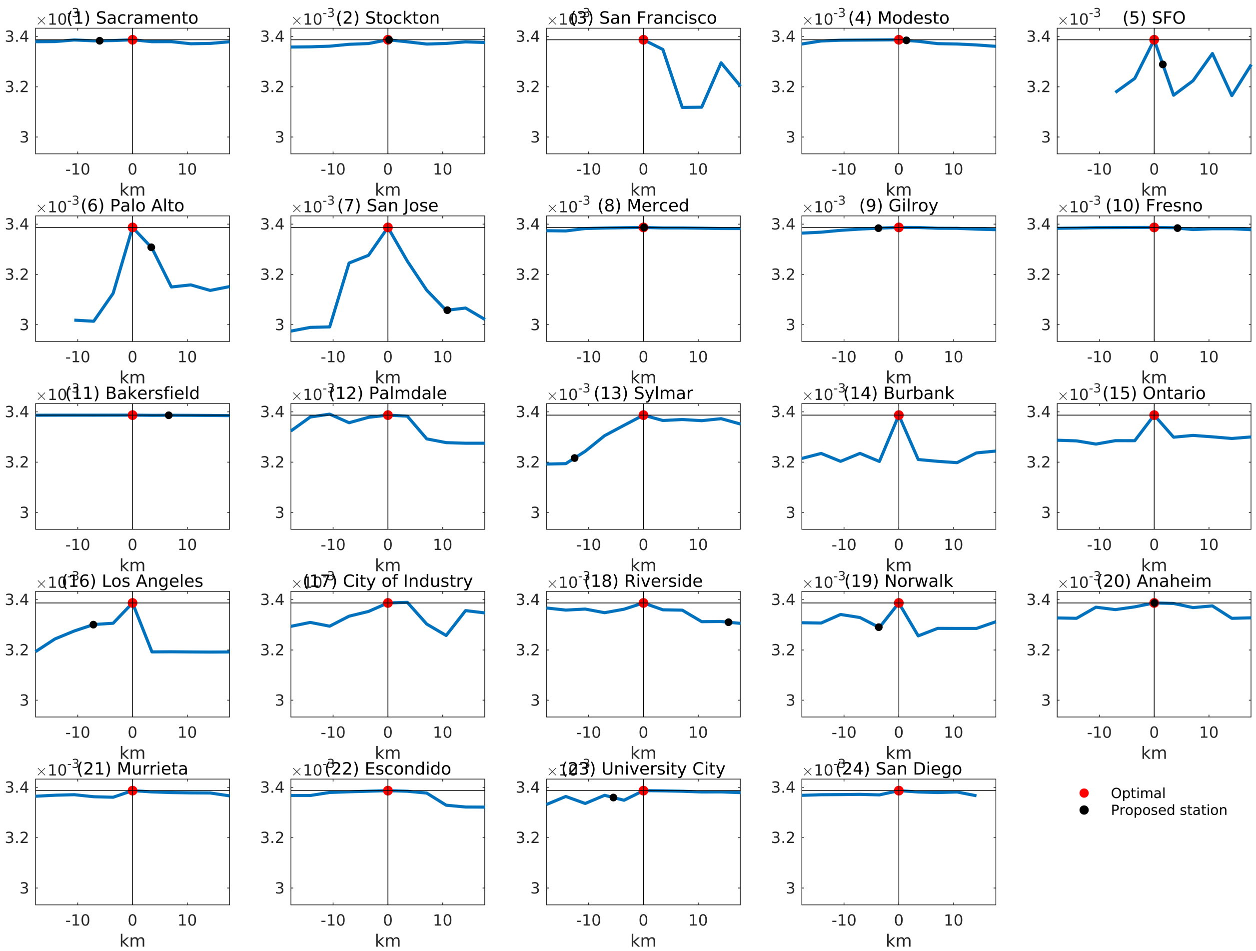
Notes: 2008 Business Plan estimates. Column (1) reports the distance in km between the optimal location and the proposed station. A positive number indicates a movement towards San Diego, negative towards San Francisco. Column (2) reports the change in aggregate welfare in basis points (%) given the 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.4: Distribution of $\lambda_U(i) + \lambda_V \partial v(i) / \partial \ln \hat{W}(i)$



Notes: 2008 Business Plan estimates. Histogram of the effective planning weights $\lambda_U(i) + \lambda_V \theta_V v(i)(1 - v(i))$ received by locations in specification (5) in Table 5. Parameters are normalized so that the population-weighted mean of the Pareto weights is 1.

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.

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

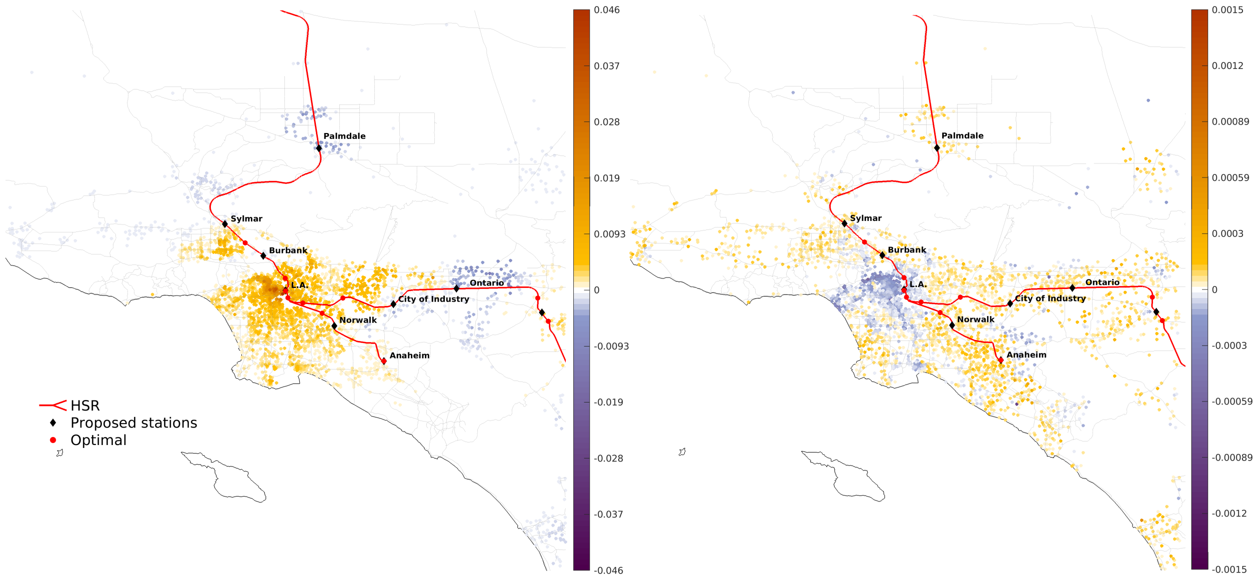


Figure A.7: Utilitarian Planner, Welfare vs. Share College in San Diego

(a) Welfare Change (b) Share College

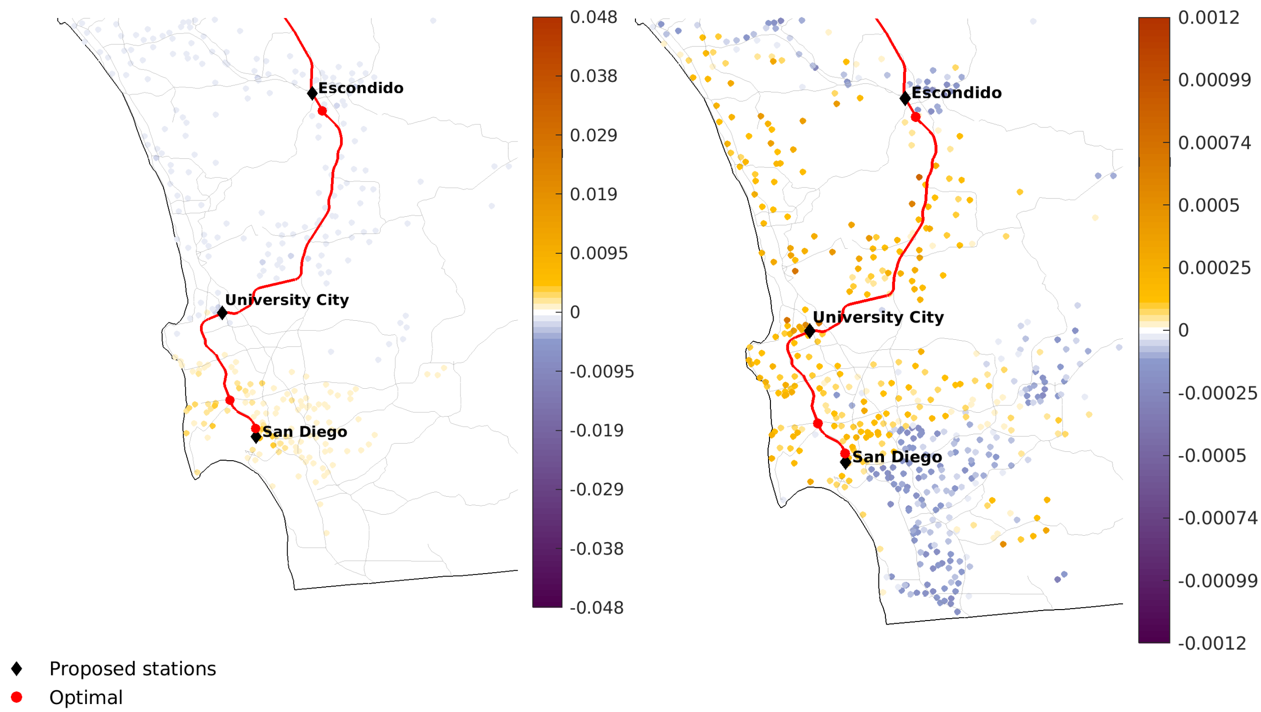


Figure A.8: Speeds over Proposed Plan and the I-5 Route Alternative

(a) Proposed HSR

(b) I5