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

September 22, 2023

Abstract

How do political preferences shape transportation policy? We study this question in the context of California's High-Speed Rail (CHSR). Combining geographic data on votes in a referendum on the CHSR with a model of its expected economic benefits, we estimate the weight of economic and non-economic considerations in voters' preferences. Then, comparing the proposed distribution of CHSR stations with alternative placements, we use a revealed-preference approach to estimate policymakers' preferences for redistribution and popular approval. While voters did respond to expected real-income benefits, non-economic factors were a more important driver of the spatial distribution of voters' preferences for the CHSR. The voter-approved CHSR would have led to modest income gains, but proposals with net income losses also would have been approved due to political preferences. For the planner, we identify strong preferences for popular approval. A politically-blind planner would have placed the stations closer to dense metro areas in California.

*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, SED, Timbergen, Toronto, UCLA ITS, UTDT, and West Coast Spatial Workshop for helpful feedback. We thank the NBER and the US DOT for financial support. Cecile Gaubert acknowledges support from NSF CAREER grant # 1941917.

1 Introduction

Transport infrastructure projects are among the largest investments at all levels of governments. Their size often makes them a focal point for public debate, with the public and policy-makers taking strong stances on their desirability. These heterogeneous views are partly driven by economic impacts, as transportation networks shape the income distribution across regions or demographic groups. However, they also depend on additional considerations. For instance, people may factor in preferences for collective goods, the environment, or loyalty to the political party championing a project. For simplicity, we refer to these non-economic considerations as the "political" component of preferences. Likewise, the policymaking process leading up to transport investments may account for redistribution to certain groups and for political motivations such as seeking public support.

These considerations matter because they shape the welfare implications of transport infrastructure, as well as the incentives of the policy-makers who implement them. How important is real income in people's preferences for transportation infrastructure, compared to the political component of preferences? How do distributional as well as electoral considerations shape the design of transport infrastructure?

We make progress on these questions by studying California's High-Speed Rail (CHSR), an electric high-speed rail designed to connect urban centers in California that is among the most expensive transport projects in U.S. history. For our purposes, a key feature is that Californians were asked to vote on this project: in the 2008 general election, a proposition on the ballot (Proposition 1A) asked Californians whether they approved 10 billion USD in state-issued bonds to initiate funding for the CHSR.[1] Detailed spatial data on Proposition 1A voting outcomes and on other simultaneous ballots across California's census tracts give us direct insights into Californians' policy preferences.

Using these voting data and an economic model, we estimate the relative weight of expected real income and political considerations in voters' preferences for the CHSR. Then, we estimate the policymakers' preferences for redistribution and for popular approval by comparing the expected impact of the initially proposed CHSR with feasible counterfactual designs with alternative station placements along the CHSR line. By revealed-preference, counterfactual station distributions were suboptimal from the policymaker's perspective; we use this insight to recover bounds on the planner's preferences. Having estimated the preferences of both the voters and the planner, we undertake counterfactuals that demonstrate the relevance of the political component of preferences in shaping the welfare effects and the design of the CHSR.

A challenge throughout the analysis is that voters' expected economic returns from the CHSR are not observed. To measure these expected returns, we develop and estimate a quantitative model of the spatial impact of the CHSR. Because we do not know how voters forecasted the

[1]The proposition passed with 52.6% of votes in favor. The CHSR is currently under construction on a delayed and costlier construction schedule than originally proposed. We discuss the background on the project in more detail in Section 2, and the way in which we introduce potential mismatches between beliefs and the proposed network in the subsequent sections.

economic impact of the CHSR, we consider a range of specifications with varying degrees of voter sophistication (e.g., whether general-equilibrium effects are considered) and optimism on the CHSR expected completion date and costs. In the spirit of the quantitative spatial frameworks of Ahlfeldt et al. (2015) and Monte et al. (2018), the model accounts for the CHSR potential impact on workers' commuting time. In addition, we include features that are specific to high-speed passenger travel. First, the CHSR connections can be used for more infrequent long-distance business or leisure travel. We observe both types of trips in the data, and we incorporate them in the model alongside commuting. Second, as the CHSR would compete with other travel modes, we incorporate a transport-mode decision (car, air, and public transit) for each route and type of travel (commuting, leisure, and business), accounting for both time and monetary costs in this choice. The model also accounts for indirect effects of transportation infrastructure beyond time or cost savings, such as its equilibrium impacts on land prices, wages, and local development as captured by urban agglomeration spillovers on productivity and amenities.

Using observed travel behavior for each purpose (commuting, leisure, and business) and mode (car, air, and public transit), we estimate key model elasticities that capture travelers' disutility from travel costs, as well as flexible tract- and purpose-specific preferences for each travel mode. Across a range of scenarios varying in voter sophistication and pessimism on costs, the quantified model reveals net aggregate real-income impacts that range from -0.18% to 0.65%. These aggregate impacts mask considerable heterogeneity; e.g., in our baseline, the most favored 10% of tracts experience real-income gains between 0.6% and 4.7% per year, while 2.7% of all tracts lose. Tracts that are closer to stations, with lower car usage, or with longer average commute times gain more from the CHSR. Thus, areas closer to larger urban centers such as San Francisco or Los Angeles benefit more, and more sparsely populated areas like Central-Valley locations stand out with lower gains.

We then estimate the relative importance of real-income effects and political considerations for voters' preferences for the CHSR. We measure the economic gains using the model above, and use a range of covariates to proxy for voters' political beliefs. Our estimation deals with several potential identification concerns, such as the correlation between the economic benefits of the CHSR and unobserved components of political preferences, and misspecification of our model of voters' forecasts. In particular, we instrument for the economic impact of the CHSR using alternate CHSR routes with random placement of stations. We also conduct the estimation using a range of alternative model specifications and sample restrictions.

Regardless of the model assumptions and identification strategies, we consistently find that voters are responsive to the expected economic impact of the CHSR. However, swaying votes is costly: the extra real-income gain from the CHSR needed to convince an extra one percent of the population of a census tract to vote in favor is between 0.2% and 0.4% at the median tract, which is large compared to the typical income effects. Moreover, the political component drives a much larger fraction of the spatial variation in preferences (and votes) than the real-income component, with a standard deviation of between 1.0 to 2.4 percentage points across tracts, which is about

5 to 6 times larger than the standard deviation of the real-income component. Furthermore, we estimate that Proposition 1A would have been approved in the absence of a preference for real income gains. Hence, due to political preferences, there is a range of costs such that the CHSR would lower aggregate real income and still be approved: CHSR proposals predicted to reduce aggregate real-income by as much as 0.3% to 0.5% uniformly across California would have still won the vote.

In a second step, we estimate the preferences of a social planner. We assume that the observed distribution of stations along the CHSR line maximizes the planner's objective function. The first component of this objective function is a weighted sum of aggregate real income across tracts, with tract-specific Pareto weights parameterized as function of their economic and demographic composition. In addition the planner may be politically minded, in the sense that it attaches a weight to the total number of votes in favor of the project.[2] To estimate these preferences, we rely on counterfactual station placements. These deviations imply a distribution of real income changes across demographics and votes which determine bounds on the planner's Pareto weights on each demographic and preference for the aggregate vote. We implement a moment-inequality method based on Andrews and Soares (2010) to compute confidence sets for these parameters, and then solve for optimal designs under alternative planner preferences.

We identify strong planner preferences for votes, as well as some preference for dense areas and for areas with a larger share of college graduates. These estimates imply a large heterogeneity in tract-specific weights, and the planner is thus far from the utilitarian benchmark. The optimal designs of the CHSR for an *apolitical* planner (i.e., without political concerns) substantially differs 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 (i.e., they strongly support the project already). Hence, the absence of electoral motives leads to redistribution towards high density locations with low expected political gains, where the marginal voter is far from the mode of political support. The utilitarian planner solution looks broadly similar to the apolitical planner, although it reallocates some stations away from high college-share areas.

This paper contributes to several literatures. First, it builds on the literature studying the real income effects of transportation infrastructure, including Donaldson (2018), Faber (2014), and Donaldson and Hornbeck (2016) among others (see Redding and Turner (2015) for a review). Methodologically, our approach develops a quantitative spatial model in the style of Redding and Rossi-Hansberg (2017) to capture the specificity of the type of long-distance travel facilitated by the CHSR. In doing so we incorporate standard model choices in the tradition of McFadden et al. (1973), and incorporate insights from studies that have estimated the impact of high-speed rails on economic outcomes, such as Zheng and Kahn (2013), Bernard et al. (2019), Borusyak and Hull

[2]This planning problem captures, in a reduced-form way, a complex political process whereby the design of infrastructure responds to both real income considerations and the overall popularity of a project. While voting over infrastructure project is unusual, policymakers' preferences for popular approval are not. These preferences may exist whenever authorities are mindful of their popularity and keep an eye on the number of constituents approving of an infrastructure project.

(2020), Dong et al. (2020), and Hayakawa et al. (2021).[3] Our paper departs from the traditional focus on market-access driven economic gains by estimating how other considerations matter in people's preferences for transportation infrastructure.

Recent studies, such as Tsivanidis (2019), Gupta et al. (2022), Severen (2021) and Tyndall (2021), among others, study spatial and distributional effects of improvements to commuter rails. Our analysis studies (expected) distributional effects stemming from both the real-income considerations and a political component of utility. Voting data allows us to estimate the strength of the component of utility associated with political values (such as party-affiliation) in real-income equivalent terms. We find that the distributional impacts from political components are several times larger than the standard impacts through market access. Recent papers, including Alder (2019), Fajgelbaum and Schaal (2020), and Allen and Arkolakis (2022), study optimal design of road infrastructure in quantitative spatial models with real-income-maximizing social planners. We estimate the preferences of a planner who has both distributional and political concerns, and study optimal placement of stations in this context. Unlike in these frameworks, our optimization problem is inherently subject to non-convexities due to the mix of substitutabilities and complementarities among possible station locations.

Second, the paper contributes to the literature in economics and political science on whether policy preferences reflect individual economic considerations. This type of question has been asked in an international trade context, in public finance, and in environmental economics.[4] Holian et al. (2013) and Holian and Kahn (2015) correlate votes in several ballot initiatives, including the CHSR Proposition 1A, with determinants of political ideology and proxies for economic geography. Kendall and Matsusaka (2021) estimate voters' preference weights over the common good versus the private returns of California ballot propositions. We use a structural model of economic gains from the CHSR, and develop an identification strategy that separates the expected economic returns from the political component of preferences.

Third, we contribute to the literature studying the political determinants of transportation infrastructure. Empirically, Knight (2005), Burgess et al. (2015) or Alder and Kondo (2020) find evidence of political distortions in transportation investments. A theoretical literature studies mechanisms that drive the political economy of transportation investments (Brueckner and Selod (2006), De Borger and Proost (2012), and Glaeser and Ponzetto (2018)). We estimate the Pareto weights of a hypothetical social planner and its preferences for popular approval by using an observed policy choice (in our case, the CHSR station placement). In doing so, we follow the work of Goldberg and Maggi (1999) in international trade, and the tradition of Bourguignon and Spadaro

[3]As in Dingel and Tintelnot (2020), our setting is sparse as it includes bilateral commuting and long-distance travel among about 64 million census tracts. After imposing bounds on total travel time, we treat the resulting sparsity as a result of sampling noise.

[4]Recent papers relying on voting data to infer preferences for international trade policy include Hicks et al. (2014), Becker et al. (2017), and Van Patten and Méndez (2022). Alesina and Giuliano (2011) review research that uses survey data to measure preferences for private vs. public value in income taxation. Kahn and Matsusaka (1997), Deacon and Schläpfer (2010), and Wu and Cutter (2011) use ballots to draw inference about demand for collective goods following Deacon and Shapiro (1975).

(2012), Jacobs et al. (2017) and Hendren (2020) in public finance. The logic of all these works is to estimate social planning weights that rationalize the observed tax schedule as optimal in order to learn about the weight of distributional or electoral concerns in the policy-making process. We apply this logic to a spatial policy, the CHSR, with complex spatial linkages.

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

2.1 Institutional Background

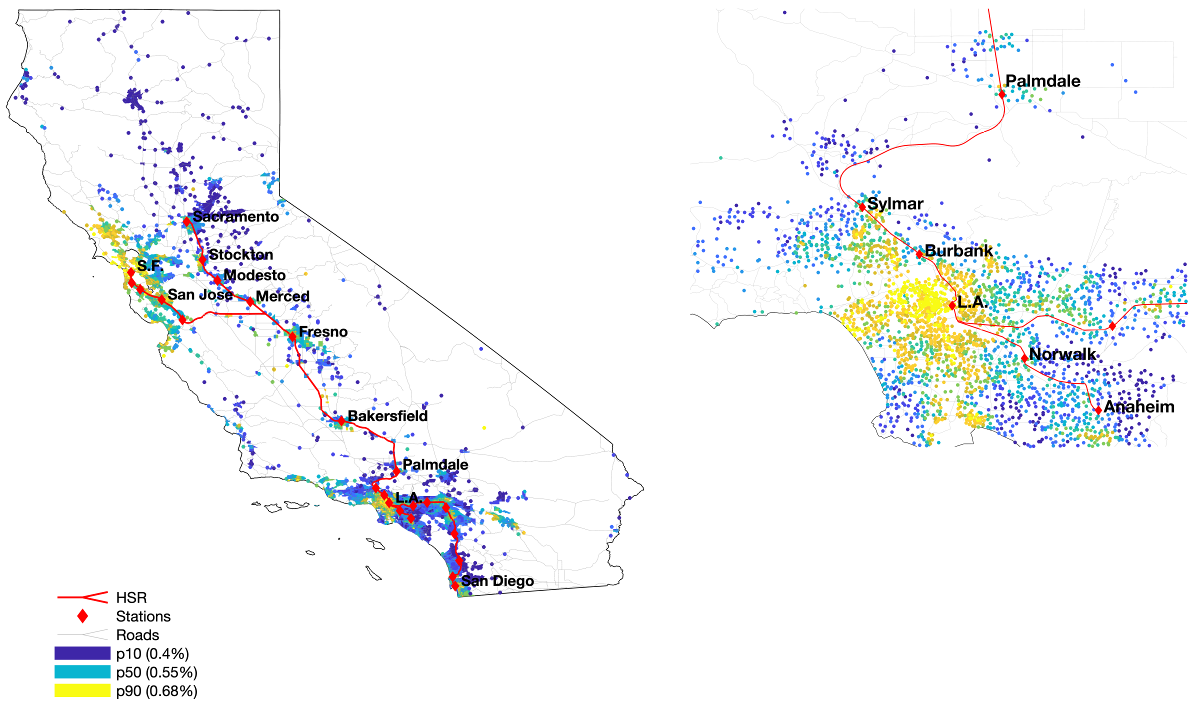
In 1996, the California state legislature established the California High Speed Rail Authority (CAHSRA), which was tasked with exploring the creation of a high-speed rail network to connect northern California, including San Francisco and Sacramento, with southern California, including Los Angeles and San Diego. In August 2008, the California 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 fund the CHSR.[5] The CHSR would be 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 2020. Phase 1 of construction would focus on the Los Angeles-San Francisco corridor, while Phase 2 would extend the network to Sacramento in the north and San Diego in the south. Proposition 1A was one of twelve measures on the ballot in the November 2008 general election.

Proposition 1A was ultimately approved by 52.6% of votes. Participation was high, equal to 94% of voters casting a vote in the 2008 presidential election. Figure 1 shows support for the CHSR, as measured by 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

[5] The proposed funding sources for the CHSR included funds from bond issuances, private investors, the federal government, and other local sources. Specifically, the nearly $10 billion in Proposition 1A bonds were expected to be coupled with up to $3 billion in local funding, $16 billion in federal funding, and a $7 billion in public-private partnership funding. In principle, this funding would finance a complete San Francisco to Anaheim system for a total of $33.6 billion in 2008 dollars.

Figure 1: CHSR Route and Proposition 1A Votes

(a) California (b) LA County



San Francisco bay area (e.g., Marin and Sonoma) show clusters of strong support, while in Los Angeles support is more concentrated in central areas.

The construction of the CHSR has suffered many delays and hurdles since its construction began in 2015. It 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 of the project have risen from around \$50 billion in 2008 (California High Speed Rail Authority, 2008) for the entire system to up to \$105 billion in 2023 for Phase I (California High Speed Rail Authority, 2022). Our analysis considers several scenarios to account for the uncertainty voters had in 2008 about both the future CHSR costs and the completion schedule.

2.2 Potential Time Savings and Costs across Routes

We use data on trips undertaken by Californians to assess the potential usage, time saving, and cost savings of the CHSR. 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. The first three rows of Appendix Table A.1 show that car was 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 and 88% of business trips.

Against this backdrop, the information presented to voters in 2008 suggested sizable time and cost effects of the CHSR. As a preliminary look of these potential gains, Table A.1 also reports summary statistics by travel mode for the percentage of all routes (origin-destination census tract pairs) that would be quicker or cheaper were the CHSR to be used on that route in combination with that mode. We consider two scenarios for costs: assuming ticket prices corresponding to the original proposal voted in 2008 (California High Speed Rail Authority, 2008), and a "pessimistic" scenario with twice-as-expensive ticket prices, which roughly corresponds to most recent updates (California High Speed Rail Authority, 2022). We defer to Appendix D for details on the construction of the travel times and costs by transport mode.

The table shows that the majority of origin-destination census tracts would be considerably faster or cheaper according to the 2008 business plan. Of course, this calculation includes routes that are seldom used. In the next sections we account for the likelihood that travelers from each origin will travel through the CHSR, and we assign a monetary equivalent value to the corresponding time savings.

3 Framework

This section gives an overview of the theoretical framework.[6] We start by presenting voters' preferences, and a model of the voting decision as driven by political and economic considerations. We then zoom in on these economic considerations, by developing a quantitative model of the impact of the CHSR on the distribution of real incomes across census tracts in California. In section 6, we discuss how we model the planner's problem when choosing the design of the CHSR.

3.1 Utility and Voting

Preferences Consider a resident ω of a location i. Her utility $u_\omega(s)$ captures both her economic and political preferences, which depend on whether an infrastructure project, in our case the California High-Speed Rail, is approved to be built ($s = Y$ for Yes) or not ($s = N$ for No):

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

The first component, $\mathbb{E}\left[\ln W(i,s) \mid \mathcal{I}_i\right]$ measures the average expected real income of residents of i when the project has status s. It incorporates economic forces through which an infrastructure project can impact real income, and is discussed in detail in Section 3.2. The expectation on $\ln W(i,s)$ is taken over shocks that may affect the economy, including shocks over fundamental economic characteristics of different locations i and uncertainty about the transportation project itself, conditional on the information $\mathcal{I}_i$ of residents of i. The second component, $\ln a(i,s)$, captures the political component of preferences on average across residents of i. Finally, the shock $\varepsilon^u_\omega(s)$

[6] A complete self-contained exposition of the model is provided in Appendix B.

captures idiosyncratic (mean-zero) variation in utility across residents of tract i stemming from either economic or political considerations.

Given (1), we define the average *utility impact* of the transport project across residents of location i as $\Delta U(i) \equiv \mathbb{E}_{\omega}[u_{\omega}(Y) - u_{\omega}(N)]$ and get:

$$\Delta U(i) = \mathbb{E}\left[\ln \hat{W}(i) \mid \mathcal{I}_i\right] + \ln \hat{a}(i). \tag{2}$$

In this expression, $\hat{W}(i) \equiv \frac{W(i,Y)}{W(i,N)}$ measures real income differences between a world with and without the planned infrastructure project, while $\hat{a}(i)$ measures the net political preferences for the project. Both components affect which locations win or lose ($\Delta U(i) \gtrless 0$). Existing research on the distributional impacts of infrastructure projects typically aims to measure $\hat{W}(i)$. One of the goals in this paper is to measure the relative importance of $\hat{W}(i)$ and $\hat{a}(i)$ in shaping this distributional impact of a transport project across locations i.

In practice, neither the real-income or the political components are directly observed. To make progress, we impose more structure. First, we assume that residents form rational expectations, so that:

$$\mathbb{E}\left[\ln \hat{W}(i) \mid \mathcal{I}_i\right] = \ln \hat{W}(i) - \epsilon_W(i), \tag{3}$$

where $\ln \hat{W}(i)$ is the real income change that would be brought by the CHSR under the actual realization of shocks, and $\epsilon_W(i)$ is a mean-zero expectational error. In the next section, we proxy for $\hat{W}(i)$ using a range of economic models with varying degrees of sophistication in the forces that they incorporate.

Second, the political component $\hat{a}(i)$ is not directly observed but can be proxied for with location-specific variables $X_k(i)$ that bear a relationship to people's political preferences, such as people's party affiliation (we discuss these proxies later on). Formally, we assume that:

$$\ln \hat{a}(i) = \sum_{k=1}^{K} \tilde{\beta}_k X_k(i) + \epsilon_a(i), \tag{4}$$

where $X_0(i) = 1$ so that $\tilde{\beta}_0$ captures a common political component of preferences, and where $\epsilon_a(i)$ captures unobserved determinants of political ideology. Given our goals, a central question is how to measure the rate $\tilde{\beta}_k$ at which a given proxy (people's political affiliation, say) can be meaningfully compared with the real-income impacts of a given project. Unlike in the case of $\ln \hat{W}(i)$, no family of models exists to translate proxies of values into a welfare metric that is comparable to real income. However, voting outcomes directly reveal this trade-off between political values and economic gains. Hence, to overcome this issue, we turn next to the voting model.

Voting Faced with a choice between Y and N, households vote "Yes" if and only if $u_{\omega}(Y) > u_{\omega}(N)$. Aggregating over individuals, given utility (1), our setup corresponds to a probabilistic voting model (Deacon and Shapiro, 1975), with the fraction of positive votes in location i defined as

$$v(i) = \Pr\left[u_\omega(Y) > u_\omega(N)\right]. \tag{5}$$

We assume that these idiosyncratic shocks are Type-I extreme-value distributed across residents, with shape parameter θ_V:

$$\Pr\left(\varepsilon_\omega^u(s) < x\right) = e^{-e^{-\theta_V x}}. \tag{6}$$

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

$$v(i) = \frac{e^{\theta_V\left(\mathbb{E}\left[\ln \hat{W}(i)|\mathcal{I}_i\right]+\ln \hat{a}(i)\right)}}{1+e^{\theta_V\left(\mathbb{E}\left[\ln \hat{W}(i)|\mathcal{I}_i\right]+\ln \hat{a}(i)\right)}}. \tag{7}$$

Combining this expression with (3) and (4), and re-writing the left-hand side as the log odds-ratio, we obtain the equation that we will bring to the data:

$$\ln\left(\frac{v(i)}{1-v(i)}\right) = \theta_V \ln \hat{W}(i) + \sum_{k=1}^{K} \beta_k X_k(i) + \epsilon(i), \tag{8}$$

where $\beta_k \equiv \theta_V \tilde{\beta}_k$ and where $\epsilon(i) \equiv -\epsilon_W(i) + \epsilon_a(i)$ includes the expectational error and the unobserved components of political ideology.[7]

Equation (8) shows the trade-off between real income and political preferences in determining votes. The parameter θ_V is the rate at which real income is translated into votes. Having identified this parameter, we can recover the $\tilde{\beta}_k$'s in (4) as $\tilde{\beta}_k = \beta_k/\theta_V$, thus revealing how a given proxy for political preferences translates into real-income equivalent terms. We can then construct $\ln \hat{a}(i)$ and compute how much it drives the heterogeneity in $\Delta U(i)$ in (2). The elasticity of votes to real income θ_V does not matter *per se* for the average utility impact of the project on a tract $\Delta U(i)$, but estimating this parameter is a crucial step to translate ideology proxies into a common real-income metric.

3.2 Economic Effects of the CHSR

We now turn to describing the model of the real income gains from the CHSR in tract i, $\ln \hat{W}(i)$.

Uncertainty and Delays The real income impacts of the CHSR are twofold. First, until the CHSR is built and operational, households expect to pay a tax t to fund the corresponding investment. Second, once the CHSR is operational, households still pay the tax but also benefit from the corresponding real income gains. The overall real income effect of the CHSR is the annualized flow combining these two streams.

At the time of voting, in 2008, the CHSR business plan stated a specific year in which the CHSR would be operational. To account for the fact that voters may have doubted that timing, we assume that they expected the CHSR to be operational no sooner than T years after the vote, and that they attributed a yearly completion probability equal to p in every year after that (until built

[7] While notation has not been introduced for it, this error term also captures that voters may use an economic model that diverges from the range of models that we use to proxy for $\hat{W}(i)$. This source of error, as well as the one we have already specified, bring about identification concerns that we tackle in the empirical section.

and operational). Voters may have also doubted the CHSR business plan's sunk and operational cost projections. We consider scenarios with varying degrees of optimism over p, T and the total cost of the CHSR as manifested in the tax rate t and ticket prices.

Assuming a yearly discount rate r, the annualized real-income effects of the CHSR, if approved, are a weighted average of $\ln(1-t)$ (the annual real-income change from paying a yearly tax t to finance to the CHSR) and $\ln \hat{V}(i)$ (the yearly tax-inclusive real-income change associated with the CHSR being operational for residents of location i):

$$\ln \hat{W}(i) = \left(1 - (1+r)^{-T} \frac{p}{r+p}\right) \ln(1-t) + (1+r)^{-T} \frac{p}{r+p} \ln \hat{V}(i). \tag{9}$$

Economic benefits from CHSR To evaluate the real-income change associated with the CHSR being operational, $\hat{V}(i)$, for residents of tract i, we follow the body of research that uses quantitative spatial models to estimate the distributional impacts of infrastructure. We tailor the model to the specifics of high-speed rail. One may worry that voters rely on simpler heuristics rather than on model predictions with fairly rich spatial interactions; therefore, we consider polar cases of such models in terms of complexity: a baseline where the CHSR only impacts time savings and cost of travel; and a full model with several equilibrium feedbacks. Throughout, we assume that $N_R(i)$ residents in location i vote on the expected returns to that location. The development of the CHSR impacts commuting and travel choices but not residential choice, although we could accommodate cases with residential mobility.[8]

The *raison d'être* of the CHSR is that it allows passengers to save time on their medium to long-distance travel. At the heart of the model, therefore, is the time potentially saved by workers on daily commutes (for workers who have especially long commutes) and on less frequent long-distance leisure trips. The CHSR also facilitates long distance business trips, making firms more productive as they connect with suppliers and customers (Bernard et al., 2019). We augment a commuting model *alla* Ahlfeldt et al. (2015) to incorporate these infrequent long-distance trips.

Our "baseline" model only captures direct effects of the CHSR. Specifically, in each census tract, residents spends income on tradeable goods, housing, commuting, and leisure trips. Each resident receives idiosyncratic preference shocks for commuting or traveling for leisure to different destinations. These activities entail a payoff (a salary when commuting or a utility flow when traveling for leisure) in exchange for a time and a monetary cost. Thus, residents make discrete choices of destinations for commuting and for leisure trips (for the latter, they also choose the number of trips taken). In this baseline, wages and the cost of housing remain constant and, as shown below, the real income gains from the CHSR amount to the net time- or cost-savings over

[8] Our framework accommodates an extension where workers can move between locations. Assume that every year workers can change location at some fixed rate and that, in this case, they draw an idiosyncratic preference shock for each possible location. Then, the net present value of the effect of the CHSR for someone living in i in 2008, $\ln \hat{W}(i)$, is still a linear function of $\ln \hat{V}(i)$, but the coefficient in front of $\ln \hat{V}(i)$ is smaller than in the baseline model (9), as voters are likely to move out and put less weight on their current tract. In this extension, the coefficient θ_V in front of $\ln \hat{W}(i)$ when running the regression (8) can be deducted from the one we obtain when running (8) in the baseline without having to rerun the regression, as they are scaled version of each other.

likely travel destinations, adjusted by the travelers' valuation of time and willingness to substitute across destinations.

The "full" model formally nests our "baseline" and captures, in addition, equilibrium impacts on wages, land rents, amenities, and productivity. Tradeable goods are produced with labor, land, and productivity-enhancing business trips. Hence, the CHSR indirectly impacts the productivity of firms and wages through its direct impact on the time and monetary cost of both commuting and business trips. In addition, firm productivity and wages are impacted by agglomeration spillovers that depend on the (endogenous) spatial distribution of worker density. Such spillovers from labor density also impact the amenity enjoyed by residents and leisure travelers. Finally, because producers compete with residents to use land, land rents respond to the development of the CHSR, capitalizing local productivity enhancements through the increased demand for productive space. However, only a share of the residents of each tract is homeowner and benefits from this capitalization (with remaining land being owned by absentee landowners). Together, changes in land rents and wages capture that the CHSR may lead to local economic development.[9]

Substitution across Modes Importantly, when deciding the destination, travelers for each purpose (commuting, business, or leisure) incorporate a choice of transport mode (car, public transit, air, or walking/biking), perceiving the different modes as imperfect substitutes. Residents of different locations vary in their preferences for different travel modes. In our "baseline", the CHSR is perceived by travelers as a perfect substitute to public-transit or air travel, potentially enhancing on those choices. In that case, when the CHSR becomes available, someone who was initially traveling by car may substitute into the CHSR but doing so implies a different choice of mode of travel, with an implied utility effect. We also show results for a case where the CHSR is perceived as a perfect substitute to traveling by car.

Key Measurement Equations We show here the key equations governing the real income impact of the CHSR. The summary equations below apply to the full model, which nests the baseline model with the appropriate parameter restrictions.

The annual real-income change for residents of tract i due to the CHSR is:[10]

$$\hat{V}(i) = \left(\frac{\hat{B}(i)}{\hat{r}(i)^{\mu_H}}\right)\left(\frac{\hat{W}_C(i)}{\hat{P}_L(i)^{\mu_L}}\right). \tag{10}$$

The first term in parenthesis includes in the numerator the change in expected income net of commuting costs and in the denominator the change in the price index for leisure trips ($\hat{P}_L(i)$), adjusted by the spending share of tract i's residents on leisure travel (μ_L). The second term in parenthesis includes in the numerator the change in residential amenities stemming from urban spillovers ($\hat{B}(i)$) and in the denominator the change in the cost of housing ($\hat{r}(i)$), adjusted by its spending share μ_H. In the baseline model, this second term in parenthesis equals one by assumption.

[9]The economic model does not include environmental effects from the CHSR. To the extent that they are valued by voters, they will be allocated to the political component of preferences in our analysis.

[10]In general, $\hat{X}$ denotes the ratio between value of variable X if the CHSR is constructed and its value if is not.

The change in income net of commuting cost, $\hat{W}_C(i)$, captures the expected time savings and change in travel costs of commuters, based on their ex ante commuting patterns, augmented by the flexibility to substitute across employment destinations and modes of travel once the CHSR is operational:

$$\hat{W}_C(i) \equiv \left(\sum_{j \in \mathcal{J}} \sum_{m \in \mathcal{M}_\mathcal{C}} \lambda_C(i,j,m) \left(\frac{\hat{I}(i,j,m)}{\hat{\tau}(i,j,m)^{\rho_C}} \right)^{\theta_C} \right)^{\frac{1}{\theta_C}}. \tag{11}$$

In this expression, $\hat{I}(i,j,m)$ is the change in disposable income conditional on commuting from one's residence i to one's workplace j using transport mode m from a set of transport modes that may be used commuting, $\mathcal{M}_\mathcal{C}$.[11] This term depends on the changes in the monetary cost of commuting from i to j by mode m as well as (in the full model only) on the changes in wage at destination and in land rent income at origin. Second, $\hat{\tau}(i,j,m)$ is the change in travel time from i to j using m, converted into a dollar-equivalent value by the elasticity ρ_C. Third, θ_C captures the extent to which residents substitute across commuting destinations travel modes when the relative appeal of destination or modes changes.[12] Finally, the weights $\lambda_C(i,j,m,N)$ on these changes are the fraction of tract-i residents that commute for work to j through mode m absent the CHSR.

The change in the leisure price index follows a similar functional form:

$$\hat{P}_L(i) \equiv \left(\sum_{j \in \mathcal{J}} \sum_{m \in \mathcal{M}_C} \lambda_L(i,j,m) \left(\frac{\hat{p}_L(i,j,m)\,\hat{\tau}(i,j,m)^{\rho_L}}{\hat{B}(j)} \right)^{-\mu_L \theta_L} \right)^{-\frac{1}{\mu_L \theta_L}}, \tag{12}$$

where now $\lambda_L(i,j,m)$ 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, $\hat{B}(j)$ are changes in amenities in the leisure destination j, and (ρ_L, θ_L) capture respectively the value of time when traveling for leisure and the substitution across destinations for leisure.[13]

Our baseline model with only direct effects assumes away endogenous changes in amenities or land rents in (10) and (12) (i.e., $\hat{B}(i) = \hat{r}(i) = 1$) and no wage changes from the reallocations triggered by the CHSR, implying that changes in disposable income in (11) stem from the tax to finance the CHSR and the change in the monetary cost of commuting.[14] As a result, in our

[11] This set of modes is travel-purpose specific because (1) air is used for long-distance travel but not for commuting in practice and (2) walking or biking amounts for a non-negligible fraction of commuting trips but is negligible for long-distance travel.

[12] Formally, as detailed in the appendix, this parameter is the inverse of the dispersion in idiosyncratic preference draws across residents of i about where to commute and how to commute there.

[13] When comparing (11) and (12), one can notice two asymmetries: in (11), the monetary cost enters as a negative additive shifter to disposable income while in (12) it enters multiplicatively; and the latter is shaped by the intensity of leisure spending $\mu_L(i)$. As detailed in the appendix, these differences arise due to the non-homothetic nature of spending on commuting : assuming a fix number of commuting days in the year, travelers spend a fix amount of money in commuting and the remaining income is divided into consumption, housing, and leisure trips. In contrast, leisure 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.

[14] More generally, the change in disposable income is defined as the change in after-tax income net of commuting costs: $\hat{I}(i,j,m) = (1 + \chi^{pre}(i,j,m))(1-t)\,\hat{y}(i,j) - \chi^{pre}(i,j,m)\,\hat{p}_C(i,j,m)$, where $\hat{y}(i,j)$ is the change in pre-tax income, $\chi^{pre}(i,j,m)$ is the share of commuting costs in disposable income for someone traveling from i to j through mode m before the CHSR is operational, and $\hat{p}_C(i,j_C,m_C)$ is the change in the monetary cost of this commuting

baseline model where only direct effects are included, we do not need any additional equations beyond (11) and (12) to measure real-income effects. In contrast, the full model with indirect effects generates additional changes in amenities $\hat{B}(i)$, land rents $\hat{r}(i)$, and disposable income $\hat{I}(i, j, m)$. In particular, wages enter as part of disposable income and change with the CHSR in the full model, as the rollout of the CHSR impacts firm productivity. Specifically, firm productivity is impacted for two reasons: first, because the CHSR allows firm to send workers on productivity-enhancing business trips with greater ease; second, because the CHSR changes the distribution of workers over space, impacting agglomeration spillovers. The full system of equations describing these forces is presented in Appendix B.8.

4 Measuring the Economic Impacts of the CHSR

In this section, we estimate the parameters of the economic model of the CHSR. We then present the real income effects of the California High-Speed rail predicted by the quantified model.

4.1 Data

We start with a brief overview of the main data used in our analysis and refer the reader to Appendix C 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 and on leisure and business trips from the California Household Travel Survey (CAHTS) 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, the Bureau of Transportation Statistics, and various rail operators. We construct times and costs of traveling by CHSR using information from the 2008 CHSR business plan. We use voting data for Proposition 1A as well as other elections in 2006 and 2008 from UC Berkeley's Statewide Database.

4.2 Gravity Estimates

To characterize the impact of the high-speed rail using the equations (10), (11) and (12), we need estimates of the travel flows absent the CHSR, $\lambda_k(i, j, m)$ for $k = C, L$ (commuting and leisure), of the substitution elasticities θ_k for $k = C, L$, and of the parameters ρ and μ_L, along with information on the times and cost shocks associated with the CHSR. In addition, when business travel is introduced in the analysis as part of the full model, we need estimates of the business

route. Since commuting costs are about 2% of disposable income, $\hat{I}(i, j, m) \approx (1 - t)\, \hat{y}(i, j)$, and in the baseline model with only direct effects we have $\hat{y}(i, j) = 1$.

travel flows, $\lambda_B(i,j,m)$, and of the parameters θ_B and μ_B. Next, we discuss the procedure to obtain these estimates.

4.2.1 Commuting Equation

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

$$\lambda_C(i,j,m) = \frac{\left(\frac{I(i,j,m)}{D_C(i,m)\tau(i,j,m)^{\rho_C}}\right)^{\theta_C}}{\sum_{j'\in\mathcal{J}}\sum_{m'\in\mathcal{M}_\mathcal{C}}\left(\frac{I(i,j',m')}{D_C(i,m')\tau(i,j',m')^{\rho_C}}\right)^{\theta_C}} \tag{13}$$

where $I(i,j,m)$ (defined in (A.4) in Appendix B) denotes the total income net of commuting costs that residents of location i would have if they were to commute to location j by transport mode m, $\tau(i,j,m)$ denotes the travel time between locations i and j by mode m, and $D_C(i,m)$ captures a systematic component of preferences of residents of location i for transport mode m. All these variables are measured prior to the construction of the high-speed rail.[15]

We now discuss some implementation details. First, we allow for three possible travel modes for commuters: $\mathcal{M}_C = \{\text{walking or biking, car, public transport}\}$. Second, we assume the labor income component of $I(i,j,m)$ 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.[16] Third, for every pair of tracts i and j and mode m, we observe commuting shares and construct yearly commuting costs and travel times as discussed in Appendix C. Fourth, 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). [17] 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, $\lambda_C(i,j,m)$ determined according to (13), by a term $error_C(i,j,m)$ that captures sampling error: $\lambda_C^{obs}(i,j,m) = \lambda_C(i,j,m) + error_C(i,j,m)$.

[15] I.e., all expressions in this section correspond to the pre-CHSR full model equilibrium, as defined in Appendix B. To save notation, we omit the index s in every variable; e.g., the expression in (13) corresponds to $\lambda_C(i,j_C,m_C,s)$ for $s=N$, as defined in (A.17).

[16] 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.

[17] 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 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. Only 5 origins have no destination tract reachable by bike or walking.

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

$$\mathbb{E}\left[\left(\lambda_C^{obs}\left(i,j|m\right)-\lambda_C\left(i,j|m\right)\right)X\left(i,j,m\right)\right]=0, \tag{14}$$

where $X(i,j,m)=\left(\ln\left(I\left(i,j,m\right)\right),\ln\tau\left(i,j,m\right)\right)'$, and $\lambda_C^{obs}\left(i,j|m\right)$ and $\lambda_C\left(i,j|m\right)$ denote the observed and model-implied shares of residents of census tract i that commute to j conditional on the transport mode m. These moments are independent from origin- and mode-specific effects $D_C\left(i,m\right)$. 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 that residents of a census tract i have for a particular transport mode m, we model the origin- and mode-specific term $D_C\left(i,m\right)$ entering (13) as a function of observed origin-specific covariates $X_C\left(i\right)$ and a vector of unknown mode-specific parameters $\Psi_C\left(m\right)$:

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

In our implementation, $X_C(i)$ 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, regardless of the destination. We use the mode m-specific moment conditions

$$\mathbb{E}\left[\left(\sum_j\lambda_C^{\text{obs}}\left(i,j,m\right)-\sum_j\lambda_C\left(i,j,m\right)\right)X_C\left(i\right)\right]=0, \tag{15}$$

where we build a sample analogue of these mode-specific moment conditions by taking an average across all origins$i$.[18]

In the first step, we obtain an estimate of θ_C equal to 2.97 (with robust standard error equal to 0.14) and an estimate of ρ_C equal to 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; it thus measures the value of time in commuting. Our estimates of θ_C and ρ_C are both 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 the US, and obtain estimates of $\theta_C=3.3$ and $\rho_C=1.34$.

[18]Since we observe data from a sample of workers residing in each census tract, we assume that any difference between $\lambda_C^{obs}(i,j|m)$ and $\lambda_C^{obs}(i,m)$ and their model implied expressions, when evaluated at the true parameter values, is due to sampling noise. Consequently, the estimates of θ_C, ρ_C, and $\{\Psi_D(m)\}_m$ that we obtain from the two sets of moment conditions described above converge to their true values as the sample size in each census tract grows arbitrarily large. An assumption of our current setting is that there is no amenity component of destination census tracts that matters for workers' commuting patterns; as long as these amenities are equally valued by all residents, regardless of their census tract of residence, one may extend our analysis to include destination-specific fixed effect that would account for such unobserved amenities.

We present the resulting second-step estimates of $\Psi_C(m)$ in Appendix Table A.2.[19] Column (1) shows that, on average, residents have a preference for commuting by public transport and, to an even greater extent, by car, relative to bike or walking. Importantly, these preferences are in addition to whichever preferences may arise from the different commuting times and monetary costs associated to each transport mode. Columns (2) to (8) further reveal differences in these preferences across census tracts. Focusing on column (8), we observe that car ownership is associated to a positive preference for commuting by car and a mild negative preference for public transport relative to bike or walking. Younger or more educated census tracts have a negative preference for commuting by either public transport or car, while the share of nonwhite residents is strongly correlated with a preference for public transport. Log median income and log population density seem both correlated with preferences for both public transport and car relative to commuting by bike or walking.

Using the estimates of θ_C, ρ_C, and $\{\Psi_C(m)\}_m$ described in this section, as well as the expression for $\lambda_C(i,j,m)$ in (13), we generate model-predicted values of the share of commuters between every pair of census tracts (i,j) by any transport mode m. As these predicted shares are functions of consistent estimators of the parameters of interest, they are themselves consistent.

4.2.2 Business and Leisure Travelers

Our modeling of both leisure and business trips implies the following expression for the share of trips that have origin in a census tract i, destination in a census tract j, and are done using a mode of transit m in a setting before the high-speed rail is implemented:[20]

$$\tilde{\lambda}_k(i,j,m) = \frac{\left(\frac{Z_k(i,j)}{D_k(i,m)\tau(i,j,m)^{\rho_k}}\right)^{\mu_k\theta_k} p_k(i,j,m)^{-\mu_k\theta_k-1}}{\sum_{j\in\mathcal{J}}\sum_{m\in\mathcal{M}_k}\left(\frac{Z_k(i,j)}{D_k(i,m)\tau(i,j,m)^{\rho_k}}\right)^{\mu_k\theta_k} p_k(i,j,m)^{-\mu_k\theta_k-1}} \tag{16}$$

for $k=L$ (leisure) or $k=B$ (business), and where $\tilde{\lambda}_k(i,j,m)$ denotes the share of trips of type k originating from census tract i that both have tract j as destination and are done by transport mode m,[21] $Z_k(i,j)$ is a quality shifter of the number of trips done from tract i to tract j, $D_k(i,m)$ captures a systematic component preferences of travelers from i by transport mode m, $\tau(i,j,m)$ denotes the travel time between locations i and j by mode m, and $p_k(i,j,m)$ is the monetary cost per round trip. For both leisure and business travel we consider three feasible modes of transport: $\mathcal{M}_B = \mathcal{M}_L = \{\text{airplane, private vehicle, public transport}\}$.[22]

Due to limitations in the size of the sample, we write the shifters $Z_k(i,j)$ and $D_k(i,m)$ as a function of observable characteristics rather than as origin-destination and origin-mode fixed effects.

[19] 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.

[20] This expression is implied by (A.18) to (A.22) in Appendix (B.4).

[21] We use $\tilde{\lambda}_k(i,j,m)$ to denote the share of trips in order to distinguish it from $\lambda_k(i,j,m)$, which we use to denote the corresponding share of travelers and is the summary statistic we use to quantify the model.

[22] In the case of leisure, the shifter is $Z_L(i,j) = q_L(i,j)\,B(j)$, equal to the product of a bilateral preference shifter and destination amenities; in the case of business travel it is $Z_B(i,j) = q_B(i,j)\,A(j)$, equal to the product of a bilateral productivity shifter and endogenous destination productivity.

Specifically, we let:

$$\left(\frac{Z_k(i,j)}{D_k(i,m)}\right)^{\mu_k\theta_k} \equiv \exp\left(\gamma_k(m) + \Psi_k X_k(j)\right), \tag{17}$$

for $k = L, B$, where $\gamma_k(m)$ is a mode and purpose-specific parameter, Ψ_k is a vector of mode-invariant but purpose-specific parameters, and $X_k(j)$ is a vector of observed characteristics. In our empirical specification for leisure travel, $X_L(j)$ includes proxies for the amenity value of the 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. For the business travel regression, $X_B(j)$ includes the share of workers in management roles in the destination tract, as well as 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. Similarly to the commuting estimation, our observed measure of $\tilde{\lambda}_k(i,j,m)$, which we denote as $\tilde{\lambda}_k^{obs}(i,j,m)$, is based on a small random sample of residents of California and, consequently, will likely differ from the corresponding true share $\tilde{\lambda}_k(i,j,m)$ due the sampling error.

As the expression in equation (16) illustrates, the parameters μ_k and θ_k are not separately identified from this estimating equation alone. We thus calibrate μ_k using external data sources. For leisure travel, we calibrate μ_L to equal 0.05, consistent with the Bureau of Labor Statistics figure on the annual share of spending on travel across U.S. households, including on transportation, food away from home, and lodging.[23] For business travel, μ_B corresponds to the share of a firm's revenue spent on its employees' business travel. We set $\mu_B = 0.015$ using information from industry reports.[24] Furthermore, the only way to separately identify the parameters θ_k and ρ_k is to use the separate variation provided by 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, we are reasonably confident of its accuracy, our measure of the monetary cost of traveling between any two census tracts is likely to suffer from substantial measurement error. Consequently, we introduce the term $p_k(i,j,m)$ merely as a control, assume that $\rho_L = \rho_B = \rho_C$ (as a reminder, ρ_C is the corresponding term entering the commuting equation), and estimate θ_k from the variation induced by the travel time term $\tau(i,j,m)$.

We follow a two-step estimation approach similar to that described in Section 4.2.1. In the first step, we identify θ_k and β_k through the following moment condition:

$$\mathbb{E}\left[\left(\lambda_k^{obs}(i,j|m) - \lambda_k(i,j|m)\right) X_k(i,j,m)\right] = 0, \tag{18}$$

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

[23] The US travel Association reports leisure travel spending in the ballpark of 800 billion USD, which as a share of US private consumption yields a similar share of 5.6%.

[24] 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.

the transport mode m,

$\hat{\rho}_k = \hat{\rho}_C = 0.75$, and α_k is an unknown nuisance parameter. The survey from which we obtain the information on leisure and business trips $\tilde{\lambda}_k^{obs}(i,j,m)$ only collects information for trips whose origin and destination tracts are at least 50 miles away; thus we build sample analogues of the moment conditions in (18) by taking an average across all modes of transport m and all pairs of origins i and destinations j that are at least 50 miles apart. Consequently, for each origin tract i, the denominator of the second term entering the moment condition above sums only over such a restricted choice set.[25]

In a second step, given the first-step estimates of θ_k, β_k, and α_k, and a normalization that imposes that the model-specific shifter $\gamma_k(m)$ equals zero for $m =$ airplane, we identify the shifter $\gamma_k(m)$ for $m =$ private vehicle and $m =$ public transport using the following two mode m-specific moment conditions

$$\mathbb{E}\left[\sum_j \tilde{\lambda}_k^{\text{obs}}(i,j,m) - \sum_j \tilde{\lambda}_k(i,j,m)\right] = 0. \tag{19}$$

We build a sample analogue of these mode-specific moment conditions by taking an average across all origins i.[26]

Our estimate of the coefficient on log travel time $-\mu_k\theta_k\rho_k$ computed according to the moment conditions in equation (18) is -1.20 (with robust standard error equal to 0.28) for leisure and -1.65 for business (robust s.e. 0.82). According to the estimates presented in Section 4.2.1, this time elasticity is -2.24 in the case of commuting, reflecting a higher disutility of time spent traveling in the case of regular trips to the workplace. In combination with our calibrated values of μ_L and μ_B and our estimate of ρ_C, the estimated coefficients on log travel time imply the substitution elasticities across destinations are $\theta_L = 31.99$ (with robust standard error equal to 7.35) and $\theta_B = 146.98$ (with robust s.e. equal to 73.38). This result reveals that, despite small substitution rates with respect to travel time, travelers in fact perceive business and leisure destinations as highly substitutable among each other because the calibrated budget share of each activity is small.

Concerning the estimates of $\gamma_k(m)$ for $k = L$ and $k = B$ and both $m =$ private vehicle and $m =$ public transport, we find that leisure travelers have a strong amenity preference for traveling by car (for $m =$ private vehicle, $\hat{\gamma}_L(m) = 4.02$ with robust standard error equal to 0.17) , while we are not able to reject the null hypothesis that they extract the same amenity value from traveling

[25] In the case of trips performed by airplane, we further classify as infeasible those for which the travel time by airplane is larger than by car. Hence, in our application, the choice set varies also by transport mode m, being smaller for $m =$ airplane than for the other two modes of transport, which share the same choice set. Specifically, the sample analogue of each of the moments described in equation (18) 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 and by public transport.

[26] As discussed in Section 4.2.1, we assume that, for both $k = L$ and $k = B$, any difference between the observed travel shares entering in the moment conditions and and their model implied expressions, when evaluated at the true parameter values, is due to sampling noise. Under this assumption, our estimates of the parameters entering equations (16) or (17) converge to their true values as the sample size grows arbitrarily large. Currently, these estimates are computed under the extra assumption that there is no unobserved amenity component of destination census tracts that matters for workers' business and leisure trips; however, as long as these amenities are equally valued by all residents, one may account for such unobserved amenities through destination-specific fixed effects.

by public transport and by air (for $m =$ public transport, it is the case that $\hat{\gamma}_L(m) = -0.26$, with robust s.e. equal to 0.23). For the case of business travelers, we also find that they have an amenity preference for traveling by car relative to air, although this one is much smaller than in the case of leisure travelers (specifically, for $m =$ private vehicle, it is the case that $\hat{\gamma}_B(m) = 0.65$, with standard error equal to 0.10). Quantitatively more important is the disutility that business travelers have for traveling by public transport relative to traveling by air (for $m =$ public transport, it is the case that $\hat{\gamma}_B(m) = -1.69$, with robust s.e. equal to 0.21).[27]

Using the estimates described in the previous paragraphs and the expression for $\tilde{\lambda}_k(i, j, m)$ that results from combining (16)and (17), we compute model-implied shares of trips As these predicted shares are functions of consistent estimators of the parameters of interest, they are themselves consistent. Furthermore, they exploit the structure of the model to generate predicted trip shares for origin and destination tracts that are less than 50 miles away, solving the corresponding missing data problem that affects our survey data on leisure and business trips.

4.3 Model Parametrization and Counterfactual Scenarios

We implement counterfactuals using the system (A.32)-(A.42) in Appendix B.8 at the census tract level, for the 7866 census tracts for which all information is available. Doing so requires information on a range of variables in a pre-CHSR equilibrium that we describe next. For each variable we choose the latest year for which data is available prior to 2020, as this was the earliest year on which the CHSR would have been operational according to the 2008 business plan, and because doing so avoids noise from Covid-19 shocks.

First, the fraction of commuters, leisure travelers, and business travelers by origin-destination and mode ($\lambda_k(i, j, m)$ for $k = C, L, B$), the origin and destination components of commuters' wages, the gravity parameters (ρ_k, θ_k), and the spending shares in leisure and business travel (μ_L, μ_B) are obtained for 2019 from the estimation and calibration steps described in the previous sections. Consistent with the previous estimation, we implement the model assuming that the available transport modes for commuting are $\mathcal{M}_C = \{\text{car,public transit,walking/biking}\}$, while for leisure and business travel they are $\mathcal{M}_k = \{\text{car,public transit,air}\}$ for $k = L, B$. Second, we retrieve information on the number of residents, the share of floor space use for housing, labor income, land-rent income, and initial commuting travel costs as described in Appendix C. Finally, for the full model, we calibrate spillover elasticities and firms' use of floorspace using existing estimates from the literature, as also explained in Appendix C.

[27]Concerning the estimates of the parameter vector Ψ_k, which includes the coefficients on the observed destination-specific shifters included in the vector $X_k(j)$,we find that leisure travelers tend to travel more to census tracts that are closer to the beach, that belong to a national park, or that have a larger share of employment in hospitality (the estimates of the corresponding parameters are always statistically different from zero at the 1% level). Conversely, the relationship between the population of a census tract and its amenity value as a destination for leisure trips is not statistically different from zero. For the case of business trips, we find that, similarly, the relationship between log population of a census tract and its appeal as a destination for business trips is not statistically different from zero, but that the share of workers employed in management positions has a positive and statistically significant correlation with the amenity value of a tract as a destination for business travelers.

Table 1: Model and Calibration Variants

	Impact of the CHSR			
	Direct time and cost impacts	Indirect impacts (land, wages, spillovers)	Completion probability and taxpayer costs	CHSR may replace:
Baseline	yes	no	$p = 1, T = 12, c = \$10b$	pub. transit, air
Full	yes	**yes**	$p = 1, T = 12, c = \$10b$	pub. transit, air
Pessimistic	yes	no	$\mathbf{p = 0.5, T = 24, c = \$130b}$	pub. transit, air
+ Car	yes	no	$p = 1, T = 12, c = \$10b$	pub. transit, air, **car**

Note: this table summarizes the differences across counterfactual types. Each of the cases "Full", "Pessimistic" and "+Car" deviate from the baseline by one criterion shown in bold. In the third column, T is the minimum number of years after the vote at which the CHSR may become operational, p is the probability that it will become operational in after year after T if it has not done so before, and c is the present discounted value of taxes Californian taxpayers expect to pay if the project is approved. In the highly optimistic baseline, c equals just the $10bn pledged by the Prop 1a. In the much more "Pessimistic" case, c equals current projections. In particular, the 2022 business plan projected between $72bn and $105bn for phase-I of CSHR (California High Speed Rail Authority, 2022). Since our counterfactuals encompass the full network (about 800 miles) instead of just phase-I (about 520 miles), we take an intermediate point of these estimates and adjust proportionally by the extra length to obtain costs of about $130bn.

Using these equations and these data, to compute the real income gains associated with the CHSR we also need the time and monetary cost shocks $\hat{\tau}(i, j, m)$ and $\hat{p}_k(i, j, m)$ for each origin-destination pair and mode of travel on which the CHSR is endogenously used by travelers, if available. In our baseline counterfactuals, we compute these time and monetary costs shocks assuming that the CHSR is a perfect substitute to public transit and to air travel. Appendix C describes how we compute time and cost of travel with and without the CHSR for each mode, and Appendix B.6 describes the endogenous decision of each type of traveler to use the CHSR. The next section describes summary statistics of this shock.

Given this parametrization, we implement counterfactuals assuming 4 configurations of model assumptions that span varying levels of complexity and optimism regarding the completion and costs of the CHSR. Our "baseline" parametrization implements the simplest model without any general equilibrium effects, in which the only benefits from CHSR accrue from time and cost savings of commuters and leisure travelers (i.e., there are no productivity-enhancing effects through business travel). In this highly optimistic baseline, the completion probability, time until completion, ticket prices, and sunk cost financed by Californian taxpayers correspond to what was spelled out in the 2008 CHSR business plan. Specifically, this assumes that the project is complete and operational in 2020 ($p = 1$ and $T = 12$), with total costs to Californian taxpayers of only the $10bn pledged by the Prop 1a.

We implement three additional models, each differing from the baseline in a different dimension: the "full" model allows for indirect impacts of the CHSR through land prices, wages, and spillovers accruing from agglomeration of employment.; the "pessimistic" variant assumes half the completion probability ($p = 0.5$) in twice the time ($T = 24$), with current projections on ticket prices (approximately twice as large as in 2008) and with costs to Californian taxpayers to cover in full current projections of capital costs; and the "+Car" variants allows CHSR to also be considered as a perfect substitute to traveling by car in the eyes of travelers. Table 1 summarizes these alternatives and provides additional details.

4.4 Time and Cost Shocks from the CHSR

Table 2 gives a sense of how many travelers are directly exposed to the CHSR shock, and by how much. The first two columns show fractions of total travelers who, before the CHSR is available, travel on routes where the CHSR would be used if available. These are lower bounds to the number of actual users in a counterfactual with the CHSR, as initial travelers on other routes may substitute into routes where the CHSR is preferred. The subsequent columns report moments from the distributions of time and cost changes across these direct winners.

The first column considers our baseline scenario where the CHSR may only replace public transit or air travel. Before the CHSR, a small fractions of public-transit or air travelers use routes where the CHSR is preferred, which follows from the initial share of these travelers being small (see Table A.1). A considerable fraction of these initial travelers directly benefit (19% of initial public-transit commuters, 14% of initial public-transit or air leisure travelers, and 43% of the initial public-transit or air business travelers). The shares grow considerably in the second column, when the CHSR may directly replace the car, because the initial share of travelers by car is large. However, comparing to the initial shares in Table A.1, only 2% of commuters via car directly benefit, while 14% of leisure and 8.6% of business travelers by car do.

The remaining columns report the median and 75 percentile of the traveler-weighted distribution

Table 2: The CHSR Shock

	% Initial Travelers Directly Better Off		Time Gain		Cost Change (Pub. Trans. or Air) 2008 CHSR Price		2X Ticket Price	
	Pub. Trans. or Air	+ Car	median	75 ptile	med	75p	med	75p
Commute	1.0%	3.1%	26' (31%)	43' (41%)	-19%	-4%	5%	22%
Leisure	0.5%	14.1%	10' (6%)	33' (25%)	-62%	-42%	-33%	-12%
Business	5.1%	12.7%	9' (4%)	27' (12%)	-63%	-57%	-33%	-25%

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 second column assumes that the CHSR may also replace the car. The third and fourth (fifth and sixth) column show moments from the traveler-weighted distribution of time changes (cost changes) across origin-destination-modes within each travel purpose, conditioning on origin-destination-modes where the CHSR is used when available.

of time gains and cost changes across all routes and travel modes that directly benefit, assuming as in the first column that the CHSR is a direct substitute to public transit or air. The direct beneficiaries save a substantial amount of time; e.g., all the commuters via public transit who directly benefit, the median saves 31 % of commute time. These direct winners also save a substantial amount in travel costs at the most optimistic 2008 projection for ticket prices. In the "pessimistic" scenario, corresponding to current projections of ticket prices, there are still considerable gains among leisure or business travelers who directly benefit, whereas for commuters the time savings come at the expense of higher commuting costs. The difference between these types of travelers is because the latter does not use air travel, and therefore only gains time when switching from the relatively cheap public transit into the CHSR.

4.5 Aggregate and Regional Real-Income Effects of the CHSR

We now compute the expected real-income effects of the CHSR, $\hat{W}(i)$ defined in (9). Table 3 summarizes the aggregate effects across the 4 model variants described in Section 4.3. In our baseline (without general-equilibrium effects and with CHSR costs as promised in the 2008 business plan), the net aggregate gain is 0.32%, which corresponds to an annual monetary gain of $143.0 in 2008 USD per worker. The "full" model, which includes productivity effects stemming from business travel, changes in land prices and spillovers, has net aggregate gains that are twice as large. The gains are also about twice as large as the baseline in "+Car" variant, which assumes that, when deciding how to substitute across modes, travelers consider the combination of driving and using the CHSR as a perfect substitute to just driving (in the baseline, using the CHSR is considered as a perfect substitute to public transit or to taking an airplane, but not to driving).[28] The pessimistic scenario (which roughly corresponds to current projections on costs and ticket prices) gives net losses.[29]

[28] 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 $135bn. The 2023 CHSR project update (California High Speed Rail Authority, 2023b) includes a benefit-cost analysis (California High Speed Rail Authority, 2023a) such that, if only "high-speed rail user benefits" (corresponding to the forces we include in the "pessimistic" case) are included (while wider economic benefits are excluded), 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 "pessimistic" case only considers direct time savings to travelers, uses total costs corresponding to the 2022 Business Plan (California High Speed Rail Authority, 2022), and 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 "pessimistic" scenario yields a loss $121bn in 2021 USD. The 2023 CHSR update also reports gains of $26bn for the Phase I when "wider economic benefits for worker and firms" are further included in addition to rail user benefits. If we run a "Pessimistic+Full" version of our model (so that wider economic benefits beyond direct time savings are included), we obtain 2021 USD gains of $33bn for the full network.

[29] Appendix table A.3 decomposes the aggregate effects into the gains stemming from commuting, leisure travel, and business trips. In the three cases without general-equilibrium effects, leisure trips account for about 10%-25% 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. Leisure trips are more important when the CHSR may directly replace the car rather than only public transit. The magnification of gains in the "full" case with general-equilibrium impacts are due almost exclusively to productivity-enhancing effects of business trips.

Table 3: Aggregate Real-Income Impacts from the CHSR

Case	Aggregate Annual Gain	2008 USD per Worker
Baseline	0.32%	$143.0
Full Model	0.65%	$292.4
Pessimistic	-0.18%	$-81.9
+ Car	0.60%	$265.6

Note: The table reports the aggregate annual gain $\hat{W}$ across the model variants described in Section 4.3, defined as the working age population weighted average across tracts of $\hat{W}(i)$ defined in (9). This number measures an annual welfare gain. To transform to 2008 USD gains, we multiply $\hat{W}(i)$ by the average nominal 2008 wage in tract i and then take the residents-weighted average across tracts.

These aggregate changes mask considerable heterogeneity. Figure 2 plots real-income effects across tracts in the baseline scenario and Figure 3 zooms into the LA county and the San Francisco Bay Area. Bright yellow tracts gain the most, and dark blue tracts lose the most. The top 10% of winning tracts gain between 0.6% and 4.7% per year, while 2.7% of all tracts lose. To gauge the drivers of heterogeneity, Appendix Table A.4 shows regressions of these tract-level real-income changes on tract-specific characteristics in the pre-CHSR equilibrium: each 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 or San Francisco county. These variables enter indirectly in the quantification through the summary statistics needed for the quantification, as described in Section 4.3. 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. As a result, among Central Valley locations, Fresno and Bakersfield stand out with relatively low gains. The tracts located in San Francisco or Los Angeles county gain over and above what these characteristics would predict, suggesting that their specific travel patterns, in particular for commuting, are more likely to adopt the CHSR.[30]

As we shall see in Section 5, despite the different aggregate effects across model variants shown in Table 3, the heterogeneity in gains across model variants is always consistent with a both i) a strong elasticity of votes to real income, and ii) a more important role for political preferences than economic effects in explaining the cross-sectional dispersion in welfare effects. Furthermore, as we will see in the Section 6, the significant gap between the gains in San Francisco or LA and the Central Valley would have been even larger in the absence of political incentives for the planner.

[30] We also find that employment gains are concentrated in the immediate vicinity of CHSR stations, in Los Angeles and San Francisco, and also in the Central Valley (e.g. Stockton, Modesto). Suburban locations in the greater San Francisco and Los Angeles areas hollow out.

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

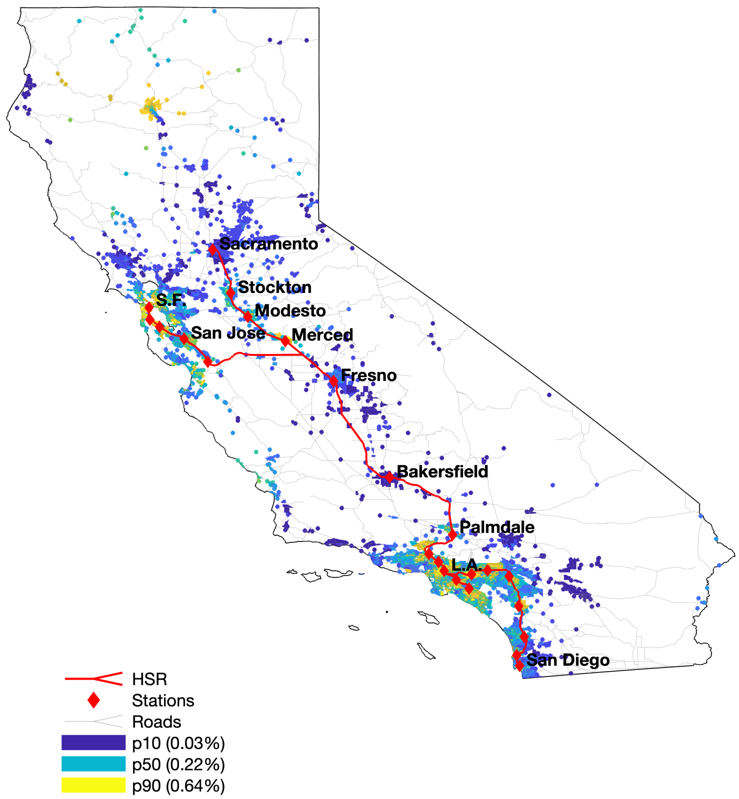
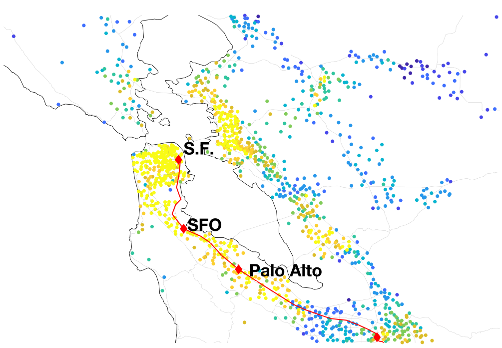
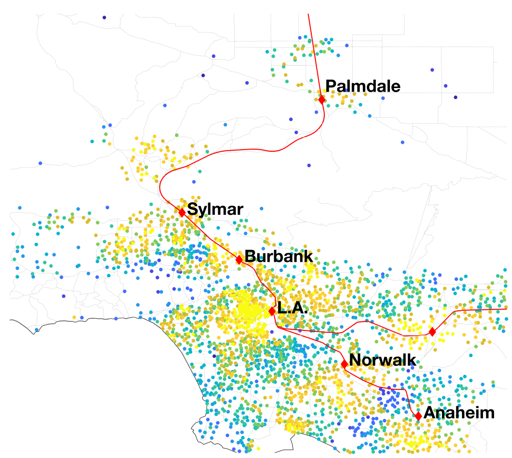


Figure 3: Spatial Distribution of Real-Income Gains from the CHSR (SF and LA)

(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.

5 Voter Preferences

Armed with the quantified model of the real-income gains of the CHSR, we now estimate the relative importance of real-income and political components of preferences in shaping voters' overall preferences for the CHSR.

5.1 Estimation Strategy

We bring (8) to the data. We aim to compute consistent estimates of θ_V and β_k. Using these estimates, we can recover the weights $\tilde{\beta}_k \equiv \beta_k/\theta_V$ on the components of political preferences defined in (4) and compute predicted distributions of vote shares in favor of the CHSR in counterfactual scenarios.

As proxies for political preferences, we consider three different sets of covariates $X_k(i)$ in (8). The first set bears a direct relationship to voters' political ideology: it includes the share of registered Democrats and the shares of votes cast in favor of two propositions, Prop 10 and Prop 1B, on alternative energy and transportation projects, respectively.[31] We interpret the vote shares on Prop 10 as proxies for the strength of the environmental concerns and the vote shares on Prop 1B as proxies for the willingness of voters to support transportation infrastructure spending in general. The second set of covariates measures several demographic characteristics of the residents of each census tract: the share of residents who are nonwhite, college-educated, and under 30 years of age. Finally, the third set of covariates includes the distanced between each census tract and both the closest CHSR station and the closest point on the CHSR railway track. These two distance measures account for the potential amenity impacts of being close to either a CHSR station (likely positive) or the CHSR railway track (likely negative), as well as for potential correlation between the location of the CHSR tracks and stations and voters political ideology that is not properly captured by the two other sets of covariates. In addition to the three sets of covariates discussed above, we control in all specifications for county fixed effects. Consequently, the identification is based on variation across census tracts within counties.

The consistency of OLS estimators of the the parameter θ_V is affected by several potential identification concerns. First, the expectational error of voters (term $\epsilon_W(i)$ discussed in Section 3.1) is included in the error term. If voters' expectations are rational, the expectational error $\epsilon_W(i)$ is correlated with the realized values of $\ln \hat{W}(i)$ across tracts, biasing the OLS estimate of θ_V towards zero (Dickstein and Morales, 2018). Addressing this issue requires an instrument that belongs to the voters' information set when they voted. We use as instrument the real income impact of the CHSR $\hat{W}(i)$ generated by our model with fundamentals from the year 2008 (our dependent variable $\hat{W}(i)$ is constructed with fundamentals from the year 2019). Assuming that voters' expectations are rational and that the 2008 fundamentals belong to their information sets,

[31] Proposition 10 ("Bonds for Alternative Fuels Initiative"), on the ballot in the same 2008 election as the CHSR, would have allowed the state to issue \$5 billion in bonds for alternative fuel projects (it failed with 40.6% in favor). Proposition 1B ("Transportation Bond Measure"), on the ballot in the 2006 midterm elections, authorized the state to issue \$19.9 billion in bonds for transportation projects (it passed with 61.4% in favor). Source: *ballotpedia.org*.

this instrument is mean independent of voters' expectational error $\epsilon_W(i)$.

Second, the error term also includes the term $\epsilon_a(i)$ in (4), which captures preferences for the CHSR of residents of a tract i that are not adequately controlled for by the observed covariates $X_k(i)$. If the CHSR favored tracts with systematically large or small values of $\epsilon_a(i)$, this term would be correlated with the real income effects of the CHSR, $\ln \hat{W}(i)$. More generally, it is possible that components of voters' political preferences that are not properly captured by our proxies are correlated with the voters' exposure to the CHSR as measured by our model. The OLS estimate of θ_V may be biased upwards or downwards as a result.[32]

To address these potential issues, we follow two IV strategies that rely on counterfactual CHSR designs. For each census tract i, both instruments equal the average real-income change $\hat{W}^{IV}(i)$ of the residents of i across 100 simulated counterfactual high-speed railway networks:

$$\hat{W}^{IV}(i) = \frac{1}{100}\sum_{n=1}^{100} \ln\left(\hat{W}^{cf}(i,n)\right), \tag{20}$$

where $\hat{W}^{cf}(i,n)$ is the model-predicted change in welfare in location i from a counterfactual CHSR design n, and where the model used to generate this welfare change is calibrated to 2008 fundamentals, so that they are also mean independent of the expectational error term $\epsilon_W(i)$.

The first IV following (20) uses CHSR designs built by randomizing the location of the 24 stations along the coordinates of the currently projected CHSR line. The second IV exploits the shape of three alternative routes that were considered in the early stages of the CHSR design process during the 1990s (US DOT, 2005), as shown in Appendix Figure A.1. One of these routes was similar, but not identical, to that which was ultimately chosen and voted upon in 2008. Our understanding of the process is that these three initially considered CHSR routes were selected on the basis of technical feasibility and cost savings.[33] Thus, these three potential routes were thus *directly* affected by the political preferences of voters in any given tract. Our second instrument uses the formula described in equation (20), where each counterfactual n corresponds to a high-speed railway network where one of these three potential routes is randomly chosen and then the 24 stations are randomly located along the randomly chosen route. As these two instrumental variables do not incorporate information on the actual location of the 24 projected stations (in the case of the first instrument) nor on the actually projected CHSR railway line (in the case of the second instrument), they may still be valid even if CHSR stations (or railway line, in the case of the second instrument) favor tracts with systematically large or small values of the unobserved term.

A third potential concern that would bias the OLS estimator of θ_V as well as invalidate all instrumental variables described above is the possibility that our economic model does not correctly

[32] It is worth highlighting that the set of covariates in our regression account for a large share of the variation in vote shares across locations; the R-squared of the OLS estimated regression is slightly above 0.9 in the specification with the largest set of covariates.

[33] There are two main potential routes from Northern to Southern California: along the coast or through the center of the state (along I-5 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 is flatter and cheaper but 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.

capture voters' forecast of the economic gains of the CHSR. It is generally hard to determine the statistical properties of this potential error in our measure of voters' expected welfare gains and, consequently, we do not attempt to find an instrumentation strategy that may deal with it. To address this concern, we check instead the robustness of our analysis to the range of alternative model specifications that we have studied in the previous section.[34]

5.2 Estimates of Voter Preferences

In Table 4, we report report OLS and IV estimates of θ_V and the β_k's relying on the baseline variant of the economic model. In the first four columns, we present OLS estimates for regression specifications that progressively account for a larger the number of regressors. In column (1), we include no covariate other than the model-implied welfare impact of the CHSR, and obtain a positive and significant estimate of θ_V. In column (2), we add the proxies for political ideology, in (3) we add variables that capture demographic composition, and in column (4) we also incorporate the log distances of each tract to both the closest station and the closest point on the CHSR railway line.

In this last OLS regression specification, the estimated value of θ_V equals 14.15 (with robust standard error equal to 1.03). According to these estimates, a larger support for the CHSR is predicted by a larger support for Obama in the 2008 general election; 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; proximity to a CHSR station; and distance from the CHSR railway line. The specification whose estimates are presented in columns (5) to (7) coincide with that whose estimates appear in column (4), but instead of presenting OLS estimates they correspond to IV estimates that use as instrumental variable either the 2008-based real income measure (in column (5)), the "Random Station" instrument in equation (20) that randomizes the location of the stations alone (in column (6)), or the "Random Path" instrument that randomizes the location of both the railway line and the stations (in column (7)). For all these three instruments, the large first-stage F-statistics suggest that these are not weak.

The estimate of θ_V increases progressively as we move from column (4) to (7). The fact that the estimate of θ_V computed using the 2008-based real income measure as instrument is larger than the OLS estimate (16.84 vs. 14.15) is consistent with the OLS estimate being downward biased as a consequence of voters' expectational errors. The fact that the estimates of θ_V computed using both the random-station and the random-path instrument are larger than that reported in column (5) may be explained as a consequence of the CHSR favoring census tracts whose residents were, for unobserved reasons, less predisposed to support the construction of the proposed CHSR.

[34] A specific source of model misspecification that none of the models studied in the previous section deals with and that none of the previous IVs directly address is the possibility that tracts situated closer to the CHSR line may face idiosyncratic economic consequences from it; e.g., economic compensation from eminent domain, or direct nuisance from proximity to the train tracks. Thus, we control for the distance to the railway tracks. We show that our estimates are also robust to dropping census tracts that are less than 5 km away the railway line.

Table 4: Estimates of Voting Equation

Inst. Var.:	None - OLS				$\ln(\hat{W}_{08})$	Random Station	Random Path
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$log(\hat{W}_{19})$	38.53^a	17.27^a	14.53^a	14.15^a	16.84^a	19.45^a	22.68^a
	(1.76)	(1.26)	(0.98)	(1.03)	(1.23)	(1.68)	(1.80)
Log-odds Dem. Sh.		0.30^a	0.38^a	0.38^a	0.38^a	0.38^a	0.39^a
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Environ.: Prop. 10		1.16^a	2.46^a	2.46^a	2.44^a	2.43^a	2.41^a
		(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Transp.: Prop. 1b		1.54^a	0.82^a	0.83^a	0.82^a	0.81^a	0.80^a
		(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)
Sh. non-White			-0.17^a	-0.17^a	-0.18^a	-0.18^a	-0.19^a
			(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Sh. College			0.74^a	0.74^a	0.73^a	0.73^a	0.72^a
			(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Sh. Under 30			0.17^a	0.18^a	0.18^a	0.18^a	0.18^a
			(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Log. Dist. Station				-0.01^a	-0.01^a	-0.01	-0.00
				(0.00)	(0.00)	(0.00)	(0.00)
Log. Dist. Rail				0.02^a	0.02^a	0.01^a	0.01^a
				(0.00)	(0.00)	(0.00)	(0.00)
F-stat					803	574	286
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.

Our results are qualitatively very similar across a range of empirical specifications, including different weighting schemes, sample selection, economic model variant, and covariate selection. Appendix Table A.6 replicates the estimates from Column (6) of Table 4 across the 4 main model variants that we have discussed in the previous section. The estimates of θ_V are smaller in the "Full" model when we incorporate general-equilibrium effects into the computation of voters' expected economic impact of the CHSR, much larger in the "pessimistic" model (because, in this case, the variance of real-income effects is much smaller), and roughly unchanged in the "+Car" version of the model. Table A.7 replicates Table A.6 using political covariates only.[35] For each model variant, the estimates of θ_V are very similar regardless of whether we use all covariates or only political covariates. A message in our next section is that the importance of political preferences in the cross-section of welfare effects is similar across these specifications.

[35]Columns (3) to (6) in Appendix Table A.5, show that the random-design IV estimates of θ_V are slightly larger when we weight each census tract either by total number of votes or participation rate in the Proposition 1A referendum. Columns (7) and (8) show that the estimates remain roughly constant when we drop the nearly 3,000 census tracts that are less than 5km away from the railway line.

5.3 Importance of Political Preferences

Armed with these estimates, we can now investigate the importance of political considerations versus real-income in driving preferences for transportation policy. The regression results above allow us to estimate the distribution of the political component of preferences as $\ln \hat{a}(i) = \sum_{k=1}^{K} \frac{\beta_k}{\theta_v} X_k(i)$. Not all the covariates $X_k(i)$ in $\ln \hat{a}(i)$ are proxies for political considerations only. We report all the results in two cases: i) using all covariates to build a proxy $\hat{a}^{all}$ for $\hat{a}$ (from the "Random-Station" IV estimates from Table A.6); and ii) using *only* political covariates to build a proxy $\hat{a}^{pol}$ for $\hat{a}$ (from the "Random-Station" IV estimates from Table A.7). Table 5 reports results for the broad definition $\hat{a}^{all}$, and Appendix Table A.8 reports results for $\hat{a}^{pol}$. The results are almost identical regardless of whether we use $\hat{a}^{all}$ or $\hat{a}^{pol}$. Hence, the political variables entering in $\hat{a}^{all}$ are the main drivers of the non-economic component of preferences.

Decomposition of Preferences We first directly decompose the utility impact of the CHSR (measured in percentage points increase in real income) into its economic and political components. Using (2) and (3), this utility change in location i is

$$\Delta U(i) = \ln \hat{W}(i) + \ln \hat{a}(i) - \epsilon_W(i).$$

As we do not observe the expectational error $\epsilon_W(i)$, we set it to zero and construct $\Delta U(i)$ under the assumption of perfect foresight about the fundamentals. We find that the CHSR has a population-weighted impact on $\Delta U(i)$ equivalent to a 1.06%- 1.00% increase in real-income (here and below, we report results for the two specifications, $\hat{a}^{pol}$ and $\hat{a}^{all}$). The actual population-weighted increase in real income $\ln \hat{W}(i)$ is 0.32%, so that the average intensity of political preferences is positive and large, corresponding to a 0.74%-0.68% increase in real income.

The second and third columns of tables 5 and A.8 show variance decompositions. The political component $\ln \hat{a}(i)$ has a standard deviation of 2.3%-2.2% across tracts, which is about 6 times larger than the real-income component $\ln \hat{W}(i)$, at 0.4%. The political component therefore drives a much larger fraction of the variation in preferences (and votes) than the real-income component. Hence, strong opponents to the project (as well as strong supporters) are typically politically motivated, rather than having a lot at stake economically.

Cost of Swaying Votes Our estimates of θ_V imply that changing votes is quite costly. As shown in the last two columns of tables 5 and A.8, at the observed distribution of votes the extra real-income gain from the CHSR that would be needed to sway an extra 1 percent of the population to vote in favor of the project has a median of 0.4% across tracts.[36] This suggests a large economic cost of swaying voters in the typical census tract.

Aggregate Vote To illustrate the various drivers of the vote, we first compute the counterfactual aggregate vote share in favor of the CHSR if voters voted only based on their real income gains

[36] The logit choice structure implies that swaying an extra 1 percentage point in favor of the CHSR requires an extra $\frac{1}{\theta_V v(i)^2 (1-v(i))}$% real-income gain from the CHSR.

Table 5: Political vs. Real Income Determinants of the Vote

	Variance Decomposition of $\Delta U(i)$			Favorable Vote if Voters Only Consider...		Cost of Swaying 1% of Vote	
	θ_V	$\sigma_{\Delta \ln a}$	$\sigma_{\Delta \ln \hat{W}}$	$\hat{W}$	$\hat{a}$	50p	90p
Baseline	19.5	2.2%	0.4%	51.4%	51.6%	0.4%	0.5%
Full	17.5	2.4%	0.4%	52.7%	50.3%	0.4%	0.6%
Pessimistic	41.6	1.0%	0.2%	47.9%	54.9%	0.2%	0.3%
+ Car	19.2	2.3%	0.6%	52.7%	50.4%	0.4%	0.6%

Note: The θ_V reported corresponds to IV estimate using the random routes instrument shown in Table A.6. The political component corresponds to all covariates. Appendix table A.8 shows the results using the more restricted proxy $\hat{a}^{pol}$.

(setting $\hat{a}(i) = 1$), or only based on their political preferences (setting $\hat{W}(i) = 1$). We find that the vote in favor of the CHSR would be 51.4%-51.6% based on political preferences only, and 51.3%-51.4% based on real income gains only. So, despite the high cost of swaying votes and the importance of political preferences for the cross-section of welfare effects, economic effects are revealed to be an important driver of the aggregate vote under optimistic beliefs. However, the fact that the CHSR would be approved in the complete absence of a preference for real income gains means that there is a range of costs such that the CHSR could lower aggregate real income and still be approved. For example, a CHSR that would lead to a uniform 0.3%-0.3% loss for all tracts in the baseline case would receive 50% favorable votes thanks to the strong political component of preferences. Under pessimistic beliefs this real-income loss that would have left the average voter indifferent increases to 0.5%-0.5%.

6 Planner's Preferences

6.1 Planner's Problem

We now formalize the problem of a hypothetical planner designing the CHSR. We use it to estimate the planner's preferences over residents of different locations (which we model as depending on the residents' demographic composition) and votes, as well as to implement optimal CHSR designs under counterfactual preferences.

We represent the observed CHSR as the optimal investment decision of a central planner. We focus on the problem of where to locate the 24 stations along the the actual rail line. Formally, the planner chooses geographic coordinates $\mathbf{d} = \{d_1, ..., d_{24}\}$ from the set $\mathscr{D}$ of coordinates that define the actual CHSR. For simplicity, we assume that the cost of building a station is the same everywhere. Each design $\mathbf{d}$ maps to a distribution of travel times and travel costs. Hence, each $\mathbf{d}$ maps to a real income $\ln W(i, Y, \mathbf{d})$ if the CHSR is approved, to a real income impact $\hat{W}(i; \mathbf{d})$ defined in (9), to an expected welfare impact $\Delta U(i; \mathbf{d})$ defined in (2), and to a share of favorable votes $v(i;\mathbf{d})$ defined in (7).[37]

Formally, the planner chooses the CHSR design $\mathbf{d}$ that solves the following optimization:

$$\mathcal{W}=\max_{\mathbf{d}\in\mathscr{D}}\mathbb{E}\left[\sum_{i=1}^{J}\Omega(i)\,N(i)\,\Delta U(i;\mathbf{d})+\lambda\sum_{i=1}^{J}N(i)\,v(i;\mathbf{d})\mid\mathcal{I}_P\right]. \tag{21}$$

The first term is a weighted sum of expected utilities across locations using location-specific (per-capita) Pareto weights $\Omega(i)$ that capture the planner's preferences for that location. The second term is the total number of California residents that approve the project. The variable $\mathcal{I}_P$ denotes the planner's information set. When $\lambda>0$, the planner attaches a positive weight to votes in favor of the CHSR. We thus allow the planner to take into account political considerations in the design of transportation infrastructure. Conditional on the project being approved, the design of the CHSR internalizes political preferences $\hat{a}(i)$ only through votes. Then, (6.1) is equivalent to:[38]

$$\mathcal{W}=\max_{\mathbf{d}\in\mathscr{D}}\mathbb{E}\left[\sum_{i=1}^{J}\Omega(i)\,N(i)\ln\hat{W}(i;\mathbf{d})+\lambda\sum_{i=1}^{J}N(i)\,v(i;\mathbf{d})\mid\mathcal{I}_P\right]. \tag{22}$$

We now consider optimality conditions. Similar to voters, we assume the planner has rational expectations. For any alternative design $\mathbf{d}^n$ different from the actual design $\mathbf{d}^0$, the objective function would have changed (to a first order approximation) according to:

$$\Delta\mathcal{W}(\mathbf{d}^n)\approx\sum_{i=1}^{J}\left(\Omega(i)+\lambda\underbrace{\theta_V v(i)(1-v(i))}_{\equiv V(i)}\right)N(i)\,\Delta\ln\hat{W}(i,\mathbf{d}^n)-\epsilon(\mathbf{d}^n)\leq 0 \tag{23}$$

where

$$\Delta\ln\hat{W}(i,\mathbf{d}^n)\equiv\ln\left(\frac{\hat{W}(i;\mathbf{d}^n)}{\hat{W}(i;\mathbf{d}^0)}\right)=\ln\left(\frac{W(i,Y,\mathbf{d}^n)}{W(i,Y,\mathbf{d}^0)}\right) \tag{24}$$

is the log difference in real income between the counterfactual and the actual CHSR designs given a realization of economic shocks, and ϵ accounts for the planner's expectational error when evaluating the impact of the high-speed rail design. Because we assume that the CHSR corresponds to the optimal choice of the planner, any deviation such as (23) must yield weakly negative returns.

Condition (23) demonstrates the trade-off between votes and real income in the planner's design. Naturally, the planner has incentives to invest more in locations with higher $\Omega(i)$ or higher returns in terms of votes. The latter is captured by the elasticity of votes to real income,

$$V(i)\equiv\frac{\partial v(i)}{\partial\ln\hat{W}(i)}=\theta_V v(i)(1-v(i)). \tag{25}$$

The voting elasticity is larger closer to median voters ($v(i)=0.5$), so that locations with very small or large political component $\hat{a}(i)$ are therefore less sensitive to real income. Next, we leverage that unchosen CHSR designs that increase real-income of a location i at the expense of some other location i' reveal an upper bound for $\Omega(i)$ relative to $\Omega(i')$. Similarly, unchosen CHSR designs

[37] An assumption implicit in this formulation is that the political preferences of voters over whether or not to build the CHSR, $\ln\hat{a}(i)$, do not depend on the design $\mathbf{d}$.

[38] This step combines (2) with (6.1) and uses the law of iterated expectations assuming that voter's information set at the time of voting is at least as rich as the planner's at the time of choosing the location of stations.

that increase (reduce) aggregate (Ω-weighted) real-income while reducing (increasing) aggregate votes reveal a lower (upper) bound for $\lambda(i)$ relative to the average of the Ω's.

6.2 Estimation

We use a moment inequality estimator based on (23) to compute confidence sets for the parameters entering the objective function of the planner's problem. We discuss key features and provide details of implementation in Appendix E.

6.2.1 Construction of Moments

First, to limit the dimensionality of the parameter vector, we write the planner's weight for each census tract i, $\Omega(i)$, as a function of observed covariates $Z_k(i)$ and a constant:

$$\Omega(i) = \beta_0 + \sum_{k=1}^{K} \beta_k Z_k(i). \tag{26}$$

Second, to build the inequalities described in (23), we construct a large number (i.e., several hundreds) of perturbations to the proposed CHSR. In each, a single station is moved to another "potential location". 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 $V(i)$ along the CHSR line.[39] 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 (23).

Figure 4 shows a perturbation where the Los Angeles station shifts towards Anaheim. The map on the left shows the real-income changes $\Delta \ln \hat{W}(i)$ defined in (24). The map on the right shows the voting elasticity $V(i)$. Downtown LA is a low-voting gradient area because it strongly supports the CHSR (a high value of $\hat{a}(i)$); 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 utilitarian 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 relative to an average of the $\Omega(i)$.

For the estimation, we build E moment inequalities (indexed by e) from these N perturbations, each defined as the sum of the inequalities $\Delta\mathcal{W}(\mathbf{d}^n)$ in (23) corresponding to (mutually exclusive) subsets $\mathcal{N}_e \subset \mathcal{N}$ of all perturbations:

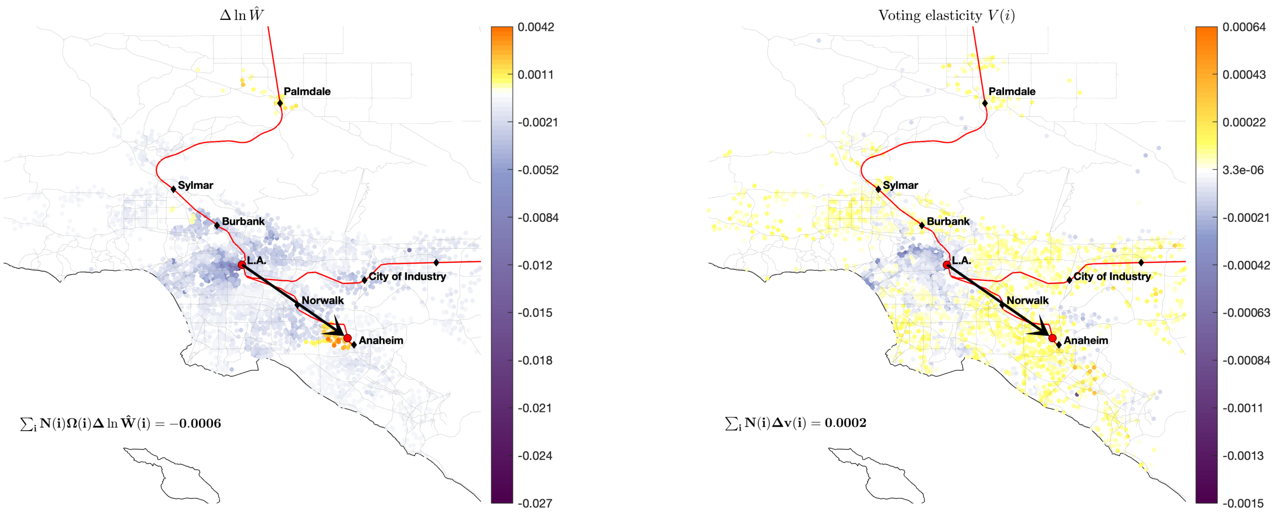
$$\sum_{n=1}^{N} \Delta\mathcal{W}(\mathbf{d}^n)\, 1\{n \in \mathcal{N}_e\} \leq 0 \text{ for } e = 1, .., E. \tag{27}$$

As long as the subset $\mathcal{N}_e$ is a function of variables that belong to the information set of the planner, the assumption of rational expectations implies that the expectational errors $\epsilon(\mathbf{d}^n)$ are averaged out, $\sum_n \epsilon(\mathbf{d}^n)\, 1\{n \in N_e\} \to 0$ as $\sum_n 1\{n \in N_e\} \to \infty$.[40] Then, using (26), we can rewrite the

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

[40] This result crucially depends on two assumptions. First, for any perturbation n, the unobserved term $\epsilon(\mathbf{d}^n)$

Figure 4: Example of Perturbation Identifying λ



moment inequality e in (27), as

$$\sum_{n=1}^{N}\sum_{i=1}^{J}\left\{\left[\beta_0+\sum_{k=1}^{K}\beta_k Z_k\left(i\right)+\lambda V\left(i\right)\right] N\left(i\right)\Delta\ln\hat{W}\left(i,\mathbf{d}^n\right)\right\} 1\left\{n\in\mathcal{N}_e\right\}\leq 0 \text{ for } e=1,..,E. \quad (28)$$

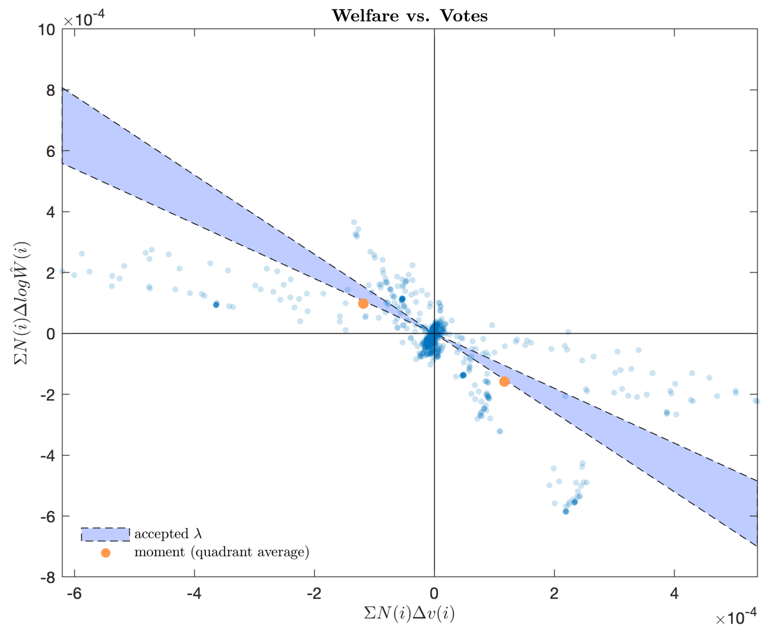
The moment inequalities in (28) form the basis of ours procedure to estimate the β_k's and λ. We compute confidence sets for these parameters following Andrews and Soares (2010). We construct the subsets of deviations $\mathcal{N}_e$ for each moment by grouping perturbations that help identify upper and lower bounds for our various parameters. This procedure identifies an admissible set of parameter values such that, for a given confidence level, we are unable to reject the hypothesis that the data was generated by a parameter vector within the set.

6.2.2 Pareto Frontier Between Votes and Welfare

To understand how we form our moment inequalities and how they identify our parameters, we discuss here a restricted case in which the planner's objective function depends exclusively on two unknown parameters: β_0, which captures a utilitarian component of preferences (and here we normalize to 1) and λ, which determines the role of votes in the planner's preferences. Figure 5 plots our the perturbations on a two-dimensional graph 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 ($\sum_{i=1}^{J} N\left(i\right)\ln\hat{W}\left(i;\mathbf{d}^n\right)$). We form four moments by grouping the perturbations that fall in the four quadrants displayed in the figure. The lower-left quadrant groups perturbations that yield lower utilitarian welfare and lower votes than the actual CHSR proposal (corresponding to the $(0,0)$ point). These perturbations trivially satisfy our inequality conditions and do not provide

is unknown to the planner at the time at which the optimal CHSR design was chosen; this is guaranteed by the assumption that $\epsilon\left(\mathbf{d}^n\right)$ 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 $\epsilon\left(\mathbf{d}^n\right)$ is sufficiently low such that the Law of Large Numbers applies.

Figure 5: Utilitarian Welfare vs. Votes



any information to identify our parameters. Similarly, perturbations in the upper-right quadrant imply greater welfare and higher votes. These perturbations do not suggest a trade-off, violate our inequality conditions, and can only be rationalized by the expectational errors.

In contrast, the perturbations in the upper-left and lower-right quadrants are informative because they suggest a trade-off between welfare and votes. The orange dots represent the average across all perturbation within each quadrant, on which our moments are built. An admissible parameter λ is such that the moment condition in (28) is satisfied. The admissible values for λ (relative to β_0) are such that $y(e) + \lambda x(e) \leq 0$, where $(x(e), y(e))$ are the coordinates of the two moments $e = 1, 2$ (orange dots). Visually, this corresponds to drawing a line through the origin with slope $-\lambda$ such that the two orange dots fall underneath. The set of admissible parameters λ is represented by the purple area on the graph.

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

6.2.3 Parameter Estimates

Table 6 shows estimates of the β's and λ across specifications that allow for covariates that include: population density, residents' average wage, and shares of college educated and non-white residents. In the estimation, the support of the parameters is not restricted and each parameter β_k

Table 6: Planner's Preferences Estimates

Observable	Pareto weight parameters β and λ					
	(1)	(2)	(3)	(4)	(5)	(6)
Density		[0.00, 1.13]				[0.00, 0.16]
Wages			[-0.15, 0.26]			[-0.15, 0.26]
Share college				[-0.13, 0.54]		[0.00, 0.59]
Share non-white					[-0.33, 0.54]	[-0.26, 0.13]
Votes	[0.46, 2.29]	[0.58, 2.89]	[0.58, 1.53]	[0.62, 1.68]	[0.37, 2.48]	[0.61, 1.68]
Constant	[1.00, 1.00]	[0.98, 1.00]	[0.99, 1.00]	[0.99, 1.00]	[0.98, 1.01]	[0.99, 1.01]

or λ 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 $\Omega(i)$ equals 1. Each parameter β_k can thus be interpreted as capturing the impact of a one-standard deviation increase on the covariate relative to the average Pareto weight.[41] 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 6 corresponds to the case discussed in Figure 5. Columns (2) to (5) allow for one extra unknown parameter β_k at a time, while column (6) reports a full estimation with all the β_k's.

The estimation reveals a strictly positive preference for votes across the various specifications. The other covariates do not appear significantly different from 0, as 0 belongs to the 95% confidence set for each parameter. However, our estimates from column (6) suggest a positive preference for census tracts with higher population density and a larger share of college-educated residents.[42] In what follows, we use column (6) as our baseline specification.

Figure A.4 in the Appendix shows the distribution of the total weights $\Omega(i) + \lambda V(i)$ affecting each tract i implied by the centroid of our confidence set. Overall, these estimates imply large variance in these Pareto weights, so that the planner is far from the utilitarian benchmark.

6.3 Counterfactual Optimal Designs

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. More specifically, we compare the actual proposed CHSR design to that preferred by an *apolitical* planner whose objective function does not depend on votes (i.e., $\lambda = 0$), and to that preferred by a *utilitarian* planner that does not attach different Pareto weights to different tracts on the basis of their demographic characteristics (i.e., $\beta_k = 0$ for some

[41] We also standardize the variable $V(i)$. As a result, the constant we estimate is $\left(\beta_0 + \lambda\overline{V}\right)$ and captures the overall utilitarian motive of the planner, including what is inherited from the voting block.

[42] Figure A.3 in Appendix E displays the topology of the confidence set from column (6) by showing pairwise projections for each combination of the parameters β_k and λ.

or all of the k).[43]

Since it is infeasible to compute counterfactual optimal designs for a large set of parameter values, we use the centroid of our 95% the confidence set. This centroid yields values of the β_k associated to average wages and to the share of non-white residents that are very close to zero; hence, for simplicity in the exposition, we set these parameter to zero and use the centroid of the confidence set conditional on these two zero parameters: $\beta = [0.00, 0.00, 0.43, 0.00]'$ for [density, wages, share college, share non-white] and $\lambda = 1.25$.[44]

Apolitical Planner To explore what the optimal CHSR design would have been if the planner had no political concerns, we define the preferences of an *apolitical* planner to be such that the Pareto weights $\Omega(i)$ are as determined by the centroid of our confidence set but $\lambda = 0$, thereby eliminating the planner's incentives to assign higher real income to locations in higher voting elasticity areas.

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 politically biased in favor of the CHSR (have a low voting elasticity due to a strong political preference for the high-speed rail). 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, Anaheim) and San Diego (Escondido) reallocate towards their corresponding metropolitan areas.

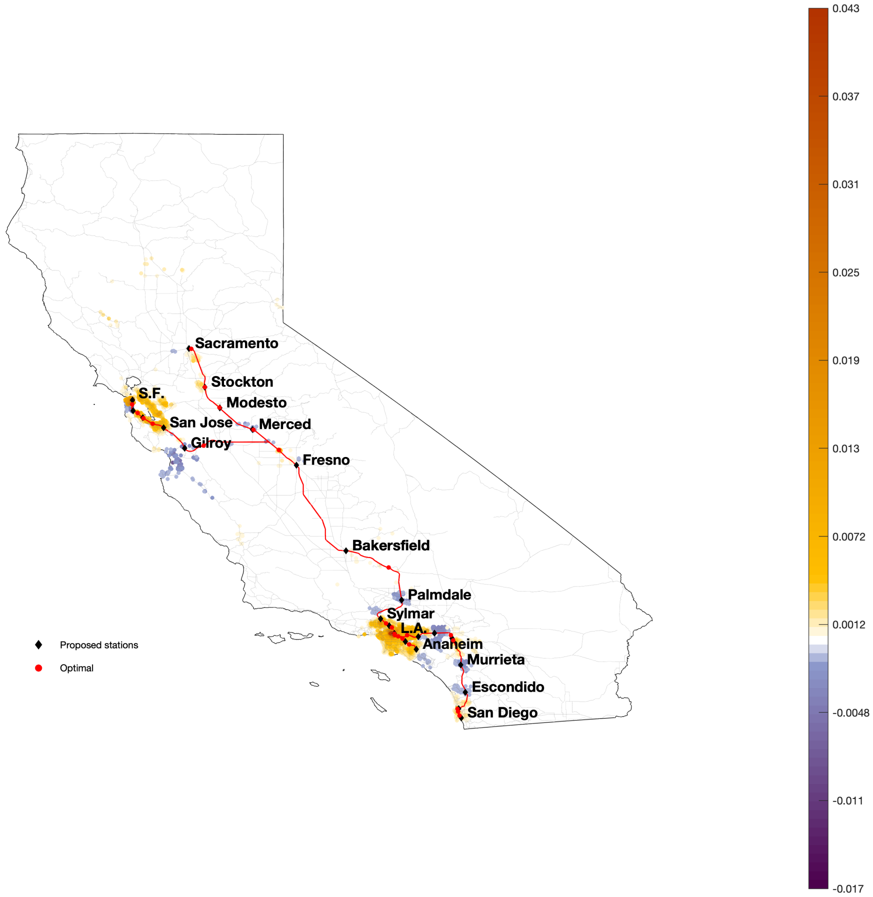
Table 7 summarizes the impact of the counterfactual CHSR designs. Relative to the original design, aggregate real income increases and the aggregate vote declines. Locations with high density or low voting elasticity experience the largest gains.[45] Appendix Figure A.6 zooms on the Los Angeles region, displaying the real income changes (left panel) and the voting elasticity (right panel). The relocation of stations towards more central 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.

[43] The optimal station placement problem is highly non-convex due to substitutabilities and complementarities across any station locations, as well to the sigmoidal function that defines the voting probabilities. We use a three-step optimization technique 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 highest possible outcome when each station is moved individually within a range of 10 km from the proposed optimum along the CHSR outline. Appendix Figure also A.5 shows aggregate welfare around each station in the optimal apolitical design, suggesting that our procedure selects a reasonable candidate for an optimum. Appendix E.2 provides additional details.

[44] Although the centroid of β_{density} is close to zero, the $\Omega(i)$ are per-capita Pareto weights, still implying a strong utilitarian motive.

[45] Additional details are provided in Table A.9 in Appendix E. Moving each station to its optimal location while keeping all other stations at the proposed plan usually raises the planner's objective function. This is not always the case, however, due to complementarities in simultaneously reallocating several stations to the optimum.

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.

Utilitarian Planner We define the utilitarian planner as having no preferences for votes ($\lambda = 0$) and constant Pareto weights ($\Omega(i) = 1$). Unsurprisingly since the utilitarian and the apolitical planners only differ in the positive weight that the latter puts on the share of college educated residents, the utilitarian design does not differ significantly from the apolitical one. In both cases, the counterfactual optimal design moves several stations closer to the core of the large metropolitan areas, where density is higher and the voting elasticity lower. Some stations do differ, however, due to a redistribution from high to low college share areas, as is illustrated in the case of Riverside in Appendix Figure A.7. The optimal utilitarian design reallocates the Riverside towards L.A. The left panel displays the distribution of real income changes implied by this relocation, and the right panel shows the standardized distribution of the college share. The main beneficiaries from eliminating redistribution from the planner's objective are areas with a low share of college-educated residents, while the high-college educated areas in the immediate surroundings of UC Riverside are negatively affected.

Table 7: Apolitical planner, Statistics

Utilitarian Component Change	0.18%			
[min max]	[-1.70%, 4.25%]			
Aggregate Vote Change	-0.18%			
[min max]	[-14.15%, 3.79%]			
ΔReal Income by Quartile	1st	2nd	3rd	4th
Density	0.01%	0.06%	0.16%	0.46%
Wages	0.27%	0.14%	0.11%	0.16%
Share college	0.19%	0.17%	0.13%	0.20%
Share non-white	0.06%	0.09%	0.19%	0.34%
Voting elasticity	0.39%	0.16%	0.10%	0.07%

Notes: Aggregate real income change is computed using the counterfactual Pareto weights under the normalization that the population-weighted mean $E[\Omega(i)] = 1$ and provided in basis points. The rows [min max] show the support across locations. The bottom five rows show the average real income change for each quartile of each covariate.

7 Conclusion

In this paper we study the determinants of transportation infrastructure projects, and specifically how important are considerations beyond private real-income gains in shaping these 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. We combine this voting data with a range of quantitative spatial models used to compute expected real income impacts of the CHSR at the time of the 2008 vote.

The empirical analysis reveals that voters did respond to the expected real-income impacts of the CHSR. However, we find that political considerations, proxied by party affiliation and by voters' opinions in other propositions, shaped the vote and the spatial distribution of welfare changes to a larger extent than real-income considerations. In fact, our analysis suggests that the CHSR project would still have won the vote even for a range of costs so large that the project would have led to net aggregate income losses.

We then turn to estimating the drivers of the CHSR design. We posit that the observed CHSR design represents the optimal choice of a hypothetical planner, whose preferences we estimate by revealed choice using the moment inequalities defined by unchosen designs. We find strong planner's preferences for winning votes, and some preferences for densely populated and college-educated areas. 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.

References

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. (2019): "Chinese roads in India: The effect of transport infrastructure on economic development," *Manuscript, Univ. North Carolina, Chapel Hill.*

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.

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

Becker, S. O., T. Fetzer, and D. Novy (2017): "Who voted for Brexit? A comprehensive district-level analysis," *Economic Policy*, 32, 601–650.

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 (2020): "Non-random exposure to exogenous shocks: Theory and applications," Tech. rep., National Bureau of Economic Research.

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

Brueckner, J. K. and H. Selod (2006): "The political economy of urban transport-system choice," *Journal of Public Economics*, 90, 983–1005.

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 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/.*

——— (2023a): "2023 Benefit Cost Analysis Presentation," *https://hsr.ca.gov/wp-content/uploads/2023/05/2023-Benefit-Cost-Analysis-Report-Presentation-v1-A11Y.pdf.*

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

De Borger, B. and S. Proost (2012): "A political economy model of road pricing," *Journal of Urban Economics*, 71, 79–92.

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

Deacon, R. T. and F. Schläpfer (2010): "The spatial range of public goods revealed through referendum voting," *Environmental and Resource Economics*, 47, 305–328.

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.

Donaldson, D. (2018): "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review*, 108, 899–934.

Donaldson, D. and R. Hornbeck (2016): "Railroads and American Economic Growth: A Market Access Approach," *The Quarterly Journal of Economics*, qjw002.

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.

Faber, B. (2014): "Trade integration, market size, and industrialization: evidence from China's National Trunk Highway System," *The Review of Economic Studies*, rdu010.

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

Glaeser, E. L. and G. A. Ponzetto (2018): "The political economy of transportation investment," *Economics of Transportation*, 13, 4–26.

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

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.

Hayakawa, K., H. Koster, T. Tabuchi, and J.-F. Thisse (2021): "High-speed rail and the spatial distribution of economic activity: Evidence from Japan's Shinkansen," .

Hendren, N. (2020): "Measuring economic efficiency using inverse-optimum weights," *Journal of public Economics*, 187, 104198.

Hicks, R., H. V. Milner, and D. Tingley (2014): "Trade Policy, Economic Interests, and Party Politics in a Developing Country: The Political Economy of CAFTA-DR1," *International Studies Quarterly*, 58, 106–117.

Holian, M. J. and M. E. Kahn (2015): "Household demand for low carbon policies: Evidence from California," *Journal of the Association of Environmental and Resource Economists*, 2, 205–234.

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

JACOBS, B., E. L. JONGEN, AND F. T. ZOUTMAN (2017): "Revealed social preferences of Dutch political parties," *Journal of Public Economics*, 156, 81–100.

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.

KENDALL, C. AND J. MATSUSAKA (2021): "The common good and voter polarization," Tech. rep., Mimeo, University of Southern California.

KNIGHT, B. (2005): "Estimating the value of proposal power," *American Economic Review*, 95, 1639–1652.

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.

REDDING, S. J. AND E. ROSSI-HANSBERG (2017): "Quantitative Spatial Economics," *Annual Review of Economics*, 9, 21–58.

REDDING, S. J. AND M. A. TURNER (2015): "Transportation Costs and the Spatial Organization of Economic Activity, vol. 5 of Handbook of Regional and Urban Economics," .

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.

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.

WU, X. AND B. CUTTER (2011): "Who votes for public environmental goods in California?: Evidence from a spatial analysis of voting for environmental ballot measures," *Ecological Economics*, 70, 554–563.

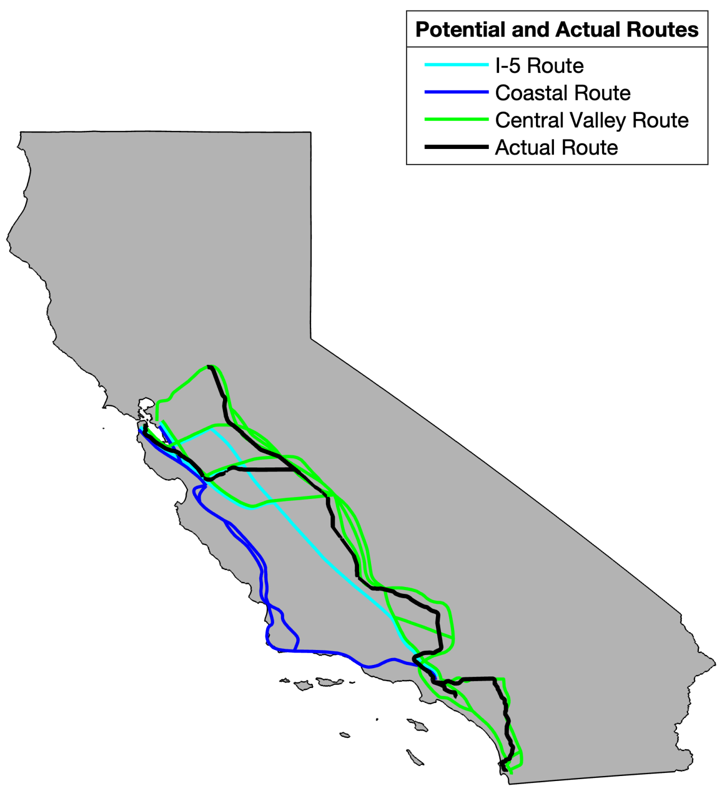
ZHENG, S. AND M. E. KAHN (2013): "China's bullet trains facilitate market integration and mitigate the cost of megacity growth," *Proceedings of the national academy of sciences*, 110, E1248–E1253.

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 Tables and Figures

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.1: Travelers by Mode and Potential Gains across Routes

		Public Transit	Car	Air
Travelers by Mode	Commute	5.3%	90.3%	0%
	Leisure	1.8%	96.4%	1.8%
	Business	2.2%	88.4%	9.4%
Time-Saving Routes	Share of routes	65%	61%	30%
	Gain (min)	327	127	33
Cost-Saving Routes ('08 CHSR Price)	Share of routes	62%	39%	40%
	Gain ($)	43	6	96
Cost-Saving Routes (2X Ticket Price)	Share of routes	44%	0%	39%
	Gain ($)	33	-	45

Source: ACS and CAHTS for fraction of commuters and leisure or business travelers, respectively. The fraction of commuters adds up to less than 100% because the complement includes walking or biking. Those fractions are 0% for long-distance (>50 miles) leisure or business trips. To compute CHSR trips and compare them with car or airplane travel, we compute the fastest route that combines road and CHSR. When comparing to car, we consider all origin-destination pairs. When comparing to air travel, we only consider routes were traveling by plane (including driving to airports) is faster than driving all the way. When comparing to public transit (defined as bus or rail), we assume that CHSR users complement their trip with public transit and that public-transit travelers use the rail station closer to origin and destination tracts. The average time and cost gains are computed among routes that are faster or cheaper, respectively. See Appendix D for data construction on travel and cost gains.

Table A.2: 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 $\hat{\rho}_C = 0.75$.

Table A.3: Gain by Type of Travel

	Baseline	Full	Pessimistic	+Car
Commute	0.29%	0.23%	-0.14%	0.44%
Leisure	0.03%	0.03%	-0.04%	0.15%
Business	-	0.38%	-	-
Total	0.32%	0.65%	-0.18%	0.60%

Note: in cases "Baseline," "Pessimistic," and "+Car," this table splits the aggregate welfare change in (9) into the commuting component $\hat{W}_C(i)$ in (11) and the leisure component $\hat{P}_L(i)$ in (12). By construction of (10), these effects are log-additive. The remaining components of (10), are constant in these cases. In the "Full" case, 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.

Table A.4: Heterogenous Gains from CHSR

	Baseline	Full	Pessimistic	+Car
	(1)	(2)	(3)	(4)
dist. to station	-0.00128***	-0.00120***	-0.000984***	-0.00527***
	(0.0000514)	(0.0000717)	(0.0000438)	(0.000121)
% public transit	0.0790***	0.0740***	0.0678***	0.0450***
	(0.00123)	(0.00172)	(0.00105)	(0.00289)
% car	0.00539***	-0.000240	0.00390***	-0.0121***
	(0.000743)	(0.00104)	(0.000634)	(0.00175)
commute time	0.0000297***	0.0000768***	0.0000181***	0.000315***
	(0.00000374)	(0.00000521)	(0.00000318)	(0.00000877)
LA fixed effect	0.00296***	0.00310***	0.00257***	-0.000135
	(0.000105)	(0.000146)	(0.0000892)	(0.000246)
SF fixed effect	0.0105***	0.0141***	0.0100***	0.00679***
	(0.000312)	(0.000435)	(0.000266)	(0.000731)
R2	0.734	0.589	0.745	0.359
N	7866	7866	7866	7866

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

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 across all residents. The LA and SF fixed effects correspond to dummies for whether the census tract is within LA county or SF county, respectively.

Table A.5: Estimates of Voting Equation, Alternative Weighting and Sample Selection Criteria

	Baseline		Weighting - Num. Votes		Weighting - Participation		Selection - $\geq$ 5km line	
Inst. Var.:	Random Station (1)	Random Path (2)	Random Station (3)	Random Path (4)	Random Station (5)	Random Path (6)	Random Station (7)	Random Path (8)
$log(\hat{W}_{19})$	19.45^a	22.68^a	27.09^a	29.93^a	22.31^a	25.51^a	20.73^a	23.73^a
	(1.68)	(1.80)	(2.22)	(2.38)	(3.75)	(3.31)	(2.29)	(2.54)
Log-odds Dem. Sh.	0.38^a	0.39^a	0.41^a	0.41^a	0.43^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	2.43^a	2.41^a	2.36^a	2.35^a	2.20^a	2.19^a	2.46^a	2.45^a
	(0.05)	(0.05)	(0.06)	(0.06)	(0.16)	(0.17)	(0.06)	(0.06)
Transp.: Prop. 1b	0.81^a	0.80^a	0.86^a	0.86^a	0.90^a	0.90^a	0.84^a	0.83^a
	(0.05)	(0.05)	(0.05)	(0.05)	(0.09)	(0.10)	(0.05)	(0.06)
Sh. non-White	-0.18^a	-0.19^a	-0.26^a	-0.27^a	-0.23^a	-0.24^a	-0.29^a	-0.30^a
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Sh. College	0.73^a	0.72^a	0.72^a	0.71^a	0.69^a	0.69^a	0.71^a	0.70^a
	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)
Sh. Under 30	0.18^a	0.18^a	0.21^a	0.20^a	0.21	0.21	0.26^a	0.26^a
	(0.03)	(0.03)	(0.03)	(0.03)	(0.16)	(0.16)	(0.03)	(0.03)
Log. Dist. Station	-0.01	-0.00	0.01^c	0.01^b	-0.01	-0.00	-0.03^a	-0.02^b
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Log. Dist. Rail	0.01^a	0.01^a	0.01^b	0.00^c	0.01	0.01	0.03^a	0.03^a
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
F-stat	574	286	743	543	241	351	294	261
Num. Obs.	7861	7861	7861	7861	7861	7861	5010	5010

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.

Table A.6: Estimates of Voting Equation, Alternative Models, All Covariates

Model:	Baseline		Full		Pessimistic		+ Car	
Inst. Var.:	Random Station (1)	Random Path (2)	Random Station (3)	Random Path (4)	Random Station (5)	Random Path (6)	Random Station (7)	Random Path (8)
$log(\hat{W}_{19})$	19.45^a	22.68^a	17.51^a	18.92^a	41.62^a	47.20^a	19.17^a	23.15^a
	(1.68)	(1.80)	(1.48)	(1.54)	(3.55)	(3.72)	(1.46)	(2.00)
Log-odds Dem. Sh.	0.38^a	0.39^a	0.39^a	0.39^a	0.38^a	0.39^a	0.40^a	0.41^a
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Environ.: Prop. 10	2.43^a	2.41^a	2.43^a	2.42^a	2.42^a	2.41^a	2.46^a	2.45^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	0.90^a	0.91^a
	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.05)	(0.04)	(0.05)
Sh. non-White	-0.18^a	-0.19^a	-0.17^a	-0.18^a	-0.18^a	-0.19^a	-0.18^a	-0.19^a
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Sh. College	0.73^a	0.72^a	0.73^a	0.73^a	0.72^a	0.72^a	0.75^a	0.74^a
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Sh. Under 30	0.18^a	0.18^a	0.17^a	0.17^a	0.18^a	0.18^a	0.16^a	0.16^a
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Log. Dist. Station	-0.01	-0.00	-0.01^c	-0.01	-0.01	-0.00	0.01^a	0.03^a
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
Log. Dist. Rail	0.01^a	0.01^a	0.02^a	0.01^a	0.01^a	0.01^a	0.01^a	0.01^a
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
F-stat	574	286	610	311	599	295	394	206
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.7: Estimates of Voting Equation, Alternative Models, Political Covariates Only

Model:	Baseline		Full		Pessimistic		+ Car	
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)
$log(\hat{W}_{19})$	17.89^a	23.57^a	15.94^a	20.62^a	39.09^a	49.63^a	15.56^a	20.92^a
	(1.65)	(1.98)	(1.49)	(1.77)	(3.55)	(4.10)	(1.33)	(2.02)
Log-odds Dem. Sh.	0.30^a	0.30^a	0.30^a	0.30^a	0.30^a	0.30^a	0.31^a	0.32^a
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Environ.: Prop. 10	1.16^a	1.11^a	1.16^a	1.11^a	1.16^a	1.12^a	1.16^a	1.11^a
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Transp.: Prop. 1b	1.54^a	1.52^a	1.57^a	1.56^a	1.54^a	1.52^a	1.62^a	1.63^a
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
F-stat	713	317	732	333	738	328	454	231
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.8: Political vs. Real Income Determinants of the Vote, Political Variables only

		Variance Decomposition of $\Delta U(i)$		Favorable Vote if Voters Only Consider...		Cost of Swaying 1% of Vote	
	θ_V	$\sigma_{\Delta \ln a}$	$\sigma_{\Delta \ln \hat{W}}$	$\hat{W}$	$\hat{a}$	50p	90p
Baseline	17.9	2.3%	0.4%	51.3%	51.4%	0.4%	0.6%
Full	15.9	2.6%	0.4%	52.5%	50.3%	0.5%	0.7%
Pessimistic	39.1	1.0%	0.2%	48.0%	54.5%	0.2%	0.3%
+ Car	15.6	2.7%	0.6%	52.2%	50.5%	0.5%	0.7%

Note: The θ_V reported corresponds to IV estimate using the random routes instrument from Table Table A.7.

B 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 $(s = N)$.

B.1 Preferences

When the CHSR status is s, the utility U_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:

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

subject to the budget constraint:

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

Expression A.1 indicates that consumers derive utility from the amenities of their place of residence $B(i, s)$ and from the consumption of tradeable commodities C, housing H_C, and leisure trips T_L, with Cobb-Douglas shares that can be location-specific. This heterogeneity captures in a reduced-form way the fact that workers in different tracts in California may have different spending patterns on leisure trips and housing, for example due to different demographic characteristics. The amenity term $B(i, s)$ may respond endogenously to the local density of economic activity as detailed below. These workers face disutility $d_C(i, j_C, m_C, s)$ from daily commuting travel . The utility that they derive from leisure trips depends negatively on time travelled $d_L(i, j_L, m_L, s)$ and positively on the quality of the destination visited, equal to a composite of an exogenous origin-destination component $q_L(i, j_L)$ (capturing, for example, that residents of some locations may on average be more likely to have relatives in some other specific location) and the destination-specific amenity $B(j_L, s)$. The last two terms of (A.1), $\varepsilon_\omega^C(j_C, m_C)$ 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 ($k = L$):

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

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

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

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

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

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

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

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

B.2 Indirect Utility and Welfare

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

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

Each resident ω makes discrete choices of destination and transport mode for both commuting and leisure to maximizes indirect utility. These choices are represented by the quadruplet $\{j_C, j_L, m_C, m_L\}$. Destinations are chosen from the set of tracts $\mathcal{J}$ while the set of transport modes 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:

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

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

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

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

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

$$V\left(i, s\right)=\frac{B\left(i, s\right)}{r\left(i, s\right)^{\mu_H(i)}} \frac{W_C\left(i, s\right)}{P_L\left(i, s\right)^{\mu_L(i)}} \tag{A.9}$$

where $W^C\left(i, s\right)$ captures average income net of commuting costs of residents of i,

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

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

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

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

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

B.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 output according to the Cobb-Douglas production function:

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

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:

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

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

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

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

where $p_B(j,j_B,m_B,s)$ is the monetary cost per roundtrip business trip. Because conditional on floor space and labor there are decreasing returns to the number of trips, we can solve for the number of trips T_B, plug them back into the term within the expectation, and then integrate over realization of idiosyncratic business shocks using standard properties of the extreme value distribution defined in (A.13).

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

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

where $\Omega(j,s)$ is an endogenous TFP term that depends on both the TFP of the location $A(j,s)$

and the distribution of business travel opportunities,

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

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

B.4 Travel Choices

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

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

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

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

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

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

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

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

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

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

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

through mode m_B is:

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

B.5 Spillovers

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

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

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

where

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

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

B.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. When $s = N$ (i.e., when the CHSR is not available), we assume the chosen route to be the fastest through the network within each mode (bike, car, public transit, airplane) and assign a monetary cost to the corresponding route.

When the CHSR is available, travelers choose (within each mode) between their preferred pre-CHSR route or using the CHSR. In our baseline analysis, the CHSR option is only available as a substitute to traveling via public transit (for commuters) or via air (for business and leisure travelers). In a robustness exercise, the CHSR is also a substitute for road travel via car for all travelers.

In making this decision of whether to travel through the CHSR, travelers trade off time and costs on whether or not to use the CHSR. Specifically, as implied by the indirect utility (A.6), when the CHSR is available within mode m ($s = Y$), a commuter chooses the time-cost pair $(\tau_C(i,j,m,Y), p_C(i,j,m,Y))$ between the pre-CHSR fastest time and cost $(\tau_C(i,j,m,N), p_C(i,j,m,N))$ and the CHSR time and cost $\left(\tau^{CHSR}(i,j,m), p^{CHSR}(i,j,m)\right)$, by solving, for each (i,j,m):

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

Similarly, as implied by the indirect utility (A.6) and by the definition of business productivity (A.16), a leisure or business traveler chooses $(\tau_k(i,j,m,Y), p_k(i,j,m,Y))$between $(\tau_k(i,j,m,N), p_k(i,j,m,N))$ and $\left(\tau^{CHSR}(i,j,m), p^{CHSR}(i,j,m)\right)$ by solving:

$$(\tau_k(\cdot,Y), p_k(\cdot,Y)) = \arg \min_{(\tau,p)\in\{(\tau_C(\cdot,N),p_C(\cdot,N)),(\tau^{CHSR},p^{CHSR})\}} p\tau^{\rho_k} \tag{A.27}$$

for $k = L, B$. Therefore, when $s = Y$ (i.e., when the CHSR is available), the times and costs $\tau_k(i,j,m,Y)$ and $p_k(i,j,m,Y)$ may be different for commuting and for leisure or business travel because, the monetary costs from using the CHSR enter asymmetrically in indirect utility depending on the travel purpose. I.e., within the same origin-destination, travelers may use CHSR for commuting and not for leisure, or vice-versa.

B.7 Equilibrium Conditions

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

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

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

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

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

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

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

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

for some constant $\kappa_\Omega(j)$ that is a function of $\mu_B(j)$ and $\mu_{H_Y}(j)$.

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

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

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

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

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

B.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). 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 B.6. When $s=Y$, then disposable income also changes with the tax rate in a common away across locations:

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

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

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

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

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

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

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

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

the change in pre-tax income is

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

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

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

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

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

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

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

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

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

where, from (A.16),

$$\hat{\Omega}(j) = \hat{A}(j) \left(\sum_{j_B \in \mathcal{J}} \sum_{m \in \mathcal{M}_B} \lambda_B(i,j_B,m_B,N) \left(\frac{\hat{A}(j_B)}{\hat{p}_B(j,j_B,m_B)\, \hat{d}_B(j,j_B,m_B)}\right)^{\theta_B \mu_B(j)}\right)^{\frac{1}{\theta_B}}. \tag{A.40}$$

From (A.23), the agglomeration component of TFP changes according to:

$$\hat{A}(j) = \left(\sum_{k \in \mathcal{J}} \frac{\tilde{N_Y}(k,N)/H(k)}{\sum_{k' \in \mathcal{J}} e^{-\rho_A^{\text{spillover}} \tau^{min}(j,k',N)} \tilde{N_Y}(k',N)/H(k')} e^{-\rho_A \tau^{min}(j,k,B)} \hat{\tilde{N}}_Y(k)\right)^{\gamma_A^{\text{spillover}}} \tag{A.41}$$

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

$$\hat{\tilde{N}}_Y(j) = \sum_i \left(\frac{\lambda_C(i,j,N)\, N_R(i)}{\tilde{N_Y}(j,s)}\right) \hat{\lambda_C}(i,j). \tag{A.42}$$

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

Welfare Changes The welfare changes in (10) to (12) follow from (A.9)-(A.11), where the endogenous amenities component $\hat{B}(i)$ satisfies a similar equation to $\hat{A}(j)$ in (A.41).

C Data Appendix

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. 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. 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) 2012-2016. The American Community Survey reports tract-to-tract data on commuting by transport mode as a part of the Census Transportation Planning Products.[2] 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 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.

[2]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.

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. 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.[3]

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

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

Construction of Additional Variables Using the previous sources we construct additional variables needed for the implementation of the model. We construct disposable income $y(i, j_C, N)$ in (A.5) using its definition as the sum of labor income and locally owned land rents. We construct the land rent component of income as $\eta(i) r(i, N) \equiv \frac{\eta_R(i) r(i) H_C(i,N)}{N_R(i)}$, where $\eta_R(i)$ is the share of homes that is owner-occupied from ACS and $r(i) H_C(i, N)$ are residential home values from ACS transformed to annualized rent-equivalent values. We measure the share of land used for residential purposes, $H_C(i, N) / H(i)$ entering in (A.32), using Zillow's ZTRAX data. We construct the disposable income $I(i, j_C, m_C, N)$ defined in (A.4) using $y(i, j_C, N)$, the round-trip commuting costs described in the next section, and $T_C = 250$ commuting trips throughout the year.

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

[3] 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.

D 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.[4] 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.[5] 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.[6] 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 for car trips, whereas we tend to somewhat under-predict 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 via bus or via rail for each origin-destination pair. The route via rail

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

assume that travelers use the train stations nearest to the origin and destination tracts as long as the distance between these stations is less than between the original tracts (otherwise assuming the bus is used). 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), along with the previously calibrated car speed from stations and centroids and transfer time of 17 minutes from car to rail stations. Fastest biking routes on the road network are constructed 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, using flight times from Google Maps, and including a total transfer time of 90 minutes at the origin airport.

We define the CHSR paths differently depending on whether it is used as a direct substitute to public transit or to car. When it is a substitute to public transit, the path and time are defined similarly to the fastest time via rail (i.e., travelers use the nearest station, with driving times from centroids to stations). When it is a substitute to car or to air travel, it is defined similarly to the fastest time via air (i.e., travelers can drive to their preferred airport). In both cases we use the transfer time at stations calibrated for public transit.

When studying the actual CHSR design, we use planned speeds between contiguous station of the CHSR network from the November 2008 High Speed Rail Authority's Business Plan (California High Speed Rail Authority, 2008). The resulting pairwise travel times between all stations closely match those reported between major stations. When studying optimal counterfactual CHSR designs by the planner, we maintain the speeds on each segment.

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, 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 17c per minute traveled at the average rail speed of 50 miles per hour. The cost of air travel between each pair of airports corresponds to the average ticket price according to the Bureau of Transportation Statistics (average one-way ticket cost of $151 across routes).[9] For the ticket price of the CHSR, we rely on the $55 ticket price from LA to San Francisco from the 2008 Business Plan (California High Speed Rail Authority, 2008) as well as an update of this number to $110 in the 2022 plan (California High Speed Rail Authority, 2022). We use these ticket prices for LA-SF in the baseline and pessimistic scenarios, and project these costs to the remaining network using the

[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 .

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

E Appendix to the Planning Problem

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

E.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

$$\sum_{i=1}^{J} \left[\beta_0 + \sum_{k=1}^{K} \beta_k Z_k(i) + \lambda V(i) \right] N(i) \Delta \ln \hat{W}(i, \mathbf{d}^n) - \epsilon(\mathbf{d}^n) \leq 0. \tag{A.43}$$

We generate these perturbations with the aim of identifying upper and lower bounds on the parameters β and λ. For each covariate $Z \in \mathcal{Z} = \{Z_1, \ldots, Z_K, V\}$, 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).

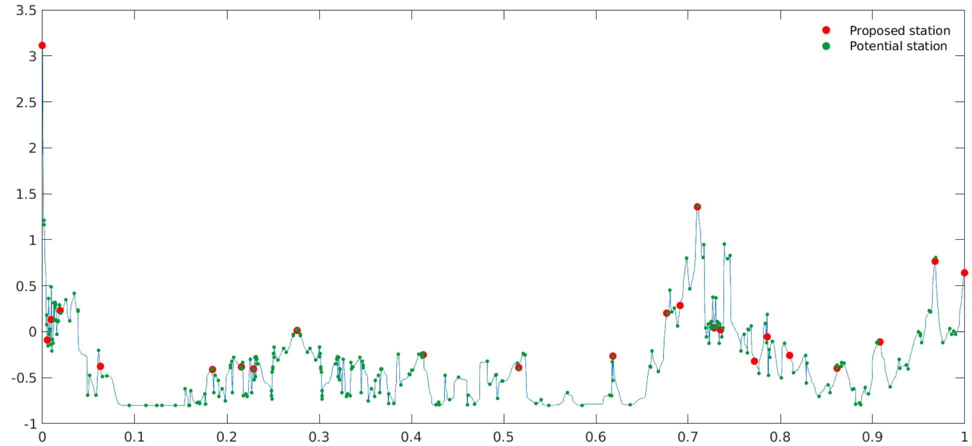
For each covariate Z, we then define a set of perturbations $\mathcal{N}(Z)$ by moving each station $s = 1 \ldots N_s$ individually by a certain number of steps $k = -N_{steps} \ldots -1$ (towards San Francisco) and $k = 1 \ldots N_{steps}$ (towards San Diego) among the set of potential locations. In our baseline specification, we choose $N_{steps} = 2$, which, for a number of stations $N_s = 24$, 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, average wages, the share of college-educated residents, the share of non-white residents and votes) we obtain $5 \times 96 = 480$ perturbations to identify 6 parameters ($\beta_0, ..., \beta_K$and λ).

Moment Conditions To lighten notation, we rewrite the welfare inequality (A.43) for perturbation j as

$$m(n; \gamma) \equiv \sum_k \gamma_k X_k(n) - \epsilon(n) \leq 0,$$

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 $\mathrm{dist}(i,i')$ 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.

where $\gamma = (\beta_0, \beta_1, \ldots, \beta_K, \lambda)$ is the vector of parameters we want to estimate and

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

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

$$\hat{m}_e(\gamma) = \frac{\sum_{n \in \mathcal{N}_e} m(n;\gamma)}{|\mathcal{N}|}$$

and associated standard deviation

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

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

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

We construct a 95% confidence set over parameter γ as

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

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 $\Omega(i)$ 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

  $$z\left(\gamma^{sphere}\right) = \sum_i \sum_k N(i)\, \gamma_k^{sphere} Z_k(i).$$

  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.

E.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 $\gamma = (\beta;\lambda)$, defined as

$$\begin{aligned}\epsilon(\mathbf{d};\gamma) = &\sum_i \Omega(i;\beta) N(i) \Delta U(i;\mathbf{d}) + \lambda \sum_i N(i) v(i;\mathbf{d}) \\ &- \mathbb{E}\left[\sum_i \Omega(i;\beta) N(i) \Delta U(i;\mathbf{d}) + \lambda \sum_i N(i) v(i;\mathbf{d}) \mid \mathcal{I}^{2008}\right].\end{aligned}$$

When computing optimal station locations for a counterfactual set of planner weight γ, $\epsilon(\mathbf{d};\gamma)$ is unknown. When optimizing, we assume that the uncertainty driving the forecast error is orthogonal to 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, i.e., that

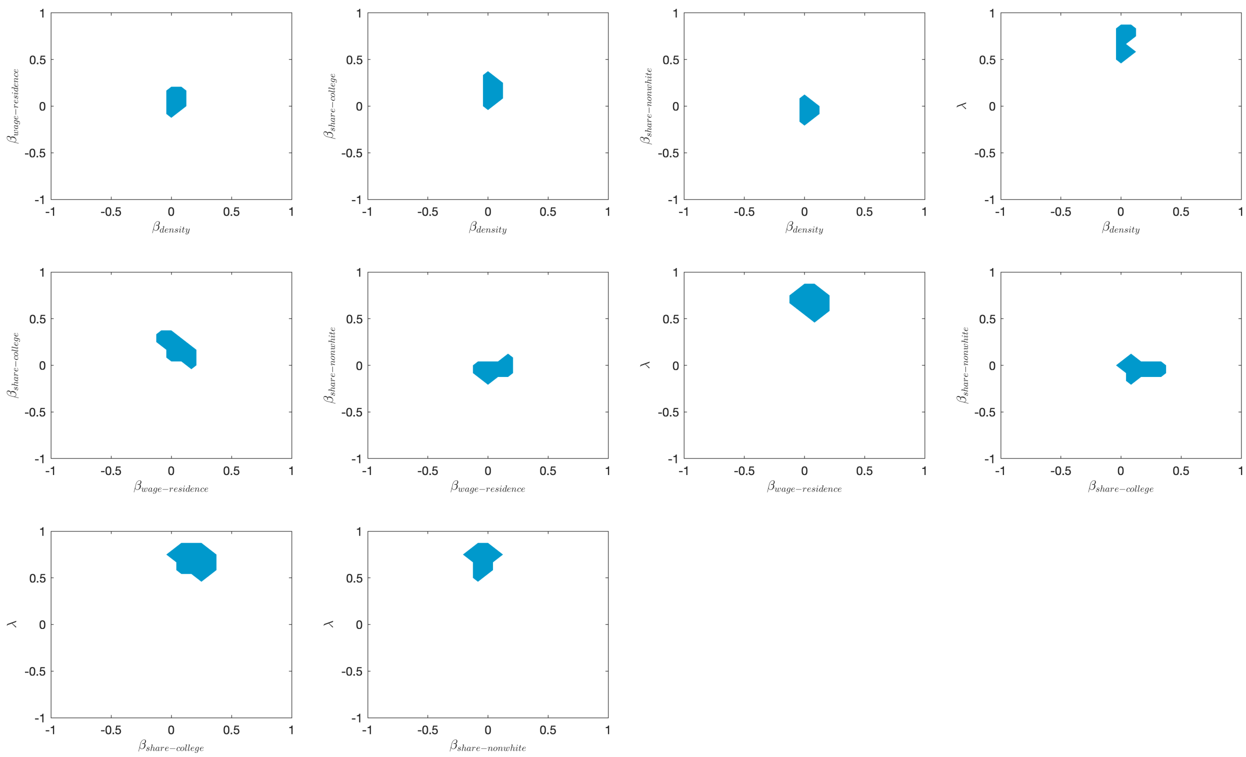
$$\sum_i \Omega\left(i;\hat{\beta}\right) N(i) \Delta U(i;\mathbf{d}) + \hat{\lambda} \sum_i N(i) v(i;\mathbf{d}) - \epsilon(\mathbf{d};\hat{\gamma}) \leq 0.$$

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

$$\epsilon(\mathbf{d};\hat{\gamma}) \equiv \max\left(0, \sum_i \Omega\left(i;\hat{\beta}\right) N(i) \Delta U(i;\mathbf{d}) + \hat{\lambda} \sum_i N(i) v(i;\mathbf{d})\right).$$

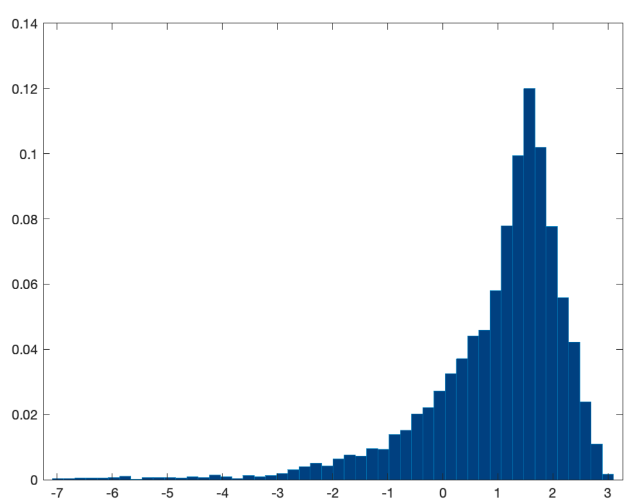
E.3 Additional Tables and Figures for the Planning Problem

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



Notes: Pairwise projections of the 95% confidence set from column (6) in Table 6 in spherical normalization (see Appendix E.1). In blue are denoted all the points $\left(\beta_k^*, \beta_j^*\right)$ such that there exists a vector $\left(\vec{\beta}, \lambda\right)$ with $\beta_k = \beta_k^*$ and $\beta_l = \beta_j^*$ which belongs to the confidence set.

Figure A.4: Distribution of $\Omega\left(i\right) + \lambda V\left(i\right)$



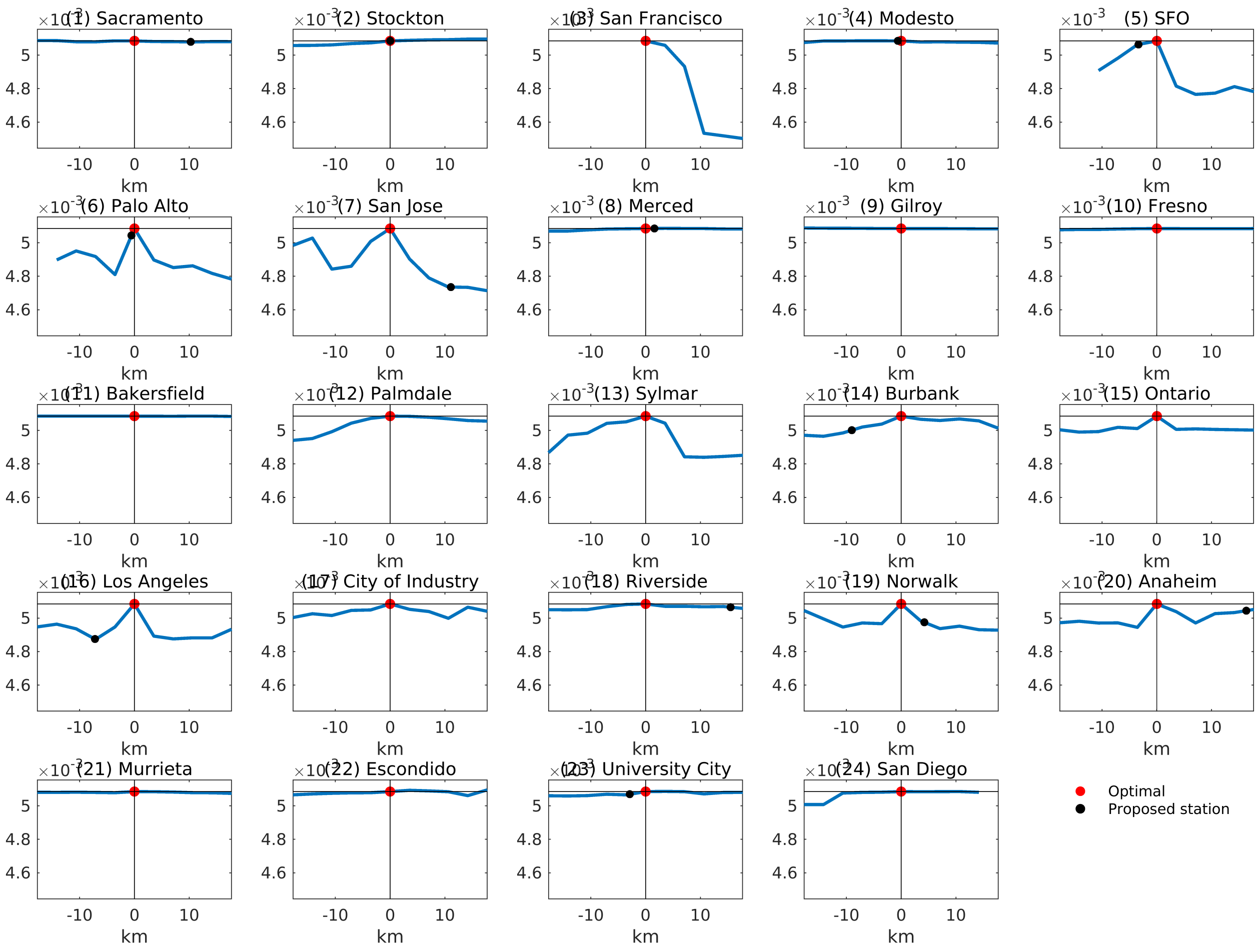
Notes: Histogram of the effective planning weights $\Omega\left(i\right) + \lambda V\left(i\right)$ received by locations in specification (6) in Table 6. Parameters are normalized so that the population-weighted mean of the Pareto weights is 1.

Table A.9: Stations in the Apolitical Planner vs. Proposed Plan

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Station	**Distance (km)**	**ΔWelfare (%)**	$\rho_{\textbf{density}}$	$\rho_{\textbf{wages}}$	$\rho_{\textbf{share college}}$	$\rho_{\textbf{share non-white}}$	$\rho_{\textbf{voting elasticity}}$
Sacramento	-3.43	5.80e-04	-0.051	-0.048	-0.047	0.0072	0.02
Stockton	-0.03	1.20e-06	-0.0093	-0.011	-0.035	-0.0086	0.016
San Francisco	7.58	3.50e-02	0.41	0.14	0.17	0.13	-0.62
Modesto	1.11	-2.00e-04	0.035	0.049	0.078	0.0019	-0.024
SFO	7.86	5.00e-03	0.03	0.064	0.057	-0.011	-0.027
Palo Alto	1.51	5.70e-03	0.13	0.086	0.11	0.15	-0.19
San Jose	-16.72	4.00e-02	0.34	0.18	0.24	0.19	-0.49
Merced	-1.45	-8.50e-05	0.014	0.029	0.036	-0.012	-0.015
Gilroy	27.22	-4.50e-03	0.0016	0.023	0.064	0.02	-0.012
Fresno	-36.69	2.00e-04	-0.0052	-0.015	-0.021	-0.036	0.0093
Bakersfield	68.54	6.30e-04	-0.061	-0.063	-0.071	-0.032	0.013
Palmdale	47.01	4.50e-03	0.17	-0.11	-0.06	0.11	-0.042
Sylmar	33.1	3.10e-02	0.46	-0.34	-0.16	0.32	-0.17
Burbank	4.34	1.10e-02	0.3	-0.16	-0.014	0.16	-0.12
Ontario	-53.43	2.40e-02	0.45	-0.35	-0.15	0.29	-0.19
Los Angeles	3.05	1.80e-03	-0.013	-0.14	-0.13	0.17	0.048
City of Industry	-16.59	1.40e-02	0.27	-0.2	-0.12	0.24	-0.072
Riverside	-5.8	7.20e-04	0.0021	0.0077	0.018	-0.0033	0.0079
Norwalk	-7.85	3.20e-03	-0.13	0.13	0.045	-0.079	0.11
Anaheim	-27.84	1.70e-02	0.38	-0.33	-0.19	0.28	-0.14
Murrieta	-44.76	3.70e-04	-0.0032	-0.073	-0.077	0.073	0.001
Escondido	35.8	3.60e-04	0.036	-0.018	0.08	0.0074	-0.024
University City	1.76	1.80e-04	0.043	0.035	0.071	-0.0031	-0.023
San Diego	-4.78	8.40e-04	0.019	0.047	0.074	-0.024	0.013

Notes: 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 real income in basis points (%) given the Pareto weights $\Omega(i)$ when *only* the corresponding station is moved individually to its optimal location while all other stations remain at proposed plan. In columns (3)-(7), ρ_X show the correlation between $\Delta \log \hat{W}(i)$ and the corresponding covariate X in that one-deviation counterfactual from the proposed plan.

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



Notes: 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.

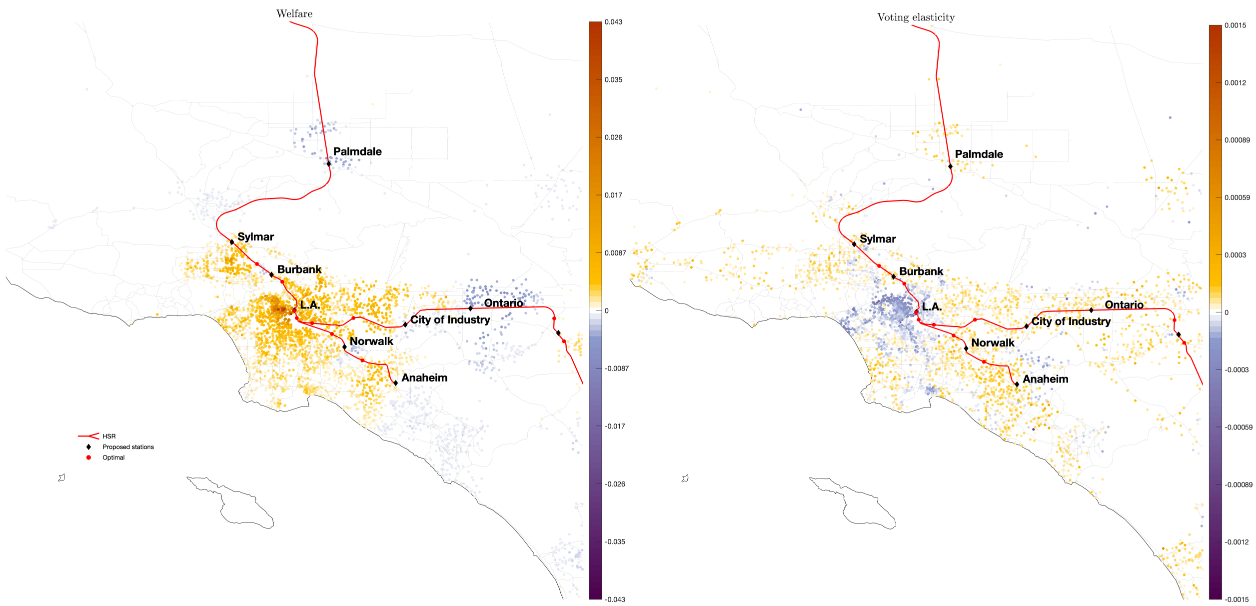


Figure A.7: Utilitarian Planner, Welfare vs. Share College in Riverside