Less Traffic, Higher Tolls: Optimal Congestion Pricing under Remote Work and Traveler Heterogeneity

David Bernal*

Last version here

Abstract

This paper studies how a persistent shift toward remote work changes optimal freeway congestion prices. I combine hourly station traffic from the California Performance Measurement System with tract-level commuting patterns, work-from-home shares, industry composition, population, and wages for five large California counties. Station traffic is allocated to residential tracts using fixed pre-pandemic commuting-route exposure, producing an estimated tract-level freeway VMT measure that aggregates exactly to observed PeMS VMT. A shift–share design based on predetermined local industry shares and national leave-California-out changes in industry remote work implies that a 10 percentage point increase in working from home reduces tract freeway VMT by 1.43 percent. A capacity-normalized BPR technology estimated by Poisson pseudo-maximum likelihood yields a congestion curvature of 3.27. These inputs produce three main pricing findings. First, holding the value of time fixed, the demand shift lowers average optimal tolls by 2.4–4.7 percent. Second, allowing the value of time to move with local wages reverses the result: optimal tolls rise by 8.0–11.4 percent. Third, income heterogeneity changes the meaning of distributional comparisons. Lower-income tracts supply 77–84 percent of modeled traffic and bear most of the marginal external damages in the common Pigouvian calculation. Under a constrained policy that imposes the same time-equivalent charge on both groups, however, the monetary rate paid by higher-income travelers is 1.6–2.0 times larger because their value of time is higher. Distributional conclusions therefore depend on whether the object of interest is a group's role in and exposure to external delay or the monetary rate it pays.

Keywords: Congestion pricing, remote work, vehicle miles traveled, value of time, shift–share instruments, distributional incidence
JEL Classification: R40, R41, D62, H23

*University of Houston, Houston, Texas, United States. E-mail: vdbernalpavas@uh.edu.
I am grateful to my advisors, Vikram Maheshri, Fan Wang, and Janet E. Kohlhase, for their guidance and support.

1 Introduction

Every additional mile driven on a congested road imposes a delay cost on other motorists. The scale of the congestion problem is substantial. The 2025 Urban Mobility Report estimates 9.8 billion hours of delay in U.S. urban areas in 2024, with delay and excess fuel use costing $269 billion in 2024 dollars (Schrank, Albert, Jha, & Eisele, 2025). A longstanding literature proposes congestion taxes that make drivers account for the delay they impose on others (Pigou, 1920; Small, Verhoef, & Lindsey, 2024; Vickrey, 1969). Implementing this principle requires assumptions about who uses the road, how demand responds, and how travelers value time. These choices determine both the efficient charge and its distribution across travelers.

My paper asks how optimal freeway congestion taxes should be designed when travel demand and the value of time differ across income groups and remote work changes commuting patterns. I develop an empirical framework that combines differences in time valuation, commuting-route exposure, and remote-work demand responses on a shared congested network. These differences matter quantitatively. Accounting for contemporaneous wages reverses the pricing implication of lower traffic, and a common charge expressed in units of travel time produces monetary rates 1.6–2.0 times larger for higher-income travelers. The framework makes this distributional choice explicit by allowing higher dollar-per-mile rates for travelers with higher values of time while preserving a common time-equivalent charge.

The model links three components. A nonlinear congestion technology maps total freeway traffic into travel time per mile on a shared network (U.S. Bureau of Public Roads, 1964). Each income group's demand for freeway travel follows a constant-elasticity specification (Yang, Purevjav, & Li, 2020). It responds to the monetary cost of travel time and tolls, while remote work and population shift demand. A travel-time cost function connects these components by converting travel time into dollars using each group's wage-based value of time. Traffic and travel costs are determined jointly, since traffic generates delay and the cost of that delay affects demand. The social cost of an additional mile includes both the driver's own time cost and the delay imposed on other travelers. I use this structure to compute optimal charges under two pricing rules, a common monetary toll and group-specific monetary tolls that correspond to a common time-equivalent charge.

I estimate congestion technology by Poisson pseudo-maximum likelihood using hourly traffic and speed data. A shift–share instrument combining 2019 tract industry shares with national industry-level remote-work changes outside California identifies demand shifts. I set time values to half hourly wages and calibrate the common price elasticity. Together, these inputs determine equilibrium traffic and optimal hourly tolls.

My first contribution is the heterogeneous-agent pricing framework. I estimate demand shifts separately for lower- and higher-income residential tracts, connect these groups to the freeway through their commuting patterns, and allow their values of time to differ. Both groups use the same network, so a change in either group's demand affects the delay experienced by everyone. The analysis builds on the heterogeneous-user congestion-pricing framework of Verhoef and Small (2004) by incorporating identified remote-work responses and observed differences in network use. It distinguishes the external delay borne by each group from the monetary toll that group pays, a distinction central to interpreting distributional comparisons (Eliasson & Mattsson, 2006; van den Berg & Verhoef, 2011).

My second contribution is to show that these economic differences materially change the taxes implied by the data. I separate changes in marginal physical delay from changes in its dollar valuation and compare pricing calculations with fixed and contemporaneous values of time. The same reduction in travel demand can imply a lower or a higher monetary tax depending on how time is valued. Related empirical work calculates congestion charges using traffic conditions and behavioral responses (Kreindler, 2024; Yang et al., 2020). My analysis incorporates observed wage changes and income-specific demand shifts into that calculation, drawing on evidence that values of time vary across people and settings (Bento, Roth, & Waxman, 2024; Lam & Small, 2001; Small, 2012).

My third contribution is to compute the optimal common time-equivalent charge and the associated monetary rates for each income group. The restriction sets the monetary toll divided by the group's value of time equal across groups, giving a transparent rule under which higher-income travelers pay more dollars per mile. Pandey, Alamri, Khoury, and Bakre (2022, Proposition 1) provide a theoretical precedent for this proportionality rule in an express-lane model. To my knowledge, my paper is the first to combine empirical estimates of congestion technology and income-specific remote-work demand shifts to compute the optimal common time-equivalent charge. Progressivity here means a higher monetary rate per mile for the higher-income group. This offers policymakers an explicit way to incorporate distributional objectives into congestion-tax design and communicate how differentiated rates arise from the pricing rule.

I apply the framework to Los Angeles, Orange, San Diego, San Francisco, and Santa Clara counties, comparing 2018–2019 with 2021–2022. The underlying freeway data contain approximately 340.6 million station-hour observations from 6,137 distinct mainline and high-occupancy-vehicle detectors over 2016–2022. The California Performance Measurement System (PeMS) provides traffic and speed observations, while the Longitudinal Employer-Household Dynamics Origin-Destination Employment Statistics (LODES) provide tract-to-tract residence–workplace counts. I use fixed pre-pandemic commuting-route exposure to allocate station traffic across residential tracts, preserving observed county totals. American Community Survey data provide

tract-level remote-work shares, industry composition, population, and earnings. The congestion regressions for the pricing years use 175,180 county-hour observations.

I identify the traffic-demand effect of remote work using differences in tracts' exposure to national changes in industry-specific remote work. Local work-from-home shares may respond to labor-market conditions, residential selection, and transportation shocks. I therefore instrument them with 2019 tract industry shares interacted with national industry changes calculated outside California, following the shift–share literature (Borusyak, Hull, & Jaravel, 2022; Goldsmith-Pinkham, Sorkin, & Swift, 2020). Tract and county-by-year fixed effects compare differently exposed tracts within the same county and year. I combine the identified demand shifts with an estimated relationship between traffic and travel time and wage-based values of time to solve for optimal hourly taxes. The common price elasticity is calibrated at 0.1. Separate morning, afternoon, and off-peak congestion curves allow the time-of-day analysis to account for differences in traffic conditions.

The main pricing result is that optimal monetary taxes rise despite lower congestion when the value of time follows contemporaneous wages. Holding the value of time at its 2018 level, the demand shift lowers average optimal taxes by 2.4–4.7 percent across the five counties. Allowing wages to change reverses the comparison, with taxes rising by 8.0–11.4 percent. The decomposition attributes a reduction of 7.3–9.4 percent of the initial tax to lower marginal physical delay and an increase of 17.0–18.8 percent to the higher value of time. The valuation component is larger, so the net monetary tax rises.

The heterogeneous-agent results show how differentiated rates shift the sharing of congestion charges across income groups. Lower-income tracts supply 77–84 percent of modeled traffic and bear most of the marginal external-delay cost in the common monetary toll calculation. Under the common time-equivalent policy, higher-income travelers pay monetary rates 1.6–2.0 times larger because their value of time is higher. Evaluated at the same traffic allocation, the higher-income group's payments above the common marginal external cost cover the gap left by the lower-income group's payments below it. This is an implicit cross-subsidy in congestion charges. The results compare rates and exposure to delay rather than net welfare gains or losses by income group.

The demand estimates imply that a 10 percentage point increase in working from home reduces allocated tract freeway vehicle miles traveled by 1.43 percent. Under the instrument's identifying assumptions, this is a causal effect of remote work on traffic demand. Research on remote work examines its effects on commuting and the locations of homes and workplaces (Delventhal, Kwon, & Parkhomenko, 2022; Loo & Huang, 2022), while industry differences in its feasibility generate unequal exposure to those changes (Dingel & Neiman, 2020). My analysis carries the identified traffic response into congestion-tax calculations for travelers sharing the same road network.

The hourly results show the largest taxes around the commuting peaks and effectively zero rates during some off-peak hours. With contemporaneous wages, the post-period hourly schedule generally lies above its pre-period counterpart. The time profile therefore preserves the distinction between lower physical congestion and a higher dollar valuation of the remaining delay.

Section 2 develops the model. Section 3 describes the data and traffic allocation. Section 4 presents the congestion and demand specifications. Section 5 reports the pricing results and income-group comparisons. Section 6 discusses the implications for congestion-tax design. The appendices provide theoretical derivations and supplementary evidence.

2 Model

Computing an optimal congestion charge requires three objects: a technology that maps traffic into travel time, a demand curve that maps generalized cost into freeway use, and a value of time that converts delay into dollars. County is indexed by c, hour of the weekly profile by h, calendar year by t, and flow by F in passenger-car-equivalent vehicle miles per hour.

2.1 Congestion technology and travel costs

Throughout this section, F denotes a generic traffic-flow level. When evaluating the model for county c, hour h, and year t, I use F_{cht}. The baseline congestion function $T_c(\cdot)$ is county-specific; travel time varies across hours and years through changes in traffic flow.

Let S_c^0 denote free-flow speed, measured in miles per hour, and K_c the engineering capacity of the observed freeway network. Free-flow speed is the speed vehicles can maintain when traffic is light and other road users do not slow them down. Following the functional form introduced by U.S. Bureau of Public Roads (1964), the BPR travel-time function per mile is

$$T_c(F) = \frac{1}{S_c^0}\left[1 + \alpha_c\left(\frac{F}{K_c}\right)^{\beta}\right], \tag{1}$$

where $\alpha_c > 0$ is a county-specific scale parameter and $\beta > 0$ controls curvature. The ratio F/K_c is dimensionless, and $T_c(F)$ is measured in hours per mile. Differentiation gives

$$T_c'(F) = \frac{\alpha_c \beta}{S_c^0 K_c}\left(\frac{F}{K_c}\right)^{\beta-1}. \tag{2}$$

Following the standard distinction between private and social congestion costs (Small et al., 2024; Walters, 1961), let $TSC_{cht}(F, T)$ denote total social travel-time cost, measured in dollars per hour, as a function of traffic flow F and travel time per mile T. This function adds the time

costs borne by all travelers in a county-hour-year. Because network travel time depends on flow, total cost along the congestion curve is $TSC_{cht}(F, T_c(F))$. Before imposing a functional form for costs, the chain rule gives

$$\frac{dTSC_{cht}(F, T_c(F))}{dF} = \underbrace{\frac{\partial TSC_{cht}}{\partial F}}_{\text{direct cost}} + \underbrace{\frac{\partial TSC_{cht}}{\partial T} \times T_c'(F)}_{\text{congestion effect}}. \tag{3}$$

The partial derivatives are evaluated at $(F, T_c(F))$, holding the other cost determinants fixed. The first term measures the direct cost of additional flow with travel time per mile unchanged. The second captures the change in total cost when additional flow raises travel time for existing traffic. Their sum is marginal social cost, measured in dollars per mile.

In the homogeneous-agent baseline, all travelers have the same value of time VOT_{ct}, measured in dollars per hour. The cost function takes the specific form

$$TSC_{cht}(F, T) = VOT_{ct} \times F \times T. \tag{4}$$

Evaluating this function at the flow and corresponding travel time for county c, hour h, and year t gives

$$TSC_{cht}(F_{cht}, T_c(F_{cht})) = VOT_{ct} \times F_{cht} \times T_c(F_{cht}).$$

Holding VOT_{ct} fixed, the first term in Equation (3) is therefore the private marginal cost (PMC) of one vehicle mile, or the time cost borne by its driver,

$$PMC_{cht}(F) = VOT_{ct} \times T_c(F) = \frac{VOT_{ct}}{S_c^0}\left[1 + \alpha_c \left(\frac{F}{K_c}\right)^{\beta}\right]. \tag{5}$$

The second term is the marginal external congestion cost ($MECC$), the value of the extra delay imposed on all existing traffic. Marginal social cost (MSC) includes both components,

$$MSC_{cht}(F) = PMC_{cht}(F) + MECC_{cht}(F), \tag{6}$$

$$MECC_{cht}(F) = VOT_{ct} F T_c'(F) = \frac{VOT_{ct}}{S_c^0}\alpha_c \beta \left(\frac{F}{K_c}\right)^{\beta}. \tag{7}$$

Thus $MECC$ is measured in dollars per mile. Appendix A derives Equations (6) and (7) from the driver's problem and the planner's problem.

I set the value of time to one-half of the local hourly wage,

$$VOT_{ct} = \theta w_{ct}, \qquad \theta = 0.5, \tag{8}$$

following the conventional benchmark summarized by Small (2012). The factor $\theta = 0.5$ is held fixed across all policy exercises. Wages are either held at their county-specific 2018 levels or updated by county and year. In the heterogeneous-agent extension, wages and values of time also differ between the lower- and higher-income groups within each county.

2.2 Demand, equilibrium, and the Pigouvian tax

Let p_{cht} denote the driver's private cost of travel time and tolls for one additional mile of freeway travel in county c, weekly hour h, and year t, measured in dollars per mile. Travel time is converted into dollars using the value of time, and the monetary toll is added to this amount,

$$p_{cht} = PMC_{cht}(F_{cht}) + \tau_{cht} = VOT_{ct} \times T_c(F_{cht}) + \tau_{cht}.$$

Here VOT_{ct} is the value of time in dollars per hour, $T_c(F_{cht})$ is travel time in hours per mile at traffic flow F_{cht}, and τ_{cht} is the toll in dollars per mile. Thus p_{cht} is neither the toll alone nor the marginal social cost. Without a toll, it equals the driver's travel-time cost. Fuel and other vehicle operating costs are excluded from p_{cht}. In the demand regression, county-by-year fixed effects absorb their effects on traffic to the extent that those effects are common to tracts in the same county and year (Equation (28)).

For the policy simulations, I adopt the constant-elasticity travel-demand specification used by Yang et al. (2020),

$$F_{cht} = A_{cht} p_{cht}^{-\eta}, \qquad \eta = 0.1, \tag{9}$$

where A_{cht} is a positive demand shifter and $-\eta$ is the elasticity of freeway use with respect to generalized monetary cost. This parsimonious form separates changes in demand from movements along the demand curve. Remote work and population shift A_{cht}, while η governs the response to a change in price. It implies $\partial \ln F_{cht} / \partial \ln p_{cht} = -\eta$, holding A_{cht} fixed, and keeps demand positive at positive prices. Constant elasticity is a maintained functional-form assumption, not a result of the remote-work estimates. I set $\eta = 0.1$ as a common calibration across counties, hours, and income groups; the empirical design estimates demand shifts rather than this price elasticity.

With a monetary toll τ_{cht}, private equilibrium satisfies

$$F_{cht} = A_{cht} \left[PMC_{cht}(F_{cht}) + \tau_{cht} \right]^{-\eta}. \tag{10}$$

Following the standard welfare-maximizing congestion-pricing condition (Small et al., 2024, Section 4.1), the planner equates inverse demand with marginal social cost. Efficient flow F^*_{cht} therefore solves

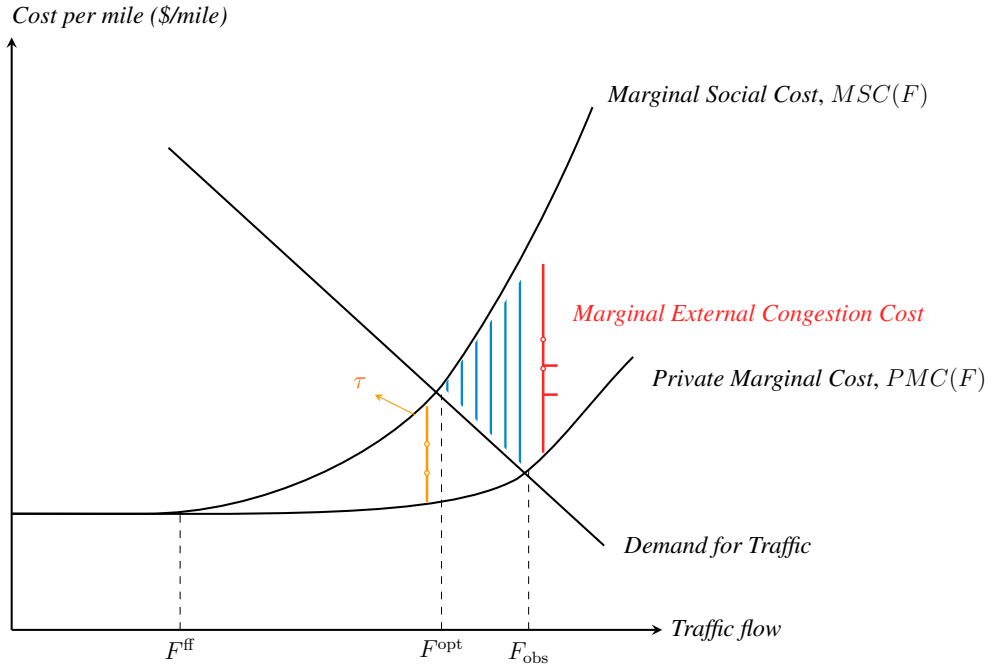
$$F^*_{cht} = A_{cht} \left[MSC_{cht}(F^*_{cht}) \right]^{-\eta}, \tag{11}$$

and is decentralized by the Pigouvian tax

$$\tau^*_{cht} = MECC_{cht}(F^*_{cht}). \tag{12}$$

The tax is not chosen to make traffic as small as possible. It is the value of the delay imposed on other travelers at the efficient flow. Raising the tax beyond Equation (12) would remove trips whose willingness to pay exceeds their social cost and would lower welfare. Figure 1 illustrates this result. Without a toll, demand intersects private marginal cost at F_{obs}. Marginal social cost lies above private marginal cost by the marginal external congestion cost, so the efficient intersection occurs at the lower flow F^{opt}. The vertical gap between the two cost curves at F^{opt} is the optimal toll, and the hatched wedge is the deadweight loss from unpriced congestion. Appendix B provides the formal first-order condition.

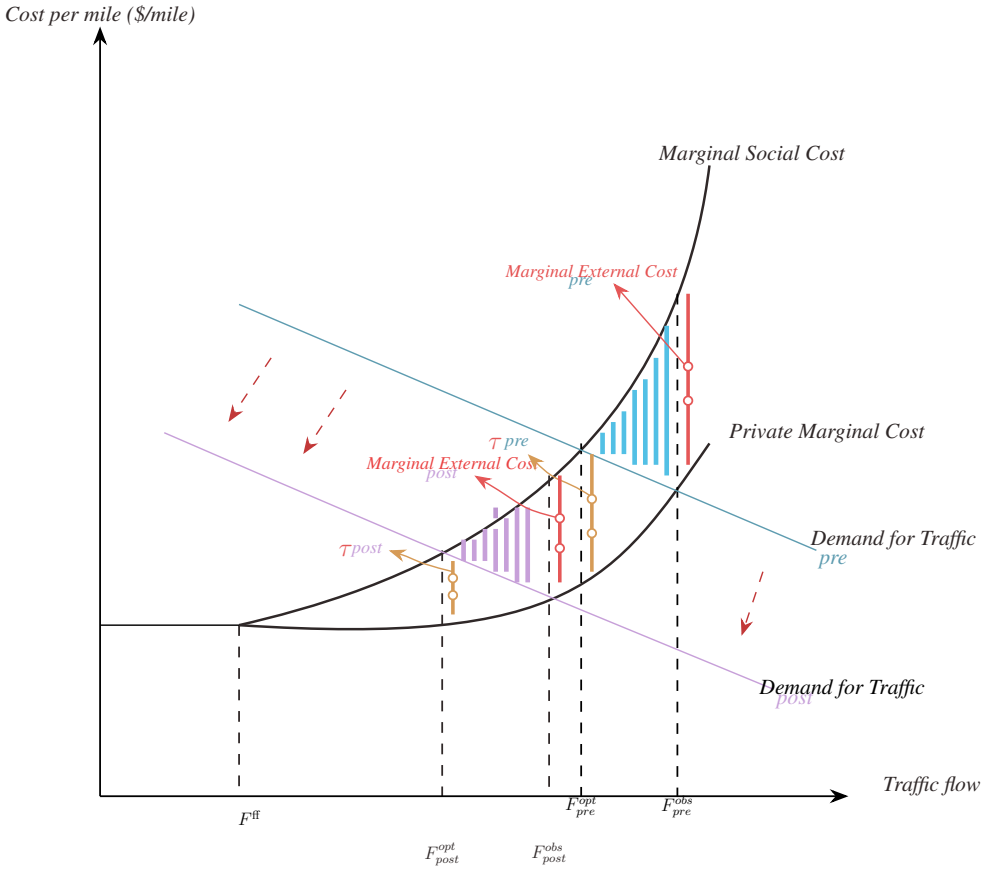
Figure 1. Private and Marginal Social Costs of Congestion



Notes: Schematic illustration of the homogeneous-traveler model. The horizontal axis is traffic flow in vehicle miles per hour and the vertical axis is dollars per vehicle mile. Inverse demand gives the marginal benefit of travel. Private marginal cost is $PMC(F) = VT_c(F)$, where V is the value of time in dollars per hour and $T_c(F)$ is travel time in hours per mile. Marginal social cost is $MSC(F) = PMC(F) + VFT'_c(F)$; the second term is marginal external congestion cost, the added delay cost borne by existing travelers. Demand intersects private cost at unpriced flow F_{obs} and social cost at efficient flow F^{opt}. The vertical gap at efficient flow is the optimal toll τ. The larger gap at observed flow is the external cost before pricing. The hatched area is surplus lost from excess travel under no toll. F^{ff} marks the illustrated free-flow region. This schematic does not report an empirical welfare estimate.

Figure 2 applies the same logic before and after a decline in travel demand. Holding the congestion technology and the value of time fixed, the inward demand shift moves both the unpriced and efficient intersections to lower flows. Because marginal external delay is lower at the new efficient flow, the optimal toll falls. This is the mechanism isolated by the fixed-value-of-time counterfactual; the empirical analysis subsequently allows wages, and therefore the monetary value attached to delay, to change as well.

Figure 2. Optimal Congestion Pricing Before and After a Demand Shift



Notes: Schematic comparison before and after a reduction in travel demand. Traffic flow is on the horizontal axis and costs are dollars per vehicle mile on the vertical axis. Blue demand represents the pre-shift situation and purple demand the post-shift situation. Private marginal cost is the traveler's own time cost per mile; marginal social cost adds the delay imposed on other travelers. Both cost curves hold road capacity, congestion technology, and the dollar value of time fixed. For each demand curve, F^{obs} denotes the no-toll intersection with private cost and F^{opt} the efficient intersection with social cost. The optimal toll τ is the vertical social–private cost gap at the corresponding efficient flow. F^{ff} marks the illustrated free-flow region. Lower demand reduces efficient traffic, marginal external delay, and the optimal toll under these fixed conditions. The diagram illustrates the fixed-value-of-time comparison and does not incorporate the wage changes used in the contemporaneous-value-of-time exercise.

2.3 Toll incidence and congestion-cost offset

Suppressing subscripts, write the perceived price as $p(\tau) = PMC(F(\tau)) + \tau$. Taking logs of the private-equilibrium condition gives

$$\ln F = \ln A - \eta \ln[PMC(F) + \tau]. \tag{13}$$

Differentiating both sides with respect to the toll,

$$\frac{1}{F}\frac{dF}{d\tau} = -\frac{\eta}{PMC(F) + \tau}\left[PMC'(F)\frac{dF}{d\tau} + 1\right]. \tag{14}$$

Collecting the terms that contain $dF/d\tau$ gives

$$\frac{dF}{d\tau} = -\frac{\eta F}{PMC(F) + \tau + \eta F PMC'(F)} < 0. \tag{15}$$

The toll reduces flow, which improves traffic conditions and lowers private time cost:

$$\frac{dPMC}{d\tau} = -\frac{\eta F PMC'(F)}{PMC(F) + \tau + \eta F PMC'(F)} < 0. \tag{16}$$

Accordingly,

$$0 < \frac{dp}{d\tau} = \frac{PMC(F) + \tau}{PMC(F) + \tau + \eta F PMC'(F)} \leq 1. \tag{17}$$

The perceived price rises less than dollar-for-dollar because part of the charge is returned to drivers through lower congestion. This does not imply that an arbitrarily large tax is efficient. Equation (17) is an incidence result at any feasible tax; Equation (12) is the welfare condition selecting the tax. At the optimum,

$$\tau^* = F^* PMC'(F^*) = MECC(F^*), \tag{18}$$

so the pass-through expression becomes

$$\begin{aligned} \left.\frac{dp}{d\tau}\right|_{\tau=\tau^*} &= \frac{PMC(F^*) + MECC(F^*)}{PMC(F^*) + MECC(F^*) + \eta MECC(F^*)} \\ &= \frac{MSC(F^*)}{MSC(F^*) + \eta MECC(F^*)} \in (0, 1). \end{aligned} \tag{19}$$

Thus the optimal toll is determined by the marginal external cost at the efficient flow, while Equation (19) describes how a marginal change in that toll is divided between a higher perceived price and lower congestion cost. This distinction matters for interpretation: a one-dollar increase in the posted toll raises the driver's generalized price by less than one dollar because the induced

reduction in traffic returns part of the charge as time savings. Pass-through therefore describes the incidence of a marginal tax change, not whether the tax itself is welfare maximizing.

2.4 Decomposition of the optimal congestion tax

The optimal tax equals marginal external congestion cost at the efficient flow. To distinguish physical delay from its monetary valuation, define

$$g_c(F) \equiv FT_c'(F), \qquad \tau_{cht}^* = VOT_{ct} g_c(F_{cht}^*). \tag{20}$$

The function $g_c(F)$ is marginal external delay in hours per mile. It measures the additional delay imposed on existing travelers by one more unit of traffic, as derived in Appendix A. Differentiating marginal external congestion cost with respect to flow and the value of time gives the continuous decomposition

$$d\tau_{cht}^* = \underbrace{VOT_{ct} g_c'(F_{cht}^*)\, dF_{cht}^*}_{\text{congestion component}} + \underbrace{g_c(F_{cht}^*)\, dVOT_{ct}}_{\text{value-of-time component}}. \tag{21}$$

Here $g_c'(F) = T_c'(F) + FT_c''(F)$. The first term captures how a small change in efficient traffic changes the tax when the value of time is held fixed. The second captures how a small change in the value of time changes the dollar valuation of external delay when traffic is held fixed.

For the empirical before–after comparison, bars denote the reported county-period tax and value of time. The corresponding effective delay index is $g_{cr} \equiv \overline{\tau}_{cr} / \overline{VOT}_{cr}$, where $r \in \{pre, post\}$. Suppressing the county subscript, the exact discrete decomposition is

$$\Delta\overline{\tau} = \underbrace{\overline{VOT}_{pre}(g_{post} - g_{pre})}_{\text{congestion component}} + \underbrace{g_{post}(\overline{VOT}_{post} - \overline{VOT}_{pre})}_{\text{value-of-time component}}. \tag{22}$$

The first term changes the effective delay index from its pre-period to its post-period value while holding the value of time at its pre-period level. The second values the post-period index using the change in the value of time. Evaluating the index at its post-period value in the second term includes the interaction between the two changes, so the decomposition is exact for finite before–after differences.

3 Data

3.1 PeMS traffic and the observed freeway network

PeMS records hourly flow, VMT, speed, lane counts, and station location for California freeways. I use mainline and HOV stations in the five study counties. Flow in passenger-car equivalents (PCE), expressed as vehicle miles traveled (VMT) per hour, is

$$F_{cht} = VMT_{cht}^{passenger} + 1.5\, VMT_{cht}^{truck}. \tag{23}$$

The factor 1.5 converts truck VMT into passenger-car equivalents. It is a transparent engineering benchmark rather than an estimated parameter.

Free-flow speed is estimated at the station level using low-flow observations in 2018 and 2019, requiring speeds between 50 and 90 miles per hour and adequate detector coverage. Mainline capacity per lane follows the standard engineering schedule $1{,}700 + 10S_s^0$ passenger cars per hour per lane when $S_s^0 \leq 70$ mph, capped at 2,400 above 70 mph. HOV capacity is set to 1,650 passenger cars per hour per lane.[1] Station capacity is multiplied by the number of lanes and the PeMS coverage length; county capacity K_c is the sum over stations, measured in PCE-VMT per hour, the same units as F_{cht}. Table 1 reports the resulting network inputs.

These capacity measures describe the freeway mileage represented by the detector network, not the administrative capacity of every road in the county. This alignment is important: the numerator F_{cht} and denominator K_c cover the same observed network. Unlike these engineering inputs, the county scale parameters $\widehat{\alpha}_c$ in Table 1 are estimated jointly with the common curvature β by Poisson pseudo-maximum likelihood using Equation (27), presented in Section 4.

3.2 Allocating station VMT to residential tracts

PeMS does not identify the residential origin of a vehicle. I therefore allocate observed station VMT to origin census tracts using pre-pandemic LODES commuting pairs and modeled freeway routes:

$$\widehat{VMT}_{it} = \sum_{s \in \mathcal{S}_{c(i)}} \left[\frac{\sum_j \overline{N}_{ij}^{pre} A_{ijs}}{\sum_{k \in \mathcal{I}_{c(i)}} \sum_j \overline{N}_{kj}^{pre} A_{kjs}} \right] \left[\sum_{h \in t} Flow_{sh} L_s \right]. \tag{24}$$

Here i indexes the residential origin tract and j the workplace destination tract. The commuting count $\overline{N}_{ij}^{pre}$ is the average number of LODES workers traveling from i to j before the pandemic. The indicator A_{ijs} equals one when the modeled freeway route between the centroids of i and j

[1] The capacity benchmarks follow the Federal Highway Administration's HPMS guidance and Freeway Management Handbook.

Table 1. Estimated Congestion Technology and Engineering Inputs

County	$\widehat{\alpha}_c$	S_c^0 (mph)	K_c (million PCE-VMT/h)	Stations	Lane-miles
Los Angeles	2.583	66.85	10.407	2,623	4,562.9
Orange	1.649	67.37	4.402	1,667	1,950.7
San Diego	2.350	67.19	5.533	891	2,388.5
San Francisco	3.260	66.82	0.406	62	171.6
Santa Clara	2.720	66.26	2.497	554	1,058.6

Notes: The sample covers hourly freeway traffic in Los Angeles, Orange, San Diego, San Francisco, and Santa Clara counties in 2018–2019 and 2021–2022. California's Performance Measurement System (PeMS) supplies the detector and traffic inputs. The conditional-mean model is $E[D_{cht} \mid F_{cht}, c] = \alpha_c (F_{cht}/K_c)^{\beta}$, where $D_{cht} = \max\{S_c^0/Q_{cht} - 1, 0\}$ is proportional delay relative to free flow, Q_{cht} is observed speed, and F_{cht} is hourly traffic. Poisson pseudo-maximum likelihood (PPML) estimates county-specific scale parameters α_c and common curvature $\widehat{\beta} = 3.270$, with heteroskedasticity-robust HC1 standard error 0.019, from 175,180 county-hour observations. Free-flow speed S_c^0 is the capacity-mile-weighted harmonic mean of station speeds under light traffic. Engineering capacity K_c sums station capacity times coverage length over mainline and high-occupancy-vehicle detectors. PCE denotes passenger-car equivalents, VMT vehicle miles traveled, and mph miles per hour; trucks receive a PCE factor of 1.5. Capacity is million PCE-VMT per hour, matching traffic units. Stations and lane-miles describe the observed engineering network rather than every road in the county.

traverses the freeway edge represented by PeMS station s. The set $\mathcal{S}_{c(i)}$ contains the stations in the county of origin tract i.

The index k runs over every candidate residential origin tract in $\mathcal{I}_{c(i)}$, including tract i. Thus the denominator is the total pre-pandemic routed-worker exposure to station s across all origin tracts in that county, and the ratio in brackets is the share assigned to tract i.

In the second bracket, $Flow_{sh}$ is observed traffic flow at station s in hour h, L_s is the freeway coverage length represented by that station in miles, and $\sum_{h \in t} Flow_{sh} L_s$ is annual station VMT. The allocation therefore does not weight every commuting pair equally. Routes that use more of the observed freeway network encounter more station exposure, while each station contributes only the VMT associated with its own coverage length. The pre-period route weights are fixed, so the pandemic movement in $\widehat{VMT}_{it}$ comes from observed PeMS traffic rather than from contemporaneous LODES commuting counts.

For every matched station, the bracketed allocation shares sum to one across origin tracts. Consequently,

$$\sum_{i \in c} \widehat{VMT}_{it} = \sum_{s \in c} VMT_{st}^{PeMS}. \tag{25}$$

Stations without a route match are allocated using fixed baseline origin-worker shares, so the county accounting identity is preserved. Equation (25) shows that the allocation preserves observed

county totals; it does not independently validate the tract-level estimates. The cross-sectional allocation comes from LODES and modeled route exposure; the aggregate time-series movement comes from observed PeMS traffic.

3.3 Remote work, industry exposure, and wages

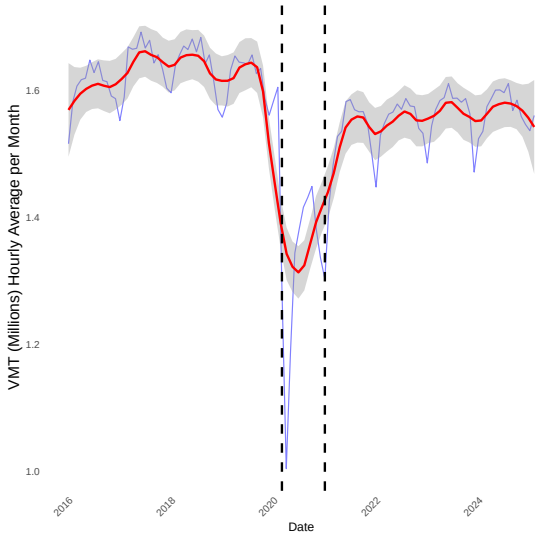
ACS five-year estimates provide tract population, work-from-home shares, unemployment, median household income, poverty, college attainment, vehicle availability, civilian employment by industry, and earnings. LODES provides the origin–destination commuting counts used in Equation (24). Section 4 introduces the shift–share design after presenting the demand equation and explains how the industry variables identify differential exposure to national changes in remote work.

Figure 3 summarizes monthly traffic conditions across the five counties. Vehicle miles traveled (VMT) fell sharply in 2020, average speed rose, and the travel time index (TTI) fell. TTI is the ratio of observed travel time to travel time at the free-flow reference speed: a value of 1.2 indicates a trip taking 20 percent longer than the reference time. Traffic subsequently recovered in part, but the persistent change in commuting behavior motivates separating the pre-period 2018–2019 from the post-period 2021–2022.

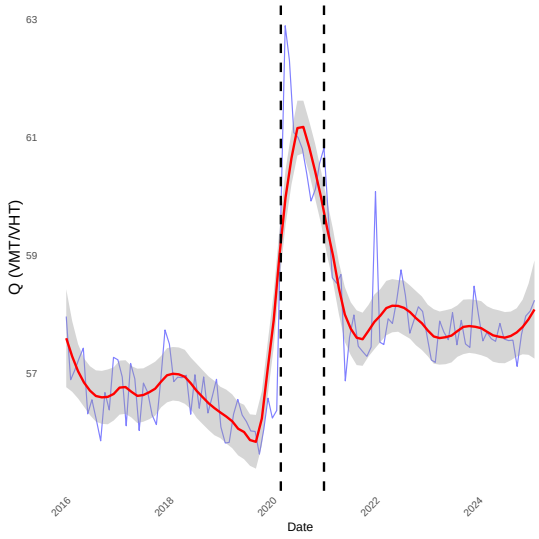
I summarize the persistent break in traffic conditions using adjusted before–after regressions that compare 2018–2019 with 2021–2022 and exclude 2020. Each regression relates log VMT, average speed, or TTI to a post-period indicator while absorbing county-by-hour-of-week and county-by-calendar-month fixed effects; standard errors are clustered by year-month. Appendix Table 7 shows that, across all hours, VMT was 7.71 percent lower, average speed was 3.08 percent higher, and TTI was 3.11 percent lower in the post period. During the weekday afternoon peak, the reduction in TTI was substantially larger, at 7.44 percent. These comparisons describe the change in traffic conditions; they are not the causal estimates of the effect of remote work developed in Section 4.

Appendix Figure 12 reports the adjusted TTI reduction separately by county. The decline ranges from 1.09 percent in Orange County to 5.25 percent in San Francisco, with Santa Clara also experiencing a relatively large reduction of 4.39 percent. This geographic heterogeneity motivates the county-specific congestion technologies used below. Appendix Figure 13 shows the corresponding increase in county work-from-home shares, while Appendix Figure 14 documents how the changes in VMT and congestion vary over the 24-hour day. Together with Figure 3, these results show that the pandemic shift affected the level, timing, and geography of freeway use.

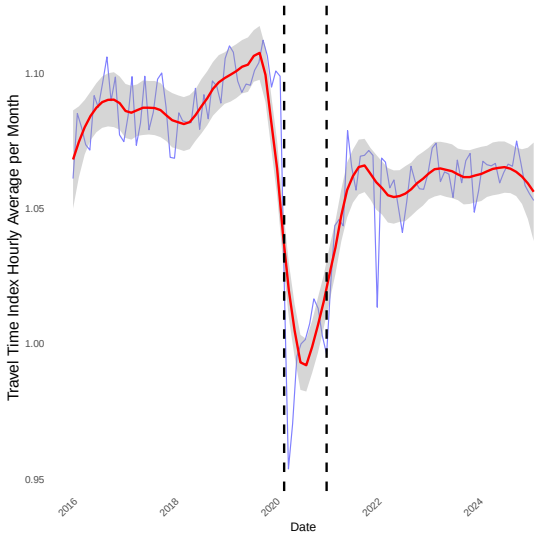
Figure 3. Freeway Traffic and Congestion



(a) Vehicle miles traveled



(b) Speed



(c) Travel time index

Notes: Source is California's Performance Measurement System (PeMS), using county-hour freeway observations for Los Angeles, Orange, San Diego, San Francisco, and Santa Clara counties from January 2016 through February 2025. Panel (a) is mean hourly county vehicle miles traveled (VMT), in millions, within each month. Panel (b) is speed in miles per hour, calculated from hourly $Q = VMT/VHT$, where VHT denotes vehicle hours traveled, and summarized using VMT weights. Panel (c) is the VMT-weighted travel time index (TTI), the ratio of travel time to the free-flow reference time; TTI of 1.2 means travel takes 20 percent longer than under free flow. Blue lines show monthly observations, red lines are locally smoothed trends, and gray bands are the smoothers' 95 percent confidence bands. Dashed vertical lines mark March 2020 and January 2021. These descriptive series include the disruption and years outside the pricing comparison, which uses 2018–2019 and 2021–2022. The 2020 fall in VMT coincides with higher speed and lower TTI.

4 Estimation

4.1 Congestion technology

Let Q_{cht} denote observed average speed. Proportional delay relative to free flow is

$$D_{cht} = \max\left\{\frac{S_c^0}{Q_{cht}} - 1, 0\right\}. \tag{26}$$

The truncation at zero treats speeds above the estimated free-flow benchmark as no measured congestion rather than negative delay. The preferred conditional-mean specification is

$$E[D_{cht} \mid F_{cht}, c] = \alpha_c \left(\frac{F_{cht}}{K_c}\right)^{\beta}, \tag{27}$$

estimated by Poisson pseudo-maximum likelihood (PPML), following the conditional-mean approach discussed by Santos Silva and Tenreyro (2006). PPML is useful here because D_{cht} is nonnegative and frequently zero; log OLS would discard zero-delay observations and mechanically select the congested part of the sample. The estimator is used as a quasi-likelihood for the conditional mean and does not require D_{cht} to be a count.

The preferred specification contains county-specific α_c and one common β. It uses 175,180 county-hour observations in 2018, 2019, 2021, and 2022. The estimated curvature is $\widehat{\beta} = 3.270$ with an HC1 standard error of 0.019. Table 1 reports the county scale estimates. Because flow and speed are jointly determined within the hour, I interpret Equation (27) as the congestion curve used for structural calibration rather than as a fully causal supply estimate.

4.2 Tract-level demand

The empirical demand equation is

$$\ln \widehat{VMT}_{it} = \xi_i + \delta_{c(i)t} + \rho WFH_{it} + \gamma \ln Population_{it} + X_{it}'\pi + \varepsilon_{it}, \tag{28}$$

where WFH_{it} is measured in percentage points, ξ_i are tract fixed effects, and $\delta_{c(i)t}$ are county-by-year fixed effects. County-by-year effects absorb shocks to traffic common to tracts in the same county and year, including the common effects of fuel and other operating costs, county labor-market conditions, and policies. The preferred specification sets X_{it} empty; a robustness specification adds unemployment, household income, poverty, college attainment, and vehicle availability.

Because remote work is endogenous, I instrument it with

$$Z_{it} = 100 \sum_{q} s_{iq,2019} \left(WFH_{q,t}^{US\setminus CA} - WFH_{q,2019}^{US\setminus CA} \right), \tag{29}$$

where $s_{iq,2019}$ is the predetermined employment share of industry q in tract i, and the term in parentheses is the national change in that industry's work-from-home rate calculated outside California. The identifying assumption is that, conditional on tract and county-by-year fixed effects, predetermined differences in local industry exposure affect tract freeway VMT through their effect on remote work rather than through a separate tract-specific channel aligned with the national industry shock.

Equation (29) is an exposure design in the terminology of Goldsmith-Pinkham et al. (2020) and a shock-based shift–share design in the terminology of Borusyak et al. (2022). The empirical argument relies on the national industry shocks being as-good-as-random with respect to the appropriately weighted tract residuals, while the 2019 shares determine differential exposure.

Table 2 presents the estimates. The preferred coefficient is -0.001441, so a 10 percentage point increase in remote work changes VMT by $100[\exp(10\widehat{\rho}) - 1] = -1.43$ percent. Adding demographic controls gives a similar effect of -1.16 percent. The clustered first-stage Wald statistics are large, and standard errors are two-way clustered by tract and year.

The estimated remote-work and population responses enter the demand shifters used in the hourly congestion-pricing simulations.

Table 2. Remote Work and Tract-Level Freeway VMT

	(1) Preferred IV	(2) IV with controls
Work-from-home share (percentage points)	-0.001441^{**}	-0.001167^{*}
	(0.000582)	(0.000486)
Effect of a 10 p.p. increase (percent)	-1.43	-1.16
First-stage coefficient	3.606	3.622
First-stage clustered Wald statistic	$1{,}006.6$	744.9
Observations	$25{,}088$	$24{,}975$
Tract fixed effects	Yes	Yes
County-by-year fixed effects	Yes	Yes
Log population	Yes	Yes
Additional demographic controls	No	Yes

Notes: The sample contains annual residential-tract observations in Los Angeles, Orange, San Diego, San Francisco, and Santa Clara counties over 2016–2022. The dependent variable is the logarithm of allocated tract freeway vehicle miles traveled (VMT), constructed from PeMS station traffic and fixed pre-pandemic LODES commuting-route exposure. PeMS is California's Performance Measurement System; LODES is the Longitudinal Employer-Household Dynamics Origin-Destination Employment Statistics. The American Community Survey supplies the work-from-home share, measured in percentage points, and tract characteristics. Both instrumental-variable (IV) specifications instrument remote work using 2019 tract industry shares interacted with national industry-specific remote-work changes calculated outside California. Both include tract and county-by-year fixed effects and log population. Column (2) adds unemployment, median household income, poverty, college attainment, and the share of households without a vehicle; its smaller sample reflects missing controls. A 10-point effect equals $100[\exp(10\widehat{\rho}) - 1]$. Parentheses contain standard errors clustered by tract and year, with seven year clusters. The first-stage coefficient measures the instrument's association with remote work; the Wald statistic tests the excluded instrument. Stars denote $^{*}p < 0.10$ and $^{**}p < 0.05$. A causal interpretation requires the instrument's identifying assumptions.

5 Results

5.1 Optimal taxes with fixed and contemporaneous values of time

Table 3 reports average optimal taxes. Under a fixed 2018 value of time, post-period taxes fall in every county. The reductions range from 2.39 percent in San Diego to 4.75 percent in San Francisco. These estimates isolate the physical effect of the demand shift: lower efficient traffic reduces marginal external delay.

Table 3. Optimal Congestion Taxes Before and After the Demand Shift

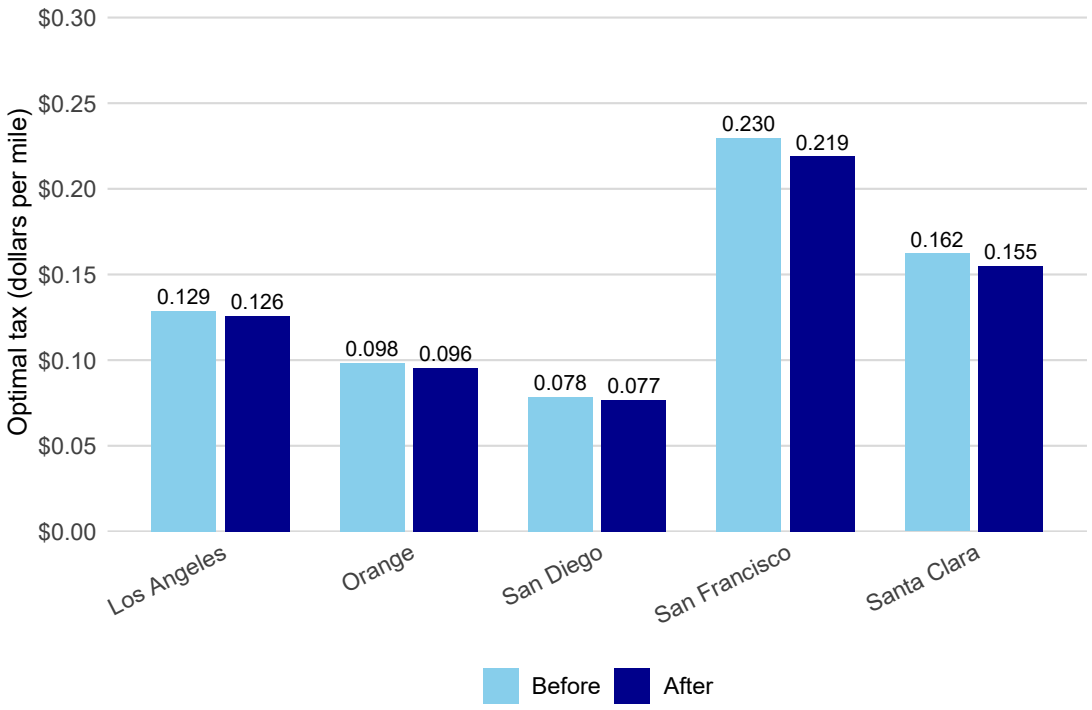
	Fixed pre-period value of time			Contemporaneous value of time		
County	Pre	Post	Change (%)	Pre	Post	Change (%)
Los Angeles	0.129	0.126	−2.52	0.130	0.145	11.38
Orange	0.098	0.096	−2.78	0.098	0.109	10.84
San Diego	0.078	0.077	−2.39	0.079	0.088	10.95
San Francisco	0.230	0.219	−4.75	0.231	0.250	8.30
Santa Clara	0.162	0.155	−4.42	0.165	0.178	7.98

Notes: Average optimal congestion taxes for Los Angeles, Orange, San Diego, San Francisco, and Santa Clara counties, in dollars per vehicle mile. Pre denotes 2018–2019 and post denotes 2021–2022; 2020 is excluded. Change is $100(\tau_{post}/\tau_{pre} - 1)$, calculated from unrounded rates. The fixed-value-of-time columns set time value to one-half of each county's 2018 median hourly wage in every year. The contemporaneous columns use one-half of the county median hourly wage in the corresponding year. Estimated annual remote-work and population demand shifts feed into hourly demand, with common calibrated price-elasticity magnitude 0.1. Each hourly equilibrium jointly determines efficient traffic and the toll, equal to the marginal delay cost imposed on other travelers. The table summarizes those schedules over the county's hourly profile. Inputs are PeMS traffic, American Community Survey wages and characteristics, and the instrumental-variable remote-work estimates. The opposing signs across wage scenarios reflect changes in the dollar valuation of delay as well as physical congestion.

Figure 4 shows the same counterfactual. The levels differ across counties because the estimated congestion curves, network capacities, baseline hourly profiles, and wages differ. Changes in work-from-home shares and population also vary across tracts and counties, producing different county demand shifts and therefore different post-period traffic and taxes. San Francisco has the largest dollar-per-mile tax even though its observed detector network is much smaller, reflecting its steep congestion technology and high value of time.

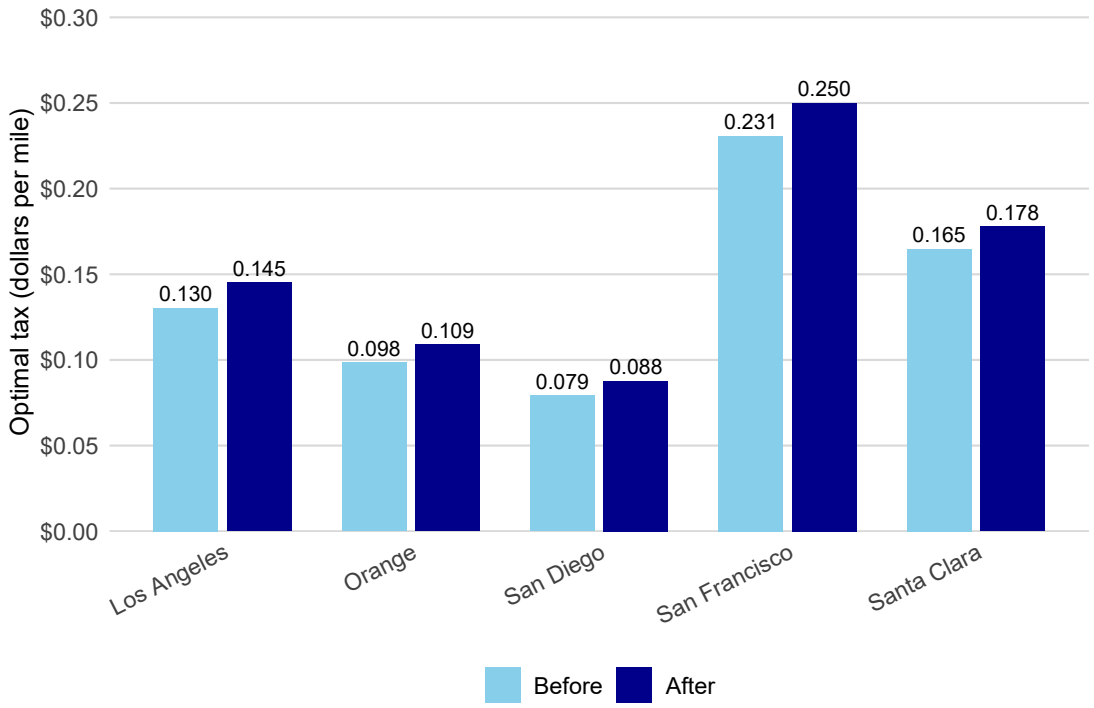
Allowing wages to change reverses the comparison. As shown in Figure 5, post-period taxes rise by 7.98 percent in Santa Clara, 8.30 percent in San Francisco, and approximately 11 percent in Los Angeles, Orange, and San Diego. The fixed-wage calculation understates the contemporaneous post-period tax by 12.4–13.4 percent.

Figure 4. Optimal Congestion Taxes with the Value of Time Fixed



Notes: Average optimal congestion taxes for Los Angeles, Orange, San Diego, San Francisco, and Santa Clara counties, comparing 2018–2019 (before) and 2021–2022 (after). Light-blue bars denote before and dark-blue bars after; labels are dollars per vehicle mile. The value of time stays fixed at one-half of each county's 2018 median hourly wage. Estimated remote-work and population changes shift annual demand, while the baseline traffic profile determines its hourly pattern. Efficient traffic and its tax are solved jointly using the county congestion curve. Each hourly toll equals the monetary delay cost imposed on other travelers at efficient flow. All calculations use calibrated price-elasticity magnitude 0.1. Bars summarize county-period average rates rather than a single hour. With time valuation fixed, the demand shift lowers taxes in every county. Source inputs are PeMS traffic, American Community Survey wages and demographics, and the instrumental-variable demand estimates.

Figure 5. Optimal Congestion Taxes with Contemporaneous Values of Time



Notes: Average optimal congestion taxes in dollars per vehicle mile for Los Angeles, Orange, San Diego, San Francisco, and Santa Clara counties, comparing 2018–2019 (before) and 2021–2022 (after). Light blue denotes before and dark blue after. The value of time is one-half of the county median hourly wage in the corresponding year, so it changes with observed wages. Estimated remote-work and population responses shift annual demand; price-elasticity magnitude is calibrated at 0.1. County congestion technology links traffic to travel time, and efficient flow and its toll are solved jointly for every hour before calculating period averages. Each toll measures the monetary delay cost imposed on other travelers at efficient flow. Post-period taxes exceed pre-period taxes because higher monetary time values outweigh lower marginal physical delay. Sources are PeMS traffic, American Community Survey wages and demographics, and the instrumental-variable demand estimates. Wage changes are observed inputs, not estimated causal effects of remote work.

5.2 Why the tax rises

Table 4 and Figure 6 apply Equation (22). The congestion component is negative in every county, ranging from -7.35 to -9.40 percent of the pre-period tax. The value-of-time component is positive and substantially larger, from 16.96 to 18.76 percent. The net effect is therefore an increase in the optimal monetary charge.

Table 4. Decomposition of the Change in the Optimal Tax

County	Congestion component (% of pre tax)	Value-of-time component (% of pre tax)	Total change (% of pre tax)
Los Angeles	-7.38	18.76	11.38
Orange	-7.80	18.63	10.84
San Diego	-7.35	18.30	10.95
San Francisco	-9.40	17.71	8.30
Santa Clara	-8.99	16.96	7.98

Notes: Exact decomposition of the change in average optimal congestion taxes for Los Angeles, Orange, San Diego, San Francisco, and Santa Clara counties from 2018–2019 to 2021–2022, using contemporaneous wages and excluding 2020. Every entry is a percentage of the county's pre-period tax. Write $\overline{\tau} = \overline{V}g$, where $\overline{V}$ is the period-average value of time and $g = \overline{\tau}/\overline{V}$ is the effective external-delay index. The congestion component is $100\overline{V}_{pre}(g_{post} - g_{pre})/\overline{\tau}_{pre}$. The valuation component is $100g_{post}(\overline{V}_{post} - \overline{V}_{pre})/\overline{\tau}_{pre}$. Using the post-period index in the second term includes the interaction, so the components sum exactly to the total change before rounding. Time values equal half county-year hourly wages; taxes come from hourly efficient demand–congestion equilibria with price-elasticity magnitude 0.1. The negative congestion component lowers taxes, but the larger positive time-valuation component raises the net monetary charge. Sources are PeMS traffic, American Community Survey wages and demographics, and the demand estimates.

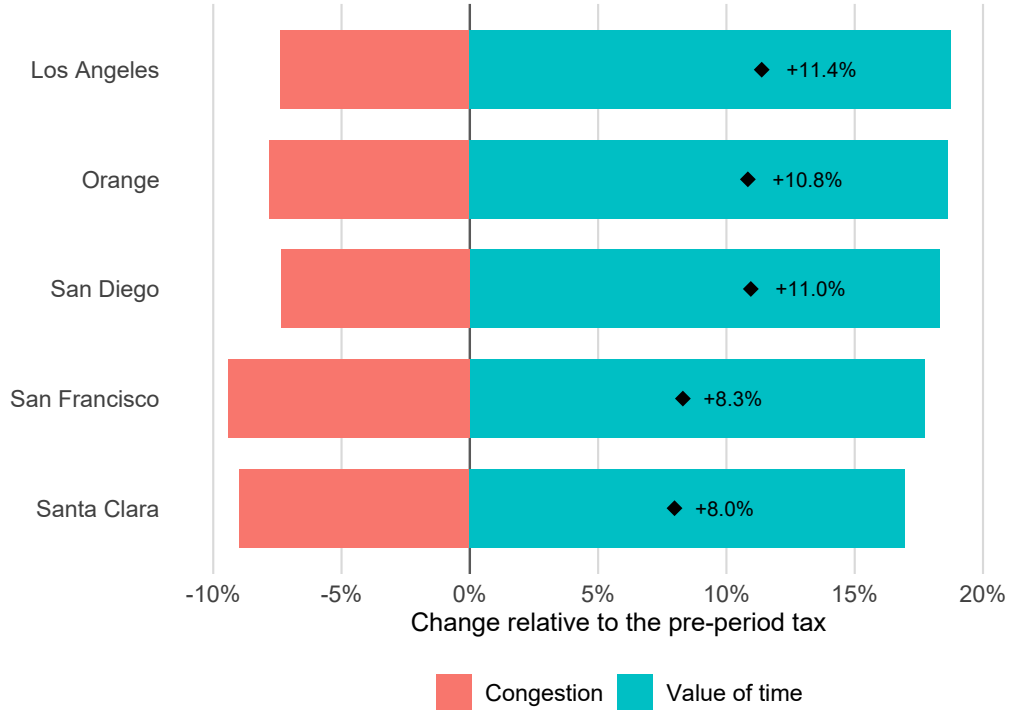
This result should not be interpreted as evidence that the pandemic increased physical congestion. It did not. Rather, the dollar value attached to a unit of marginal delay rose by more than the marginal delay itself fell. The sign difference between the fixed- and changing-wage exercises is exactly the distinction between physical congestion and its monetary social cost.

5.3 Time of day

The hourly profile is economically important because the marginal external cost is highly nonlinear in flow. I re-estimate Equation (27) separately for weekday morning peak hours (6:00–8:59), afternoon peak hours (15:00–17:59), and all other weekday hours. Within each period, the tax is still solved for every hour rather than assigned as a single period-wide rate.

Figure 7 plots the contemporaneous-wage hourly schedule. Taxes are nonnegative in every county and period: several off-peak hourly rates round to zero, and the model never predicts a negative tax or subsidy. The largest rates occur around the commuting peaks. San Francisco

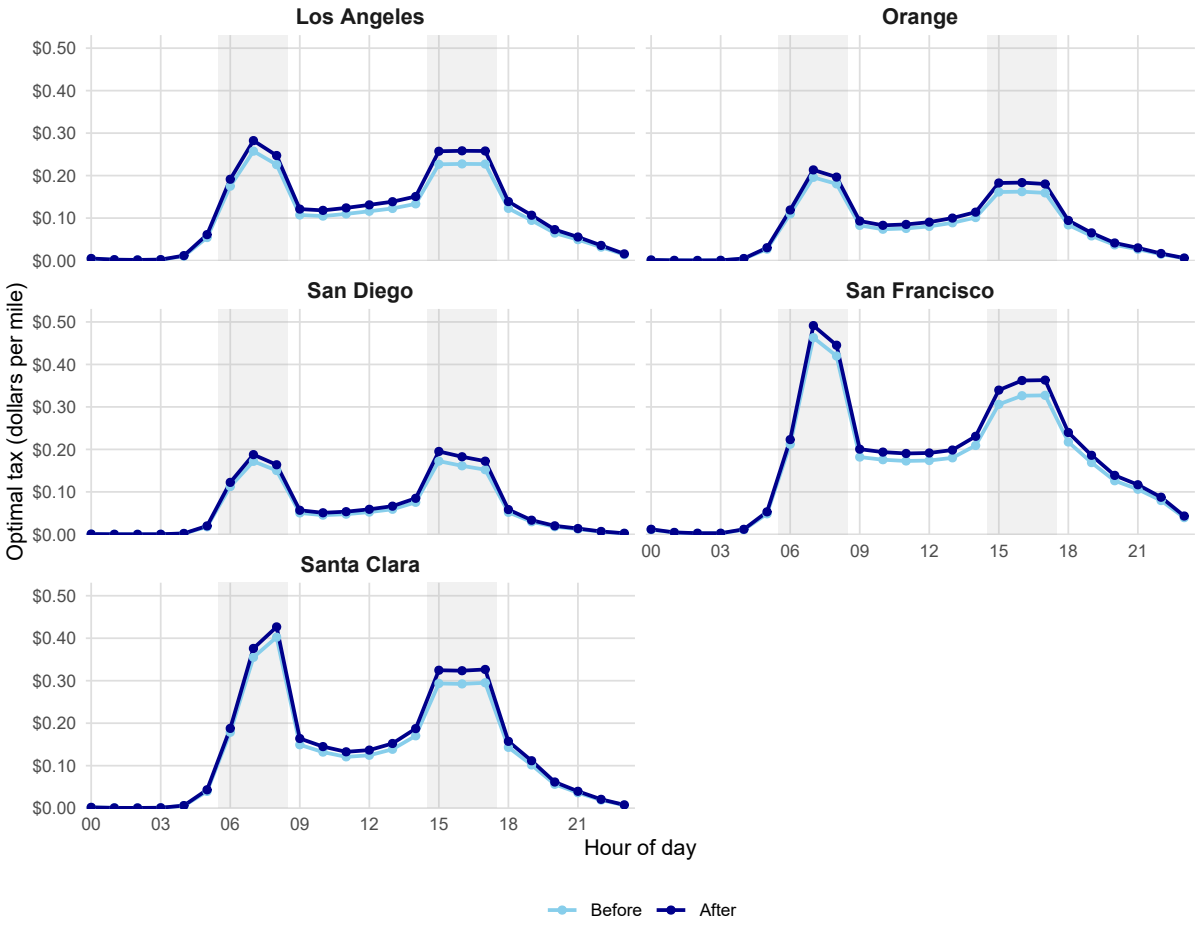
Figure 6. Decomposition of the Change in the Optimal Tax



Notes: Exact decomposition of changes in county-period optimal taxes between 2018–2019 (before) and 2021–2022 (after) for Los Angeles, Orange, San Diego, San Francisco, and Santa Clara counties, using contemporaneous wages. Components are percentages of each county's pre-period tax. Salmon bars are the congestion component, which changes the effective marginal-delay index while holding the pre-period value of time fixed. Turquoise bars are the valuation component, which values the post-period delay index using the change in the value of time. The effective index is the ratio of the period's average tax to its average value of time. Black diamonds show the net percentage change. The two components sum to the diamond, up to rounding; evaluating the valuation term at the post-period index includes the interaction between changes. Negative congestion components reflect lower marginal delay. The larger positive valuation components explain rising monetary taxes. Hourly rates come from the demand–congestion equilibrium with price-elasticity magnitude 0.1.

and Santa Clara have especially pronounced morning profiles, while Los Angeles and San Diego exhibit relatively large afternoon rates. The post-period schedule is generally above the pre-period schedule because higher values of time dominate lower marginal delay.

Figure 7. Optimal Congestion Tax over the Weekday Hourly Profile



Notes: Weekday optimal congestion taxes for Los Angeles, Orange, San Diego, San Francisco, and Santa Clara counties. Each panel is one county, the horizontal axis is clock hour (00–23), and the vertical axis is dollars per vehicle mile. Light-blue lines and points represent 2018–2019 and dark blue 2021–2022. Each year's value of time equals half its county median hourly wage. Gray bands identify the morning peak, 6:00–8:59, and afternoon peak, 15:00–17:59. Separate congestion curves are estimated for those periods and all other weekday hours, but efficient flow and the tax are solved separately for each hour using price-elasticity magnitude 0.1. Estimated annual remote-work and population demand shifts apply across hours within a county-year. Curves summarize the resulting hourly schedules within each period. Peak travel produces the largest external-delay charges; higher contemporaneous time values generally raise the post-period schedule. Inputs are PeMS traffic, American Community Survey wages and demographics, and the instrumental-variable demand estimates.

The period averages in Appendix Figure 9 range from roughly 4 to 14 cents per mile off peak in the pre-period and from 15 to 37 cents per mile during peak periods. With contemporaneous

wages, post-period changes are positive in all 15 county–time-of-day combinations (five counties and three times of day), ranging from 5.85 to 13.54 percent.

5.4 Income heterogeneity and group-specific tax rates

I divide tracts into groups $g \in \{L, H\}$. A tract belongs to the higher-income group when its 2018 median household income is above the household-weighted 80th percentile in its county; the remaining tracts form the lower-income group. This classification is held fixed over time. Let $\widetilde{w}_{gct}$ be the employed-person-weighted median of tract median annual earnings within county-income group gc, divided by 2,080 hours.[2] The group value of time is

$$V_{gct} = \theta \widetilde{w}_{gct}, \qquad \theta = 0.5. \tag{30}$$

Thus V_{gct} is group g's value of travel time in dollars per hour in county c and year t; it is one-half of the group's estimated hourly wage, not its annual household income.

The computation begins by estimating the tract demand equation separately for each group:

$$\ln \widehat{VMT}_{it} = b_g WFH_{it} + \rho_g \ln Population_{it} + \mu_i + \delta_{c(i)t} + e_{it}, \qquad i \in \mathcal{I}_{gc}. \tag{31}$$

As in the pooled specification, WFH_{it} is instrumented with the leave-California-out shift–share measure. Let A_{gcht} denote the demand shifter for income group g in county c, weekly hour h, and year t. It incorporates the group's baseline traffic and its estimated demand changes associated with remote work and population. Demand can therefore evolve differently across groups because both their estimated responses and their observed changes in remote work and population differ. Congestion depends on the groups' combined use of the road, $F_{cht} = F_{Lcht} + F_{Hcht}$.

As in the heterogeneous-user congestion-pricing framework of Verhoef and Small (2004), marginal external cost accounts for the values of time of all travelers exposed to the added delay. In the shared-network model here, this implies a common monetary Pigouvian toll. The superscript P denotes the unconstrained Pigouvian allocation; it is a policy label, not the generalized price p

[2]The conversion assumes 40 hours of work per week for 52 weeks, giving $40 \times 52 = 2{,}080$ hours per year. This is a common annual-hours assumption used to convert earnings into an hourly wage, not a measure of each worker's actual hours.

used in Equation (9). Thus F^P_{gcht} is efficient flow supplied by group g, F^P_{cht} is their total, and τ^P_{cht} is the common first-best monetary toll in dollars per mile. These objects solve

$$F^P_{gcht} = A_{gcht}\left[V_{gct}T_c(F^P_{cht}) + \tau^P_{cht}\right]^{-\eta}, \qquad F^P_{cht} = \sum_g F^P_{gcht}, \tag{32}$$

$$\tau^P_{cht} = T_c'(F^P_{cht}) \sum_g V_{gct}F^P_{gcht}. \tag{33}$$

Every additional vehicle mile creates the same physical delay, regardless of which group supplies it, so Equation (33) is the same dollar-per-mile charge for both groups. The additive terms $\tau^{contrib}_{gcht} = T_c'(F^P_{cht})V_{gct}F^P_{gcht}$ sum to τ^P_{cht} and attribute the external cost to the travelers who bear the added delay. They are contributions to the common toll, not rates levied on each group. Appendix Figure 11 reports this decomposition.

Pandey et al. (2022, Proposition 1) provide a theoretical precedent for a common time-equivalent charge. Their deterministic express-lane model uses discounts that imply a common toll-to-value-of-time ratio across traveler groups. To construct group-specific rates in this paper, I impose the same proportionality restriction, so both groups face the same toll measured in travel time, κ_{cht} hours per mile, converted into money using the group value of time,

$$\tau^V_{gcht} = V_{gct}\kappa_{cht}. \tag{34}$$

Here κ_{cht} is the common time-equivalent charge: it adds the same number of travel hours per mile to each group's perceived time cost. It is not itself a dollar toll. Multiplication by V_{gct} converts it into the group-specific monetary rate τ^V_{gcht}, where the superscript V denotes the value-of-time-proportional policy. The generalized monetary price is therefore $V_{gct}[T_c(F_{cht}) + \kappa_{cht}]$.

Pandey et al. (2022) examine this rule in express-lane simulations. To my knowledge, this paper is the first to use empirical estimates of congestion technology and income-specific remote-work demand shifts to compute the optimal common time-equivalent congestion charge. Computing the common time-equivalent charge itself is an empirical contribution of the paper, together with calculating the corresponding monetary tolls for each income group and examining how both change with remote work and wages.

With the common elasticity η, the constrained optimum and the two group flows are the joint solution to

$$F^\kappa_{gcht} = A_{gcht}\left\{V_{gct}\left[T_c(F^\kappa_{cht}) + \kappa^*_{cht}\right]\right\}^{-\eta}, \qquad F^\kappa_{cht} = \sum_g F^\kappa_{gcht}, \tag{35}$$

$$\kappa^*_{cht} = F^\kappa_{cht}T_c'(F^\kappa_{cht}). \tag{36}$$

The superscript κ on flow denotes equilibrium under this constrained schedule, while κ^*_{cht} is the welfare-maximizing common time charge within that policy class. Substituting the BPR derivative from Equation (2) makes the tax computed in each county-hour explicit:

$$\kappa^*_{cht} = \frac{\alpha_c \beta}{S^0_c} \left(\frac{F^{\kappa}_{cht}}{K_c} \right)^{\beta}, \tag{37}$$

$$\tau^V_{gcht} = \frac{V_{gct}}{S^0_c} \alpha_c \beta \left(\frac{F^{\kappa}_{cht}}{K_c} \right)^{\beta}. \tag{38}$$

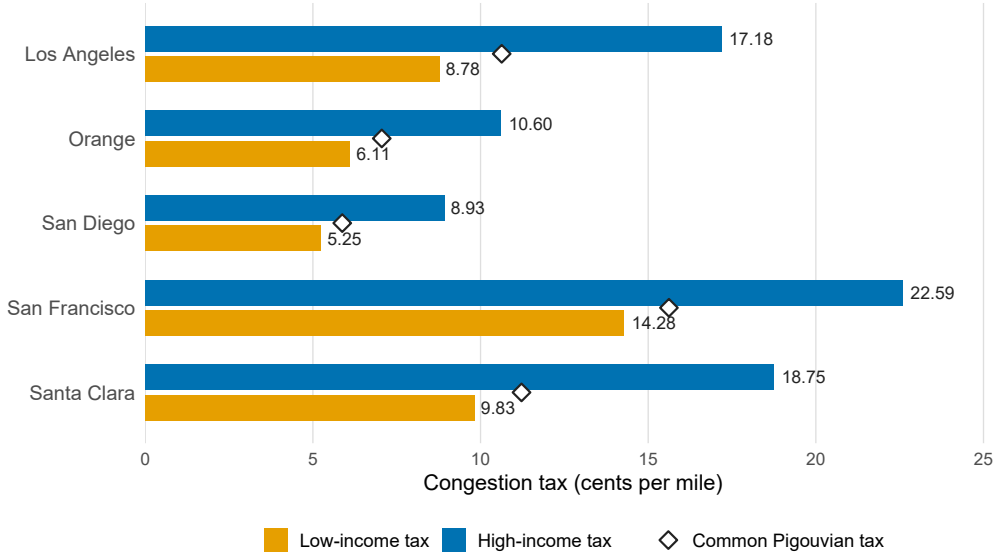
Equations (35)–(38) are solved simultaneously, rather than assigning each group a separate road or holding total traffic fixed. Within the same county-hour, both groups face the same κ^*_{cht}, so

$$\frac{\tau^V_{Hcht}}{\tau^V_{Lcht}} = \frac{V_{Hct}}{V_{Lct}}. \tag{39}$$

This proportional schedule is not the unconstrained first-best common monetary toll in Equation (33); it is the optimum within the restricted class $\tau_g = V_g \kappa$. Appendix E derives both results from welfare.

Figure 8 shows the constrained rates in 2021–2022 using contemporaneous wages; Table 5 compares them with 2018–2019. Appendix Figure 10 reports both periods under fixed and contemporaneous wages. Lower-income tracts account for approximately 77–84 percent of equilibrium flow and therefore for most of the common Pigouvian externality. Nonetheless, their proportional rates are lower because Equation (39) maps the common time charge into dollars through group-specific values of time. In the post-period, higher-income rates are 1.6–2.0 times lower-income rates. For example, the Los Angeles rates are 17.2 and 8.8 cents per mile, respectively. With contemporaneous wages, the post-period rate rises in every county and for both groups: by 10.8–13.7 percent for lower-income tracts and 5.9–10.3 percent for higher-income tracts.

Figure 8. Value-of-Time-Proportional Taxes after the Demand Shift



Notes: Optimal group-specific monetary tolls for Los Angeles, Orange, San Diego, San Francisco, and Santa Clara counties in 2021–2022, using contemporaneous wages. Orange bars show lower-income rates and blue bars higher-income rates; labels and the horizontal axis are cents per vehicle mile. Higher-income tracts are above the household-weighted county 80th percentile of 2018 median household income, with group membership fixed thereafter. Group value of time V_g equals half estimated hourly earnings in the corresponding year. The common time-equivalent charge κ is hours per mile, and each group's monetary rate is $\tau_g^V = V_g\kappa$. Both groups' demands and shared congestion are solved jointly using income-specific remote-work and population shifts and calibrated price-elasticity magnitude 0.1. Diamonds mark the unconstrained common monetary Pigouvian toll, computed at that policy's own equilibrium. Higher monetary rates for higher-income travelers reflect their higher time values, not estimates of net welfare effects. Appendix Figure 10 shows both periods and wage scenarios.

Table 5. Value-of-Time-Proportional Taxes by Income Group

	Lower-income tracts			Higher-income tracts		
County	Pre	Post	Change (%)	Pre	Post	Change (%)
Los Angeles	0.079	0.088	10.84	0.159	0.172	8.14
Orange	0.054	0.061	12.41	0.100	0.106	6.19
San Diego	0.046	0.052	13.17	0.081	0.089	10.25
San Francisco	0.126	0.143	13.73	0.213	0.226	5.91
Santa Clara	0.088	0.098	11.77	0.176	0.187	6.73

Notes: Average dollar-per-vehicle-mile rates under the common time-equivalent policy for Los Angeles, Orange, San Diego, San Francisco, and Santa Clara counties. Pre is 2018–2019 and post 2021–2022, excluding 2020; changes use unrounded rates. Higher-income tracts exceed the household-weighted county 80th percentile of 2018 median household income, and all other tracts form the lower-income group. Membership stays fixed. Each group's value of time V_g equals one-half of its employed-person-weighted median hourly earnings proxy, updated by year. The policy charges $\tau_g^V = V_g\kappa$, where κ is a common time-equivalent charge in hours per mile. At the constrained optimum, $\kappa = F^\kappa T_c'(F^\kappa)$, with F^κ the groups' total equilibrium traffic and T_c travel time per mile. Both demands and congestion are solved jointly using income-specific remote-work and population demand shifts and common price-elasticity magnitude 0.1. These entries are rates actually paid by each group, rather than components of a common monetary toll. Higher-income rates are larger because group values of time are higher. Inputs are PeMS, LODES commuting exposure, and American Community Survey income, earnings, and remote work.

6 Conclusion

The paper's main contribution is an empirical congestion-pricing framework that combines income-group differences in remote-work exposure, estimated demand responses, traffic shares, and the value of time. I estimate remote-work responses separately for lower- and higher-income tracts and use their observed changes in remote work and population to construct group-specific demand shifts, rather than imposing the same shock on all travelers. Both groups' demands and congestion are then solved jointly on the shared freeway network.

The shift–share demand estimates imply that a 10 percentage point increase in the share working from home reduces allocated tract freeway VMT by 1.43 percent. Under the design's identifying assumptions, this is a causal effect of remote work on traffic demand. The group-specific estimates carry differences in remote-work exposure and behavior into the pricing exercise.

The pricing results show that less congestion can coexist with higher optimal monetary tolls. Holding the value of time at its 2018 level, taxes fall by 2.4–4.7 percent after the demand shift. Using contemporaneous wages, they instead rise by 8.0–11.4 percent. The decomposition shows that the increase in the dollar value of time outweighs the decline in marginal physical delay. The heterogeneous-agent exercise further shows that lower-income tracts supply most modeled traf-

fic, while higher-income travelers face monetary rates 1.6–2.0 times larger under the constrained common time-equivalent schedule. These rate comparisons are distinct from net welfare effects by income group.

Heterogeneity matters because a demand shock to one group changes travel time for everyone on the shared network. The framework captures both the group's own demand adjustment and the resulting change in delay borne by other travelers. It therefore connects income-specific remote-work responses to traffic composition and congestion prices, rather than representing the demand shift through a single average traveler.

The estimated tax rates depend on the data and assumptions used in the analysis and should be interpreted as policy benchmarks rather than precise prescriptions for implementation. The paper provides a framework that policymakers can use when setting congestion taxes to account for changes in traffic demand, the value of time, and differences across travelers.

For example, the heterogeneous-agent exercise illustrates an implicit cross-subsidy under the common time-equivalent charge. Evaluated at the same traffic allocation, higher-income travelers pay more per mile than the common marginal external congestion cost, while lower-income travelers pay less. The higher-income group's payments above this benchmark cover the gap left by the lower-income group's payments, allowing lower-income traffic to face a lower monetary rate. This interpretation concerns the sharing of congestion charges across groups.

References

Bento, A., Roth, K., & Waxman, A. R. (2024). *The value of urgency: Evidence from real-time congestion pricing* (Working Paper No. 26956). Cambridge, MA: National Bureau of Economic Research.

Borusyak, K., Hull, P., & Jaravel, X. (2022). Quasi-experimental shift-share research designs. *The Review of Economic Studies*, *89*(1), 181–213. doi: 10.1093/restud/rdab030

Delventhal, M. J., Kwon, E., & Parkhomenko, A. (2022). JUE insight: How do cities change when we work from home? *Journal of Urban Economics*, *127*, 103331. doi: 10.1016/j.jue.2021.103331

Dingel, J. I., & Neiman, B. (2020). How many jobs can be done at home? *Journal of public economics*, *189*, 104235.

Eliasson, J., & Mattsson, L.-G. (2006). Equity effects of congestion pricing: Quantitative methodology and a case study for stockholm. *Transportation Research Part A: Policy and Practice*,

40(7), 602–620.

Goldsmith-Pinkham, P., Sorkin, I., & Swift, H. (2020). Bartik instruments: What, when, why, and how. *American Economic Review*, *110*(8), 2586–2624. doi: 10.1257/aer.20181047

Kreindler, G. (2024). Peak-hour road congestion pricing: Experimental evidence and equilibrium implications. *Econometrica*, *92*(4), 1233–1268. doi: 10.3982/ECTA18422

Lam, T. C., & Small, K. A. (2001). The value of time and reliability: Measurement from a value pricing experiment. *Transportation Research Part E: Logistics and Transportation Review*, *37*(2–3), 231–251.

Loo, B. P. Y., & Huang, Z. (2022). Spatio-temporal variations of traffic congestion under work from home (WFH) arrangements: Lessons learned from COVID-19. *Cities*, *124*, 103610. doi: 10.1016/j.cities.2022.103610

Pandey, V., Alamri, B. G., Khoury, H., & Bakre, M. (2022, September). *Equitable dynamic pricing of express lanes* (Final Report No. CATM-2022-R9-NCAT). Center for Advanced Transportation Mobility, North Carolina Agricultural and Technical State University. Retrieved from `https://rosap.ntl.bts.gov/view/dot/64671`

Pigou, A. C. (1920). *The economics of welfare*. London: Macmillan.

Santos Silva, J. M. C., & Tenreyro, S. (2006). The log of gravity. *The Review of Economics and Statistics*, *88*(4), 641–658. doi: 10.1162/rest.88.4.641

Schrank, D., Albert, L., Jha, K., & Eisele, B. (2025, August). *2025 urban mobility report* (Tech. Rep.). Texas A&M Transportation Institute. Retrieved from `https://static.tti.tamu.edu/tti.tamu.edu/documents/mobility-report-2025.pdf`

Small, K. A. (2012). Valuation of travel time. *Economics of transportation*, *1*(1-2), 2–14.

Small, K. A., Verhoef, E. T., & Lindsey, R. (2024). *The economics of urban transportation* (3rd ed.). Routledge. doi: 10.4324/9781315157375

U.S. Bureau of Public Roads. (1964). *Traffic assignment manual for application with a large, high speed computer* (Tech. Rep.). Washington, DC: U.S. Department of Commerce, Bureau of Public Roads, Office of Planning, Urban Planning Division.

van den Berg, V. A. C., & Verhoef, E. T. (2011). Congestion tolling in the bottleneck model with heterogeneous values of time. *Transportation Research Part B: Methodological*, *45*(1), 60–78.

Verhoef, E. T., & Small, K. A. (2004). Product differentiation on roads: Constrained congestion pricing with heterogeneous users. *Journal of Transport Economics and Policy*, *38*(1), 127–156.

Vickrey, W. S. (1969). Congestion theory and transport investment. *American Economic Review*, *59*(2), 251–260.

Walters, A. A. (1961). The theory and measurement of private and social cost of highway congestion. *Econometrica*, *29*(4), 676–699. doi: 10.2307/1911814

Yang, J., Purevjav, A.-O., & Li, S. (2020). The marginal cost of traffic congestion and road pricing: Evidence from a natural experiment in beijing. *American Economic Journal: Economic Policy*, *12*(1), 418–453. doi: 10.1257/pol.20170195

A Derivation of Private, External, and Social Costs

The distinction between private and social congestion costs follows the classical analysis of Walters (1961) and the textbook treatment in Small et al. (2024, Chapter 3). The derivation below applies this framework to the travel-time costs in the paper's model.

This appendix suppresses the county, hour, and year indices for readability. Let F denote modeled freeway flow in passenger-car-equivalent vehicle miles per hour, $T(F)$ travel time in hours per mile, and V the value of time in dollars per hour. Thus V corresponds to VOT_{ct} in the main text. A prime denotes differentiation with respect to flow. The abbreviations PMC, $MECC$, and MSC mean private marginal cost, marginal external congestion cost, and marginal social cost, respectively, all in dollars per mile. Aggregate travel-time cost $C(F)$ is

$$C(F) = V \times F \times T(F), \tag{40}$$

measured in dollars per hour. A nonatomic driver treats $T(F)$ as given. Her private cost of one additional vehicle mile is therefore

$$PMC(F) = V \times T(F). \tag{41}$$

The planner recognizes that an additional unit of flow changes the travel time experienced by every unit already on the road:

$$\begin{aligned} MSC(F) &= \frac{dC(F)}{dF} && (42)\\ &= V \times [T(F) + F \times T'(F)] && (43)\\ &= \underbrace{V \times T(F)}_{PMC(F)} + \underbrace{V \times F \times T'(F)}_{MECC(F)}. && (44) \end{aligned}$$

This is the origin of Equations (6) and (7): the first term is the entrant's own time cost and the second is the time cost imposed on all existing flow.

For the capacity-normalized Bureau of Public Roads (BPR) function (U.S. Bureau of Public Roads, 1964), let S^0 be free-flow speed in miles per hour, K network capacity in the same units as F, $\alpha > 0$ the congestion scale parameter, and $\beta > 0$ the curvature parameter. Travel time is

$$T(F) = \frac{1}{S^0} \times \left[1 + \alpha \times (F/K)^{\beta}\right], \tag{45}$$

the derivative is

$$T'(F) = \frac{\alpha \times \beta}{S^0 \times K} \times (F/K)^{\beta-1}. \tag{46}$$

Multiplying by F gives

$$F \times T'(F) = \frac{\alpha \times \beta}{S^0} \times (F/K)^{\beta}. \tag{47}$$

Therefore,

$$PMC(F) = \frac{V}{S^0} \times \left[1 + \alpha \times (F/K)^{\beta}\right], \tag{48}$$

$$MECC(F) = \frac{V}{S^0} \times \alpha \times \beta \times (F/K)^{\beta}, \tag{49}$$

$$MSC(F) = \frac{V}{S^0} \times \left[1 + \alpha \times (1+\beta) \times (F/K)^{\beta}\right]. \tag{50}$$

All three expressions are dollars per mile. Capacity is essential for interpreting the flow ratio, while free-flow speed determines the conversion from a time index to hours per mile.

B Efficient Flow, Optimal Toll, and Pass-Through

The welfare argument follows the standard static congestion-pricing model presented in Small et al. (2024, Section 4.1). Social surplus is the benefit from travel net of its resource costs. Toll payments are transfers from travelers to the toll recipient, so they are not deducted again as a social cost.

This appendix suppresses the county, hour, and year indices for readability. Let F denote modeled traffic flow, $T(F)$ travel time in hours per mile, and V the value of time in dollars per hour. Let $P(F)$ denote inverse demand, or marginal willingness to pay for freeway travel, in dollars per mile.[3] Write A for the demand shifter, η for the common price-elasticity magnitude, and τ for the monetary toll per mile. Private marginal cost is $PMC(F) = V \times T(F)$, marginal external congestion cost is $MECC(F) = V \times F \times T'(F)$, and marginal social cost is $MSC(F) = PMC(F) + MECC(F)$. A prime denotes a flow derivative and a star an efficient value.

Let $F^{ref} > 0$ be a fixed reference flow. Net social surplus $W(F)$, up to an additive constant, is

$$W(F) = \int_{F^{ref}}^{F} P(x)\, dx - V \times F \times T(F). \tag{51}$$

Here x is the variable of integration used to add the marginal benefits of successive units of travel. All travelers experience the travel time at total flow F, so total time cost is $V \times F \times T(F)$. A positive reference makes welfare finite under the calibrated constant-elasticity demand, whose

[3] In the main text's notation, $P_{cht}(F) = (A_{cht}/F)^{1/\eta}$ is the inverse of $F_{cht} = A_{cht} p_{cht}^{-\eta}$. Here A_{cht} is the demand shifter and $\eta = 0.1$ is the absolute value of the price elasticity. At a private equilibrium, $P_{cht}(F_{cht}) = p_{cht} = PMC_{cht}(F_{cht}) + \tau_{cht}$. The function $P(F)$ therefore denotes willingness to pay, whereas p_{cht} denotes the generalized price faced by drivers.

integral from zero diverges when $\eta \leq 1$. The reference changes neither welfare comparisons nor the optimality condition.

To find efficient flow, differentiate Equation (51) with respect to F, holding V and F^{ref} fixed. The fundamental theorem of calculus gives

$$\frac{d}{dF}\left[\int_{F^{ref}}^{F} P(x)\,dx\right] = P(F).$$

For total travel-time cost, both the amount of travel F and the time per mile $T(F)$ change with flow. The product rule gives

$$\frac{d}{dF}\left[V \times F \times T(F)\right] = V \times \left[\frac{dF}{dF} \times T(F) + F \times \frac{dT(F)}{dF}\right] = V \times T(F) + V \times F \times T'(F).$$

Subtracting the cost derivative from the benefit derivative yields

$$W'(F) = P(F) - V \times T(F) - V \times F \times T'(F).$$

The planner sets this derivative equal to zero at the optimal flow F^* and rearranges the terms, giving

$$P(F^*) = V \times [T(F^*) + F^* \times T'(F^*)] = MSC(F^*). \tag{52}$$

Private equilibrium under a tax τ satisfies $P(F) = V \times T(F) + \tau$. Evaluating this condition at F^* and comparing it with Equation (52) proves that

$$\tau^* = V \times F^* \times T'(F^*) = MECC(F^*). \tag{53}$$

For the constant-elasticity demand $F = A \times [PMC(F) + \tau]^{-\eta}$, log differentiation yields

$$\frac{1}{F} \times \frac{dF}{d\tau} = -\eta \times \frac{PMC'(F) \times (dF/d\tau) + 1}{PMC(F) + \tau}, \tag{54}$$

which rearranges to Equation (15). Because $PMC'(F) = V \times T'(F)$, the optimal tax also satisfies $\tau^* = F^* \times PMC'(F^*)$. Substitution into Equation (17) gives

$$\begin{aligned} \left.\frac{dp}{d\tau}\right|_{\tau=\tau^*} &= \frac{PMC(F^*) + MECC(F^*)}{PMC(F^*) + MECC(F^*) + \eta \times MECC(F^*)} \\ &= \frac{MSC(F^*)}{MSC(F^*) + \eta \times MECC(F^*)}. \end{aligned} \tag{55}$$

Here $p = PMC(F) + \tau$ is the generalized private price in dollars per mile, so $dp/d\tau$ measures how much of a marginal toll increase reaches the driver after the resulting congestion relief. The expression lies strictly between zero and one whenever congestion is present. It characterizes the marginal incidence of the optimal toll; it is not a separate rule for selecting its level.

C Tract Freeway VMT and the Shift–Share Design

In this appendix, i and k index residential origin tracts, j workplace destination tracts, s freeway detector stations, h observed hours, and t calendar years. The county containing tract i is $c(i)$; $\mathcal{I}_{c(i)}$ is its set of candidate origin tracts and $\mathcal{S}_{c(i)}$ its observed stations. The outcome $\widehat{VMT}_{it}$ is annual vehicle miles traveled allocated to tract i, not a direct count of that tract's trips. The pre-period commuting count $\overline{N}_{ij}^{pre}$ comes from the Longitudinal Employer-Household Dynamics Origin-Destination Employment Statistics (LODES). The route indicator A_{ijs} equals one when the modeled route from i to j passes station s; it is not the demand shifter A_{cht}. $Flow_{sh}$ denotes observed vehicles per hour and L_s station coverage length in miles. Station observations come from California's Performance Measurement System (PeMS).

Equation (24) combines two distinct sources of variation. The numerator $\sum_j \overline{N}_{ij}^{pre} A_{ijs}$ measures the pre-pandemic commuting exposure of origin tract i to station s. The denominator repeats the same calculation over every candidate origin k in the county. For any matched station, summing that ratio over i therefore gives one. The second bracket, $\sum_{h\in t} Flow_{sh} L_s$, is the station's observed annual VMT contribution. Summing Equation (24) over origin tracts consequently reproduces total PeMS VMT for the county, as stated in Equation (25).

The indices i and j uniquely identify an origin–destination pair; no redundant county index is needed inside the commuting-pair sums. County enters only through the origin-tract set $\mathcal{I}_{c(i)}$ and the station set $\mathcal{S}_{c(i)}$. Because the LODES counts and route indicators are fixed in the pre-period, they determine only how observed station VMT is distributed across tracts. Year-to-year traffic movement comes from PeMS. The resulting $\widehat{VMT}_{it}$ should therefore be interpreted as exposure-allocated freeway VMT associated with tract i; changes in unobserved route choice within an origin tract are not measured.

For the first-stage regression, WFH_{it} is the tract's work-from-home share in percentage points, $Population_{it}$ its population, and Z_{it} the shift–share instrument combining 2019 tract industry employment shares with national industry-specific changes in remote work calculated outside California. The first stage is

$$WFH_{it} = \xi_i^F + \delta_{c(i)t}^F + \pi Z_{it} + \gamma_F \ln Population_{it} + X_{it}'\lambda + u_{it}. \tag{56}$$

Here ξ_i^F and $\delta_{c(i)t}^F$ are tract and county-by-year fixed effects in the first stage; the superscript F labels that stage, not traffic flow. The coefficients π and γ_F capture the instrument and log-population relationships. X_{it} is the optional vector of demographic controls, λ its coefficient vector, and u_{it} the regression error; the prime on X_{it} denotes transpose. The leave-California-out construction limits mechanical correlation between local remote-work measurement and the national industry shocks.

D Time-of-Day Congestion Technology

Let c index counties, h observed hours, and t calendar years. F_{cht} is traffic flow in passenger-car-equivalent vehicle miles per hour, Q_{cht} is observed average speed in miles per hour, S_c^0 is county free-flow speed in miles per hour, and K_c is network capacity in the same units as F_{cht}. Proportional delay above free flow is

$$D_{cht} = \max\left\{S_c^0/Q_{cht} - 1, 0\right\}.$$

Thus $D_{cht} = 0.2$ means travel time per mile is 20 percent above the free-flow benchmark, while speeds above that benchmark are assigned zero delay. This is a dimensionless delay measure, not travel time in hours. Period r denotes morning peak (6:00–8:59), afternoon peak (15:00–17:59), or all other weekday hours. I estimate

$$E[D_{cht} \mid F_{cht}, c, r] = \alpha_{cr}\left(\frac{F_{cht}}{K_c}\right)^{\beta_r}. \tag{57}$$

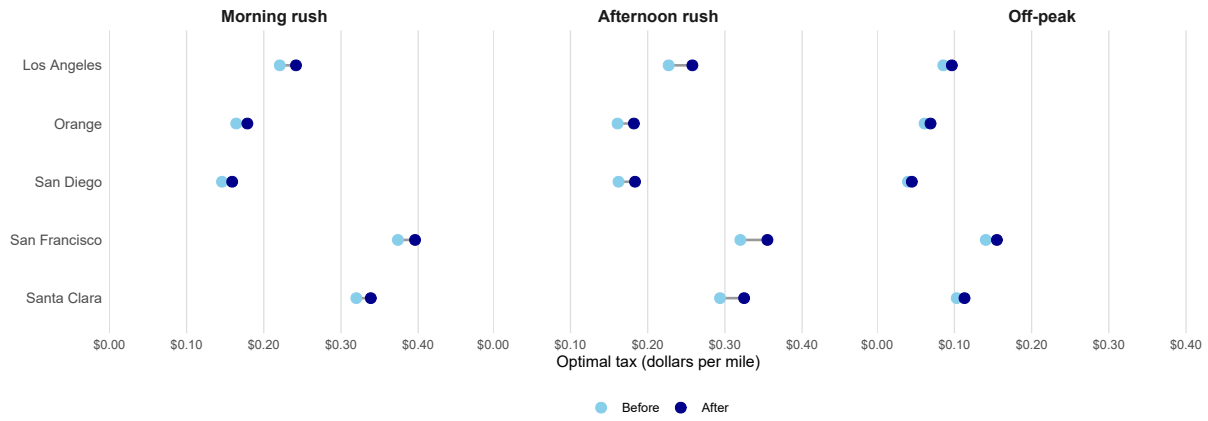
Here $E[\cdot \mid \cdot]$ is a conditional mean, $\alpha_{cr} > 0$ is a county-period congestion scale, and $\beta_r > 0$ is a curvature parameter common to counties within period r. Each period's Bureau of Public Roads (BPR) curve is fitted by Poisson pseudo-maximum likelihood (PPML). Capacity and free-flow speed remain county-specific engineering inputs, while the BPR scale and curvature can differ across periods. This allows the same flow-capacity ratio to produce different expected delay at different times of day, reflecting systematic changes in bottlenecks and network conditions.

Table 6. Congestion Curvature by Time of Day

Period	Hours	$\widehat{\beta}$	HC1 standard error
Morning peak	6:00–8:59	4.240	0.040
Afternoon peak	15:00–17:59	2.394	0.089
Off-peak	Other weekday hours	2.538	0.021

Notes: Separate weekday congestion regressions for Los Angeles, Orange, San Diego, San Francisco, and Santa Clara counties using hourly PeMS data in 2018–2019 and 2021–2022. PeMS is California's Performance Measurement System. Each row estimates $E[D_{cht} \mid F_{cht}, c] = \alpha_{cr}(F_{cht}/K_c)^{\beta_r}$ by Poisson pseudo-maximum likelihood, where r denotes the displayed time period, F_{cht} is traffic in passenger-car-equivalent vehicle miles per hour, and K_c is engineering capacity in the same units. Proportional delay is $D_{cht} = \max\{S_c^0/Q_{cht} - 1, 0\}$, with S_c^0 free-flow speed and Q_{cht} observed speed. The scale varies by county and period; curvature β_r is common across counties within a period. Morning and afternoon regressions each use 15,645 observations, and off-peak uses 93,870. HC1 standard errors allow heteroskedasticity. Larger curvature means delay rises more sharply with traffic relative to capacity. These curves determine the period-specific technology used to solve hourly taxes rather than a single tax assigned to all hours in a period.

Figure 9. Optimal Congestion Taxes by Time Period



Notes: Average weekday optimal taxes for Los Angeles, Orange, San Diego, San Francisco, and Santa Clara counties under contemporaneous values of time. Panels summarize morning peak (6:00–8:59), afternoon peak (15:00–17:59), and all remaining weekday hours. Axes are dollars per vehicle mile. Light-blue points denote 2018–2019 and dark-blue points 2021–2022; points are model-implied rates rather than confidence intervals. Separate congestion curves are estimated for the three periods. Within each period, efficient traffic and its tax are solved for each hour using annual remote-work and population demand shifts, price-elasticity magnitude 0.1, and a value of time equal to half the corresponding county-year hourly wage. The toll is marginal external congestion cost at efficient flow. Taxes are generally larger at peak hours and rise after the demand shift in all 15 county–time-of-day combinations. Sources are PeMS, the American Community Survey, and the demand estimates.

E Welfare Derivation for Income-Group Taxes

The welfare accounting follows the heterogeneous-user congestion-pricing framework of Verhoef and Small (2004), in which demand and the value of time differ across users sharing congested roads. Here I apply that accounting to two income groups on a common network and derive the pricing conditions under the paper's demand specification.

This appendix suppresses the county, hour, and year indices for readability. Groups $g \in \{L, H\}$ denote lower- and higher-income residential tracts, classified using the county's household-weighted 80th percentile of 2018 tract median household income. Group g has flow F_g, inverse demand $P_g(F_g) = (A_g/F_g)^{1/\eta}$ in dollars per mile, demand shifter A_g, and value of time V_g in dollars per hour. Both groups share the price-elasticity magnitude $\eta = 0.1$. Total modeled flow is $F = \sum_g F_g$, measured in passenger-car-equivalent vehicle miles per hour. Travel time $T(F)$ is in hours per mile and $T'(F)$ is its flow derivative. Each V_g equals one-half of the group's hourly earnings proxy.

Let $F_g^{ref} > 0$ be fixed reference flows and x a dummy flow variable. Welfare W, net of travel-time cost and up to an additive constant, is

$$W(F_L, F_H) = \sum_g \int_{F_g^{ref}}^{F_g} P_g(x)\, dx - T(F) \times \sum_g V_g \times F_g. \tag{58}$$

Positive reference flows keep these integrals finite for $\eta = 0.1$ without changing the welfare derivatives. Define $M \equiv \sum_g V_g \times F_g$, the value-of-time-weighted flow exposed to congestion. The welfare derivative for one additional vehicle mile supplied by group q is

$$\frac{\partial W}{\partial F_q} = P_q(F_q) - \underbrace{V_q \times T(F)}_{\text{private time cost}} - \underbrace{T'(F) \times M}_{\text{external congestion cost}}. \tag{59}$$

The index q denotes the group supplying the additional mile. The private term differs with the entrant's value of time. The external term does not depend on the entrant's identity: one additional vehicle mile changes travel time for all existing users. A common monetary toll yields the private condition $P_q(F_q) = V_q \times T(F) + \tau^P$. Comparing it with Equation (59) gives

$$\tau^P = T'(F^P) \times M^P = T'(F^P) \times \sum_g V_g \times F_g^P, \tag{60}$$

where superscript P denotes the common monetary Pigouvian policy, F_g^P its group flows, $F^P = \sum_g F_g^P$, and $M^P = \sum_g V_g \times F_g^P$. The toll τ^P is in dollars per mile. This proves Equation (33). For accounting purposes, this common toll can be written as the sum

$$\tau^P = \sum_g \tau_g^{contrib}, \qquad \tau_g^{contrib} = T'(F^P) \times V_g \times F_g^P. \tag{61}$$

Each contribution measures the portion of marginal external damage borne by group g. It is not the tax rate paid by an entrant from that group.

The distinction between unrestricted and constrained pricing with heterogeneous users is discussed in Small et al. (2024, Section 4.2.3). The particular restriction considered here is $\tau_g = V_g \times \kappa$, where τ_g is the monetary toll paid by group g in dollars per mile and κ is a common time-equivalent toll in hours per mile. The same proportionality restriction underlies the deterministic-lane-choice discount rule in Pandey et al. (2022, Proposition 1). Here I derive the welfare-maximizing charge for the paper's elastic group demands with a common price elasticity. The private equilibrium condition and constant-elasticity group demand are

$$P_g(F_g) = V_g \times [T(F) + \kappa], \qquad F_g = A_g \times V_g^{-\eta} \times [T(F) + \kappa]^{-\eta}. \tag{62}$$

Because both groups share the same elasticity η, their flow shares do not change with a marginal change in κ, holding the demand shifters and values of time fixed:

$$\frac{F_g}{F} = \frac{A_g \times V_g^{-\eta}}{\sum_\ell A_\ell \times V_\ell^{-\eta}}, \qquad \frac{dF_g}{d\kappa} = \frac{F_g}{F} \times \frac{dF}{d\kappa}. \tag{63}$$

Here $\ell \in \{L, H\}$ is a summation index over the two income groups, while g denotes the group whose share is being calculated. Thus the denominator is $A_L \times V_L^{-\eta} + A_H \times V_H^{-\eta}$. Differentiating Equation (58) along the equilibrium response to κ gives

$$\begin{aligned}
\frac{dW}{d\kappa} &= \sum_g P_g(F_g) \times \frac{dF_g}{d\kappa} - T(F) \times \sum_g V_g \times \frac{dF_g}{d\kappa} - T'(F) \times M \times \frac{dF}{d\kappa} \\
&= \kappa \times \sum_g V_g \times \frac{dF_g}{d\kappa} - T'(F) \times M \times \frac{dF}{d\kappa} \\
&= M \times \frac{dF}{d\kappa} \times \left[\frac{\kappa}{F} - T'(F)\right].
\end{aligned} \tag{64}$$

The last equality uses Equation (63), which implies $\sum_g V_g \times (dF_g/d\kappa) = (M/F) \times (dF/d\kappa)$. Because $M > 0$ and $dF/d\kappa < 0$, welfare rises with the time toll below $F \times T'(F)$ and falls above it. The constrained optimum therefore satisfies

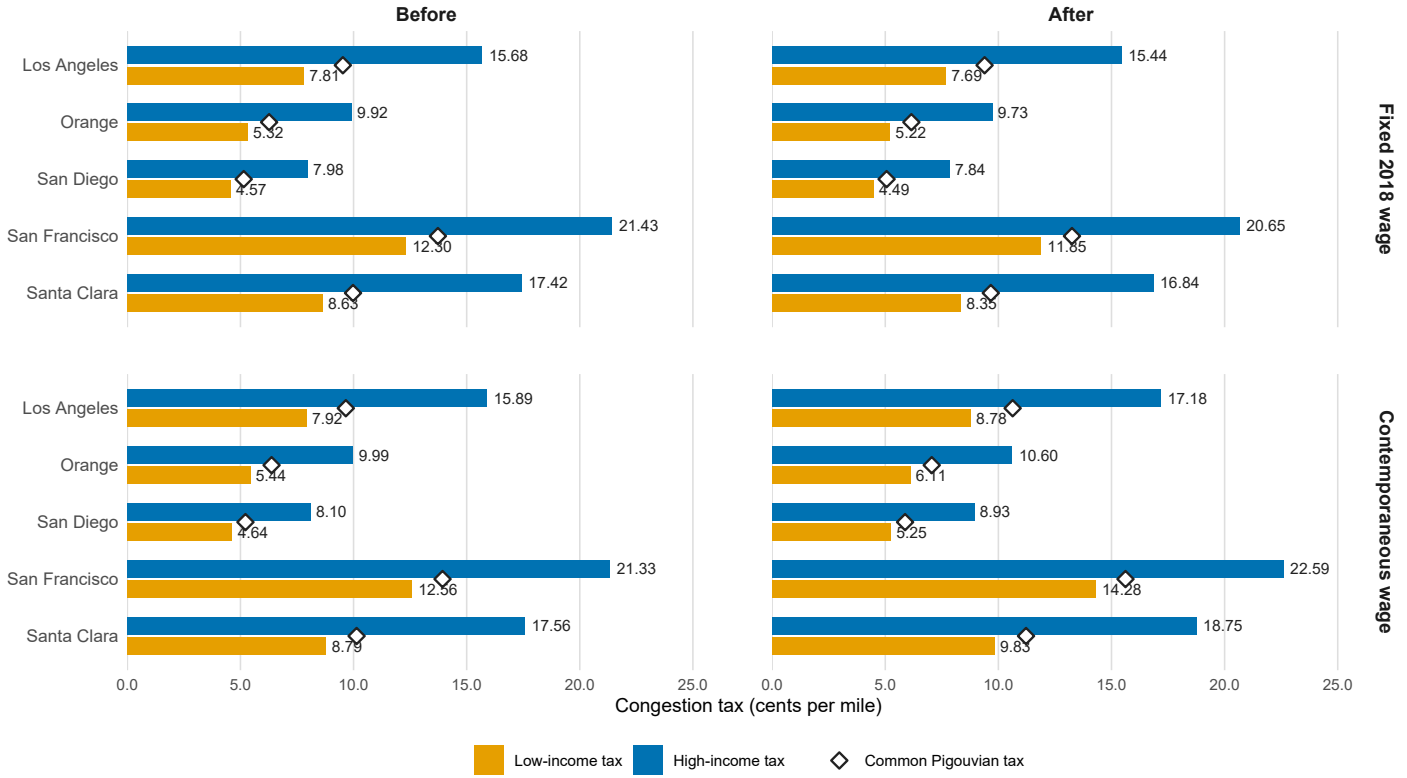
$$\kappa^* = F^\kappa \times T'(F^\kappa). \tag{65}$$

Here F^κ denotes total equilibrium flow under the constrained time-equivalent policy, and a star denotes its optimal time charge. This establishes Equation (36). For the Bureau of Public Roads (BPR) technology, $T(F) = (1/S^0) \times [1 + \alpha \times (F/K)^\beta]$, where S^0 is free-flow speed in miles per hour, K is capacity in flow units, α is the congestion scale, and β is curvature. The monetary-rate superscript V labels the value-of-time-proportional policy. Its rates satisfy

$$\kappa^* = \frac{\alpha \times \beta}{S^0} \times \left(\frac{F^\kappa}{K}\right)^\beta, \qquad \tau_g^V = V_g \times \kappa^*. \tag{66}$$

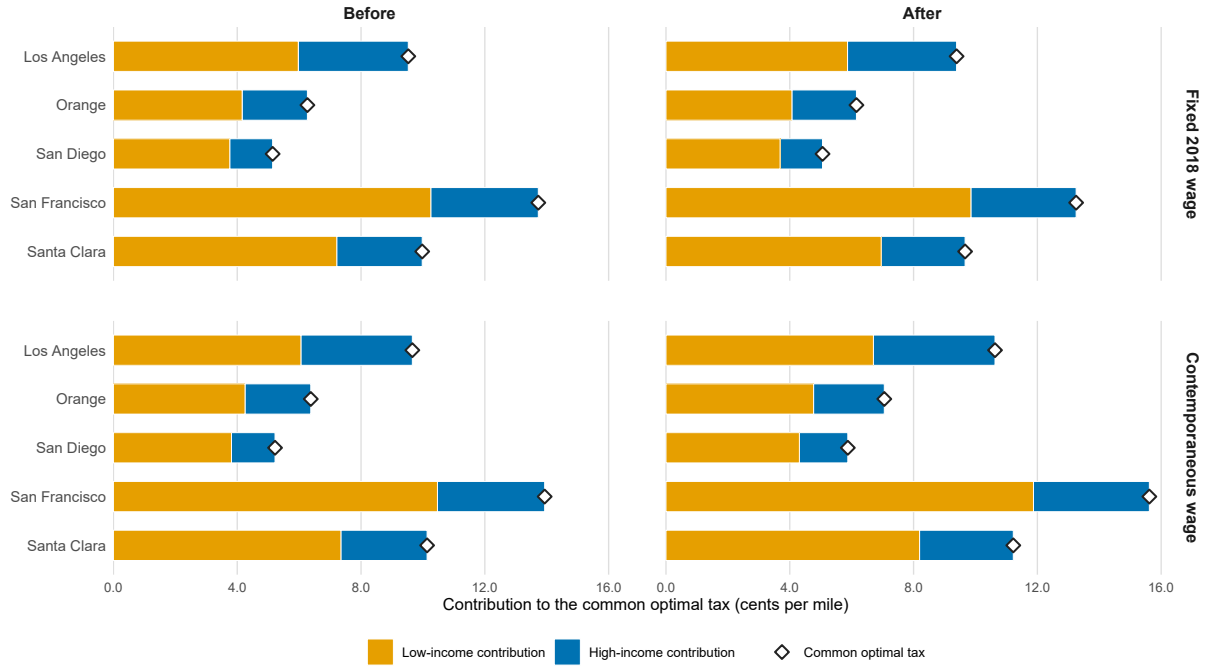
The fixed-wage exercise holds V_g at its 2018 group value in every year; the contemporaneous exercise updates it using that group's year-specific wage. If group elasticities differ, Equation (63) no longer holds, and the constrained optimum must instead be obtained from the unrestricted derivative in the second line of Equation (64).

Figure 10. Value-of-Time-Proportional Taxes by Period and Wage Scenario



Notes: Group-specific monetary rates under the common time-equivalent policy for Los Angeles, Orange, San Diego, San Francisco, and Santa Clara counties. Columns compare 2018–2019 (before) and 2021–2022 (after). Upper panels hold each group's wage at its 2018 value; lower panels use contemporaneous wages. Orange bars are lower-income rates and blue bars higher-income rates; numbers and axes are cents per vehicle mile. Higher-income tracts are above the household-weighted county 80th percentile of 2018 median household income, with membership fixed thereafter. Group value of time V_g is half estimated hourly earnings. The common time charge κ is hours per mile and the monetary rate is $\tau_g^V = V_g\kappa$. Both demands and congestion are solved jointly with price-elasticity magnitude 0.1. Diamonds show the unconstrained common monetary Pigouvian toll from its own equilibrium. Higher-income monetary rates are larger under the proportional policy because time values are higher. These are toll-rate comparisons rather than estimates of net welfare changes by group.

Figure 11. Income-Group Contributions to the Common Optimal Tax



Notes: Decomposition of the common monetary Pigouvian toll into marginal delay damages borne by lower- and higher-income travelers in Los Angeles, Orange, San Diego, San Francisco, and Santa Clara counties. Columns compare 2018–2019 (before) and 2021–2022 (after). Upper panels fix group wages at 2018 levels and lower panels use contemporaneous wages. Orange and blue segments are lower- and higher-income components; axes are cents per vehicle mile. Rows run from Los Angeles at the top through Orange, San Diego, San Francisco, and Santa Clara. At the common-toll efficient allocation $F^P = \sum_g F_g^P$, group g's component is $T_c'(F^P)V_g F_g^P$, where V_g is its value of time and F_g^P its traffic. Segments sum to the diamond, the common monetary toll. Higher-income tracts are above the household-weighted county 80th percentile of 2018 median household income; remaining tracts form the lower-income group. Time values equal half hourly earnings and price-elasticity magnitude is 0.1. Segments allocate exposure to external-delay damages, not separate tax rates charged to groups.

F Adjusted Before–After Evidence

Let c denote county, h an observed hour, and t its calendar date, which determines the year and month. Vehicle miles traveled (VMT) measure hourly traffic, Q is average speed in miles per hour, and the travel time index (TTI) is observed travel time divided by the free-flow reference time. Higher TTI means slower travel relative to that benchmark. For each outcome $Y_{cht} \in \{VMT_{cht}, Q_{cht}, TTI_{cht}\}$, I estimate

$$\log Y_{cht} = \delta Post_t + \mu_{c,\omega(h,t)} + \lambda_{c,m(t)} + \varepsilon_{cht}, \tag{67}$$

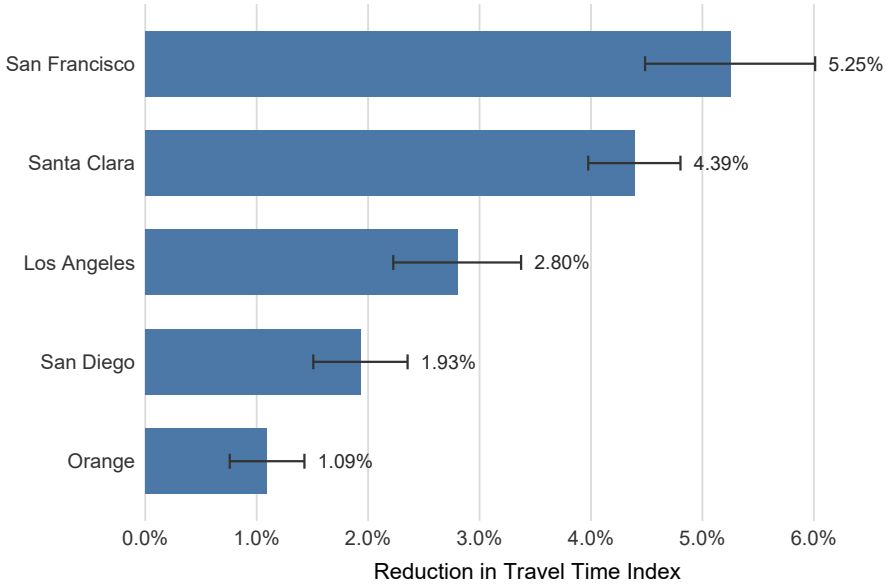
where $Post_t$ equals one in 2021–2022 and zero in 2018–2019, $\mu_{c,\omega(h,t)}$ denotes county-by-hour-of-week fixed effects, and $\lambda_{c,m(t)}$ denotes county-by-calendar-month fixed effects. The coefficient δ is the adjusted post-period change in log outcome, ε_{cht} is the regression error, $\omega(h,t)$ identifies the hour of the week, and $m(t)$ the calendar month. Here μ and λ denote the two sets of fixed effects, not demand parameters. Standard errors are clustered by year-month. The fixed effects ensure that the post-period comparison holds constant each county's recurring weekly traffic profile and seasonal pattern. Equation (67) remains a descriptive comparison because the post indicator can capture any contemporaneous change, not only remote work.

Table 7. Adjusted Changes in Freeway Traffic and Congestion

	All hours	Afternoon peak	Weekends
Panel A: Log vehicle miles traveled			
Post period	-0.080***	-0.039***	-0.071***
	(0.009)	(0.005)	(0.011)
Implied change (%)	-7.71	-3.87	-6.81
Panel B: Log average speed			
Post period	0.030***	0.078***	0.008***
	(0.002)	(0.006)	(0.003)
Implied change (%)	3.08	8.07	0.76
Panel C: Log Travel Time Index			
Post period	-0.032***	-0.077***	-0.009***
	(0.002)	(0.006)	(0.002)
Implied change (%)	-3.11	-7.44	-0.92
Observations	175,180	15,645	50,020
County × hour-of-week fixed effects	Yes	Yes	Yes
County × calendar-month fixed effects	Yes	Yes	Yes

Notes: Hourly PeMS freeway observations for Los Angeles, Orange, San Diego, San Francisco, and Santa Clara counties, comparing 2018–2019 with 2021–2022 and excluding 2020. PeMS is California's Performance Measurement System. Each entry labeled Post period is the coefficient on an indicator equal to one in 2021–2022 in a separate log-outcome regression. Panel A uses vehicle miles traveled, Panel B average speed in miles per hour, and Panel C the travel time index, travel time divided by its free-flow reference. Each regression includes county-by-hour-of-week and county-by-calendar-month fixed effects. Parentheses contain standard errors clustered by year-month (48 clusters). Afternoon peak is weekdays 15:00–17:59; weekends include all Saturday and Sunday hours. Implied changes use $100[\exp(\widehat{\delta}) - 1]$ applied to the unrounded coefficient. Negative changes in traffic and the time index, together with positive speed changes, indicate reduced congestion. These are adjusted descriptive comparisons rather than causal remote-work effects. Stars denote $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Figure 12. Adjusted Post-Pandemic Reduction in Congestion by County

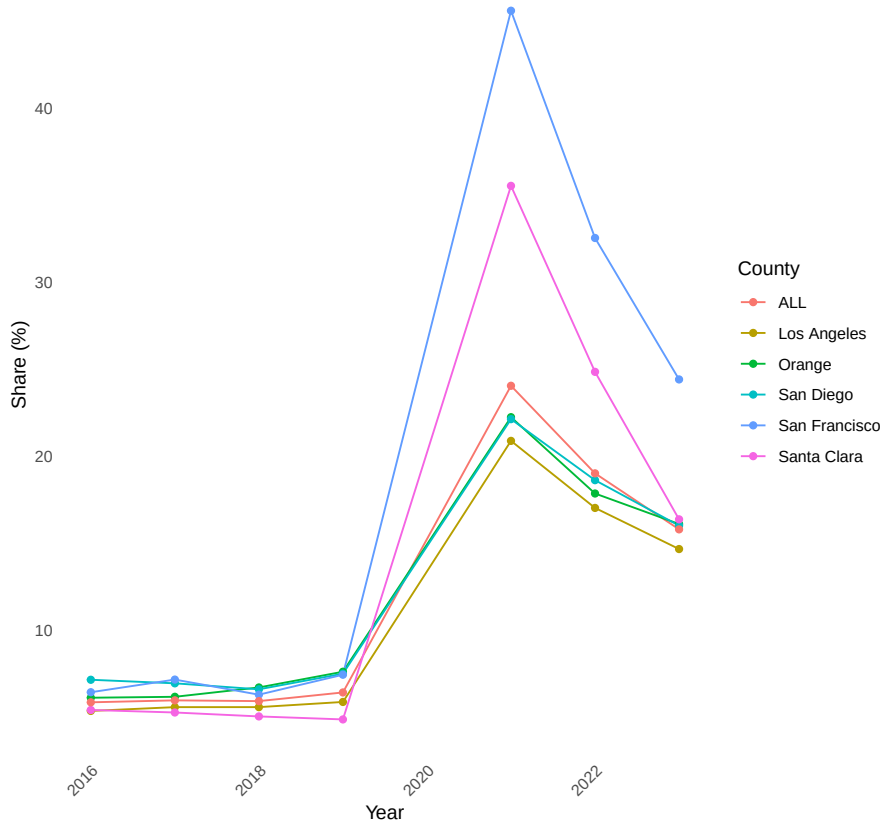


Notes: Adjusted descriptive reductions in the travel time index (TTI) for Los Angeles, Orange, San Diego, San Francisco, and Santa Clara counties between 2018–2019 and 2021–2022, excluding 2020. TTI is travel time divided by free-flow reference time, so a smaller value means less congestion. Each county's log-TTI regression includes a post-period indicator, hour-of-week fixed effects, and calendar-month fixed effects. Bars convert the post coefficient $\widehat{\delta}_c$ to a reduction, $100[1 - \exp(\widehat{\delta}_c)]$, so positive values mean less congestion. Whiskers are 95 percent confidence intervals based on standard errors clustered by year-month. The horizontal axis is percent and county names label bars. Source is hourly PeMS freeway traffic. These are adjusted before–after changes, not causal remote-work estimates. San Francisco has the largest estimated reduction and Orange the smallest.

G Additional Descriptive Figures

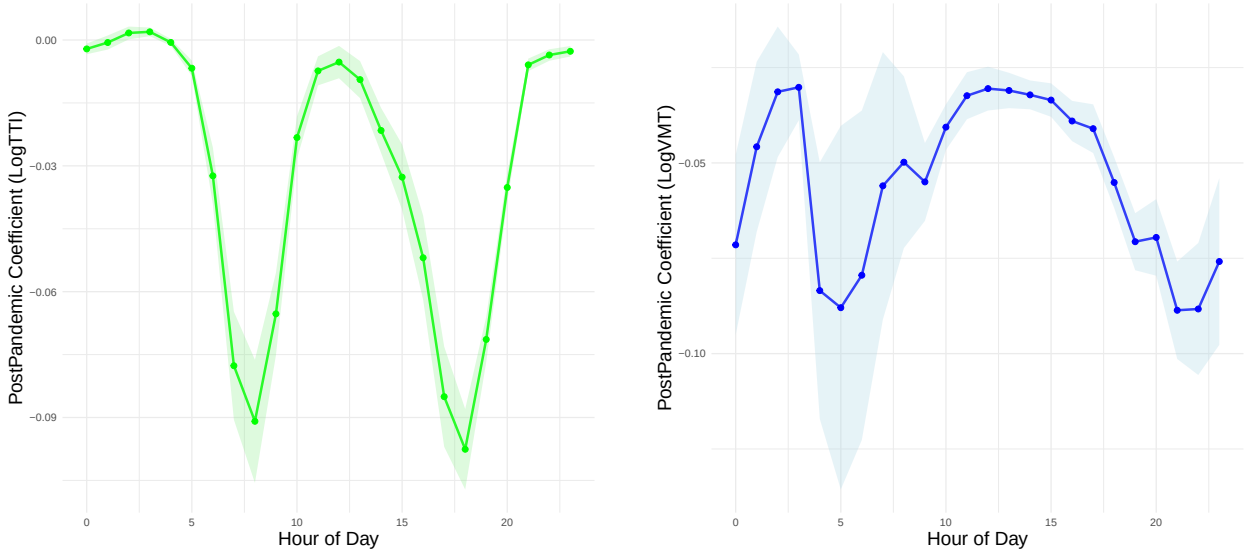
The American Community Survey (ACS) supplies the work-from-home shares. PeMS, California's Performance Measurement System, supplies hourly freeway traffic. VMT denotes vehicle miles traveled, and TTI denotes the travel time index, observed travel time relative to the free-flow reference time. The hourly panels report descriptive before–after regression coefficients for traffic and congestion at each clock hour.

Figure 13. Work-from-Home Shares by County



Notes: Source is the American Community Survey one-year county estimates for 2016–2019 and 2021–2023. The vertical axis is the percent of workers aged 16 or older who usually worked from home and the horizontal axis is survey year. Colored lines and points identify Los Angeles, Orange, San Diego, San Francisco, and Santa Clara counties. The series labeled ALL pools workers across these five counties, dividing total home workers by total workers; it does not denote the United States or all of California. Lines connect observed years, with no 2020 estimate plotted. These county-level descriptive shares are distinct from the tract-level five-year survey estimates used in the demand regressions. The figure shows the persistent rise in remote work after 2019 and substantial variation across counties.

Figure 14. Hourly Before–After Changes in Traffic and Congestion



(a) Travel Time Index

(b) Vehicle Miles Traveled

Notes: Descriptive before–after regressions using hourly PeMS freeway data for Los Angeles, Orange, San Diego, San Francisco, and Santa Clara counties. The underlying 2018–2022 sample defines before as dates preceding March 19, 2020, and after as dates following January 24, 2021, omitting the intervening disruption. Observations are averaged across counties for each calendar date and clock hour. For each hour separately, log outcome is regressed on an after-period indicator and a constant. Panel (a), in green, uses the travel time index (TTI), travel time relative to its free-flow reference. Panel (b), in blue, uses vehicle miles traveled (VMT). Points and lines are post-period log coefficients; shaded bands are conventional 95 percent ordinary-least-squares confidence intervals. A coefficient b corresponds to a percent change of $100[\exp(b) - 1]$. Negative values indicate less congestion or traffic. These unadjusted descriptive regressions differ from the full-year, fixed-effects comparisons elsewhere and do not identify remote-work effects.