# Designing High-Occupancy Toll Lanes: A Game-Theoretic Analysis

Zhanhao Zhang[a,*], Ruifan Yang[a], Manxi Wu[b]

[a] Operations Research and Information Engineering, Cornell University

[b] Department of Civil and Environmental Engineering, University of California, Berkeley

[*] Corresponding author

*Extended abstract submitted for presentation at the 12th Triennial Symposium on Transportation Analysis conference (TRISTAN XII)*
*June 22-27, 2025, Okinawa, Japan*

---

Keywords: High occupancy toll lane design, Non-atomic games, Equilibrium analysis with heterogeneous preferences, Comparative statics.

## 1 Introduction

High Occupancy Toll (HOT) lanes have been successfully implemented in states like California, Texas, and Washington to reduce traffic congestion by incentivizing carpooling and generating revenue for infrastructure. These lanes provide access to vehicles that meet a minimum occupancy requirement, as well as to others willing to pay a toll. In this article, we develop a game-theoretic model to analyze the strategic decisions of travelers, who choose between using HOT lanes by either paying the toll or carpooling, or opting for ordinary lanes. While previous studies have examined drivers' equilibrium behavior and HOT lane design (Konishi & il Mun (2010), Yang & Huang (1999), Lou *et al.* (2011), Zhong *et al.* (2020) and many others), they often assume simplified latency functions or homogeneous values of time and/or carpool disutilities. There also lacks complete characterization of equilibrium structure and data validation.

Our model addresses this gap by considering heterogeneous traveler preferences, specifically a continuous distribution of values of time and carpooling disutility, and focuses on optimizing HOT lane design to achieve various goals. We provide a complete equilibrium characterization, using general travel time cost functions and preference distributions, and conduct comparative statics to assess how changes in system parameters affect traffic flows and costs. We also extend our equilibrium analysis to the setting with multiple highway segments. To connect theory with practice, we apply our model to real-world data from the I-880 HOT lane in California. Using inverse optimization, we estimate traveler preferences and demand, and compute optimal HOT lane designs under realistic conditions, demonstrating how HOT lane systems can be tailored for different objectives, such as minimizing congestion or maximizing revenue.

## 2 Model and equilibrium analysis

**2.1 A basic model.** Consider a highway segment consisting of *ordinary* and *high occupancy toll* (HOT) lanes, where the ordinary lane is toll-free, and the high occupancy toll lane is accessible to vehicles that either pay the toll $\tau \in \mathbb{R}_{\geq 0}$ or meet the minimum occupancy requirement with passenger size higher or equal to $A \geq 2$. A central planner (e.g. transportation authority) determines the toll price $\tau$, the minimum occupancy requirement $A$, and the allocation of road capacity $\rho \in [0,1]$ between HOT lanes and ordinary lanes. The latency function of the HOT lanes $\ell_{\mathrm{h}}(x_{\mathrm{h}}, \rho)$ and the ordinary lanes $\ell_{\mathrm{o}}(x_{\mathrm{o}}, 1-\rho)$ are both increasing in the lane flow $x_{\mathrm{h}}$ and $x_{\mathrm{o}}$, respectively. The two types of lanes have equal free flow travel time $\ell_{\mathrm{h}}(0,\rho) = \ell_{\mathrm{o}}(0, 1-\rho)$.

Travelers are non-atomic agents with total demand of $D > 0$ and action set $A = \{\text{toll}, \text{pool}, \text{o}\}$, where toll (resp. pool) is the action of taking the HOT lanes by paying the toll price (resp. meeting occupancy requirement), and o is to take the ordinary lane. Agents have heterogeneous value of time $\beta \in \mathrm{B} = [0, \bar{\beta}]$, and heterogeneous carpool disutility $\gamma \in \Gamma = [0, \bar{\gamma}]$. Agents' preference parameters $(\beta, \gamma)$ are continuously distributed with probability density function $f : \mathrm{B} \times \Gamma \to \mathbb{R}$ such that $f(\beta, \gamma) > 0$ for all $(\beta, \gamma)$ and $\int_{\mathrm{B}\times\Gamma} f(\beta, \gamma) d\beta d\gamma = 1$.

An agent's strategy $s$ is a mapping from their preference parameters $(\beta, \gamma)$ to a pure strategy in action set $A$, denoted as $s : \mathrm{B} \times \Gamma \to A$. Given $s$, we can compute the population strategy distribution as $\sigma = (\sigma_a)_{a\in A}$, where $\sigma_a = \frac{1}{D} \int_{\mathrm{B}\times\Gamma} f(\beta, \gamma) \mathbb{1}\{s(\beta, \gamma) = a\} d\beta d\gamma$ is the fraction of agents who choose each action $a \in A$. The flow on each type of lanes induced by $\sigma$ is $x_{\mathrm{h}} = \left(\sigma_{\text{toll}} + \frac{\sigma_{\text{pool}}}{A}\right) D,\ x_{\mathrm{o}} = \sigma_{\mathrm{o}} D$. The cost of each agent with preference parameters $(\beta, \gamma)$ for choosing actions toll (resp. pool, o) is given by: $C_{\text{toll}}(\sigma, \beta, \gamma) = \beta \cdot \ell_{\mathrm{h}}(x_{\mathrm{h}}, \rho) + \tau$ (resp. $C_{\text{pool}}(\sigma, \beta, \gamma) = \beta \cdot \ell_{\mathrm{h}}(x_{\mathrm{h}}, \rho) + \gamma,\ C_{\mathrm{o}}(\sigma, \beta, \gamma) = \beta \cdot \ell_{\mathrm{o}}(x_{\mathrm{o}}, 1 - \rho)$), where $\beta \cdot \ell_{\mathrm{h}}(x_{\mathrm{h}}, \rho)$ (resp. $\beta \cdot \ell_{\mathrm{o}}(x_{\mathrm{o}}, 1 - \rho)$) represents the cost of enduring the latency on the HOT (resp. ordinary) lanes. The toll payment and the carpool disutility is added for action toll and pool, respectively.

**2.2 Equilibrium characterization.** A strategy profile $s^* : \mathrm{B}\times\Gamma \to A$ is a *Wardrop equilibrium* if no agent has incentive to deviate. Formally,

$$s^*(\beta, \gamma) = a, \quad \Rightarrow \quad C_a(\sigma^*, \beta, \gamma) = \mathrm{argmin}_{a'\in A} C_{a'}(\sigma^*, \beta, \gamma),\ \forall (\beta, \gamma) \in \mathrm{B} \times \Gamma,$$

where $\sigma^*$ is the associated equilibrium strategy distribution. We next present a complete equilibrium characterization. Our result shows that equilibrium exhibits *two qualitatively different regimes* depending on the game parameters. In particular, the threshold value of latency difference $\ell_\delta^\dagger := \ell_{\mathrm{o}}(\sigma^\dagger, 1 - \rho) - \ell_{\mathrm{h}}(\sigma^\dagger, \rho)$ is central in separating different equilibrium regimes, where

$$\sigma^\dagger_{\text{toll}} = 0, \quad \sigma^\dagger_{\text{pool}} = \int_0^{\bar{\beta}} \int_0^{\min\{\tau, \bar{\gamma}\}\beta/\bar{\beta}} f(\beta, \gamma) d\gamma d\beta, \quad \sigma^\dagger_{\mathrm{o}} = 1 - \sigma^\dagger_{\text{pool}}.$$

We present the complete equilibrium characterization in Theorem 1 that can be used to compute equilibrium strategy distribution $\sigma^*$.

**Theorem 1 (Equilibrium regimes)** *The game has a unique Wardrop equilibrium.*
*Regime A: The toll price $\tau$ is relatively high, i.e. $\tau \geq \min\left\{\bar{\gamma}, \bar{\beta}\ell_\delta^\dagger\right\}$. No agent takes HOT lanes by paying the toll, i.e. $\sigma^*_{\text{toll}} = 0$. Furthermore,*

$$\sigma^*_{\text{pool}} = \begin{cases} \int_0^{\bar{\beta}} \int_0^{\ell_\delta(0, \sigma^*_{\text{pool}}, 1-\sigma^*_{\text{pool}}, \rho)\beta} f(\beta, \gamma) d\gamma d\beta, & \text{if } \bar{\beta}\ell_\delta^\dagger \leq \bar{\gamma}, \\ 1 - \int_0^{\bar{\gamma}} \int_0^{\gamma/\ell_\delta(0, \sigma^*_{\text{pool}}, 1-\sigma^*_{\text{pool}}, \rho)} f(\beta, \gamma) d\beta d\gamma, & \text{if } \bar{\beta}\ell_\delta^\dagger > \bar{\gamma}, \end{cases} \qquad \sigma^*_{\mathrm{o}} = 1 - \sigma^*_{\text{pool}}.$$

*Regime B: The toll price $\tau$ is relatively low, i.e. $0 < \tau < \min\left\{\bar{\gamma}, \bar{\beta}\ell_\delta^\dagger\right\}$. All three actions are taken, and $\sigma^*$ is the unique solution that satisfies the following equations:*

$$\sigma^*_{\text{toll}} = \int_\tau^{\bar{\gamma}} \int_{\tau/\ell_\delta(\sigma^*, \rho)}^{\bar{\beta}} f(\beta, \gamma) d\beta d\gamma,\ \sigma^*_{\text{pool}} = \int_0^\tau \int_{\gamma/\ell_\delta(\sigma^*, \rho)}^{\bar{\beta}} f(\beta, \gamma) d\beta d\gamma,\ \sigma^*_{\mathrm{o}} = 1 - (\sigma^*_{\text{toll}} + \sigma^*_{\text{pool}}).$$

**Theorem 2 (Comparative statics)** *Equilibrium changes with $(\rho, \tau)$ as follows:*

| | Fix $\tau$ increase $\rho$ | Fix $\rho$ increase $\tau$ |
|---|---|---|
| $\sigma^*_{\mathrm{o}}$ | Decreasing | Either Direction |
| $\sigma^*_{\text{toll}}$ | Increasing | Non-Increasing |
| $\sigma^*_{\text{pool}}$ | Increasing | Non-Decreasing |
| $\ell_\delta(\sigma^*, \rho)$ | Increasing | Non-Decreasing |

**2.3 Multi-segment extension.** We now extend the basic model and equilibrium analysis to multiple segments $e \in [E] := 1, \ldots, E$, and multiple occupancy levels $m \in [M] := 1, \ldots, M$. The toll price of segment $e \in [E]$ with occupancy level $m$ is $\tau_{e,m} \geq 0$. Agents split the toll price evenly. For each pair of $i \leq j \in [E]$, agent population (denoted as $(i, j)$) with demand $D^{ij}$ enter the highway from the beginning of segment $i$ and leave at the ending of segment $j$. Agents in each population $(i, j)$ decides their carpool size $a_{\text{occu}} \in [M]$ and whether to take the HOT lane $(a_e = \text{h})$ or the ordinary lane $(a_e = \text{o})$ in each segment $e \in [i : j]$. The heterogeneous preference of agents includes their value of time $\beta \in [0, \overline{\beta}]$ and carpool disutilities $\gamma := (\gamma_m)_{m \in [M]}$, where $\gamma_m \in [0, \overline{\gamma}_m]$ is the disutilities of occupancy level $m$. We define agents' strategy and Wardrop equilibrium following the basic model.

We next provide conditions under which equilibrium is unique in the multi-segment setting. An agent's equilibrium strategy $s^{*ij}(\beta, \gamma)$ depends on the segment latency of each lane, but only through the difference between them. We define $\delta = (\delta_e)_{e \in E}$ as latency difference vector, where $\delta_e$ is the latency of the ordinary lane exceeding that of HOT lane in segment $e$. Given $\delta$ and $\tau$, we compute the unique best response of all agents, and the aggregate best response lane flow vector $x(\delta)$. We define $\Phi_e(\delta) = \ell_{o,e}(x_{o,e}(\delta)) - \ell_{h,e}(x_{h,e}(\delta))$ as the latency cost difference of $e$ induced by best responses given $\delta$, and $\Phi(\delta) = (\Phi_e(\delta))_{e \in E}$. We observe a one-to-one correspondence between the fixed point solution of $\Phi(\delta) = \delta$, the equilibrium latency cost difference $\delta^*$, and the equilibrium strategy $s^*$. The following theorem shows that equilibrium exists and is unique in the multi-segment model when $\Phi(\cdot)$ satisfies certain condition. The function $\Phi(\cdot)$ plays a central role in equilibrium computation, which is omitted due to space limit.

**Theorem 3** *Given any toll price vector $\tau$, Wardrop equilibrium $s^*$ exists. Moreover, $s^*$ is unique if the Jacobian matrix $\nabla \Phi(\delta)$ does not have 1 as its eigenvalue for any $\delta$.*

# 3 Empirical study of HOT design on California I-880

**3.1 Cost and preference distribution calibration.** We calibrate multi-segment model using real data of I-880 from Dixon Landing Rd and Lewelling Blvd (Fig. 1) from 5am to 8pm $t \in [T]$ on workdays $n \in [N]$ between March 1st 2021 and August 31st 2021. We use traffic sensor data (PEMS) to calibrate the latency function of ordinary and HOT lanes based on the Bureau of Public Roads (BPR) function. We apply *inverse optimization* to estimate the hourly population demand and preference distribution using toll price (every 5min) and HOT vehicle occupancy data (daily aggregate) requested from Caltrans. For tractability, we evenly grid the preference parameter vector set into $K$ subsets. We estimate $d := (d_k^{ij,t})_{i \leq j \in [E], t \in [T], k \in [K]}$, where $d_k^{ij,t}$ is the product of population demand $D^{ij}$ at time $t$ and the fraction of agents with preference parameters in subset $k$. We estimate $d$ as the vector such that the equilibrium vehicle flow $x^{*t,n}(d)$ is close to the observed flows $\hat{x}^{t,n}$ and the average daily equilibrium occupancy level distribution on HOT lanes $y_m^{*n}(d)$ is close to the observed ones $\hat{y}_m^n$:

$$\min_d \sum_{n \in [N]} \sum_{t \in [T]} \sum_{e \in [E]} \left( \left( x_{\text{o},e}^{*t,n}(d) - \hat{x}_{\text{o},e}^{t,n} \right)^2 + \left( x_{\text{h},e}^{*t,n}(d) - \hat{x}_{\text{h},e}^{t,n} \right)^2 \right) + \sum_{n \in [N]} \sum_{m \in [M]} \left( y_m^{*n}(d) - \hat{y}_m^n \right)^2 ,$$

where the equilibrium flow $x^*$ and derived equilibrium occupancy distribution $y^*$ of each hour and day are computed using the equilibrium analysis in the multi-segment model. Fig. 2 shows that the observed occupancy level distribution (dotted lines) is closely aligned with the equilibrium distribution (curved lines) given the calibrated vector $d^*$.

**3.2 Optimal toll design.** We compute the optimal toll design to achieve each of the following four objectives: minimize the total agent travel time, minimize the total vehicle driving time, maximize the total toll revenue, and minimize the total cost (driving time, toll price and carpool disutility) of all agents. Figure 3 demonstrates the optimal toll prices (red curves) in each hour

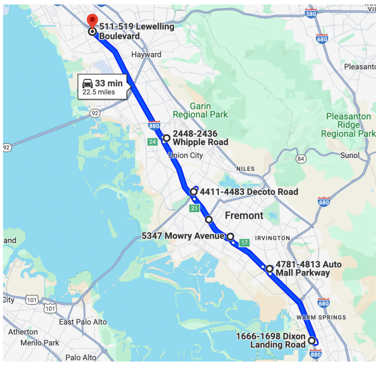



Figure 1 – *I880 HOT segments.*

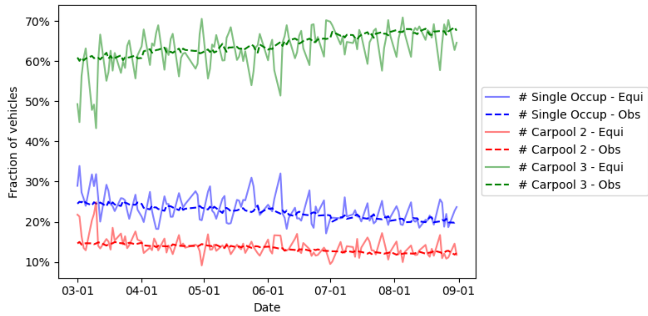
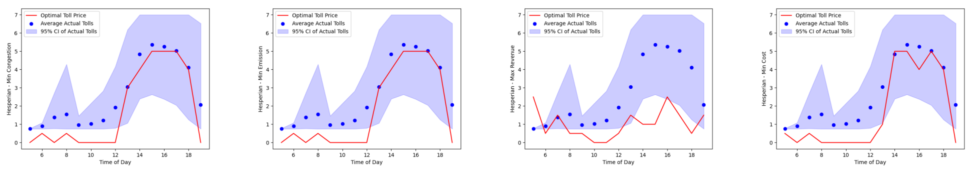



Figure 2 – *Comparison between equilibrium and actual percentage of travelers with each occupancy level.*

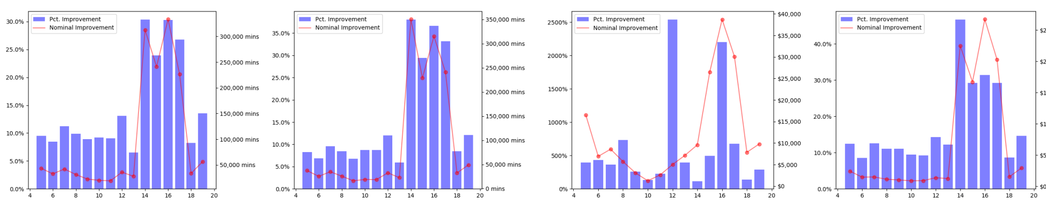
from 5 am to 8 pm, and compare that to the actual prices from the 2021 operations (blue dots for the mean value and blue shaded regions for the 95% confidence interval).

For space limit, we only illustrate the optimal toll prices for the most congested segment Hesperian Blvd even though we have computed the optimal toll on all segments. We find that the optimal toll price is lower than the current ones during non-rush hours and about the same during rush hours for minimizing total agent travel time, vehicle driving time and total cost. The optimal toll price for revenue maximization is consistently lower than the current toll price as well as the optimal price associated with the other three objectives. This is because charging a high toll price leads agents to either carpool or take the ordinary lane, leaving fewer agents willing to pay the toll. We illustrate that the optimal toll prices achieves significant improvements in percentage (blue bars) and numerical value (red curves) of each objective summed over all five segments in Figure 4.



(a) *Agent Time* (b) *Vehicle Time* (c) *Revenue* (d) *Cost*

Figure 3 – *Optimal hourly toll price of each objective and actual price on Hesperian segment.*



(a) *Agent Time* (b) *Vehicle Time* (c) *Revenue* (d) *Cost*

Figure 4 – *Total improvement of each objective by optimal toll price.*

# References


Konishi, Hideo, & il Mun, Se. 2010. Carpooling and congestion pricing: HOV and HOT lanes. *Regional Science and Urban Economics*, **40**(7), 173–186.

Lou, Yingyan, Yin, Yafeng, & Laval, Jorge A. 2011. Optimal dynamic pricing strategies for high-occupancy/toll lanes. *Transportation Research Part C: Emerging Technologies*, **19**(1), 64–74.

Yang, Hai, & Huang, Hai-Jun. 1999. Carpooling and congestion pricing in a multilane highway with high-occupancy-vehicle lanes. *Transportation Research Part A: Policy and Practice*, **33**(2), 139–155.

Zhong, Lin, Zhang, Kenan, Nie, Yu (Marco), & Xu, Jiuping. 2020. Dynamic carpool in morning commute: Role of high-occupancy-vehicle (HOV) and high-occupancy-toll (HOT) lanes. *Transportation Research Part B: Methodological*, **135**(5), 98–119.