COLUMBIA UNIVERSITY
Center for Pricing and Revenue Management

WORKING PAPER SERIES: NO. 2015-01

Revenue Maximizing Dynamic Tolls for Managed Lanes: A Simulation Study

Caner Gocmen
Nomis Solutions

Robert Phillips
Columbia University, Nomis Solutions

Garrett van Ryzin,
Columbia University

2015

http://www.cprm.columbia.edu

Revenue Maximizing Dynamic Tolls for Managed Lanes: A Simulation Study

Caner Göçmen*
Robert Phillips†
Garrett van Ryzin‡

October 28, 2015

Abstract

In recent years, public-private partnership schemes for highway construction have become increasingly popular. In a typical private-public partnership, a private company builds additional lanes on existing highways in return for the right to charge a toll on the additional lane for a specified period of time and to keep all or part of the resulting revenue. We address the question of how an operator should set and update tolls in order to maximize expected revenue when drivers have access to a free alternative. We address this problem through stochastic simulation of a freeway with both toll lanes and free lanes. We assume that drivers choose whether to travel on the toll (*managed)* lane or the free (*unmanaged*) lane based on the current congestion in each lane and on the current toll. We use a mesoscopic traffic model to represent the traffic dynamics in each lane and calibrate the model using data from the SR 91 highway in Orange County, California. Our baseline is a *myopic policy* in which the operator sets tolls to maximize expected revenue from each vehicle. We compare this policy with time-of-use policies that can anticipate the likely pattern of future demand and consider both *non-adaptive* policies which cannot update the toll based on current conditions and *adaptive* policies which can. We find that the best-performing policies raise tolls prior to anticipated peaks in order to divert traffic to the unmanaged lanes and thereby increase congestion on those lanes and decrease congestion on the managed lanes – an approach we call *jam-and-harvest.* When a peak is present, the myopic policy compares poorly to non-adaptive policies that anticipate expected demand but do not adapt to current conditions. We confirm and extend these observations using simplified stylized models.

*Nomis Solutions
†Columbia Business School and Nomis Solutions
‡Columbia Business School

In an era of increasing traffic congestion and limited budgets for infrastructure improvement, the idea of allowing a private company to foot the bill for a highway project in return for a share of future toll revenues is tempting to many public transportation agencies. In the past five years a handful of such projects have been built in the United States. Examples include the LBJ Freeway (near Dallas Texas) and the 495 Express (in Northern Virginia near Washington DC). A number of similar projects have been proposed or are under construction. As of 2012, 32 states and Puerto Rico had passed legislation enabling such public-private partnerships for highway construction (Perez et al. (2012)). The vast majority of these projects consist of a *managed lane scheme* in which drivers have a choice between a number of *managed lanes* for which a usage toll is charged and a number of *unmanaged lanes* that are always free to use. (A managed lane scheme is sometimes called a *high occupancy and toll* (HOT) scheme. Under this nomenclature, the toll lanes are called the *HOT lanes* and the free lanes are called the *general purpose* or *GP* lanes.) In a managed lane scheme, the driver always has the choice of choosing the parallel unmanaged lanes if she does not want to pay the toll – this distinguishes managed lane schemes from traditional toll roads in which the only alternative available to the driver is to take a different route. The motivation of an arriving driver to choose the managed lanes is the possibility of less congestion and faster travel time than if she chooses the unmanaged lanes. In this paper we consider the problem of setting and updating the tolls for a managed lane in order to maximize expected revenue.

A driver approaching the managed lanes is informed of the current toll by large digital signs. These are placed sufficiently far from the entrance that the driver has the time to decide whether to chose the managed or unmanaged lanes. The frequency of toll changes is governed by regulation – for current projects the minimum interval between toll changes ranges from three to five minutes. Managed lane schemes use a gantry-based toll system in which a transponder is used to identify vehicles at the point of entry. If an entering vehicle does not carry a transponder, the license plate is photographed and optical character recognition technology is used to identify the license plate number of the vehicle. There are no toll booths and there is no need for vehicles to slow down as they enter the managed lanes. Vehicle owners are typically billed monthly for the tolls accumulated in the prior month.

In this paper, we consider the problem faced by an operator who is seeking to set the tolls of a set of managed lanes over time in order to maximize revenue. Because the incremental cost incurred by an additional vehicle using the managed lanes is essentially zero, maximizing revenue is equivalent to maximizing short-run profit. In managed lane schemes, the current traffic conditions on both the managed and unmanaged lanes are continually monitored by sensors and the operator can use this information in setting the toll. In practice, operators will also have additional information available on factors such as weather or lane closures that could influence the revenue-maximizing

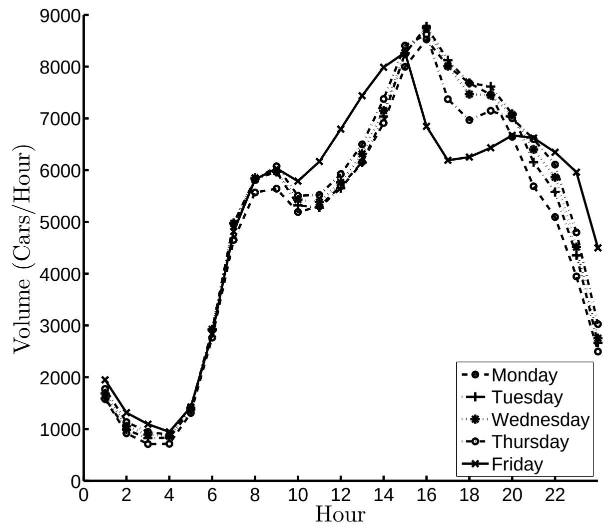


Figure 1: Average hourly volumes for SR 91 Eastbound between January 2009 - July 2011.

toll, however for simplicity assume that the only information available to the operator the travel time difference between the managed and unmanaged lanes.

An important characteristic of managed lane projects is that demand – particularly weekday demand – follows predictable patterns. Managed lane projects are usually built to alleviate congestion in areas that experience either a morning rush hour or an afternoon rush hour or both. The California State Route 91 (SR 91) in Orange County, California was one of the first managed lane projects in the United States. Average Eastbound traffic for summer weekdays is shown in Figure 1. On each weekday, Eastbound traffic has a small morning peak between about 7:00 and 9:30 AM followed by a much higher afternoon peak between 2:00 and 6:00 PM. Westbound traffic has a higher morning peak and a much lower afternoon peak. The patterns vary somewhat by day of the week but are generally stable.

We consider two broad categories of pricing policies. Under a *non-adaptive policy*, tolls are published at least one day prior to the day of operation. Tolls may vary by time-of-day and/or day-of-week but they cannot be changed based on current conditions. Non-adaptive pricing has historically been the norm for managed lane schemes such as the SR 91. The average hourly tolls for the Westbound managed lanes for the SR 91 for weekdays in June, 2011 are shown in Figure 2. Tolls can vary dramatically over the course of a day – in this case from a low of $1.30 to a high of $9.75. The fact that the strong traffic peaks evident in Figure 1 persist in spite of such dramatic differences in tolls across the day implies that there are a large number of drivers who are either unwilling or unable to change departure times away from periods characterized by both high congestion *and* high tolls.

Under *adaptive pricing* tolls can be changed periodically in response to current conditions. We consider two types of adaptive policies. In the myopic policy, the toll is set that maximizes the expected revenue from vehicles arriving in the next time interval. In a linear adjustment policy, a time-of-use policy is used as a baseline for tolls. If the travel-time differential between the managed and unmanaged lanes is greater than expected the toll is increased above the baseline: if the differential is less than expected, tolls are decreased.

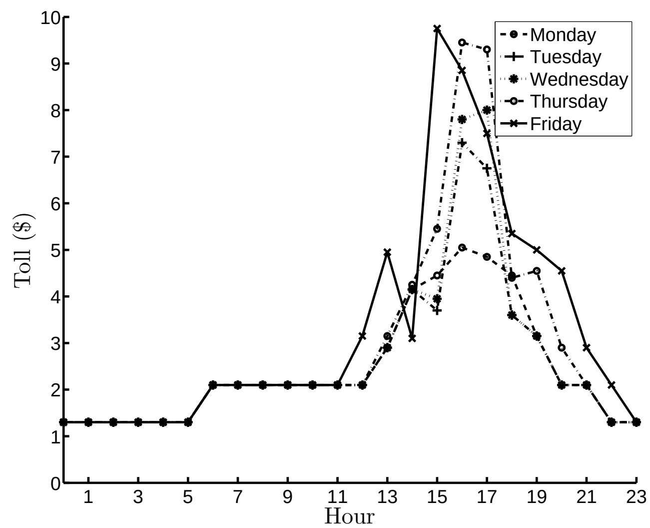


Figure 2: Average Westbound weekday hourly tolls for the SR 91 managed lanes in June, 2011. Source:www.912expresslanes.com/schedules.asp

One of our key findings is that the majority of the benefits from actively managing tolls comes from carefully setting and updating tolls around peaks. In particular, a revenue-maximizing operator should increase tolls very high in anticipation of a peak. This encourages drivers to take the unmanaged lanes which has the dual effects of decreasing congestion in the managed lanes and increasing congestion in the unmanaged lanes. These two effects combine to make the managed lanes more attractive to future drivers, enabling the controller to charge higher tolls than would have been otherwise possible. In this situation, we find that non-adaptive time-of-use policies can significantly outperform the adaptive myopic policy which considers only the current states of the managed and unmanaged lanes in setting the toll. More generally, we find that successful policies tend to charge higher tolls in all situations than the myopic policy. However, in cases in which total traffic demand is low or declining, the uplift of such policies relative to the myopic policy is small. On the other hand, when traffic demand is relatively high and increasing – as it would be entering a peak – both intelligent time-of-use and linear adjustment policies can yield substantially more revenue than the myopic policy.

Up to the last few years, the majority of managed lane schemes were operated by public entities who were not trying to maximize profit. The criteria used for setting such tolls was often to keep traffic moving freely in the managed lanes subject to some constraints such as maximum tolls. Chung and Recker (2011) provide an overview of the approaches taken by public managed lane operators in the United States. However revenue-maximizing toll policies are very different from those that seek to maintain free-flow in the managed lanes. The problem of setting revenue-maximizing tolls for a managed lane scheme has received very little attention in the literature. To our knowledge, the only exception is Yang (2012). However that paper not consider that congestion on the unmanaged lanes is influenced by the tolls on the managed lanes – which we find to be a very important consideration in setting the revenue-maximizing tolls. Göçmen (2013) in his thesis derives some structural results for the case in which traffic in each lane is modeled as a queue and demand is stationary. While this provides insight into the problem, it does not apply directly to the real-world cases of interest in which traffic dynamics are more complex and demand is not stationary.

In the remainder of this paper, we describe our simulation model and how the underlying modules were calibrated to the SR 91 data. We then compare the performance of different toll-setting policies using data from both the Eastbound and the Westbound SR 91 traffic. A key finding is that policies that anticipate future traffic and set tolls accordingly outperform the myopic policy that sets tolls based only on current conditions. The advantage of policies that incorporate forecasts of future traffic demand is particularly important immediately before peak periods. To better understand the performance of different policies, we used a set of stylized models in which we could vary the size and duration of peak demand to see how these features influenced performance. In the last section, we discuss our results and their implications for toll operators as well as potential extensions.

1 Simulation Model Description

Our simulation represents a highway that is 10 miles long and consists of five unmanaged lanes and two managed lanes with a single entry and exit point. This is approximately the length of the managed lanes for the SR 91. We assume that traffic arrives at a random rate. The arrival rate is exogenous, that is, it is not influenced by the current toll or traffic. Arriving vehicles choose either the managed lanes or the unmanaged lanes based on the travel time difference between the managed and unmanaged lanes and the toll. A consumer choice model estimates the fraction of arriving vehicles that choose the managed lanes based on the current toll.

A schematic diagram of the discrete time simulation model shown in Figure 3. In each time increment, the demand generation module determines how much new traffic arrives to the system.

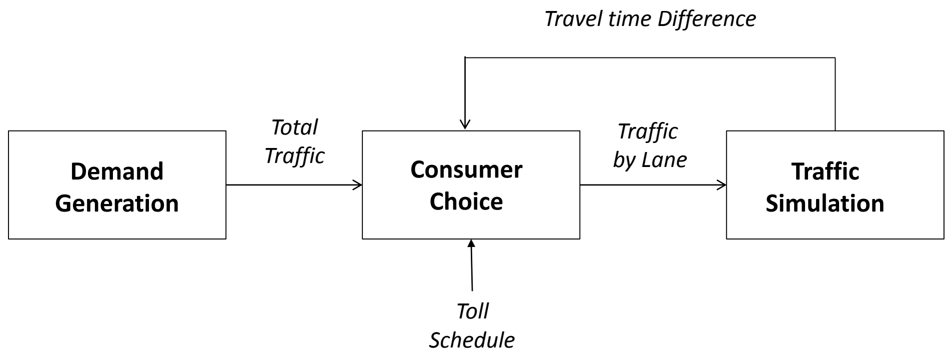


Figure 3: Modules in the Simulation Model.

Based on the current toll and time differential between the managed and unmanaged lanes, the consumer choice module determines the proportions of the arriving traffic that choose the managed and unmanaged lanes. The traffic module uses that information to update the traffic on the managed and unmanaged lanes. This information is then passed to the consumer choice module which determines the fraction of the incoming traffic that chooses the managed and the unmanaged lanes. This process repeats itself until the stopping time for the simulation is reached. In the numerical study, we start with an empty highway and simulate the system for a whole day.

We calibrated the simulation model using publicly data on the SR 91 available from the California Freeway Performance Measurement System (PeMS) [1]. We used data from the 10 mile managed lane section of the SR 91 that runs from the SR 55 interchange to the Riverside County line. These managed lanes are known as the "SR 91 Express Lanes" and are operated by the Orange County Transportation Authority who uses a time-of-use based tolling schedule for the managed lanes which it updates every few months. In setting the tolls, the primary stated goal of the authority is to maintain free flow speed on the managed lanes (The Orange County Transportation Authority, 2013).

[1] http://pems.dot.ca.gov/

1.1 Demand Generation Module

The demand generation module computes sample paths of total traffic demand that are based on both the mean traffic arrival rate patterns shown in Figure 1 and the serial correlation of traffic. Total traffic arriving to the system is initially generated as hourly demands which are then distributed to demands at five-minute intervals. Starting from midnight, hourly demand is generated by an autoregressive model of order three:

$$Y_t = \beta_t + \alpha_{1,t}Y_{t-1} + \alpha_{2,t}Y_{t-2} + \alpha_{3,t}Y_{t-3} + \varepsilon_t, \tag{1}$$

where Y_t is the traffic volume for hour t; $\beta_t, \alpha_t^1, \alpha_t^2$ and α_t^3 are coefficients and ε_t is a normally distributed error term. We estimated the parameters of (1) using Ordinary Least Squares (OLS) applied to historic SR 91 total traffic demands. The resulting parameter estimates by hour are shown in Tables 7 and 8 for Eastbound and Westbound traffic respectively.

In the next step we distribute the hourly traffic into five-minute intervals. For this purpose we used five-minute traffic volume data for July 2011. We omitted the first week of July due to the Independence Day holiday. For each hour, we calculated the fraction of hourly demand occuring in each 5-minute interval. By averaging those fractions across all days in our dataset, we calculated the average proportion of hourly demand each 5-minute interval for each hour. More details of the regression and the results are given in the Appendix.

1.2 Consumer Choice Module

The Consumer Choice Module takes in total highway demand at five minute intervals produced by the demand generation module and allocates the demand between the managed lanes and the unmanaged lanes based on the current travel-time differential and toll. Our approach is similar to the models described in (Xu, 2009; Yin and Lou, 2009). We use historical data on lane choice for the SR 91 to estimate the parameters of a consumer choice model as in Liu et al. (2004) and Liu et al. (2007).

Denote the expected travel time savings from choosing the managed lanes at time t by $\Delta T(t)$ and the toll by $p(t)$. Let $U_{in}^{(k)}(t) = g_{ik}(t) + \varepsilon_{in}$ denote the utility that driver n receives at time t by choosing alternative $i = u, m$, where u and m denote the unmanaged and managed lanes, respectively. The index k denotes the structure used for the deterministic part of the utility function. The term ε_{in} accounts for the unobserved component of driver n's utility from choosing alternative i. For the unmanaged lanes. Without loss of generality we set $g_{uk}(t)$ to zero. We considered three different formulae for driver utility:

$$g_{m1}(t) = \beta_T(t)\Delta T(t) + \beta_p(t)p(t),$$
$$g_{m2}(t) = \beta_T(t)\log(\Delta T(t)) + \beta_p(t)p(t),$$
$$g_{m3}(t) = \beta_T(t)(\Delta T(t))^2 + \beta_p(t)p(t).$$

The first formula corresponds to the case in which a driver's utility increases linearly with the expected time savings. The second corresponds to the case in which drivers get less sensitive to the expected travel time savings as it increases, and the third formula corresponds to the case where they become more sensitive. In all cases we allow β_T and β_p to vary by period.

For both lanes, we assume that the random term in the utility function is independently and identically distributed across drivers according to a Type I Extreme Value distribution. After evaluating the utility of both alternatives, each driver chooses the alternative that provides the highest utility. The probability that a driver chooses alternative i at time t is given by the logit function (Ben-Akiva and Lerman, 1985)

$$P_i^{(k)}(t) = \frac{e^{g_{ik}(t)}}{1 + e^{g_{mk}(t)}}. \tag{2}$$

We estimated β_T and β_p using maximum likelihood estimation (MLE). We use the same VDS sensor data that was used to calibrate the demand generation module. We analyzed the lane choice decisions of Eastbound commuters on SR 91 from Monday through Friday during the last two weeks of July 2011. Traffic in this direction has an afternoon peak as shown in Figure 1.

Figure 4 shows the minimum, maximum and average hourly time savings observed in our dataset. Not surprisingly, the expected travel time savings is highest during the afternoon peak. There is also significant variation between the time savings observed throughout this two week period. Figure 5 shows the managed lane share of traffic for Eastbound traffic. The managed lane share peaks in the afternoon when both congestion and tolls are at their highest. During the off-peak hours, the managed lanes command a very low share of the traffic passing through this segment of the highway.

Estimation of the model parameters, as well as the model parameters themselves and the fit for different models are described in the appendix. The model that showed the best fit with sensible coefficient signs was $g_{m3}(t) = \beta_T(t)(\Delta T(t))^2 + \beta_p(t)p(t)$, which is the function we use in the simulation model. Note that this model implies that the utility of the managed lines rises at an increasing rate with the time difference.

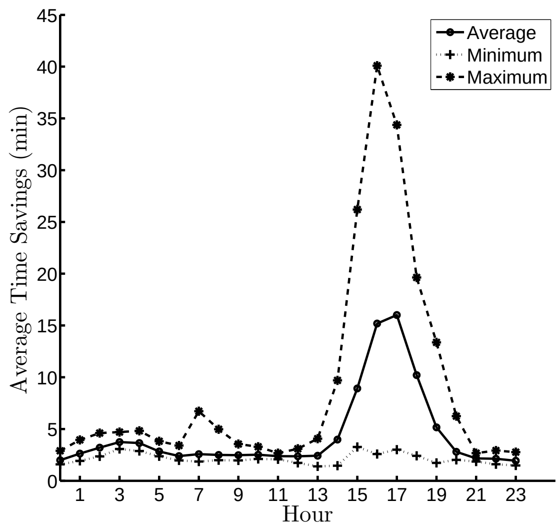


Figure 4: Average hourly time savings on the SR 90 Eastbound.

1.3 Traffic Module

The traffic module takes in demand for both the managed and unmanaged lanes at five minute intervals as calculated by the consumer choice module and calculates new travel times for the managed and unmanaged lanes. The resulting travel time differential is fed back to the consumer choice model as an input to the allocation of vehicles between managed and unmanaged lanes in the next time period. As vehicle density increases in a set of lanes, the speed of the vehicles in those lanes will decrease. Figure 6 shows a scatter plot of the weekday speed and density for the unmanaged lanes in the SR 91 during the first four weeks of July 2011. At low density – less than approximately 35 vehicles/mile/lane – vehicles can pass freely and the average speed is at or close to the "free-flow speed" for the highway. As traffic density increases above this level, average speed begins to drop. At approximately 60-70 vehicles/mile/lane, average speed tends to stabilize at or near the so-called "jam speed".

Traffic simulation models can be categorized as macroscopic, microscopic, and mesocopic, Macroscopic simulation works at the level of flows and does not represent the behavior of individual vehicles. Macroscopic simulation is not appropriate for our purposes because it cannot track lane choice by vehicles based on current tolls and congestion. Mesoscopic and microscopic simulations keep track of individual vehicles. Microscopic simulation models the behavior of individual drivers and how they change their behavior with changing road conditions. MITSIM, VISSIM and PARAMICS are some of the most well known microscopic simulation models (Olstam and Tapani, 2004). However, the level of detail that microscopic simulations can capture comes at a heavy

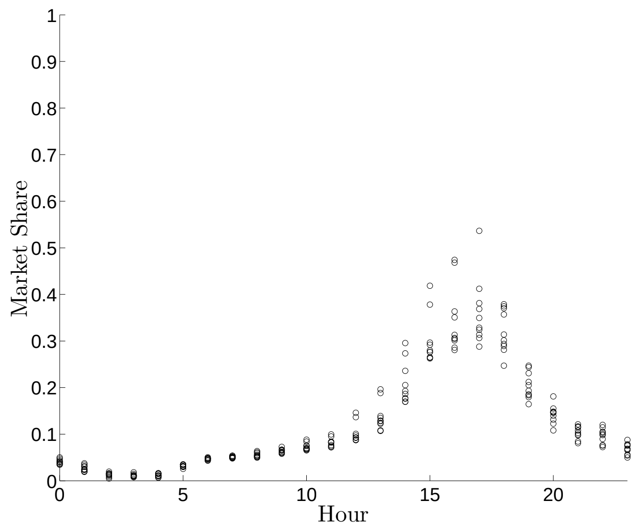


Figure 5: Hourly market share of the managed lanes on the SR 90 Eastbound.

cost: there are many more parameters that need to be estimated such as a desired speed for each driver, propensity to pass, desired following distance, etc. In addition, the time required to run and the time required to run a microscopic simulaiton model is orders of magnitude greater. For this reason, microscopic simulation is typically used to model the details of individual intersections and is not well suited to our purposes.

Mesoscopic simulation is typically used to model traffic behavior over stretches of highway of the length that we are considering. In mesoscopic simulations the road is divided into segments and at each time step the vehicles are moved from one segment to another based on speed, flow, and density relationships. Mesoscopic models have proven effective at realistically representing the behavior of traffic over longer stretches while requiring relatively few parameters to be calibrated. We based the logic in the traffic module on the mesoscopic simulations used in the DYNASMART (Jayakrishnan et al., 1994) and the DynaMIT (Ben-Akiva et al., 2002) models.

We divide the highway into nine segments with equal length L. The segment length was chosen as the distance that a vehicle moving at the free-flow speed would traverse during one time increment. The number of lanes in each segment is denoted by w and the average space that a vehicle occupies (including its headway at jam density) is denoted by ℓ. As a result, the physical capacity of each segment is wL/ℓ. We split each segment into two parts: moving and queuing. Vehicles that are queued up to join the next segment will be in the queuing part and the remainder of the vehicles in the segment will be in the moving part. The lengths of both parts are dynamic and depend on the number of vehicles in the queuing part. Given that there are n_q vehicles queued, the length of the queuing part is $n_q\ell/w$. Accordingly, the length of the moving part is $L - n_q\ell/w$.

At each time step in our simulation, the vehicles in the moving part traverse the segment at a speed $v(k)$ calculated according to the speed-density relationship

$$v(k) = \begin{cases} 66.8 - 0.14k, & \text{if } k \leq 25, \\ 15.0 + 69.33(1 - (k/100)^{2.22})^{7.69}, & \text{if } k > 25, \end{cases} \tag{3}$$

where k is the density of the moving part of the segment at the beginning of the time step, All speeds are in miles-per-hour (mph) and densities are in vehicles/mile/lane unless otherwise noted. After a vehicle has traversed a segment it has two options. If there is space in the next segment, it passes on to that segment and travels on that segment for the remainder of the time step. Otherwise it joins a queue and the end of the segment. The parameters of the model were estimated using historic data from the SR 91 as described in the appendix.

At each time increment, the movement of vehicles is calculated in three steps. In the first step, vehicles in the moving segment move according to the speed-density relationships in (3), and the vehicles that reach the end of the segment move into a queue to await transition into the next segment. We update the position of each vehicle starting from the one that is closest to the highway's end and move towards the beginning of the highway. In the next step, vehicles move from the queues at the end of each segment to the next segment. Vehicles are allowed to pass into the next segment until it reaches jam density. The last step involves moving the vehicles that just changed segments. If a vehicle waited in the queue for at least one iteration, then it is moved according to the prevailing moving part speed of its previous segment. If the vehicle joined the queue in that iteration, it completes its movement by traveling the amount it was not allowed to complete before joining the queue.

Once the vehicles are moved, we calculate the expected travel times for each segment. The expected travel time for each segment consists of the time it takes for a vehicle to traverse the moving part of a segment (T_m), and the waiting time in the queue (T_w). Let q denote the number of vehicles waiting in the queuing part of a segment, then,

$$T_m = \frac{L - n_q \times \ell \times w}{v(k)},$$

and

$$T_w = n_q/d,$$

where d is the moving average of the discharge rates observed in the previous periods. Those times are calculated for each segment and are in turn used to calculate the expected travel time for both

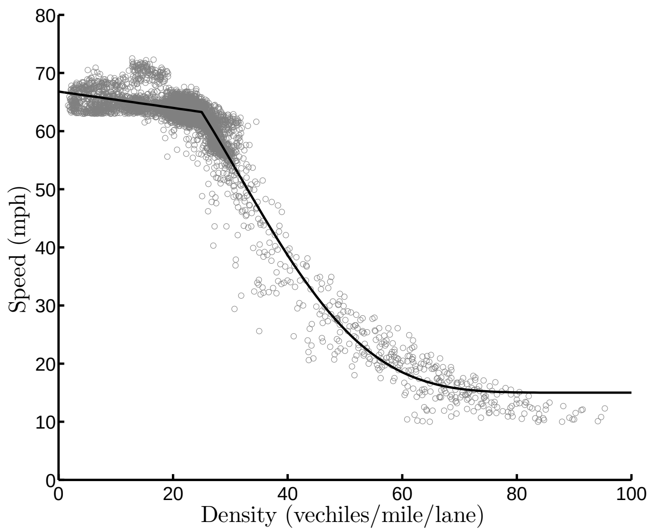


Figure 6: Speed-density relationship for SR 91.

parts of the highway.

This approach to modeling traffic flow is identical in structure to DYNASMART. It is also similar to DynaMIT, although we use a slightly different approach to calculate moving times and waiting times. These models have been used in many studies and have been extensively validated (Han et al., 2006; Roelofsen, 2012; Ben-Akiva et al., 2010).

We calibrated our model to the historic speed-density relationship for the unmanaged lanes in the SR 91 as shown in Figure 6. The black curve in this figure is the speed-density relationship that we fit to the underlying. The parameters of this speed-density relationship are shown in Table **??** and details of the calibration can be found in the Appendix. Since the current SR 91 policy sets tolls to encourage free-flow conditions in the managed lanes, there is no data for more congested conditions in the managed lanes. For this reason, we use the same speed-density relationship in Equation 3 for both parts of the highway. Since the managed lanes run parallel to the unmanaged lanes and are of similar quality, this is a reasonable assumption.

2 Numerical Studies

We used the simulation model described in the previous section to compare the expected revenue generated by different policies. We first considered two cases based on the SR 91 traffic. In the first case, we calibrated the simulation model using data from the weekday Eastbound SR 91. The

Eastbound SR 91 has a very pronounced afternoon peak and a much weaker morning peak. In the second case, we calibrated the simulation model using the weekday Westbound SR 91, which displays a strong morning peak and a weaker afternoon peak. These two cases provided a reasonable comparison of revenue generation among the policies considered but do not provide much insight into the drivers of revenue. In order to understand how tolling and revenue is influenced by various aspects of demand, we constructed a series of highly stylized demand models that allowed us to vary the height and duration of peak demand among other parameters. In this section, we first discuss the different tolling policies that we tested, then describe the results on the Eastbound and Westbound SR 91 data followed by the results on the stylized demand models.

2.1 Policies

The tolling policies considered include *adaptive* policies in which the tolls can be adjusted in response to current conditions as well as *non-adaptive* policies in which tolls are established in advance and do not change based on current conditions. In each case, we consider different intervals over which the toll can change. We use a discrete time approach in which we divide the planning horizon into T intervals, and t denotes the interval index. The number of vehicles that arrive in an interval is random, and is denoted by the random variable $D(t)$. We assume that the toll stays constant over each interval. This is realistic; all current managed lane schemes enforce a minimum interval between toll changes – for example, five minutes in the case of the LBJ Project. The number of vehicles in the lanes and their locations are denoted by $x_i(t)$ for $i = u, m$.

The revenue maximization problem for a non-adaptive policy is

$$\begin{aligned} \max \ & \sum_{t=1}^{T} \mathbb{E}\left[D(t)P_m^{(k)}(t)\right] p(t) \\ \text{s.t. } & x_i(t+1) = f_i(D(t), p(t), x_m(t), x_u(t)), \forall t \in \{1, \ldots, T-1\}, \ i = \{u, m\}, \\ & p(t) \geq 0, \forall t \in \{1, \ldots, T\}, \end{aligned}$$

where the mapping $f_i(.)$, for i in $\{u, m\}$, updates the list of vehicles and their locations in every period. The solution to this problem is a *Time-of-Use* policy since it tells the toll manager how much to charge at each point in time independent of the real-time state of the system.

The discrete-time counterpart of the adaptive policy is

$$\begin{aligned}
\max \ & \sum_{t=1}^{T} \mathbb{E}\left[D(t)P_m^{(k)}(t)\right] p(t, x_u(t), x_m(t)) \\
\text{s.t. } & x_i(t+1) = f_i(D(t), p(t, x_u(t), x_m(t)), x_m(t), x_u(t)), \forall t \in \{1, \ldots, T-1\}, \ i = \{u, m\}, \\
& p(x_u(t), x_m(t)) \geq 0, \forall t \in [0, T].
\end{aligned}$$

In each case, determination of the optimal policy given the underlying parameters is intractable. For this reason, we use various well-established approaches to approximate the parameters of the optimal policies based on the data.

2.1.1 Myopic Pricing

The *Myopic* Policy is an adaptive policy that sets the toll that maximizes the expected revenue rate for the next period based on the current travel time difference:

$$p(t) = \text{argmax}_{p \geq 0} P_m^{(k)}(t)p.$$

We use Brent's method (Brent, 2002) which is a robust derivative-free approach to estimate the myopic toll for every period. Because the myopic policy is a simple and easy to understand policy, we use it as the baseline for policy comparisons.

2.1.2 Time-of-Use Tolling

We use *Time-of-Use Tolling* to refer to a non-adaptive policy that specifies a toll for each time period based on the anticipated demand in each period, and the parameters of the consumer choice and the traffic flow modules. Determining the optimal Time-of-Use Policy is a stochastic optimization problem. Such problems are typically solved by iterative algorithms that sequentially update a trial solution using the stochastic gradient of the objective function:

$$x_{k+1} = x_k + a_k g_k(x_k),$$

where k is the iteration index, $x_k \in \mathbb{R}^n$ denotes the current solution, $a_k \in \mathbb{R}_+^n$ is the updating step size that decreases in k, and $g_k(.)$ is the stochastic gradient estimate.

Since no direct measurements of the gradient are available in our case, we employ the finite differences stochastic approximation (FDSA) method that estimates the gradient by calculating the difference quotient one-by-one for each decision variable using the Monte Carlo method. Kiefer and Wolfowitz (1952) introduced this method for univariate optimization problems and Blum (1954) extended it to the multivariate case. Let ω denote the outcome of a random process, and $y(x, \omega)$ be a function whose value depends on $x \in \mathbb{R}^n$ and the realization of the random outcome ω. Using the FDSA method the gradient estimate of $y(.)$ for each iteration is obtained by

$$(\hat{g}_k(x_k))_i = \sum_{j=1}^{m} \frac{y(x_k + e_i c_k, \omega_{kj}) - y(x_k - e_i c_k, \omega_{ki})}{2c_k}, \ \forall i = 1, \ldots, n,$$

where c_{ki} is a small positive scalar that decreases in k, and $e_i \in \mathbb{R}^n$ is the unit vector in direction i. For the Eastbound Case, we used the sequences: $c_k = 0.5/k^{1/6}$, $a_k = a/(A + k)$ with $a = 1$ and $A = 100$ when $n_{\max} = 1000$, and $a = 5$ and $A = 500$ when $n_{\max} = 5000$. For the Westbound Case, we used the sequences $c_k = 1/k^{1/6}$ and $a_k = 0.5/(200 + k)$, and perform 1000 iterations in the stochastic approximation procedure. The resulting parameters can be found in Table 16 of Appendix A.4.

Typically, smoothness and the differentiability of the objective function are required to establish convergence (Spall, 2003). In our problem the objective function is not tractable and we can assert neither smoothness nor differentiability. Thus, convergence is not guaranteed. Furthermore, due to the ill-behaved nature of the problem, the final set of decision variables may depend on the initial starting points. So, the Time-of-Use policies that are generated by the stochastic approximation procedures are heuristics. We stop the algorithm after a predetermined number of iterations denoted by $n_{\max}$.

2.1.3 Linear Adjustment Policy (LAP)

The *Linear Adjustment Policy* is an adaptive policy that takes a set of base time-of-use tolls $\{\bar{p}(t)\}_{t=1}^{T}$ and travel time savings $\{\Delta\bar{T}(t)\}_{t=1}^{T}$ as inputs. Every time the toll is updated, it compares the current travel time savings to the base values. If the system is more congested than expected, the policy increases the toll relative to the base toll. If there is less congestion than expected, the policy decreases the toll. The form of this policy is given by

$$p(t, \Delta T(t)) = \bar{p}(t) + \alpha^{+}(t)(\Delta T(t) - \Delta\bar{T}(t))^{+} - \alpha^{-}(t)(\Delta\bar{T}(t) - \Delta T(t))^{+}, \tag{4}$$

where $\alpha^{+}(t)$ and $\alpha^{-}(t)$ are positive scalars.

For the Linear Adjustment Policy, we start with a set of time-of-use tolls calculated as described above. Then, we estimate the adjustment factors using the FDSA method. Details of the calculation of the adjustment factor can be found in the Appendix in Section A.4.

2.2 Eastbound SR 91

In this example we analyze the Eastbound traffic scenario. The calibration of the demand generation and consumer choice model components for this direction were described earlier in Sections 1.1 and 1.2. We generated 1,000 sample paths for the traffic demand and performed our analysis on the same set of sample paths in every case.

Table 1 reports the average revenues and the 90% confidence intervals for the Myopic Policy with different update intervals. From the results we can see that there is a slight but consistent decrease in the expected revenues as the update interval increases. However, since all confidence intervals contain zero we cannot conclude that this decrease is statistically significant. Thus, the tolling frequency does not appear to have a significant effect on the performance of the Myopic Policy.

Figure 7 shows the average myopic toll (60 min. tolling interval) and the average hourly traffic load. During the off-peak hours, the average toll is relatively stable. During the peak hours, the toll increases as the congestion build-ups in the unmanaged lanes, and later decreases to its off-peak value.

We now consider Time-of-Use policies. We explored the performance of a hourly time-of-use tolling schedule to match the real-life implementations of such policies. Before starting the stochastic approximation procedure, we obtained two different starting points. For the first one we assumed that the demand is deterministic and equal to its certainty equivalent (CE) values. In the second case, we optimized over hundred randomly drawn sample paths – an approach known as*sample average approximation* SAA. We used the Nelder-Mead nonlinear optimization heuristic, and we tried one hundred different random starting points in each case.

We set the upper bound on the toll to \$100. Figure 8(a) depicts the tolls obtained through the stochastic approximation procedure and the average hourly traffic load. The 2-tuple in the legend indicates the starting point and the number of iterations performed, respectively. The second part of the figure reports the market shares of the managed lanes for the tolling schedules given in the first

Tolling Interval	1 min.	5 min.	10 min.	15 min.	20 min.	30 min.	60 min.
Avg. Rev.	\$125,157	\$125,095	\$124,859	\$124,647	\$124,547	\$124,510	\$124,511
C.I. Lower Bound	-\$3,500.61	-\$3,254.63	-\$2,219.23	-\$1,184.97	-\$626.42	-\$473.75	-
C.I. Upper Bound	\$2,207.17	\$2,086.27	\$1,522.01	\$912.19	\$554.71	\$475.85	-

Table 1: Average revenues and confidence intervals for the Myopic Policy.

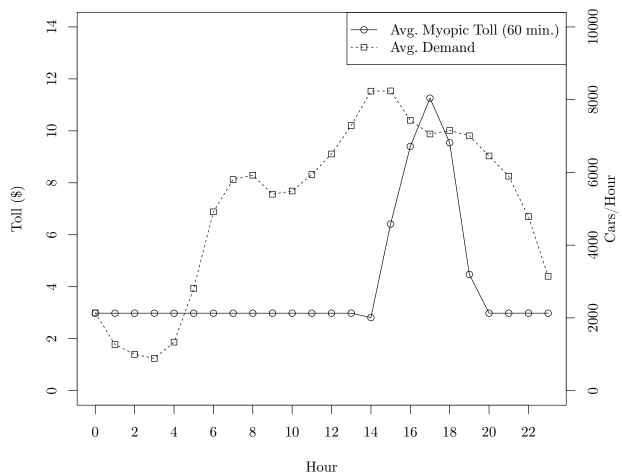


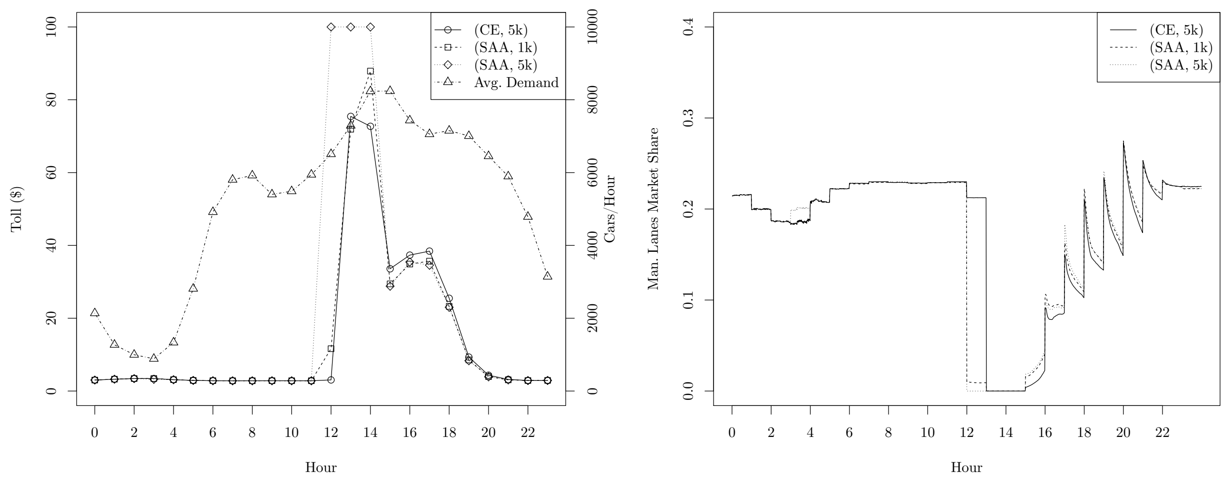
Figure 7: Myopic tolls and mean hourly demand.

part of the figure. The structure of all three policies are very similar. When the traffic load is low, the tolls are also quite low and stable in the region of $3. A few hours before the peak arrival traffic is observed, the tolls go up to very high levels and effectively divert all arrivals into the unmanaged lanes. By diverting almost all arriving vehicles into the unmanaged lanes, the toll operator achieves two goals: he reserves capacity in the managed lanes for the peak hours and increases congestions in the unmanaged lanes. These two effects combine to increase the attractiveness of the managed lanes during the peak hours – which enables the operator to extract more revenue from arriving traffic just when the volume of arrivals is highest. We term this a *jam and harvest* approach. From Table 2 and Figure 9 we can see that this approach translates into substantial revenue improvements over the Myopic Policy. When the Time-of-Use Policy sets its tolls high, a minuscule amount of revenue is earned since almost all drivers choose the unmanaged lanes. By forgoing the revenue in this period of time, the operator earns substantially more revenues when the *jamming* period ends and the *harvest* period begins.

We note that, in several cases the policies recommend a toll at the upper bound of $100. This would suggest that, during these periods, the optimal policy is to forgo almost all revenue from incoming

Static Policy	(CE, 5k)	(SAA, 1k)	(SAA, 5k)
Avg. Rev.	$153,086.51	$152,029.13	$151,532.68
% Imp. over Myopic (1 min. tolling update)	22.38%	21.53%	21.13%

Table 2: Performance of the static Time-of-Use Policies for the Eastbound Case.



(a) Time-of-use tolls and average hourly traffic load.

(b) Market share of managed lanes. (1 min. granularity)

Figure 8: Time-of-use tolls and the corresponding market share of managed lanes for the Eastbound case.

traffic in favor of diverting that traffic to the unmanaged lanes in order to increase congestion in those lanes.

Almost 70% of the daily revenues come between the hours of 4-8 pm Thus, we calibrate the Linear Adjustment Policy to those hours. We use the (CE, 5k) Time-of-Use Policy to form the base tolls and time savings since this policy resulted in the highest expected revenue among the non-adaptive policies considered. We allow α^- and α^+ to vary hourly. From the results in Table 15, we can see that α^- is always much higher than α^+. This is a result of the non-linearity of the speed-density relationship shown in Figure 6; travel times get increasingly more sensitive to density as traffic increases up to the jam density. The operator realizes substantially more revenue when the time differential is high. Because of the serial correlation of traffic demand, higher demand now is a strong indicator that future demand is likely to be higher so raising current tolls is likely to increase the future time-differential enabling even higher revenue later. On the other hand, if traffic is lower than expected, it is likely that future traffic is also lower and there is a motivation to reduce current tolls towards the Myopic Policy tolls in order to generate higher revenue from current traffic.

Table 3 shows the average revenues for different tolling intervals for the Linear Adjustment Policy along with the 90% confidence intervals for the revenue differences compared to the Linear Adjustment Policy with a 60 minute tolling interval. In this case, there appears to be no advantage to increasing the frequency of updates. Furthermore, from Figure 9 we can see that the Linear Adjustment Policy results in tolls that are similar in structure to the Time-of-Use Policy.

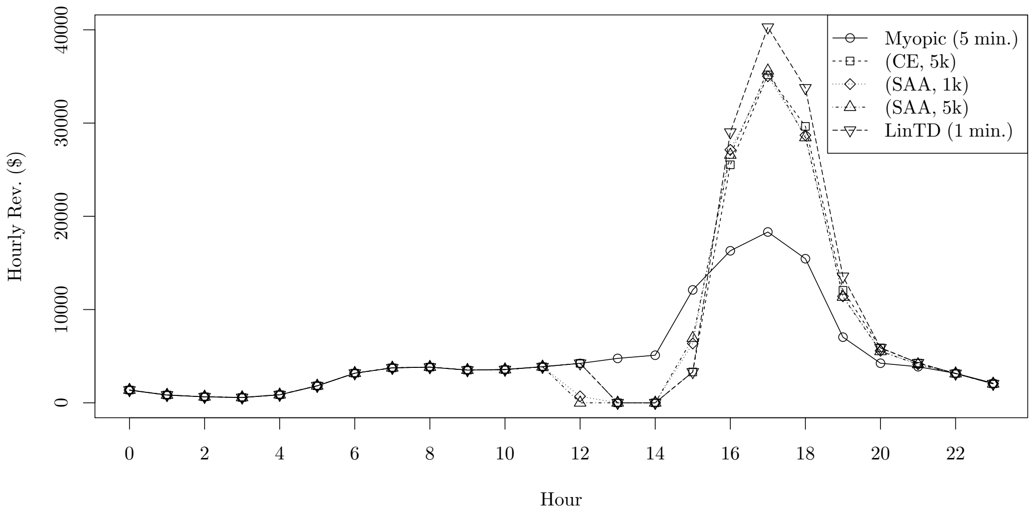


Figure 9: Average hourly revenues from different policies for the Eastbound Case.

Tolling Interval	1 min.	5 min.	10 min.	15 min.	20 min.	30 min.	60 min.
Avg. Rev.	\$167,338	\$167,328	\$167,456	\$167,595	\$168,058	\$167,119	\$167,709
% Imp. over Policy (1 min.)	33.7%	33.7%	33.8%	33.9%	34.3%	33.35%	34.0%
C.I. Lower Bound	-\$14,473.18	-\$8,684.09	-\$10,190.90	-\$7,006.63	-\$8,920.62	-\$9,970.29	–
C.I. Upper Bound	\$15,213.38	\$9,444.61	\$10,695.74	\$7,233.73	\$8,221.04	\$11,149.70	–

Table 3: Average revenues, percent improvement over Myopic Policy with 1 minute period and confidence intervals for the Linear Adjustment Policy for the Eastbound case.

	Myopic 1 min.	Time-of-Use CE, 5k	LAP 20 min.	Computational Bound
Avg. Rev.	\$125,157	\$153,087	\$168,058	\$179,444
Rel. Gap	30.25%	14.69%	6.34%	–

Table 4: Summary of policies and comparison to the computational upper bound for the Eastbound case.

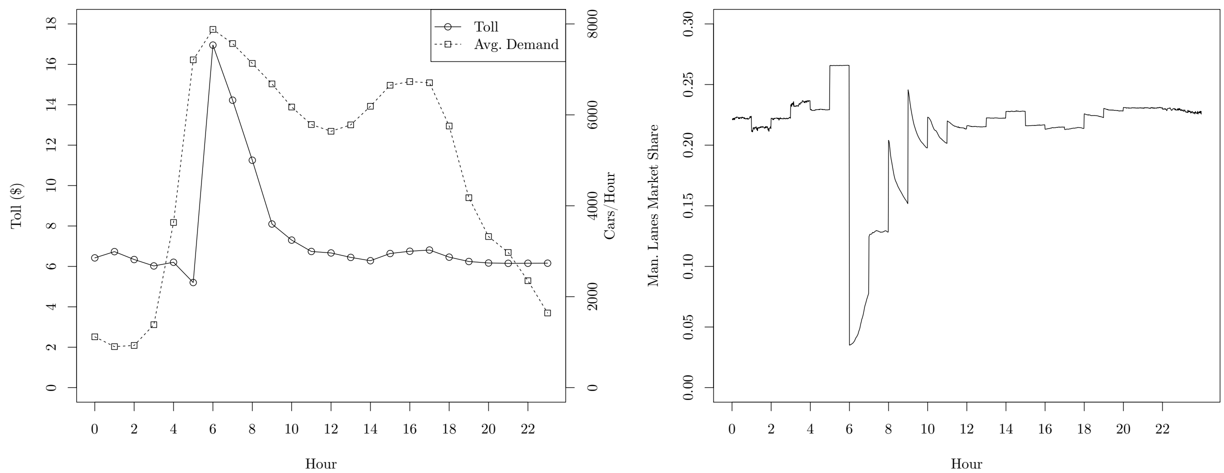
So far we analyzed each policy separately. Now, we compare their performance to each other and also to a computational upper bound where the operator is assumed to know the whole traffic pattern for each day and can set tolls that maximize revenue given that knowledge. The computational upper bound is obtained by computing the revenue-maximizing tolls for each sample path and then averaging them. Consistent with the SR 91 policy, we assumed that tolls could be changed only hourly.

Table 4 shows the expected revenue from three of the policies compared to the computational upper bound for the Eastbound Case. The Myopic Policy generates the least revenue of the three. We attribute this to the inability of the Myopic Policy to set tolls that anticipate future traffic demand. The Time-of-Use Policy outperforms the Myopic Policy by more than 20%. This uplift comes almost entirely from additional peak revenue resulting from the jam and harvest policy. Making the policy adaptive using linear adjustments adds an additional 10% of revenue. Furthermore, the Linear Adaptive Policy achieves 93.7% of the computational upper bound, which is impressive given that the computational upper bound assumes full knowledge of demand for each sample path.

2.3 Westbound SR 91 Results

As shown in Figure 1, Westbound traffic experiences a higher morning peak and a lower afternoon peak than Eastbound traffic. Furthermore, the larger morning peak is less pronounced than the afternoon peak for the Eastbound traffic. Parameters for the Westbound demand model were generated in the same fashion as the Eastbound case and can be found in the appendix. The same methods were used to estimate the Time-of-Use tolls and the adjustment parameters as in the Eastbound case.

Figure 10 shows the average hourly demand, the time-of-use tolls and the corresponding market share of managed lanes for the Westbound case. The Time-of-Use Policy again seeks to jam the unmanaged lanes by setting the toll very high between 6-7 AM in anticipation of the morning peak. As a result, the managed lanes' market share dips and most of the incoming vehicles choose the unmanaged lanes, which in turn creates increased congestion in the unmanaged lanes.



(a) Time-of-use tolls and average hourly traffic load.

(b) Market share of managed lanes. (1 min. granularity)

Figure 10: Time-of-use tolls and the market share of managed lanes for the Westbound case.

The revenues obtained from different policies in the Westbound Case are given in Table 5. Consistent with the Eastbound Case, we can see that increasing the tolling frequency for adaptive policies does not result in significant additional revenue. The Time-of-Use and Linear Adjustment policies both outperform the Myopic Policy and the Linear Adjustment Policy outperforms the Time-of-Use Policy. However, these policies generate less additional revenue relative to the Myopic Policy than in the Eastbound Case. The gap between the Linear Adjustment Policy, and the computational upper bound is also higher. A potential explanation for these two observations stems from the traffic load being spread out more evenly compared to the Eastbound case. As a result, the operator does not have the same scope for increasing unmanaged lane congestion by diverting traffic into the managed lanes. Furthermore, the mean traffic demand in the Eastbound case appears to be closer to a threshold on the magnitude of peak demand at which the value of Time-of-Use pricing changes dramatically. This means that more of the sample paths fall on either side of this threshold and the value of full knowledge of demand is greater than when the peak was more pronounced in the Eastbound Case. We will make this discussion more concrete in the next section.

2.4 Stylized Models

Our analysis of the SR 91 Westbound and Eastbound traffic showed that both adaptive and non-adaptive time-of-use policies generated significantly more revenue than myopic pricing. In all Time-of-Use policies, the recommended tolls had a jam and harvest character – that is, they raise

Policy	Myopic		Time-of-Use	LinTD		Comp. Bound
Tolling Interval	1 min.	60 min.	60 min.	1 min.	60 min.	60 min.
Revenue	\$167,325	\$167,031	\$171,831	\$181,454	\$181,680	\$210,698
Rel. Gap (vs. Comp. Bound)	20.59%	20.73%	18.45%	13.88%	13.77%	–
% Imp. over Myopic (1 min. tolling update)	–	–	2.69%	8.44%	8.58%	25.92%

Table 5: Revenues from different policies for the Westbound case.

tolls in advance of an anticipated peak in order to divert more traffic into the unmanaged lanes. This has the dual effect of increasing congestion in the unmanaged lanes and reducing congestion in the managed lanes, thereby making the managed lanes relatively more attractive during the peak which enables the controller to set higher tolls during the peak and the resulting increased revenue more than makes up for the reduced off-peak revenue. This suggests that a critical characteristic of a successful policy is the ability to anticipate peaks and set tolls accordingly.

While the character of the Time-of-Use policies was similar for both the Eastbound and Westbound cases, the gain over the Myopic Policy was significantly different. In particular, the magnitude of the gain was considerably greater on the Eastbound traffic pattern, which had a very high afternoon peak, than on the Westbound traffic pattern in which the peak was less pronounced. This raises the question of the extent to which the results depend upon the magnitude and duration of peak demand relative to off-peak demand. To address this question, we created some simple stylized models of traffic demand.

2.5 Stylized Models

Each of the stylized models considers a 24-hour day and has the general form shown in Figure 11. Traffic at the beginning of the day is set at the off-peak level. Traffic then rises linearly through a transition period to a peak. Traffic demand remains at the peak level for a number of hours, then decreases linearly through another transition period back to off-peak levels. We assume that both transition periods are of equal length. To create different scenarios, we vary the peak demand, the off-peak demand, the length of the peak, and the length of the transition periods.

In each scenario, we assumed deterministic arrival rates and used the same traffic model and choice models described in Sections 1.3 and 1.2. We calculated the tolls and revenues generated by both the non-adaptive Time-of-Use Policy and the Myopic Policy. Since traffic demand is deterministic, we did not test any adaptive models. We used the Nelder-Mead heuristic with 20 different randomly chosen starting points to compute the Time-of-Use tolls.

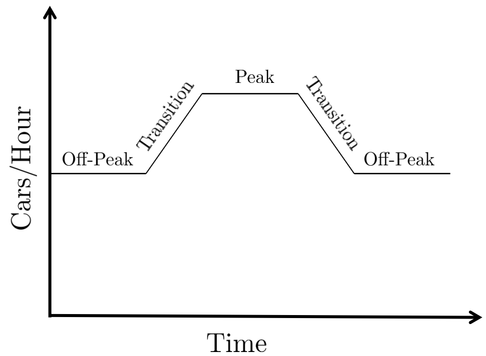


Figure 11: Demand structure of the stylized deterministic model.

Table 6 reports the gap between Time-of-Use and Myopic policies for different combinations of the settings. The Time-of-Use Policy outperforms the Myopic Policy in all cases. This supports the idea that anticipating peaks and pricing accordingly is a characteristic of any successful policy. The relative benefits of time-of-use pricing generally (but not uniformly) increase as a function of peak hourly demand, length-of-peak, and transition length. The most striking characteristic of the relative performance is the existence of some sharp transitions. When peak hourly demand is less than 8,000 cars/hour, the additional benefits from the Time-of-Use Policy are always less than 3.0%, regardless of the settings of the other parameters. When demand rises above 9,000 vehicles per hour and the length of the peak is two or three periods, the benefits from the Time-of-Use Policy increase substantially.

The results in Table 6 show a number of these sharp transitions, notably from 8,000 to 9,000 vehicles per hour when the peak length is three and from 9,000 to 10,000 cars/hour when the peak length is two. These transitions result from the non-linearity of the speed-volume relationship combined with the nature of the jam-and-harvest policy. When total peak demand – measured as a combination of peak vehicles/hour and length of peak – is relatively low, the controller is unable to divert enough vehicles into the unmanaged lanes to significantly influence the travel time differential. In this case, the additional benefit from time-of-use pricing is low. However, when peak hourly/demand is high, the controller can significantly influence travel time differential by using high tolls early to divert vehicles into the unmanaged lanes. This increases congestion in the unmanaged lanes and reduces congestion in the managed lanes relative to the Myopic Policy, which enables the controller to generate higher revenue during the peak. The benefits of a jam and harvest policy in this cases are substantial – over 50% when peak demand is high.

The existence of these thresholds is a likely explanation of the difference in the revenues from the Time-of-Use and the Linear Adjustment Policies relative to the computational upper bound on revenue as noted in the previous section. If the mean peak demand is significantly high, than the

Table 6: Revenue gap between Time-of-Use and Myopic polices for different traffic patterns in the stylized deterministic model in Figure 11.

Transition Length	**0**			**1**			**2**		
Length of Peak	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
Peak Hourly Dem.									
7000	1.37%	1.24%	1.09%	1.41%	1.51%	1.61%	0.67%	1.23%	0.95%
8000	0.95%	1.22%	1.08%	1.09%	1.11%	0.98%	1.23%	1.01%	1.36%
9000	1.28%	4.70%	21.63%	1.64%	7.70%	17.53%	2.24%	9.14%	23.79%
10000	2.35%	22.28%	37.70%	2.93%	23.40%	38.39%	4.41%	34.74%	53.25%

(a) Off-peak demand is 4000 vehicles/hour.

Transition Length	**0**			**1**			**2**		
Length of Peak	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
Peak Hourly Dem.									
7000	2.53%	2.29%	2.69%	2.28%	2.21%	2.63%	2.45%	2.33%	2.45%
8000	2.75%	2.49%	2.65%	2.42%	1.81%	1.84%	2.29%	2.20%	1.97%
9000	2.80%	6.33%	12.94%	3.28%	9.09%	21.39%	3.75%	10.67%	31.85%
10000	3.84%	22.79%	40.63%	8.44%	25.96%	40.17%	13.13%	38.66%	54.97%

(b) Off-peak demand is 5000 vehicles/hour.

Transition Length	**0**			**1**			**2**		
Length of Peak	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
Peak Hourly Dem.									
7000	1.09%	1.16%	1.27%	0.85%	1.03%	0.97%	1.23%	1.20%	1.38%
8000	1.19%	0.99%	0.58%	1.04%	1.25%	0.96%	0.67%	0.48%	0.80%
9000	3.63%	5.40%	14.13%	3.92%	7.90%	17.83%	5.31%	11.60%	41.17%
10000	4.00%	17.50%	36.90%	4.67%	34.65%	54.69%	24.66%	50.81%	66.29%

(c) Off-peak demand is 6000 vehicles/hour.

vast majority of sample paths will fall in the region in which Time-of-Use policies are effective – corresponding to the parameters in Table 6 for which the benefits of the Time-of-Use Policy are high, then using the same jam and harvest policy every day will be effective at capturing much of the available revenue for each sample path. If, on the other hand, the mean demand is close to the threshold, then jam and harvest will be very effective on days when demand is above the threshold but counter-productive on days when demand falls below the threshold. In this case, knowledge of the full pattern of demand would be much more valuable than if demand is consistently above or below the threshold.

3 Discussion

A key insight from this work is that the most important aspect of maximizing revenue from managed lanes is managing around peaks. If traffic is always low, no policy can do appreciably better than the Myopic Policy of individually maximizing the expected revenue from each entering vehicle. However, in the presence of peaks, policies that anticipate future traffic demand and set tolls accordingly outperform the Myopic Policy in every case studied. In particular, the best-performing policies raise tolls well above the Myopic Policy in the periods before the peak in order to divert traffic from the managed to the unmanaged lanes. This increases congestion in the unmanaged lanes while reducing congestion in the managed lanes and increases the relative attractiveness of the managed lanes to arriving traffic. This allows the toll operator to charge higher rates in the peak than would have profitable under the Myopic Policy, more than making up for any revenue lost in the off-peak period. These jam and harvest policies are very robust and outperform the Myopic Policy not only using both realistic Eastbound and Westbound traffic demand from the SR 91 but also in highly stylized models whenever there is a substantial peak.

For each of the cases studied with uncertain demand, the non-adaptive Time-of-Use Policy outperformed the Myopic Policy and the adaptive Linear Adjustment Policy outperformed both. The revenue from each policy is shown along with the computed upper bound for both Eastbound and Westbound traffic in Figure 12. For the Eastbound traffic, the greatest gain was from the Time-of-Use policy with a smaller relative benefit from making this policy adaptive. The opposite pattern is observed for Westbound traffic. We conjecture that the higher gain for the Time-of-Use for the Eastbound traffic is due to the more pronounced peak than in the Westbound traffic.

We showed using deterministic stylized models that the benefit from a Time-of-Use Policy over a Myopic Policy is highly sensitive to the length and magnitude of the peak. When the peak demand is small and/or of short duration relative to the off-peak demand, Time-of-Use policies gain little

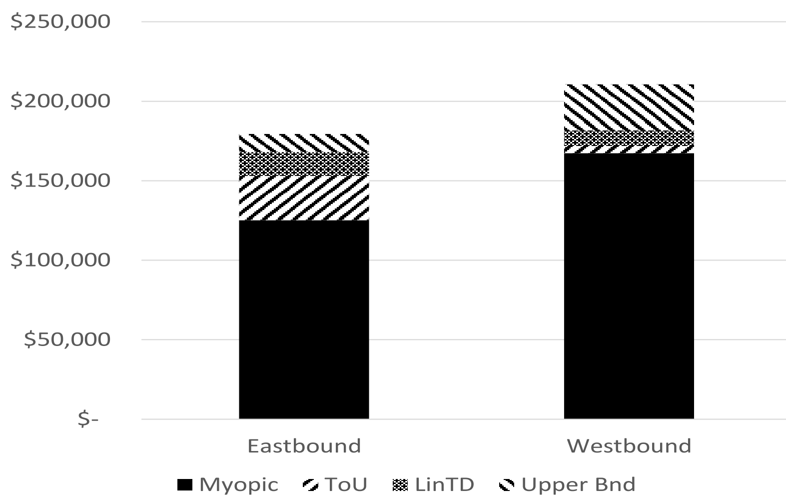


Figure 12: Revenue generated by different policies for Eastbound and Westbound traffic shown with the computational upper bound for revenue in each direction.

over a Myopic Policy. However, when peak demand is high and/or of long duration, the gain from a Time-of-Use Policy can be 50% or more. Furthermore, the data shows sharp transitions between a regime in which Time-of-Use provides relatively small additional benefits (5% or less) and a regime in which the benefits are 20% or higher. We attribute the existence of these transitions to the nature of the "jam and harvest" policy combined with the non-linearity of the underlying speed-density relationship.

Policies that sought to maximize revenue consistently set tolls that were higher than the myopic policy. The difference can be attributed in part to the existence of two *negative externalities* associated with a vehicle that chooses the managed lanes. The first is the additional congestion in the managed lanes created by that vehicle. The second negative externality is that, by choosing the managed lanes, the vehicle is not contributing to congestion in the unmanaged lanes. The combination of these two effects means that each vehicle that chooses the managed lanes rather than the unmanaged lanes leads to a reduced time differential, making the managed lanes less attractive to future arrivals. In this sense, the economics similar to *Paris Metro Pricing* as described by Odlyzko (1999). In Paris Metro Pricing, two identical units of capacity – such as train cars – are priced differently with the idea that some customers will be willing to pay more to use the less congested alternative. To the extent that customers prefer less crowded capacity, PMP can improve both revenue and customer satisfaction compared to pricing both units identically. While the fundamental idea of tolling for managed lanes is similar, the calculation of tolls is far more complicated due to the dynamic and non-linear relationship between traffic and the time-differential.

A somewhat counter-intuitive result from the simulations was that the time interval at which tolls were updated did not have a strong influence on revenue in the range from one minute to one hour. For example, for the SR 91 adjusting the the Myopic Policy every minute resulted in a gain of only .5% relative to adjusting the toll every hour. Similarly, the Linear Adjustment Policy showed no significant gain from being adjusted at intervals of a minute versus intervals of an hour. These results need further validation, however, they suggest that frequent toll changes may not be necessary to capture the majority of the revenue available.

Our model was parameterized using cleansed data that excluded exceptional situations such as lane closures, traffic accidents and unforeseen weather conditions. For this reason, our analysis almost certainly underestimates the value of adaptive policies relative to non-adaptive policies. Specifically, exceptional events are more likely to lead to higher congestion rather than lower congestion in the unmanaged lanes. An adaptive policy would detect this difference and adjust the tolls accordingly, extracting more revenue than a non-adaptive policy. However, in reality, exceptional conditions will typically be managed directly by the operator who can use knowledge about the nature and expected duration of the exceptional conditions to set and update the tolls.

Our analysis has assumed that traffic demands were generated exogenously – that is, that total demand in each time period was not influenced by either the toll or the travel-time differential. In actuality, some drivers may have the flexibility to change their travel plans – either by choosing a different departure time or a different route – in order to avoid high tolls and/or heavy congestion. From the initial paper by Wardrop (1952), extensive research has been performed on *traffic equilibrium models* which assume that at least some drivers will choose their routes in order to minimize travel time given congestion and/or tolls. Incorporation of such strategic behavior on the part of drivers may change some of the results and is a topic for on-going research. However, we note that the existence of predictable traffic jams at "rush hours", often combined with high tolls on many urban highways suggests that many drivers are unable or unwilling to change their routes and/or departure times easily. We suspect that the incorporation of strategic behavior on the part of some subset of drivers would not change the qualitative nature of the optimal tolling policies.

Acknowledgments

The authors would like to acknowledge the very detailed comments by Philipp Afeche that significantly improved the paper.

References

Ben-Akiva, M., M. Bierlaire, H. Koutsopoulos, R. Mishalani. 2002. *Real Time Simulation of Traffic Demand-Supply Interactions Within DynaMIT*, chap. 2. Kluwer Academic Publishers, 19–36.

Ben-Akiva, Moshe, Haris N Koutsopoulos, Constantinos Antoniou, Ramachandran Balakrishna. 2010. *Traffic Simulation with DynaMIT*, chap. 10. Springer, 363–398.

Ben-Akiva, Moshe, Steven Lerman. 1985. *Discrete choice analysis: theory and application to predict travel demand*. MIT press.

Blum, Julius R. 1954. Multidimensional stochastic approximation methods. *The Annals of Mathematical Statistics* **25**(4) 737–744.

Brent, Richard P. 2002. *Algorithms for minimization without derivatives*. Dover Publications.

Chung, C. L., W. Recker. 2011. State-of-the-art assessment of toll rates for high-occupancy and toll lanes. *Proceedings of the Transportation Research Board 90th Annual Meeting*.

Göçmen, Caner. 2013. Pricing managed lanes. Columbia University.

Han, Lee D, Fang Yuan, Shih-Miao Chin, Holing Hwang. 2006. Global optimization of emergency evacuation assignments. *Interfaces* **36**(6) 502–513.

Jayakrishnan, R., H. S. Mahmassani, T. Hu. 1994. An evaluation tool for advanced traffic information and management systems in urban networks. *Transporation Research C* **2** 129–147.

Kiefer, Jack, Jacob Wolfowitz. 1952. Stochastic estimation of the maximum of a regression function. *The Annals of Mathematical Statistics* **23**(3) 462–466.

Lam, Terence, Kenneth Small. 2001. The value of time and reliability: measurement from a value pricing experiment. *Transportation Research Part E: Logistics and Transportation Review* **37**(2) 231–251.

Liu, Henry X, Xiaozheng He, Will Recker. 2007. Estimation of the time-dependency of values of travel time and its reliability from loop detector data. *Transportation Research Part B: Methodological* **41**(4) 448–461.

Liu, Henry X, Will Recker, Anthony Chen. 2004. Uncovering the contribution of travel time reliability to dynamic route choice using real-time loop data. *Transportation Research Part A: Policy and Practice* **38**(6) 435–453.

Odlyzko, Andrew. 1999. Paris metro pricing for the internet. *Proceedings of the First ACM Confrence on Electronic Commerce*. 140–147.

Olstam, Johan Janson, Andreas Tapani. 2004. *Comparison of Car-following models*. Swedish National Road and Transport Research Institute.

Perez, Benjamin G., Charles Fuhs, Colleen Gants, Reno Giordano, David H. Ungemah. 2012. Priced managed lane guide. Tech. Rep. FHWA-HOP-13-007, US Department of Transportation, Federal Highway Adminsitration.

Roelofsen, Mark. 2012. Dynamic modelling of traffic management scenarios using dynasmart. Tech. rep., University of Twente.

Spall, James C. 2003. *Introduction to stochastic search and optimization: estimation, simulation, and control*, vol. 64. Wiley-Interscience.

Sullivan, Edward. 2000. Continuation study to evaluate the impacts of the sr 91 value-priced express lanes. Tech. rep., Cal Poly State University.

The Orange County Transportation Authority. 2013. The sr-91 express lanes. `http://www.91expresslanes.com/`. Accessed on 6.2..2013.

Wardrop, J. G. 1952. Some theoretical aspects of road traffic research. *Proceedings of the Institute of Civil Engineers* **11**(1) 325–378.

Xu, Shunan. 2009. Development and test of dynamic congestion pricing model. Master's thesis, Massachusets Institute of Technology.

Yang, Li. 2012. Stochastic traffic flow modeling and optimal congestion pricing. Ph.D. thesis, University of Michigan.

Yin, Yafeng, Yingyan Lou. 2009. Dynamic tolling strategies for managed lanes. *Journal of Transportation Engineering* **135**(2) 45–52.

A Model Calibration

A.1 Demand Generation Module

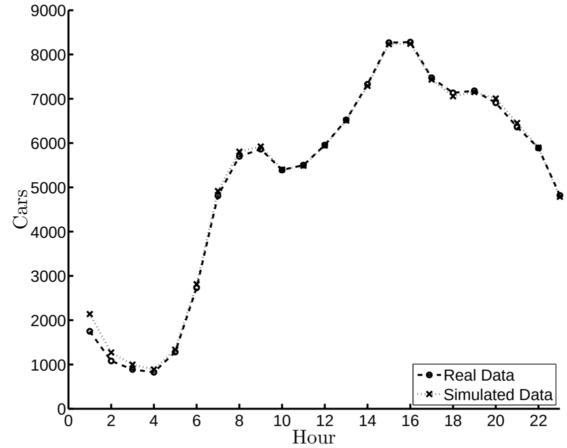
To calibrate the hourly demand generator we used the hourly flow data for the SR 91 from January 2009 to August 2011. We combined the volume information from the managed and unmanaged lanes to calculate the total volume of traffic using the highway. Traffic data for managed lanes comes from VDS 1208156 and for unmanaged lanes we use the data from VDS 1208147. Figure 1 depicts the average hourly traffic volumes for each day of the week for Monday through Friday for both Eastbound and Westbound traffic on the SR 91. The demand pattern is similar for all days so we used an average across days to create a mean demand for each hour. Based on this data, we used Ordinary Least Squares to estimate the parameters of Equation 1 using Ordinary Least Squares (OLS). The resulting parameter estimates by hour are shown in Tables 7 and 8 for Eastbound and Westbound traffic respectively.

t	β_t	α_t^1	α_t^2	α_t^3	Std. Dev. of Residual (ε_t)
0	116.94	0.67	-0.01	-0.03	180.51
1	152.92	0.73	-0.19	0.04	159.63
2	259.67	0.73	-0.09	0.00	83.25
3	303.68	0.71	-0.20	0.06	62.11
4	379.54	1.67	-0.38	-0.12	130.38
5	288.02	2.35	-0.20	-0.44	165.62
6	792.27	1.59	-0.11	-0.23	204.16
7	1091.10	0.77	0.35	-0.05	195.74
8	1540.70	0.47	0.37	-0.06	220.13
9	1818.52	0.56	0.04	0.01	354.20
10	197.21	1.37	-0.14	-0.22	270.91
11	810.85	1.01	-0.02	-0.05	287.93
12	452.39	0.95	-0.12	0.20	315.61
13	2122.24	1.10	-0.24	-0.10	388.58
14	4729.05	1.13	-0.28	-0.49	519.94
15	6479.56	1.14	-0.25	-0.89	734.43
16	2618.14	0.87	-0.11	-0.20	644.00
17	1599.20	0.78	0.00	-0.04	538.35
18	1371.06	0.69	-0.01	0.12	513.08
19	3602.23	0.73	-0.27	0.01	528.49
20	4118.83	0.78	-0.29	-0.15	505.42
21	1082.86	1.00	-0.12	-0.11	424.01
22	212.01	1.09	0.00	-0.26	565.98
23	-280.13	0.90	-0.16	0.01	442.78

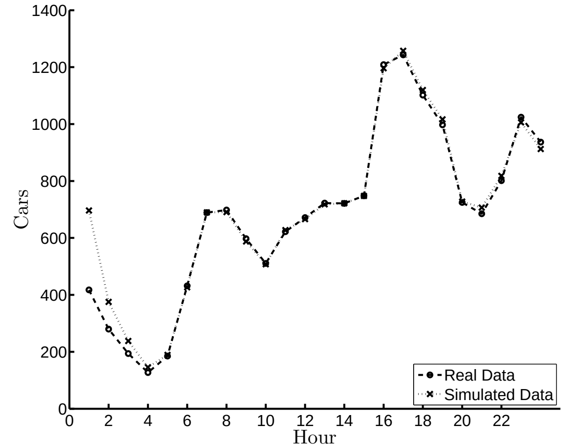
Table 7: Hourly demand model parameters for the Eastbound direction.

t	β_t	α_t^1	α_t^2	α_t^3	Std. Dev. of Residual (ε_t)
0	692.42	0.10	0.09	0.01	180.51
1	218.53	0.69	0.11	-0.12	159.63
2	310.82	0.89	-0.13	-0.02	83.25
3	881.90	1.12	-0.44	-0.11	62.11
4	1626.93	3.23	-1.71	-1.00	130.38
5	1565.48	1.85	0.33	-1.65	165.62
6	2309.88	0.96	0.21	-1.52	204.16
7	856.04	0.63	0.21	0.05	195.74
8	1339.35	0.66	-0.11	0.23	220.13
9	2555.56	0.43	0.06	0.08	354.20
10	2736.31	0.50	-0.02	0.04	270.91
11	1855.47	0.83	-0.10	-0.07	287.93
12	1030.70	0.85	0.10	-0.14	315.61
13	30.46	0.80	0.14	0.07	388.58
14	-350.95	0.98	-0.12	0.27	519.94
15	813.96	1.07	-0.04	-0.09	734.43
16	323.55	1.03	0.04	-0.12	644.00
17	73.59	0.65	0.36	-0.03	538.35
18	-277.89	0.60	-0.03	0.33	513.08
19	282.95	0.94	-0.37	0.14	528.49
20	417.60	1.03	-0.18	-0.05	505.42
21	295.30	0.94	-0.07	-0.03	424.01
22	-216.51	0.90	0.07	-0.07	565.98
23	390.16	0.97	-0.26	-0.08	442.78

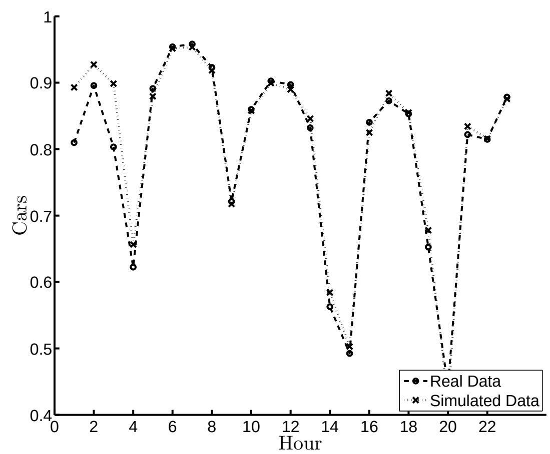
Table 8: Hourly demand model parameters for the Westbound direction.



(a) Comparison of means



(b) Comparison of standard deviations



(c) Comparison of autocorrelations with a lag of one

Figure 13: Demand model validation

In order to test the validity of this approach we generated 1000 sample paths and compared their statistics to the real-life traffic data. Figure 13 depicts the hourly means, standard deviations and autocorrelation (with a lag of one) for both the data and the sample paths on the Eastbound traffic. As can be seen from the figure, the statistics of the generated demand matched those of the real-life data quite well. Similar fits were obtained for the Westbound traffic (not shown). Tables 7 and 8 gives the parameters for the fitted demand model for Eastbound and Westbound traffic respectively.

To start generating hourly loads from midnight and onwards, we need starting values for the hours 21, 22 and 23. For simplicity, we sampled the demand for these three hours from normal distributions whose means, standard deviations and pairwise correlations match their real-life values. These statistics are shown in Tables 9 and 10.

	Eastbound		Westbound	
	Mean	Std. Dev.	Mean	Std. Dev.
Hour 21	5887.20	862.36	2958.30	529.14
Hour 22	4940.18	1142.69	2339.71	506.18
Hour 23	3351.17	1103.87	1633.21	356.85

Table 9: Mean and standard deviation of traffic volume for the hours used to start the demand generation module.

	Eastbound	Westbound
Hours 21 & 22	0.78	0.93
Hours 21 & 23	0.62	0.77

Table 10: Correlations between hours used to start the demand generation module.

A.2 Consumer Choice Module

The SR 91's policy leads to high correlation between the tolls and time savings. To estimate β_T and β_p we use an approach similar to Lam and Small (2001) in which we exploit the variation in time savings during the peak hours when the tolls do not change very much. Figure 14 plots the average ratio of expected time savings to tolls in five minute intervals. There is significant variation in that ratio during both peak and early morning hours. For our analysis we choose the afternoon peak hours, specifically between the hours of 2 PM and 8 PM, since the traffic volume is significantly higher compared to the early morning hours.

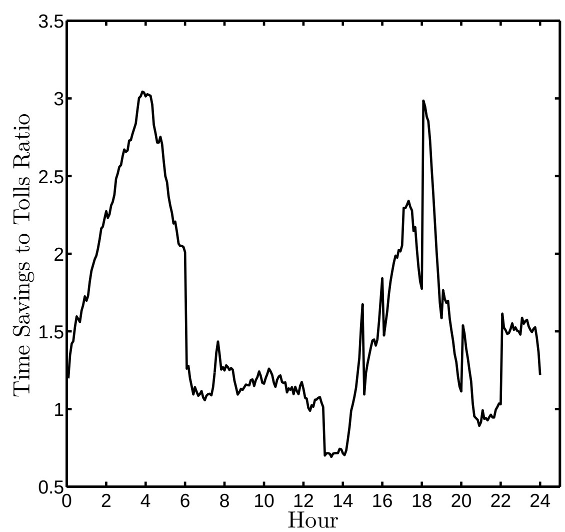


Figure 14: The average ratio of time savings to tolls.

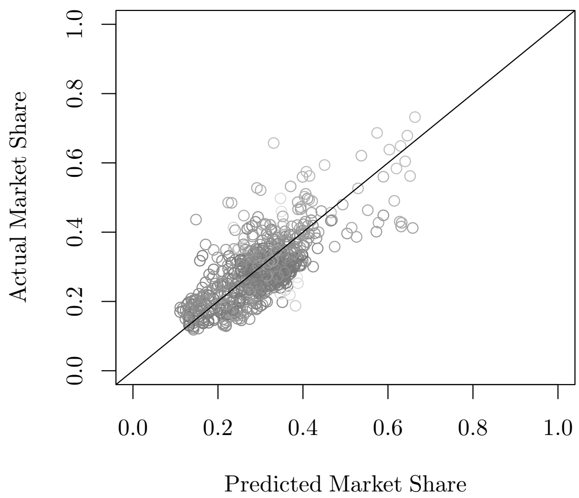
Now, we are in a position to use MLE to estimate β_T and β_p. The dataset we use contains the number of vehicles that chose the unmanaged and managed lanes between 2 PM and 8 PM in five minute granularity. It also contains the expected time savings the vehicles choosing the managed lanes enjoyed as well as the tolls they paid. There are 390,310 vehicles that chose the unmanaged and 140,931 that chose the managed lanes in the dataset.

In the estimation procedure we treat the aggregated five minute market share of the managed lanes as the dependent variable. The weight of each observation is equal to the total number of vehicles that pass through the unmanaged and managed lanes in that five minute time frame.

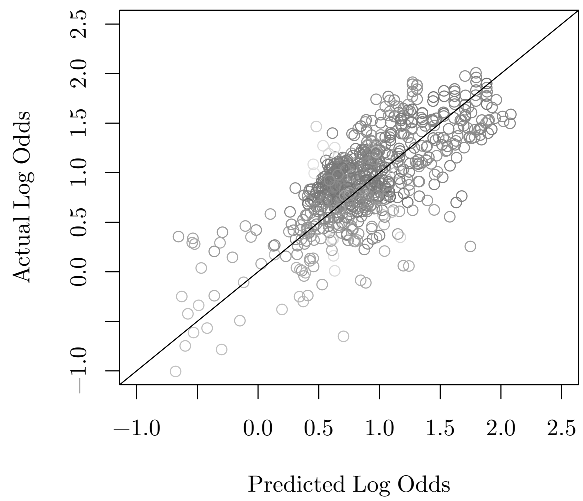
We evaluated the performance of the three different structures introduced for g_{uk}. For each structure, we tested four models which correspond to cases where the coefficients are allowed to vary over time or kept fixed. In the former case, different values for coefficients are estimated on a hourly basis. The coefficients for a given point in time are determined by taking the weighted average of the estimates corresponding to the current and upcoming hour, where the weights are obtained proportionally. For example, the coefficients at 2:20 PM would be the weighted averages of the coefficients for 2:00 PM and 3:00 PM with the weights 2/3 and 1/3, respectively. As discussed in the next, we considered three possible transformations for the travel-time savings (ΔT), $\ln(\Delta)$, ΔT, and ΔT^2. The resulting coefficients and LLE's are shown in Tables 11, 12 and 13. We chose Model 3-2, which corresponds to the case of $k = 3$, as it has the best fit among models with the correct signs on the coefficients. For the hours outside of the 2-8 PM range, we use the coefficients corresponding to 8 PM

The goodness-of-fit plots for Model 3-2 are given in Figure 15. The color of each point in the scatter plots depicts its weight. The darker a point is, the more weight it has. From the plots we can see that the model fits reasonably well to the data and there is no evident bias.

We did not account for high occupancy vehicles in our choice model. Under SR 91's tolling policy, vehicles that have at least three occupants (HOV3+) can use the managed lanes for free except between 4 PM-6 PM when they have to pay 50% of the toll. According to traffic counts performed on the eastbound direction of SR 91 between 3:30 PM5 and 5:30 PM, only 3.7% of the total vehicles that entered into the unmanaged and managed lanes were in the HOV3+ category (Sullivan, 2000). Because this percentage is so low, we do not feel that omitting the HOV3+ category significantly influenced our results.



(a) Actual vs. predicted market shares



(b) Actual vs. predicted log of odds

Figure 15: Consumer choice model goodness-of-fit plots.

A.3 Traffic Module

We calibrated the traffic model using the first four weeks of July, 2011. This period was chosen because there was no rain. The data was obtained through PeMS for VDS 1208147 with an aggregation level of 5 minutes. We deleted approximately 17% of the observations as outliers resulting from lane closures and accidents, resulting in 5,760 observations.

We set the jam density to 100/vehicles/mile/lane because there were only three observations with densities greater than 100. We removed the observations from our data set. We set the minimum speed $v_{\min}$ to 15 mph in accordance with the average speed observed when density reached 100 vehicles/mi/ln. We set the simulation time-step to a minute. The moving average of the last five periods gave a fairly accurate representation of the discharge rate from each segment under congestion. As shown in Equation 3, to fit the historic speed and density observations, we divided the data into two regions, one in which the density is below 25 vcl/mi/ln and one in which the density is greater than 25 but less than 100 vcl/mi/ln. We fit a line to the first part and a geometric-decay model with the form shown in Equation 3 to the second region. We breakpoint density of 25 vcl/mi/ln minimized the sum of squared residuals from both regions of the fitted model. We used OLS regression to estimate the parameters for the first (linear) regime and NLS regression to estimate the parameters in the second (nonlinear) regime. After the estimation procedure we readjusted β_2 to ensure continuity at the breakpoint density. Table **??** shows the parameters that were estimated using this procedure.

A.4 Numerical Study Data and Results

In the stochastic approximation procedure we used the same c_k sequence as we did in the calibration of the Time-of-Use Policy. Table 14 shows the parameters used in the sequence a_k. We performed 1000 iterations for each parameter. We used the same updating intervals as in the Myopic Policy and calibrated a different set of parameters for each one. The starting points were obtained through the application of Brent's method over 100 randomly chosen sample paths. Tables 15 and 16 show the results for the Eastbound and Westbound traffic respectively.

Variable	Model 1-1	Model 1-2	Model 1-3	Model 1-4
Toll	-0.2460^{***}		-0.2253^{***}	
	(0.0011)		(0.0012)	
Toll - Hour 14		-0.4781^{***}		-0.3999^{***}
		(0.004)		(0.0067)
Toll - Hour 15		-0.2352^{***}		-0.3156^{***}
		(0.0022)		(0.0043)
Toll - Hour 16		-0.1595^{***}		-0.1611^{***}
		(0.0018)		(0.0028)
Toll - Hour 17		-0.1444^{***}		-0.1670^{***}
		(0.0018)		(0.0034)
Toll - Hour 18		-0.1770^{***}		-0.1495^{***}
		(0.0022)		(0.0040)
Toll - Hour 19		-0.3354^{***}		-0.2958^{***}
		(0.003)		(0.0048)
Toll - Hour 20		-0.4520^{***}		-0.3999^{***}
		(0.0051)		(0.0074)
Time Savings	0.0567^{***}	0.0357^{***}		
	(5e-04)	(5e-04)		
Time Savings - Hour 14			-0.1969^{***}	-0.0453^{***}
			(0.0041)	(0.0065)
Time Savings - Hour 15			0.0534^{***}	0.0922^{***}
			(0.0014)	(0.0025)
Time Savings - Hour 16			0.0588^{***}	0.0348^{***}
			(8e-04)	(0.0013)
Time Savings - Hour 17			0.0690^{***}	0.0462^{***}
			(8e-04)	(0.0014)
Time Savings - Hour 18			0.0503^{***}	0.0241^{***}
			(8e-04)	(0.0016)
Time Savings - Hour 19			-0.0092^{***}	0.0155^{***}
			(0.0014)	(0.0023)
Time Savings - Hour 20			-0.1217^{***}	-0.0201^{***}
			(0.0038)	(0.0051)
Log Likelihood	-16294.6593	-7443.8744	-9295.4047	-6943.2609

$^{***}p < 0.01$

Table 11: Parameter estimates for models with untransformed variables.

Variable	Model 2-1	Model 2-2	Model 2-3	Model 2-4
Toll	-0.1874^{***}		-0.2488^{***}	
	(0.0016)		(0.0019)	
Toll - Hour 14		-0.4792^{***}		-0.5423^{***}
		(0.0042)		(0.0081)
Toll - Hour 15		-0.2186^{***}		-0.1438^{***}
		(0.0027)		(0.0060)
Toll - Hour 16		-0.1237^{***}		-0.1528^{***}
		(0.0021)		(0.0057)
Toll - Hour 17		-0.1000^{***}		-0.2540^{***}
		(0.0021)		(0.0074)
Toll - Hour 18		-0.1340^{***}		-0.1128^{***}
		(0.0027)		(0.0055)
Toll - Hour 19		-0.3128^{***}		-0.3009^{***}
		(0.0036)		(0.0061)
Toll - Hour 20		-0.4491^{***}		-0.3859^{***}
		(0.0052)		(0.0096)
log(Time Savings)	0.1395^{***}	0.0988^{***}		
	(0.0044)	(0.0050)		
log(Time Savings) - Hour 14			-0.5207^{***}	0.3300^{***}
			(0.0141)	(0.0260)
log(Time Savings) - Hour 15			0.1559^{***}	-0.1408^{***}
			(0.0077)	(0.0176)
log(Time Savings) - Hour 16			0.4856^{***}	0.2072^{***}
			(0.0073)	(0.0175)
log(Time Savings) - Hour 17			0.5443^{***}	0.5396^{***}
			(0.0067)	(0.0207)
log(Time Savings) - Hour 18			0.3453^{***}	0.0456^{***}
			(0.0060)	(0.0131)
log(Time Savings) - Hour 19			-0.0380^{***}	0.0779^{***}
			(0.0072)	(0.0137)
log(Time Savings) - Hour 20			-0.4543^{***}	-0.1193^{***}
			(0.0154)	(0.0270)
Log Likelihood	-22175.6920	-9437.6691	-10465.0480	-9022.4410
Num. obs.	720	720	720	720

$^{***}p < 0.01$

Table 12: Parameter estimates for models with log. of time savings.

Variable	Model 3-1	Model 3-2	Model 3-3	Model 3-4
Toll	-0.1954^{***}		-0.1882^{***}	
	(8e-04)		(8e-04)	
Toll - Hour 14		-0.4547^{***}		-0.4153^{***}
		(0.004)		(0.0045)
Toll - Hour 15		-0.1994^{***}		-0.2700^{***}
		(0.0021)		(0.0028)
Toll - Hour 16		-0.1360^{***}		-0.1307^{***}
		(0.0016)		(0.0019)
Toll - Hour 17		-0.1136^{***}		-0.1083^{***}
		(0.0015)		(0.0020)
Toll - Hour 18		-0.1340^{***}		-0.1345^{***}
		(0.0019)		(0.0027)
Toll - Hour 19		-0.2859^{***}		-0.2887^{***}
		(0.0028)		(0.0037)
Toll - Hour 20		-0.4290^{***}		-0.4110^{***}
		(0.0050)		(0.0056)
(Time Savings)^2	0.0016^{***}	0.0010^{***}		
	(1e-04)	(1e-05)		
(Time Savings)^2 - Hour 14			-0.0235^{***}	-0.0029^{***}
			(6e-04)	(6e-04)
(Time Savings)^2 - Hour 15			0.0033^{***}	0.0045^{***}
			(1e-04)	(1e-04)
(Time Savings)^2 - Hour 16			0.0011^{***}	$6e-04^{***}$
			(2e-05)	(0.0000)
(Time Savings)^2 - Hour 17			0.0018^{***}	$9e-04^{***}$
			(3e-05)	(0.0000)
(Time Savings)^2 - Hour 18			0.0018^{***}	$9e-04^{***}$
			(4e-05)	(0.0000)
(Time Savings)^2 - Hour 19			$-9e-04^{***}$	0.0012^{***}
			(1e-04)	(1e-04)
(Time Savings)^2 - Hour 20			-0.0082^{***}	-0.0014^{***}
			(3e-04)	(3e-04)
Log Likelihood	-15642.3036	-7174.4661	-12972.6198	-6309.6336
Num. obs.	720	720	720	720

$^{***}p < 0.01$

Table 13: Parameters estimates for models with time savings squared.

	α^+				α^-			
	Hour 16	Hour 17	Hour 18	Hour 19	Hour 16	Hour 17	Hour 18	Hour 19
a	0.5	0.5	0.5	1	5	2	5	10
A	100	100	100	100	100	100	100	100
α	1.5	1.8	1.5	1.5	1.8	1.5	1.5	1.5

Table 14: Parameters of a_k used in the calibration of LinTD for the Eastbound example.

Tolling Interval	1 min.	5 min.	10 min.	15 min.	20 min.	30 min.	60 min.
Hour 16	(0.44,1.36)	(0.35,1.17)	(0.36,1.26)	(0.22,1.11)	(0.34,1.13)	(0.48,1.04)	(0.44,1.25)
Hour 17	(0.62,1.60)	(0.52,1.39)	(0.53,1.37)	(0.47,1.53)	(0.52,1.94)	(0.58,1.81)	(0.51,1.91)
Hour 18	(0.54,1.59)	(0.44,1.05)	(0.45,1.67)	(0.39,1.46)	(0.47,2.61)	(0.41,2.29)	(0.37,2.58)
Hour 19	(0.24,1.57)	(0.3,4.60)	(0.01,1.43)	(0.20,3.44)	(0.14,3.73)	(0.34,3.94)	(0.10,1.41)

Table 15: Stochastic approximation procedure results for the (α^+, α^-) pairs for the Eastbound example.

Tolling Interval	1 min.	60 min.
Hour 5	(0.15,3.5)	(1.38,3.45)
Hour 6	(0.18,2.58)	(1.35,1.89)
Hour 7	(0.25,2.50)	(0.77,4.17)
Hour 8	(0.42,4.88)	(0.57,3.76)
Hour 9	(0.28,2.49)	(0.38,4.70)
Hour 10	(0.25,1.63)	(0.28,1.95)
Hour 11	(0.23,1.53)	(0.18,2.19)
Hour 12	(0.38,1.47)	(0.39,2.71)
Hour 13	(0.28,1.67)	(0.36,3.00)
Hour 14	(0.29,1.76)	(0.23,3.44)
Hour 15	(0.35,1.37)	(0.24,2.53)
Hour 16	(0.41,2.45)	(0.57,2.56)
Hour 17	(0.28,1.57)	(0.59,0.45)
Hour 18	(0.17,1.74)	(0.23,1.42)

Table 16: Stochastic approximation procedure results for the (α^+, α^-) pairs for the Westbound example.