

IS THE CURB 80% FULL OR 20% EMPTY? ASSESSING THE IMPACTS OF SAN FRANCISCO'S PARKING PRICING EXPERIMENT

Preprint of manuscript published in

Transportation Research Part A: Policy and Practice

<http://dx.doi.org/10.1016/j.tra.2014.02.016>

March 2014

Adam Millard-Ball

University of California, Santa Cruz
Santa Cruz, California 95064

Rachel R. Weinberger

Nelson\Nygaard Consulting Associates
New York, NY 10001

Robert C. Hampshire

Carnegie Mellon University
Pittsburgh, Pennsylvania 15217

ABSTRACT

The city of San Francisco is undertaking a large-scale controlled parking pricing experiment. San Francisco has adopted a performance goal of 60% to 80% occupancy for its metered parking. The goal represents an heuristic performance measure intended to reduce double parking and cruising for parking, and improve the driver experience; it follows a wave of academic and policy literature that calls for adjusting on-street parking prices to achieve similar occupancy targets. In this paper, we evaluate the relationship between occupancy rules and metrics of direct policy interest, such as the probability of finding a parking space and the amount of cruising. We show how cruising and arrival rates can be simulated or estimated from hourly occupancy data. Further, we evaluate the impacts of the first two years of the San Francisco program, and conclude that rate changes have helped achieve the City's occupancy goal and reduced cruising by 50%.

1. INTRODUCTION

Parking management has been a vexing problem for cities since the invention of the automobile. One concern is excess travel, congestion, air pollution and greenhouse gas (GHG) emissions that are caused by drivers searching for available parking – an activity colloquially known as *cruising*. Studies of cruising date to 1927, and some researchers have estimated that upwards of 30%, and maybe as much as 50%, of traffic on a given downtown street is comprised of people searching for a parking spot (Shoup 2006; see also Schaller Consulting 2006; Shoup 2008; Arnott & Rowse 1999). In particular, Shoup (2005) estimated that cruising in one small area of Los Angeles produced 3,600 miles of excess travel each day – equivalent to two round trips to the Moon each year.

One common response by cities has been to require developers to provide “sufficient” off-street spaces to accommodate expected demand for free parking and to provide municipal garages to make up for shortages at the curb. Indeed, many cities in the United States had introduced off-street parking requirements for new development by the mid-1960s, and such policies have worked to prevent cruising in low-density, suburban and newly developed areas. But, in denser urban centers, competition at the curb has persisted, frequently combined with off-street surpluses. Hence, the wisdom of such parking requirements has recently come into question. Moving away from a policy of copious, free off-street parking places an additional premium on curb management. If off-street parking is provided at a higher cost than on-street alternatives, drivers rationally choose to cruise, driving the increased need for curb management (Shoup 2006; Arnott & Inci 2006).

The need for on-street management has long been recognized. In 1935, the magazine *Popular Mechanics* reported that Oklahoma City, the first city to do so, introduced parking meters in order to enable customers to a commercial area to find a space more easily (Popular Mechanics 1935); and a 1956 book by the United States Bureau of Public Roads recommends maintaining a curb occupancy rate of no more than 85–90% in order to mitigate cruising (Bureau of Public Roads 1956). During the 1970s and 1980s, many cities lost sight of this need. Unable to foster the political will to effectively manage their curbs these cities suffered double parking; parking in loading zones, bus stops and in front of fire hydrants; and increased cruising. However, there has been a recent wave of interest in price-based curb management to mitigate these problems. Cities as varied as San Francisco, Seattle, Pasadena, Budapest, Mexico, D.F. and Seoul have set parking occupancy performance standards and adjusted prices to meet the performance goals.

Among these cities, San Francisco has piloted and carefully documented an extensive and innovative parking management program (*SFpark*). One of the hallmarks of *SFpark* is the availability of occupancy

data, which we use in this paper to address two related questions. First, we examine the theoretical and empirical relationships between parking performance standards – typically expressed as keeping average occupancy within a certain range – and the outcomes of policy interest, such as the driver’s experience and the amount of cruising. Second, we provide an evaluation of the effectiveness of *SFpark*. We do not attempt to separate the effects of the individual components of *SFpark*, such as information provision and price changes. Rather we examine travel behavior impacts of the program as a whole.

The contributions of this paper are theoretical, methodological and empirical. On the theoretical side, we show how different measures of parking performance, such as occupancy, relate to the driver experience and cruising. On the methodological side, we develop techniques to estimate cruising, arrival rates and the probability that a block is full from aggregated occupancy data. We then apply these theoretical and methodological tools to an empirical analysis of a large-scale controlled dynamic pricing experiment, *SFpark*. While one might expect that increased prices would reduce occupancy and cruising, the magnitude of any impacts is not obvious *ex ante* – particularly since price adjustments are small and are not immediately visible to drivers.

The following section reviews existing work on cruising and measuring the performance of parking systems, and then shows how this work can be informed by insights from the large literature on queueing theory. We then introduce the empirical setting of San Francisco, and describe the *SFpark* program and our sources of data. Subsequent sections discuss the simulation model of cruising that we calibrate using data from *SFpark*, and present the model results in terms of cruising and other measures of parking performance. We conclude by discussing policy extensions and directions for future research.

2. UNDERSTANDING PARKING PERFORMANCE

2.1 Cruising Research

Cruising is a long-standing concern for cities. For example, Shoup (2006) identifies 16 empirical cruising studies conducted between 1927 and 2001. In more recent years, numerous cruising studies have been added to this collection. Empirical studies rely on some kind of driver survey (Schaller Consulting 2006), videotaping (King 2010), or driving and searching by car (Shoup 2006). This last technique has been criticized as adding to cruising, thus changing the fundamental terrain of the study. Driver surveys generally stop people at intersections to ask if they are seeking parking, or ask people emerging from their cars about their experience finding a parking place. Studies that rely on video – or other visual techniques – may be the most robust. They count vehicles passing an open space and infer that the inverse of one

plus the number of vehicles to pass an open space is equal to the proportion of traffic that is looking for parking.

The other class of cruising studies is theoretical; this work is well represented in the economics literature. Most analyses conclude that cruising is due to misallocation of resources and should be eliminated (Arnott & Inci 2006; Button 2006). An extension of Arnott & Inci (2006) is due to Arnott & Rowse (2009) who look specifically at spatial competition between curb parking and garages. One study suggests that street parking should be priced equivalent to the marginal cost of providing an additional off-street space (Calthrop et al. 2000).

In contrast to a marginal cost pricing approach, Shoup (2006) recommends that on-street prices be set to achieve an average occupancy level of 85%, with an explicit rationale of eliminating cruising. Though Anderson and de Palma (2004) question the premise that cruising should be entirely eliminated, the Shoup rule-of-thumb requires that price vary both throughout the day and across different blocks, in order to achieve the occupancy goal. The rule-of-thumb has gained wide policy traction and, as noted above, occupancy targets have been introduced in places including Budapest, Seattle and Mexico, D.F. A similar approach with a slightly lower occupancy target has been adopted as part of *SFpark*. A recent empirical analysis of the San Francisco case by Pierce and Shoup (2013) shows a relatively elastic demand for parking on blocks where meter prices have been adjusted. Unfortunately, that work fails to address the endogeneity of the price adjustments, which complicates our ability to interpret their findings (Millard-Ball et al. in press).

In spite of widespread acceptance of the 85% or similar occupancy standard, very few studies have sought to analyze the heuristic empirically. The existing literature – primarily based on theory (Qian & Rajagopal 2013; D’Acierno et al. 2011; Bagloee et al. 2012); stated preference surveys (Kelly & Clinch 2006; Barata et al. 2011; Simićević et al. 2012; Ahmadi Azari et al. 2013; Simićević et al. 2013); and simulation models (Benenson et al. 2008; Martens & Benenson 2008) – suggests that cruising would decrease as prices adjust to achieve the target occupancy. However, it is also possible that rate changes would merely displace cruising drivers to other, perhaps non-metered, blocks, leaving total cruising constant, or even have the unintended effect of increasing trip-making. An increase in trip-making and attendant vehicle travel would occur if higher prices lead to parking spaces being occupied by a larger number of short-stay trips, instead of fewer trips by long-stay commuters. Indeed, one stream of work (Glazer & Niskanen 1992) concludes that, in the presence of strategic drivers, parking occupancy can increase when the per-unit time price of parking increases.

Simulation models have also focused on thresholds and other non-linearities in occupancy and cruising (Gallo et al. 2011; Levy et al. 2013; Martens & Benenson 2008; Benenson et al. 2008). In Levy et al. (2013), the authors identify changes in the system dynamics at about 85% but find a much greater impact above 92%.

There has been little interplay, however, between empirical studies that attempt to estimate cruising, and the stated-preference, theoretical and simulation studies that explore the dynamics of an urban parking system. In this paper, we build on both streams of research. We develop a theoretical model of the parking system, and use this model in conjunction with a large, robust dataset to empirically derive estimates of arrival rates, parking availability and cruising.

2.2 Beyond Average Occupancy

Average occupancy is a convenient, intuitive measure of parking performance, and is used by many of the simulation studies and other research discussed above. Unfortunately, average occupancy does not directly translate into two key measures of policy interest – how easy it is for drivers to find parking (the driver experience), and how much cruising occurs. In fact, depending on the period of averaging, occupancy measures could actually obscure insight into the driver experience.

First, we argue that average occupancy is not the most policy-relevant variable, as driver behavior is not guided by average occupancy on a block. Rather, it is guided by price and availability; the driver is most interested in whether or not there is available parking, and at what price. As long as there is an available space at a price the driver finds acceptable, it matters little if the block is 5%, 67%, or 95% full. Similarly, no cruising occurs if there is an available space, again, regardless of average occupancy. While there is certainly a relationship between average occupancy and the probability of finding a parking space, it is the latter measure that is relevant for policy. Average occupancy relates to the system average, whereas the likelihood of finding a space (or, equivalently, one minus the probability that a block is full) corresponds to the driver's experience.

It is important to note that the relationship between the average occupancy and the probability of finding a place to park is non-linear. This observation is consistent with fundamental results in queueing theory (see Kleinrock 1976), which establish that the probability of an arriving customer finding a system full is not the same as the time-averaged number of available resources (i.e., average occupancy). For example, an increase in average occupancy from 30% to 40% likely has no impact on the probability of finding a space, in sharp contrast to an increase from 90% to 100%. This implies that the marginal change in the

probability of finding a block full as a function of average utilization grows at an increasing rate. Formally, the relationship between average occupancy and the probability of a full block is convex.

Second, the nonlinear relationship between average occupancy and the probability of a block being full will vary depending on the length of the period over which occupancy is averaged. Consider, for example, a block with 85% average occupancy. The longer the averaging period, the more likely is this average to include instances of 100% occupancy. At the other extreme, if a one-minute average is computed using 60 one-second-level observations, an 85% average usually means that the block is never full. Given this intuition and the convexity of the relationship, by Jensen's inequality (Rudin 1987), we would expect the probability of a full block under the two-week average metric, i.e., as the average converges to expected occupancy, to be higher than under the hourly average metric, even when the average occupancies evaluate to the same value. Empirical evidence on the impact of time averaging is discussed later in the paper. This issue is important because it highlights the importance of the time scale over which average occupancy targets are set.

Third, more people will be trying to park at high-demand times. Thus more people are exposed to crowded conditions even if crowding is experienced for less time. The problem is best illustrated by the case where a block is empty for half the time, fills up very rapidly, and remains full, during which time drivers continue to arrive but are forced to seek parking elsewhere. Objectively, this block has a time-averaged occupancy rate of ~50%, yet only one user experiences it as 50% full. The vast majority of parkers, or would-be parkers, arrives after the block is full and experience it at 100% occupancy. While the occupancy target may thus be met, the user experience may still leave something to be desired.

Queueing theory provides established results specifying the connection between time averages and user averages, and the conditions under which the two averages are equal (Wolff 1982; Melamed & Whitt 1990). To formalize the experience of the *typical driver*:

- Assume that drivers arrive randomly to park at a block according to a non-stationary Poisson process, $\{N(t)|t > 0\}$ with an arrival rate function, $\lambda(t)$, which is a function of the time of day.
- The number of arriving drivers up to time T is denoted by the random variable $N(T)$, and the random arrival time of the i^{th} driver is denoted by T_i .
- Define X_t as a Bernoulli random variable denoting if the block is full at time t .

In this stylized setting, a block at time t can be either full, $X_t = 1$, or not full, $X_t = 0$. The driver-average occupancy (shown below on the left) is the policy-relevant variable, while the time-average occupancy (shown below on the right) is the commonly used heuristic. Formally, these can be respectively stated:

$$E \left[\frac{1}{N(T)} \sum_{i=1}^{N(T)} 1\{X_{T_i} = 1\} \mid N(T) > 0 \right] \quad \text{and} \quad \frac{1}{T} E \left[\int_0^T 1\{X_s = 1\} ds \right]$$

These are equal only under very particular conditions, hence we expect the time-average and driver-average perspective of the system to be different (Wolff 1982). The expected parking occupancy, as experienced by the *typical driver* is computed by weighting the occupancy by the arrival rate (Massey 2002). This is formalized in Eq. 1, which represents the proportion of arriving drivers that find the block full.

$$E \left[\frac{1}{N(T)} \sum_{i=1}^{N(T)} 1\{X_{T_i} = 1\} \mid N(T) > 0 \right] = \frac{E \left[\int_0^T 1\{X_s = 1\} \lambda(s) ds \right]}{\int_0^T \lambda(s) ds}. \quad (1)$$

The empirical analysis developed in the remainder of this paper addresses all three of these issues. We use queueing theory as the basis to derive an empirical relationship between average occupancy and the probability that a driver finds a block full, and between average occupancy and the number of blocks cruised. We analyze how the length of the averaging period – hourly or biweekly – affects the form of this relationship. And we demonstrate how occupancy data can be used to estimate the arrival rate of drivers looking for parking. By weighting the occupancy data by arrival rates, we focus on the driver average, rather than the time average.

3. EMPIRICAL SETTING AND DATA

3.1 Overview of SFpark

SFpark is a large-scale smart parking initiative, which was developed to improve the management of on- and off-street parking in the City of San Francisco.¹ A unique feature of SFpark is that it includes both pilot (treatment) and control areas. While SFpark is administered by the San Francisco Municipal Transportation Agency (SFMTA), the United States Department of Transportation (USDOT) helped

¹ The information in this section is derived from the policy and evaluation reports on the SFpark website (www.sfpark.org, last accessed January 16, 2014), and personal communications with staff at the San Francisco Municipal Transportation Agency.

finance state-of-the-art technology, including new parking meters and in-street sensors that communicate data regularly to a data management system. The system includes two new ways to accept payment, credit card and pay-by-phone, in addition to the previous payment options of cash and parking cards.

One of the key elements of *SFpark* is the use of pricing to reduce the number of drivers cruising for an on-street parking space. This is evidenced by the program's slogan, "Circle Less, Live More." Decreasing the number of cruising drivers is directly connected to the primary public policy goals of interest: reduced traffic congestion and pollution. Several other public policy goals follow from reduced cruising, including safer streets for bikers and pedestrians, and more reliable public transit schedule adherence².

Parking prices under *SFpark* are adjusted regularly in an attempt to bring average occupancy within the target range of 60% to 80%. This range is slightly lower than Shoup's (2005) recommendation of an 85% occupancy target; the rationale of *SFpark* is that an occupancy rate of 60-80% averaged over a two-week evaluation period considers variance within the evaluation period, which may include moments where occupancy exceeds 85% and may even reach 100%. When average occupancy in the evaluation period³ on a given block in a given time band (e.g. weekdays 7am to noon) reaches 80% or more, the hourly cost to park is increased by 25 cents. When average occupancy is below 60%, the rate is reduced by 25 cents (50 cents if it is below 30%). Rates are constrained to a maximum of \$6.00 per hour, and a minimum of 25 cents per hour.

There are 256 on-street blocks (i.e., pairs of opposing block faces or street segments) in the pilot and thus eligible for rate adjustments. In addition the program comprises one surface lot and 14 parking garages, which are not considered in this paper. There are three time bands: (meter opening time to noon; noon to 3pm; and 3pm to close), over two day types (weekday and weekend), allowing for six possible price regimes.⁴ These different block, day-type and time-band combinations create more than 1,500 possible on-street price adjustments at each rate adjustment on the 256 pilot blocks.

During the first two years of *SFpark* operations, the period evaluated in this paper, there were ten rate changes. On 5% of pilot block/day-type/time-band combinations, the meter rate was never adjusted – indicating that average occupancy was in the target range. On 37%, the meter prices were adjusted

² See <http://sfpark.org/about-the-project/benefits/>, last accessed January 16, 2014.

³ The length and nature of the evaluation period has been refined during the *SFpark* pilot. Initially, two weeks of data were used to compute a separate average by time band, giving an average of 10 days of data for weekdays and 2-4 days of data for weekends (depending on whether meters operate on Sundays in a given area). More recently, Mondays and Fridays have been excluded from the data, meaning that the weekday average uses six days of data. Holidays and special events are also excluded.

⁴ An additional timeband was introduced for evening parking in March 2013 in one neighborhood. This is not considered in the analysis here.

upward or stayed the same at each adjustment. These blocks experienced average occupancy above the target level during at least one rate change period. On a further 37%, the rate was adjusted downwards or stayed the same, indicating segments that were under the target occupancy during at least one rate change period. On the remaining 21%, rates were adjusted up at least once and down at least once. These fluctuating blocks may merit additional attention in future research. After the 10th rate change in April 2013, nine blocks had reached the \$6 per hour cap, and 179 were at the \$0.25 per hour minimum.

3.2 Data Sources

The data used for this study comprise variations on the in-road parking sensor data collected by SFMTA. Some of the analysis is based on hourly average occupancy rates on each block. These hourly data, provided by SFMTA, span the period March 1, 2011 to April 14, 2013. This represents a 5-month baseline period followed by 10 rate changes. In addition, for a subset of the two-year period, we have captured instantaneous occupancy data from the SFMTA website's application programming interface (API).

3.2.1 SFMTA Hourly Data

The hourly data contain information for metered on-street spaces on 408 blocks, plus one off-street lot which we do not consider in this paper. Of the 408 blocks, 256 are pilot blocks and 55 blocks form a study control group designated by SFMTA. Sensors were installed in the pilot and control areas but only the pilot areas are subject to price changes. In the control areas, meter prices have been held fixed throughout the period at \$2 per hour.

Sensor data are available for 97 additional blocks, which we do not include in the analysis unless stated. Some of these additional blocks were not considered for rate changes because they fall under the jurisdiction of the Port of San Francisco rather than SFMTA; the Port pursues its own on-street pricing policies. Others were not considered for rate changes because sensor installation was not yet complete, or because smart meters (which allow rate changes to be transmitted wirelessly to the meters, avoiding the need for meter-by-meter manual adjustments) had not yet been installed. This group of blocks without smart meters are almost all located in the downtown area where metered spaces are restricted to commercial loading for most of the day. Figure 1 shows estimated arrival rates by block and illustrates the geographic distribution of these three categories of block.

Each data point specifies the date, the hour, the number of metered and commercial loading spaces on the block, and the parking rate in effect. Most important for this analysis, the SFMTA hourly data set also includes parking occupancy averaged over the hour. Each data point combines parking occupancy on

both sides of the street, which makes most sense for one-way streets and represents a minor data limitation for our analysis with respect to two-way streets (discussed in more detail below). The entire data set comprises 7,064,098 data points over approximately a two-year period. Unless otherwise stated, we only consider the 2,422,901 data points during metered hours, or (in the case of the figures) a more limited subset of 1,682,208. This limited subset only consists of blocks with a continuous record of occupancy data and rate adjustments over the two-year period, and excludes blocks where sensors or smart meters were installed at a later date, where sensors were removed for repaving, or where similar issues led to gaps in data.

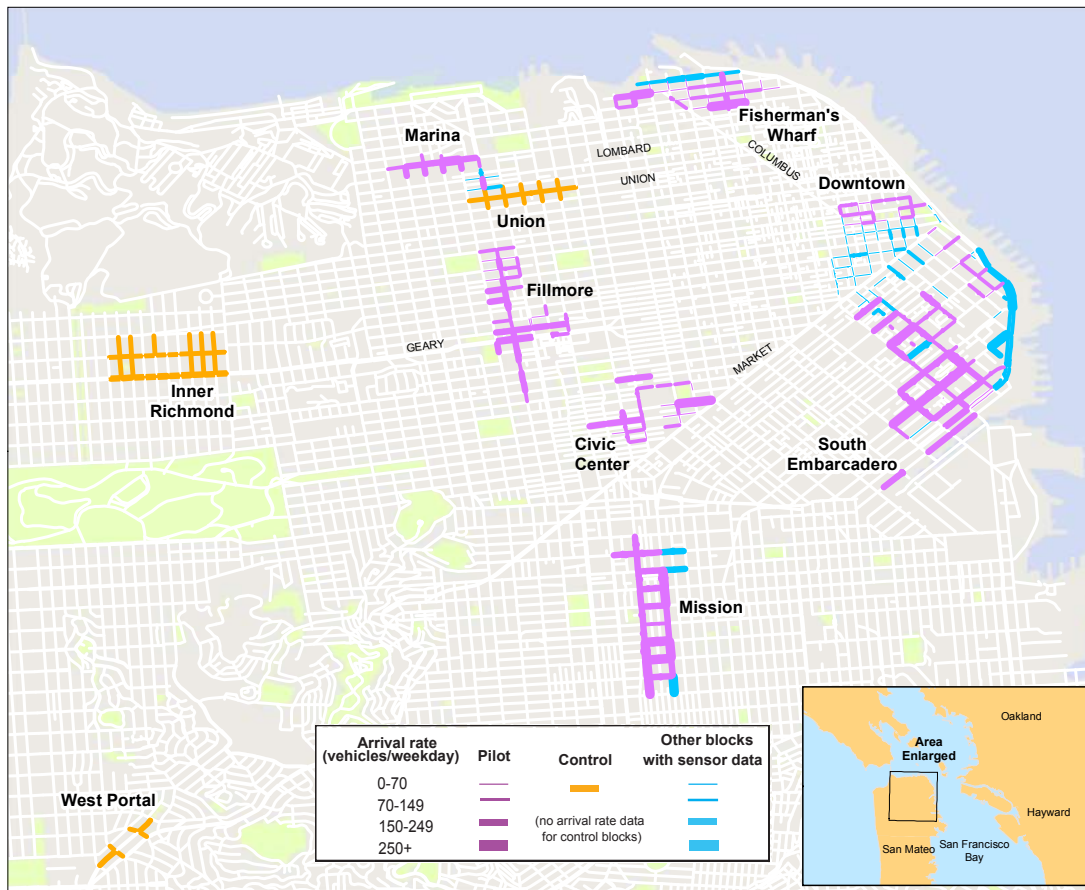


Figure 1 Estimated Arrival Rates by Block (vehicles per weekday).
 Arrival rates are estimated assuming a Skellam distribution, as discussed in Section 4.2.

3.2.2 Web API Snapshots

The second data set was collected by developing a web application that interacts with the *SFpark* API, which is made available by SFMTA to mobile phone and web application designers as well as any other interested party. These data provide snapshots of parking availability and capacity for each side of the street (i.e., block faces) which we aggregate to both sides of the street (i.e., blocks) in order to match the hourly data. Here, we use data for 340 blocks⁵, collected at approximately 5-minute intervals.

The two data sets contain overlapping periods of observations from January 1, 2012 to February 14, 2012. The API data set comprises 4,664,469 block-level data points during this period, which reduces to 1,730,770 when we limit the data to observations during metered hours. The API snapshot data set was joined to the hourly data, and occupancy was cross-validated. During this validation process, 3.4% of observations from the API data were discarded due to inconsistencies with the hourly data. In most cases, these inconsistencies simply reflect sampling error given that the API provides about 12 snapshots per hour while the hourly data averages the second-by-second occupancy data from the sensors.

3.2.3 Analysis Sample

The core sample that we use for the regression analysis consists of all 311 blocks – 55 control blocks and 256 pilot blocks. For the descriptive charts and tables, we weight the data, which reduces the sample to 306 – 55 control blocks plus 251 of the 256 pilot blocks for which API data (and thus arrival rate estimates, which we use as weights) are available. The weighted results are not used for the regression analysis for computational reasons, but as we are assessing changes over time the weighting is less critical.

Where stated, we also include data from the 97 other blocks in the SFMTA dataset, of which 89 also have API data. These additional blocks are included where we estimate the relationship between average occupancy and the probability that a block is full, rather than where we evaluate *SFpark*'s impacts.

Where stated, we further restrict the sample to 262 blocks – the 55 control blocks plus 207 of the pilot blocks where sensor data is available for the entire two-year period. (On some blocks, sensors were installed at a later date, or failed or were removed for construction work.) This restricted sample is used where we graphically illustrate trends over time, in order to keep the composition of the sample consistent.

⁵ While there are 408 blocks in the SFMTA hourly data set, we only captured information on 340 through the API. The API does not provide information for the 55 control blocks. On the remaining missing blocks, street repaving meant that no data was available during the period of data capture, or sensors were installed at a later date.

4. MODELING PARKING AVAILABILITY AND CRUISING

Our empirical analysis entails three main stages. First, using the API data at five-minute intervals, we determine the empirical relationship between the probability of a block being full and its average occupancy. This relationship is of intrinsic interest given that the probability of being full is the measure of direct policy relevance, while performance measures and data collection have typically focused on average occupancy. The probability that an arriving driver finds a block full is also needed for our simulations. Second, using the API data, we estimate an arrival rate for each one-hour interval on each block, in order to weight our results to reflect the driver experience. Third, we simulate the amount of cruising, and compare these estimates to manual surveys conducted by SFMTA.⁶

In all cases, the underlying model for a block is a stationary Markovian multiserver queue (Kleinrock 1976), which assumes n identical resources (parking spaces) and that drivers searching for parking arrive on a block at random via a Poisson process with rate λ . If there is a parking space available, then a driver parks on a first-come, first-served basis. Once parked, each driver's parking duration is random with an exponential distribution of mean $1/\mu$ that is independent of the other drivers' parking durations. The model further assumes that cruising drivers search until they are able to park without giving up. Under this model, in steady state the probability that an arriving driver finds no available parking spaces is derived using the Erlang C formula whose functional form is defined in Eqs 2 and 3:

$$\gamma_n(\rho) \equiv \frac{(n\rho)^n}{n!} \cdot \frac{1}{1-\rho} \cdot P\{N=0\}, \quad (2)$$

where

$$P\{N=0\} = \left(\sum_{k=0}^{n-1} \frac{(n\rho)^k}{k!} + \frac{1}{1-\rho} \frac{(n\rho)^n}{n!} \right)^{-1} \quad (3)$$

is the probability that no cars are parked. The average proportion of time that each parking space is occupied is defined as:

$$\rho \equiv \frac{\lambda}{\mu n}. \quad (4)$$

⁶ In principle, the probability that a block is full could be obtained directly from the sensor data. However, only hourly average occupancy data are retained by SFMTA, and we only have 5-minute API data for a subset of blocks and dates. Similarly, arrival rates are not computed by SFMTA, and in any case would not include the latent arrivals that are estimated by our method.

Thus, the Erlang C formula expresses the relationship between the average occupancy and the probability that a driver finds an available parking space. This model provides a theoretical foundation that enables our empirical strategy. While we have assumed the first-come, first-serve service discipline, the mean number of cruising drivers in the system, the mean waiting time and the Erlang C formula remain valid for service in random order (e.g. if a newly arrived driver “jumps the queue” and beats other cruising drivers to an open spot).

In our empirical estimations, the arrival rates λ_{it} vary over each of N blocks $i = 1 \dots N$ and T time periods $t = 1 \dots T$. In general, each time period t represents a one-hour period on a weekday, Saturday or Sunday/holiday. If each block is a stationary Markovian multiserver queue, then the parked cars departing a block also follow a Poisson process with rate λ_{it} . The independence of the departure process from the average parking time in this model is due to a foundational queueing theory result called Burke’s theorem (Burke 1956).

4.1 Empirical Estimation of Parking Availability

The relationship between the probability of a block being full (Pr[full]) and hourly average occupancy is calibrated from the joined hourly occupancy and five-minute API data. Pr[full] is calculated for each hourly period as the proportion of API observations where the block is full.

The Erlang C formula given in Eq. (2) provides a deterministic relationship between the “true” average occupancy ρ , the number of spaces on each block n , and Pr[full]. While hourly average occupancy $\hat{\rho}$ is observed, this is not the same as ρ . Rather, ρ is a function of the true arrival rate, length of stay and size of block, while $\hat{\rho}$ is the observed outcome following a random process of arrivals. Since Eq. (2) is nonlinear, we cannot simply substitute $\hat{\rho}$ for ρ ; the bias is analogous to that in measurement error models (Cameron & Trivedi 2005). Instead, we estimate the log-linear regression equation in Eq. (5),

$$\log(\text{Pr}[\text{full}]_j) = \log(\gamma_n(\hat{\rho}_j)) + \beta_0 + \beta_1 n_j + \beta_2 \log n_j + \beta_3 n_j \hat{\rho}_j + \varepsilon_j \quad (5)$$

where for each data point j : $\gamma_n(\hat{\rho}_j)$ is the Erlang C formula defined in Eq. 2; n is the number of spaces on the block; $\hat{\rho}$ is the observed occupancy on the block averaged over that hour; Pr[full] is the percentage of API observations on the block during that hour where no space is available; ε is a mean zero error term; and $\beta_0, \beta_1, \beta_2, \beta_3$ are estimated coefficients.⁷ The first term in Eq. (5) comprises the deterministic component of the logged Erlang C formula. The remaining terms provide a flexible parametric

⁷ We estimate the coefficients via iteratively reweighted least squares, using the glm function in the R statistical software package.

component that captures both the effects of measurement error and other departures from the assumptions of the Markovian multiserver queue model.

Table 1 shows the coefficients of the regression model.⁸ Figure 2 (left panel) shows the fit of the regression model predictions against the actual data. As can be seen, the fit is very close across multiple block sizes – a paramount consideration given that the purpose of this regression is entirely predictive. Figure 2 (right panel) shows the smoothed predictions for various block sizes.

The relationship between block size (number of spaces) and Pr[full] can also be seen in Figure 2. For any given hourly average occupancy, Pr[full] decreases as the number of spaces increases. This makes intuitive sense and suggests that a uniform occupancy target, across all block sizes, may be inappropriate from a policy perspective. On a block with only one space, the hourly average occupancy is the same as Pr[full] – the relationship would be a 45-degree line. For very large blocks, a parker has a good chance of finding a space even at an occupancy level of 90% or more.

The parameters estimated in the regression model are applied to the full dataset to estimate Pr[full] for each data point in the hourly occupancy data set – including data points which do not overlap with the API data. In other words, Pr[full] is estimated as a function of the number of spaces on a block and the hourly average occupancy using Eq. 5.

	Estimate	Standard Error
β_0 (intercept)	0.125	0.00287
β_1 (number of spaces)	-1.095	0.00291
β_2 (log number of spaces)	-0.0180	0.00281
β_3 (no. of spaces x hourly average occupancy)	1.094	0.00292
N = 135,153 AIC = -319,052		

Table 1 Predictive Model to Estimate Probability that Block is Full

⁸ Alternative models with a subset of the four coefficients were also estimated. The model in Eq. (5) / Table 1 was selected based on the lowest AIC value, a common model selection criterion. The addition of block fixed effects may give a better fit, but would eliminate the ability to generate predictions for out-of-sample blocks (i.e. blocks in control areas and other blocks missing from the API data). As it is, the model generates very close predictions.

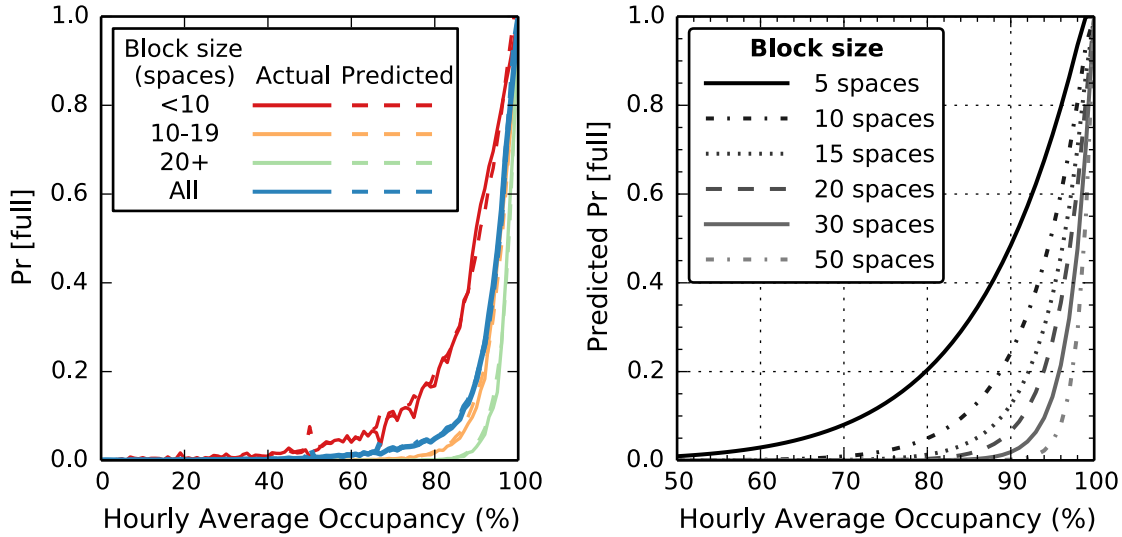


Figure 2 Predicted Probability of a Block Being Full.

Left panel: actual vs. predicted probability. Right panel: predictions for different block sizes

4.2 Estimation of Arrival Rates

We estimate arrival rates in order to weight the simulation results by the level of demand on a block, and thus generate results that are more representative of the driver experience (see Eq. 1). Estimating arrival rates, rather than taking them directly from the data, is necessary for two reasons. First, in the case where blocks are full, the data are censored. In other words, once a block is full, further arrivals are unserved (i.e., the driver cannot park), and are therefore not observed. Second, the granularity of the available data causes a naïve estimate to miss many instances of parking turnover. Suppose, for example, the API data show an occupancy of six vehicles on a block, rising to seven vehicles in the subsequent snapshot five minutes later. One vehicle could have arrived. Or two vehicles could have arrived and one departed. Or three could have arrived and two departed, and so on.

The assumption of a Poisson process for arrivals and departures in the model in Eq. (2) enables the underlying arrival and departure rates to be recovered as the parameters of a Skellam or Poisson Difference distribution. Formally, if $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\gamma)$, then $X - Y \sim \text{Skellam}(\lambda, \gamma)$. In this case, we do observe the quantity $X - Y$ as the change in occupancy between each five-minute snapshot. Both parameters are identified by assuming that arrival and departure rates are constant on each block within each hour-long period.

The Skellam distribution has several well-known properties (Skellam 1948) and the parameters can be estimated by the method of moments or maximum likelihood (Alzaid & Omair 2010). In the present context, the censoring necessitates a modified approach. We therefore maximize the log likelihood function in:

$$\log L(\lambda, \gamma; c_i, m_i) = \sum_{i=1}^N \log \left[f_{SK}(c_i; \lambda, \gamma) + \sum_{j=m_i}^{\infty} \sum_{k=1}^{\infty} f_P(j+k; \lambda) f_P(j-c_i; \gamma) g(j+k, c_i, m_i) \right] \quad (6)$$

where f_{SK} and f_P respectively denote the Skellam and Poisson probability mass distribution functions, given the arrival rate λ and departure rate γ ; c_i denotes the observed change in occupancy between each pair of 5-minute observations i ; and m_i denotes the maximum potential change (i.e., the number of available spaces at the start of the 5-minute period.)

The first term in Eq. (6) represents the likelihood provided by the Skellam distribution, which by definition sums probabilities over all possible combinations of uncensored arrivals j and departures $j - c_i$ that would yield the observed change in occupancy c_i . The remaining terms represent the likelihood summed over all possible combinations of arrivals $j \geq c_i$ and censored arrivals $k \geq 1$. For each possible combination j, k , the likelihood is the product of the probabilities of observing the number of arrivals $j + k$ and departures $j - c_i$, each of which is given by a Poisson distribution.

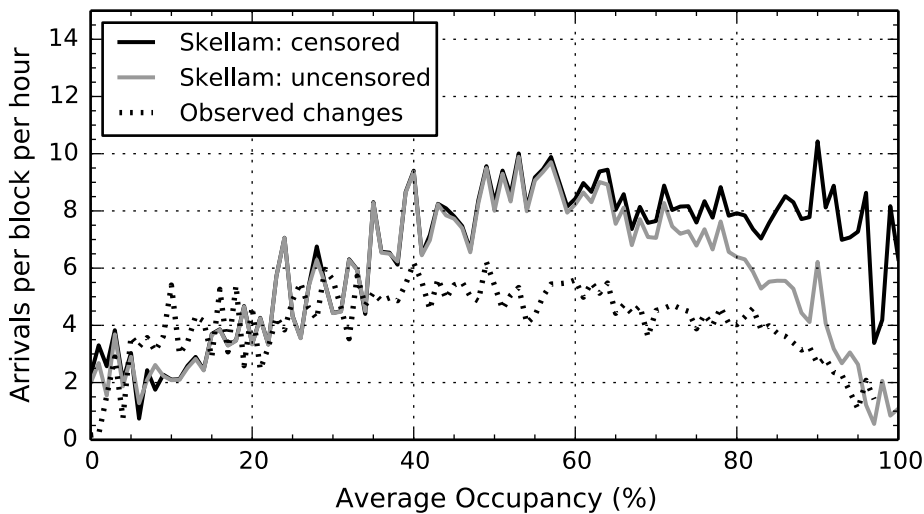
The term g denotes the probability that the combination of arrivals and departures is feasible. For example, take the case of a block with a net occupancy change of zero, one available space, one uncensored arrival and one censored arrival, i.e. $c_i = 0, m_i = 1, j = 1$ and $k = 1$. There are two arrivals and one departure, but they can only occur in the order (a, a, d). With the orders (d, a, a) or (a, d, a), the second arrival will not be censored. We estimate g using a Monte Carlo simulation over 1,000 random permutations of each combination of c, m, j and k . This approach approximates the complex closed form likelihood function of the transient distribution of the underlying multiserver queue (see Kleinrock 1976).

Computationally, the maximum likelihood is estimated iteratively, using the method of moments for the uncensored Skellam distribution (Alzaid & Omair 2010) to provide the starting values. The arrival and departure rates are estimated separately for 72 one-hour intervals on each block, corresponding to each hour of the day on weekdays, Saturdays and Sundays/holidays.

Figure 3 plots the estimated arrival weights as a function of average occupancy on each block. For comparison, two alternative methods of estimating the arrival weights are illustrated. At moderate occupancies, estimates based on the uncensored Skellam distribution are similar to those from the

censored distribution. However, at occupancy levels above about 85%, the two sets of estimates diverge, and the censored estimates are more plausible (there is no drop in arrival rates at high occupancies). Using observed arrivals only (e.g., a change in occupancy from 6 to 8 over a 5-minute period represents two observed arrivals) results in artificially low arrival rates, particularly at high occupancies, as most arrivals are not observed. Because arrivals are unobserved it appears that the arrival rate simply drops off substantially once a block is full, or near full. Obviously there is no intuition for this drop off; indeed the observed drop in arrivals should suggest that drivers continue to arrive but cannot be accommodated, thus they contribute to the cruising problem that the program attempts to address.

Figure 1, which showed the distribution of pilot and control areas, also shows the distribution of arrivals, plotting the blocks with widths proportional to the arrival rate. (Control blocks, for which no API data are available, are shown with a uniform width.) Interestingly, there are no clear geographic patterns, other than low arrival rates on some side streets and others with just a handful of metered spaces, which is to be expected given that arrival rates are estimated on a per-block rather than per-space basis.



Sample: all 340 blocks in the API dataset

Figure 3 Comparison of Alternative Arrival Weight Estimators

4.3. Cruising Simulation

For each hourly observation in the SFMTA dataset, a single cruising simulation is run.⁹ A parker ‘arrives’ at each block within each hourly period, and finds a space on that block with probability $1 - \text{Pr}[\text{full}]$. If a space is found, then the number of blocks cruised is recorded as zero. Otherwise, the parker randomly

⁹ Multiple simulations could be run and the average number of blocks cruised taken. However, our approach preserves the full range of the distribution.

selects an intersecting block,¹⁰ and finds a space on that block with $1 - \text{Pr}[\text{full}]$. Thus, the parker proceeds via a random walk through the neighborhood until either a space is found, or the cap of 30 blocks cruised is reached (and by assumption, the parker gives up or parks off-street). Note that blocks (including the original block) can be visited multiple times. We assume a first-come, first-served system; relaxing this assumption would not affect mean cruising times, but would slightly lengthen the tail of the distribution.¹¹

This approach has several advantages over alternative methods of assessing cruising. Most importantly, it allows cruising to be estimated for every hour on every block, rather than the small sample possible with manual methods. It also avoids potential selection bias whereby cruising surveys may focus on busy blocks at busy times. The downsides are primarily as follows:

1. Blocks are chosen at random from the set of blocks that intersect the parker's current location. Refinements to this process could take account of the direction of travel, one-way streets and other restrictions, but we do not consider these here.
2. Intra-hour correlations between neighboring blocks are not considered. The simulation only considers blocks within the same hourly observation period (e.g. 10-11AM on June 22, 2011). However, it is likely that $\text{Pr}[\text{full}]$ is spatially correlated within each hour. For example, if at 10:10AM a particular block is full, the neighboring block is more likely to be full at that precise time than would be expected from the occupancy averaged over the hour.
3. Only streets with sensor data and general metered parking are included. Thus, cruising on residential side streets that are not equipped with sensors and other unmetered blocks is not accounted for.
4. The simulation assumes that a parker can take advantage of available spaces on either side of the street. On a one-way street, this is realistic (except on wide streets at times of heavy traffic volume). On two-way streets, this may require illegal U-turns, except on streets with perpendicular parking. In practice, many motorists do make U-turns to secure a vacant space on the opposite side of the street, particularly if available parking is scarce. However, the assumption will underestimate cruising from more law-abiding drivers.

¹⁰ Some blocks on the fringes of a pilot or control area have few intersecting blocks with sensor data. In these instances, the block's nearest neighbors (e.g. parallel blocks) are added to ensure that at least four blocks are available to be chosen at random as the parker cruises. Also note that the randomly selected block is constrained to be within 400m (~1/2 mile) of the original block, ensuring that the parker cannot cruise too far from his or her presumed destination.

¹¹ A related issue is that it takes some time for drivers to cruise along a block, and that a vacant space may be taken by another driver in the meantime. However, we interpret $\text{Pr}[\text{full}]$ as the steady state probability, and thus it is equally likely that a space is vacated ahead of the driver.

Points 2 and 4 and, to the extent that residential streets with free 2-hour parking have higher occupancy rates, point 3 are likely to mean that our estimates of cruising under count and are, therefore, a lower bound on cruising.

5. RESULTS

In this section we evaluate the first two years of *SFpark* by applying the Erlang C-based estimate of the probability of a full block, the estimation methodology for the arrival rates (used only for weighting the data), and the cruising simulation. Unless stated, all results are weighted by arrival rate.

5.1 Evaluating *SFpark* Average Occupancy Targets

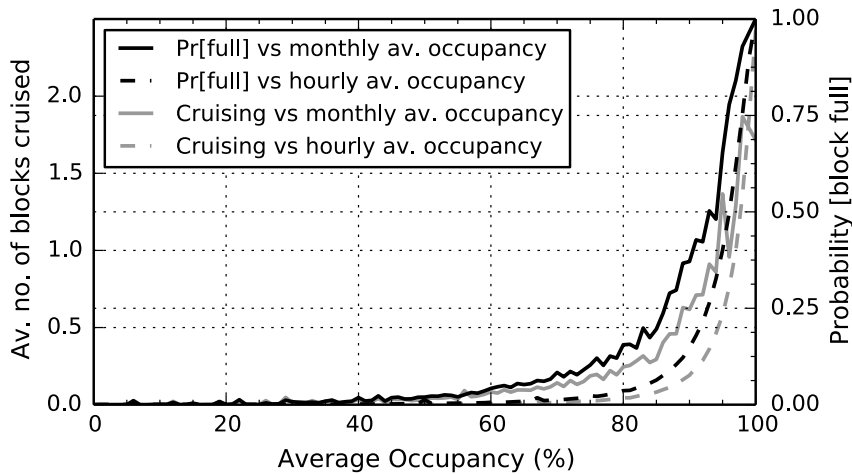
Figure 4 plots the relationship between average occupancy and two related metrics: the probability that a block is full (black lines), and the average number of blocks cruised per parking attempt (gray lines). The probability that a block is full is estimated from the Erlang C-based regression model. The average number of blocks cruised is estimated from the cruising simulations. The latter is a function of $\text{Pr}[\text{full}]$, but also depends on the spatial correlation of occupancy, i.e. the probability of finding a space on neighboring blocks if the desired block is full.

There are two points of particular note from the chart. First, both the probability that a block is full and cruising have nonlinear relationships with average occupancy. Below 90% occupancy, there is almost no cruising – even if no space is available on a particular block, a driver is likely to find space on the next block visited. Above 90% occupancy, however, the expected cruising distance increases dramatically. The probability of a block being full also increases more rapidly at higher average occupancies. Hence, we see that moving from 90% to 85% average occupancy has much more impact on cruising than moving from 85% to 80%.

Second, it matters how average occupancy is computed. For any given average occupancy, further aggregation (solid lines) results in a higher probability that a block is full and implies more cruising than a lower aggregation would indicate (dotted lines). This occurs because of the nonlinear (convex) shape of the curves in Figure 4; the reasons were discussed above in Section 2, *Understanding Parking Performance*. Intuitively, a longer averaging period will include more observations with higher-than-average occupancies where both cruising and $\text{Pr}[\text{full}]$ are much higher, as well as lower-than-average occupancies where cruising and $\text{Pr}[\text{full}]$ are more similar to those at the mean.

A direct implication of this second finding is that average occupancy targets should not be set without reference to the period over which the average is calculated. If a two-week period of averaging is used, as

in the case of *SFpark*, then a lower occupancy target may be appropriate to ensure availability and achieve a given level of cruising. If a day or two of data are used (perhaps as in a small town with fewer data collection resources and fewer parking spaces), then a higher occupancy target may be appropriate.. Another implication is that the variance in occupancy matters as well as the mean. The higher the variance, the greater the probability of full will be for any given mean occupancy.



Sample: 395 pilot, control and other blocks with sensor data; metered hours only

Figure 4 Length of occupancy averaging period vs. measures of parking performance

5.2 Changes in Occupancy and Cruising Under *SFpark*

We evaluate changes over time in occupancy and cruising through both a descriptive analysis (Figures 5 and 6; Table 2) and a regression analysis (Table 3). In both cases, we compare changes in pilot areas (where rates have been adjusted) to those in control areas (where sensor data exists but no rate changes have been made). This analysis is intended to illustrate the application of the metrics developed earlier in this paper, and to provide initial evidence for the overall impacts of *SFpark*.

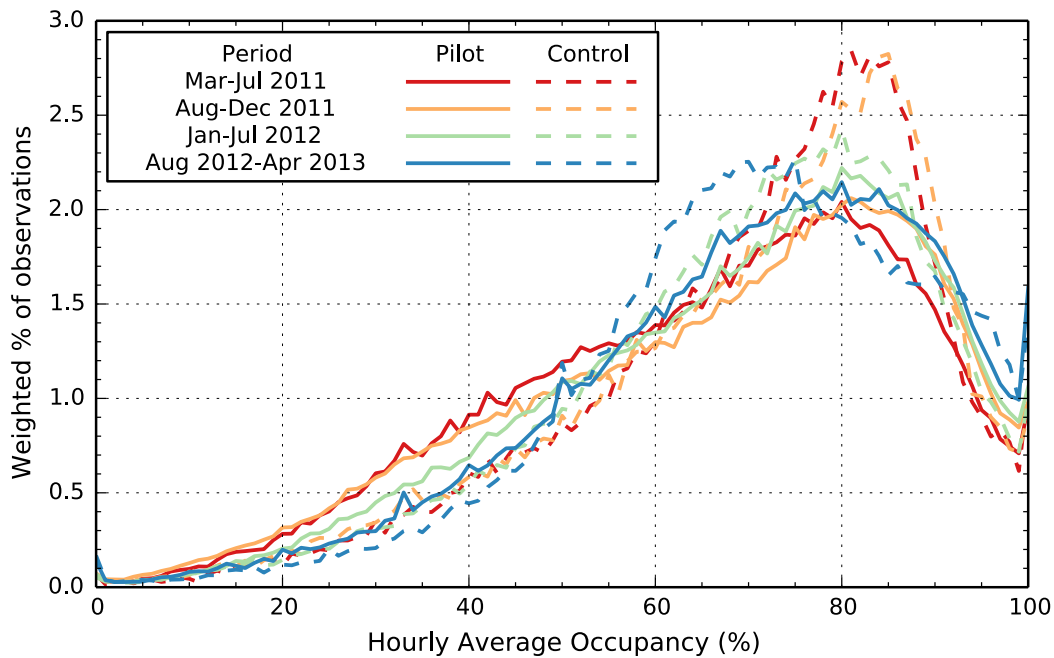
5.2.1 Descriptive Results

Figure 5 plots the changes in the distribution of hourly average occupancy for four periods: the baseline – i.e. March to July 2011, which was before the advent of any rate changes, and three subsequent periods. There has been little visually discernible change in the distribution of occupancy over time. Despite significant rate changes on some blocks, both upwards and downwards, hourly average occupancy has hardly budged. Indeed, there is a slight increase in the number of blocks at very high occupancies, including 100%. However, the control areas have experienced an even greater increase in this regard, suggesting that citywide trends of increasing occupancy – perhaps due to economic growth – may be important.

The plot of the distribution also shows that parking availability is generally good within the SFpark study area. Very few blocks are fully occupied, and the majority of the system operates at less than 80% occupancy. These data are shown in Table 2 where it is easy to see that the control areas have more blocks with average occupancies of 81% to 90% while the pilot areas are more likely to have blocks in the 96% to 100% occupancy range.

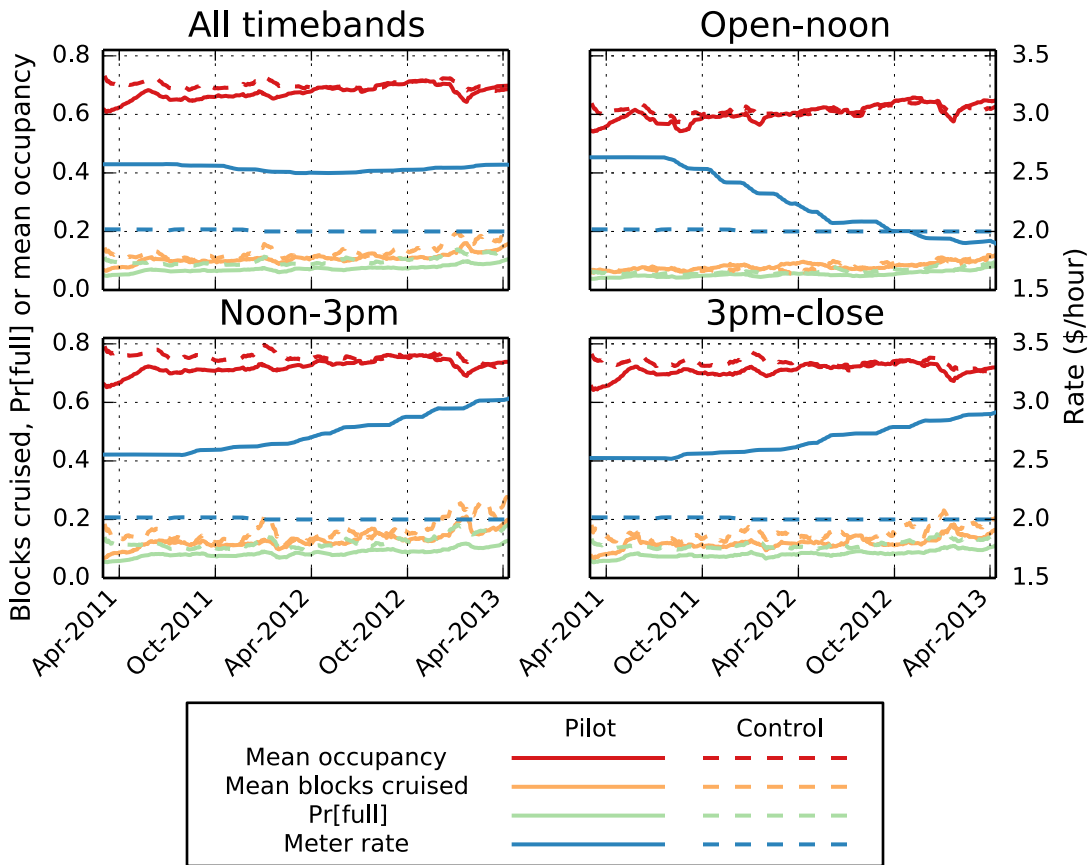
The left axis of Figure 6 shows changes in three metrics: hourly average occupancy, the probability that a block is full, and average blocks cruised. On the right axis, we show the average meter rate per hour. As in the previous plots, the period of analysis is March 2011 through April 2013, and observations are weighted by the arrival rates, so that blocks that have more arrivals and more latent demand are weighted more highly. Results for the pilot areas are shown with solid lines, and the control areas with dashed lines.

As would be expected given that there has been little change in the distribution of hourly average occupancy, there are few clear trends evident from the charts. Parking rates (blue lines) have come down in the morning period, and increased in other periods. However, these rate changes have had little discernible impact on cruising, hourly average occupancy or the probability that a given block is full. Average hourly occupancy and the probability that a block is full have remained almost constant or increased slightly since May 2011.



Sample: 262 blocks with continuous sensor record; weekdays and metered hours only

Figure 5 Changes over time in distribution of hourly average occupancy



Sample: 262 blocks with continuous sensor record; weekdays and metered hours only

Figure 6 Changes in occupancy, cruising and meter rates over time

Rate Adjustment Period										
Baseline (Mar-Jul '11)	Period 1 (Aug-Oct '11)	Period 2 (Oct-Dec '11)	Period 3 (Dec '11- Feb '12)	Period 4 (Feb-Mar '12)	Period 5 (Mar-May '12)	Period 6 (May-Aug '12)	Period 7 (Aug-Oct '12)	Period 8 (Oct '12- Jan '13)	Period 9 (Jan-Mar '13)	Period 10 (Mar '13 – Apr '13)
Control	23.1%	25.0%	23.4%	24.7%	24.7%	25.7%	21.4%	23.5%	26.7%	27.4%
0 to 59%	38.9%	35.5%	35.9%	38.0%	39.7%	42.9%	43.9%	40.8%	41.0%	40.2%
60 to 80%	14.1%	13.5%	13.7%	13.9%	12.5%	8.9%	10.0%	9.3%	7.9%	8.0%
81 to 85%	12.0%	12.1%	13.3%	11.6%	11.1%	8.2%	8.9%	8.4%	7.8%	7.3%
86 to 90%	7.0%	7.1%	8.3%	6.7%	6.9%	7.9%	8.3%	8.7%	8.1%	8.1%
91 to 95%	3.7%	3.5%	4.2%	4.1%	4.0%	5.0%	6.0%	7.5%	6.5%	6.6%
96 to 99%	1.2%	1.1%	1.2%	1.1%	1.0%	1.4%	1.6%	1.8%	2.0%	2.5%
100%										
Pilot										
0 to 59%	36.8%	33.1%	34.0%	32.1%	29.2%	29.5%	25.4%	31.6%	30.2%	28.9%
60 to 80%	34.8%	32.3%	33.6%	33.2%	34.4%	36.3%	37.3%	35.7%	37.7%	36.6%
81 to 85%	9.5%	9.8%	10.4%	10.4%	11.1%	10.1%	11.2%	9.5%	8.9%	9.0%
86 to 90%	8.3%	9.3%	9.6%	10.3%	10.6%	9.6%	10.7%	8.9%	8.7%	9.3%
91 to 95%	6.1%	7.4%	7.9%	7.8%	8.3%	8.2%	8.8%	7.3%	7.5%	8.2%
96 to 99%	3.6%	4.4%	4.6%	4.8%	5.1%	5.2%	5.2%	4.8%	4.8%	5.5%
100%	0.9%	1.1%	1.0%	1.3%	1.2%	1.2%	1.3%	2.2%	2.1%	2.5%

Notes: SFMTA target range shown in green. Early rate adjustments were slightly staggered across different neighborhoods (generally over a single week), and so dates are approximate.

Sample: 262 blocks with continuous sensor record; metered hours only

Table 2 Distribution of occupancies

5.3 Regression Analysis

To formally test the impacts of the first ten rate adjustments, we use two dependent variables in the regression analysis. The first is the number of percentage points by which occupancy falls outside the 60-80% range. (Almost identical results are obtained using a model where the dependent variable is the absolute difference from 70% occupancy, the center of the target range.) The second dependent variable is the number of blocks cruised. The different regression models include neighborhood- or block-level fixed effects to account for unobserved characteristics of each neighborhood or block, such as the number of nearby destinations; and to control for time variation using some combination of fixed effects for date, month, day of week and time of day. The key identifying assumption is that citywide trends, such as economic activity and gasoline prices, affect pilot and control groups in the same way. One of the models is a negative binomial specification, which is suitable given the large number of zeros in the dependent variables (recall that a zero represents an observation within the target 60-80% range).¹²

The key independent variable of interest is the number of rate changes. As shown in Table 3, the results indicate a small but (in most cases) statistically significant impact of each rate change on occupancy, with these individually small impacts adding up to larger cumulative effects over the first two years. On average, each rate change brings a block 0.1 to 0.2 percentage points closer to the 60-80% range, and thus the average impact is 1-2 percentage points after ten rate changes.

The impacts on cruising are smaller, and in some of the models are not statistically significant at conventional levels. However, these signs are still encouraging, and suggest that each rate change reduces the average search distance for parking by 0.007 to 0.017 blocks – or 0.07 to 0.17 blocks after the tenth rate change. Given that the mean number of blocks cruised in the pilot area is 0.13, this represents a reduction in cruising of more than 50% after two years of *SFpark*, compared to the situation without the *SFpark* program.

A more flexible regression specification uses dummy variables for each rate change, which relaxes the assumption of the models in Table 3 that impacts are linear in the number of rate changes. Each dummy variable is zero in control areas and in pilot areas indicates whether an observation falls within a given rate adjustment period. The marginal effects of these dummy variables are plotted in Figure 7. The qualitative conclusions remain unchanged – small impacts after each rate change, which in cumulative terms amount to substantive changes after ten rate changes.

¹² Due to difficulties in achieving convergence, neighborhood- rather than block-level fixed effects are used in the negative binomial model for occupancy. This may explain why the estimates are somewhat lower than the OLS models, and suggests that the OLS model results are preferred.

The positive results from the regression models come in spite of the fact that availability does not appear to have improved in the SFpark pilot areas (Table 2 and Figure 5). The interpretation here is that the first two years of SFpark were characterized by a rebounding local economy, which is likely to have increased parking demand throughout the city. Even if parking availability and cruising worsened in the pilot areas, they worsened more in the control areas. Thus, our results can be interpreted as the impact of SFpark relative to a counterfactual situation where SFpark was not implemented, rather than to the impacts relative to the pre-implementation period. It is also important to note that our results indicate the combined effect of all the components of SFpark over time, including real-time driver information and new payment options as well as rate changes.

One concern in the analysis might be that the control areas are fundamentally different from the pilot areas, making them a questionable point of comparison. The pilot areas include the central business district (Downtown, South Embarcadero and Civic Center) as well as the tourist-oriented Fisherman’s Wharf. Dropping these four areas from the analysis ensures that both the remaining pilot areas and all control areas are characterized by neighborhood commercial development surrounded by primarily residential areas. The restriction reduces the magnitude of the estimated coefficients in the occupancy models to about a quarter of those in Table 3, while the estimates for cruising remain similar to those based on the entire sample. Moreover, not all estimates are statistically significant at conventional levels. However, while the conclusions are less definitive, the evidence still points to a substantive impact of SFpark in the first two years.

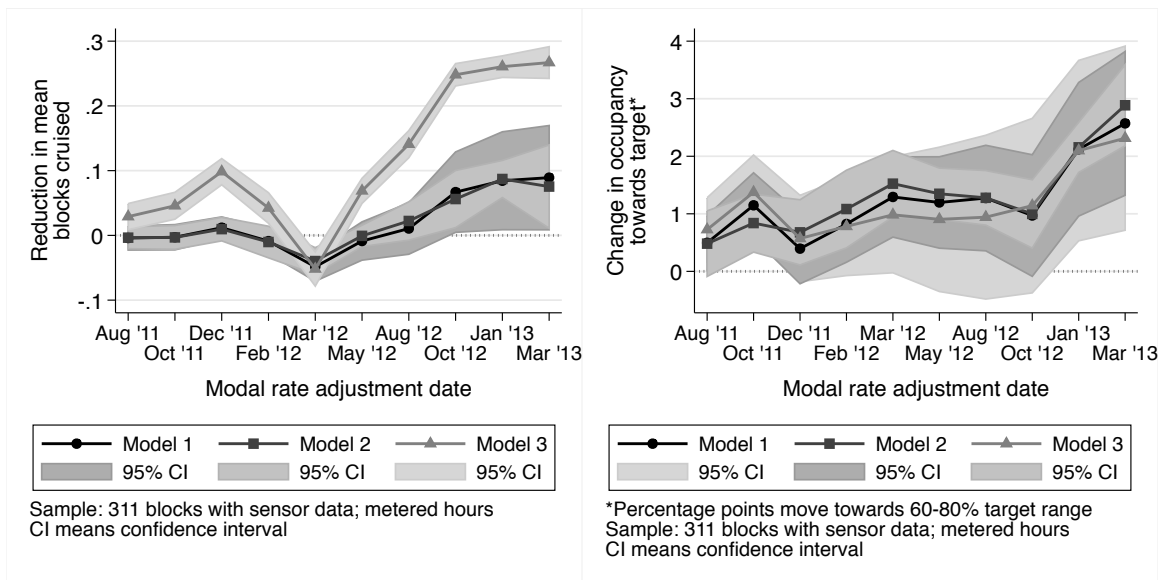


Figure 7 Impacts of rate changes over time on cruising (*left*) and occupancy (*right*)
Note: Marginal effects rather than coefficients are shown. Y-axis is inverted compared to Table 3.

Dependent Variable	Occupancy (distance from target range) ¹			Blocks cruised		
	1	2	3	4	5	6
Specification	OLS	OLS	Negative binomial	OLS	OLS	Negative binomial
Number of rate changes	-0.2059 (0.0632)	-0.2133 (0.0193)	-0.0888 ² (0.1023)	-0.0067 (0.0037)	-0.0067 (0.0011)	-0.0170 ² (0.0008)
Capacity (spaces per block) ³			-0.0144 (0.0025)			
Fixed effects:	Block, month, day of week, hour	Block, date, hour	Neighborhood, month, day of week, hour	Block, month, day of week, hour	Block, date, hour	Block, month, day of week, hour
Standard errors	Clustered by block	Clustered by date	Clustered by block	Clustered by block	Clustered by date	Observed information matrix
N	1,953,150	1,953,150	1,953,150	1,953,037	1,953,037	1,942,109

(1) Dependent variable: Percentage points by which observation falls outside [60%,80%] interval for occupancies

(2) Marginal effect of each rate change, rather than coefficient, is shown.

(3) In other models, capacity is captured in the block fixed effects

Standard errors in parentheses. Sample: 311 blocks with at least partial sensor data; metered hours only

Table 3 Regression models to identify impacts of rate adjustments

5.4 The Spatial and Temporal Patterns of Cruising

Given that reducing cruising is one of the central goals of *SFpark*, it is useful to analyze patterns of cruising in more detail. Table 4 shows the overall distribution of cruising. Recall that the simulations assume that parkers can take advantage of vacant spaces on either side of the street. If parkers are restricted to one side of the street (thus cutting the effective capacity of each block in half), the mean number of blocks cruised increases by about 50% in both pilot and control areas. Note that because of the nonlinear relationship between occupancy and cruising, the number of blocks cruised will usually be a more volatile metric than average occupancy, as it is highly dependent on the number of blocks that are full at a given time and their spatial correlation.

The data in Table 4 suggest that cruising is limited during metered hours,¹³ with a weighted mean of 0.13 blocks cruised (equivalent to 14 meters, or just a few seconds, plus the distance driven part-way along the block where the driver finds a space) for each driver arriving and looking for parking. It is important to

¹³ The overall level of cruising during non-metered hours is very similar. However, as shown later in the paper (see Figure 6), this conceals a large amount of cruising immediately after meters cease to operate, coupled with much lower levels of cruising in the early hours of the morning.

stress that this represents a lower bound on cruising for several reasons. First, as noted above, cruising would increase by about 50% if parkers could only take advantage of vacant spaces on one side of the street. Second, even if a vacant space is available, it may not be taken by a parker – perhaps due to an impending tow-away restriction or a 30-minute time limit. Third, our simulation assumes that drivers follow a random path and are unimpeded by one-way streets and turn restrictions. However, the overall message in qualitative terms is that cruising is perhaps primarily a perception problem, or is localized on under-priced or unmetered streets or at unmetered times. Indeed, *SFpark* has begun to address these issues through improved driver information, through installing new meters and through extending hours of operation.

Table 4 also shows the comparison to SFMTA’s manual cruising surveys. The manual survey data are gathered by data collectors following a prescribed route by bicycle or car, and recording the location of the first vacant metered parking space they find. The manual surveys find considerably more cruising than our simulations, and in particular, many more searches that result in long searches for parking. The reasons for the differences are likely to be as follows:

1. Weighting by the arrival rate substantially reduces the amount of cruising. This is because the probability of a block being full is much lower on longer blocks, and these longer blocks have higher arrival rates. For comparison, the unweighted mean number of blocks cruised is 0.20.
2. The manual surveyors only search for parking on the right-hand side of the street on two-way streets, while our simulations allow for illegal U-turns. As noted above, restricting the search to one-side of the street would increase the simulated mean number of blocks cruised by about 50%.
3. Our simulations do not account for intra-hour correlations across blocks. For example, if the first block is full at a particular time, the next block is more likely to be full, even if hourly average occupancies are moderate.
4. The manual surveys take into account temporary restrictions such as construction, which the sensor data may not account for.

Overall, the comparison to the manual surveys suggests that the simulations in pilot areas provide a reasonable but lower-bound estimate of the amount of cruising. It is unclear why cruising is much higher in the manual surveys in the control areas, but the non-random nature of the manual surveys, where the prescribed survey route begins on the same block and follows the same prescribed route each time, is likely to play a role. Even if a vacant space is clearly visible on a side street, the surveyors will ignore it, thus inflating estimates of cruising compared to a more realistic model of driver behavior. In the most frequently sampled control area (Inner Richmond), the starting block has high average occupancy and a small number of metered spaces. In other words, the manual surveys are likely to provide an upper-bound estimate of the amount of cruising, while our simulation provides a lower-bound estimate.

Blocks Cruised	Weighted % of simulations		% of SFMTA manual surveys	
	Pilot Areas	Control Areas	Pilot Areas	Control Areas
0 (parking found on initial block)	91.7%	88.7%	76.6%	60.3%
1	5.8%	8.5%	11.0%	18.5%
2	1.5%	1.9%	5.1%	7.3%
3	0.5%	0.5%	1.3%	3.3%
4	0.2%	0.2%	2.0%	3.7%
5	0.1%	0.1%	1.1%	1.2%
6 or more	0.2%	0.1%	3.0%	5.8%
Mean blocks cruised	0.13	0.16	0.75	1.25

Sample: 306 blocks with at least partial sensor data; metered hours only

Table 4 Distribution of Cruising (metered hours)

It is also instructive to examine the pattern of cruising and other metrics over the course of the day. The previous sections reported results for metered hours only, but Figure 8 shows an extended simulation for all hours of the day. (When meters are not operational, the simulation assumes that parkers take the first available space, whether or not it is metered.) The figure plots the data for each pilot and control district separately, due to the different temporal patterns and meter operating hours across districts. The shaded areas represent typical meter operating hours.

In all districts except downtown and the tourist-oriented Fisherman’s Wharf, both the probability that a block is full and the amount of cruising spike immediately after metered hours end. Price-sensitive parkers may be delaying their arrival until free parking is available, or perhaps searching for free parking on residential blocks earlier in the day when meters are in operation. The implication is that cruising is of greater concern outside metered hours, and possibly on non-metered blocks.

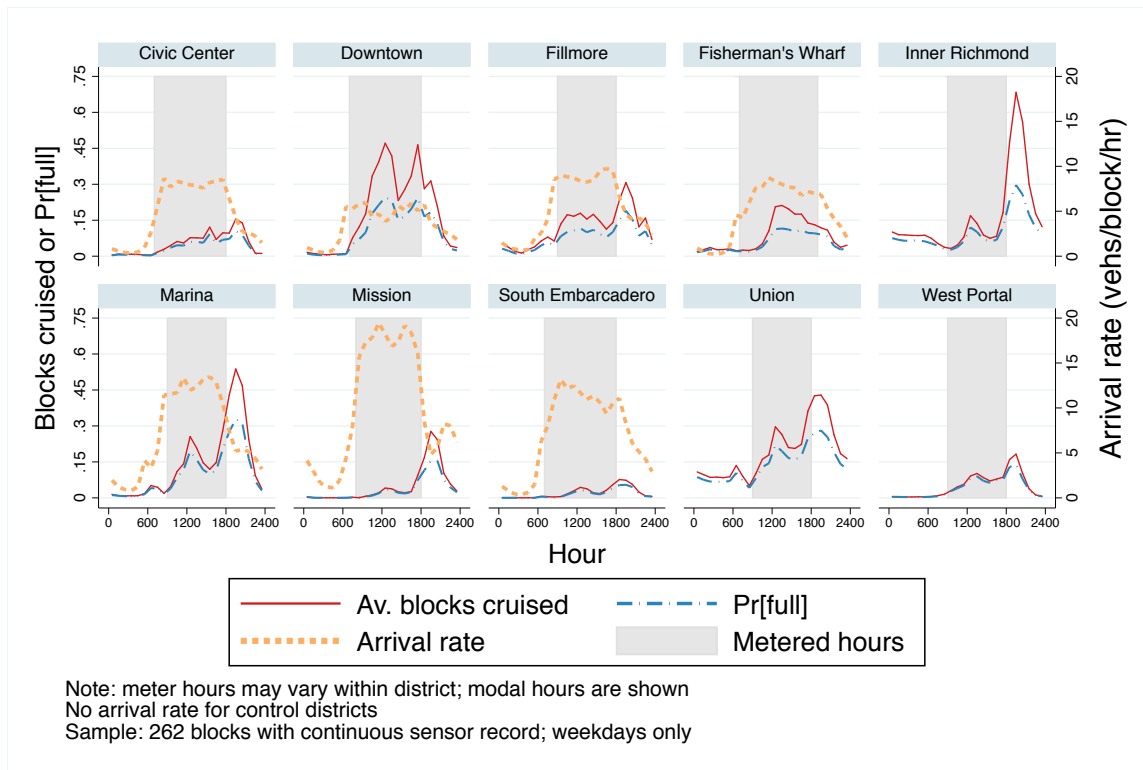


Figure 8 Cruising, Pr[full] and arrival rate by district and time of day

Finally, it is worth noting another important implication of the non-linearities in occupancy, probability of a block being full and the related cruising. We showed in Table 2 and Figures 5-6 that the distribution of occupancies has changed very little. In the first two years of *SFpark*, a very small number of streets were highly problematic and likely contributed disproportionately to cruising in San Francisco. As a small percentage of the sample, the incidence of high-occupancy blocks may seem trivial – 5% of weighted observations have occupancy greater than 96%, and less than 2% have 100% occupancy. At the same time, there were more than 85,000 metered hours over the course of the first two years when blocks had an average occupancy of 100%, and more than 186,000 with an average occupancy of 97% or greater.

Figure 9 shows the locations of the 43 blocks that accounted for 50% of the hours where occupancy exceeds 96% during metered times. These blocks also account for almost half of the cruising.

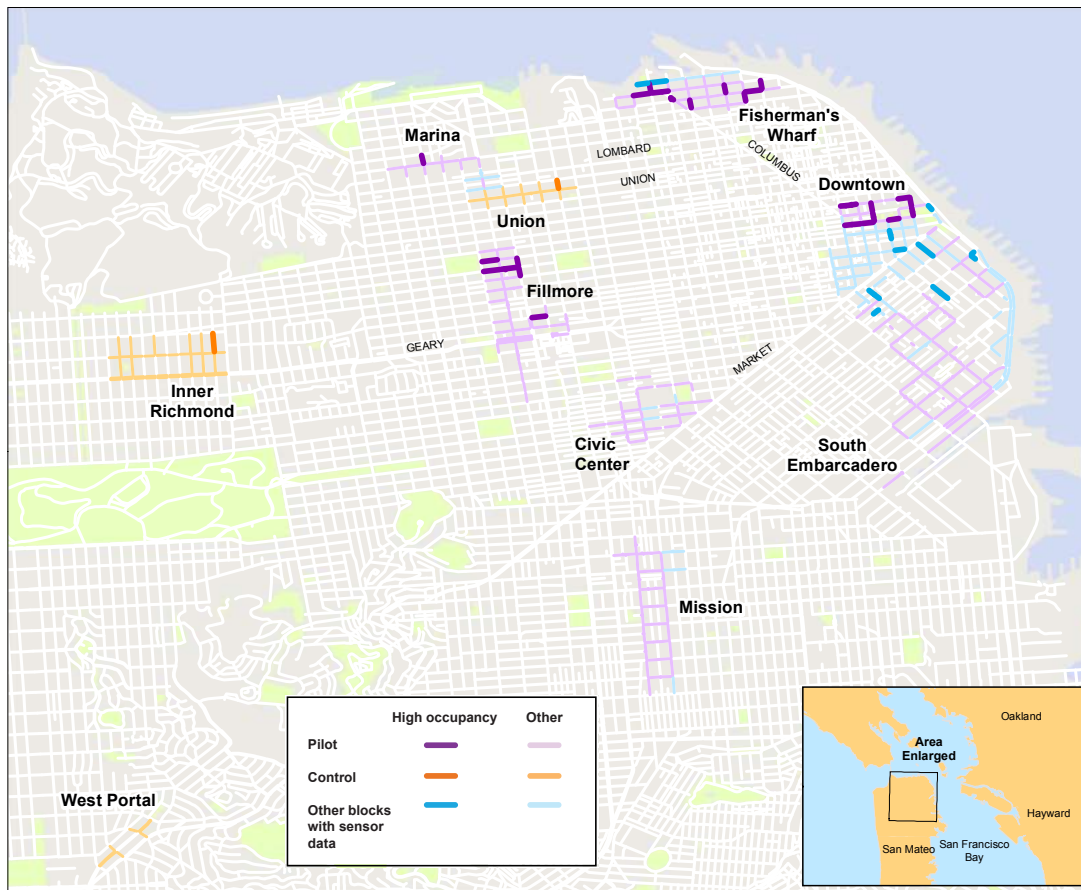


Figure 9 Distribution of high-occupancy blocks
(blocks that account for 50% of hours with occupancy greater than 96%)

6. DISCUSSION OF RESULTS AND CONCLUSION

In 2011, San Francisco embarked on the most ambitious parking management experiment in the United States, if not the world. In this paper, we estimate that the program is slowly achieving its goal of moving occupancy into its target range of 60-80%. We also estimate that *SFpark* has reduced cruising for parking by about 50%, although these results are less certain. The impacts of each rate change are individually small, but after the ten rate changes evaluated in this paper, the changes are shown to be cumulatively substantive. Gradual impacts over time also suggest that it has taken time for rates to come up to market-changing levels and for parkers to adjust to rate changes, learning where to find blocks with cheaper and more abundant parking. This is consistent with findings elsewhere, where we and others are unable to detect any short-term impact of *SFpark* after each individual rate change (Millard-Ball et al. in press; Chatman & Manville 2013).

It should also be noted that these impacts come against a background of other policy changes implemented under *SFpark* – relaxing time limits, providing more information to users and improving other aspects of the parker experience – and a rebounding economy. These changes would normally be expected to increase demand and reduce availability, making the positive results in this paper even more notable. Moreover, the results also come in spite of likely latent demand on some blocks that are persistently full, or more-than-full due to the presence of double or other forms of illegal parking.

The strength of our simulation approach allows for cruising to be assessed throughout the metered system rather than remaining restricted to manual surveys, which, at best, only spot-check cruising. The approach is also generalizable to other cities with similar data, and the conclusions are also generalizable more broadly. There are also some shortcomings to consider, particularly related to our need to estimate the arrival rate from aggregated data. While our estimate is theoretically robust, it is based on five-minute observations and could be refined with finer grain data.

In addition to the policy-relevant impacts, the *SFpark* experiment is already providing a large volume of useful data that allows the merits of different performance metrics to be tested, and the relationships between different metrics to be examined. It is important to bear in mind that average occupancy is not the metric of interest when determining impacts on congestion and air pollution. In this paper, we show how to estimate the relationships between average occupancy; the probability that a driver can find a parking space; the amount of cruising; and actual and latent parking demand in terms of the number of arrivals to a block.

As we discuss, 85% occupancy has been widely promoted in the literature as an optimum performance standard. Our analysis, consistent with that of others, shows this to be a reasonable threshold. Below 85% occupancies, drivers can usually find a parking space, while the probability of finding a space on streets with occupancies above 85% goes quickly to zero (Figure 2). More succinctly, with continuous occupancies below 95% there is virtually no cruising. However, the precise impacts of the performance standard will vary depending on the size of the block and the length of the period over which occupancy is averaged. The fewer spaces on the block and the longer the period of averaging, the lower the occupancy standard needed to achieve a given availability to the parker and a given level of cruising.

A simple rule of thumb such as 80% or 85% is useful in data-poor settings (i.e., almost everywhere except San Francisco at present). But the analysis based on *SFpark* data allows policy makers, even in data-poor settings to make more nuanced decisions with respect to performance standards and rate setting. A

performance standard that varies by block, or is based on the probability of the block being full rather than average occupancy, may be a more policy-relevant basis for changing rates.

Future research should certainly consider the impacts of performance-based parking management initiatives over a longer time frame, as the parking system moves towards a new equilibrium. Moreover, there are several wider issues of interest that are beyond the scope of this paper. *SFpark* and similar programs in other cities provide a rich source of data to analyze the spatial pattern of parking changes and the interactions between on- and off-street parking. Determining price elasticities and changes in consumer surplus are also potential avenues for investigation, along with *SFpark*'s impacts on wider issues of policy interest such as vehicle travel, retail revenue, transit operating speeds and traffic safety.

Finally, we have not investigated the bigger question of parking availability on trip making. As cities see lower cruising due to better curb management it is also conceivable that there is more turnover which could imply more trip making. More trip making could mean more efficient use of the road network but could also imply different impacts on congestion (perhaps on different parts of the system) and greater emissions associated with greater vehicle travel. On the other hand, the greatest impacts of a program such as *SFpark* on vehicle travel may be the hardest to measure. If performance-based pricing can succeed in improving not only parking availability but in creating the perception that parking is easy to find, then potential long-term benefits lie in defusing political pressure for additional off-street parking, and increasing the competitive advantage of urban neighborhoods.

ACKNOWLEDGEMENTS

We are grateful to SFMTA staff, particularly Jay Primus, Leslie Bienenfeld, Stephanie Nelson and Alex Demisch for providing access to *SFpark* data, and for numerous clarifications on the data and price adjustment methodology. We also thank SFMTA staff, Allen Greenberg and members of the USDOT evaluation team for helpful comments on earlier drafts. The conclusions do not necessarily reflect the views of SFMTA or USDOT. We are grateful to Emma McDonnell at UC Santa Cruz for producing the maps. This work is partially supported by the National Science Foundation under Grant No. CMMI #1055832 and US DOT University Transportation Center TRT #DTRT12GUTC11.

REFERENCES

- Ahmadi Azari, K. et al., 2013. Modelling demand under parking and cordon pricing policy. *Transport Policy*, 25, pp.1–9.
- Alzaid, A.A. & Omair, M.A., 2010. On the Poisson Difference Distribution. Inference and Applications. *Bulletin of the Malaysian Mathematical Sciences Society*, 8(33), pp.17–45.

- Anderson, S.P. & de Palma, A., 2004. The economics of pricing parking. *Journal of Urban Economics*, 55(1), pp.1–20.
- Arnott, R. & Inci, E., 2006. An integrated model of downtown parking and traffic congestion. *Journal of Urban Economics*, 60(3), pp.418–442.
- Arnott, R. & Rowse, J., 2009. Downtown parking in auto city. *Regional Science and Urban Economics*, 39(1), pp.1–14.
- Arnott, R. & Rowse, J., 1999. Modeling Parking. *Journal of Urban Economics*, 45(1), pp.97–124.
- Bagloee, S.A., Asadi, M. & Richardson, L., 2012. Methodology for Parking Modeling and Pricing in Traffic Impact Studies. *Transportation Research Record: Journal of the Transportation Research Board*, 2319, pp.1–12.
- Barata, E., Cruz, L. & Ferreira, J.-P., 2011. Parking at the UC campus: Problems and solutions. *Cities*, 28(5), pp.406–413.
- Benenson, I., Martens, K. & Birfir, S., 2008. PARKAGENT: An agent-based model of parking in the city. *Computers, Environment and Urban Systems*, 32(6), pp.431–439.
- Bureau of Public Roads, 1956. *Parking Problems in Cities*, Washington, D.C.
- Burke, P.J., 1956. The Output of a Queuing System. *Operations Research*, 4(6), pp.699–704.
- Button, K., 2006. The political economy of parking charges in “first” and “second-best” worlds. *Transport Policy*, 13(6), pp.470–478.
- Calthrop, E., Proost, S. & Van Dender, K., 2000. Parking Policies and Road Pricing. *Urban Studies*, 37(1), pp.63–76.
- Cameron, A.C. & Trivedi, P.K., 2005. *Microeconometrics: methods and applications*, Cambridge: Cambridge University Press.
- Chatman, D. & Manville, M., 2013. *Too little, too soon? A preliminary evaluation of congestion-based parking pricing in San Francisco*,
- D’Acierno, L. et al., 2011. A Multimodal Multiuser Approach for Analysing Pricing Policies in Urban Contexts. *Journal of Applied Sciences*, 11(4), pp.599–609.
- Gallo, M., D’Acierno, L. & Montella, B., 2011. A multilayer model to simulate cruising for parking in urban areas. *Transport Policy*, 18(5), pp.735–744.
- Glazer, A. & Niskanen, E., 1992. Parking fees and congestion. *Regional Science and Urban Economics*, 22(1), pp.123–132.
- Kelly, J.A. & Clinch, J.P., 2006. Influence of varied parking tariffs on parking occupancy levels by trip purpose. *Transport Policy*, 13(6), pp.487–495.
- King, D., 2010. Estimating Environmental and Congestion Effects from Cruising for Parking. In *Paper presented at Transportation Research Board Annual Meeting*. Washington, D.C.
- Kleinrock, L., 1976. *Queueing Systems*, Wiley.
- Levy, N., Martens, K. & Benenson, I., 2013. Exploring cruising using agent-based and analytical models of parking. *Transportmetrica A: Transport Science*, 9(9), pp.773–797.
- Martens, K. & Benenson, I., 2008. Evaluating Urban Parking Policies with Agent-Based Model of Driver Parking Behavior. *Transportation Research Record: Journal of the Transportation Research Board*, 2046, pp.37–44.

- Massey, W.A., 2002. The Analysis of Queues with Time-Varying Rates for Telecommunication Models. *Telecommunication Systems*, 21(2-4), pp.173–204.
- Melamed, B. & Whitt, W., 1990. On Arrivals That See Time Averages. *Operations Research*, 38(1), pp.156–172.
- Millard-Ball, A., Weinberger, R.R. & Hampshire, R.C. (in press), Evaluating the Impacts of Performance-Based Parking. *Journal of the American Planning Association*.
- Pierce, G. & Shoup, D., 2013. Getting the Prices Right. *Journal of the American Planning Association*, 79(1), pp.67–81.
- Popular Mechanics, 1935. Coin-in-Slot Parking Meter Brings Revenue to City. *Popular Mechanics*, p.519.
- Qian, Z. & Rajagopal, R., 2013. Optimal Parking Pricing in General Networks with Provision of Occupancy Information. *Procedia - Social and Behavioral Sciences*, 80, pp.779–805.
- Rudin, W., 1987. *Real and Complex Analysis*, McGraw Hill.
- Schaller Consulting, 2006. *Curbing Cars: Shopping, Parking and Pedestrian Space in SoHo*, New York.
- Shoup, D.C., 2006. Cruising for parking. *Transport Policy*, 13(6), pp.479–486.
- Shoup, D.C., 2008. *Driven to Excess: What under-priced curbside parking costs the Upper West Side*, New York.
- Shoup, D.C., 2005. *The high cost of free parking*, Washington, DC: Planners Press.
- Simićević, J. et al., 2012. Defining parking price based on users' attitudes. *Transport Policy*, 23, pp.70–78.
- Simićević, J., Vukanović, S. & Milosavljević, N., 2013. The effect of parking charges and time limit to car usage and parking behaviour. *Transport Policy*, 30, pp.125–131.
- Skellam, J.G., 1948. A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials. *Journal of the Royal Statistical Society. Series B (Methodological)*, 10(2), pp.257–261.
- Wolff, R.W., 1982. Poisson Arrivals See Time Averages. *Operations Research*, 30(2), pp.223–231.