Welfare Effects of Shifting Public Transit Demand*

Robert W. Hahn
University of Oxford
Carnegie Mellon University

Robert D. Metcalfe
Columbia University
NBER

Eddy H.F. Tam
King's College London

August 8, 2025

Abstract

Understanding the demand for public transit is important for efficient and equitable transportation policy. We develop novel estimates of peak and off-peak price elasticities for urban mass transit demand in the San Francisco Bay Area using a large natural experiment and a natural field experiment that both have exogenous price subsidies. We then estimate the welfare impacts for these price subsidies to shift demand using a sufficient statistics approach. Our analysis suggests that off-peak subsidies can increase welfare, but the positive effects are reduced when consumers take the decisions of others into account compared to when they do not. We also find a large variation in the welfare impacts of shifting travel to different time periods, which is explained by differences in demand and congestion characteristics. Finally, we show that the targeting of subsidies can increase welfare, but need not do so if the regulator does not have accurate information on demand.

*We would like to thank Ryan Greene-Roesel at Bay Area Rapid Transit (BART) and the BART team for partnering with us to conduct this research and taking a leading role developing the experiments reported in this paper. We also thank the Federal Highways Agency and the San Francisco County Transportation Authority (especially Joe Castiglione and Camille Guiriba) and BART for funding the program. We greatly appreciate the thoughtful comments of Leo Bursztyn, Thomas Chaney, Raj Chetty, Nathan Hendren, Jonathan Libgober, Victor Mylonas, Matthew Tarduno, Shosh Vasserman, Reed Walker, Zi Yang Kang and seminar participants at Harvard University, University of California, Berkeley, and the University of Southern California. We also thank those individuals at Metropia who helped to implement the field experiment, especially Yi-Chang Chiu and Ali Arian. The usual caveat applies.

1 Introduction

Many infrastructure systems—from electricity grids to public transportation—face capacity constraints and steep costs during periods of peak demand. In these settings, shifting consumption to off-peak periods can improve efficiency, reduce congestion, and lower external harms such as pollution (Boiteux, 1960). Time-based pricing is increasingly used as a policy lever to encourage such shifts, but its effectiveness depends critically on how consumers respond to price variation across time.

This paper focuses on urban public mass transit, where peak-load pricing has the potential to substantially improve economic welfare and address equity concerns. Understanding the demand for public transit is critical for designing effective transportation systems and meeting climate policy goals. Transit systems play a central role in reducing road congestion, improving air quality, and lowering greenhouse gas emissions by shifting travelers away from private vehicles (Glaeser, 2007). In urban areas, where transport accounts for a significant share of emissions and local pollution, increasing transit ridership can yield substantial environmental and public health benefits (Anderson, 2014). Moreover, efficient transit systems support equitable access to jobs and services, making them a focal point for inclusive urban planning (Duranton and Turner, 2012; Small, Verhoef and Lindsey, 2024). However, to design pricing and investment policies that achieve these objectives, policymakers require credible estimates of how riders respond to fare changes and service characteristics across time and space. Without this information, attempts to expand or optimize transit systems risk being inefficient or regressive.

We develop a simple model to estimate the welfare impacts of peak-shifting. While we apply the model to peak travel, it can be applied generally to analyze general problems in dynamic pricing. The model allows prices to diverge from marginal cost and incorporates externalities. The demand elasticity estimates for peak and off-peak travel are based on a large natural experiment and a natural field experiment. The experiments were conducted by the Bay Area Rapid Transit (BART) in California.[1] To the best of our knowledge, this is the first analysis to develop demand elasticity estimates for mass public transit when there is experimental variation in prices.

Both experiments allow us to obtain causal estimates of the price elasticities of demand. The first, which includes 17,500 BART riders (7% of BART riders) and over 3.6 million trip sessions, is a natural experiment with a price change in the off-peak period for six months.

[1] BART is a rail and subway system that serves the greater San Francisco area. It is the urban public transit system that Daniel McFadden used to estimate various discrete choice models that contributed to his Nobel prize work (McFadden, 1974; Domencich and McFadden, 1975; McFadden et al., 1977; McFadden, Tye and Train, 1977; McFadden, 2001).

Riders were staggered into the pricing program over the course of six weeks. Prices were reduced an hour before the peak period (630am-730am) and an hour after the peak period (830am-930am). The second is a natural field experiment with randomized prices across 1,900 riders over four months (i.e., 77,000 trip sessions). The experiment incentivized treated customers to shift their riding demand into the least congested train within a 40 minute time window from their usual departure time by randomly lowering prices in the new time window. With detailed consumer-level data on demand and individual characteristics, both of these experiments yield own and cross price elasticities, which are used in the welfare analysis. Across both experiments, we have train-level data on the weight of the train and the time the consumer enters and leaves a station, allowing us to measure congestion. We estimate the welfare impacts of the price subsidies in the experiments using a sufficient statistics approach (Chetty, 2009).

The price reductions in both experiments influenced consumer choice in the expected direction. They reduced travel in the period that did not receive the subsidy, and increased travel in the period that received the subsidy. We estimated an own price elasticity for the period in which consumers received the subsidy and a cross-price elasticity for the period in which they did not. The own and cross price elasticities are around -0.86 and 0.44 respectively in the natural experiment and -0.94 and 0.54 respectively in the field experiment. These estimates suggest some departure time flexibility for BART users. Other non-experimental elasticity estimates for the BART system as a whole are comparable to our own elasticity for off-peak demand (see, e.g., McFadden (1974)). But our own elasticity estimate is a little higher in absolute value than the -0.6 to -0.75 range in the review of estimates in the US by Litman (2004) and Holmgren (2007), and those estimated in Mexico by Davis (2021).[2]

Our two experiments offer different insights into the impact of providing subsidies at particular times for economic welfare. We find that the subsidy in the natural experiment generally increases welfare. In the base case, we show that the marginal value of public funds (MVPF) is 1.6 in trying to shift travel to the off-peak period from the peak period (730am-830am).[3] This means that for each additional dollar of net cost to the government, the marginal benefit to all parties is $1.60. We also calculate the net benefits per dollar of subsidy in the natural experiment, and find that the net benefits are $0.36 per dollar of subsidy expenditure.[4]

[2]We have not found other well-identified *cross-price* elasticity estimates for peak and off-peak in public transit, and thus do not provide a comparison

[3]See Hendren (2020), Finkelstein and Hendren (2020), and Hahn et al. (2024) for reviews of this literature. Our MVPF of 1.6 is quite favorable in comparison to the marginal cost of public funds of around 0.8 to 1 through changes in the linear tax rate (Hendren, 2020) and the MVPFs of other transit subsidies, such as EV subsidies (Hahn et al., 2024).

[4]We measure net benefits as total dollar gains from a discrete subsidy change. The MVPF, by contrast, is a marginal, unitless metric. Both help compare policies.

In the field experiment, we had price variation across more time periods and show that the welfare benefits varied dramatically across routes and time periods due to differences in demand and congestion characteristics. In particular, we show that net benefits per person in the field experiment are sometimes less than zero. We also show that some MVPFs in the field experiment are less than one, which means that the willingness to pay associated with the policy is less than the net cost to the government.[5]

We do several sensitivity analyses to examine the robustness of our main findings. In particular, we examine three sets of behavioral assumptions about consumer response to see how they would change our welfare results. The first relates to how welfare results would change if we scaled up the subsidy to all BART customers, but allowed for some customers to be less responsive to the subsidy than others, perhaps because of selection effects. The second relates to measuring the impact on consumer welfare under the assumption that consumers take the full impact of others' behavior into account in making their transit choices–a case we refer to as endogenous congestion.[6] In both the scaling and endogenous congestion simulations, the effect on per capita welfare is to reduce the net benefits from the subsidy by up to 47%. The third assumption relates to the level and shape of the congestion damage function. We find that for our parameter estimates the shape of the damage function does not appear to have a large effect on our welfare results.

A final set of findings on welfare relates to how subsidies should be targeted. We examined this issue in two ways. First, using our natural experiment, we examined how net benefits per person might change if the most congested route were targeted. We found that targeting users of the most congested route increases average net benefits by 15%. Second, our entire natural field experiment is focused on the issue of targeting. While the field experiment shows that targeting can be effective in shifting demand, the main takeaway is that these shifts may not always increase welfare. A key reason that targeting does not necessarily increase welfare is because the machine learning algorithm used for targeting in the field experiment only takes into account information on crowding. It does not take into account information on unobserved demand characteristics, which are also central for determining the welfare impact of the subsidy. Our experiments suggest that targeting may require a substantial amount of information on economic fundamentals, such as demand, to increase the likelihood that welfare will actually increase with targeting.

Our paper relates to three important research areas. First, our empirical analysis builds on work in sufficient statistics. Our paper is most closely related to Jacobsen et al. (2020).

[5]We also do a calculation on the optimal subsidy and find that the optimal subsidy may be substantially higher than the actual subsidy based on the natural experiment.

[6]This concept could apply to any externality where consumers consider the impact of others' behavior in their decisions.

These authors examine how sufficient statistics can be applied to externality-correcting policies; however, they assume markets are perfectly competitive and that consumers take the externality as given, and do not adjust their behavior in response to the level of the externality. We relax those assumptions in Jacobsen et al. (2020), but retain the assumption that consumers are fully informed.[7]. Our model also relates to work by Kreindler (2024) on estimating endogenous congestion, but our model differs in terms of its focus on urban mass transit and its theoretical underpinnings.

Second, we build on a large literature on congestion and the welfare effects of peak pricing schemes. It has long been argued that many sectors could benefit from peak-load pricing (Vickrey, 1963; Mohring, 1972; Glaister, 1974; Train, 1977; Glaister and Lewis, 1978; Winston, 1985; Jansson, 1993; Parry and Small, 2009). There have been attempts to estimate the welfare implications of changes in peak or off-peak prices for mass transit (*e.g.*, Parry and Small (2009); Joskow and Wolfram (2012); Knittel and Sandler (2013); Winston (2013); Levin, Lewis and Wolak (2017); Kreindler (2024); Tarduno (2022)).[8] Moreover, there appears to be a dearth of well-identified estimates of demand elasticities and cross elasticities for mass transit that have been used to estimate welfare effects. For instance, Parry and Small (2009) estimate the welfare effects of urban transit subsidies using a structural approach, but note that "[l]ittle information is available about shifts of transit riders across time periods" (p. 17). Our empirical analysis helps to address this gap in the literature.[9]

Third, we add to the literature on optimal targeting. Several papers examine how targeting price changes or interventions could increase welfare (Rodrik, 1987; Chassang et al., 2012; Dupas, 2014; Allcott and Taubinsky, 2015; Cohen, Dupas and Schaner, 2015; Alatas et al., 2016; Byrne, Martin and Nah, 2022; Polyakova and Ryan, 2019; Farhi and Gabaix, 2020; Gerarden and Yang, 2023). Our field experiment targeted subsidies using detailed data

[7]We do not address issues of estimating welfare with adverse selection, information asymmetries, market structure issues and behavioral agents. For examples of such work, see Bundorf, Levin and Mahoney (2012); Hendren (2013); Weyl and Fabinger (2013); Mahoney and Weyl (2017); Farhi and Gabaix (2020); Ito, Ida and Tanaka (2023). Several studies that focus on externalities and internalites are also related to our work, including Allcott, Mullainathan and Taubinsky (2014); Piketty, Saez and Stantcheva (2014); Allcott and Taubinsky (2015); Allcott, Lockwood and Taubinsky (2019). These studies fit within the general framework developed by Kleven (2021). See Donaldson (2025) for a recent review of the transportation literature and sufficient statistics.

[8]There have been many papers that analyze the impact of peak-time pricing on overall demand and welfare (Borenstein, 2005; Wolak, 2007; Joskow and Wolfram, 2012; Jessoe and Rapson, 2014; Ito, Ida and Tanaka, 2018; Burkhardt, Gillingham and Kopalle, 2023; Fowlie et al., 2021; Williams, 2022). Our paper builds on this work by developing an analytically tractable approach for estimating welfare changes of peak and off-peak pricing in the presence of externalities, where the externalities can affect demand.

[9]Other papers that estimate changes in subway demand with a change in prices include Davis (2021), who estimates the price elasticity of demand for subways in Mexico. Davis' analysis differs from ours in that he focuses on the impact of general price decreases in subway systems, and not on differences in peak and off-peak travel aimed at reducing an externality. Another paper by Yang and Long Lim (2018), estimates the impact of temporary free rides on early morning subway demand in Singapore, and finds a meaningful shift to earlier trains.

on congestion and historical consumer behavior, which is typical of many machine learning algorithms in transport (Tizghadam et al., 2019) and other markets (Mullainathan and Spiess, 2017; Athey, 2018; Knittel and Stolper, 2025; Burlig et al., 2020; Christensen et al., 2024; Aiken et al., 2022). However, not having information on price elasticities of demand for travel and consumers' willingness to shift their travel times can limit the effectiveness of targeting subsidies using machine learning. For instance, we show in some time periods that targeted pricing using a machine learning algorithm actually lowered welfare.

The paper is structured as follows: Section 2 provides an overview of the theory that allows us to estimate welfare impacts of the experiment. Section 3 develops demand estimates for the natural experiment and section 4 develops demand estimates for the field experiment. Section 5 implements our welfare analysis. Finally, Section 6 concludes.

2 Overview of the Theory

We develop a theoretical model that allows us to estimate the welfare effects of a subsidy using empirically well-identified travel responses as sufficient statistics, without imposing strong structure on the preferences of transit riders.[10] The model identifies key demand parameters of interest that are obtained from our experiments. In addition, we consider how demand is affected when consumers take the behavior of others into account (which we define as the case of "endogenous congestion") and when they do not (which we define as the case of "exogenous congestion"). Finally, we briefly describe how we derive the optimal subsidy.

The basic model has two types of agents, identical consumers and a firm (e.g., a utility). A representative consumer maximizes utility over peak and off-peak travel, and a numeraire good. We assume that the consumer has a well-behaved utility function and quasi-linear preferences. In addition to the consumers, a regulated utility is assumed to minimize the cost of meeting demand. Consumer choices generate an externality that affects the utility, which can be thought of as congestion or delay in our application.

We allow prices to deviate from their marginal private cost (which is consistent with pricing in many regulated markets), and we allow for subsidizing particular consumers. The price subsidy is aimed at encouraging off-peak travel in our first experiment (experiment 1), and shifting particular individuals into less crowded trains in our second experiment (experiment 2). The primary aim of the subsidy is to reduce an externality, which can be thought of as crowding on a subway or delay. In our application, we consider both.

[10] We present the details and derivations of our welfare model in appendix A1 of Hahn, Metcalfe and Tam (2023), and do not include into the manuscript due to space constraints of the journal.

The key results from this model are intuitive. The welfare effect of a subsidy depends on how consumers respond to the subsidy, and the extent to which prices deviate from the marginal social cost. If the price is above the marginal social cost in a particular period, and travel increases as a result of the policy intervention (e.g., a subsidy), then it contributes to an increase in welfare. Summing the welfare changes across each period gives the overall change in welfare.

The equation we use to estimate the per capita welfare effect of an off-peak subsidy is given by:

$$\frac{1}{N}(W(s')-W(0)) = (p-MSC_1)\frac{dx_1}{ds}s' + (p-\frac{1}{2}s'-MSC_2)\frac{dx_2}{ds}s' \qquad (1)$$

where N is the number of consumers; $W(s')$ is the economic welfare of a subsidy, s'; and $W(0)$ is the welfare associated with no subsidy.[11] $\frac{dx_1}{ds}$ and $\frac{dx_2}{ds}$ are the demand responses to the subsidy in the peak and off-peak periods, respectively. These are assumed to be constant. The price of the good is p, assumed to be constant in both periods. The marginal social cost in periods 1 and 2 are MSC_1 and MSC_2, respectively. Marginal social cost is the sum of the marginal private cost (MPC), which is assumed to be constant across periods, and the marginal external cost, MEC_1 and MEC_2, which can vary across periods.

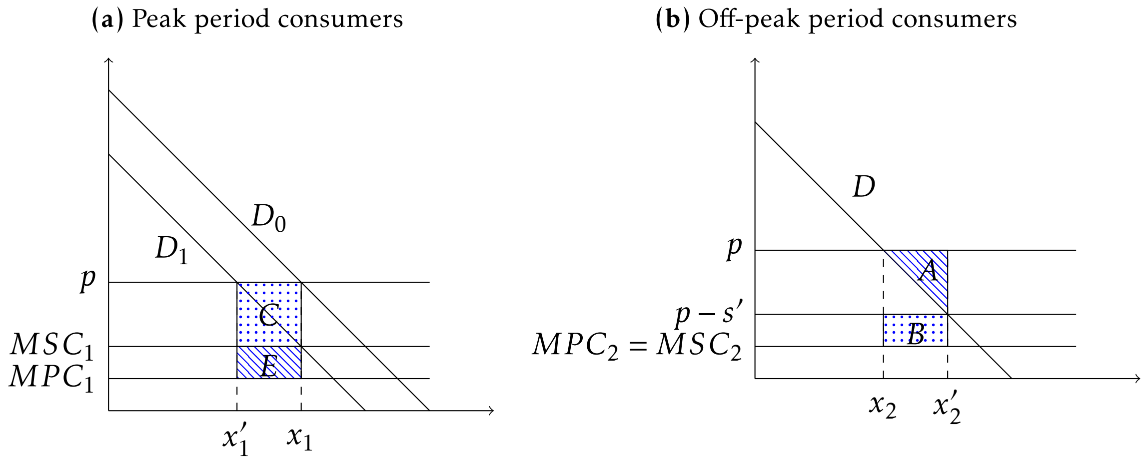
The derivation of equation (1) makes use of the envelope theorem for the consumer maximization problem. When the off-peak subsidy increases, consumers may alter their travel demand because of an equilibrium change in the subsidy, the price or the congestion level, but the resulting changes in consumption do not have direct impact on consumer utility. This is consistent with the sufficient statistics approach (Harberger, 1964; Chetty, 2009; Jacobsen et al., 2020).

The expression for welfare, equation (1), lends itself to a graphical interpretation. Suppose, for simplicity, that demand is linear in the peak and the off-peak market and that p exceeds MSC_1. Figure 1a shows the impact on welfare in the peak period of moving from a subsidy of 0 to a subsidy of s'. The effect of the subsidy is to shift demand from D_0 to D_1, with a reduction in consumption from x_1 to x_1'. The net welfare loss is given by the negative of rectangle C (which corresponds to $(p-MSC_1)\frac{dx_1}{ds}s'$ in equation (1)). It consists of a decline in producer profits (C+E) and a reduction in congestion costs, E. The resulting change in welfare is –(C+E) + E = -C.[12] That is, welfare declines in the peak market with the introduction of the subsidy in this example.

[11] We define the change in welfare as the change in producer surplus plus consumer surplus, assuming the subsidy is a lump sum transfer.

[12] We ignore any direct changes in consumer utility resulting from the subsidy in Figure 1a and Figure 1b below because of the envelope theorem.

Figure 1: Graphical illustration of welfare impact of off-peak subsidy



Notes: In Panel (a), the welfare loss from the subsidy is given by rectangle C in this example. It reflects a combination of the fiscal externality change due to a reduction in producer profits, -(C+E), and a reduction in the congestion externality (E).; In panel (b), the increase in welfare is given by A+B. See text for details.

The graph for the off-peak market is shown in Figure 1b. The off-peak demand is given by D. The subsidy price is reduced from p to $p-s$, with an increase in demand from x_2 to x'_2. Because there is no change in congestion in the off-peak period (it is presumed to be zero), we can focus on the net change in financial flows from the government to compute the change in welfare. The change in profits for producers from the additional units sold $(x'_2 - x_2)$ is given by the rectangle $B + 2A$. This means the government can reduce its expenditures to cover BART's losses by that amount. The change in the fiscal externality as the subsidy is gradually increased from 0 to s' is given by triangle A, which increases the government's cost.[13] The welfare change in the off-peak period is the difference between the increase in producer profits (B+2A) less the change in the fiscal externality (A), giving A+B (which corresponds to $(p-\frac{1}{2}s'-MSC_2)\frac{dx_2}{ds}s'$ in equation (1). Adding this off-peak welfare change (A+B) to the change in the peak period (-C) gives A+B-C, which corresponds to equation (1).

2.1 Modeling the response to congestion: endogenous and exogenous

An important issue for welfare is how consumers respond to the subsidy. We consider the two polar cases of exogenous congestion and endogenous congestion. The two cases can provide insights into the welfare impacts of how interventions scale. The exogenous congestion case may be more appropriate when estimating the welfare impacts of a small-scale subsidy, such as a pilot. The endogenous congestion case may be more appropriate when measuring the welfare impacts of a relatively large-scale intervention, such as providing a subsidy to all

[13]This triangle represents the change in subsidy expenditure resulting from the off-peak consumption change.

customers.[14]

The cases of endogenous congestion and exogenous congestion will typically result in different values for $\frac{dx_1}{ds}$ and $\frac{dx_2}{ds}$ in equation (1), and thus, different empirical measures of the welfare impact of the subsidy. If, for example, a rider takes others' choices to shift to off-peak into account, fewer riders may shift to off-peak for a given subsidy.

To see how the two cases relate formally, define the peak and off-peak demand response of riders to congestion as $\frac{\partial x_1(p,s,e)}{\partial e} = v_{1e}$ and $\frac{\partial x_2(p,s,e)}{\partial e} = v_{2e}$. In the exogenous congestion case, $v_{1e} = v_{2e} = 0$ (i.e., riders would not adjust their demand in response to a change in the level of congestion (holding constant prices p and subsidy s); In the endogenous case, $v_{1e} \neq 0$ or $v_{2e} \neq 0$. We assume that the price effect of the off-peak subsidy is to increase off-peak demand ($\frac{\partial x_2}{\partial s} > 0$) and reduce peak demand ($\frac{\partial x_1}{\partial s} < 0$). If a higher level of congestion reduces peak demand and increases off-peak demand when price is held constant ($v_{1e} < 0, v_{2e} > 0$), the equilibrium change in peak and off-peak travel in response to an off-peak subsidy ($\frac{dx_1}{ds}$) and ($\frac{dx_2}{ds}$) would be smaller in *magnitude* with endogenous congestion than with exogenous congestion (when $v_{1e} = v_{2e} = 0$) (see appendix A.10 for a formal proof).

The intuition is that consumers will be less responsive to the price change when they take into account the actual level of congestion on the network. With endogenous congestion, people not only observe the subsidy and react, but also react to the levels of congestion, which have gone down. They therefore do not switch as much in the endogenous case because the effective price of congestion has decreased (compared to the exogenous case).

2.2 Optimal level of the subsidy.

We derive the optimal level of the off-peak subsidy for the natural experiment. The optimal level of the off-peak subsidy can be found by setting $\frac{dW(s)}{ds} = 0$. This implies the optimal subsidy, s^*, is given by: $s^* = (p - MSC_1)\frac{\frac{dx_1}{ds}}{\frac{dx_2}{ds}} + (p - MSC_2)$ if $(p - MSC_1)\frac{dx_1}{ds} + (p - MSC_2)\frac{dx_2}{ds} > 0$, otherwise $s^* = 0$. This formulation assumes that the envelope theorem holds and the slopes of the demand curves are constant with respect to a change in the subsidy.[15]

3 Natural Experiment

To estimate the welfare impacts of the subsidy, we will need causal estimates of the impact of a subsidy on peak and off-peak travel. This section explains how we derive these estimates for

[14]This issue applies in many settings, as List (2020, 2022) notes.
[15]In principle, one could accommodate changing demands with a more complicated formula.

the natural experiment. We provide a brief overview of the BART network (3.1), a description of the price subsidy and natural experiment (3.2), a discussion of our identification strategy (3.3), and present the key results on peak and off-peak demand (3.4).

3.1 The BART network

The natural experiment was conducted with BART, which is a large public transit system serving the San Francisco Bay area in California.[16] BART is the fifth-busiest heavy rail rapid transit system in the United States. Figure A5 in Appendix B shows the network map.[17] The prices on BART are comparable to those of other U.S. commuter rail systems and are higher than those of most subways, especially for long trips. The minimum price is $1.95 (except for San Mateo County trips) under 6 miles (9.7 km), and the maximum one-way price including all possible surcharges is $15.70, the journey between San Francisco International Airport and Oakland International Airport (2017 prices).[18] Figure A6 in Appendix B plots the distribution of fare by routes. The mean fare for non-airport routes is $4.16, with a median fare of $4.25. The mean and median fare for airport route is $9.60.

3.2 The Perks Treatment

The treatment changed the relative peak to off-peak prices, and BART called this the *Perks program*. To be eligible for the new pricing program, customers had to have a Clipper ID number.[19] We used Clipper card data for our study. The card records the amount paid, the time the person entered the departing station, and the time they left the arrival station. Linking each user with their Clipper card was necessary to give them the pricing subsidy. This subsidy was based on the frequency, timing, and length of their trips.

The Perks program was advertised over a four week period through direct outreach at

[16]The heavy rail elevated and subway system connects San Francisco and Oakland with Alameda, Contra Costa, and San Mateo counties across 112 miles of track connecting 46 stations. BART has an average of 423,000 weekday passengers and 124.2 million annual passengers in fiscal year 2017 (BART, 2017),

[17]BART has five rapid transit lines; most of each line's length is on track shared with other lines. Trains on each line run approximately every 15 minutes on weekdays and 20 minutes during evenings, weekends and holidays; stations on the section of track between Daly City and West Oakland are served by four lines and therefore have 16 trains an hour on each track. BART service begins around 4:00 am on weekdays, 6:00 am on Saturdays, and 8:00 am on Sundays.

[18]The price is based on a formula that takes into account both the length and speed of the trip. A surcharge is added for trips traveling through the Transbay Tube, to Oakland International Airport, to San Francisco International Airport, and/or through San Mateo County, a county that is not a member of the San Francisco Bay Area Rapid Transit District. The farthest possible trip, from Pittsburg/Bay Point to Millbrae, costs less because of the $4 additional charge added to SFO trips and $6 additional charge added to Oakland trips.

[19]Most regular BART customers have a Clipper card, which is a contactless smart card accepted on all major Bay Area public transit agencies, and may be used in lieu of a paper ticket.

stations, media coverage, and employer partnerships. There was some natural time variation in some stations receiving ads before others, and we use this staggered variation later in this section. An example of the ads for direct outreach is shown in Figure A5b in Appendix B. The left hand side of the figure is an example of the basic ad. The right hand side provides a fuller description of the program with eligibility criteria. The program started on August 23, 2016 and ran until the end of February 2017.

The price subsidy changes based on the time the journey starts. For every mile traveled between 630am and 730am or 830am and 930am, the consumer would get a 11.25% reduction in the off-peak price. The peak time at BART was between 730am and 830am. As part of the program to incentivize people to take part, it also gave a modest 2.25% reduction in the peak price. A relative price difference between peak and off-peak price of 9% was thus created by the Perks program. That is the exogenous change in prices that we will leverage to estimate demand. To receive the subsidy, participants needed to have an active PayPal account that used the same email address as their BART Perks account.

3.3 Identification

Identification of $\frac{dx_1}{ds}$ and $\frac{dx_2}{ds}$ comes directly from this natural experiment. We obtain identification of peak shifting due to changes in off-peak prices using two strategies. First, we compare the self-selected Perks users with the rest of the network using a difference-in-differences framework (we have data on all individual rides on BART before and during the natural experiment). We can determine the share of rides that are in the peak and off-peak hours and use this to obtain a causal estimate of the Perks program on demand. We test for parallel trends in peak and off-peak shares before the Perks program began, and find that the peak and off-peak trends for the Perks and non-Perks users are statistically indistinguishable.

Second, we use the fact that not all BART customers were brought onto Perks at the same time. As mentioned above, some stations advertised the subsidy program before others, so we have quasi-random variation in when people saw the advertisements to the Perks program. The program officially started on August 23th, 2016, but 18.9% of the sample were enrolled in the program after one month from the start of the program. This is our ‘late’ sample. Comparing ‘early’ versus ‘late’ enrollers allows us to estimate the impact of Perks on the number of BART trips, as opposed to just the shares of peak and off-peak travel. Again, we find that the trends for early and late enrollers are quite similar prior to the treatment starting (i.e., they are statistically indistinguishable from each other). We estimate the cross-price elasticity using a difference-in-differences approaches to estimate our welfare model.

We have individual level trip data for the 17,545 BART users who signed up for Perks for the six months before the Perks program began (February 2016) and six months during the Perks program (up to February 2017). This is a total of 2,119,588 individual journeys on the BART network for twelve months. We also have data on the rest of the BART network at 15 minute intervals for the same time period which corresponds to 72,098,160 individual rides for non-Perks customers.

Because prices were not perfectly randomized across BART users, a potential issue with our causal estimates could arise if those who signed up for the program knew that they were more likely to reduce their peak demand during the program. While we cannot rule out this possibility, the near perfect parallel trends in both identification strategies suggest that this is unlikely. To address any concerns, we provide an external check on the results obtained here with our field experiment described in section 4, which randomizes prices across BART consumers. Importantly, we will compare the elasticities from the field experiment to those generated from the quasi-experiments considered here. We show that they are similar.

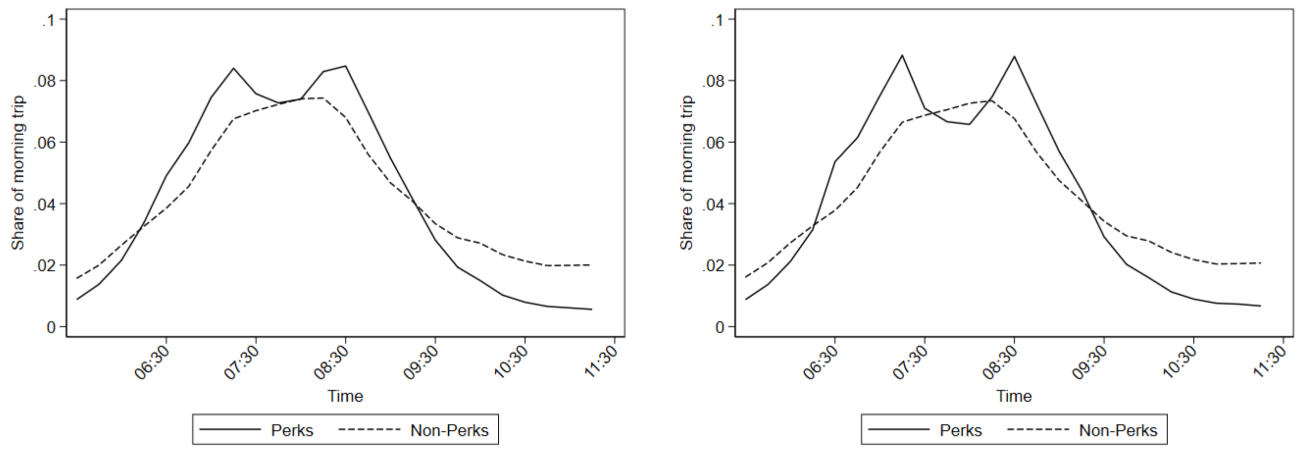
3.4 Results

This results section is based on the two identification strategies presented above. Before we present the results, we describe the data of the Perks users and how it compares to the BART non-Perks users. For the rest of the customers using BART, we know what rides (time, day, stations) have been taken in the same time frame as our opt-in sample, but we can not identify individual accounts.

We first compare the distribution of travel demand for our 17,545 BART users in the Perks program with the demand from non-Perks users. Figure 2 below shows the raw distribution of rides taking place in five minute intervals before and during the Perks program. Panel (a) shows demand before the Perks program and panel (b) shows demand during the Perks program. For both panels (a) and (b), the y-axis is the share of trips. The figure reveals that the overall shape of the dashed line for the non-Perks sample is very similar in panels (a) and (b). In contrast, the line for the Perks sample appears to be steeper and higher for the shoulder hours in the morning (i.e., the subsidized hours). This provides some visual evidence that the increase in demand in the shoulder hours from the Perks commuters during the Perks program does not happen with the rest of the BART network customers during the Perks program. This evidence suggests that once customers received the new price changes, they shifted some of their demand to the shoulder off-peak hours from out the peak hour.

Figure 2: Comparing Perks program participants with the rest of the BART network

(a) Before Perks **(b)** During Perks



Note: The figure plots the average daily share of trips taken by the participants in the first experiment in each time interval from 530am-1130am. It also plots on the same figure the average share of daily trips at each time interval taken by the rest of the BART users. The sample for Panel (a) includes April 1, 2016 to August 22, 2017 for all weekdays. The sample for Panel (b) includes August 23, 2016 to February 28, 2017 for all weekdays.

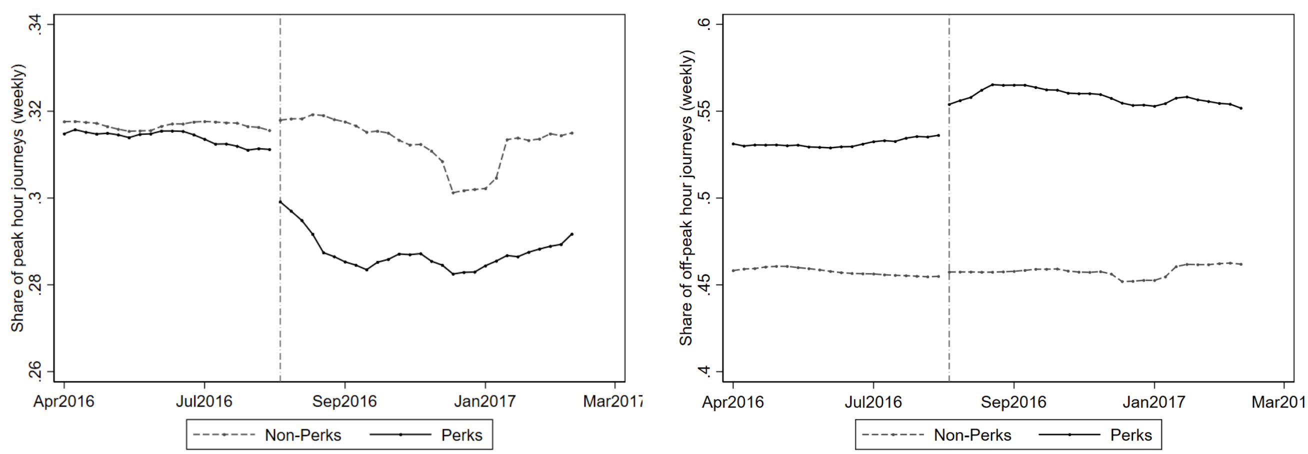
3.4.1 Impact of Prices on Demand: Empirical Approach 1

For our first identification approach in comparing the Perks users to the rest of the BART network users, we must ensure that we have parallel trends in the peak and off-peak shoulder hours for the Perks BART users and the non-Perks BART users before the experiment began. This is the necessary condition for our difference-in-differences approach. Figure 3 presents the daily share of peak hour weekday journeys for the Perks and non-Perks users before and during the experiment. This figure shows that there is a very similar trend in the raw peak hour demand data for the Perks users (black line) and the non-Perks users (the gray line) before the Perks program started on August 23, 2018. The trends for the two groups are not statistically significantly different (see the econometric test of pre-trends in peak and off-peak demand in the Perks and non-Perks samples in Figure A8). This allows us to estimate the impact of the price change on peak demand using a difference-in-differences framework.

It is also clear from Figure 3 that there is a very similar trend in the shoulder off-peak hour demand data between the Perks users (black line) and other users (the gray line) before the Perks program started. The trends for the two groups are not statistically significantly different before July 2016, as suggested in Figure 3 (see the test of pre-trend in Figure A8). This allows us to estimate the impact of the price change on shoulder demand using a difference-in-differences framework.

Figure 3: Share of peak and off-peak shoulder hour demand for Perks users versus the rest of the BART network users (non-Perks)

(a) Peak hour **(b)** Shoulder off-peak hour



Note: Panel (a) shows the five week moving average of the share of peak-hour demand (730am-830am) among all trips from 5.30am-10.30am for Perks and non-Perks users. Panel (b) shows the moving average of the share of shoulder off-peak demand (630am-730am and 830am-930am) among all trips from 5.30am-10.30am for Perks and non-Perks users. The sample includes weekdays from April 2016 to March 2017. The vertical line represents the beginning of the first experiment on August 23, 2016.

Given these two parallel trends in peak and off-peak shoulder hours before the Perks program started for Perks and non-Perks users, we can now estimate the following panel fixed effects model:

$$y_{gt} = \beta_1 * Perks_t + \beta_2 * Treated_g * Perks_t + \beta_3 Treated_g + \gamma_t + \epsilon_{it} \tag{2}$$

where y_{gt} is the share of peak hour (730am - 830am) trips in the morning from 5:30am-10:30am on date t for group g. Treated$_g$ is an indicator for Perks users with the omitted group being non-Perks users. Perks$_t$ is an indicator for the period in which the Perks program is implemented from August 23, 2016 to February 28, 2017. γ_t is a date fixed effect that controls for unobserved changes in travel patterns such as a date-specific service disruption. Our sample period covers April 1, 2016, to February 28, 2017. It covers five months before the Perks program started and a six month period when the Perks program was running. Our null hypothesis is that the difference-in-difference estimator, β_2, is equal to zero for both the peak period and the off-peak shoulder hour.

Table 1 presents the estimates of equation 2. Column (1) reports estimates on the demand share of peak hour. We find that comparing with other users, Perks users reduce their share of peak hour trips by 2.1 percentage points during the program period. Column (2) reports

Table 1: Difference-in-difference estimates - Participants and rest of network

Outcome: Share of all trips in morning		
	All dates	
	Peak Hour	Off-peak Shoulder Hour
	(1)	(2)
Treated users × Perks Period (β_2)	-0.021***	0.026***
	(0.0029)	(0.0023)
Treated users (β_3)	-0.0030***	0.0744***
	(0.0008)	(0.0016)
Observations	466	466

Note: Includes daily observations for Perks users and non-Perks users. Sample includes weekdays from April 1, 2016 to February 28, 2017, excluding public holidays. Standard errors in parentheses are clustered by week.

estimates of the share of off-peak shoulder hours. We find that using non-Perks users as the reference group, Perks users have a 2.6 percentage point increase in the share of off-peak shoulder hour trips during the program period compared to before.

We also estimate an alternative specification that controls for differences between Perks and non-Perks users in terms of the routes they take (defined by the entry and exit station). We compute the weekly share of peak hour (and off-peak shoulder hour) trips for Perks users and for non-Perks users for each individual route in our data. This specification allows for some route has sparse number of rides at daily level. We then estimate equation 2 using the weekly share of peak hour (and off-peak shoulder hour) trips for each route and group as outcome, controlling for route-group fixed effects (i.e., the route-specific travel pattern for Perks and non-Perks users). Table A14 reports the estimates. We find very similar effects as those shown in Table 1.

We next analyze whether the effect of the subsidy applies to the two busiest stations in the BART network, Montgomery Street and Embarcardero. In Table A16 we split the routes into two directions - those that travel from the east of the two stations, thereby crossing the Transbay tube, and those that travel from the west of the two stations. We find that the treatment effect in shifting riders from the peak to the off-peak hour is much larger for the congested routes travelling from the east of the Transbay exiting at Montgomery Street and Embarcardero, compared to those travelling from the west that are less congested.

We also examine the change in demand for peak and off-peak demand when the Perks incentives are removed, which allows us to understand if the incentives created a persistent change in behavior. After six months of the price subsidy program, the program ended. We have BART demand data for those that signed up versus those that did not for six months after the program concluded. We find that while the peak to off-peak shift is lower after the program, there is a persistent and significant treatment effect. We find that the long-term change in peak hour demand without the subsidy was -0.9% (see appendix Table A9). One possible explanation is that people change their demand during the experiment, and once the price changes have been removed, it is costly for users to change their habits or schedules to revert back to their pre-treatment levels. Another explanation is that people may find better ways to commute that increase their welfare, which is supported by evidence from London and Singapore (Larcom, Rauch and Willems, 2017; Yang and Long Lim, 2018).

Despite parallel trends prior to Perks, the customers who selected into Perks may have known prior to Perks that they might want to change their early morning commute time in the future. This is impossible to test with our data but given how strong the parallel trends are, we find this outcome is unlikely. Another caveat here is that we only observe shares of demand and not absolute demand (i.e., number of trips). This is where the next section and identification strategy helps us.

3.4.2 Impact of Prices on Demand: Empirical Approach 2

For the second empirical approach, we estimated how the Perks price program changed demand in the first month in the program, when 18.9% of users had not already signed up (through the staggered roll-out of the advertisement). We define 'early' enrolled Perks users as those who enrolled in Perks from August 23 to September 2, 2016, where our 'late' control group consist of users who eventually enrolled between Oct 1 to Nov 5, 2016. Therefore in our treatment time period (August 23 to September 31, 2016), all the late enrolled Perks users would have not enrolled in the program yet and thus constitute an effective control group for the travel behavior in the absence of the Perks program.[20]

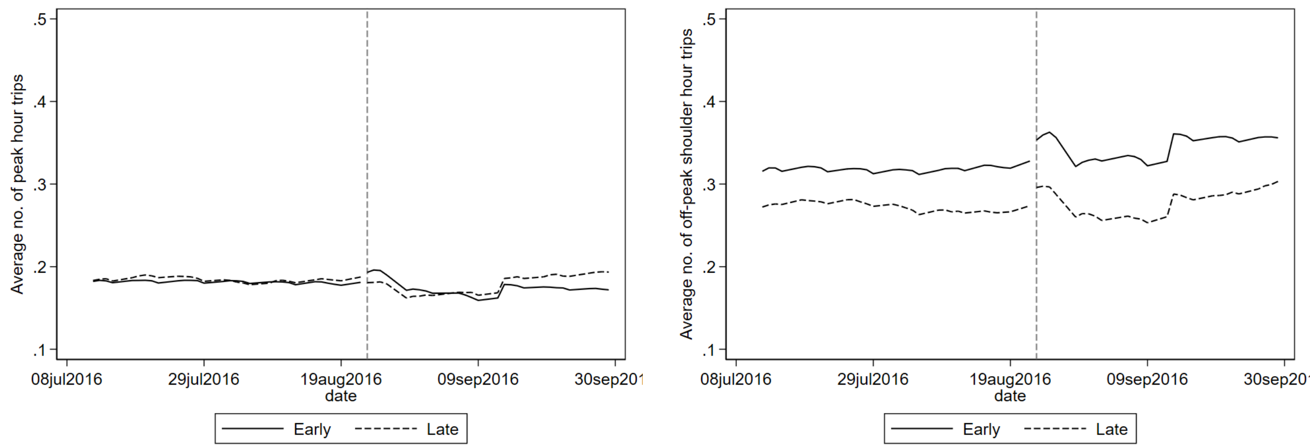
A benefit of using early versus late Perks users is that we can analyze absolute changes in demand at the individual level as a result of the price change. Moreover, both early and late customers all enrolled into the price change, so unobserved characteristics are even less likely to drive the change in peak demand. In addition, the approach of comparing an early treatment group to a late control group is suggested by Callaway and Sant'Anna (2021) as a way to overcome some of the issues in two-way fixed effects models. Figure A4b shows the cumulative distribution for BART enrollment as a function of time.

We can analyze whether the pre-Perks trend before the Perks program is the same for the early and late enrollers. Figure 4 suggests that the trend on peak and shoulder hours before

[20]Our analysis is robust to the time period selected for early versus late. We present in Appendix B Table A15 results that define late enrolled users as those enrolled between Oct 15 to Nov 5.

Figure 4: Early and late enrolled

(a) Peak hour **(b)** Off-peak shoulder hour



Note: The sample includes trips from July 12, 2016 through September 30, 2016. The figure plots the two week moving average for off-peak shoulder hour trips (630am-730am and 830-930am) and peak hour trips (730am-830am) of early enrolled and late enrolled users. The vertical line represents the date August 23, 2016, when the Perks program started.

the Perks program is the same. In Figure A9 we test for the difference of the pre-trend by each of the week, we find that for peak hour trips, the pre-trend for the two groups are not statistically significant. For off-peak shoulder hour trips, we find no systematic difference in the trend for the two groups as well.

We can now estimate the following equation:

$$y_{it} = \beta_1 * E_i * Perks_t + \sigma_i + \mu_t + \epsilon_{it} \tag{3}$$

where y_{it} measures the number of trips during Peak hour on date t for user i. E_i is an indicator for individual i being enrolled early from August 23 to September 2, 2016. $Perks_t$ is an indicator for the perks program period. σ_i is an individual fixed effect and μ_t is a date fixed effect. Our null hypothesis is that the difference-in-difference estimator, β_1, is equal to zero in both the peak hour and the off-peak shoulder hour.

Table 2 presents the estimates for equation 3. Columns (1)-(2) report the estimates on the number of trips per day in the time window of peak hour and off-peak hours respectively. We find that compared with the late enrolled users, during the first month of the program the daily number of trips during the peak hour by early enrollers decreases by 0.007; for the off-peak shoulder hours the daily number of trips by early enrollers increases by 0.015. Column

Table 2: Average treatment effects by time period: Early vs. late comparison

	OLS		
	Outcome: Number of trips daily		
	Peak	Off-peak shoulder (Either)	Total (5.30-10.30am)
	(1)	(2)	(3)
Early enrolled users × Perks Period (β_1)	-0.0071**	0.0151***	0.0040
	(0.0031)	(0.0038)	(0.0044)
Observations	600148	600148	600148
Date FE	x	x	x
User FE	x	x	x
Sample mean (before Perks)	0.18	0.31	0.57

Note: The sample includes daily trips from June 1 to October 1, 2016, excluding August 23 to September 2. The Perks period is defined as dates after September 2. Early enrollers are defined as those who enrolled between August 23 and September 2. The control group includes those who enrolled between October 1 and November 5, 2016. Standard errors are clustered at the user-week level. The sample means are for late enrolled users in periods before the Perks period.

(3) reports the estimates on the total number of trips in the morning.[21] We find that the Perks program does not increase the total number of trips taken by early enrollers.

To convert these numbers to percentages, we need to know the baseline number of trips. The baseline number of trips for the late enrollers during September is 0.178 and 0.273 during the peak and shoulder hours respectively. The DiD estimate of -0.00705 for the peak hour represents a 4% decrease in peak demand. The DiD estimate of 0.0151 for the off-peak shoulder hours represents a 5.5% increase in off-peak shoulder hour demand.

To check whether the treatment effect is robust to different specifications, we relax the specific definition of early versus late enrollment. Instead, we make use of the fact that the activation time into treatment for each user differs. We estimate the change in travel behavior of each user before and after their Perks program activation using the following equation:

$$y_{it} = \eta_1 PerksActive_{it} + \mu_t + \sigma_i + \epsilon_{it} \tag{4}$$

where $PerksActive_{it}$ is an indicator that turns on when user i activated their Perks status on date t. σ_i are user fixed effects and μ_t are time fixed effects. As above, our null hypothesis is that the difference-in-difference estimator, η_1, is equal to zero in both the peak hour and the off-peak shoulder hours.

Table 3 presents the estimates for equation 4 and shows that the effects of the price sub-

[21]The estimates correspond closely to the estimates in share in Table 1. The share of peak (off-peak) trips implied among all trips from 5.30am-10.30am is 0.303 (0.535), with reduction of trips of -0.007 (increase of 0.0151), and the share of peak (off-peak) trip during the program would be 0.284 (0.566), which is an implied change of shares of -0.019 (0.0304).

Table 3: Average treatment effects by time period: staggered activation

	OLS		
	Outcome: Number of daily trips		
	Peak	Off-peak shoulder (Either)	Total (5.30-10.30am)
	(1)	(2)	(3)
Active status level	-0.00713***	0.0244***	0.0130***
	(0.00221)	(0.00389)	(0.00454)
Observations	1426285	1426285	1426285
Date FE	x	x	x
User FE	x	x	x
Sample mean (before Perks)	0.17	0.30	0.56

Note: Sample includes dates from July 1, 2016 through November 6, 2016. Standard errors are clustered by users and dates.

sidies on peak and off-peak demand are very similar to our staggered specification. We find that the after activating their Perks user status, peak hour journeys decreased by 0.007 per day for each user and bonus hour journeys increased by 0.024. The latter result is significantly larger than that from the early versus late enrolment specification. We also estimate the treatment effect following De Chaisemartin and d'Haultfoeuille (2020) and Callaway and Sant'Anna (2021) and report the estimates in Appendix Table A18. Using these approaches does not alter our estimates, providing more confidence in our identification strategy.

3.5 Own and cross-price elasticity of demand for peak and off-peak travel

Given our estimates above, we can now estimate the own and cross-price elasticity of demand for peak and off-peak travel. We estimate $\hat{\eta}$ from equation 4 from estimating the demand effects in our two difference-in-difference specifications. We estimate the causal impact of changing relative prices on the demand for rides taken per day in peak or off-peak periods (we denote η_j as the estimates for the equation where the outcome is travel in period j).

We define the relative price of off-peak, p_2, as the ratio of price per mile of off-peak to that of peak. The estimated marginal price of a mile demanded on BART is \$0.112 in the control condition, where there is no difference between peak or off-peak prices. With no price subsidy, this relative price of off-peak is $p_2 = \frac{0.112}{0.112} = 1$, and it applies to the control group. For the treatment group, the price of one mile on BART is reduced by 9.2%. We calculate the average off-peak shoulder hour trip taken per day by rider before the program as $\bar{x_2}$. Our estimate on own-price elasticity of off-peak hour equals to $\epsilon_{22} = \frac{p_2}{\Delta p_2}\frac{\eta_2}{\bar{x_2}}$. Our estimate on cross-price elasticity of peak hour demand to off-peak hour price equals to $\epsilon_{22} = \frac{p_2}{\Delta p_2}\frac{\eta_1}{\bar{x_1}}$. Our estimates for own-price elasticity of off-peak demand is -0.86, and the cross-price elasticity of peak demand with respect to off-peak price is 0.44.[22] We find that this own-price elasticity does not meaningfully vary by whether we examine consumers who shift to the early off-peak hour as opposed to the later off-peak hour (see appendix Table A17).

We find that the cross-price elasticities of demand change in ways that one would think based on economic theory. For example, the closer the consumer typically travels to the off-peak hour before the subsidy, the greater their own-price response (see appendix Figure A10).

We also find that the more congested the network, the greater the cross-price elasticity (see Table A16 in the appendix). For example, we examine the shift for any route exiting at Montgomery Street and Embarcadero (the two most congested stations of the BART network). Demand from the east (i.e., eastbound Transbay tube trips that are in the busy corridor) towards these two stations shift more (change in absolute share) than that from the west (i.e., westbound transbay tube trips). While all of these heterogeneous effects go in the direction that we expect, they are still correlations and not the causal estimates of changing these parameters on the cross-price elasticities of demand. Nevertheless, we can estimate welfare changes with these different samples in the BART network.

4 Field Experiment

In this section, we estimate the change in peak and off-peak demand from the field experiment. We use these demand estimates to develop a causal estimate of the impact of the subsidy on welfare in the next section. Here, we provide a brief overview of the field experiment and a discussion of our identification strategy, and then present the key results on demand. We then compare the estimates with those obtained in experiment 1.

4.1 Identification

The identification of price elasticities in this field experiment comes from different variation than in the natural experiment. First, we control the assignment mechanism of price changes in this field experiment, whereas we did not control assignment in the natural experiment. The clear advantage of this experiment was the ability to randomize some BART customers

[22]The program offered 5 points per mile of off-peak trip travelled. Each point has monetary value of \$2.52/1000, therefore the subsidy provides a monetary value of \$5*2.52/1000 = 0.0126 per mile in off-peak. The subsidy also gives 1 mile per peak trip travelled. This implies a new relative price of off-peak as $\frac{0.112-0.0126}{0.112-0.00252} = 0.9079$. Therefore, the relative price change is $\frac{\Delta p_2}{p_2} = -0.0921$. From Table 3, the estimate for η_2 is 0.0244. The average off-peak trip per day is $\bar{x}_2$ is 0.307, therefore, $\epsilon_{22} = -0.86$ (s.e. = 0.137). As $\bar{x_1} = 0.176$, and η_1 from Table 3 is -0.00713, $\epsilon_{12} = 0.44$ (s.e. = 0.136).

into the control group where prices did not change and some BART customers into the treatment group where prices were changed.

Second, in contrast to the natural experiment, which focused on generally shifting riders from peak (730am-830am) to off-peak (630am-730am and 830am-930am) travel, this field experiment focuses on shifting riders to less crowded trains in time periods adjacent to their normal commute. For example, a rider may receive a subsidy to shift their travel from their usual time period to an adjacent twenty minute interval, when a less crowded train is available at their departing station.

Third, in comparison with the natural experiment, this field experiment differed in terms of how price changes were determined. Prices were customized based on the individual's previous BART demand and location. Prices were also updated at least monthly based on changes in congestion levels and other factors. In contrast, the natural experiment offered the same incentives to all participants for the duration of the pilot.[23]

Overall, the general size of the incentives, the way they are paid to consumers, and the type and selection of consumers are very similar across both experiments. The price changes started in mid-December 2018 and ended in March 2019. We could not examine any persistence in travel behavior in this experiment as those in the control group received offers in April 2019. Customers registered for the price subsidies via BART's website and the BART Official Mobile Application.

In terms of defining our field experiment, we class it as a hybrid between a natural and a framed field experiment, as defined by Harrison and List (2004). While we have selection of people into the experimental sample who wanted some incentives, people did not know that they were part of an experiment and did not know that they would get price subsidies for changing their commute time.[24] Therefore, we believe that there is little selection on peak and off-peak price elasticities due to the opaque advertising in the field experiment about what incentives would be given.

4.2 Recruitment

Participants were recruited from two sources. First, twelve people employed by BART distributed flyers advertising that people could sign up for the program at the Embarcadero, Montgomery and Civic Center stations in downtown San Francisco from 8:00 – 9:30 AM on December 13th, 2018, These three stations were chosen because they are the most congested

[23]Price subsidies were paid out through gift cards as opposed to PayPal (as in the natural experiment).

[24]The program was advertised as giving monetary payments for completing surveys and traveling on the BART on evening and weekends. See the recruitment flyer in Figure A11.

stations in downtown San Francisco. Second, a subset of those who participated in Perks natural experiment opted in to receiving notifications from BART about future incentive programs. We analyzed the travel histories of these individuals, and identified customers who made at least 50 percent of their BART trips from downtown Embarcadero, Montgomery, or Civic Center stations. An email invitation to join the Perks Phase II program was sent to a random subset of qualifying individuals. Emails were sent in batches between November 20 and December 12, 2018. The overall recruitment led to 1900 customers, of which 63% were recruited through flyers and 36% through email. See appendix Table A19 for the summary statistics for our sample.

4.3 Price changes

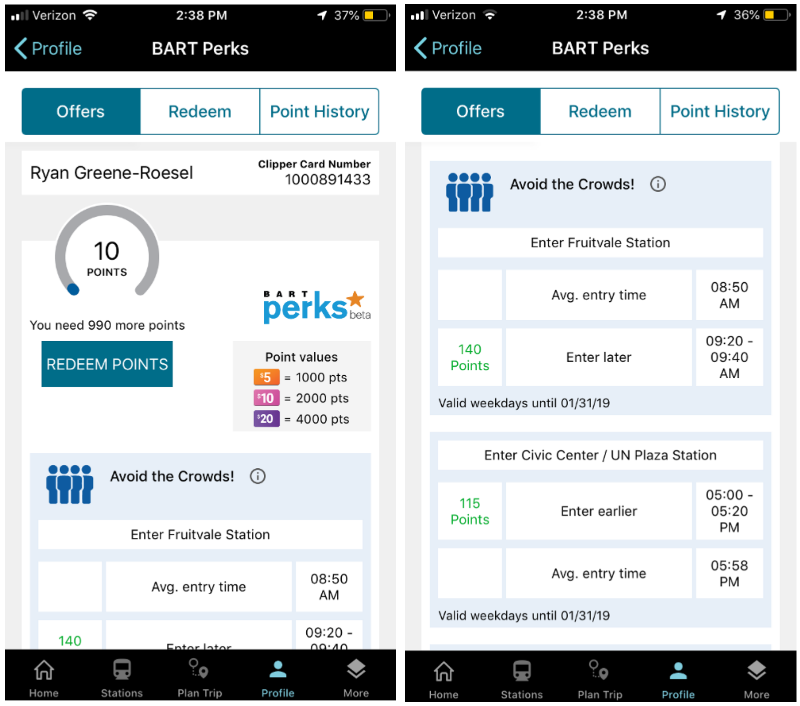
Experimental subjects in the treatment group were offered a price reduction of around 25% per ride in an adjacent time window to their current commute time. The price subsidy had two components: a random term and a deterministic term. The latter was based on how far the time interval was from their usual departure time. The price changes were targeted to each customer's frequently-visited station at a specific 20-minute time window during the morning and/or evening commute period. The time window where the subsidy was offered was up to 40 minutes before or after their typical entry time, and was optimized to reduce overall crowding on the BART network. Users could receive up to four commute-related point offers (shift early AM, shift late AM, shift early PM, shift late PM).

The price subsidies were determined by a machine learning algorithm that drew upon a participant's travel history and a predictive model for crowding provided by the transit company Metropia. Figure 5 provides an example of an offer to shift commuting time. The price subsidies were offered in terms of points that were to be redeemed by the consumer.

To identify the correct time windows to reduce overall BART congestion, the Metropia algorithm identified a user's typical departure time and then calculated the predicted crowding reduction benefit of shifting this user to one of the adjacent time periods. If no benefit would occur from shifting the user, then no offer would be shown. Similarly, no offer was shown if achieving the crowding reduction would require the user to make shifts beyond 40 minutes from their average departure time.[25] The consumer was emailed the price change at the beginning of the month and the price change was constant for one month, so we are estimating one-month elasticities in this field experiment (in comparison with six month elasticities from the natural experiment). During the course of the field experiment, about $23,000 in

[25]This interval was based on feedback from user focus groups, which found that individuals did not wish to make large shifts in their commute.

Figure 5: View of the incentives offer on the app



peak shifting price subsidies were given to BART consumers.

4.4 Results

Our analysis of the field experiment consists of two parts. First, we provide the overall treatment effects of the incentives on shifting ride demand later or earlier. Second, we provide a more precise estimate of the welfare effects of shifting travel into or out of specific 20 minute time periods.

To estimate the average treatment effect of the price subsidies for shifting demand earlier or later than the usual travel time, we employ a linear regression of whether a rider travels at the subsidized time period based on the subsidy they receive as the dependent variable.

For the shift early subsidies, we estimate the following equation:

$$y_{it}^{SE} = \alpha + \beta T_{it}^{SE} + \epsilon_{it} \tag{5}$$

Here, y_{it}^{SE} is an indicator variable for rider i who receives the shift early subsidy and travels at the subsidized time on date t. T_{it}^{SE} is an indicator variable for rider i in the treated group who receives the subsidy (while it is zero for control group riders who receive a hypothetical subsidy). The parameter β captures the overall effect of the shift early subsidies on travel at the subsidized time.

Similarly, we define another indicator variable u_{it}^{SE} to capture whether a rider received a shift early offer but still travels at their usual travel time on date t. We estimate the following equation:

$$u_{it}^{SE} = \gamma + \delta T_{it}^{SE} + \mu_{it} \tag{6}$$

Here, u_{it}^{SE} represents the indicator variable for rider i who received a shift early offer but still travels at their usual time on date t. T_{it}^{SE} is the treatment indicator, indicating whether rider i received a shift early offer on date t. The parameter δ measures the average effect of the subsidy in reducing travel at the usual time for riders.

We follow a similar approach to estimate the effects of the shift late subsidies, using indicator variables y_{it}^{SL} and u_{it}^{SL} to capture riders traveling at the subsidized time and their usual time, respectively.

In Figure 6, we present the likelihood of riders traveling at their usual time or the subsidized time, comparing the treatment and control groups. The results show a clear effect of the subsidies in shifting riders from their usual travel time to the subsidized time. Specifically, for the shift early subsidies, the estimated equation (5) suggests that the subsidies increased the probability of riders traveling at the subsidized time (compared to the control group) by 0.034 (s.e. 0.009), or 3.4%. Equation (6) suggests that the subsidies reduced the probability of riders traveling at their usual time by 0.035 (s.e. 0.016), or 3.5%.

For the subsidies that shift riders to a later time, the analysis indicates an increase in the probability of riders traveling at the subsidized time by 0.022 (s.e. 0.008), or 2.2%, and a reduction in the probability of riders traveling at their usual time by 0.046 (s.e. 0.017), or 4.6%. These estimates demonstrate that the subsidies successfully shifted transit demand from congested time periods to non-congested time periods, resulting in meaningful and significant changes in travel behavior.

Our second part of analysis of the field experiment leverages the detailed part of the experiment, where we have varying off-peak and peak time windows for each consumer, so a flexible specification is required to account for this. We will also turn these estimates into welfare calculations. We estimate the following equation:

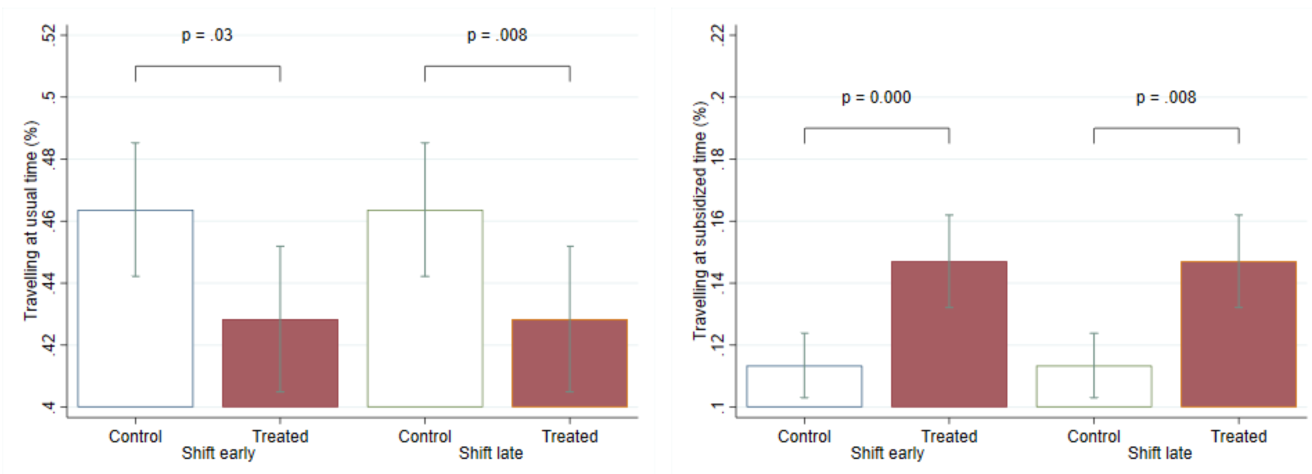
$$y_{it}^{k} = \beta_{k1} S_k + \beta_{k2} S_{-k} + \delta_{kjt} + \delta_{-kjt} + \mu_{ia} A_t + \mu_{id} + \mu_{iq} + \gamma_t + \epsilon_{it} \tag{7}$$

where y_{it}^{k} is the number of trips that start in time period k by individual i on date t. Our sample includes both the treatment and control groups, and both time periods (before and

Figure 6: Field Experiment: Change in demand for subsidized time and usual travel time

(a) Travel at usual more congested time

(b) Travel at subsidized less congested time



Note: The figure presents the share of days riders travelling at their usual time (Panel (a)), and the subsidized time (Panel (b)), by the treatment and control riders, and by the period they receive shift early or shift late subsidies. Each bar represents the share of days the outcome (travelling at the usual time or the subsidized time) takes the value of one for the respective group. The red bar(s) represent the average outcome for treated riders and the white bar(s) represent that outcome for the control riders for the respective groups. The p-values above the pairs of bars indicates the statistical significance for the differences between the treated and control riders in the outcome variable, estimated by the regression equations (the treatment coefficients are 0.0336 (0.00928), 0.0217 (0.0082), -0.0353 (0.0163) and -0.0458 (0.0171) respectively for the shift early usual time, shift late usual time, shift early subsidized time and shift late subsidized time). Standard errors are clustered at the rider level and are in parentheses. The sample includes dates when riders receive a shift early (or late) offer or a hypothetical shift early (or late) offer for the control group riders. The experimental period begins December 2018 and goes through March 2019.

during the experiment). S_k is the monetary value of the points subsidy offered to individual i for time period k, regardless of the "usual travel time" of individual i (e.g., if k is 8am, S_k is any subsidy for 8am, whether it is for someone who usually travels at 7.20am, 7.40am, 8.20am or 8.40am). $S_k = 0$ for the treated group in the pre-treatment period, for the control group, and for those who are not offered a subsidy for time period k. β_{k1} is therefore the own-price demand effect for time period k.

S_{-k} is the value of the subsidy received by individual i for any time period outside of time period k, targeted for someone who usually travels at time k. It is the value of the subsidy offered to individual i who usually travels at time period k that may shift them away from time period k (e.g., if k is 8am, S_{-k} are subsidies for 8.20am/8.40am/7.40am/7.20am that shift people away from 8am). For individuals who do not normally travel in time k, $S_{-k} = 0$ for any subsidy they were offered.[26] β_{k2} is therefore the generic cross-price effect on the demand for time period k when a price subsidy was offered in time periods other than k.

δ_{kjt} is a time-varying fixed effect for the individuals who normally travel in time j and have received a subsidy for time period k (i.e., those who were actually chosen by the Metropia machine learning algorithm to travel on a less congested train within a 40 minute window of their usual travel time). Each type of rider, jk, includes riders from the treated group and the control group. For the treated group it is defined by the price subsidy offered (those who normally travel in time j who received a subsidy for time period k), and for the control group, it is defined using the hypothetical offer that would have been given to the control group if they were in the treatment group (those who normally travel in time j who received a hypothetical subsidy for time period k).[27] The fixed effects therefore control for any time-varying factors that affect the travel pattern of similar types of riders in the treatment and control group. For the control group, $S_k = 0$. Hence, the identification of S_k comes from a comparison of riders of similar "types" as defined by Metropia offers in both the treatment and control groups.[28]

δ_{-kjt} is the time-varying fixed effect for the individuals who normally travel in time k and were offered a subsidy for time period j. Each type of rider, jk, includes both riders from the

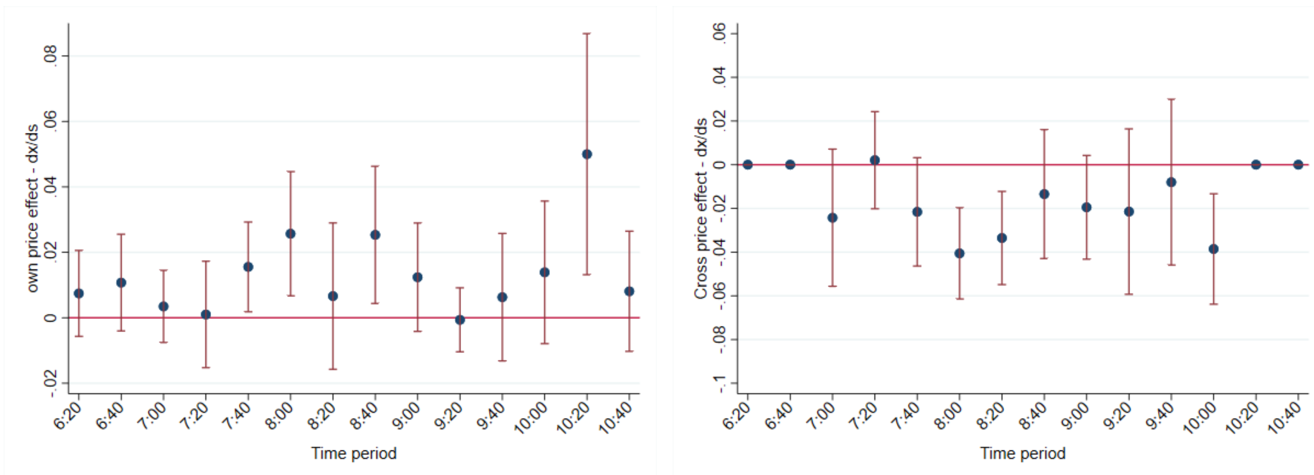
[26] Specifically, let S_m^k be the subsidy that a targeted traveler who usually travels in time period k and subsidize travel in time period m. As in our setup the subsidy shift traveller earlier and later by 20 or 40 minutes interval, if we index the time interval according to the order of time of the day, $S_m^k = 0$ if $|k - m| > 2$. S_{-k} is defined as $S_{-k} = S_{k-2}^k + S_{k-1}^k + S_{k+1}^k + +S_{k+2}^k$. As each traveller has a specific usual travel time k, and receive subsidy for only one time period other than k at each period t, it is equivalent to $S_{-k} = \max\{S_{k-2}, S_{k-1}, S_{k+1}, S_{k+2}\}$.

[27] The algorithm created hypothetical offers for all consumers, but we only gave the offers to the treated group and the control group was unaware of such offers.

[28] Equivalently, these are the fixed effect coefficients for a time invariant indicator D_{kj} interacted with time dummies, where the indicator D_{kj} is 1 if the rider i in the treatment group has ever received subsidy of type S_{kj} in the experiment period, or if the rider i from the control group has ever received a hypothetical subsidy of type S_{kj}.

Figure 7: Field Experiment: Demand Parameter Estimates

(**a**) Change in subsidized time (**b**) Change in usual time



Note: The figure plots the estimates of equation 7. The sample include participants in the field experiment. Panel (a) plots the estimate for the effect of a small change in the subsidy at time t on the demand at time t. Panel (b) plots the estimate for the effect of a small change in subsidy for time t, where the usual travel time is time j, on the demand at time j. Standard errors are robust to autocorrelation of 14 days. The average journey price in the control group is \$3.99.

treated and control groups. For the control group, $S_{-k} = 0$. Hence, our identification for S_{-k} comes from a comparison of riders of similar type in the treatment and control groups that were identified as having similar offers.

μ_{id} is a fixed effect for individual-day of the week. It controls for individual travel pattern on each day of the week. μ_{iq} is an individual*year quarter fixed effect, which allows for an individual's travel pattern to change in each quarter. μ_{ia} is an individual fixed effect that takes a value of one for the weeks beginning November 12 and 19, 2018, in which the air quality in San Francisco was a historically poor level due to wildfires. It controls for any idiosyncratic response of each individual during those two weeks. γ_t is the date fixed effect. Our null hypotheses are that the price subsidy in the targeted time period has a zero impact on demand and that the cross-price effect on demand is zero.

Figure 7 plots the demand estimates, β_{k1}, for each of the time periods in equation 7. Panel (a) plots the demand change in the time period that subsidies were offered. We find, in general, that the estimates for βs are positive, meaningful, and statistically significant, suggesting riders are responsive to price at relatively narrow time intervals (as compared to the wider window in experiment 1). The time periods with the greatest demand shifts toward that period are 740am-8am, 8am-820am, 840am-9am, and 1020am-1040am. The average own-price elasticity of demand from peak to off-peak is -.939 . However, there are some time periods where people are less elastic in their demand, such as before 740am and after 9am (excluding the 1020am-1040am time period).

Panel (b) plots β_{k2} of equation 7, which is the change in the usual travel time for the consumer. In general, we find negative and statistically significant coefficients, suggesting that riders were reducing travel in their usual time and shifted into the subsidized time period (panel (a)). The average cross-price elasticity of demand from peak to off-peak is .543 .[29] It seems that consumers are more elastic between 7am and 10am, but less elastic before 7am and after 10am.

From both panels in Figure 7, it is clear that elasticities vary across time periods. This variation in elasticities results from a combination of individual preferences and BART system effects. For example, the own-price and cross-price elasticity at 8-820am are -1.45 and 0.30 respectively.

For those individuals who did not switch at any time period, we sent them a survey asking them why they did not switch earlier or later.[30] We found that for those who would not switch earlier, the most chosen reasons were personal preference (55%) and employer would not allow it (29%), and that for those who would not switch later, the most chosen reasons were also personal preference (41%) and employer would not allow it (40%). This suggests possible unobserved reasons for not switching demand to an earlier or later time period.

The previous analysis allows us to compare the elasticities from the field experiment to those from the natural experiment. The own-price elasticity of demand and cross-price elasticity of demand from the natural experiment were -0.86 (se = .14) and 0.44 (se = .14) respectively. Given the respective estimates from the field experiment are -0.94 and 0.54, we cannot reject that the estimates from both experiments are equal for both the own-price elasticity and the cross price elasticity.

[29] The cross-price elasticity for each type of treatment subsidy (a subsidy for those who normally travel in period j' receiving a subsidy for time period k') is calculated as $\epsilon_{j'k'} = \frac{dx_{j'}}{dS'_{-j}} * \frac{p}{x'_j}$ where $x_{j'}$ is the average travel (in number of trips) for those who normally travel in time period j' who received subsidy in time period k', $\frac{dx_{j'}}{dS_{-j'}}$ is the estimate of the coefficient for S_{-k} in equation (7) for the time period $k = j'$, and $p = 3.99$. The average elasticity is arithmetic average of the elasticities $\epsilon_{j'k'}$all the treatment subsidies (include both shifting early and late).

[30] Survey implementation and results can be found at BART Perks Phase II Evaluation Report (2019).

5 Welfare analysis

In this section we use the theory presented in section 2 and estimates of demand from the natural and field experiments to derive the change in welfare. We also present an alternative welfare calculation using the marginal value of public funds (MVPF). We present an overview of the key data used in the calculation in section 5.1. We then present welfare estimates for the natural experiment in section 5.2. We extend this analysis to allow for differences in demand across groups (in this case, customers that received or did not receive a subsidy) in section 5.2.2. Finally we estimate the MVPFs for the natural experiment in section 5.3 and the field experiment in section 5.4.

The welfare analysis yields several key findings. First, providing an off-peak BART subsidy improves welfare under both our model and the MVPF framework (Finkelstein and Hendren, 2020). In the base case, the average net benefit per subsidy dollar is \$0.36, and the MVPF is 1.6—meaning each additional government dollar yields \$1.60 in marginal benefit.[31] Welfare gains stem from increased ridership and a shift from peak to off-peak use, easing congestion. Second, accounting for riders' responses to congestion (*i.e.*, endogenous congestion) raises welfare, though less than under exogenous congestion. This is because fewer peak-period trips (smaller welfare loss) and more off-peak trips (larger gain) occur with exogenous congestion.

Third, targeting congested routes boosts per capita welfare. Gains rise by 15% when subsidies are focused on the highly congested Transbay route. Fourth, the optimal subsidy may be significantly larger—up to eight times the experimental level—yielding up to four times the welfare. These results depend on assumptions about demand's functional form. Finally, narrow time-based targeting of subsidies can raise welfare, but not always. Our experiments indicate such strategies require detailed demand information to be effective.

5.1 Natural Experiment: Data

We analyze the welfare effect of experiment 1 using equation (1) in section 2. The key parameters used in equation (1) in the "base case" are presented in Table 4, and discussed in more detail in appendix Table A11.

Table 4 shows that the assumption about the response to congestion matters for the demand response to the subsidy in both the peak and off-peak periods. Even though peak travel declines with the introduction of the subsidy in both cases, the decline is about 15% less with endogenous congestion than with exogenous congestion. More people travel during the peak

[31] All MVPFs are computed assuming a zero initial subsidy.

Table 4: Key parameters used in our base case welfare analysis

Parameter	Exogenous congestion	Endogenous congestion	Percent Change
$\frac{dx_1}{ds}$	-0.0194	-0.0166	-14.68
$\frac{dx_2}{ds}$	0.0664	0.0625	-5.91
p	\$3.99	\$3.99	-
s'	\$0.25	\$0.25	-
MPC	\$1.89	\$1.89	-
MEC_2	0	0	-
MSC_1	\$3.22	\$3.22	-
MSC_2	\$1.89	\$1.89	-

Notes: Demand decreases in the peak period and increases in the off-peak period with the introduction of the subsidy. The magnitude of the decrease is smaller with endogenous congestion in the peak period, and the magnitude of the increase is smaller with endogenous congestion in the off-peak period. See discussion in the text for sources of the parameter estimates. Demand is measured as the average number of trips taken per weekday for the average rider.

period when they take into account how others will respond to congestion. Similarly, fewer people travel during the off-peak period with endogenous congestion (see the discussion in section 2 and Appendix A.3 for details of the empirical estimation). In both the exogenous case and the endogenous case, the subsidy increases overall ridership. The change in overall ridership, measured by $\frac{dx_1}{ds} + \frac{dx_2}{ds}$, is 0.047 trips per day in the exogenous case and 0.046 trips per day in the endogenous case. Thus, the endogenous case has a two percent lower increase in overall ridership than the exogenous case.

5.2 Natural Experiment: Welfare estimates

5.2.1 Estimates for the Experimental Sample

We consider the welfare impacts of two subsidies in detail: one is a general subsidy for all riders in the experiment 1, and the second is a targeted subsidy for riders on what we call the Transbay route. We define the Transbay route to include trains that lead to the Transbay Corridor[32], and also trains that travel on the Transbay route westbound. This route is in the top 10 percent of congested routes for BART in the morning.[33] We conclude by briefly considering the welfare effects of an optimal subsidy.

[32]See "Transbay Corridor Core Capacity Program Train Control Modernization Project", BART.

[33]This includes routes that depart or arrive at stations: West Oakland, Embarcadero, Bay Fair, San Leandro, Coliseum, Fruitvale, Lake Merritt, Orinda, Rockridge, MacArthur, 19th St/Oakland, 12th St/Oakland City Centre, and Ashby, towards the Transbay tube direction.

Table 5 presents the main welfare results for the base case subsidy and a targeted subsidy for the Transbay route. First, we consider the base case. We measure welfare in dollars per person per weekday. The overall welfare effect for the base case is the combination of the welfare loss associated with reduced peak travel and the welfare gain associated with more off-peak travel. For both the exogenous and the endogenous congestion cases, this net gain is positive. A subsidy of $0.25 per trip in the off-peak period results in a net welfare gain of about $0.03 per rider per day for the exogenous congestion case (row 3 of Table 5).

Table 5: Welfare impacts for the base case and a targeted subsidy

	Parameter	Base case			Targeted subsidy		
		Exogenous congestion	Endogenous congestion	Percent Change	Exogenous congestion	Endogenous congestion	Percent Change
(1)	Welfare change associated with peak travel change	-0.0037	-0.0032	-14.7	0.0006	0.0005	-14.7
(2)	Welfare change associated with off-peak travel change	0.0328	0.0309	-5.9	0.0328	0.0309	-5.9
(3)	Total Welfare change	0.0291	0.0277	-4.8	0.0334	0.0314	-6.1

Notes: Welfare is measured in dollars per person per weekday. Overall welfare increases with the subsidy. The decline in ridership in the peak period reduces welfare, and the increase in ridership in the off-peak period increases welfare. See text for details. The total welfare change is calculated from equation (1).

Welfare in the endogenous case declines by 4.8% compared with the exogenous case because of differences in demand response (see row 3 of the table). This percentage change masks more significant differences in actual travel patterns in the peak and off-peak periods (see the percentage change in row 1 and row 2 for the base case in Table 5).[34]

The welfare analysis for the Transbay route is similar to the base case qualitatively. The main point of this analysis is to illustrate that there could be potential gains from targeting. The average net benefits are about 15% higher when targeting Transbay riders under the assumption of exogenous congestion.[35]

Average net benefits per dollar of direct subsidy, which sometimes is used as a measure of welfare by decision makers, is also higher.[36] We estimate average net benefits by taking

[34]As a sensitivity check, we calculated welfare using a discrete choice model based on a conditional logit (see Appendix A.12). Consumers choose between peak, off-peak, or no BART travel. This illustrative model shows welfare declines but stays positive. The decline partly reflects that the logit model implies little net change in ridership, unlike the OLS model, which shows a large increase. We prefer the OLS specification because it avoids assumptions about alternative travel modes for which we lack data.

[35]Take the average welfare gain of .0334 for Transbay and divide by .0291 for the base case. Note that this percentage is higher in the endogenous case.

[36]The measure is similar to that used in the tax literature that estimates the welfare change per dollar of tax revenue. See, *e.g.*, Saez, Slemrod and Giertz (2012).

the total welfare change and dividing by the average subsidy expenditure per person per weekday.[37] For the base case, we find that an essential difference between the analysis of the Transbay route and our analysis of the base case is the level of congestion, which is higher. This, in turn, increases the MSC associated with the Transbay route during the peak period. The average peak period congestion on the Transbay route is 108 passengers per car, while it is 57 passengers per car for all other routes.[38] We also find that delay is more severe on the Transbay route than other routes. Using equation (1), we obtain a welfare estimate of providing a subsidy that exclusively focuses on the Transbay route during the morning (see appendix A.4 for details of the calculation). For the Transbay route, the marginal social cost per peak period trip is $4.12. This is higher than the marginal social cost when we use the average level of congestion across all routes ($3.22). The welfare effect of a $0.25 subsidy for an off-peak trip is $0.033 per person per day in the case of exogenous congestion, and $0.031 in the case of endogenous congestion. These welfare estimates are greater than the welfare effects estimated in the base case. The difference is due to the higher marginal social cost in the peak period for the Transbay route.

With additional assumptions about the functional form of demand, we can compute the optimal subsidy using equation (2.2).[39] We provide details of this computation in appendix A10. Assuming linear demand, we find that the optimal subsidy is about eight times higher than the actual subsidy ($1.87 in the exogenous case and $1.90 in the endogenous case versus an actual subsidy of $0.25 per trip). Total welfare is about the same for the exogenous and endogenous cases, but is about four times higher than welfare associated with the actual subsidy ($.12 vs. $.03). The reason that total welfare is similar in the endogenous and exogenous cases is because the optimal subsidy is slightly higher in the endogenous case, but the overall demand increase ($\frac{dx_1}{ds} + \frac{dx_2}{ds}$) is slightly lower.

We also considered what would happen with changes in the shape of the MEC function. In particular, we allow for the MEC function for peak period travel to be an increasing linear and non-linear function of density, e, instead of a constant function of e. Our main conclusion is that in this particular example, the results do no change much in moving from a constant MEC function to a linear MEC function or in moving from a constant MEC function to a quadratic MEC function for the peak period. We discuss our methodology and present the empirical results in appendix A.7.

The preceding analysis of welfare changes in Table 5 does not explicitly consider how

[37]See appendix A.5 for details.

[38]The units of the congestion measure used in our calculation is passengers per square meter in an average car. See appendix Table A11 for details.

[39]We analyze the optimal off-peak subsidy as a second-best solution, since prices deviate from marginal social cost in both markets. In principle, we could compute optimal subsidies for both, but lack data on peak-period demand response to price.

changes in other modes of travel could affect our calculation. In Appendix A.5, we present a calculation of how intermodal substitution could affect the welfare from the subsidy. In general, we argue that welfare would improve because of decreases in emissions and reduced congestion that are not priced in the market, supporting previous research by Anderson (2014). These pollution reduction benefits may be compared with the benefits of the subsidy from the experiment, which are about $2 million per year. We derive this estimate by multiplying the average benefit per person per weekday by the number of weekdays per year by the number of BART riders. See Appendix A.5 for details of the calculation.[40]

5.2.2 Estimates for the BART Network

We extend our theoretical framework to allow us to estimate the welfare benefits when riders may differ in their responsiveness to a subsidy. Formally, we model two types of riders with different demand characteristics. We allow for differences in demand characteristics because people participating in the experiment may, for example, be more sensitive to price changes on average than those who do not. In appendix A.11, we show that this can lead to a straightforward modification of equation 1, where the average per capita welfare change for all riders depends on the demand characteristics of both groups and their respective sizes. For example, when demand characteristics are the same for the two groups, average per capita welfare benefits from the subsidy are the same.

To empirically implement this framework, we first explored how the sample in the natural experiment differed from the general BART population. As discussed below, the sample does not appear to be that different. Nonetheless, because there may be unobserved differences between populations, we present a sensitivity analysis.

A key question is how participants in the natural experiment may differ from the general population that uses BART. We relied on a survey administered by BART to address this question. Our main finding is that, based on the survey, the observable characteristics of our sample and the BART customer sample are no different. For instance, the percent of individuals who have household income less than $50,000 in our natural experiment sample was 13%, versus 14% for all those using BART who commute downtown. The percent of individuals who are BIPOC in the natural experiment sample was 57%, versus 64% for all those using BART who commute downtown.

However, there may be unobserved demand differences between those who signed up to the natural experiment and the rest of the BART network. To address this issue, we compared

[40] We do not analyze how the introduction of the BART network affects pollution. Previous research suggests this could be important (Gendron-Carrier et al., 2022).

our base case results with a case in which the rest of the BART network is assumed to be one-half as responsive to the price subsidy as participant in the experiment. In the base case, recall that average net benefit per dollar of direct subsidy was $0.36. If the demand by those not in the experiment is one-half that of people in the experiment, this measure drops by about 46% to $0.2. The reason for the large drop is that other people not in the experiment represent about 93.5 percent of the entire network, and thus are the major factor driving the per capita welfare calculation. This analysis highlights the need to understand the demand characteristics of different groups of riders to develop a more precise estimate of the welfare impacts of a large-scale subsidy.

5.3 Natural Experiment: MVPF estimates

The preceding section used a measure of welfare based on our welfare model. It assumes that the method of financing the subsidy is a lump sum tax. In contrast, the MVPF calculation for the subsidy is silent on the source of the revenues that fund the intervention. The MVPF benefits measure is given in terms of the change in net benefits per net additional dollar of government cost. It is, thus, not a welfare measure in the sense of providing an estimate of net benefits (*i.e.*, benefit minus costs). However, if we assume that the net costs to the government are funded by a lump sum tax, it is possible to show that the MVPF measure and the measure for welfare we obtain in our basic model are related.[41] The intuition is that a lump sum tax converts the net cost to the government into a cost to society.

Equation 8 describes our MVPF calculation.

$$MVPF = \frac{\text{Inframarginal benefit from the subsidy + Change in congestion benefits}}{\text{Change in direct subsidy cost + Change in the transfer to BART}} \tag{8}$$

In our application, the benefit to BART users is the numerator and the net cost to the government of the subsidy is the denominator. The numerator consists of two terms: inframarginal benefits of the subsidy for those off-peak users who did not switch from the peak, and the benefits to all peak travelers from reduced congestion.[42] The denominator also consists of two terms: the direct subsidy costs that are paid to BART consumers, and the change in the transfer to BART resulting from the change in ridership.

The two terms in the numerator are positive. The subsidy provides benefits to those who do not change their usage patterns and there are benefits (in terms of reduced congestion) for

[41] See appendix A.9 for details.

[42] We assume the net benefits to those who shift from peak to off-peak travel as a result of the subsidy are negligible or zero because of the envelope theorem, and do not include this term in net benefits. If we included some part of these benefits, it would have the effect of increasing the MVPF.

those who continue to travel during the peak period. The two terms in the denominator have different signs. The direct subsidy cost to the government increases as a result of introducing the subsidy; however, in our particular example, discussed in more detail below, BART's revenues profits actually increase as a result of the subsidy, which means that the transfer from the government to BART, which is needed to cover BART's losses actually decreases. Thus, the second expression in the numerator is negative. This situation arises because ridership increases as a result of the intervention, and price exceeds marginal cost for each ride. BART's profits thus increase from the increased ridership, and the transfer needed by the government to keep BART's profits at a particular level decreases.

Table 6 shows the MVPF calculation for the cases of exogenous and endogenous congestion. In both cases the MVPF is close to 1.6. This means that for each additional dollar of net cost to the government, the marginal benefit to all parties is $1.60. This is quite favorable in comparison to EV or hybrid subsidies (Hahn et al., 2024). The 95% confidence interval for the MVPF, obtained by bootstrapping both the peak and off-peak demand response estimations, ranges from about 1.4 to 1.7 for both cases.[43]

Table 6: MVPF's for the subsidy

	Exogenous Congestion	Endogenous Congestion
Numerator		
1. Inframarginal benefits	0.307	0.307
2. Congestion benefits	0.0258	0.0220
Numerator total	0.3328	0.329
Denominator		
1. Change in direct subsidy cost	0.307	0.307
2. "Change in transfer to BART"	-0.0987	-0.0965
Denominator Total	0.2083	0.211
MVPF = Numerator/Denominator and error bounds	1.598 [1.479, 1.716]	1.563 [1.444, 1.681]

Notes: The MVPF is computed for a subsidy of $1 per trip in the off-peak period using equation (8). The VOT is set equal to 50% of the wage rate. The 95% confidence interval is calculated by bootstrapping the estimation for the demand responses with 100 replications. Details of the calculation are provided in appendix Table A12 .

There are two general points to be made about the MVPF approach based on our example.

[43]If we assume riders not in the experiment were half as sensitive to the subsidy as those in the experiment then the MVPF drops from 1.6 to 1.26.

First, the subsidy in this case involves more than simply the direct subsidy to passengers. Because we are dealing with a regulated entity, one must also consider the impact of the subsidy on the revenues needed to support the regulated entity – in this case, BART. Second, more work needs to be done in this area to rank various subsidies and pricing interventions using the MVPF framework in the regulatory domain, and connecting them to tax policies that might be used to assess their effectiveness.

5.4 Field Experiment: MVPF Estimates

There are two main results. First, targeting results in many MVPFs that exceed one (meaning the increase in willingness to pay exceeds the net cost to the government), but in some that do not. This is because targeting parameters are set without reference to changes in demand, and these changes are critical for welfare. Second, we explain the subtle relationship between a change in the value of time and a change in welfare.

In experiment 2, thirty groups receive a subsidy in 12 time periods that are 20 minutes long. The subsidies are to encourage travel that is 20 minutes or 40 minutes away from the rider's usual travel period of 20 minutes. A group is defined by their subsidized time period and their usual period of travel. Because of data limitations, we assume that the targeted subsidy for a specific group only affects travel in the targeted period and the usual travel time period.

We identify each targeted group, j, by the usual travel interval and the subsidized interval, which can be 20 or 40 minutes before or after the usual travel time. For example, one group is riders who usually travel from 8:20 am to 8:40 am and receive a subsidy to travel between 8.40 am and 9 am.

We focus on evaluating the welfare impact of the subsidies for experiment 2 using the MVPF. We consider the case in which the riders take congestion as exogenous in their decision making. In calculating the MVPF for subsidies in experiment 2, we allow for the possibility that the marginal external cost is positive in the targeted interval. We calculate the MVPF for each subsidy using the following formula:

$$MVPF_{jt} = \frac{x_{jt} - MEC_{j1} * \frac{dx_{j1}}{ds_{jt}} - MEC_{jt} * \frac{dx_{jt}}{ds_{jt}}}{x_{jt} - (p - MPC) * (\frac{dx_{j1}}{ds_{jt}} + \frac{dx_{jt}}{ds_{jt}})} \tag{9}$$

$MVPF_{jt}$ is the MVPF of the subsidy given to group j for the targeted time interval t; p is the price for a trip; MEC_{j1} is the marginal external cost for the route taken by group j on the usual time before the subsidy, and MEC_{jt} is the marginal external cost for the same route in the time targeted by the subsidy. x_{jt} is the average number of trips taken by group j in the targeted travel time. The numerator consists of two parts: the inframarginal benefits for those who receive the subsidy in time t (the first term) plus the change in external congestion costs (the last two terms). The denominator consists of two parts as well: the direct costs to the government of the subsidy for customers (the first term) plus the change in profits to BART (the second term).

We measure the congestion level using the passenger density on the route travelled by group j, at the usual travel time and also at the subsidized time. We also consider the travel delay in estimating the marginal external cost in each time period. This allows us to calculate MEC_{j1} and MEC_{jt}.

To calculate the marginal external cost, we need a measure of the value of time (see equation (A1)). Using survey data from the sample, the median wage for a single person household is $50 dollar per hour. This wage estimate is higher than the median wage used to evaluate experiment 1, which was based on all San Francisco residents. We consider a value of time equal to 50% of the wage, as we did before, and also one equal to 75% of the wage. This allows us to calculate the marginal external and social cost using equation (A1), as we did for experiment 1.

The impact of a change in the value of time cannot be signed apriori. A change in the value of time enters into the analysis through its impact on the MEC in both periods. If the change in the value of time entered into the calculation in only one period, the comparative statics would be straightforward. However, when we allow for a change in the MEC in both periods, the comparative statics depends on the MEC *and* demand parameters in both periods, and this effect cannot be signed a priori. We find that in 75.7 percent of cases, an increase in value of time from 50 to 75 percent increases welfare, and in 24.3 percent of cases an increase in value of time decreases the welfare effect.[44]

A comparison of the MVPFs from the two experiments suggests that the average MVPF from the field experiment is lower than the MVPF obtained in the natural experiment. This may at first seem unusual because one might expect net benefits to be higher in the case of targeting (experiment 2) than with a general off-peak subsidy (experiment 1). However, there are several possible reasons for this finding. First, the MVPF provides a measure of marginal and not total net benefits. Second, the two experiments are measuring different things. The

[44]The precise formula for measuring the impact of VOT on MVPF is obtained by differentiating equation (9) with respect to value of time. It is given by $-(\frac{dMEC_{j1}}{dVOT}\frac{dx_{j1}}{ds_{jt}} + \frac{dMEC_{jt}}{dVOT}\frac{dx_{jt}}{ds_{jt}})MVPF_{jt}$, where VOT represent value of time, and $\frac{dMEC_{j1}}{dVOT}$ and $\frac{dMEC_{jt}}{dVOT}$ are the change in each of the MEC with respect to VOT, that depends on the crowding level during the time period, and how the delay relates to the number of passengers. See appendix A.1 for the derivation and further discussion of the empirical results.

natural experiment measures the marginal benefit of moving people from peak to off-peak, and the field experiment measures the marginal benefit of moving from one relatively narrow time interval to an adjacent time interval, which is not necessarily off-peak. Finally, the estimate of congestion used in the field experiment varies across all of the relatively small time intervals, whereas we assume for simplicity in our natural experiment that there is one average level of congestion that applies to the peak period, and there is no congestion in the off-peak period.

6 Conclusion

Using a large natural experiment and a natural field experiment, we estimate peak and off-peak price elasticities for urban mass transit demand in San Francisco. Both settings allow for causal identification of demand. We use these estimates to assess the welfare impacts of price subsidies applying a sufficient statistics approach. Off-peak subsidies can raise welfare, though gains are smaller when consumers internalize others' decisions. Welfare effects vary across time periods, reflecting differences in demand and congestion. Targeting subsidies can further improve welfare, but only if the regulator has accurate demand information. The experiments offer different insights. In the natural experiment, the subsidy generally improves welfare, with a base-case MVPF of 1.6 and average net benefits of \$0.36 per dollar of subsidy. In the field experiment, MVPFs vary more depending on the time period targeted.

We highlight three directions for future research. First, further experiments—especially on congestion pricing—could help estimate key elasticities and improve external validity. These should explore both the intensive margin (greater use of an existing mode) and the extensive margin (mode switching or structural urban changes) as discussed in Desmet and Rossi-Hansberg (2013); Severen (2019); Tsivanidis (2019); Heblich, Redding and Sturm (2020); Barwick et al. (2024). Second, following Kleven (2021), it is worth considering the relative merits of different types of models in measuring the impact of peak pricing for both marginal and non-marginal changes. Third, as we noted, targeting does not always increase welfare when decision makers lack accurate demand data. Further empirical and theoretical work could help clarify when targeting yields meaningful gains.

References

Aiken, Emily, Suzanne Bellue, Dean Karlan, Chris Udry, and Joshua E Blumenstock. 2022. "Machine learning and phone data can improve targeting of humanitarian aid." *Nature*, 603(7903): 864–870.

Alatas, Vivi, Ririn Purnamasari, Matthew Wai-Poi, Abhijit Banerjee, Benjamin A Olken, and Rema Hanna. 2016. “Self-targeting: Evidence from a field experiment in Indonesia.” *Journal of Political Economy*, 124(2): 371–427.

Allcott, Hunt, and Dmitry Taubinsky. 2015. “Evaluating behaviorally motivated policy: Experimental evidence from the lightbulb market.” *American Economic Review*, 105(8): 2501–38.

Allcott, Hunt, Benjamin B Lockwood, and Dmitry Taubinsky. 2019. “Regressive sin taxes, with an application to the optimal soda tax.” *Quarterly Journal of Economics*, 134(3): 1557–1626.

Allcott, Hunt, Sendhil Mullainathan, and Dmitry Taubinsky. 2014. “Energy policy with externalities and internalities.” *Journal of Public Economics*, 112: 72–88.

Anderson, Michael L. 2014. “Subways, strikes, and slowdowns: The impacts of public transit on traffic congestion.” *American Economic Review*, 104(9): 2763–96.

Athey, Susan. 2018. “The impact of machine learning on economics.” In *The economics of artificial intelligence: An agenda*. 507–547. University of Chicago Press.

Barwick, Panle Jia, Shanjun Li, Andrew Waxman, Jing Wu, and Tianli Xia. 2024. “Efficiency and equity impacts of urban transportation policies with equilibrium sorting.” *American Economic Review*, 114(10): 3161–3205.

Boiteux, Marcel. 1960. “Peak-load pricing.” *The Journal of Business*, 33(2): 157–179.

Borenstein, Severin. 2005. “The long-run efficiency of real-time electricity pricing.” *Energy Journal*, 26(3).

Bundorf, M Kate, Jonathan Levin, and Neale Mahoney. 2012. “Pricing and welfare in health plan choice.” *American Economic Review*, 102(7): 3214–48.

Burkhardt, Jesse, Kenneth T Gillingham, and Praveen K Kopalle. 2023. “Field experimental evidence on the effect of pricing on residential electricity conservation.” *Management Science*, 69(12): 7784–7798.

Burlig, Fiona, Christopher Knittel, David Rapson, Mar Reguant, and Catherine Wolfram. 2020. “Machine learning from schools about energy efficiency.” *Journal of the Association of Environmental and Resource Economists*, 7(6): 1181–1217.

Byrne, David P, Leslie A Martin, and Jia Sheen Nah. 2022. “Price discrimination by negotiation: A field experiment in retail electricity.” *The Quarterly Journal of Economics*, 137(4): 2499–2537.

Callaway, Brantly, and Pedro HC Sant'Anna. 2021. "Difference-in-differences with multiple time periods." *Journal of Econometrics*, 225(2): 200–230.

Chassang, Sylvain, Padró I Miquel, Erik Snowberg, et al. 2012. "Selective trials: A principal-agent approach to randomized controlled experiments." *American Economic Review*, 102(4): 1279–1309.

Chetty, Raj. 2009. "Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods." *Annual Review of Economics*, 1(1): 451–488.

Christensen, Peter, Paul Francisco, Erica Myers, Hansen Shao, and Mateus Souza. 2024. "Energy efficiency can deliver for climate policy: Evidence from machine learning-based targeting." *Journal of Public Economics*, 234: 105098.

Cohen, Jessica, Pascaline Dupas, and Simone Schaner. 2015. "Price subsidies, diagnostic tests, and targeting of malaria treatment: evidence from a randomized controlled trial." *American Economic Review*, 105(2): 609–45.

Davis, Lucas W. 2021. "Estimating the price elasticity of demand for subways: Evidence from Mexico." *Regional Science and Urban Economics*, 87: 103651.

De Chaisemartin, Clément, and Xavier d'Haultfoeuille. 2020. "Two-way fixed effects estimators with heterogeneous treatment effects." *American Economic Review*, 110(9): 2964–96.

Desmet, Klaus, and Esteban Rossi-Hansberg. 2013. "Urban accounting and welfare." *American Economic Review*, 103(6): 2296–2327.

Domencich, Thomas A, and Daniel McFadden. 1975. "Urban travel demand-a behavioral analysis."

Donaldson, Dave. 2025. "Transport Infrastructure and Policy Evaluation." In *Handbook of Regional and Urban Economics, vol. 6*. Elsevier North Holland Amsterdam.

Dupas, Pascaline. 2014. "Short-run subsidies and long-run adoption of new health products: Evidence from a field experiment." *Econometrica*, 82(1): 197–228.

Duranton, Gilles, and Matthew A Turner. 2012. "Urban growth and transportation." *Review of Economic Studies*, 79(4): 1407–1440.

Farhi, Emmanuel, and Xavier Gabaix. 2020. "Optimal taxation with behavioral agents." *American Economic Review*, 110(1): 298–336.

Finkelstein, Amy, and Nathaniel Hendren. 2020. "Welfare analysis meets causal inference." *Journal of Economic Perspectives*, 34(4): 146–67.

Fowlie, Meredith, Catherine Wolfram, Patrick Baylis, C Anna Spurlock, Annika Todd-Blick, and Peter Cappers. 2021. "Default effects and follow-on behaviour: Evidence from an electricity pricing program." *The Review of Economic Studies*, 88(6): 2886–2934.

Gendron-Carrier, Nicolas, Marco Gonzalez-Navarro, Stefano Polloni, and Matthew A Turner. 2022. "Subways and urban air pollution." *American Economic Journal: Applied Economics*, 14(1): 164–96.

Gerarden, Todd D, and Muxi Yang. 2023. "Using targeting to optimize program design: evidence from an energy conservation experiment." *Journal of the Association of Environmental and Resource Economists*, 10(3): 687–716.

Glaeser, Edward L. 2007. "The economics approach to cities." National bureau of economic research.

Glaister, Stephen. 1974. "Generalised consumer surplus and public transport pricing." *Economic Journal*, 84(336): 849–867.

Glaister, Stephen, and Davis Lewis. 1978. "An integrated fares policy for transport in London." *Journal of Public Economics*, 9(3): 341–355.

Goldszmidt, Ariel, John A List, Robert D Metcalfe, Ian Muir, V Kerry Smith, and Jenny Wang. 2020. "The Value of Time in the United States: Estimates from Nationwide Natural Field Experiments." National Bureau of Economic Research.

Hahn, Robert W, Nathan Hendren, Robert D Metcalfe, and Ben Sprung-Keyser. 2024. "Welfare Analysis of Policies Impacting Climate Change."

Hahn, Robert W, Robert D Metcalfe, and Eddy Tam. 2023. "Welfare estimates of shifting peak travel." NBER Working Paper 31629.

Harberger, Arnold C. 1964. "The measurement of waste." *American Economic Review*, 54(3): 58–76.

Harrison, Glenn W, and John A List. 2004. "Field experiments." *Journal of Economic literature*, 42(4): 1009–1055.

Haywood, Luke, and Martin Koning. 2015. "The distribution of crowding costs in public transport: New evidence from Paris." *Transportation Research Part A: Policy and Practice*, 77: 182–201.

Heblich, Stephan, Stephen J Redding, and Daniel M Sturm. 2020. "The making of the modern metropolis: evidence from London." *Quarterly Journal of Economics*, 135(4): 2059–2133.

Hendren, Nathaniel. 2013. "Private information and insurance rejections." *Econometrica*, 81(5): 1713–1762.

Hendren, Nathaniel. 2020. "Measuring economic efficiency using inverse-optimum weights." *Journal of Public Economics*, 187: 104198.

Holmgren, Johan. 2007. "Meta-analysis of public transport demand." *Transportation Research Part A: Policy and Practice*, 41(10): 1021–1035.

Ito, Koichiro, Takanori Ida, and Makoto Tanaka. 2018. "Moral suasion and economic incentives: Field experimental evidence from energy demand." *American Economic Journal: Economic Policy*, 10(1): 240–67.

Ito, Koichiro, Takanori Ida, and Makoto Tanaka. 2023. "Selection on welfare gains: Experimental evidence from electricity plan choice." *American Economic Review*, 113(11): 2937–2973.

Jacobsen, Mark R, Christopher R Knittel, James M Sallee, and Arthur A Van Benthem. 2020. "The use of regression statistics to analyze imperfect pricing policies." *Journal of Political Economy*, 128(5): 1826–1876.

Jacobsen, Mark R, James M Sallee, Joseph S Shapiro, and Arthur A Van Benthem. 2023. "Regulating untaxable externalities: Are vehicle air pollution standards effective and efficient?" *The Quarterly Journal of Economics*, 138(3): 1907–1976.

Jansson, Kjell. 1993. "Optimal public transport price and service frequency." *Journal of Transport Economics and Policy*, 33–50.

Jessoe, Katrina, and David Rapson. 2014. "Knowledge is (less) power: Experimental evidence from residential energy use." *American Economic Review*, 104(4): 1417–38.

Joskow, Paul L, and Catherine D Wolfram. 2012. "Dynamic pricing of electricity." *American Economic Review*, 102(3): 381–85.

Kleven, Henrik J. 2021. "Sufficient Statistics Revisited." *Annual Review of Economics*, 13: 515–538.

Knittel, Christopher R, and Ryan Sandler. 2013. "The welfare impact of indirect pigouvian taxation: Evidence from transportation." National Bureau of Economic Research.

Knittel, Christopher R, and Samuel Stolper. 2025. "Using machine learning to target treatment: The case of household energy use." *The Economic Journal*, ueaf028.

Kreindler, Gabriel. 2024. "Peak-Hour Road Congestion Pricing: Experimental Evidence and Equilibrium Implications." *Econometrica*, 92(4): 1233–1268.

Larcom, Shaun, Ferdinand Rauch, and Tim Willems. 2017. "The benefits of forced experimentation: striking evidence from the London underground network." *Quarterly Journal of Economics*, 132(4): 2019–2055.

Levin, Laurence, Matthew S Lewis, and Frank A Wolak. 2017. "High frequency evidence on the demand for gasoline." *American Economic Journal: Economic Policy*, 9(3): 314–47.

List, John A. 2020. "Non est disputandum de generalizability? A glimpse into the external validity trial." National Bureau of Economic Research.

List, John A. 2022. *The voltage effect: How to make good ideas great and great ideas scale.* Currency.

Litman, Todd. 2004. "Transit price elasticities and cross-elasticities." *Journal of Public Transportation*, 7(2): 3.

Mahoney, Neale, and E Glen Weyl. 2017. "Imperfect competition in selection markets." *Review of Economics and Statistics*, 99(4): 637–651.

McFadden, Daniel. 1974. "The measurement of urban travel demand." *Journal of Public Economics*, 3(4): 303–328.

McFadden, Daniel. 2001. "Economic choices." *American Economic Review*, 91(3): 351–378.

McFadden, Daniel, Antti Talvitie, Stephen Cosslett, Ibrahim Hasan, Michael Johnson, Fred Reid, and Kenneth Train. 1977. "Demand model estimation and validation." *Urban Travel Demand Forecasting Project, Phase*, 1.

McFadden, Daniel, William B Tye, and Kenneth Train. 1977. *An application of diagnostic tests for the independence from irrelevant alternatives property of the multinomial logit model.* Institute of Transportation Studies, University of California Berkeley, CA.

Mohring, Herbert. 1972. "Optimization and scale economies in urban bus transportation." *The American Economic Review*, 62(4): 591–604.

Mullainathan, Sendhil, and Jann Spiess. 2017. "Machine learning: an applied econometric approach." *Journal of Economic Perspectives*, 31(2): 87–106.

Parry, Ian WH, and Kenneth A Small. 2009. "Should urban transit subsidies be reduced?" *American Economic Review*, 99(3): 700–724.

Piketty, Thomas, Emmanuel Saez, and Stefanie Stantcheva. 2014. "Optimal taxation of top labor incomes: A tale of three elasticities." *American Economic Journal: Economic Policy*, 6(1): 230–71.

Polyakova, Maria, and Stephen P Ryan. 2019. "Subsidy targeting with market power." National Bureau of Economic Research.

Rodrik, Dani. 1987. "Policy targeting with endogenous distortions: Theory of optimum subsidy revisited." *Quarterly Journal of Economics*, 102(4): 903–911.

Saez, Emmanuel, Joel Slemrod, and Seth H Giertz. 2012. "The elasticity of taxable income with respect to marginal tax rates: A critical review." *Journal of Economic Literature*, 50(1): 3–50.

Severen, Christopher. 2019. "Commuting, labor, and housing market effects of mass transportation: Welfare and identification." *Review of Economics and Statistics*, 1–99.

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

Small, Kenneth A, Erik T Verhoef, and Robin Lindsey. 2024. *The economics of urban transportation.* Routledge.

Tarduno, Matthew. 2022. "For Whom the Bridge Tolls: Congestion, Air Pollution, and Second-Best Road Pricing." *Unpublished manuscript*.

Tizghadam, Ali, Hamzeh Khazaei, Mohammad HY Moghaddam, and Yasser Hassan. 2019. "Machine learning in transportation."

Train, Kenneth. 1977. "Optimal transit prices under increasing returns to scale and a loss constraint." *Journal of Transport Economics and Policy*, 185–194.

Train, Kenneth E. 2009. *Discrete choice methods with simulation.* Cambridge university press.

Tsivanidis, Nick. 2019. "Evaluating the impact of urban transit infrastructure: Evidence from bogota's transmilenio." *Unpublished manuscript*.

Vickrey, William S. 1963. "Pricing in urban and suburban transport." *American Economic Review*, 53(2): 452–465.

Weyl, E Glen, and Michal Fabinger. 2013. "Pass-through as an economic tool: Principles of incidence under imperfect competition." *Journal of Political Economy*, 121(3): 528–583.

Williams, Kevin R. 2022. "The welfare effects of dynamic pricing: Evidence from airline markets." *Econometrica*, 90(2): 831–858.

Winston, Clifford. 1985. "Conceptual developments in the economics of transportation: an interpretive survey." *Journal of Economic Literature*, 23(1): 57–94.

Winston, Clifford. 2013. "On the performance of the US transportation system: Caution ahead." *Journal of Economic Literature*, 51(3): 773–824.

Wolak, Frank A. 2007. "Residential customer response to real-time pricing: The Anaheim critical peak pricing experiment."

Yang, Nan, and Yong Long Lim. 2018. "Temporary incentives change daily routines: Evidence from a field experiment on Singapore's subways." *Management Science*, 64(7): 3365–3379.

A ONLINE APPENDIX

A.1 Measuring the marginal external cost of a trip for the natural experiment

To estimate the marginal social cost of a trip in the peak period (MSC_1), we need an estimate of marginal external cost (MEC_1). We define congestion, e, as crowding (or density) and delay. Crowding is measured as the number of passengers per square meter in a train car in the peak period. The reason we chose this measure is because reducing crowding was an objective in both experiments and it is an important externality. We used BART train-level crowding data to calculate the average density of a train car.

The marginal external cost during the peak period is defined as $MEC_1 = \frac{\partial u}{\partial e}\frac{\partial e}{\partial x_1}$. It is the product of the effect of a congestion change on utility and the effect of an increase in peak rides on congestion (crowding measured in density). We assume that an additional peak period trip has a constant effect on congestion that does not vary with the level of subsidy. Moreover, the value placed on incremental congestion in the peak period by the representative rider is also assumed to be constant. This implies MEC_1 is also constant.

We can derive the marginal external cost in the peak period with some additional assumptions. We assume passengers are distributed evenly across all train cars in the system, therefore $n = \frac{Nx_1}{k}$ where k is number of train cars during peak period, n is the number of passengers per car, and $e = \frac{n}{a} = \frac{Nx_1}{ka}$ is the density in the car with a as the size of a car.[45] Under this assumption, the effect of a peak period trip on density in a train car is $\frac{\partial e}{\partial x_1} = \frac{e}{x_1}$.

We define the effect of a congestion change on utility that results from crowding and delay in journey time as ϕ_c and ϕ_d respectively.

A rider's disutility from crowding depends on the level of congestion. We assume a minute on a train with congestion level e gives rider disutility equivalent to that of spending $T(e)$ minutes on a train with $e = 0$. (i.e., $T(e)$ is the "time-multiplier" in Haywood and Koning (2015). Therefore, the disutility from crowding for a train ride of t minutes with congestion level e equals $t * T(e)$ in terms of time, and $VOT * t * T(e)$ in monetary terms, where VOT is the value of time for a rider. If riders on average take x_1 trips a day on a congested train (e.g., in the peak period), the disutility from crowding per day is $VOT * t * T(e) * x_1$.

We assume that a marginal increase in density e increases the (per minute) disutility from crowding by a constant factor k (i.e. $T^{'}(e) = k > 0$).The change in utility resulting from a marginal increase in congestion therefore equals $\phi_c = VOT * t * k * x_1$, holding VOT, t and x_1 constant.[46] Similarly, a rider may be delayed when congestion increases because it takes longer for all passengers to board the train. We assume an increase in density increases journey time by a constant, o. We can empirically estimate o from our data (see below). For each trip, this results in a monetary loss of $VOT * o$. Since each rider takes x_1 peaks trip per day, $\phi_d = VOT * o * x_1$. Following Haywood and Koning (2015), and adding the effect from delay, we define

$$MEC_1 = \frac{\partial u}{\partial e}\frac{\partial e}{\partial x_1} = (\phi_c + \phi_d)\frac{\partial e}{\partial x_1} = VOT * t * k * e + VOT * o * e \tag{A1}$$

where the second equality is a first order approximation.

We assume the value of time is half the hourly wage (Small, 2012) or 75% of the hourly wage (Goldszmidt et al., 2020), which implies VOT = \$12.5 or VOT = \$18.68 respectively (the median wage in San Francisco in 2016 was \$24.9).[47] We calculate the average in-train time in the peak period as 37 minutes from our BART gate entry and exit data.

[45] We relax this assumption in our field experiment. See discussion below.

[46] As our model in section 2 is in terms of representative rider, the fact that ϕ_c, the change in utility resulting from a marginal increase in congestion depends on the number of peak ride per person x_1, is equivalent to that the marginal change in congestion density is increasing with the total number of rides in the system. We assume that the marginal external cost of a ride in off-peak to be zero in section 2, this could be considered as either because T'(e) to be very low at a low level of e, i.e. a change of density has no impact on the utility of a rider per ride because of preferences. Otherwise, the marginal external cost in the off-peak period could be lower than the peak-period because there are fewer riders experience following equation A1, and it would only be approximate zero if the density/traffice in the off-peak period is close to zero.

[47] Metropolitan and Non-metropolitan Area Occupational Employment and Wage Estimates, San Francisco-Oakland-Hayward, CA, US Bureau of Labor Statistics, May 2016

The average density in the peak period, e, is 1.13 passengers per square meter, estimated from our novel BART train crowding data. We use a time-multiplier of 0.11 from Haywood and Koning (2015). Thus, $VOT * t * k * e = \$0.954$.

We find that an extra passenger per car results in a delay in departure time relative to its schedule time of 1.38 seconds, or 0.023 minute (see Table A1). This implies a unit increase in density increases journey time by 1.59 minutes on average, i.e., $o = 0.026$ (= 1.59/60mins). Thus, $VOT * o * e = \$0.37$. As $d = \frac{n}{a}$, where d is density, n is number of person per car, a is the size of the car (68.93 sqm), let t be the journey time, $\frac{\partial t}{\partial d} = a\frac{\partial t}{\partial n} = \frac{1.38}{60} * 68.93$, which implies $MEC_1 = \$\ 1.33$ per peak hour trip.

To calculate the effect of an increase in density on delay (i.e., o), we use train-level data from December 2016 to March 2017. The data contains the scheduled and actual departure information for each train at each station, for all trains that run for the day. We estimate the following specification using the train and station level data

$$d_{ijkdt} = \alpha + \beta n_{ijkdt} + \mu_{ij} + \gamma_t + \eta_d + \epsilon_{ikdt} \tag{A2}$$

where d_{ijkdt} is the difference between the actual departure (door closing) time with the scheduled departure (door closing) time, for train k on date d scheduled at time t departing from station i towards station j. n_{ijkdt} is the number of passenger per car on the train. To control for any systematic delay at the station or route level, we control for the station-direction specific fixed effect μ_{ij}. We control for 20 minute time interval fixed effects γ_t, and date specific fixed effects η_d that account for unobserved factors that affect delay on specific dates or time. As robustness check, we control for any planned delay in terms of scheduled run time between the previous station and station i for the train k (X_{ijkdt}).

The number of passenger per car for a train at a station i is related to the net number of passengers getting on the train (i.e., number of people boarding the train net number of people getting off the train) at all the stations before i, plus that at station i. To address potential confounding factors at the route-time level, e.g., unexpected events or accidents happening at the route-level or at the stations before i that affects both the delay time and the density, we use the number of passenger boarding the train cars at station i as instrument for the passenger per car on the train at station i (before departure). Our estimate therefore exploit the variation in the number of boarding passengers at different stations and time to identify the effect of congestion on delay. Table A1 presents the estimates of β in equation A2.

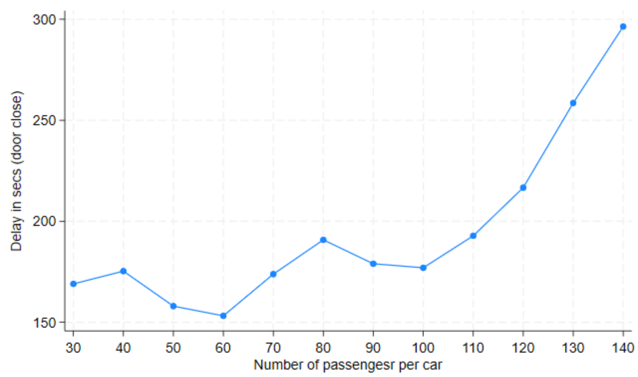
An increase in density in a train car, related to an extra passenger boarding at a specific station, could be associated with additional delay from: i) a passenger boarding at the station of origin, ii) an increase in in-car crowding that slows down boarding of other passengers at stations between the origin and the destination, and iii) a passenger leaving the train at the destination. Figure A1 plots the relationship between the number of passengers per car and the time delay of the train during peak time. We estimate equation A2 exploiting empirically the variation in the number of passengers boarding at each station.

Table A1: The effect of density on train delay

	Outcome: Train-station delay (sec) Sample - Peak period trains			
	(1)	(2)	(3)	(4)
	OLS	OLS	IV	IV
Passenger per car	0.636^{***}	0.504^{***}	1.370^{***}	1.388^{***}
	(0.153)	(0.180)	(0.344)	(0.343)
Observations	72181	71540	72181	71540
Train-station specific controls	N	Y	N	Y

Notes: The table presents the regression results with the outcome as the difference between a schedule departure time (door closing) and the actual departure time (door closing) on the passenger per train car on the train. Controls includes date fixed effect. All columns control for date fixed effects, time-specific train fixed effects and station-line fixed effects.

Figure A1: Passenger density and delay



Notes: The figure plot the delay of train (in secs) by the number of passengers per train car (bins width is 10).

A.2 Estimation of the marginal cost per trip

We estimate the marginal private cost of a BART trip using information on BART's operating costs. We combine various sources of BART financial and operations information to compute the marginal private cost of an individual BART trip. Our measure of operating expenses includes the cost of maintenance.[48]

Assuming the total operating cost of BART depends only on the total number of passenger trips, i.e., $C(Nx_1, Nx_2) = C(Nx_1 + Nx_2)$, the relationship of marginal cost to average cost with respect to the number of passenger trips can be estimated using the equation:

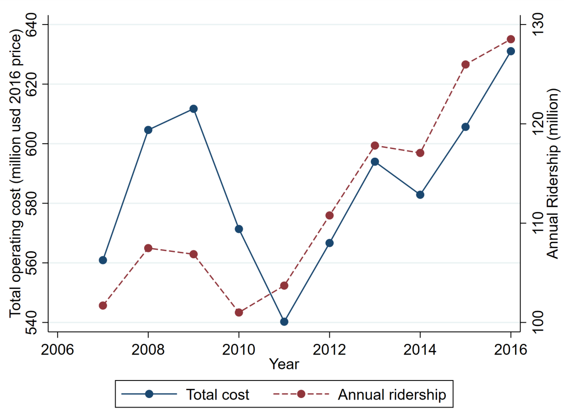
$$C_t = \alpha + \beta N x_t + \epsilon_t \tag{A3}$$

[48] p. 4-20, BART SRTP-CIP.

where C_t is the total operating cost of BART in year t, Nx_t is the annual number of total trip in the system, and β represent the marginal private cost per BART trip.

From 2007 to 2016, the cost per passenger increases from \$4.47 to \$4.91 in nominal terms.[49] In real terms, it declines from \$ 5.66 to \$ 4.91,[50] while the annual ridership increases from 101 to 128 million.[51] Figure A2 suggests that there is closer relationship between annual ridership and total operating cost. Table A2 column (1) provide estimate of the above equation. The implied marginal cost in 2016 is \$1.89. The estimate is robust to controlling for yearly average energy price per kwh in San Francisco.

Figure A2: Cost per trip and annual number of trips



Note: The graph plots the total operating cost and annual ridership from 2007 to 2016, in 2016 prices.

Table A2: Estimation of marginal cost

	(1)	(2)
	Outcome: Total Cost	
Annual ridership	1.898**	1.695
	(0.724)	(1.184)
Energy price control	N	Y
Observations	10	10

Note: The table represent estimate of equation A3. Total operating and annual rider ship data from from BART annual financial report. Energy price is annual average price per kwh (in 2016 price), data source from U.S. Bureau of Labor Statistics, Electricity Per KWH in San Francisco-Oakland-Hayward, CA (CBSA) [APUS49B72610], retrieved from FRED, Federal Reserve Bank of St. Louis.

A.3 Estimation of feedback effect from congestion

In this section, we present the methodology and the empirical estimates on the demand response to congestion. When the level of congestion increases or decreases, this may affect the

[49] Figure 3-2 BART SRTP-CIP.

[50] Deflated by San Francisco CPI provided by Bureau of Labor Statistics (https://www.bls.gov/regions/west/data/consumerpriceindex_sanfrancisco_table.pdf).

[51] Figure 3-3 BART SRTP-CIP.

travel demand by consumer, holding other factors constant. We need to estimate this demand parameter to empirically evaluate and scale-up the welfare consequences of small-scale price subsidies when considering the endogenous congestion case.

In the endogenous case, we need to estimate two important parameters, $v_{11} = -\frac{\partial x_1}{\partial e}\frac{\partial e}{\partial x_1}$ and $v_{21} = -\frac{\partial x_2}{\partial e}\frac{\partial e}{\partial x_1}$, which are the demand responses of peak and off-peak travel with respect to a marginal change in peak-hour traffic in the aggregate. Intuitively, it measures how individual demand responds to (expected) aggregate congestion changes. In this section we estimate the demand response with respect to the aggregate travel, via congestion as the channel.

We use the travel data of participants in the natural experiment to estimate these parameters. We exploit variation in naturally occurring congestion changes, and that some riders may be more affected by congestion at certain stations, to identify how riders responds to congestion. Our strategy focus on variation in naturally occurring congestion at some part of the system, e.g., idiosyncratic events or conferences near the stations that lead to an increase in demand to travel to the area (therefore increasing congestion), and control for potential confounding factors that affect the demand on all parts of the system, e.g., weather.

The majority of morning peak hour journey in BART uses two exit stations, Embarcadero and Montgomery Street. Riders who regularly uses these busy stations are more likely to be affected by congestion at these stations, and may adjust their travel demand accordingly when (expected) congestion at these stations changes. We estimate how congestion at these two exit stations in the morning affects the peak and off-peak demand for riders who regularly exit at these two stations. We use other riders who do not regularly use these two stations as the control group (in the period before the price subsidies were introduced).

The treatment variable is the aggregate number of rides that exit at these two station on a given day (*i.e.*, comparing naturally occurring busy and quiet days), interacted with a dummy variable in which the rider regularly uses these two stations (comparing with riders who do not regularly uses these two stations). This is effectively a difference-in-differences strategy.

The aggregate number of rides that exit at the two stations proxy for the (expected) level of congestion at the two stations on each day. We assume that riders who travel regularly on route that exit at these two stations have correct expectation on the congestion at these two stations. For example, a rider who travels regularly would be able to infer from public information the level of congestion at the stations they uses regularly. Under this assumption, our strategy estimate the demand respond to a change in the congestion level that are relevant to the riders' trip.

To control for unobserved factors that affects both demand and overall usage of BART system, e.g., weather or events that disrupt or affect travel demand for the whole BART system, we use riders who do not regularly exit at the two stations as the control group. We assume that the demand from riders who do not travel regularly at these stations would not depend on the congestion level on the station that they do not travel with.

Specifically, we estimate the following equation

$$y_{ijt} = \beta Busy_i * PeakBusy_t + \mu_t + \gamma_i + \epsilon_t \tag{A4}$$

where y_{it} is number of peak hour trip by user j on date t. $Busy_i$ is an indicator for user i regularly exit at Montgomery street and Embarcadero station in morning, the two busiest stations in BART. $PeakBusy_t$ is the total number of trips (started) in peak hour that exited at the two busy stations, in the whole BART system. γ_i and μ_t are rider and date fixed effects.

Table A3: Peak hour trip by riders who regularly exit at Montgomery street and Embarcadero, compared with others

	Number of trip			
	(1)	(2)	(3)	(4)
	Peak	Peak	Off-peak	Off-peak
Rider who regularly uses Embarcadero and Mont. St	-0.00617*	-0.00635*	0.00858**	0.00877**
× Peak hour rides exiting at E&M (10,000)	(0.00371)	(0.00372)	(0.00365)	(0.00366)
N	1807471	1807471	1807471	1807471
Rider FE	Y	Y	Y	Y
Date FE	N	Y	N	Y

Notes: The table present the estimates of equation A4. Sample includes dates that riders have at least 1 trip for the estimation of intensive margin response. Standard errors robust to 7 days of auto-correlation.

Table A3 reports the estimate of equation (A4).The sample include the travel demand of the Perks program participants from April 2016 to February 2017. There are 17,185 riders in the sample, among them, 7,477 use the Embarcadero and Montegomery Street stations regularly–this is defined as these users exiting at one of the two stations with above median percentage (median is around 97%) among their morning trips (5.30-10.30am) in the sample period. We find that an increase in 10,000 peak hour riders exiting at the two busiest station reduces the peak demand of riders who uses these routes regularly (among the participants of experiment 1) by -0.00635 per trip per day at peak hour. We also find that their demand for off-peak trip increase by 0.00858 trip per day at off-peak hour.

The above are estimates of the effect of changes in aggregate rides on trip demand per person - $\frac{\partial x_1}{\partial e}\frac{\partial e}{\partial N x_1}$ and $\frac{\partial x_2}{\partial e}\frac{\partial e}{\partial N x_1}$. For the welfare analysis in section 4, we scaled the aggregate rides by the total number of BART riders, which gives the feedback effect of the increase in peak rides per person on the peak and off-peak demand $\frac{\partial x_1}{\partial e}\frac{\partial e}{\partial x_1}$ and $\frac{\partial x_2}{\partial e}\frac{\partial e}{\partial x_1}$. Using $N = 271{,}341$, we find $v_{11} = -0.172$ and $v_{21} = 0.237$.We use the estimates of these two parameters for our welfare analysis in section 5.

A.4 Welfare effect of the natural experiment: Transbay and other routes

This section presents the welfare calculations for a targeted subsidy of $0.25 for the Transbay route (i.e., a heavily congested route) and all other routes. The key point is that welfare increases by 23% percent when targeting the Transbay route compared with all other routes.

The welfare calculation assumes that all network effects of the subsidy are captured by our demand parameters derived in section (3), and the two systems can be treated independently. In terms of the welfare equation (equation (1)), we are measuring differences in demand and marginal social cost across the two sets of routes.

The first three rows of the table provide key parameters. MEC_1 and MSC_1 represent the marginal external cost and marginal social cost in the peak period. Congestion is a key determinant of MEC_1 as explained in section A.1. Congestion is measured as the number of passengers per square meter in a train car (see Appendix Table A11 for details). The equilibrium level of congestion depends on whether consumers take congestion into account. The congestion during the peak period is about 0.4% percent higher in the case of endogenous congestion. The welfare benefit associated with congestion alleviation is smaller in the endogenous case because of a range of factors, including a weaker demand response.

The third and fourth rows of the table summarize welfare changes associated with peak and off peak travel. The total welfare change (representing the sum of the welfare changes for the peak and off-peak travel) is given in the final row for various cases. In comparison, the MVPF for the Transbay routes are 1.6818 (95% CI: 1.5602, 1.803) in exogenous congestion case and 1.6336 (1.515, 1.751) in endogenous congestion case respectively.

Table A4: Welfare effect of the subsidy for the Transbay route and other routes

	Transbay routes			Other routes		
Parameter	Exo. cong.	Endo. cong.	% Diff	Exo. cong.	Endo. cong.	% Diff
MEC_1	$2.23	$2.23	.	$0.93	$0.93	.
MSC_1	$4.12	$4.12	.	$2.82	$2.82	.
Congestion	1.518	1.5243	0.4	0.8116	0.815	0.4
Welfare change with change in						
Peak travel	0.0006	0.0005	-14.7	-0.0057	-0.0048	-14.7
Off-peak travel	0.0328	0.0309	-5.9	0.0328	0.0309	-5.9
Total	0.0334	0.0314	-6.1	0.0271	0.026	-4.1

A.5 Additional discussion for welfare analysis

This section explains how we include the benefits from reduced externalities from other modes of transit in an illustrative calculation. It analyzes the potential impact of other travel modes on the estimation of welfare from section 5. The basic point is that considering possible travel shifts from other more polluting modes to mass transit could increase the welfare estimate. We estimate this increase in welfare could be 15.4 percent.

While our model in section 2 does not explicitly consider other modes of transport, it is possible that other modes of transport, such as automobiles or bus, could affect welfare. Using the sufficient statistics framework, if the price paid by the consumer is equal to the marginal social cost in the other mode of transport, then there would be no change in welfare resulting from the consumption change in that mode.

If these other modes of transport are not priced optimally, the change in consumption in the other modes would have an impact on the welfare. Our estimate of the welfare effect of the subsidy could be biased downward if the market prices in these other modes are below their marginal social cost. This is because we find net increases in travel on BART with the introduction of the subsidy. Formally, we consider the case where $\frac{dx_1}{ds} + \frac{dx_2}{ds} > 0$, which implies, on average, that there are participants who demand more BART rides after receiving the subsidy. This is consistent with our empirical estimation in the natural experiment.

We provide an illustrative calculation for the case of emission reductions that could result from the mode shift to BART.[52] For simplicity, we assume that total travel remains unchanged and an increase in a BART trip reduces automobile travel by the same amount in terms of miles traveled. Then, the net increase in welfare from shifting travel is given by the following formula: $\left(\frac{dx_1}{ds} + \frac{dx_2}{ds}\right)s'\phi \geq 0$, where ϕ_r is the externality measured in dollars per trip.

Using the parameters in Table A5, we calculate that the externality associated vehicle travel per mile associated with air pollution is \$0.0055 per mile of automobile travel. With an average BART journey of 14.6 miles per trip, this implies a net increase in BART trips is associated with a reduction of external cost of \$0.08 ($\phi = 0.08$). We present the welfare estimates with and without the benefits from emission reductions for the natural experiment in Table A6 - it shows that welfare benefits increase by about 15.8 percent in the endogenous congestion case and 15.4 percent in the exogenous case when mode shifting is accounted for.

Table A5: Parameters for vehicle externality

	CO	HC	NO_x	CO_2
Tier 2 Federal Exhaust Standards (2007-2016), grams per mile	3.40	0.10	0.14	
Emission per mile (grams)	1.70	0.05	0.07	404
Damage per mile (\$)	0.0020	0.0008	0.0027	0.0206
Total damage per mile (\$)	0.0261			
Average damage per trip (avg. trip length, 14.60 miles)	0.38			

Note: This table shows the 'average damage per trip' parameter used to calculate the externality associated with each extensive margin response per trip from vehicle. It assumes "New vehicle emissions are typically 40 to 50 percent of exhaust standards" as in Jacobsen et al. (2023). It does not consider variation in vehicle ages or types.

[52] It is possible that there could be congestion benefits as well for automobile travelers. We do not include such benefits because we do not have good data on their magnitude for shifts in particular time periods. If, for example, shifts in automobile travel to BART occurred during periods in which there was little congestion on highways, the benefits from reduced congestion would be small.

Table A6: Welfare impacts of emission benefits for the base case and a targeted subsidy

	Parameter	Base case		Targeted subsidy	
		Exogenous congestion	Endogenous congestion	Exogenous congestion	Endogenous congestion
(1)	Total Welfare change	0.0291	0.0277	0.0334	0.0314
(2)	Total Welfare change including vehicle externality	0.0336	0.0321	0.0379	0.0358
	Percent change (%)	15.4	15.8	13.5	14

Notes: Welfare is measured in dollars per person per weekday. The total welfare change is calculated from eq. (1) in the row(1); and it is calculated with an adjustment factor of $\left(\frac{dx_1}{ds} + \frac{dx_2}{ds}\right)s'\phi$ assuming $\phi = 0.38$, $s' = 0.25$.

A.6 Additional estimates on welfare impact

In this section we provide a numerical example for how we calculate the annual welfare benefits if all BART riders receive the price subsidy used in the natural experiment.

Calculation of welfare impact We provide details on how we calculate the annual welfare impact of the subsidy for two cases that help bound the welfare estimates. First we compute the total benefits from the experiment in the case of exogenous congestion if all riders in BART received the subsidy and had the same elastcities as those in the natural experiment.[53] Second we compute the total benefits from the experiment in the case of exogenous congestion if only those riders in BART receive the subsidy, and other riders do not change their behavior.

To compute annual benefits from rolling out the subsidy to all riders, we multiply three numbers – the number of riders, the average daily benefits for those riders, and the number of days in a year. Our estimate of total riders in BART is $N = 271,341$. The average welfare impact per person per weekday is \$0.029 in the base case. This gives an annual welfare estimate of $2,052,966$ (i.e., $0.029 * 271,341 * 260$).

We next compute annual benefits from rolling out the subsidy to a subset of riders who actually participate in the experiment. We use eq. (A9), which considers two types of riders. In addition, we make the assumption that riders not in the experiment do not change their behavior. We do this to illustrate the idea in the extreme case, and thus provide a plausible lower bound for the welfare result. In terms of eq. (A9), this means that $k = 0$ (i.e., those riders not in the experiment do not change their behavior.) Applying the equation, we let $n_a = 17,545$, $n_b = N - n_a = 253,796$ and other parameters are the same as in the previous case. This gives an average welfare impact of \$0.00188 per person per day. The aggregate welfare impact for the program rolled out for 6 months (130 days) is calculated as $0.00188 * 271,341 * 130 = \$66,373$. If the subsidy program is rolled out for the whole year, the overall welfare impact would be twice that or 0.18 million.

[53]The calculation of the endogenous congestion case is similar, except we use the numbers for demand responsiveness associated with that case.

Calculation of net welfare change per dollar of subsidy Under the assumption of exogenous congestion, the net welfare increase for the natural experiment is \$0.0291 per person per day. We calculate the subsidy expenditure in this case as $x_2' * s$, where x_2' is the off-peak travel per day per person with the subsidy, and $s = 0.25$. In this case, $x_2' = x_2 + \frac{dx_2}{ds}s = 0.32$. This implies that the net welfare benefit per dollar of subsidy expenditure is $\frac{0.0291}{0.32*0.25} = 0.36$. If we allow for the two types of consumers to differ in their demand response, and assume that other BART riders are half as responsive to price as the participants in experiment 1, the net welfare benefit is 0.016 per person per day. $x_2' = x_2 + (\frac{n_a}{n_a+n_b} + \frac{n_b}{n_a+n_b}0.5) * \frac{dx_2^b}{ds}s = 0.32$. Thus, the net welfare benefit per dollar of subsidy is 0.196.

A.7 Natural experiment welfare estimates when the marginal social cost varies

A key assumption in the body of the paper is that the marginal social cost is constant in the peak period and in the off-peak period. In this appendix, we consider the empirical implications of relaxing that assumption for the peak period. We compare the welfare result using different functional forms for the MEC function. Our main conclusion is that in this particular example, the results do no change much in moving from a constant MEC function to a linear or quadratic MEC function.

A.7.1 Allowing the marginal social cost to vary with the level of congestion

Following the model in Section 2 and equation (1), we derive the formula in the case when the MSC is a non-linear function of congestion, and discuss how that allows us to estimate the impact of MSC_1 varying with e. We present the derivation here.

When MEC_1 varies with the level of e, we have $MSC_1(e) = MPC + MEC_1(e)$. We can rewrite the second term in the final expression in (1) as $\int_0^{s'} MSC_1(e)\frac{dx_1}{ds}ds = MPC\frac{dx_1}{ds}s' + \frac{dx_1}{ds}\int_0^{s'} MEC_1(e)ds$. One way to simplify the preceding expression is to define an average MEC_1 over the range of the subsidy as $\overline{MEC}_1 = \int_0^{s'} MEC_1(e)ds\frac{1}{s}$. It can then be rewritten as:

$$MPC\frac{dx_1}{ds}s' + \frac{dx_1}{ds}s'\int_0^{s'} MEC_1(e)ds\frac{1}{s'} = MPC\frac{dx_1}{ds}s' + \frac{dx_1}{ds}s'\overline{MEC}_1 \quad \text{(A5)}$$

(A5) is general, and thus allows for a non-linear functional form for MEC_1. The difference between the second term in equation (A5) associated with the marginal social cost, compared with that of equation (1), is that the constant MEC_1 is replaced with $\overline{MEC}_1$ for the calculation of MSC_1. The other terms in equation (A5) would be the same as in equation (1).

Estimating $\overline{MEC}_1$ requires knowledge of both the shape of the $MEC_1(e)$ function and also how congestion varies with the subsidy. In particular, $\overline{MEC}_1 = \int_{e(0)}^{e'} MEC_1(e)\frac{1}{\frac{de}{ds}}de\frac{1}{s'}$, where e'

and $e(0)$ are the congestion at the subsidy level s' and before the subsidy respectively, e.g. if we assume $MEC_1(e) = \beta_0 + \beta_1 e + \beta_2 e^2$ and $\frac{de}{ds}$ is constant, we could calculate $\overline{MEC}_1$ numerically.

We can derive a simpler formula if we assume MEC_1 is linear. Suppose we assume that the marginal external cost can be written as $MEC_1(e) = \alpha_0 + \alpha e$, where α_0 is a non-negative constant and α is a positive constant. The second term in equation (1), associated with the marginal social cost in the peak period, becomes $\int MSC_1(e)\frac{dx_1}{ds}ds = MPC\frac{dx_1}{ds}s' + \frac{dx_1}{ds}\alpha_0 s' + \frac{dx_1}{ds}\alpha\int_0^{s'} e(s)ds$. If we assume $\frac{de}{ds} = \frac{de}{dx_1}\frac{dx_1}{ds}$ is approximately constant, i.e. $\frac{de(s)}{ds} \approx \frac{e(s')-e(0)}{s'}$, we can simplify the expression by noting $\int e(s)ds = \frac{1}{2}\left(e\left(s'\right)+e(0)\right)s' = \bar{e}s'$, where $\bar{e}$ is the average of the congestion levels before and after the subsidy change. This implies $\int MSC_1(e)\frac{dx_1}{ds}ds = \left(MPC + \alpha_0 + \alpha\bar{e}\right)\frac{dx_1}{ds}s' = MSC_1(\bar{e})\frac{dx_1}{ds}s'$. Substituting this preceding expression into Equation (1) yields:

$$\frac{1}{N}\left(W\left(s'\right) - W(0)\right) = \left(p - MSC_1(\bar{e})\right)\frac{dx_1}{ds}s' + \left(p - \frac{1}{2}s' - MSC_2\right)\frac{dx_2}{ds}s' \qquad \text{(A6)}$$

The only difference between equation (1) and equation (A6) is that $MSC_1(\bar{e})$ replaces MSC_1.

A.7.2 Empirical implementation

Linear MEC Consider the case of a linear MEC function that varies with e. We follow the derivation above for expression (A.7.1), and the formula for the MEC function we presented in appendix A.1 where we illustrate the constant MEC case.

In the constant MEC case, we evaluate the MEC function at the initial value of $e(0)$. For the linear MEC function, following the derivation in equation (A6), we use the empirical representation of the MEC function in appendix A.1. We evaluate this function at the average congestion level before and after the subsidy, i.e., $\bar{e} = \frac{1}{2}\left(e' + e(0)\right)$, where e' is the density at subsidy level s', and $e(0)$ is the density before the subsidy. The key difference between the upward sloping linear MEC case and the constant MEC case is that we allow the MEC value to vary as the subsidy is implemented.[54]

The empirical MEC function in the linear case is based on data from (Haywood and Koning, 2015). They estimate T′(e), the rate of change of the time multiplier with respect to the density (see Appendix A.1). That paper uses a contingent valuation approach to estimate how the willingness to pay for reduced congestion varies with the level of congestion. It is based on the subway system in Paris in 2010-2011, and is thus not necessarily applicable to other areas and time periods. Thus, this calculation should be viewed as illustrative.

MEC with a more flexible functional form As a robustness check, we also implement the welfare analysis with an MEC function that is more flexible than the linear form. Specifically, we estimate the function that determines the time multiplier T(e), using the data points taken

[54] We assume α_0, the constant term in the linear MEC case, is zero. That is, the marginal external cost is zero when there is no congestion.

from Haywood and Koning (2015) by applying ordinary least squares and allowing for a quadratic functional form. (Haywood and Koning (2015) fit a linear function, which is the value of $T'(e)$ we used for the constant and linear case)[55]. The time multiplier T(e) measures how much time one is willing to travel on a crowded train with density e, relative to 1 unit of time spent on a train with no congestion.

The slope of the fitted function gives the valuation of a marginal change in density, $T'(e) = 0.075 + 0.0106e$. This gives a new MEC function $MEC_1(e) = VOT * t * (0.075 + 0.0106e) * e + VOT * o * e$. We use the constant values described in Table A11 for VOT, t, and o. The resulting functional form for $MEC_1(e)$ is quadratic in e. For comparison, in the linear case, we assume $T'(e)$ is a constant. For the quadratic MEC case, the welfare change associated with the congestion externality in the peak period is derived from eq. (1). $\int_0^{s'} MEC_1(e)\frac{dx_1}{ds}ds = \frac{dx_1}{ds}\int_0^{s'} MEC_1(e)ds = \frac{dx_1}{ds}\int_{e(0)}^{e'} MEC_1(e)\frac{1}{\frac{de}{ds}}de$ where e' and $e(0)$ is the density at subsidy level s' and before the subsidy. We integrate the term $\int_{e'}^{e''} MEC_1(e)\frac{1}{\frac{de}{ds}}de$ numerically, assuming $\frac{de}{ds}$ to be constant for simplicity.[56]. Equivalently, as discussed in the earlier subsection, we could calculate an "average" MEC, as $\frac{1}{s'}\int_0^{s'} MEC_1(e)ds$, and substitute it in place of MEC_1 in eq. (1).

Table A7 shows the results for three cases: a constant MEC function (the base case), a linear MEC function, and a quadratic MEC function. The main takeaway from Table A7 is that the welfare numbers are similar across all three cases. This reflects the facts that: 1. the MEC is only one component of welfare; and 2. The average values are not that different across all three cases (1.33, 1.298, 1.1 for the constant, linear and quadratic MEC case respectively).

Table A7: Results comparing welfare change with three different MEC functions

Welfare change	MEC constant			MEC Linear function			MEC Quadratic function		
with travel change in	Exo.	Endo.	% Diff.	Exo.	Endo.	% Diff.	Exo.	Endo.	% Diff.
Peak	-0.0037	-0.0032	-14.7	-0.00389	-0.00332	-15.2	-0.0049	-0.0041	-15.1
Off-peak	0.0328	0.0309	-5.9	0.0328	0.0309	-5.9	0.0328	0.0309	-5.9
Total	0.0291	0.0277	-4.8	0.0289	0.0276	-4.7	0.0279	0.0267	-4.3

Notes: Numbers are calculated for the base case for the natural field experiment. Col (1)-(2), (4)-(5) and (7)-(8) present the welfare estimates under assumption that the MEC function is constant, a linear function and a quadratic function respectively. Col. (1), (4), (7) ((2), (5), (8)) present welfare estimates under the assumption of exogenous (endogenous) congestion. Col. (3), (6), and (9) present the percentage change between the two cases.

The total welfare change for the exogenous case is about \$0.029 per person per day with a constant MEC function, \$0.029 for a linear MEC function and \$0.028 for a quadratic MEC function. The results are also similar for endogenous congestion (.028, .028 and .027). Overall welfare percentage decreases in welfare (between exogenous and endogenous congestion are

[55]The data points for (T(e), e) are: (1, 0), (1, 1), (1.05, 2), (1.18, 2.5), (1.26, 3), (1.4, 4) and (1.57, 6).

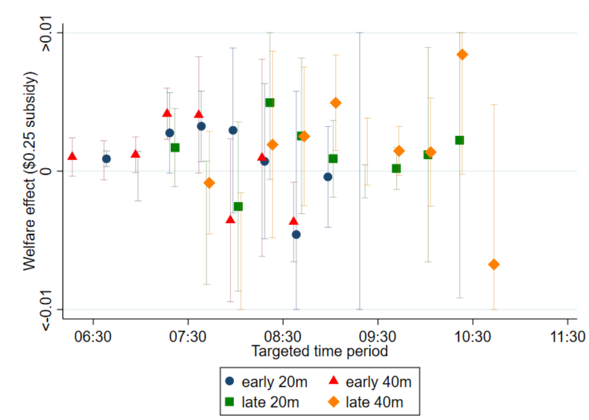
[56]We assume $\frac{de}{ds} = \frac{\partial e}{\partial x_1}\frac{dx_1}{ds}$ is constant, which is a reasonable approximation when the subsidy is small. See Appendix A.1 for more details.

also similar, on the order of 5%. It is worth pointing out that these assumptions are quite sensitive to our assumptions about the MEC function. Because there is a paucity of data on this issue, we believe more research would be useful to inform policy makers.

A.8 Welfare estimates for the natural field experiment using the welfare model

In the body of the paper, we used MVPF as an index of welfare with different values for the value of time (VOT). We found that MVPFs could be above or below one, depending on the particular field experiment. Figure A3 shows a similar calculation using a measure of net benefits per person per weekday with a $.25 subsidy, using the welfare model summarized in section 2. The key point of the figure is that some estimates for net benefits are positive and some are negative. The figure assumes that VOT is 50% of the wage. The same qualitative finding emerges in the case where VOT is 75% of the wage rate.

Figure A3: Welfare effect of the subsidy in the field experiment (with sufficient statistics approach)



Note: This figure shows the welfare effect (per subsidy recipient) of a targeted subsidy. The x-axis represents the start time of the targeted time period. The shape of each point shows the usual travel time of the group of riders (20 or 40 mins. earlier or later than the targeted time). The value of time is assumed to be 50% of the wage.

A.9 Comparing the Welfare Model with the MVPF calculation

In this section, we compare the expression for $\frac{dW(s)}{ds}$ derived using our welfare model (Equation A3 in the appendix A1 of Hahn, Metcalfe and Tam (2023) with an expression we derive for $\frac{dW(s)}{ds}$ using the MVPF approach.

Lemma 1: If the net cost to the government is funded by a lump sum tax, then the expression for $\frac{dW(s)}{ds}$ using sufficient statistics is the same as the expression for $\frac{dW(s)}{ds}$ that we derive below for the MVPF.

Proof: A general expression for the MVPF is $\frac{x_2 - MEC\frac{dx_1}{ds}}{x_2 + s\frac{dx_2}{ds} - (p - MPC)(\frac{dx_1}{ds} + \frac{dx_2}{ds})}$ where we use the same notation introduced in section 2, and assume that this applies to N individuals. The numerator represents the willingness to pay and the denominator represents the net cost to the government. If we assume that this net cost is funded by a lump sum tax, then we can represent the difference between benefits and costs at the margin using the MVPF approach as the numerator minus the denominator in the MVPF expression. That is,

$$\frac{dW(s)}{ds} = (x_2 - MEC\frac{dx_1}{ds}) - (x_2 + s\frac{dx_2}{ds} - (p - MPC)(\frac{dx_1}{ds} + \frac{dx_2}{ds}))$$
$$= (p - MPC - MEC_1)\frac{dx_1}{ds} + (p - s - MPC)\frac{dx_2}{ds} \quad \text{(A7)}$$

The second line follows from cancelling x_2 and the third from factoring. Noting that $MPC + MEC_1 = MSC_1$ and $MPC = MSC_2$ yields $(p - MSC_1)\frac{dx_1}{ds} + (p - s - MSC_2)\frac{dx_2}{ds}$. This is the same expression as equation (A3) in Hahn, Metcalfe and Tam (2023) , except that equation is given in per capita terms. Multiplying equation (A3) in Hahn, Metcalfe and Tam (2023) by N gives the expression here, which proves Lemma 1.

In short we have shown conditions under which the MVPF calculation yields the same result for measuring the impact of a change in the subsidy as our model that uses sufficient statistics. The general result on welfare also follows (eq. 1) if we use the same assumptions.

A.10 Comparing the demand response with exogenous congestion and endogenous congestion

In this section, we argue that the demand response to a subsidy is smaller in *absolute value* for both the peak and off-peak periods with endogenous congestion than exogenous congestion.

Claim: The claim comes in two parts. 1. If $\frac{\partial x_1}{\partial s} < 0$ and $\frac{\partial x_1}{\partial e} < 0$, the response for peak demand in the endogenous congestion case is smaller in absolute value than that for the exogenous congestion case; and 2. In addition, if $\frac{\partial x_2}{\partial e} > 0$, the change in off-peak rides from a subsidy in the endogenous congestion case is also smaller in absolute value than that for the exogenous congestion case.

Consider the first part of the claim. We begin by computing the total derivative of peak travel with respect to a change in the subsidy. Starting with the peak demand function, $x_1(p, s, e(x_1))$, and differentiating with respect to s, yields $\frac{dx_1}{ds} = \frac{\partial x_1}{\partial s} + v_{11}\frac{dx_1}{ds}$, where $v_{11} = \frac{\partial x_1}{\partial e}\frac{\partial e}{\partial x_1} = v_{1e}\frac{\partial e}{\partial x_1}$. In particular, the total change ($\frac{dx_1}{ds}$) equals the sum of the direct price effect of the subsidy ($\frac{\partial x_1}{\partial s}$) , and the feedback effect from congestion, $v_{11}\frac{dx_1}{ds}$. Solving for $\frac{dx_1}{ds}$, we have $\frac{dx_1(p,s,e))}{ds} = \frac{\frac{\partial x_1}{\partial s}}{1 - v_{11}}$ To sign v_{11}, note that because $v_{1e} < 0$ and we assume $\frac{\partial e}{\partial x_1} > 0$ (that is, increases in peak demand contribute to congestion), we have $v_{11} < 0$. This implies that $|\frac{dx_1}{ds}| < |\frac{\partial x_1}{\partial s}|$.

We now prove the second part of the claim. Taking the total derivative of the off-peak demand function $x_2(p, s, e(x_1))$ with respect to s gives $\frac{dx_2}{ds} = \frac{\partial x_2}{\partial s} + \frac{\partial x_2}{\partial e}\frac{\partial e}{\partial x_1}\frac{dx_1}{ds}$. We use the expression for $\frac{dx_1}{ds}$ from above, which yields: $\frac{dx_2}{ds} = \frac{\partial x_2}{\partial s} + \frac{\partial x_2}{\partial e}\frac{\partial e}{\partial x_1}\frac{\frac{\partial x_1}{\partial s}}{1-v_{11}}$. Letting $v_{21} = \frac{\partial x_2}{\partial e}\frac{\partial e}{\partial x_1}$, we have $\frac{dx_2(p,s,e)}{ds} = \frac{\partial x_2}{\partial s} + \frac{v_{21}\frac{\partial x_1}{\partial s}}{1-v_{11}}$. Since $\frac{\partial x_2}{\partial e} > 0$, $v_{21} > 0$. This implies that the second term in the above equation is negative (provided that $\frac{dx_1}{ds} < 0$). This yields the result: $|\frac{dx_2}{ds}| < |\frac{\partial x_2}{\partial s}|$. In other words, the increase in ridership with an off peak subsidy is less with endogenous congestion than with exogenous congestion.

These results also allow us to compare the travel in peak and off-peak periods in the endogenous congestion case with the exogenous congestion case. Let x_1^o and x_2^o be the peak and off-peak travel before the subsidy. Furthermore, let x_1' and x_2' be the peak and off-peak travel after the subsidy with exogenous congestion; and $\tilde{x}_1, \tilde{x}_2$ be the peak and off-peak travel after the subsidy with endogenous congestion case. Noting $x_1' = x_1^0 + \frac{\partial x_1}{\partial s}s'$ and $\tilde{x}_1' = \tilde{x}_1^0 + \frac{dx_1}{ds}s'$, we have $x_1' < \tilde{x}_1$ because $\frac{dx_1}{ds} > \frac{\partial x_1}{\partial s}$. Using a similar argument, $x_2' > \tilde{x}_2$ because $\frac{dx_2}{ds} < \frac{\partial x_2}{\partial s}$.

A.11 Welfare analysis of Perks subsidy with different demand responses

Riders who join the Perks program may have different demand responses than the BART riders who do not. We provide a theoretical model below that allows for different demand responses for these two groups in estimating the welfare gains of a subsidy. Allowing for different price responses enables us to put bounds on what the welfare effects could be as we scale these price subsidies up to the entire BART ridership.

Consider two types of riders, one who joins the Perks program (a) and one who does not (b). There are n_a type a riders and n_b type b riders, exogenously given. Define the "average" consumption, x_1 and x_2 as, $x_k = \mu_a x_k^a + \mu_b x_k^b$, where $\mu_j = \frac{n_j}{N}$ is the share of population of type j, x_1^j, x_2^j are peak and off-peak consumption for type j for $k \in \{1,2\}$, and $N = n_a + n_b$. We consider the welfare impact of a single uniform subsidy for both type of riders. Welfare is given by the sum of utility of both types of riders, minus the transfer to BART and the subsidy cost [57].

Congestion is determined by the technology, with $e = E(n_a x_1^a + n_b x_1^b)$. The impact of the subsidy on congestion is given by $\frac{de}{ds} = \frac{\partial e}{\partial x_1}(\mu_a \frac{dx_1^a}{ds} + \mu_b \frac{dx_1^b}{ds})$. Define $MEC_1 = -(\mu_a \frac{\partial u_a}{\partial e} + \mu_b \frac{\partial u_b}{\partial e})\frac{\partial e}{\partial x_1}$ that represents the the marginal external cost of an increase in total peak rides, and substituting the expressions into the marginal welfare change expression yields:

$$\frac{dW}{ds} = n_a(p - c - MEC_1)\frac{dx_1^a}{ds} + n_a(p - c - s)\frac{dx_2^a}{ds} + n_b(p - c - MEC_1)(\frac{dx_1^b}{ds}) + n_b(p - c - s)\frac{dx_2^b}{ds}. \quad \text{(A8)}$$

where c is the marginal cost of the cost function $c = C_1(Nx_1, Nx_2) = C_2(Nx_1, Nx_2) = MPC$.

The first two terms represent the change in welfare associated with a change in type a demand in the peak and off-peak markets, and the second two terms represent the change in

[57] i.e. $W(s) = n_a V_a(p,s,e) + n_b V_b(p,s,e) - \min C(Nx_1, Nx_2) - Np(x_1 + x_2) - n_a x_1^a s - n_b x_1^b s$, where $V_a(p,s,e)$ and $V_b(p,s,e)$ are the indirect utility functions

welfare associated with a change in type b demand in the peak and off-peak markets. Thus, the average per capita welfare change from a change depends on the demand characteristics of both groups and their respective sizes. Equation (A8) can be seen as a generalization of equation (1) in the text, which measures the welfare change for one group that received the subsidy. Here, we are measuring the welfare change for two groups.

We can simplify equation (A8) if we assume that the demand response in both markets is similar for those consumers who do not join Perks. Specifically, assume that type b riders have a demand response that is proportional to type a in both markets. Specifically, we let $\frac{dx_j^b}{ds} = k\frac{dx_j^a}{ds}$ where k is between 0 and 1, for $j \in \{1,2\}$. The total welfare change per N population is then:

$$\frac{1}{N}(W(s') - W(0)) = \mu((p - c - MEC_1)\frac{dx_1^a}{ds}s' + (p - c - \frac{1}{2}s')\frac{dx_2^a}{ds}s') \tag{A9}$$

where $\mu = \frac{n_a + n_b k}{N}$ is between 0 and 1. The formula here for two groups is very similar to the formula for welfare change for one group in the text (see equation 1), with the difference being we have added the factor μ. Note that when the demand responses of the two groups are identical (k=1), the two formulas are the same.

A.12 Modelling the subsidy with a discrete choice model

In this subsection, we use a discrete choice model to estimate and evaluate the welfare impact of the subsidy. We find that the welfare impact of the subsidy is still likely to be positive, but smaller in magnitude that value we estimate in the main text. The discrete choice model we use for estimation (conditional logit) is often used in the transport literature. The calculation here is meant to be illustrative.

A.12.1 Consumer demand model

We assume BART riders choose among three travel options - peak or off-peak period on BART, and an outside option e.g. driving, denoted as $j = 0, 1, 2$ respectively. The utility function for rider i for option j is $u_{ij} = v_{ij} + \epsilon_{ij}$. v_{ij} is the representative utility, and ϵ_{ij} is unobserved random error term. We assume the consumer has a quasi-linear representative utility, $v_{i1} = Z_i - p + \zeta + \phi(e)$, $v_{i2} = Z_i - p + s + \zeta$ and $v_{i0} = Z_i - p_d + \zeta_0$, where Z_i is the exogenous income, p is the price of a BART journey, s is the amount of subsidy for a off-peak BART journey. ζ is the utility the consumer enjoy when completing the BART trip, and ζ_0 is the utility when completing the trip with the outside option, with price p_d. The congestion during peak time on BART generates a cost of $\phi(e) < 0$, which depends on the level of congestion e, with $\phi'(e) < 0$.

We further specify how congestion affects utility by decomposing its effect into that from delay and crowding $\phi(e) = -VOT * t(e) + \gamma e$. The first term assumes the effect on travel time $t(e)$ on utility is proportional to the consumer's value of time (VOT). The second term shows how a higher level of congestion affect utility, and assumes that crowding is proportional to the level of congestion e, with $\gamma < 0$.

Consumer i chooses the option that gives the highest utility. Following the literature (Train, 2009), we assume that ϵ_{ij} follows a joint distribution with density function $f(\epsilon_{i1}, \epsilon_{i2}, \epsilon_{i0})$ - specifically, for a logit demand function, we assume that ϵ_{ij} are all identical and independently distributed in the Gumbel disbritubion. [58] The probability that the consumer i selects option j is $P_{ij} = \frac{e^{v_{ij}}}{\sum_j e^{v_{ij}}}$, with the total demand $P_j = \sum_i P_{ij}$.

A.12.2 Supply, Congestion technology and Welfare

We assume that the frequency and routes of the trains are exogenous. BART has the the following profit function: $\pi = p(P_1 + P_2) - MC * (P_1 + P_2)$, where MC is the marginal private cost of an additional ride.[59] The net loss (or gain) is financed through lump-sum transfer implicitly to the consumer income Z_i. The number of rides in the equilibrium depends on the equilibrium level of congestion e.

The effect of aggregate peak demand P_1 on congestion e is given by $e = \Phi(P_1(p,s,e))$ - in the case of public transit, with the number and size of trains fixed, we define Φ as $e = \frac{P_1(p,s,e)}{ka}$ where k is the number of train cars and a is the area of the car. In equilibrium, $e^* = \frac{P_1(p,s,e^*)}{ka}$.

Welfare is the sum of consumer expected utility for all i, netting out the cost of the subsidy, plus producer profit.[60] With no fixed cost of production, the welfare W (measured as dollars per day) is $W = \sum_i (ln(exp^{v_{i1}} + exp^{v_{i2}} + exp^{v_{i0}})) + (p - MC)(P_1 + P_2) - sP_2$.

The marginal welfare effect of increasing the subsidy, s, is $\frac{dW}{ds} = \left(\frac{d\phi_k}{ds} \sum_i \frac{e^{v_{i1}}}{\sum_j exp^{v_{ij}}} + P_2\right) - P_2 - s\frac{dP_2}{ds} + (p - MC)(\frac{dP_1}{ds} + \frac{dP_2}{ds})$, where the first term in the first bracket is the change in the externality with respect to the subsidy (multiplied by quantity of peak travel), and the second term in the first bracket is the monetary gain for those who were travelling during off-peak. The welfare change for a subsidy s' can be calculated similarly to that in the main text by integrating $\frac{dW}{ds}$ from 0 to s'. In principle, P_1, P_2, $\frac{dP_1}{ds}$ and $\frac{dP_2}{ds}$ all varies with s. For simplicity, we approximate the welfare change by assuming all terms in the preceding equation are constant other than s itself. The change is given by: $\Delta W \approx (\phi'(e)\frac{\partial e}{\partial P_1} P_1 \frac{dP_1}{ds})s' - \frac{1}{2}s'^2\frac{dP_2}{ds} + (p - MC)(\frac{dP_1}{ds} + \frac{dP_2}{ds})s'$, which is analogous to the welfare formula in the main text if we rewrite $-\phi_i'(e)\frac{\partial e}{\partial P_{1,i}} P_{1,i}$ as MEC_i.

A.12.3 Empirical implementation

We estimate a conditional logit model of a rider's choice between peak hour, off-peak hour and no trip on BART, using the specification $u_{j,it} = \beta_j + \beta_1 p_{j,it} + \epsilon_{jt}$, where $p_{j,it}$ is the subsidy-

[58]The density for each ϵ_{ij} is $f(\epsilon_{ij}) = e^{-\epsilon_{ij}} - e^{e^{-\epsilon_{ij}}}$

[59]We could have added a fixed cost component, e.g. $F(n,a)$ that represents fixed cost and could be specified by a regulator. This would not change the analysis, so we omit it for simplicity.

[60]The expected utility of agent i is $E(\max\{u_{i1}, u_{i2}, u_{i0}\}) = ln(exp^{v_{i1}} + exp^{v_{i2}} + exp^{v_{i0}})$

inclusive price of choice j for person i in period t. The price for the peak hour is assumed to be $3.99, for the off-peak hour it is $3.99 plus the $0.0126 subsidy per mile with an average distance of 14.6 miles. The price for the outside option is assumed to be the price of driving, calculated by the distance between the entry and exit station times the average cost of 53.5 cents per miles (IRS standard rate). Table A8 presents the estimates of β_1, including a specification that controls for date fixed effects and the average daily peak trip 2 weeks before the sample period. We are not able to include congestion or travel time in the estimation given our data, as we do not have information on the date-specific travel time or congestion information for the outside option (e.g., driving). The implied estimates of $\frac{dP_1}{ds}$ and $\frac{dP_2}{ds}$ from

Table A8: Estimates of discrete choice (Conditional logit) model

	(1)	(2)	(3)
Price	-1.122***	-1.377***	-1.412***
	(0.0344)	(0.0446)	(0.0889)
Control		Y	Y
Date FE			Y
N	2,080,896	1773063	1773063

Notes: The table presents estimates of the conditional logit model for the travel choices outcomes: peak-hour, off-peak hour and no BART trips. The sample includes morning trips, the same as that in Table 3. Col (2)-(3) includes average daily peak trip 2 weeks before the sample period as control. Standard errors are in parentheses.

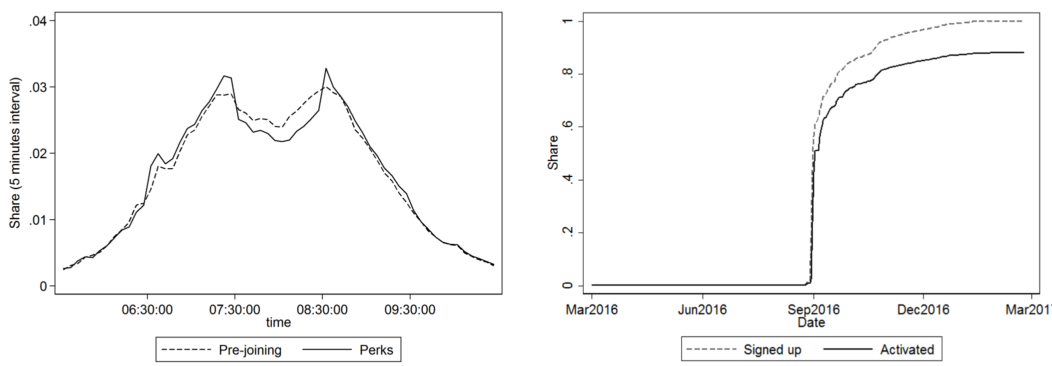
column 3 is -0.22 and 0.22, respectively. We then use the value for the MEC described in the earlier section, assuming that VOT = 0.5 x wage, and apply it in the welfare equation above together with the demand parameters estimated from the conditional logit model. Using these demand estimates, the welfare effect of a $0.25 change in the subsidy per trip is 0.0058 per person per day.

The welfare effect using the discrete choice model is smaller than that presented in Section 5.2. The smaller estimate likely arises because the discrete choice estimates yield an almost 1 for 1 substitution between the peak hour and off-peak hour trips given our specification and the data. That is, the net impact of the subsidy on total trips is close to zero. When the price is above marginal social cost, a net increase in BART usage with subsidies contributes positively to the welfare change. The estimates from the conditional logit model may therefore under-estimate this gain from a net change in ridership.

B Additional Figures and Tables

Figure A4: Natural experiment - Participants characteristics and enrollment

(**a**) Travelling time of participants (**b**) Program enrollment



Note: Panel (a): The figure plots the density of trip by its starting time for participants in the first experiment. Sample includes trips taken with start time between 5.30-10.30am on Mon-Fri, from Mar 2016-Feb 2017 by 17,788 participants with 2,099,895 trips. Bin width is 5 minutes interval.; Panel (b): The graph plots the cumulative share of users signed up, and who had activated.

Table A9: Natural experiment - Long term effect DiD

	Outcome: Share of all trip in morning	
	Peak Hour	Bonus Hour
	(1)	(2)
Treated users × Perks period	-0.0211***	0.0255***
	(0.00133)	(0.000963)
Treated users × Post-Perks period	-0.00946***	0.00135
	(0.000909)	(0.000927)
Treated users	-0.00297***	0.0744***
	(0.000554)	(0.000644)
Observations	636	636

Note: Daily observation for Perks users and the rest of network. Sample includes 1 Apr 2016-30 Jun 2017, weekdays excluding public holidays. Robust standard error reported in parenthesis.

Table A10: Welfare and congestion impacts for the optimal subsidy

Parameter	Exogenous cong.	Endogenous cong.	% Change
Welfare change associated with			
Peak travel change	-0.028	-0.0239	-14.7
Off-peak travel change	0.1448	0.1362	-5.9
Total Welfare change	0.1168	0.1123	-3.8
Congestion	0.6621	0.6874	3.8
Optimal subsidy	1.875	1.896	1.1

Notes: The table calculates the optimal subsidy using equation 2.2 in section 5.2, for the natural experiment. It uses parameters described in Table A11 for the baseline case.

Table A11: Additional description for parameters for welfare calculation 1

Parameter and description	Value	Source
s' ($ / trip)	$0.25	Average subsidy per off-peak trip in the Perks experiment.
$\frac{dx_2^*}{ds}$ (change in quantity/ $)	0.0664 (exogenous) 0.0625 (endogenous)	1st Perks experiment ($\frac{dx_2}{ds} = -\epsilon_{22}\frac{x_2}{p}$, $\epsilon_{22} = -0.86$; see Section 3)
$\frac{dx_1^*}{ds}$ (change in quantity/ $)	-0.0194 (exogenous) -0.0166 (endogenous)	1st Perks experiment ($\frac{dx_1}{ds} = -\epsilon_{12}\frac{x_1}{p}$, $\epsilon_{12} = 0.44$; see Section 3)
v_{11} (feedback effect from congestion on peak-ride)	-.172	Detail of estimation in appendix A2
v_{21} (feedback effect from congestion on off-peak ride)	0.237	Detail of estimation in appendix A2
MEC_1 (the external cost is measured in $/peak-trip)	-0.537	See appendix A1 for formula.
t (In-vehicle time)	37 (minutes)	Calculated form BART journey data (gate entry to gate exit)
$Tm'(d)$ (Time multiplier)	0.11 (minutes/passengers per sq m)	Haywood and Koning (2015)
Value of time (VOT)	= 0.116 per minute (= $7 per hour)	Small (2012) - value of time as half of wage. minimum wage in SF in 2017 = $14
d (density in car, number of passengers per sq m)	1.13 (= 77.9/68.93)	Weighted average number of passengers per car/car size (weighted by number of passengers in car)
n (average passengers per car in peak hour)	41.389	Calculated form BART crowding data
a (Car size, sq m)	68.93 (= 70 (feet)*10.6(feet)* 0.3048^2)	We take type B car as representative car, where the length is 70', with width 10.6'
$\frac{de}{ds}$ (change in crowding)	-0.24 (exogenous) -020 (endogenous)	1st Perks experiment.
p (average fare per journey)	$3.99	Average off-peak fare per journey paid by riders in the 1st experiment.
$\frac{dC}{dNx_1}$, $\frac{dC}{dNx_2}$ ($/journey)	$1.89	Calculated using cost and ridership information from BART. See Appendix A5.
N	271,341(people)	Average number of morning peak trip in BART per day (47,756) / peak-trip per rider per days (estimated from data of participants in the first Perks experiment) (0.176)

Notes: The table summarizes the key parameters used in the welfare calculation in the natural experiment.

Table A12: Additional description for parameters for MVPF calculation of experiment 1

Description	Value	Source
Inframarginal benefit (x_2)	0.307	1st experiment
Congestion benefit ($-\frac{dx_1}{ds} * MEC_1$)	0.0258	1st experiment; see Appendix A.3
Change in direct subsidy cost (x_2)	0.307	1st experiment
Change in transfer to BART ($-(p - MPC)(\frac{dx_1}{ds} + \frac{dx_2}{ds})$)	-0.0987	1st experiment estimates
Marginal private cost (MPC)	1.89	See Table A11
Price per trip (p)	\$3.99	See Table A11
Marginal external cost (peak period) (MEC_1)	1.33	See Table A11

Figure A5: BART - the natural experiment

(a) BART system map

(b) Example of ads to motivate sign-up in the natural experiment

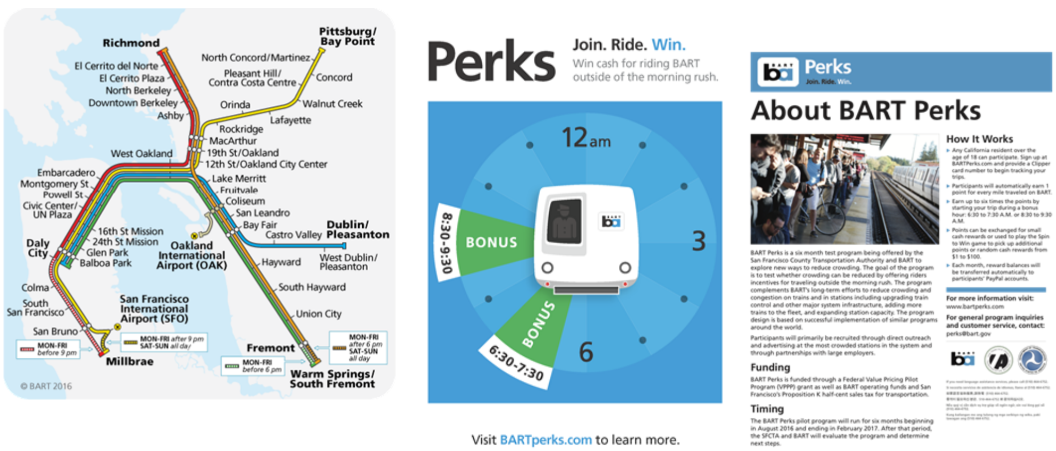
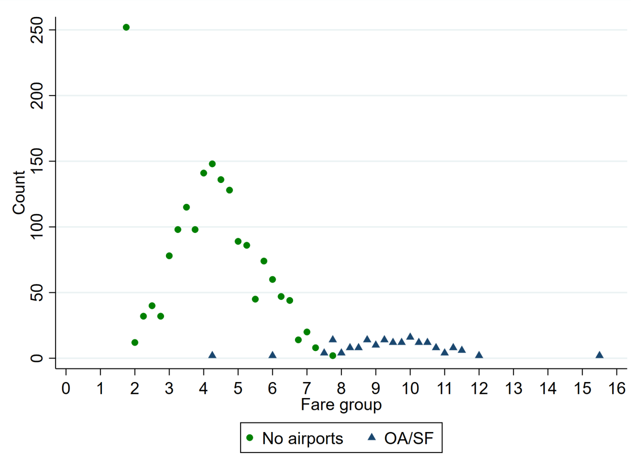


Figure A6: Distribution of fare for each route on BART



Notes: The figure plots the distribution of fare for each route on BART. The diamond markers represent the fare distribution for routes excluding those to or from the San Francisco International Airport or Oakland International Airport. The triangle markers represent the distribution for routes that are to or from the San Francisco International Airport or Oakland International Airport. (2017 prices)

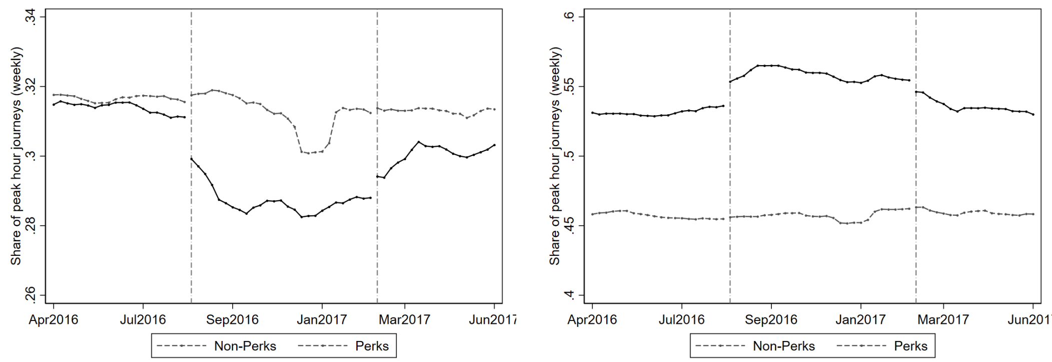
Table A13: Parameters for welfare analysis of the field experiment

				VOT (% of wage)			
Time				75%	50%	75%	50%
Usual (t_1)	Target (t_2)	$\frac{dx_1}{ds}$	$\frac{dx_2}{ds}$	MEC1, 2	MEC1,2	MVPF	MVPF
0700	0620	-0.02	0.01	1.13, 0.30	0.05, -0.50	1.28	1.06
	0640		0.01	1.15, 0.58	0.06, -0.31	1.17	1.04
	0720		0.00	1.51, 1.90	0.31, 0.56	1.13	1.03
	0740		0.02	1.55, 1.97	0.34, 0.62	1.18	0.96
0740	0700	-0.02	0.00	1.31, 0.45	0.17, -0.40	1.33	1.06
	0720		0.00	1.86, 1.72	0.54, 0.45	1.18	1.05
	0800		0.03	2.15, 2.38	0.73, 0.89	0.81	0.91
	0820		0.01	2.15, 2.71	0.73, 1.11	1.37	1.11
0800	0720	-0.04	0.00	1.67, 1.23	0.41, 0.12	1.50	1.12
	0740		0.02	1.82, 1.61	0.51, 0.38	1.23	1.07
	0820		0.01	2.04, 2.46	0.66, 0.94	1.34	1.10
	0840		0.03	2.07, 1.91	0.68, 0.57	1.48	1.18
0820	0740	-0.03	0.02	2.08, 1.53	0.69, 0.32	1.35	1.14
	0800		0.03	2.52, 2.09	0.98, 0.69	1.15	1.07
	0840		0.03	2.39, 2.03	0.89, 0.65	1.22	1.10
	0900		0.01	2.01, 1.06	0.64, 0.01	1.72	1.28
0840	0800	-0.01	0.03	2.37, 2.39	0.88, 0.89	0.45	0.79
	0820		0.01	1.99, 2.14	0.63, 0.72	1.07	1.02
	0900		0.01	1.76, 1.20	0.48, 0.10	1.10	1.06
0900	0820	-0.02	0.01	2.01, 2.82	0.64, 1.18	1.19	1.04
	0840		0.03	2.55, 3.10	1.00, 1.37	0.71	0.85
	0940		0.01	1.28, 0.19	0.16, -0.57	1.34	1.09
0920	0840	-0.02	0.03	1.04, 1.72	0.00, 0.45	0.49	0.72
	0900		0.01	1.56, 1.94	0.34, 0.59	1.06	1.00
	0940		0.01	0.93, 0.28	-0.08, -0.52	1.11	1.01
	1000		0.01	0.77, -0.16	-0.19, -0.81	1.32	1.12
0940	1000	-0.01	0.01	-0.02, -0.27	-0.71, -0.88	1.08	1.14
	1020		0.05	0.18, -0.29	-0.58, -0.89	0.58	-0.06
1000	1020	-0.04	0.05	-0.22, -0.39	-0.85, -0.96	1.30	1.42
	1040		0.01	-0.32, -0.69	-0.92, -1.16	0.92	0.69

Note: The table presents the parameter estimates for experiment 2. Col. (1) indicates the usual travel time of the subsidized group, and col. (2) indicates the subsidized time. Col. (3) and (4) are the demand response estimated from experiment 2. Col. (5) indicate the marginal external cost of the usual and subsidized travel time, when value of time is 75% of wage. Col. (6) indicate the marginal external cost for the usual and subsidized travel time, when the value of time is 50% of wage. Col. (7) and (8) indicate the calculated MVPF of the subsidy. The *MEC* data are based on train load data provide by BART.

Figure A7: Long term - Share of peak hour journeys

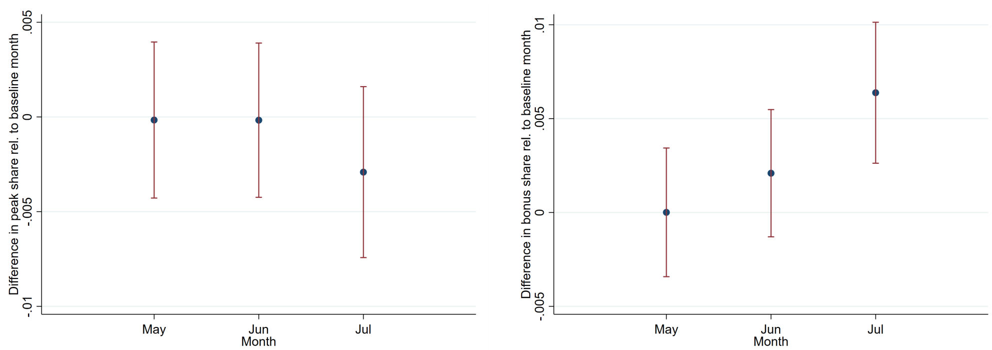
(a) Peak hour **(b)** Off-peak shoulder hour



Notes: The graph plot trend of weekly average peak hour (panel (a)) and off-peak shoulder hour (panel (b)) journey share in the morning (6:30am-7:30am and 8:30am-9:30am among all journeys in 5:30am-10am), for weekdays. Sample include 1 Apr 2016-30 Jun 2017, excluding public holidays.

Figure A8: Test of pre-trend

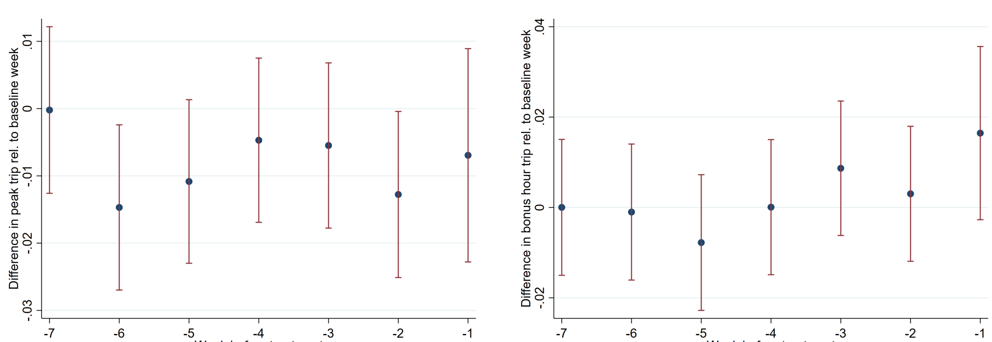
(a) Share of peak hour **(b)** Share of off-peak shoulder hour



Notes: The figure plots the test of pre-trend with respect to Figure 3. It estimate a variant of eq. 2, $y = \beta_1 * Treated * May + \beta_2 * Treated * Jun + \beta_3 * Treated * Jul + \eta_g + \gamma_t + \epsilon_{gt}$. y_{gt} is the share of peak hour trips (a) or off-peak shoulder hour trips (b) from 5.30-10.30am on date t for group g. May, Jun and Jul are indicators for May, June and July 2016 respectively. Sample includes dates from Apr 2016 to 23 Jul 2016. β_1, β_2 and β_3 represent respectively the change of difference between the Perks participants and the rest of the network, from April to the respective month.

Figure A9: Test of pre-trend

(a) Peak hour trips **(b)** Off-peak shoulder hour



Note: The figure plots the test of pre-trend - it estimates a variant of eq. 3, $y_{it} = \sum_k \beta_k * Treated * D_k + \sigma_i + \mu_t + \epsilon_{it}$. y_{it} is the number of trip during peak hour (a) or bonus hour (b). D_k is an indicator of each week before the treatment in 23 Aug. Sample includes dates from 2016 week 27 to week 34. β_k represent respectively the change of difference between the early enrolled participants and the late enrolled participants, from week 27 to the respective week of the year. Baseline/omitted week is 2016 week 27.

Table A14: Route level difference-in-differences estimate

	Outcome: Share of all trip in morning	
	Peak Hour	off-peak shoulder hour
	(1)	(2)
Treated users × Perks Period	-0.0332***	0.0225***
	(0.00428)	(0.00416)
Observations	97457	95432
Route-Perks user FE	x	x

Notes: The table reports estimation of equation 2 at route level at weekly level. The outcome is weekly observation for each route in the BART network for Perks users or the rest of network. Sample includes 1 Aprl 2016 to 28 Feb 2017, all weekdays excluding specific public holidays. Standard error clustered at route-group level. All columns control for week fixed effects, and route × Perks user fixed effects.

Table A15: Average treatment effects by time period: Early vs. late comparison with alternative definition

	OLS (late enrolled users: Oct 15 to Nov 5) Outcome: Number of trips daily		
	Peak	Off-peak shoulder (Either)	Total (5.30-10.30am)
	(1)	(2)	(3)
Early enrolled users (β_1)	-0.00817**	0.0125***	0.00105
× Perks Period	(0.00355)	(0.00428)	(0.00497)
Observations	582017	582017	582017
Sample mean (before Perks)	0.18	0.31	0.57

Note: The sample includes daily trips from June 1 to October 15, 2016, excluding August 23 to September 2. The Perks period is defined as dates after September 2. Early enrollers are defined as those who enrolled between August 23 and September 2. The control group includes those who enrolled between October 15 and November 5, 2016. All columns controlled for date fixed effects and user fixed effects. Standard errors are clustered at user-week level. The sample means are for late enrolled users in periods before the Perks period.

Table A16: Route level difference-in-differences estimate - by direction of travel towards Montgomery Street and Embarcadero

	Outcome: Share of all trip in morning			
	Peak hour		off-peak shoulder hour	
	(1)	(2)	(3)	(4)
Treated useres × Perks period	-0.0213***	-0.00323	0.0216***	-0.00108
	(0.00485)	(0.00707)	(0.00531)	(0.00744)
N	4793	1416	4791	1421
Route-Perks user FE and Week FE	x	x	x	x
Sample	Westbound	Eastbound	Westbound	Eastbound

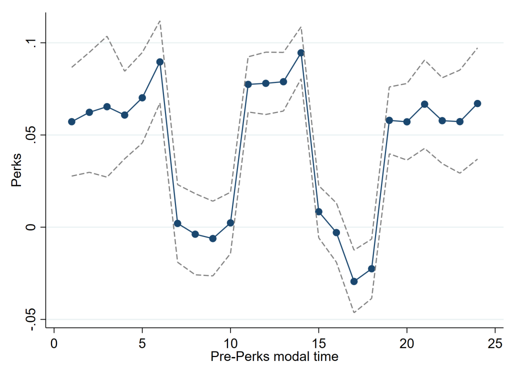
Notes: The table reports estimation of equation 2 at route level at weekly level. The outcome is weekly observation for each route in the BART network for Perks users or the rest of network. Sample includes 1 April 2016 to 28 Feb 2017, all weekdays excluding specific public holidays. Standard error clustered at route-group level. Sample includes routes exiting at Montgomery Street and Embarcadero. Col. (1) and (3) report the estimate for the sample with origin travelling westbound towards Montgomery Street and Embarcadero, and Col. (2) and (4) for the sample with origin travelling eastbound towards Montgomery Street and Embarcadero.

Table A17: Treatment effects on early or late off-peak shoulder hour: staggered activation

	Outcome: Number of daily trips	
	Off-peak shoulder hour (early)	Off-peak shoulder hour (late)
	(1)	(2)
Active status level	0.0175***	0.00992***
	(0.00261)	(0.00289)
Observations	821643	821643
Baseline mean	0.291	0.254
Implied own-price elasticity	-0.653	-0.424

Note: Sample includes dates from 1 Jul-6 Nov 2016. Standard errors clustered by users and dates. Early off-peak shoulder hour refers to trip between 6.30-7.30am; late off-peak shoulder hour refers to trip between 8.30-9.30am. All columns control for date and user fixed effects.

Figure A10: Implied treatment effect on off-peak shoulder hour travel by modal travel time in pre-treatment period



Notes: The figure plots the treatment effect on off-peak shoulder hour travel implied by the estimation of equation (4) with additional interaction of the term $PerksActive_{it}$ with indicators of the modal travel time of user i in period before perks. The modal travel time are defined by 15 minutes intervals.

Figure A11: Example of ads to motivate sign-up in the field experiment

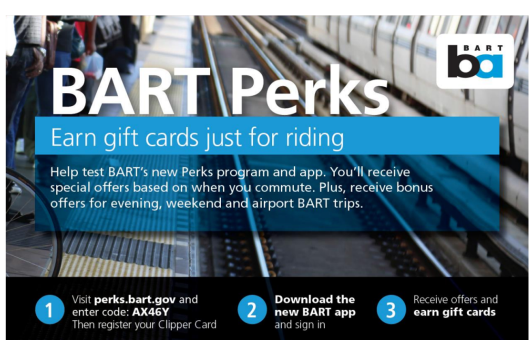
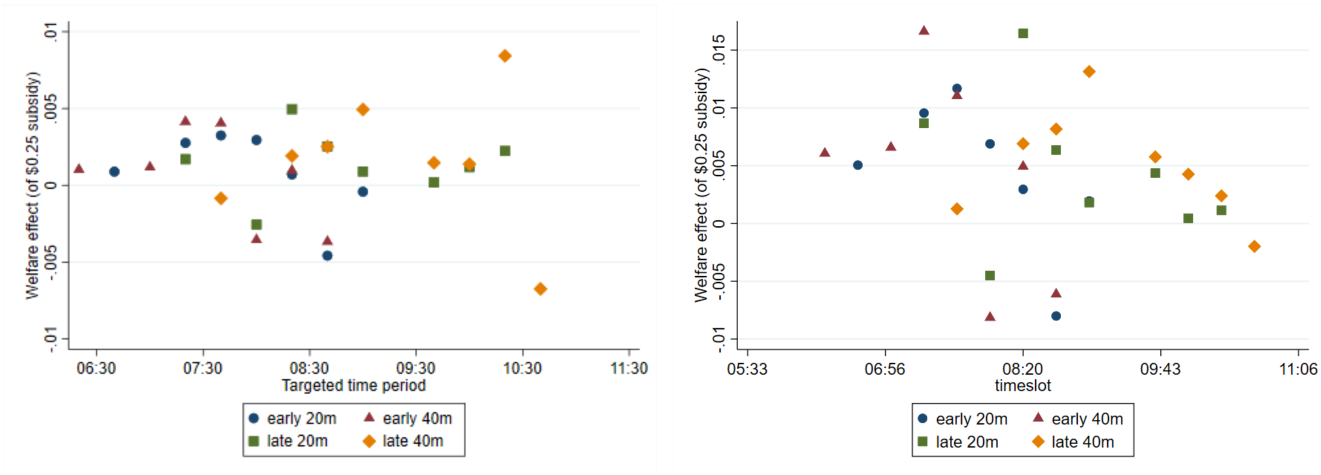


Figure A12: Welfare effect of subsidy in the field experiment – by subsidized time

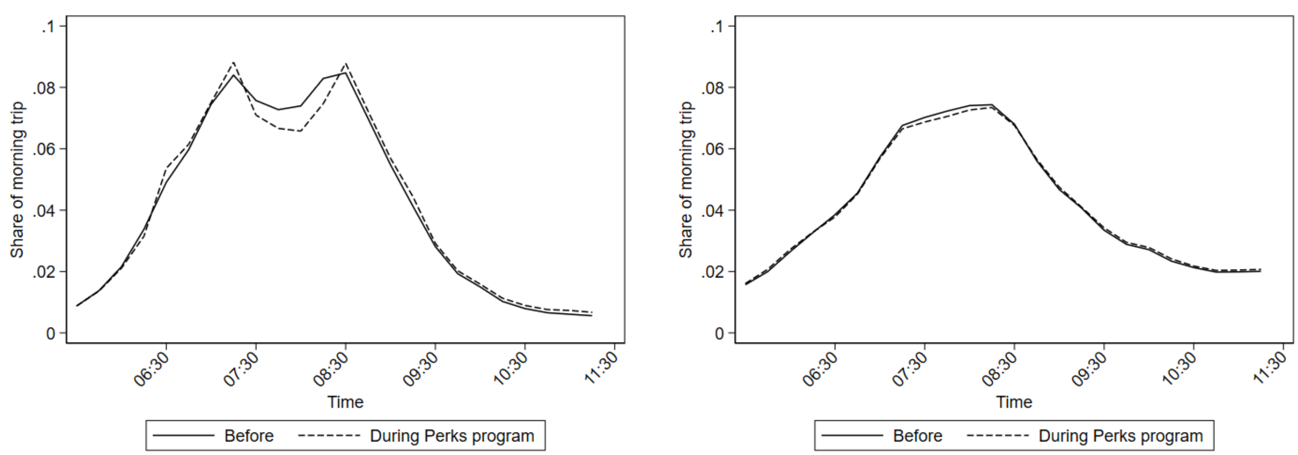
(**a**) VOT - 50% of wages (**b**) VOT - 75% of wages



Note: This figure shows that welfare change of a targeted subsidy can be positive or negative. The figure plots the welfare effect of the subsidy per rider using equation 1". The x-axis represents the beginning time of the targeted time period. The shape of each point shows the usual travel time of the group of riders, which is 20 or 40 minutes earlier or later than the targeted time. The value of time is assumed to be 50% of the wage.

Figure A13: Comparing Perks program participants with the rest of the BART network

(**a**) Perks participants (**b**) Rest of the BART network



Note: The figure plots the average daily share of trips taken by the participants in the first experiment in each time interval from 530am-1130am. It also plots on the same figure the average share of daily trips at each time interval taken by the rest of the BART users. The sample for Before Perks period includes April 1, 2016 to August 22, 2017 for all weekdays. The sample for Perks peiod includes August 23, 2016 to February 28, 2017 for all weekdays.

Table A18: Alternative DiD estimator - staggered activation for the natural experiment

	Outcome: Number of trips daily			
	Peak	Off-peak shoulder (Either)	Other	Total
Treatment effect	-.0158	.00540	-.00495	-.0153
(Instantaneous)	(0.0038)	(0.00513)	(0.00317)	(0.00523)
ATT (weekly data)	-0.0137	0.00590	-0.0074	-0.0155
	(0.004)	(0.0052)	(0.0029)	(0.0054)
Observations	1426285	1426285	1426285	1426285

Note: The table present the alternative DiD estimator for the treatment effect of subsidy in experiment 1. The first row report the estimate of the instantaneous treatment effect of activation in experiment 1 following (De Chaisemartin and d'Haultfoeuille, 2020). Standard errors are calculated with 50 bootstrap replication. The second row report the average treatment effect following Callaway and Sant'Anna (2021) estimated with daily data averaged at weekly level, using the "not yet treated" as control group.

Table A19: Summary statistics - field experiment

Time Period	6.20	6.40	7.00	7.20	7.40	8.00	8.20
Average no. of trip (before exp.)	0.020	0.028	0.064	0.100	0.106	0.107	0.083
Share of user-date received experiment offer							
Shift early	0.022	0.028	0.041	0.045	0.037	0.026	0.009
Shift late offers				0.003	0.012	0.024	0.03
Time Period	8.40	9.00	9.20	9.40	10.00	10.20	10.40
Average no. of trip (before exp.)	0.054	0.030	0.017	0.012	0.008	0.005	0.003
Share of user-date received experiment offer							
Shift early offers	0.001	0.000	0.000				
Shift late offers	0.021	0.024	0.026	0.017	0.009	0.005	0.004
Observations	196083						

Notes: The table report summary statistics for experiment 2. In each of the top and bottom panel, row (1) report the participants average daily number of trip by 20 minutes time interval in period before the experiment. Rows (2) and (3) report on average during the experiment period, the share of user-date received offer by the incentivized time period in the morning. Rows. (2) (and (3)) report the average share of user-date received an offer that the incentivized offer is earlier (later) than their usual travel time.

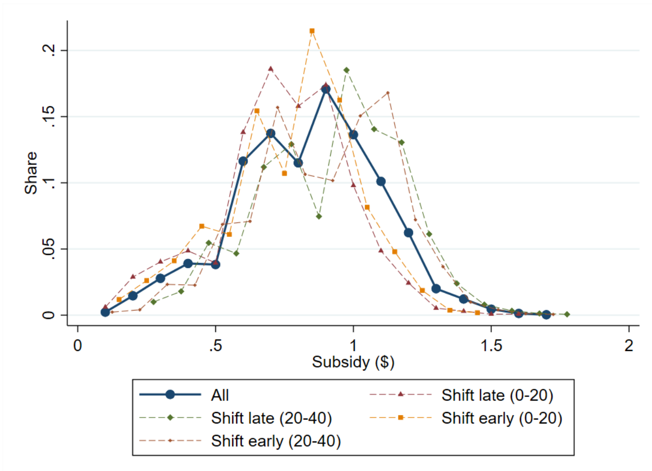
Table A20: Journey characteristics

	(1) All BART	(2) Perks pre	(3) Early pre	(4) Late pre
Distance (mile)	14.60	15.22	15.07	13.77
	(9.05)	(9.76)	(9.48)	(9.33)
Fare ($)	3.89	4.05	4.04	3.90
	(1.31)	(1.30)	(1.29)	(1.26)
Duration (s)	NA	2226.83	2216.75	2071.28
	NA	(1094.26)	(1073.71)	(1082.17)
N		23,770	13,481	2,225

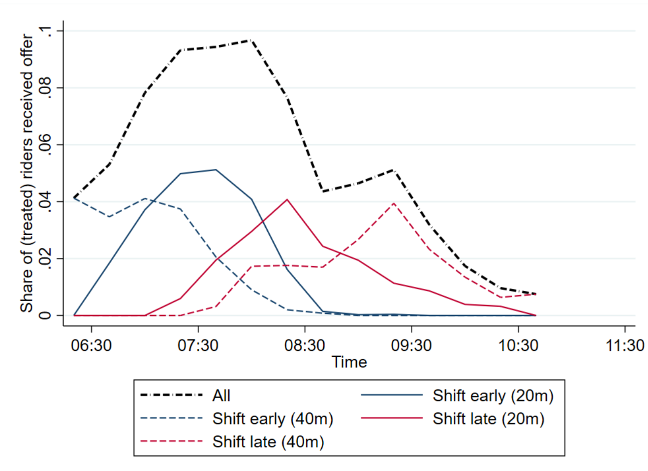
Notes: The table report the average journey characteristics for all BART morning trips, and for the Perks sample. Column (1) report the journey characteristics for all BART trips between Jun-Aug 2016. Columns (2) present the journey characteristics for the Perks sample for the period between Jun-Aug 2016. Column (3) and (4) present the characteristics for the Early and late enrolled perks riders, respectively during Jun-Aug 2016.

Figure A14: Subsidy in experiment

(a) Amount

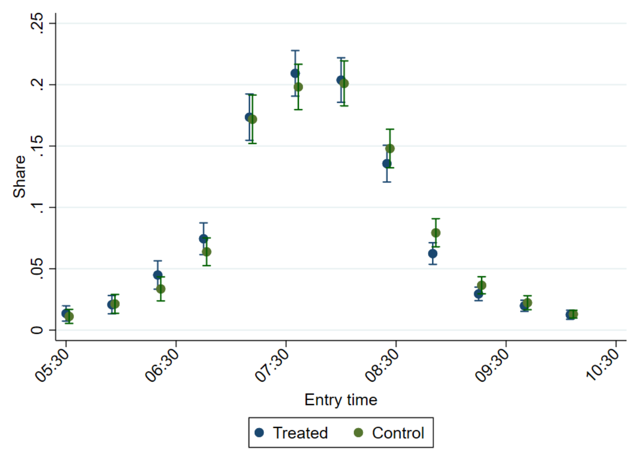


(b) Targetted time



Notes: Panel (a): The figure plots the distribution of subsidy amount in experiment 2 for the morning subsidies. The y-axis represent the share of subsidy given with the amount (in bin width of 0.1$), among all subsidies in the respective type; Panel (b): This figure plot the share of user-date under each of the types of offers, during the program period among the treated users (a_j).

Figure A15: Comparison of treated and control group before experiment



Notes: The figure plots the share of entry time for the trip among all trips between 5.30am-10.30am, for the treated and control group respectively before the experiment. It also plot the 95% confidence interval for the share clustering at user level.

Table A21: Decomposition of the staggering enrollment specification

	Rush Hour	Bonus Hour	Total	(Weight)
Overall	-0.007	0.024	0.014	
	(0.001)	(.0016969)	(0.002)	
Timing groups	-0.008	0.025	0.014	0.783
Never v timing	-0.003	0.021	0.014	0.217

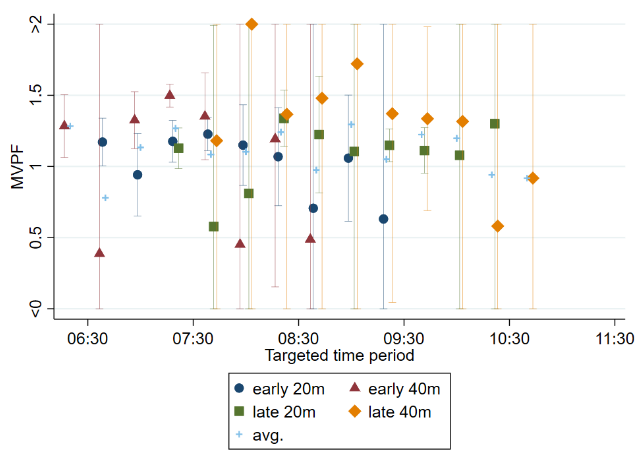
Notes: the table present the Goodman-Bacon decomposition for the staggering enrollment specification. The first row present the overall DiD estimates on a fully balanced sub-sample. The second and third row present the estimates by decomposition. The fourth column present the weights.

Table A22: Comparison of treatment and control group for experiment

	(1) Distance	(2) Fare
Treated	1155.9	0.0771
	(737.3)	(0.0648)
N	145066	149933
Mean	26459.42	4.40

Notes: The table presents a regression of the outcomes on a treatment indicator for the field experiment, for the period three months before the experiment.

Figure A16: MVPF of subsidy in experiment 2 – by subsidized time (VOT - 75% of wages)

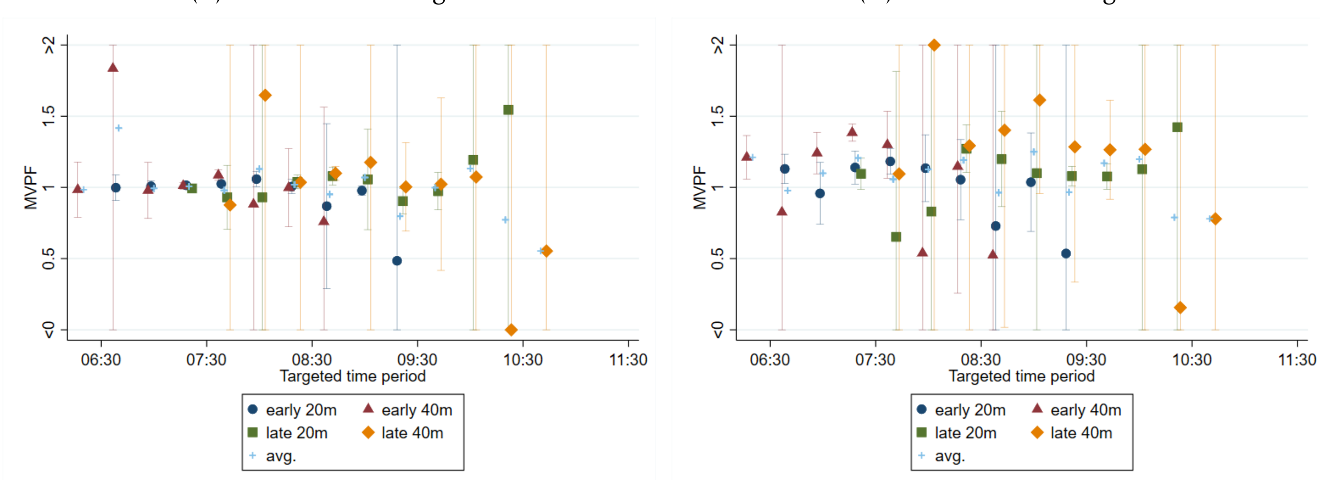


Note: This figure shows that MVPF of a targeted subsidy. p is assumed to 3.99 and does not differ across time period/subsidy groups; similarly for $mpc = 1.89$. ϕ_1 and ϕ_2 are the MEC for time period 1 and 2 (usual and targeted time) respectively. The x-axis represents the beginning time of the targeted time period. The shape of each point shows the usual travel time of the group of riders, which is 20 or 40 minutes earlier or later than the targeted time. The value of time is assumed to be 50% of the wage.

Figure A17: MVPF of subsidy in experiment 2 - including mode shifting

(a) VOT - 50% of wages

(b) VOT - 75% of wages



Note: This figure shows the MVPF of a targeted subsidy. See text for details. The x-axis represents the beginning time of the targeted time period. The shape of each point shows the usual travel time of the group of riders, which is 20 or 40 minutes earlier or later than the targeted time. The value of time is assumed to be 50% of the wage in panel (a) and 75% of the wage in panel (b).