Using Big Data to Estimate Consumer Surplus: The Case of Uber

Peter Cohen, Robert Hahn, Jonathan Hall,
Steven Levitt, and Robert Metcalfe[1]

August 30, 2016

Abstract

Estimating consumer surplus is challenging because it requires identification of the entire demand curve. We rely on Uber's "surge" pricing algorithm and the richness of its individual level data to first estimate demand elasticities at several points along the demand curve. We then use these elasticity estimates to estimate consumer surplus. Using almost 50 million individual-level observations and a regression discontinuity design, we estimate that in 2015 the UberX service generated about $2.9 billion in consumer surplus in the four U.S. cities included in our analysis. For each dollar spent by consumers, about $1.60 of consumer surplus is generated. Back-of-the-envelope calculations suggest that the overall consumer surplus generated by the UberX service in the United States in 2015 was $6.8 billion.

Keywords: consumer surplus, demand estimation, discontinuity design, peer-to-peer markets, Uber.
JEL: D1, D2, D4, L1.

[1] We are grateful to Josh Angrist, Keith Chen, Joseph Doyle, Hank Farber, Alan Krueger, Greg Lewis, Jonathan Meer, and Glen Weyl for helpful comments and discussions. We are also grateful to Mattie Toma for excellent research assistance. Affiliations: Cohen: Uber; Hahn: Smith School and Institute for New Economic Thinking, University of Oxford; Hall: Uber Levitt: University of Chicago Department of Economics and National Bureau of Economic Research; Metcalfe: University of Chicago Department of Economics and Becker Friedman Institute.

1. Introduction

For over 250 years, economists have recognized the importance of consumer surplus when making welfare calculations.[2] Consumer surplus (and the closely related concepts of equivalent variation and compensating variation) is a critical input to many economic policies, such as antitrust analysis, the valuation of non-market goods, and measuring the value of innovation (e.g., Williamson 1968, Willig 1976, Bresnahan 1986).

In practice, however, obtaining convincing empirical estimates of consumer surplus has proven to be extremely challenging. We typically observe only the equilibrium point that balances supply and demand. Variations in that equilibrium across time and space are generally the result of a combination of supply-driven and demand-driven shocks and thus are of little use in this regard. A large body of economic research focuses on demand estimation (see, for instance, Deaton 1986). The key to estimating demand elasticities is to isolate exogenous shifts in the supply curve, holding demand factors constant. In recent years, a great deal of work has focused on the development of new techniques for generating demand estimates in differentiated product markets (Baker and Bresnahan 1988, Berry et al. 1995, Nevo 2000, Petrin 2002).[3] This strand of the literature focuses on overcoming the data limitations that are often present in standard economic settings, such as the absence of individual level data, unobservable product characteristics, and unobservable consumer characteristics.

Existing empirical explorations of demand almost always generate local estimates of demand elasticities. These elasticities describe how consumers are likely to respond to small variations around the equilibrium price. Local elasticities, however, are not sufficient for estimating consumer surplus. To compute consumer surplus one needs to integrate the area under the demand curve, which requires knowledge of the quantity demanded for each possible price. Typically, there are no direct estimates of elasticities far from the equilibrium price, necessitating a strong functional form assumption (e.g., iso-elastic demand) to produce consumer surplus estimates.

In this paper we exploit the remarkable richness of the data generated by Uber, and in particular its low-cost product UberX, to generate consumer surplus estimates that require less restrictive identifying assumptions than any other prior research that we are aware of.

[2] The concept of consumer surplus, or "utilite relativé," was first introduced in 1844 by French engineer Jules Dupuit. Alfred Marshall later independently reintroduced and named the concept in his 1890 publication Principles of Economics (Houghton, 1958; Svoboda, 2008).
[3] In differentiated product markets such as those studied by Berry et al. (1995) and Nevo (2000), one needs not only instruments for price, but also instruments that shift market shares through a channel other than prices (Berry and Haile 2014).

UberX is an app-based service that algorithmically matches drivers to consumers seeking rides (see uber.com).[4] A critical feature of Uber is that it uses real-time pricing ("surge" pricing) to equilibrate local, short-term supply and demand. A consumer wishing to take a particular trip can face prices ranging from the base price (what we call the "no surge" or "1.0x" price) to five or more times higher, depending on local market conditions. Importantly, we observe detailed information not only for every trip taken using Uber, but also, critically, when a consumer searches for a ride using Uber *without* ultimately deciding to make a request. We, thus, observe the price offered to the consumer, and whether she accepts or rejects that offer. This information is crucial in our strategy for estimating demand.

If the degree of surge pricing faced by a consumer on a given trip were generated at random, then all that would be required to trace out a demand curve would be to compute the share of Uber searches culminating in a ride at each level of surge pricing. With randomization, if 70 percent of searches lead to a transaction at the base price, but only 63 percent of searches lead to trips when the price is ten percent higher (1.1x surge), then we could assume that people who received surge 1.0x would also have requested trips at a rate of 63 percent, had they been quoted the 1.1x price. This would imply that the elasticity of demand would be one on this part of the demand curve (i.e. a 10 percent reduction in the share of people who accept the offer--from a 70% purchase rate to a 63% purchase rate--is associated with a 10 percent increase in price).[5] Similar comparisons of ride completion rates at higher prices would trace out demand over whatever range of prices consumers faced. Combining these elasticity estimates with the actual quantity purchased at 1.0x surge yields the demand curve for customers offered 1.0x surge, as well as an associated consumer surplus.

In practice, the surge price that consumers face is not random; it reflects local demand and supply conditions. There is, however, a component of Uber pricing that is largely random from a consumer's perspective. Uber calculates each surge price to an arbitrary number of decimal places, but consumers are presented with discrete price increments (e.g., the lowest surge price is 1.2x, or 20 percent higher than the base price) to facilitate a simple, easy user experience.[6] Market conditions are nearly identical when the algorithm suggests a surge of 1.249x and when it

---

[4] The rampant growth of "peer-to-peer" transactions and the "sharing" economy have had a profound impact on many industries in recent years. A burgeoning economic literature is devoted to this topic (e.g. Cramer (2016), Cullen and Farronato (2014), Einav et al. (2015), Fraiberger and Sundararajan (2015), Hall and Krueger (2015)).

[5] In this example, the 63 percent of consumers who demonstrated a willingness to pay of 1.1x, reveal that had they only been asked to pay the base price, they would have received a consumer surplus of at least 10 percent of that base price. The 7 percent of customers who refuse to transact at 1.1x, by this same logic, reveal a consumer surplus when transacting at the base price that is less than 10 percent of the base price.

[6] For example, Uber might estimate that the appropriate multiplier is 1.61809, but for easy interpretation, they would charge the customer 1.6x.

suggests a surge of 1.251x, but in one case consumers face a 1.2x surge and in the other case they face a 1.3x surge. This provides the opportunity for regression discontinuity (RD) analysis, which allows us to estimate local elasticities of demand across the full range of surge prices.[7] A complicating factor in our analysis is that the expected wait time a consumer faces systematically changes at the price discontinuity. We observe the expected wait time of the customer in our data, so we can control for this factor in our analysis. Additionally, the expected wait time algorithm used by Uber is continuous, but is rounded to whole minutes when presented to customers. This allows us to use an RD design for identifying the causal impact of expected wait time on purchases and thus to more convincingly purge any impact of wait time differences from our price elasticity estimates.

Using a sample of nearly 50 million UberX consumer sessions, which represents the first 24 weeks in 2015 from Uber's four biggest U.S. markets, we estimate demand elasticities for Uber's most used service ("UberX").[8] Empirically, three basic facts emerge. First, our estimated demand elasticities are similar regardless of the sources of variation that we use in the estimation or the set of included controls, suggesting that our results are robust. Second, demand is quite inelastic. Our methodology estimates a set of price elasticities, most of which are between -.4 and -.6. Third, the elasticity of demand varies somewhat (but perhaps less than expected) as a function of observable characteristics such as time of day, user experience with Uber, or the presence of close substitutes.

These estimated Marshallian price elasticities form the basis of our consumer surplus calculations, but further assumptions are required. To compute consumer surplus, one needs to know how consumers would have responded had they faced a higher price. We do not directly observe this in the data. Instead, what we observe is how price responsive consumers are when market conditions dictated a higher price. The set of sessions with high surge prices may, however, differ systematically in their price responsiveness from those who see low surge.[9] We deal with this complication in two ways. First, we use propensity score methods to identify a subset of sessions that saw high surge prices, but whose observable characteristics (e.g., location, time of day, day of the week, and past usage of Uber) match the pool of sessions that face no surge. Second, we redo our estimates eliminating from the sample all observations where there is a positive local demand shock. The prices charged depend on the interplay of supply and

---

[7] Additionally, there are Uber business rules that sometimes cause prices to be far below what the surge algorithm recommends, allowing us also to analyze consumer behavior when the differences between the surge level and the surge generator are larger.

[8] The data were chosen in conjunction with Uber to be large enough to be representative, while not revealing information that may have more business sensitivity.

[9] Note that the same consumer opening the app under different circumstances may have different likelihoods of making a purchase and different sensitivities to price. Thus, we focus on sessions as our unit of analysis, not individuals.

demand. If the supply of drivers is low, prices can be high even though the number of requests in a given time and place are not out of the ordinary. Price spikes driven by idiosyncratically low supply are likely to provide a better counterfactual than those triggered by unusually high demand. Neither the propensity-score methods nor eliminating positive demand shocks materially affects the consumer surplus results.

We obtain large estimates of the consumer surplus generated by UberX. We compute the dollar value of consumer surplus from UberX rides taken in Uber's four biggest U.S. markets in 2015 (Chicago, Los Angeles, New York, and San Francisco) to be roughly $2.88 billion (SE=$122 million) annually. This is more than six times Uber's revenues from UberX in those cities.[10] In 2015, these cities accounted for around 42.6% of UberX US gross bookings. If we assume that consumer surplus is proportional to gross bookings, we can extrapolate to an estimate of $6.76 billion in consumer surplus from UberX in the U.S. The estimated consumer surplus is approximately 1.57 times as large as consumer expenditures on rides taken at base pricing. That is, for each $1 spent on an UberX ride at 1.0x, we estimate the consumer receives $1.57 in extra surplus. These estimates of consumer surplus are large relative to the likely gains or losses experienced by taxi drivers as a consequence of Uber's entrance into the market (Cramer 2016).

From a public policy perspective, our consumer surplus estimates have two shortcomings. First, they are derived from short-run demand elasticities, but any policy decision is likely to be interested in long-run consequences.[11] Second, our estimates miss the consumer surplus associated with other ride-sharing products (both those offered by Uber and by other ride-sharing companies), as well as consumer benefit or harm resulting from responses of the taxi cab industry to Uber's entry. We discuss in the concluding section what economic theory tells us about mapping from the numbers we are able to credibly estimate to the numbers that are of greatest economic interest.

Although very different methodologically from our paper, Buchholz (2016) is the most similar prior work in terms of goals. Buchholz (2016) estimates a dynamic spatial equilibrium model of New York City taxi cabs to assess the efficiency cost of existing regulations. He concludes that efficient two-part tariff pricing and a directed matching technology would deliver welfare gains of over $2 billion annually in New York.[12] Also in the basic spirit of our work in estimating

---

[10] This assumes a 25% commission of gross fares reserved for Uber. This percentage likely overstates Uber's actual average commission rate on the trips represented.

[11] The price variation we exploit is highly transient, and the set of competitors is fixed. Thus, the appropriate interpretation of our estimate is roughly, if Uber's system malfunctioned and Uber were therefore unavailable for a day, how much would consumers suffer? (The answer would be 1/365th of our annual consumer surplus number, or about $18 million.)

[12] More specific to the literature on taxis, there are a very rich set of papers that have attempted to understand the supply side of the taxi market (see Camerer et al., 1997; Farber, 2005, 2008, 2015; Crawford and Meng, 2011).

welfare impacts on consumers are Petrin (1999), Nevo (2000), Brynjolfsson et al. (2003), Goolsbee and Petrin (2004), Mortimer (2007), Crawford and Yurukoglu (2012), Quan and Williams (2014), and Crawford et al. (2015).

The remainder of the paper is structured as follows. Section 2 provides background on Uber. Section 3 describes the data and identification approaches underlying our estimates of demand elasticities. Section 4 presents the estimation results along with a series of sensitivity analyses. Section 5 explores the set of assumptions necessary to translate the demand estimates into consumer surplus and the conclusions we reach based on these calculations. Section 6 concludes with a discussion of the economic implications and interpretations of our findings.

2. Background on Uber

Uber is a technology company founded in 2009, which created a smart phone application that matches and handles payments between consumers seeking rides and Uber's "driver-partners." Uber's service has proven extremely popular, growing dramatically in terms of both geography and volume.[13]

To use Uber, a consumer downloads the app onto her smartphone for free. When seeking a ride, the consumer opens the Uber app and sees something akin to the screenshot in Figure 1. There is a map of the local area, a display of driver-partners in the area available to provide rides, and an estimate of how many minutes it will take the nearest vehicle to reach the consumer's location. Uber offers a number of different products, as shown near the bottom of the screen in Figure 1. The user is able to scroll between those products. If a consumer places an order, driver-partners are sequentially given the opportunity to accept that order until one does so. That driver-partner then picks up the rider and drops her off at her desired location. Uber defines a user product-session in which the user opens the app, culminating either in the user ordering a ride or electing not to order a ride.[14] Throughout this interaction, Uber records all actions taken on the app as well as certain background information relevant to the transaction. These data are collected and stored regardless of whether or not the session ends with a purchase.

[13] City governments have had varying reactions to Uber's operations, which dramatically alter the status-quo of transportation. Historically, for-hire transportation has been heavily regulated, usually via a taxi medallion system (see Frankena and Pautler, 1984). The incumbent taxi cab providers, not surprisingly, have been hostile to Uber's presence.

[14] A session is Uber's best attempt to identify a consumer's decision as to whether or not to make a purchase. For simplicity with our focus on UberX, we use "session" to refer to a product-session of UberX interaction unless otherwise specified. Technically, if the rider interacts with another Uber product at a similar time, this creates a separate product-session. A session is defined as a period of, not necessarily continuous, use between opening the app to UberX and either requesting a ride or ceasing to use the app for a period of 30mins. Thus, multiple closings and openings of the app in a short period of time do not generate many sessions. The median elapsed time of a session is 31 seconds.

Uber offers several products, which differ in terms of the types and size of cars, whether the ride is shared with other passengers, and the price. Our focus is on UberX, the core product that represented almost 80 percent of all Uber rides during the time period of our sample. With UberX, a rider summons a driver-partner who drives her own private vehicle and delivers the rider to the desired location without stops.[15]

Uber's base pricing system has components similar to standard cab pricing systems in which each city and product has a fare defined by price per mile, price per minute, a fixed fee, and a minimum total fare.[16] In contrast to regulated cabs, Uber also utilizes a dynamic pricing system, called surge pricing, on many of its products.[17] Uber's surge algorithm monitors rider demand and available driver supply and institutes a multiplier on the base price when demand outstrips supply at the base price.[18] This pricing system helps increase supply at times of high demand, and allocate rides to riders who value them most highly (Hall et al. 2015).

Around 21% of UberX sessions in our dataset have some surge price exceeding 1.0x. Figure 2 presents the observed distribution of UberX sessions in our data for surge prices greater than 1.0x. The number of sessions between 1.2x (the lowest surge) and 1.5x are roughly equally common. 1.5x is somewhat overrepresented due to less than perfect accuracy in excluding sessions where the surge price was altered by Uber business rules. Beyond that point, the number of sessions decreases monotonically with the level of surge. Roughly 4.1 percent of surge sessions involve surge levels at or above 3.0x; 0.65 percent involve a surge greater than or equal to 4.0x.

Although consumers are only shown a limited number of discretized surge levels, the algorithm generates a continuous measure of surge (which we call the "surge generator" or "generator"). This proves to be extremely useful for our identification strategy. Two customers who have nearly identical surge generators (i.e., face nearly identical market conditions), but who happen to be on opposite sides of a pre-defined cut-off, face discretely different surge prices.

---

[15] UberXL is identical to UberX, except that the requested vehicle must accommodate at least six passengers. UberBlack and UberSUV work just like UberX and UberXL respectively, except that the driver-partners are Transportation Charter Permit (TCP) licensed and the vehicles driven meet a higher set of standards. UberPool differs from the products above in that multiple riders are picked up en route by the same driver-partner, increasing the expected travel time and changing the user experience. UberTaxi differs from the other products in that it only summons licensed taxi drivers in licensed taxis, and the fares match the regulated taxi fares in the city.

[16] During the period of this study, each rider is also charged a $1 Safe Ride Fee, since renamed to Booking Fee, used specifically for ensuring the safety of riders and driver-partners on the Uber system. The Safe Ride Fee is not included in the base fare (and hence is unaffected by surge). An example fare can be found in Appendix 2.

[17] UberPool is priced according to a different system which is beyond the scope of this paper.

[18] Surge prices are always greater than or equal to 1.0, i.e. the price is never lowered below the base fare, even when market conditions suggest it should.

Figure 3 shows the frequency of surge by hour of the day and day of week for UberX, displayed as a heat map. The darker the color, the greater the share of trips involving surge pricing. Surge pricing is most likely to occur late at night (especially on weekends) and during the morning rush hour, but even at those times surge is invoked in less than half of the sessions. Surge pricing is less tightly linked to intuitions about demand than might be naively expected. This is because driver-partners seeking more revenue adjust their labor supply in response to predictable demand shocks.[19]

3. Data, Identifying price elasticities, and estimation

We focus our analysis on UberX because it is the only Uber product that has both substantial scale and frequent surge pricing, which is central to our identification strategy. Our primary data set covers all UberX sessions in Uber's four largest U.S. markets (San Francisco, New York City, Chicago, and Los Angeles) over the period January 1, 2015 to June 17, 2015.

Our unit of analysis is a customer session, a company-defined measure that captures a particular consumer trying to order a particular ride. There are approximately 54 million UberX sessions in our raw data. For each session, we observe the surge price, underlying surge generator (which the consumer does not see), Uber-defined geographic region, time and date, anonymized rider id, Uber's prediction of expected wait time (consumers see wait time in minutes), product, and the ultimate decision of the rider whether to request a car.[20] Uber has a variety of business rules in place that override the surge algorithm in certain cases, primarily to prevent riders and drivers from experiencing very sharp price changes. In our main analysis, we discard these rider sessions, which make up roughly 11.5% percent of the sample, leaving approximately 48 million observations in our base sample.[21] Summary statistics for this sample are presented in Table 1. We present statistics for the entire sample (column 1), as well as for three mutually exclusive and exhaustive subsets of the data: rider sessions with baseline pricing, (i.e. surge equal to one, in column 2), moderate surge (surge between one and two, in column 3) and high surge (column 4).

[19] High surge prevalence in the early morning is consistent with compensating differentials for driver-partners who have higher disutility from working those hours (or perhaps associated with the clientele seeking late night rides). Also consistent with this hypothesis is that driver-partners earn more on average on weekends than weekdays (both because of higher surge and more rides per hour.) There are some dimensions across which drivers cannot easily substitute, for example, across cities. Substantial differences in surge persist along those dimensions. For example, 14.4 percent of all New York City sessions involve surge, compared to 24.7 percent of Chicago sessions.

[20] Our methodology treats a car request as a purchase. Although there are times where either the driver or the rider decides to subsequently to cancel the car, we do not consider these cases as meaningfully informing our interpretation of the rider's decision to purchase.

[21] We will, however, use those excluded observations for supplementary analysis.

A number of patterns emerge in Table 1. The entries in the top row highlight the wide range of prices that consumers are exposed to due to surge. Relative to baseline prices in Column 2, the same trip costs a consumer roughly 50 percent more on average with moderate surge (Column 3) and is two and one half times more expensive on average when high surge is in effect (Column 4). Expected wait times are not highly correlated with surge. Although surge kicks in when demand is high relative to supply (implying long expected waits), high surge reduces demand and increases supply, equilibrating wait times. Purchase rates, shown in the third row, decline from 62 percent in Column 2 to 39 percent in Column 4 as surge rises. In percentage terms, however, purchase rates decline less than prices rise. This pattern foreshadows the consistently inelastic demand estimates found throughout the paper.

The remaining rows of Table 1 describe the distribution of our observations across city, time of the day and week, by the number of Uber trips made by the consumer. Our sample is split relatively equally across the four cities, with high-surge trips relatively less frequent in New York and more frequent in Chicago. As shown in Figure 4 above, surge most frequently occurs during rush hour and during weekend "party" hours. Most of the observations in our sample are associated with frequent Uber riders. The bulk of our data comes from frequent users of the product (more than eight rides in the period), and these frequent users are overrepresented during high surge periods.

Identifying Price Elasticities

If surge prices were randomly assigned to sessions, then identification of price elasticities would be straightforward: a simple regression of purchases on price, properly specified, would produce unbiased estimates. While we report such estimates for purposes of comparison, the surge varies systematically with observable factors (as shown in Table 1), and no doubt, with unobservable factors as well.

Our identification approach exploits the discontinuous pricing induced by Uber's business rules regarding surges. As noted earlier, although Uber generates a continuous measure of surge, actual prices charged are limited to a discrete set of points: 1.0, 1.2, 1.3, …, 4.8, 4.9, etc.[22] A surge generator value of 1.249 leads to a surge price of 1.2x whereas a value of 1.251 triggers a 1.3x surge. Thus, discrete pricing leads to discontinuous jumps in prices for sessions with arbitrarily small differences in the underlying demand and supply conditions that determine the surge generator calculation. Using RD methods applied to small windows around these jump

[22] For reasons that are not completely clear to us, Uber does not employ a surge of 1.1x. While surge multipliers in some unusual situations may rise above 5.0, we use only sessions at or below 5.0 in our analysis.

points, we hope to identify price variation that is more plausibly viewed as exogenous than other price fluctuations, which reflect a mix of demand and supply factors.

Figures 4 provides a visual example of our identification strategy, showing how purchase rates vary as a function of the surge generator over the range 1.15 to 1.35. The horizontal axis is the surge generator. The vertical axis is the percent of all sessions that convert into purchases within each bucket. Each point in the graph represents the aggregation of all observations falling into a particular generator bucket, with each bucket having a width of 0.0025. The vertical line at 1.250 represents the point at which price jumps discontinuously from 1.2x to 1.3x. To the left of that vertical line, all customers face a price of 1.2x; to the right of the line price is 1.3x. Moving from left to right towards the vertical line in the figure, purchase rates rise slightly, before dropping sharply at the price discontinuity. To the right of the vertical line, purchase rates continue a slight positive trend.[23] The purchase rate falls approximately 3 percent at the discontinuity; price rises by 8.3 percent (the percent difference between 1.2x and 1.3x), for an implied price elasticity of roughly -0.36 (i.e., -3/8.3).

Figure 5 displays the relationship between purchase rates and price discontinuities across a broader range of surge values. Once again, the surge generator is on the horizontal axis and the purchase rate is on the vertical axis. Each bar in the figure aggregates all sessions within a .01 surge generator window. Red bars lie just to the left of price discontinuities; yellow bars are just to the right of price discontinuities. All intermediate price ranges are shown in grey. In each of the 13 cases, the red bars are higher than the adjacent yellow bars, indicating that purchase rates fall as price discontinuously jumps. In contrast, at any given price level (the gray bars between a yellow and a red bar), purchase rates exhibit no obvious pattern of increase or decline.

To estimate a demand curve, one needs to hold everything else constant, other than price. To test whether this is indeed the case at our surge-driven price discontinuities, we estimate equations of the form:

$$\begin{aligned} Outcome = \alpha + \theta * Window * Post + \beta_2 * Window + \beta_3 * (1 - Window) * Post \\ + \beta_4 * Wait + \beta_5 * (1 - Post) * Generator + \beta_6 * Post * Generator \\ + \varepsilon \end{aligned} \quad (1)$$

[23] The slight positive trend moving from left to right is not a consistent pattern over the range of surge generator levels.

where *Outcome* represents an outcome of interest (e.g. was a purchase made, was the session during rush hour, etc.). *Window* is an indicator as to whether the observation lies close to a price discontinuity, *Post* is an indicator taking the value of one when the observation is to the right of the price discontinuity, *Wait* is the expected wait time, *Generator* is the continuous measure of surge produced by the Uber algorithm.[24]

Table 2 reports the results of these regressions for a wide range of outcomes. Each row represents a separate regression on a different dependent variable. We report only the key coefficient θ, which captures the average difference in the outcome variable just after a discontinuity versus just before (using a window on either side of the discontinuity of 0.01), controlling for other factors. The first row corresponds to the request rate (i.e. the share of observations in which a purchase is made). As would be expected given the results in Figures 4 and 5, request rates exhibit a clear decline at the price threshold where the discontinuity occurs: (a coefficient of -.0201 with a t-statistic of 20). Note, however, that expected wait times also drop discontinuously at the surge thresholds: customers wait an average of .129 minutes (i.e., approximately 8 seconds) less after the threshold. This result is not surprising; it occurs mechanically as a consequence of purchases declining at the threshold as well as due to increased incentive for drivers to make pick-ups in this area. When fewer people make purchases, there are more open cars available to pick up other customers, reducing wait times.[25] Failure to take these differences in wait times into account will bias our price elasticities towards zero, since higher prices are correlated with lower wait times. Empirically, we deal with this complication in two ways. First, we control for expected wait times in some of our specifications. Second, in other specifications we not only control for expected wait times, but also instrument for expected wait times to deal with concerns of endogeneity. Conveniently, Uber's measure of expected wait time is measured in seconds, but consumers are only shown wait times rounded to whole minutes. Thus, we can exploit the sharp discontinuities in wait times displayed to consumers to identify their sensitivity to waiting, just like we use the price discontinuities to identify sensitivity to price.[26] In other words, we identify the causal impact of longer expected wait times using only the variation in wait times that is generated by the sharp jumps that occur around the thresholds where reported wait times jump discretely. The t-statistic on our instrument in the first stage regression is around 20 (detail in Appendix: First-stage expected wait time regression).

---

[24] When our outcome is expected wait time, we exclude the control for wait time in the regression.

[25] It is also true that the supply of drivers responds positively to higher surge prices, but even without that behavioral response, we would expect to see wait times lower after the discontinuities.

[26] Uber primarily stores wait times as minutes. Accessing wait times in seconds turns out to be time consuming and difficult. With substantial effort, we were able to pull two weeks of this data. As it turns out, the estimated impact of longer wait times on purchase rates is virtually identical when estimated using OLS or when we instrument for wait time using discontinuities, so instrumenting has no discernable impact on either the coefficient on wait time in a price regression or on our price elasticity estimates.

In contrast to wait times, the other observable characteristics vary only weakly at the price thresholds, as would be expected.

Estimating price elasticities

In principle, we can estimate price elasticities simply using the observed discontinuities in purchase behavior around the price jumps. We also estimate specifications that control for a range of observable characteristics that might influence demand. We run a separate regression for each price discontinuity, which is identical to equation (1) above with purchase as the outcome, but with the addition of fixed effects for city and our eight time of the week indicators:

$$\begin{aligned} Purchase = {} & \alpha + \theta * Window * Post + \beta_2 * Window + \beta_3 * (1 - Window) \\ & * Post + \beta_4 * Wait + \beta_5 * (1 - Post) * Generator + \beta_6 * Post \\ & * Generator + FE(city) + FE(hour\ and\ day) + \varepsilon \end{aligned} \tag{2}$$

where *Purchase* is an indicator variable for whether or not a session ends in a purchase, *Window* is an indicator as to whether the observation lies within 0.01 of a price discontinuity, *Post* is an indicator taking the value of one when the observation is to the right of the price discontinuity, *Wait* is the expected wait time, *Generator* is the continuous measure of surge produced by the Uber algorithm. Fixed effects for city and our eight "*Hour and day"* categories are also included. For each price discontinuity, we include all sessions that are quoted the price falling on either side of the discontinuity (e.g., for the price discontinuity between 1.5x and 1.6x surge pricing, all observations at 1.5x and 1.6x are included). The price elasticities are identified only using the observations within the window around the discontinuity; the remaining observations are nonetheless useful for pinning down the other parameters estimated.

The coefficient of interest is θ which captures the drop in purchase rates at the price discontinuity, controlling for other factors and allowing for a different linear trend in the generator on the two sides of the threshold. To transform these values into elasticities, we use the definition of a price elasticity:

$$Elasticity = (\%\Delta\ in\ Quantity / \%\Delta\ in\ Price) = ((\theta / Purchase\ Rate) / \%\Delta\ in\ Price) \tag{3}$$

where *Purchase Rate* is the share of sessions that result in purchases at a given price. Since the purchase rate and percent change in price are directly observed in the data, there is a direct mapping between θ and the price elasticity.

Table 3 presents these elasticity estimates. Each row in the table corresponds to a different surge-driven price discontinuity.[27] For instance, the first row reflects the discontinuity where prices jump from 1.0x to 1.2x; the second row is the jump from 1.2x to 1.3x. At higher surge levels, we have fewer observations and consequently our estimates become less precise. Therefore, we report pooled estimates for all discontinuities from 1.9 to 2.3, 2.4 to 3.0 and 3.1 to 5.0[28] At these higher surge levels, we generate separate estimates around each discontinuity, and then we present an inverse variance-weighted average of the estimates. Each column in Table 3 corresponds to a different specification. For purposes of comparison, Column 1 reports naive results from regressing quantity on price, ignoring the obvious simultaneity between those two variables, i.e. Column 1 does not exploit the RD identification strategy used in all of the remaining columns. Column 2 shows the results from our RD approach, absent any other controls. Columns 3 adds expected wait time as a control. Column 4 instruments for expected wait time.[29] Column 5 adds controls for time of week and city as controls to the specification in Column 4.

Three notable patterns emerge in Table 3. The first pattern is that there is little systematic change in the coefficients as controls are introduced or from instrumenting for expected wait time. At low surge levels (e.g., up to 1.5x), estimates that use all the variation in the data (column 1) are on average slightly smaller in absolute magnitude than those identified only off of the price discontinuities (column 1 versus column 2). Controlling for expected wait time also slightly increases the estimates (column 3 versus column 2). Instrumenting for wait time and adding other controls (columns 4 and 5) has little impact on the coefficients compared with RD estimates that add expected wait time as a control (Column 3). The second striking pattern in the table is the degree to which demand is price *inelastic* along the length of the demand curve. Of the fifty price elasticities estimated in the table, only three carry a coefficient greater than one in

[27] Note that our specification differs somewhat from the standard setting in which price, rather than degree of surge, would be the right-hand-side variable. Since surge is proportional to the fare absent surge, the longer the trip, the greater the dollar increase in the fare as surge increases. There are also differences in base fares per mile across cities in the sample. Thus, the elasticity estimates we report here represent a mixture across these various types of trips. We explore the impact on elasticities and consumer surplus of treating cities as distinct markets below.

[28] The particular groupings of the elasticity estimates do not materially impact our conclusions.

[29] As noted earlier, because it is extremely labor intensive to retrieve the continuous measure of expected wait time, we were only able to access two weeks worth of those data. This continuous measure of wait time is our instrument for the discrete version that is seen by customers. Consequently, we estimate the first stage equation for expected wait times only in that subset of the data and then impose that coefficient on the full sample using a two-step procedure. Columns 4 and 5 thus are estimated using all of the observations in the data.

absolute value. The median point estimate is -0.51; the mean point estimate is -0.57. The third pattern in the table is that elasticities are quite precisely estimated at low prices, but become less precise at higher prices, because there are many more observations at low prices and % price changes are much larger.

Table 4 explores how the estimated price elasticities vary across subsets of the data. To facilitate comparisons across subgroups, we compute a single price elasticity estimate that is an inverse variance-weighted average of the estimates reported in Column 5 of Table 3 The top row of Table 4 reflects the whole sample (i.e. all the data in Table 3), for which we find a demand elasticity of -0.55 (se=.02). The next four rows of the table report elasticities separately for the four cities in our sample. Interestingly, demand is estimated to be substantially less elastic in Los Angeles (-0.33, se=.05) than in the other three cities. The next eight rows of the table partition the hours of the week into eight mutually exclusive and collectively exhaustive categories.[30] We observe little variation in price elasticity along this dimension. Demand is estimated to be most elastic during the day on weekends (-0.66, se=0.05) and least elastic during non-rush hour times of the day on weekdays (-0.46, se=0.06). When we divide consumers by prior experience with Uber, the more frequent users (e.g., those taking more than 3 rides in a period) are the most elastic.

4. Turning price elasticities into consumer surplus estimates

Armed with price elasticity estimates at various points along the demand curve, in this section we describe our methodology for mapping these into consumer surplus. We first carry out that mapping and ignore the fact that the estimates provided by the data are not precisely the estimates we require. This gives us a baseline, but potentially biased, estimate. We then consider three potential problems with our approach, present approaches to dealing with those problems, and make conjectures regarding the likely sign and size of the bias caused by these data shortcomings.

Measuring consumer surplus

Transforming price elasticity estimates into an overall measure of consumer surplus requires additional assumptions. A thought experiment is helpful in this regard. Consider first the set of transactions that took place with no surge. To estimate consumer surplus via price changes, we would want to offer a menu of prices, identifying the price at which the consumer becomes indifferent to making the purchase, holding everything else constant. The sum of the differences between that willingness to pay and the price the consumer pays is the consumer surplus

[30] The details of this time grouping are described in Appendix 1.

associated with those transactions. One would then carry out the same exercise for the transactions done at 1.2x surge, but using 1.2x as the base price, and so on, for each transacted price. The sum of these calculations would be total consumer surplus from UberX.[31]

What we actually observe in the data differs from this idealized thought experiment in at least three ways. First, we can only use our RD design to measure price elasticities at a handful of discrete points. Second, the price elasticities we see are derived from the population of sessions that appear near the threshold of the price change. That population of sessions is not necessarily the same as the population of sessions exposed to other prices. For instance, surge pricing is more common during rush hours and late at night; the willingness to pay for an Uber ride may vary systematically over the course of the day. Third, in the thought experiment above, everything else is held constant as prices change. In reality, other demand determinants like the availability of taxis might differ between low surge and high surge situations.

We explore each of those potential problems below, but for the time being we start by simply assuming them away to generate a first pass estimate. For our first-pass estimates, with respect to a changing composition of sessions as surge changes and the possibility that outside options might also change, we begin by assuming those two concerns are unimportant, simply using the estimated elasticities from Table 2 and assuming that these elasticities are unbiased. To deal with discrete elasticity estimates, we follow the obvious path of assuming that the locally estimated elasticities also apply nearby. For instance, although our RD price elasticity going from 1.2x to 1.3x is identified off of variation from observations with generators between 1.24x and 1.26x, we will assume the same elasticity holds for sessions ranging from 1.20x to 1.30x. As noted above, we revisit these concerns below after explaining our basic methodology.

Estimates using the basic methodology

Our basic methodology is straightforward. We start with the set of sessions who made a purchase at 1.0x surge. Assuming that the relevant price elasticity for this group at 1.2x surge is the price elasticity estimated using RD and reported in Figure 3, we compute the number of sessions who would have continued to buy at 1.2x surge. Multiplying the price difference (20 percent) by the average fare actually paid ($13.3) multiplied by the quantity willing to pay 1.2x (in this case, 104MM trips) yields an estimate of the amount of consumer surplus generated up to the price 1.2x for those who paid 1.0x. Starting at that price and quantity, we then carry out an

[31] Note that there are sessions who are only offered, say, 1.5x surge by Uber, who do not transact at that price, but would have made a purchase had they been offered, say, 1.2x surge. These individuals are (rightly) ignored in the consumer surplus calculation because they didn't transact and thus received no consumer surplus from Uber. These individuals might, however, be of interest in the creation of hypothetical counterfactuals had Uber priced differently.

analogous calculation moving between 1.2x and 1.3x. That yields the consumer surplus for the same set of sessions, but associated with the consumer willingness to pay between 1.2x and 1.3x. We repeat that process until we reach 4.8x.[32] Beyond that point, where we have no elasticity estimates, we make the most conservative assumption, namely that no consumer is willing to pay above 4.9x.[33] Summing all of these surplus estimates yields the consumer surplus for sessions who faced 1.0x pricing. Figure 6 displays the outcome of this first step in the exercise visually.

We then repeat the exact same process outlined in the previous paragraph, but using sessions where a purchase was *actually* made at 1.2x (instead of 1.0x in the previous paragraph). Given the simplifying assumptions of our basic methodology, these sessions have the same price elasticities and thus an identically shaped demand curve, but they paid 1.2x, so their consumer surplus starts only at that point. The same approach is used for those who paid 1.3x, 1.4x, …, up to 4.9x.

Carrying out this exercise, we can extrapolate our estimates in these four cities from the first 24 weeks of 2015 using volume figures from all of 2015 to reach an overall consumer surplus estimate of $2.88 billion[34], with a bootstrapped standard error estimate of $122 million. If we assume that consumer surplus per trip in our sample (the first 24 weeks of 2015 for four large cities) is typical for the broader set of UberX consumers, than our estimates imply that UberX generated $6.76 billion in consumer surplus in the United States in 2015. We now turn our attention to possible biases induced by the simplifying assumptions made above in estimating consumer surplus.

Sensitivity of the results to incorporating continuous elasticity estimates

Our basic methodology above assumes a functional form for the elasticities that is linear with jumps at the RD points. A more nuanced approach predicts the elasticity at each point along the surge generator by fitting a curve using a flexible functional form. Doing this using a fifth-order polynomial yields consumer surplus estimates that are generally similar to those presented above. The two curves are nearly identical, and the high order polynomial would imply an estimate of 1.1% less consumer surplus than our preferred method.

---

[32] For CS, we inverse variance weight the less precise estimates created at high surge as described in the elasticity section.

[33] If instead, we assume that the price elasticity associated with our highest level of surge holds beyond that point, our estimate of consumer surplus increases 34 percent.

[34] As noted in Table 2, different cities have different elasticities. If, instead of applying the blended elasticities to all fares, we apply each city's elasticities to the fares from that city and then sum the consumer surplus, we arrive at an estimate of $2.68B, implying $1.46 of consumer surplus per dollar of fare. A further analysis would be to do a similar sensitivity analysis grouping by rides with a similar fare, rather than by city, but this analysis is not feasible with current data availability.

<u>Sensitivity of the results to the possibility that the composition of the sessions changes over the surge distribution</u>

As reported in Table 1 earlier in the paper, there are observable dimensions along which sessions systematically differ with the degree of surge. For example, the average surge at 1am on Saturday night is meaningfully higher than the average during the middle of the work day. The primary concern surrounding observable differences in the sample as a function of surge is that the price elasticities might vary along these dimensions.

We adopt two different approaches to deal with this issue. Our first approach is to use matching techniques to reweight the data so that the observable characteristics at a particular surge level match as closely as possible to the baseline population we are interested in. Say, for instance, we want to construct a counterfactual in which the set of sessions used to estimate the price elasticity at 1.5x mirrors the observable characteristics of the sessions that actually saw 1.0x surge.

To accomplish this matching on observables, we partition the data into 9,216 possible cells that represent the intersection of 144 geographic areas X 8 parts of the week X 4 measures of intensity of Uber use X 2 measures of whether the observation is before or after the calendar mid-point of our sample to capture possible changes over time given the rapid growth of Uber and the dynamic environment in which it operates. We observe sessions in 7,627 of these cells. We compute the share of sessions falling into each bucket b when surge is 1.0x (call this $S_{1.0x}$) and also the share in each of the buckets in the neighborhood of the 1.5x cutoff (call this $S_{1.5x}$), where b indexes the buckets. The resulting weight given to each observation in bucket b when estimating the price elasticity at the 1.5x cutoff is $S_{1.0x}/S_{1.5x}$. This ensures that the weighted number of observations per bucket in the vicinity of the 1.5 cutoff matches the number of observations in that same bucket at 1.0x surge.[35] Figure 8 provides a comparison of the estimated demand curves from our basic methodology and those using the propensity score matching technique. Until surge reaches 1.5x, the two curves look quite similar. Beyond that point, the re-weighted demand curve is substantially steeper. This implies that users who face high surge, but who look like no-surge users, are more likely to convert than the typical high-surge user. These results suggest that our basic approach understates consumer surplus by 10.4 percent.

[35] A complication arises, of course, when there are no observations in this example in a particular bucket around the 1.5x threshold. Our baseline procedure in this scenario is to estimate the elasticity without this bucket, which assumes that sessions in this bucket would have behaved like the sessions we actually observe near the 1.5x threshold.

Our second approach to addressing the problem that sessions that see high surge may systematically differ from those with low surge exploits the fact that pricing is a function of both demand and supply conditions. Even when levels of demand are completely ordinary, consumers can face high prices if supply is unusually low. Price changes caused by a shift in supply holding demand constant are likely to provide a more compelling counterfactual for constructing consumer surplus estimates than those driven by spikes in demand.[36] We estimate separate regressions of the form below in data where each observation is a location in a particular five minute time period:

$$Elasticity = (\%\Delta\ in\ Quantity / \%\Delta\ in\ Price) = ((\theta / Purchase\ Rate) / \%\Delta\ in\ Price) \quad (4)$$

where the dependent variable is the number of sessions initiated by consumers within a given five minute time period t in one of the 783 mutually exclusive and exhaustive geographic units in our sample.[37] On the right hand side, we include as indicators the week of the year and the hour of the week (24 hours x 7 days per week). The fitted value of this regression is the expected number of consumer sessions originating in a particular time and place. Under the hypothesis that price elasticities may systematically differ during demand spikes, we then explore two different subsets of the data. First, we exclude any observation where the observed number of sessions was greater than the number of sessions our model predicted.[38] Second, we exclude observations in the top and bottom quartile of demand shocks, i.e. we keep only the middle half of the data. As would be expected, Table 5 shows that high surge activity is greatly reduced in these two subsets. Compared to the full data set, the mean level of surge in our two subsets is about half as high (1.14 versus roughly 1.07). The effects are most extreme for high surge levels: our two subsets each comprise about half of the total observations, but capture less than 20 percent of the surges above 2.0 and less than 10 percent of surges above 3.0. It is rare, empirically, to achieve high surge levels based on supply shocks alone.

Figure 9 compares our baseline demand estimates to those derived using these two subsets of the data. Up until a surge of roughly 2.0x, the three curves look similar. Beyond that point, but especially at the very highest surge levels, the estimates based on the subsets become extremely

---

[36] If the RD approach we use is valid, then demand shocks are a useful form of variation for identifying price elasticities. The concern here is not the usual one of endogeneity. Rather, the concern is that during demand spikes consumer preferences differ from typical demand periods. In other words, using demand shocks we can get consistent estimates of the price sensitivity for the population facing those prices; that population may not, however, be representative of the typical consumer.

[37] We treat each update to the geography of a unit as a separate unit. The number of geographic units at a particular time is substantially smaller.

[38] Exploration of the most extreme positive demand shocks confirmed the unusual nature of the circumstances, e.g. rainstorms so severe that widespread flooding occurred, major concert and sporting events, etc.

noisy. The implied consumer surplus using only below average demand shocks is 11.1% percent greater than the baseline: when we use the middle demand shocks only, the estimate is 5.9% percent greater.

Sensitivity of the results to the possibility that other (mostly unobservable) demand determinants covary with the degree of surge

The assumption underlying a demand curve is that all else is held constant as the price of the good in question changes. We do not, in our setting, have good measures of the other demand determinants, such as the availability of taxi cabs, how hard it is raining, whether or not the person has an umbrella, or how costly it will be to the consumer if he or she get to their location ten minutes late. Most likely, the paucity of outside options will make demand for Uber more inelastic. If high surge is correlated with bad outside options for users, then biased elasticity estimates will likely lead us to overstate consumer surplus.[39]

Although we cannot deal directly with this issue because we do not observe outside options in our data, we are able to address this potential confound in three ways. The first approach is the standard one -- to control for an increasing number of dummies and interactions and to observe how this affects the estimates. Additional controls had little impact on the coefficients in Table 3. Moreover, when we include dummies for the nearly 8,000 partitions described above, the estimated elasticities do not systematically change.

The specific structure of the Uber surge algorithm and business rules provide two other ways to shed light on the possible biases induced by unobservable factors, which are very different from the approach of adding more and more controls. The first of these exploits the continuous nature of Uber's surge algorithm combined with the discontinuous pricing jumps. The other exploits a business rule imposed by Uber that leads a subset of sessions to face prices that differ sharply from what the surge algorithm judges to be optimal. We deal with these two different approaches in turn.

Do users who experience worse market conditions, but the same prices, have different purchase likelihoods?

[39] It is possible, also, that the number and type of consumers who open the Uber app will be a function of outside options, further complicating this analysis. If Uber users on the extensive margin are more price sensitive, and those users disproportionately appear when outside options are bad and surge is high, then the empirical bias could actually go in the other direction.

Market conditions differ across consumers who are quoted the same price. Consumers quoted 1.3x surge have underlying surge generator values ranging from 1.25 to 1.35. To the extent that outside options are worse on average at high surge relative to low surge, that same pattern should be evident within these more narrow windows as well: the sessions at 1.35 should have worse outside options than those at 1.25. And if this is the case, then one would expect a higher share of purchases by those at 1.35 than for those at 1.25. The key to this analysis is that we see consumers with known differences in underlying supply/demand conditions, but who all face the same price. In typical market settings, we do not get to observe this sort of variation, because prices equilibrate supply and demand at each point. It arises in this case only because Uber holds prices fixed over a range of market conditions. If prices are not held constant, we lose our ability to isolate the influence of outside options.

To empirically test the concept in the preceding paragraph, we run regressions with an indicator for whether a purchase is made as the dependent variable, with indicators for each price level on the right hand side, along with Uber's continuous measure of surge. Because of the inclusion of the price-level indicators, the identification of the continuous measure comes only from consumers who face different market conditions, but the same price.[40, 41] The estimated coefficient on the continuous surge measure is 0.0104 (SE=0.00). So, indeed, by this metric, it appears that outside options are slightly worse at higher levels of surge. The magnitude of the effect is quite small. The degradation of outside options moving from 1.0x to 3.0x would lead purchase rates, holding actual price constant, to increase by 2.08 percent, implying that our consumer surplus estimates above may be upward biased by roughly 3-4 percent.

Do individuals presented with prices well below what is implied by demand conditions behave differently than others facing those prices?

In the discussion above, we exploited relatively small differences in demand conditions across sessions offered the same price. Uber business rules generate a different natural experiment that leads, on occasion, to consumers who face very different supply/demand conditions to face the same price. Specifically, Uber limits the extent to which each consecutive update to the price can raise the surge price. At the beginning of a sharp demand/supply imbalance, the surge price can only increase by .5 in the first increase, .6 in the second increase, etc. As a consequence, there are cases in which sessions see a price of 1.5x, but the surge generator characterizes the market setting as warranting a surge price of 2x, 3x, or even higher. If the outside options at

[40] Essentially this regression tells us whether, in Figure 5, there is a systematic trend in the height of the gray bars that fall within each pair of red and yellow bars.
[41] For computational feasibility, this regression is estimated using all sessions with surge > 1 and 15% of sessions where surge = 1. For specification, see Appendix: Within surge level estimator specification.

higher surge are worse, then we would expect that purchase rates of sessions facing high surge conditions *but artificially low prices of 1.5x* will be higher than the purchase rates of sessions who see prices of 1.5x because that is what market conditions warrant.

Figure 10 shows virtually no relationship between underlying demand conditions and purchase rates in those cases where price is artificially restricted by internal business rules. While cases where the underlying model is changing prices rapidly may not perfectly represent standard scenarios, this does suggest that there is little or no bias from this channel in the consumer surplus estimates above.[42]

4. Conclusion

This paper exploits the remarkable richness of Uber data to investigate the impact that Uber's introduction has had on consumer welfare. Our approach exploits the fact that Uber (1) has detailed session-level data, even when no purchase is made, (2) varies prices with market conditions, and (3) has business rules that generate sharp discontinuities in the prices that like customers face. We find that consumer demand is inelastic, despite the existence of what would seem to be reasonably close substitutes (competitors, taxis, public transportation, driving one's self). Inelastic demand translates into large consumer surplus estimates: roughly $2.88 billion dollars in 2015 for the four cities in our sample, or $6.76 billion if extrapolated to all UberX trips in the U.S. for that year. This estimate of consumer surplus is two times larger than the revenues received by driver-partners and six times greater than the revenue captured by Uber after the driver-partner's share is removed.

The consumer surplus estimates we generate correspond to a short-run demand curve because they are identified off of short-run price shocks. One day's worth of consumer surplus, by our estimates, is about $18 million. If Uber were to unexpectedly disappear for a day, that is how much consumers would lose in surplus. From a public policy perspective, however, our measure of consumer surplus generally would not be the estimate of greatest interest. Economic theory helps one move from the number we estimate to the relevant elasticity for a given policy question. If, for instance, one wanted to know how consumers would be affected if Uber disappeared permanently, a long-run elasticity would be more appropriate, as consumers would find substitutes and other firms would enter the market. If, however, one wanted to examine the impact of regulators banning ride-sharing altogether in an existing geography or delaying the entry of ride-sharing in a new market, then it is quite possible our estimates are far too low. The

[42] One caveat on this conclusion is that many more riders (relatively to available supply) are opening the app for high generators, causing higher ETAs. However, from 1.5x to 4.0x, a 166% increase in the estimated market price, ETAs only increase by ~26%. Given the coefficients on ETA in purchase regressions, this change in ETA would not have dramatic effects on purchase rates.

demand curve we estimate takes as given both competitor offerings and the presence of other Uber products. Without such products, demand for UberX would likely be more inelastic, thus giving rise to higher losses in consumer surplus than our estimation procedure would suggest.

The emphasis in demand estimation over the last two decades has been on methodological advances that allow researchers to overcome the inherent limitations in the sorts of data that have typically been available. While recognizing the immense contributions of that work, this paper also points to a second path forward: one in which better data are the key to deeper insights. Massive changes that are taking place in the economy in terms of the availability of transaction-level data, the increased use of sophisticated pricing tools by firms, and the growing openness of firms to randomized experiments. All of these forces point towards a future world in which data richness transforms our understanding of firms and consumers.

References

Baker, Jonathan B. and Timothy F. Bresnahan. 1988. "Estimating the residual demand curve facing a single firm." *International Journal of Industrial Organization* 6(3): 283-300.

Berry, Steven and Philip Haile. 2014. "Identification in Differentiated Products Markets Using Market Level Data." *Econometrica* 82(5): 1749-1797.

Berry, Steven, James Levison, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 60(4): 889-917.

Bresnahan, Timothy. 1986. "Measuring the Spillovers from Technical Advance: Mainframe Computers in Financial Services." *American Economic Review* 76(4): 742-755.

Brynjolfsson, Erik, Michael D. Smith, and Yu (Jeffrey) Hu. 2003. "Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers." *Management Science* 49(11): 1580-1596.

Buchholz, Nicholas. 2016. "Spatial Equilibrium, Search Frictions, and Efficient Regulation in the Taxi Industry." (Job market paper, UT Austin).

Camerer, Colin F., Linda Babcock, George Loewenstein and Richard Thaler. 1997. "Labor supply of New York City cab drivers: One day at a time." *Quarterly Journal of Economics* 112(2), 408-441.

Cramer, Judd. 2016. "The Effect of Uber on the Wages of Taxi and Limo drivers." (Job market paper: Princeton University).

Crawford, Gregory S., Robin S. Lee, Michael D. Whinston, and Ali Yurukoglu. 2015. "The Welfare Effects of Vertical Integration in Multichannel Television Markets." NBER Working Paper No. 21832.

Crawford, Vincent P. and Juanjuan Meng. 2011. "New York City Cab Drivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational Expectations Targets for Hours and Income." *American Economic Review* 101(5): 1912-32.

Crawford, Gregory S. and Ali Yurukoglu. 2012. "The Welfare Effects of Bundling in Multichannel Television Markets". *American Economic Review* 102(2): 643-85.

Cullen, Zoë and Chiara Farronato. 2014. "Outsourcing Tasks Online: Matching Supply and Demand on Peer-to-Peer Internet Platforms." (Job market paper: Stanford University).

Deaton, Angus. 1986. "Demand analysis." *Handbook of Econometrics* 3: 1767-1839.

Einav, Liran, Chiara Farronato, and Jonathan Levin. 2015. "Peer-to-Peer Markets." Stanford Institute for Economic Policy Research. Discussion Paper No. 15-029.

Farber, Henry. 2005. "Is Tomorrow Another Day? The Labor Supply of New York City Cab Drivers." *Journal of Political Economy* 113(1): 46-82.

Farber, Henry. 2008. "Reference Dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers." *American Economic Review* 98(3): 1069-1082.

Farber, Henry. 2015. "Why You Can't Find a Taxi in the Rain and Other Labor Supply Lessons from Cab Drivers." *Quarterly Journal of Economics.* 130(4): 1975-2026.

Fraiberger, Samuel and Arun Sundararajan. 2015. "Peer-to-Peer Rental Markets in the Sharing Economy." NYU Stern School of Business Research Paper.

Gains from Direct Broadcast Satellites and the Competition with Cable TV." *Econometrica* 72(2): 351-381.

Hall, Jonathan, Cory Kendrick, and Chris Nosko. 2015. "The Effects of Uber's Surge Pricing: A Case Study". unpublished.

Hall, Jonathan V. and Alan B. Krueger. 2015. “An Analysis of the Labor Market for Uber’s Driver-Partners in the United States.” Working paper #587, Princeton University Industrial Relations Section.

Houghton, R.W. 1958. “A Note on the Early History of Consumer's Surplus.” *Economica* 25(97): 49-57.

Mortimer, Julie Holland. 2007. “Price discrimination, copyright law, and technological innovation: Evidence from the introduction of DVDs.” *Quarterly Journal of Economics* 122(3): 1307-1350.

Nevo, Aviv. 2000. “Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry.” *RAND Journal of Economics* 31(3): 395-421.

Petrin, Amil. 2002. Quantifying the Benefits of New Products: The Case of the Minivan, *Journal of Political Economy*. 110(2): 705-729.

Quan, Thomas W., and Kevin R. Williams. 2014. “Product Variety, Across Market Demand Heterogeneity, and the Value of Online Retail”. unpublished.

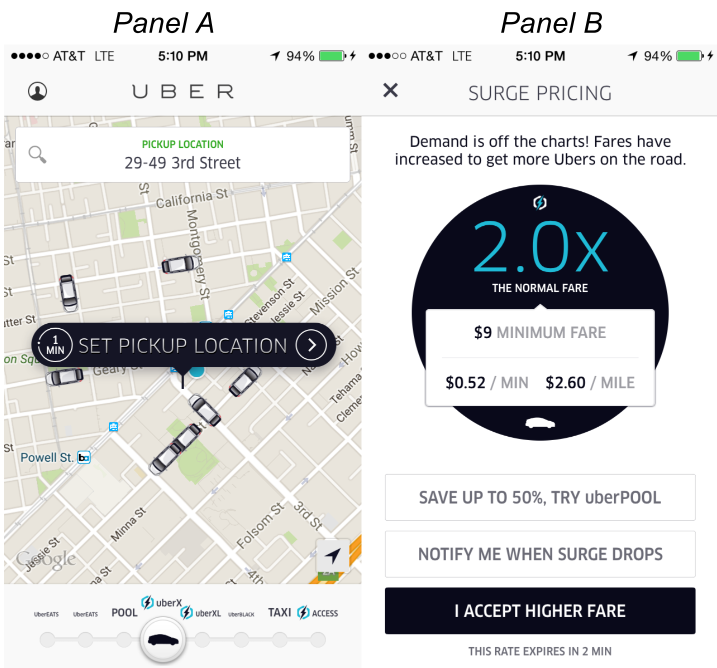
Svoboda, M. 2008. “History and Troubles of Consumer Surplus,” *Prague Economic Papers* 17(3): 230-242.

Williamson, Oliver. 1968. “Economies as Antitrust Defense: The Welfare Trade-Offs.” *American Economic Review* 58(1): 18-36.

Willig, Robert D. 1976. “Consumer’s Surplus Without Apology.” *American Economic Review.* 66(4): 589-597.

Figures

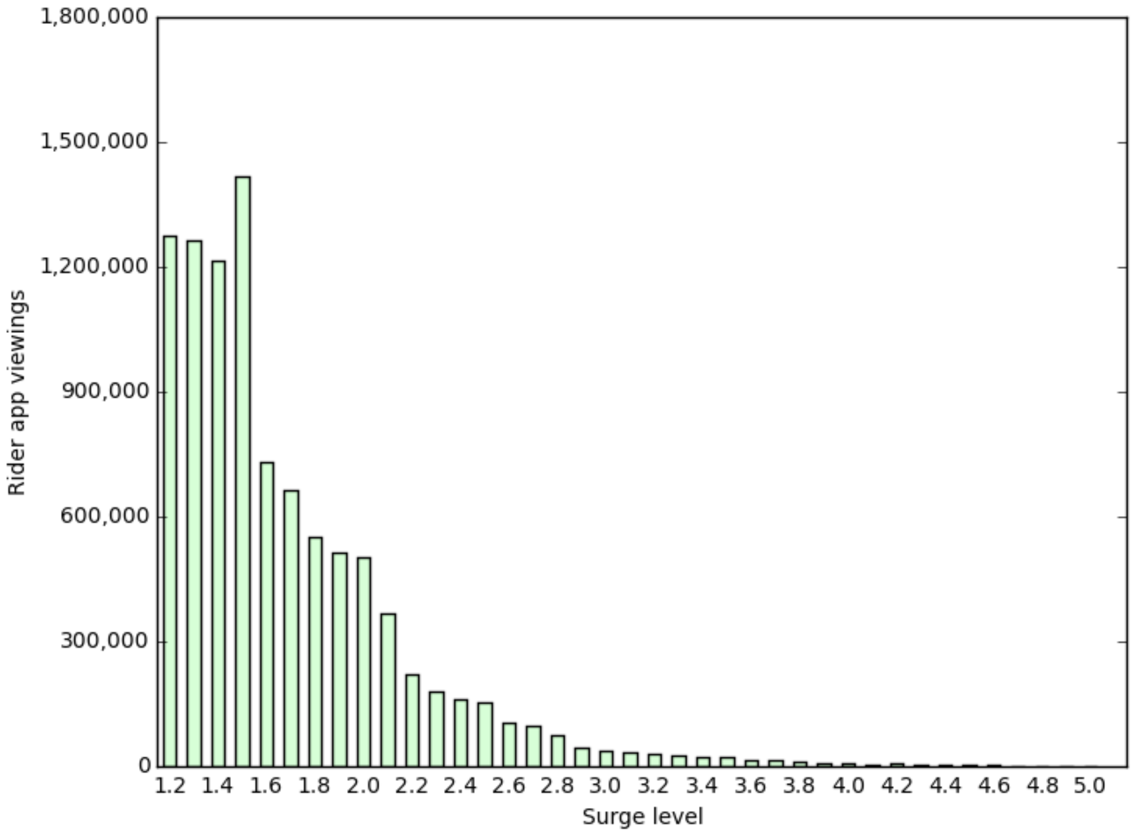
Figure 1: Uber mobile application request screens



Note: These figures illustrate what the Uber app looks like[43] when a rider is requesting transportation. Panel A depicts the period preceding a request when users are asked to choose a product and set a pick-up location. Panel B depicts the confirmation screen where users are presented with a surge price when applicable.

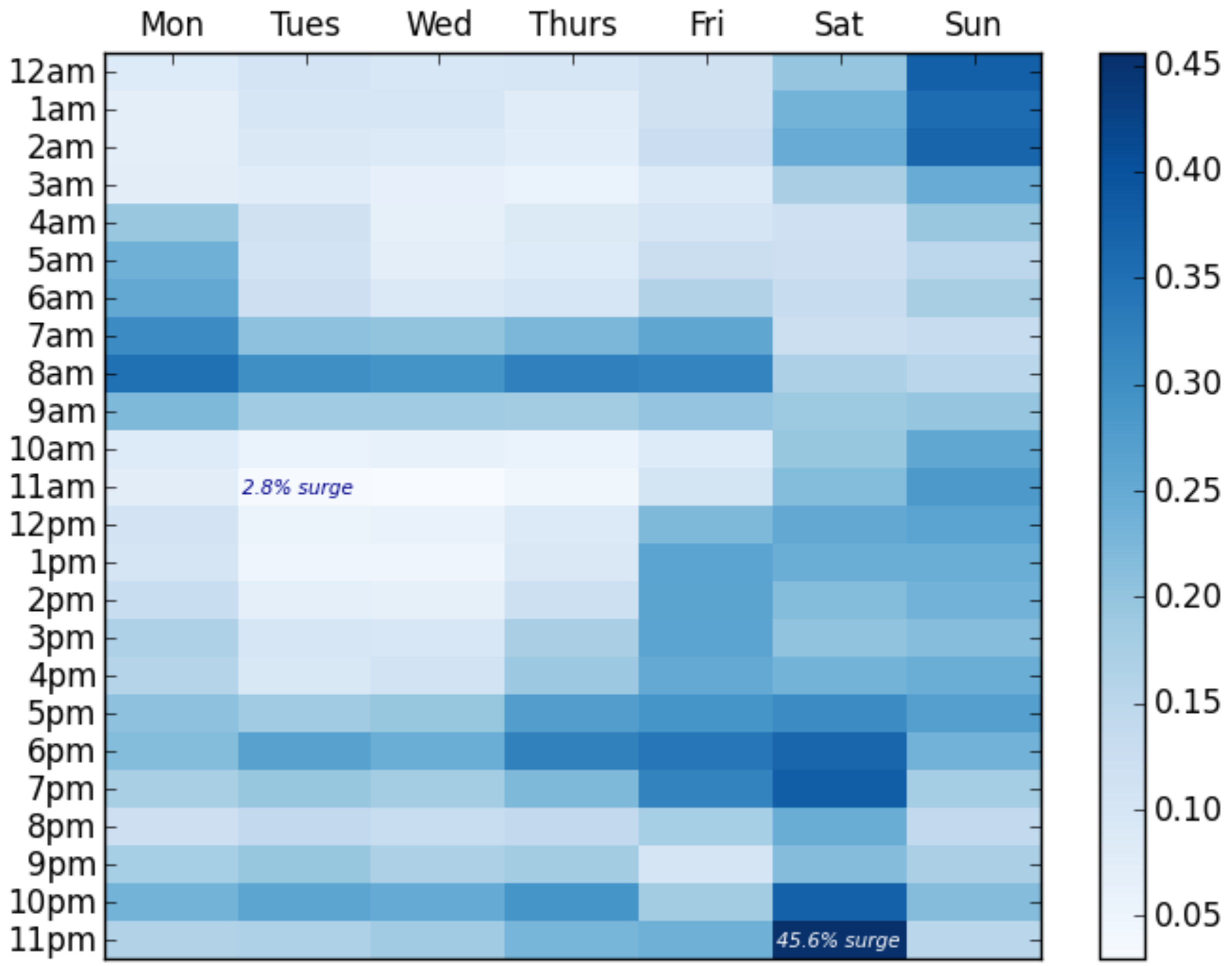
[43] These pictures are representative. In practice, there are variations across geography and time in terms of visual layout and exact product availability.

Figure 2: Distribution of surge price sessions for surge prices greater than 1.0x



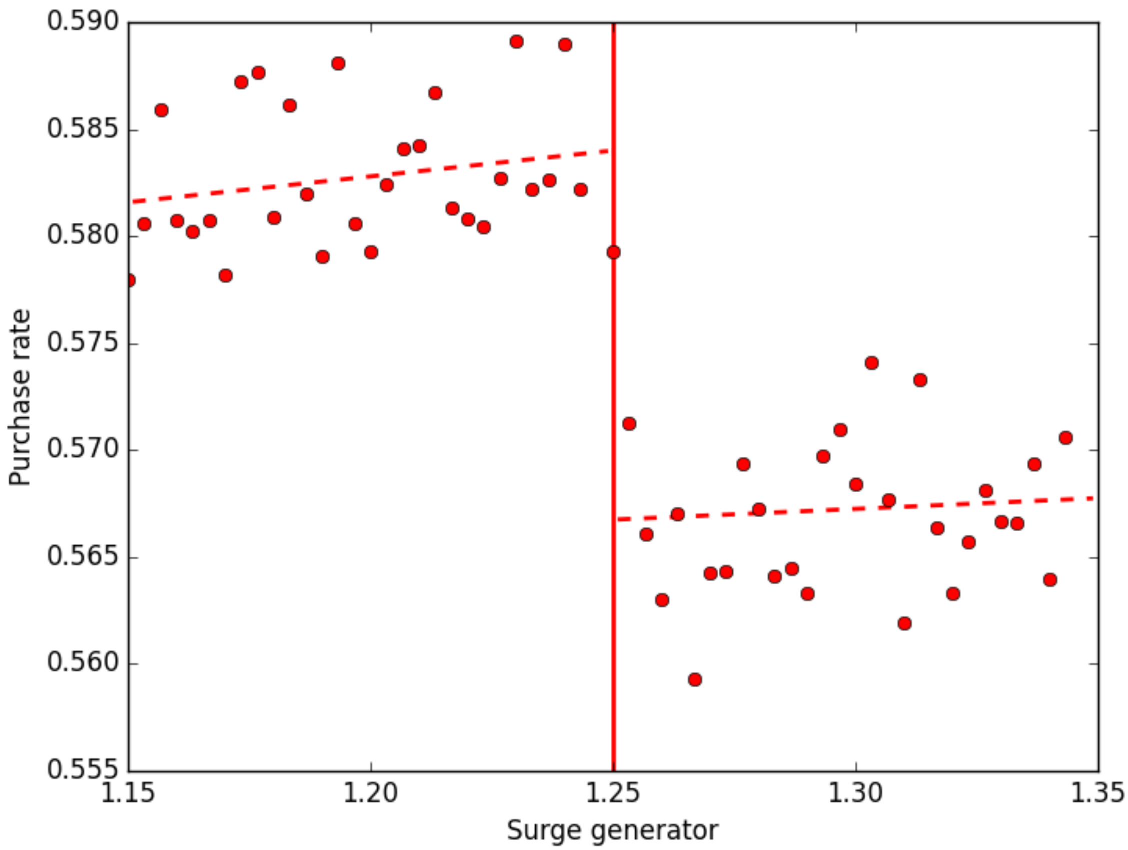
Note: This figure presents the number of observed UberX surge prices by surge level. Rides with no surge are excluded. Surge price notation is abbreviated. For example 1.2 in the graph corresponds to a surge price of 1.2x.

Figure 3: Heat map of the percent of sessions with surge by hour of the week



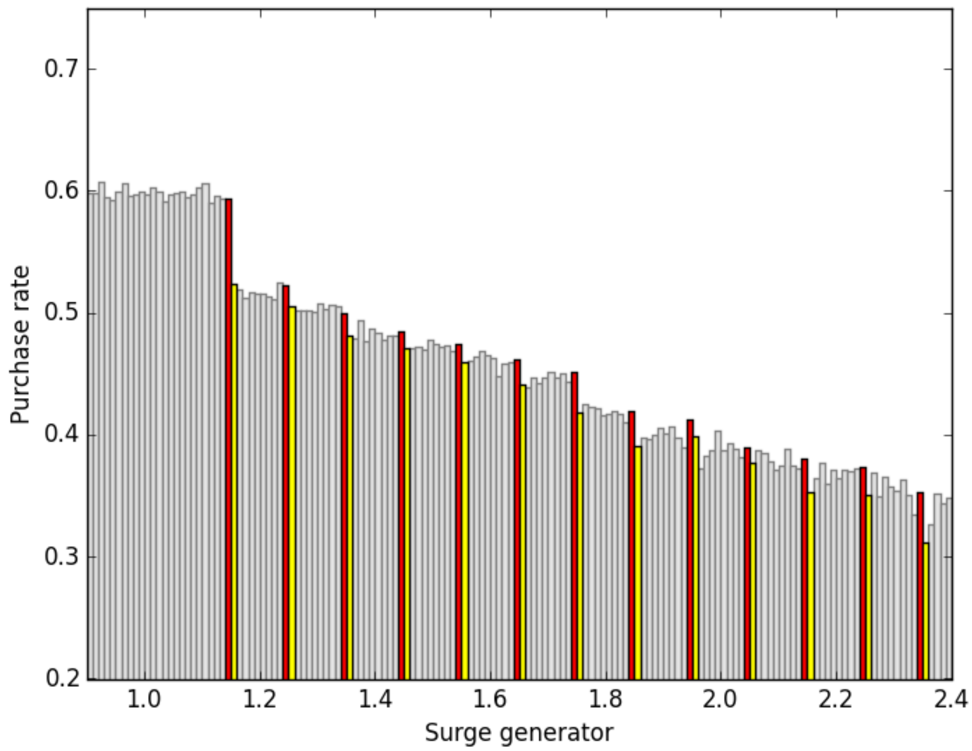
Note: This figure shows the frequency of surges by hour of day and day of week for UberX. Darker rectangles identify times and days when riders are more likely to face surge pricing. Tuesday at 11am represents the time and day combination when surge pricing is least common, and Saturday at 11pm represents the time and day combination when surge pricing is most common.

Figure 4: Example of purchase rate changes at price discontinuity



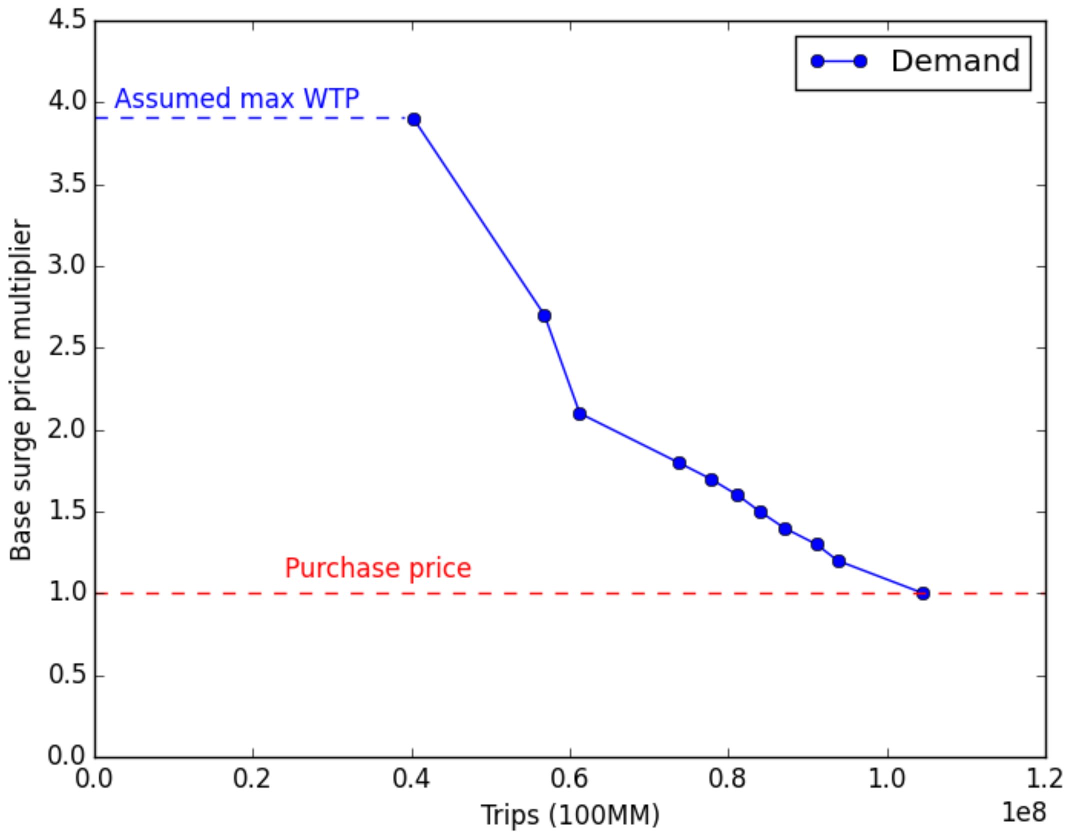
Note: This figure illustrates how purchase rates vary as a function of the surge generator over the range 1.15x to 1.35x. The vertical line when the surge generator equals 1.25 identifies the point at which the surge price changes from 1.2x to 1.3x.

Figure 5: Request rate drops at pricing discontinuities



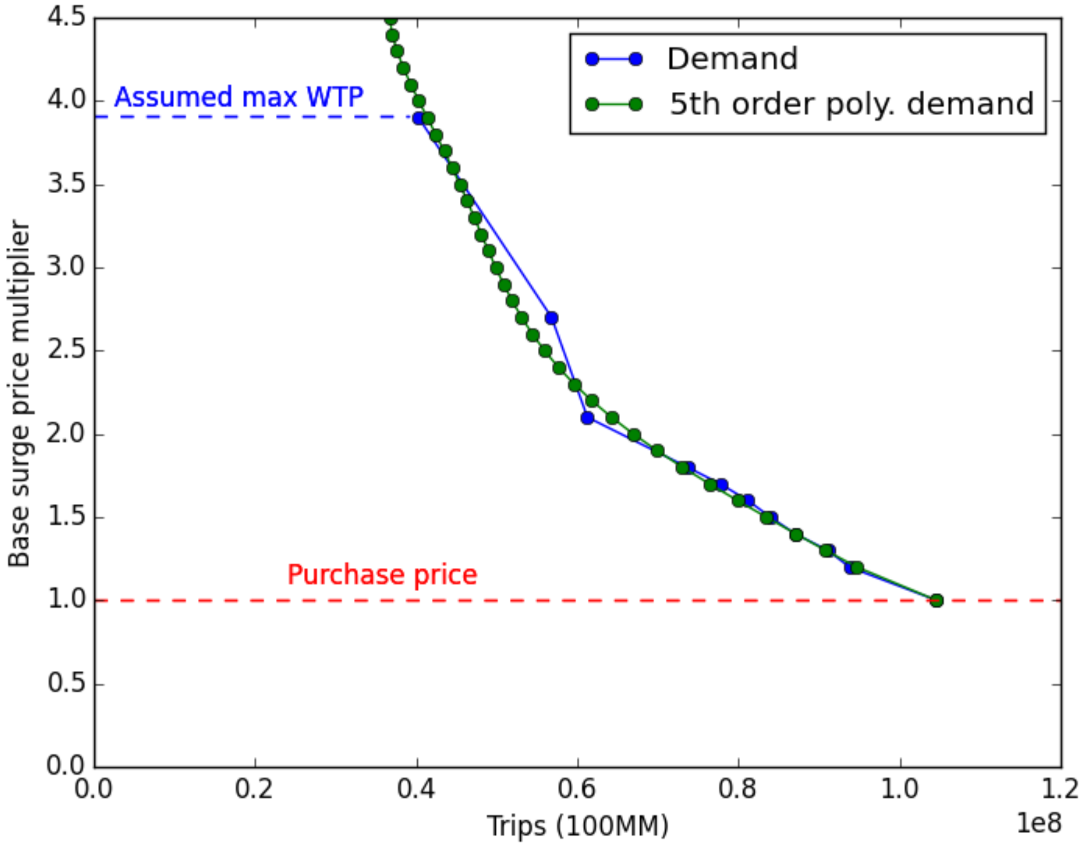
Note: This figure illustrates how purchase rates vary as a function of the surge generator when the surge generator is less than 2.4x. Red bars identify all observations within .01 units to the left of a price discontinuity. Yellow bars identify all observations within .01 units to the right of a price discontinuity. All observations not within these windows are depicted in gray.

Figure 6: Visual representation of demand curve for transactions at 1.0x



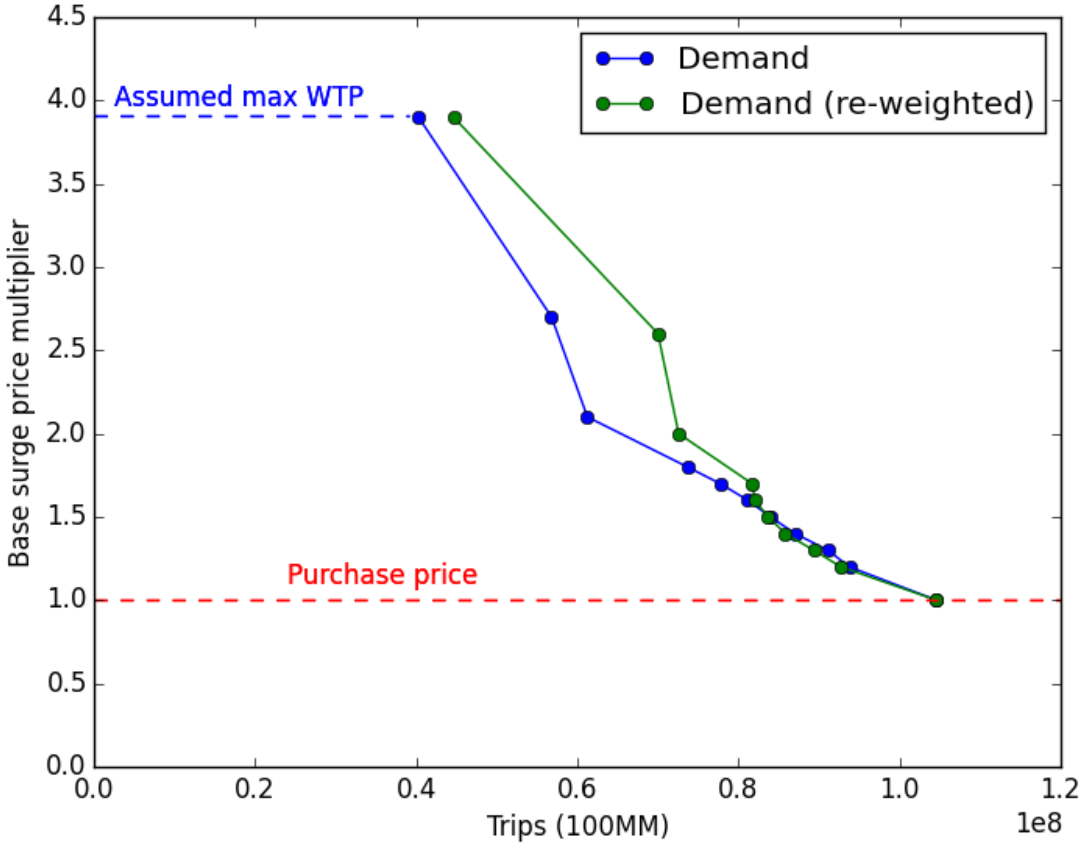
Note: This figure presents a piecewise linear demand curve with jumps at each price discontinuity. The curve is generated from the underlying elasticities estimated for each price discontinuity and for consumers facing transactions at 1.0x.

Figure 7: Comparison of 5th order smooth and linear segmentation demand curves



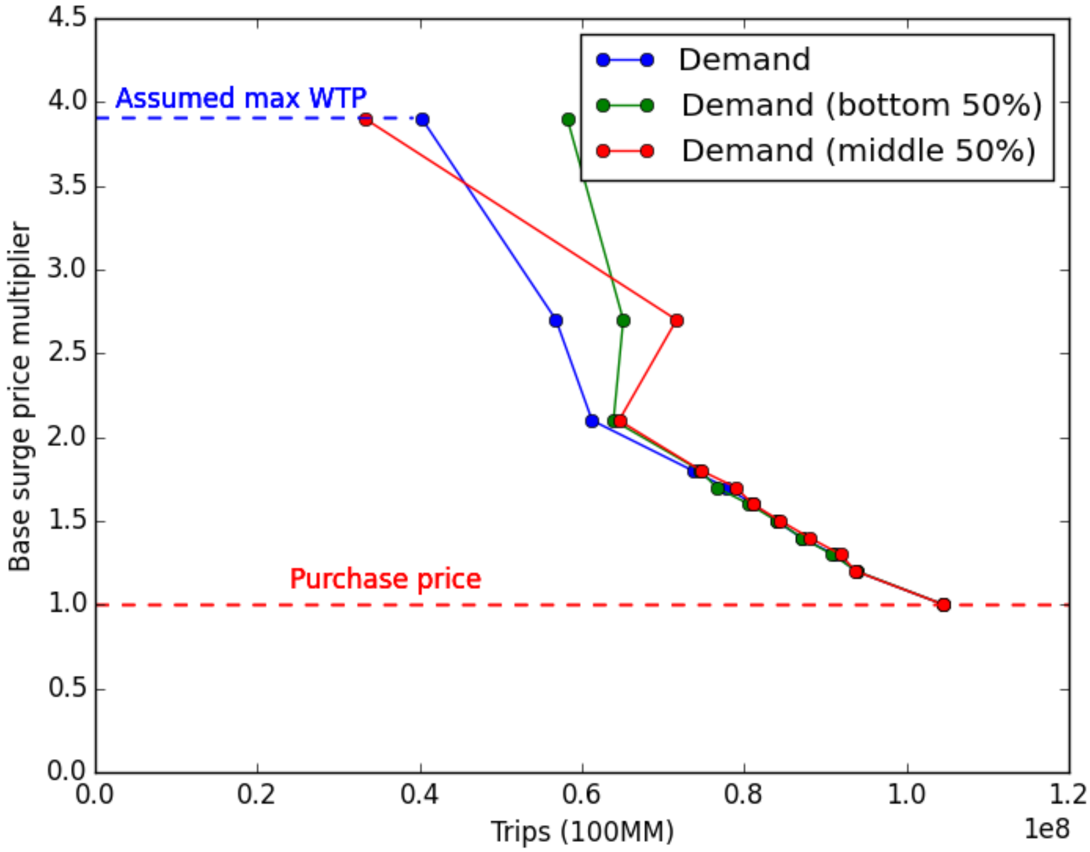
Note: This figure presents two demand curves generated via different approaches. The blue demand curve (also presented in Figure 6) is piecewise linear with jumps at each price discontinuity while the green demand curve is derived by fitting a 5^{th} *order polynomial to the elasticity estimates.*

Figure 8: Elasticity estimates with and without matching on observables



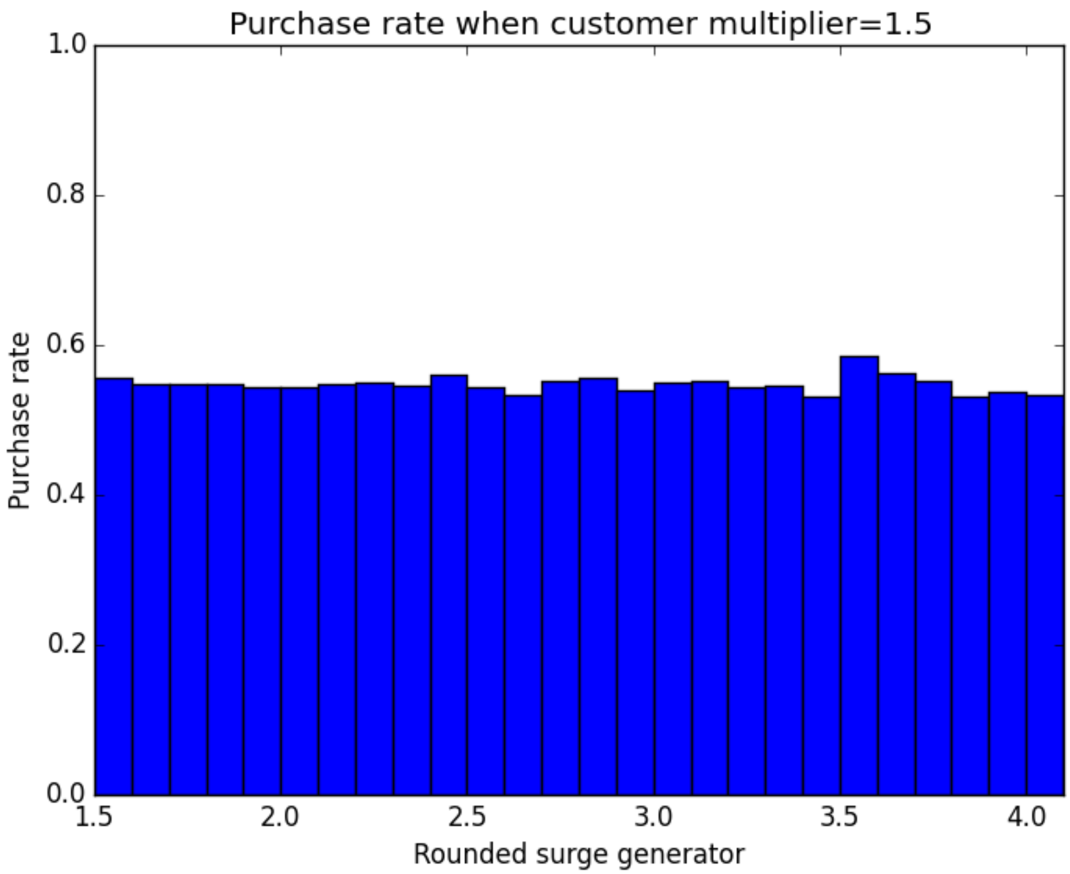
Note: This figure presents two demand curves generated via different approaches. The blue demand curve (also presented in Figures 6 and 7) is linear with jumps at each price discontinuity while the green demand curve is based on elasticity estimates derived from data that were re-weighted to match the distribution of observables found at price 1.0x.

Figure 9: Estimated demand curves with and without extreme demand shocks



Note: This figure presents three demand curves generated via different approaches. The blue demand curve (also presented in Figures 6, 7, and 8) is linear with jumps at the price discontinuities. The green demand curve restricts the analysis to observations where demand is below the median level, while the red demand curve restricts the analysis to observations where demand is around the median level.

Figure 10: Purchase rates as a function of underlying market conditions when the surge price is artificially restricted to 1.5x



Note: This figure presents the purchase rate by surge generator where the actual price observed by the rider is constrained to be 1.5x.

Tables

Table 1: Summary of sessions data

Table 1: Summary of Sessions Data

	Full Data	Surge = 1	$1 < \text{Surge} \leq 2.0$	Surge > 2.0
Surge	1.141	1.000	1.509	2.531
Expected wait time	4.118	4.205	3.731	4.046
Purchase rate	59%	62%	53%	39%
City				
Chicago	22%	20%	29%	32%
Los Angeles	25%	26%	20%	24%
New York	29%	31%	21%	19%
San Francisco	24%	22%	30%	25%
Time of Day				
Evening rush	8%	8%	10%	13%
Morning rush	6%	6%	7%	14%
Slow nighttime	12%	13%	10%	8%
Weekday day	23%	25%	15%	12%
Weekday evening	14%	15%	13%	10%
Weekend day	15%	14%	18%	17%
Weekend evening	6%	6%	7%	6%
Weekend event	15%	14%	20%	20%
Rides in Period				
1 ride in period	6%	6%	4%	4%
$1 <$ rides in period ≤ 3	9%	10%	8%	7%
$3 <$ rides in period ≤ 8	14%	15%	13%	13%
> 8 rides in period	71%	70%	75%	75%
Sessions	47469440	37667052	8135793	1666595

This table presents summary statistics for the analysis sample. Column (1) presents statistics for the entire sample. The remaining columns present statistics for three mutually exclusive and exhaustive subsets of the data: Column (2) presents rider sessions with baseline pricing; Column (3) presents rider sessions facing a moderate surge between 1.0x and 2.0x; and Column (4) presents rider sessions facing a surge greater than 2.0x.

Table 2: Observables Measured on Both Sides of the Price Discontinuity

	Post-Discontinuity
Request rate	-0.0201
	(0.0011)
Expected wait time	-0.1293
	(0.0142)
City	
Chicago	0.0073
	(0.0058)
Los Angeles	-0.0078
	(0.0044)
New York	0.0031
	(0.0023)
San Francisco	-0.0023
	(0.0054)
Time of Day	
Evening rush	0.0032
	(0.0037)
Morning rush	0.0031
	(0.0024)
Slow nighttime	-0.0048
	(0.0028)
Weekday day	0.0009
	(0.0033)
Weekday evening	-0.0003
	(0.0003)
Weekend day	0.0028
	(0.0036)
Weekend evening	-0.0012
	(0.0021)
Weekend event	-0.0023
	(0.0052)
Rides in Period	
1 ride in period	0.0007
	(0.0004)
$1 <$ rides in period ≤ 3	-0.0004
	(0.0005)
$3 <$ rides in period ≤ 8	0.0008
	(0.0006)
> 8 rides in period	-0.0012
	(0.0012)

This table presents OLS estimates for differences in key observables across price discontinuities. Standard errors are reported in parentheses.

Table 3: Estimated Price Elasticities at various Points along the Demand Curve

Table 3: Raw and RD Elasticity Estimates

Surge Threshold	(1)	(2)	(3)	(4)	(5)
1.2	-0.26	-0.43	-0.52	-0.52	-0.52
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
1.3	-0.32	-0.31	-0.35	-0.36	-0.34
	(0.01)	(0.04)	(0.04)	(0.04)	(0.04)
1.4	-0.42	-0.47	-0.53	-0.53	-0.58
	(0.01)	(0.05)	(0.05)	(0.05)	(0.05)
1.5	-0.42	-0.47	-0.50	-0.50	-0.49
	(0.02)	(0.05)	(0.05)	(0.05)	(0.05)
1.6	-0.33	-0.33	-0.43	-0.45	-0.50
	(0.02)	(0.06)	(0.06)	(0.06)	(0.06)
1.7	-0.62	-0.60	-0.66	-0.68	-0.68
	(0.03)	(0.08)	(0.08)	(0.08)	(0.08)
1.8	-0.73	-0.80	-0.85	-0.88	-0.89
	(0.03)	(0.10)	(0.10)	(0.10)	(0.10)
1.9 - 2.3	-0.77	-0.99	-1.02	-1.06	-1.01
	(0.02)	(0.07)	(0.07)	(0.07)	(0.07)
2.4 - 3.0	-0.37	-0.34	-0.38	-0.39	-0.25
	(0.05)	(0.17)	(0.17)	(0.17)	(0.17)
3.1 - 5.0	-0.72	-0.61	-0.75	-0.78	-0.65
	(0.14)	(0.46)	(0.46)	(0.46)	(0.46)
Source of identification	All variation	RD only	RD only	RD only	RD only
Control for wait time	No	No	Yes	Yes	Yes
Instrument for wait time	No	No	No	Yes	Yes
Additional controls	No	No	No	No	Yes

This table presents price elasticity estimates for each discontinuous jump in price. The surge prices on the left-most column indicate the higher of the two surge prices represented at the threshold e.g. the estimate created using the threshold between 1.2 and 1.3 would be labelled 1.3. Column (1) reports the raw elasticity estimates; Column (2) reports RD elasticity estimates; Column (3) reports RD estimates with expected wait time as a control; Column (4) reports RD estimates where expected wait time serves as an instrument; Column (5) reports instrumented RD estimates with additional controls for city and time of week. Standard errors are reported in parentheses. In each specification, the bottom three rows are the inverse variance weighted estimates taken from multiple discontinuities.

Table 4: Price Elasticities for Various Sub-Populations

	Elasticity
All periods	-0.5463
	(0.0209)
City	
Chicago	-0.6618
	(0.0367)
Los Angeles	-0.3270
	(0.0458)
New York	-0.6084
	(0.0514)
San Francisco	-0.5244
	(0.0378)
Time of Day	
Evening rush	-0.4994
	(0.0659)
Morning rush	-0.5206
	(0.0776)
Slow nighttime	-0.5254
	(0.0675)
Weekday day	-0.4611
	(0.0557)
Weekday evening	-0.5418
	(0.0571)
Weekend day	-0.6615
	(0.0540)
Weekend evening	-0.5388
	(0.0775)
Weekend event	-0.5480
	(0.0433)
Rides in Period	
1 ride in period	-0.4374
	(0.1398)
$1 <$ rides in period ≤ 3	-0.3822
	(0.0914)
$3 <$ rides in period ≤ 8	-0.4603
	(0.0642)
> 8 rides in period	-0.5702
	(0.0229)

This table presents elasticity estimates for sub-populations within the data. The RD regression that generates these estimates instruments for expected wait time and includes all relevant controls; this specification corresponds to that presented in Column (5) of Table 3. Standard errors are reported in parentheses.

Table 5: Comparison of surge with and without positive demand shocks included

Table 5: Comparison of Surge in General and Demand Neutral Populations

	Mean Surge	90th Percentile	Sessions above 2.0	Sessions above 3.0
Full data	1.141	1.5	1,666,595	217,934
Middle demand data	1.070	1.3	290,328	13,334
Low demand data	1.071	1.3	253,931	11,266

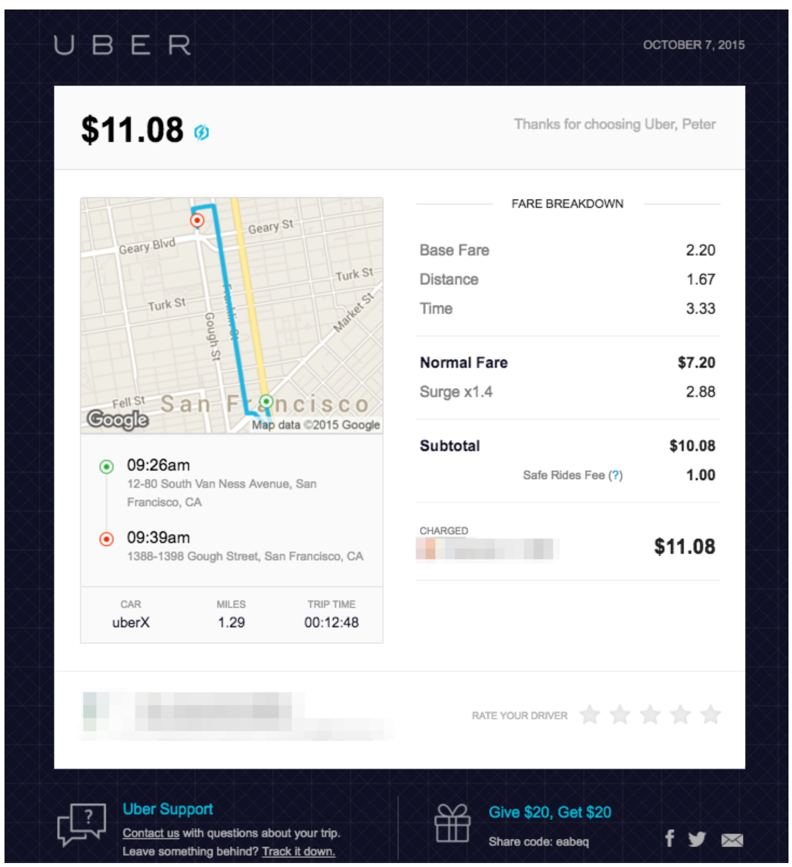
This table presents statistics for the full sample and for rides with observed demand around and below predicted demand. Predictions are generated using OLS estimates for the effect of week and hour of week on demand by location. Column (1) presents the mean surge in the given locations; Column (2) presents the surge at the 90th percentile; Column (3) presents the number of sessions with surge > 2.0; Column (4) presents the number of sessions with surge > 3.0.

Appendices

Time category definitions

Morning Rush = Monday - Friday, between 5:00am and 7:00am
Weekday day = Monday - Friday, between 6:00am and 5:00pm (excluding both rush hours)
Evening Rush = Monday - Friday, between 5:00pm and 7:00pm
Weekday Evening = Monday - Friday, between 7:00pm to 11:00pm
Weekend day = Saturday - Sunday, between 6:00am and 5:00pm
Weekend Evening = Saturday - Sunday, between 6:00pm to 11:00pm
Bar Hours = Thursday 11:00pm to 11:59pm, Friday 12:00am to 3:00am and 11:00pm to 11:59pm, Saturday 12:00am to 3:00am and 11:00pm to 11:59pm, Sunday 12:00am to 3:00am

Example UberX fare



Note: This figure presents the email riders receive upon completion of their ride which details the fare breakdown and additional information about the trip.

First stage expected wait time regression

To estimate the effect of wait times on purchase rates we use the following two specifications, which mirror specifications (4) and (5) in Table 3:

$$Purchase \sim \alpha + \ \beta_1 * Surge + \ \delta_2 * Wait\ Time + \ \varepsilon$$

$$Purchase \sim \alpha + \ \beta_1 * Surge + \ \delta_2 * Wait\ Time + \ \beta_3 * FE(city) + \beta_4 * FE(day\ and\ hour) + \ \varepsilon$$

$$Purchase \sim \alpha + \ \beta 1 * Surge + \theta 2 * Wait\ Time + \beta 3 * FEcity + \beta$$

In both cases, we instrument for both Surge and Wait Time (measured in minutes). The surge instrument is equal to -.5 when the generator is below a threshold by less than .01, .5 when the generator is above a threshold by less than .01, and 0 otherwise. The wait time instrument follows the same pattern but uses a 3 second window (e.g. if the wait time estimate in seconds was only 2 seconds below the threshold for a higher Wait Time (in minutes), then the instrument would equal -.5. The coefficient of interest in the first specification, δ_2 , is estimated to be -.0169 (SE = .0008) and in the second specification $\theta_2$is estimated to be -.01724 (SE = .0447).

Within surge level estimator specification

Using the full range surges within the data, though sampling according to footnote 40, we estimate the following regression to understand the relationship between the surge generator and purchase rates, holding surge constant:

$$Purchase \ = \ \alpha + \ \theta * generator + \beta_2 * FE(surge) + \varepsilon$$

where **θ** is the coefficient of interest.