The Paper of How: Estimating Treatment Effects Using the Front-Door Criterion*

Marc F. Bellemare† Jeffrey R. Bloem‡ Noah Wexler§

January 9, 2024

Abstract

We illustrate the use of Pearl's (1995) front-door criterion with an application in which the assumptions for point identification hold with observational data. For identification, the front-door criterion leverages exogenous mediator variables on the causal path. After a preliminary discussion of the identification assumptions behind and the estimation framework used for the front-door criterion, we present an empirical application. In our application, we look at the effect of deciding to share an Uber or Lyft ride on tipping by exploiting the algorithm-driven exogenous variation in whether one actually shares a ride conditional on authorizing sharing, the full fare paid, and origin–destination fixed effects interacted with two-hour interval fixed effects. We find that most of the observed negative relationship between choosing to share a ride and tipping is driven by customer selection into sharing rather than by sharing itself. In the appendix, we explore the consequences of violating the identification assumptions for the front-door criterion.

Keywords: Front-Door Criterion, Causal Inference, Causal Identification, Treatment Effects, Ride-Hailing

JEL Codes: C13, C18, R40, D90

---

*We thank three anonymous reviewers as well as the editor Climent Quintana-Domeque for comments which have significantly improved this manuscript. We also thank Chris Auld, David Childers, Carlos Cinelli, Dave Giles, Paul Glewwe, Adam Glynn, Paul Hünermund, Guido Imbens, Jason Kerwin, Dan Millimet, Judea Pearl, Bruce Wydick, J. Wesley Burnett, conference participants at the annual meeting of the Canadian Economics Association, the Causal Data Science Meeting, and the Latin American, North American Winter, and North American Summer Meetings of the Econometric Society as well as seminar participants at Berkeley, Michigan State, the Montréal Methods Workshop, Idaho, New Mexico, Tennessee, and the World Bank for useful comments and suggestions. All remaining errors are ours. The code for all figures and tables is available at `www.marcfbellemare.com/wordpress/research`.

†Corresponding Author. Distinguished McKnight University Professor, Distinguished University Teaching Professor, and Northrop Professor, Department of Applied Economics, University of Minnesota, 1994 Buford Avenue, Saint Paul, MN 55108, Email: `mbellema@umn.edu`.

‡Research Fellow, International Food Policy Research Institute, 2101 I ST NW, Washington, DC, 20005. Email: `j.r.bloem@cgiar.org`.

§Ph.D. Student, Humphrey School of Public Affairs, University of Minnesota. 301 19th Ave. S, Minneapolis, MN 55455, Email: `wexle059@umn.edu`.

1 Introduction

We illustrate the use of Pearl's (1995, 2000) front-door criterion (FDC) with an application using observational data in which the required assumptions *for point-identification* hold. The directed acyclic graph (DAG) in Figure 1 illustrates the front-door criterion setup, where the reduced-form relationship between outcome variable Y and treatment variable X is biased because of the presence of unobserved confounders U, which cause both X and Y.[1]

FIGURE 1: The Front-Door Criterion

$$\begin{tikzcd}
X \arrow[r, "\gamma"] & M \arrow[r, "\delta"] & Y \\
& U \arrow[ul] \arrow[ur] &
\end{tikzcd}$$

Pearl's insight is that when there exists a mediator variable M on the causal path from X to Y and that mediator is not directly caused by U, it is possible to estimate the causal effect of X on Y.[2] This is done by (i) estimating the effect γ of X on M (which is identified because the unobserved confounders in U cause X but not M), (ii) estimating the effect δ of M on Y conditional on X (which is identified because the unobserved confounders in U cause Y but not M), and (iii) multiplying the estimates $\hat{\gamma}$ and $\hat{\delta}$ by each other. This last step yields the causal effect of X on Y, which we label $\hat{\beta}$ in keeping with convention. Intuitively, the front-door criterion estimates a causal effect because it decomposes a reduced-form relationship that is not causally identified into two causally identified relationships.

Despite its relative simplicity, economists have been reluctant to incorporate the front-door criterion in their empirical toolkit.[3] Anecdotally, that resistance appears to stem from

[1]For readers unfamiliar with DAGs, directed arrows (i.e., $\rightarrow$) represent causal relationships between variables (Heckman and Pinto 2015). In a DAG, $X \rightarrow Y$ simply means that $Y = f(X, e_Y)$, where e_Y is an error independent of either X or Y. In a DAG, the causal relationship $X \rightarrow Y$ flowing from X to Y need not be parametric or linear (Morgan and Winship 2015), but we focus in this paper on parametric, linear relationships, both for brevity and because linear regression is the workhorse of applied economics, as noted by Imbens (2015).

[2]Although the literature often refers to M as a mechanism, we will refer to it as mediator in this paper. Our view is that the "mechanism" terminology is more accurate when discussing the theoretical framework underlying an application of the front-door criterion, and the "mediator" terminology is more accurate when discussing the statistical setup of the front-door criterion, as we do in this paper. Pearl (1993) originally called M a "mediating instrumental variable" and described M as an "exogenously-disturbed mediator."

[3]One exception to this reluctance is the use of insights from the front-door criterion to project the long run effects of programs based on a shorter-run outcome, or "surrogate index" (Athey et al. 2019). Variants of

the fact that finding an empirical application where the required assumptions for point identification credibly hold has thus far proven elusive (see, e.g., Imbens 2020; Gupta et al. 2020).

In his writings on the front-door criterion, Pearl repeatedly provides the same example of an empirical application. In his canonical example, X is a dummy variable for whether one smokes, Y is a dummy variable for whether one develops lung cancer, and M is the accumulation of tar in one's lungs (Pearl 1995; Pearl 2000; Pearl and Mackenzie 2018). But some have pointed out that if (i) smoking has a direct effect on lung cancer, independent of tar accumulation or (ii) both tar accumulation and lung cancer are caused by alternative sources, such as a hazardous work environment, then this canonical example violates the necessary front-door criterion identifying assumptions (see, e.g., Imbens 2020). The only extant published social science applications of the front-door criterion are by Glynn and Kashin (2017; 2018).[4] We build on these previous contributions, although it is important to note that the authors of those previous studies themselves admit that the necessary assumptions required for point identification with the front-door criterion approach do not hold.[5]

Our contribution is threefold. First and foremost, we present an application of the front-door criterion using observational data where the necessary assumptions for point-identification hold.[6] In that application, we estimate the effect of authorizing a shared Uber or Lyft ride on tipping behavior. The mere authorizing of a shared ride does not guarantee a shared ride. Rather, whether one gets to share a ride or not depends on many supply- and demand-side factors, such as how many other customers in one's vicinity are going in the same direction at the same time, and how many drivers are available at that time. We argue that conditional on distinct two-hour time slot fixed effects, origin-

this type of approach include efforts to estimate the long-run effect of Head Start (Kline and Walters 2016) or early-childhood education (Garcia et al. 2020).

[4]In Glynn and Kashin (2018) the authors apply the front-door criterion approach to estimate the effects of attending a job training program on earnings. In Glynn and Kashin (2017) the authors also apply the front-door criterion difference-in-differences approach to evaluate the effect of an early in-person voting program on voter turnout. Both applications closely approximate, but ultimately do not exactly replicate existing experimental estimates.

[5]Specifically, Glynn and Kashin (2018) write, "As we discuss in detail below, the assumptions implicit in [the front-door criterion] graph will not hold for job training programs, but this presentation clarifies the inferential approach." In Glynn and Kashin (2017), the authors develop a difference-in-difference extension to the front-door criterion approach which requires an exclusion restriction and a parallel trends assumption specifically for their empirical setting where the necessary conditions for the front-door criterion do not hold. This previous work is helpful in partially identifying the treatment effect of interest, thereby establishing reasonable bounds on effect estimates. In contrast, the treatment effects we estimate here are point-identified.

[6]In the Appendix A3, we also show an example that uses simulated data to illustrate an application of the front-door criterion in which we know the true effect.

destination fixed effects, and the interaction of the two-hour time slot fixed effects and the origin-destination fixed effects, we effectively control for those supply- and demand-side factors, which makes whether one gets to actually share a ride exogenous.[7,8] In our ride-sharing application, due to one-sided non-compliance inherent in the data generating process—that is, there is no variation in the mediator when the treatment is not taken up; in other words, a customer who chooses not to share a ride will always ride by herself and never share—the estimates obtained both with the front-door criterion approach and with a naïve OLS approach will estimate treatment effects that we characterize as estimates of an average treatment effect on the treated (ATT). This follows Glynn and Kashin (2018) and addresses some of the ambiguity discussed by those authors about the use of the term ATT in settings using observational data.

Exploiting this exogenous mediator for identification, we find that the observed negative relationship between authorizing a shared ride and tipping is almost entirely explained by customer selection into treatment. In other words, our finding first dispels the notion, common among Lyft and Uber drivers, that the decision to share a ride (rather than the type of person making that decision) leads to lower tips on shared rides (Bowman 2019; Harrington 2019). This finding suggests that ride-hailing companies like Lyft and Uber might be able to increase their drivers' tips on shared rides by making shared rides the default, thereby exploiting status quo bias (Samuelson and Zeckhauser 1986). Given that the 1.5 to 2 million Lyft or Uber drivers in the United States comprise between 0.9 and 1.2 percent of the US labor force (Berry 2021) and that ride-hailing apps are increasingly replacing old-fashioned taxicabs, our insights are broadly important for labor markets. In addition, with many settings in which more frugal consumers can choose a cheaper option associated with lower tipping percentages (e.g., daily specials in restaurants, happy hours in bars), our finding is relevant more broadly for the economics of tipping (Azar 2020).

In our application, we are not able to randomly assign whether someone authorizes shared rides (i.e., X) on Uber or Lyft and this choice is clearly endogenous to tipping

[7] Each distinct two-hour time slot in our data has its own fixed effect. Thus, the midnight to 2 AM time slot on January 1 has a fixed effect that is different from the midnight to 2 AM time slot on January 2.

[8] This battery of fixed effects allows the variation of whether one gets to share a ride conditional on having authorized a shared ride to be exogenous in situations where, for example, a flight out of O'Hare gets canceled and most of its passengers decide to head to Midway at the same time by hailing rides on Lyft or Uber. Perhaps more importantly, because origin-time or destination-time fixed effects are included in the interaction of origin–destination fixed effects interacted and two-hour time slot fixed effects, the same battery of fixed effects also allows the variation of whether one gets to share a ride conditional on having authorized a shared ride to be exogenous in situations where, for example, a baseball game ends at Wrigley Field and those in attendance all head to different neighborhoods at the same time by hailing rides on Lyft or Uber.

behavior (i.e., Y). Riders who authorize sharing are likely drawn to do so by the lower fare, and so they are also likely to tip less. Additionally, there may be specific times and locations where hailing an Uber or Lyft costs more due to surge pricing. We apply the front-door criterion to this application by making use of the following fact: Once a passenger authorizes a shared a ride, they will not necessarily share a ride. We can therefore exploit the exogenous variation—conditional on the full fare paid and the battery of fixed effects described above—in whether or not a passenger actually shares a ride (i.e., M) once they authorize sharing. In that case, the front-door criterion can credibly estimate the causal effect of authorizing shared rides on tipping behavior (i.e., Y).[9]

Our application highlights a class of potential applications where the front-door criterion can be used to estimate important economic parameters. This class of applications for the front-door criterion consists of cases where there is exogenous non-compliance in some mechanism M that lies on the causal path between X and Y. Our ride-sharing application motivates other applications in the tech industry where users endogenously choose X and an algorithm determines some mechanism M before an outcome of interest Y is realized. In labor economics, Hull (n.d.) presents the idea of using the front-door criterion to estimate the effect of a worker having a criminal record X on employment Y, where a background check M may suffer from either a type I (i.e., false positive) or type II (i.e., false negative) error.[10] In agricultural economics, one can think of applications where endogenous variation in the use of some input X (e.g., fertilizer) causes output Y (i.e., yield) to change via some agro-ecological mechanism M (e.g., evapotranspiration) driven by exogenous weather conditions. In each of these potential applications—which span the economics of the tech industry, labor, and agriculture—the endogenous relationship between X and Y is mediated by a mechanism M that includes some form of non-compliance.

Second, though this is not a methodological paper, linear regression remains the workhorse of empirical economics, and because an explanation of how to use the front-door criterion in the context of linear regression has so far been lacking in the literature, we briefly explain how to estimate treatment effects with the front-door criterion in the con-

[9] One may be concerned that the fare differential between (more expensive) solo rides and (less expensive) shared rides is a channel from X to Y in Figure 1 because consumers may feel more generous the higher the fare differential. We explain in section 3 how our data and method rule out that channel. Further, our use of the tip percentage of full fare as a dependent variable theoretically avoids this problem. As we show in section 4, the results of these models are consistent with those of models that use an indicator for tipping (i.e., tipping at the extensive margin) and those that use the actual tip amount as outcomes (i.e., tipping at the intensive margin) respectively.

[10] Similar to our ride-sharing application, Hull (n.d.) notes that the plausibility of the exogenous variation in M, in practice, may be supported by conditioning on available covariates or leveraging some policy change (e.g., a "ban-the-box" initiative).

text of linear regression prior to presenting our application to ride-hailing data.[11] Estimation relies on Pearl's three identification assumptions and an additional, empirically verifiable data requirement which, depending on the form of non-compliance between the treatment and the mediator, determines whether the front-door criterion estimates the average treatment effect (ATE), the average treatment effect on the treated (ATT), or the average treatment effect on the untreated (ATU).[12]

Third, and importantly for applied researchers interested in using the front-door criterion in their own work, we explore in the appendix what happens in general when the necessary assumptions for the front-door criterion to identify the effect of X on Y fail to hold. Specifically, we examine what happens when (i) there are multiple mediators, some of which may be omitted from the estimation specification (Appendix A9), and (ii) the assumption of strict exogeneity of M is violated (Appendix A10).

The remainder of this paper is organized as follows. In section 2, we introduce the necessary point-identification assumptions of the front-door criterion and present a brief "how-to" for economists wishing to broaden their empirical toolkit by incorporating the front-door criterion. Section 3 presents our main empirical illustration using real-world data. In section 4, we conclude by reflecting on possible future applications that could use front-door criterion to estimate key economic parameters. In the Appendix we explore and discuss departures from some of the assumptions underpinning the front-door criterion and how these departures might influence applied research implementing the front-door criterion.

2 The Front-Door Criterion: Identification and Estimation

We begin by presenting the front-door criterion for those readers who may not be familiar with it (readers who are already familiar with the front-door criterion may wish to skip directly to the application in Section 3). We first present and expand upon the necessary identification assumptions noted by Pearl (1995, 2000), and demonstrate how these assumptions recover the causal effect of X on Y in the presence of unobserved confounders U, as shown in Figure 1 above. We then explain how to estimate treatment effects using the front-door criterion in a linear regression context.

[11] Although Morgan and Winship (2015) dedicate part of a chapter to the front-door criterion, they only indirectly present a regression-based approach to the front-door criterion.

[12] Although this contribution builds on the previous work by Chalak and White (2011), we clarify several details. For example, we show that the data requirement that $P(X_i|M_i) > 0$, as noted in Pearl (2000), is only necessary for the front-door criterion to estimate the ATE. In cases when this data requirement does not hold, the front-door criterion can estimate either the ATT or the ATU.

2.1 Identification

We are interested in estimating the causal effect of X on Y in Figure 1 above. Recall that with observational data, estimating the causal effect is complicated by the presence of unobserved confounders U, which give rise to the identification problem. Given the validity of a number of identifying assumptions, however, the front-door criterion approach pictured in Figure 1 allows for the recovering of the causal effect of X on Y.

As discussed in Pearl (1995, 2000), the front-door criterion requires that there exists a vector M which satisfies some assumptions relative to X and Y. In this discussion, we assume M is a single variable, and we further assume that M, X, and Y are binary variables for ease of exposition.

Let $Y(m, x)$ be the potential outcome if $M = m$ and $X = x$, and let $M(x)$ be the potential outcome if $X = x$. The assumptions for identification are as follows.

Assumption 1. $Y(m, 1) = Y(m, 0), \forall\, m \in \{0, 1\}$.

This assumption states that the only way in which X influences Y is through M. In Figure 1, this means that there should be no arrows bypassing M between X and Y. In Pearl's terminology, M should intercept all directed paths from X to Y.

Assumption 2. *$M(1)$ and $M(0)$ are independent of X.*

This assumption states that the relationship between X and M is not confounded by unobserved variables. That is, the coefficient γ in Figure 1 is identified. In Pearl's terminology, there can be no back-door path between X and M.

Assumption 3. $E[Y(m, x)|M, X = x] = E[Y(m, x)|X = x], \forall\, m \in \{0, 1\}$.

This assumption states that conditional on X, the relationship between M and Y is not confounded by unobserved variables. That is, the coefficient δ in Figure 1 is identified. In Pearl's terminology, every back-door path between M and Y has to be blocked by X.[13] In Appendix A1, we present Pearl's proof of identification using the front-door criterion (Pearl 1995) with additional explanations to help the reader's intuition.

In later writings on the front-door criterion, Pearl (2000) discusses an additional criterion for identification. That condition states that no matter what the value of the mediator M is for unit i, that unit must have a non-zero probability of getting treated. That is, $P(X|M) > 0$. In Pearl's canonical example of the relationship between smoking X

[13]This assumption also excludes the possibility of a "recanting witness," whereby X induces an unmeasured association between M and Y because a confounder of M is itself caused by exposure to X (Tchetgen Tchetgen and VanderWeele 2014; Naimi 2015).

and lung cancer Y, this condition implies that the amount of tar in the lungs of smokers M must be the result not only of smoking, but also of other factors (e.g., exposure to environmental pollutants), and that tar be absent from the lungs of some smokers (say, because of an extremely efficient tar-rejecting mechanism). This requirement, which we call the full-support requirement, is necessary for the front-door criterion to estimate the ATE. Failure to meet this condition due to a lack of full support can result in estimating alternative parameters such as the average ATT or ATU.

To see this, consider the case where both X and M are binary variables. The full support requirement, $P(X|M) > 0$, states that we must observe that $P(X = 1|M = 1) > 0$, $P(X = 1|M = 0) > 0$, $P(X = 0|M = 1) > 0$, and $P(X = 0|M = 0) > 0$ to estimate the ATE with the front-door criterion. If this condition fails to hold, then the front-door criterion will not estimate the ATE. If a slightly weaker variation of this condition holds, however, then the front-door criterion will estimate either the ATT or the ATU. In particular, as is the case in our application, we observe that $P(X = 1|M = 1) > 0$, $P(X = 1|M = 0) > 0$, and $P(X = 0|M = 0) > 0$ but not $P(X = 0|M = 1) > 0$ due to one-sided non-compliance. Therefore, because in this case, we only observe individuals with $M = 1$ if $X = 1$, then the front-door criterion estimates the ATT. We present a brief mathematical sketch of this idea in Appendix A2.

2.2 Estimation

We now discuss how to estimate treatment effects using the front-door criterion. As stated above, our goal is to estimate the causal effect of X on Y in Figure 1. When the necessary identification assumptions for the front-door criterion to hold, we can estimate the causal effect by using the following linear regression-based approach. Let

$$M_i = \kappa + \gamma X_i + \omega_i \tag{1}$$

and

$$Y_i = \lambda + \delta M_i + \phi X_i + \nu_i. \tag{2}$$

In Equation 1, following Assumption 2 which states that the only way in which X influences Y is through M, the relationship between X and M is identified. In Equation 2, Y is the outcome variable, which is related to X only through M. This follows assumptions 1 and 3, which together imply that the only way X influences Y is through M and conditional on X the relationship between M and Y is not confounded. Therefore, estimating Equations 1 and 2 and multiplying coefficient estimates $\hat{\delta}$ and $\hat{\gamma}$ by each other yields β, the causal effect of X on Y.

At this point, it is important to make a few clarifying remarks. First, we focus here on the context of linear regression because linear regression is the approach favored by the majority of applied economists.[14] We note, however, that although we have written Equations 1 and 2 as linear equations, directed acyclic graphs such as the one in Figure 1 impose no such linear relationships on their constituent variables, nor do they impose that the relationships be parametric.[15] Therefore the front-door criterion is nonparametrically identifiable, and linear regression is but one way to estimate treatment effects using the front-door criterion. In Appendix A8, for instance, we present nonparametric estimation results for our real-world example

Second, the necessary point-identification Assumptions 1 through 3 lead to estimating δ and γ, and via multiplication, β, the causal effect of X on Y. Given an additional conditional ignorability assumption (Rosenbaum and Rubin 1983), these equalities can be achieved by conditioning on a vector of control variables. This is akin to conditional excludability in instrumental variable estimation (Angrist and Kruger 1995), and we illustrate this result directly in Section 3.

Finally, in our applications, we estimate the front-door criterion using a seemingly unrelated regressions (SUR) framework (Zellner 1962). Although the SUR framework is not strictly necessary to estimate treatment effects using the front-door criterion, it does have some useful features, such as ease of computation of the overall treatment effect. For each set of results, we present both heteroskedasticity-robust standard errors as well as standard errors clustered at the level of origin-destination pair interacted with two-hour time slots (i.e., the same level as our most granular fixed effects), the latter being consistent with the recommendations in Abadie et al. (2023).

3 Empirical Application: Ride Sharing and Tipping

We now illustrate the use of the front-door criterion with an application using observational data wherein the required assumptions hold for point-identification. Additionally, in Appendix A3, we present and discuss simulation results where we know the true effect and show that the front-door criterion recovers this effect when the required assumptions for point-identification hold. These simulation results support the results in the previous

[14]That said, note that the linear estimator will not be consistent without conditions beyond those described in the previous section. A sufficient condition is linearity of the conditional expectation functions $E(Y|M, X)$ and $E(M|X)$.

[15]See Morgan and Winship (2015) for an introduction to directed acyclic graphs as they are used in causal inference, and see Pearl (2000) for an in-depth treatment.

section and serve as a baseline example that enables our investigation of departures from the ideal case, which we present in the Appendix.

Using publicly available data on over 95 million Lyft and Uber rides in Chicago during the calendar year of 2019, we apply the front-door criterion to estimate the effect of authorizing a shared ride on tipping at both the extensive (i.e., whether the passenger tips) and intensive margins (i.e., how much the passenger tips), as well as on tipping as a proportion of the fare paid. After discussing our data, we explain how the necessary conditions for the front-door criterion to yield a consistent point estimate hold in this setting after conditioning on relevant observed variables and a battery of time-and-place fixed effects that include origin–destination fixed effects, two-hour time slot fixed effect, and their interactions.

We find that naïve regressions overestimate the magnitude of the effect of authorizing sharing on both tipping margins because of selection into treatment.[16] Our application illustrates a broader principle: shared rides are cheaper and less convenient services compared to solo rides, and so there is a tradeoff between fare and (expected) inconvenience which in principle affects tipping.

The front-door criterion is particularly useful in this context because it can help reliably rule out the effect of selection into a less convenient service on tipping, conditional on the lower fare. As we show, a naïve OLS specification that conditions on observables—even one that conditions on the fares of the two competing services—still overestimates the effect due to bias associated with more frugal customers selecting the cheaper service. Because the front-door criterion estimates unbiased treatment effects, it can also indirectly show the extent to which the gap in tipping between sharing-authorized rides and solo rides is associated with rider self-selection into authorizing shared rides. This information may be useful to drivers making marginal decisions about whether to pick up potential passengers who have authorized sharing.

[16]Using an earlier version of the same data we use here, Harrington (2019) reports that when riders opt to share rides with another passenger, they are less likely to tip. In the context of the potential outcomes model (Rubin 2005), this finding is problematic because it fails to account for selection into treatment. Customers who are frugal are both less likely to tip and more likely to authorize sharing, enticed by lower fares. Alternatively, some consumers may authorize shared rides because it is better for the environment, and those same consumers might also be more likely to be socially conscious and tip drivers because of the precarity of jobs in the gig economy. To effectively infer the effect of authorizing a shared ride on tipping, one must deal with the endogeneity associated with selection of people with a lower propensity to tip into authorizing shared rides. An unbiased effect would capture the difference in tipping if authorizing a shared ride was randomly distributed across all passengers, no matter their proclivity to tip.

3.1 Data

Our data include 95.6 million dedicated (i.e. standard, single-transaction) Uber and Lyft rides and sharing-authorized Uber and Lyft rides taken from January 1 to December 31, 2019. The data come from the Chicago Department of Business Affairs and Consumer Protection's Transportation Network Providers Data Portal.[17] Each observation represents a single transaction on either app. These data show whether the passenger authorized a shared ride (i.e., X), whether the passenger actually shared a ride with another paying customer (i.e., M), and the passenger's tipping behavior at both the extensive and intensive margins (i.e, Y).[18]

These data include the base fare (rounded to the nearest \$2.50) and tip amount (rounded to the nearest \$1.00).[19] For the extensive margin of tipping, our dependent variable is a dummy variable for whether a passenger tips or not. For the intensive margin, we use the observed tip value.[20] For tip as a percentage of the fare, we simply divide observed tip value by the full fare paid.[21] As we discuss in more detail below, we control for the observed base fare in all of our analysis (i.e., in the naïve OLS as well as in both stages of front-door criterion estimation).

The data also include detailed information on ride time-and-place, including the origin and destination community area. We interact origin–destination fixed effects with each unique two-hour time slot across the year to generate a rich set of time–place cell fixed effects. Given that we have 5,929 origin-destination pairs (77 areas as origins, and 77 areas as destinations), 4,380 distinct time slots (i.e., 365 days, each with 12 two-hour slots) and that we interact each origin-destination pair with all time slots, we end up with $25,969,020 + 5,929 + 4,380 = 25,979,329$ fixed effects. Obviously, since the number of fixed effects equals about a quarter of our sample size, many "bins" (i.e., distinct cells in the data defined by values of the explanatory variables) are empty; see Appendix A4 for

[17] The data are one of a few publicly available data sets on transportation network company trips and have been collected since November 2018. They can be downloaded via the City of Chicago's website.

[18] The data show the number of overlapping sharing-authorized rides a given ride occurred within. Specifically, this field counts how many individual passengers were transported between any two points in time during which the car was occupied by passengers. Any number over one indicates that a ride was shared with at least one other passenger.

[19] We discuss the measurement error introduced by these rounding schemes below, when interpreting our results. We also drop observations with fare level under \$2.50 and over \$50 to analyze a reasonable range of fares.

[20] To account for the high number of zero observations (indicating that a passenger did not tip), and because we would ideally want to take the logarithm of tip value, we apply the inverse hyperbolic sine (i.e., arcsinh) transformation, a log-like transformation which allows to keep zero-valued observations, before calculating elasticities (Bellemare and Wichman 2019; see Card et al. 2020). Following Mullahy and Norton (2022) we also show results with the non-transformed level of the dependent variable in Appendix A5.

[21] We also apply the inverse hyperbolic sine transformation to this outcome variable.

TABLE 1: Summary Statistics

	Ride Type	Total Charge ($) Mean	Total Charge ($) Std. Dev.	Tip ($) Mean	Tip ($) Std. Dev.	Tipped (Dummy) Mean	Tipped (Dummy) Std. Dev.	Observations N	Observations % Total
Full Sample	Dedicated	13.388	(7.605)	0.592	(1.455)	0.204	(0.403)	76,980,046	80.5%
	Sharing Authorized	9.686	(5.269)	0.181	(0.698)	0.087	(0.282)	18,690,403	19.5%
Sharing Authorized	Shared	9.827	(5.365)	0.175	(0.683)	0.086	(0.280)	13,085,093	70.0%
	Not Shared	9.356	(5.024)	0.193	(0.731)	0.092	(0.288)	5,605,310	30.0%

a histogram of the distribution of observations within each bin. Finally, we also control for the full fare paid by the rider, i.e., the sum of the fare value and any additional charges levied on a rider.[22] We use the full fare instead of separating fare and additional charges because riders observe an itemized sum of the two fare components when selecting rides and when deciding on a tip value. Table 1 shows summary statistics.

3.2 Conceptual Framework

After opening their preferred ride-hailing app, a passenger is shown a menu of available services and given the choice to take a guaranteed solo ride (i.e., UberX or Lyft) or to authorize sharing (i.e., UberPool or LyftLine). This decision reflects rider preferences across a price–convenience trade-off. Passengers observe the fare they would be charged for each service, with fares for both services charged up front and calculated according to the probability a given passenger ends up sharing a hailed vehicle with a co-rider. Sharing-authorized rides are discounted relative to the single-passenger "base fare" such that (sharing-authorized) rides more likely to overlap with another passenger's trip are cheaper relative to the base fare. Though cheaper, a shared ride is also likely to be more inconvenient and time-consuming than a solo ride. This is because if a sharing-authorized ride ends up matched with another passenger's trip, the driver will likely have to make a detour to pick-up or drop off the co-passenger. The possibility of a detour reduces the utility of sharing-authorized rides, a phenomenon dubbed the "detour penalty" by Young et al. (2020). Additionally, sharing-authorized rides may be less desirable because riders simply want to spend their travel time alone (except for the driver) instead of with a co-rider. Figure 2 illustrates this decision-making process.

Thus, a rider's choice to authorize a shared ride is determined by their preferences along the price–convenience tradeoff (i.e., by whether they gather more utility from a guaranteed reduction in fare with pooling or a likely increase in convenience with a solo ride) and by the fare discount itself. The fare discount is determined by the time and place

[22] Additional charges are levied by the City of Chicago and, in 2019, were specifically set to increase the cost of taking rides in high-volume areas, viz. any ride with an origin or destination within the downtown area or a special area (Navy Pier, McCormick Place, or the airports). This variable is not rounded.

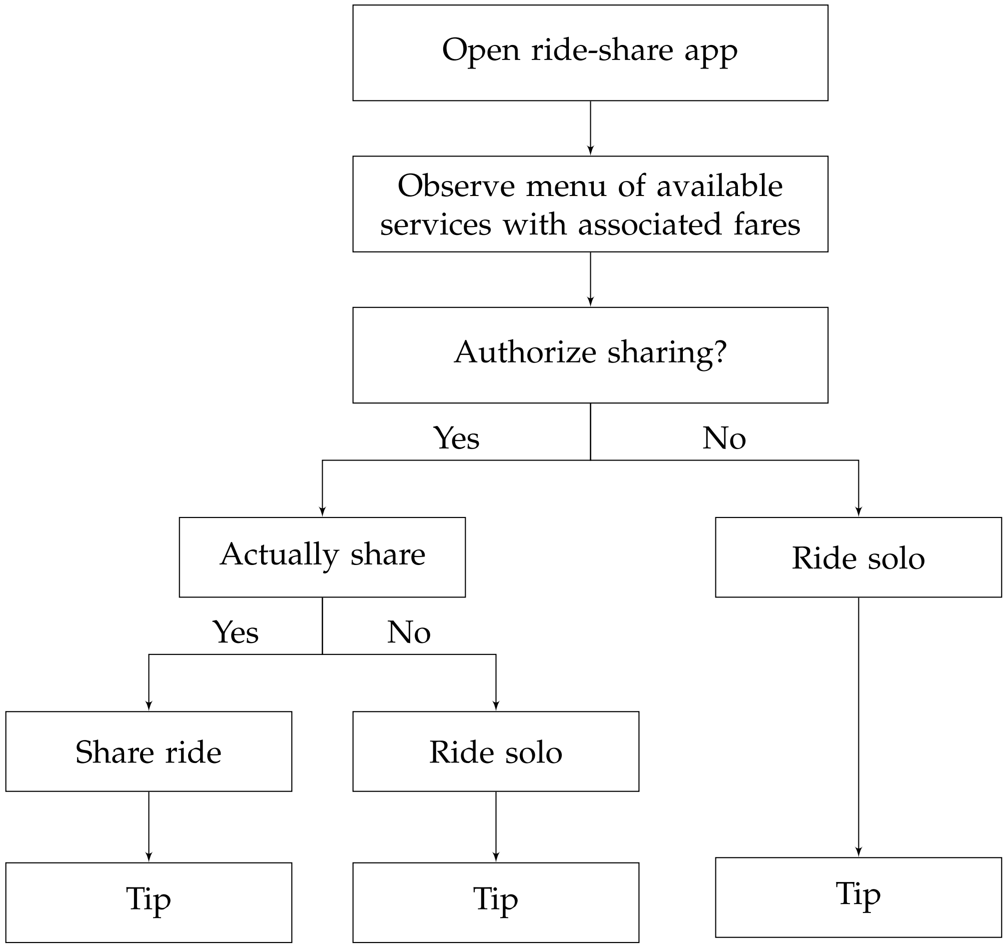


FIGURE 2: A flowchart illustrating the timing of a rider's decision to authorize sharing with another passenger on Uber or Lyft.

of the ride and may be correlated with rider utility if riders travelling between similar locations at similar times hold similar preferences.

Notably, this price–convenience tradeoff only arises if a sharing-authorized ride actually ends up being shared, since the discount associated with a shared ride only applies in cases when a shared-authorized ride is actually shared with a co-rider.[23] A considerable portion of sharing-authorized rides end up solo simply because at a given time and place, there are no other unmatched riders who have also authorized sharing and are attempting a similar trip.[24] It is likely that time and place also affects the likelihood a sharing-authorized ride ends up shared, since different settings are correlated with higher demand for transportation. Thus, they may be also correlated with rider preferences. The

[23]Uber's terms of use note this explicitly and state that, "Matching with a co-rider is based on time of day, request volumes, and driver supply and is determined algorithmically." See `https://www.uber.com/us/en/ride/uberx-share/`.

[24]In our data, roughly 30% of sharing authorized rides end up not being shared.

fare discount is also algorithmically determined to capture this likelihood and indeed has been found to positively affect willingness to authorize sharing (Uber 2018; Hou et al. 2020).

How much a rider tips is thus a function of (i) their preferences, as frugal riders are less likely to tip generously or at all, and (ii) which fare (i.e., solo or sharing-authorized) is selected, since tip payments are customarily a percentage of total payment, and (iii) time, place, and time–place effects, because factors such as traffic, trip length, and weather affect tipping.[25] Tipping is also motivated by a desire to avoid unpleasant feelings of guilt and embarrassment (Azar 2020). The guilt associated with not tipping or tipping a low amount, however, may be less when a passenger knows another rider may tip—behavior which is reminiscent of the free-rider problem (Boyes et al. 2006).

3.3 Empirical Setup

We apply the front-door criterion in this context to remove bias associated with selection into sharing-authorization, i.e., the effect caused by frugal riders hoping to pay less on fares also being less likely to tip generously or at all. Though our M variable (i.e., whether a passenger who authorized a shared ride actually shares a ride) is not strictly exogenous to either our X (i.e., whether a passenger authorizes a shared ride) or Y (i.e., tipping behavior) variables, we argue that it is *conditionally* exogenous to both those variables. We condition on (i) full fare (fare level plus any additional charges) with (ii) origin–destination fixed effects, two-hour time slot fixed effects, and origin–destination-time slot fixed effects to ensure that the remaining variation in M is no longer plagued by endogenous variation. In Appendix A4, we show a histogram of the number of ride observations within each fixed effect cell. On average, there are ten rides per fixed effed cell.

There are three sources of endogeneity in this setup. First, customers choose their own fare as part of their choice over X. In the United States, tipping norms tend to involve tipping a fixed percentage of full fare. Thus, as lower fares (more specifically, the absolute size of the difference in fares associated with opting to share) induce riders to authorize sharing, the charged fare is endogenous to a rider's decision on X and their tip behavior Y. To deal with this source of endogeneity, we control for each ride's full fare in both stages of our front-door criterion estimation. Recall that riders observe their fare options

[25] Examining 40 million UberX (i.e., solo) rides during the summer of 2017, Chandar et al. (2019) find that "demand-side" factors that capture an individual consumer's propensity to tip explain more of the variation in tipping than "supply-side" factors such as driver or ride quality. By examining only solo rides, however, Chandar et al. (2019) do not examine a key correlate of tipping: whether a passenger opts to share a ride.

before they select on X. One may be compelled to argue that fare (and the fare differential) serves as another mediator between X and Y, but it is not, because it is a factor that leads a rider to decide whether to authorize sharing, i.e., a variable that directly causes X instead of a variable whereby the effect of X on Y is mediated. Once one authorizes sharing (i.e., once X is chosen), the fare is locked in. One may also argue, because the full fare includes only the fare paid and not the two fare options associated respectively with a shared ride and a solo ride, that the fare paid is a function of X, and thus may be correlated with M, thereby violating the required identifying assumptions for the front-door criterion. Although this is conceptually possible, we argue that this correlation and associated bias are likely very small in this application conditional on the inclusion of the time-place fixed effects in the regression. Indeed in Appendix A6, we present estimation results without fare included as a control variable. We find that whether the fare is included or not makes almost no quantitative difference in the first stage of the front-door criterion setup, but it makes a difference in the second stage because people tip on the basis of the fare actually paid. The overall qualitative result, however, is robust to the omission of fare as a control variable.

Second, it is likely that tipping behavior differs across time and space. It is plausible that tipping differs between origin–destination community area pairs, because this is likely correlated with a rider's socio-economic status and trip purpose. Similarly, it is possible that tipping behavior differs by time within the same origin-destination pair. For instance, a rush-hour ride could lead a rider to want to tip less by virtue of being longer and slower than the same ride late at night. Thus, we control for origin–destination fixed effects interacted with each unique two-hour time slots between 12:00 AM January 1st and 11:59 PM December 31st 2019. These fixed effects help control for confounding factors determined by time and geographic location of the trip origin and destination, including major events that affect both demand for rides and traffic. For instance, the end of a Chicago Cubs game would increase both ride demand and traffic in the community areas containing and surrounding Wrigley Field. In this scenario, the base fare discount associated with authorizing a shared ride may be larger and the likelihood of actually sharing a ride with another passenger may be larger as well. Simultaneously, trip quality may be worse because of longer travel times. Accounting for origin–destination fixed effects interacted with two-hour time bins helps account for any potential correlation between the base fare (F), the likelihood a ride is actually shared with another passenger (M), and other latent trip quality factors.

Finally, riders who are generally more frugal are more likely to authorize sharing because it offers a guaranteed lower fare. These customers are also less likely to tip generously, or at all. For example, customers who are more likely to authorize sharing likely tip a lower percentage of the total fare compared to less frugal riders.[26] This source of endogeneity differs from the first source mentioned, both qualitatively and empirically. The first "fare-driven" source of endogeneity occurs with lower fares inducing lower tips for riders with the same tipping behavior (as manifested in the fixed percentage of fare they usually tip). As described above, this source of endogeneity is dealt with by controlling for the selected fare of each ride in regressions. By contrast, this third source of endogeneity is due to customers with unobserved preferences for cost-saving authorizing into sharing and also tipping less.

Because this third source of endogeneity is unobservable in our data, it cannot be dealt with through traditional back-door conditioning. It can be dealt with, however, using the front-door criterion. Indeed, although sharing-authorized rides are endogenous to tipping behavior, whether a passenger actually ends up sharing a ride with another is exogenous conditional on our set of time–place cell fixed effects and the full fare. Including both time–place cell fixed effects and (to a much lesser degree) full fare in the first stage rules out confounding due to conditions such as high demand or traffic that cause selection into authorization and make sharing authorized trips more likely to end up shared. Including the same control variables in the second stage helps ensure that time-and-place factors that affect M and Y are accounted for.

Conditional on these variables we uphold the assumption of ignorability (Rosenbaum and Rubin 1983). The only path through which sharing authorization (i.e., X) will affect tipping (i.e., Y) is through whether a ride is actually shared (i.e., M).[27] This mediator is relevant because tipping variation in ride-hailing and taxi settings can be affected by demand-side factors such as rider experience, mood, and social preferences (Chandar et al. 2019).[28]

Our estimation strategy consists of estimating the following equations.

$$\text{Naïve: } Y_i = \alpha_2 + \beta_2 X_i + \rho_2 F_i + \boldsymbol{G_i'\theta_2} + \epsilon_{2i} \tag{3}$$

[26]We show this empirically in the next subsection by using tip as a percentage of fare paid as our third dependent variable. As expected, the front-door criterion estimates a much lower effect of authorizing sharing on the tip share than the naiv̈e specification, suggesting that people who tip a lower share select into sharing-authorized rides.

[27]We include the same exact set of controls in both stages of front-door criterion estimation. This is because the exact same controls are necessary to uphold conditional exogeneity of both X and M. In other applications, it may not be necessary to include identical sets of controls if conditional exogeneity at different stages of estimation can be upheld through conditioning on different sets of observables.

[28]The battery of fixed effects we deploy in our application also prevents the "recanting witness" problem, whereby authorizing a shared ride would induces an unmeasured association between whether one gets to share a ride and tipping because a confounder of whether one gets to share a ride is itself caused by exposure to treatment.

$$\text{FDC First Stage: } M_i = \kappa_2 + \gamma_2 X_i + \tau_2 F_i + \boldsymbol{G_i'\sigma_2} + \omega_{2i} \tag{4}$$

$$\text{FDC Second Stage: } Y_i = \lambda_2 + \delta_2 M_i + \phi_2 X_i + \pi_2 F_i + \boldsymbol{G_i'\iota_2} + \nu_{2i} \tag{5}$$

where Y_i now represents tipping at either the extensive or intensive margin, F_i is the full fare of each trip, computed as the sum of base fare level and additional charges, and $\boldsymbol{G_i}$ is a vector of time-and-place cell fixed effects (i.e., origin–destination pairs interacted with each two-hour time slot during the year). Additionally, X_i is our treatment variable, which indicates whether a passenger authorized ride-sharing, and M_i indicates whether the ride was actually shared with another passenger.

We estimate the two front-door criterion equations by seemingly unrelated regression (Zellner 1962) to account for the potential correlation between equations 4 and 5. To recover the effect of X on Y estimated by the front-door criterion, we simply multiply the coefficient estimates $\hat{\gamma}_2$ and $\hat{\delta}_2$ by each other. In the tables below, we report heteroskedasticity-robust standard errors as well as standard errors clustered at the level of the origin–destination pair interacted with two-hour time slot fixed effects. Because the effect of interest is a nonlinear combination of coefficients, standard errors for that effect are obtained using the delta method.

In this empirical application, the identifying assumptions follow those discussed in Section 2. First, the only way in which X influences Y is through M given that we condition out the backdoor path represented by F. That is, given that we condition on the full fare F in both stages, when a rider authorizes sharing (X), the only way X can ever influence tipping behavior (Y) is if the passenger actually gets to share a ride (M).[29] Second, the assumption that $Cov(X, \omega_2) = 0$ is supported by the fact that M is determined by X and the app algorithm, which we approximate with our fixed effects. Third, $Cov(M, \nu_2) = 0$ given that M is as good as random conditional on X, F, and $\boldsymbol{G}$.[30] Finally, given that we have one-sided non-compliance between treatment and mediator (e.g., if $M_i = 1$ then $X_i = 1$ and if $M_i = 0$ then $X_i = 1$ or $X_i = 0$), the full-support requirement is not satisfied, and the front-door criterion estimates the ATT in this application.

3.4 Results

Table 2 shows results for tipping at the extensive margin. In this case, the naïve specification estimates that authorizing sharing reduces the probability a rider will tip by 6.8 per-

[29] Recall that the price-convenience tradeoff only arises if a sharing-authorized ride is actually shared since the discount associated with a shared ride only applies in cases when a shared-authorized ride is actually shared with a co-rider.

[30] For instance, Uber's terms of use explicitly state that "[m]atching with a co-rider is based on time of day, request volumes, and driver supply and is determined algorithmically."

cent. The front-door criterion, however, estimates that authorizing sharing reduces tipping probability by only 0.84 percent, an estimate almost a full order of magnitude smaller than the naïve estimate. The difference between the two estimates suggests that much of the naïve specification simply captures endogeneity of selection into treatment, despite the naïve specification controlling for the same exact variables as the front-door criterion specification. Notably, this means that the naïve specification generates biased estimates because it is unable to account for stingier customers being attracted to the cheaper service and also less likely to tip. The front-door criterion estimation deals with this endogeneity by leveraging the conditionally exogenous mediator of actually sharing—the only qualitative variable determined after a rider observes the fares of each service and makes a selection over whether to authorize sharing.

Table 3 shows results for tipping at the intensive margin. These results mirror those for tipping at the extensive margin. The naïve specification finds that authorizing sharing reduces the tip amount by about 5.5 percent. The front-door criterion presents a much lower, yet still statistically significant effect. The treatment effect calculated using the front-door criterion finds that authorizing sharing reduces the tip amount by about 0.57 percent, an effect that is an order magnitude smaller than that of the naïve estimate.

Finally, in Table 4 we estimate results on the tip as a share of the fare as the dependent variable with the same right-hand side variables as specified in equations 4 and 5. This iteration of our application is motivated by the observation that some passengers may simply tip based on some fixed percentage of the ride fare. Again, we find that the naïve estimate, which is likely biased due to selection into X, is roughly an order of magnitude larger than the front-door criterion estimate. As described in the previous section, the naïve specification is biased because it does not account for more frugal customers demonstrating a lower average tip share, not just a lower average tip payment. Specifically, the naïve specification finds that authorizing sharing reduces the tip share by over 9 percent. The front-door criterion, on the other hand, estimates that authorizing sharing reduces the tip share by about 1.1 percent.[31] This reaffirms the relevance of the third source of endogeneity mentioned in the last subsection; the naïve specification is biased due to selection of riders who are more likely to pay a lower tip share, endogeneity that also biases the two other naïve specifications in Tables 2 and 3.

This application demonstrates the usefulness of the front-door criterion. By exploit-

[31] In Appendix A6, we present estimation results without fare included as a control to show that whether fare is included or not makes almost no difference in the first stage of the front-door criterion setup, but it makes a difference in the second stage because people tip on the basis of the fare actually paid. Similarly, in Appendix A7, we present estimation results without time-place fixed effects. We find omitting these fixed effects makes almost no difference.

TABLE 2: Estimation Results for Tipping at the Extensive Margin

	Naïve	Front-Door	
Variables	Tipped	Shared Trip	Tipped
	(1)	(2)	(3)
Sharing Authorized (X)	-0.0628***	0.6769***	-0.0550***
	(0.0001)	(0.0002)	(0.0002)
	[0.0001]	[0.0004]	[0.0003]
Shared Trip (M)	–	–	-0.0115***
			(0.0002)
			[0.0002]
Full Fare (F)	0.0050***	-0.0064***	0.0049***
	(0.00001)	(0.00001)	(0.00003)
	[0.00002]	[0.0001]	[0.0001]
Intercept	0.1306***	0.0851***	0.1316***
	(0.0002)	(0.0002)	(0.0005)
	[0.0003]	[0.0013]	[0.0010]
Treatment Effect	-0.0628***	-0.0078***	
	(0.0001)	(0.0001)	
	[0.0001]	[0.0002]	
Elasticity	-6.764%***	-0.836%***	
	(0.0001)	(0.0001)	
	[0.0001]	[0.0002]	
Observations	95,670,449	95,670,449	
R^2	0.1165	0.7297	0.1165

Notes: Both specifications control for a linear function with full fare (fare level + additional charges) and origin–destination–date–two-hour-time cell fixed effects. The front-door criterion specification is estimated using seemingly unrelated regression. Heteroskedasticity-robust standard errors are in parentheses and standard errors clustered by origin–destination-two hour time cell are in brackets. Standard errors for the front-door criterion treatment effect and elasticity are computed using the delta method. *** p<0.01, ** p<0.05, * p<0.1.

TABLE 3: Estimation Results for Tipping at the Intensive Margin

Variables	Naïve	Front-Door	
	arcsinh(Tip)	Shared Trip	arcsinh(Tip)
	(1)	(2)	(3)
Sharing Authorized (X)	-0.0958***	0.6769***	-0.0858***
	(0.0002)	(0.0002)	(0.0003)
	[0.0002]	[0.0004]	[0.0005]
Shared Trip (M)	–	–	-0.0147***
			(0.0003)
			[0.0005]
Full Fare (F)	0.0148***	-0.0064***	0.0147***
	(0.00003)	(0.00002)	(0.0001)
	[0.00004]	[0.0001]	[0.0002]
Intercept	0.1233***	0.0851***	0.1245***
	(0.0004)	(0.0002)	(0.0011)
	[0.0005]	[0.0013]	[0.0026]
Treatment Effect	-0.0958***	-0.0099***	
	(0.0002)	(0.0002)	
	[0.0002]	[0.0003]	
Elasticity	-5.527%***	-0.574%***	
	(0.0001)	(0.0002)	
	[0.0001]	[0.0003]	
Observations	95,670,449	95,670,449	
R^2	0.1545	0.7297	0.1545

Notes: Both specifications control for a linear function in full fare (fare level + additional charges) and origin–destination–date–two-hour-time cell fixed effects. The front-door criterion specification is estimated using seemingly unrelated regression. Heteroskedasticity-robust standard errors are in parentheses and standard errors clustered by origin-destination-two hour time cell are in brackets. Standard errors for the front-door criterion treatment effect and elasticity are computed using the delta method. *** p<0.01, ** p<0.05, * p<0.1.

ing the conditional exogeneity of actual ride matching once sharing is authorized, we can estimate the effect of authorizing a shared ride on tipping. As it turns out, the treatment effect estimated by the front-door criterion is considerably lower—by about one order of magnitude—than that estimated by a naïve specification at both the intensive and extensive margins of tipping as well as for tipping as a percentage of the fare.

For illustrative purposes, and to show that the front-door criterion does not require linear regression or even a parametric specification, in Appendix A8 we show results using an alternative nonparametric estimation method—simple in-sample conditional expectations and their product—that omits control variables. The results are qualitatively similar to the main results presented in this section. The naïve effect estimated in columns 1 and 2 is comparable to that in column 1 of Tables 2 and 3. In that case, the nonparametrically estimated effects in Appendix Table A.11 are roughly double the parametrically estimated effects in Tables 2, 3, and 4. The first-stage (i.e., whether a trip is shared conditional on authorizing a shared ride) estimates are nearly identical across parametric and nonparametric specifications at around 0.700 in Table A.11 and 0.677 in Tables 2, 3, and 4. Finally, the effects estimated with the front-door criterion are very close to one another when comparing columns 1, 2, and 3 of Table A.11 with the front-door criterion estimates in Tables 2, 3, and 4.

Our results have a few limitations. First, the data do not differentiate between Uber and Lyft rides. Though it is likely that Uber and Lyft employ different algorithms for setting fares once a rider opts to authorize sharing, we are confident that our strategy of conditioning on fixed effects adequately takes care of this issue. Alternatively, one might be worried that people choose between Uber and Lyft on the basis of tipping: whereas Lyft always allowed tipping, Uber introduced tipping only in 2017, and at least one of us initially preferred Lyft over Uber precisely because Lyft allowed tipping. But this is exactly an example of the selection problems the front-door criterion allows dealing with, as it allows effectively controlling for a consumer's "type." Moreover, by 2019, tipping was a well-known feature of both services.

Additionally, we do not observe the exact tip or fare payments, observing rounded values instead.[32] This means that in columns 2 and 3 of Table 3, we are dealing with two sources of classical measurement error. The first is classical measurement error in fare level,[33] which is a control variable in both columns 2 and 3. While measurement error

[32]This mainly challenges the calculation of the effect of authorizing sharing on tipping percentage, a potentially interesting causal relationship which we do not explore because the dependent variable (i.e., tipping percentage) would have to be calculated on the basis of two variables measured with error.

[33]For example, because fares are rounded at the nearest \$2.50 in the data, a reported \$15 ride's true fare could lie anywhere in the $(\$13.75, \$16.25)$ interval.

TABLE 4: Estimation Results for Tip as a Fraction of Fare

	Naïve	Front-Door	
Variables	arcsinh(Tip/Fare)	Shared Trip	arcsinh(Tip/Fare)
	(1)	(2)	(3)
Sharing Authorized (X)	-0.0175***	0.6769***	-0.0155***
	(0.00003)	(0.0002)	(0.00005)
	[0.00004]	[0.0004]	[0.0001]
Shared Trip (M)	–	–	-0.0031***
			(0.00005)
			[0.00007]
Full Fare (F)	-0.0004***	-0.0064***	0.0004***
	(0.000003)	(0.00002)	(0.000007)
	[0.000004]	[0.0001]	[0.00003]
Intercept	0.0473***	-0.0851***	0.0475***
	(0.00005)	(0.0002)	(0.0001)
	[0.00006]	[0.0012]	[0.0003]
Treatment Effect	-0.0175***	-0.0021***	
	(0.00003)	(0.00003)	
	[0.00004]	[0.00005]	
Elasticity	-9.015%***	-1.064%***	
	(0.0001)	(0.0002)	
	[0.0002]	[0.0002]	
Observations	95,670,449	95,670,449	
R^2	0.0867	0.7297	0.0867

Notes: Both specifications control for a linear function in full fare (fare level + additional charges) and origin–destination–date–two-hour-time cell fixed effects. The front-door criterion specification is estimated using seemingly unrelated regression. Heteroskedasticity-robust standard errors are in parentheses and standard errors clustered by origin-destination-two hour time cell are in brackets. Standard errors for the front-door criterion treatment effect and elasticity are computed using the delta method. *** p<0.01, ** p<0.05, * p<0.1.

in a control variable can bias our estimate of the coefficient of interest, we show results that omit fare as a control variable in Appendix A6, and those results show that estimates do not change sensibly whether or not we include fare as a control variable. The second source of measurement error is classical measurement error in tipping amount, i.e., the dependent variable, in column 3. This is in principle more problematic because classical measurement error in the dependent variable leads to less precise estimates. Though this would be worrisome in a small sample because it could lead to a type II error (i.e., we would fail to reject the null hypothesis that the coefficient on M is equal to zero), this is not an issue in our a sample of over 95 million observations—we indeed reject the null hypothesis that the coefficient on M in column 3 is equal to zero.

Finally, in Appendix A11, we show results using the front-door criterion approach that omits the X variable from the second-stage regression. This leads to biased results that are closer to those of our naïve regressoins than to those of our (correctly specified) front-door criterion regressions. These results illustrate the idea that conditioning on X is essential for the the front-door criterion to yield a credible estimate.

4 Conclusion

We have focused on the application of Pearl's (1995, 2000) front-door criterion. Because the goal of most research in applied economics is to answer questions of the form "What is the causal effect of X on Y?," economists should welcome the addition of techniques that allow answering such questions to their empirical toolkit. Yet economists have been reluctant to incorporate the front-door criterion in that toolkit.

First, we explain how to use the front-door criterion in the context of linear regression, which remains the workhorse of applied economics. Second, we present an application of the front-door criterion using observational data where the necessary assumptions for point-identification hold. Finally, in an effort to help overcome economists' resistance to incorporating the front-door criterion in their empirical toolkit, we look at what happens when the assumptions underpinning the front-door criterion are violated, and present these results in Appendix A9 (multiple mediators M) and in Appendix A10 (violation of strict exogeneity of the mediator M). Our implementation of the front-door criterion in this paper is, first and foremost, intended to be illustrative. There are a number of additional techniques and methods that could improve the estimation of causal effects such as the use of non-linear and nonparametric approaches, which can be particularly useful in the presence of multiple mediators, to estimation and machine-learning methods. The suitability of these approaches, of course, depends on the given empirical setting.

Ultimately, the front-door criterion is most promising in cases where units of observations are selected into treatment on the basis of unobservables which also affect the outcome, but for which treatment intensity or non-compliance to the treatment can argued to be (as good as) randomly assigned. In these cases, the front-door criterion can be a useful tool for applied researchers interested in causal inference with observational data. When selection into treatment is endogenous but there exists an exogenous mediator whereby the treatment causes the outcome, the front-door criterion can identify the causal effect of treatment on outcome. Several possible applications where the front-door criterion could be used to estimate important economic parameters include the following, data availability notwithstanding:

1. *Technology*: Our ride-sharing application directly motivates other applications in technology where internal administrative data, coupled with intimate knowledge of the factors included in algorithms that determine a user's experience, could be applied within the framework of the front-door criterion to estimate treatment effects. In particular, the front-door criterion could be fruitful in cases where users can endogenously choose X (e.g., opening an app or starting a session), but the algorithm then determines some mechanism M (e.g., a new feature, an information treatment, or a promotion) before an outcome of interest Y (e.g., engagement measures, amount spent) is realized. For further discussion on the use of DAGs to estimate causal effects in the context of Lyft's ecosystem, see Rich and Manek (2022).

2. *Labor I*: Hull (n.d.) documents a potential application for the front-door criterion where a researcher is interested in estimating the effect of a worker having a criminal record X on employment Y, where a background check M may suffer from both type I and type II errors. Given the existence of both types of error, this application would meet the full support requirement necessary for the front-door criterion to estimate the ATE. Hull (n.d.) notes that in practice, researchers may be concerned that the errors are truly as good as random and may want to condition on available covariates or exploit some unexpected policy change (e.g., a ban-the-box initiative). Our ride-sharing application provides a useful example of conditioning on available covariates and fixed effects in order to ensure excludability of M.

3. *Labor II*: The contribution by Glynn and Kashin (2018) highlights the use of the front-door criterion to study the effect of a job training program X on employment outcomes Y when only some of the workers actually show up to the job training program M. Similar to our ride-sharing application, this application does not meet the

full support requirement because of one-sided non-compliance between the treatment X and the mechanism M and, as in the ride-sharing application in this paper, the front-door criterion estimates the ATT. Glynn and Kashin (2018) are careful to point out that the mechanism may not be fully excludable in their application, but that the front-door criterion produces more credible estimates than a naïve regression. Therefore, it may be fruitful to consider cases where the credibility of this application could be improved by conditioning on available covariates or leveraging unexpected events (e.g., in-person gathering restrictions due to the COVID-19 pandemic).

4. *Agriculture*: The front-door criterion could be used to leverage agro-ecological mechanisms embedded within biological systems to estimate causal effects relevant to the field of agricultural economics. It is common to find differences in estimated yield gains between experimental agronomic trials and farmers using the same inputs under real-life conditions (Laajaj et al. 2020). The front-door criterion could exploit evapotranspiration (i.e., the agro-ecological mechanism whereby water moves from soil through the plant and then evaporates in the air) that may be based on as-good-as-random variation in weather conditions to estimate the effect of the use of some input X (e.g., fertilizer, seeds, etc.) on some output Y (e.g., yield, etc.) in a real-world setting.

5. *Health*: Hünermund and Bareinboim (2023) discuss possible applications of the front-door criterion where a medical test or screening result M provides a noisy signal for the presence of some medical condition X. An example raised by the authors is that of congenital anomalies which are routinely tested via ultrasound screenings during pregnancy, but often exhibit both type I and type II errors. These errors may occur at random or may be related to measurable health characteristics (i.e., BMI, blood pressure, etc.). If the only way that some medical condition influences medical decision making Y is via detection on a test or screening, then the necessary assumptions for the front-door criterion are satisfied.

These applications each fall under a class of potential applications for the front-door criterion where there is exogenous non-compliance in some mechanism M that lies on the causal path between X and Y. Our goal is that the application and discussion in this paper inspire the use of the front-door criterion to answer the types of questions embedded in these and other potential applications.

References

Abadie, A., Athey, S., Imbens, G.W., and Wooldrdige, J.M. (2023) "When Should You Adjust Standard Errors for Clustering?," *Quarterly Journal of Economics*, vol. 138, no. 1, pp. 1-35.

Acharya, A., Blackwell, M., and Sen, M. (2016) "Explaining Causal Findings Without Bias: Detecting and Assessing Direct Effects," *American Political Science Review*, vol. 110, no. 3, pp. 512-529.

Angrist, J. and Kruger, A. (1995) "Split-Sample Instrumental Variables Estimates of the Return to Schooling," *Journal of Business and Economic Statistics*, vol. 13, issue, 2, pp. 225—235.

Athey, S., Chetty, R., Imbens, G., and Kang, H. (2019) "The Surrogate Index: Combining Short-Term Proxies to Estimate Long-Term Treatment Effects More Rapidly and Precisely," *NBER Working Paper No. 26463*.

Azar, O.H. (2020) "The Economics of Tipping," *Journal of Economic Perspectives*, vol. 34, no. 2, pp. 215-236.

Bellemare, M.F., and Wichman, C.J. (2020) "Elasticities and the Inverse Hyperbolic Sine Transformation," *Oxford Bulletin of Economics and Statistics*, vol. 82, no. 1, pp. 50-61.

Berry, M. (2021) "How Many Uber Drivers Are There?," `https://therideshareguy.com/how-many-uber-drivers-are-there` last accessed June 22, 2021.

Bowman, C. (2019) "10 Things I Wish I Knew Before I Started Driving for Uber and Lyft," `https://www.businessinsider.com/uber-lyft-drivers-job-advice-car-2019-8` last accesssed June 22, 2021.

Boyes, W.J., Mounts, W.S., and Sowell, C. (2006) "Restaurant Tipping: Free-Riding, Social Acceptance, and Gender Differences," *Journal of Applied Social Psychology*, vol. 34, no. 12, pp. 2616-2625.

Card, D., DellaVigna, S., Funk, P., and Iriberri, N. (2020) "Are Referees and Editors in Economics Gender Neutral?," *Quarterly Journal of Economics*, vol. 135, no. 1, pp. 269-327.

Chalak, K. and White, H. (2011) "Viewpoint: An extended class of instrumental variables for the estimation of causal effects," *Canadian Journal of Economics*, vol. 44, issue, 1, pp. 1-51.

Chandar, B., Gneezy, U., List, J.A., and Muir, I. (2019) "The Drivers of Social Preferences: Evidence From a Nationwide Tipping Experiment," NBER Working Paper no. 26380.

Conley, T.G., C.B. Hansen, and P.E. Rossi (2012) "Plausibly Exogenous," *Review of Economics and Statistics*, vol. 94, no. 1, pp. 260-272.

Davidson, R. and MacKinnon, J. G. (2004) *Econometric Theory and Methods*, Oxford University Press, New York.

Fulcher, I.R. and Shpitser, I., Marealle, S. and Tchetgen, E.J.T. (2020) "Robust inference on population indirect causal effects: the generalized front door criterion," *Journal of the Royal Statistical Society, Statistical Methodology, Series B*, vol. 82, issue 1, pp. 199-214.

Garcia, J.L., Heckman, J.J., Leaf, D.E., and Prados, M.J. (2020) "Quantifying the Life-Cylce Benefits of an Influential Early-Childhood Program," *Journal of Political Economy*, vol. 128, no. 7, pp. 2502-2541.

Glynn, A.N. and Kashin, K. (2017) "Front-Door Difference-in-Differences Estimators," *American Journal of Political Science*, vol. 61, no. 4, pp. 989-1002.

Glynn, A.N. and Kashin, K. (2018) "Front-Door Versus Back-Door Adjustment With Unmeasured Confounding: Bias Formulas for Front-Door and Hybrid Adjustments With Application to a Job Training Program," *Journal of the American Statistical Association*, vol. 113, no. 523, pp. 1040-1049.

Gupta, S., Z.C. Lipton, and Childers, D. (2020) "Estimating Treatment Effects with Observed Confounders and Mediators," Working Paper, Carnegie Mellon University.

Haavelmo, T. (1943) "The Statistical Implications of a System of Simultaneous Equations," *Econometrica*, vol. 11, no. 1, pp. 1–12.

Harrington, R. (2019) "Want To Get a Tip As An Uber Driver? Don't Pick-Up A Shared Ride," https://www.compassred.com/data-journal/want-to-get-a-tip-as-an-uber-driver-dont-pick-up-a-shared-ride last accessed May 26, 2020.

Heckman, J., Pinto, R., and Savelyev, P. (2013) "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes," *American Economic Review*, vol. 103, no. 6, pp. 20052–2086.

Heckman, J. and Pinto, R. (2015) "Causal Analysis After Haavelmo," *Econometric Theory*, vol. 31, pp. 115-151.

Hemel, D. J. (2017) "Pooling and Unpooling in the Uber Economy," *University of Chicago Legal Forum* vol. 2017, pp. 265-286.

Hull, P. (n.d.) "A Napkin Problem 'Use Case,'"
`https://www.dropbox.com/s/wpykyiwvc5p4qes/napkin.pdf?dl=0`.

Hünermund and Bareinboim (2023) "Causal inference and data fusion in econometrics, *The Econometrics Journal*, March 2023.

Imai, K., Keele, L., and Yamamoto, T. (2010) "Identification, Inference and Sensitivity Analysis for Causal Mediation Effects," *Statistical Science*, vol. 25, no. 1, pp. 51-71.

Imai, K., Keele, L., Tingley, D., and Yamamoto, T. (2011) "Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational

Studies," *American Political Science Review*, vol. 105, no. 4, pp. 765-789.

Imbens, G.W. (2015), "Matching Methods in Practice: Three Examples," *Journal of Human Resources*, vol. 50, no. 2, pp. 373-419.

Imbens, G.W. (2020) "Potential Outcome and Directed Acyclic Graph Approaches to Causality: Relevance for Empirical Practice in Economics," *Journal of Economic Literature*, forthcoming.

Imbens, G.W., and Angrist, J.D. (1994) "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, vol 62, no. 2, pp. 467-475.

Kline, P. and Walters, C.R. (2016) "Evaluating Public Programs with Close Substitutes: The Case of Head Start," *The Quarterly Journal of Economics*, vol. 131, issue 4, pp. 1795-1848.

Laajaj, R., Macours, K., Masso, C., Thuita, M., and Vanlauwe, B. (2020) "Reconciling yield gains in agronomic trials with returns under African smallholder conditions," *Scientific Reports*, vol. 10, no. 1, pp. 1-15.

Morgan, S.L., and Winship, C. (2015) *Counterfactuals and Causal Inference*, Cambridge University Press, Cambridge, United Kingdom.

Mullahy, J. and Norton, E.C. (2022) *Why Transform Y? A Critical Assessment of Dependent-Variable Transformations in Regression Models for Skewed and Sometimes-Zero Outcomes*, NBER Working Paper, No. 30735.

Naimi, A.I. (2015) "Invited Commentary: Boundless Science—Putting Natural Direct and Indirect Effects in a Clearer Empirical Context," *American Journal of Epidemiology* vol. 182, issue 2, pp. 109–114.

Pearl, J. (1993) "Mediating Instrumental Variables," *Technical Report R-210*.

Pearl, J. (1995) "Causal Diagrams for Empirical Research," *Biometrika*, vol. 82, no. 4, pp. 669-688.

Pearl, J. (2000) *Causality: Models, Reasoning, and Inference*, Cambridge University Press, Cambridge, United Kingdom.

Pearl, J. and Mackenzie D. (2018) *The Book of Why: The New Science of Cause and Effect*, Basic Books: New York, NY.

Rich, D. and Manek, S. (2022) "Causal Forecasting at Lyft (Part 1)", Available online: `https://eng.lyft.com/causal-forecasting-at-lyft-part-1-14cca6ff3d6d`.

Rosenbaum, P. and Rubin, D. (1983) "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, vol. 70, pp. 41-55.

Rubin, D.B. (2005) "Causal Inference Using Potential Outcomes: Design, Modeling, Decisions," *Journal of the American Statistical Association*, vol. 100, no. 469, pp. 322-331.

Samuelson, W. and Zeckhauser, R. (1986) "Status Quo Bias in Decision-Making," *Journal of Risk and Uncertainty*, vol. 1, issue 1, pp. 7–59.

Strotz, R.H. and Wold, H.O.A. (1960) "Recursive versus nonrecursive systems: An attempt at synthesis," *Econometrica*, vol. 28, pp. 417–427.

Tchetgen Tchetgen, E.M. and VanderWeele, T.J. (2014) "Identification of Natural Direct Effects When a Confounder of the Mediator Is Directly Affected by Exposure," *Epidemiology*, vol. 25, issue 2, pp. 282–291.

Young, M., Farber, S., Palm, M. (2020) "The true cost of sharing: A detour penalty analysis between UberPool and UberX trips in Toronto." *Transportation Research Part D.*, vol. 87. 102450.

Zellner A.(1962) "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association*, vol. 57, issue 298, pp. 348—368.

Appendix

Supporting Appendix for "The Paper of How: Estimating Treatment Effects Using the Front-Door Criterion." This Appendix includes the following Sections:

- Appendix A1 presents Pearl's proof of identification for the front-door criterion.
- Appendix A2 clarifies the full support requirement and its implications on treatment effect parameters.
- Appendix A3 presents simulation results under the ideal case where all of the identification assumptions hold.
- Appendix A4 illustrates the number of ride observations are in each fixed effect cell.
- Appendix A5 reports results with the dependent variable in levels rather than transformed by the inverse hyperbolic sine function.
- Appendix A6 reports results from our empirical application without the full fare included as a control variable.
- Appendix A7 reports results from our empirical application without the time and place fixed effects included in the regression specification.
- Appendix A8 reports results from our empirical application using a non-parametric estimation approach.
- Appendix A9 discusses a departure from the 'ideal case' where we have multiple mediators.
- Appendix A10 discusses another departure from the 'ideal case' where we have violations of strict exogeneity of the mediator.
- Appendix A11 reports results from our empirical application without conditioning on X (i.e., sharing authorization) in the second stage of the front-door criterion estimation.

A1. Pearl's Front-Door Criterion Proof of Identification

We now summarize Pearl's (1995) proof and provide additional explanation to help the reader's intuition along the way by deriving the front-door criterion estimand. In what follows we use $P(M|X)$ as shorthand for $P(M = 1|X = 1) - P(M = 1|X = 0)$, $P(Y|M,X)$ as shorthand for $P(Y = 1|M = 1, X) - P(Y = 1|M = 0, X)$, and $P(Y|do(X))$ as shorthand for $P(Y = 1|do(X) = 1) - P(Y = 1|do(X) = 0)$. Our aim is to compute $P(Y|do(X)$ with observable variables, where $P(Y|do(X))$ represents the causal effect of X on Y. Pearl (2000) introduces $do(\cdot)$ as shorthand for an intervention that sets the variable between parentheses to a specific value. Thus, $P(Y|do(X = x))$ denotes the probability of $Y = 1$ when X is set equal to x by researcher intervention, or when X is manipulated and everything else is held constant. $P(Y|do(X))$ should be read as the causal effect of X on Y. This should be contrasted with $P(Y|X)$, which may not represent the causal effect of X on Y due to the presence of the unobserved confounder U.

Our goal here is to restate $P(Y|do(X))$ using only the observed variables M, X, and Y while leveraging Assumptions 1 through 3. The first step is to compute $P(M|do(X))$. Under Assumption 2, the lack of a back-door path between X and M implies the relationship between X and M is identified. When that assumption holds, we can write

$$P(M|do(X)) = P(M|X), \tag{A.1}$$

given that in this case, the unobserved confounder U affecting X but not M makes the two sides of Equation A.1 equivalent.

The second step is to compute $P(Y|do(M))$. Here we cannot set $do(M = m)$ because there is a back-door path from M to Y via X. To block this path we use Assumption 3. Conditional on X, the relationship between M and Y is not confounded by unobserved variables. In that case, by controlling for and summing over all observations, indexed by i, X_i of X, we can write

$$P(Y|do(M)) = \sum_X P(Y|X, do(M)) \times P(X|(M)) \tag{A.2}$$

where the right-hand-side of Equation A.2 involves two expressions involving $do(M)$. The second term on the right-hand-side of Equation A.2 can be reduced to $P(X)$ because, as stated by Assumption 1, the only way in which X influences Y is through M. The first term on the right-hand-side of Equation A.2 can be expressed as $P(Y|X, M)$ because, as stated by Assumption 3, conditional on X, the relationship between M and Y is not confounded. Therefore, we can write

$$P(Y|do(M)) = \sum_X P(Y|X, M) \times P(X). \tag{A.3}$$

The third and last step is to combine the two effect estimates, $P(M|do(X))$ from Equation A.1 and $P(Y|do(M))$ from Equation A.2, in order to compute $P(Y|do(X))$—the causal effect of X on Y.

To start with, we express $P(Y|do(X))$ in terms of $do(X)$ by controlling for and summing over all observations, indexed by i, M_i of M. This allows us to write

$$P(Y|do(X)) = \sum_M P(Y|M, do(X)) \times P(M|do(X)). \tag{A.4}$$

Assumption 3 allows us to rewrite M as $do(M)$ in the first term on the right-hand-side of Equation A.4. Since, conditional on X, the relationship between M and Y is not confounded, the variation in M is conditionally exogenous. Additionally, as stated by Assumption 1, the only way in which X influences Y is through M, and so we can remove $do(X)$ from the second term on the right-hand side of Equation A.4. Said differently, M should have no effect on X, because X causes M and not vice versa in Figure 1. Therefore, we can rewrite the first term on the right-hand side of Equation A.4 as

$$P(Y|M, do(X)) = P(Y|do(M), do(X)) = P(Y|do(M)). \tag{A.5}$$

Recall that Equation A.3 states that $P(Y|do(M)) = \sum_X P(Y|X, M) \times P(X)$ and Equation A.1 states that $P(M|do(X)) = P(M|X)$. Therefore, plugging Equation A.3 into Equations A.4 and A.5, and plugging Equation A.1 into Equation A.4 gives us the front-door criterion estimand as originally derived by Pearl (1995). That estimand is such that

$$P(Y|do(X)) = \sum_M P(M|X) \times \sum_{X'} P(Y|X', M) \times P(X'). \tag{A.6}$$

A2. Full Support and Treatment Effect Parameters

The following illustrates why the full support requirement is necessary for the front-door criterion to estimate the ATE. This sketch also shows the type of treatment parameter the front-door criterion estimates when the full support requirement fails to hold in specific ways.

In the set-up represented by Figure 1, we have:

$$(Y^0, Y^1)X|U \tag{A.7}$$

where Y^0 and Y^1 are potential outcomes and U is an unobserved confounder. If we were able to condition on U, then estimating the causal effect of X on Y would not require the front-door criterion method.

The ATE is defined as the average of the ATT and the ATU weighted by the probability an observation chooses X=1:

$$ATE = ATT[P(X=1)] + ATU[1 - P(X=1)] \tag{A.8}$$

Further note that the ATT is defined as follows:

$$ATT = E[Y^1 - Y^0|X=1, U] \tag{A.9}$$

Additionally note that the ATU is defined as follows:

$$ATU = E[Y^1 - Y^0|X=0, U] \tag{A.10}$$

Given that the three necessary assumptions for the front-door criterion to hold, and both X and M are binary variables, we can estimate the ATT as follows:

$$\begin{aligned} ATT = E[Y|X=1, M=0]P(M=0|X=1) + E[Y|X=1, M=1]P(M=1|X=1) - \\ E[Y|X=1, M=0]P(M=0|X=0) - E[Y|X=1, M=1]P(M=1|X=0) \end{aligned} \tag{A.11}$$

Similarly, the ATU is defined as follows:

$$\begin{aligned} ATU = E[Y|X=0, M=0]P(M=0|X=1) + E[Y|X=0, M=1]P(M=1|X=1) - \\ E[Y|X=0, M=0]P(M=0|X=0) - E[Y|X=0, M=1]P(M=1|X=0) \end{aligned} \tag{A.12}$$

The full support requirement, $P(X|M) > 0$, states that we must observe $P(X = 1|M = 1) > 0$, $P(X = 1|M = 0) > 0$, $P(X = 0|M = 1) > 0$, and $P(X = 0|M = 0) > 0$ to estimate the ATE with the front-door criterion. This can be seen in equations A.11 and A.12 where we must observe some average value of Y for each of these four possible combinations of both X and M, for the front-door criterion to estimate the ATE. In particular, the term $E[Y|X = 0, M = 1]P(M = 1|X = 1)$ is undefined.

If, however, weaker variations of the full support requirement (i.e., partial support) holds, then the front-door criterion will estimate either the ATT or the ATU. In particular, as is the case in our ride-sharing application, if we observe that $P(X = 1|M = 1) > 0$, $P(X = 1|M = 0) > 0$, and $P(X = 0|M = 0) > 0$ but $P(X = 0|M = 1) = 0$ due to one-sided non-compliance, then the ATU will be undefined and the front-door criterion estimates the ATT.

A3. Ideal Case Simulation Results

Our simulation setup is as follows. Let $U_i \sim N(0,1)$, $Z_i \sim U(0,1)$, $\epsilon_{Xi} \sim N(0,1)$, $\epsilon_{Mi} \sim N(0,1)$, and $\epsilon_{Yi} \sim N(0,1)$ for a sample size of $N = 100,000$ observations.[34] Then, let

$$X_i = 0.5U_i + \epsilon_{Xi}, \tag{A.13}$$

$$M_i = Z_i X_i + \epsilon_{Mi}, \tag{A.14}$$

and

$$Y_i = 0.5M_i + 0.5U_i + \epsilon_{Yi}. \tag{A.15}$$

This satisfies Pearl's (1995, 2000) three point-identification assumptions for the front-door criterion to be able to estimate the effect of X on Y, viz. (i) the only way in which X influences Y is through M, (ii) the relationship between M and X is not confounded by U, since U only affects X and not M, and (iii) conditional on X, the relationship between M and Y is not confounded by U. This simulation also meets the full support requirement (i.e., $P(X|M) > 0$) and so the front-door criterion will estimate the ATE. By substituting Equation A.14 into Equation A.15, it should be clear that the true effect is equal to 0.250 in our simulations.

To show that the front-door criterion estimates the causal effect of X on Y, we estimate three specifications. The first specification, which we refer to as our benchmark specification because it generates an unbiased estimate of the causal effect by virtue of controlling for the unobserved confounder U, estimates

$$Y_i = \alpha_0 + \beta_0 X_i + \zeta_0 U_i + \epsilon_{0i}, \tag{A.16}$$

where, because both X_i and U_i are included on the right-hand-side, $E(\hat{\beta}_0) = \beta$.

Second, we estimate a naïve specification. The naïve specification differs from the benchmark specification in Equation A.16 by failing to control for the presence of the unobserved confounder.

The last specification, which we refer to as our front-door specification, estimates

$$M_i = \kappa_0 + \gamma_0 X_i + \omega_{0i} \tag{A.17}$$

$$Y_i = \lambda_0 + \delta_0 M_i + \phi_0 X_i + \nu_{0i} \tag{A.18}$$

[34]We allow Z_i to be a random coefficient here to be consistent with subsequent simulations discussed later in this paper, but the results of our simulation exercise do not hinge on the coefficient on X_i being a random coefficient in Equation A.15.

TABLE A.1: Simulation Results—Ideal Case

	Benchmark	Naïve	Front-Door		Direct Effect
Variables	Y	Y	M	Y	Y
	(1)	(2)	(3)	(4)	(5)
Treatment (X)	0.252***	0.454***	0.507***	0.200***	-0.003
	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)
Mediator (M)	–	–	–	0.502***	0.500***
				(0.003)	(0.003)
Confounder (U)	0.499***	–	–	–	0.501***
	(0.004)				(0.004)
Intercept	-0.004	-0.005	-0.004	-0.003	-0.003
	(0.004)	(0.004)	(0.003)	(0.004)	(0.003)
Treatment Effect	0.252***	0.454***	0.254***		–
	(0.004)	(0.003)	(0.002)		
Observations	100,000	100,000	100,000		100,000

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The front-door equations in columns (3) and (4) are estimated by seemingly unrelated regressions. The standard error for the front-door criterion treatment effect is estimated by the delta method.

where the unobserved confounder U_i does not appear anywhere, but because the necessary assumptions for the front-door criterion to hold, $E(\hat{\gamma}_0 \cdot \hat{\delta}_0) = \beta$.

Column 1 of Table A.1 shows estimation results for Equation A.16, our benchmark specification. Column 2 shows estimation results for our naïve specification. Columns 3 and 4 show estimation results respectively for the front-door specification in Equations A.17 and A.18, respectively. The line labeled "Treatment Effect" shows estimates of the treatment effect for each of those three specifications. Unsurprisingly, the estimates in columns 1 and 2 differ markedly, as the former controls for U_i but the latter does not: $\hat{\beta}$ is equal to 0.252 in the benchmark case, but it is near double that at 0.454 in the naïve case.

Given the derivations above, it should also be unsurprising that the estimate generated by multiplying the coefficient on treatment in column 3 by the coefficient on the mediator in column 4 is equal to 0.254. Assuming the estimate in column 1 is not correlated with the estimate computed from columns 3 and 4, the benchmark and front-door criterion estimates are statistically identical. In both cases, the estimated treatment effects is not statistically different from its true value of 0.250.

Column 5 in Table A.1 serves an illustrative purpose. It shows that conditional on the mediator (M) and the unobserved confounder (U), the coefficient on the treatment (X) is statistically indistinguishable from zero. This result highlights the "no direct effect" assumption that is implied by Assumptions 1 through 3.

Finally, we slightly alter our simulation's data generating process to show that it is possible to use the front-door criterion approach when the mediator M is not strictly, but only conditionally exogenous to the relationship between X and Y. In this setup, we add C_i, an observed confounder that captures both selection into treatment and into the mechanism, while also affecting outcome. Let $C_i \sim N(0,1)$. Then, let

$$X_i = 0.5U_i + 0.5C_i + \epsilon_{Xi}, \tag{A.19}$$

$$M_i = Z_i X_i + 0.3C_i + \epsilon_{Mi}, \tag{A.20}$$

and

$$Y_i = 0.5M_i + 0.5U_i + 0.15C_i + \epsilon_{Yi}. \tag{A.21}$$

In this case, our benchmark specification controls for both unobserved and observed confounders. Thus, it estimates

$$Y_i = \alpha_1 + \beta_1 X_i + \rho_1 C_i + \zeta_1 U_i + \epsilon_{1i}, \tag{A.22}$$

where the right-hand-side of Equation A.22 includes both the observed confounder C_i and the unobserved confounder U_i. Similar to the previous simulation, the naïve specification differs from the benchmark specification by failing to control for the presence of the unobserved confounder, U_i.

Finally, the front-door criterion estimates

$$M_i = \kappa_1 + \gamma_1 X_i + \tau_1 C_i + \omega_{1i} \tag{A.23}$$

$$Y_i = \lambda_1 + \delta_1 M_i + \phi_1 X_i + \pi_1 C_i + \nu_{1i} \tag{A.24}$$

where both Equations A.23 and A.24 include C_i on the right hand side.

Table A.2 presents results of this simulation with 100,000 observations. Column 1 displays results of the benchmark regression, with the coefficient on X indicating the "true" effect of 0.251. Unsurprisingly the naïve specification overestimates the effect and the front-door criterion estimation approach yields an unbiased estimate. These results highlight that, given an additional conditional ignorability assumption (Rosenbaum and Rubin 1983), the front-door criterion approach can point-identify treatment effects by conditioning on a vector of control variables. In this simulation we show that when we are able to adequately control for a confounding variable influencing M, we are able to recover the correct effect with the front-door criterion approach. The topic of a conditionally exoge-

TABLE A.2: Simulation Results—Conditionally Exogenous Mediator

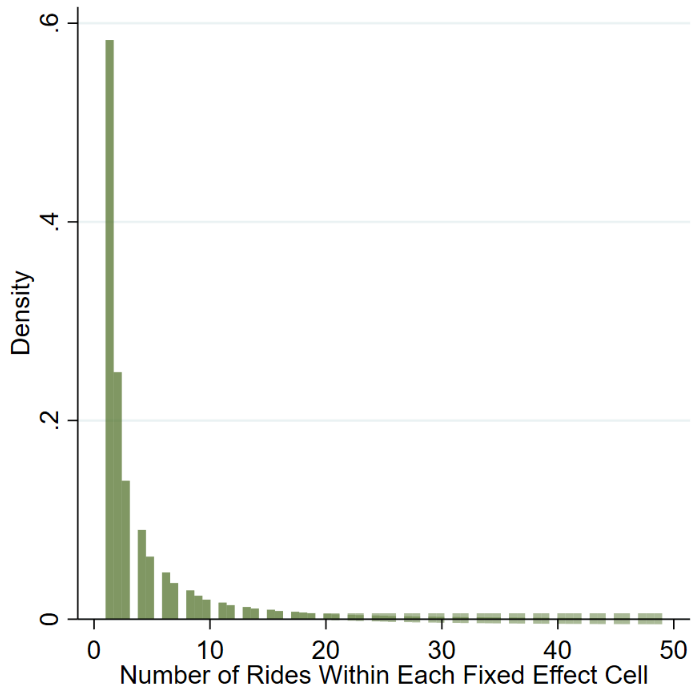
Variables	Benchmark	Naïve	Front-Door	
	Y	Y	M	Y
	(1)	(2)	(3)	(4)
Treatment (X)	0.251***	0.452***	0.499***	0.200***
	(0.004)	(0.003)	(0.003)	(0.003)
Mediator (M)	–	–	–	0.504***
				(0.003)
Observed Confounder (C)	0.303***	0.204***	0.304***	0.052***
	(0.004)	(0.004)	(0.004)	(0.004)
Unobserved Confounder (U)	0.499***	–	–	–
	(0.004)			
Intercept	-0.004	-0.005	-0.004	-0.003
	(0.004)	(0.004)	(0.003)	(0.004)
Treatment Effect	0.251***	0.452***	0.251***	
	(0.004)	(0.003)	(0.002)	
Observations	100,000	100,000	100,000	

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The front-door equations in columns (3) and (4) are estimated by seemingly unrelated regressions. The standard error for the front-door criterion treatment effect is estimated by the delta method.

nous M variable raises questions about contexts in which we cannot adequately control for the confounding variable influencing M and thus we violate the strict exogeneity assumption. In Section 4 we directly investigate violations of the strict exogeneity assumption in which confounding cannot be dealt with through conditioning.

A4. Number of Observations per Fixed Effect Cell

FIGURE A.1: Histogram of Number of Observations per Fixed Effect Cell



A5. Dependent Variable in Levels

TABLE A.3: Estimation Results for Level of Tip

Variables	Naïve	Front-Door	
	Tip	Shared Trip	Tip
	(1)	(2)	(3)
Sharing Authorized (X)	-0.1401***	0.6769***	-0.1314***
	(0.0003)	(0.0002)	(0.0006)
Shared Trip (M)	–	–	-0.0129***
			(0.0006)
Full Fare (F)	0.0386***	-0.0064***	0.0385***
	(0.0001)	(0.00001)	(0.0002)
Intercept	0.0502***	0.0851***	0.0512***
	(0.0009)	(0.0002)	(0.0028)
Treatment Effect	-0.1401***	-0.0088***	
	(0.0001)	(0.0004)	
Elasticity	-5.352%***	-0.334%***	
	(0.0001)	(0.0001)	
Observations	95,670,449	95,670,449	
R^2	0.1055	0.7297	0.1055

Notes: Both specifications control for a linear function in full fare (fare level + additional charges) and origin–destination–date–two-hour-time cell fixed effects. The front-door criterion specification is estimated using seemingly unrelated regression. Heteroskedasticity-robust standard errors are in parentheses and standard errors clustered by origin-destination-two hour time cell are in brackets. Standard errors for the front-door criterion treatment effect and elasticity are computed using the delta method. *** p<0.01, ** p<0.05, * p<0.1.

TABLE A.4: Estimation Results for Level of Tip Percentage

	Naïve	Front-Door	
Variables	Tip %	Shared Trip	Tip %
	(1)	(2)	(3)
Sharing Authorized (X)	-0.0180***	0.6769***	-0.0159***
	(0.00003)	(0.0002)	(0.0001)
Shared Trip (M)	–	–	-0.0031***
			(0.0001)
Full Fare (F)	0.0005***	-0.0064***	-0.0004***
	(0.0001)	(0.00001)	(0.0002)
Intercept	0.0489***	0.0851***	0.0491***
	(0.00005)	(0.0002)	(0.0001)
Treatment Effect	-0.0180***	-0.0021***	
	(0.00003)	(0.00004)	
Elasticity	-0.0890***	-0.105%***	
	(0.0001)	(0.0002)	
Observations	95,670,449	95,670,449	
R^2	0.1055	0.7297	0.1055

Notes: Both specifications control for a linear function in full fare (fare level + additional charges) and origin–destination–date–two-hour-time cell fixed effects. The front-door criterion specification is estimated using seemingly unrelated regression. Heteroskedasticity-robust standard errors are in parentheses and standard errors clustered by origin-destination-two hour time cell are in brackets. Standard errors for the front-door criterion treatment effect and elasticity are computed using the delta method. *** p<0.01, ** p<0.05, * p<0.1.

A6. Ride-Hailing Application Results without Full Fare Control

TABLE A.5: Results for Tipping at the Extensive Margin, Omitting Fare as a Control

	Naïve	Front-Door	
Variables	Tipped	Shared Trip	Tipped
	(1)	(2)	(3)
Sharing Authorized (X)	-0.0812***	0.7005***	-0.0671***
	(0.0001)	(0.0001)	(0.0002)
Shared Trip (M)	–	–	-0.0202***
			(0.0002)
Intercept	0.1973***	-0.0001***	0.1973***
	(0.00005)	(0.00001)	(0.00005)
Treatment Effect	-0.0812***	-0.0141***	
	(0.00003)	(0.0001)	
Elasticity	-8.748%***	-1.521%***	
	(0.0001)	(0.0001)	
Observations	95,670,449	95,670,449	
R^2	0.1150	0.7266	0.1151

Notes: Both specifications control for origin–destination–date–two-hour-time cell fixed effects and exclude fare controls. The front-door criterion specification is estimated using seemingly unrelated regression. Robust standard errors in parentheses. Standard errors for the front-door criterion treatment effect and elasticity are computed using the delta method. *** p<0.01, ** p<0.05, * p<0.1.

TABLE A.6: Results for Tipping at the Intensive Margin, Omitting Fare as a Control

	Naïve	Front-Door	
Variables	arcsinh(Tip)	Shared Trip	arcsinh(Tip)
	(1)	(2)	(3)
Sharing Authorized (X)	-0.1505***	0.7005***	-0.1220***
	(0.0001)	(0.0001)	(0.0003)
Shared Trip (M)	–	–	-0.0407***
			(0.0003)
Intercept	0.1973***	-0.0001***	0.3214***
	(0.00005)	(0.00001)	(0.0001)
Treatment Effect	-0.1505***	-0.0285***	
	(0.0001)	(0.0001)	
Elasticity	-6.457%***	-1.224%***	
	(0.0001)	(0.0001)	
Observations	95,670,449	95,670,449	
R^2	0.1501	0.7266	0.1501

Notes: Both specifications control for origin–destination–date–two-hour-time cell fixed effects and exclude fare controls. The front-door criterion specification is estimated using seemingly unrelated regression. Robust standard errors in parentheses. Standard errors for the front-door criterion treatment effect and elasticity are computed using the delta method. *** p<0.01, ** p<0.05, * p<0.1.

TABLE A.7: Results for Tip as a Fraction of Fare, Omitting Fare as a Control

	Naïve	Front-Door	
Variables	arcsinh(Tip/Fare)	Shared Trip	arcsinh(Tip/Fare)
	(1)	(2)	(3)
Sharing Authorized (X)	-0.0160***	0.7005***	-0.0144***
	(0.00002)	(0.0001)	(0.00005)
Shared Trip (M)	–	–	-0.0023***
			(0.00005)
Intercept	0.0419***	-0.0001***	0.0419***
	(0.00001)	(0.00001)	(0.00001)
Treatment Effect	-0.0160***	-0.0016***	
	(0.0001)	(0.0003)	
Elasticity	-7.946%***	-0.802%***	
	(0.0001)	(0.0002)	
Observations	95,670,449	95,670,449	
R^2	0.0865	0.7266	0.0865

Notes: Both specifications control for origin–destination–date–two-hour-time cell fixed effects and exclude fare controls. The front-door criterion specification is estimated using seemingly unrelated regression. Robust standard errors in parentheses. Standard errors for the front-door criterion treatment effect and elasticity are computed using the delta method. *** p<0.01, ** p<0.05, * p<0.1.

A7. Ride-Hailing Application without Time-Place FEs

TABLE A.8: Results for Tipping at the Extensive Margin, Omitting Fixed Effects

Variables	Naïve	Front-Door	
	Tipped	Shared Trip	Tipped
	(1)	(2)	(3)
Sharing Authorized (X)	-0.0982***	0.7015***	-0.0926***
	(0.0001)	(0.0001)	(0.0002)
Shared Trip (M)	–	–	-0.0081***
			(0.0002)
Intercept	0.1374***	-0.0050***	0.1373***
	(0.0001)	(0.00002)	(0.0001)
Treatment Effect	-0.0982***	-0.0057***	
	(0.0001)	(0.0001)	
Elasticity	-10.5802%***	-0.612%***	
	(0.0001)	(0.0001)	
Observations	95,670,449	95,670,449	
R^2	0.0231	0.6526	0.0232

Notes: Both specifications control for fare and exclude origin–destination–date–two-hour-time cell fixed effects. The front-door criterion specification is estimated using seemingly unrelated regression. Robust standard errors in parentheses. Standard errors for the front-door criterion treatment effect and elasticity are computed using the delta method. *** p<0.01, ** p<0.05, * p<0.1.

TABLE A.9: Results for Tipping at the Intensive Margin, Omitting Fixed Effects

	Naïve	Front-Door	
Variables	arcsinh(Tip)	Shared Trip	arcsinh(Tip)
	(1)	(2)	(3)
Sharing Authorized (X)	-0.1507***	0.7015***	-0.0926***
	(0.0001)	(0.0001)	(0.0002)
Shared Trip (M)	–	–	-0.0181***
			(0.0002)
Intercept	0.1029***	-0.0050***	0.1029***
	(0.0002)	(0.00003)	(0.0002)
Treatment Effect	-0.1507***	-0.0127***	
	(0.0001)	(0.0001)	
Elasticity	-8.6952%***	-0.7327%***	
	(0.0001)	(0.0001)	
Observations	95,670,449	95,670,449	
R^2	0.0515	0.6526	0.0515

Notes: Both specifications control for fare and exclude origin–destination–date–two-hour-time cell fixed effects. The front-door criterion specification is estimated using seemingly unrelated regression. Robust standard errors in parentheses. Standard errors for the front-door criterion treatment effect and elasticity are computed using the delta method. *** p<0.01, ** p<0.05, * p<0.1.

TABLE A.10: Results for Tip as a Percentage of Fare, Omitting Fixed Effects

	Naïve	Front-Door	
Variables	arcsinh(Tip/Fare)	Shared Trip	arcsinh(Tip/Fare)
	(1)	(2)	(3)
Sharing Authorized (X)	-0.0246***	0.7015***	-0.0226***
	(0.00002)	(0.0001)	(0.00004)
Shared Trip (M)	–	–	-0.0028***
			(0.00004)
Intercept	0.0438***	-0.0050***	0.0438***
	(0.00002)	(0.00003)	(0.00002)
Treatment Effect	-0.0246***	-0.0020***	
	(0.0001)	(0.00003)	
Elasticity	-12.6782%***	-1.0275%***	
	(0.0001)	(0.0001)	
Observations	95,670,449	95,670,449	
R^2	0.0100	0.6526	0.0100

Notes: Both specifications control for fare and exclude origin–destination–date–two-hour-time cell fixed effects. The front-door criterion specification is estimated using seemingly unrelated regression. Robust standard errors in parentheses. Standard errors for the front-door criterion treatment effect and elasticity are computed using the delta method. *** p<0.01, ** p<0.05, * p<0.1.

A8. Ride-Hailing Application without Controls Estimated Nonparametrically

In this section, we show results using an alternative nonparametric estimation method. Specifically, we estimate simple in-sample conditional expectations and their product (as in Pearl and McKenzie 2018, for example) while omitting control variables for the sake of simplicity in this illustration. The results are qualitatively similar to the main results presented in Section 3.

TABLE A.11: Nonparametric Results for Tipping (All Margins) Omitting All Controls

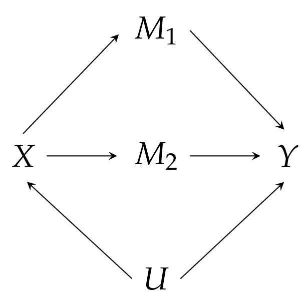
	(1) Tipped (Extensive Margin)	(2) arcsinh(Tip) (Intensive Margin)	(3) arcsinh(Tip/Fare) (Intensive Margin)
$\hat{\beta}_{\text{Naïve}} = E[Y\|X=1] - E[Y\|X=0]$	-0.1167	-0.2145	-0.2459
$\hat{\gamma} = E[M\|X=1]$	0.7001	0.7001	0.7001
$\hat{\delta} = E[Y\|M=1, X=1] - E[Y\|M=0, X=0]$	-0.0057	-0.0010	-0.0029
Naïve (i.e., $\hat{\beta}_{\text{Naïve}}$)	-0.1167	-0.2459	-0.2459
FDC (i.e., $\hat{\beta}_{\text{FDC}} = \hat{\gamma} \times \hat{\delta}$)	-0.0040	-0.0070	-0.0020
Observations	95,670,449	95,670,449	95,670,449

Notes: Each row reports different nonparametric estimates obtained by computing in-sample conditional expectations or product thereof.

A9. Multiple Mediators

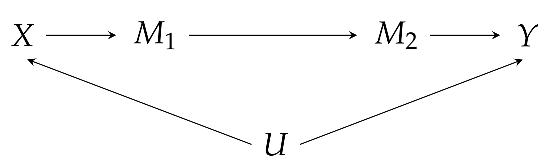
Pearl's (1995, 2000) canonical treatment of the front-door criterion assumes that M is a single variable, and not a vector of mediator variables. In this section of the Appendix we consider how to implement the front-door criterion in cases where we have multiple mediators.

FIGURE A.2: Multiple Mediators—Case 1



There are two basic cases in which we can imagine multiple mediators. Of course, one can imagine more complicated cases that combine these two basic cases, and we will illustrate one such more complicated case. For illustrative purposes, we will examine these three cases separately. In the first case, as shown in Figure A.2, the multiple mediators are independent from each other. Specifically, a path flows from X to both M_1 and M_2, and additionally, a path flows from both M_1 and M_2 to Y. In this case, M_1 and M_2 together intercept all directed paths from X to Y and meet the requirement Assumption 1.[35] By simply examining Figure A.2 it is clear that omitting either M_1 or M_2 from the estimation will violate Assumption 1, since the single mediator does not intercept all directed paths from X to Y.

FIGURE A.3: Multiple Mediators—Case 2



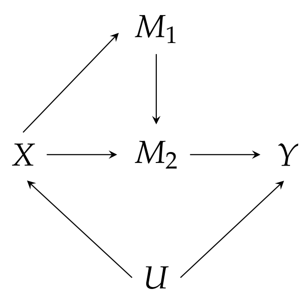
In the second case, as shown in Figure A.3, the multiple mediators both lie on the same path between X and Y. Specifically, a path flows from X to M_1, from M_1 to M_2, and finally from M_2 to Y. In this case, either M_1 or M_2 intercept all directed paths from X to Y and meet the requirement of Assumption 1. In contrast to the previous case, omitting

[35] See ch. 10 in Morgan and Winship (2015) for a similar discussion. This case, where M_1 and M_2 together intercept all directed paths from X to Y is similar to the surrogate index of Athey et al. (2019).

either M_1 or M_2 from the estimation will not violate Assumption 1, since both mediators individually intercept all directed paths from X to Y. Therefore the front-door criterion approach will recover the treatment effect when using either only M_1, only M_2, or both M_1 and M_2 as mediators in the front-door criterion estimation.

In the third, more complicated case, as shown in Figure A.4, one of the two mediators depends on both X and the other mediator. A path flows from X to M_1 and to M_2, from M_1 to M_2, and finally from M_2 to Y. In this case, either only M_2 intercepts all directed paths from X to Y and meet the requirement of Assumption 1. Omitting M_1 from the estimation will violate Assumption 1. Therefore the front-door criterion approach will recover the treatment effect when using either only M_2, or both M_1 and M_2 as mediators in the front-door criterion estimation.

FIGURE A.4: Multiple Mediators—Case 3



We now show simulation results that demonstrate the consequences of multiple mediators for each of these three cases. We start with the first case, where multiple mediators lie on different paths from X to Y. Our simulation setup is as follows. Let $U_i \sim N(0,1)$, $\epsilon_{Xi} \sim N(0,1)$, $Z_{1i} \sim U(0,1)$, $Z_{2i} \sim U(0,1)$, $\epsilon_{M1i} \sim N(0,1)$, $\epsilon_{M2i} \sim N(0,1)$, and $\epsilon_{Yi} \sim N(0,1)$ for a sample size of $N = 100,000$ observations. Then, let

$$X_i = 0.5U_i + \epsilon_{Xi}, \tag{A.25}$$

$$M_{1i} = Z_{1i}X_i + \epsilon_{M1i}, \tag{A.26}$$

$$M_{2i} = Z_{2i}X_i + \epsilon_{M2i}, \tag{A.27}$$

and

$$Y_i = 0.5M_{1i} + 0.5M_{2i} + 0.5U_i + \epsilon_{Yi}. \tag{A.28}$$

As illustrated in Figure A.2, this fully satisfies Pearl's (1995, 2000) three assumptions for the front-door criterion to be able to estimate the causal effect of X on Y. This simulation also meets the full support requirement (i.e., $P(X|M) > 0$) and so the front-door criterion will estimate the ATE. By substituting Equations A.26 and A.27 into Equation A.28, we can see that the true effect is equal to 0.500 in our simulations.

We estimate several specifications. The first baseline specification estimates

$$Y = \alpha_3 + \beta_3 X_i + \zeta_3 U_i + \epsilon_{3i}, \tag{A.29}$$

where, because both X and U are included on the right-hand-side, $E(\hat{\beta}_3) = \beta$.

The second specification estimates

$$M_{1i} = \kappa_3 + \gamma_3 X_i + \omega_{3i}, \tag{A.30}$$

$$M_{2i} = \pi_3 + \rho_3 X_i + \eta_{3i}, \text{ and} \tag{A.31}$$

$$Y_i = \lambda_3 + \delta_3 M_{1i} + \tau_3 M_{2i} + \phi_3 X_i + \nu_{3i}, \tag{A.32}$$

where the unobserved confounder U does not appear anywhere. The small difference in the case of multiple independent mediators is because the causal effect of X on Y is calculated by adding two products together. That is, $\hat{\beta} = (\hat{\gamma}_3 \cdot \hat{\delta}_3) + (\hat{\rho}_3 \cdot \hat{\tau}_3)$.

Column 1 of Table A.12 shows our benchmark estimation results for Equation A.29. Column 2 shows estimation results for the naïve version of Equation A.29 which omits the unobserved confounder U. Columns 3, 4, and 5 show front-door criterion estimation results using the specification outlined in Equations A.30 to A.32, respectively. Again, the estimates in columns 1 and 2 are quite different. While the treatment effect estimate is equal to 0.501 in the benchmark case, it is much larger, at 0.703, in the naïve case.

Given the derivations above, it should be unsurprising that the front-door criterion approach accurately estimates the causal effect of X on Y. The front-door criterion approach first multiplies the coefficient on X in column 3 by the coefficient on M_1 in column 5. Next, the front-door criterion approach multiplies the coefficient on X in column 4 by the coefficient on M_2 in column 5. Finally, these two products are summed to estimate the causal effect. Assuming the estimate in column 1 is not correlated with the estimate computed from columns 3 through 5, the two effect estimates are statistically identical. In both cases, the estimated effect is not statistically different from its true value of 0.500. Finally, in column 6, the direct effect of treatment conditional on M_1, M_2, and U is statistically indistinguishable from zero. More interesting, however, is investigating and interpreting estimates when we erroneously omit one of the mediators (say, for example, M_2) from the front-door criterion estimation. In this case, we no longer can correctly assume no "direct effect" of X on Y since there is a directed path independent of M_1 via M_2. This violates Assumption 1 above. When we omit M_2 from the front-door criterion estimation

TABLE A.12: Simulation Results—Multiple Mediators, Case 1

	Benchmark	Naïve	Front-Door			Direct Effect	Biased Front-Door		Direct Effect
Variables	Y	Y	M_1	M_2	Y	Y	M_1	Y	Y
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treatment (X)	0.501***	0.703***	0.497***	0.502***	0.204***	0.001	0.497***	0.457***	0.254***
	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)
Mediator (M_1)	–	–	–	–	0.498***	0.500***	–	0.495***	0.496***
					(0.003)	(0.003)		(0.004)	(0.003)
Mediator (M_2)	–	–	–	–	0.499***	0.499***	–	–	–
					(0.003)	(0.003)			
Confounder (U)	0.498***	–	–	–	–	0.501***	–	–	0.500***
	(0.004)					(0.004)			(0.004)
Intercept	-0.002	-0.003	-0.005	0.002	-0.002	-0.004	-0.005	-0.001	0.001
	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Treatment Effect	0.501***	0.703***		0.498***		–	0.246***		–
	(0.004)	(0.004)		(0.003)			(0.002)		
Observations	100,000	100,000		100,000		100,000	100,000		100,000

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The front-door equations in columns (3) and (4) are estimated by seemingly unrelated regressions. The standard error for the front-door criterion treatment effect is estimated by the delta method.

the estimated effect (shown in columns 7 and 8) is 0.246, considerably smaller than the true effect. Column 9 shows that the "direct effect" is 0.254.

The foregoing shows the consequences of omitting a mediator when using the front-door criterion approach. With that said, the effect estimated in the biased front-door criterion estimation in columns 7 and 8 of Table A.12 can be interpreted as the "indirect effect" of X on Y via M_1 and independent of M_2 (Imai et al. 2010, Acharya et al. 2016). In the literature on causal mediation analysis, the total causal effect is framed as the aggregation of both the direct and indirect effects (Imai et al. 2011). This effect, estimated using the front-door criterion approach, is similar in spirit to the "population intervention indirect effect" of Fulcher et al. (2020).

A very common approach for estimating indirect or mediating effects is to simply condition on potential mediating variables (Acharya et al. 2016). Despite the popularity of this approach, conditioning on potential mediating variables can lead to biased estimation—specifically in the case when omitted variables are affected by the treatment X and affect both the potential mediating variable M and the outcome Y (see, e.g., Acharya et al. 2016; Imai et al. 2010; Morgan and Winship 2015). If the assumptions in Section 2.1 hold coupled with the ability to relax Assumption 1—that the M intercepts all paths from X to Y—then the front-door criterion approach allows for valid estimation of indirect effects. Of course, whether or not the indirect effect is a parameter of interest for applied researchers will ultimately depend on the specific application and research question.

We now turn to the second case and demonstrate the consequences (or lack thereof) of multiple mediators of the sort illustrated in Figure 1, where multiple mediators lie on the same path from X to Y. Our simulation setup is as follows. Let $U \sim N(0,1)$, $\epsilon_X \sim N(0,1)$, $Z_1 \sim U(0,1)$, $Z_2 \sim U(0,1)$, $\epsilon_{M1} \sim N(0,1)$, $\epsilon_{M2} \sim N(0,1)$, and $\epsilon_Y \sim N(0,1)$ for a sample size of $N = 100,000$ observations. Then, let

$$X_i = 0.5U_i + \epsilon_{Xi}, \tag{A.33}$$

$$M_{1i} = Z_{1i}X_i + \epsilon_{M1i}, \tag{A.34}$$

$$M_{2i} = Z_{2i}M_{1i} + \epsilon_{M2i}, \tag{A.35}$$

and

$$Y_i = 0.5M_{2i} + 0.5U_i + \epsilon_{Yi}. \tag{A.36}$$

Similar to the previous simulation analysis, we estimate several specifications. The first specification estimates the true effect by controlling for the confounder U. The second specification estimates the treatment effect using the front-door criterion approach. As the results in Table A.13 show, effects estimated with the front-door criterion approach in this case are statistically invariant whether either or both M_1 and M_2 are included in the estimation procedure.

TABLE A.13: Simulation Results—Multiple Mediators, Case 2

	Benchmark	Naïve	Front-Door (Both)			Front-Door (M_1 only)		Front-Door (M_2 only)	
Variables	Y	Y	M_1	M_2	Y	M_1	Y	M_2	Y
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treatment (X)	0.127***	0.326***	0.495***	0.245***	0.201***	0.496***	0.120***	0.245***	0.202***
	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)
Mediator (M_1)	–	–	–	–	0.003	–	0.254***	–	–
					(0.004)		(0.004)		
Mediator (M_2)	–	–	–	–	0.502***	–	–	–	0.503***
					(0.003)				(0.003)
Confounder (U)	0.501***	–	–	–	–	–	–	–	–
	(0.004)								
Intercept	-0.001	0.004	0.002	0.004	0.002	0.002	0.004	0.004	0.002
	(0.004)	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)
Treatment Effect	0.127***	0.326***		0.125***		0.126***		0.123***	
	(0.004)	(0.004)		(0.002)		(0.002)		(0.002)	
Observations	100,000	100,000		100,000		100,000		100,000	

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The front-door equations in columns (3) and (4) are estimated by seemingly unrelated regressions. The standard error for the front-door criterion treatment effect is estimated by the delta method.

Finally, we turn to the third, more complicated, case and illustrate the case of multiple mediators of the when the one of the two mediators depends on both X and the other mediator, a situation that can be expressed with the following DAG.

Our simulation setup is as follows. Let $U \sim N(0,1)$, $\epsilon_X \sim N(0,1)$, $Z_1 \sim U(0,1)$, $Z_2 \sim U(0,1)$, $Z_3 \sim U(0,1)$, $\epsilon_{M1} \sim N(0,1)$, $\epsilon_{M2} \sim N(0,1)$, and $\epsilon_Y \sim N(0,1)$ for a sample size of $N = 100,000$ observations. Then, let

$$X_i = 0.5U_i + \epsilon_{Xi}, \tag{A.37}$$

$$M_{1i} = Z_{1i}X_i + \epsilon_{M1i}, \tag{A.38}$$

$$M_{2i} = Z_{2i}M_{1i} + Z_{3i}X_i + \epsilon_{M2i}, \tag{A.39}$$

and

$$Y_i = 0.5M_{2i} + 0.5U_i + \epsilon_{Yi}. \tag{A.40}$$

Similar to the previous simulation analysis, we estimate several specifications. The first specification estimates the true effect by controlling for the confounder U. The second specification estimates the treatment effect using the front-door criterion approach. As the results in Table A.14 show, effects estimated with the front-door criterion approach differ based on which of the two mediators are included in the regressions. If we include both $M1$ and $M2$, then the front-door criterion yields unbiased estimates of the treatment effect. If we include only $M1$, then the front-door criterion yields biased estimates of the treatment effect because the first front-door criterion assumption fails to hold. Instead, if we include only $M2$, then the front-door criterion yields unbiased estimates of the treatment effect.

TABLE A.14: Simulation Results—Multiple Mediators, Case 3

Variables	Benchmark	Naïve	Front-Door (Both)			Front-Door (M_1 only)		Front-Door (M_2 only)	
	Y	Y	M_1	M_2	Y	M_1	Y	M_2	Y
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treatment (X)	0.374***	0.571***	0.496***	0.743***	0.200***	0.496***	0.446***	0.743***	0.200***
	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)
Mediator (M_1)	–	–	–	–	0.001	–	0.25***	–	–
					(0.004)		(0.004)		
Mediator (M_2)	–	–	–	–	0.498***	–	–	–	0.499***
					(0.003)				(0.003)
Confounder (U)	0.494***	–	–	–	–	–	–	–	–
	(0.004)								
Intercept	0.000	-0.001	0.002	-0.003	0.000	-0.002	-0.002	-0.003	0.000
	(0.004)	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)
Treatment Effect	0.374***	0.571***		0.370***		0.124***		0.370***	
	(0.004)	(0.004)		(0.003)		(0.002)		(0.003)	
Observations	100,000	100,000		100,000		100,000		100,000	

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The front-door equations in columns (3) and (4) are estimated by seemingly unrelated regressions. The standard error for the front-door criterion treatment effect is estimated by the delta method.

A10. Violations of Strict Exogeneity

Together, Assumptions 2 and 3 imply that the mediator M is excludable. More formally, the strict exogeneity of M implies that $P(U|M,X) = P(U|X)$ and $P(Y|X,M,U) = P(Y|M,U)$. In this sub-section, building on the work of Glynn and Kashin (2018), we examine violations of this assumption. Again, we do this with a simulation analysis.

Our simulation setup is the same as in the Appendix, except that here we allow for the endogeneity of M. Let $U_i \sim N(0,1)$, $Z_i \sim U(0,1)$, $\epsilon_{Xi} \sim N(0,1)$, $\epsilon_{Mi} \sim N(0,1)$, and $\epsilon_{Yi} \sim N(0,1)$ for a sample size of $N = 100,000$ observations. Then, let

$$X_i = 0.5U_i + \epsilon_{Xi}, \tag{A.41}$$

$$M_i = Z_i X_i + \Gamma U_i + \epsilon_{Mi}, \tag{A.42}$$

and

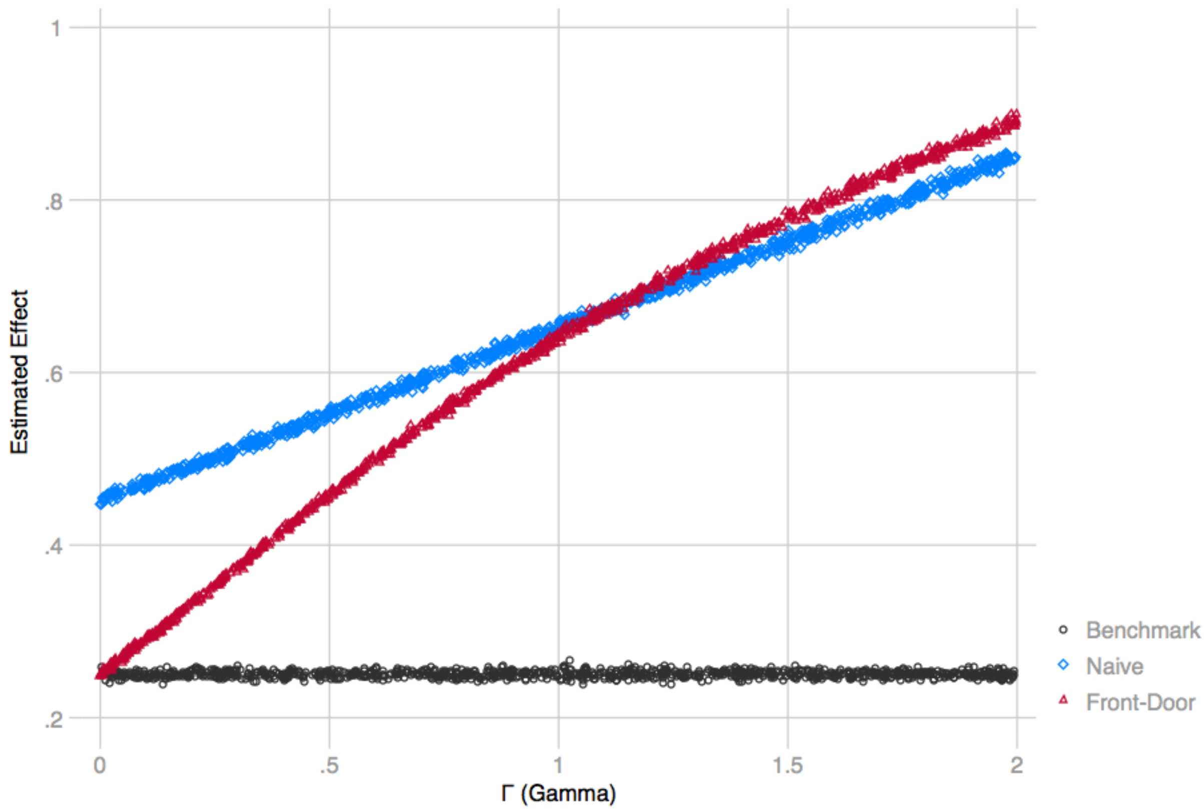
$$Y_i = 0.5M_i + 0.5U_i + \epsilon_{Yi}. \tag{A.43}$$

The critical difference here is that now, when defining M in equation A.42, U is included on the right-hand-side. The parameter Γ defines the strength of the relationship between U and M. In this simulation analysis we let $\Gamma \sim U(0,2)$. By permitting values of Γ to vary allows the degree of endogeneity in our simulations to vary. Again, this simulation meets the full support requirement (i.e., $P(X|M) > 0$) and so the front-door criterion will estimate the ATE.

We show these simulation results graphically. Figure A.5 illustrates how having an endogenous mediator influences the credibility of using the front-door criterion approach. This figure shows estimated effects for three estimation approaches. First, the benchmark estimates (black circles), which include the confounder U on the right-hand-side of the regression equation, accurately estimates the true effect of 0.250. Second, the naïve estimates (blue diamonds), which omits the confounder U from the regression equation, consistently overestimates the effect. The size of this bias is increasing in the strength of the endogeneity of M. This is because, as Γ increases, the influence of the confounder U in the relationship between X and Y increases. Finally, the front-door estimates (red triangles) are estimated as described in equations A.17 and A.18 in section 2.

Once again, a few remarks are in order. First, and rather unsurprisingly, it is only when the degree of endogeneity of M is negligible (i.e., when Γ is infinitesimally close to zero) that the front-door criterion approach accurately estimates the true effect. Second, although the front-door criterion approach produces biased treatment effect estimates, when M is weakly endogenous (i.e., when $\Gamma > 0$ but still relatively small), these estimates

FIGURE A.5: The Consequences of an Endogenous Mediator



Notes: This figure illustrates simulation results using 1,000 replications from each estimation approach. The vertical axis represents the estimated effect. The horizontal axis represents the Gamma parameter, representing the degree of endogeneity, from equation 32. The benchmark estimates (black circles) all accurately estimate the true effect of 0.250. The naive estimates are shown in blue diamonds and the front-door estimates are shown in red triangles.

are less biased than the naïve estimates. Third, when M is strongly endogenous the front-door criterion approach produces estimates that are worse—that is, more biased—than the naïve estimates.

These details lead to an important discussion for applied researchers who may want to implement the front-door criterion approach in their given empirical setting. In many cases, strict exogeneity of M may be debatable. Indeed, outside of an experimental setting, convincingly arguing that $P(U|M,X) = P(U|X)$ and $P(Y|X,M,U) = P(Y|M,U)$ will likely be challenging. That said, however, if applied researchers can convincingly argue that the degree of endogeneity of M is relatively weak—that M is not strictly exogenous but that it is plausibly exogenous (Conley et al., 2012), so to speak—then the front-door criterion approach will produce more reliable estimates compared to the naïve approach which consists in regressing Y on an endogenous X. On the other hand, when the endogeneity of M is obviously relatively strong, using the front-door criterion approach could lead to more bias in estimates than the naïve approach. Specifically in our simulation set-up, the front-door criterion estimates begin to become just as biased as the naïve estimates when Γ is equal to one. In the way we have defined our variables, this means that the direct effect of U on M is about twice as strong as the indirect effect of U on M via X. Of course when using real-world data, when we cannot observe U, testing the specific size of these relationships is impossible. In all practical settings, the case for the exogeneity of M will rely on careful reasoning based on the given empirical setting.

A11. Ride-Hailing Application FDC Results without Controlling for X in Stage 2

TABLE A.15: Estimation Results for Tipping at the Extensive Margin

Variables	Naïve	Front-Door	
	Tipped	Shared Trip	Tipped
	(1)	(2)	(3)
Sharing Authorized (X)	-0.0628***	0.6769***	–
	(0.0001)	(0.0002)	
Shared Trip (M)	–	–	-0.0594***
			(0.0002)
			[0.0002]
Full Fare (F)	0.0050***	-0.0064***	0.0055***
	(0.00001)	(0.00001)	(0.00004)
Intercept	0.1306***	0.0851***	0.1194***
	(0.0002)	(0.0002)	(0.0005)
Treatment Effect	-0.0628***	-0.0403***	
	(0.0001)	(0.0001)	
Elasticity	-6.764%***	-4.335%***	
	(0.0001)	(0.0001)	
Observations	95,670,449	95,670,449	
R^2	0.1165	0.7297	0.1157

Notes: Both specifications control for a linear function in full fare (fare level + additional charges) and origin–destination–date–two-hour-time cell fixed effects. The front-door criterion specification is estimated using seemingly unrelated regression. Heteroskedasticity-robust standard errors are in parentheses. Standard errors for the front-door criterion treatment effect and elasticity are computed using the delta method. Standard errors for the front-door criterion treatment effect and elasticity are computed using the delta method. *** p<0.01, ** p<0.05, * p<0.1.

TABLE A.16: Estimation Results for Tipping at the Intensive Margin

Variables	Naïve	Front-Door	
	arcsinh(Tip)	Shared Trip	arcsinh(Tip)
	(1)	(2)	(3)
Sharing Authorized (X)	-0.0958***	0.6769***	–
	(0.0002)	(0.0002)	
Shared Trip (M)	–	–	-0.0895***
			(0.0003)
Full Fare (F)	0.0148***	-0.0064***	0.0157***
	(0.00003)	(0.00002)	(0.0001)
Intercept	0.1233***	0.0851***	0.1056***
	(0.0004)	(0.0002)	(0.0010)
Treatment Effect	-0.0958***	-0.0606***	
	(0.0002)	(0.0002)	
Elasticity	-5.527%***	-3.498%***	
	(0.0001)	[0.0002]	
Observations	95,670,449	95,670,449	
R^2	0.1545	0.7297	0.1538

Notes: Both specifications control for a linear function in full fare (fare level + additional charges) and origin–destination–date–two-hour-time cell fixed effects. The front-door criterion specification is estimated using seemingly unrelated regression. Heteroskedasticity-robust standard errors are in parentheses. Standard errors for the front-door criterion treatment effect and elasticity are computed using the delta method. Standard errors for the front-door criterion treatment effect and elasticity are computed using the delta method. *** p<0.01, ** p<0.05, * p<0.1.

TABLE A.17: Estimation Results for Tip as a Fraction of Fare

Variables	Naïve	Front-Door	
	arcsinh(Tip/Fare)	Shared Trip	arcsinh(Tip/Fare)
	(1)	(2)	(3)
Sharing Authorized (X)	-0.0175***	0.6769***	–
	(0.00003)	(0.0002)	
Shared Trip (M)	–	–	-0.0165***
			(0.00004)
Full Fare (F)	-0.0004***	-0.0064***	-0.0002***
	(0.000003)	(0.00002)	(0.000008)
Intercept	0.0473***	-0.0851***	0.0441***
	(0.00005)	(0.0002)	(0.0001)
Treatment Effect	-0.0175***	-0.0112***	
	(0.00003)	(0.00003)	
Elasticity	-9.015%***	-5.757%***	
	(0.0001)	(0.0002)	
Observations	95,670,449	95,670,449	
R^2	0.0867	0.7297	0.0857

Notes: Both specifications control for a linear function in full fare (fare level + additional charges) and origin–destination–date–two-hour-time cell fixed effects. The front-door criterion specification is estimated using seemingly unrelated regression. Heteroskedasticity-robust standard errors are in parentheses. Standard errors for the front-door criterion treatment effect and elasticity are computed using the delta method. *** p<0.01, ** p<0.05, * p<0.1.