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

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

March 9, 2022

Abstract

We present the first application of Pearl's (1995) front-door criterion wherein the assumptions for point identification plausibly hold with observational data. For identification, the front-door criterion leverages exogenous mediator variables on the causal path. After explaining the identification assumptions and estimation framework of the front-door criterion, we present empirical applications. In our core application we estimate the effect of authorizing a shared Uber or Lyft ride on tipping by leveraging the plausibly exogenous variation in whether one actually shares a ride with a stranger conditional on authorizing sharing, on the full fare paid, and on origin–destination fixed effects interacted with two-hour interval fixed effects. We find that most of the observed negative effect on tipping is driven by selection. We then explore the consequences of violating the identification assumptions.

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

JEL Codes: C13, C18, R40, D90

*We 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 and North American Winter Meetings of the Econometric Society as well as seminar participants at UC Berkeley, Michigan State University, the Montréal Methods Workshop, the University of Idaho, the University of New Mexico, and the World Bank for useful comments and suggestions. This paper was supported in part by the US Department of Agriculture, Economic Research Service. The findings and conclusions in this manuscript are those of the authors and should not be construed to represent any official US Department of Agriculture or US Government determination or policy. All remaining errors are ours.

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

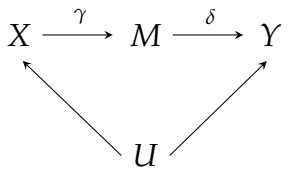
‡Research Economist, United States Department of Agriculture, Economic Research Service, MS9999, Beacon Facility, P. O. Box 419205, Kansas City, MO 64141, Email: Jeffrey.Bloem@usda.gov.

§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 present the first application of Pearl's (1995, 2000) front-door criterion (FDC) to observational data in which the required assumptions *for point-identification* plausibly hold.[1] The directed acyclic graph (DAG) in Figure I illustrates the FDC 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.[2]

FIGURE I: The Front-Door Criterion



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 average treatment effect (ATE) of X on Y.[3] 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 ATE of X on Y, which we label $\hat{\beta}$ in keeping with convention. Intuitively, the FDC estimates the ATE because it decomposes a reduced-form relationship that is not causally identified into two causally identified relationships.[4]

[1] This is not the first application of the FDC in which assumptions for identification plausibly hold *in general*. As we discuss below, Glynn and Kashin (2017, 2018) present applications of the FDC wherein partial identification was feasible.

[2] 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 simplicity and because that is what the majority of applied economists are familiar with.

[3] 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."

[4] Some readers may be tempted to use the mediator M as an instrument for the endogenous treatment X. We explain in the Supplemental Appendix why this is problematic.

Despite its relative simplicity, economists have been reluctant to incorporate the front-door criterion in their empirical toolkit.[5] Anecdotally, that resistance appears to stem from 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). We provide such an application and address the following questions: How can the front-door criterion be used in the context of linear regression? And what happens when the necessary identification assumptions for the front-door criterion to estimate an ATE do not hold strictly?

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 on 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 FDC identifying assumptions (see, e.g., Imbens 2020). Consequently, the adoption of the FDC has been slow among applied researchers. The only extant published social science applications of the FDC are by Glynn and Kashin (2017; 2018).[6] 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 FDC approach do not hold.[7]

Our contribution is threefold. First, because linear regression remains the workhorse of applied economics and because an explanation of how to use the front-door criterion in a regression context has so far been lacking in the literature, we briefly explain how to estimate treatment effects with the front-door criterion in the context of linear regression.[8]

[5] 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 shorter-run outcomes or "surrogate index" (Athey et al. 2019). Variants of 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).

[6] In Glynn and Kashin (2018) the authors apply the FDC approach to estimate the effects of attending a job training program on earnings. In Glynn and Kashin (2017) the authors also apply the FDC 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.

[7] Specifically, Glynn and Kashin (2018) write, "As we discuss in detail below, the assumptions implicit in [the FDC] 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 FDC approach which requires an exclusion restriction and a parallel trends assumption specifically for their empirical setting where the necessary conditions for the FDC 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.

[8] Although Morgan and Winship (2015) dedicate part of a chapter to the FDC, they only indirectly present

Estimation relies on Pearl's three identification assumptions and an additional empirically verifiable requirement of one-sided noncompliance.[9]

Second, we present two examples of the FDC in practice. One uses simulated data to show an ideal application of the FDC—one in which we know the true ATE.[10] Our second example is the core contribution of this paper because it presents the first application of the FDC to observational data where the necessary assumptions for point-identification of the ATE plausibly hold. In that application, we estimate the effect on tipping behavior of authorizing a ride-hailing app such as Lyft or Uber to overlap a ride with another paying passenger. When authorizing a shared ride, a shared ride is not guaranteed. Rather, whether one shares a ride or not depends on a number of 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–destination fixed effects, and the interaction of two-hour time slot fixed effects and origin–destination fixed effects, we effectively control for those supply- and demand-side factors, which makes whether one gets to actually share a ride or not plausibly exogenous.[11] This battery of fixed effects serves to exogenize whether one gets to share a ride conditional on having authorized a shared ride 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 soaked up by the interaction of origin–destination fixed effects interacted and two-hour time slot fixed effects, the same battery of fixed effects also serves to exogenize whether one gets to share a ride conditional on having authorized a shared ride 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.

Exploiting this plausibly exogenous mediator for identification, we find that the observed negative relationship between authorizing a shared ride and tipping is almost entirely explained by selection into treatment. In other words, our finding first dispels the notion, common among Lyft and Uber drivers, that it is the decision to share a ride (rather than the type of person making that decision) that leads to lower tips on shared rides (Bowman 2019; Harrington 2019). Second, our finding also suggests that ride-hailing

a regression-based approach to the FDC.

[9] This contribution builds on the previous work by Chalak and White (2011).

[10] This simulated example serves as the basis for our third contribution below, where we explore departures from the necessary assumptions for the front-door criterion to yield the average treatment effect of X on Y.

[11] 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 represents fixed effect different from the midnight to 2 AM time slot on January 2.

companies like Lyft and Uber could significantly 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. 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 to the economics of tipping more broadly (Azar 2020).

In our application—as is the case in many applied settings—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 behavior (i.e., Y). Riders who authorize sharing are likely drawn to do so by the lower fare and likely also hold a higher propensity 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 FDC 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).[12]

Third, and importantly for applied researchers interested in using the front-door criterion in their own work, we explore what happens when the necessary assumptions for the front-door criterion to identify the ATE of X on Y fail to hold. Specifically, we look at what happens when (i) there are multiple mediators, some of which may be omitted from the estimation specification, (ii) the violation of the strict exogeneity of M assumption, or (iii) the date requirement that $P(X_i|M_i) > 0$ as stated in Pearl (2000) is violated.

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 two empirical illustrations, one using simulated

[12]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 I because consumers may feel more generous the higher the fare differential. We explain in section 3.2 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 4.2, the results of these models reflect those 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.

data and the other using real-world data. In section 4, we explore departures from some of the assumptions underpinning the FDC. We conclude in section 5 by offering practical recommendations for using the FDC in empirical research.

2 The Front-Door Criterion: Identification and Estimation

We begin this section by formally introducing the front-door criterion (FDC) estimand. We first present and expand upon the necessary point-identification assumptions noted by Pearl (1995, 2000), and demonstrate how these assumptions recover the ATE of X on Y with the presence of unobserved confounders U, as shown in Figure I above. We then offer our first contribution by explaining how to estimate treatment effects using the FDC in a linear regression context.

2.1 Identification

We are interested in estimating the ATE of X on Y in Figure I above. Recall that with observational data, estimating the ATE 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 FDC approach pictured in Figure I allows recovering the ATE of X on Y.

As discussed in Pearl (1995, 2000), the FDC 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. $Y(m, 1) = Y(m, 0)$.

This assumption states that the only way in which X influences Y is through M. In Figure I, 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 I is identified. In Pearl's terminology, there can be no back-door path between X and M.

Assumption 3. $E[Y(m,x)|M, X = x] = E[Y(m,x)|X = x]$.

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 I is identified. In Pearl's terminology, every back-door path between M and Y has to be blocked by X.[13]

We now summarize Pearl's (1995) proof and provide additional explanation to help the reader's intuition along the way by deriving the FDC estimand. In what follows we use $\gamma = P(M|X)$ as shorthand for $P(M = 1|X = 1) - P(M = 1|X = 0)$, $\delta = P(Y|M, X)$ as shorthand for $P(Y = 1|M = 1, X) - P(Y = 1|M = 0, X)$, and $P(Y|\check{X})$ as shorthand for $P(Y = 1|\check{X} = 1) - P(Y = 1|\check{X} = 0)$. Our aim is to compute $P(Y|\check{X})$ with observable variables, where $P(Y|\check{X})$ represents the ATE of X on Y. Pearl (1995) introduces $\check{X}$ ("X check") as shorthand for an intervention that sets the checked variable to a specific value.[14] Thus, $P(Y|\check{X} = x)$ 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 (Haavelmo 1943; Strotz and Wold 1960; Heckman et al. 2013). $P(Y|\check{X})$ should be read as the ATE of X on Y. This should be contrasted with $P(Y|X)$, which may not represent the ATE of X on Y due to the presence of the unobserved confounder U.

As shown in Figure I, observing $\check{X}$ is complicated by the presence of the unobserved confounder U. Therefore, our goal here is to restate $P(Y|\check{X})$ using only the observed variables M, X, and Y while leveraging Assumptions 1 through 3.

The first step is to compute $P(M|\check{X})$. 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

$$P(M|\check{X}) = P(M|X), \tag{1}$$

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

The second step is to compute $P(Y|\check{M})$. Here we cannot set $\check{M} = M$ 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

[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).

[14]Readers familiar with Pearl's $do(\cdot)$ operator will have noted that $\check{X}$ as defined here is equivalent to $do(x)$. We avoid using the $do(\cdot)$ so as to avoid burdening the reader with new notation.

can write

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

where the right-hand-side of Equation 2 involves two expressions involving $\check{M}$. The second term on the right-hand-side of Equation 2 can be reduced to $P(X)$ because, as stated by Assumption 1, the only way in which X influences Y is through M.[15] The first term on the right-hand-side of Equation 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

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

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

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

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

Assumption 3 allows us to rewrite M as $\check{M}$ in the first term on the right-hand-side of Equation 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 $\check{X}$ from the first term on the right-hand side of Equation 4. Said differently, M should have no effect on X, because X causes M and not vice versa in Figure I. Therefore, we can rewrite the first term on the right-hand side of Equation 4 as

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

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

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

[15] Since M is a descendent of X in Figure I, any exogenous variation in M will not influence X.

In later writings on the FDC, Pearl (2000) discusses an additional condition for identification, a data requirement which can be directly verified, and which thus need not be assumed. 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, $P(X_i|M_i) > 0$. In Pearl's canonical example of the relationship between smoking X 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). We discuss this condition in more detail in Section 4.

2.2 Estimation

We now discuss how to estimate treatment effects using the FDC. As stated above, our goal is to estimate the ATE of X on Y in Figure I, which is defined as $P(Y|\check{X})$ and is not equivalent to $P(Y|X)$ because of the presence of unobserved confounders U. When the necessary identification assumptions for the FDC hold, we can estimate the ATE by using the following linear regression-based approach. Let

$$M_i = \kappa + \gamma X_i + \omega_i \tag{7}$$

and

$$Y_i = \lambda + \delta M_i + \phi X_i + \nu_i. \tag{8}$$

In Equation 7, 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, since $Cov(X, \omega) = 0$. In Equation 8, Y is the outcome variable, which is related to X only through M. In this case, following 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, since $Cov(M, \nu) = 0$. Therefore, estimating Equations 7 and 8 and multiplying coefficient estimates $\hat{\delta}$ and $\hat{\gamma}$ by each other estimates β, the ATE of X on Y.

At this point, it is important to note a few things for clarity. First, we focus here on the context of linear regression because linear regression is the approach favored by the majority of applied economists.[16] We note, however, that although we have written Equations 7 and 8 as linear equations, directed acyclic graphs such as the one in Figure I impose no such linear relationships on their constituent variables, nor do they impose

[16]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 $E(M|X)$.

that the relationships be parametric.[17] Therefore the FDC is nonparametrically identifiable, and linear regression is but one way to estimate treatment effects using the FDC. Indeed in our real-world application, we present nonparametric estimation results in the Supplemental Appendix.

Second, the necessary point-identification Assumptions 1 through 3 lead to $Cov(X, \omega) = 0$ in Equation 7 and $Cov(M, \nu) = 0$ in Equation 8. This allows estimating δ and γ, and via multiplication, β, the ATE 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). We will illustrate this result directly in Section 3.

Finally, in our applications we estimate the FDC using a seemingly unrelated regressions (SUR) framework (Zellner 1962). Although the SUR framework is not necessary to estimate treatment effects using the FDC, it does have some useful features, such as ease of computation.

3 Empirical Illustration

We first show empirical results using simulated data. We then demonstrate the first empirical application of the FDC to observational data wherein the required assumptions plausibly hold for point-identification of the average treatment effect. Additionally, in the Supplemental Appendix, we replicate the experimental estimates of Beaman et al. (2013) using the FDC approach.

3.1 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.[18] Then, let

$$X_i = 0.5U_i + \epsilon_{Xi}, \tag{9}$$

$$M_i = Z_i X_i + \epsilon_{Mi}, \tag{10}$$

[17] 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.

[18] 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 11.

and

$$Y_i = 0.5M_i + 0.5U_i + \epsilon_{Yi}. \tag{11}$$

This satisfies Pearl's (1995, 2000) three point-identification assumptions for the FDC to be able to estimate the average treatment 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 setup also satisfies Pearl's additional data requirement that $P(X_i|M_i) > 0$. By substituting Equation 10 into Equation 11, it should be clear that the true ATE is equal to 0.250 in our simulations.

To show that the FDC estimates the ATE 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 ATE by virtue of controlling for the unobserved confounder U, estimates

$$Y_i = \alpha_0 + \beta_0 X_i + \zeta_0 U_i + \epsilon_{0i}, \tag{12}$$

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

Second, we estimate a naïve specification. The naïve specification differs from the benchmark specification in Equation 12 by failing to control for the presence of the unobserved confounder.

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

$$M_i = \kappa_0 + \gamma_0 X_i + \omega_{0i} \tag{13}$$

$$Y_i = \lambda_0 + \delta_0 M_i + \phi_0 X_i + \nu_{0i} \tag{14}$$

where the unobserved confounder U_i does not appear anywhere, but because the necessary assumptions for the FDC to identify the ATE hold, $E(\hat{\gamma}_0 \cdot \hat{\delta}_0) = \beta$, i.e., the true ATE.

Column 1 of Table I shows estimation results for Equation 12, 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 13 and 14, respectively. The line labeled "Estimated ATE" shows estimates of the ATE for each of those three specifications. Unsurprisingly, the estimates of the ATE 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 ATE estimate

TABLE I: Simulation Results—Ideal Case

Variables	Benchmark	Naïve	Front-Door		Direct Effect
	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)
Estimated ATE	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 ATE is estimated by the delta method.

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 ATE in column 1 is not correlated with the ATE computed from columns 3 and 4, the benchmark and front-door ATEs are statistically identical. In both cases, the estimated ATE is not statistically different from its true value of 0.250.

Column 5 in Table I 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 FDC 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

$$X_i = 0.5U_i + 0.5C_i + \epsilon_{Xi}, \tag{15}$$

$$M_i = Z_i X_i + 0.3C_i + \epsilon_{Mi}, \tag{16}$$

TABLE II: 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)
Estimated ATE	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 ATE is estimated by the delta method.

and

$$Y_i = 0.5M_i + 0.5U_i + 0.15C_i + \epsilon_{Yi}. \tag{17}$$

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

$$Y_i = \alpha_1 + \beta_1 X_i + \rho_1 C_i + \zeta_1 U_i + \epsilon_{1i}, \tag{18}$$

where the right-hand-side of Equation 18 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 FDC estimates

$$M_i = \kappa_1 + \gamma_1 X_i + \tau_1 C_i + \omega_{1i} \tag{19}$$

$$Y_i = \lambda_1 + \delta_1 M_i + \phi_1 X_i + \pi_1 C_i + \nu_{1i} \tag{20}$$

where both Equations 19 and 20 include C_i on the right hand side.

Table II 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" ATE of 0.251. Unsurprisingly the naïve specification overestimates the ATE and the FDC estimation approach yields an unbiased estimate of the ATE. These results highlight that, given an additional conditional ignorability assumption (Rosenbaum and Rubin 1983), the FDC 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 ATE with the FDC approach. The topic of a conditionally exogenous 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.

3.2 Ride Sharing and Tipping Behavior

Using publicly available data on over 95 million Lyft and Uber rides in Chicago during calendar year 2019, we apply the FDC to estimate the ATE 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 a proportion of the fare paid. After discussing the data we use in this application, we explain how the necessary conditions for the FDC to yield a consistent point estimate of the ATE plausibly 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,[19] two-hour time slot fixed effect,[20] and their interactions.

We find that naïve regressions overestimate the magnitude of the ATE of authorizing sharing on both tipping margins because of selection into treatment.[21] Our application

[19]Chicago is divided in 77 community areas. We code a unique fixed effect for each origin-destination dyad among those community areas.

[20]We code a unique fixed effect for each distinct two-hour time slot in 2019.

[21]Using an earlier version of the same data we use here, Harrington (2019) reports that when riders opt to share rides with another passenger using the Lyft Line or UberPool services, 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 ATE 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 ATE would capture the difference in tipping if UberPool or Lyft Line authorization was randomly assigned across all passengers,

illustrates a broader principle: UberPool and Lyft Line are cheaper and less convenient services compared to their solo-ride counterparts, and so there is a tradeoff between fare and (expected) inconvenience which in principle affects tipping.

The FDC is particularly useful in this context because it is able to reliably rule out the effect of selection into a lower convenience 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 ATE due to bias associated with more frugal customers, whose proclivity to tip is lower, selecting the cheaper service. Because the FDC estimates an unbiased ATE, 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.

3.2.1 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.[22] 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).[23]

These data provide the base fare (rounded to the nearest \$2.50) and tip amount (rounded to the nearest \$1.00).[24] For the extensive margin of tipping, our dependent variable is a dummy variable for whether a passenger tips at all. For the intensive margin, we use the observed tip value.[25] For tip as a percentage of the fare, we simply divide observed

no matter their proclivity to tip.

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

[23]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.

[24]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.

[25]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.,

TABLE III: Summary Statistics

		Total Charge ($)		Tip ($)		Tipped (Dummy)		Observations	
	Ride Type	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	N	% 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%

tip value by the full fare paid.[26] 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 FDC estimation). These 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. Finally, we also control for the full fare paid by the rider, i.e., the sum of the rounded fare value and any additional charges levied on a rider.[27] 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 III shows summary statistics.

3.2.2 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 (UberX or traditional Lyft) or to authorize sharing (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 another stranger. 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

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).

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

[27] 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 airports). This variable is not rounded.

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 stranger. Figure II illustrates this decision-making process.

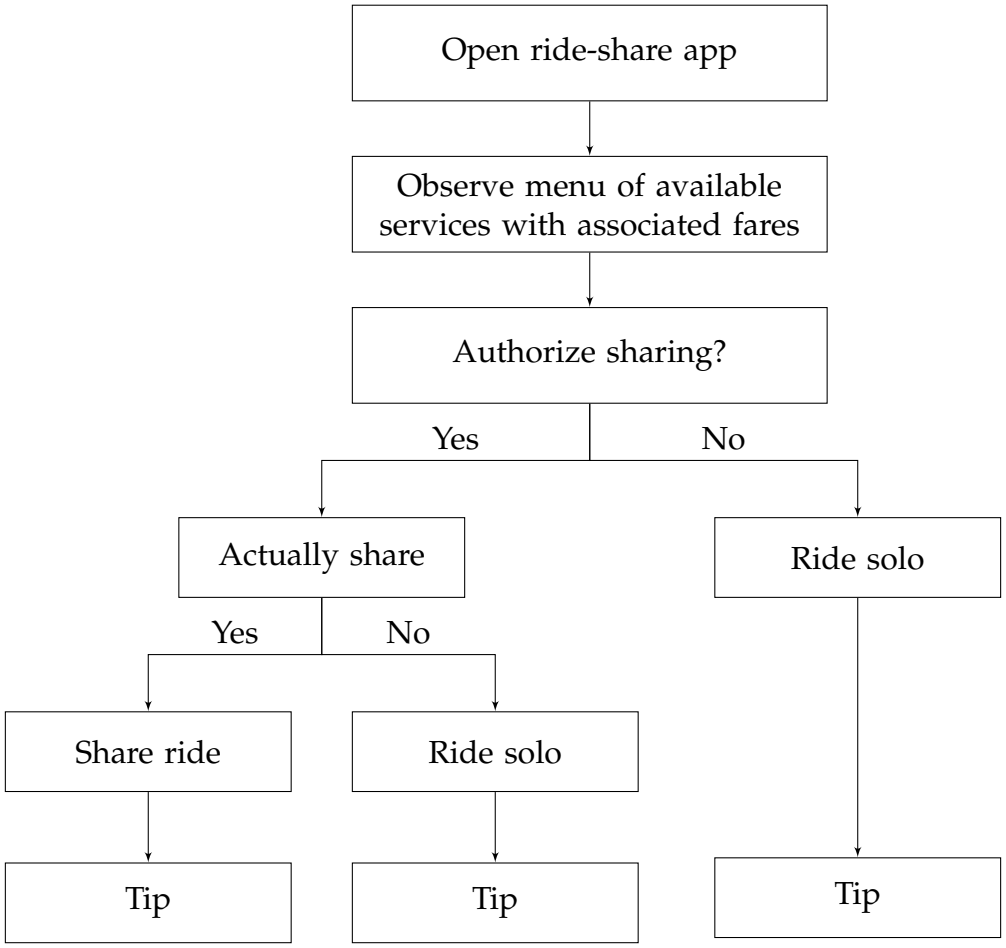


FIGURE II: A flowchart illustrating the timing of a rider's decision to authorize sharing with another passenger on Uber or Lyft.

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 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. A considerable portion of sharing-authorized rides end up solo simply because at a given time and place, there are not other unmatched riders who

have also authorized sharing and are attempting a similar trip.[28] 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 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.[29] 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).

3.2.3 Empirical Setup

We apply the FDC 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 contaminated by endogenous variation.

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

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

[29] 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.

Y. To deal with this source of endogeneity, we control for each ride's full fare in both stages of our FDC estimation. Recall that riders observe their fare options *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 FDC. 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 the Supplemental Appendix, 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 FDC 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 stingy riders.[30] 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 FDC. Indeed, although sharing-authorized rides are endogenous to tipping behavior, whether a passenger actually ends up sharing a ride with another is plausibly 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 plausibly affect Y are accounted for.

Conditional on these variables we plausibly 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).[31] 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).[32]

[30] 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 FDC estimates a much lower ATE of authorizing sharing on the tip share than the naïve specification, suggesting that people who tip a lower share select into sharing-authorized rides.

[31] We include the same exact set of controls in both stages of FDC 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.

[32] 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.

Our estimation strategy consists of estimating the following equations.

$$\text{Naïve: } Y_i = \alpha_2 + \beta_2 X_i + \rho_2 F_i + \boldsymbol{G_i'\theta_2} + \epsilon_{2i} \tag{21}$$

$$\text{FDC First Stage: } M_i = \kappa_2 + \gamma_2 X_i + \tau_2 F_i + \boldsymbol{G_i'\sigma_2} + \omega_{2i} \tag{22}$$

$$\text{FDC Second Stage: } Y_i = \lambda_2 + \delta_2 M_i + \phi_2 X_i + \pi_2 F_i + \boldsymbol{G_i'\iota_2} + \nu_{2i} \tag{23}$$

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 FDC equations by seemingly unrelated regression (Zellner 1962) to account for the potential correlation between equations 22 and 23. To recover the ATE of X on Y estimated by the FDC, we simply multiply the coefficient estimates $\hat{\gamma_2}$ and $\hat{\delta_2}$ by each other. Because the ATE is a nonlinear combination of coefficients, standard errors for the ATE estimated by the FDC 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). 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 our fixed effects allow simulating. Third, $Cov(M, \nu_2) = 0$ given that M is as good as random conditional on X, F, and $\boldsymbol{G}$. Finally, $P(X_i|M_i) > 0$ is satisfied because if $M_i = 1$ then $X_i = 1$ and if $M_i = 0$ then $X_i = 1$ or $X = 0$.

3.2.4 Results

Table IV 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 6.7 percent. The FDC, however, estimates that authorizing sharing reduces tipping probability by only 0.8 percent, an ATE almost a full order of magnitude smaller than the naïve estimate. The difference between the two ATE 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 FDC specification. Notably,

TABLE IV: 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***	-0.0550***
	(0.0001)	(0.0002)	(0.0002)
Shared Trip (M)	–	–	-0.0115***
			(0.0002)
Full Fare (F)	0.0050***	-0.0064***	0.0049***
	(0.00001)	(0.00001)	(0.00003)
Intercept	0.1306***	0.0851***	0.1316***
	(0.0001)	(0.0002)	(0.0005)
Estimated ATE	-0.0628***	-0.0078***	
	(0.0001)	(0.0001)	
Elasticity	-6.764%***	-0.836%***	
	(0.0001)	(0.0001)	
Observations	95,670,449	95,670,449	
R^2	0.1165	0.7297	0.1165

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. FDC specification estimated using seemingly unrelated regression. Robust standard errors in parentheses. Standard errors for the FDC ATE and ATE elasticity computed using the delta method. *** p<0.01, ** p<0.05, * p<0.1.

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 FDC estimation deals with this endogeneity by leveraging the conditionally exogeneous 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 V 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 4.1 percent. The FDC presents a much lower, yet still statistically significant effect. The ATE calculated using the FDC finds that authorizing sharing reduces the tip amount by about 0.4 percent, an effect that is an order magnitude of the naïve ATE estimate.

Finally, in Table VI we also estimate results on the tip as a share of the fare as the de-

TABLE V: 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)
Shared Trip (M)	–	–	-0.0147***
			(0.0003)
Full Fare (F)	0.0148***	-0.0064***	0.0147***
	(0.00003)	(0.00001)	(0.0001)
Intercept	0.1233***	0.0851***	0.1245***
	(0.0004)	(0.0002)	(0.0011)
Estimated ATE	-0.0958***	-0.0099***	
	(0.0002)	(0.0002)	
Elasticity	-4.107%***	-0.426%***	
	(0.0001)	(0.0001)	
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. FDC specification estimated using seemingly unrelated regression. Robust standard errors in parentheses. Standard errors for the FDC ATE and ATE elasticity computed using the delta method. *** p<0.01, ** p<0.05, * p<0.1.

pendent variable with the same right-hand side variables as specified in equations 22 and 23. 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 FDC 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 8.6 percent. The FDC, on the other hand, estimates that authorizing sharing reduces the tip share by about 1.0 percent.[33] 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 IV and V.

This application demonstrates the usefulness of the FDC. By leveraging the conditional exogeneity of actual ride matching once sharing is authorized, we can estimate the ATE of authorizing a shared ride on tipping. As it turns out, the FDC estimates lower ATEs than those estimated by the naïve specification, at both the intensive and extensive margins of tipping as well as for tipping as a percentage of the fare. This result suggests that if a researcher could have conducted a randomized controlled trial wherein passengers are randomly assigned to either a dedicated (i.e., single-passenger) or sharing-authorized ride in Chicago in July 2019, she would have estimated effects of sharing-authorized rides on tipping that are similar to ours.

For illustrative purposes, and to show that the FDC does not require linear regression or even a parametric specification, in the Supplemental Appendix 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 ATE presented in columns 1 and 2 is comparable to that in column 1 of Tables IV and V. In that case, the nonparametrically estimated ATEs in Supplemental Appendix Table XV are roughly double the parametrically estimated ATEs in Tables IV, V, and VI. 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 XV and 0.677 in Tables IV,

[33] In the Supplemental Appendix, 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 FDC setup, but it makes a difference in the second stage because people tip on the basis of the fare actually paid. Similarly, in the Supplemental Appendix, we present estimation results without time-place fixed effects. We find omitting these fixed effects makes almost no difference.

TABLE VI: 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***	-0.0155***
	(0.00003)	(0.0002)	(0.00005)
Shared Trip (M)	–	–	-0.0031***
			(0.0001)
Full Fare (F)	-0.0004***	-0.0064***	0.0004***
	(0.000003)	(0.00002)	(0.00003)
Intercept	0.0473***	-0.0851***	0.0475***
	(0.00005)	(0.0002)	(0.0001)
Estimated ATE	-0.0175***	-0.0021***	
	(0.00003)	(0.00003)	
Elasticity	-8.679%***	-1.024%***	
	(0.0001)	(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. FDC specification estimated using seemingly unrelated regression. Robust standard errors in parentheses. Standard errors for the FDC ATE and ATE elasticity computed using the delta method. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

V, and VI. Finally, the ATEs for the FDC are very close to one another when comparing columns 1, 2, and 3 of Table XV with the FDC ATEs in Tables IV, V, and VI.

The data have a few weaknesses. 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 FDC allows dealing with, as it allows effectively controlling for a consumer's "type." Additionally, 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.[34] This means that in columns 2 and 3 of Table V, we are dealing with two sources of classical measurement error. The first is classical measurement error in fare level,[35] which is a control variable in both columns 2 and 3. We are not worried about this source of measurement error because the coefficient on fare level, which is thus biased toward zero, is not directly of interest in our analysis. 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.

4 Departures from the Ideal Case

Having discussed how to identify and estimate ATEs with the FDC in section 2, and having illustrated the use of the FDC to estimate ATEs using both simulation and real-world data in section 3, we now turn to investigate what happens when some of the assumptions required for the FDC to identify an ATE fail to hold. To do so, we look in

[34] 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.

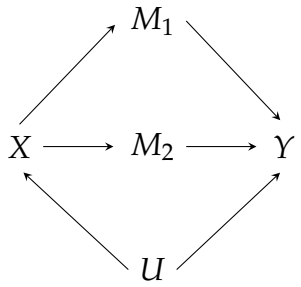
[35] For example, because fares are rounded at the nearest \$2.50 in the data, a reported \$15 ride's true fare could lie anywhere in the $(\$13.75, \$16.25)$ interval.

turn at what happens with multiple mediators, when the mediator is no longer strictly exogenous, and when $P(X_i|M_i) \not> 0$.

4.1 Multiple Mediators

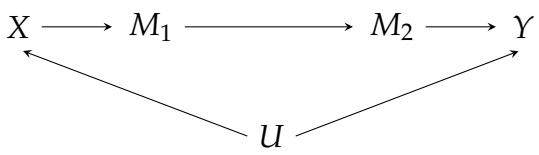
Pearl's (1995, 2000) canonical treatment of the front-door criterion assumes that M is a single variable, and not a vector of mediator variables. Consequently, in the empirical examples in section 3, we considered cases where the mediator, M, is defined by a single variable rather than a vector. In this sub-section we consider how to implement a case where we have multiple mediators.

FIGURE III: 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. For illustrative purposes, however, we will examine these two cases separately. In the first case, as shown in Figure III, 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.[36] By simply examining Figure III 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 IV: Multiple Mediators—Case 2



[36]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).

In the second case, as shown in Figure IV, 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 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 FDC approach will recover the ATE when using either only M_1, only M_2, or both M_1 and M_2 as mediators in the FDC estimation. This point should be obvious based on conceptual reasoning, but a simulation showing this result can be found in the Supplemental Appendix.

We now show simulation results that demonstrate the consequences of multiple mediators of the sort illustrated in Figure III, 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

$$X_i = 0.5U_i + \epsilon_{Xi}, \tag{24}$$

$$M_{1i} = Z_{1i}X_i + \epsilon_{M1i}, \tag{25}$$

$$M_{2i} = Z_{2i}X_i + \epsilon_{M2i}, \tag{26}$$

and

$$Y_i = 0.5M_{1i} + 0.5M_{2i} + 0.5U_i + \epsilon_{Yi}. \tag{27}$$

As illustrated in Figure III, this fully satisfies Pearl's (1995, 2000) three assumptions for the FDC to be able to estimate the average treatment effect of X on Y. By substituting Equations 25 and 26 into Equation 27, it should be obvious to the reader that the true ATE is equal to 0.500 in our simulations.

Similar to the previous simulation analysis, we estimate several specifications. The first baseline specification estimates

$$Y = \alpha_3 + \beta_3 X_i + \zeta_3 U_i + \epsilon_{3i}, \tag{28}$$

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

The second specification estimates

$$M_{1i} = \kappa_3 + \gamma_3 X_i + \omega_{3i}, \tag{29}$$

$$M_{2i} = \pi_3 + \rho_3 X_i + \eta_{3i}, \text{ and} \tag{30}$$

$$Y_i = \lambda_3 + \delta_3 M_{1i} + \tau_3 M_{2i} + \phi_3 X_i + \nu_{3i}, \tag{31}$$

where the unobserved confounder U does not appear anywhere. The small difference in the case of multiple independent mediators is the true ATE is calculated by adding two products together, $E[(\hat{\gamma}_3 \cdot \hat{\delta}_3) + (\hat{\rho}_3 \cdot \hat{\tau}_3)] = \beta$.

Column 1 of Table VII shows our benchmark estimation results for Equation 28. Column 2 shows estimation results for the naïve version of Equation 28 which omits the unobserved confounder U. Columns 3, 4, and 5 show FDC estimation results using the specification outlined in Equations 29 to 31, respectively. Again, the estimates of the ATE in columns 1 and 2 are quite different. While the ATE 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 FDC approach accurately estimates the ATE. The FDC approach first multiplies the coefficient on X in column 3 by the coefficient on M_1 in column 5. Next, the FDC 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 ATE. Assuming the ATE in column 1 is not correlated with the ATE computed from columns 3 through 5, the two ATEs are statistically identical. In both cases, the estimated ATE 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 FDC 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 FDC estimation the estimated ATE (shown in columns 7 and 8) is 0.246, considerably smaller than the true ATE. Column 9 shows that the "direct effect" is 0.254.

The foregoing shows the consequences of omitting a mediator when using the FDC approach to estimate the ATE. With that said, the effect estimated in the biased FDC estimation in columns 7 and 8 of Table VII 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 FDC 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

TABLE VII: 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)
Estimated ATE	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 ATE is estimated by the delta method.

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).[37] 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 FDC 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.

4.2 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 section 3, 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

$$X_i = 0.5U_i + \epsilon_{Xi}, \tag{32}$$

$$M_i = Z_i X_i + \Gamma U_i + \epsilon_{Mi}, \tag{33}$$

and

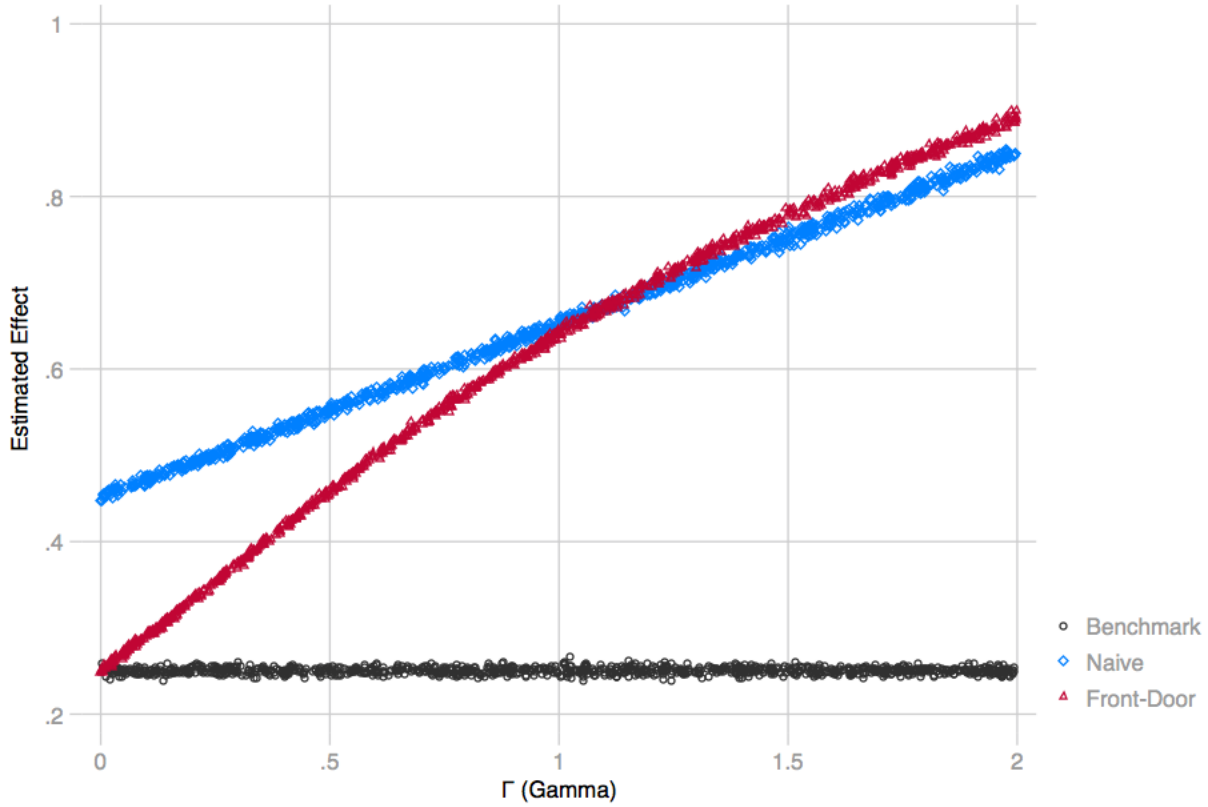
$$Y_i = 0.5M_i + 0.5U_i + \epsilon_{Yi}. \tag{34}$$

The critical difference here is that now, when defining M in equation 33, 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.

We show these simulation results graphically. Figure V illustrates how having an endogenous mediator influences the credibility of using the FDC approach to estimate the ATE. 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 ATE of 0.250. Second, the naïve estimates (blue diamonds), which omits the confounder U from the regression equation,

[37] Also see the discussion of "collider" bias in Morgan and Winship (2015, p.81).

FIGURE V: 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 ATE of 0.250. The naive estimates are shown in blue diamonds and the front-door estimates are shown in red triangles.

consistently overestimates the ATE. 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 13 and 14 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 FDC approach accurately estimates the ATE. Second, although the FDC approach produces biased ATE estimates, when M is weakly endogenous (i.e., when $\Gamma > 0$ but still relatively small), these estimates are less biased than the naïve estimates. Third, when M is strongly endogenous the FDC approach produces estimates of the ATE 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 FDC 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 FDC approach will produce more reliable estimates of the ATE compared to the naïve approach which consists in regressing Y on an endogenous X.[38] On the other hand, when the endogeneity of M is obviously relatively strong, using the FDC approach could lead to more bias in estimates of the ATE than the naïve approach. Specifically in our simulation set-up, the FDC 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.

4.3 Non-Compliance as a Necessary Condition

Recall that, in addition to Assumptions 1 to 3 in section 2.1 for the FDC to identify the average treatment effect, Pearl (2000) imposes a condition on the data, namely that $P(X_i|M_i) > 0$. This condition requires that for every value of the mediator M, the likeli-

[38] Our real-world illustration in the previous section exemplifies a scenario in which a mediator is plausibly exogenous conditional on observed confounders.

hood that an observation will receive treatment X is nonzero. In other words, no matter what value the mediator takes, it has to be the case that an observation has a nonzero probability of receiving the treatment (e.g., $X = 1$).

This requirement is satisfied in our core empirical applications, but it fails to hold in the re-analysis of the Beaman et al. (2013) randomized controlled trial in the Supplemental Appendix.[39] In that application, the authors randomly allocated fertilizer to rice farmers in Mali. One group of farmers received the full recommended dose of fertilizer, a second group received half the recommended dose, and a third group received no fertilizer. We defined the treatment (e.g., $X = 1$) if a farmer received any free fertilizer, and zero otherwise. The mediator captured the intensity of treatment (e.g., $M = 1$ if the farmer received the full dose, $M = 0.5$ if the farmer received the half dose, and $M = 0$ if the farmer received no fertilizer). Therefore, in this case, a farmer who received no fertilizer (e.g., $M = 0$) had a probability of receiving treatment equal to zero (e.g., $X = 0$). Additionally, a farmer who received some fertilizer (e.g., $M = 1$ or $M = 0.5$) had a probability of receiving treatment equal to one (e.g., $X = 1$). Thus, in this application, $P(X_i|M_i) \not> 0$, which is a violation of Pearl's additional condition (Pearl 2000).[40] In this case, because treatment is randomly assigned, one only need to omit the treatment variable X from estimation in Equation 8 for the method we outline in section 2 to recover the correct ATE.[41]

Our work for this paper, however, uncovered the following fact: the condition that $P(X_i|M_i) > 0$ is sufficient but not necessary. The necessary condition implied by Pearl (2000) is more precisely characterized as non-compliance between treatment and the mediator. We demonstrate this detail using simulated data in Table VIII by presenting a case where $P(X_i|M_i) > 0$ does not hold, but where the FDC nevertheless continues to estimate the true ATE because there is exogenous non-compliance between the treatment and the mechanism.

Our simulation set up here differs slightly from those previously discussed. Our goal here is to generate data that meet the three point-identification assumptions discussed in section 2.1, treatment is endogenous, but where $P(X_i|M_i) \not> 0$. We specifically let $U_i \sim$

[39] In our simulations, we impose this requirement in the data generating process. In our real-world tipping application, our data show that rides will not be actually shared unless a passenger first authorizes sharing.

[40] Though it is obvious how experimental settings may naturally lead to cases where $P(X_i|M_i) = 0$, violations of this assumption are not the exclusive preserve of experimental research designs. Indeed, it is not difficult to imagine observational research designs where only those subjects who select into receiving a given treatment can actually receive that treatment in nonzero amounts. Therefore, this discussion remains relevant for observational research settings.

[41] We document our re-analysis using the Beaman et al. (2013) data in the Supplemental Appendix. We also show that when the treatment is exogenous and when $P(X_i|M_i) \not> 0$, implementing the FDC method as discussed in section 2 will lead to biased estimates of the ATE.

TABLE VIII: Simulation Results when $P(X_i|M_i) \not> 0$ with Endogenous Treatment

	Benchmark	Naïve	Front-Door	
Variables	Y	Y	M	Y
	(1)	(2)	(3)	(4)
Treatment (X)	0.743***	1.078***	1.503***	0.322***
	(0.009)	(0.008)	(0.002)	(0.021)
Mediator (M)	–	–	–	0.503***
				(0.013)
Confounder (U)	0.503***	–	–	–
	(0.008)			
Intercept	-0.005	0.173***	0.000	0.173***
	(0.005)	(0.004)	(0.001)	(0.004)
Estimated ATE	0.743***	1.078***	0.755***	
	(0.006)	(0.008)	(0.019)	
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 ATE is estimated by the delta method.

$B(1, 0.5)$, $Z_{Xi} \sim B(1, 0.5)$, $Z_{Mi} \sim B(1, 0.5)$, $\epsilon_{Yi} \sim N(0, 1)$ for a sample size of $N = 100,000$ observations. Then, let

$$X_i = Z_{Xi} U_i \tag{35}$$

$$M_i = Z_{Mi} X_i + X_i \tag{36}$$

$$Y_i = 0.5 M_i + 0.5 U_i + \epsilon_{Yi} \tag{37}$$

This generates data where the binary treatment X is endogenous to the outcome Y. The mediator M takes three values and is strictly a function of the treatment and can be considered akin to treatment intensity. That is, for treated units (i.e., $X = 1$), $M = 1$ or $M = 2$, and for untreated units (i.e., $X = 0$), $M = 0$.

Table VIII shows even when $P(X_i|M_i) \not> 0$, the FDC nevertheless continues to estimate the true ATE. This is because the necessary data requirement of non-compliance persists in this example. In the example using simulated data that we show in Table VIII, exogenous variation in the mediator comes from non-compliance in whether $M = 1$ or $M = 2$, when $X = 1$. Therefore, although the $P(X_i|M_i) > 0$ is a sufficient condition for the FDC, it is not necessary. The true necessary condition is non-compliance between the treatment and the mechanism.

5 Conclusion

We focus 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 FDC in that toolkit.

We focus here first on explaining how to use the front-door criterion in the context of linear regression, which remains the workhorse of applied economics. Second, we present two empirical examples: one using simulated data, and one relying on observational data on Uber and Lyft rides in Chicago between January 1 and December 31, 2019. Our observational example is, to our knowledge, the first application of the front-door criterion to observational data where the necessary assumptions for point-identification plausibly 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 what can be done about it in practice.

Our results lead to the following recommendations for applied work:

1. Because the FDC estimand is a nonlinear combination of two estimated coefficients, standard errors can be computed either by the delta method or by bootstrapping. In small samples, bootstrapping should be preferred to the delta method (Davidson and MacKinnon 2004).

2. When the treatment operates through more than one mediator, the average treatment effect is the sum of the indirect effects, defined by the effect of the treatment on outcome through each mediator (Imai et al. 2010; Acharya et al. 2016).

3. When the mediator is no longer strictly exogenous, the usefulness of the FDC depends on the degree of exogeneity of the mediator. In cases where the mediator is only plausibly—but not strictly—exogenous (Conley et al., 2012), the estimate of the ATE obtained by the FDC is closer to the true value of the ATE than the estimate of the ATE obtained by a naïve regression of outcome on treatment. In cases where the mediator is deemed to be strongly endogenous, the estimate of the ATE obtained by the FDC is further from the true value of the ATE than the estimate of the ATE obtained by a naïve regression of outcome on treatment.

4. 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 meth-

ods that could improve the estimation of causal effects such as the use of non-linear and nonparametric approaches to estimation and machine-learning methods. The suitability of these approaches, of course, depends on the given empirical setting.

5. The FDC 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.

6. Finally, while we have focused on estimation of the ATE, nothing prevents applied economists from estimating other types of treatment effects (e.g., average treatment effects on the treated or on the untreated) when relying on the FDC as a research design, as this can be done using weighted regression methods.

Ultimately, the front-door criterion is a useful tool for applied researchers interested in causal inference with observational data. When selection into treatment is endogenous but there exists a plausibly exogenous mediator whereby the treatment causes the outcome, the front-door criterion can be argued to credibly identify the causal effect of treatment on outcome.

In discussions during early work on this article, one of us came up with the following thought experiment:

> Imagine instrumental variables estimation had never been developed. You meet up with a colleague for lunch one day, and she starts describing a new method for estimating treatment effects, which requires that a given variable only affect your outcome of interest through your treatment variable, which can only really be guaranteed when that variable is randomly assigned. Oh, and the treatment effect thus obtained only affects a subset of your sample, and it can be impossible to tell *what* that subset is! Odds are you would be skeptical, and that you would argue that the method just described is a non-starter. Yet we have learned to live with instrumental variables estimation and all of its imperfections.

Just as the economics profession has learned to live with how difficult it is to find valid instruments and with the limitations of instrumental variables estimation, the economics profession can learn to live with the difficulty inherent in finding a mediator variable that satisfies the front-door criterion's identification assumptions as well as with that method's limitations. Similar to instrumental variables estimation, the front-door criterion is an additional empirical tool that can help us learn about the world in specific settings.

References

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.

Beaman, L., Karlan, D., Thuysbaert, B., and Udry, C. (2013) "Profitability of Fertilizer: Experimental Evidence from Female Rice Farmers in Mali," *American Economic Review*, vol. 103, no. 3, pp. 381-386.

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.

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.

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. (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.

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

Naimi, A.I. (2015) "Invited Commentary: Boundless ScienceâĂŤPutting 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.

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) âĂIJThe true cost of sharing: A detour penalty analysis bet6ween UberPool and UberX trips in Toronto.âĂİ *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.

Supplemental Appendix

A1. Mediators as Instruments

In this Supplemental Appendix we discuss the idea of using the exogenous mediator M, which satisfies the FDC requirements and assumptions, as an instrumental variable. Recall the usual instrumental variable (IV) setup required for the local average treatment effect (LATE) theorem to hold (Imbens and Angrist 1994): The treatment variable X is endogenous to the outcome of interest Y, but the econometrician has access to an instrumental variable Z which (i) is independent, (ii) satisfies the exclusion restriction, and is (ii) relevant. At the outset, it may seem tempting to simply use the mediator M as an instrument instead of using the front-door criterion. We discuss here why this is not advisable for at least two reasons.

First, if the mediator M satisfies the FDC requirements and assumptions, then the outcome Y given the treatment X depends on the value of M. Thus, the exclusion restriction necessary for the identification of the LATE in the instrumental variable estimation approach is violated.

Second, it is not clear how to define the complier sub-sample estimated with the mediator M as an instrument in the LATE framework. Assume, for illustrative purposes, that both the treatment X and mediator M are binary variables. Further assume, as is the case in our ride-sharing empirical application, that we have one-sided noncompliance. So, if $M = 1$ then $X = 1$ and if $M = 0$ then $X = 1$ or $X = 0$. In this case, by construction, there are zero never-takers and zero defiers in the sample. Rather, the entire sample are either compliers or always takers with one-sided non-compliance. In the LATE framework instrumental variable estimates are the average treatment effect on the compilers and the always takers are differenced out of the LATE estimate. The compilers in the FDC set up are those who when $M = 1$ then $X = 1$ and when $M = 0$ then $X = 0$. The identifying variation in the FDC set up comes from one-sided non-compliance; the fact that not everyone with $X = 1$ has $M = 1$. That non-compliant variation is differenced out and not included in the LATE estimate. This discussion demonstrates why the front door criterion allows for the identification of treatment effects in setters where other methods are not applicable.

A2. Ride-Hailing Application Results without Full Fare Control

TABLE IX: Results for Tipping at the Extensive Margin, Omitting Fare as a Control

Variables	Naïve	Front-Door	
	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)
Estimated ATE	-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. FDC specification estimated using seemingly unrelated regression. Robust standard errors in parentheses. Standard errors for the FDC ATE and ATE elasticity computed using the delta method. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

TABLE X: Results for Tipping at the Intensive Margin, Omitting Fare as a Control

Variables	Naïve	Front-Door	
	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)
Estimated ATE	-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. FDC specification estimated using seemingly unrelated regression. Robust standard errors in parentheses. Standard errors for the FDC ATE and ATE elasticity computed using the delta method. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

TABLE XI: Results for Tip as a Fraction of Fare, Omitting Fare as a Control

Variables	Naïve	Front-Door	
	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)
Estimated ATE	-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. FDC specification estimated using seemingly unrelated regression. Robust standard errors in parentheses. Standard errors for the FDC ATE and ATE elasticity computed using the delta method. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

A3. Ride-Hailing Application without Time-Place FEs

TABLE XII: 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)
Estimated ATE	-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. FDC specification estimated using seemingly unrelated regression. Robust standard errors in parentheses. Standard errors for the FDC ATE and ATE elasticity computed using the delta method. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

TABLE XIII: Results for Tipping at the Intensive Margin, Omitting Fixed Effects

Variables	Naïve	Front-Door	
	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)
Estimated ATE	-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. FDC specification estimated using seemingly unrelated regression. Robust standard errors in parentheses. Standard errors for the FDC ATE and ATE elasticity computed using the delta method. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

TABLE XIV: Results for Tip as a Percentage of Fare, Omitting Fixed Effects

Variables	Naïve	Front-Door	
	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)
Estimated ATE	-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. FDC specification estimated using seemingly unrelated regression. Robust standard errors in parentheses. Standard errors for the FDC ATE and ATE elasticity computed using the delta method. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

A4. 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 XV: 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 ATE (i.e., $\hat{\beta}_{\text{Naïve}}$)	-0.1167	-0.2459	-0.2459
FDC ATE (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.

A5. Multiple Mediators–Case 2

We now show the results of a simulation that demonstrate the consequences (or lack thereof) of multiple mediators of the sort illustrated in Figure I, 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

$$X_i = 0.5U_i + \epsilon_{Xi}, \tag{38}$$

$$M_{1i} = Z_{1i}X_i + \epsilon_{M1i}, \tag{39}$$

$$M_{2i} = Z_{2i}M_{1i} + \epsilon_{M2i}, \tag{40}$$

and

$$Y_i = 0.5M_{2i} + 0.5U_i + \epsilon_{Yi}. \tag{41}$$

As illustrated in Figure I, this fully satisfies Pearl's (1995, 2000) three criteria for the FDC to be able to estimate the average treatment effect of X on Y. By substituting equation 39 into equation 40 and substituting equation 40 into Equation 41, it should be immediately obvious to the reader that the true ATE is equal to 0.125 in our simulations.

Similar to the previous simulation analysis, we estimate several specifications. The first specification estimates the true ATE by controlling for the confounder U. The second specification estimates the ATE using the FDC approach. As the results in Table VII show, estimates of the ATE with the FDC approach in this case are statistically invariant whether either or both M_1 and M_2 are included in the estimation procedure.

TABLE XVI: Simulation Results—Multiple Mediators, Case 2

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)	0.127***	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)
Estimated ATE	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 ATE is estimated by the delta method.

A6. Real-World Application: Experimental Replication

This section illustrates the FDC using the results of an experimental study by Beaman et al. (2013). In Table XVII, we replicate results from Beaman et al. (2013), who conduct a randomized controlled trial with rice farmers in Mali. Starting from the full sample, units of observations are either assigned to a treatment group or a control group, with treated units receiving fertilizer and control units receiving no fertilizer.

In this application, we exploit as a mediator the fact that treatment intensity varies at random within the treatment group to illustrate the FDC in practice. About half of the treatment-group observations receive half of the prescribed amount of fertilizer, the remainder of the treatment-group observations receiving the full prescribed amount of fertilizer, and the control-group observations receiving none of the prescribed amount of fertilizer.

As one would expect from the derivations in Section 2, the results in Table XVII show that the ATEs obtained by the FDC are all statistically indistinguishable from the benchmark ATEs. For example, considering the average rate of fertilizer use among the control group is 0.32, the benchmark estimate (in column 1) suggests that receiving free fertilizer increases the use of fertilizer over twofold. The FDC approach roughly replicates (in column 2) this benchmark estimate. The similarity between the benchmark and FDC estimates persist for the the quantity of fertilizer use (columns 3 and 4) and fertilizer expenses (columns 5 and 6). The results show that receipt of free fertilizer leads to increases in the use of fertilizer at both the extensive and intensive margins and reduces fertilizer expenses.

TABLE XVII: Empirical Illustration — Rice Production and Fertilizer Use in Mali

	Use of Fertilizer		Fertilizer Quantity		Fertilizer Expenses	
	(1)	(2)	(3)	(4)	(5)	(6)
Benchmark	0.639***		27.24***		-2,717.1***	
	(0.033)		(3.568)		(464.6)	
Front-Door		0.603***		26.64***		-2,605.3***
		(0.030)		(3.002)		(389.7)
Observations	378	378	378	378	373	373

Notes: Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All columns include the same control variables as in Beaman et al. (2013). Columns (1), (3), and (5) represent benchmark OLS estimates of the ATEs of receiving fertilizer on the outcome variables. The estimates in columns (1), (3), and (5) replicate the findings of Beaman et al. (2013) except that the original study differentiates between two treatment groups defined by intensity of treatment. Columns (2), (4), and (6) represent seemingly unrelated regression estimates of the front-door criterion ATEs. Standard errors in columns (2), (4), and (6) are estimated by the delta method.

A few remarks are in order. First, the real-world results in this section are most useful for highlighting the potential of the FDC approach in estimating treatment effects in settings where each of the conditions hold. Of course, since Beaman et al. (2013) assign treatment experimentally, the FDC approach is not necessary to estimate treatment effects in that context.

Second, in that real-world experimental case, there is no need to condition on the treatment variable (i.e., X) when estimating the effect of the mediator (i.e., M) on the outcome (i.e., Y) since the random assignment of treatment already removes any back-door path between Y and M. In fact, needlessly conditioning on the treatment variable in an experimental setting leads to bias in the front-door estimate.

A7. Results when $P(X_i|M_i) \not> 0$ with Exogenous Treatment

In section 4.3 we discussed Pearl's additional condition, or data requirement, for the FDC method: that $P(X_i|M_i) > 0$. Through the use of simulated data, we demonstrated that this condition is sufficient, but not necessary. In this section, we show results using the experimental data of Beaman et al. (2013) to further highlight this detail.

The results in Table XVIII revisit the data in Table XVII. Here, however, odd-numbered columns report the correct ATEs, and even-numbered columns report biased ATEs. These results show that when treatment is exogenous, it is not necessary to include treatment into the second-stage FDC regression because there are no back-door paths from U to X. In fact, conditioning on treatment could lead to biased estimates when using the FDC method as we see in the even columns in Table XVIII.

TABLE XVIII: Over-Controlling for the Treatment — Fertilizer Use in Mali

	Use of Fertilizer		Fertilizer Quantity		Fertilizer Expenses	
	(1)	(2)	(3)	(4)	(5)	(6)
Benchmark Front-Door ATE	0.603***		26.64***		-2,605.3***	
	(0.030)		(3.002)		(389.7)	
Over-Controlled Front-Door ATE		0.009		17.580***		-882.72
		(0.055)		(5.920)		(776.90)
Observations	378	378	378	378	377	377

Notes: Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All columns include the same control variables as in Beaman et al. (2013). Columns (1), (3), and (5) represent benchmark seemingly unrelated regression FDC estimates of the ATEs of receiving fertilizer on the outcome variables. Columns (2), (4), and (6) represent seemingly unrelated regression estimates of the front-door criterion ATEs which over-control for treatment in the outcome regression of the FDC setup. Standard errors are estimated by the delta method.