The Paper of How

Estimating Treatment Effects Using the Front-Door Criterion

Marc F. Bellemare, Jeffrey R. Bloem, and Noah Wexler

January 28, 2021

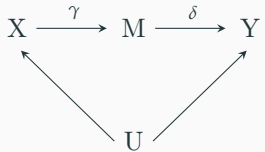
Ateliers méthodologiques de Montréal—Montréal Methods Workshop

Introduction

We present the first application of Judea Pearl's (1995, 2000) front-door criterion (FDC) to observational data in which the required assumptions for point-identification plausibly hold.

The directed acyclic graph (DAG) in Figure 1 illustrates the FDC setup.

Figure 1: 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.

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.

Despite its relative simplicity, economists have been reluctant to incorporate the front-door criterion in their empirical toolkit.

Anecdotally, that resistance appears to stem from the fact that finding a convincing empirical application has thus far proven elusive (see, e.g., Imbens 2020; Gupta et al. 2020).

We provide such an application and further 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).

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 credible identification with the FDC approach do not hold.

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

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.

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. We find that the observed negative correlation between choosing to share a ride and tipping is almost entirely explained by selection into treatment—a finding relevant to the economics of tipping (Azar 2020).

In our application, we are not able to randomly assign whether someone authorizes shared rides (i.e., X) on Uber or Lyft.

Additionally, since the base fare of shared rides is typically less than that of solo rides, this choice is clearly endogenous to tipping behavior (i.e., Y). Once a passenger chooses to share a ride, however, they will not necessarily share a ride.

We can therefore exploit the exogenous variation—conditional on fare level and date, hour, day of the week–hour, and origin–destination fixed effects—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).

Third, and most importantly for a crowd of applied researchers such as this one, 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 estimation,

(ii) the assumption of strict exogeneity of M is violated, and

(iii) the treatment is completely defined by the mediator.

Outline

- The Front-Door Criterion: Identification and Estimation
  - Identification
  - Estimation
- Empirical Illustrations
  - Simulations
  - Application: Uber and Lyft Rides in Chicago
- Departures from the Ideal Case
  - Multiple Mechanisms
  - Violations of Strict Exogeneity
  - Treatment Totally Defined by Mechanism
- Conclusion

Identification and Estimation

Identification

We are interested in estimating the ATE of X on Y in Figure 1 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 1 allows for an unbiased point estimate of the ATE of X on Y.

As discussed in Pearl (1995, 2000), the FDC requires that there exists a variable M which satisfies the following assumptions relative to X and Y:

Assumption 1: The only way in which X influences Y is through M.

Assumption 2: The relationship between X and M is not confounded by unobserved variables.

Assumption 3: Conditional on X, the relationship between M and Y is not confounded by unobserved variables.

As in Pearl (1995), in the paper we derive the FDC estimand in three steps, aiming to compute $P(Y|\check{X})$ with observable variables, where $P(Y|\check{X})$ represents the ATE of X on Y.

As shown in Figure 1, 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 exploiting Assumptions 1 through 3.

After some algebra, the FDC estimand is such that

$$\mathrm{P}(\mathrm{Y}|\check{\mathrm{X}}) = \sum_{\mathrm{M}} \mathrm{P}(\mathrm{M}|\mathrm{X}) \times \sum_{\mathrm{X}'} \mathrm{P}(\mathrm{Y}|\mathrm{X}', \mathrm{M}) \times \mathrm{P}(\mathrm{X}'). \tag{3.1}$$

In later writings on the FDC, Pearl (2000) discusses an additional condition for identification, a data requirement which can be directly be 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 has to have a nonzero probability of getting treated, and thus the mediator M cannot be entirely defined by the treatment X.

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 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).

Estimation

Our goal is to estimate the ATE of X on Y in Figure 1, 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 is by using the following approach.

Let

$$M_i = \kappa + \gamma X_i + \omega_i \tag{3.2}$$

and

$$Y_i = \lambda + \delta M_i + \phi X_i + \nu_i. \tag{3.3}$$

In the paper, we explain how estimating Equations 3.2 and 3.3 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, but the FDC is nonparametrically identifiable, and linear regression is but one way to estimate treatment effects using the FDC.

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

Empirical Illustration

Simulations

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.

Then, let

$$X_i = 0.5U_i + \epsilon_{Xi}, \tag{4.1}$$

$$M_i = Z_i X_i + \epsilon_{Mi}, \tag{4.2}$$

and

$$Y_i = 0.5M_i + 0.5U_i + \epsilon_{Yi}. \tag{4.3}$$

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 4.2 into Equation 4.3, it should be obvious 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 two 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

$$\mathrm{Y_i} = \alpha_0 + \beta_0 \mathrm{X_i} + \zeta_0 \mathrm{U_i} + \epsilon_{0\mathrm{i}}, \tag{4.4}$$

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

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

$$\mathrm{M_i} = \kappa_0 + \gamma_0 \mathrm{X_i} + \omega_{0\mathrm{i}} \tag{4.5}$$

$$\mathrm{Y_i} = \lambda_0 + \delta_0 \mathrm{M_i} + \phi_0 \mathrm{X_i} + \nu_{0\mathrm{i}} \tag{4.6}$$

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

Lastly, we estimate a naïve specification, one that is similar to the benchmark specification in Equation 4.4, but which fails to control for the presence of the unobserved confounder.

Table 1: Simulation Results

	Benchmark	Naïve	Front-Door		Direct Effect
Variables	Y	Y	M	Y	Y
	(1)	(2)	(3)	(4)	(5)
Treatment (X)	0.252***	0.454***	0.507***	0.200***	-0.003
	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)
Mechanism (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.

Application: Ride Sharing and Tipping Behavior

Using publicly available data on over 890,000 Uber and Lyft rides in Chicago between June 30 and September 30, 2019, we use 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).

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

We show that the necessary conditions for the FDC to yield a consistent ATE apply in this scenario after conditioning on relevant observed variables.

Although sharing-authorized rides are not determined exogenously, whether a passenger actually ends up sharing a ride with another is plausibly exogenous conditional on several observable factors.

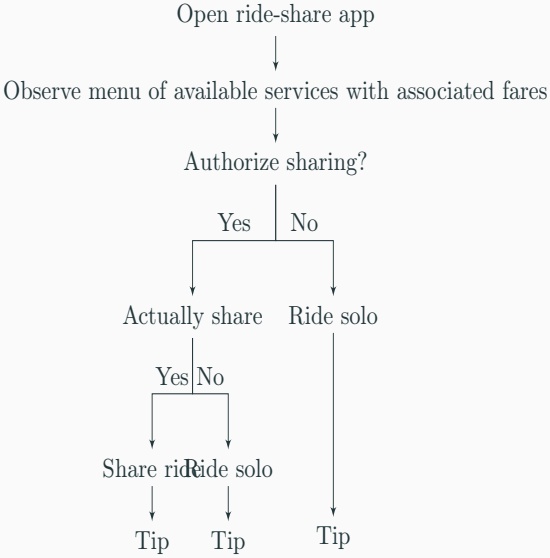


Figure 2: A flowchart illustrating the timing of a rider's decision to authorize sharing with another passenger on Uber or Lyft.

Our data include 890,000 dedicated (i.e. standard, "single-transaction" UberX and Lyft rides) and sharing-authorized Uber and Lyft rides taken within the city limits of Chicago from June 30 to September 30, 2019.

The data come from the Chicago Department of Business Affairs and Consumer Protection's Transportation Network Providers Data Portal.

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).

These data provide the base fare (rounded to the nearest $2.50) and tip amount (rounded to the nearest $1.00).

For the extensive margin of tipping, our dependent variable is a binary variable capturing whether a passenger tips.

For the intensive margin, we use the inverse hyperbolic sine of the observed tip value (Bellemare and Wichman 2020, Card and DellaVigna 2020).

Additionally, we generate several sets of fixed effects from observed time and geographic indicators, including for each origin–destination pair of community areas in Chicago. Summary statistics are provided in Table 2.

Table 2: Summary Statistics

		Fare ($)		Tip ($)		Tipped (Dummy)		Observations	
	Ride Type	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	N	% Total
Full Sample	Dedicated	10.842	(6.884)	0.659	(1.602)	0.215	(0.411)	750,883	84.2%
	Sharing Authorized	9.055	(5.401)	0.208	(0.815)	0.092	(0.289)	140,446	15.8%
Sharing Authorized	Shared	9.332	(5.628)	0.207	(0.791)	0.092	(0.289)	88,372	62.9%
	Not Shared	8.583	(4.957)	0.209	(0.853)	0.091	(0.288)	52,074	37.1%

Though our M variable (i.e., whether a passenger who authorized a shared ride actually get to share a ride) is not strictly exogenous to either our X (i.e., whether a passenger authorizes a shared ride) and Y (i.e., tipping behavior) variables, we argue that it is conditionally exogenous to both those variables.

Indeed, we first condition on a ride's fare level, exploiting app algorithms that set fares according to the likelihood a ride is ultimately shared.

This helps control for the propensity that any given ride is shared.

Although this variable does help control for endogenous factors embedded in the app's algorithms, the rounding of these data to the nearest $2.50 makes conditioning on fare level less precise.

Second, to further condition away potential endogeneity between the likelihood a (sharing-authorized) ride is actually shared on the one hand and tipping on the other hand, we control for date, hour, day of the week–hour, and origin–destination fixed effects.

Between controlling for fare level and those several layers of fixed effects, we plausibly uphold the assumption of ignorability (Rosenbaum and Rubin 1983). In other words, although M may not be strictly exogenous to X and Y, our controls ensure that M is plausibly exogenous to X and Y in this application.

Our estimation strategy consists of estimating the following equations.

$$\text{Naïve: } \mathrm{Y_i} = \alpha_2 + \beta_2 \mathrm{X_i} + \rho_2 \mathrm{F_i} + \mathrm{T_i'}\boldsymbol{\chi}_2 + \mathrm{G_i'}\boldsymbol{\theta}_2 + \epsilon_{2\mathrm{i}} \tag{4.7}$$

$$\text{FDC First Stage: } \mathrm{M_i} = \kappa_2 + \gamma_2 \mathrm{X_i} + \tau_2 \mathrm{F_i} + \mathrm{T_i'}\boldsymbol{\xi}_2 + \mathrm{G_i'}\boldsymbol{\sigma}_2 + \omega_{2\mathrm{i}} \tag{4.8}$$

$$\text{FDC Second Stage: } \mathrm{Y_i} = \lambda_2 + \delta_2 \mathrm{M_i} + \boldsymbol{\phi}_2 \mathrm{X_i} + \pi_2 \mathrm{F_i} + \mathrm{T_i'}\boldsymbol{\psi}_2 + \mathrm{G_i'}\boldsymbol{\iota}_2 \mathrm{G_i} + \nu_{2\mathrm{i}} \tag{4.9}$$

where Y now represents tipping at either the extensive or intensive margin, $\mathrm{F_i}$ is the fare level, $\mathrm{T_i}$ is a vector of time fixed effects (i.e., date, hour, and day of the week–hour fixed effects), and $\mathrm{G_i}$ is a vector of geographic fixed effects (i.e., origin-destination pairs). Additionally, $\mathrm{X_i}$ is our treatment variable, which indicates whether a passenger authorized ride-sharing, and $\mathrm{M_i}$ 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 the two equations 4.8 and 4.9.

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.

Table 3: The Effect of Authorizing Sharing on Tipping at the Extensive Margin

Variables	Naïve		Front-Door
	Tipped	Shared Trip	Tipped
	(1)	(2)	(3)
Sharing Authorized (X)	-0.073***	0.621***	-0.067***
	(0.001)	(0.001)	(0.001)
Shared Trip (M)	–	–	-0.009***
			(0.002)
Intercept	0.114***	0.0882***	0.115***
	(0.006)	(0.003)	(0.006)
Estimated ATE	-0.073***		-0.006***
	(0.001)		(0.001)
Elasticity	-5.8%***		-0.5%***
	(0.001)		(0.001)
Observations	891,329		891,329
R-squared	0.051	0.614	0.051

Notes: Both specifications control for fare level, date, hour, day of the week–hour, and origin–destination fixed effects. FDC specification estimated using seemingly unrelated regression. Standard errors for the FDC ATE and ATE elasticity computed using the delta method. *** p<0.01, ** p<0.05, * p<0.1.

Table 4: The Effect of Authorizing Sharing on Tipping at the Intensive Margin

Variables	Naïve		Front-Door
	arcsinh(Tip)	Shared Trip	arcsinh(Tip)
	(1)	(2)	(3)
Sharing Authorized (X)	-0.127***	0.621***	-0.119***
	(0.002)	(0.001)	(0.002)
Shared Trip (M)	–	–	-0.013***
			(0.003)
Intercept	0.113***	0.0882***	0.114***
	(0.010)	(0.003)	(0.010)
Estimated ATE	-0.127***		-0.008***
	(0.002)		(0.001)
Elasticity	-3.6%***		-0.2%***
	(0.001)		(0.001)
Observations	891,329		891,329
R-squared	0.084	0.614	0.084

Notes: Both specifications control for fare level, date, hour, day of the week–hour, and origin–destination fixed effects. FDC specification estimated using seemingly unrelated regression. Standard errors for the FDC ATE and ATE elasticity computed using the delta method. *** p<0.01, ** p<0.05, * p<0.1.

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.

Additionally, we do not observe the exact tip or fare payments, observing rounded values instead. This means that in columns 2 and 3 of Table 4, we are dealing with two sources of classical measurement error.

The first is classical measurement error in fare level, which is a control variable in both columns 2 and 3.

We are not worried about this source of measurement error because it merely biases the coefficient on fare level—a control variable whose coefficient is not directly of interest in our analysis—toward zero.

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 890,000 observations—we indeed reject the null hypothesis that the coefficient on M in column 3 is equal to zero.

Departures from the Ideal Case

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 turn at what happens with multiple mechanisms, when the mechanism is no longer strictly exogenous, and when the treatment is totally defined by the mechanism.

Multiple Mechanisms

Pearl's (1995, 2000) canonical treatment of the front-door criterion assumes that M is a single variable, and not a vector of mechanism variables.

Consequently, in our empirical examples, we considered cases where the mechanism M is defined by a single variable rather than a vector.

Here, we consider how to implement a case where we have multiple mechanisms.

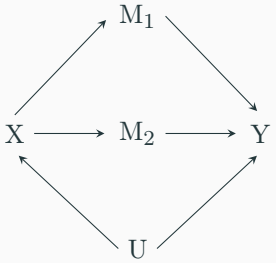
There are two basic cases in which we can imagine multiple mechanisms. Of course, one can imagine more complicated cases that combine these two basic cases.

For illustrative purposes, however, we will examine these two cases separately.

In the first case, the multiple mechanisms 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.

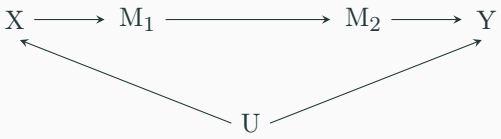
Figure 3: Multiple Mechanisms—Case 1



In this case, M_1 and M_2 together intercept all directed paths from X to Y and meet the requirement of condition (i). By simply examining Figure 3 it is clear that omitting either M_1 or M_2 from the estimation will violate condition (i), since the single mechanism does not intercept all directed paths from X to Y.

In the second case, the multiple mechanisms 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.

Figure 4: Multiple Mechanisms—Case 2



In this case, either M_1 or M_2 intercept all directed paths from X to Y and meet the requirement of condition (i).

In contrast to the previous case, omitting either M_1 or M_2 from the estimation will not violate condition (i), since both mechanisms 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 mechanisms in the FDC estimation.

We now show simulation results that demonstrate the consequences of multiple mechanisms, where multiple mechanisms lie on different paths from X to Y.

Our simulation setup is as follows. Let $\mathrm{U_i} \sim \mathrm{N}(0,1)$, $\epsilon_{\mathrm{Xi}} \sim \mathrm{N}(0,1)$, $\mathrm{Z_{1i}} \sim \mathrm{U}(0,1)$, $\mathrm{Z_{2i}} \sim \mathrm{U}(0,1)$, $\epsilon_{\mathrm{M1i}} \sim \mathrm{N}(0,1)$, $\epsilon_{\mathrm{M2i}} \sim \mathrm{N}(0,1)$, and $\epsilon_{\mathrm{Yi}} \sim \mathrm{N}(0,1)$ for a sample size of $\mathrm{N} = 100,000$ observations. Then, let

$$\mathrm{X_i} = 0.5\mathrm{U_i} + \epsilon_{\mathrm{Xi}}, \tag{5.1}$$

$$\mathrm{M_{1i}} = \mathrm{Z_{1i}X_i} + \epsilon_{\mathrm{M1i}}, \tag{5.2}$$

$$\mathrm{M_{2i}} = \mathrm{Z_{2i}X_i} + \epsilon_{\mathrm{M2i}}, \tag{5.3}$$

and

$$\mathrm{Y_i} = 0.5\mathrm{M_{1i}} + 0.5\mathrm{M_{2i}} + 0.5\mathrm{U_i} + \epsilon_{\mathrm{Yi}}. \tag{5.4}$$

This 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 Equations 5.2 and 5.3 into Equation 5.4, it should be obvious that the true ATE is equal to 0.500 in our simulations.

Similar to the previous simulation analysis, we estimate two specifications.

The first specification estimates

$$\mathrm{Y} = \alpha_1 + \beta_1 \mathrm{X_i} + \zeta_1 \mathrm{U_i} + \epsilon_{1\mathrm{i}}, \tag{5.5}$$

where, because both X and U are included on the right-hand-side, $\mathrm{E}(\hat{\beta}_1) = \beta$, i.e., the true ATE.

The second specification estimates

$$\mathrm{M}_{1\mathrm{i}} = \kappa_1 + \gamma_1 \mathrm{X}_\mathrm{i} + \omega_{1\mathrm{i}} \tag{5.6}$$

$$\mathrm{M}_{2\mathrm{i}} = \pi_1 + \rho_1 \mathrm{X}_\mathrm{i} + \xi_{1\mathrm{i}} \tag{5.7}$$

$$\mathrm{Y}_\mathrm{i} = \lambda_1 + \delta_1 \mathrm{M}_{1\mathrm{i}} + \tau_1 \mathrm{M}_{2\mathrm{i}} + \phi_1 \mathrm{X}_\mathrm{i} + \nu_{1\mathrm{i}} \tag{5.8}$$

where the unobserved confounder U does not appear anywhere.

The small difference in the case of multiple independent mechanisms is the true ATE is calculated by adding two products together, $\mathrm{E}[(\hat{\gamma}_1 \cdot \hat{\delta}_1) + (\hat{\rho}_1 \cdot \hat{\tau}_1)] = \beta$.

Table 5: Simulation Results—Multiple Mechanisms, 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)
Mechanism (M_1)	–	–	–	–	0.498***	0.500***	–	0.495***	0.496***
					(0.003)	(0.003)		(0.004)	(0.003)
Mechanism (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.

Violations of Strict Exogeneity

Together, conditions (ii) and (iii) imply that the mechanism 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, 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

$$X_i = 0.5U_i + \epsilon_{Xi}, \tag{5.9}$$

$$M_i = Z_i X_i + \Gamma U_i + \epsilon_{Mi}, \tag{5.10}$$

and

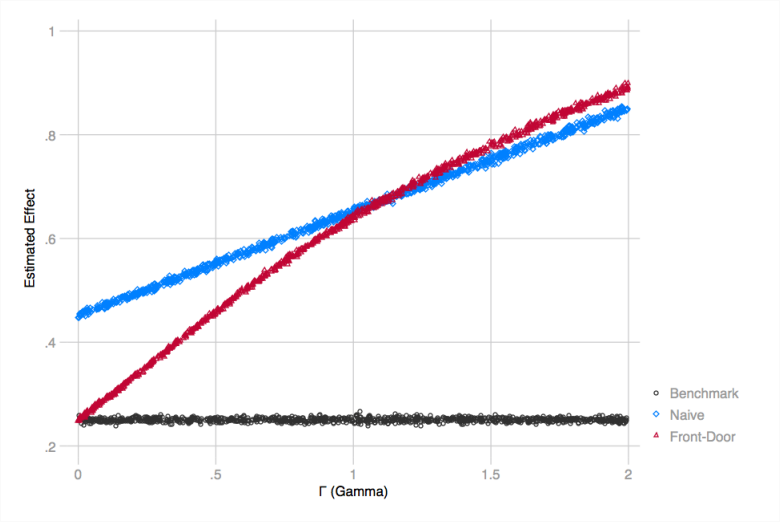
$$Y_i = 0.5M_i + 0.5U_i + \epsilon_{Yi}. \tag{5.11}$$

The critical difference here is that now, when defining M in equation 5.10, 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 $\Gamma \sim \mathrm{U}(0, 2)$. By permitting the values of Γ to vary allows the degree of endogeneity in our simulations to vary.

Figure 5: The Consequences of an Endogenous Mechanism



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, when M is weakly endogenous (i.e., when $\Gamma > 0$ but still relatively small) the FDC approach produces biased estimates of the ATE, but 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.

In many cases, strict exogeneity of M may be debatable.

Indeed, outside of an experimental setting, convincingly arguing that $\mathrm{P}(\mathrm{U}|\mathrm{M},\mathrm{X}) = \mathrm{P}(\mathrm{U}|\mathrm{X})$ and $\mathrm{P}(\mathrm{Y}|\mathrm{X},\mathrm{M},\mathrm{U}) = \mathrm{P}(\mathrm{Y}|\mathrm{M},\mathrm{U})$ 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.

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.

Of course when using real-world data, when we cannot observe U, testing the specific size of these relationships is impossible.

In nearly all practical settings, the case for the exogeneity of M will rely on careful reasoning based on the given empirical setting.

Treatment Totally Defined by Mechanism

Recall that in addition to the three assumptions in section 2.1 for the FDC to identify the average treatment effect, Pearl (2000) makes a fourth assumption, namely that $P(X_i|M_i) > 0$.

This assumption implies that for every value of the mechanism M, the likelihood that an observation will receive treatment X is nonzero.

In other words, the treatment cannot be totally defined by the mechanism.

We explore this in the appendix, but preliminary empirical work for this paper uncovered the following fact: It is only when there are no unobserved confounders that $P(X_i|M_i) = 0$ is a problem.

In such cases, one only need to omit the treatment variable X from estimation to recover the correct ATE.

When there are unobserved confounders, the method applies lock, stock, and barrel.

Why does the treatment need to be omitted from Equation 3.3 when the assumption that $P(X_i|M_i) > 0$ is violated and there are no unobserved confounders?

In such cases, the variation in X is already accounted for in the variation in M.

Indeed, when $M_i > 0$, we know $X_i = 1$, and when $M_i = 0$, we know $X_i = 0$.

Conclusion

We have focused on the application of Pearl's (1995, 2000) front-door criterion.

Because the goal of most research in applied economics nowadays 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 June 30 and September 30, 2019.

Our observational example is, to our knowledge, the first application of the front-door criterion to observational data where the necessary assumptions 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:

- 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).

- When the treatment operates through more than one mediator, the average treatment effect is the sum of the mediated average treatment effects (MATEs), defined by the effect of the treatment on outcome through each mediator. A MATE is akin to (and a special version of) an “indirect effect" in the causal mediation analysis literature (Imai et al. 2010; Acharya et al. 2016).

- 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.

- 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.

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 single, 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.