Supplement: Front-Door and Back-Door Adjustment for ATE

Adam N. Glynn* Konstantin Kashin†

October 18, 2017

1 Front-door and Back-door Adjustment

For an outcome Y and a treatment/action A, we define the potential outcome under a generic treatment as $Y(a_1)$ and the potential outcome under control as $Y(a_0)$. The ATE is defined as $E[Y(a_1)] - E[Y(a_0)]$. In what follows we discuss the asymptotic bias in estimating $E[Y(a_0)]$ and the ATE.

1.1 Asymptotic Bias in Estimating $E[Y(a_0)]$

As in VanderWeele and Arah (2011), we assume that $E[Y(a_0)]$ can be equated to the expectations over observed outcomes by conditioning on observed covariates X and unobserved covariates U. For simplicity in presentation we assume that X and U are discrete, such that

$$\mu_0 = E[Y(a_0)] = \sum_x \sum_u E[Y|a_0, x, u] \cdot P(u|x) \cdot P(x), \tag{1}$$

*Department of Political Science, Emory University, 1555 Dickey Drive, Atlanta, GA 30322. aglynn@emory.edu

†Data Scientist, Institute for Quantitative Social Science, Harvard University, 1737 Cambridge Street, Cambridge, MA 02138. kkashin@fas.harvard.edu

but continuous variables can be easily accommodated.[1] We also assume that probabilistic assignment holds such that there is a positive probability of both a_1 and a_0 for all values of U and X among all units (Rubin, 2010).

If we have measured a set of post-treatment variables M, the front-door adjustment can be written as the following:

$$\mu_0^{fd} = \sum_x \sum_m P(m|a_0, x) \sum_a E[Y|a, m, x] \cdot P(a|x) \cdot P(x), \tag{2}$$

where these sums are taken over values of x, m, and a with positive probability. The asymptotic bias for $E[Y(a_0)]$ can be written as the following (see Appendix S.A.1 for a proof):

$$\begin{aligned} B_0^{fd} &= \mu_0^{fd} - \mu_0 \\ &= \sum_x P(x) \sum_m \sum_u P(m|a_0, x) \sum_a E[Y|a, m, x, u] \cdot P(u|a, m, x) \cdot P(a|x) \\ &- \sum_x P(x) \sum_m \sum_u P(m|a_0, x, u) \cdot E[Y|a_0, m, x, u] \cdot P(u|x) \end{aligned} \tag{3}$$

In this supplement as in the paper, we will use the term bias to refer to asymptotic bias. The bias is zero when the following two conditions hold:

Assumption 1 (Y is mean independent of A conditional on M, X, and U)**.** *$E[Y|a, m, x, u] = E[Y|a_0, m, x, u]$ for all a, m, x, u.*

Assumption 2 (U is independent of M conditional on A and X)**.** *$P(m|a_0, x) = P(m|a_0, x, u)$ for all m, x, u and $\sum_a P(u|a, m, x) \cdot P(a|x) = P(u|x)$ for all m, x, u.*

The result for μ_1 is analogous. Therefore, as demonstrated in Pearl (1995), it is possible for the front-door approach to provide an unbiased estimator of ATE, even when there is

[1] We also note that for formulas of this type throughout the paper, when any of the densities take the value zero (e.g., $P(u|x, a_1) = 0$ or $P(x|a_1) = 0$) we mean the entire term to be zero.

an unmeasured common cause of A and Y. However, note that unlike the presentation in Pearl (1995, 2000, 2009), the presentation here does not require the definition of potential outcomes beyond those originally used to define the ATE.

1.2 Asymptotic Bias for ATE

The front-door formula ATE can be written as:

$$\mu_1^{fd} - \mu_0^{fd} = \sum_x P(x) \sum_m [P(m|a_1, x) - P(m|a_0, x)] \sum_a E[Y|a, m, x] \cdot P(a|x), \tag{4}$$

with the bias written as the following (see proof in Appendix S.A.1):

$$\begin{aligned} B_{ATE}^{fd} &= \mu_1^{fd} - \mu_0^{fd} - (\mu_1 - \mu_0) \\ &= \sum_x P(x) \sum_m [P(m|a_1, x) - P(m|a_0, x)] \sum_a \sum_u E[Y|a, m, x, u] \cdot P(u|a, m, x) \cdot P(a|x) \\ &- \sum_x P(x) \sum_u \sum_m \{[P(m|a_1, x, u) - P(m|a_0, x, u)] E[Y|a_1, m, x, u]\} P(u|x) \\ &- \sum_x P(x) \sum_u \sum_m \{[E[Y|a_1, m, x, u] - E[Y|a_0, m, x, u]] P(m|a_0, x, u)\} P(u|x) \end{aligned} \tag{5}$$

Note that the last line is zero when Assumption 1 holds. When Assumption 2 holds as well, B_{ATE}^{fd} can be shown to be zero similarly to B_0^{fd}.

In order to compare the bias in the front-door approach to the standard back-door approach, we write the back-door formula for ATE based on observed covariates as the following:

$$\mu_1^{bd} - \mu_0^{bd} = \sum_x P(x)[E[Y|a_1, x] - E[Y|a_0, x]], \tag{6}$$

and the bias of the back-door formula as the following (see Appendix S.A.2 for a proof),

which is very similar to the formula presented in VanderWeele and Arah (2011):

$$
\begin{aligned}
B^{bd}_{ATE} &= \mu^{bd}_1 - \mu^{bd}_0 - (\mu_1 - \mu_0) \\
&= \sum_x P(x) \sum_u \{[P(u|a_1, x) - P(u|x)] \cdot [E[Y|a_1, x, u] - E[Y|a_0, x, u]]\} \\
&- \sum_x P(x) \sum_u \{[P(u|a_1, x) - P(u|a_0, x)] \cdot [E[Y|a_0, x, u]\}
\end{aligned} \tag{7}
$$

There are two important general facts to note about the comparison between B^{fd}_{ATE} and B^{bd}_{ATE}. First, it is quite possible that the front-door ATE bias will be smaller than the back-door ATE bias even when the aforementioned front-door independence conditions do not hold exactly. Second, because both estimators are defined within levels of the observed covariates X, it is possible to form hybrid estimators that utilize the front-door estimator for some values of X and the back-door estimator for other values of X. In order to develop some intuition about when the front-door approach would be preferred to the back-door approach (perhaps within a level of X), we next consider the special case of linear Structural Equation Models with constant effects (SEMs) and a scalar M.

2 Special Case: Linear Structural Equation Models

If we assume additive linear models with constant effects for Y and M, then:

$$E[Y|a, m, x] - E[Y|a, m', x] = \kappa(m - m'), \tag{8}$$

which is constant in a and x, and:

$$E[M|a_1, x] - E[M|a_0, x] = \lambda(a_1 - a_0), \tag{9}$$

which is constant in x. This allows us to write the front-door ATE as the following (proof in Appendix S.B.1):

$$\mu_1^{fd} - \mu_0^{fd} = \kappa\lambda(a_1 - a_0) \tag{10}$$

Therefore, when we assume additive linear models, the front-door formula for ATE simplifies to a product of multiple regression coefficients. If we also assume that Y is an additive linear model in a, x, and u, then $E[Y|a_1, x, u] - E[Y|a_0, x, u] = \tau(a_1 - a_0)$ and the ATE simplifies as well:

$$\begin{aligned} \mu_1 - \mu_0 &= \sum_x \sum_u \tau(a_1 - a_0) \cdot P(u|x) \cdot P(x) \\ &= \tau(a_1 - a_0) \end{aligned} \tag{11}$$

In order to present the ATE bias in the front-door approach, it will be helpful to present a simplified linear structural equation model with constant effects for these variables. This is defined by the path diagram in Figure 1. For simplicity in presentation, independent error terms have been removed from the graph, we have assumed that there are no measured conditioning variables, and we have assumed that A, M, U, and Y are scalars. Note that when $a_1 - a_0 = 1$, the ATE τ can be written as the following for this model:

$$\tau = \alpha\beta + \gamma \tag{12}$$

When $a_1 - a_0 = 1$, the front-door formula is the following (see Appendix S.B.1):

$$\mu_1^{fd} - \mu_0^{fd} = \alpha\beta + \alpha\theta e \frac{V(U|A)}{V(M|A)} + \beta e \delta \frac{V(U)}{V(A)} + e^2 \delta\theta \frac{V(U)}{V(A)} \frac{V(U|A)}{V(M|A)} \tag{13}$$

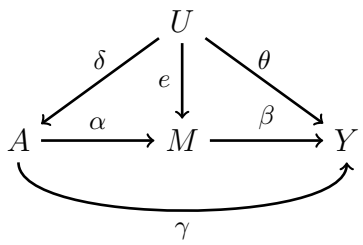

Figure 1: SEM

and the difference between the front-door formula and the ATE is the following:

$$B_{ATE}^{fd} = \alpha\theta e\frac{V(U|A)}{V(M|A)} + \beta e\delta\frac{V(U)}{V(A)} + e^2\delta\theta\frac{V(U)}{V(A)}\frac{V(U|A)}{V(M|A)} - \gamma \tag{14}$$

Therefore, the front-door formula will equal the ATE when the first three terms equal γ. In other words, when the bias for the indirect effect equals the direct effect. A special case of this is the situation when $e = 0$ and $\gamma = 0$, and this can itself be seen as an example of the front-door criterion within the context of SEMs.

For comparison, the back-door formula and bias can be written as the following (see Appendix S.B.2):

$$\mu_1^{bd} - \mu_0^{bd} = \alpha\beta + \gamma + (\beta e\delta + \theta\delta)\frac{V(U)}{V(A)} \tag{15}$$

$$B_{ATE}^{bd} = (\beta e\delta + \theta\delta)\frac{V(U)}{V(A)} \tag{16}$$

When comparing the back-door and front-door bias within SEMs, we first notice that both share the $\beta e\delta\frac{V(U)}{V(A)}$ terms. This represents the $A \leftarrow U \rightarrow M \rightarrow Y$ path. The key comparison is between the bias terms unique to the front-door formula ($\alpha\theta e\frac{V(U|A)}{V(M|A)} + e^2\delta\theta\frac{V(U)}{V(A)}\frac{V(U|A)}{V(M|A)} - \gamma$)

and the bias term unique to the back-door formula ($\theta\delta\frac{V(U)}{V(A)}$). Roughly speaking, the front-door bias can be smaller than the back-door bias when e and γ are small or when the front-door bias terms cancel. Notice as well that the front-door and back-door bias will be equal when $\theta = 0$ and $\gamma = 0$, which is equivalent to saying that there is no direct effect from A to Y or from U to Y. Therefore, another general case where the front-door will be preferred to the back-door is when U is largely mediated by M, and the bias from the common term is ameliorated by the direct effect ($\beta e\delta\frac{V(U)}{V(A)} - \gamma$).

S.A Asymptotic Bias Proofs

S.A.1 Front-door Bias

The large-sample bias for the front-door formula of $E[Y(a_0)]$ can be derived as the following:

$$\begin{aligned}
B_0^{fd} &= \mu_0^{fd} - \mu_0 \\
&= \sum_x \sum_m P(m|a_0, x) \sum_a E[Y|a, m, x] \cdot P(a|x) \cdot P(x) - \sum_x \sum_u E[Y|a_0, x, u] \cdot P(u|x) \cdot P(x) \\
&= \sum_x \sum_m P(m|a_0, x) \sum_a \sum_u E[Y|a, m, x, u] \cdot P(u|a, m, x) \cdot P(a|x) \cdot P(x) \\
&- \sum_x \sum_u \sum_m E[Y|a_0, m, x, u] \cdot P(m|a_0, x, u) \cdot P(u|x) \cdot P(x) \\
&= \sum_x P(x) \sum_m \sum_u P(m|a_0, x) \sum_a E[Y|a, m, x, u] \cdot P(u|a, m, x) \cdot P(a|x) \\
&- \sum_x P(x) \sum_m \sum_u P(m|a_0, x, u) \cdot E[Y|a_0, m, x, u] \cdot P(u|x)
\end{aligned} \tag{17}$$

The large-sample bias for the front-door formula for ATE can be derived as the following:

$$
\begin{aligned}
B_{ATE}^{fd} &= \mu_1^{fd} - \mu_0^{fd} - (\mu_1 - \mu_0) \\
&= \sum_x P(x) \sum_m [P(m|a_1, x) - P(m|a_0, x)] \sum_a E[Y|a, m, x] \cdot P(a|x) \\
&- \sum_x P(x) \sum_u \{E[Y|a_1, x, u] - E[Y|a_0, x, u]\} \cdot P(u|x) \\
&= \sum_x P(x) \sum_m [P(m|a_1, x) - P(m|a_0, x)] \sum_a \sum_u E[Y|a, m, x, u] \cdot P(u|a, m, x) \cdot P(a|x) \\
&- \sum_x P(x) \sum_u \sum_m \{E[Y|a_1, m, x, u] \cdot P(m|a_1, x, u) - E[Y|a_0, m, x, u] \cdot P(m|a_0, x, u)\} \cdot P(u|x) \\
&+ \sum_x P(x) \sum_u \sum_m P(m|a_0, x, u) \cdot E[Y|a_1, m, x, u] - \sum_x P(x) \sum_u \sum_m P(m|a_0, x, u) \cdot E[Y|a_1, m, x, u] \\
&= \sum_x P(x) \sum_m [P(m|a_1, x) - P(m|a_0, x)] \sum_a \sum_u E[Y|a, m, x, u] \cdot P(u|a, m, x) \cdot P(a|x) \\
&- \sum_x P(x) \sum_u \sum_m [P(m|a_1, x, u) - P(m|a_0, x, u)] \cdot E[Y|a_1, m, x, u] \cdot P(u|x) \\
&- \sum_x P(x) \sum_u \sum_m \{E[Y|a_1, m, x, u] - E[Y|a_0, m, x, u]\} \cdot P(m|a_0, x, u) \cdot P(u|x)
\end{aligned}
\tag{18}
$$

S.A.2 Back-door Bias

The back-door formula for ATE based on the observed covariates is the following:

$$
\mu_1^{bd} - \mu_0^{bd} = \sum_x P(x) \cdot \{E[Y|a_1, x] - E[Y|a_0, x]\},
$$

and the large sample bias of the back-door formula is the following:

$$
\begin{aligned}
B^{bd}_{ATE} &= \mu^{bd}_1 - \mu^{bd}_0 - (\mu_1 - \mu_0) \\
&= \sum_x P(x) \cdot \{E[Y|a_1, x] - E[Y|a_0, x]\} \\
&- \sum_x P(x) \sum_u \{E[Y|a_1, x, u] - E[Y|a_0, x, u]\} \cdot P(u|x) \\
&= \sum_x P(x) \sum_u \{E[Y|a_1, x, u] \cdot P(u|a_1, x) - E[Y|a_0, x, u] \cdot P(u|a_0, x)\} \\
&- \sum_x P(x) \sum_u [E[Y|a_1, x, u] - E[Y|a_0, x, u]] \cdot P(u|x)
\end{aligned}
$$

Adding and subtracting $\sum_x P(x) \sum_u P(u|a_1, x) \cdot E[Y|a_o, x, u]$:

$$
\begin{aligned}
&= \sum_x P(x) \sum_u \{E[Y|a_1, x, u] - E[Y|a_0, x, u]\} \cdot P(u|a_1, x) \\
&- \sum_x P(x) \sum_u [E[Y|a_0, x, u] \cdot [P(u|a_1, x) - P(u|a_0, x)] \\
&- \sum_x P(x) \sum_u \{E[Y|a_1, x, u] - E[Y|a_0, x, u]\} \cdot P(u|x) \\
&= \sum_x P(x) \sum_u \{E[Y|a_1, x, u] - E[Y|a_0, x, u]\} \cdot [P(u|a_1, x) - P(u|x)] \\
&- \sum_x P(x) \sum_u E[Y|a_0, x, u] \cdot [P(u|a_1, x) - P(u|a_0, x)]
\end{aligned} \tag{19}
$$

S.B Linear SEM Proofs

S.B.1 Front-door Formula

When writing the front-door formula for ATE within linear SEMs, note that $\sum_m [P(m|a_1, x) - P(m|a_0, x)] = 0$, so if we choose a reference value of m', then we can include the quantity

$-\sum_a E[Y|a,m',x]\cdot P(a|x)\cdot P(x)$ which is constant in m. If we further assume additive linear models for Y and M, then $E[Y|a,m,x]-E[Y|a,m',x]=\kappa(m-m')$, which is constant in a and x, and $E[M|a_1,x]-E[M|a_0,x]=\lambda(a_1-a_0)$ which is constant in x.

$$
\begin{aligned}
\mu_1^{fd}-\mu_0^{fd} &= \sum_x P(x)\sum_m [P(m|a_1,x)-P(m|a_0,x)]\sum_a E[Y|a,m,x]\cdot P(a|x)\\
&= \sum_x P(x)\sum_m [P(m|a_1,x)-P(m|a_0,x)]\sum_a \{E[Y|a,m,x]-E[Y|a,m',x]\}\cdot P(a|x)\\
&= \sum_x P(x)\sum_m [P(m|a_1,x)-P(m|a_0,x)]\sum_a \kappa(m-m')\cdot P(a|x)\\
&= \sum_x P(x)\sum_m [P(m|a_1,x)-P(m|a_0,x)]\kappa(m-m')\\
&= \sum_x P(x)\sum_m [P(m|a_1,x)-P(m|a_0,x)]\kappa m\\
&= \sum_x P(x)\kappa\sum_m [mP(m|a_1,x)-mP(m|a_0,x)]\\
&= \sum_x P(x)\kappa\{E[M|a_1,x]-E[M|a_0,x]\}\\
&= \sum_x P(x)\kappa\lambda(a_1-a_0)\\
&= \kappa\lambda(a_1-a_0)
\end{aligned}
\tag{20}
$$

Therefore, when we assume additive linear models, the front-door formula for ATE simplifies to a product of multiple regression coefficients. If we also assume that Y is an additive linear model in a, x, and u, then $E[Y|a_1,x,u]-E[Y|a_0,x,u]=\tau(a_1-a_0)$ and the ATE simplifies as well.

We can express κ and λ in terms of covariances:

$$\begin{aligned}\kappa &= \frac{Cov(Y, M|A)}{V(M|A)} \\ &= \beta + \theta e \frac{V(U|A)}{V(M|A)}\end{aligned} \tag{21}$$

$$\begin{aligned}\lambda &= \frac{Cov(M, A)}{V(A)} \\ &= \alpha + e\delta \frac{V(U)}{V(A)}\end{aligned} \tag{22}$$

Within the linear SEM the following covariance relationships hold (we omit uncorrelated errors in these expressions as is typically done with SEM graphs since they do not affect the derivations):

$$\begin{aligned}Cov(Y, M|A) &= Cov(\beta M + \gamma A + \theta U, M|A) \\ &= \beta Cov(M, M|A) + \theta Cov(U, M|A) \\ &= \beta V(M|A) + \theta Cov(U, \alpha A + eU|A) \\ &= \beta V(M|A) + \theta e V(U|A)\end{aligned} \tag{23}$$

$$\begin{aligned}Cov(M, A) &= Cov(\alpha A + eU, A) \\ &= \alpha V(A) + eCov(U, A) \\ &= \alpha V(A) + eCov(U, \delta U) \\ &= \alpha V(A) + e\delta V(U)\end{aligned} \tag{24}$$

Therefore, when $a_1 - a_0 = 1$, the front-door formula is

$$\begin{aligned}\mu_1^{fd} - \mu_0^{fd} = \lambda\kappa &= (\alpha + e\delta\frac{V(U)}{V(A)})(\beta + \theta e\frac{V(U|A)}{V(M|A)}) \\ &= \alpha\beta + \alpha\theta e\frac{V(U|A)}{V(M|A)} + \beta e\delta\frac{V(U)}{V(A)} + e^2\delta\theta\frac{V(U)}{V(A)}\frac{V(U|A)}{V(M|A)}\end{aligned} \tag{25}$$

and the difference between the front-door formula and the ATE is the following:

$$\begin{aligned}B_{ATE}^{fd} = \lambda\kappa - \tau &= \alpha\beta + \alpha\theta e\frac{V(U|A)}{V(M|A)} + \beta e\delta\frac{V(U)}{V(A)} + e^2\delta\theta\frac{V(U)}{V(A)}\frac{V(U|A)}{V(M|A)} \\ &- \alpha\beta + \gamma \\ &= \alpha\theta e\frac{V(U|A)}{V(M|A)} + \beta e\delta\frac{V(U)}{V(A)} + e^2\delta\theta\frac{V(U)}{V(A)}\frac{V(U|A)}{V(M|A)} - \gamma\end{aligned} \tag{26}$$

S.B.2 Back-door Formula

The back-door formula and bias can be described in terms of the following covariance relationships:

$$\begin{aligned}Cov(Y, A) &= Cov(\beta M + \gamma A + \theta U, A) \\ &= \beta Cov(M, A) + \gamma Cov(A, A) + \theta Cov(U, A) \\ &= \beta(\alpha V(A) + e\delta V(U)) + \gamma V(A) + \theta Cov(U, eU) \\ &= \beta(\alpha V(A) + e\delta V(U)) + \gamma V(A) + \theta\delta V(U)\end{aligned} \tag{27}$$

$$\begin{aligned}\mu_1^{bd} - \mu_0^{bd} &= \frac{Cov(Y, A)}{V(A)} \\ &= \alpha\beta + \gamma + (\beta e\delta + \theta\delta)\frac{V(U)}{V(A)} \\ B_{ATE}^{bd} &= (\beta e\delta + \theta\delta)\frac{V(U)}{V(A)}\end{aligned} \tag{28}$$

References

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82:669–710. 2, 3

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 1 edition. 3

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2 edition. 3

Rubin, D. B. (2010). Reflections stimulated by the comments of Shadish (2010) and West and Thoemmes (2010). *Psychological Methods*, 15(1):38–46. 2

VanderWeele, T. J. and Arah, O. A. (2011). Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology*, 22(1):42–52. 1, 4