Simulation: Causal Mediation with Selection 1 September 2024
Senan Hogan-Hennessy, seh325@cornell.edu Cornell University

This document investigates a system where a randomised measure Z affects an outcome Y via two channels: directly $Z \to Y$, and indirectly via a mediator $D(Z) \to Y$.

Causal mediation methods decompose the effect of Z into indirect effects, the proportion of effect going through the $D(Z) \to Y$ channel, and direct effects, the $Z \to Y$ channel. Conventional methods assume that D is randomly assigned, conditional on Z and other observed covariates $\boldsymbol{X}_i$; this assumption is unlikely to hold in observation settings, such as relying on quasi-experimental variation in Z.

This document simulates a system where D is not randomly assigned, but is the result of Roy-style selection (based on treatment gains) involving observed selection factors $\boldsymbol{X}_i$ and unobserved U_i. It shows how conventional estimators, controlling only for observed $\boldsymbol{X}_i$ behave under different assumptions about the distribution of U_i.

1 Notation

Write Y_i for the observed outcome value e.g., long-run income, for individuals $i = 1, \dots, N$. Suppose Y_i is the outcome of two binary variables, $Z_i = 0, 1$ which is assigned randomly, and $D_i = 0, 1$ which individuals **select into** based on which Z value they receive. The researcher observes D_i, Y_i, but not their respective potential outcomes:

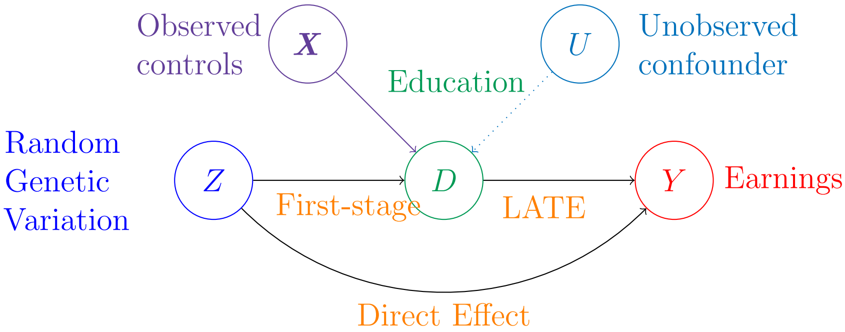
$$\begin{aligned}
D_i &= Z_i D_i(1) + (1 - Z_i) D_i(0), \\
&= \begin{cases} D_i(1), & \text{if } Z_i = 1 \\ D_i(0), & \text{if } Z_i = 0 \end{cases} \\
Y_i &= Z_i Y_i(1, D_i(1)) + (1 - Z_i) Y_i(0, D_i(0)) \\
&= \begin{cases} Y_i(1,1), & \text{if } Z_i = 1, D_i(1) = 1 \\ Y_i(1,0), & \text{if } Z_i = 1, D_i(1) = 0 \\ Y_i(0,1), & \text{if } Z_i = 0, D_i(0) = 1 \\ Y_i(0,0), & \text{if } Z_i = 0, D_i(0) = 0 \end{cases} .
\end{aligned}$$

In my empirical work, Z is a binary version of the gene score for education (differenced from parents' values, EA score), $D_i(Z_i)$ is a choice to complete higher education, and Y_i a measure of long-run income. $\boldsymbol{X}_i$ is demographic information, gender, age, and every measure of socio-economic standing available; U_i is covariates the **researcher wants to control for, but does not observe** in the data they have.

1.1 Direct and Indirect Effects

Causal mediation aims to decompose the reduced form effect of $Z \to Y$ into two separate pathways: indirectly through D, and directly absent D.

Figure 1: Structural Causal Graph of the Triangular System, $Z \to D \to Y$.



$$
\begin{aligned}
\text{Reduced Form:} \quad & \mathbb{E}\left[Y_i(1, D_i(1)) - Y_i(0, D_i(0))\right] = \mathbb{E}\left[Y_i \mid Z_i = 1\right] - \mathbb{E}\left[Y_i \mid Z_i = 0\right] \\
\text{Indirect Effect, } D(Z) \to Y: \quad & \mathbb{E}\left[Y_i(Z_i, D_i(1)) - Y_i(Z_i, D_i(0))\right] \\
\text{Direct Effect, } Z \to Y: \quad & \mathbb{E}\left[Y_i(1, D_i(Z_i)) - Y_i(0, D_i(Z_i))\right]
\end{aligned}
$$

The reduced form is the average effect of EA score on later-life earnings; the indirect effect is the effect of EA score operating purely through increased education; the direct effect is the effect of EA score operating absent education.

2 A Regression Framework for Direct and Indirect Effects

Inference for direct and direct effects can be written in a regression framework, showing how correlation between the error term and the mediator persistently biases estimates.

To motivate a regression framework, write $Y_i(Z, D)$ as a sum of observed factors Z_i, $\boldsymbol{X}_i$ and unobserved factors.

$$
Y_i(Z_i, 0) = \mu_0(Z_i; \boldsymbol{X}_i) + U_{0,i}, \quad Y_i(Z_i, 1) = \mu_1(Z_i; \boldsymbol{X}_i) + U_{1,i}
$$

μ_0, μ_1 are unknown functions, $U_{0,i}, U_{1,i}$ are mean zero error terms with unknown distributions, independent of $Z_i, \boldsymbol{X}_i$ — but possibly correlated with D_i.

$$
\begin{aligned}
Y_i &= Z_i Y_i(1, D_i(1)) + (1 - Z_i) Y_i(0, D_i(0)) \\
&= Y_i(0, D_i(0)) + Z_i \left[Y_i(1, D_i(1)) - Y_i(0, D_i(0))\right] \\
&= \underbrace{\mu_{D_i(0;\boldsymbol{X}_i)}(0)}_{\text{Intercept}} + \underbrace{Z_i \left[\mu_{D_i(1)}(1; \boldsymbol{X}_i) - \mu_{D_i(0)}(0; \boldsymbol{X}_i)\right]}_{\text{Regressor}} \\
&\quad + \underbrace{U_{D_i(0),i} + Z_i \left(U_{D_i(1),i} - U_{D_i(0),i}\right)}_{\text{Error term, mean zero}} \\
&=: \phi_i + \varphi_i Z_i + \epsilon_i
\end{aligned}
$$

$$
\implies \mathbb{E}\left[Y_i \mid Z_i\right] = \mathbb{E}\left[\phi_i\right] + \mathbb{E}\left[\varphi_i\right] Z_i + \mathbb{E}\left[\epsilon_i\right], \text{ and thus unbiased estimates since } Z_i \perp\!\!\!\perp \varphi_i, \epsilon_i.
$$

Z_i is assumed randomly assigned, independent of potential outcomes, so that $U_{0,i}, U_{1,i} \perp\!\!\!\perp Z_i$. Thus, the reduced form regression $Z \to Y$ leads to unbiased estimates.

The same cannot be said of the regression that estimates direct and indirect effects, without further assumptions.

$$\begin{aligned}
Y_i = &\ Z_i D_i Y_i(1,1) \\
&+ (1 - Z_i) D_i Y_i(0,1) \\
&+ Z_i (1 - D_i) Y_i(1,0) \\
&+ (1 - Z_i)(1 - D_i) Y_i(0,0) \\
= &\ Y_i(0,0) \\
&+ Z_i \left[Y_i(1,0) - Y_i(0,0) \right] \\
&+ D_i \left[Y_i(0,1) - Y_i(0,0) \right] \\
&+ Z_i D_i \left[Y_i(1,1) - Y_i(1,0) - (Y_i(0,1) - Y_i(0,0)) \right]
\end{aligned}$$

And so Y_i can be written as a regression equation in terms of the observed factors and error terms.

$$\begin{aligned}
Y_i = &\ \mu_0(0; \boldsymbol{X}_i) \\
&+ Z_i \left[\mu_0(1; \boldsymbol{X}_i) - \mu_0(0; \boldsymbol{X}_i) \right] \\
&+ D_i \left[\mu_1(0; \boldsymbol{X}_i) - \mu_0(0; \boldsymbol{X}_i) \right] \\
&+ Z_i D_i \left[\mu_1(1; \boldsymbol{X}_i) - \mu_0(1; \boldsymbol{X}_i) - (\mu_1(0; \boldsymbol{X}_i) - \mu_0(0; \boldsymbol{X}_i)) \right] \\
&+ U_{0,i} + D_i \left(U_{1,i} - U_{0,i} \right) \\
=: &\ \alpha_i + \beta_i D_i + \gamma_i Z_i + \delta_i Z_i D_i + \varepsilon_i
\end{aligned}$$

$\alpha_i, \beta_i, \delta_i$ are the relevant direct effect under $D_i = 1$, indirect effect under $Z_i = 1$, δ_i the interaction effect, and ε_i the remaining error term. Collecting for the expressions of the direct and indirect effects:[1]

$$\begin{aligned}
\mathbb{E}\left[Y_i(Z_i, D_i(1)) - Y_i(Z_i, D_i(0)) \right] &= \mathbb{E}\left[(\beta_i + Z_i \delta_i) \times (D_i(1) - D_i(0)) \right] \\
\mathbb{E}\left[Y_i(1, D_i(Z_i)) - Y_i(0, D_i(Z_i)) \right] &= \mathbb{E}\left[\gamma_i + \delta_i D_i \right]
\end{aligned}$$

By assumption $Z_i \perp\!\!\!\perp \gamma_i, \varepsilon_i$, so that the regression only gives unbiased estimates if D_i is also conditionally random: $D_i(z) \perp\!\!\!\perp \varepsilon_i \mid \boldsymbol{X}_i$.

2.1 Selection into Education

Conventional causal mediation work point identifies the indirect and direct effects by additionally **assuming that D_i is randomly assigned**, conditional on $\{\boldsymbol{X}_i, Z_i\}$ — known as sequential ignorability (Imai et al., 2010).

$$Y_i(z, d) \perp\!\!\!\perp D_i(z') \mid Z_i = z, \boldsymbol{X}_i, \quad \text{for all } z, z', d = 0, 1$$

[1] These equations have simpler expressions after assuming constant treatment effects; I have avoided this as having compliers, and controlling for observed factors $\boldsymbol{X}_i$ only makes sense in the case of heterogeneous treatment effects.

In the education context, point identifying direct and indirect effects requires the **researcher controls for all sources of selection-into-education**.

While this assumption may hold true in two-way randomised experiments (e.g., in a laboratory or two-way RCT), it is unlikely to hold in the case of quasi-experimental variation in Z, or when modelling education as a mediator — absent a separate identification strategy for education D. To expand this point in an econometric selection-into-treatment framework, suppose selection follows a Roy model, where individual i weighs the costs and benefits of completing education.

$$D_i(Z_i) = \mathbb{1}\left\{ \underbrace{C_i(Z_i)}_{\text{Costs}} \leq \underbrace{Y_i(Z_i, 1) - Y_i(Z_i, 0)}_{\text{Gains}} \right\}$$

Education choice $D_i(z)$ is clearly related to $Y_i(z, d)$ in this model, so let's see what the equation looks like in terms of sequential ignorability. As above, decompose costs into observed and unobserved factors.

$$C_i(Z_i) = \mu_C(Z_i; \boldsymbol{X}_i) + U_{C,i}$$

And so we can write the first-stage selection equation in full.

$$D_i(Z_i) = \mathbb{1}\left\{ \underbrace{U_{C,i} + U_{0,i} - U_{1,i}}_{\text{Unobserved}} \leq \underbrace{\mu_1(Z_i; \boldsymbol{X}_i) - \mu_0(Z_i; \boldsymbol{X}_i) - \mu_C(Z_i; \boldsymbol{X}_i)}_{\text{Observed}} \right\}$$

Sequential ignorability, where $Y_i(z, d) \perp\!\!\!\perp D_i(z') \mid \boldsymbol{X}_i$, would then require that $\mathbb{E}\left[U_{0,i} - U_{1,i} \mid D_i\right] = 0$ — no unobserved selection! This is unlikely to hold true, unless there is another identification strategy for D_i — in addition to the one used for Z_i.

3 Simulation

This simulation assumes that

1. $\Pr(Z_i = 1) = \frac{1}{2}$ for every individual, so that Z_i is randomly assigned.
2. $U_{0,i}, U_{1,i} \sim \text{BivarNormal}(\rho, 0, 0, \sigma_0, \sigma_1)$, and $U_C = 0$ for simplicity.
3. $N = 1,000$
4. Observed covariates $\boldsymbol{X}_i = [X_i^1]$ is composed of $X_i^1 \sim \text{N}(0, 1)$.

The observed part of potential outcomes, $\mu_D(Z; \boldsymbol{X}_i)$, are simulated in a linear system, with $\boldsymbol{X}_i \sim N(5, 1)$ and the following definitions.

$$\begin{aligned}
\mu_0(0; \boldsymbol{X}_i) &= \beta_0 \boldsymbol{X}_i && = \boldsymbol{X}_i \\
\mu_1(0; \boldsymbol{X}_i) &= \beta_1 \boldsymbol{X}_i && = 2\boldsymbol{X}_i \\
\mu_0(1; \boldsymbol{X}_i) &= \beta_0 \boldsymbol{X}_i + \gamma_0 && = \boldsymbol{X}_i + 0.5 \\
\mu_1(1; \boldsymbol{X}_i) &= \beta_1 \boldsymbol{X}_i + \gamma_1 && = 2\boldsymbol{X}_i + 1 \\
\mu_C(0; \boldsymbol{X}_i) &= 5 \\
\mu_C(1; \boldsymbol{X}_i) &= 3.75
\end{aligned}$$

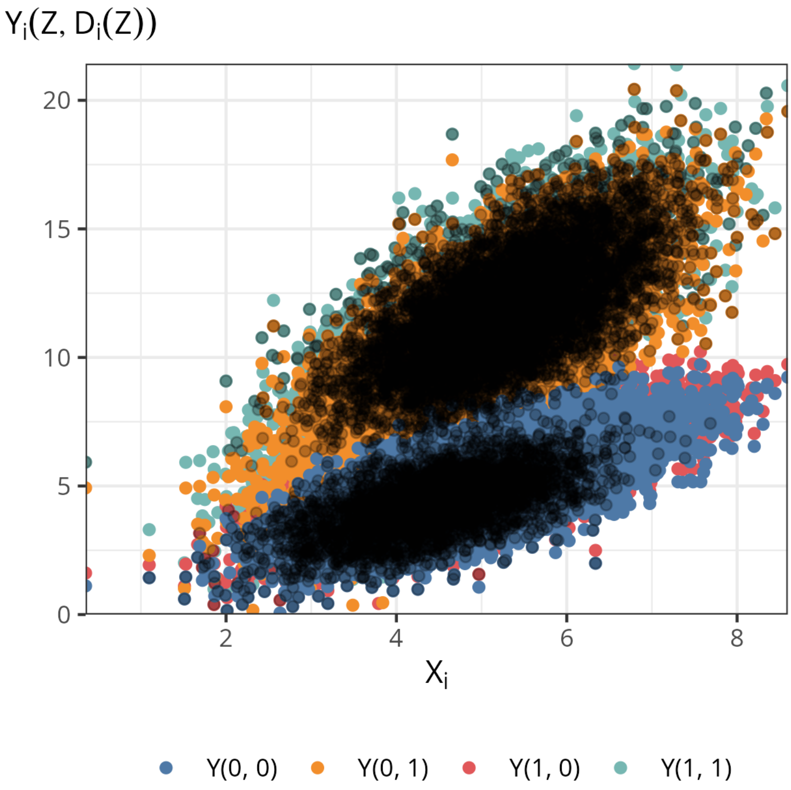
These values have the following properties, relevant to this system:

- There are compliers i.e., $0 < \Pr(D_i(0) < D_i(1))$ since gains to education do not always outweigh costs
- There are no defiers i.e., $0 = \Pr(D_i(0) > D_i(1))$ since opportunity costs of education are higher in $Z_i = 1$
- $\text{Corr}(U_{i,0}, U_{i,1}) = \rho > 0$ indicates positive selection into education, where those with higher incomes more often take education (independently of gains)
- $\sigma_1 \neq \sigma_0$ indicates heteoskedasicity in D_i, where error term variance is correlated with D_i.

What does this system look like?

$$\begin{aligned}
Y_i(Z_i, 0) = \beta_0 \boldsymbol{X}_i + \gamma_0 Z_i + U_{0,i}, \quad Y_i(Z_i, 1) = \beta_1 \boldsymbol{X}_i + \gamma_1 Z_i + U_{1,i} \\
D_i(Z_i) = \mathbb{1}\left\{\mu_C(Z_i; \boldsymbol{X}_i) + U_{C,i} \leq Y_i(Z_i, 1) - Y_i(Z_i, 0)\right\} \\
\implies Y_i = \beta_0 \boldsymbol{X}_i + \gamma_0 Z_i + \left[(\beta_0 - \beta_1)\boldsymbol{X}_i\right] D_i + (\gamma_1 - \gamma_0) Z_i D_i + U_{0,i} + D_i\left(U_{1,i} - U_{0,i}\right) \\
= \boldsymbol{X}_i + 0.5 Z_i + \boldsymbol{X}_i D_i + 0.5 Z_i D_i + \underbrace{U_{0,i} + D_i\left(U_{1,i} - U_{0,i}\right)}_{\text{Correlated error term}} \\
\mathbb{E}\left[Y_i(Z_i, 1) - Y_i(Z_i, 0)\right] = (\beta_1 - \beta_0)\mathbb{E}\left[\boldsymbol{X}_i\right] + (\gamma_1 - \gamma_0)\mathbb{E}\left[Z_i\right] + \mathbb{E}\left[U_{1,i} - U_{0,i}\right] = 5.25 \\
\mathbb{E}\left[Y_i(1, D_i(Z_i)) - Y_i(0, D_i(Z_i))\right] = \gamma_0 + (\gamma_1 - \gamma_0)\mathbb{E}\left[D_i\right] = 0.8313
\end{aligned}$$

Figure 2: Simulated Outcomes, with $\rho, \sigma_0, \sigma_1 = 3/4, 1, 2$.



Note: The transparent black dots are overlaid realised Y_i values. See the first equation for an explanation of how $Y_i(0,1)$ is only realised for always-takers, $D_i(0) = 1$.

3.1 Varying the Parameter Values

There are three values that define the system, mimicking the famous sample selection model of Heckman (1974, 1979):

Parameter	Equation	Explanation
ρ	$\mathrm{Corr}(U_{i,0}, U_{i,1})$	Correlation between $D_i = 1$ and $D_i = 0$ error terms
σ_0	$\mathrm{Var}(U_{i,0})^{\frac{1}{2}}$	Standard deviation of $D_i = 0$ error terms
σ_1	$\mathrm{Var}(U_{i,1})^{\frac{1}{2}}$	Standard deviation of $D_i = 1$ error terms

This simulation file varies the values of ρ, σ_0, σ_1 to investigate how the bias in conventional mediation estimates behaves under different assumptions of the unobserved error values U_0, U_1.

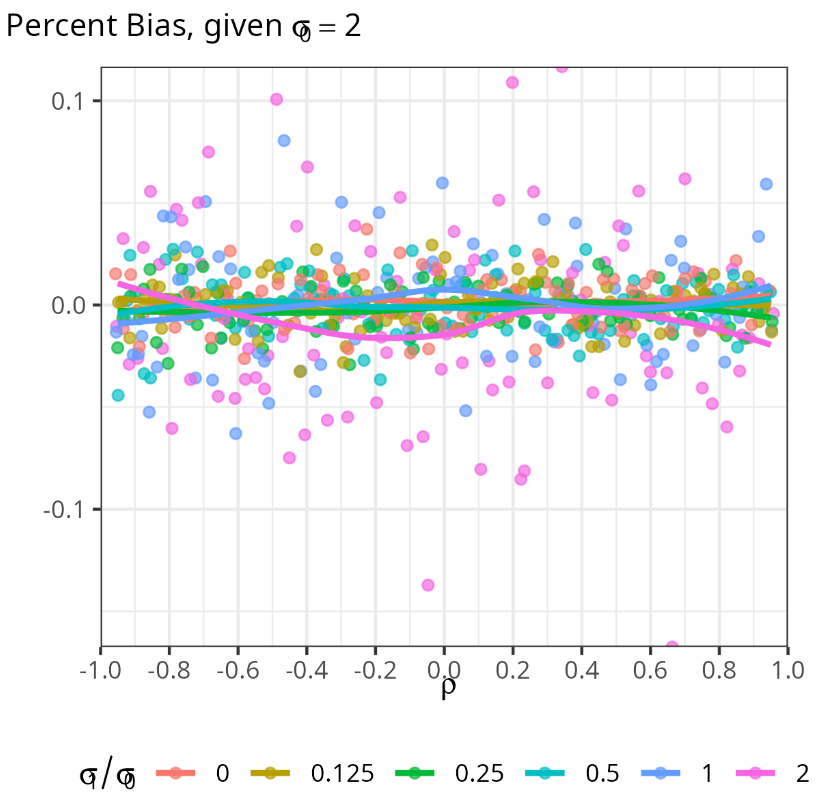
3.2 Bias in the Reduced Form Estimate

I expected the following relationship between these parameter values, and the bias in estimating the **reduced form effect**, $\mathbb{E}[Y_i \mid Z_i = 1] - \mathbb{E}[Y_i \mid Z_i = 0]$.

- Increasing both σ_0, σ_1 reduces precision
- $\sigma_1/\sigma_0 \neq 1$ indicates heteroskedasticity along D_i (not bias)
- Changing ρ has no effect on bias (may affect precision).

This is generally confirmed by the simulation, in Figure 3.

Figure 3: Bias in Reduced Form Estimates in Simulated Data, across different ρ, σ_0, σ_1 values.



Note: This figure shows the percent bias in the regression $Y_i = \phi + \theta Z_i + \boldsymbol{\zeta}_i' \boldsymbol{X}_i + \eta_i$, where the y-axis is $(\widehat{\theta}_{\text{OLS}} - \theta)/\theta$, given θ the true value of the reduced form effect.

3.3 Bias in the Direct and Indirect Effect Estimates

I expected the following relationship between these parameter values, and the bias in estimating the **Direct Effect** $Z \to Y: \quad \mathbb{E}[Y_i(1, D_i(Z_i)) - Y_i(0, D_i(Z_i))]$

- This estimate relies on estimating $D \to Y$ by selection-on-observables, so $\rho > 0$ indicates unobserved selection into treatment and downwards biases estimates

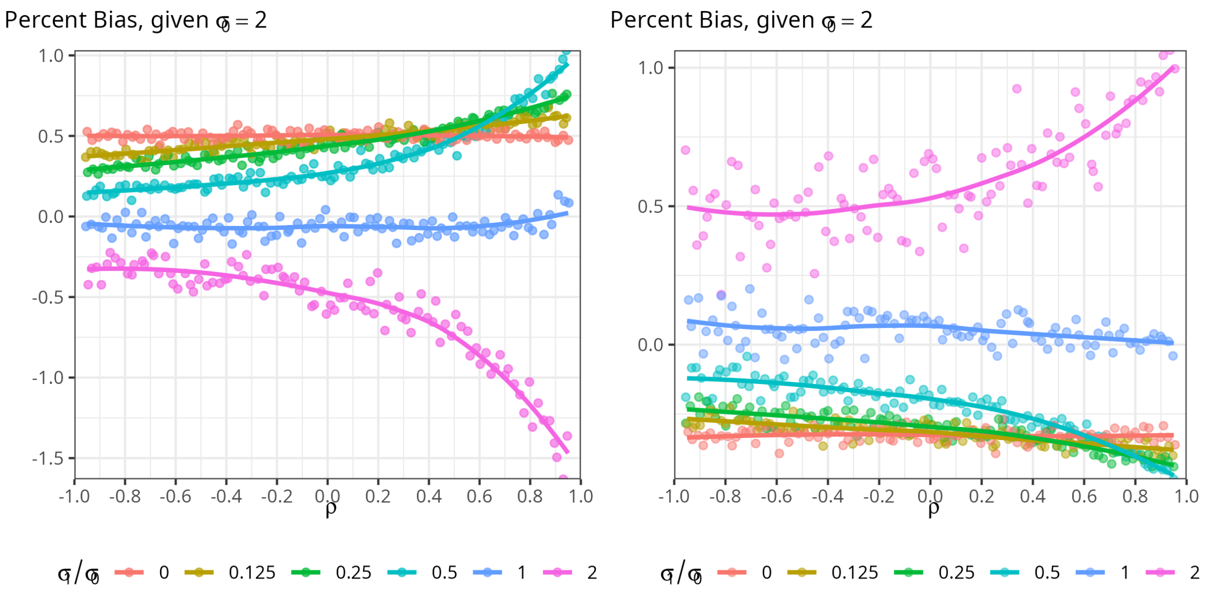
- σ_0, σ_1 have ambiguous effects on bias, beyond heteroskedasticity for inference.

I expected the following relationship between these parameter values, and the bias in estimating the **Indirect Effect** $D(Z) \to Y: \ \mathbb{E}\left[Y_i(Z_i, D_i(1)) - Y_i(Z_i, D_i(0))\right]$.

- This estimate relies on estimating $D \to Y$ by selection-on-observables, so $\rho > 0$ indicates unobserved selection into treatment and upwards biases estimates
- σ_0, σ_1 have ambiguous effects on bias, beyond heteroskedasticity for inference.

Figure 4: Bias of Point Estimates in Simulated Data, across different ρ, σ_0, σ_1 values.

(a) Direct Effect Estimates. **(b)** Indirect Effect Estimates.



Note: This figure shows the percent bias in the regression $Y_i = \alpha + \beta D_i + \gamma Z_i + \delta Z_i D_i + \boldsymbol{\zeta}_i' \boldsymbol{X}_i + \varepsilon_i$, where the y-axis is the corresponding OLS estimate for direct or indirect minus then divided by the true value of the reduced form effect.

4 A Control Function Solution(?)

I have shown above that the mediation equations without sequential ignorability take the following form, with first-stage error term $U_i = -(U_{1,i} - U_{0,i} - U_{C,i})$ and non-parametric regressor $\mu = \mu_1 - \mu_0 - \mu_C$.

$$D_i(Z_i) = \mathbb{1}\left\{U_i \leq \mu(Z_i; \boldsymbol{X}_i)\right\}$$
$$Y_i = \alpha_i + \gamma_i Z_i + \beta_i D_i + \delta_i Z_i D_i + U_{0,i} + D_i\left(U_{1,i} - U_{0,i}\right)$$

The control function approach solves identification in this exact problem. The classic Heckman (1979) approach does so by maximum likelihood with errors $U_{1,i}, U_{0,i}$ assumed normal. **This approach works exactly in the simulation above, i.e. with simulated normal errors (and even heterogeneous treatment effects).**

Newer semi-parametric approaches use a two-step approach to avoid assuming the distribution of the error terms (Newey et al., 1999; Imbens and Newey, 2009). The identifying assumption is that error terms in the first and second-stages are correlated, so that first-stage predicted residuals control for endogeneity in the second-stage.

$$\widehat{U}_i = D_i - \widehat{\mathbb{E}\left[D_i \mid \boldsymbol{X}_i, Z_i\right]} = \widehat{f_D}(\mu(Z_i\ \boldsymbol{X}_i))$$
$$Y_i = \alpha_i + \beta_i D_i + \gamma_i Z_i + \delta_i Z_i D_i + \widehat{U}_i D_i + \varepsilon_i$$

This assumption holds exactly in the Roy model, with perfectly correlated errors (minus costs variation).

4.1 Discussion:

I don't see any modern applied work using control function estimators. . . .

The control function approach assumes the error terms in the first-stage selection equation are informative for the errors in the second-stage outcome equation; this is trivial in the Roy model, though not the only first-stage selection consistent with the approach. It may make sense for me to write exclusively in a structural setting using the Roy model, and hold off on considering this approach more generally.

I have concerns:

- I only see highly technical econometric theory papers taking the control function approach
- The control function approach here replaces one assumption (D_i randomised) for another (correlated error terms).
- The second assumption is consistent (inspired by) the Roy model; the first assumption is inconsistent with a general labour/natural experiment setting
- Is this approach straying too far into the "structural world" for an applied project?

A Appendix

A.1 Things to look into

Newest thought: Use a semi-parametric two-step control function estimator to get estimates of the direct and indirect effects.

A.1.1 Thought on Sensitivity Analysis

If the above two-step control function works, then controlling for $\boldsymbol{X}_i$ in the second stage is irrelevant, except for precision (i.e., magnitude of standard errors). So estimates with varying inclusion of controls in $\boldsymbol{X}_i$ should be unbiased, even if less precise.

This should be investigated, showing the two-step control function estimates sequentially adding controls in $\boldsymbol{X}_i$ and that there is no general trend (other than more precise estimates).

A.1.2 Explaining Compliance

Sequential ignorability assumes that all levers of selection are controlled for in observed factors $\boldsymbol{X}_i$. The next step is getting a measure of how much compliance is unexplained, which is equivalent to how large U_i is in the outcome equation in the Roy model.

The first option is to measure how much compliance is explained by $\boldsymbol{X}_i$.

$$\text{Var}\left(D_i(1) - D_i(0)\right) = \underbrace{\text{Var}\left(\mathbb{E}\left[D_i(1) - D_i(0) \,|\, X_i\right]\right)}_{\text{Compliance explained by } \boldsymbol{X}_i} + \underbrace{\mathbb{E}\left[\text{Var}\left(D_i(1) - D_i(0) \,|\, |\boldsymbol{X}_i\right)\right]}_{\text{Compliance unexplained}}$$

$$\implies R_U^2 = 1 - \frac{\text{Var}\left(\mathbb{E}\left[D_i(1) - D_i(0) \,|\, X_i\right]\right)}{\text{Var}\left(D_i(1) - D_i(0)\right)}$$

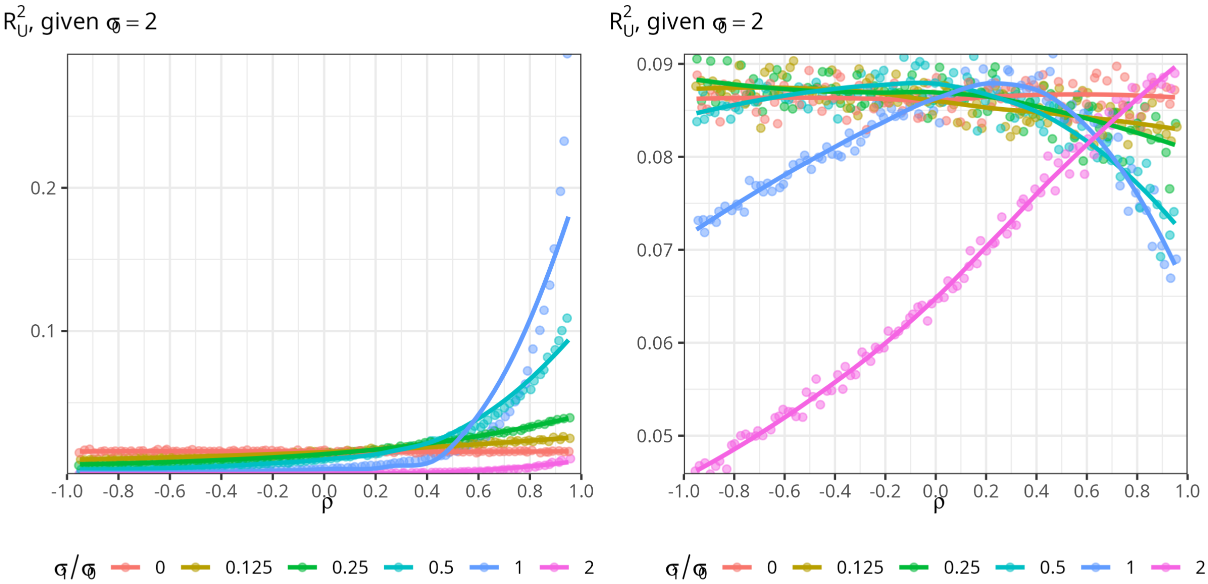
The second option is to measure how much variation in observed D_i is explained by $\boldsymbol{X}_i$, in the spirit of Altonji et al. (2005).

$$\text{Var}\left(D_i\right) = \underbrace{\text{Var}\left(\mathbb{E}\left[D_i \,|\, X_i\right]\right)}_{\text{Var}(D_i) \text{ explained by } \boldsymbol{X}_i} + \underbrace{\mathbb{E}\left[\text{Var}\left(D_i \,|\, |\boldsymbol{X}_i\right)\right]}_{\text{Var}(D_i) \text{ unexplained}}$$

$$\implies R_U^2 = 1 - \frac{\text{Var}\left(\mathbb{E}\left[D_i \,|\, X_i\right]\right)}{\text{Var}\left(D_i\right)}$$

Figure A1: Simulated R_U^2 Values.

(a) $R_U^2 = 1 - \frac{\mathrm{Var}(\mathbb{E}[D_i(1) - D_i(0) \,|\, X_i])}{\mathrm{Var}(D_i(1) - D_i(0))}$. (b) $R_U^2 = 1 - \frac{\mathrm{Var}(\mathbb{E}[D_i \,|\, X_i])}{\mathrm{Var}(D_i)}$.



Note: This figure shows the true values of R_U^2 in each simulation, based on bivariate normal error terms

Current thoughts: The idea of using R_U^2 seems not useful, given recent thoughts on using a two-step control function estimator Propose a hypothesis test, based on an estimated R_U^2 values, which (if violated) tests sequential ignorability (maybe only if selection is a Roy model). If $H_0 : R_U^2 = 0$ is rejected, then motivates the use of a control function estimator of the direct and indirect effects, instead of sequential ignorability estimates.

References

Altonji, J. G., Elder, T. E., and Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of political economy*, 113(1):151–184. 10

Heckman, J. (1974). Shadow prices, market wages, and labor supply. *Econometrica: journal of the econometric society*, pages 679–694. 6

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, pages 153–161. 6, 8

Imai, K., Keele, L., and Yamamoto, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*, pages 51–71. 3

Imbens, G. W. and Newey, W. K. (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, 77(5):1481–1512. 9

Newey, W. K., Powell, J. L., and Vella, F. (1999). Nonparametric estimation of triangular simultaneous equations models. *Econometrica*, 67(3):565–603. 9