Regression for Causal Inference

Jason Seawright

j-seawright@northwestern.edu

Jan. 16, 2024

Regression in Multi-Method Research

The Potential Outcomes Framework

- We are interested in the effects of a dichotomous treatment (i.e., independent variable).

The Potential Outcomes Framework

- We are interested in the effects of a dichotomous treatment (i.e., independent variable).
- This variable can be written as $D_i = (t, c)$.

The Potential Outcomes Framework

- For a given case, i, we either observe $D_i = t$ or $D_i = c$. If $D_i = t$, let us denote the value of the dependent variable as $y_{i,t}$. If $D_i = c$, let us denote the value of the dependent variable as $y_{i,c}$

The Potential Outcomes Framework

- The causal effect of D on y is:
  - $y_{i,t} - y_{i,c}$

The Average Treatment Effect

- Sometimes, we are interested in developing an estimate of the effect of D on y in some population, from which we have a random sample (or even the whole population) split randomly into treatment and control cases.

The Average Treatment Effect

- Sometimes, we are interested in developing an estimate of the effect of D on y in some population, from which we have a random sample (or even the whole population) split randomly into treatment and control cases.
- Here, interest focuses on the "average treatment effect":
  - $E(y_{i,t}) - E(y_{i,c})$

Experiments and Causal Inference

- Under random assignment, the set of cases where $D_i = t$ produces a random sample from the population of y_t. Likewise, the set of cases where $D_i = c$ produces a random sample from the population of y_c. Thus:
    - $E(y_{i,t}|D_i = t) = E(y_{i,t})$

Experiments and Causal Inference

- Under random assignment, the set of cases where $D_i = t$ produces a random sample from the population of y_t. Likewise, the set of cases where $D_i = c$ produces a random sample from the population of y_c. Thus:
  - $E(y_{i,t}|D_i = t) = E(y_{i,t})$
  - $E(y_{i,c}|D_i = c) = E(y_{i,c})$

Experiments and Causal Inference

- Under random assignment, the set of cases where $D_i = t$ produces a random sample from the population of y_t. Likewise, the set of cases where $D_i = c$ produces a random sample from the population of y_c. Thus:
  - $E(y_{i,t}|D_i = t) = E(y_{i,t})$
  - $E(y_{i,c}|D_i = c) = E(y_{i,c})$
  - $E(y_{i,t}) - E(y_{i,c}) = E(y_{i,t}|D_i = t) - E(y_{i,c}|D_i = c)$

Regression in Experiments

Recall that D_i as an indicator of treatment assignment in an experiment. Let's change the coding, such that D_i equals 1 for treatment cases and 0 for control cases. Then we can write:

$$Y_i = D_i Y_{i,t} + (1 - D_i) Y_{i,c} \tag{1}$$

Regression in Experiments

Recall that D_i as an indicator of treatment assignment in an experiment. Let's change the coding, such that D_i equals 1 for treatment cases and 0 for control cases. Then we can write:

$$Y_i = D_i Y_{i,t} + (1 - D_i) Y_{i,c} \tag{1}$$

$$Y_i = Y_{i,c} + D_i (Y_{i,t} - Y_{i,c}) \tag{2}$$

Regression in Experiments

$$Y_i = \bar{Y}_{i,c} + D_i(\bar{Y}_{i,t} - \bar{Y}_{i,c}) + [Y_{i,c} - \bar{Y}_{i,c} + D_i(Y_{i,t} - \bar{Y}_{i,t} - Y_{i,c} + \bar{Y}_{i,c})] \tag{3}$$

Regression in Experiments

Suppose we use OLS to estimate:

$$Y_i = \beta_0 + \beta_1 D_i + \epsilon_i$$

Will this work?

Regression in Experiments

$$\hat{\beta}_1 = \frac{cov(D, Y)}{var(D)} = \frac{\sum(D - \bar{D})(Y - \bar{Y})}{\sum(D - \bar{D})^2}$$

Regression in Experiments

$$\hat{\beta}_1 = \frac{cov(D, Y)}{var(D)} = \frac{\sum(D - \bar{D})(Y - \bar{Y})}{\sum(D - \bar{D})^2}$$

Let π equal the proportion of cases assigned to the treatment group.

Regression in Experiments

$$\hat{\beta}_1 = \frac{cov(D, Y)}{var(D)} = \frac{\sum(D - \bar{D})(Y - \bar{Y})}{\sum(D - \bar{D})^2}$$

Let π equal the proportion of cases assigned to the treatment group.

$$\hat{\beta}_1 = \frac{\sum(D - \pi)(Y - \bar{Y})}{N\pi(1 - \pi)}$$

Regression in Experiments

$$\hat{\beta}_1 = \frac{\sum_{D_i=1}(1-\pi)(Y-\bar{Y}) - \sum_{D_i=0}\pi(Y-\bar{Y})}{N\pi(1-\pi)}$$

Regression in Experiments

$$\hat{\beta}_1 = \frac{\sum_{D_i=1}(1-\pi)(Y-\bar{Y}) - \sum_{D_i=0}\pi(Y-\bar{Y})}{N\pi(1-\pi)}$$

$$\hat{\beta}_1 = \frac{\sum_{D_i=1}(Y_i - E(Y_i))}{N\pi} - \frac{\sum_{D_i=0}(Y_i - E(Y_i))}{N(1-\pi)}$$

Regression in Experiments

$$\hat{\beta}_1 = \frac{\sum_{D_i=1} Y_i}{N\pi} - \frac{\sum_{D_i=0} Y_i}{N(1-\pi)}$$

Regression in Experiments

$$\hat{\beta}_1 = \frac{\sum_{D_i=1} Y_i}{N\pi} - \frac{\sum_{D_i=0} Y_i}{N(1-\pi)}$$

$N\pi$ is just the number of cases where $D_i = 1$, and $N(1-\pi)$ is the number of cases where $D_i = 0$. So the last expression simplifies to:

$$\hat{\beta}_1 = \bar{Y}_{i,D_i=1} - \bar{Y}_{i,D_i=0}$$

Regression in Experiments

Multivariate regression for experiments is *not* guaranteed to be unbiased in the way that bivariate regression is. Multivariate regression can be (even very badly) biased if:

Regression in Experiments

Multivariate regression for experiments is *not* guaranteed to be unbiased in the way that bivariate regression is. Multivariate regression can be (even very badly) biased if:

- Some control variable W is included in the model that is in any way caused by the treatment, or

Regression in Experiments

Multivariate regression for experiments is *not* guaranteed to be unbiased in the way that bivariate regression is. Multivariate regression can be (even very badly) biased if:

- The causal effect of interest is highly heterogeneous across categories of the control variables.

Regression for Causal Inference

$$
\begin{aligned}
Y_i &= E(Y_{i,c}) + D_i\{E(Y_{i,t}) - E(Y_{i,c})\} + [Y_{i,c} - E(Y_{i,c})] \\
&\quad + D_i([Y_{i,t} - E(Y_{i,t})] - [Y_{i,c} - E(Y_{i,c})]) \\
&= \mu_0 + D_i(\mu_1 - \mu_0) + \{\nu_0 + D_i(\nu_1 - \nu_0)\}
\end{aligned}
$$

Regression for Causal Inference

- A collection of control variables $\mathbb{X}$ will allow regression to produce an unbiased estimate of $(\mu_1 - \mu_0)$ when:

Regression for Causal Inference

- A collection of control variables $\mathbb{X}$ will allow regression to produce an unbiased estimate of $(\mu_1 - \mu_0)$ when:
  1. D is uncorrelated with $\{\nu_0 + D_i(\nu_1 - \nu_0)\}$ within each group defined by $\mathbb{X}$, and

Regression for Causal Inference

- A collection of control variables $\mathbb{X}$ will allow regression to produce an unbiased estimate of $(\mu_1 - \mu_0)$ when:
  1. D is uncorrelated with $\{\nu_0 + D_i(\nu_1 - \nu_0)\}$ within each group defined by $\mathbb{X}$, and
  2. the residual causal effect is not correlated with $\mathbb{X}$, and

Regression for Causal Inference

- A collection of control variables $\mathbb{X}$ will allow regression to produce an unbiased estimate of $(\mu_1 - \mu_0)$ when:
  1. D is uncorrelated with $\{\nu_0 + D_i(\nu_1 - \nu_0)\}$ within each group defined by $\mathbb{X}$, and
  2. the residual causal effect is not correlated with $\mathbb{X}$, and
  3. a fully flexible parameterization of $\mathbb{X}$ is used.

Regression for Causal Inference

$$Y_i = \mu_0 + D_i(\mu_1 - \mu_0) + \{\nu_0 + D_i(\nu_1 - \nu_0)\}$$

$$Y_i = \mu_0 + D_i(\mu_1 - \mu_0) + \mathbb{X}_i\beta + \{\nu_0^* + D_i(\nu_1^* - \nu_0^*)\}$$

Regression for Causal Inference

- These conditions imply that:

Regression for Causal Inference

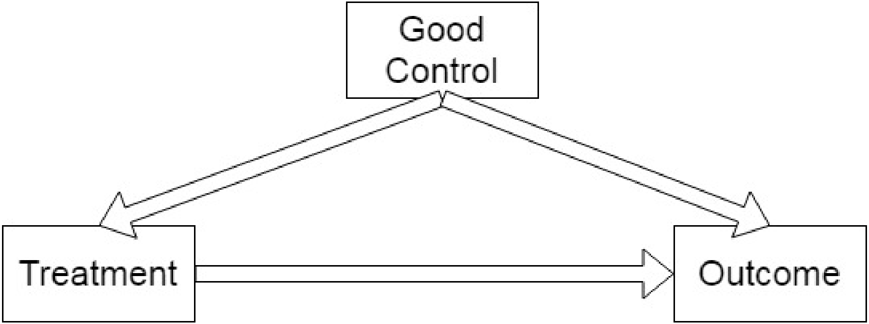
- These conditions imply that:
  1. No element of $\mathbb{X}$ is on any causal path from D to y, and

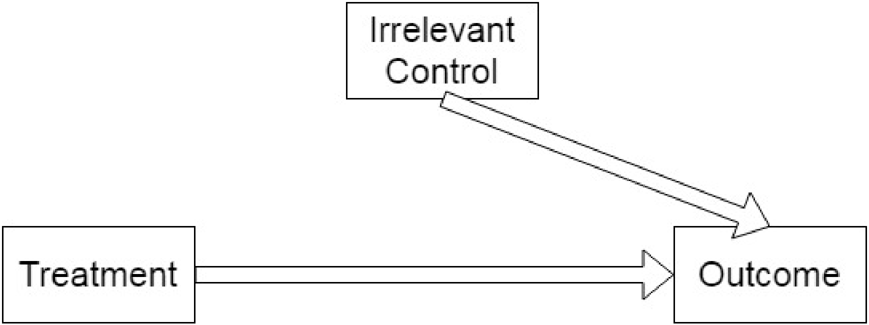
Regression for Causal Inference

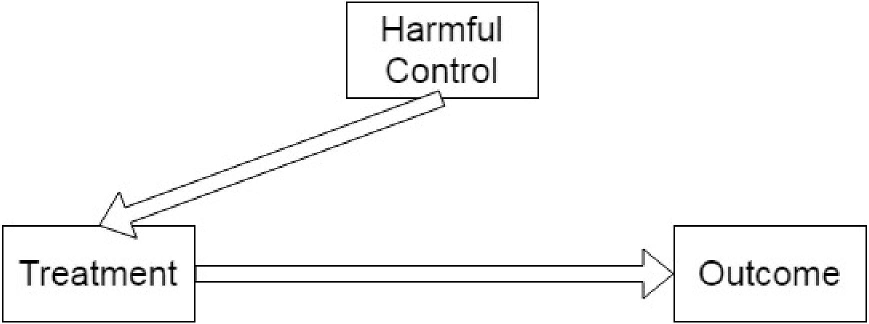
- These conditions imply that:
  1. No element of $\mathbb{X}$ is on any causal path from D to y, and
  2. no element of $\mathbb{X}$ is caused by D or any of its unmeasured causes *and* some other unmeasured cause of y, and

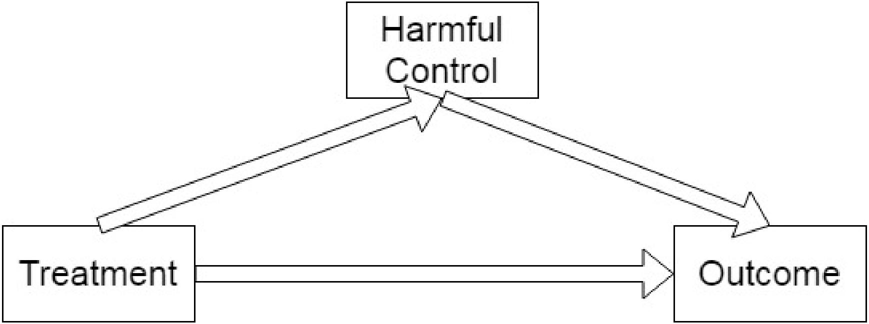
Regression for Causal Inference

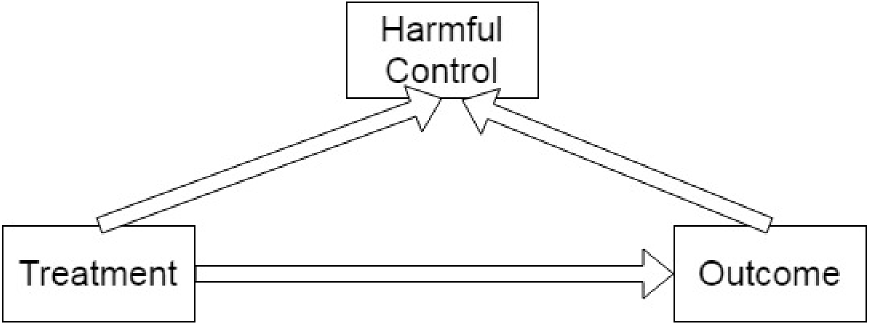
- These conditions imply that:
  1. No element of $\mathbb{X}$ is on any causal path from D to y, and
  2. no element of $\mathbb{X}$ is caused by D or any of its unmeasured causes *and* some other unmeasured cause of y, and
  3. all causes of D that are also causes of y have some element of $\mathbb{X}$ somewhere on the causal path from the unmeasured initial cause to either D or y.


Good Control
Treatment
Outcome


Irrelevant Control
Treatment
Outcome


Harmful
Control
Treatment
Outcome


Harmful
Control
Treatment
Outcome


Harmful
Control
Treatment
Outcome

Measurement Error

AJPS AMERICAN JOURNAL of POLITICAL SCIENCE

The Legacy of Political Violence across Generations

Noam Lupu Vanderbilt University
Leonid Peisakhin New York University–Abu Dhabi

Table A8. *Effects of first-generation victimization on third-generation attitudes and behaviors*

Dependent variable	Victimization effect	SE	Obs.	Families	R^2
Political identities					
In-group attachment	0.058*	0.022	957	298	0.015
Victimhood	0.080*	0.023	801	286	0.028
Threat perception	0.970*	0.047	951	298	0.009
Radicalism					
Support for radical Islam	0.009	0.058	645	267	0.000
Religiosity	0.049	0.031	988	298	0.005
Crimean Tatar issues					
Support for CT leaders	0.178*	0.050	724	285	0.032
Celebrate CT holiday	0.038*	0.019	947	297	0.009
Attitudes toward Russia					
Support for Chechen rebels	0.101*	0.021	639	270	0.046
Support for annexation	-0.074*	0.026	778	285	0.025
Pro-Russia vote choice	-0.089*	0.044	964	296	0.008
Political engagement					
Turnout	0.094*	0.042	974	298	0.010
Willingness to participate	0.120*	0.044	727	287	0.015
[Past participation]	0.080*	0.040	996	298	0.007

* $p<0.05$

Notes: Linear regression estimates with standard errors clustered by family. Constant terms not shown.

Table A9. *Effects of first-generation victimization on third-generation identities, attitudes, and behaviors, with pre-deportation controls*

Dependent variable	Victimization effect	SE	Obs.	Families	R^2
Political identities					
In-group attachment	0.052*	0.026	699	214	0.058
Victimhood	0.074*	0.026	605	210	0.120
Threat perception	0.163*	0.053	687	214	0.033
Radicalism					
Support for radical Islam	-0.002	0.066	467	191	0.059
Religiosity	0.033	0.037	715	214	0.033
Crimean Tatar issues					
Support for CT leaders	0.187*	0.061	528	207	0.057
Celebrate CT holiday	0.039	0.024	686	214	0.064
Attitudes toward Russia					
Support for Chechen rebels	0.101*	0.024	469	195	0.087
Support for annexation	-0.064*	0.028	569	205	0.107
Pro-Russia vote choice	-0.087	0.055	707	213	0.051
Political engagement					
Turnout	0.096	0.050	706	214	0.054
Willingness to participate	0.096	0.050	534	207	0.031

* $p<0.05$

Notes: Linear regression estimates with standard errors clustered by family. Constant terms and pre-deportation control variables (pre-Soviet household wealth, dekulakized, Soviet opposition, pre-Soviet religiosity, pre-deportation region, and deportation republic) not shown.

Table A10. *Effects of first-generation victimization on third-generation identities, attitudes, and behaviors, with third-generation demographic controls*

Dependent variable	Victimization effect	SE	Obs.	Families	R^2
Political identities					
In-group attachment	0.054^*	0.022	940	298	0.030
Victimhood	0.072^*	0.024	788	285	0.045
Threat perception	0.095^*	0.047	934	298	0.017
Radicalism					
Support for radical Islam	0.016	0.057	635	267	0.017
Religiosity	0.054	0.031	970	298	0.028
Crimean Tatar issues					
Support for CT leaders	0.164^*	0.050	710	285	0.045
Celebrate CT holiday	0.037	0.019	927	297	0.018
Attitudes toward Russia					
Support for Chechen rebels	0.102^*	0.021	628	269	0.058
Support for annexation	-0.071^*	0.026	765	285	0.038
Pro-Russia vote choice	-0.091^*	0.043	944	296	0.024
Political engagement					
Turnout	0.096^*	0.041	956	298	0.028
Willingness to participate	0.126^*	0.042	709	286	0.065

* $p<0.05$

Notes: Linear regression estimates with standard errors clustered by family. Constant terms and demographic control variables (wealth index, education, age, gender, and marital status) not shown.

Table A6. *Endogeneity tests*

Variable	Linear regression	Ordered probit
Pre-Soviet household wealth	0.077	0.079
	(0.068)	(0.072)
Dekulakized	0.040	0.064
	(0.152)	(0.158)
Soviet opposition	0.117	0.116
	(0.092)	(0.097)
Pre-Soviet religiosity	0.171	0.159
	(0.119)	(0.124)
Pre-deportation region		
Southwest (reference group)	--	--
	--	--
Southeast	-0.030	-0.039
	(0.235)	(0.248)
Northwest	0.137	0.172
	(0.248)	(0.238)
Northeast	0.046	0.038
	(0.457)	(0.510)
Deportation republic		
Uzbekistan (reference group)	--	--
	--	--
Other Central Asia	-0.269	-0.280
	(0.312)	(0.336)
Russia	-0.425	-0.479
	(0.252)	(0.285)
Observations	212	212
R^2/Pseudo-R^2	0.041	0.015

* $p<0.05$

Notes: Estimates with robust standard errors in parentheses. Constant term not shown.

Heterogeneity

- If the causal effect is not constant across all cases, regression will not give a consistent estimate of the average treatment effect.

Heterogeneity

- If the causal effect is not constant across all cases, regression will not give a consistent estimate of the average treatment effect.
- Instead, it estimates a covariance-adjusted weighted average of cases' treatment effects.

Aronow and Samii 2016

$$\hat{\beta} \xrightarrow{p} \frac{\mathrm{E}[w_i \tau_i]}{\mathrm{E}[w_i]}, \text{ where } w_i = (D_i - \mathrm{E}[D_i | X_i])^2$$

Aronow and Samii 2016

$$\mathrm{E}\left[w_i \,|\, X_i\right] = \mathrm{Var}\left[D_i \,|\, X_i\right],$$

Aronow and Samii 2016

FIGURE 1 **Example of nominal and effective samples from Jensen (2003)**

Note: On the left, the shading shows countries in the nominal sample for Jensen (2003) estimate of the effects of regime type on FDI. On the right, darker shading indicates that a country contributes more to the effective sample, based on the panel specification used in estimation.

Weights for Correct Causal Inference

It might sometimes be possible to get sample average treatment effects by weighting cases by $1/w_i$. But this may break down if some cases have weights equal to, or very close to, zero.

Weights for Correct Causal Inference

As long as all weights are strictly greater than zero, then the population causal inference can be derived from a regression where all control variables are centered around their means and where all possible interaction terms between control variables and the treatment are included.

Is Democracy Good for the Economy?

Democracy and Growth: A Case Study

1. Which dependent variable should we use?
   - For this literature, there isn't much ambiguity: economic growth is the outcome of interest.

Democracy and Growth: A Case Study

2. How should the dependent variable be measured, and perhaps transformed?

- Once again, there is little controversy here: annual measures or period-averages.

Democracy and Growth: A Case Study

2. How should the dependent variable be measured, and perhaps transformed?

- Once again, there is little controversy here: annual measures or period-averages.
- And yet...
    - Should we focus on short-term or long-term growth?
    - Should we use a transformation to limit the influence of extreme years, or not?

Democracy and Growth

3. Which independent variables should we use?

- Scholars have not achieved any degree of consensus regarding the proper set of control variables for a democracy-and-growth regression.

Democracy and Growth

3. Which independent variables should we use?

Table 7.4. Alternative sets of control variables

	Barro (1997)	Feng (1997)	Durham (1999)	Gasiorowski (2000)	Przeworski et al. (2000)	Kurzman et al. (2002)	Baum and Lake (2003)
Capital stock					X		
Education/literacy	X	X	X	X		X	X
GDP	X	X	X	X		X	X
Government consumption	X		X			X	
Inflation	X	X		X			
Investment		X	X	X		X	X
Labor supply				X	X		X
Life expectancy	X					X	X
Money supply growth				X			
Peaceful unrest				X			
Population growth	X		X			X	
Population size							X
Regime change		X					
Region dummies	X		X				
Rule of law	X						
Terms of trade	X						
Trade		X	X				
Violent unrest				X		X	
Wage growth				X			

Democracy and Growth

3. Which independent variables should we use?

- Consider education.

Democracy and Growth

3. Which independent variables should we use?

- Consider education.
  - Education, strangely enough, has a negative relationship with economic growth.

Democracy and Growth

3. Which independent variables should we use?

- Consider education.
  - Education, strangely enough, has a negative relationship with economic growth.
  - Education has a positive relationship with democracy.

Democracy and Growth

3. Which independent variables should we use?

- Consider education.
    - In a growth regression controlling only for lagged GDP, democracy has a significant coefficient of -0.00151.

Democracy and Growth

3. Which independent variables should we use?

- Consider education.
  - In a growth regression controlling only for lagged GDP, democracy has a significant coefficient of -0.00151.
  - In a similar regression also controlling for primary school attainment, democracy has a significant coefficient of 0.04371.

Democracy and Growth

3. Which independent variables should we use?

- Consider education.
  - Is education a cause or a consequence of democracy?

Democracy and Growth

4. What functional form should the equation take?

- In practice, the chosen functional form is almost always linear, but sometimes quadratic.
- Interaction terms are rarely considered.
- In particular, it is generally assumed that democracy has a single, universal causal effect on economic growth.

Democracy and Growth

5. Should observations be treated as independent? If not, then how?

Democracy and Growth

5. Should observations be treated as independent? If not, then how?

- Many studies assume total independence. Some assume independence conditional on constant country and/or period effects.

Democracy and Growth

5. Should observations be treated as independent? If not, then how?

- Many studies assume total independence. Some assume independence conditional on constant country and/or period effects.
- Models exist that could consider diffusion of economic growth patterns along geographic or trade-network lines, for example.

6. What kind of interpretations can the resulting coefficients reasonably be given?

- Strictly speaking, the coefficient estimates represent the slope, in one direction, of the best-fitting hyperplane through the observed data.

Democracy and Growth

6. What kind of interpretations can the resulting coefficients reasonably be given?

- Analysts virtually always interpret the coefficients as if they provided a description of causal counterfactuals applicable to each country in the data set.

Democracy and Growth

6. What kind of interpretations can the resulting coefficients reasonably be given?

- Less ambitious substantive interpretations are probably appropriate, but it is hard to know exactly what those interpretations should be. More theoretical development and more attention to mechanisms might help us guess what the various available coefficient estimates might mean.