UNIVERSITY OF MANNHEIM
School of Social Sciences

Quantitative Methods in Political Science: Linear Regression: Statistical Control & Causality

Thomas Gschwend | Oliver Rittmann | Viktoriia Semenova | David M. Grundmanns

Week 5 - 6 October 2021

What we plan to do today

Multiple Regression

Causality and Statistical Control

Example: Confounders in OLS Estimation

The Course

Roadmap

- Understand and model stochastic processes
- Understand statistical inference
- Implement it mathematically and learn how to estimate it
  - OLS
  - Maximum Likelihood
- Implement it using software
  - R
  - Basic programming skills

Multiple Regression

Regression Modeling: Prediction vs Causal Explanation

Multiple regression analysis allows us to add covariates $X_2, ..., X_k$ on top of X_1 in a regression of Y:

$$Y = \underbrace{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k}_{\hat{Y}} + \epsilon$$

This can serve two distinct purposes:

- Prediction: A bivariate regression of Y on X_1 does not yield accurate predictions of Y. We need additional covariates $X_2, ..., X_k$ to minimize the prediction error $\hat{Y} - Y$.
- Causal explanation: A bivariate regression of Y on X_1 does not yield an unbiased estimate of the true effect β_1 of X_1 on Y. We need to adjust for additional covariates to minimize the bias $\hat{\beta}_1 - \beta_1$.

Estimation is the same for both purposes. It is the rationale underlying model specification that differs.

The Multiple Linear Regression Model

- The general multiple linear regression model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \epsilon$$

  - The error term, ϵ, contains factors other than $X_1, \ldots, X_k$ that affect Y. We assume that all factors in the unobserved error term are uncorrelated with the explanatory variables.
- Estimation approach is the same as in the two-variable case, i.e., minimize the sum of squared residuals (and we get $k+1$ *normal equations*):

$$\min_{\hat{\beta}_0, \ldots, \hat{\beta}_k} \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \ldots - \hat{\beta}_k x_{ik})^2$$

Interpretation of Coefficients

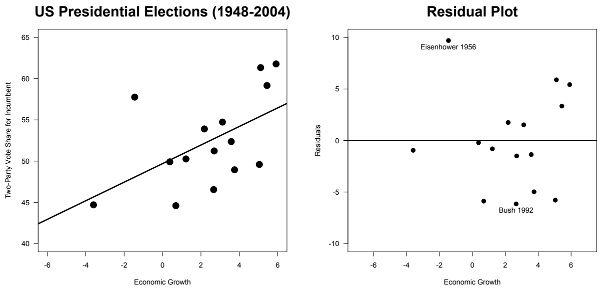
- Interpretation of regression coefficients:

$$\begin{aligned}
\hat{\beta}_0 &= \text{Predicted value of } Y \text{ when all } X\text{'s equal to zero.} \\
\hat{\beta}_1 &= \text{On average, a one-unit change in } X_1 \text{ leads to a } \hat{\beta}_1\text{-unit change in } Y, \\
&\quad \text{holding everything else (i.e., all other } X\text{'s) constant.} \\
&\dots \\
\hat{\beta}_k &= \text{On average, a one-unit change in } X_k \text{ leads to a } \hat{\beta}_k\text{-unit change in } Y, \\
&\quad \text{holding everything else (i.e., all other } X\text{'s) constant.}
\end{aligned}$$

- Note that in $k + 1$ dimensional space, a fitted multiple regression model no longer defines a line, but a hyperplane.
- This precludes visual representation of the data for $k > 2$.
- For $k = 2$, OLS means fitting a least squares plane that best fits the cloud of data points in a three-dimensional space.

Multiple Linear Regression: Example

- Let's go back to our US presidential election dataset.
- We used economic growth to predict the two-party vote share:



- Can you think of an alternative explanation for success of the presidential incumbent party?

Multiple Linear Regression: Example

- Let's consider presidential popularity in addition to economic growth.
- Popularity for presidents prior to election ranges between 31% and 74%.
- Our model then becomes:

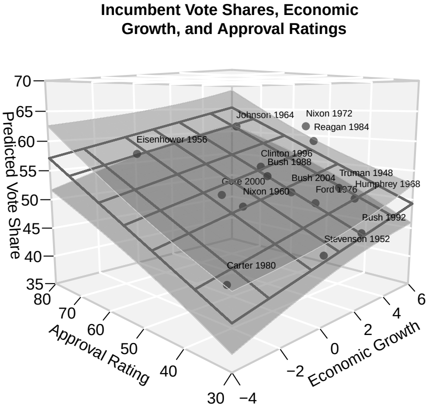
$$VoteShare = \beta_0 + \beta_1 Growth + \beta_2 Approval + \epsilon$$

- Results of the regression of vote share on growth and approval rating:

Variable	Estimate	SE
Constant	34.83	2.77
Growth	0.81	0.27
Approval	0.32	0.06
R^2 = 0.81		
Obs. = 15		

Multiple Linear Regression: Example

$$\widehat{VoteShare} = 34.83 + 0.81 \cdot Growth + 0.32 \cdot Approval$$



Statistical Control

So what is *statistical control* and how do we get an effect of approval of 0.32 independent of growth?

Our Strategy: We will purge away *Growth* from *VoteShare* and *Approval* and then regress them onto one another. This yields an effect of *Approval* on *VoteShare* which is independent of *Growth*.

- What is the effect of approval rating on vote share given growth?
  1. Let's start by running our simple regression of vote share on growth:

DV: VoteShare	Estimate	SE
Constant	49.70	1.75
Growth	1.13	0.49

- The residuals in this model are the part of *VoteShare* unexplained by growth.
- In other words, we remove the effect of growth from vote share variable.

Multiple Regression: Statistical Control

- Now, we also remove the effect of growth from approval rating
  2. Let's regress approval rating on growth:

DV: Approval	Estimate	SE
Constant	46.54	4.67
Growth	0.98	1.32

- The residuals in this model are the part of *Approval* unexplained by growth.
- In other words, we remove the effect of growth from approval rating.

Statistical Control

3. Now, we regress the residuals from step (1) on the residuals from step (2):

DV: Residuals 1	Estimate	SE
Constant	0.00	0.66
Residuals 2	0.32	0.06

- Having controlled for the effect of growth (by removing it from both the vote share and then approval rating), we can compute the unconfounded (or independent) effect of approval ratings on vote shares.
- This procedure gives us exactly the same coefficient (and SE) which we get in a multiple regression model!

DV: VoteShare	Estimate	SE
Constant	34.83	2.77
Growth	0.81	0.27
Approval	0.32	0.06

- We can now see how multiple regression coefficients can be said to statistically control for (or to be *independent of*) the effects of other variables in the model.

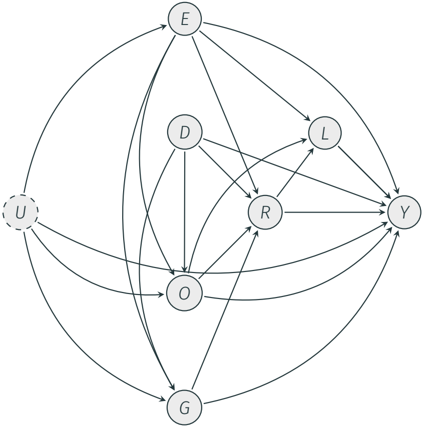
Causality and Statistical Control

Causal Questions

- Causal questions focus on the effect of a presumed cause X: "Does exposure to 'fake news' increase beliefs in false claims?"
- Causal questions typically imply comparisons of factual and unobservable counterfactual states for each observation: What would *i* believe had they (not) been exposed to fake news?
- The statistical solution: Compare average beliefs between those exposed and those not exposed to fake news.
- The problem: Structural differences between those exposed and those not exposed to fake news (e.g., due to self-selection).
- Design-based inference uses randomization; it puts the manipulation of X (*exposure to 'fake news'*) into the hands of the researcher.
- Model-based inference uses theoretical models about the relationship between variables $\rightarrow$ need to accurately translate the assumed theoretical model into a statistical model!

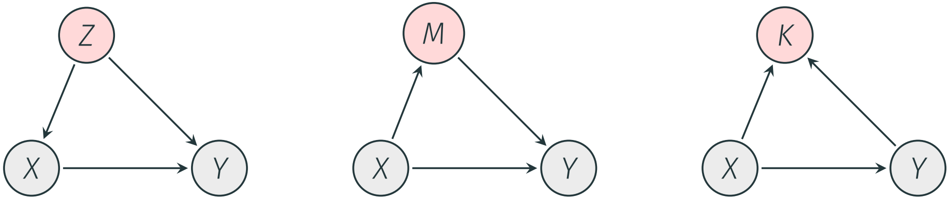
Causal Graphs (DAGs)

- Directed acyclical graphs (DAGs) allow us to express theoretical assumptions about causal relationships between (potentially correlated) variables given a theoretical model.
- Variables are depicted as nodes. They may be observable (solid nodes) or unobservable (dashed nodes).
- Variables are linked by directed arrows. These represent assumed directional relationships between variables.
- Arrows link ancestors to descendants.
- Note: The absence of arrows represents the assumption of no effect or association.
- Causal paths follow directed arrows that connect one node with another. A causal graph is acyclical if no causal path starting from an ancestor node runs back to this ancestor node.



Causal Graphs and Statistical Control

- Causal paths can be open or blocked. Statistical control can open or block paths, depending on whether a variable is:
  1. A confounder Z induces a non-causal association between X and Y. Controlling for a confounder blocks the non-causal path $X \leftarrow Z \rightarrow Y$.
  2. A mediator M captures a specific mechanism that translates a causal effect of X on Y. Controlling for a mediator blocks the *indirect* causal effect $X \rightarrow M \rightarrow Y$.
  3. A collider K is variable that is a descendant of both X and Y. Controlling for a collider opens the non-causal path $X \rightarrow K \leftarrow Y$.

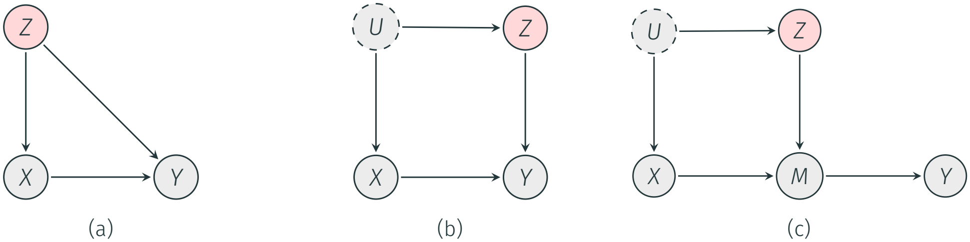


(a) Confounder (b) Mediator (c) Collider

Good Controls: Confounders

What (not) to control for?

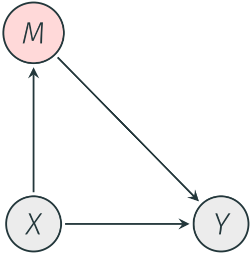
Rule 1: Control for confounders to block non-causal paths.



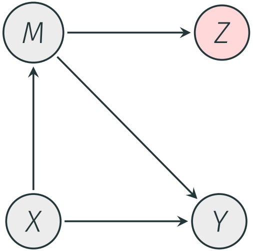
Note: Depending on the application, we may also speak of omitted variable bias or selection bias when referring to a model that fails to account for confounders.

(Mostly) Bad Controls: Mediators

Rule 2: Controlling for mediators blocks indirect causal effects $X \to M \to Y$. If your estimand is the total effect of X on Y, controlling for M will introduce post-treatment bias, i.e., you will only get the residual direct effect $X \to Y$. This also happens when you control for descendants of M.



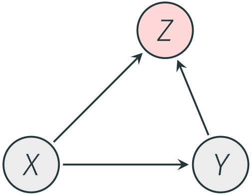
(a)



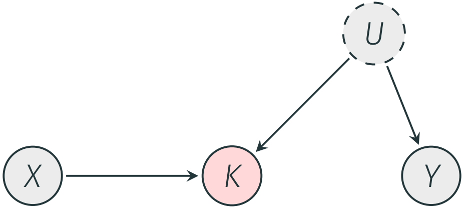
(b)

Bad Controls: Colliders

Rule 3: Controlling for a collider opens non-causal paths between X and Y. Controlling for variables that are descendants of both X and Y should thus be avoided.



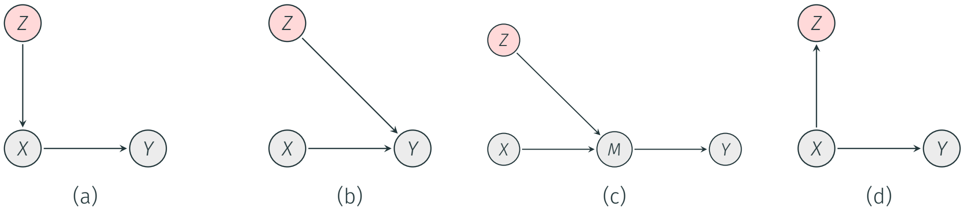
(a)



(b)

Neutral Controls

Rule 4: Neutral controls can be ancestors of X, of Y, or of mediators M. They can also be 'post-treatment variables' (i.e., descendants of X) as long as they neither open non-causal paths between X and Y nor block causal paths from X to Y. Neutral controls neither induce nor alleviate bias but may affect the precision of our estimates.



Bad Controls: Neutral Ancestors of X in Misspecified Models & Descendants of Y

Rule 5: Controlling for seemingly neutral ancestors of X will amplify bias in misspecified models.
Rule 6: Controlling for descendants of Y induces bias.

(a) (b)

Example: Confounders in OLS Estimation

Causality and Multiple Regression: An Example

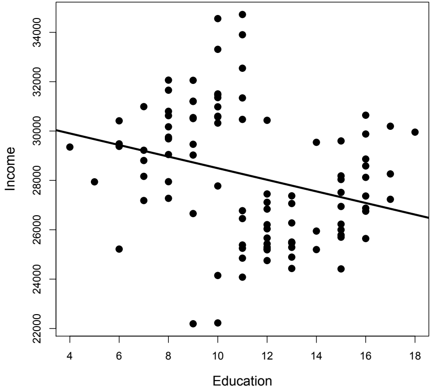
- Let's investigate income levels as a function of education and gender.
- Suppose we run two models and get the following results:

Variable	Model 1	Model 2
Education	-234.90	610.42
	(81.43)	(75.35)
Female		-6853.97
		(480.62)
Intercept	30843.22	25010.16
	(960.49)	(684.32)
Obs.	100	100

- Once we introduce "Female", the sign of the "Education" coefficient reverses.
- What is going on? and Which model is the correct one?

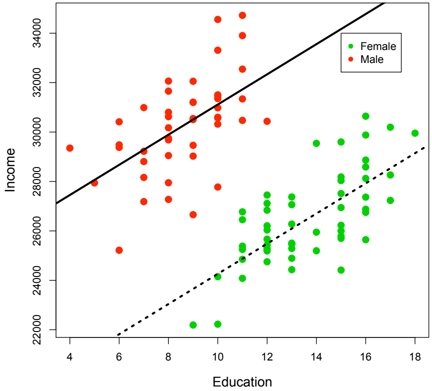
Causality and Multiple Regression: An Example

- As always, visual inspection of the data can be helpful.



Causality and Multiple Regression: An Example

- As always, visual inspection of the data can be helpful. Controlling for “Female” changes things. Within groups of “Female” the relationship of “Education” and “Income” is positive



Causality and Multiple Regression: An Example

- Gender and education are related. If we regress income on education alone, we arrive at a biased assessment of the effect of education on income because of an omitted variable, i.e., gender.
- Gender is a confounder and cause omitted variable bias in model 1.
- Since education and gender are correlated, and because gender is omitted from the model (1), part of the effect of gender on income is mistakenly attributed to education.
- The predicted income for females and males based on model (2) are:

$$\begin{aligned}\widehat{Income}_{Female} &= 25010.16 - 6583.97 \cdot 1 + 610.42 \cdot Education \\ &= 18426.19 + 610.42 \cdot Education\end{aligned}$$

$$\begin{aligned}\widehat{Income}_{Male} &= 25010.16 - 6583.97 \cdot 0 + 610.42 \cdot Education \\ &= 25010.16 + 610.42 \cdot Education\end{aligned}$$

Side Note: These Four Ways to Denote the Linear Model Formula...

1. Scalar form:
   $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i$ for all $i = 1, ..., N$
2. Row-vector form:
   $y_i = \mathrm{x}_\mathrm{i}' \beta + \epsilon_i$ for all $i = 1, ..., N$
3. Column-vector form:
   $\mathrm{y} = \beta_0 + \beta_1 \mathrm{x}_1 + \beta_2 \mathrm{x}_2 + \beta_3 \mathrm{x}_3 + \epsilon$
4. Matrix form:
   $\mathrm{y} = \mathrm{X}\beta + \epsilon$

...Are in Fact Equivalent

$$\underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}}_{N\times 1} = \underbrace{\begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N,1} & \cdots & x_{N,k} \end{bmatrix}}_{N\times(k+1)} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}}_{(k+1)\times 1} + \underbrace{\begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_N \end{bmatrix}}_{N\times 1}$$

$$\underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}}_{N\times 1} = \underbrace{\begin{bmatrix} \beta_0 \cdot 1 + \beta_1 \cdot x_{1,1} + \cdots + \beta_K \cdot x_{1,k} \\ \vdots \\ \beta_0 \cdot 1 + \beta_1 \cdot x_{N,1} + \cdots + \beta_K \cdot x_{N,k} \end{bmatrix}}_{N\times 1} + \underbrace{\begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_N \end{bmatrix}}_{N\times 1}$$