POLS 6481. Research Design and Quantitative Methods II

Lecture 9. ~~Rescaling,~~ Partialling, and Mediation

Readings: Wooldridge, *Introductory Econometrics 5e,* 3.2f

Iacobucci, *Mediation Analysis*, 1–23

Outline:

4A. ‘Partialling Out’

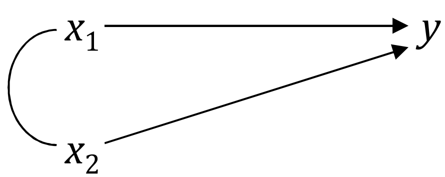
4B. Mediation

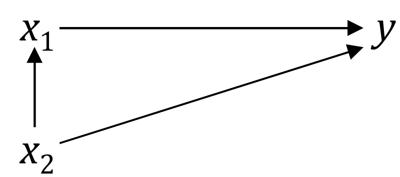
5. Example: *Campaign Expenditures and Congressional Elections*

POLS 6481. Research Design and Quantitative Methods II
Lecture 9. ~~Rescaling,~~ Partialling, and Mediation

4. Interference

There's a big difference between the following two hypothetical situations:





In the left figure, x_1 and x_2 are correlated

In the right figure, x_2 causes x_1 / x_1 mediates x_2

Identifying which situation you are confronting requires substantive knowledge and a theory

"Temporality" can help us distinguish between exogenous and endogenous variables

https://en.wikipedia.org/wiki/Bradford_Hill_criteria

"Granger causality" is a method of using time series data to distinguish causes and effects

https://en.wikipedia.org/wiki/Granger_causality

POLS 6481. Research Design and Quantitative Methods II
Lecture 9. ~~Rescaling,~~ Partialling, and Mediation
Readings: Wooldridge, *Introductory Econometrics 5e,* 3.2f

4A. 'Partialling Out'

Consider two-regressor linear model … $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$

… and suppose x_2 has a <u>causal</u> effect on x_1 $x_{1i} = \gamma_0 + \gamma_2 x_{2i} + e_i$

Solve for $\widehat{\gamma_1}$ and $\widehat{\gamma_0}$, and then compute $\widehat{x_{1i}} = \widehat{\gamma_0} + \widehat{\gamma_2} x_{2i}$

Then define the residual from this equation: $\widehat{r_{1i}} = x_{1i} - \widehat{x_{1i}}$

It must be the case that $\widehat{r_{1i}}$ is uncorrelated to x_{2i} (this can be checked easily).

It must be the case that $\widehat{r_{1i}}$ is conditionally uncorrelated to disturbance u_i in population regression function – conditional on x_{2i} – as long as the model is correctly specified.

Therefore, replacing x_{1i} with $\widehat{r_{1i}}$ in the sample regression function will yield unbiased estimates of slope coefficients and no multicollinearity – check the VIFs – and the coefficients are directly interpretable.

POLS 6481. Research Design and Quantitative Methods II

Lecture 9. ~~Rescaling,~~ Partialling, and Mediation

Readings: Iacobucci, *Mediation Analysis,* 1–23

4B. Mediation

Consider two-regressor linear model … $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$

… and suppose x_2 has a <u>causal</u> effect on x_1 $x_{1i} = \gamma_0 + \gamma_2 x_{2i} + e_i$

Estimate the following three models:

find $\widehat{\gamma_2}$ and $se(\widehat{\gamma_2})$ using: $\widehat{x_{1i}} = \widehat{\gamma_0} + \widehat{\gamma_2} x_{2i}$

find $\widehat{\beta_1}$ and $\widehat{\beta_2}$ and $se\left(\widehat{\beta_2}\right)$ using: $\widehat{y_i} = \widehat{\beta_0} + \widehat{\beta_1} x_{1i} + \widehat{\beta_2} x_{2i}$

Find the total effect of x_2 on y, i.e., $\widetilde{\beta_2}$, using: $\widetilde{y_i} = \widetilde{\beta_0} + \widetilde{\beta_2} x_{2i}$

The indirect effect of x_2 on y equals: $\widetilde{\beta_2} - \widehat{\beta_2} = \widehat{\gamma_2} \cdot \widehat{\beta_1}$

Sobel's z tests the statistical significance

of the indirect effect of x_1 on y:

$$z = \frac{\widehat{\gamma_2} \cdot \widehat{\beta_1}}{\sqrt{\widehat{\gamma_2}^2 \cdot var\left(\widehat{\beta_1}\right) + \widehat{\beta_1}^2 \cdot var(\widehat{\gamma_2}) + var\left(\widehat{\beta_1}\right) \cdot var(\widehat{\gamma_2})}}$$

POLS 6481. Research Design and Quantitative Methods II
Lecture 9. ~~Rescaling,~~ Partialling, and Mediation
Readings: Iacobucci, *Mediation Analysis,* 1–23

4B. Mediation

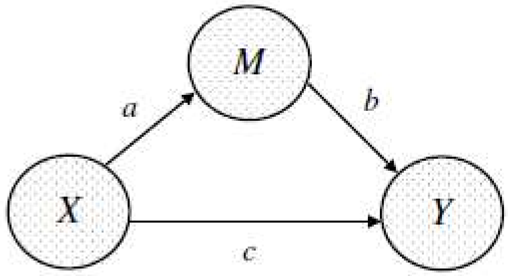


FIGURE 1 Simple, standard trivariate mediation: X = independent variable; M = mediator; Y = dependent variable.

$$M = \beta_1 + aX + \varepsilon_1$$

$$Y = \beta_2 + cX + \varepsilon_2$$

$$Y = \beta_3 + c'X + bM + \varepsilon_3$$

$$z = \frac{a \times b}{\sqrt{b^2 s_a^2 + a^2 s_b^2}}$$

5. Example: *Campaign Expenditures and Congressional Elections*

To investigate the effects of campaign spending, data were obtained on spending by the incumbent and challenger in 274 contested campaigns for the U.S. House of Representatives during the 2014 midterm election. Open-seat races and uncontested elections were dropped.

The dependent variable is *dv*, the Democratic House candidate's share of the vote.

The key independent variables are *ratio*, the natural log of the Democratic candidate's share of spending, and *dvp*, Barack Obama's share of the vote in the 2012 presidential election.

The dataset is *USHouseElections2014.csv*. The R script is *Lecture 9 House2014.R*.

Here is the correlation matrix:

	dv	*ratio*	*dvp*
dv	1.000		
ratio	.757	1.000	
dvp	.934	.685	1.000

5. Example: *Campaign Expenditures and Congressional Elections*

Because *dv* (y), *dvp* (x_2), and *ratio* (x_1) all are positively correlated, the simple regression of *dv* on *ratio* will exaggerate the effect of campaign spending on election outcomes.

In other words, the omitted variable bias in $\widetilde{\beta_1}$ is positive.

A multiple regression of *dv* on *ratio* and *dvp* will suffer from modest multicollinearity,

and some of the effect of *dvp* on *dv* will be inaccurately attributed to *ratio*.

If we use partialling to create $\widehat{r_{1i}}$, then the coefficient on *ratio* will decrease (from $\widetilde{\beta_1}$ to $\widehat{\beta_1}$)

while the coefficient on *dvp* will increase (from $\widehat{\beta_2}$ to $\widetilde{\beta_2}$)

meanwhile, multicollinearity vanishes (i.e., all VIF = 1)

5. Example: *Campaign Expenditures and Congressional Elections*

Because *dv* (y), *dvp* (x_2), and *ratio* (x_1) all are positively correlated, the simple regression of *dv* on *ratio* will exaggerate the effect of campaign spending on election outcomes.

In other words, the omitted variable bias in $\widetilde{\beta_1}$ is positive.

A multiple regression of *dv* on *ratio* and *dvp* will suffer from modest multicollinearity,

and some of the effect of *dvp* on *dv* will be inaccurately attributed to *ratio*.

If we use a mediation model, the indirect effect will equal $\widetilde{\beta_2} - \widehat{\beta_2}$;
Sobel's z statistic will show that this indirect effect is statistically significant.

The total effect of *dvp* on *dv* is equal to: $\widetilde{\beta_2} = \widehat{\beta_2} + \left[\widehat{\gamma_2} \cdot \widehat{\beta_1}\right]$

$= \text{direct effect} + \text{indirect/mediated effect}$