Inference and interaction effects

Felipe Balcazar

NYU

October, 2022

NEW YORK UNIVERSITY

Modernization hypothesis

$$\begin{tikzcd}
\text{Conflict} \arrow[d] \arrow[dr] & \\
\text{GDP} \arrow[r] & \text{Democracy}
\end{tikzcd}$$

- Does economic prosperity lead to democratization?

$$Dem_{it} = \alpha + \beta GDP_{it} + \delta X_{it} + \mu_i + \varepsilon_{it}.$$

How can we can make questions

- Is the the effect of the treatment (main independent variable) different from zero?
  - *Null hypothesis* (H_0): $\beta = 0$.
  - The alternative (H_a): $\beta \neq 0$.

How can we can make questions

- Is the the effect of the treatment (main independent variable) different from zero?
  - *Null hypothesis* (H_0): $\beta = 0$.
  - The alternative (H_a): $\beta \neq 0$.

		Null hypothesis	
		TRUE	FALSE
Findings	Reject null	Type I error (α)	Correct decision
	Accept null	Correct decision	Type II error (β)

- α is what is called "significance." Probability of type I error.
- β is what is called "power." Probability of type II error.

How can we can make questions

- Is the the effect of the treatment (main independent variable) different from zero?
  - *Null hypothesis* (H_0): $\beta = 0$.
  - The alternative (H_a): $\beta \neq 0$.

		Null hypothesis	
		TRUE	FALSE
Findings	Reject null	False positive (α)	Correct decision
	Accept null	Correct decision	False negative (β)

- α is what is called "significance." Probability of false positive.
- β is what is called "power." Probability of false negative.

How can we can make questions

- We do not want to reject the null when it is true: *type I error* - i.e., we do not want false positives.

- The probability for this must be small, less than α.
  - α is the statistical significance (usually 0.05 or 5%).

- We do not want to accept the null when it is false: *type II error* - i.e., we do not want false negatives.
  - If the number of unit of observations you have is too small (below 42) you will be underpowered.
  - Important: I talk about the number of unit of observations, not the number of observations!

Is the effect statistically significant?

- To test this hypothesis we compute a t-statistic:

$$\hat{t} = \frac{\widehat{\beta} - H_0}{SE}.$$

Is the effect statistically significant?

- To test this hypothesis we compute a t-statistic:

$$\hat{t} = \frac{\widehat{\beta} - 0}{SE}.$$

Is the effect statistically significant?

- To test this hypothesis we compute a t-statistic:

$$\hat{t} = \frac{\widehat{\beta} - 0}{SE}.$$

- Then we ask $Pr(|t| \geq |\hat{t}|) \equiv$ p-value.
  - p-value is the probability of type I error.
- If p-value $< \alpha$, we reject the null hypothesis.
  - Don't worry, the computer does all of this for you.
  - You probably have seen stars in papers.

Is the effect statistically significant?

- To test this hypothesis we compute a t-statistic:

$$\hat{t} = \frac{\widehat{\beta} - 0}{SE}.$$

- Then we ask $Pr(|t| \geq |\hat{t}|) \equiv$ p-value.
  - p-value is the probability of type I error.
- If p-value $< \alpha$, we reject the null hypothesis.
  - Don't worry, the computer does all of this for you.
  - You probably have seen stars in papers.

Eyeballing the answer to the same question

- Use a confidence interval!
- Focus on the statistical significance (α),
- So $(1-\alpha) \times 100$ is the confidence that H_0 is true.
  - 90% confidence when $\alpha = 0.1$.
    $$H_0 \in [\widehat{\beta} - 1.65SE; \widehat{\beta} + 1.65SE]$$
  - 95% confidence when $\alpha = 0.05$.
    $$H_0 \in [\widehat{\beta} - 1.96SE; \widehat{\beta} + 1.96SE]$$
  - 99% confidence when $\alpha = 0.01$.
    $$H_0 \in [\widehat{\beta} - 2.56SE; \widehat{\beta} + 2.56SE]$$

Statistical significance

- *Null hypothesis* (H_0): $\beta = 0.0$.
- The alternative (H_a): $\beta \neq 0.0$.

- If we reject the null hypothesis then $\widehat{\beta}$ is statistically significant - that is statistically different from zero.
  - 90% confident that is not zero if

$$0 \notin [\widehat{\beta} - 1.65SE; \widehat{\beta} + 1.65SE]$$

  - 95% confident that is not zero if

$$0 \notin [\widehat{\beta} - 1.96SE; \widehat{\beta} + 1.96SE]$$

  - 99% confident that is not zero if

$$0 \notin [\widehat{\beta} - 2.56SE; \widehat{\beta} + 2.56SE]$$

Statistical significance (II)

- *Null hypothesis* (H_0): $\beta = 0.0$.
- The alternative (H_a): $\beta \neq 0.0$.

- Only for this type of hypothesis we can also check the absolute value of the $\widehat{\beta}$ against the SE:
  - 90% confident that is not zero if
$$\frac{|\widehat{\beta}|}{1.65} > SE$$
  - 95% confident that is not zero if
$$\frac{|\widehat{\beta}|}{1.96} > SE$$
  - 99% confident that is not zero if
$$\frac{|\widehat{\beta}|}{2.56} > SE$$

Robust statistical inference

- Gauss-Markov assumptions: $Var(\varepsilon_{it}) = \sigma^2$.
  - Homoscedasticity - necessary assumption for BLUE.
  - Unbiased estimator for standard error.
- Not satisfied when $Var(e) = g(X)$, where $g(\cdot)$ is some function. This happens almost always!
  - Heteroscedasticity.
  - Standard error is biased! (Usually deflated.)
- Solutions:
  - Robust standard errors.
  - Clustered robust standard errors
    - Some treated units may react similarly (groups).

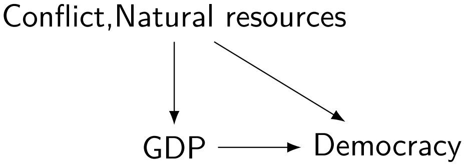
Robust statistical inference

- Gauss-Markov assumptions: $Var(\varepsilon_{it}) = \sigma^2$.
  - Homoscedasticity - necessary assumption for BLUE.
  - Unbiased estimator for standard error.
- Not satisfied when $Var(e) = g(X)$, where $g(\cdot)$ is some function. This happens almost always!
  - Heteroscedasticity.
  - Standard error is biased! (Usually deflated.)
- Solutions:
  - Robust standard errors.
  - Clustered robust standard errors
    - Some treated units may react similarly (groups).

Robust statistical inference

- Gauss-Markov assumptions: $Var(\varepsilon_{it}) = \sigma^2$.
  - Homoscedasticity - necessary assumption for BLUE.
  - Unbiased estimator for standard error.
- Not satisfied when $Var(e) = g(X)$, where $g(\cdot)$ is some function. This happens almost always!
  - Heteroscedasticity.
  - Standard error is biased! (Usually deflated.)
- Solutions:
  - Robust standard errors.
  - Clustered robust standard errors
    - Some treated units may react similarly (groups).

Modernization hypothesis



- Does economic prosperity lead to democratization?

Interaction effects

- The treatment may affect some units but not other! Or it may affect them differently!

Interaction effects

- The treatment may affect some units but not other! Or it may affect them differently!
- Countries rich in natural resources may experience a resource curse.

Interaction effects

- The treatment may affect some units but not other! Or it may affect them differently!
- Countries rich in natural resources may experience a resource curse.
  - Oil can increase incentives to capture the state.
  - Absence of natural resources may create incentives to increase productivity.

- The treatment may affect some units but not other! Or it may affect them differently!
- Countries rich in natural resources may experience a resource curse.
    - Oil can increase incentives to capture the state.
    - Absence of natural resources may create incentives to increase productivity.
- Therefore the treatment can have *heterogenous* effects!

Interaction effects: in theory

Consider

$$D_{it} = \begin{cases} 1 & \text{if country } i \text{ has high GDP,} \\ 0 & \text{otherwise.} \end{cases}$$

and

$$X_{it} = \begin{cases} 1 & \text{if country } i \text{ big oil producer,} \\ 0 & \text{otherwise.} \end{cases}$$

Interaction effects: in theory

Consider

$$D_{it} = \begin{cases} 1 & \text{if country } i \text{ has high GDP,} \\ 0 & \text{otherwise.} \end{cases}$$

and

$$X_{it} = \begin{cases} 1 & \text{if country } i \text{ big oil producer,} \\ 0 & \text{otherwise.} \end{cases}$$

- The effect of having high GDP for countries with natural resources:

$$E(Y_i|D_i = 1, X_i = 1) - E(Y_i|D_i = 0, X_i = 1)$$

- The effect of having high GDP for countries without natural resources:

$$E(Y_i|D_i = 1, X_i = 0) - E(Y_i|D_i = 0, X_i = 0)$$

Consider

$$D_{it} = \begin{cases} 1 & \text{if country } i \text{ has high GDP,} \\ 0 & \text{otherwise.} \end{cases}$$

and

$$X_{it} = \begin{cases} 1 & \text{if country } i \text{ big oil producer,} \\ 0 & \text{otherwise.} \end{cases}$$

- The effect of having high GDP for countries with natural resources:

$$E(Y_i|D_i = 1, X_i = 1) - E(Y_i|D_i = 0, X_i = 1)$$

- The effect of having high GDP for countries without natural resources:

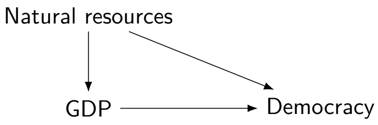
$$E(Y_i|D_i = 1, X_i = 0) - E(Y_i|D_i = 0, X_i = 0)$$

We say that X moderates the effect of the treatment.

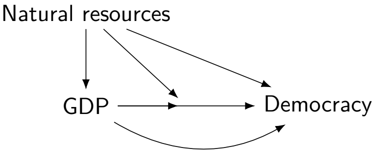
- When there is a variable X that moderates the effect of the treatment, we call this element a *moderator*.

Interaction effects: controls and confounders

- When there is a variable X that moderates the effect of the treatment, we call this element a *moderator*.



- When there is a variable X that moderates the effect of the treatment, we call this element a *moderator*.



Interaction effects: regression form

$$Y_{it} = \alpha + \beta_1 D_{it} + \beta_2 X_{it} + \beta_3 D \times X_{it} + \mu_i + \gamma_t + \varepsilon_{it}.$$

- β_1 is the effect of the treatment conditional on $X = 0$:

$$E(Y_{it}|D_{it} = 1, X_{it} = 0) - E(Y_{it}|D_{it} = 0, X_{it} = 0)$$

- $\beta_1 + \beta_3$ is the effect of the treatment conditional on $X = 1$:

$$E(Y_{it}|D_{it} = 1, X_{it} = 1) - E(Y_{it}|D_{it} = 0, X_{it} = 1)$$

- β_3 is the additional effect of the treatment for $X = 1$.

$$Y_{it} = \alpha + \beta_1 D_{it} + \beta_2 X_{it} + \beta_3 D \times X_{it} + \mu_i + \gamma_t + \varepsilon_{it}.$$

- β_1 is the effect of the treatment conditional on $X = 0$:

$$E(Y_{it}|D_{it} = 1, X_{it} = 0) - E(Y_{it}|D_{it} = 0, X_{it} = 0)$$

- $\beta_1 + \beta_3$ is the effect of the treatment conditional on $X = 1$:

$$E(Y_{it}|D_{it} = 1, X_{it} = 1) - E(Y_{it}|D_{it} = 0, X_{it} = 1)$$

- β_3 is the additional effect of the treatment for $X = 1$.

Interaction effects: regression form

$$Y_{it} = \alpha + \beta_1 D_{it} + \beta_2 X_{it} + \beta_3 D \times X_{it} + \mu_i + \gamma_t + \varepsilon_{it}.$$

- β_1 is the effect of the treatment conditional on $X = 0$:

$$E(Y_{it}|D_{it} = 1, X_{it} = 0) - E(Y_{it}|D_{it} = 0, X_{it} = 0)$$

- $\beta_1 + \beta_3$ is the effect of the treatment conditional on $X = 1$:

$$E(Y_{it}|D_{it} = 1, X_{it} = 1) - E(Y_{it}|D_{it} = 0, X_{it} = 1)$$

- β_3 is the additional effect of the treatment for $X = 1$.

$$Y_{it} = \alpha + \beta_1 D_{it} + \beta_2 X_{it} + \beta_3 D \times X_{it} + \mu_i + \gamma_t + \varepsilon_{it}.$$

- β_1 is the effect of the treatment conditional on $X = 0$:

$$E(Y_{it}|D_{it} = 1, X_{it} = 0) - E(Y_{it}|D_{it} = 0, X_{it} = 0)$$

- $\beta_1 + \beta_3$ is the effect of the treatment conditional on $X = 1$:

$$E(Y_{it}|D_{it} = 1, X_{it} = 1) - E(Y_{it}|D_{it} = 0, X_{it} = 1)$$

- β_3 is the additional effect of the treatment for $X = 1$.

Interaction effects: interpretation

$$Y_{it} = \alpha + \beta_1 D_{it} + \beta_2 X_{it} + \beta_3 D \times X_{it} + \mu_i + \gamma_t + \varepsilon_{it}.$$

- X need not to be binary!
- The key to interpret an interaction effect:
  - Interpret $\widehat{\beta}_1$ directly, thus we hold X at zero ($X = 0$).
  - Hold X constant then add up: $\widehat{\beta}_1 + \widehat{\beta}_3 X$.

Interaction effects: interpretation

$$Y_{it} = \alpha + \beta_1 D_{it} + \beta_2 X_{it} + \beta_3 D \times X_{it} + \mu_i + \gamma_t + \varepsilon_{it}.$$

- X need not to be binary!
- The key to interpret an interaction effect:
  - Interpret $\widehat{\beta}_1$ directly, thus we hold X at zero ($X = 0$).
  - Hold X constant then add up: $\widehat{\beta}_1 + \widehat{\beta}_3 X$.

Interaction effects: interpretation

$$Y_{it} = \alpha + \beta_1 D_{it} + \beta_2 X_{it} + \beta_3 D \times X_{it} + \mu_i + \gamma_t + \varepsilon_{it}.$$

- X need not to be binary!
- The key to interpret an interaction effect:
  - Interpret $\widehat{\beta}_1$ directly, thus we hold X at zero ($X = 0$).
  - Hold X constant then add up: $\widehat{\beta}_1 + \widehat{\beta}_3 X$.

$$Y_{it} = \alpha + \beta_1 D_{it} + \beta_2 X_{it} + \beta_3 D \times X_{it} + \mu_i + \gamma_t + \varepsilon_{it}.$$

- X need not to be binary!
- The key to interpret an interaction effect:
  - Interpret $\widehat{\beta}_1$ directly, thus we hold X at zero ($X = 0$).
  - Hold X constant then add up: $\widehat{\beta}_1 + \widehat{\beta}_3 X$.

Interaction effects: regression form

$$\begin{aligned} Dem_{it} = &\alpha + \beta_1 GDP_{it} + \beta_3 GDP_{it} \times Oil_{it} \\ &+ \beta_2 Oil_{it} + \delta Conflict_{it} + \varepsilon_{it}. \end{aligned}$$

- β_1 is the effect of GDP conditional on $Oil = 0$:
- $\beta_1 + \beta_3$ is the effect of the treatment conditional on $Oil = 1$:
- β_3 is the additional effect of GDP for a country with oil.

- We can use an interaction to test for robustness.
  - That is one relevant use for interaction effects.

Interaction effects: why?

- We can use an interaction to test for robustness.
  - That is one relevant use for interaction effects.
- For our purposes also:
  - X does not need to be all else equal to interpret β_1 and β_3 as long as D is "all es equal."
  - If assignment of D is all else equal this opens many potential questions.
    - Does the the moderator, moderates the effect of my treatment?
    - These questions require theoretical knowledge and of the case being studied.

Interaction effects: why?

- We can use an interaction to test for robustness.
  - That is one relevant use for interaction effects.
- For our purposes also:
  - X does not need to be all else equal to interpret β_1 and β_3 as long as D is "all es equal."
  - If assignment of D is all else equal this opens many potential questions.
    - Does the the moderator, moderates the effect of my treatment?
    - These questions require theoretical knowledge and of the case being studied.

The blessing of bad geography? (Nunn and Puga, 2012)

- Rugged terrain can be a curse:
  - Hard to develop agriculture.
  - Higher costs of transportation.
  - Costlier to tax and to provide public goods.
  - Breeding ground for militias.
- But it can also be a "blessing":
  - It can reduce expropriation by despotic ruler.
  - In the past, it protected people from being enslaved.
- For African people fleeing the slave trade, rugged terrain was a positive advantage.

Ruggedness was a blessing... or was it?

TABLE 1.— THE DIFFERENTIAL EFFECT OF RUGGEDNESS IN AFRICA

	Dependent Variable: Log Real GDP per Person, 2000					
	(1)	(2)	(3)	(4)	(5)	(6)
Ruggedness	−0.203	−0.196	−0.203	−0.243	−0.193	−0.231
	(0.093)**	(0.094)**	(0.094)**	(0.092)***	(0.081)**	(0.077)***
Ruggedness × I^{Africa}	0.393	0.404	0.406	0.414	0.302	0.321
	(0.144)***	(0.146)***	(0.138)***	(0.157)***	(0.130)**	(0.127)**
I^{Africa}	−1.948	−2.014	−1.707	−2.066	−1.615	−1.562
	(0.220)***	(0.222)***	(0.325)***	(0.324)***	(0.295)***	(0.415)***
Diamonds		0.017				0.028
		(0.012)				(0.010)***
Diamonds × I^{Africa}		−0.014				−0.026
		(0.012)				(0.011)**
% Fertile soil			0.000			−0.002
			(0.003)			(0.003)
% Fertile soil × I^{Africa}			−0.008			−0.009
			(0.006)			(0.007)
% Tropical climate				−0.007		−0.009
				(0.002)***		(0.002)***
% Tropical climate × I^{Africa}				0.004		0.006
				(0.004)		(0.004)
Distance to coast					−0.657	−1.039
					(0.177)***	(0.193)***
Distance to coast × I^{Africa}					−0.291	−0.194
					(0.360)	(0.386)
Constant	9.223	9.204	9.221	9.514	9.388	9.959
	(0.143)***	(0.148)***	(0.200)***	(0.164)***	(0.134)***	(0.195)***
Observations	170	170	170	170	170	170
R^2	0.357	0.367	0.363	0.405	0.421	0.537

Coefficients are reported with robust standard errors in brackets. ***, **, and * indicate significance at the 1%, 5%, and 10% levels.

Things to keep in mind (Very important!)

- An interaction variable X must be pretreatment.
- It does not matter if X suffers from omitted variables bias.
  - If D is as good as random the variation we exploit comes from D, what X does is to condition your treatment effect on a value of X.
- This means that you can only interpret $\widehat{\beta}_1$, $\widehat{\beta}_3 \times X$ and $\widehat{\beta}_1 + \widehat{\beta}_3 \times X$ causally.
- If D suffers from selection/omitted variable bias, then $\widehat{\beta}_1$, $\widehat{\beta}_3 \times X$ and $\widehat{\beta}_1 + \widehat{\beta}_3 \times X$ are also biased.
- If X is as good as random, but not D, then only $\widehat{\beta}_3 \times X$ can be interpreted causally.

Things to keep in mind (Very important!)

- An interaction variable X must be pretreatment.
- It does not matter if X suffers from omitted variables bias.
    - If D is as good as random the variation we exploit comes from D, what X does is to condition your treatment effect on a value of X.
- This means that you can only interpret $\widehat{\beta}_1$, $\widehat{\beta}_3 \times X$ and $\widehat{\beta}_1 + \widehat{\beta}_3 \times X$ causally.
- If D suffers from selection/omitted variable bias, then $\widehat{\beta}_1$, $\widehat{\beta}_3 \times X$ and $\widehat{\beta}_1 + \widehat{\beta}_3 \times X$ are also biased.
- If X is as good as random, but not D, then only $\widehat{\beta}_3 \times X$ can be interpreted causally.

Things to keep in mind (Very important!)

- An interaction variable X must be pretreatment.
- It does not matter if X suffers from omitted variables bias.
  - If D is as good as random the variation we exploit comes from D, what X does is to condition your treatment effect on a value of X.
- This means that you can only interpret $\widehat{\beta}_1$, $\widehat{\beta}_3 \times X$ and $\widehat{\beta}_1 + \widehat{\beta}_3 \times X$ causally.
- If D suffers from selection/omitted variable bias, then $\widehat{\beta}_1$, $\widehat{\beta}_3 \times X$ and $\widehat{\beta}_1 + \widehat{\beta}_3 \times X$ are also biased.
- If X is as good as random, but not D, then only $\widehat{\beta}_3 \times X$ can be interpreted causally.

Things to keep in mind (Very important!)

- An interaction variable X must be pretreatment.
- It does not matter if X suffers from omitted variables bias.
  - If D is as good as random the variation we exploit comes from D, what X does is to condition your treatment effect on a value of X.
- This means that you can only interpret $\widehat{\beta}_1$, $\widehat{\beta}_3 \times X$ and $\widehat{\beta}_1 + \widehat{\beta}_3 \times X$ causally.
- If D suffers from selection/omitted variable bias, then $\widehat{\beta}_1$, $\widehat{\beta}_3 \times X$ and $\widehat{\beta}_1 + \widehat{\beta}_3 \times X$ are also biased.
- If X is as good as random, but not D, then only $\widehat{\beta}_3 \times X$ can be interpreted causally.

Things to keep in mind (Very important!)

- An interaction variable X must be pretreatment.
- It does not matter if X suffers from omitted variables bias.
  - If D is as good as random the variation we exploit comes from D, what X does is to condition your treatment effect on a value of X.
- This means that you can only interpret $\widehat{\beta}_1$, $\widehat{\beta}_3 \times X$ and $\widehat{\beta}_1 + \widehat{\beta}_3 \times X$ causally.
- If D suffers from selection/omitted variable bias, then $\widehat{\beta}_1$, $\widehat{\beta}_3 \times X$ and $\widehat{\beta}_1 + \widehat{\beta}_3 \times X$ are also biased.
- If X is as good as random, but not D, then only $\widehat{\beta}_3 \times X$ can be interpreted causally.

Things to keep in mind (Very important!)

- An interaction variable X must be pretreatment.
- It does not matter if X suffers from omitted variables bias.
  - If D is as good as random the variation we exploit comes from D, what X does is to condition your treatment effect on a value of X.
- This means that you can only interpret $\widehat{\beta}_1$, $\widehat{\beta}_3 \times X$ and $\widehat{\beta}_1 + \widehat{\beta}_3 \times X$ causally.
- If D suffers from selection/omitted variable bias, then $\widehat{\beta}_1$, $\widehat{\beta}_3 \times X$ and $\widehat{\beta}_1 + \widehat{\beta}_3 \times X$ are also biased.
- If X is as good as random, but not D, then only $\widehat{\beta}_3 \times X$ can be interpreted causally.