Regression and controls

Felipe Balcazar

NYU

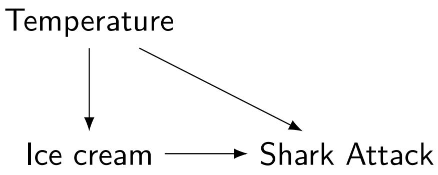
October, 2021

Recall: shark attacks

Ice cream $\longrightarrow$ Shark Attack

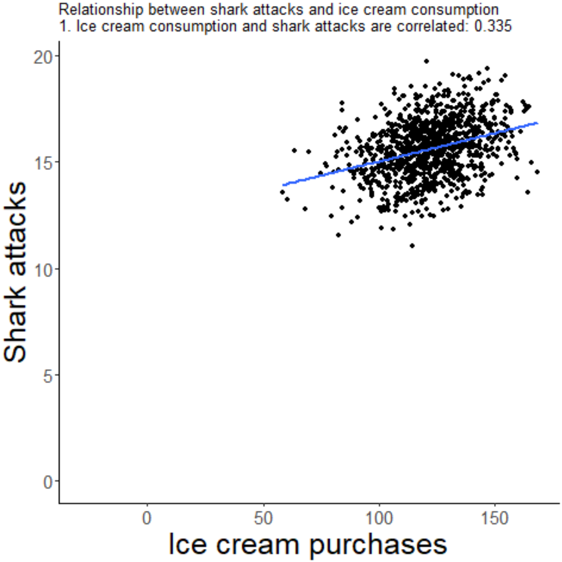
- We ask whether icecream causes shark attacks.
- But the temperature causes both.

Recall: shark attacks



- We ask whether icecream causes shark attacks.
- But the temperature causes both.

Correlation and a regression line



What did we see in the previous graph?

- The graph suggests a positive correlation to begin with.
- The regression line had a positive slope:

$$SharkAttacks_i = \widehat{\alpha} + \widehat{\beta} IceCream_i, \quad \widehat{\beta} > 0$$

- Once we control for confounders the slope is zero; i.e., no actual relationship (i.e., $\beta = 0$).
- Does the slope of the regression line capture the average treatment effect?

What did we see in the previous graph?

- The graph suggests a positive correlation to begin with.
- The regression line had a positive slope:

$$SharkAttacks_i = \widehat{\alpha} + \widehat{\beta}IceCream_i, \quad \widehat{\beta} > 0$$

- Once we control for confounders the slope is zero; i.e., no actual relationship (i.e., $\beta = 0$).
- Does the slope of the regression line capture the average treatment effect?

What did we see in the previous graph?

- The graph suggests a positive correlation to begin with.
- The regression line had a positive slope:

$$SharkAttacks_i = \widehat{\alpha} + \widehat{\beta} IceCream_i, \quad \widehat{\beta} > 0$$

- Once we control for confounders the slope is zero; i.e., no actual relationship (i.e., $\beta = 0$).
- Does the slope of the regression line capture the average treatment effect?

What did we see in the previous graph?

- The graph suggests a positive correlation to begin with.
- The regression line had a positive slope:

$$SharkAttacks_i = \widehat{\alpha} + \widehat{\beta} IceCream_i, \quad \widehat{\beta} > 0$$

- Once we control for confounders the slope is zero; i.e., no actual relationship (i.e., $\beta = 0$).
- Does the slope of the regression line capture the average treatment effect?

$$Y_i = \alpha + \beta D_i + \varepsilon_i.$$

- Y_i is the *outcome* for individual i.
- D_i is the the value of the *treatment* for i.
    - i is the *unit of observation*—the level at which your outcome is defined.

The method for causal inference: regression

$$Y_i = \alpha + \beta D_i + \varepsilon_i.$$

- Y_i is the *outcome* for individual i.
- D_i is the the value of the *treatment* for i.
  - i is the *unit of observation*—the level at which your outcome is defined.
- α is the intercept, the constant.
- β is the slope; the *effect of the treatment*.
- ε is an error term with mean zero $E(\varepsilon) = 0$; in it there could be other things that explain Y and random noise.

The method for causal inference: regression

$$Y_i = \alpha + \beta D_i + \varepsilon_i.$$

- Y_i is the *outcome* for individual i.
- D_i is the the value of the *treatment* for i.
  - i is the *unit of observation*—the level at which your outcome is defined.

- α is the intercept, the constant.
- β is the slope; the *effect of the treatment*.
- ε is an error term with mean zero $E(\varepsilon) = 0$; in it there could be other things that explain Y and random noise.

- Assume for now nothing in ε_i affects D_i. That is $\varepsilon_i | D_i = \varepsilon_i$.

The method for causal inference: regression

$$Y_i = \alpha + \beta D_i + \varepsilon_i.$$

- Y_i is the *outcome* for individual i.
- D_i is the the value of the *treatment* for i.
  - i is the *unit of observation*—the level at which your outcome is defined.
- α is the intercept, the constant.
- β is the slope; the *effect of the treatment.*
- ε is an error term with mean zero $E(\varepsilon) = 0$; in it there could be other things that explain Y and random noise.
- Assume for now nothing in ε_i affects D_i. That is $\varepsilon_i | D_i = \varepsilon_i$.

$$Y_i = \alpha + \beta D_i + \varepsilon_i.$$

- Y_i is the *outcome* for individual i.
- D_i is the the value of the *treatment* for i.
  - i is the *unit of observation*—the level at which your outcome is defined.
- α is the intercept, the constant.
- β is the slope; the *effect of the treatment*.
- ε is an error term with mean zero $E(\varepsilon) = 0$; in it there could be other things that explain Y and random noise.
- Assume for now nothing in ε_i affects D_i. That is $\varepsilon_i | D_i = \varepsilon_i$.

The method for causal inference: regression

$$Y_i = \alpha + \beta D_i + \varepsilon_i.$$

- Y_i is the *outcome* for individual i.
- D_i is the the value of the *treatment* for i.
  - i is the *unit of observation*—the level at which your outcome is defined.
- α is the intercept, the constant.
- β is the slope; the *effect of the treatment*.
- ε is an error term with mean zero $E(\varepsilon) = 0$; in it there could be other things that explain Y and random noise.
- Assume for now nothing in ε_i affects D_i. That is $\varepsilon_i | D_i = \varepsilon_i$.

Let's unpack the previous expression

$$D_i = \begin{cases} 1 & \text{if unit } i \text{ received the treatment,} \\ 0 & \text{otherwise.} \end{cases}$$

and

$$Y_i = \alpha + \beta D_i + \varepsilon_i.$$

Let us restrict our attention to those individuals whose $D_i = 1$.

$$\begin{aligned}
(Y_i|D_i = 1) =& \alpha + \beta + \varepsilon_i|D_i = 1 \\
E(Y_i|D_i = 1) =& E(\alpha + \beta + [\varepsilon_i|D_i = 1]) \\
=& \alpha + \beta + E(\varepsilon_i|D_i = 1) \\
=& \alpha + \beta + E(\varepsilon_i) \\
E(Y_i|D_i = 1) =& \alpha + \beta
\end{aligned}$$

Let's unpack the previous expression

$$D_i = \begin{cases} 1 & \text{if unit } i \text{ received the treatment,} \\ 0 & \text{otherwise.} \end{cases}$$

and

$$Y_i = \alpha + \beta D_i + \varepsilon_i.$$

Let us restrict our attention now to those individuals whose $D_i = 0$.

$$\begin{aligned} (Y_i|D_i = 0) =& \alpha + \varepsilon_i|D_i = 0 \\ E(Y_i|D_i = 0) =& E(\alpha + [\varepsilon_i|D_i = 0]) \\ =& \alpha + E(\varepsilon_i|D_i = 0) \\ =& \alpha + E(\varepsilon_i) \\ E(Y_i|D_i = 0) =& \alpha \end{aligned}$$

Let's unpack the previous expression

We obtain a simple system of equations with two unknowns!

$$E(Y_i|D_i = 1) = \alpha + \beta$$
$$E(Y_i|D_i = 0) = \alpha$$

Remember

- $\widehat{\alpha} = E(Y_i|D_i = 0)$.
- $\widehat{\beta} = E(Y_i|D_i = 1) - E(Y_i|D_i = 0)$.

Unpacking $\widehat{\beta}$

$$\widehat{\beta} = E(Y_i|D_i = 1) - E(Y_i|D_i = 0)$$
$$= \text{Avg. treatment effect} + \text{selection bias.}$$

- $\widehat{\beta}$ is the naive comparison.
- Only when there is no selection bias, $\widehat{\beta}$ is the ATE.
- What is selection bias in a regression equation?

Selection bias is omitted variable bias

$$
\begin{tikzcd}
X \arrow[d] \arrow[dr] & \\
D \arrow[r] & Y
\end{tikzcd}
$$

$$Y_i = \alpha + \beta D_i + \underbrace{\delta X_i + \epsilon_i}_{\varepsilon_i}.$$

Selection bias is omitted variable bias

$$
\begin{tikzcd}
X \arrow[d] \arrow[dr] & \\
D \arrow[r] & Y
\end{tikzcd}
$$

$$Y_i = \alpha + \beta D_i + \underbrace{\delta X_i + \epsilon_i}_{\varepsilon_i}.$$

- It is no longer the case that $E[\varepsilon_i | D_i = 1] \neq E[\varepsilon_i] = 0$ or $E[\varepsilon_i | D_i = 0] \neq E[\varepsilon_i] = 0$.

Selection bias is omitted variable bias

$$\begin{tikzcd} X \arrow[d] \arrow[dr] & \\ D \arrow[r] & Y \end{tikzcd}$$

$$Y_i = \alpha + \beta D_i + \underbrace{\delta X_i + \epsilon_i}_{\varepsilon_i}.$$

- It is no longer the case that $E[\varepsilon_i | D_i = 1] \neq E[\varepsilon_i] = 0$ or $E[\varepsilon_i | D_i = 0] \neq E[\varepsilon_i] = 0$.
- Selection bias only comes from confounders (X).

What happens if we don't control for confounders?

Let us restrict our attention to those individuals whose $D_i = 1$.

$$
\begin{aligned}
(Y_i | D_i = 1) =& \alpha + \beta + \varepsilon_i | D_i = 1 \\
E(Y_i | D_i = 1) =& E(\alpha + \beta + [\varepsilon_i | D_i = 1]) \\
=& \alpha + \beta + E(\varepsilon_i | D_i = 1)
\end{aligned}
$$

What happens if we don't control for confounders?

Let us restrict our attention to those individuals whose $D_i = 1$.

$$\begin{aligned} (Y_i|D_i = 1) &= \alpha + \beta + \varepsilon_i|D_i = 1 \\ E(Y_i|D_i = 1) &= E(\alpha + \beta + [\varepsilon_i|D_i = 1]) \\ &= \alpha + \beta + E(\varepsilon_i|D_i = 1) \end{aligned}$$

Let us restrict our attention now to those individuals whose $D_i = 0$.

$$\begin{aligned} (Y_i|D_i = 0) &= \alpha + \varepsilon_i|D_i = 0 \\ E(Y_i|D_i = 0) &= E(\alpha + [\varepsilon_i|D_i = 0]) \\ &= \alpha + E(\varepsilon_i|D_i = 0) \end{aligned}$$

What happens if we don't control for confounders?

Let us restrict our attention to those individuals whose $D_i = 1$.

$$\begin{aligned}(Y_i|D_i = 1) =& \alpha + \beta + \varepsilon_i|D_i = 1\\ E(Y_i|D_i = 1) =& E(\alpha + \beta + [\varepsilon_i|D_i = 1])\\ =& \alpha + \beta + E(\varepsilon_i|D_i = 1)\end{aligned}$$

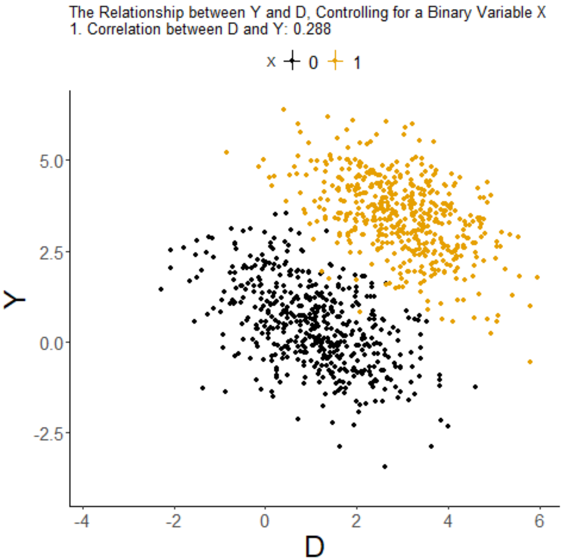
Let us restrict our attention now to those individuals whose $D_i = 0$.

$$\begin{aligned}(Y_i|D_i = 0) =& \alpha + \varepsilon_i|D_i = 0\\ E(Y_i|D_i = 0) =& E(\alpha + [\varepsilon_i|D_i = 0])\\ =& \alpha + E(\varepsilon_i|D_i = 0)\end{aligned}$$

Solving the system of equations:

$$\widehat{\beta} = \underbrace{E(Y_i|D_i = 1) - E(Y_i|D_i = 0)}_{ATE} + \underbrace{[E(\varepsilon_i|D_i = 1) - E(\varepsilon_i|D_i = 0)]}_{\text{selection bias}}.$$

The effect of a confounder: illustration



What happens if we control for ALL confounders?

Let us restrict our attention to those individuals whose $D_i = 1|X_i$.

$$\begin{aligned}
(Y_i|D_i = 1, X_i) =& \alpha + \beta(D_i = 1|X_i) + \epsilon_i|(D_i = 1|X_i) \\
E(Y_i|D_i = 1, X_i) =& \alpha + \beta E(D_i = 1|X_i) + E(\epsilon_i|(D_i = 1|X_i)) \\
=& \alpha + \beta E(D_i = 1|X_i) + E(\epsilon_i) \\
E(Y_i|D_i = 1, X_i) =& \alpha + \beta E(D_i = 1|X_i)
\end{aligned}$$

What happens if we control for ALL confounders?

Let us restrict our attention to those individuals whose $D_i = 1|X_i$.

$$
\begin{aligned}
(Y_i|D_i = 1, X_i) =& \alpha + \beta(D_i = 1|X_i) + \epsilon_i|(D_i = 1|X_i) \\
E(Y_i|D_i = 1, X_i) =& \alpha + \beta E(D_i = 1|X_i) + E(\epsilon_i|(D_i = 1|X_i)) \\
=& \alpha + \beta E(D_i = 1|X_i) + E(\epsilon_i) \\
E(Y_i|D_i = 1, X_i) =& \alpha + \beta E(D_i = 1|X_i)
\end{aligned}
$$

Let us restrict our attention now to those individuals whose $D_i = 0|X_i$.

$$
\begin{aligned}
(Y_i|D_i = 0, X_i) =& \alpha + \beta(D_i = 0|X_i) + \epsilon_i|(D_i = 0|X_i) \\
E(Y_i|D_i = 0, X_i) =& \alpha + \beta E(D_i = 0|X_i) + E(\epsilon_i|(D_i = 0|X_i)) \\
=& \alpha + \beta E(D_i = 0|X_i) + E(\epsilon_i) \\
E(Y_i|D_i = 0, X_i) =& \alpha + \beta E(D_i = 0|X_i)
\end{aligned}
$$

What happens if we control for ALL confounders?

Let us restrict our attention to those individuals whose $D_i = 1|X_i$.

$$\begin{aligned}
(Y_i|D_i = 1, X_i) =& \alpha + \beta(D_i = 1|X_i) + \epsilon_i|(D_i = 1|X_i) \\
E(Y_i|D_i = 1, X_i) =& \alpha + \beta E(D_i = 1|X_i) + E(\epsilon_i|(D_i = 1|X_i)) \\
=& \alpha + \beta E(D_i = 1|X_i) + E(\epsilon_i) \\
E(Y_i|D_i = 1, X_i) =& \alpha + \beta E(D_i = 1|X_i)
\end{aligned}$$

Let us restrict our attention now to those individuals whose $D_i = 0|X_i$.

$$\begin{aligned}
(Y_i|D_i = 0, X_i) =& \alpha + \beta(D_i = 0|X_i) + \epsilon_i|(D_i = 0|X_i) \\
E(Y_i|D_i = 0, X_i) =& \alpha + \beta E(D_i = 0|X_i) + E(\epsilon_i|(D_i = 0|X_i)) \\
=& \alpha + \beta E(D_i = 0|X_i) + E(\epsilon_i) \\
E(Y_i|D_i = 0, X_i) =& \alpha + \beta E(D_i = 0|X_i)
\end{aligned}$$

Solving the resulting system of equations we have

$$\widehat{\beta} = [E(Y_i|D_i = 1, X_i) - E(Y_i|D_i = 0, X_i)].$$

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!
- A drug maybe improve your health if you have an ailment, but it may be useless or even harmful otherwise! It may also be more effective if you have certain physical characteristics.

Interaction effects

- The treatment may affect some units but not other! Or it may affect them differently!
- A drug maybe improve your health if you have an ailment, but it may be useless or even harmful otherwise! It may also be more effective if you have certain physical characteristics.
- Benzodiazepines are used for seizures:
  - They are helpful for people with epilepsy.
  - They may cause ataxia in people with pre-existing abnormalities.
  - In some other they may not have and effect.

Interaction effects

- The treatment may affect some units but not other! Or it may affect them differently!
- A drug maybe improve your health if you have an ailment, but it may be useless or even harmful otherwise! It may also be more effective if you have certain physical characteristics.
- Benzodiazepines are used for seizures:
  - They are helpful for people with epilepsy.
  - They may cause ataxia in people with pre-existing abnormalities.
  - In some other they may not have and effect.
- Therefore the treatment can have heterogenous effects!

Consider

$$D_i = \begin{cases} 1 & \text{if unit } i \text{ is treated with a drug,} \\ 0 & \text{otherwise.} \end{cases}$$

and

$$X_i = \begin{cases} 1 & \text{if unit } i \text{ has abnormality,} \\ 0 & \text{otherwise.} \end{cases}$$

Consider

$$D_i = \begin{cases} 1 & \text{if unit } i \text{ is treated with a drug,} \\ 0 & \text{otherwise.} \end{cases}$$

and

$$X_i = \begin{cases} 1 & \text{if unit } i \text{ has abnormality,} \\ 0 & \text{otherwise.} \end{cases}$$

- The effect of the drug name for people with the abnormality:

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

- The effect of the drug name for people without the abnormality:

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

Interaction effects: in theory

Consider

$$D_i = \begin{cases} 1 & \text{if unit } i \text{ is treated with a drug,} \\ 0 & \text{otherwise.} \end{cases}$$

and

$$X_i = \begin{cases} 1 & \text{if unit } i \text{ has abnormality,} \\ 0 & \text{otherwise.} \end{cases}$$

- The effect of the drug name for people with the abnormality:

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

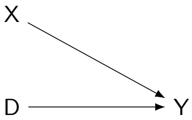
- The effect of the drug name for people without the abnormality:

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

We say that X moderates the effect of the treatment.

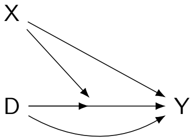
- When there is a variable X that moderates the effect of the treatment, we call this element a *moderator*.

Case 1: If the moderator X is a pre-treatment control but not a confounder:



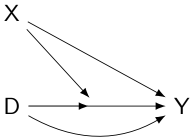
- When there is a variable X that moderates the effect of the treatment, we call this element a *moderator*.

Case 1: If the moderator X is a pre-treatment control but not a confounder:

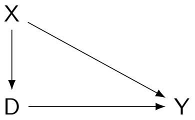


- When there is a variable X that moderates the effect of the treatment, we call this element a *moderator*.

Case 1: If the moderator X is a pre-treatment control but not a confounder:



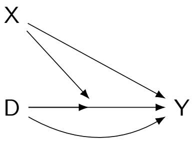
Case 2: If the moderator X is a pre-treatment confounder:



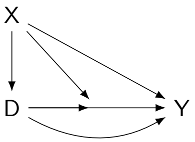
Interaction effects: controls and confounders

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

Case 1: If the moderator X is a pre-treatment control but not a confounder:



Case 2: If the moderator X is a pre-treatment confounder:



$$Y = \alpha + \beta_1 D + \beta_2 X + \beta_3 D \times X + \varepsilon.$$

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

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

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

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

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

Interaction effects: regression form

$$Y = \alpha + \beta_1 D + \beta_2 X + \beta_3 D \times X + \varepsilon.$$

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

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

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

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

- β_3 is the additional effect of the treatment for $X = 1$.
- Neither Y nor X need to be binary!

Interaction effects: regression form

$$Y = \alpha + \beta_1 D + \beta_2 X + \beta_3 D \times X + \varepsilon.$$

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

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

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

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

- β_3 is the additional effect of the treatment for $X = 1$.
- Neither Y nor X need to be binary!

Interaction effects: regression form

$$Y = \alpha + \beta_1 D + \beta_2 X + \beta_3 D \times X + \varepsilon.$$

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

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

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

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

- β_3 is the additional effect of the treatment for $X = 1$.
- Neither Y nor X need to be binary!

Interaction effects: regression form

$$Y = \alpha + \beta_1 D + \beta_2 X + \beta_3 D \times X + \varepsilon.$$

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

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

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

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

- β_3 is the additional effect of the treatment for $X = 1$.
- Neither Y nor X need to be binary!

$$Y = \alpha + \beta_1 D + \beta_2 X + \beta_3 D \times X + \varepsilon.$$

- We can use an interaction to test for robustness (Nunn and Puga).
    - That is one relevant use for interaction effects.

$$Y = \alpha + \beta_1 D + \beta_2 X + \beta_3 D \times X + \varepsilon.$$

- We can use an interaction to test for robustness (Nunn and Puga).
  - That is one relevant use for interaction effects.
- For our purposes also:
  - X can be a political variable that moderates the effect of a treatment.
  - X does not need to be all else equal to interpret β_1 and β_3 as long as D is assigned all else equal.
  - If assignment of D is all else equal this opens many potential questions.
    - Does the this political variable moderate the effect of my treatment?
    - These questions require a lot of theoretical knowledge and of the case being studied.

Interaction effects: why?

$$Y = \alpha + \beta_1 D + \beta_2 X + \beta_3 D \times X + \varepsilon.$$

- We can use an interaction to test for robustness (Nunn and Puga).
  - That is one relevant use for interaction effects.
- For our purposes also:
  - X can be a political variable that moderates the effect of a treatment.
  - X does not need to be all else equal to interpret β_1 and β_3 as long as D is assigned all else equal.
  - If assignment of D is all else equal this opens many potential questions.
    - Does the this political variable moderate the effect of my treatment?
    - These questions require a lot of theoretical knowledge and of the case being studied.

Interaction effects: why?

$$Y = \alpha + \beta_1 D + \beta_2 X + \beta_3 D \times X + \varepsilon.$$

- We can use an interaction to test for robustness (Nunn and Puga).
  - That is one relevant use for interaction effects.

- For our purposes also:
  - X can be a political variable that moderates the effect of a treatment.
  - X does not need to be all else equal to interpret β_1 and β_3 as long as D is assigned all else equal.
  - If assignment of D is all else equal this opens many potential questions.
    - Does the this political variable moderate the effect of my treatment?
    - These questions require a lot of theoretical knowledge and of the case being studied.

Interaction effects: interpretation

$$Y = \alpha + \beta_1 D + \beta_2 X + \beta_3 D \times X + \varepsilon.$$

- Neither Y nor X need 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 $\widehat{\beta}_1 + \widehat{\beta}_3 X$.

$$Y = \alpha + \beta_1 D + \beta_2 X + \beta_3 D \times X + \varepsilon.$$

- Neither Y nor X need 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

$$Y = \alpha + \beta_1 D + \beta_2 X + \beta_3 D \times X + \varepsilon.$$

- Neither Y nor X need 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

$$Y = \alpha + \beta_1 D + \beta_2 X + \beta_3 D \times X + \varepsilon.$$

- Neither Y nor X need 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 $\widehat{\beta}_1 + \widehat{\beta}_3 X$.

Things to keep in mind

- An interaction variable X must be pretreatment.
- It does not matter if X suffers from omitted variables bias.
- This is because if D is randomly assigned 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.

Things to keep in mind

- An interaction variable X must be pretreatment.
- It does not matter if X suffers from omitted variables bias.
- This is because if D is randomly assigned 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.

Things to keep in mind

- An interaction variable X must be pretreatment.
- It does not matter if X suffers from omitted variables bias.
- This is because if D is randomly assigned 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.

Things to keep in mind

- An interaction variable X must be pretreatment.
- It does not matter if X suffers from omitted variables bias.
- This is because if D is randomly assigned 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.

Things to keep in mind

- An interaction variable X must be pretreatment.
- It does not matter if X suffers from omitted variables bias.
- This is because if D is randomly assigned 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.

Non-linear effects of the treatment

$$
\begin{tikzcd}
X \arrow[d] \arrow[dr] & \\
D \arrow[r] & Y
\end{tikzcd}
$$

$$Y_i = \alpha + f(D_i) + \delta X_i + \varepsilon_i.$$

- The $f(D)$ may be non-linear in D.

$$Y_i = \alpha + \beta_1 D_i + \beta_2 D_i^2 + \delta X_i + \varepsilon_i.$$

Non-linear effects of the treatment

$$\begin{tikzcd} X \arrow[d] \arrow[dr] & \\ D \arrow[r] & Y \end{tikzcd}$$

$$Y_i = \alpha + f(D_i) + \delta X_i + \varepsilon_i.$$

- The $f(D)$ may be non-linear in D.

$$Y_i = \alpha + \beta_1 D_i + \beta_2 D_i^2 + \delta X_i + \varepsilon_i.$$

- Maybe theoretical insights suggest the effect of the treatment is non-monotonic!

Canon transfers in Peru and children test scores

- Canon transfers are resources redistributed from mining taxes for public spending.
- More public spending in education could improve children's learning.

Canon transfers in Peru and children test scores

- Canon transfers are resources redistributed from mining taxes for public spending.
- More public spending in education could improve children's learning.
- Selection bias concerns:
    - Economic activity $\rightarrow$ mining $\rightarrow$ canon.
    - Districts with more political power secure more Canon transfers.

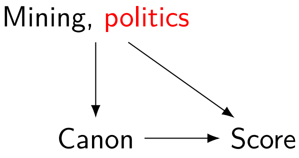
Canon transfers in Peru and children test scores

- Canon transfers are resources redistributed from mining taxes for public spending.
- More public spending in education could improve children's learning.
- Selection bias concerns:
  - Economic activity $\rightarrow$ mining $\rightarrow$ canon.
  - Districts with more political power secure more Canon transfers.
- But...
  - Control for mining production.
  - Canon transfer rule is fixed for the period of analysis.
  - Control for the fixed impact of politics at the district level.

Canon transfers in Peru and children test scores

$$
\begin{tikzcd}
\text{Mining, politics} \arrow[d] \arrow[dr] & \\
\text{Canon} \arrow[r] & \text{Score}
\end{tikzcd}
$$

- Unit of observation: student.



- Unit of observation: student.
- Politics is determined before the period of analysis (they control for it).

Canon transfers in Peru and children test scores

$$\begin{array}{ccc} \text{Mining, } \color{red}{\text{politics}} & & \\ \downarrow & \searrow & \\ \text{Canon} & \longrightarrow & \text{Score} \end{array}$$

- Unit of observation: student.
- Politics is determined before the period of analysis (they control for it).
- Can you think of an unobservable confounder?

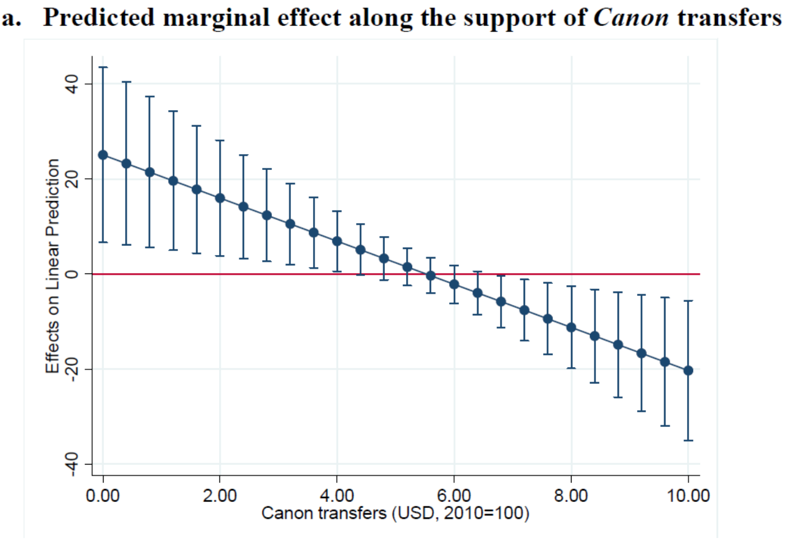
Canon seems to be a blessing unless it is too large

	Dependent variable: Test scores				
	(1)	(2)	(3)	(4)	(5)
	Panel A. Math scores				
Canon	7.2844	22.8601**	23.0187**	25.1328***	22.5513**
	(4.6782)	(9.5349)	(9.6193)	(9.4648)	(9.4570)
Canon squared		−2.0713**	−2.0797**	−2.2729***	−2.0215**
		(0.8379)	(0.8408)	(0.8257)	(0.7853)
Mining production			−0.0222	−0.0111	−0.0364
			(0.0453)	(0.0422)	(0.0452)
Canon (lag)					−2.0750
					(6.4640)
Canon squared (lag)					0.3651
					(0.6222)
Mining production (lag)					−0.0325
					(0.0204)
R-squared	0.1506	0.1507	0.1507	0.1604	0.0063
Number of students	2,072,351	2,072,351	2,072,351	2,012,798	2,072,351
Student controls	N	N	N	Y	N
School controls	N	N	N	Y	N

Note: Robust standard errors clustered at the district level in parentheses. * Significant at ten percent; ** significant at five percent; *** significant at one percent. Mining and *Canon* correspond to the value of mining production per-capita and *Canon* per-capita, in thousands of USD at constant prices of 2010. All regressions include fixed effects at the district level and by year. School characteristics include: school day (full-day, half-day morning or half-day afternoon); school administration type (public/private), and school type (one-teacher school/full grade). Student characteristics include: gender and mother language (Spanish/other) of the children.
Source: Authors' calculations based on data from Peru's Ministries of Education, of Finance, of Energy and Mines, and INEI.

- The regression is: $Y = \alpha + \beta_1 D + \beta_2 D^2 + \delta X + \varepsilon$.

Canon seems to be a blessing unless it is too large



- The total effect is: $\beta_1 D + \beta_2 D^2$.

Results are related to public expenditure on education

	Dependent variable: Test scores				
	(1)	(2)	(3)	(4)	(5)
	Panel A. Math scores				
Canon	7.2844	22.8601**	23.0187**	25.1328***	22.5513**
	(4.6782)	(9.5349)	(9.6193)	(9.4648)	(9.4570)
Canon squared		−2.0713**	−2.0797**	−2.2729***	−2.0215**
		(0.8379)	(0.8408)	(0.8257)	(0.7853)
Mining production			−0.0222	−0.0111	−0.0364
			(0.0453)	(0.0422)	(0.0452)
Canon (lag)					−2.0750
					(6.4640)
Canon squared (lag)					0.3651
					(0.6222)
Mining production (lag)					−0.0325
					(0.0204)
R-squared	0.1506	0.1507	0.1507	0.1604	0.0063
Number of students	2,072,351	2,072,351	2,072,351	2,012,798	2,072,351
Student controls	N	N	N	Y	N
School controls	N	N	N	Y	N

Note: Robust standard errors clustered at the district level in parentheses. * Significant at ten percent; ** significant at five percent; *** significant at one percent. Mining and *Canon* correspond to the value of mining production per-capita and *Canon* per-capita, in thousands of USD at constant prices of 2010. All regressions include fixed effects at the district level and by year. School characteristics include: school day (full-day, half-day morning or half-day afternoon); school administration type (public/private), and school type (one-teacher school/full grade). Student characteristics include: gender and mother language (Spanish/other) of the children.
Source: Authors' calculations based on data from Peru's Ministries of Education, of Finance, of Energy and Mines, and INEI.

The crux of the empirical researcher

- When we use observational data we may not observe all confounders.
  - We may be able to control for observable/measurable ones.
- We can find clever ways to provide a very good sense of the direction of the causal effect.
- To interpret the magnitude is not always possible although many do, with a grain of salt.

The crux of the empirical researcher

- When we use observational data we may not observe all confounders.
  - We may be able to control for observable/measurable ones.
- We can find clever ways to provide a very good sense of the direction of the causal effect.
- To interpret the magnitude is not always possible although many do, with a grain of salt.

- When we use observational data we may not observe all confounders.
  - We may be able to control for observable/measurable ones.
- We can find clever ways to provide a very good sense of the direction of the causal effect.
- To interpret the magnitude is not always possible although many do, with a grain of salt.

- Robustness means two things in general:
  - The direction of the effect doesn't change.
  - The magnitude of the effect doesn't change.
  - Results are more robust when it is the latter.
  - Effects must remain statistically significant.

Robustness and clever ways to address the issue

- Robustness means two things in general:
  - The direction of the effect doesn't change.
  - The magnitude of the effect doesn't change.
  - Results are more robust when it is the latter.
  - Effects must remain statistically significant.
- Robustness tests can be of many kinds:
  - Introduce one confounder at a time.
  - Drop observations with replacement.
  - Use another measure of the treatment and (or) control.

Robustness and clever ways to address the issue

- Robustness means two things in general:
  - The direction of the effect doesn't change.
  - The magnitude of the effect doesn't change.
  - Results are more robust when it is the latter.
  - Effects must remain statistically significant.
- Robustness tests can be of many kinds:
  - Introduce one confounder at a time.
  - Drop observations with replacement.
  - Use another measure of the treatment and (or) control.
- Robustness tests can be also clever exercises!
  - Theoretical prior may tell you where you should observe the effect and where you shouldn't.
  - This is called a placebo test!

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

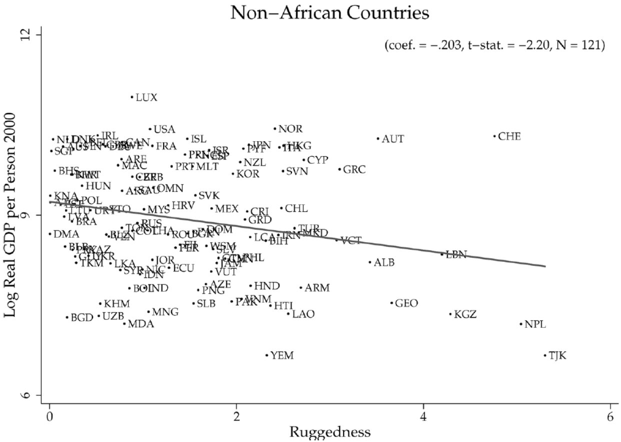
- Rugged terrain can be a curse:
  - Hard to develop agriculture.
  - Higher costs of transportation.
  - Costlier to tax and to provide public goods.
  - Breeding ground for militias.
- But it can also be a "blessing":
  - It can reduce expropriation by despotic ruler.
  - In the past, it protected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

It is difficult to estimate a causal relationship but...

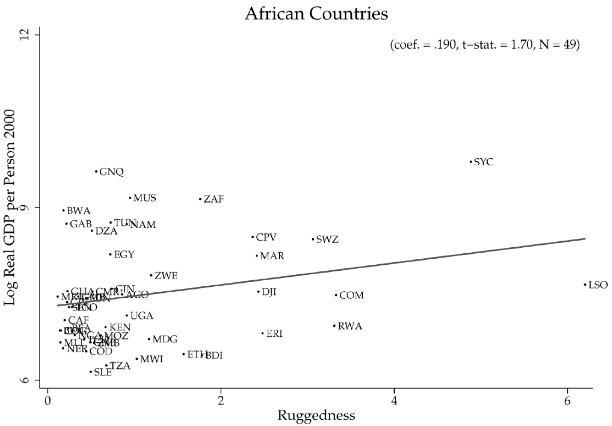
The authors carry a number of robustness tests and placebos tests

- Placebo: The blessing only exists for Africa.
- Robustness tests:
    1. with Respect to Omitted Geographical Variables.
    2. with Respect to Alternative Income and Ruggedness Measures
    3. with Respect to Influential Observations
- Then they test what characteristics of African countries may drive their results.

Rugged terrain can be a curse



But it seems to be a blessing in Africa



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.

Robustness: omitted geographical variables

TABLE 2.—ROBUSTNESS WITH RESPECT TO INFLUENTIAL OBSERVATIONS

	Dependent Variable: Log Real GDP per Person, 2000				
	Omit 10 Most Rugged	Omit 10 Smallest	Omit if $\lvert DFBETA \rvert > 2/\sqrt{N}$	Using ln(Ruggedness)	Box-Cox Transformation of Ruggedness
	(1)	(2)	(3)	(4)	(5)
Ruggedness	−0.202	−0.221	−0.261	−0.171	−0.249
	(0.083)**	(0.083)***	(0.068)***	(0.051)***	(0.075)***
Ruggedness × I^{Africa}	0.286	0.188	0.223	0.234	0.333
	(0.133)**	(0.099)*	(0.116)*	(0.119)**	(0.142)**
I^{Africa}	−1.448	−1.465	−1.510	−1.083	−1.139
	(0.454)***	(0.405)***	(0.406)***	(0.394)***	(0.391)***
All controls	Yes	Yes	Yes	Yes	Yes
Observations	160	160	164	170	170
R^2	0.520	0.545	0.564	0.527	0.533

Coefficients are reported with robust standard errors in brackets. ***, **, and * indicate significance at the 1%, 5%, and 10% levels. All regressions include a constant, and our full set of control variables: diamonds, diamonds × I^{Africa}, % fertile soil, % fertile soil × I^{Africa}, % tropical climate, % tropical climate × I^{Africa}, distance to coast, and distance to coast × I^{Africa}. Coefficients and standard errors for the control variable are reported in the online appendix.

Robustness: different measures

Table 8: Differential effect of ruggedness for Africa, alternative income and ruggedness measures

		Dependent variable:			
		Log real GDP per person 2000 (World Bank)	Log real GDP per person 2000 (Maddison)	Log real GDP per person 1950 (Maddison)	Log real GDP per person 1950–2000 average (Maddison)
Ruggedness measure:	Ruggedness	0.321	0.250	0.284	0.284
		(0.127)**	(0.113)**	(0.129)**	(0.123)**
	Average slope	0.098	0.076	0.083	0.084
		(0.044)**	(0.040)*	(0.047)*	(0.045)*
	Local std. dev. of elevation	1.105	0.835	0.919	0.922
		(0.459)**	(0.414)**	(0.460)**	(0.443)**
	% highly rugged land	0.017	0.014	0.017	0.017
		(0.006)***	(0.006)**	(0.006)***	(0.006)***
	Pop.-weighted ruggedness	0.726	0.664	0.393	0.531
		(0.220)***	(0.206)***	(0.192)**	(0.190)***
	Observations	170	159	137	137

Notes: Each entry of the table reports the coefficient and robust standard errors for Ruggedness $\cdot I^{\text{Africa}}$ from the specification of column (6) in table 1 in the main text, estimated using an alternative income measure as the dependent variable and an alternative measure of ruggedness. The alternative income measure is reported above the corresponding column, and the alternative ruggedness measure is reported to the left of the corresponding row. ***, **, and * indicate significance at the 1, 5, and 10 percent levels.

Checks: differential effects

Table 3.—Considering Differential Effects of Ruggedness by Characteristics Prevalent in Africa

	Dependent Variable: Log Real GDP per Person, 2000				
	(1)	(2)	(3)	(4)	(5)
Ruggedness	−0.259	−0.322	−0.374	−0.386	−0.543
	(0.101)**	(0.160)**	(0.161)**	(0.176)**	(0.179)***
Ruggedness × I^{Africa}	0.357	0.400	0.360	0.399	0.435
	(0.130)***	(0.155)***	(0.140)**	(0.203)**	(0.135)***
I^{Africa}	−1.814	−1.977	−1.818	−1.740	−1.994
	(0.213)***	(0.223)***	(0.218)***	(0.337)***	(0.216)***
Ruggedness × % tropical climate	Yes	No	Yes	Yes	Yes
% Tropical climate	Yes	No	Yes	Yes	Yes
Ruggedness × % fertile soil	No	Yes	Yes	Yes	Yes
% Fertile soil	No	Yes	Yes	Yes	Yes
Ruggedness × colonizer FEs	No	No	No	Yes	No
Colonizer FEs	No	No	No	Yes	No
Ruggedness × legal origin FEs	No	No	No	No	Yes
Legal origin FEs	No	No	No	No	Yes
Observations	170	170	170	170	170
R^2	0.404	0.363	0.408	0.430	0.559

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

Checks: differential effects

Table 4.—Differential Effects of Ruggedness across Regions within Africa

	Dependent Variable: Log Real GDP per Person, 2000				
	(1)	(2)	(3)	(4)	(5)
Ruggedness	−0.203	−0.203	−0.203	−0.203	−0.203
	(0.093)**	(0.093)**	(0.093)**	(0.093)**	(0.093)**
Ruggedness × I^{Africa}	0.312	0.408	0.409	0.406	0.448
	(0.159)**	(0.161)**	(0.147)***	(0.147)***	(0.179)**
I^{Africa}	−1.735	−1.844	−2.008	−2.046	−2.054
	(0.291)***	(0.229)***	(0.230)***	(0.222)***	(0.232)***
Ruggedness × $I^{\text{West Africa}}$	0.532				
	(0.154)***				
$I^{\text{West Africa}}$	−0.635				
	(0.283)**				
Ruggedness × $I^{\text{East Africa}}$		0.162			
		(0.274)			
$I^{\text{East Africa}}$		−0.760			
		(0.532)			
Ruggedness × $I^{\text{Central Africa}}$			0.575		
			(1.197)		
$I^{\text{Central Africa}}$			0.020		
			(0.597)		
Ruggedness × $I^{\text{North Africa}}$				−0.404	
				(0.131)***	
$I^{\text{North Africa}}$				1.465	
				(0.241)***	
Ruggedness × $I^{\text{South Africa}}$					−0.200
					(0.195)
$I^{\text{South Africa}}$					0.592
					(0.519)
Constant	9.223	9.223	9.223	9.223	9.223
	(0.144)***	(0.144)***	(0.144)***	(0.144)***	(0.144)***
Observations	170	170	170	170	170
R^2	0.367	0.368	0.359	0.375	0.363

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

Temporary page!

LaTeX was unable to guess the total number of pages correctly. A
there was some unprocessed data that should have been added t
the final page this extra page has been added to receive it.
If you rerun the document (without altering it) this surplus page
will go away, because LaTeX now knows how many pages to expe
for this document.