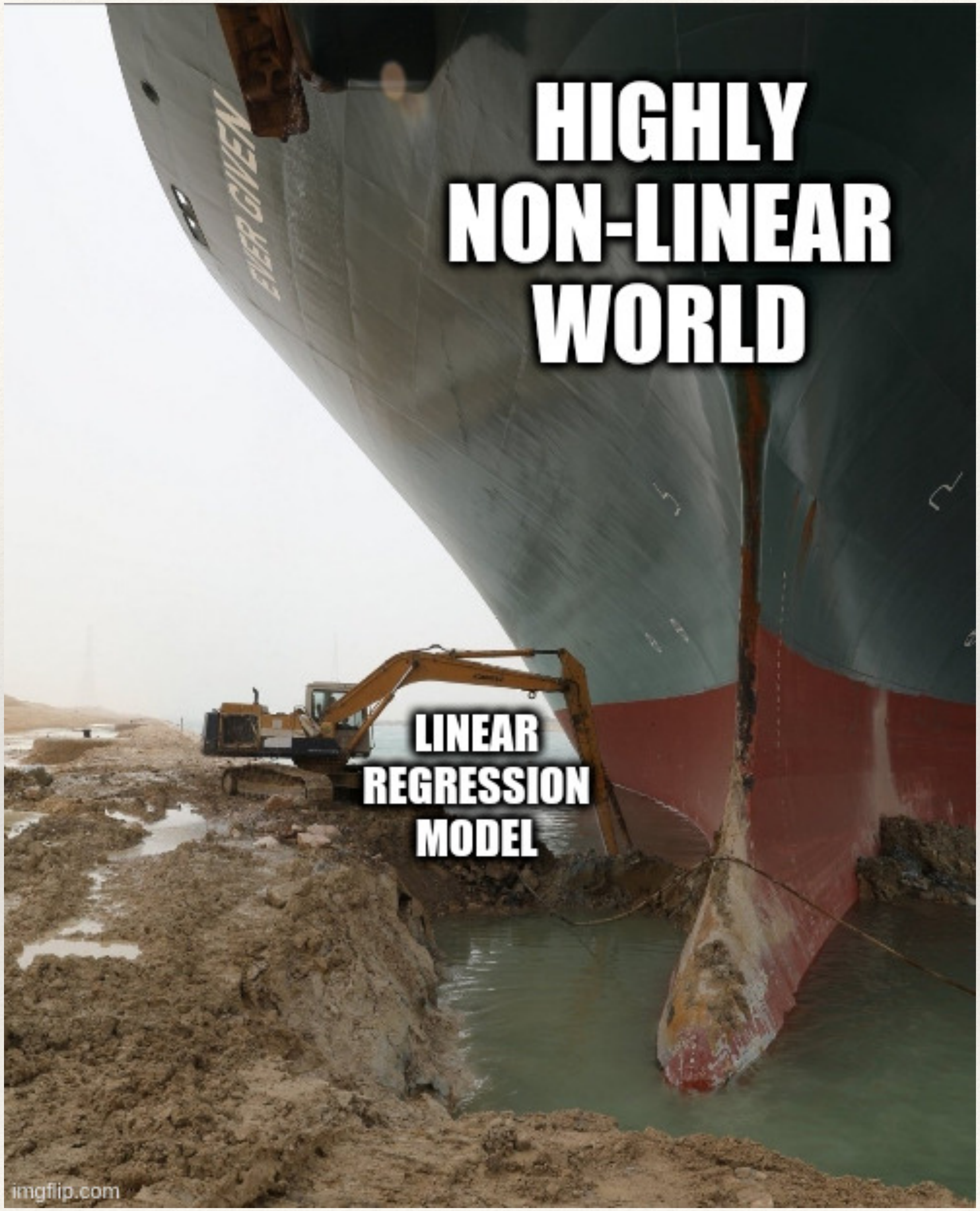
Interactions

Introduction to Statistics



The Plan for Today

The Plan for Today

- **Recap of Multiple Regression**

The Plan for Today

- **Recap of Multiple Regression**
  - Tying up some loose ends from last class, and a bit more on ’controlling’.

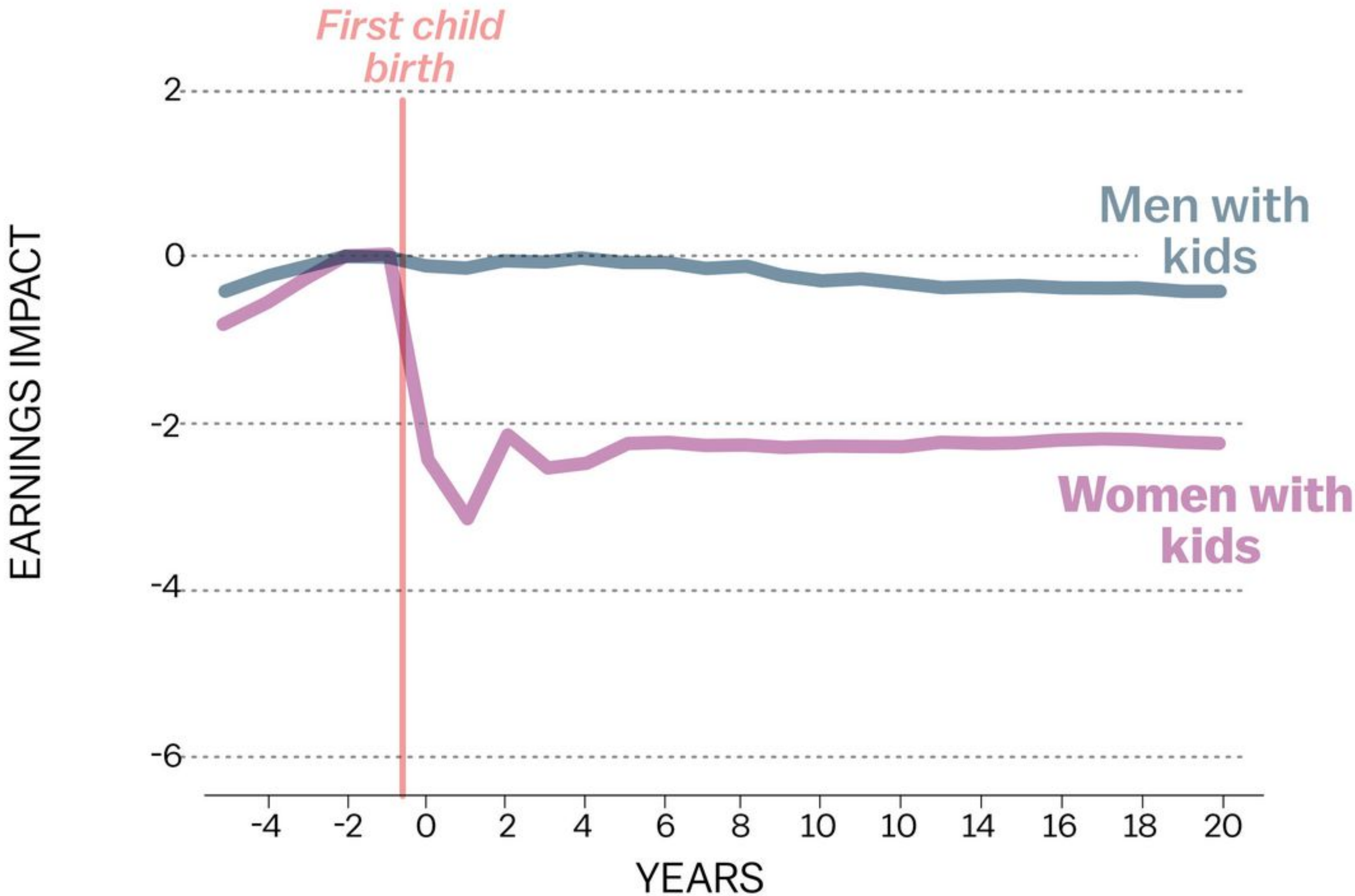
The Plan for Today

- **Recap of Multiple Regression**
  - Tying up some loose ends from last class, and a bit more on ’controlling’.
- **Interactions**

The Plan for Today

- **Recap of Multiple Regression**
  - Tying up some loose ends from last class, and a bit more on ’controlling’.
- **Interactions**
  - Intuition: what’s the effect of parenthood on earnings? Well, *depends*.

Women's earnings drop significantly after having a child. Men's don't.



Source: "Children and gender inequality: Evidence from Denmark," National Bureau of Economic Research

The Plan for Today

- **Recap of Multiple Regression**
  - Tying up some loose ends from last class, and a bit more on ’controlling’.
- **Interactions**
  - Intuition: what’s the effect of parenthood on earnings? Well, *depends*.

The Plan for Today

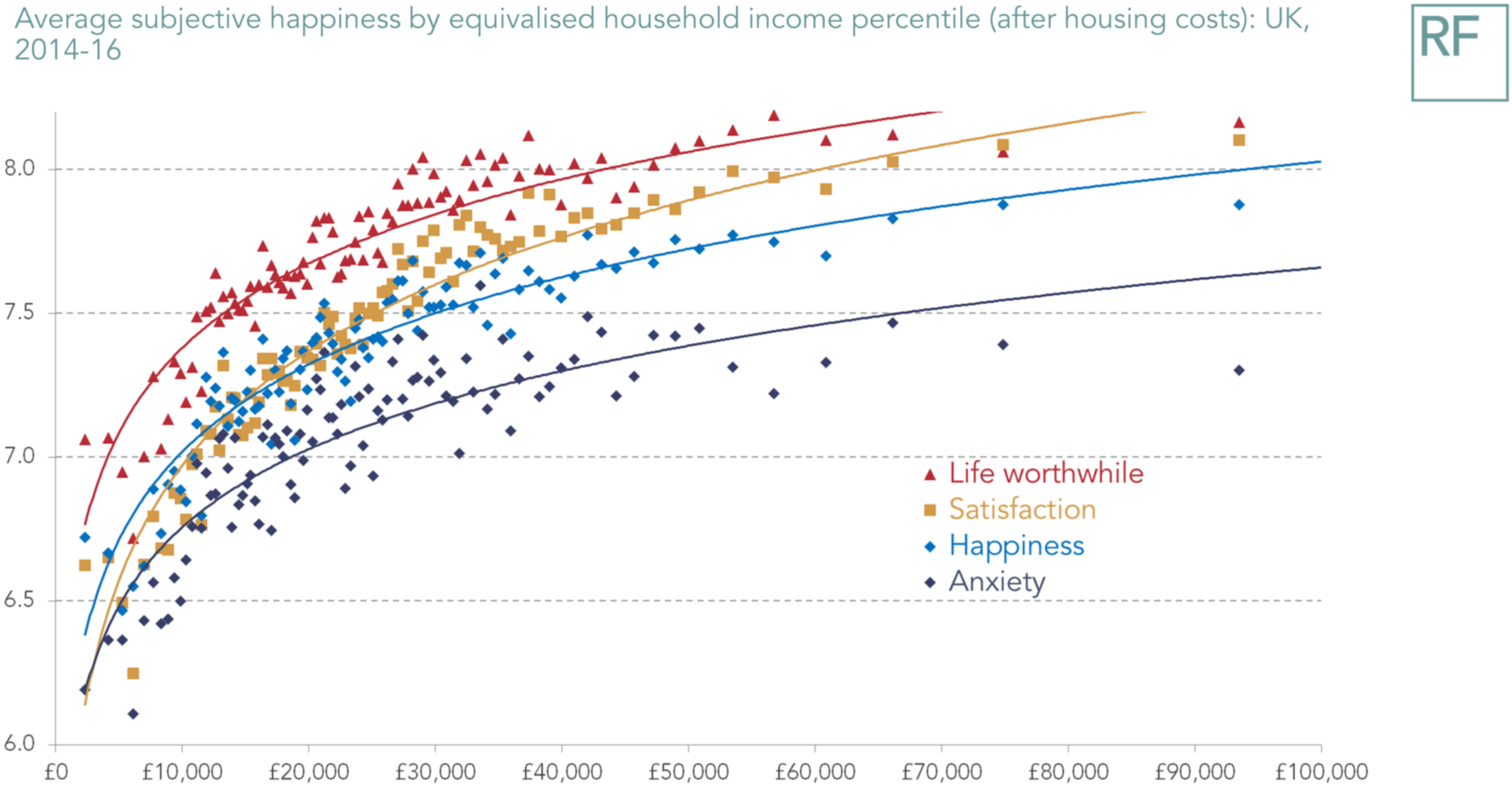
- **Recap of Multiple Regression**
  - Tying up some loose ends from last class, and a bit more on 'controlling'.
- **Interactions**
  - Intuition: what's the effect of parenthood on earnings? Well, *depends*.
- **Non-linearities**

The Plan for Today

* **Recap of Multiple Regression**
  * Tying up some loose ends from last class, and a bit more on ’controlling’.
* **Interactions**
  * Intuition: what’s the effect of parenthood on earnings? Well, *depends*.
* **Non-linearities**
  * Intuition: does money buy you happiness? Well, *depends*.

The Plan for Today

Average subjective happiness by equivalised household income percentile (after housing costs): UK, 2014-16



Notes: Each dot represents the average level of well-being for a percentile of household income (measured after housing costs), ranging from percentile 1 on the far left of the chart to percentile 100 on the far right. The lines are logarithmic lines of best fit.
Source: RF analysis of DWP, *Family Resources Survey*; pooled data for 2014-15 to 2016-17

The Plan for Today

- **Recap of Multiple Regression**
  - Tying up some loose ends from last class, and a bit more on ’controlling’.
- **Interactions**
  - Intuition: what’s the effect of parenthood on earnings? Well, *depends*.
- **Non-linearities**
  - Intuition: does money buy you happiness? Well, *depends*.

The Plan for Today

* **Recap of Multiple Regression**
  * Tying up some loose ends from last class, and a bit more on ’controlling’.
* **Interactions**
  * Intuition: what’s the effect of parenthood on earnings? Well, *depends*.
* **Non-linearities**
  * Intuition: does money buy you happiness? Well, *depends*.
* In the meantime, **visualisation, visualisation, visualisation**

The Plan for Today

* **Recap of Multiple Regression**
  * Tying up some loose ends from last class, and a bit more on ’controlling’.
* **Interactions**
  * Intuition: what’s the effect of parenthood on earnings? Well, *depends*.
* **Non-linearities**
  * Intuition: does money buy you happiness? Well, *depends*.
* In the meantime, **visualisation, visualisation, visualisation**
  * With complex models, plots are much clearer than regression tables.

Regression: Recap

Multiple Linear Regression with OLS

Multiple Linear Regression with OLS

* Our model of reality:

Multiple Linear Regression with OLS

- Our model of reality:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \ldots \beta_p X_p + \epsilon$$

Multiple Linear Regression with OLS

- Our model of reality:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \ldots \beta_p X_p + \epsilon$$

Multiple Linear Regression with OLS

- Our model of reality:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \ldots \beta_p X_p + \epsilon$$

- Where each β_j represents the average increase in Y associated with a one-unit increase in X_j **holding the other variables constant**.

Multiple Linear Regression with OLS

- Our model of reality:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \ldots \beta_p X_p + \epsilon$$

- Where each β_j represents the average increase in Y associated with a one-unit increase in X_j **holding the other variables constant**.
- How do we pick the coefficients?

Multiple Linear Regression with OLS

- Our model of reality:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \ldots \beta_p X_p + \epsilon$$

- Where each β_j represents the average increase in Y associated with a one-unit increase in X_j **holding the other variables constant**.
- How do we pick the coefficients?
- The most common method (not the only one!) is **Ordinary Least Squares (OLS)** — choose the combination of coefficients that **minimise the sum of squared residuals**.

Multiple Linear Regression with OLS

Multiple Linear Regression with OLS

- What are residuals? They are the difference between…

Multiple Linear Regression with OLS

- What are residuals? They are the difference between…
  - The **observed values** of Y, that is $Y_1, Y_2, Y_3, Y_4 \ldots Y_n$

Multiple Linear Regression with OLS

- What are residuals? They are the difference between…
  - The **observed values** of Y, that is $Y_1, Y_2, Y_3, Y_4 \ldots Y_n$
  - And the **fitted values** $\hat{Y}$ (that is $\hat{Y}_1, \hat{Y}_2, \hat{Y}_3, \hat{Y}_4 \ldots \hat{Y}_n$) that we get at with out prediction line $\hat{Y} = \hat{\alpha} + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 \ldots \hat{\beta}_p X_p$.

Multiple Linear Regression with OLS

* What are residuals? They are the difference between…
  * The **observed values** of Y, that is $Y_1, Y_2, Y_3, Y_4 \ldots Y_n$
  * And the **fitted values** $\hat{Y}$ (that is $\hat{Y}_1, \hat{Y}_2, \hat{Y}_3, \hat{Y}_4 \ldots \hat{Y}_n$) that we get at with out prediction line $\hat{Y} = \hat{\alpha} + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 \ldots \hat{\beta}_p X_p$.
* Each observation i will have its own residual $\hat{\epsilon}_i = Y_i - \hat{Y}_i$

Multiple Linear Regression with OLS

- What are residuals? They are the difference between…
  - The **observed values** of Y, that is $Y_1, Y_2, Y_3, Y_4 \ \dots \ Y_n$
  - And the **fitted values** $\hat{Y}$ (that is $\hat{Y}_1, \hat{Y}_2, \hat{Y}_3, \hat{Y}_4 \ \dots \ \hat{Y}_n$) that we get at with out prediction line $\hat{Y} = \hat{\alpha} + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 \dots \hat{\beta}_p X_p$.
- Each observation i will have its own residual $\hat{\epsilon}_i = Y_i - \hat{Y}_i$

- So OLS will choose $Y = \hat{\alpha} + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 \dots \hat{\beta}_p X_p + \hat{\epsilon}$ so that $\sum_{i=1}^{n} \hat{\epsilon}_i^{\,2} = \sum_{i=1}^{n} (Y - \hat{Y}_i)^2$ is minimised.

Multiple Linear Regression with OLS

	Dependent variable:
	Life Satisfaction (0–10)
Age	0.013*** (0.004)
Income Decile	0.163*** (0.019)
Female	0.288*** (0.100)
Religiosity (0–10)	0.022 (0.017)
Years of Education	−0.003 (0.014)
Divorced	−0.354 (0.299)
Single	−0.118 (0.131)
Widowed	−0.412** (0.189)
Constant	5.713*** (0.321)
Observations	1,601
R^2	0.078
Adjusted R^2	0.073
Residual Std. Error	1.947 (df = 1592)
F Statistic	16.778*** (df = 8; 1592)
Note:	*p<0.1; **p<0.05; ***p<0.01

Multiple Linear Regression with OLS

Multiple Linear Regression with OLS

* With OLS, we also estimate:

Multiple Linear Regression with OLS

* With OLS, we also estimate:
* The **standard errors of the coefficient**, which represents the (estimated) standard deviation of the sampling distribution of the coefficient, obtained through a (hypothetically) repeated sampling process, yielding different estimated coefficients every time.

Multiple Linear Regression with OLS

* With OLS, we also estimate:
* The **standard errors of the coefficient**, which represents the (estimated) standard deviation of the sampling distribution of the coefficient, obtained through a (hypothetically) repeated sampling process, yielding different estimated coefficients every time.
* The ***p*-value of the coefficient**, which represents the probability of obtaining a coefficient at least as extreme as the one estimated in our sample, under the null hypothesis that in the population there's no relationship between X and Y, conditional on covariates.

Multiple Linear Regression with OLS

* With OLS, we also estimate:
* The **standard errors of the coefficient**, which represents the (estimated) standard deviation of the sampling distribution of the coefficient, obtained through a (hypothetically) repeated sampling process, yielding different estimated coefficients every time.
* The ***p*-value of the coefficient**, which represents the probability of obtaining a coefficient at least as extreme as the one estimated in our sample, under the null hypothesis that in the population there's no relationship between X and Y, conditional on covariates.
* The **adjusted R-squared**, which quantifies the extent to which the model as a whole explains variation in the outcome variable.

Multiple Linear Regression with OLS

```
Call:
lm(formula = life_satisf ~ age + income_decile + female + religiosity +
    years_education + marital_status, data = ess)

Residuals:
    Min      1Q  Median      3Q     Max
-8.1662 -0.8452  0.2721  1.2738  3.8794

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               5.712586   0.320715  17.812  < 2e-16 ***
age                       0.013353   0.003510   3.804 0.000148 ***
income_decile             0.163156   0.019339   8.437  < 2e-16 ***
female                    0.287897   0.099643   2.889 0.003914 **
religiosity               0.022203   0.016572   1.340 0.180513
years_education          -0.003186   0.014112  -0.226 0.821429
marital_status divorced -0.353683   0.299287  -1.182 0.237480
marital_status single   -0.118078   0.130715  -0.903 0.366491
marital_status widowed  -0.412239   0.188733  -2.184 0.029090 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.947 on 1592 degrees of freedom
  (603 observations deleted due to missingness)
Multiple R-squared:  0.07776,   Adjusted R-squared:  0.07312
F-statistic: 16.78 on 8 and 1592 DF,  p-value: < 2.2e-16
```

OLS Assumptions

OLS Assumptions

1. Linearity

OLS Assumptions

1. Linearity

   * The model **in the population** (the 'true' model) can be written as a linear combination of variables and coefficients: $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 \ldots \beta_p X_p + \epsilon$.

OLS Assumptions

1. Linearity
   - The model **in the population** (the 'true' model) can be written as a linear combination of variables and coefficients: $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 \ldots \beta_p X_p + \epsilon$.
2. Random Sampling

OLS Assumptions

1. Linearity
   * The model **in the population** (the 'true' model) can be written as a linear combination of variables and coefficients: $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 \ldots \beta_p X_p + \epsilon$.
2. Random Sampling
   * We have a **random sample** of n observations, following the population model.

OLS Assumptions

1. Linearity

   * The model **in the population** (the 'true' model) can be written as a linear combination of variables and coefficients: $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 \ldots \beta_p X_p + \epsilon$.

2. Random Sampling

   * We have a **random sample** of n observations, following the population model.

3. No Perfect Collinearity

OLS Assumptions

1. Linearity

   * The model **in the population** (the 'true' model) can be written as a linear combination of variables and coefficients: $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 \ldots \beta_p X_p + \epsilon$.

2. Random Sampling

   * We have a **random sample** of n observations, following the population model.

3. No Perfect Collinearity

   * In the sample, none of the independent variables are **constant**, and there are no **exact linear relationships** between independent variables.

OLS Assumptions

1. Linearity
   - The model **in the population** (the 'true' model) can be written as a linear combination of variables and coefficients: $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 \ldots \beta_p X_p + \epsilon$.
2. Random Sampling
   - We have a **random sample** of n observations, following the population model.
3. No Perfect Collinearity
   - In the sample, none of the independent variables are **constant**, and there are no **exact linear relationships** between independent variables.
4. Zero Conditional Mean (Exogeneity)

OLS Assumptions

1. Linearity
   - The model **in the population** (the 'true' model) can be written as a linear combination of variables and coefficients: $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 \ldots \beta_p X_p + \epsilon$.
2. Random Sampling
   - We have a **random sample** of n observations, following the population model.
3. No Perfect Collinearity
   - In the sample, none of the independent variables are **constant**, and there are no **exact linear relationships** between independent variables.
4. Zero Conditional Mean (Exogeneity)
   - The error term has a **mean of zero** and is **unrelated to any of the *X*s**. *Many potential violations in practice:* omitted variable bias, non-linear relationships, reverse causality.

OLS Assumptions

1. Linearity
   * The model **in the population** (the 'true' model) can be written as a linear combination of variables and coefficients: $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 \ldots \beta_p X_p + \epsilon$.
2. Random Sampling
   * We have a **random sample** of n observations, following the population model.
3. No Perfect Collinearity
   * In the sample, none of the independent variables are **constant**, and there are no **exact linear relationships** between independent variables.
4. Zero Conditional Mean (Exogeneity)
   * The error term has a **mean of zero** and is **unrelated to any of the *X*s**. *Many potential violations in practice:* omitted variable bias, non-linear relationships, reverse causality.

If assumptions 1–4 are satisfied, our OLS coefficient estimates are unbiased

OLS Assumptions

1. Linearity

   * The model **in the population** (the 'true' model) can be written as a linear combination of variables and coefficients: $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 \ldots \beta_p X_p + \epsilon$.

2. Random Sampling

   * We have a **random sample** of n observations, following the population model.

3. No Perfect Collinearity

   * In the sample, none of the independent variables are **constant**, and there are no **exact linear relationships** between independent variables.

4. Zero Conditional Mean (Exogeneity)

   * The error term has a **mean of zero** and is **unrelated to any of the *X*s**. *Many potential violations in practice:* omitted variable bias, non-linear relationships, reverse causality.

If assumptions 1–4 are satisfied, our OLS coefficient estimates are unbiased

* We also assume 5. Homoskedasticity and 6. Normality, rushed through last time…

Homoskedasticity

Homoskedasticity

- Default Standard Errors are computed assuming the population regression has constant variance (**homoskedasticity**) across values of the *X*s.

Homoskedasticity

* Default Standard Errors are computed assuming the population regression has constant variance (**homoskedasticity**) across values of the *X*s.
* We may diagnose that this is likely not the case (**heteroskedasticity**) from plotting the residuals against the independent variable.

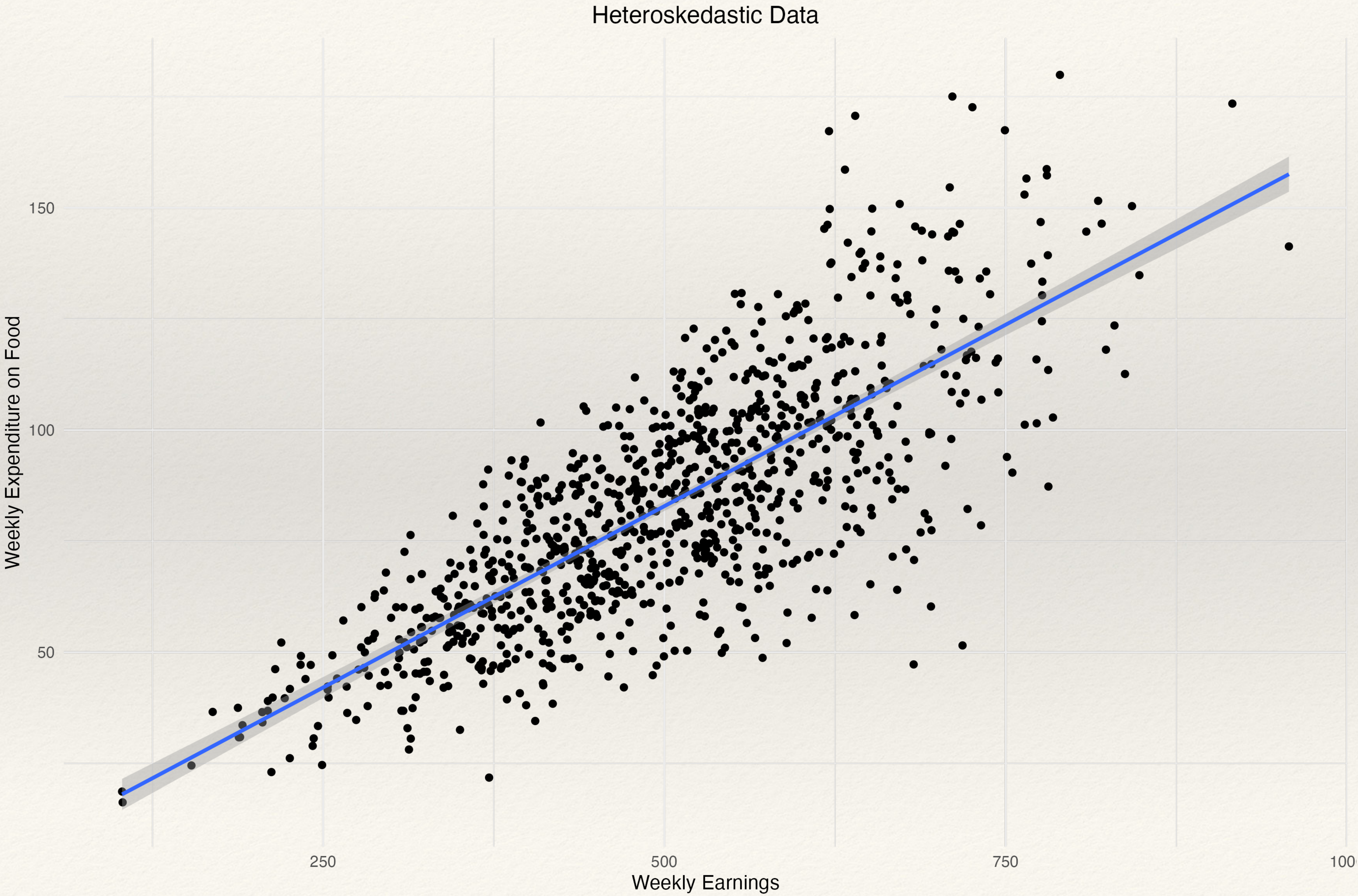
Homoskedasticity

* Default Standard Errors are computed assuming the population regression has constant variance (**homoskedasticity**) across values of the *X*s.
* We may diagnose that this is likely not the case (**heteroskedasticity**) from plotting the residuals against the independent variable.
* Biases standard errors, but not coefficients.

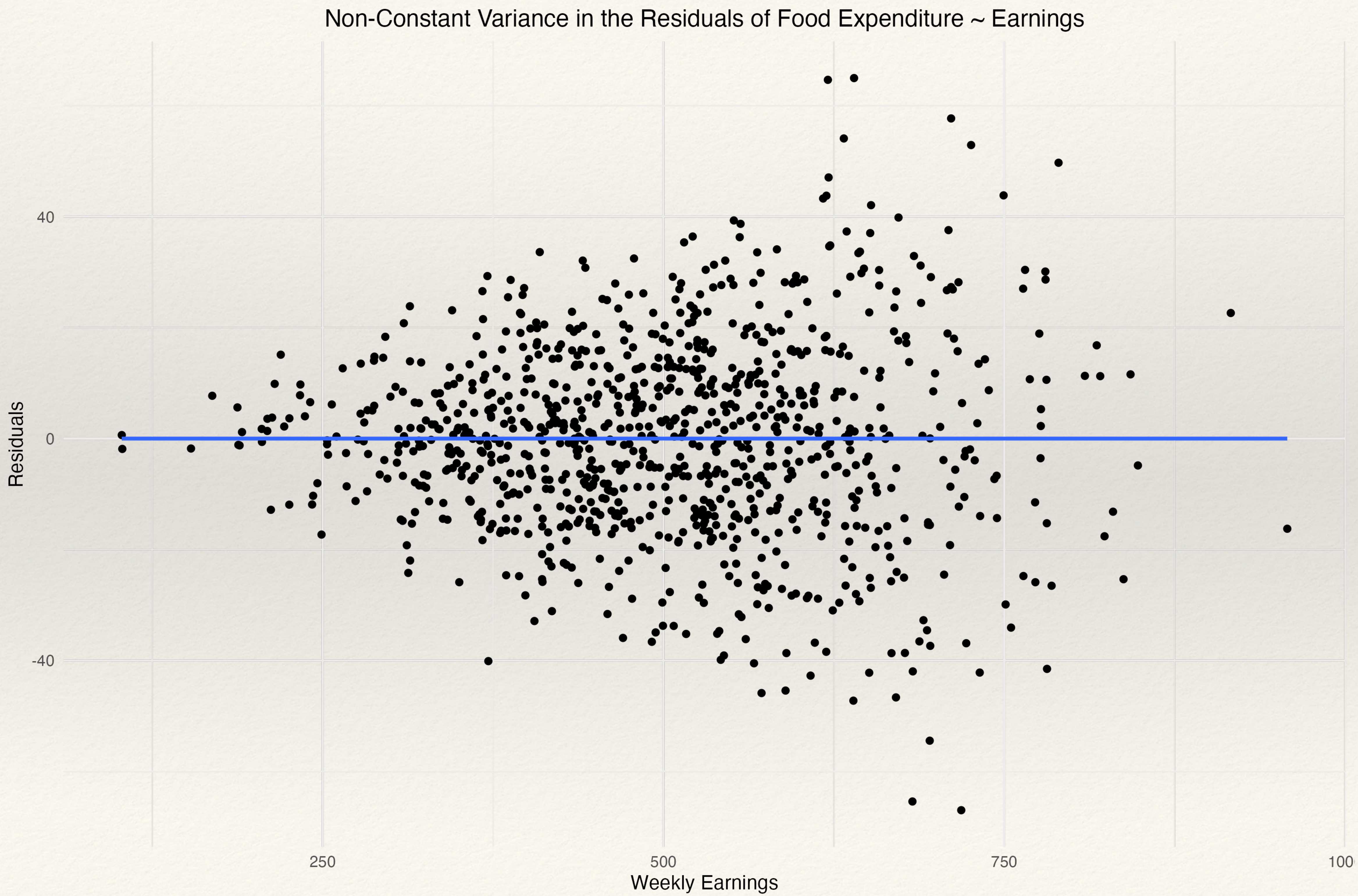
Homoskedasticity

- Default Standard Errors are computed assuming the population regression has constant variance (**homoskedasticity**) across values of the *X*s.
- We may diagnose that this is likely not the case (**heteroskedasticity**) from plotting the residuals against the independent variable.
- Biases standard errors, but not coefficients.
- One popular fix: **heteroskedasticity-consistent standard errors** (more conservative).

Violation of Homoskedasticity Assumption



Violation of Homoskedasticity Assumption



Normality of the Error Term

Normality of the Error Term

- The error term is independent of the explanatory variables (zero conditional mean), has constant variance (homoskedasticity) and is **normally distributed (normality).**

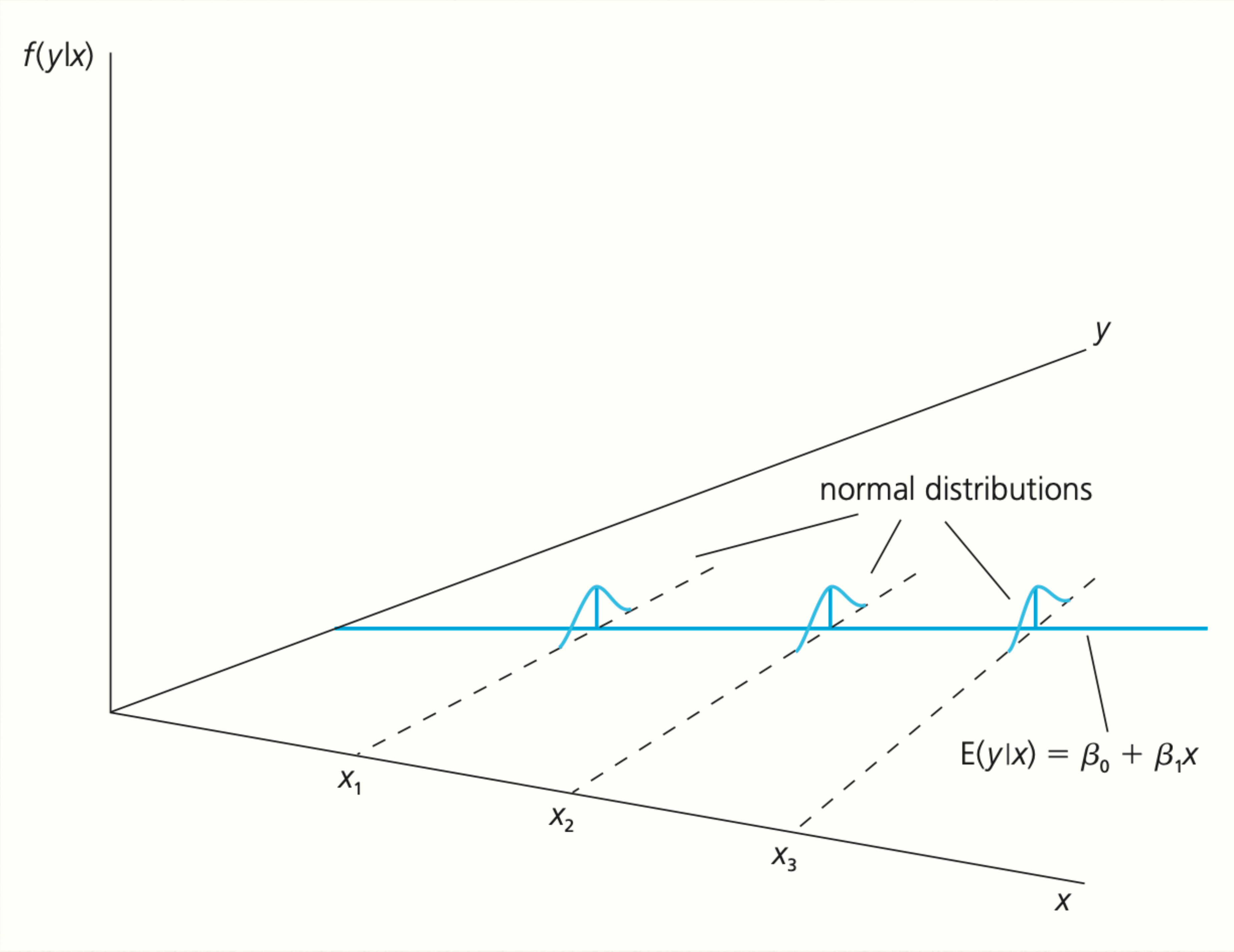
Normality of the Error Term

* The error term is independent of the explanatory variables (zero conditional mean), has constant variance (homoskedasticity) and is **normally distributed (normality).**
* To calculate the *t*-statistic and the *p*-value, we need to know the full sampling distribution of the estimate. This depends on (unobserved) population errors.

Normality of the Error Term

* The error term is independent of the explanatory variables (zero conditional mean), has constant variance (homoskedasticity) and is **normally distributed (normality).**
* To calculate the *t*-statistic and the *p*-value, we need to know the full sampling distribution of the estimate. This depends on (unobserved) population errors.
* Useful to assume that they are normally distributed (as we model them as 'random').

Normality of the Error Term



Normality of the Error Term

Normality of the Error Term

* Least worrisome of the OLS assumptions:

Normality of the Error Term

- Least worrisome of the OLS assumptions:
    - Only affects inferential statistics, not coefficients or S.E.

Normality of the Error Term

- Least worrisome of the OLS assumptions:
  - Only affects inferential statistics, not coefficients or S.E.
  - With correct model specifications, your residuals will vary (approximately) randomly. **In large samples**, this will give you a normal distribution. But **no guarantee in small samples**.

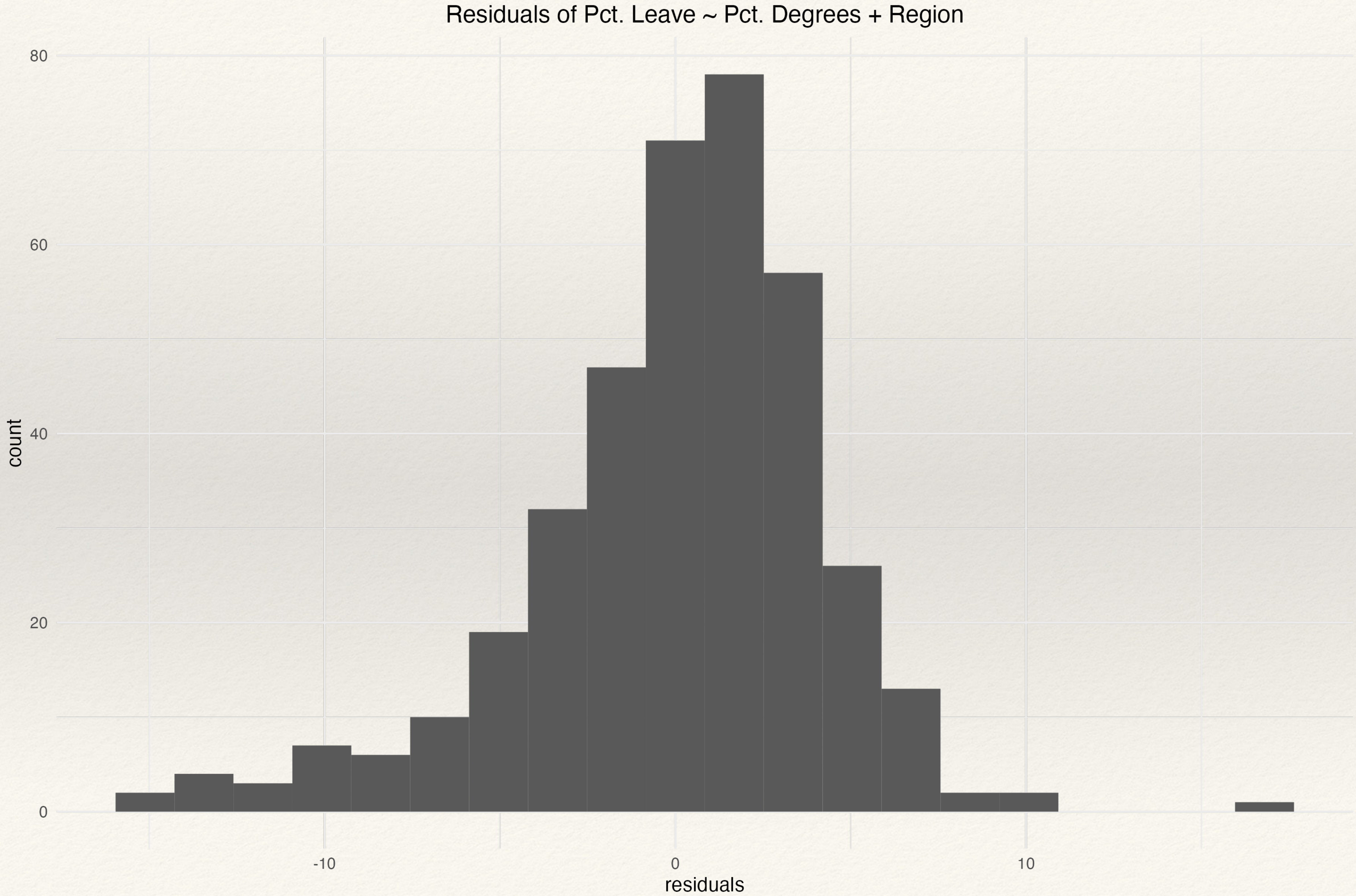
Normality of the Error Term

* Least worrisome of the OLS assumptions:
  * Only affects inferential statistics, not coefficients or S.E.
  * With correct model specifications, your residuals will vary (approximately) randomly. **In large samples**, this will give you a normal distribution. But **no guarantee in small samples**.
  * Non-normal errors are usually the result of linearity assumption not holding. If you fix that, things are usually fine.

Normality of the Error Term

- Least worrisome of the OLS assumptions:
  - Only affects inferential statistics, not coefficients or S.E.
  - With correct model specifications, your residuals will vary (approximately) randomly. **In large samples**, this will give you a normal distribution. But **no guarantee in small samples**.
  - Non-normal errors are usually the result of linearity assumption not holding. If you fix that, things are usually fine.
  - Visual check: histogram of residuals.

Normality of the Error Term



What Variables Should I Control For?

What Variables Should I Control For?

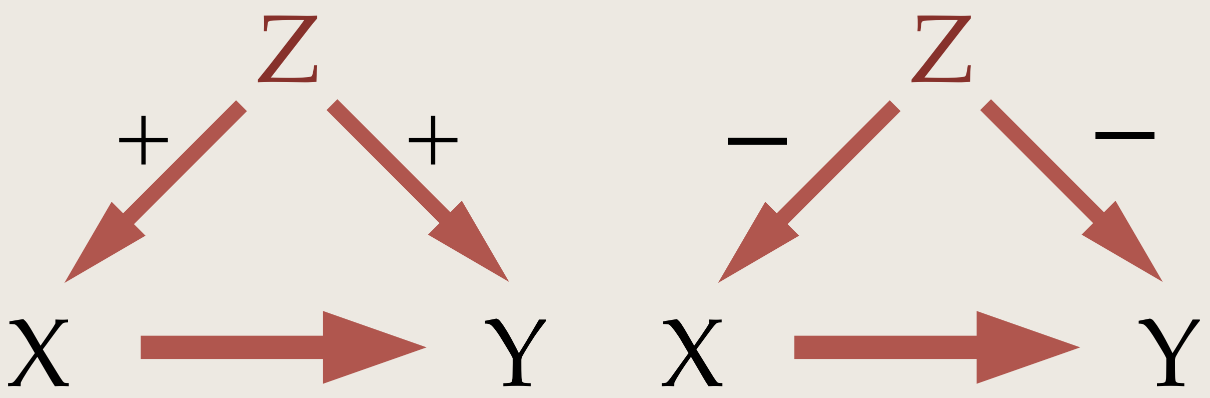
- Goal of ‘controlling’: *accounting for omitted variable bias*.

What Variables Should I Control For?

- Goal of ‘controlling’: *accounting for omitted variable bias*.
- Visually, close ’back doors’ to the causal path $X \rightarrow Y$

What Variables Should I Control For?

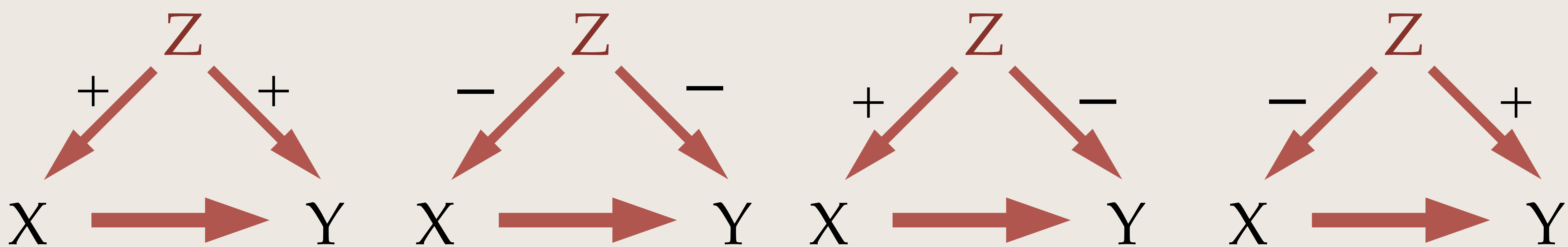
- Goal of ‘controlling’: *accounting for omitted variable bias*.
- Visually, close ’back doors’ to the causal path $X \rightarrow Y$



Without controlling for *Z*, the ATE of *X* on *Y* is positively biased

What Variables Should I Control For?

- Goal of 'controlling': *accounting for omitted variable bias*.
- Visually, close 'back doors' to the causal path $X \rightarrow Y$

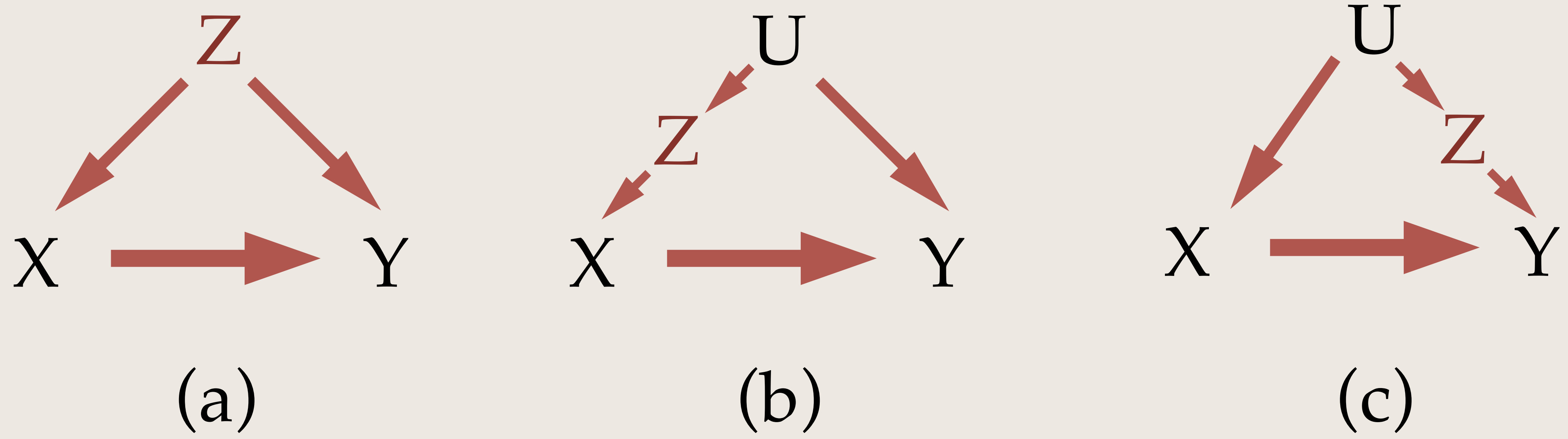


Without controlling for *Z*, the ATE of *X* on *Y* is positively biased

Without controlling for *Z*, the ATE of *X* on *Y* is negatively biased

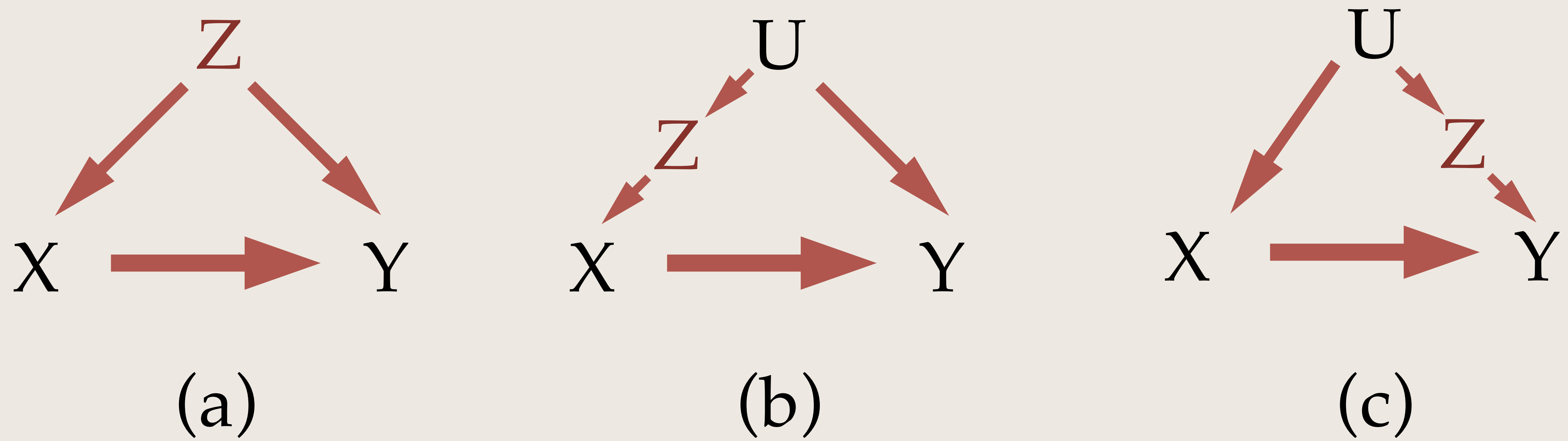
What Variables Should I Control For?

What Variables Should I Control For?



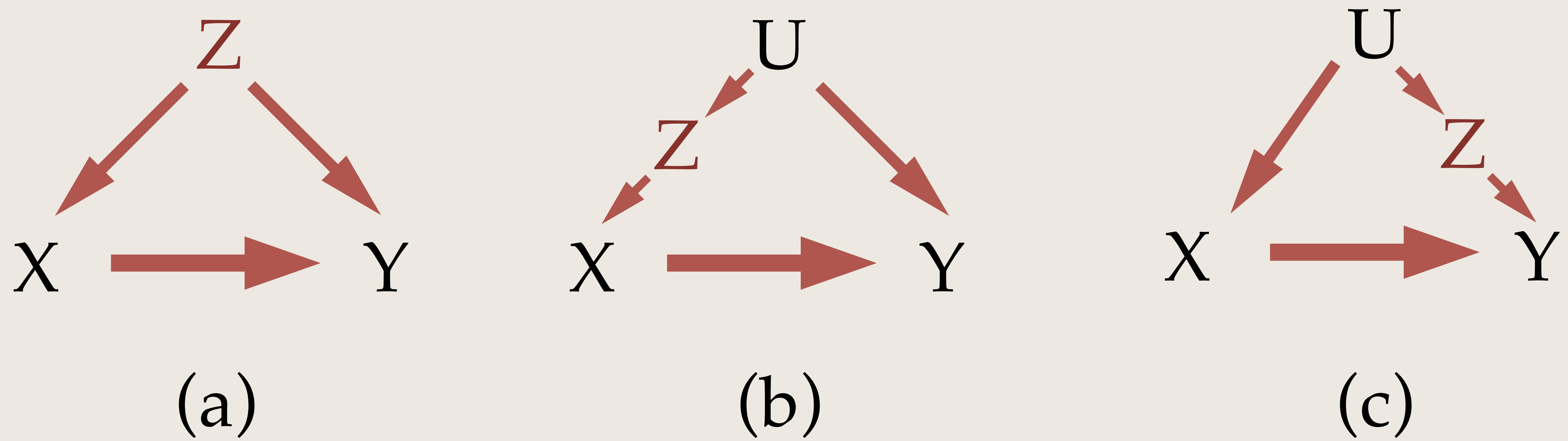
* Adapted from Cinelli et al (2022)

What Variables Should I Control For?



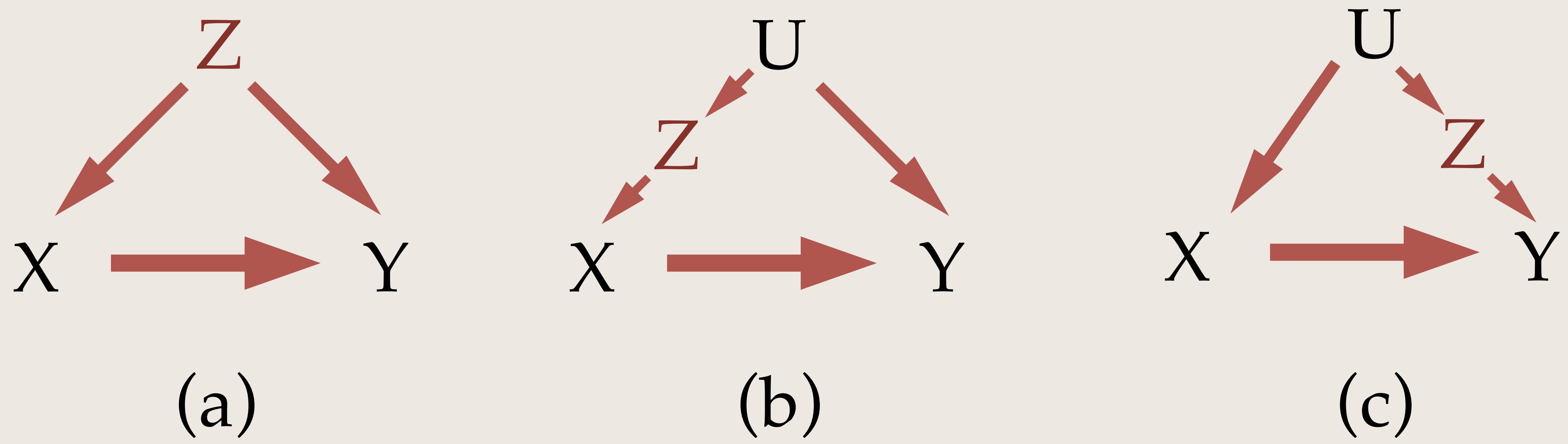
- Adapted from Cinelli et al (2022)
  - **Back-door criterion**: Z is a 'good control' if

What Variables Should I Control For?



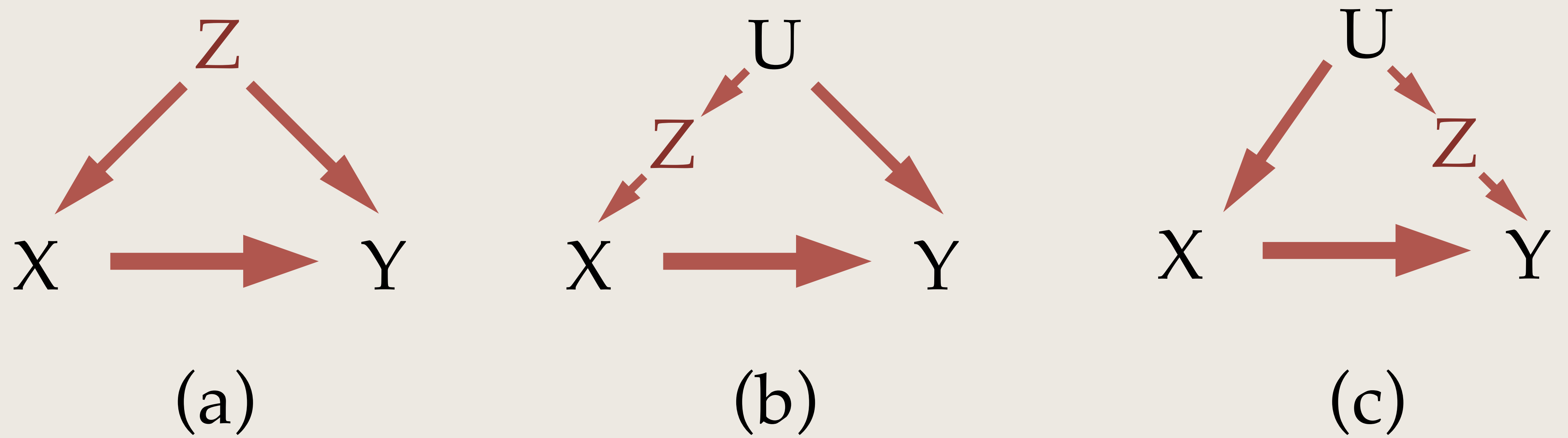
* Adapted from Cinelli et al (2022)
    * **Back-door criterion**: Z is a 'good control' if
        1. Z is not a descendant of X (not **post-treatment**), and

What Variables Should I Control For?



- Adapted from Cinelli et al (2022)
  - **Back-door criterion**: Z is a 'good control' if
    1. Z is not a descendant of X (not **post-treatment**), and
    2. Z blocks a path between X and Y **that contains an arrow into X**.

What Variables Should I Control For?



- Adapted from Cinelli et al (2022)
  - **Back-door criterion**: Z is a 'good control' if
    1. Z is not a descendant of X (not **post-treatment**), and
    2. Z blocks a path between X and Y **that contains an arrow into X**.
       - i.e. Z is a **common cause** of X and Y (*a*) or is the **mediator** of the relationship between an unobserved common cause U and either X or Y (respectively, *b* and *c*) .

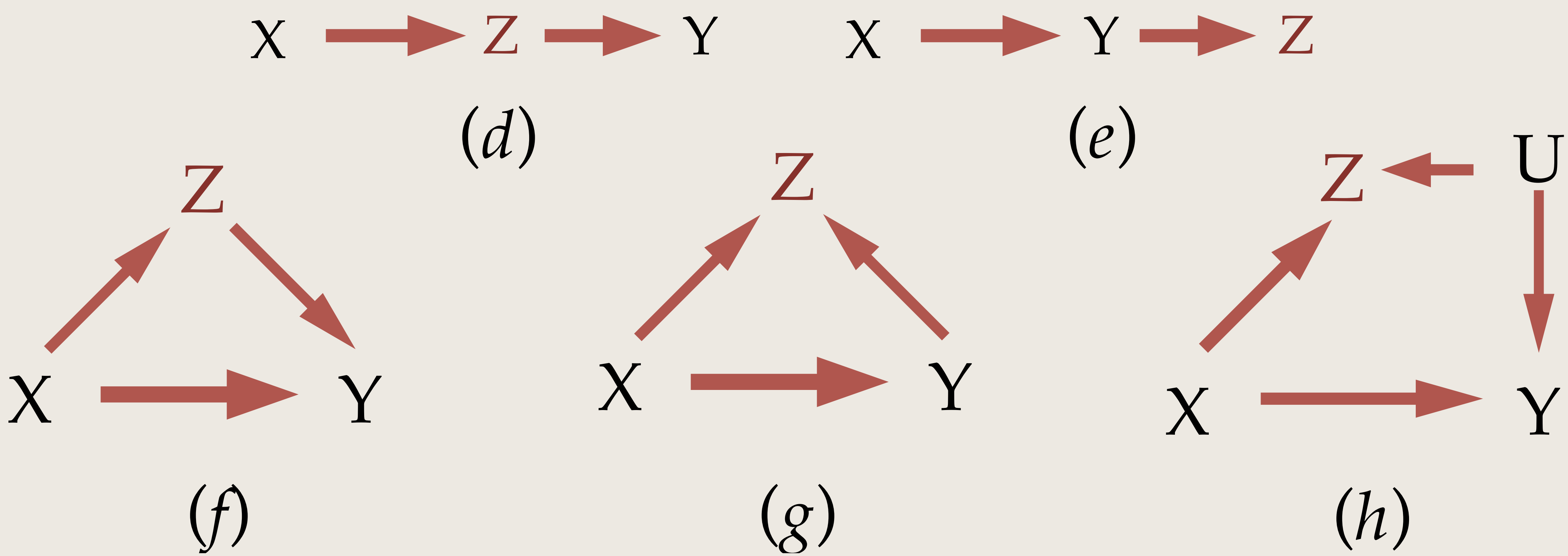
What Variables Should I **Not** Control For?

What Variables Should I **Not** Control For?

- If Z descends from of X (post-treatment variable): **bad idea**.

What Variables Should I **Not** Control For?

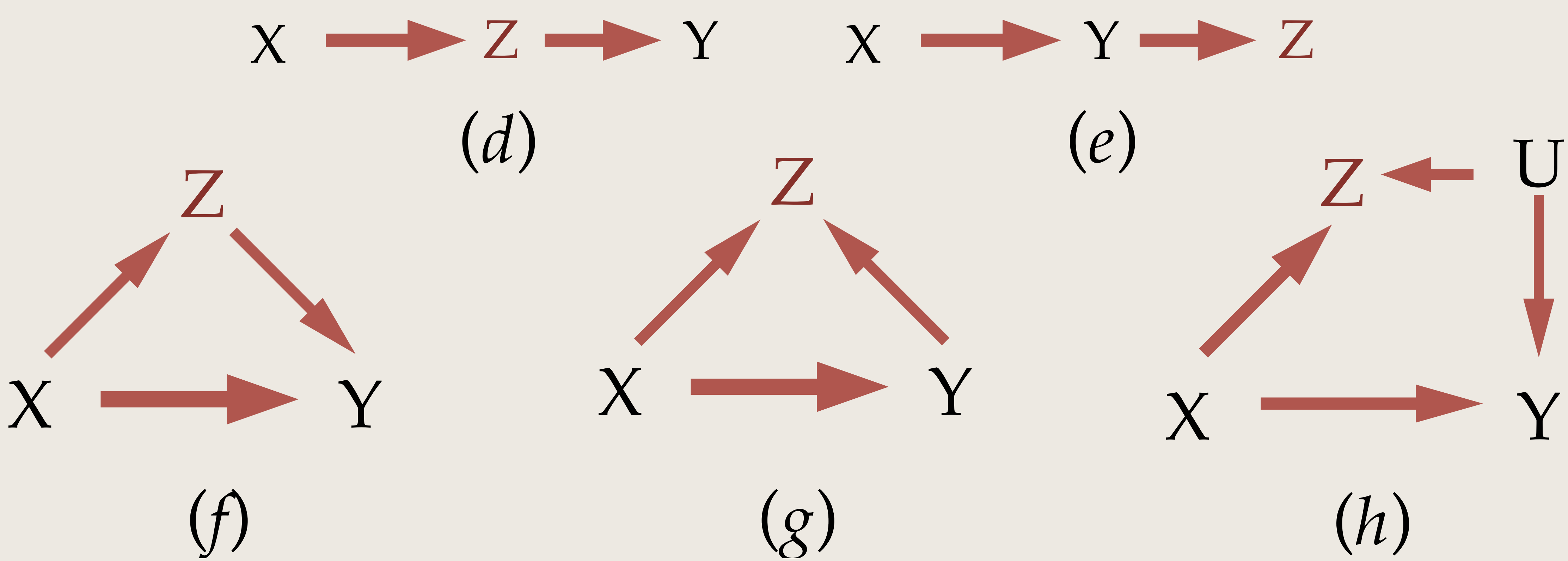
* If Z descends from of X (post-treatment variable): **bad idea**.



* Adapted from Cinelli et al (2022)

What Variables Should I **Not** Control For?

- If Z descends from of X (post-treatment variable): **bad idea**.
- These can: (1) **block the causal path** $X \rightarrow Y$ (*d*), (2) are **effects** of the outcome (*e*), or (3) **open a backdoor path** to a previously unbiased causal path (*f*, *g* and *h*).

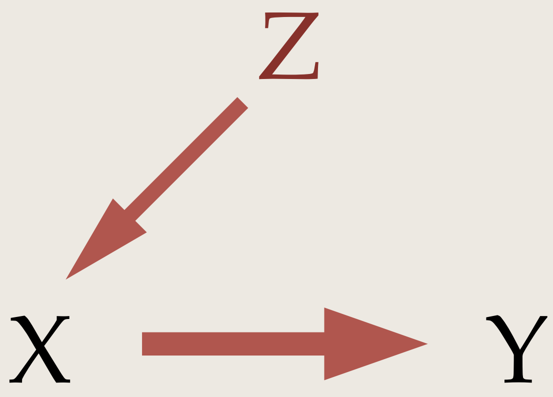


- Adapted from Cinelli et al (2022)

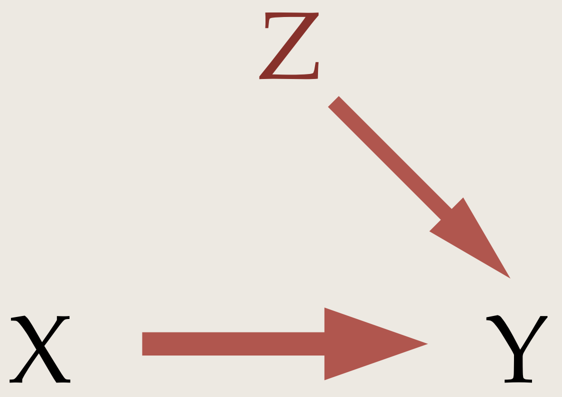
Control for **all** pre-treatment variables?

Control for **all** pre-treatment variables?

- Usually pre-treatment variables are good (*a*, *b* and *c*) or **neutral** (*i* and *j*).



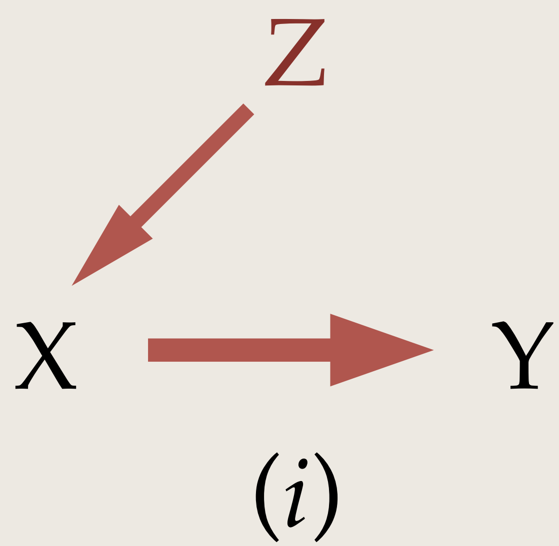
(*i*)

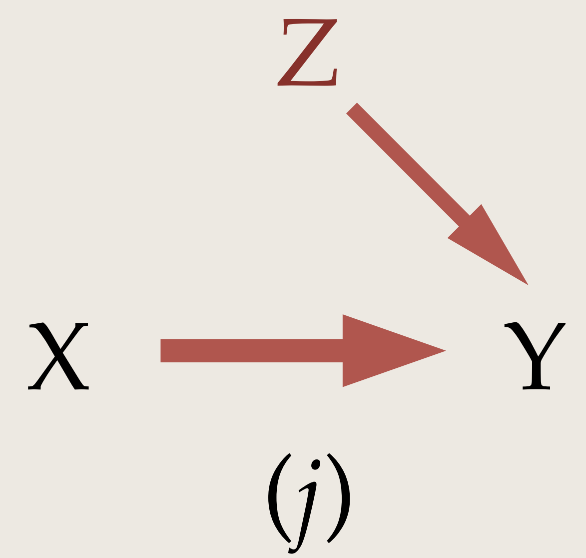


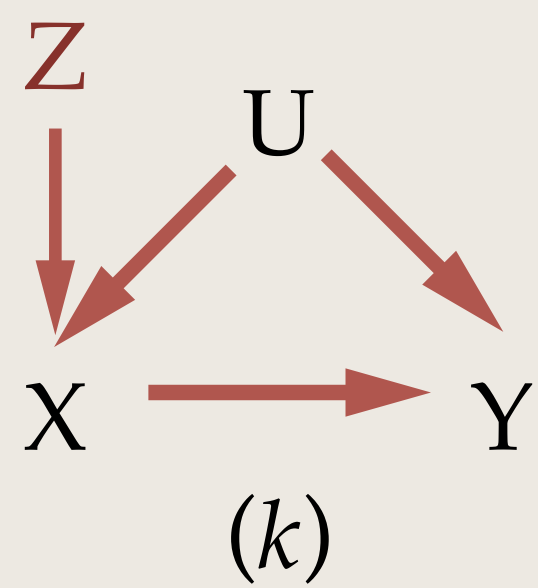
(*j*)

Control for **all** pre-treatment variables?

- Usually pre-treatment variables are good (*a*, *b* and *c*) or **neutral** (*i* and *j*).
- But **in presence of unobserved confounders**, 'pointless' control can make existing bias worse (*k*).

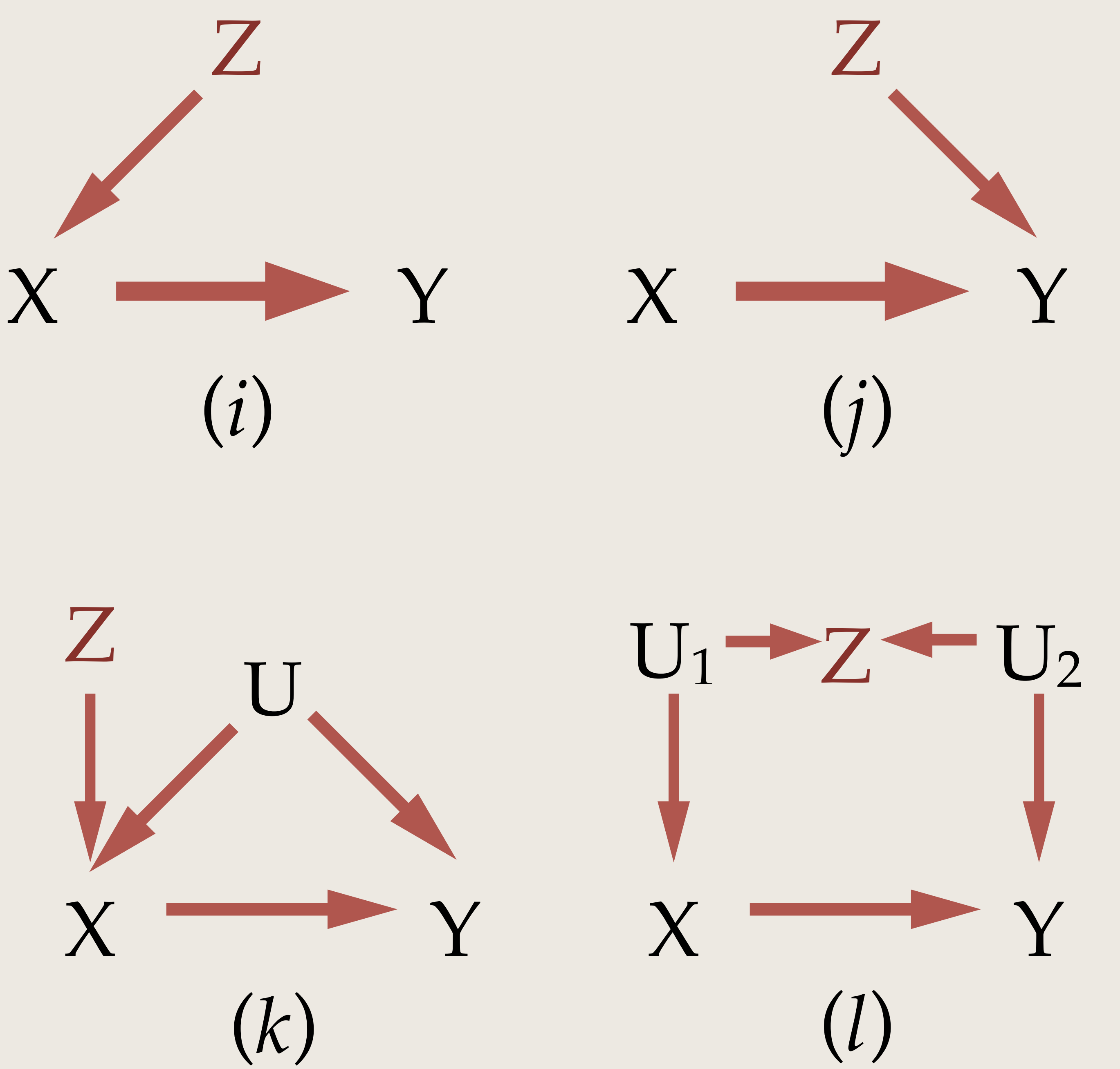






Control for **all** pre-treatment variables?

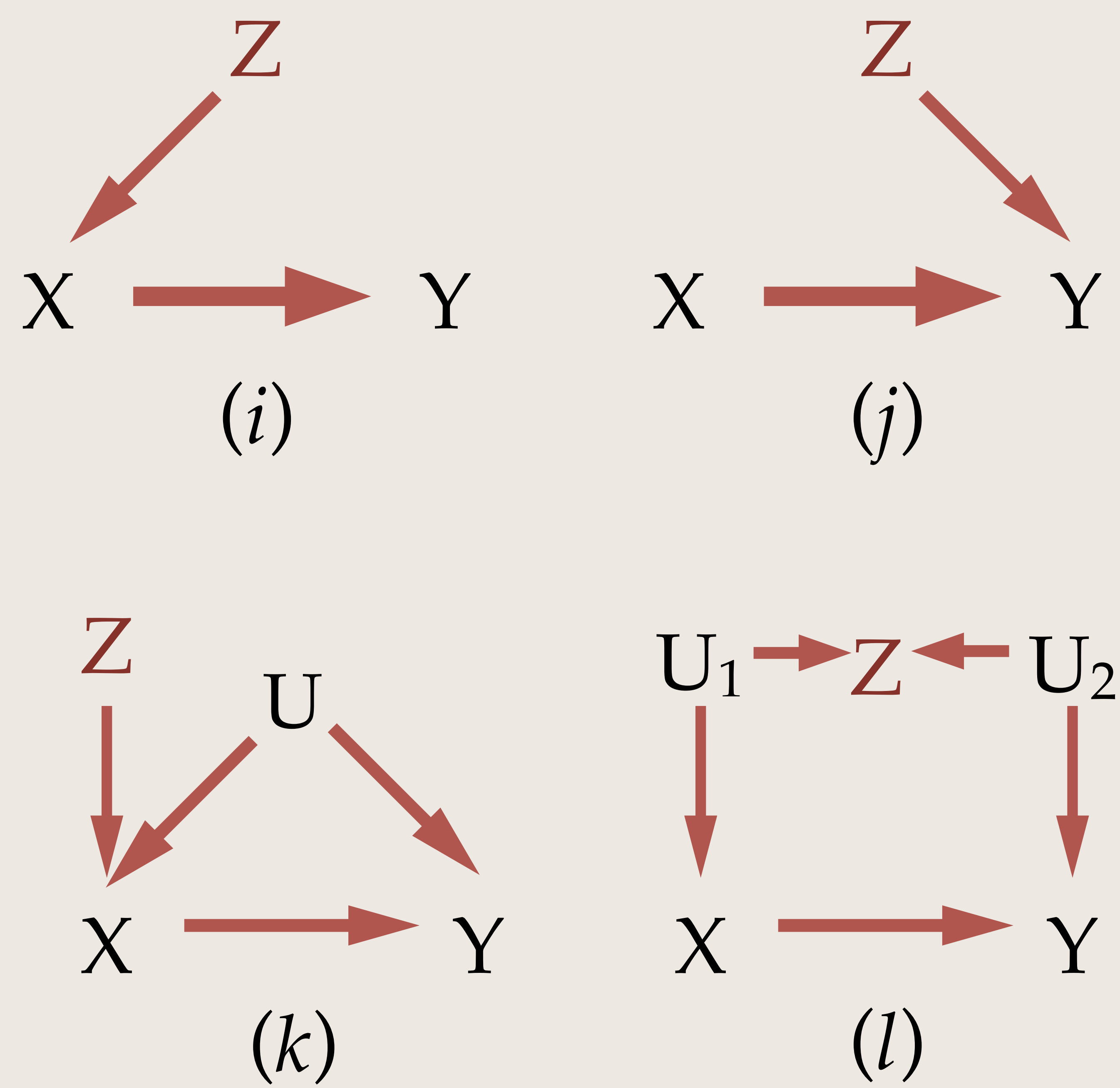
* Usually pre-treatment variables are good (*a*, *b* and *c*) or **neutral** (*i* and *j*).
* But **in presence of unobserved confounders**, 'pointless' control can make existing bias worse (*k*).
* Also, they can be a problem if they **open a backdoor path** (*l*, collider bias).



Adapted from Cinelli et al (2022)

Control for **all** pre-treatment variables?

- Usually pre-treatment variables are good (*a*, *b* and *c*) or **neutral** (*i* and *j*).
- But **in presence of unobserved confounders**, 'pointless' control can make existing bias worse (*k*).
- Also, they can be a problem if they **open a backdoor path** (*l*, collider bias).
- Bottom line: **theory** should inform your choice of controls, not data availability.
-



Adapted from Cinelli et al (2022)

Interactions

Example

Example

- Are graduates more worried about climate change?

Example

- Are graduates more worried about climate change?
  - $\text{Climate Worry} = \alpha + \beta\,\text{Degree} + \epsilon$

Example

- Are graduates more worried about climate change?
  - Climate Worry $= \alpha + \beta$ Degree $+ \epsilon$
- What's a possible confounder?

Example

- Are graduates more worried about climate change?
  - Climate Worry $= \alpha + \beta$ Degree $+ \epsilon$
- What's a possible confounder?
  - **Ideology**? Left-wingers are more likely to go to university, and being left-wing makes you worry about climate.

Example

- Are graduates more worried about climate change?
  - $\text{Climate Worry} = \alpha + \beta\,\text{Degree} + \epsilon$
- What's a possible confounder?
  - **Ideology**? Left-wingers are more likely to go to university, and being left-wing makes you worry about climate.
  - Ideology may be partly endogenous to education, but for now let's make peace with that, and fit:

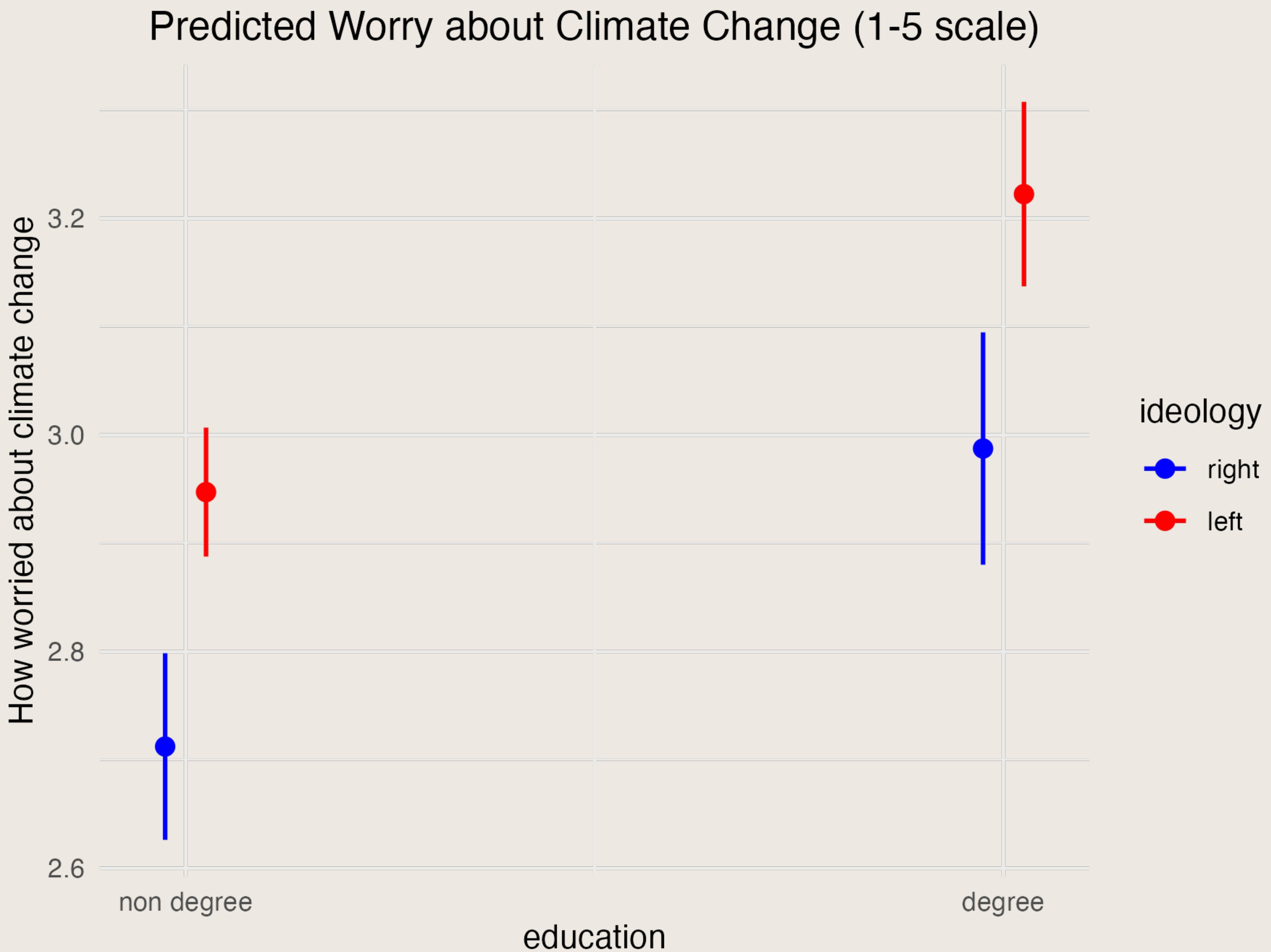
Example

- Are graduates more worried about climate change?
  - $\text{Climate Worry} = \alpha + \beta\,\text{Degree} + \epsilon$
- What's a possible confounder?
  - **Ideology**? Left-wingers are more likely to go to university, and being left-wing makes you worry about climate.
  - Ideology may be partly endogenous to education, but for now let's make peace with that, and fit:
  - $\text{Climate Worry} = \alpha + \beta_1\,\text{Degree} + \beta_2\,\text{Left} + \epsilon$

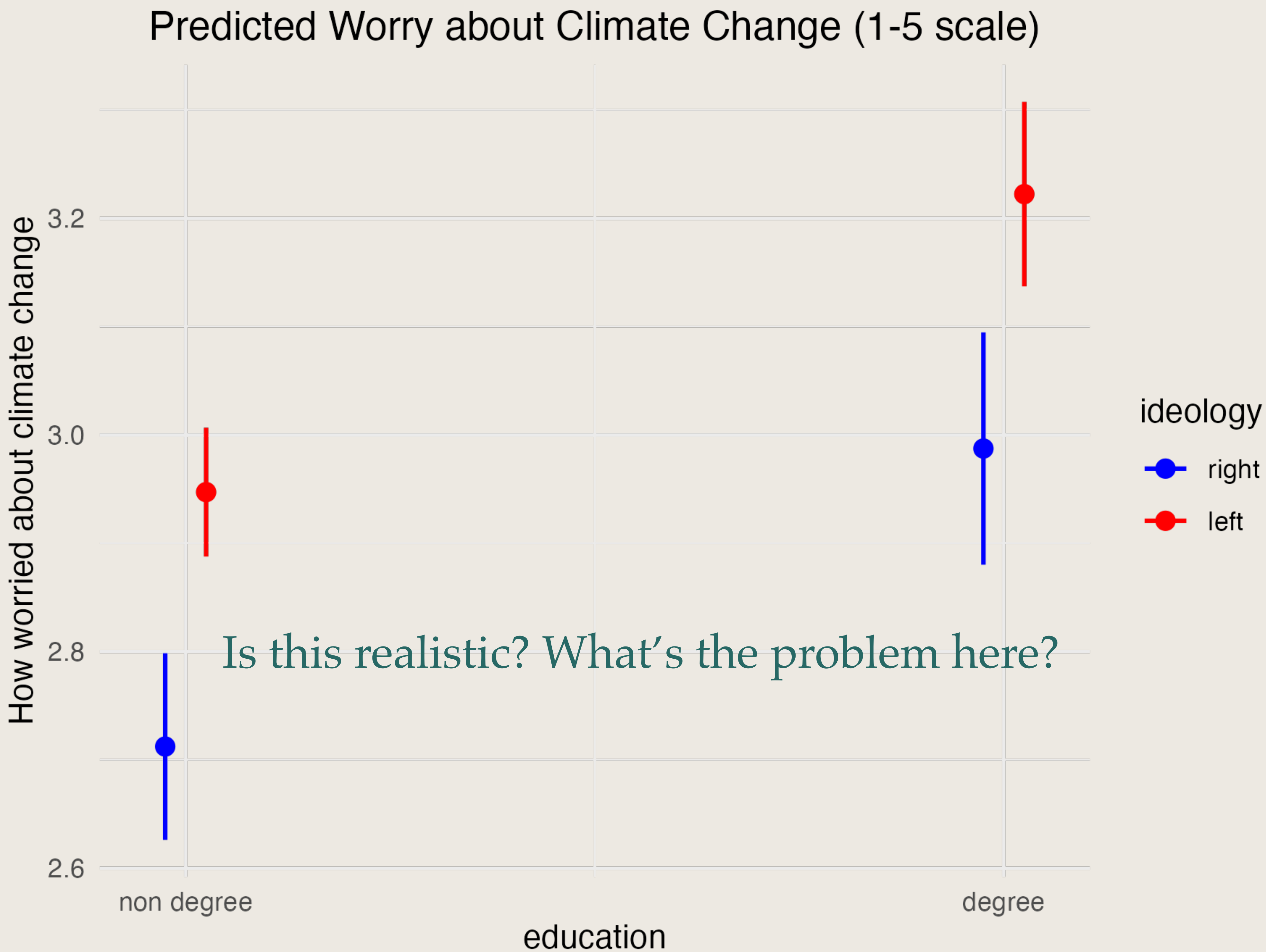
Example: Regression Table

	Dependent variable:
	wrclmch
educationdegree	0.275***
	(0.049)
ideologyleft	0.235***
	(0.049)
Constant	2.712***
	(0.044)
Observations	1,699
R2	0.031
Adjusted R2	0.030
Residual Std. Error	0.923 (df = 1696)
F Statistic	27.511*** (df = 2; 1696)
Note:	*p<0.1; **p<0.05; ***p<0.01

Example: Predicted Values Plot

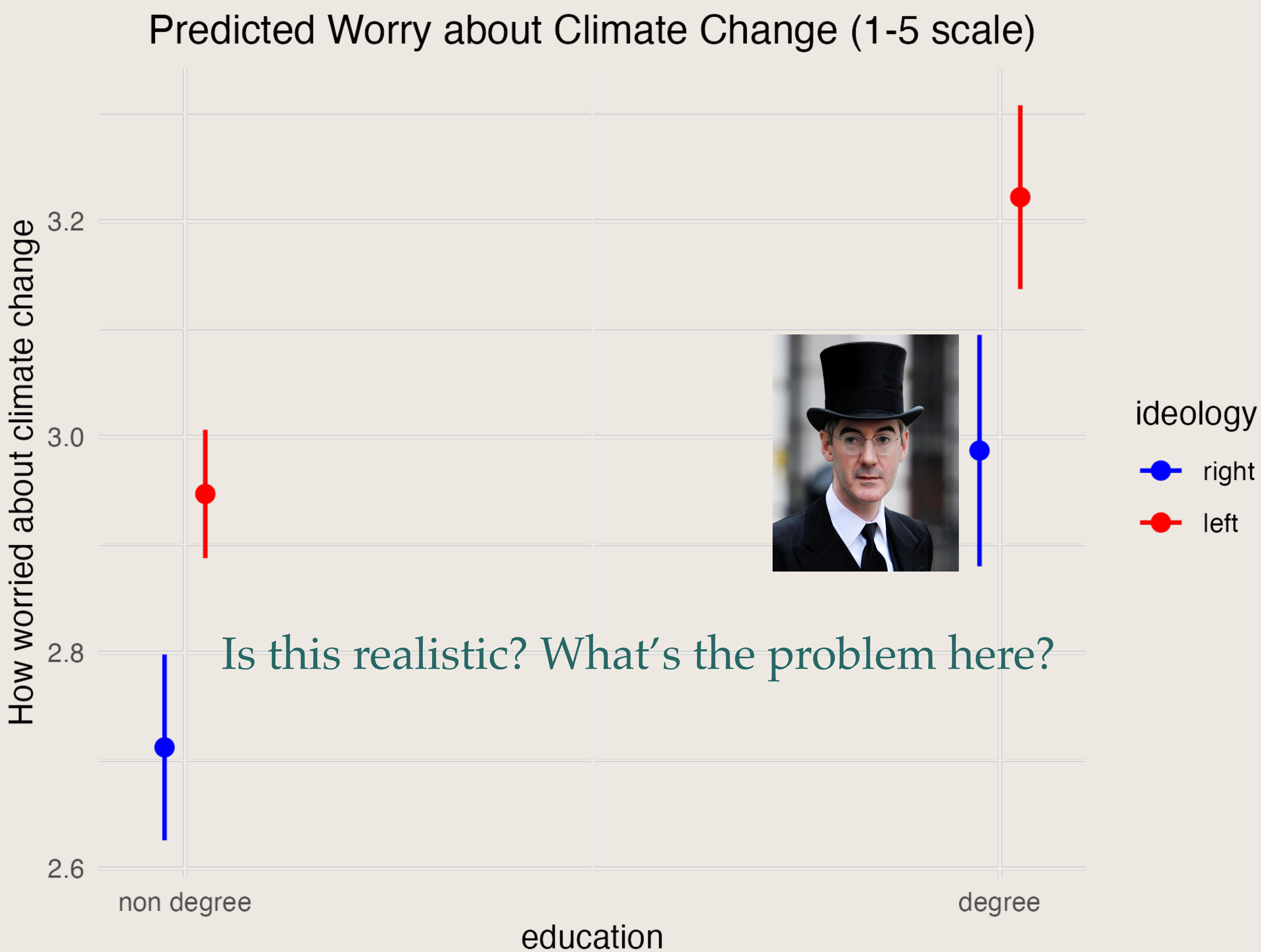


Example: Predicted Values Plot



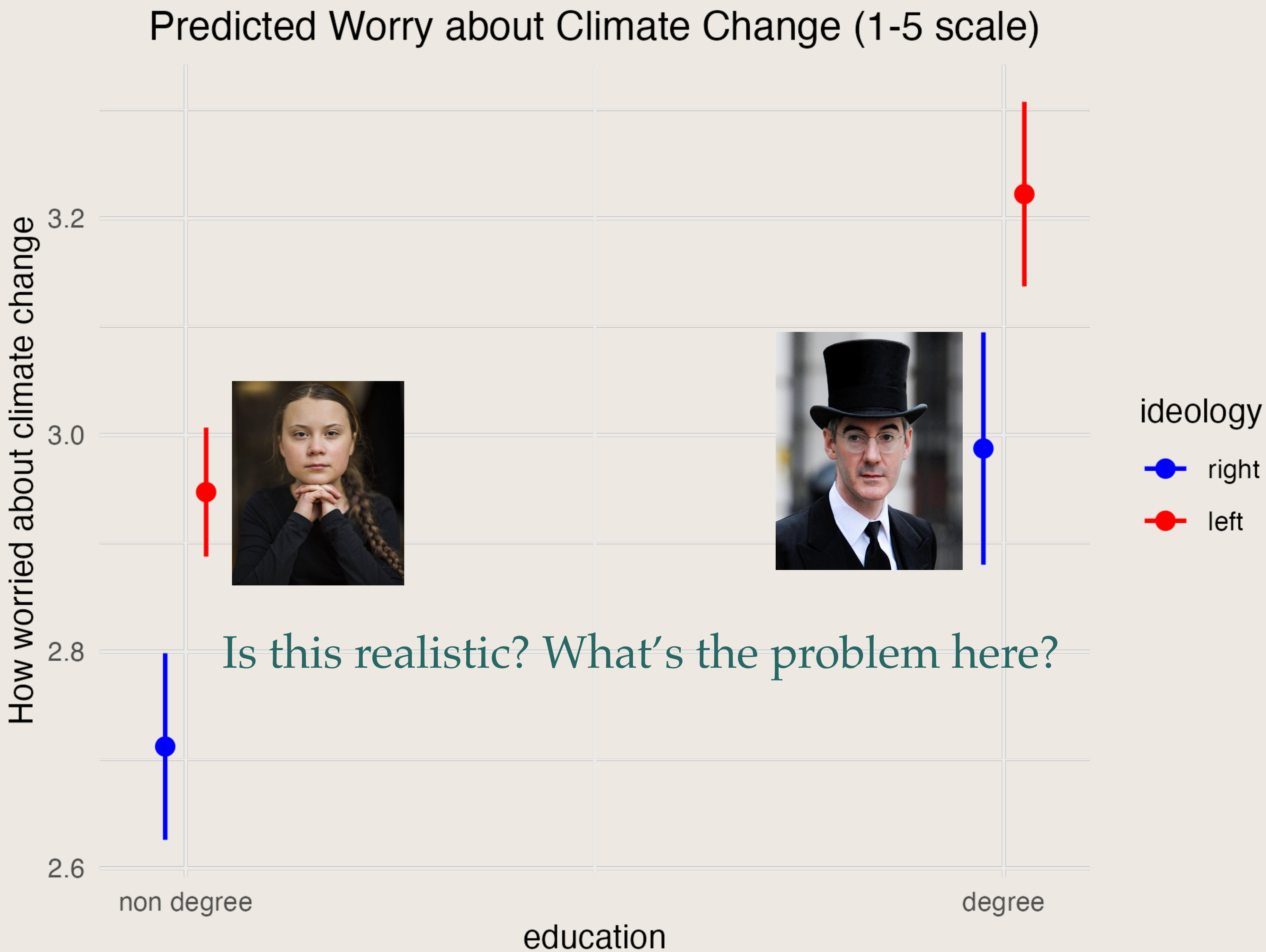
Is this realistic? What's the problem here?

Example: Predicted Values Plot



Is this realistic? What's the problem here?

Example: Predicted Values Plot



Is this realistic? What's the problem here?

Solution: Interaction Term

Solution: Interaction Term

$$\text{Climate Worry} = \alpha + \beta_1 \text{ Degree} + \beta_2 \text{ Left} + \beta_3 (\textbf{Degree} \times \textbf{Left}) + \epsilon$$

Solution: Interaction Term

$$\text{Climate Worry} = \alpha + \beta_1 \text{ Degree} + \beta_2 \text{ Left} + \beta_3 (\textbf{Degree} \times \textbf{Left}) + \epsilon$$

	Dependent variable:
	Climate Worry (1–5)
Intercept	2.793*** (0.05)
Degree	−0.012 (0.09)
Left	0.121** (0.06)
Degree × Left	0.398*** (0.11)

Solution: Interaction Term

$$\text{Climate Worry} = \alpha + \beta_1 \text{ Degree} + \beta_2 \text{ Left} + \beta_3 (\textbf{Degree} \times \textbf{Left}) + \epsilon$$

	Dependent variable:
	Climate Worry (1–5)
Intercept	2.793*** (0.05)
Degree	−0.012 (0.09)
Left	0.121** (0.06)
Degree × Left	0.398*** (0.11)

	Degree = 0	Degree = 1
Left = 0		
Left = 1		

Solution: Interaction Term

Climate Worry $= \alpha + \beta_1$ Degree $+ \beta_2$ Left $+\beta_3$**(Degree $\times$ Left)** $+ \epsilon$

Dependent variable:

	Climate Worry (1–5)
Intercept	2.793^{***} (0.05)
Degree	−0.012 (0.09)
Left	0.121^{**} (0.06)
Degree × Left	0.398^{***} (0.11)

	Degree = 0	Degree = 1
Left = 0	2.793	
Left = 1		

- If Degree = 0 and Left = 0, then

$$\hat{Y} = \alpha + \beta_1(0) + \beta_2(0) + \beta_3(0 \times 0) = \alpha$$

Solution: Interaction Term

$$\text{Climate Worry} = \alpha + \beta_1 \text{ Degree} + \beta_2 \text{ Left} + \beta_3(\textbf{Degree} \times \textbf{Left}) + \epsilon$$

	Dependent variable:
	Climate Worry (1–5)
Intercept	2.793*** (0.05)
Degree	−0.012 (0.09)
Left	0.121** (0.06)
Degree × Left	0.398*** (0.11)

	Degree = 0	Degree = 1
Left = 0	2.793	
Left = 1		

Solution: Interaction Term

Climate Worry $= \alpha + \beta_1$ Degree $+ \beta_2$ Left $+\beta_3$**(Degree $\times$ Left)** $+ \epsilon$

	Dependent variable:
	Climate Worry (1–5)
Intercept	2.793*** (0.05)
Degree	−0.012 (0.09)
Left	0.121** (0.06)
Degree × Left	0.398*** (0.11)

	Degree = 0	Degree = 1
Left = 0	2.793	2.781
Left = 1		

- If Degree = 1 and Left = 0, then

$$\hat{Y} = \alpha + \beta_1(1) + \beta_2(0) + \beta_3(1 \times 0) = \alpha + \beta_1$$

Solution: Interaction Term

$$\text{Climate Worry} = \alpha + \beta_1 \text{ Degree} + \beta_2 \text{ Left} + \beta_3(\textbf{Degree} \times \textbf{Left}) + \epsilon$$

	Dependent variable: Climate Worry (1–5)
Intercept	2.793^{***} (0.05)
Degree	−0.012 (0.09)
Left	0.121^{**} (0.06)
Degree × Left	0.398^{***} (0.11)

	Degree = 0	Degree = 1
Left = 0	2.793	2.781
Left = 1		

Solution: Interaction Term

$$\text{Climate Worry} = \alpha + \beta_1 \text{ Degree} + \beta_2 \text{ Left} + \beta_3(\textbf{Degree} \times \textbf{Left}) + \epsilon$$

	Dependent variable:
	Climate Worry (1–5)
Intercept	2.793*** (0.05)
Degree	−0.012 (0.09)
Left	0.121** (0.06)
Degree × Left	0.398*** (0.11)

	Degree = 0	Degree = 1
Left = 0	2.793	2.781
Left = 1	2.914	

- If Degree = 0 and Left = 1, then

$$\hat{Y} = \alpha + \beta_1(0) + \beta_2(1) + \beta_3(0 \times 1) = \alpha + \beta_2$$

Solution: Interaction Term

$$\text{Climate Worry} = \alpha + \beta_1 \text{ Degree} + \beta_2 \text{ Left} + \beta_3\textbf{(Degree} \times \textbf{Left)} + \epsilon$$

	Dependent variable:
	Climate Worry (1–5)
Intercept	2.793*** (0.05)
Degree	−0.012 (0.09)
Left	0.121** (0.06)
Degree × Left	0.398*** (0.11)

	Degree = 0	Degree = 1
Left = 0	2.793	2.781
Left = 1	2.914	

Solution: Interaction Term

Climate Worry $= \alpha + \beta_1$ Degree $+ \beta_2$ Left $+\beta_3$**(Degree × Left)** $+ \epsilon$

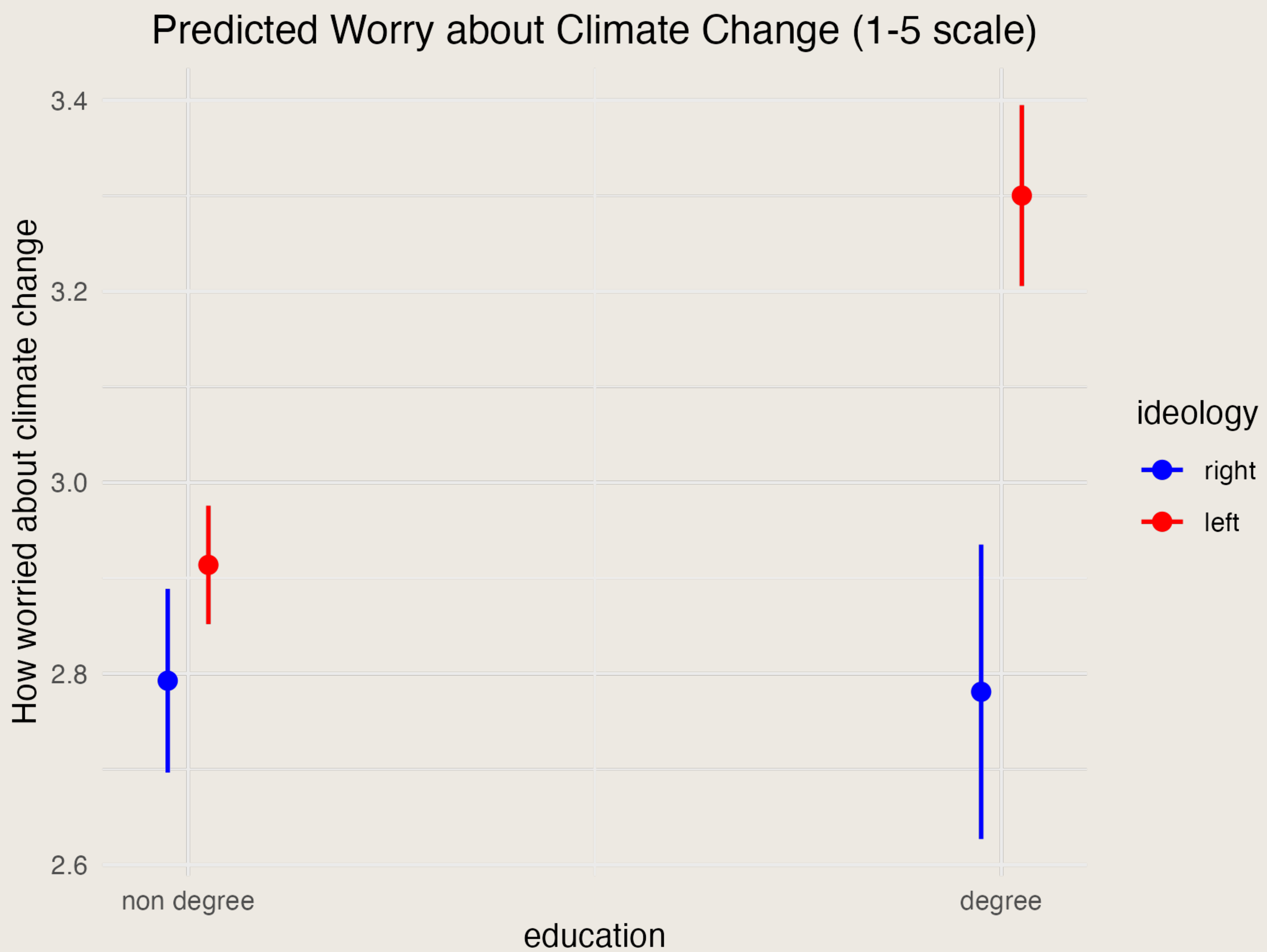
	Dependent variable:
	Climate Worry (1–5)
Intercept	2.793^{***} (0.05)
Degree	−0.012 (0.09)
Left	0.121^{**} (0.06)
Degree × Left	0.398^{***} (0.11)

	Degree = 0	Degree = 1
Left = 0	2.793	2.781
Left = 1	2.914	3.312

- If Degree = 0 and Left = 0, then

$$\hat{Y} = \alpha + \beta_1(1) + \beta_2(1) + \beta_3(1 \times 1) = \alpha + \beta_1 + \beta_2 + \beta_3$$

Solution: Interaction Term



Interaction Terms in R

```
Call:
lm(formula = wrclmch ~ education + ideology + education * ideology,
    data = ess)

Residuals:
     Min       1Q   Median       3Q      Max
-2.30028 -0.79261  0.08619  0.21898  2.21898

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                   2.79261    0.04900  56.997  < 2e-16 ***
educationdegree              -0.01159    0.09257  -0.125  0.90036
ideologyleft                  0.12120    0.05829   2.079  0.03776 *
educationdegree:ideologyleft  0.39805    0.10906   3.650  0.00027 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9192 on 1695 degrees of freedom
  (260 observations deleted due to missingness)
Multiple R-squared:  0.03898,	Adjusted R-squared:  0.03727
F-statistic: 22.91 on 3 and 1695 DF,  p-value: 1.533e-14
```

Interaction Terms in R

Interaction Terms in R

- Note, in R you will get the same result if you run:

Interaction Terms in R

* Note, in R you will get the same result if you run:

```
lm(wrclmch ~ education + ideology + education*ideology, data = ess)

lm(wrclmch ~ education*ideology, data = ess)
```

Interaction Terms in R

- Note, in R you will get the same result if you run:

```
lm(wrclmch ~ education + ideology + education*ideology, data = ess)

lm(wrclmch ~ education*ideology, data = ess)
```

- This is **a really good feature** of `lm()`. Whenever you have interaction terms, you always want to control for the parent terms (*education* and *ideology*) as well as the interaction term.

Interaction Terms in R

- Note, in R you will get the same result if you run:

```
lm(wrclmch ~ education + ideology + education*ideology, data = ess)

lm(wrclmch ~ education*ideology, data = ess)
```

- This is **a really good feature** of `lm()`. Whenever you have interaction terms, you always want to control for the parent terms (*education* and *ideology*) as well as the interaction term.
- There is a way of telling R to include only the interaction term (*education* × *ideology*), but it's best you don't know because this is **wrong** 99% of the times.

Interpreting Interaction Terms

	Dependent variable:
	Climate Worry (1–5)
Intercept	2.793*** (0.05)
Degree	−0.012 (0.09)
Left	0.121** (0.06)
Degree × Left	0.398*** (0.11)

Interpreting Interaction Terms

- We call 'Left' the **moderator**, because it moderates the effect of our **treatment** (Degree).

	Dependent variable:
	Climate Worry (1–5)
Intercept	2.793^{***} (0.05)
Degree	-0.012 (0.09)
Left	0.121^{**} (0.06)
Degree × Left	0.398^{***} (0.11)

Interpreting Interaction Terms

- We call 'Left' the **moderator**, because it moderates the effect of our **treatment** (Degree).
- The coefficient for the **treatment** (Degree) is the effect of the variable when the **moderator** (Left) is zero.

	Dependent variable:
	Climate Worry (1–5)
Intercept	2.793^{***} (0.05)
Degree	-0.012 (0.09)
Left	0.121^{**} (0.06)
Degree × Left	0.398^{***} (0.11)

Interpreting Interaction Terms

- We call ‘Left’ the **moderator**, because it moderates the effect of our **treatment** (Degree).
- The coefficient for the **treatment** (Degree) is the effect of the variable when the **moderator** (Left) is zero.
- The coefficient for the **moderator** (Left) is the effect of the variable when the **treatment** (Degree) is zero.

	Dependent variable:
	Climate Worry (1–5)
Intercept	2.793^{***} (0.05)
Degree	−0.012 (0.09)
Left	0.121^{**} (0.06)
Degree × Left	0.398^{***} (0.11)

Interpreting Interaction Terms

	Dependent variable:
	Climate Worry (1–5)
Intercept	2.793*** (0.05)
Degree	−0.012 (0.09)
Left	0.121** (0.06)
Degree × Left	0.398*** (0.11)

Interpreting Interaction Terms

- The coefficient for the interaction term represents the **difference in the effect of 'Degree'** as we move from **Left = 0 to Left = 1**.

	Dependent variable:
	Climate Worry (1–5)
Intercept	2.793*** (0.05)
Degree	−0.012 (0.09)
Left	0.121** (0.06)
Degree × Left	0.398*** (0.11)

Interpreting Interaction Terms

- The coefficient for the interaction term represents the **difference in the effect of 'Degree'** as we move from **Left = 0 to Left = 1**.
- Statistical significance (*p*-value) of the interaction tests against the null that the effect of the treatment is the same across subgroups.

	Dependent variable:
	Climate Worry (1–5)
Intercept	2.793^{***} (0.05)
Degree	-0.012 (0.09)
Left	0.121^{**} (0.06)
Degree × Left	0.398^{***} (0.11)

Interpreting Interaction Terms

- The coefficient for the interaction term represents the **difference in the effect of 'Degree'** as we move from **Left = 0 to Left = 1**.
- Statistical significance (*p*-value) of the interaction tests against the null that the effect of the treatment is the same across subgroups.
- Here: large and significant — we do have an important interaction.

	Dependent variable:
	Climate Worry (1–5)
Intercept	2.793^{***} (0.05)
Degree	−0.012 (0.09)
Left	0.121^{**} (0.06)
Degree × Left	0.398^{***} (0.11)

Categorical Moderators with More Levels

Categorical Moderators with More Levels

- What about the Centrists? Recode Ideology as a **three-category** variable. Now, the model is:

Categorical Moderators with More Levels

- What about the Centrists? Recode Ideology as a **three-category** variable. Now, the model is:
- Climate Worry $= \alpha + \beta_1$ Degree $+ \beta_2$ Left $+ \beta_3$ **Centrist** $+$ β_4 (Degree $\times$ Left) $+ \beta_5$ **(Degree $\times$ Centrist)** $+ \epsilon$

Categorical Moderators with More Levels

- What about the Centrists? Recode Ideology as a **three-category** variable. Now, the model is:
- Climate Worry $= \alpha + \beta_1$ Degree $+ \beta_2$ Left $+ \beta_3$ **Centrist** $+$ β_4 (Degree $\times$ Left) $+ \beta_5$ **(Degree $\times$ Centrist)** $+ \epsilon$
- In R, just pass the categorical variable:

Categorical Moderators with More Levels

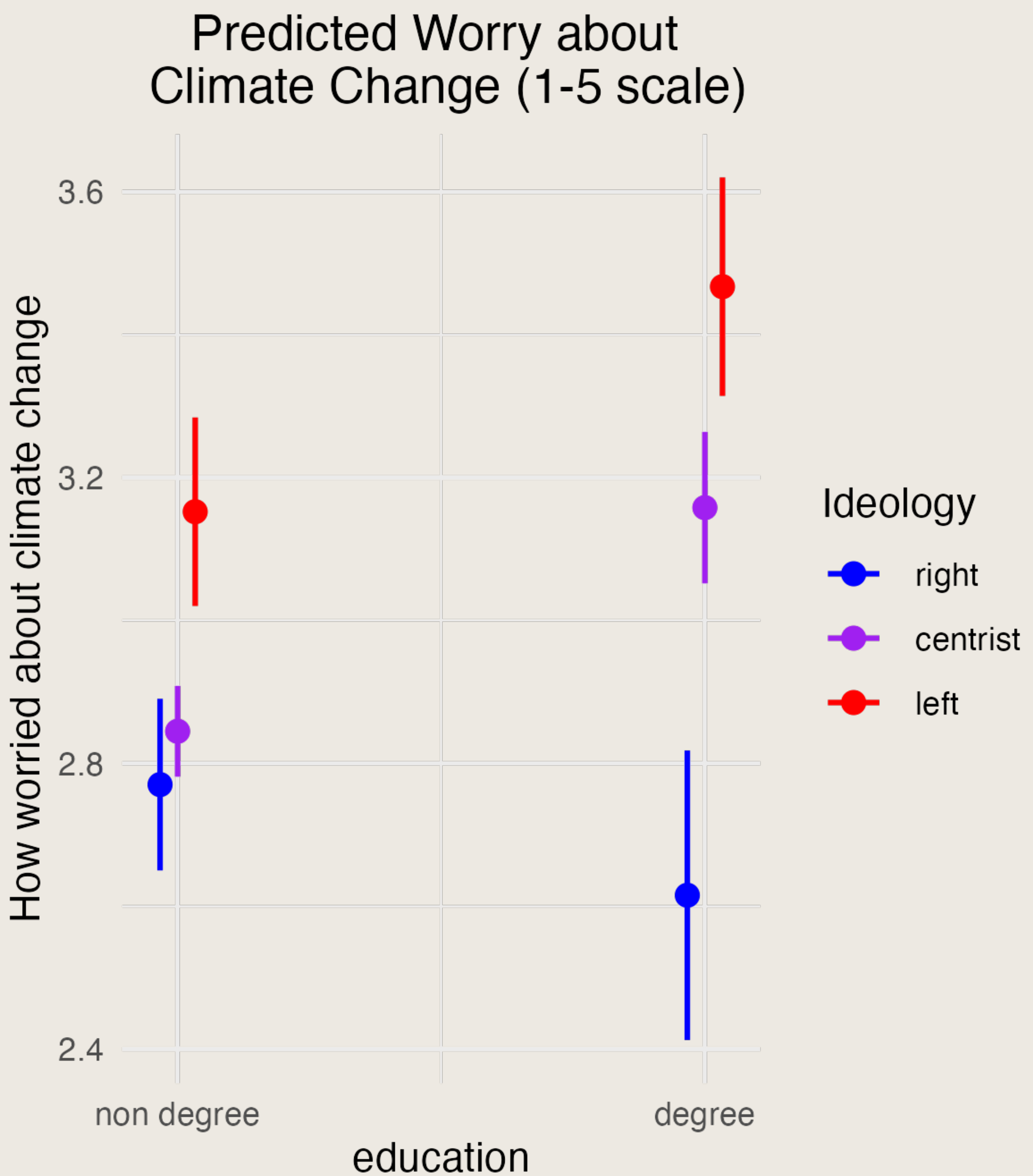
- What about the Centrists? Recode Ideology as a **three-category** variable. Now, the model is:
- Climate Worry $= \alpha + \beta_1$ Degree $+ \beta_2$ Left $+ \beta_3$ **Centrist** $+$ β_4 (Degree $\times$ Left) $+\beta_5$ **(Degree $\times$ Centrist)** $+ \epsilon$
- In R, just pass the categorical variable:

```
lm(wrclmch ~ education + ideo_group + education*ideo_group, data = ess)

# or equivalently

lm(wrclmch ~ education*ideo_group, data = ess)
```

Categorical Moderators with More Levels

	Dependent variable:
	Climate Worry (1–5)
Intercept	2.770*** (0.061)
Degree	−0.155 (0.120)
Centrist	0.075 (0.069)
Left	0.382***(0.091)
Degree × Centrist	0.468*** (0.136)
Degree × Left	0.470***(0.148)
Observations	1,699
Adjusted R^2	0.052

Note: *p<0.1; **p<0.05; ***p<0.01



Continuous Moderators

Continuous Moderators

- What if we want to measure ideology with a 0-10 scale?

Continuous Moderators

* What if we want to measure ideology with a 0-10 scale?

$$\text{Worry} = \alpha + \beta_1 \text{Degree} + \beta_2 \text{R-L Scale} + \beta_3 (\text{Degree} \times \text{R-L Scale}) + \epsilon$$

Continuous Moderators

* What if we want to measure ideology with a 0-10 scale?

$$\text{Worry} = \alpha + \beta_1\text{Degree} + \beta_2\text{R-L Scale} + \beta_3(\text{Degree} \times \text{R-L Scale}) + \epsilon$$

* β_1 is the estimate for the effect of 'Degree' on 'Worry' **when 'R-L Scale' is zero** (i.e. for the most right-wing).

Continuous Moderators

- What if we want to measure ideology with a 0-10 scale?

$$\text{Worry} = \alpha + \beta_1\text{Degree} + \beta_2\text{R-L Scale} + \beta_3(\text{Degree} \times \text{R-L Scale}) + \epsilon$$

- β_1 is the estimate for the effect of 'Degree' on 'Worry' **when 'R-L Scale' is zero** (i.e. for the most right-wing).
- β_2 is the predicted change in 'Worry' associated with of a **one-unit increase** in 'R-L Scale' when 'Degree' is zero (i.e. for non-graduates).

Continuous Moderators

- What if we want to measure ideology with a 0-10 scale?

$$\text{Worry} = \alpha + \beta_1\text{Degree} + \beta_2\text{R-L Scale} + \beta_3(\text{Degree} \times \text{R-L Scale}) + \epsilon$$

- β_1 is the estimate for the effect of 'Degree' on 'Worry' **when 'R-L Scale' is zero** (i.e. for the most right-wing).
- β_2 is the predicted change in 'Worry' associated with of a **one-unit increase** in 'R-L Scale' when 'Degree' is zero (i.e. for non-graduates).
- β_3 is tricky: it's the change in the effect of 'Degree' on 'Worry' **as we increase the value of 'L-R Scale' by one unit**. Easier to interpret significance and direction, use plots to show effect size.

Continuous Moderators

Continuous Moderators

$$\text{Worry} = \alpha + \beta_1 \text{ Degree} + \beta_2 \text{R-L Scale} + \beta_3 (\text{R-L Scale} \times \text{Degree}) + \epsilon$$

Continuous Moderators

$$\text{Worry} = \alpha + \beta_1 \text{ Degree} + \beta_2 \text{R-L Scale} + \beta_3 (\text{R-L Scale} \times \text{Degree}) + \epsilon$$

	Dependent variable:
	Climate Worry (1–5)
Intercept	
Degree	
R-L Scale	
Degree × R-L Scale	

Continuous Moderators

$$\text{Worry} = \alpha + \beta_1 \text{ Degree} + \beta_2 \text{R-L Scale} + \beta_3 (\text{R-L Scale} \times \text{Degree}) + \epsilon$$

	Dependent variable:
	Climate Worry (1–5)
Intercept	2.544*** (0.075)
Degree	
R-L Scale	
Degree × R-L Scale	

Continuous Moderators

$$\text{Worry} = \alpha + \beta_1 \text{ Degree} + \beta_2 \text{R-L Scale} + \beta_3 (\text{R-L Scale} \times \text{Degree}) + \epsilon$$

β_1 = effect of 'Degree' on 'Worry' when 'R-L Scale' is zero

	Dependent variable:
	Climate Worry (1–5)
Intercept	2.544*** (0.075)
Degree	−0.116 (0.142)
R-L Scale	
Degree × R-L Scale	

Continuous Moderators

$$\text{Worry} = \alpha + \beta_1 \text{ Degree} + \beta_2 \text{R-L Scale} + \beta_3 \text{ (R-L Scale} \times \text{Degree)} + \epsilon$$

β_1 = effect of 'Degree' on 'Worry' when 'R-L Scale' is zero

β_2 = effect of a one-unit increase in 'R-L Scale' on 'Worry' when 'Degree' is zero

	Dependent variable:
	Climate Worry (1–5)
Intercept	2.544*** (0.075)
Degree	−0.116 (0.142)
R-L Scale	0.068***(0.014)
Degree × R-L Scale	

Continuous Moderators

$$\text{Worry} = \alpha + \beta_1 \text{ Degree} + \beta_2 \text{R-L Scale} + \beta_3 \text{ (R-L Scale} \times \text{Degree)} + \epsilon$$

β_1 = effect of 'Degree' on 'Worry' when 'R-L Scale' is zero

β_2 = effect of a one-unit increase in 'R-L Scale' on 'Worry' when 'Degree' is zero

β_3 = change in the effect of 'Degree' on 'Worry' as we increase the value of 'L-R Scale' by one unit.

	Dependent variable:
	Climate Worry (1–5)
Intercept	2.544*** (0.075)
Degree	−0.116 (0.142)
R-L Scale	0.068***(0.014)
Degree × R-L Scale	0.068***(0.025)

Visualising Continuous Moderators (1)

Visualising Continuous Moderators (1)

- One solution: pick **some representative values of the moderator** and show predicted values of Y across treatment conditions.

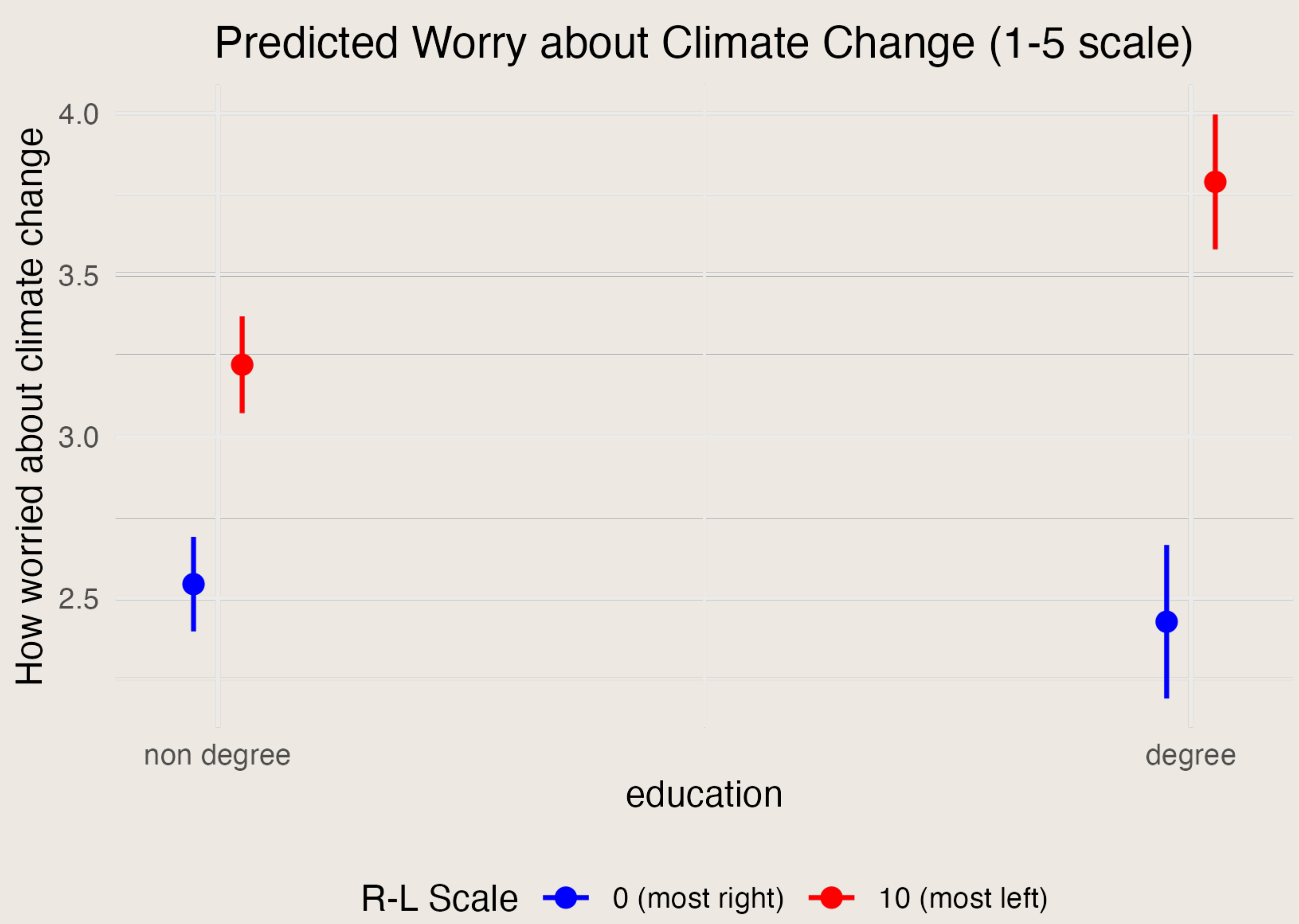
Visualising Continuous Moderators (1)

- One solution: pick **some representative values of the moderator** and show predicted values of Y across treatment conditions.
- Some options:

Visualising Continuous Moderators (1)

- One solution: pick **some representative values of the moderator** and show predicted values of *Y* across treatment conditions.

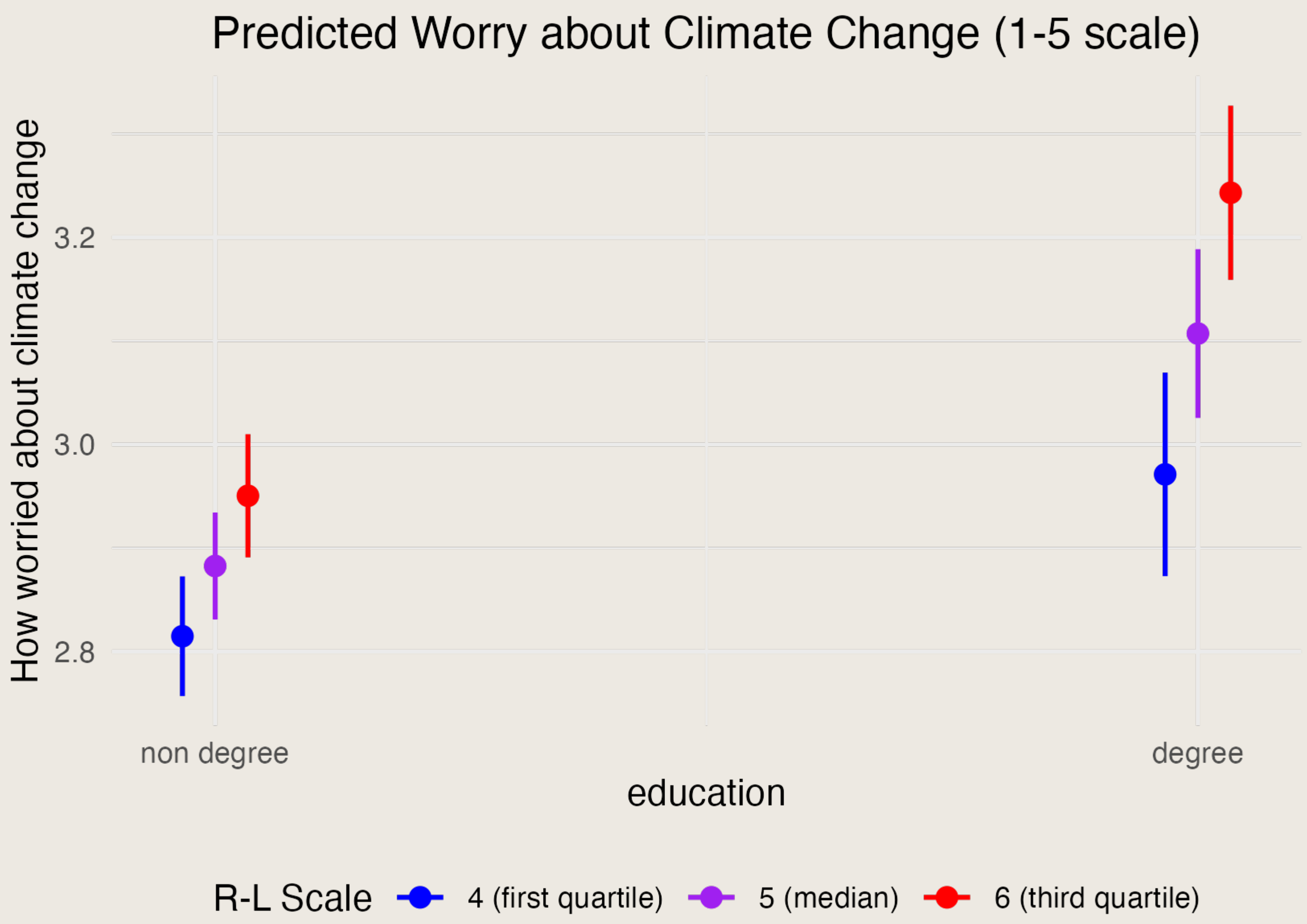
- Some options:
  - Minimum and Maximum value.



Visualising Continuous Moderators (1)

* One solution: pick **some representative values of the moderator** and show predicted values of Y across treatment conditions.

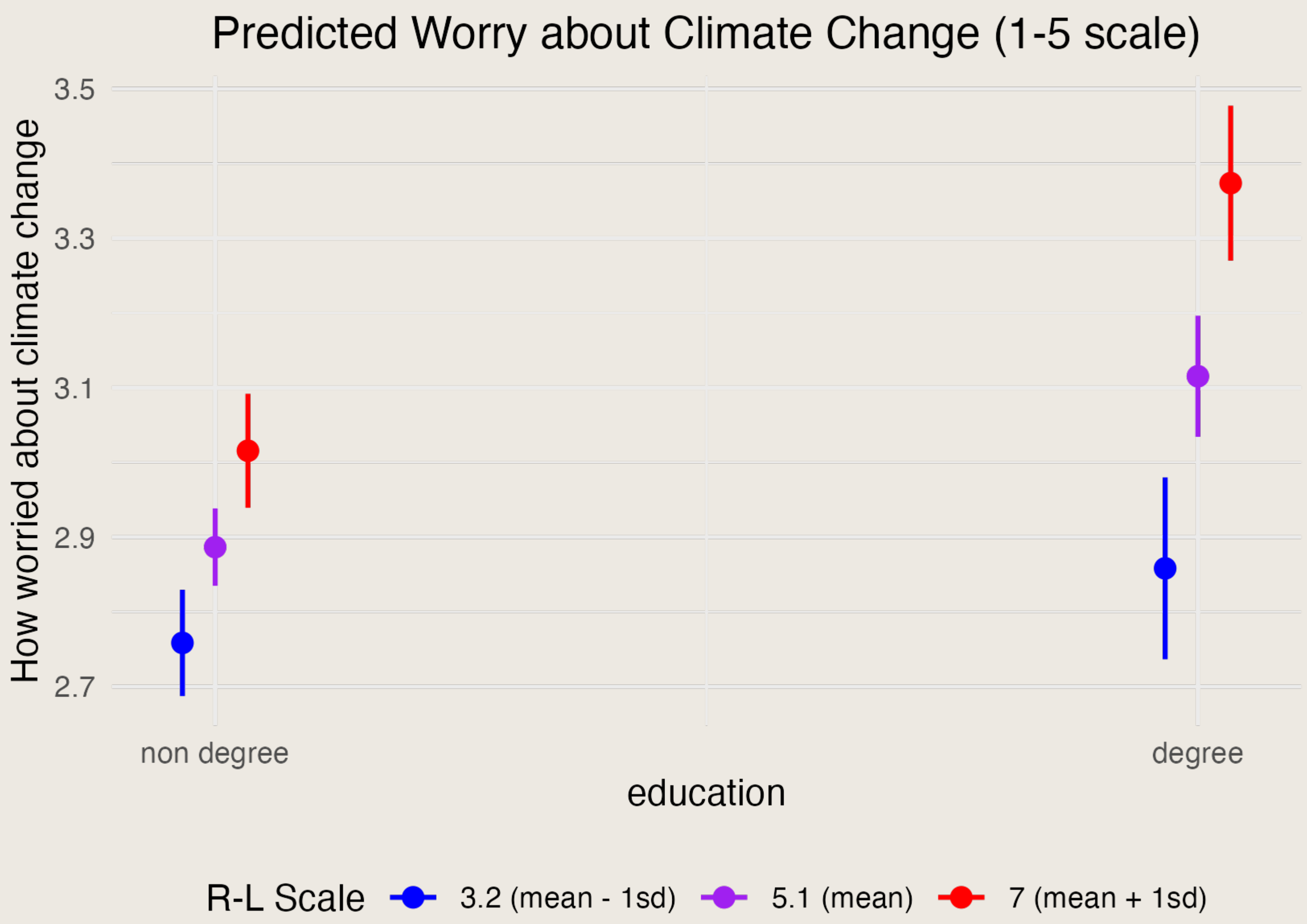
* Some options:
  * Minimum and Maximum value.
  * Quartiles of the distribution.



Visualising Continuous Moderators (1)

- One solution: pick **some representative values of the moderator** and show predicted values of *Y* across treatment conditions.

- Some options:
  - Minimum and Maximum value.
  - Quartiles of the distribution.
  - Mean *plus* and *minus* one std. deviation.



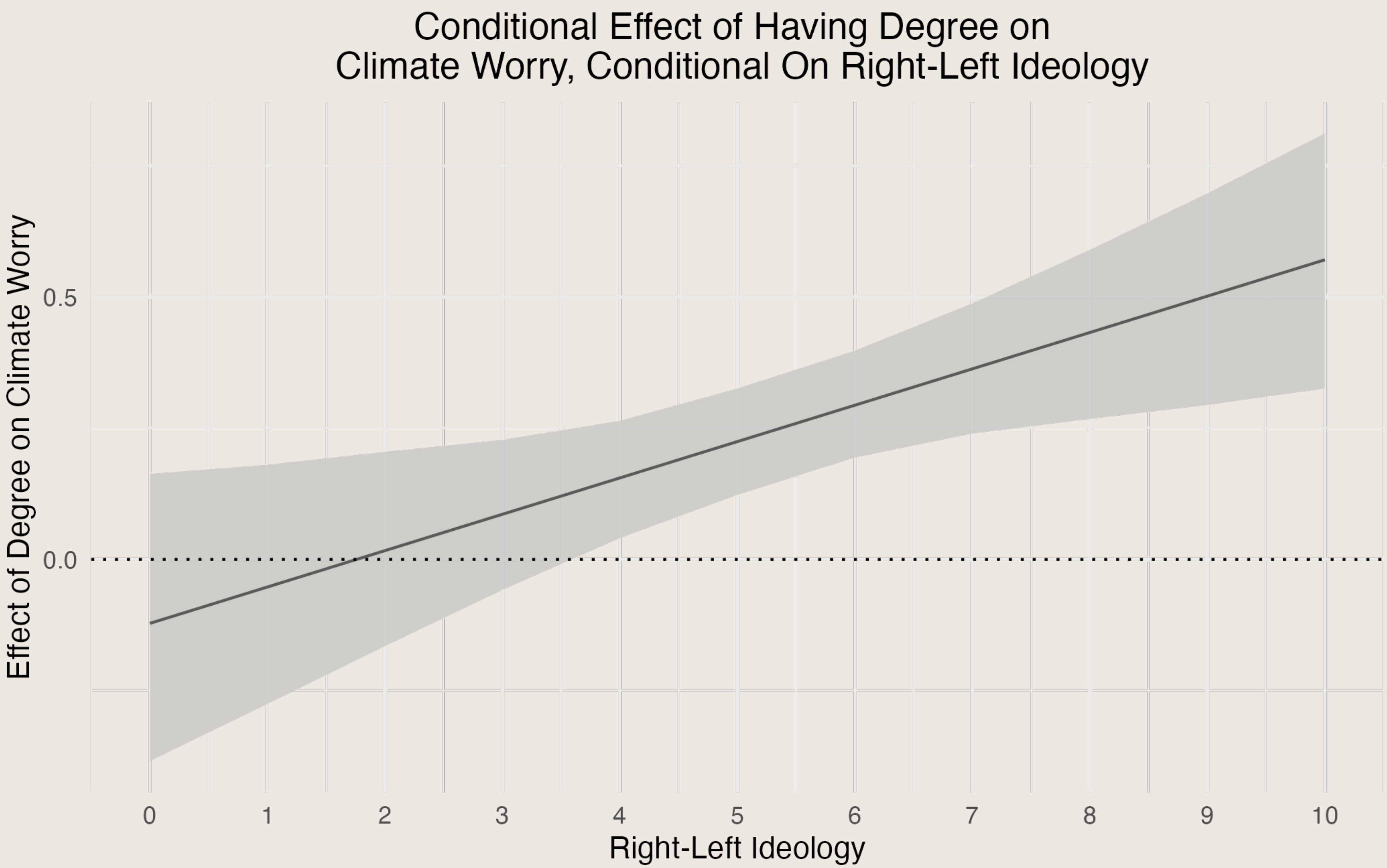
Visualising Continuous Moderators (2)

Visualising Continuous Moderators (2)

- A second solution: plot the effect of the treatment (Y-axis) by the value of the moderator (X-axis). This is known as a *conditional effect plot*.

Visualising Continuous Moderators (2)

- A second solution: plot the effect of the treatment (Y-axis) by the value of the moderator (X-axis). This is known as a *conditional effect plot*.



Continuous Treatment and Moderator

Continuous Treatment and Moderator

* What if we want to measure education as an interval variable? For instance, ‘years of education’. Same set-up:

Continuous Treatment and Moderator

* What if we want to measure education as an interval variable? For instance, 'years of education'. Same set-up:

$$\text{Worry} = \alpha + \beta_1 \text{EduYears} + \beta_2 \text{R-L Scale}$$
$$+ \beta_3(\text{EduYears} \times \text{R-L Scale}) + \epsilon$$

Continuous Treatment and Moderator

- What if we want to measure education as an interval variable? For instance, 'years of education'. Same set-up:

$$\text{Worry} = \alpha + \beta_1 \text{EduYears} + \beta_2 \text{R-L Scale} + \beta_3(\text{EduYears} \times \text{R-L Scale}) + \epsilon$$

- Both linear coefficients refer to effect of a one-unit change.

Continuous Treatment and Moderator

- What if we want to measure education as an interval variable? For instance, 'years of education'. Same set-up:

$$\text{Worry} = \alpha + \beta_1 \text{EduYears} + \beta_2 \text{R-L Scale}$$
$$+ \beta_3(\text{EduYears} \times \text{R-L Scale}) + \epsilon$$

- Both linear coefficients refer to effect of a one-unit change.
- The interaction term's coefficient is the estimated **change in the effect of one year of education** on Climate Worry, associated with a **one-point increase in the R-L scale**.

Continuous Moderators

Continuous Moderators

$$\text{Worry} = \alpha + \beta_1 \text{EduYears} + \beta_2 \text{R-L Scale} + \beta_3 (\text{R-L Scale} \times \text{EduYears}) + \epsilon$$

Continuous Moderators

$$\text{Worry} = \alpha + \beta_1 \text{EduYears} + \beta_2 \text{R-L Scale} + \beta_3 (\text{R-L Scale} \times \text{EduYears}) + \epsilon$$

	Dependent variable:
	Climate Worry (1–5)
Intercept	
Edu Years	
R-L Scale	
Edu Years × R-L Scale	

Continuous Moderators

$$\text{Worry} = \alpha + \beta_1 \text{EduYears} + \beta_2 \text{R-L Scale} + \beta_3 (\text{R-L Scale} \times \text{EduYears}) + \epsilon$$

	Dependent variable:
	Climate Worry (1–5)
Intercept	2.622*** (0.246)
Edu Years	
R-L Scale	
Edu Years × R-L Scale	

Continuous Moderators

$$\text{Worry} = \alpha + \beta_1 \text{EduYears} + \beta_2 \text{R-L Scale} + \beta_3 (\text{R-L Scale} \times \text{EduYears}) + \epsilon$$

β_1 = effect of one additional Year of Education when 'R-L Scale' is zero

Dependent variable:

	Climate Worry (1–5)
Intercept	2.622*** (0.246)
Edu Years	−0.008 (0.018)
R-L Scale	
Edu Years × R-L Scale	

Continuous Moderators

$$\text{Worry} = \alpha + \beta_1 \text{EduYears} + \beta_2 \text{R-L Scale} + \beta_3(\text{R-L Scale} \times \text{EduYears}) + \epsilon$$

β_1 = effect of one additional Year of Education when 'R-L Scale' is zero

β_2 = effect of a one-point increase in 'R-L Scale' on 'Worry' when Years of Education is zero

	Dependent variable:
	Climate Worry (1–5)
Intercept	2.622*** (0.246)
Edu Years	−0.008 (0.018)
R-L Scale	−0.018(0.045)
Edu Years × R-L Scale	

Continuous Moderators

$$\text{Worry} = \alpha + \beta_1 \text{EduYears} + \beta_2 \text{R-L Scale} + \beta_3 (\text{R-L Scale} \times \text{EduYears}) + \epsilon$$

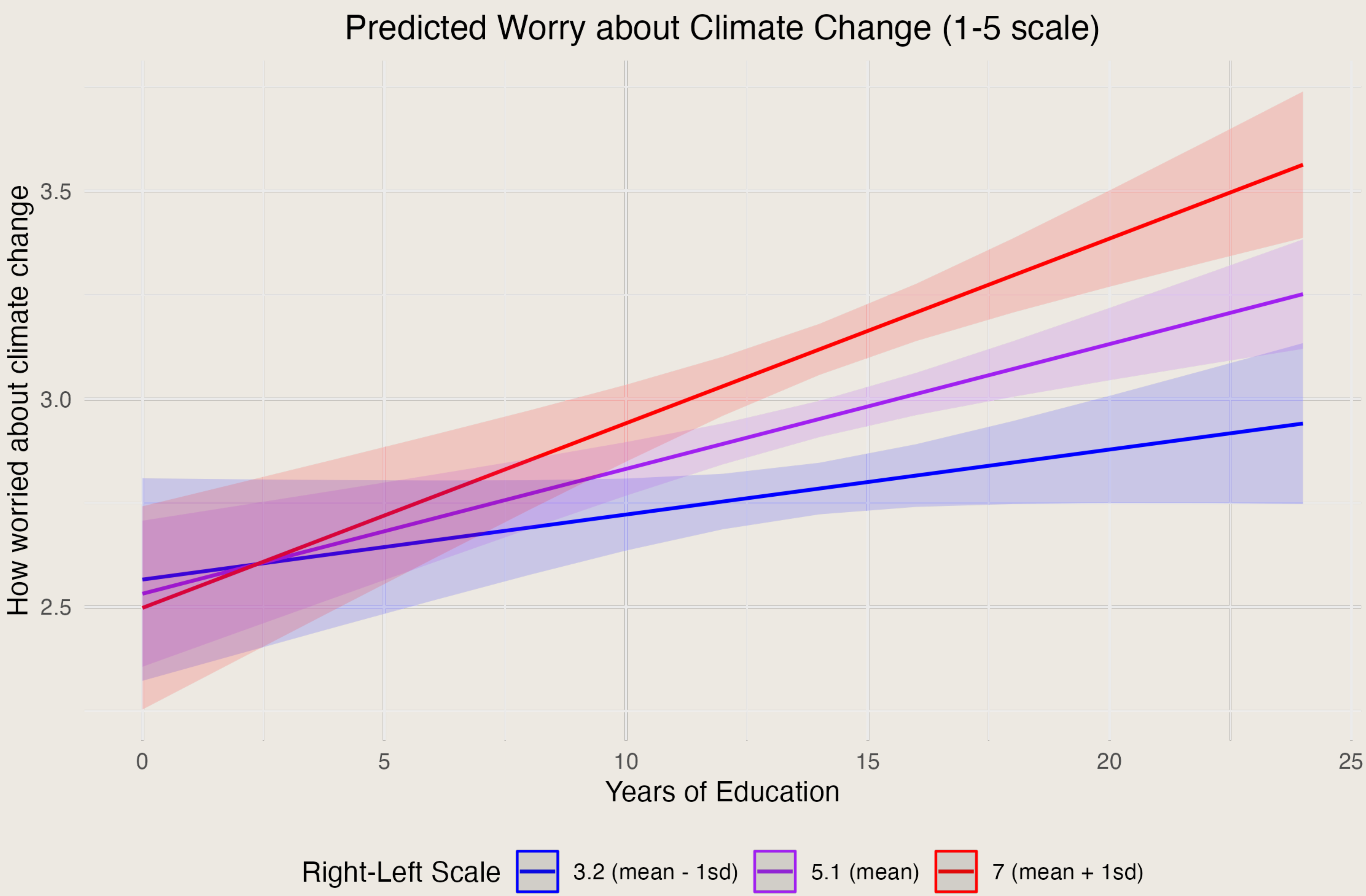
β_1 = effect of one additional Year of Education when 'R-L Scale' is zero

β_2 = effect of a one-point increase in 'R-L Scale' on 'Worry' when Years of Education is zero

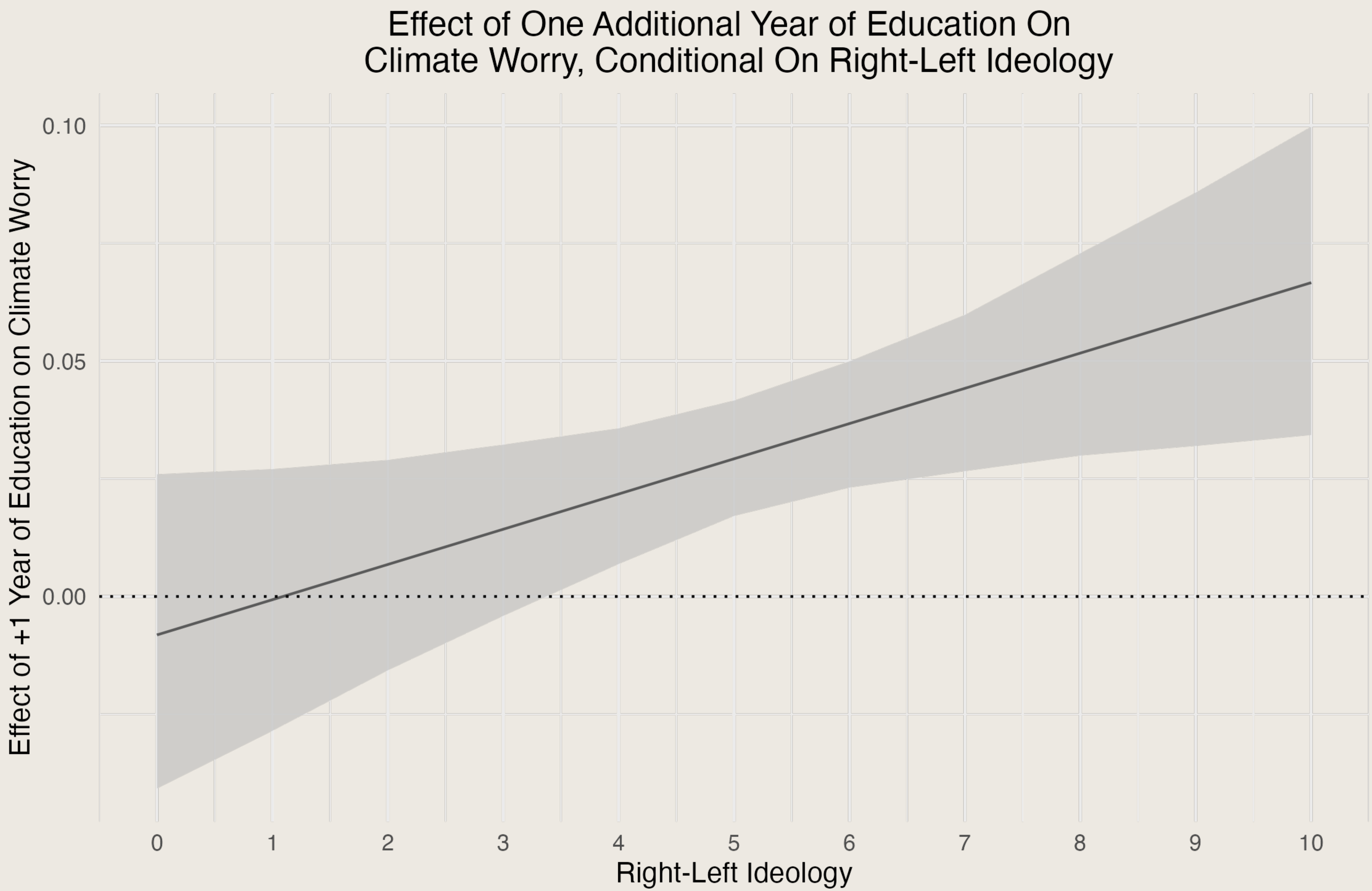
β_3 = change in the effect of one additional Year of Education on 'Worry' as we increase the value of 'L-R Scale' by one point.

	Dependent variable:
	Climate Worry (1–5)
Intercept	2.622*** (0.246)
Edu Years	−0.008 (0.018)
R-L Scale	−0.018(0.045)
Edu Years × R-L Scale	0.008***(0.003)

Predicted Values Plot



Conditional Effects Plot



Interaction Terms: Handle with Care

Interaction Terms: Handle with Care

- Always **include both the ‘parent’ terms** in a model with an interaction. `lm()` forces you to do that, thankfully.

Interaction Terms: Handle with Care

- Always **include both the 'parent' terms** in a model with an interaction. `lm()` forces you to do that, thankfully.
- It follows that moderators appear in your formula as covariates: therefore, for causal interpretation, you should **use variables that are plausibly pre-treatment**.

Interaction Terms: Handle with Care

- Always **include both the 'parent' terms** in a model with an interaction. `lm()` forces you to do that, thankfully.
- It follows that moderators appear in your formula as covariates: therefore, for causal interpretation, you should **use variables that are plausibly pre-treatment**.
- Software and math do not distinguish between treatment and moderator: the models we've just seen could be just as good to get at the effect of ideology on climate worry, conditional on education.

Interaction Terms: Handle with Care

- Always **include both the ‘parent’ terms** in a model with an interaction. `lm()` forces you to do that, thankfully.
- It follows that moderators appear in your formula as covariates: therefore, for causal interpretation, you should **use variables that are plausibly pre-treatment**.
- Software and math do not distinguish between treatment and moderator: the models we’ve just seen could be just as good to get at the effect of ideology on climate worry, conditional on education.
- It’s up to you to **interpret things correctly**.

Interaction Terms: Handle with Care

Interaction Terms: Handle with Care

* You should have a strong theoretical reason to use an interaction term. **Don't be this person**:

Interaction Terms: Handle with Care

- You should have a strong theoretical reason to use an interaction term. **Don't be this person**:
  - "I spent a year collecting all these data and I got a null result. Maybe the treatment works differently for men and women. **Let's try adding an interaction for gender.**"

Interaction Terms: Handle with Care

- You should have a strong theoretical reason to use an interaction term. **Don't be this person**:
  - "I spent a year collecting all these data and I got a null result. Maybe the treatment works differently for men and women. **Let's try adding an interaction for gender.**"
  - "Nothing. Maybe it's race? Nope. Hair colour? Nada. Maybe it's a triple interaction — treatment × race × gender? Maybe the treatment only works for people born in odd years."

Interaction Terms: Handle with Care

- You should have a strong theoretical reason to use an interaction term. **Don't be this person**:
  - "I spent a year collecting all these data and I got a null result. Maybe the treatment works differently for men and women. **Let's try adding an interaction for gender.**"
  - "Nothing. Maybe it's race? Nope. Hair colour? Nada. Maybe it's a triple interaction — treatment × race × gender? Maybe the treatment only works for people born in odd years."
- Potentially **infinite** combinations of interaction terms. You will get 'lucky' and find something significant at some point.

Interaction Terms: Handle with Care

Interaction Terms: Handle with Care

- Temptation for ‘fishing’ with interactions is particularly strong also because interactions tend to be **noisy**.

Interaction Terms: Handle with Care

* Temptation for ‘fishing’ with interactions is particularly strong also because interactions tend to be **noisy**.
* Our main effects are already noisy, because they’re estimated with uncertainty.

Interaction Terms: Handle with Care

- Temptation for ‘fishing’ with interactions is particularly strong also because interactions tend to be **noisy**.
- Our main effects are already noisy, because they’re estimated with uncertainty.
- Interactions estimate a difference between two noisy things. So they’re **even noisier**. Surprisingly big effects could pop up because of a few outliers.

Interaction Terms: Handle with Care

* Temptation for 'fishing' with interactions is particularly strong also because interactions tend to be **noisy**.
* Our main effects are already noisy, because they're estimated with uncertainty.
* Interactions estimate a difference between two noisy things. So they're **even noisier**. Surprisingly big effects could pop up because of a few outliers.
* You need very large sample sizes to estimate an interaction effect precisely (16× larger than for a main effect).

Interaction Terms: Handle with Care

Interaction Terms: Handle with Care

- More on pitfalls of interactions:

Interaction Terms: Handle with Care

- More on pitfalls of interactions:
- Brambor, T., Clark, W., and Golder, M. (2006) “Understanding interaction models: Improving empirical analyses.” *Political Analysis* 14(1), 63-82.

Interaction Terms: Handle with Care

- More on pitfalls of interactions:
- Brambor, T., Clark, W., and Golder, M. (2006) "Understanding interaction models: Improving empirical analyses." *Political Analysis* 14(1), 63-82.
- Hainmueller, J., Mummolo, J., & Xu, Y. (2019). "How much should we trust estimates from multiplicative interaction models? Simple tools to improve empirical practice." *Political Analysis*, 27(2), 163-192.

Interaction Terms: Handle with Care

- More on pitfalls of interactions:
- Brambor, T., Clark, W., and Golder, M. (2006) "Understanding interaction models: Improving empirical analyses." *Political Analysis* 14(1), 63-82.
- Hainmueller, J., Mummolo, J., & Xu, Y. (2019). "How much should we trust estimates from multiplicative interaction models? Simple tools to improve empirical practice." *Political Analysis*, 27(2), 163-192.
- Gelman, A. (2023) "You need 16 times the sample size to estimate an interaction than to estimate a main effect, explained", blogpost in *Statistical Modeling, Causal Inference, and Social Science*.

Check if you understand (1)

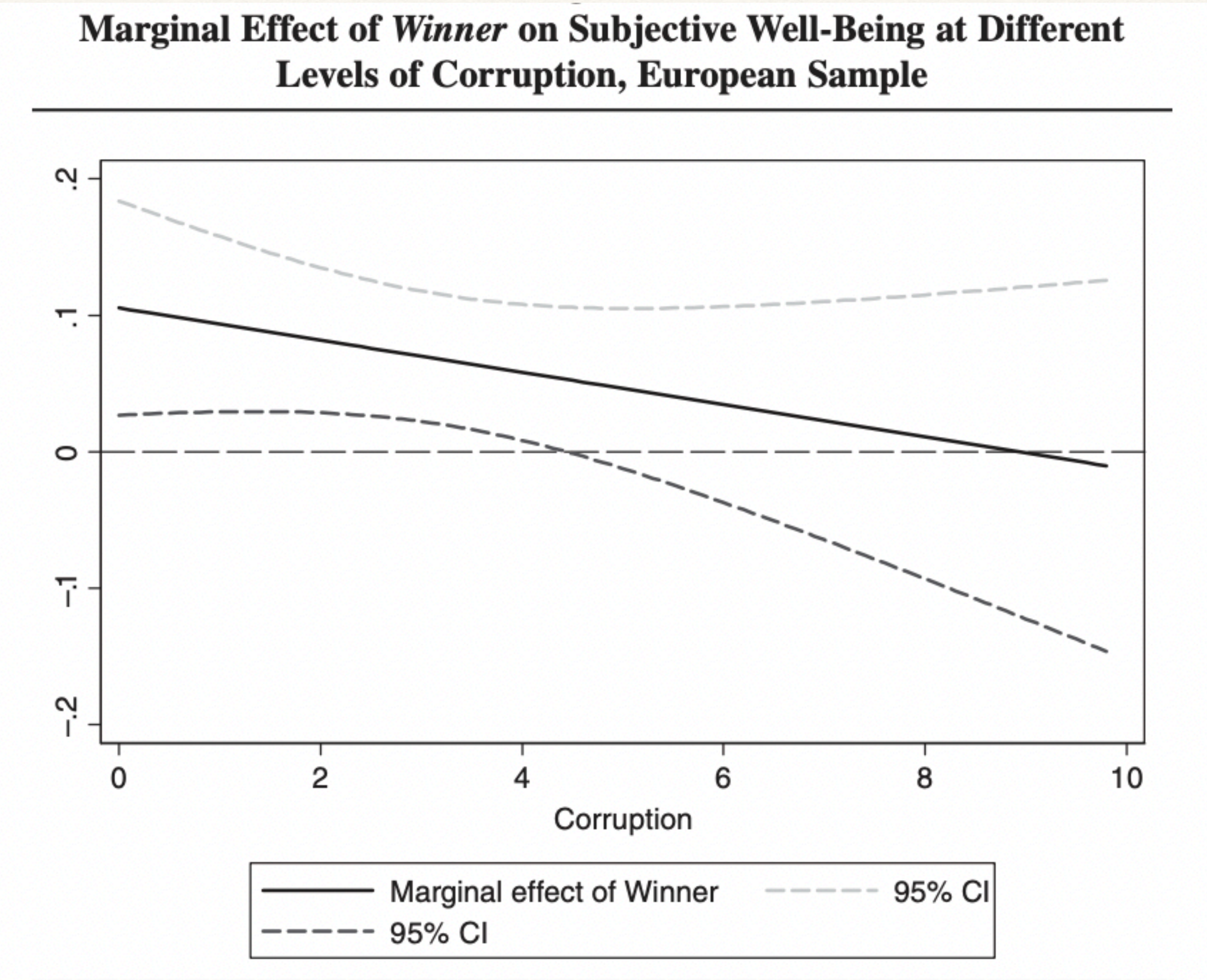
- Does 'winning' (i.e. voting for the party that forms the government) make people feel happier?

	Random Intercept, Interaction
Winner	.101*** (.021)
Corruption	–.079*** (.029)
Winner*Corruption	–.014** (.007)
Nonvoter	–.034** (.018)
Left-right self-placement	.018*** (.003)
Constant	3.166*** (.522)
Variance components	
Country	.018*** (.006)
Individual	.435*** (.005)
–2 log likelihood	26,133.8
N at Level 1	12,996
N at Level 2	16

- Margit Tavits (2008) Representation, Corruption, and Subjective Well-Being, *CPS*.

Check if you understand (1)

* Does 'winning' (i.e. voting for the party that forms the government) make people feel happier?



* Margit Tavits (2008) Representation, Corruption, and Subjective Well-Being, *CPS*.

Check if you understand (2)

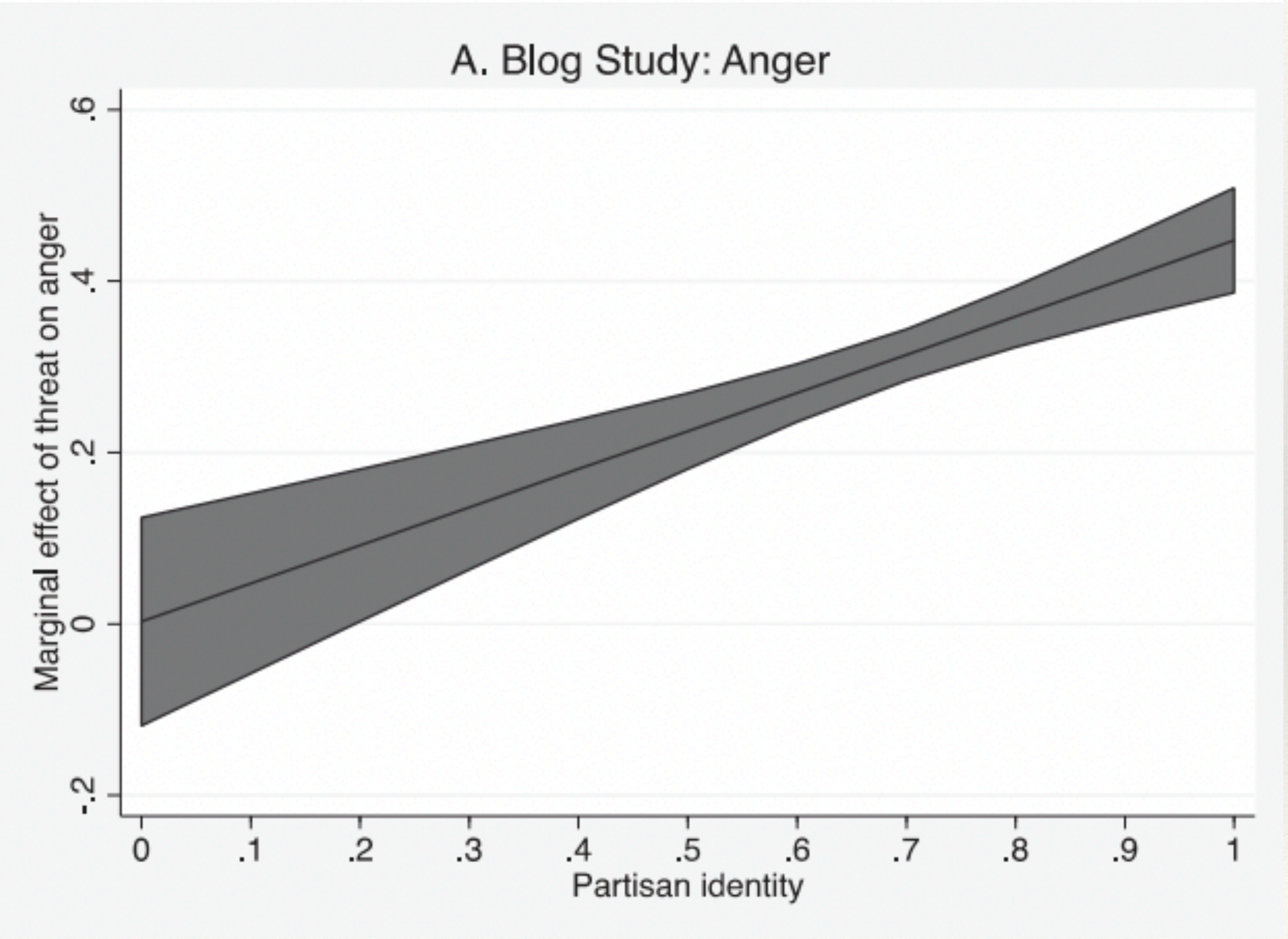
- Does telling people their party is going to lose the next election (*threat* treatment vs *reassurance* control) make them angrier?

	Anger and Party Threat	
	1	2
Partisan strength	–.01 (.03)	.01 (.03)
Partisan identity	–	–.07 (.07)
Party threat/reassurance	.26 (.06)***	.03 (.08)
Partisan strength × threat/reassurance	.10 (.04)**	–.01 (.04)
Partisan identity × threat/reassurance	–	.44 (.09)***
Ideological issue intensity	.06 (.05)	.07 (.05)
Ideological intensity × threat/reassurance	–.03 (.07)	–.03 (.07)
Knowledge	–.19 (.10)*	–.19 (.09)**
Gender (male)	–.04 (.02)**	–.03 (.02)*
Education	–.05 (.04)	–.04 (.04)
Age (decades)	.01 (.01)	.00 (.01)
Constant	.42 (.11)***	.46 (.11)***
Adj. R^2	*0.22*	*0.24*
N	*1482*	*1482*

Huddy, L., Mason, L., & Aarøe, L. (2015). Expressive partisanship: Campaign involvement, political emotion, and partisan identity. *APSR*, *109*(1), 1-17.

Check if you understand (2)

- Does telling people their party is going to lose the next election (*threat* treatment vs *reassurance* control) make them angrier?



Huddy, L., Mason, L., & Aarøe, L. (2015). Expressive partisanship: Campaign involvement, political emotion, and partisan identity. *APSR*, *109*(1), 1-17.

Non-Linearities

Dealing with Non-Linearities

Dealing with Non-Linearities

* **Polynomial terms (main focus today).** Introducing as regressors a variable and powers of the same variable (usually: squared, but you can add cubed, fourth power etc.).

Dealing with Non-Linearities

* **Polynomial terms (main focus today).** Introducing as regressors a variable and powers of the same variable (usually: squared, but you can add cubed, fourth power etc.).
    * $Y = \alpha + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \epsilon$

Dealing with Non-Linearities

- **Polynomial terms (main focus today).** Introducing as regressors a variable and powers of the same variable (usually: squared, but you can add cubed, fourth power etc.).
  - $Y = \alpha + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \epsilon$
- **Variable transformations (if there's time).** Commonly, taking the natural logarithm of the variables to reduce their skew.

Dealing with Non-Linearities

- **Polynomial terms (main focus today).** Introducing as regressors a variable and powers of the same variable (usually: squared, but you can add cubed, fourth power etc.).
  - $Y = \alpha + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \epsilon$
- **Variable transformations (if there's time).** Commonly, taking the natural logarithm of the variables to reduce their skew.
  - $Y = \alpha + \beta \log(X) + \epsilon$

Dealing with Non-Linearities

* **Polynomial terms (main focus today).** Introducing as regressors a variable and powers of the same variable (usually: squared, but you can add cubed, fourth power etc.).
    * $Y = \alpha + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \epsilon$
* **Variable transformations (if there's time).** Commonly, taking the natural logarithm of the variables to reduce their skew.
    * $Y = \alpha + \beta \log(X) + \epsilon$
* Both approaches are consistent with linearity assumptions: regression are still 'linear in the βs'.

Second-Degree Polynomial

Second-Degree Polynomial

* You might remember from high-school calculus the formula for a parabola: $y = ax^2 + bx + c$

Second-Degree Polynomial

- You might remember from high-school calculus the formula for a parabola: $y = ax^2 + bx + c$
- A regression curve with the second-order polynomial of X has the same functional form: $\hat{Y} = \hat{\alpha} + \hat{\beta}_1 X + \hat{\beta}_2 X^2$.

Second-Degree Polynomial

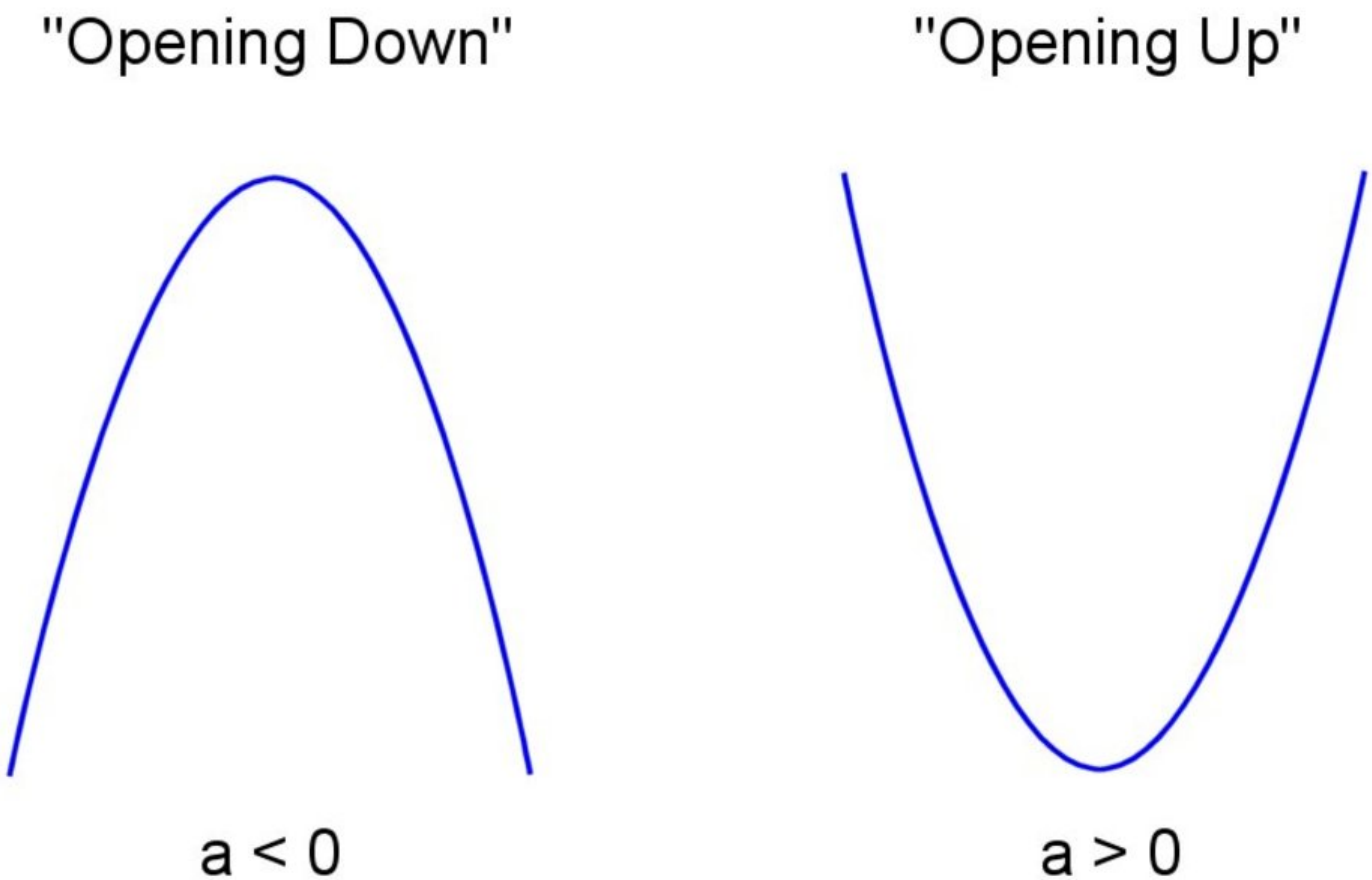
- You might remember from high-school calculus the formula for a parabola: $y = ax^2 + bx + c$
- A regression curve with the second-order polynomial of X has the same functional form: $\hat{Y} = \hat{\alpha} + \hat{\beta}_1 X + \hat{\beta}_2 X^2$.
- Characteristics of a parabolic curve:

Second-Degree Polynomial

* You might remember from high-school calculus the formula for a parabola: $y = ax^2 + bx + c$
* A regression curve with the second-order polynomial of X has the same functional form: $\hat{Y} = \hat{\alpha} + \hat{\beta}_1 X + \hat{\beta}_2 X^2$.
* Characteristics of a parabolic curve:
* It is **U-shaped** ('opening up') if $\beta_2 > 0$. It is **n-shaped** ('opening down') if $\beta_2 < 0$.

Second-Degree Polynomial

- You might remember from high-school calculus the formula for a parabola: $y = ax^2 + bx + c$
- A regression curve with the second-order polynomial of X has the same functional form: $\hat{Y} = \hat{\alpha} + \hat{\beta}_1 X + \hat{\beta}_2 X^2$.
- Characteristics of a parabolic curve:
- It is **U-shaped** ('opening up') if $\beta_2 > 0$. It is **n-shaped** ('opening down') if $\beta_2 < 0$.
- It has **one** bend, known as its vertex, given by $-\dfrac{\beta_1}{2\beta_2}$



The coefficient of x^2 determines whether the parabola opens up or down

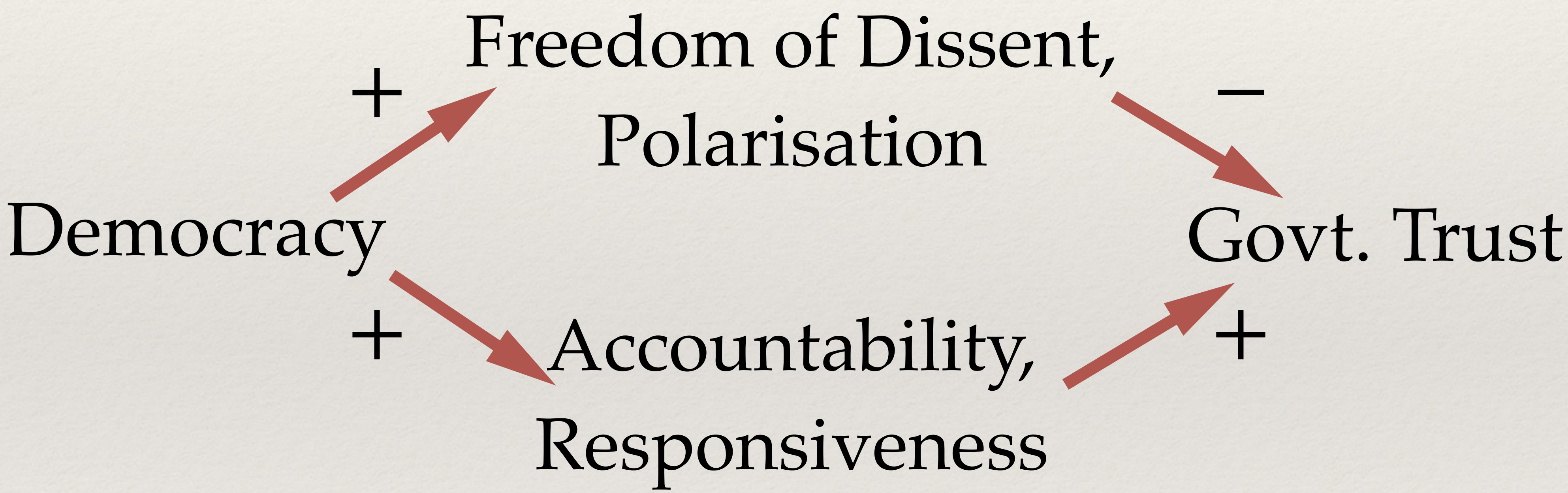
Example

Example

- Does **democracy** increase or decrease **trust in government**?

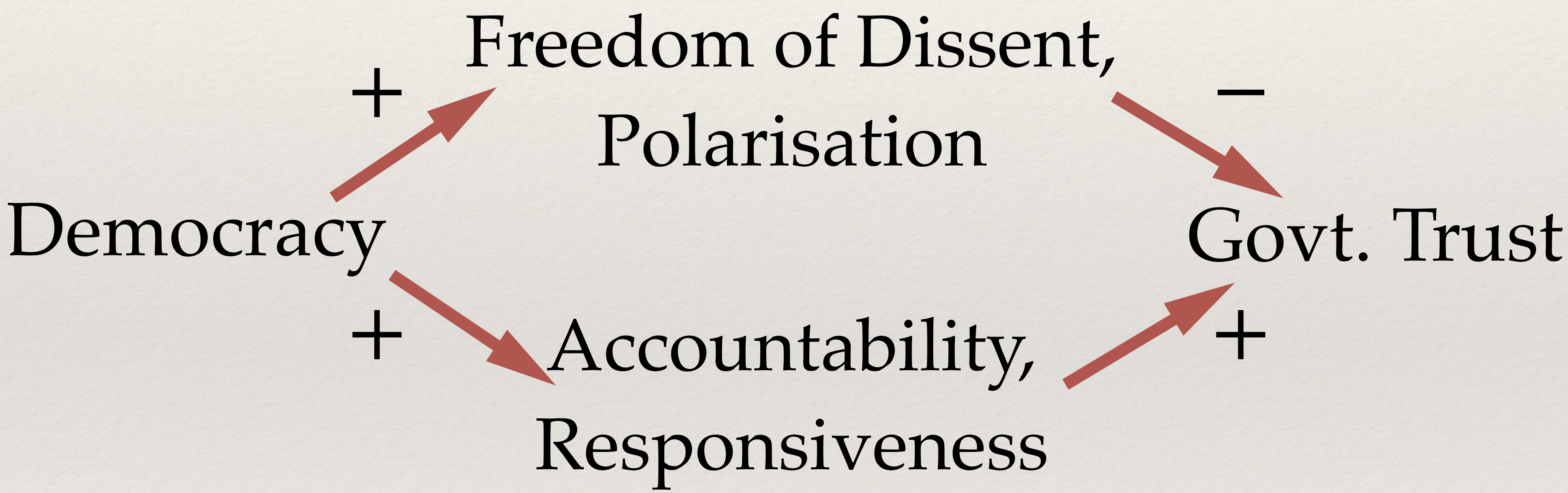
Example

* Does **democracy** increase or decrease **trust in government**?



Example

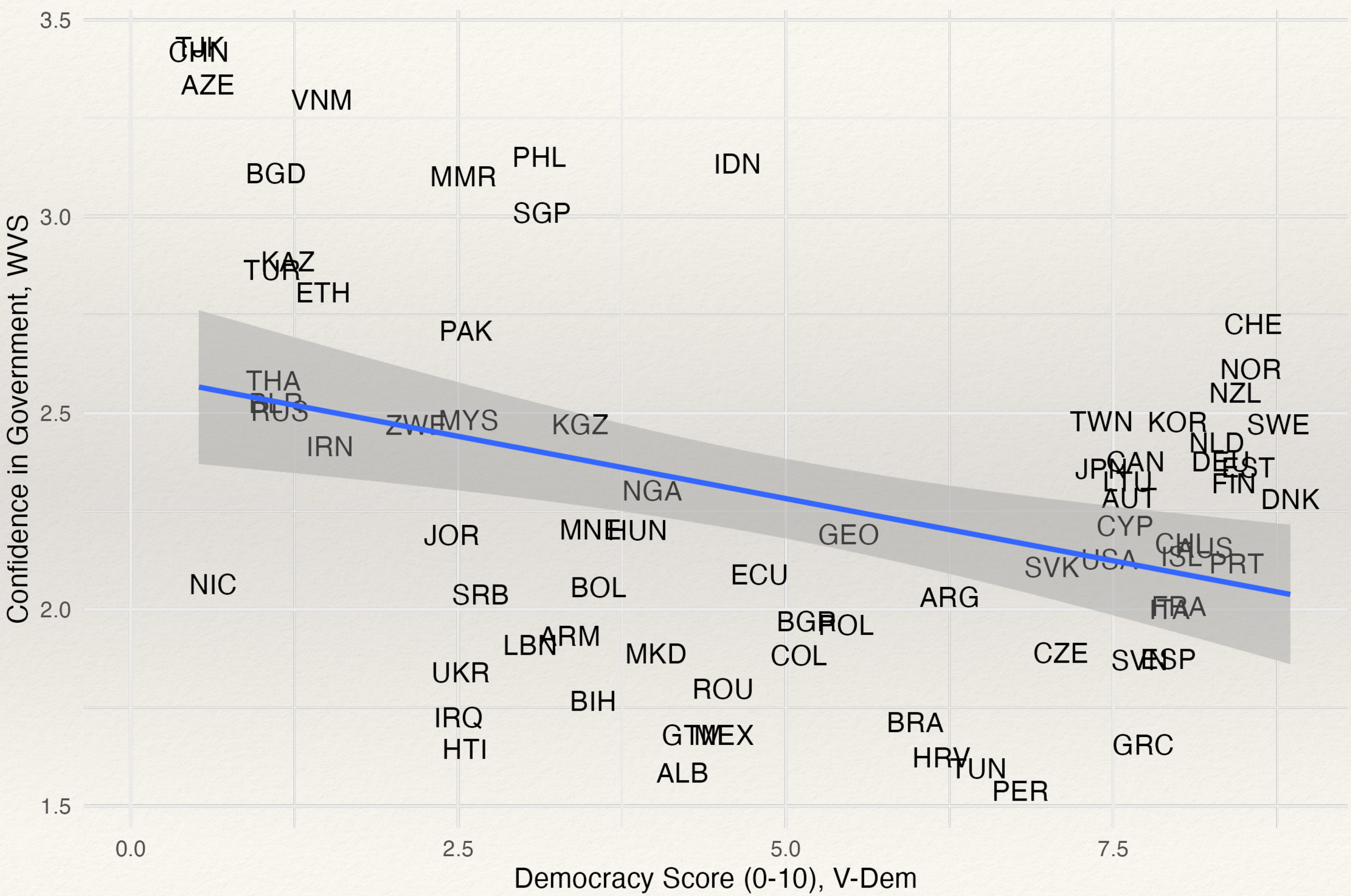
- Does **democracy** increase or decrease **trust in government**?



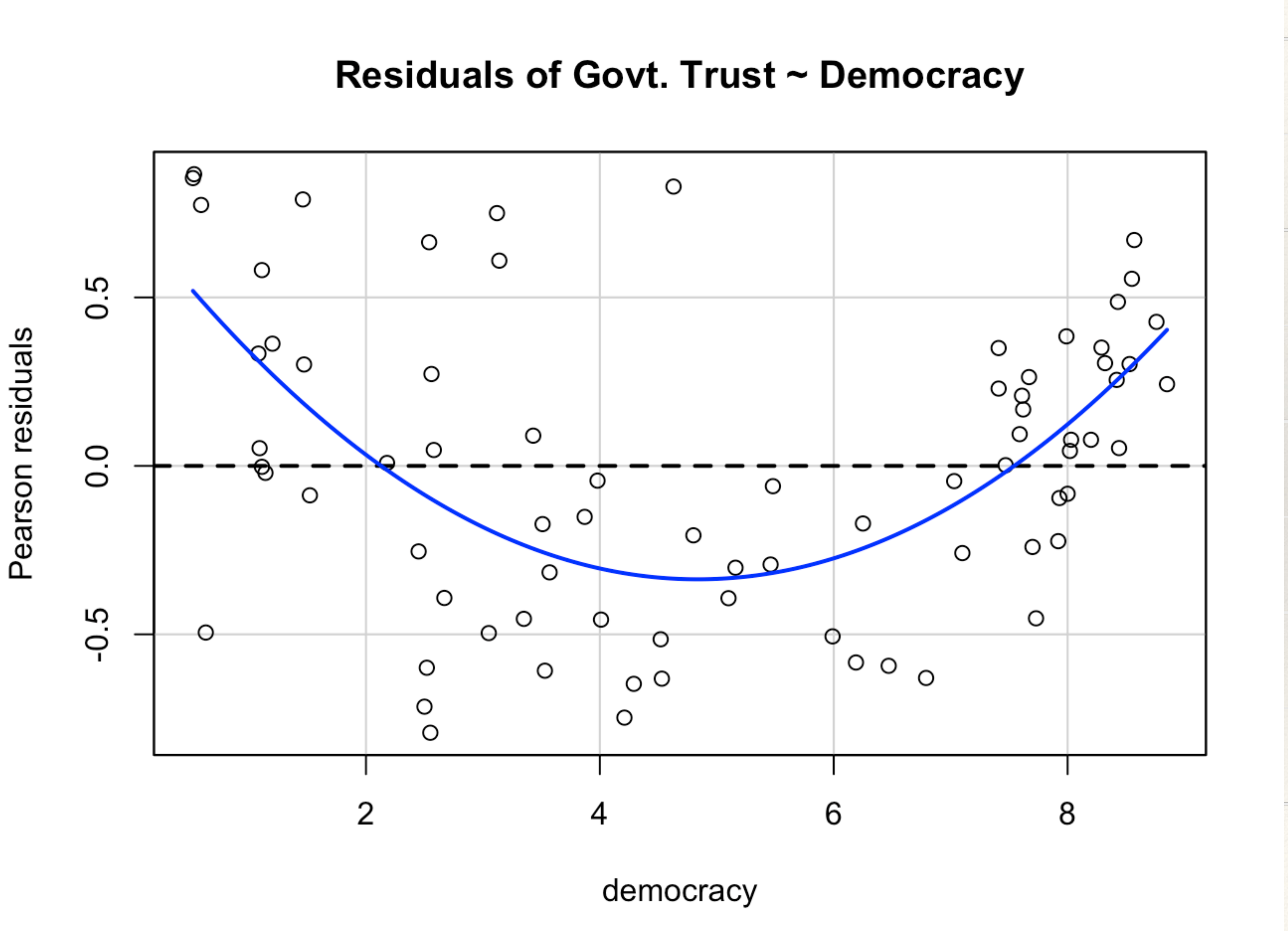
- We gather data on **Democracy** (0-10 scale) from V-Dem, and on the average country-level **Trust in Government** (1 = none at all, 4 = a great deal) from the World Values Survey (WVS).

$$\text{Govt. Trust} = \alpha + \beta_1 \text{Democracy} + \epsilon$$

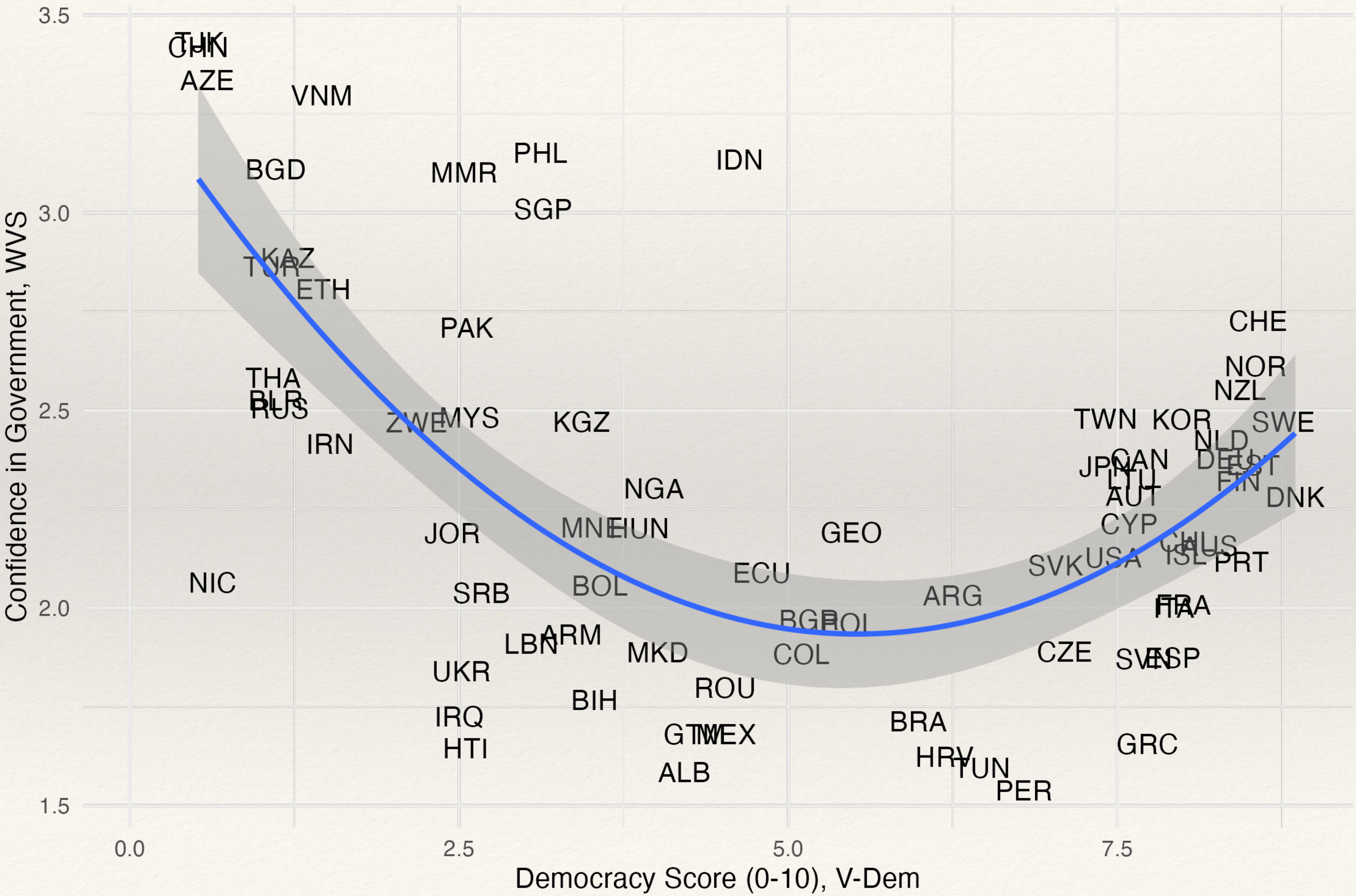
$\text{Govt. Trust} = \alpha + \beta_1 \text{Democracy} + \epsilon$



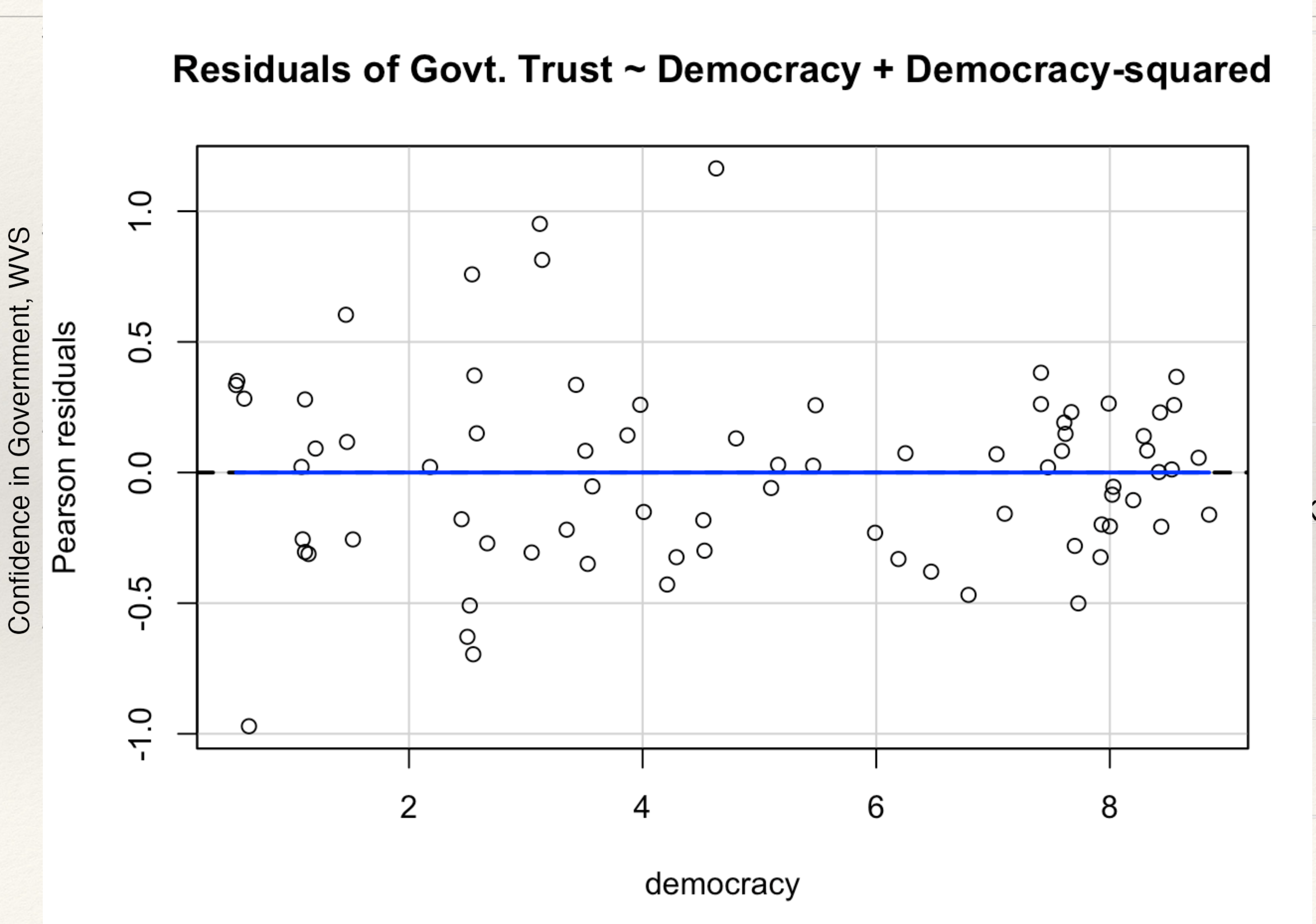
$\text{Govt. Trust} = \alpha + \beta_1 \text{Democracy} + \epsilon$



$\text{Govt. Trust} = \alpha + \beta_1 \text{Democracy} + \beta_2 \text{Democracy}^2 + \epsilon$

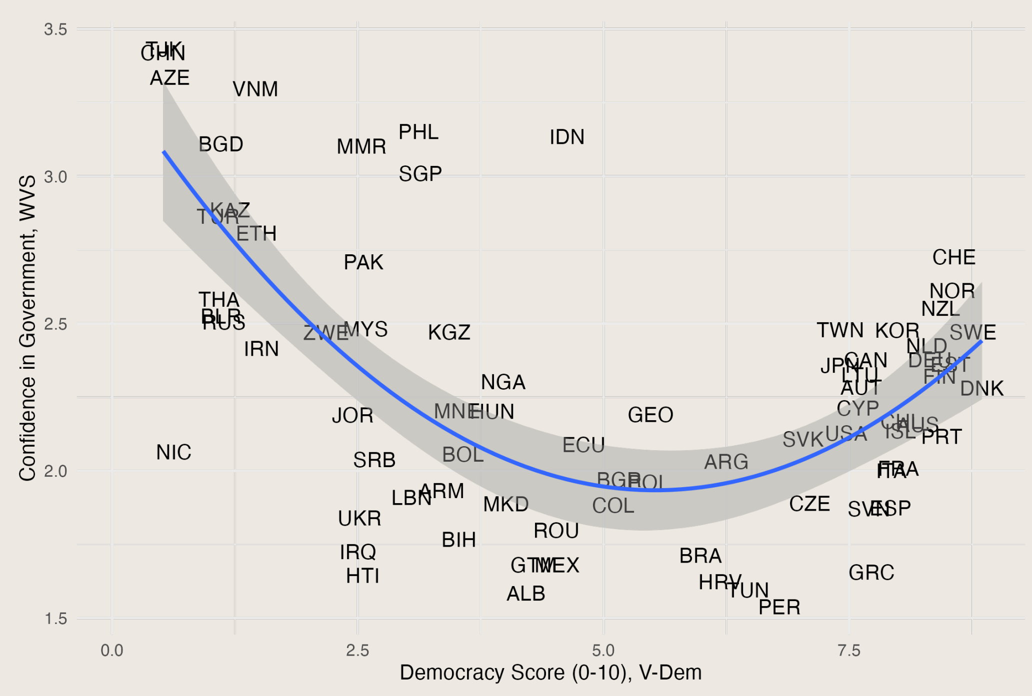


$$\text{Govt. Trust} = \alpha + \beta_1\text{Democracy} + \beta_2\text{Democracy}^2 + \epsilon$$



Second-Degree Polynomial: Coefficients

	Dependent variable:
	Govt. Trust (1–4)
Intercept	3.337*** (0.152)
Democracy	-0.508*** (0.076)
Democracy²	0.046*** (0.008)

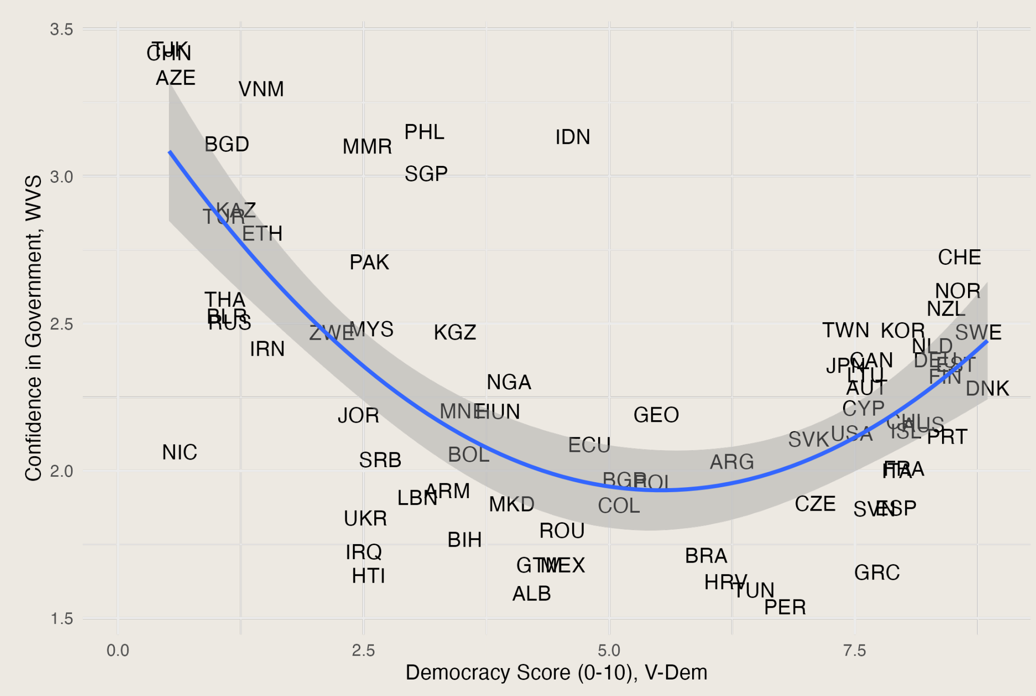


Second-Degree Polynomial: Coefficients

- **Sign** of β_2: if $\beta_2 > 0$, U-shaped curve, if $\beta_2 < 0$, n-shaped curve.

Dependent variable:

	Govt. Trust (1–4)
Intercept	3.337*** (0.152)
Democracy	-0.508*** (0.076)
Democracy²	0.046*** (0.008)

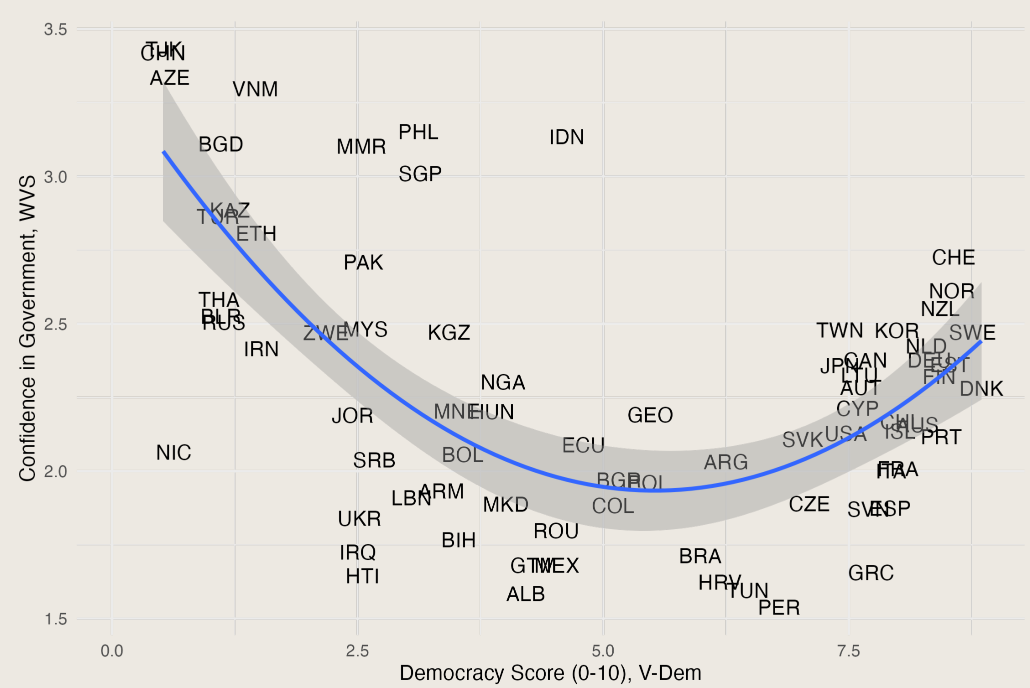


Second-Degree Polynomial: Coefficients

- **Sign** of β_2: if $\beta_2 > 0$, U-shaped curve, if $\beta_2 < 0$, n-shaped curve.
- **Significance** of β_2: tests against the null that the relationship is linear.

Dependent variable:

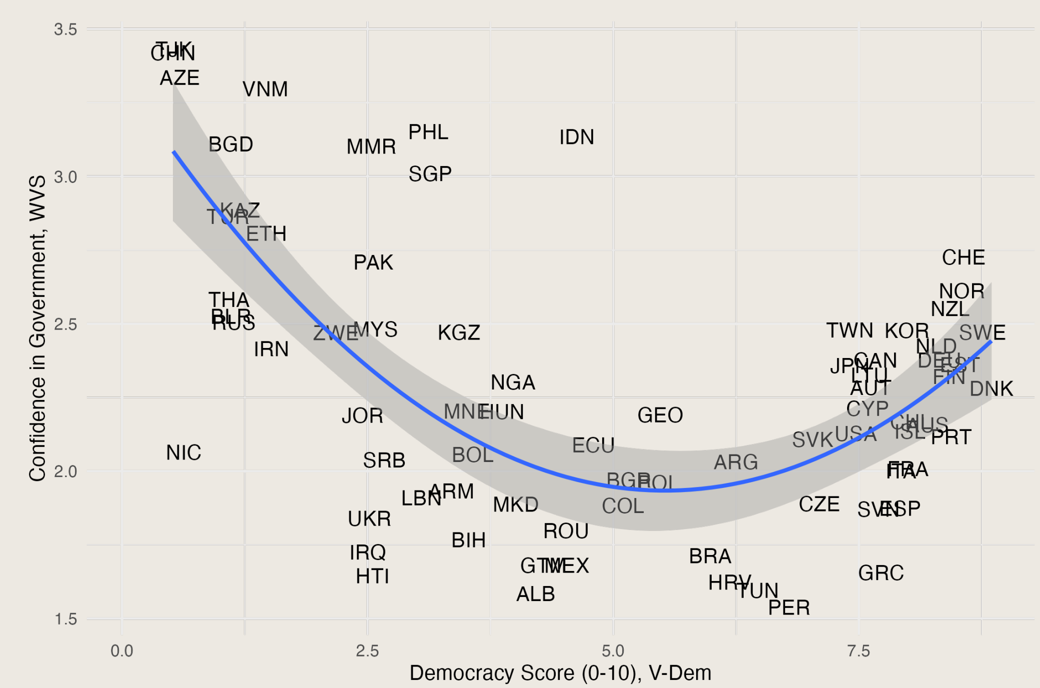
	Govt. Trust (1–4)
Intercept	3.337*** (0.152)
Democracy	-0.508*** (0.076)
Democracy²	0.046*** (0.008)



Second-Degree Polynomial: Coefficients

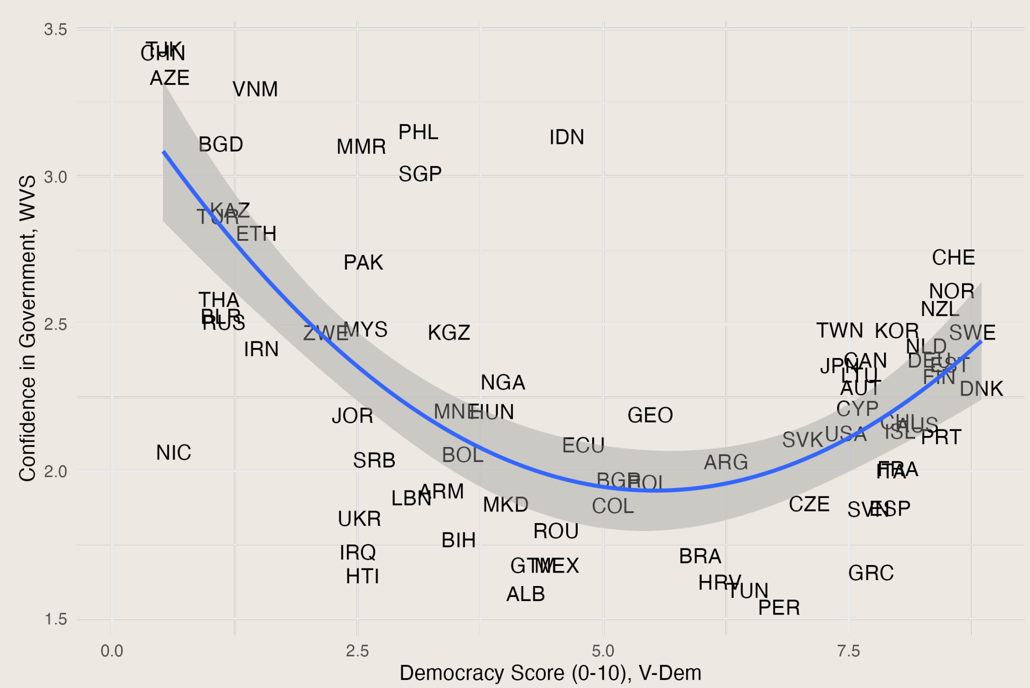
- **Sign** of β_2: if $\beta_2 > 0$, U-shaped curve, if $\beta_2 < 0$, n-shaped curve.
- **Significance** of β_2: tests against the null that the relationship is linear.
- **Vertex:** $-\beta_1/(2\beta_2)$. This is where sign of the relationship changes — **may fall outside the observed range of** X.

	Dependent variable:
	Govt. Trust (1–4)
Intercept	3.337*** (0.152)
Democracy	-0.508*** (0.076)
Democracy²	0.046*** (0.008)



Second-Degree Polynomial: Coefficients

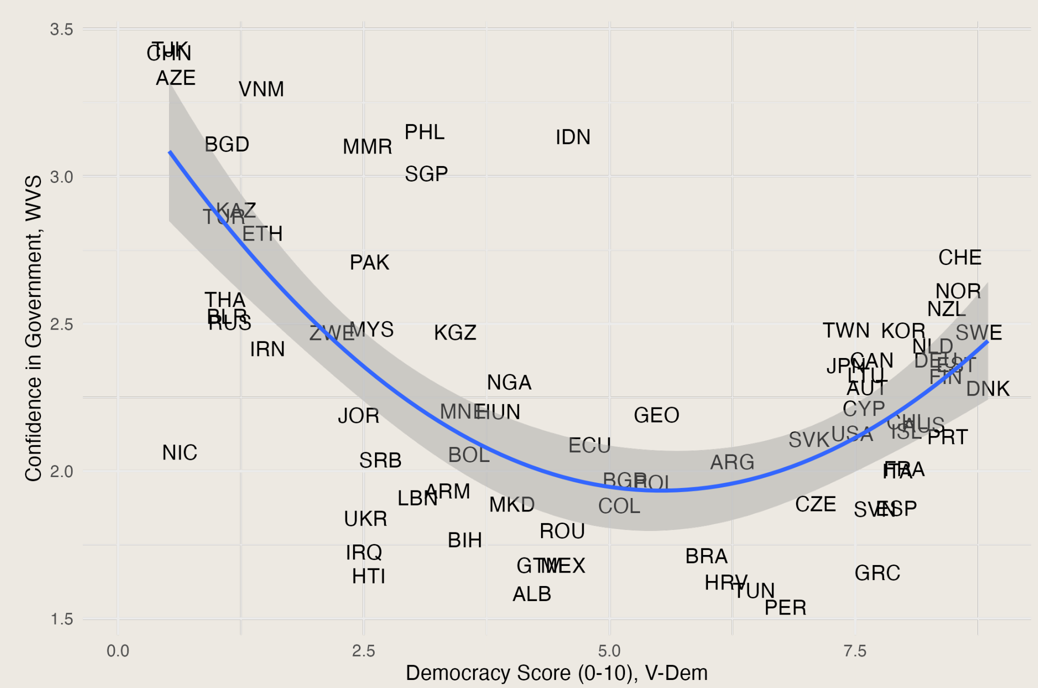
	Dependent variable:
	Govt. Trust (1–4)
Intercept	3.337*** (0.152)
Democracy	-0.508*** (0.076)
Democracy^2	0.046*** (0.008)



Second-Degree Polynomial: Coefficients

- Usual interpretation of effect size **doesn't work**: "holding all else constant, a one-unit increase in X is associated with a β_1 increase in Y."

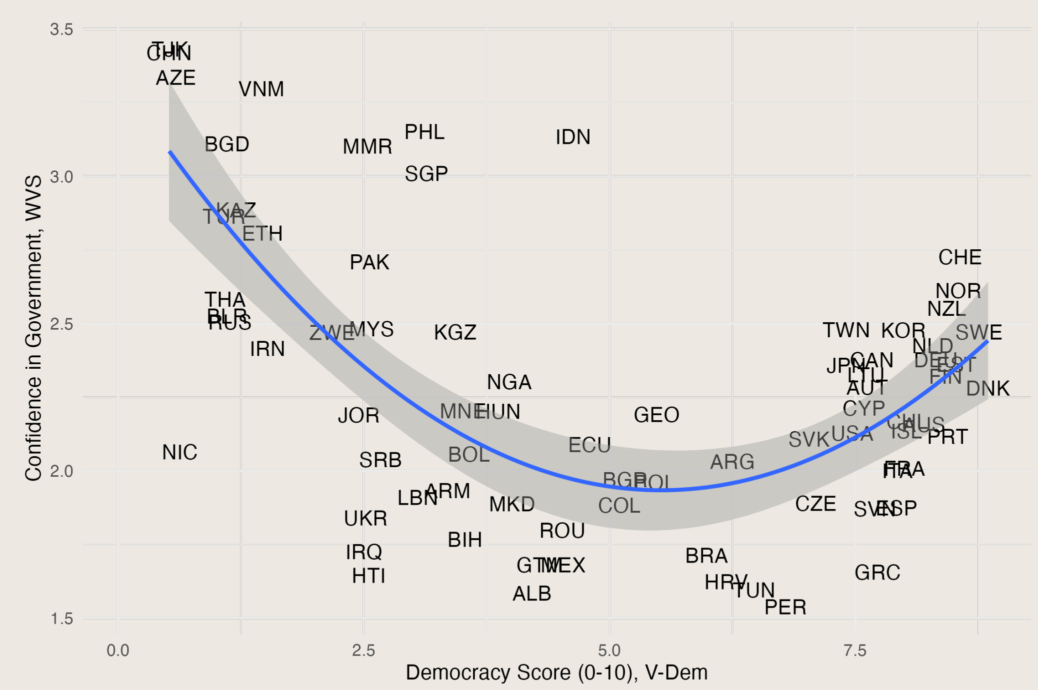
	Dependent variable:
	Govt. Trust (1–4)
Intercept	3.337*** (0.152)
Democracy	-0.508*** (0.076)
Democracy²	0.046*** (0.008)



Second-Degree Polynomial: Coefficients

- Usual interpretation of effect size **doesn't work**: "holding all else constant, a one-unit increase in X is associated with a β_1 increase in Y."
- We **can't hold all else constant**. If we increase X, we also increase X^2.

	Dependent variable:
	Govt. Trust (1–4)
Intercept	3.337*** (0.152)
Democracy	-0.508*** (0.076)
Democracy²	0.046*** (0.008)

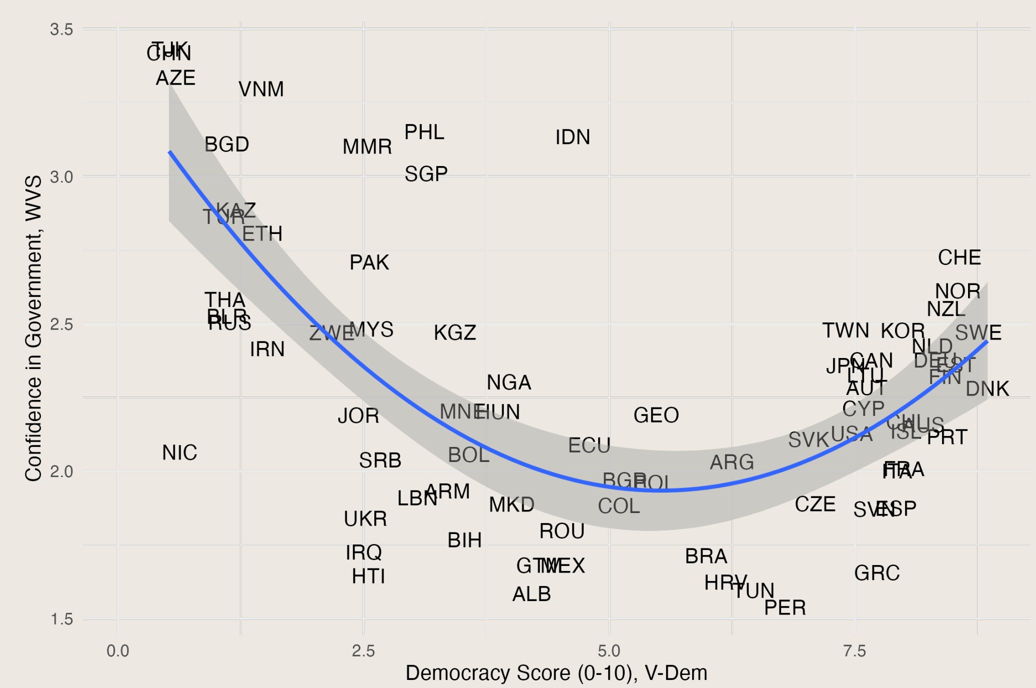


Second-Degree Polynomial: Coefficients

- Usual interpretation of effect size **doesn't work**: "holding all else constant, a one-unit increase in X is associated with a β_1 increase in Y."
- We **can't hold all else constant**. If we increase X, we also increase X^2.
- At each value X the predicted **rate of change** in Y varies.

Dependent variable:

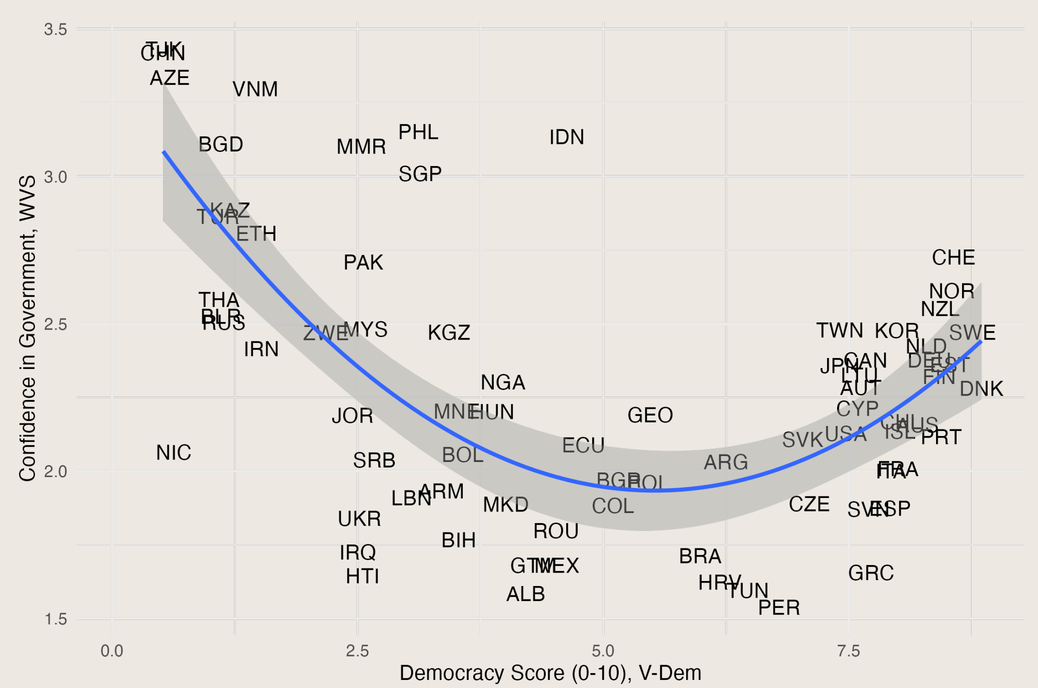
	Govt. Trust (1–4)
Intercept	3.337*** (0.152)
Democracy	-0.508*** (0.076)
Democracy²	0.046*** (0.008)



Second-Degree Polynomial: Coefficients

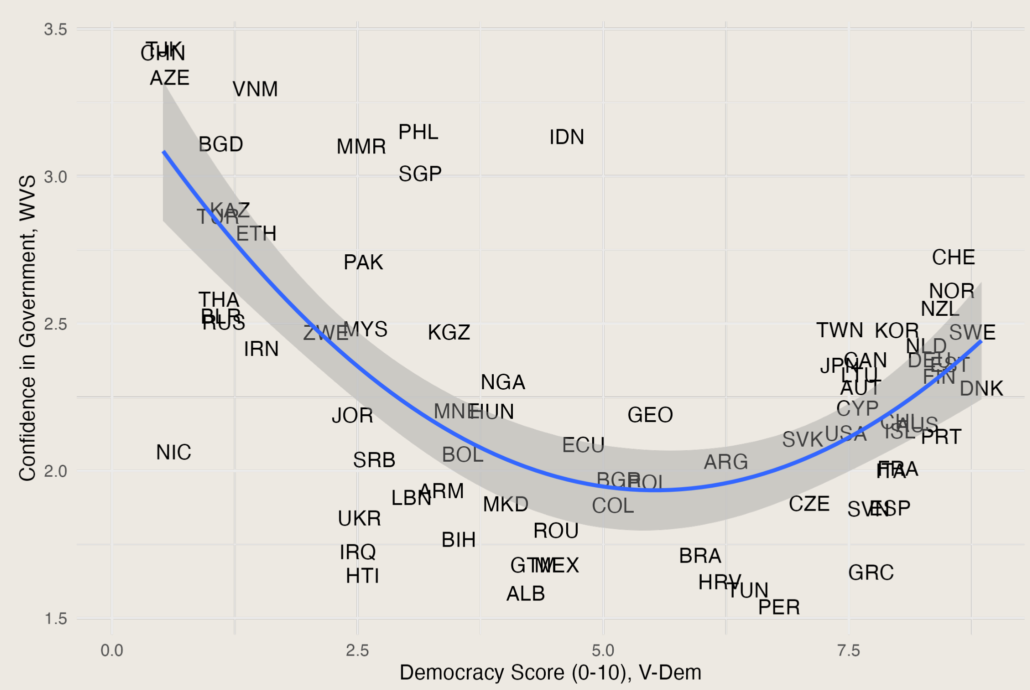
- Usual interpretation of effect size **doesn't work**: "holding all else constant, a one-unit increase in X is associated with a β_1 increase in Y."
- We **can't hold all else constant**. If we increase X, we also increase X^2.
- At each value X the predicted **rate of change** in Y varies.
- Polynomial variable coefficients β_1 and β_2 mean little on their own, they must be interpreted together

	Dependent variable:
	Govt. Trust (1–4)
Intercept	3.337*** (0.152)
Democracy	-0.508*** (0.076)
Democracy²	0.046*** (0.008)



Second-Degree Polynomial: Coefficients

	Dependent variable:
	Govt. Trust (1–4)
Intercept	3.337*** (0.152)
Democracy	-0.508*** (0.076)
Democracy²	0.046*** (0.008)

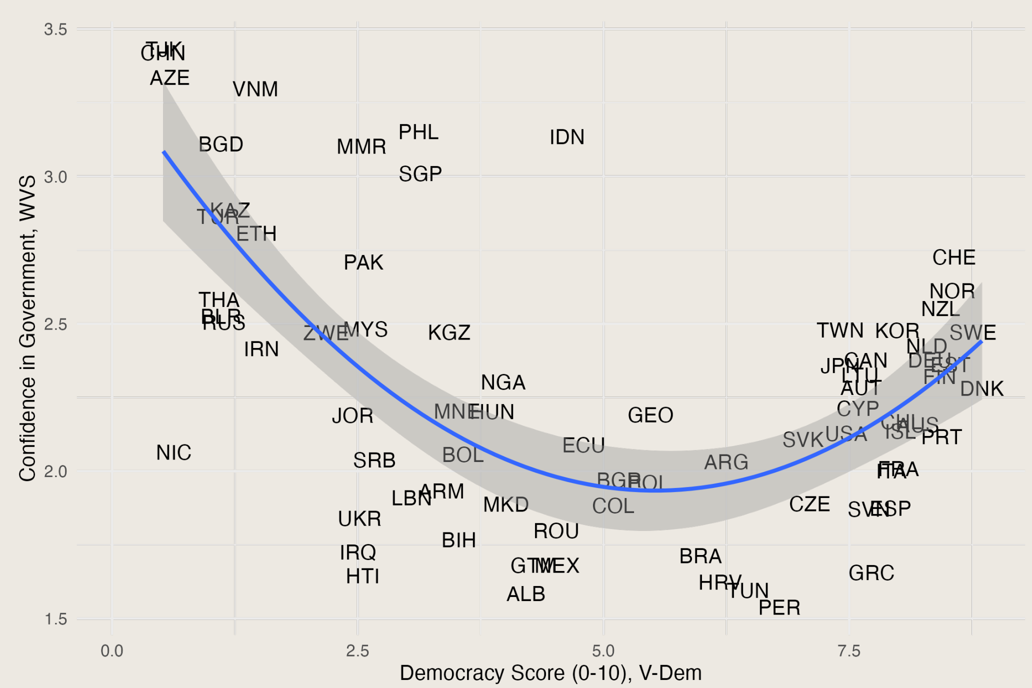


Second-Degree Polynomial: Coefficients

- **Instantaneous rate of change**, expressed by the **derivative**. The derivative of $\hat{Y} = \alpha + \beta_1 X + \beta_2 X^2$ in X is $\beta_1 + 2\beta_2 X$.

Dependent variable:

	Govt. Trust (1–4)
Intercept	3.337*** (0.152)
Democracy	-0.508*** (0.076)
Democracy²	0.046*** (0.008)

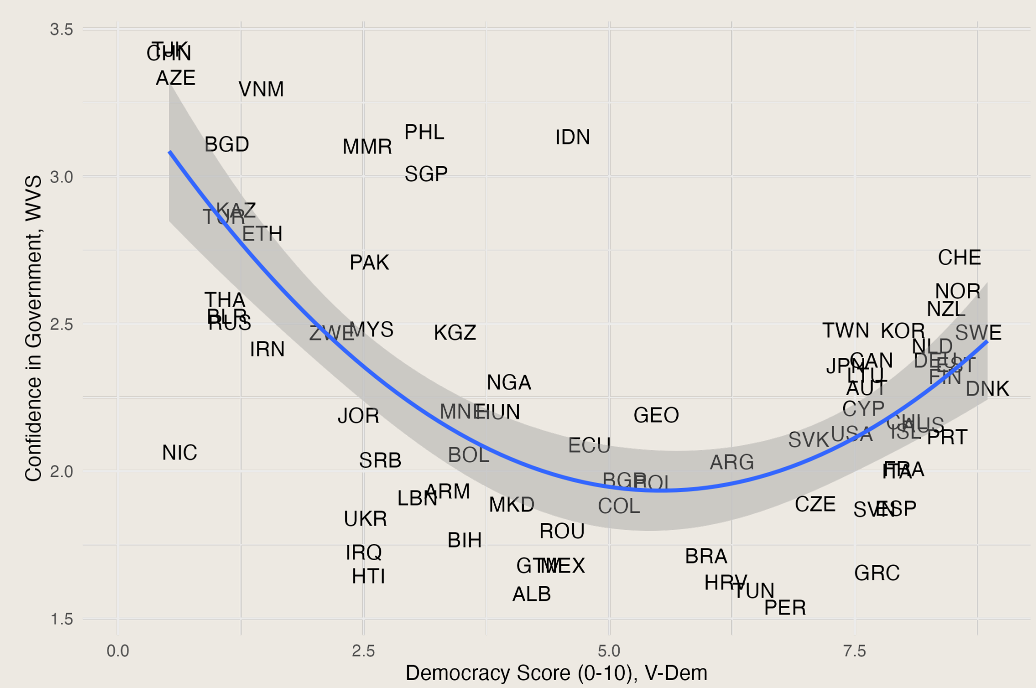


Second-Degree Polynomial: Coefficients

- **Instantaneous rate of change**, expressed by the **derivative**. The derivative of $\hat{Y} = \alpha + \beta_1 X + \beta_2 X^2$ in X is $\beta_1 + 2\beta_2 X$.

- In our model, $-0.508 + 0.092 \times$ Democracy

	Dependent variable:
	Govt. Trust (1–4)
Intercept	3.337*** (0.152)
Democracy	-0.508*** (0.076)
Democracy²	0.046*** (0.008)

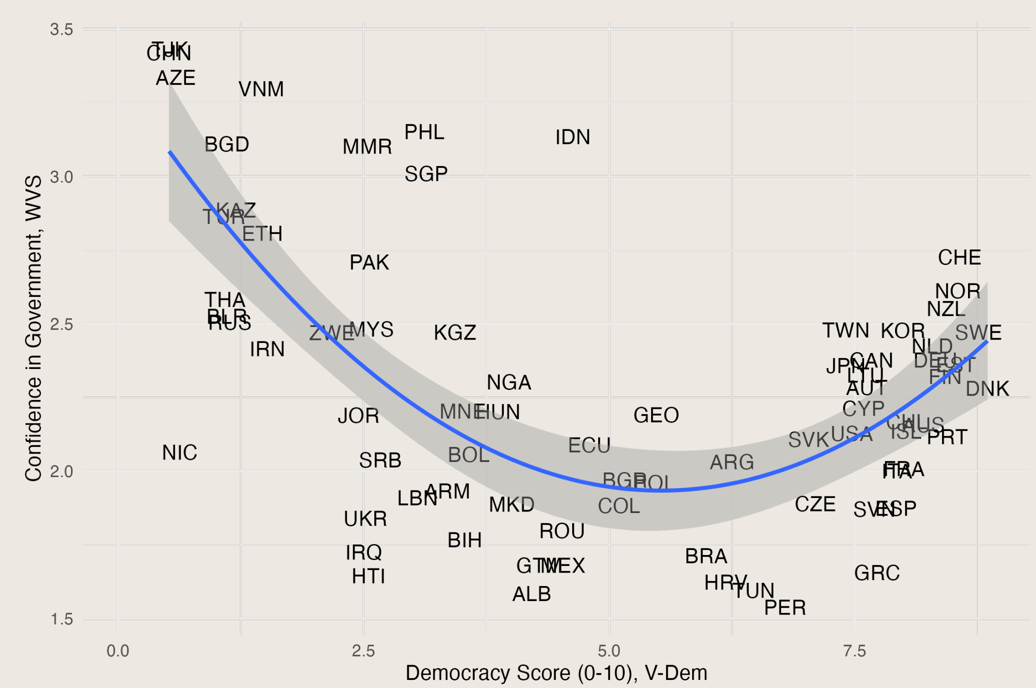


Second-Degree Polynomial: Coefficients

- **Instantaneous rate of change**, expressed by the **derivative**. The derivative of $\hat{Y} = \alpha + \beta_1 X + \beta_2 X^2$ in X is $\beta_1 + 2\beta_2 X$.
- In our model, $-0.508 + 0.092 \times$ Democracy
- Rate of change if Democracy = 1:

Dependent variable:

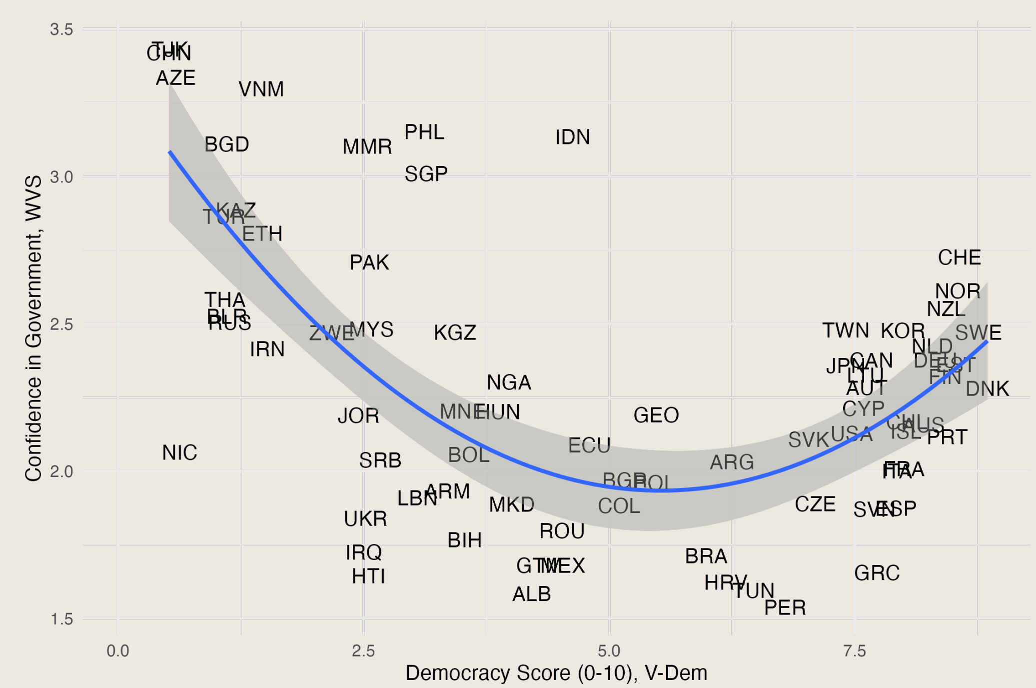
	Govt. Trust (1–4)
Intercept	3.337*** (0.152)
Democracy	-0.508*** (0.076)
Democracy²	0.046*** (0.008)



Second-Degree Polynomial: Coefficients

- **Instantaneous rate of change**, expressed by the **derivative**. The derivative of $\hat{Y} = \alpha + \beta_1 X + \beta_2 X^2$ in X is $\beta_1 + 2\beta_2 X$.
- In our model, $-0.508 + 0.092 \times$ Democracy
- Rate of change if Democracy = 1:
  - $-0.508 + 0.092 \times 1 = -0.416$

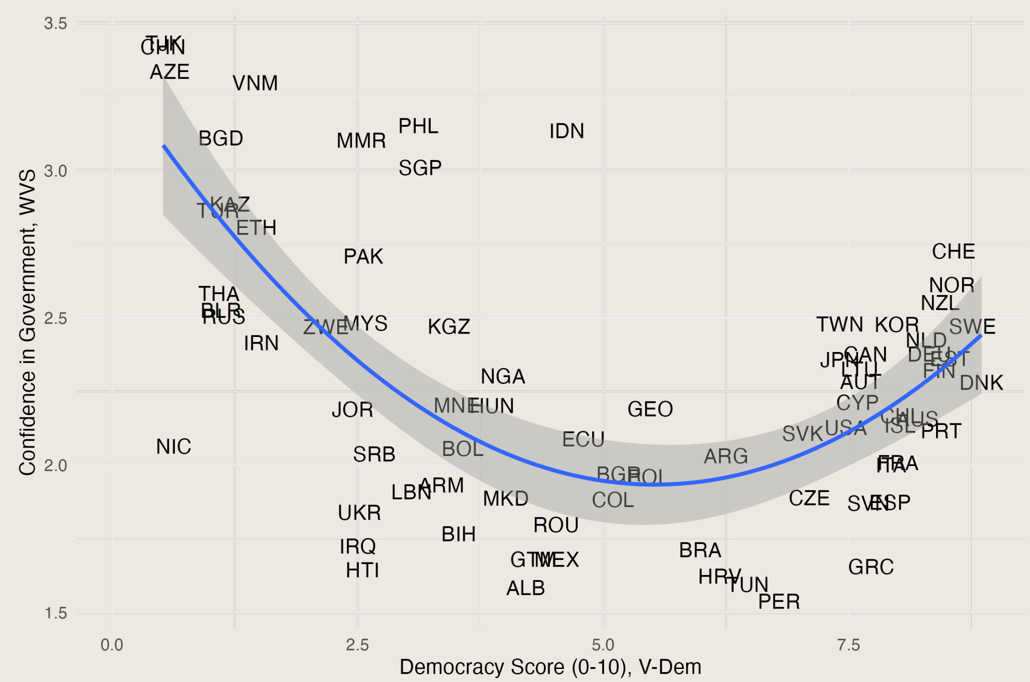
	Dependent variable:
	Govt. Trust (1–4)
Intercept	3.337*** (0.152)
Democracy	-0.508*** (0.076)
Democracy²	0.046*** (0.008)



Second-Degree Polynomial: Coefficients

- **Instantaneous rate of change**, expressed by the **derivative**. The derivative of $\hat{Y} = \alpha + \beta_1 X + \beta_2 X^2$ in X is $\beta_1 + 2\beta_2 X$.
- In our model, $-0.508 + 0.092 \times$ Democracy
- Rate of change if Democracy = 1:
  - $-0.508 + 0.092 \times 1 = -0.416$
- Rate of change in Democracy = 5:

	Dependent variable:
	Govt. Trust (1–4)
Intercept	3.337*** (0.152)
Democracy	-0.508*** (0.076)
Democracy²	0.046*** (0.008)

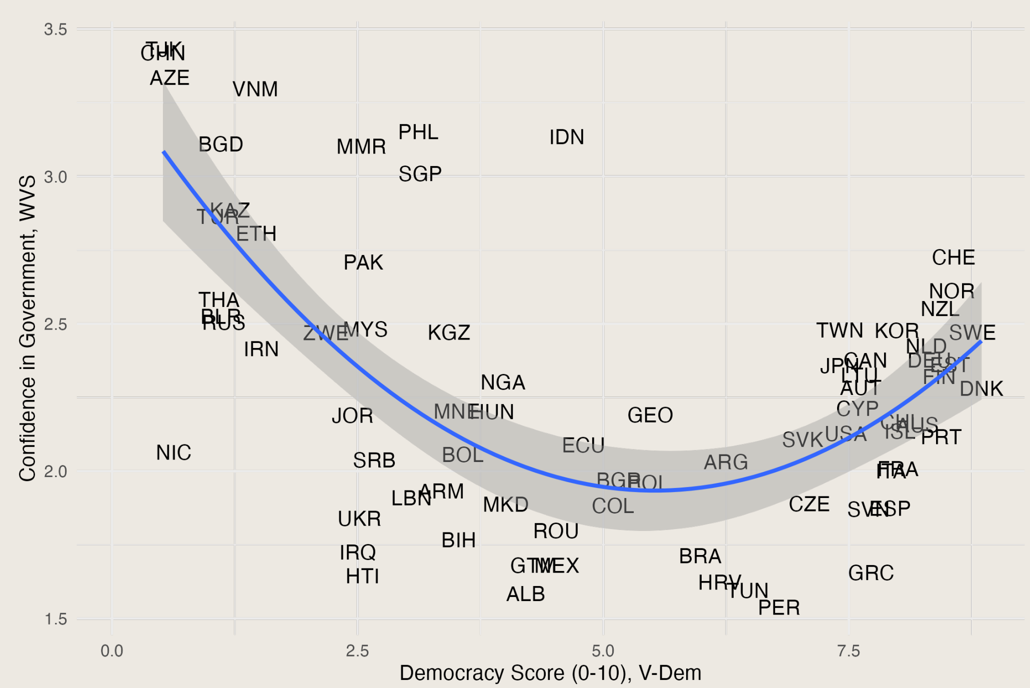


Second-Degree Polynomial: Coefficients

- **Instantaneous rate of change**, expressed by the **derivative**. The derivative of $\hat{Y} = \alpha + \beta_1 X + \beta_2 X^2$ in X is $\beta_1 + 2\beta_2 X$.
- In our model, $-0.508 + 0.092 \times$ Democracy
- Rate of change if Democracy = 1:
  - $-0.508 + 0.092 \times 1 = -0.416$
- Rate of change in Democracy = 5:
  - $-0.508 + 0.092 \times 5 = -0.048$

Dependent variable:

	Govt. Trust (1–4)
Intercept	3.337*** (0.152)
Democracy	-0.508*** (0.076)
Democracy²	0.046*** (0.008)

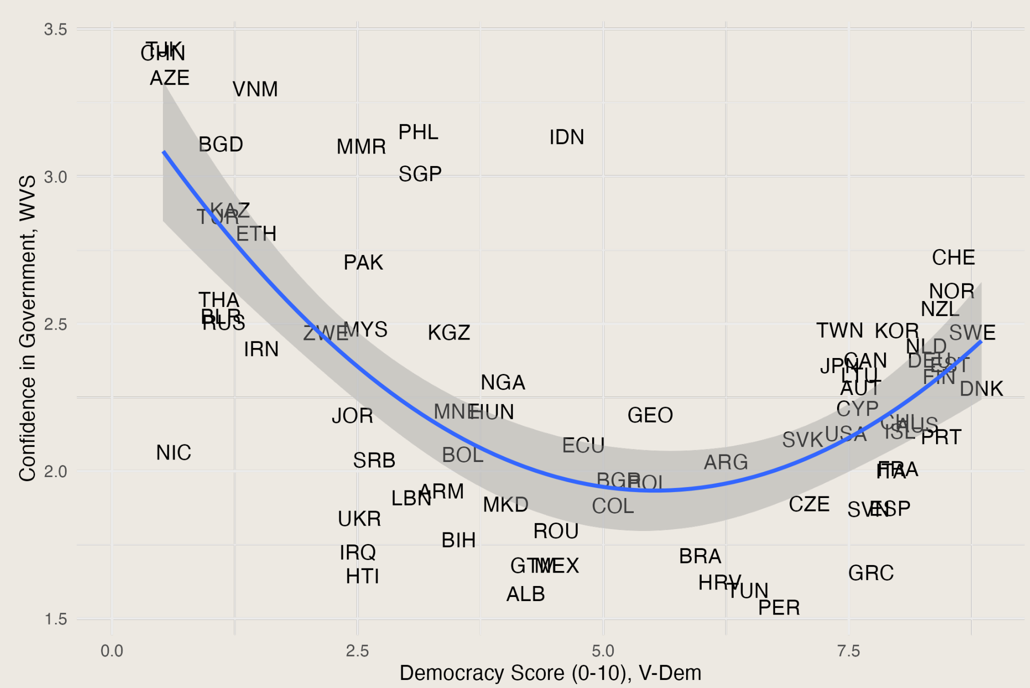


Second-Degree Polynomial: Coefficients

- **Instantaneous rate of change**, expressed by the **derivative**. The derivative of $\hat{Y} = \alpha + \beta_1 X + \beta_2 X^2$ in X is $\beta_1 + 2\beta_2 X$.
- In our model, $-0.508 + 0.092 \times$ Democracy
- Rate of change if Democracy = 1:
  - $-0.508 + 0.092 \times 1 = -0.416$
- Rate of change in Democracy = 5:
  - $-0.508 + 0.092 \times 5 = -0.048$
- Rate of change in Democracy = 8:

Dependent variable:

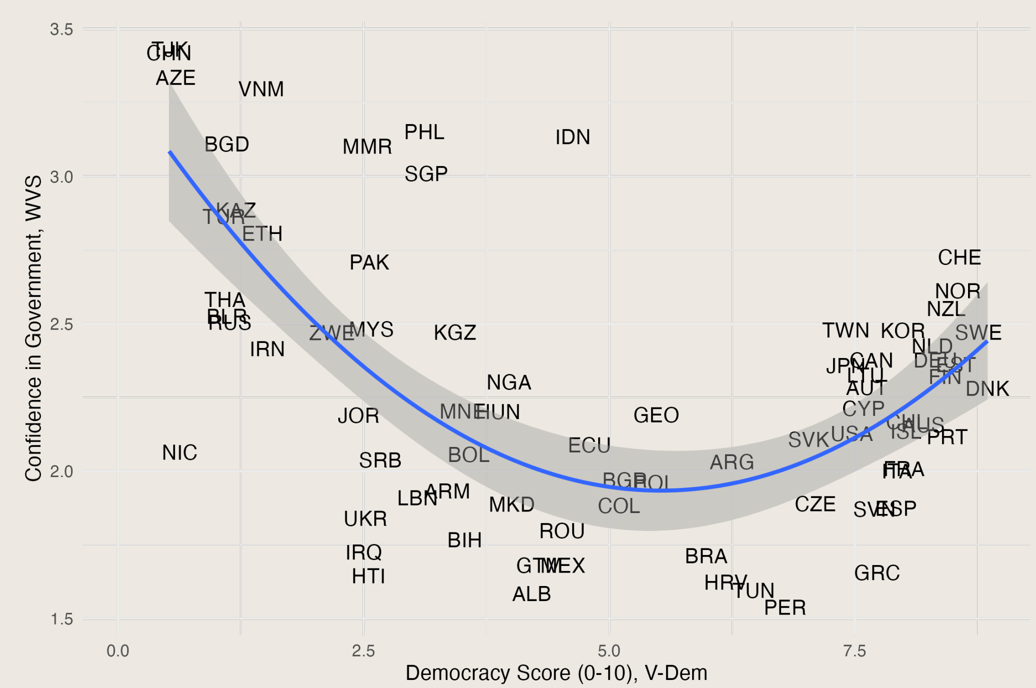
	Govt. Trust (1–4)
Intercept	3.337*** (0.152)
Democracy	-0.508*** (0.076)
Democracy²	0.046*** (0.008)



Second-Degree Polynomial: Coefficients

- **Instantaneous rate of change**, expressed by the **derivative**. The derivative of $\hat{Y} = \alpha + \beta_1 X + \beta_2 X^2$ in X is $\beta_1 + 2\beta_2 X$.
- In our model, $-0.508 + 0.092 \times$ Democracy
- Rate of change if Democracy $= 1$:
  - $-0.508 + 0.092 \times 1 = -0.416$
- Rate of change in Democracy $= 5$:
  - $-0.508 + 0.092 \times 5 = -0.048$
- Rate of change in Democracy $= 8$:
  - $-0.508 + 0.092 \times 8 = +0.228$, etc.

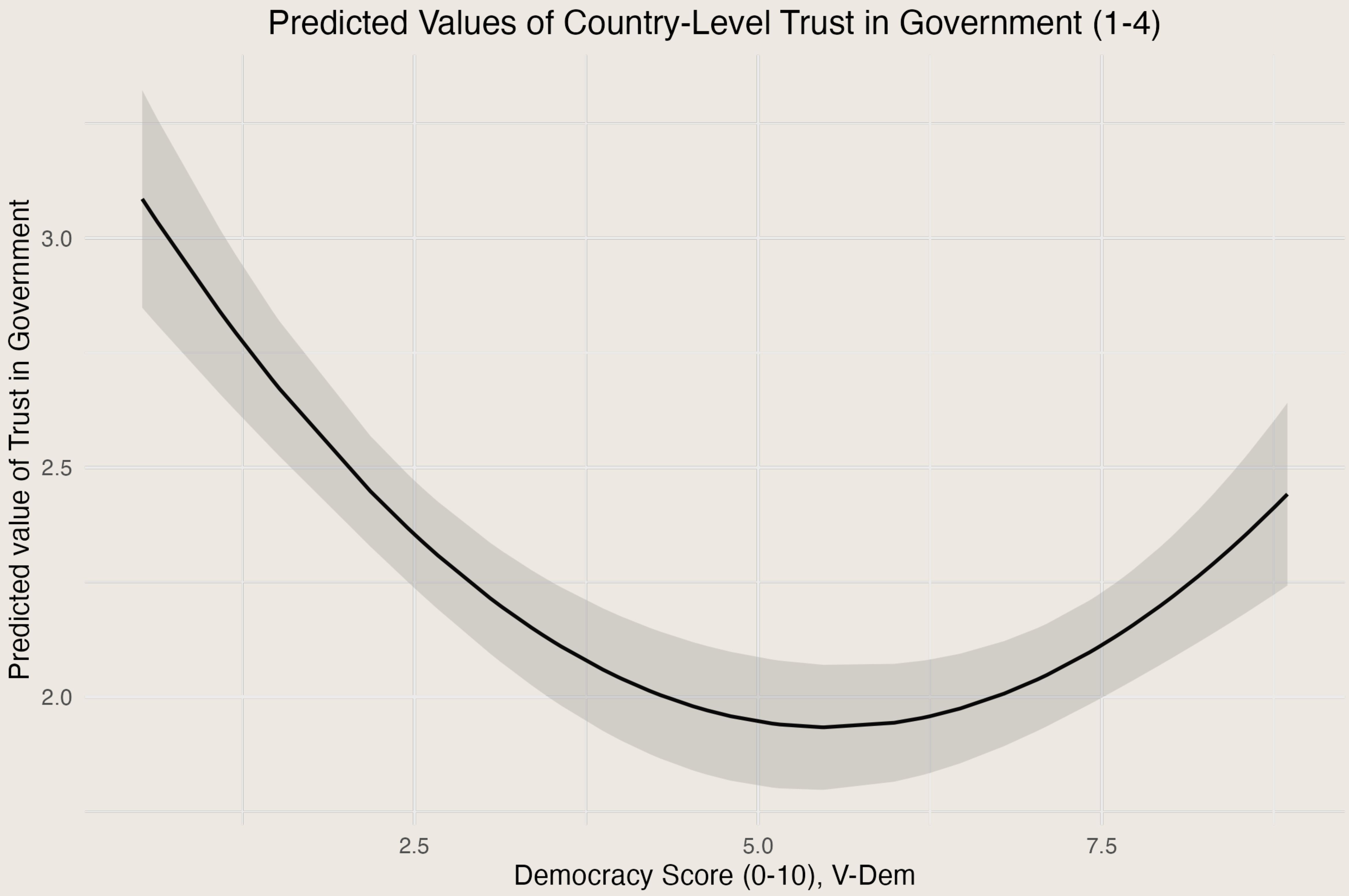
	Dependent variable:
	Govt. Trust (1–4)
Intercept	3.337*** (0.152)
Democracy	-0.508*** (0.076)
Democracy²	0.046*** (0.008)



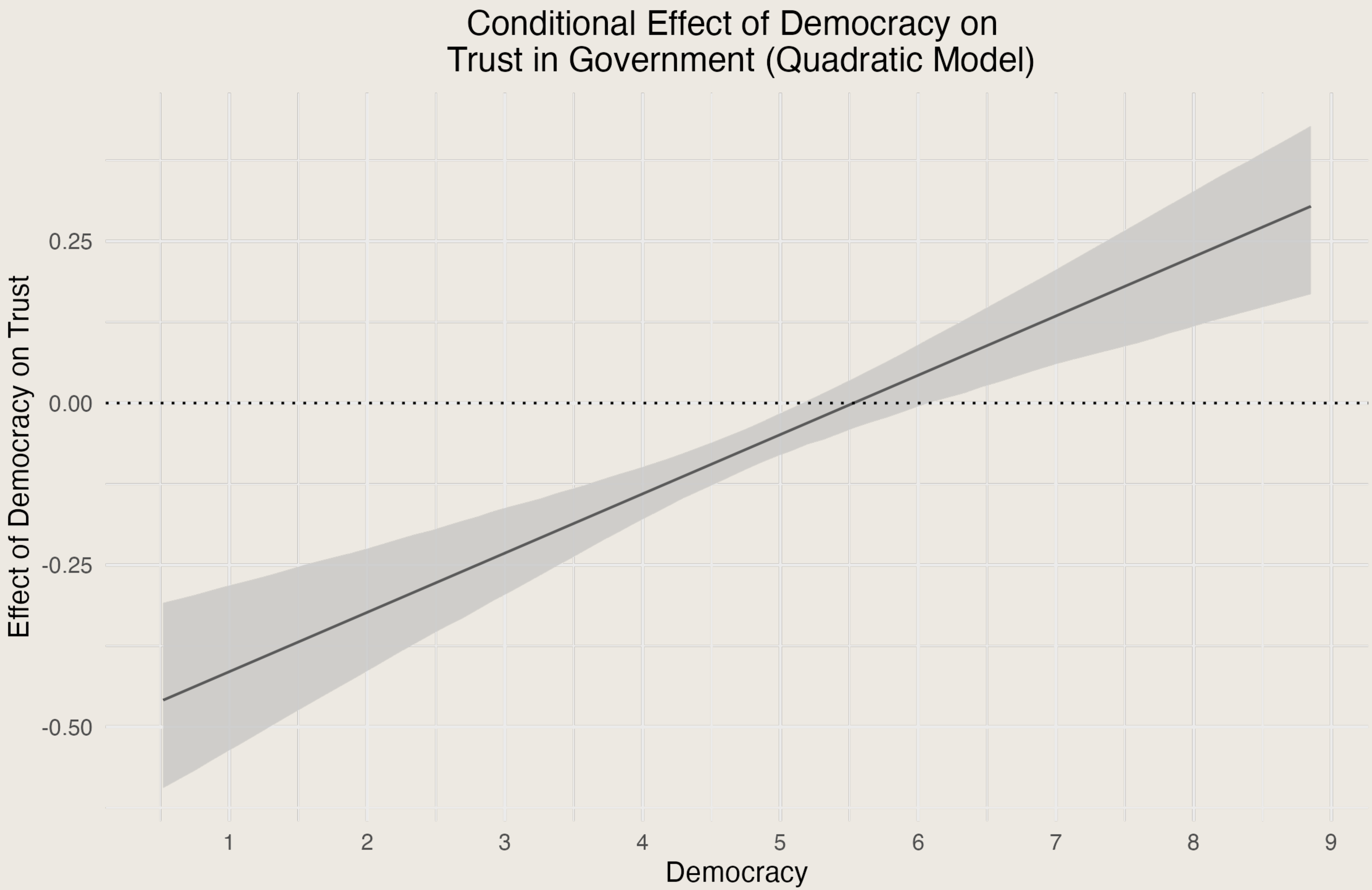
Polynomial Terms in R

```
> model1 <- lm(conf_goverment ~ democracy + I(democracy^2), data = qog)
> stargazer(model1, type = "text", single.row = TRUE)


===============================================
                        Dependent variable:
                    ---------------------------
                          conf_goverment
-----------------------------------------------
democracy                -0.508*** (0.076)
I(democracy2)            0.046*** (0.008)
Constant                 3.337*** (0.152)
-----------------------------------------------
Observations                     76
R2                              0.417
Adjusted R2                     0.401
Residual Std. Error        0.366 (df = 73)
F Statistic           26.076*** (df = 2; 73)
===============================================
Note:               *p<0.1; **p<0.05; ***p<0.01
```

Visualisation: Predicted Values Plot



Visualisation: Conditional Effect Plot



Check if you understand

- How does a leader's time in office affect spending in Chinese counties?

Dependent Variable: Annual Growth Rate of Expenditures Per Capita Explanatory Variables	Party Secretary Model Coefficient (Standard Error)	
$(\text{Time in office})^2$	−0.3946**	−0.4860**
	(0.1728)	(0.2049)
Time in office	2.4793**	3.1624**
	(1.0212)	(1.2252)
Annual growth rate of revenues per capita	0.2493***	0.2589***
	(0.0142)	(0.0166)
Annual growth rate of subsidies per capita		0.1411***
		(0.0092)

- Guo, G. (2009). China's local political budget cycles. *American Journal of Political Science, 53*(3), 621-632.

Higher-Order Polynomials

Higher-Order Polynomials

* You can add higher-order terms (X^3, X^4, etc.) to model more complex non-linearities. In general, a polynomial of order n corresponds to a curve with $n - 1$ bends.

Higher-Order Polynomials

- You can add higher-order terms (X^3, X^4, etc.) to model more complex non-linearities. In general, a polynomial of order n corresponds to a curve with $n - 1$ bends.
- You always want to **include lower-order terms**. E.g., if you want to have X^3, you should also have X^2 and X.

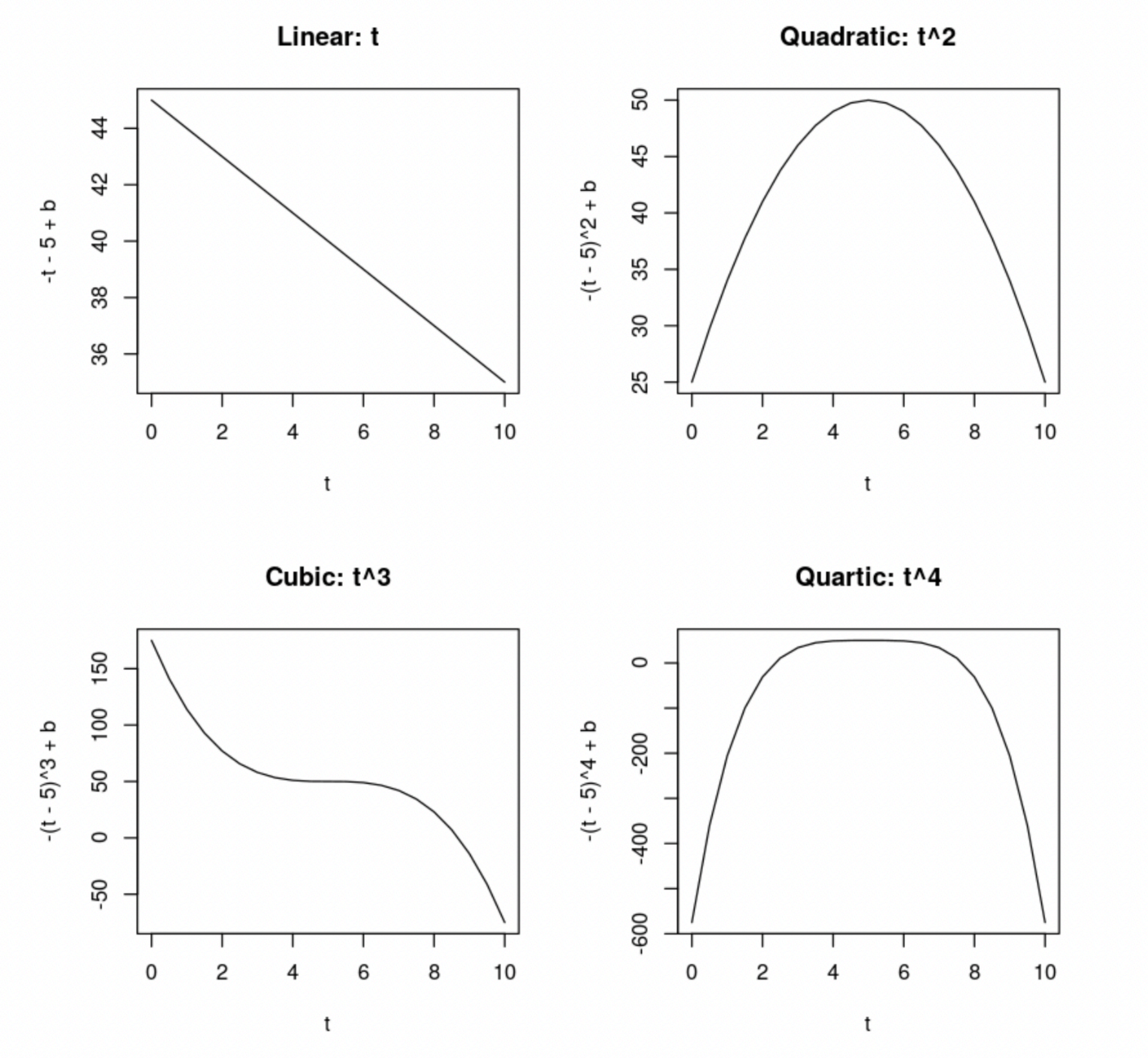
Higher-Order Polynomials

- You can add higher-order terms (X^3, X^4, etc.) to model more complex non-linearities. In general, a polynomial of order n corresponds to a curve with $n-1$ bends.
- You always want to **include lower-order terms**. E.g., if you want to have X^3, you should also have X^2 and X.
- If a quadratic term doesn't improve the model, it's unlikely a cubic term will do, and so on. In practice, it **(almost) never makes sense to go beyond a cubic**.

Higher-Order Polynomials

* You can add higher-order terms (X^3, X^4, etc.) to model more complex non-linearities. In general, a polynomial of order n corresponds to a curve with $n - 1$ bends.
* You always want to **include lower-order terms**. E.g., if you want to have X^3, you should also have X^2 and X.
* If a quadratic term doesn't improve the model, it's unlikely a cubic term will do, and so on. In practice, it **(almost) never makes sense to go beyond a cubic**.
* Interpretation gets trickier. Use visualisation tools to get a sense of what you're fitting.

Higher-Order Polynomials



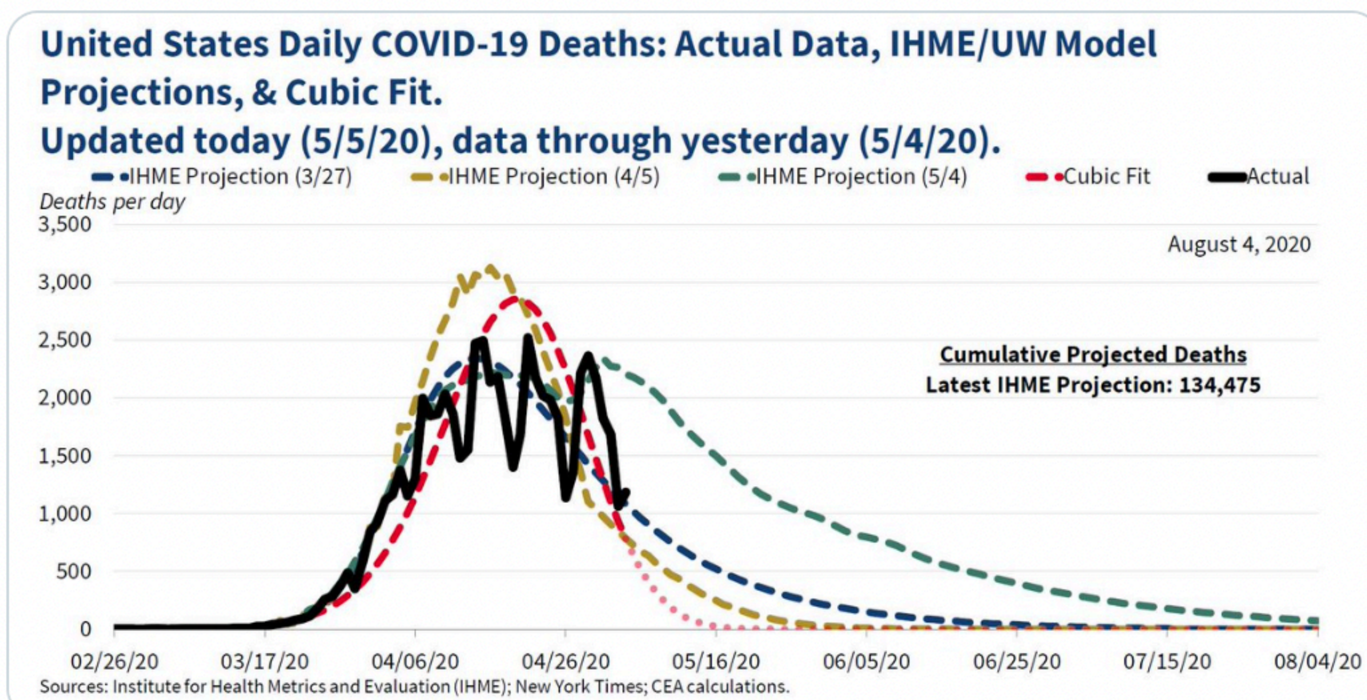
Higher-Order Polynomials: Handle with Care

Higher-Order Polynomials: Handle with Care

CEA45 Archived
@WhiteHouseCEA45 · **Follow**

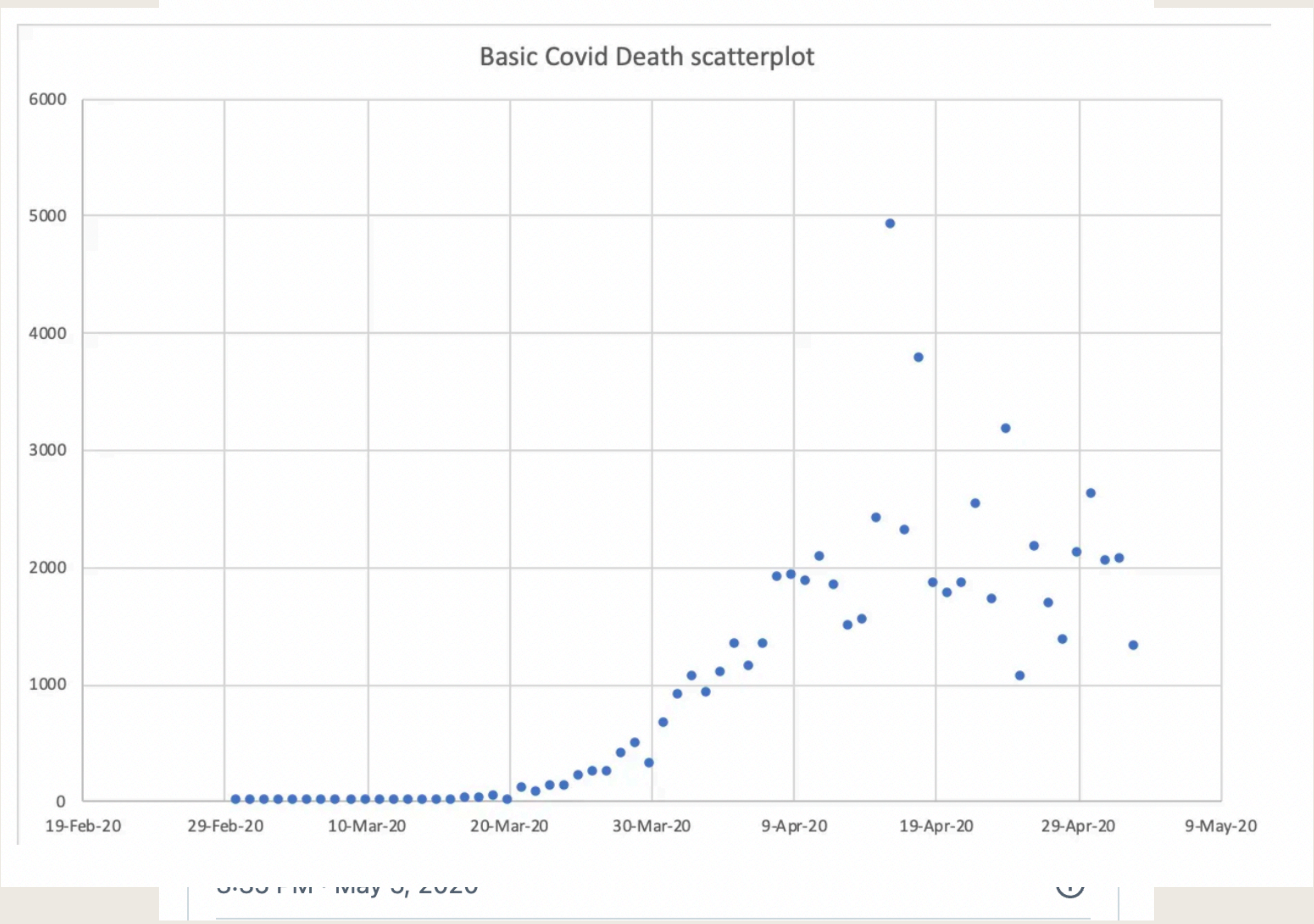
Replying to @WhiteHouseCEA45

To better visualize observed data, we also continually update a curve-fitting exercise to summarize COVID-19's observed trajectory. Particularly with irregular data, curve fitting can improve data visualization. As shown, IHME's mortality curves have matched the data fairly well.

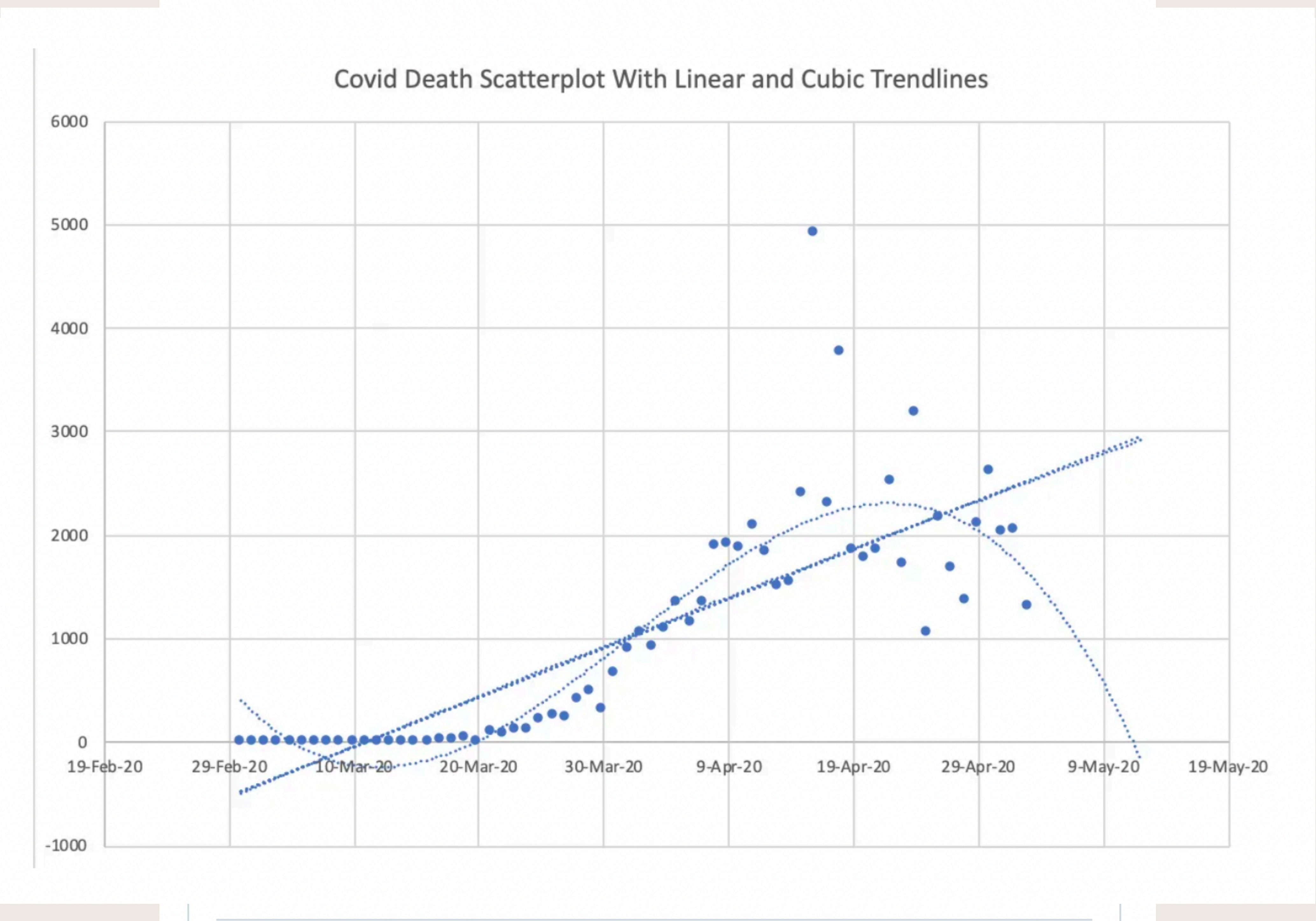


3:35 PM · May 5, 2020

Higher-Order Polynomials: Handle with Care



Higher-Order Polynomials: Handle with Care



Log-Transformations

Log-Transformations

- Useful when dealing with variables that are **positive** and **right-skewed**:

Log-Transformations

- Useful when dealing with variables that are **positive** and **right-skewed**:
  - **Income**: lots of people around the median income, and a handful of mega-rich.

Log-Transformations

- Useful when dealing with variables that are **positive** and **right-skewed**:
  - **Income**: lots of people around the median income, and a handful of mega-rich.
  - **Population:** 50% of countries below 10m people (10^7). Then there's China and India, with 1bn people (10^9).

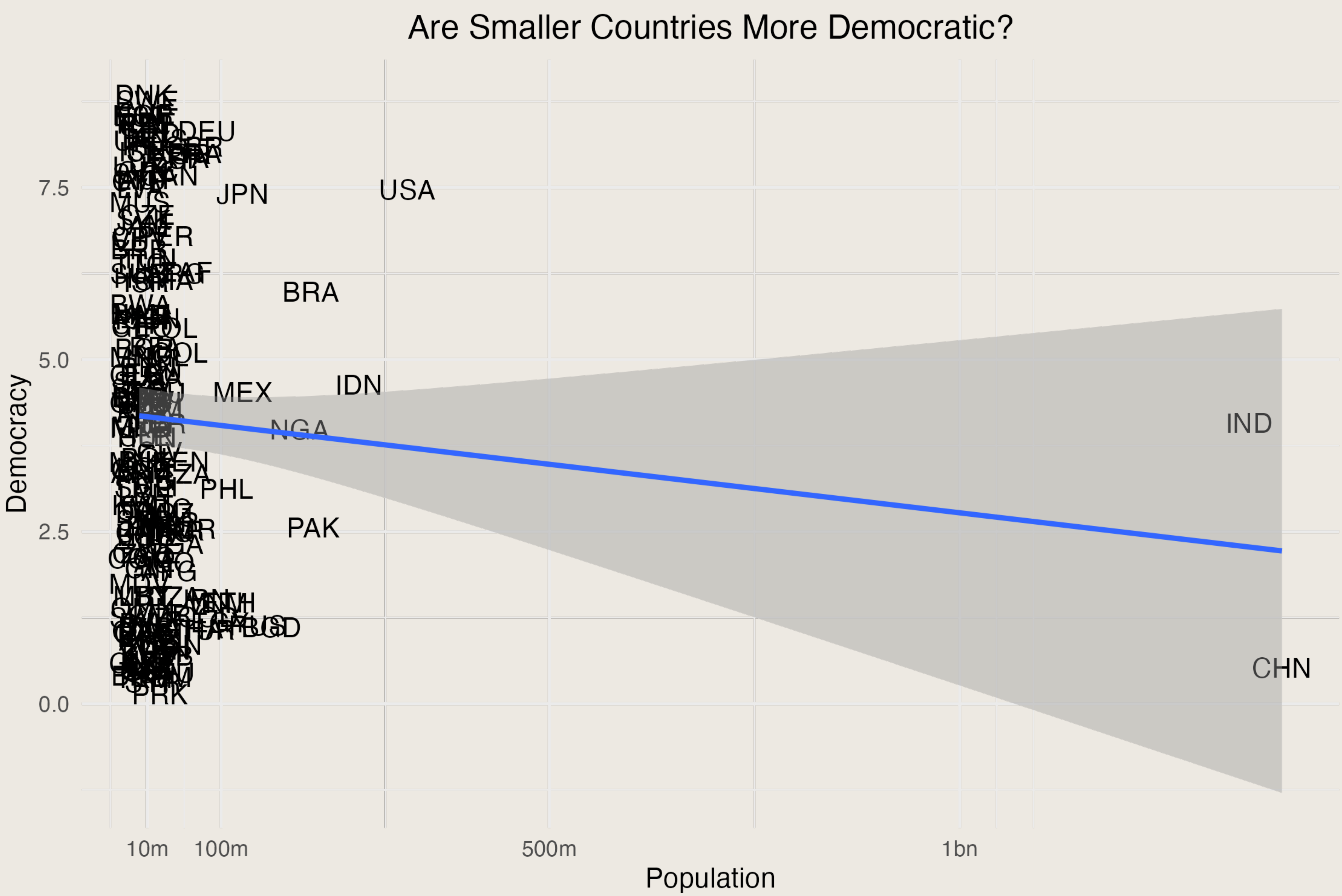
Log-Transformations

- Useful when dealing with variables that are **positive** and **right-skewed**:
  - **Income**: lots of people around the median income, and a handful of mega-rich.
  - **Population:** 50% of countries below 10m people (10^7). Then there's China and India, with 1bn people (10^9).
  - **GDP per capita:** 80% of countries below $50k. Then, there's Luxembourg, Singapore and Qatar (> $125k).

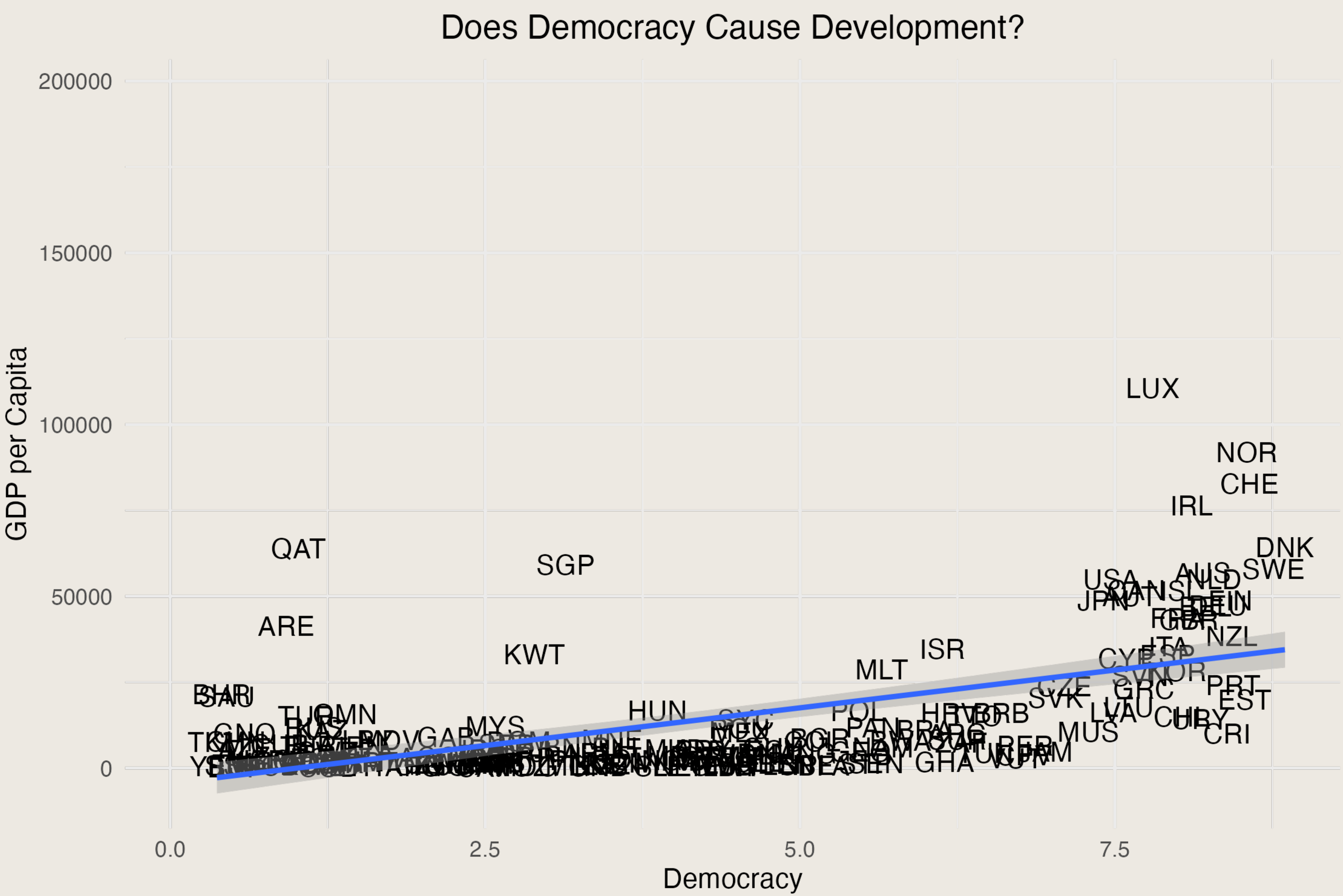
Log-Transformations

- Useful when dealing with variables that are **positive** and **right-skewed**:
  - **Income**: lots of people around the median income, and a handful of mega-rich.
  - **Population:** 50% of countries below 10m people (10^7). Then there's China and India, with 1bn people (10^9).
  - **GDP per capita:** 80% of countries below $50k. Then, there's Luxembourg, Singapore and Qatar (> $125k).
- Linear relationships are unlikely with these variables as your predictors, outcomes or both.

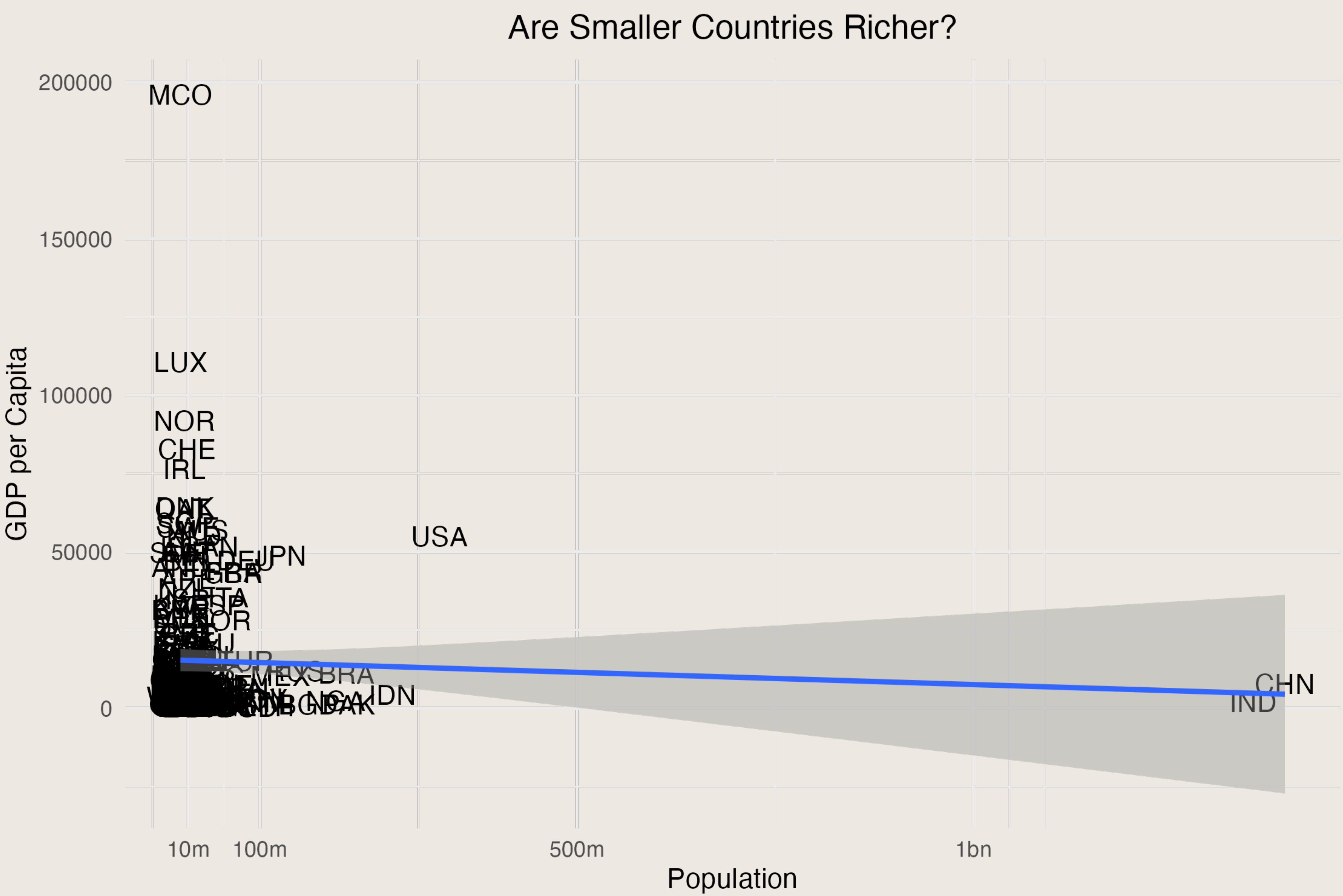
Log-Transformations



Log-Transformations



Log-Transformations



Log-Transformations

Log-Transformations

- We can **unskew** these variables by taking their natural logarithm (notated as **log**, or **ln**). *Reminder*:

Log-Transformations

- We can **unskew** these variables by taking their natural logarithm (notated as **log**, or **ln**). *Reminder*:
  - If $\log(a) = b$, then $e^b = a$, where $e \approx 2.71828$.

Log-Transformations

- We can **unskew** these variables by taking their natural logarithm (notated as **log**, or **ln**). *Reminder*:
  - If $\log(a) = b$, then $e^b = a$, where $e \approx 2.71828$.
- How it works in practice:

Log-Transformations

* We can **unskew** these variables by taking their natural logarithm (notated as **log**, or **ln**). *Reminder*:
    * If $\log(a) = b$, then $e^b = a$, where $e \approx 2.71828$.
* How it works in practice:
    * $\log(1) = 0$

Log-Transformations

- We can **unskew** these variables by taking their natural logarithm (notated as **log**, or **ln**). *Reminder*:
  - If $\log(a) = b$, then $e^b = a$, where $e \approx 2.71828$.
- How it works in practice:
  - $\log(1) = 0$
  - $\log(10) \approx 2.30$

Log-Transformations

- We can **unskew** these variables by taking their natural logarithm (notated as **log**, or **ln**). *Reminder*:
  - If $\log(a) = b$, then $e^b = a$, where $e \approx 2.71828$.
- How it works in practice:
  - $\log(1) = 0$
  - $\log(10) \approx 2.30$
  - $\log(100) \approx 4.60$

Log-Transformations

- We can **unskew** these variables by taking their natural logarithm (notated as **log**, or **ln**). *Reminder*:
  - If $\log(a) = b$, then $e^b = a$, where $e \approx 2.71828$.
- How it works in practice:
  - $\log(1) = 0$
  - $\log(10) \approx 2.30$
  - $\log(100) \approx 4.60$
  - $\log(1000) \approx 6.91$

Log-Transformations

- We can **unskew** these variables by taking their natural logarithm (notated as **log**, or **ln**). *Reminder*:
  - If $\log(a) = b$, then $e^b = a$, where $e \approx 2.71828$.
- How it works in practice:
  - $\log(1) = 0$
  - $\log(10) \approx 2.30$
  - $\log(100) \approx 4.60$
  - $\log(1000) \approx 6.91$
  - $\log(10^6) \approx 13.82$

Log-Transformations

- We can **unskew** these variables by taking their natural logarithm (notated as **log**, or **ln**). *Reminder*:
  - If $\log(a) = b$, then $e^b = a$, where $e \approx 2.71828$.
- How it works in practice:
  - $\log(1) = 0$
  - $\log(10) \approx 2.30$
  - $\log(100) \approx 4.60$
  - $\log(1000) \approx 6.91$
  - $\log(10^6) \approx 13.82$
  - (Careful: you can't take logs of zero or negative numbers!)

Log-Transformations: Example

Log-Transformations: Example

* Famous paper: Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). **The colonial origins of comparative development: An empirical investigation**. *American Economic Review*, *91*(5), 1369-1401.

Log-Transformations: Example

- Famous paper: Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). **The colonial origins of comparative development: An empirical investigation**. *American Economic Review*, *91*(5), 1369-1401.
- Argument: Colonial powers set up **extractive institutions** in places where they faced high mortality rates (due to e.g. diseases). Where they can settle easily, they set up **growth-inducing institutions**, like property rights. Long-run growth is thus related to initial conditions faced by settlers:

Log-Transformations: Example

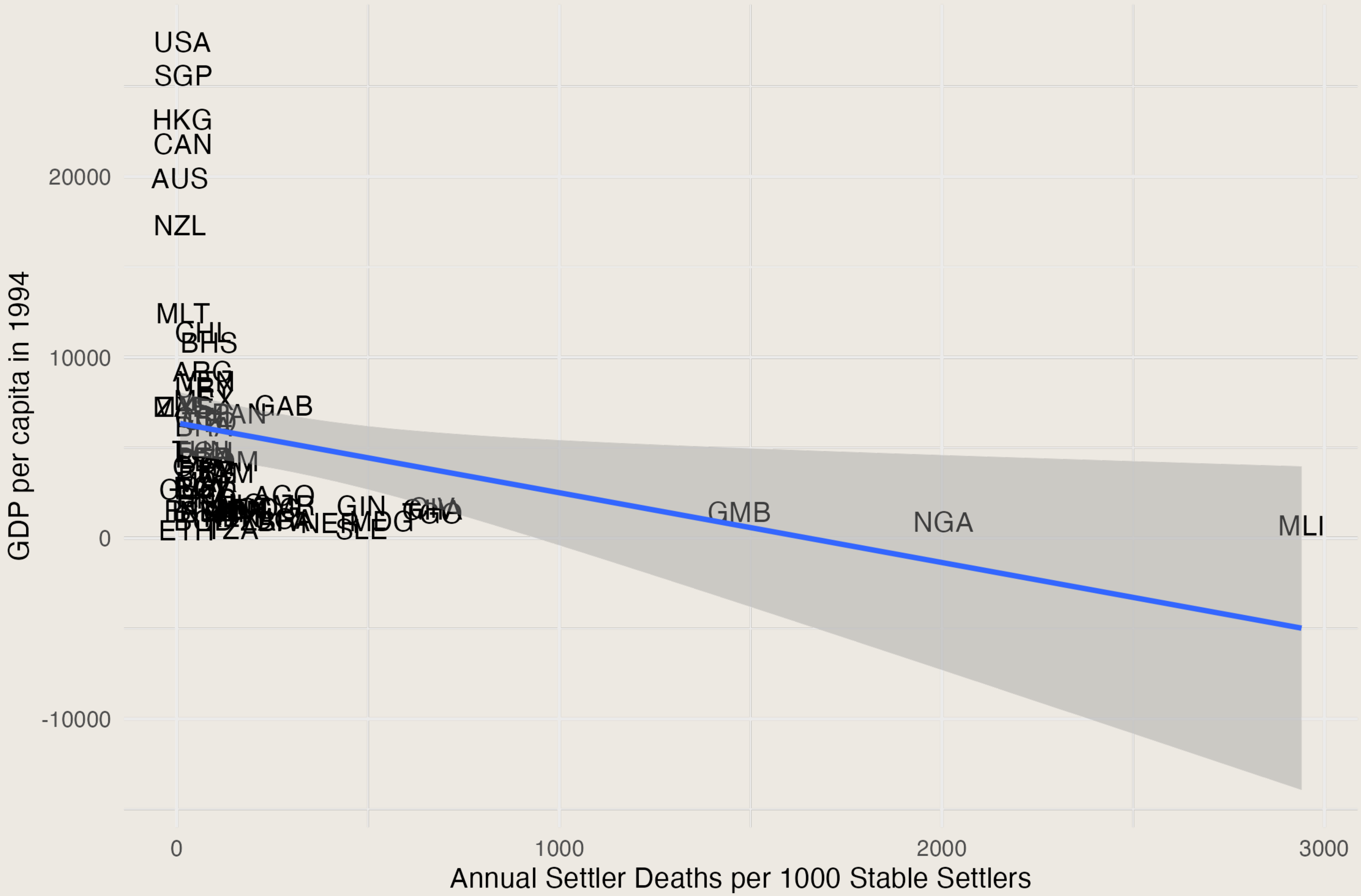
- Famous paper: Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). **The colonial origins of comparative development: An empirical investigation**. *American Economic Review*, *91*(5), 1369-1401.

- Argument: Colonial powers set up **extractive institutions** in places where they faced high mortality rates (due to e.g. diseases). Where they can settle easily, they set up **growth-inducing institutions**, like property rights. Long-run growth is thus related to initial conditions faced by settlers:

- $\text{GDP in 1995} = \alpha + \beta \text{Settler Mortality} + \epsilon$

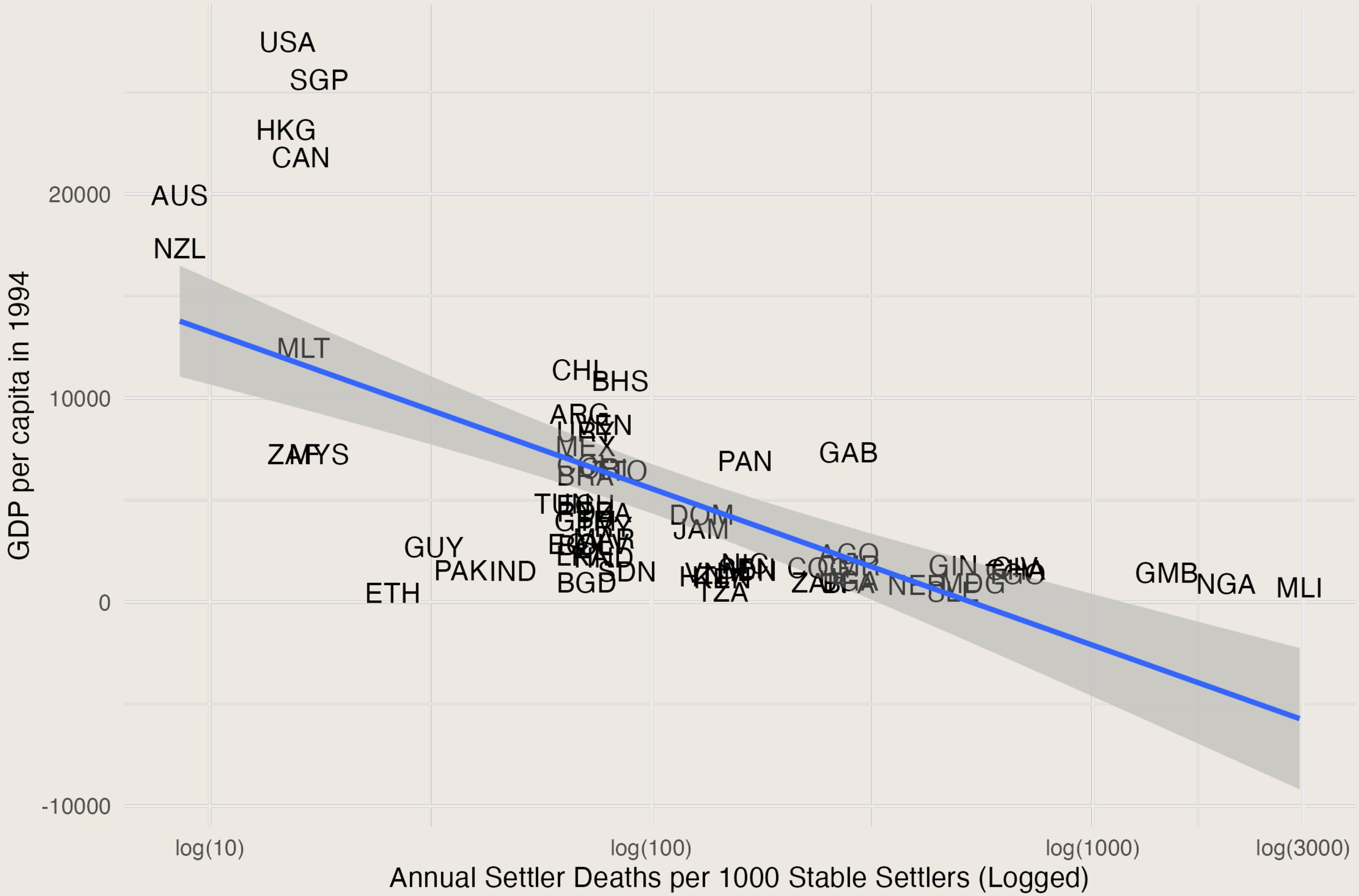
Log-Transformations: Example

$$\text{GDP in 1995} = \alpha + \beta\text{Settler Mortality} + \epsilon$$



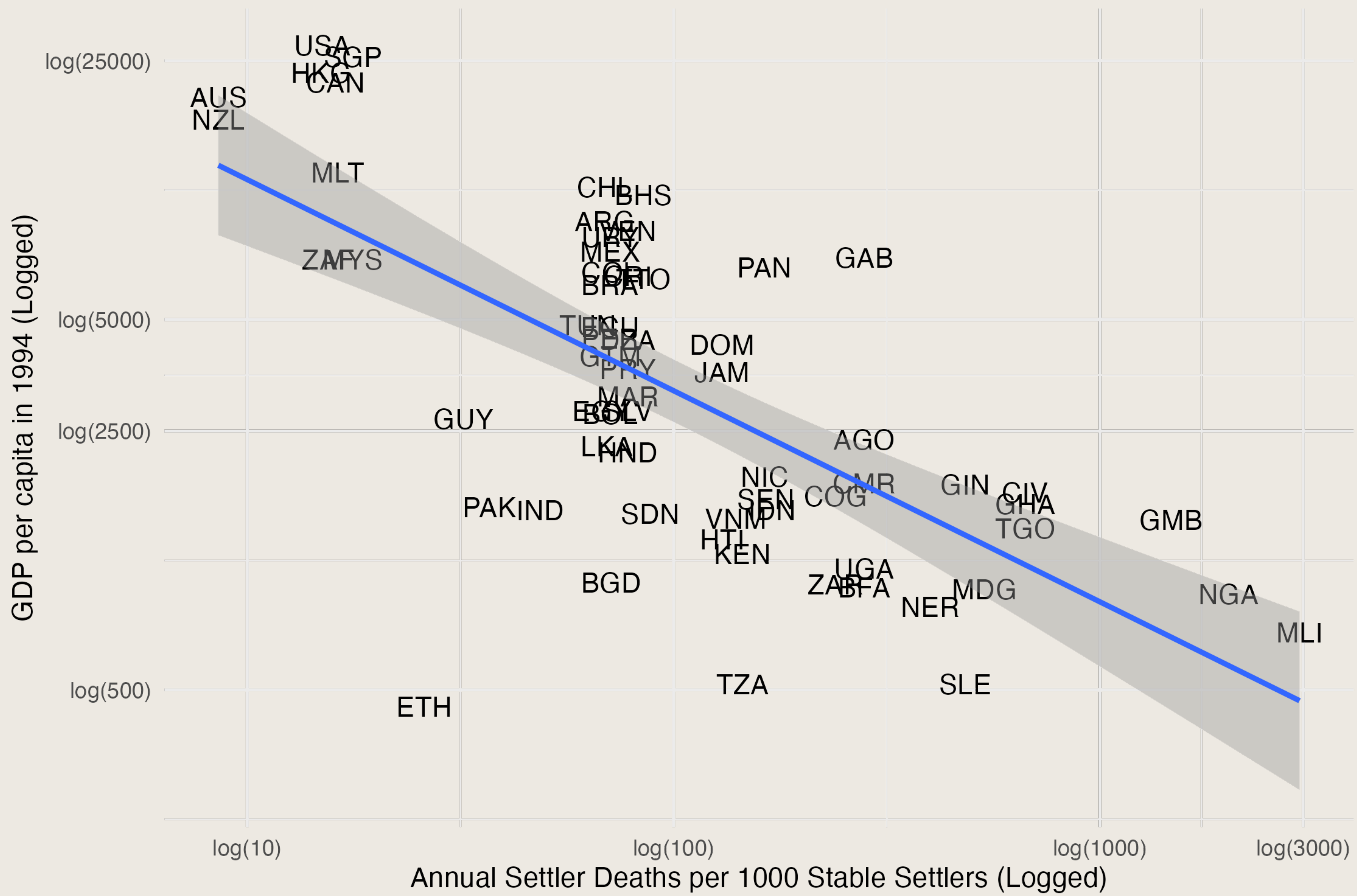
Log-Transformations: Example

$$\text{GDP in 1995} = \alpha + \beta \log(\text{Settler Mortality}) + \epsilon$$



Log-Transformations: Example

$$\log(\text{GDP in 1995}) = \alpha + \beta \log(\text{Settler Mortality}) + \epsilon$$



Log Coefficients: Interpretation

Log Coefficients: Interpretation

- Interesting property of logarithms: can interpret the coefficients in terms of **percentage** change (an approximation, valid only for small increases).

Log Coefficients: Interpretation

* Interesting property of logarithms: can interpret the coefficients in terms of **percentage** change (an approximation, valid only for small increases).
* **Level-Level** model $Y = \alpha + \beta X + \epsilon$

Log Coefficients: Interpretation

- Interesting property of logarithms: can interpret the coefficients in terms of **percentage** change (an approximation, valid only for small increases).
- **Level-Level** model $Y = \alpha + \beta X + \epsilon$
  - One-unit change in $X \rightarrow Y$ predicted to change by β

Log Coefficients: Interpretation

- Interesting property of logarithms: can interpret the coefficients in terms of **percentage** change (an approximation, valid only for small increases).
- **Level-Level** model $Y = \alpha + \beta X + \epsilon$
  - One-unit change in $X \rightarrow Y$ predicted to change by β
- **Level-Log** model $Y = \alpha + \beta \log(X) + \epsilon$

Log Coefficients: Interpretation

- Interesting property of logarithms: can interpret the coefficients in terms of **percentage** change (an approximation, valid only for small increases).
- **Level-Level** model $Y = \alpha + \beta X + \epsilon$
  - One-unit change in $X \rightarrow Y$ predicted to change by β
- **Level-Log** model $Y = \alpha + \beta \log(X) + \epsilon$
  - 1% change in $X \rightarrow Y$ predicted to change by $(\beta/100)$

Log Coefficients: Interpretation

- Interesting property of logarithms: can interpret the coefficients in terms of **percentage** change (an approximation, valid only for small increases).
- **Level-Level** model $Y = \alpha + \beta X + \epsilon$
  - One-unit change in $X \rightarrow Y$ predicted to change by β
- **Level-Log** model $Y = \alpha + \beta \log(X) + \epsilon$
  - 1% change in $X \rightarrow Y$ predicted to change by $(\beta/100)$
- **Log-Level** model $\log(Y) = \alpha + \beta X + \epsilon$

Log Coefficients: Interpretation

- Interesting property of logarithms: can interpret the coefficients in terms of **percentage** change (an approximation, valid only for small increases).
- **Level-Level** model $Y = \alpha + \beta X + \epsilon$
  - One-unit change in $X \rightarrow Y$ predicted to change by β
- **Level-Log** model $Y = \alpha + \beta \log(X) + \epsilon$
  - 1% change in $X \rightarrow Y$ predicted to change by $(\beta/100)$
- **Log-Level** model $\log(Y) = \alpha + \beta X + \epsilon$
  - One-unit change in $X \rightarrow Y$ predicted to change by $\beta \times 100\%$

Log Coefficients: Interpretation

- Interesting property of logarithms: can interpret the coefficients in terms of **percentage** change (an approximation, valid only for small increases).
- **Level-Level** model $Y = \alpha + \beta X + \epsilon$
  - One-unit change in $X \rightarrow Y$ predicted to change by β
- **Level-Log** model $Y = \alpha + \beta \log(X) + \epsilon$
  - 1% change in $X \rightarrow Y$ predicted to change by $(\beta/100)$
- **Log-Level** model $\log(Y) = \alpha + \beta X + \epsilon$
  - One-unit change in $X \rightarrow Y$ predicted to change by $\beta \times 100\%$
- **Log-Log** model $\log(Y) = \alpha + \beta \log(X) + \epsilon$

Log Coefficients: Interpretation

- Interesting property of logarithms: can interpret the coefficients in terms of **percentage** change (an approximation, valid only for small increases).
- **Level-Level** model $Y = \alpha + \beta X + \epsilon$
  - One-unit change in $X \rightarrow Y$ predicted to change by β
- **Level-Log** model $Y = \alpha + \beta \log(X) + \epsilon$
  - 1% change in $X \rightarrow Y$ predicted to change by $(\beta/100)$
- **Log-Level** model $\log(Y) = \alpha + \beta X + \epsilon$
  - One-unit change in $X \rightarrow Y$ predicted to change by $\beta \times 100\%$
- **Log-Log** model $\log(Y) = \alpha + \beta \log(X) + \epsilon$
  - 1% change in $X \rightarrow Y$ changes by $\beta\%$

Log Coefficients: Interpretation

```
> model1 <- lm(log(gdp_per_capita) ~ log(settler_mortality), data = colonialism)
> stargazer(model1, type = "text", single.row = TRUE)

==================================================
                           Dependent variable:
                       ---------------------------
                           log(gdp_per_capita)
--------------------------------------------------
log(settler_mortality)      -0.570*** (0.078)
Constant                    10.700*** (0.374)
--------------------------------------------------
Observations                        64
R2                                0.464
Adjusted R2                       0.456
Residual Std. Error          0.773 (df = 62)
F Statistic             53.766*** (df = 1; 62)
==================================================
Note:                  *p<0.1; **p<0.05; ***p<0.01
```

Wrap-Up: Non-Linearities

Wrap-Up: Non-Linearities

- Polynomial terms are a very flexible tool:

Wrap-Up: Non-Linearities

- Polynomial terms are a very flexible tool:
  - Unlike logs, they can handle changes in effect direction over the range of the predictor, and negative values.

Wrap-Up: Non-Linearities

* Polynomial terms are a very flexible tool:
  * Unlike logs, they can handle changes in effect direction over the range of the predictor, and negative values.
  * Including higher-order terms comes with the risk of overfitting. **Theory** should inform model specification.

Wrap-Up: Non-Linearities

* Polynomial terms are a very flexible tool:
  * Unlike logs, they can handle changes in effect direction over the range of the predictor, and negative values.
  * Including higher-order terms comes with the risk of overfitting. **Theory** should inform model specification.
* Log-transformation are used more narrowly:

Wrap-Up: Non-Linearities

- Polynomial terms are a very flexible tool:
  - Unlike logs, they can handle changes in effect direction over the range of the predictor, and negative values.
  - Including higher-order terms comes with the risk of overfitting. **Theory** should inform model specification.
- Log-transformation are used more narrowly:
  - Non-linearities produced by skewed, positive variables.

Wrap-Up: Non-Linearities

- Polynomial terms are a very flexible tool:
  - Unlike logs, they can handle changes in effect direction over the range of the predictor, and negative values.
  - Including higher-order terms comes with the risk of overfitting. **Theory** should inform model specification.
- Log-transformation are used more narrowly:
  - Non-linearities produced by skewed, positive variables.
  - **Assume proportional relationships**: halving X has approximately the same effect size on Y as doubling X.

Thank you for your kind attention!

Leonardo Carella

leonardo.carella@nuffield.ox.ac.uk