1

2

3

4

How can we minimise the chance of making **mistakes** when creating our research design?

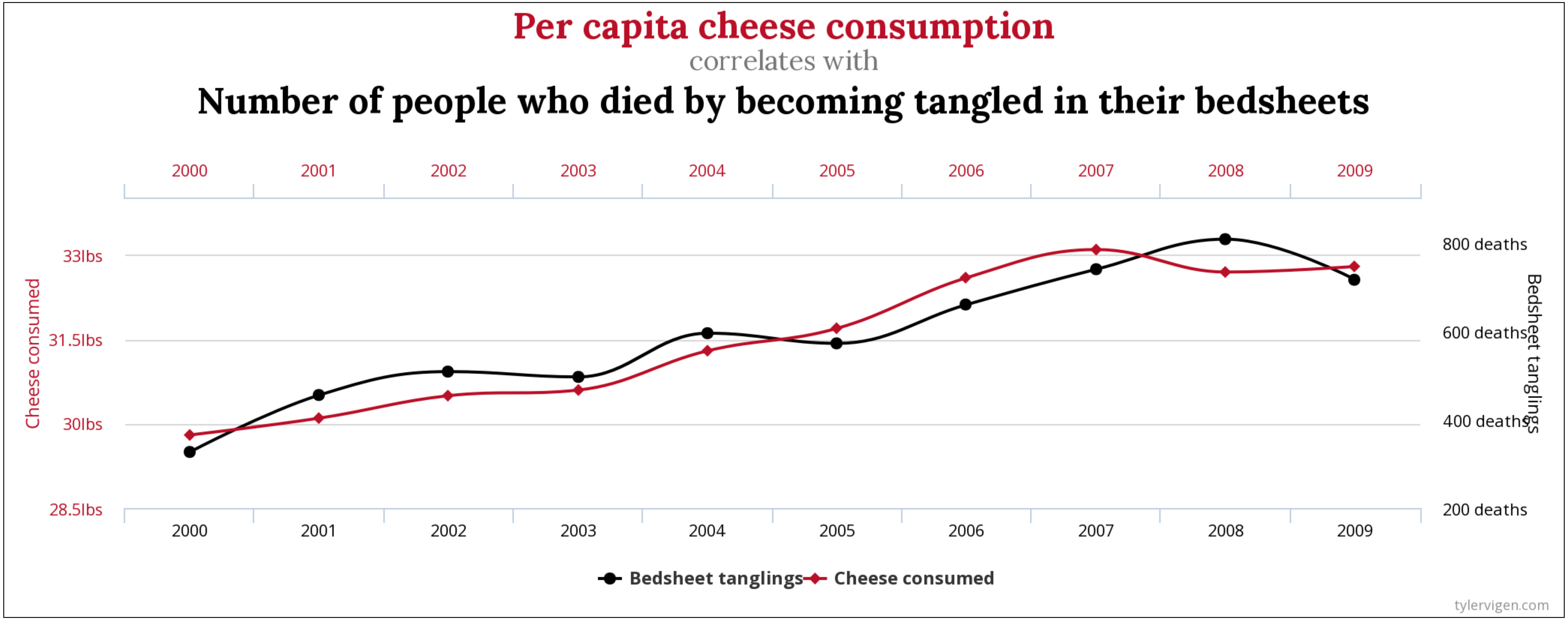
What theoretical, empirical, and simply human factors should we be aware of?


1

1

1. Is there a credible mechanism connecting X and Y?
2. Can we rule out Y causing X (endogeneity)?
3. Is there covariation between X and Y?
4. Have we controlled for potential spuriousness (Z)?

1



Source: https://tylervigen.com/spurious-correlations

1

It is a mistake to think there is a causal link when it could be because of **chance** or a **third factor**.

1

"A **third variable problem** occurs when an observed correlation between two variables can actually be explained by a third variable that has not been accounted for."

Y	**X**	**Z**
# dogs	# fire hydrants	# people
# shark attacks	Ice cream sales	Temperature
Total natural disaster damage	# volunteers showing up to a natural disaster	Size of the natural disaster
Conflict	Trade	State capacity

Sources: https://www.statology.org/third-variable-problem/

1

Questions to ask yourself:

Does **X** cause **Y**?

Does **Y** cause **X**?

Do they **both** affect each other?

1

Democratic history $\longleftrightarrow$ Individual support for democracy

POLITICAL ECONOMY

On the endogeneity of political preferences: Evidence from individual experience with democracy

Nicola Fuchs-Schündeln*† and Matthias Schündeln*†

Democracies depend on the support of the general population, but little is known about the determinants of this support. We investigated whether support for democracy increases with the length of time spent under the system and whether preferences are thus affected by the political system. Relying on 380,000 individual-level observations from 104 countries over the years 1994 to 2013, and exploiting individual-level variation within a country and a given year in the length of time spent under democracy, we find evidence that political preferences are endogenous. For new democracies, our findings imply that popular support needs time to develop. For example, the effect of around 8.5 more years of democratic experience corresponds to the difference in support for democracy between primary and secondary education.

Popular support for democracy is critical to the success of a democracy, especially an emerging democracy (*1*, *2*). Will support increase over time when a democracy emerges and the population gains experience with democracy? If so, how quickly? Or are democratic attitudes deeply ingrained in individuals, such that they are hard to change? The latest wave of democratizations in the world, which started in December 2010 in a movement often collectively referred to as the "Arab Spring," and the subsequent struggles of these countries provide a recent illustration of the importance of these questions. However, a study that uses a clean identification strategy based on an experimental or quasi-experimental setup to identify the causal effect of accumulating experience with democracy on support for democracy in a broad set of countries—or more generally, a study that identifies endogenous preferences for political systems—is missing from the literature.

Indeed, recent research suggests that economic preferences are shaped by individual experiences with markets (*3*). In particular, preferences regarding fairness, preferences for redistribution, and other types of preferences related to economic behavior vary across societies in a way that correlates with market characteristics (*4*, *5*). A causal interpretation of these correlations and the view that economic preferences are endogenous is founded in theoretical arguments (*6*–*8*) and is empirically supported by research based on experimental or quasi-experimental settings, such as the end of communism in Eastern Europe or the stock market return experiences accumulated over a lifetime (*9*–*11*).

Regarding the endogeneity of political preferences, research has so far shown a positive correlation between experience with political systems and political preferences at the country level (*12*), a positive correlation between attitudes toward democracy and currently living under a democratic system (*13*), and that a longer democratic experience lowers the probability of exit from democracy and increases the probability of exit from autocracy (*12*). However, a causal influence of experience with democracy on the support for democracy, which would imply endogeneity of preferences, cannot be established from these correlations. The correlations could (partly) be due to reverse causality (i.e., countries have a democratic history precisely because the electorate supports democratic values); or a third, possibly unobserved, variable, such as historic events or economic conditions, could determine both individuals' support for democracy and the political system in place.

Here, we exploited within-country variation at the individual level in experience with a democratic regime to establish a plausibly causal impact of experience with democracy on preferences for democracy, and thereby contribute to a better understanding of the endogeneity of political preferences. Because we control for country-year fixed effects, the observed differences in attitudes toward democracy do not simply reflect a reaction to differences in the current quality of institutions or political environments, but, under the minimal and plausible identifying assumption that we state below, constitute a change in intrinsic preferences due to differences in the length of exposure to democracy. For example, if democratic institutions or economic conditions improve with the length of time spent under democracy, this might increase the support for democracy directly and not through intrinsic preferences, but it would be captured in our specification by the country-year fixed effects, which control for all country-level unobservables that are specific to a country in a given year. Any remaining correlation between experience with democracy and support for democracy can therefore confidently be attributed to a change in preferences.

Goethe University Frankfurt, 60320 Frankfurt, Germany.
*Both authors contributed equally to this work. **†Corresponding author. E-mail: fuchs@wiwi.uni-frankfurt.de (N.F.-S.); schuendeln@wiwi.uni-frankfurt.de (M.S.)**

5. L. J. Tranvik *et al.*, *Limnol. Oceanogr.* **54**, 2298–2314 (2009).
6. J. J. Elser *et al.*, *Ecol. Lett.* **10**, 1135–1142 (2007).
7. J. R. Webster *et al.*, *Freshw. Biol.* **41**, 687–705 (1999).
8. J. B. Wallace, S. L. Eggert, J. L. Meyer, J. R. Webster, *Science* **277**, 102–104 (1997).
9. D. A. Walther, M. R. Whiles, *J. N. Am. Benthol. Soc.* **30**, 357–373 (2011).
10. J. R. Webster *et al.*, *Verh. Int. Ver. Theor. Angew. Limnol.* **27**, 1337–1340 (2000).
11. K. Suberkropp, V. Gulis, A. D. Rosemond, J. P. Benstead, *Limnol. Oceanogr.* **55**, 149–160 (2010).
12. Materials and methods are available as supplementary materials on *Science* Online.
13. S. G. Fisher, G. E. Likens, *Ecol. Monogr.* **43**, 421–439 (1973).
14. M. O. Gessner, E. Chauvet, *Ecol. Appl.* **12**, 498–510 (2002).
15. J. P. Benstead *et al.*, *Ecology* **90**, 2556–2566 (2009).
16. J. L. Greenwood, A. D. Rosemond, J. B. Wallace, W. F. Cross, H. S. Weyers, *Oecologia* **151**, 637–649 (2007).
17. V. Gulis, K. Suberkropp, *Freshw. Biol.* **48**, 123–134 (2003).
18. C. J. Tant, A. D. Rosemond, M. R. First, *Freshw. Sci.* **32**, 1111–1121 (2013).
19. S. A. Thomas *et al.*, *Limnol. Oceanogr.* **46**, 1415–1424 (2001).
20. N. A. Griffiths *et al.*, *Ecol. Appl.* **19**, 133–142 (2009).
21. L. Boyero *et al.*, *Ecol. Lett.* **14**, 289–294 (2011).
22. J. B. Wallace, J. R. Webster, T. F. Cuffney, *Oecologia* **53**, 197–200 (1982).
23. J. B. Wallace, T. F. Cuffney, B. S. Goldowitz, K. Chung, G. J. Lugthart, *Verh. Int. Ver. Theor. Angew. Limnol.* **24**, 1676–1680 (1991).
24. V. Ferreira *et al.*, *Biol. Rev.* 10.1111/brv.12125 (2015).
25. J. S. Kominoski *et al.* (2015); available at www.esajournals.org/doi/abs/10.1890/14-1113.1.
26. D. J. Conley *et al.*, *Science* **323**, 1014–1015 (2009).
27. J. E. Allgeier, A. D. Rosemond, C. A. Layman, *J. Appl. Ecol.* **48**, 96–101 (2011).
28. J. R. Webster, *J. N. Am. Benthol. Soc.* **26**, 375–389 (2007).
29. R. B. Alexander, R. A. Smith, *Limnol. Oceanogr.* **51**, 639–654 (2006).
30. T. J. Battin *et al.*, *Nat. Geosci.* **1**, 95–100 (2008).
31. W. K. Dodds, *Trends Ecol. Evol.* **22**, 669–676 (2007).

ACKNOWLEDGMENTS

We thank H. Weyers, N. Taylor, R. Hilten, S. Dye, J. Coombs, and K. Norris for their assistance in maintaining the enrichments and associated data collection and analyses; C. Tant and J. Greenwood for conducting the N+P litterbag studies; and S. Eggert for assisting in data collection and analysis and helping develop sampling protocols. T. Mccallister illustrated Fig. 1 (photo credit: PMB). The manuscript was improved by comments from R. Hall, E. Rosi-Marshall, F. Ballantyne, S. Altizer, A. Helton, A. Huryn, M. Paul, J. Davis, R. Sponseller, C. Song, and anonymous reviewers. J. Maerz contributed ideas and logistical help associated with the N×P experiment; S. Wenger, R. Hall, C. Song, and D. Hall provided statistical advice; D. Leigh and J. Hepinstall-Cymerman provided spatial data; and J. Webster provided site information. We are grateful to A. Helton for conducting the network-scale extrapolation. Data are available in the supplementary materials in *Science* Online. This research leveraged logistical support from the Coweeta Long Term Ecological Research Program at the University of Georgia, which is supported by the National Science Foundation Division of Environmental Biology (NSF DEB grant 0823293). The order of authors after the first author is alphabetical; funding for these experiments was provided in NSF grants DEB-9806610, 0318063, 0918894, 0918904, and 0919054 from the Ecosystem Studies Program to A.D.R., J.P.B., V.G., K.S., J.B.W., and others (J. Maerz, above; M. Black, University of Georgia; and P. Mulholland, Oak Ridge National Laboratory).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/347/6226/1142/suppl/DC1
Materials and Methods
Supplementary Text
Tables S1 to S7
References (*32*–*50*)

7 November 2014; accepted 27 January 2015
10.1126/science.aaa1958

Table 1. Determinants of support for democracy. Question E117 asks whether "having a democratic political system" is "a very good, fairly good, fairly bad, or very bad way of governing this country." Question E123 asks whether the respondent agrees strongly, agrees, disagrees, or disagrees strongly with the statement "Democracy may have problems but it's better than any other form of government." Robust standard errors (in parentheses) are clustered at the country-year level. The omitted age category is older than 60 years; the omitted education category is no education. Columns 1 to 5 show coefficients from ordered probit estimations, column 6 from a probit estimation.

Determinant	Basis of support for democracy					
	World Values Survey					Afrobarometer
	IW index (2003) (1)	IW index (2003) (2)	IW index (2003) (3)	Question E117 (4)	Question E123 (5)	Bratton (2004) (6)
Country democratic at time of survey	0.339**	0.335**				
	(0.141)	(0.142)				
Country's democratic capital	0.063**	0.040				
	(0.030)	(0.030)				
Individual's democratic capital		0.021***	0.021***	0.018***	0.021***	0.021***
		(0.005)	(0.005)	(0.003)	(0.004)	(0.006)
Age 11–20	−0.162***	−0.066*	−0.053	−0.057**	−0.080**	−0.095***
	(0.044)	(0.036)	(0.035)	(0.024)	(0.040)	(0.029)
Age 21–30	−0.101***	−0.023	−0.011	−0.090***	−0.063*	−0.044*
	(0.039)	(0.032)	(0.032)	(0.020)	(0.035)	(0.024)
Age 31–40	−0.041	0.007	0.014	−0.069***	−0.047	0.049**
	(0.031)	(0.026)	(0.026)	(0.017)	(0.030)	(0.022)
Age 41–50	0.001	0.023	0.031	−0.039***	−0.022	0.078***
	(0.025)	(0.023)	(0.023)	(0.015)	(0.027)	(0.021)
Age 51–60	0.038**	0.048***	0.051***	−0.026**	−0.001	0.089***
	(0.019)	(0.018)	(0.018)	(0.012)	(0.021)	(0.020)
Male	0.049***	0.050***	0.050***	0.063***	0.042***	0.194***
	(0.011)	(0.011)	(0.011)	(0.008)	(0.012)	(0.015)
Primary education	0.073**	0.067**	0.067**	0.029*	0.011	0.215***
	(0.033)	(0.033)	(0.034)	(0.017)	(0.031)	(0.022)
Secondary education	0.250***	0.244***	0.233***	0.162***	0.098**	0.448***
	(0.043)	(0.043)	(0.043)	(0.022)	(0.042)	(0.036)
Postsecondary education	0.529***	0.523***	0.518***	0.374***	0.275***	0.562***
	(0.053)	(0.052)	(0.051)	(0.029)	(0.051)	(0.045)
Country fixed effects	Yes	Yes				
Year fixed effects	Yes	Yes				
Country-year fixed effects			Yes	Yes	Yes	Yes
Observations	82,990	82,990	82,990	228,901	92,565	149,035
Number of countries	56	56	56	79	57	31
Survey waves (WVS)	3–5	3–5	3–5	3–6	3–5	
Rounds (Afrobarometer)						1–5
Years covered	1994–2006	1994–2006	1994–2006	1994–2013	1994–2006	1999–2013

*P < 0.1, **P < 0.05, ***P < 0.01.

1

Before we can even think about running analyses, we need to think **theoretically** about the myriad possible relationships between the outcome we are trying to explain (Y) and the factors (X's) that could affect it.

Ask yourself the following questions:

> Is there a **credible mechanism** connecting X to Y?
>
> Is there a real risk of **endogeneity**?
>
> Is there significant **covariation** between X and Y to explain?
>
> Have we thought about potential **spurious** factors (Z's)?

2

2

Previously discussed issues

Issue	Example
Links between concepts and proxy measurements	Democracy and Polity IV
Raw numbers vs. ratio variables	GDP & GDP per capita
Raw numbers vs. percentages	Natural resource rents & rents as a % of government expenditure
Raw numbers vs. indices	Trafficking victims vs. trafficking index
Mean vs. median vs. mode	Individual salaries
Levels of analysis	System/country/province/city/individual

2

Perfect multicollinearity definition: "when there is an exact linear relationship between any two or more of a regression model's independent variables." (Kellstedt and Whitten 2018: 243)

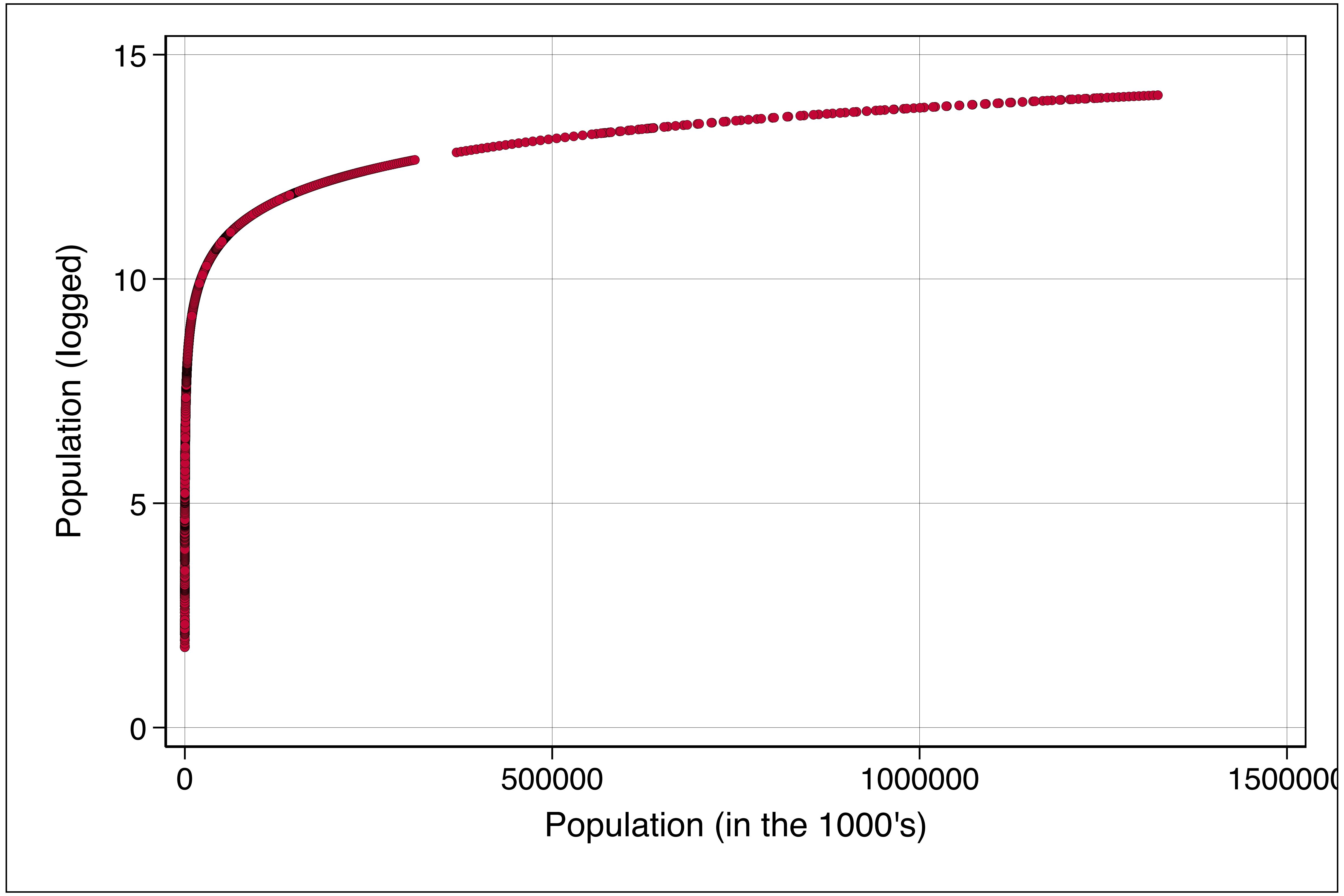
Multicollinearity is "usually the result of a small number of cases relative to the number of parameters we are estimating, limited independent variable values, or model mis-specification." (Kellstedt and Whitten 2018: 246)

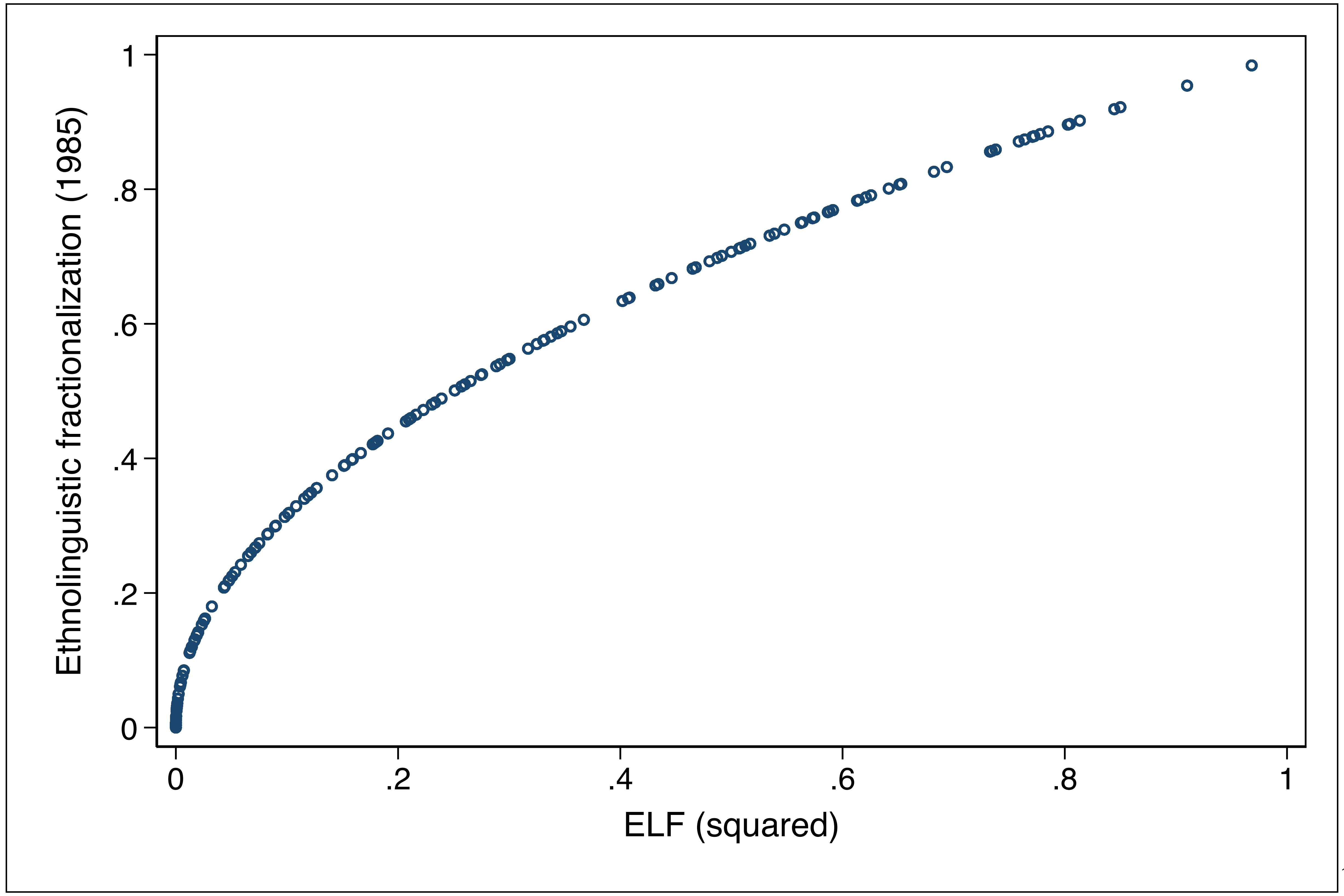
If there are two variables that are perfectly multi-collinear, one will be dropped.

Think theoretically if both variables are capturing the **same underlying trait** of the sample you are using.

2

Scholars often **transform** their variables for theoretical or practical reasons. **Why**?

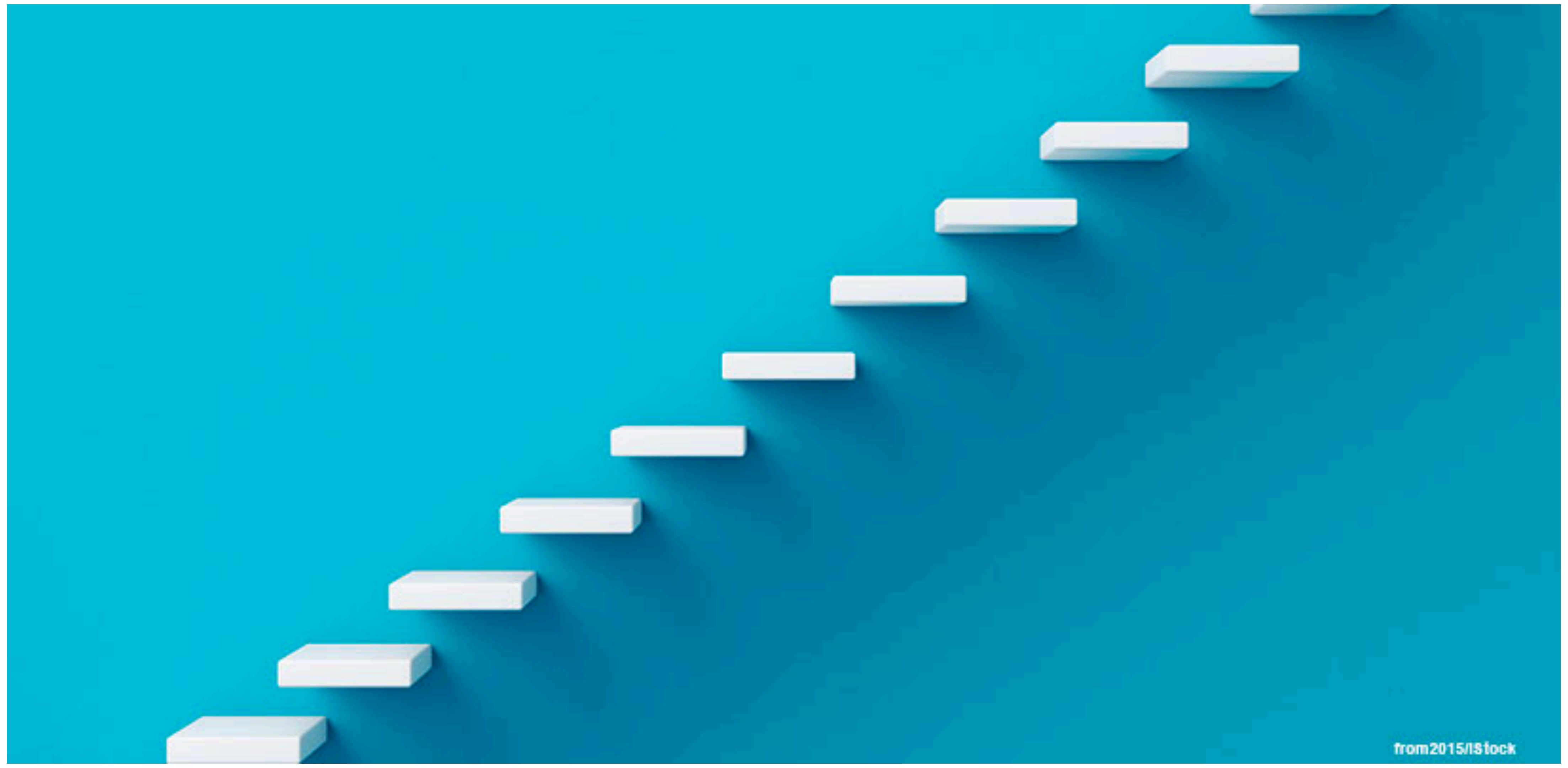




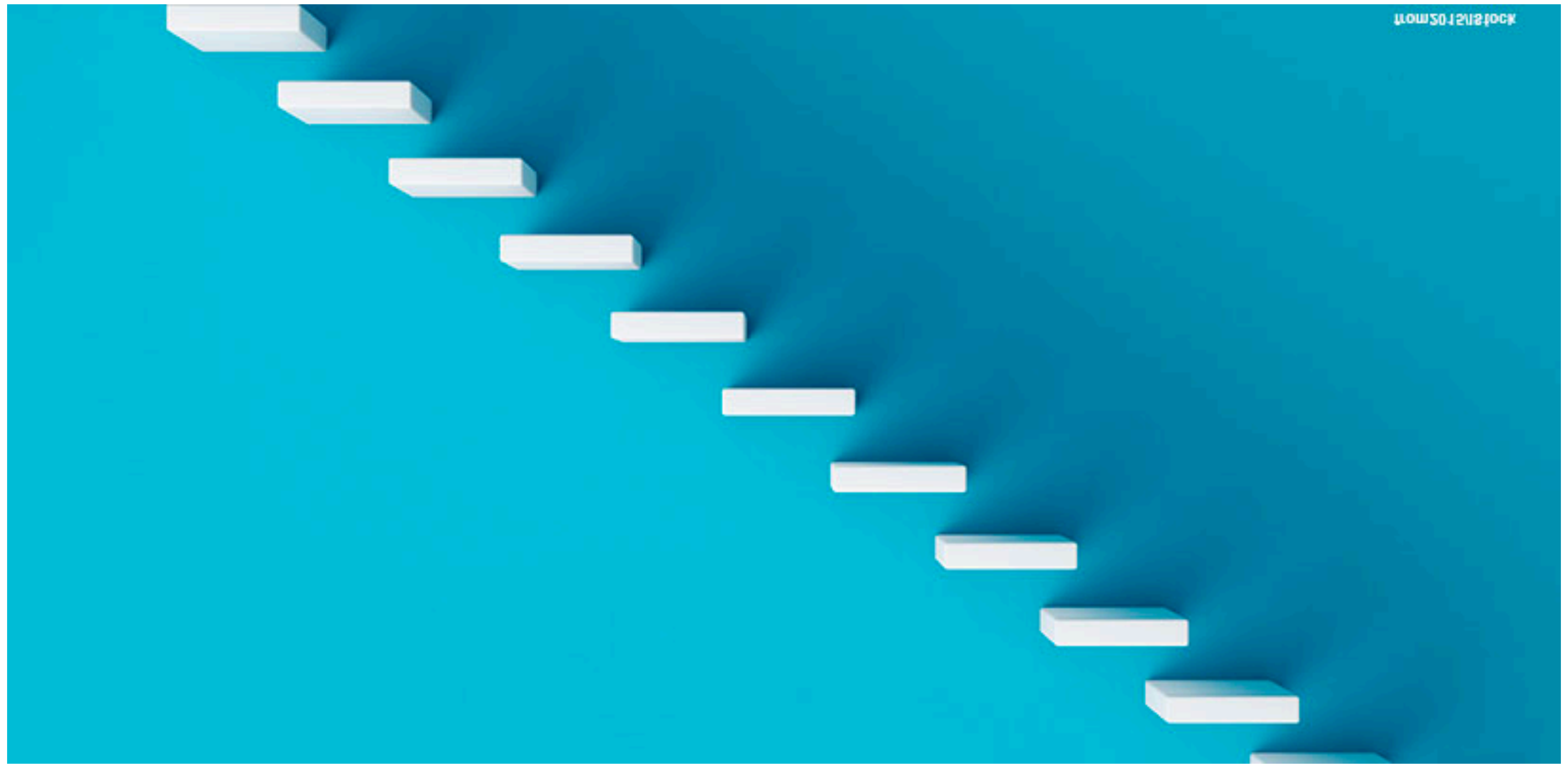
2

A regression approach in which you automatically specify a final model through trial and error of **adding** or **subtracting** independent variables according to some model fit criterion.

Forward selection



Backward elimination



2

Stepwise regression can lead to **overfitting**.

It will explain the current data but may not do well with **new data**.

It can inflate **accuracy** estimates and **statistical significance**.

2

If we include 20 variables in a model, then **on average** we will find one statistically significant relationship.

Most variables include **missing data**. The more variables you include, the smaller your sample becomes.

Some variables may do well with **prediction** but have only tenuous theoretical links.

Humans can only conceptualise a **small number of moving parts** at the same time.

2

Conflict Management and Peace Science, 22:327–339, 2005

ISSN: 0738-8942 print / 1549-9219 online
DOI: 10.1080/07388940500339167

Routledge
Taylor & Francis Group

Let's Put Garbage-Can Regressions and Garbage-Can Probits Where They Belong

CHRISTOPHER H. ACHEN

Department of Politics
Princeton University
Princeton, New Jersey, USA

Many social scientists believe that dumping long lists of explanatory variables into linear regression, probit, logit, and other statistical equations will successfully "control" for the effects of auxiliary factors. Encouraged by convenient software and ever more powerful computing, researchers also believe that this conventional approach gives the true explanatory variables the best chance to emerge. The present paper argues that these beliefs are false, and that without intensive data analysis, linear regression models are likely to be inaccurate. Instead, a quite different and less mechanical research methodology is needed, one that integrates contemporary powerful statistical methods with deep substantive knowledge and classic data–analytic techniques of creative engagement with the data.

Keywords regression analysis, linearity, data analysis, rule of three, monotonicity

Sometimes you can see a lot just by looking.
—attributed to former New York Yankees catcher Yogi Berra

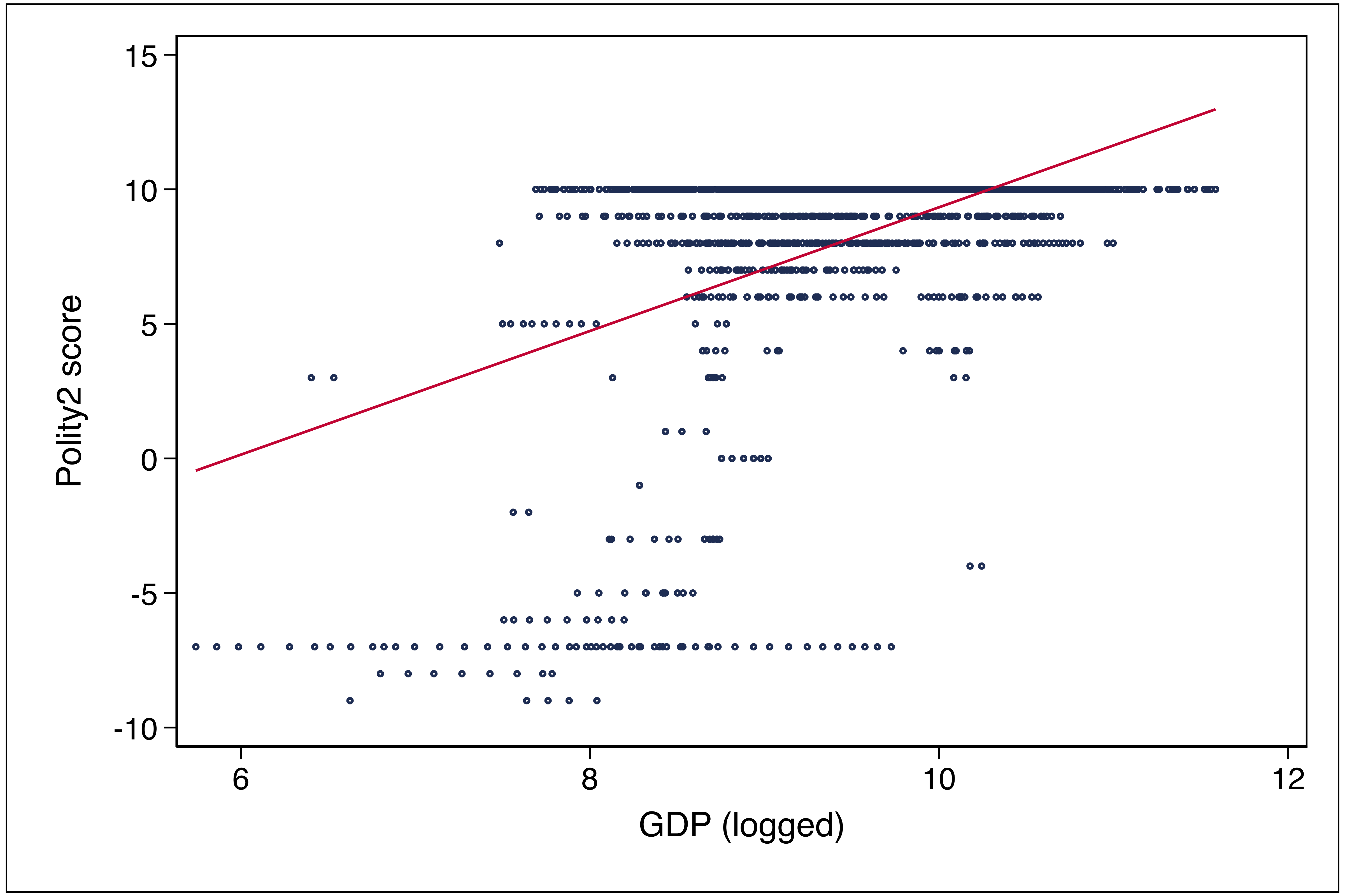
Political researchers have long dreamed of a scientifically respectable theory of international politics. International peace and justice are painfully difficult to achieve, and some of the obstacles have an intellectual character. We do not understand what we most need to know.

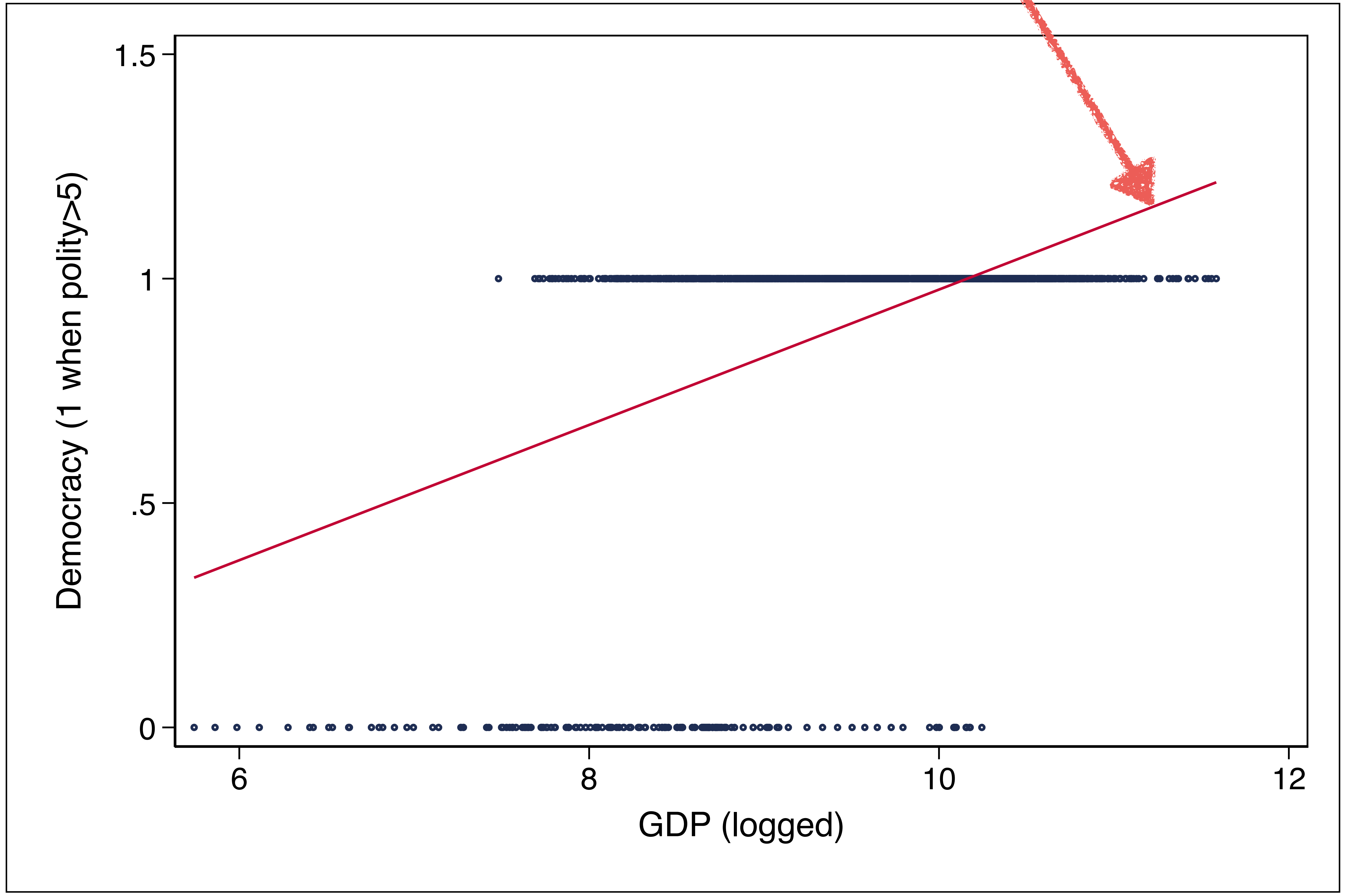
In this quest, humanistic, interpretive, and historical methodologies have been profoundly valuable for more than two millennia. They have taught us most of what we know about international politics, and without question we will need their continuing insights for additional progress. Yet these traditional approaches encounter conceptual knots in

In sum, we need to abandon mechanical rules and procedures. "Throw in every possible variable" won't work; neither will "rigidly adhere to just three explanatory variables and don't worry about anything else." Instead, the research habits of the profession need greater emphasis on classic skills that generated so much of what we know in quantitative social science: plots, crosstabs, and just plain looking at data. Those methods are simple, but sophisticatedly simple. They often expose failures in the assumptions of the elaborate statistical tools we are using, and thus save us from inferential errors. Doing that kind

2

Is it bad that this is above democracy=1?





2

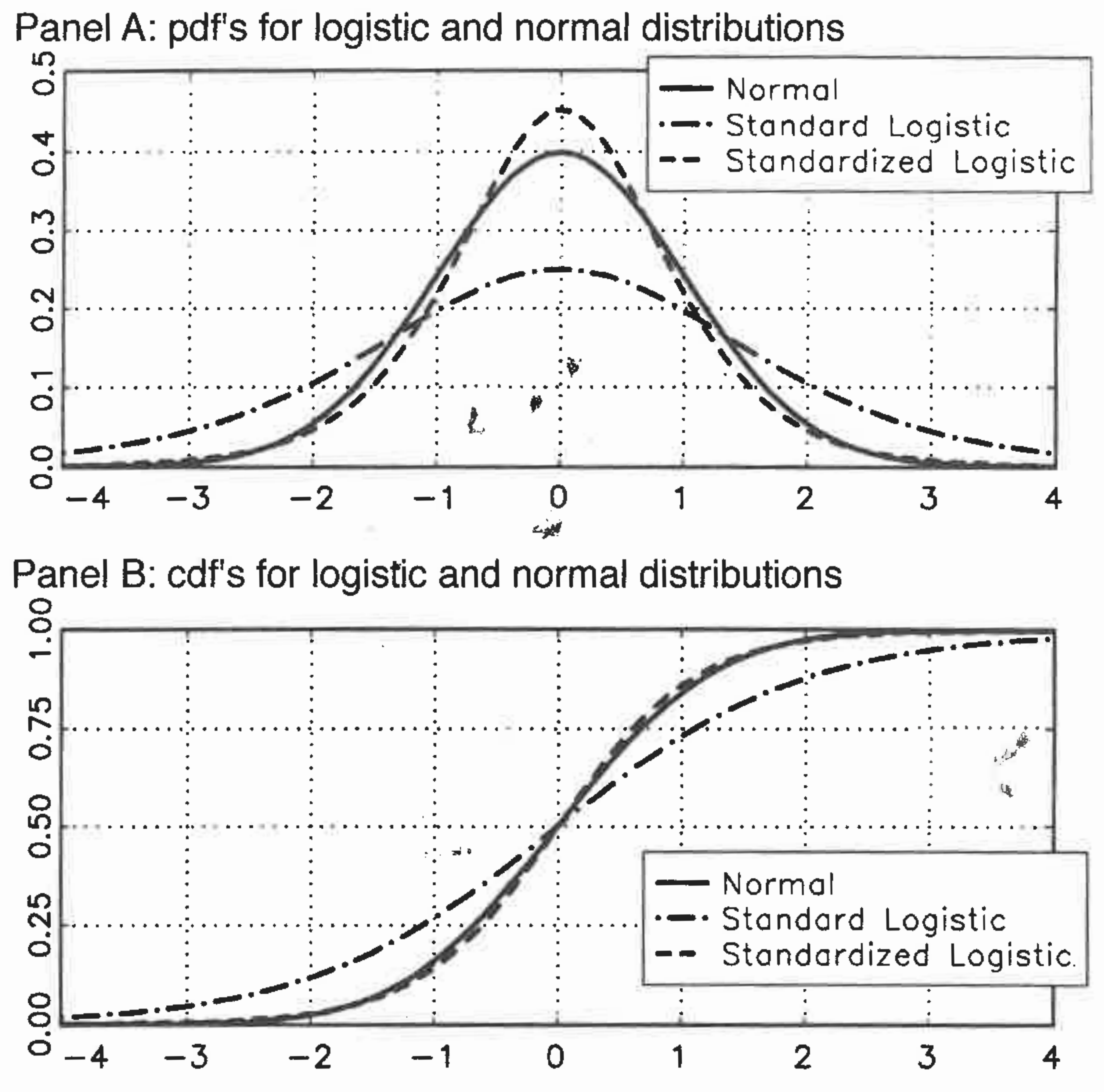


Figure 3.3. Normal and Logistic Distributions

Source: Long (1997: 43, 63)

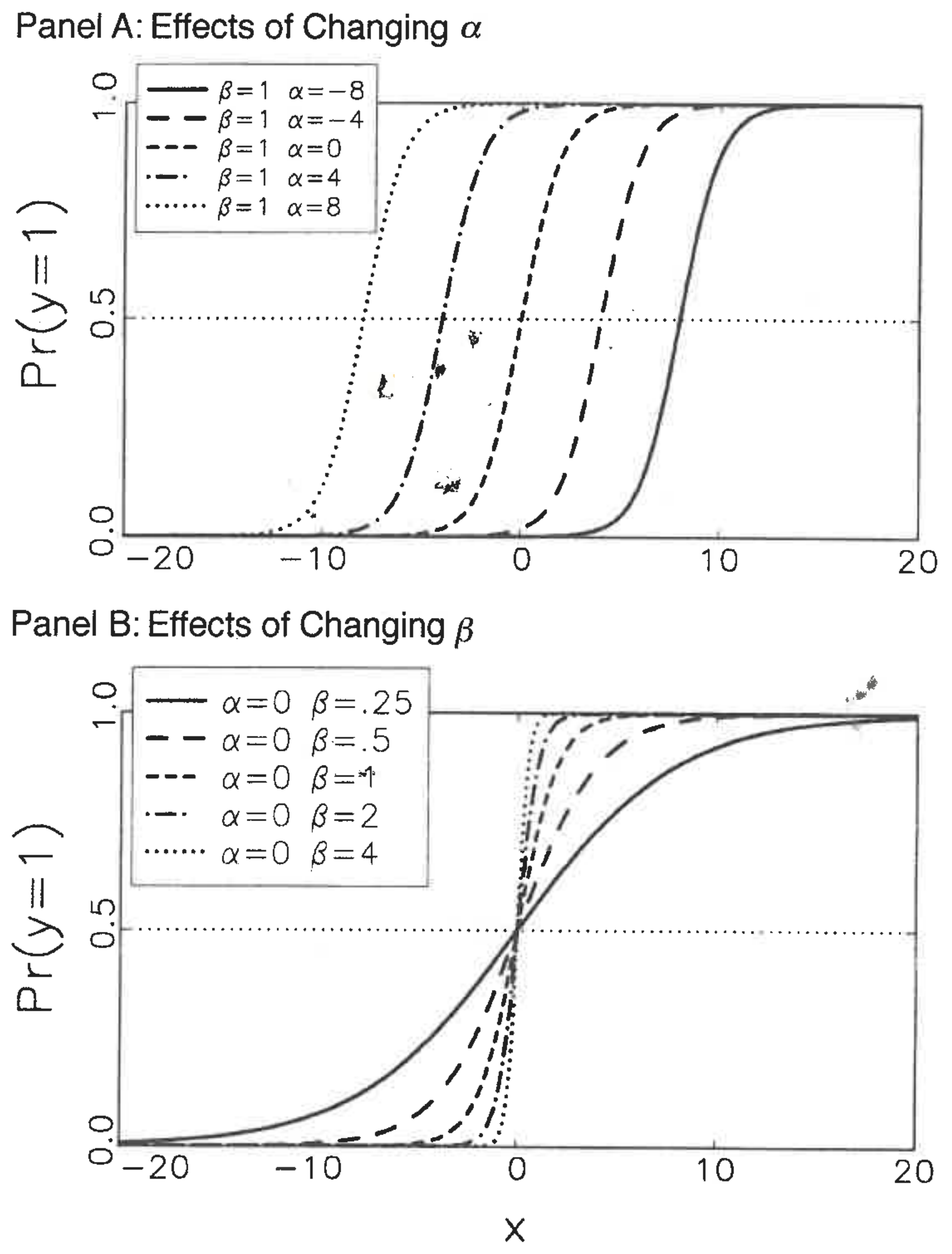


Figure 3.8. Effects of Changing the Slope and Intercept on the Binary Response Model: $\Pr(y=1\,|\,x) = F(\alpha + \beta x)$

2

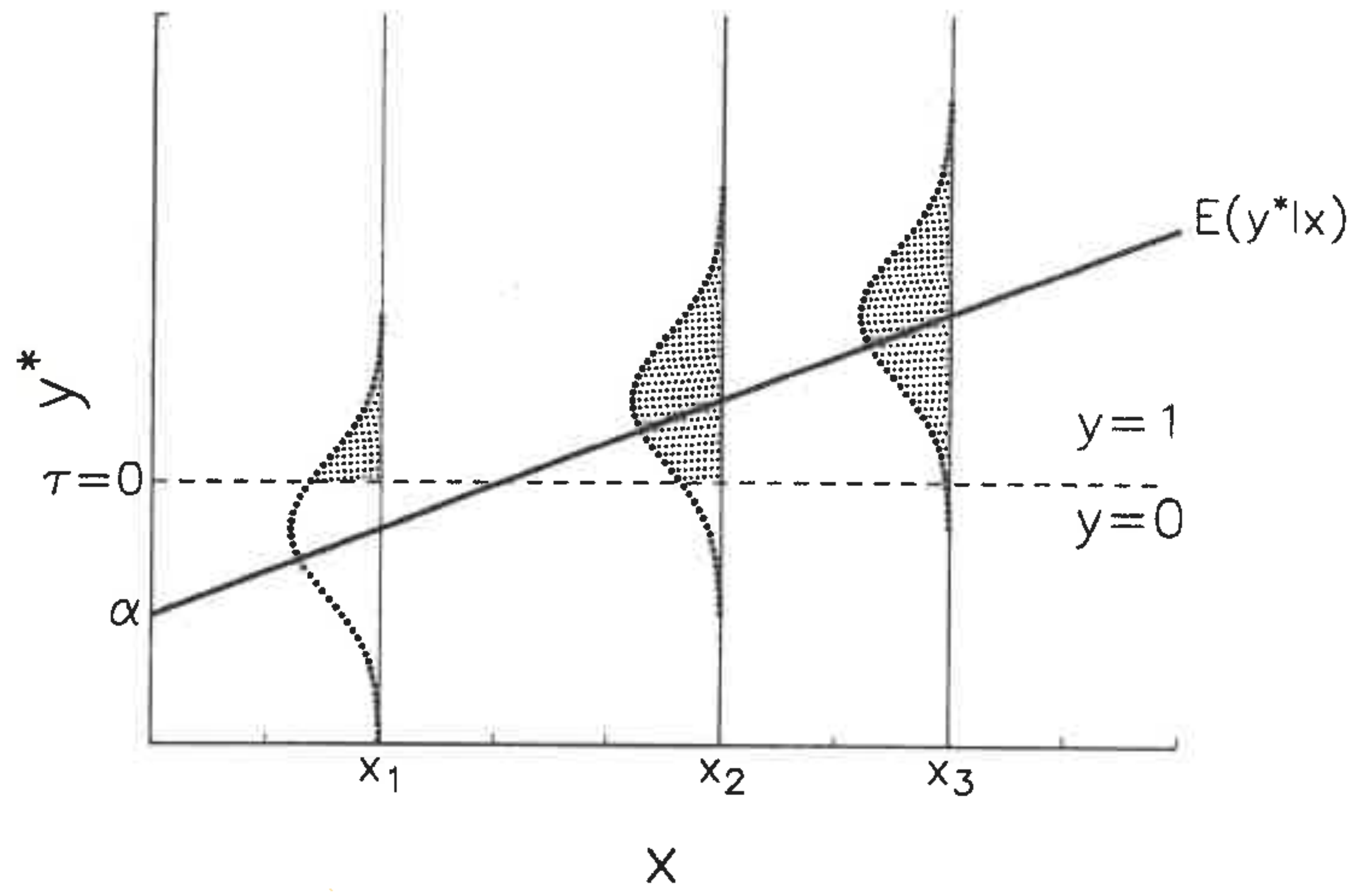


Figure 3.2. The Distribution of y^* Given x in the Binary Response Model

$$y^* = \mathbf{X}\beta + u$$

Where

$$y_i = \begin{cases} 1, & if \;\; y_i^* > \kappa \\ 0, & if \;\; y_i^* \leq \kappa \end{cases}$$

Source: Long (1997: 41, 63)

2

Model	Maximum likelihood function
Logit	$\text{Ln}\, L = \sum_{i=1}^{n} y_i \ln\left(\frac{1}{1+e^{(X\beta)}}\right) + \ (1\text{-}\ y_i)\ \ln\left(1\ -\ \frac{1}{1+e^{(X\beta)}}\right)$
Probit	$\text{Ln}\, L = \sum_{i=1}^{n} y_i \ln\Phi(X\beta) + \ (1\text{-}\ y_i)\ln\Phi(X\beta)$

2

Scholars engage in a daily balancing act when deciding:

Which variables to include

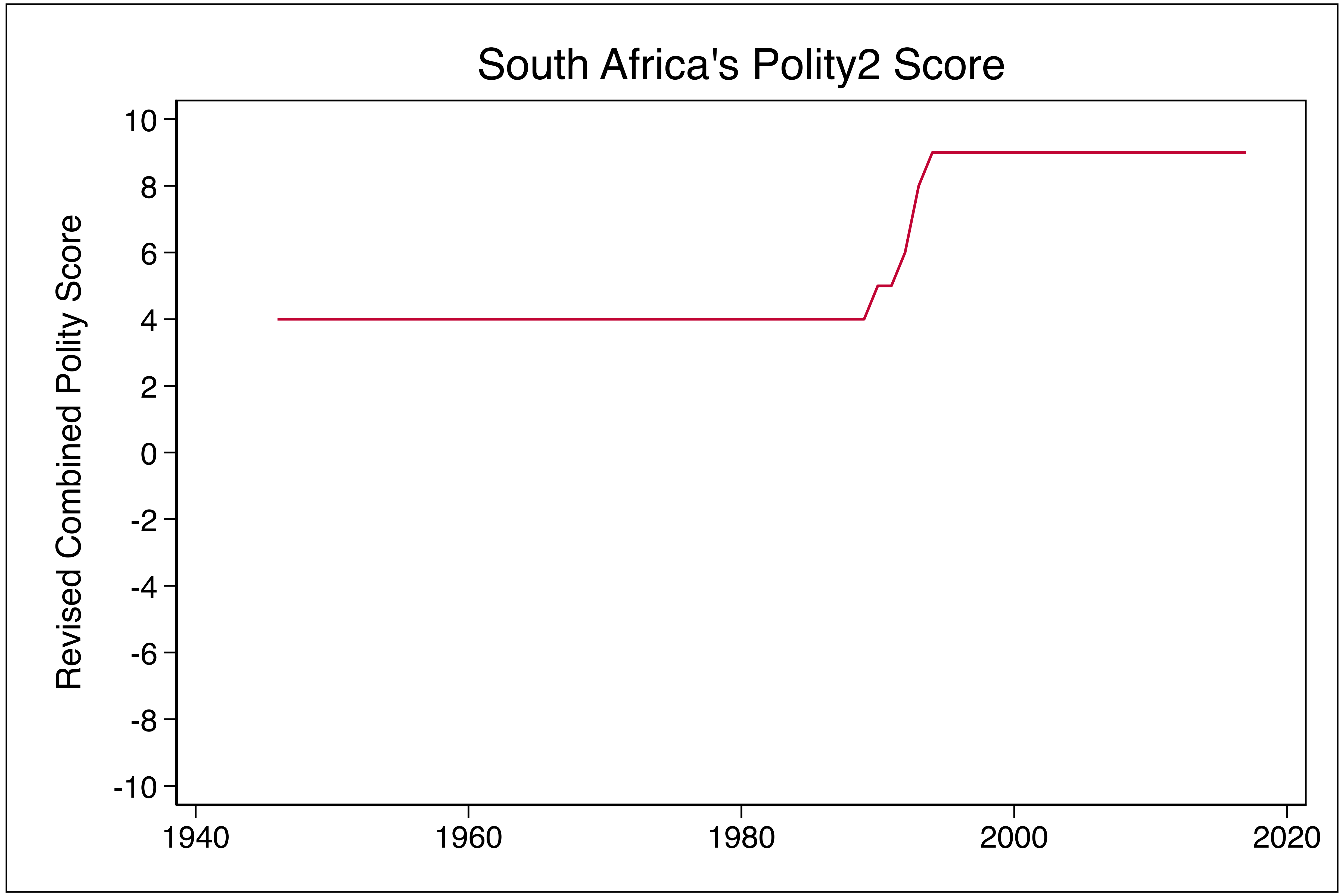
In **what** form should we include them

How to estimate our models

And **which** model is appropriate for the distribution of our Y.

3

3



Data source: Center for Systemic Peace (https://www.systemicpeace.org/polityproject.html)

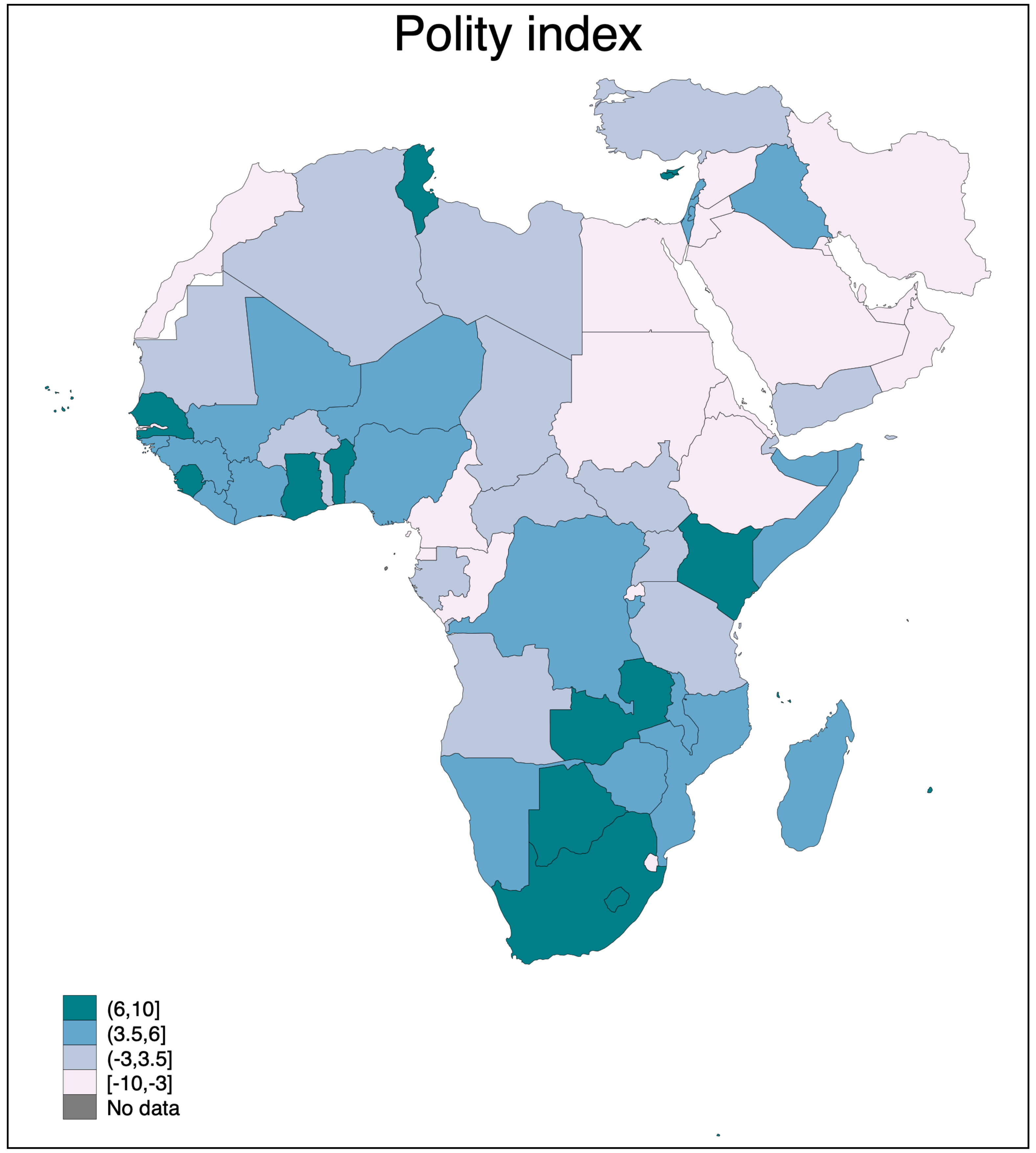
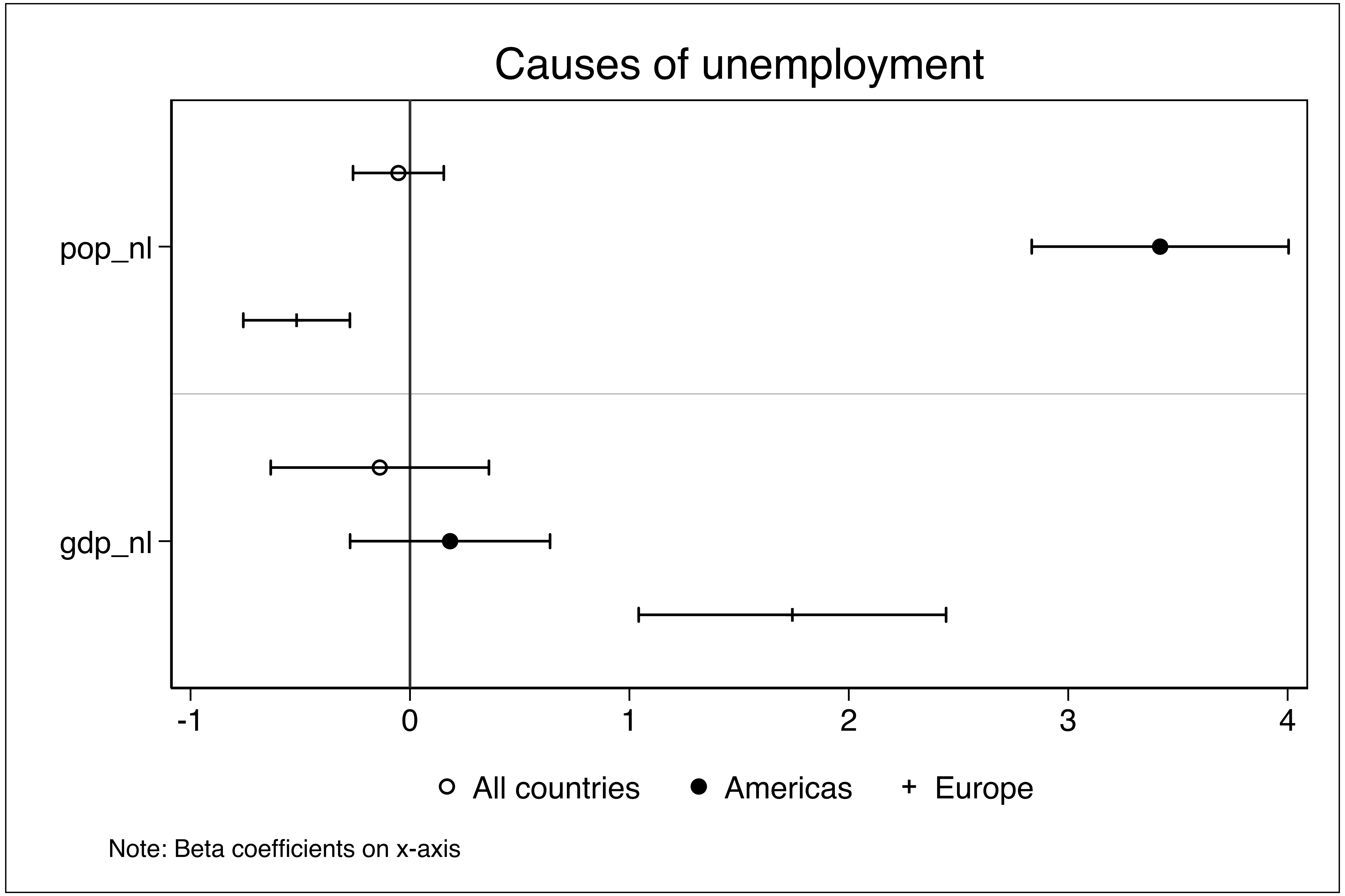


Image source: https://www.stathelp.se/en/spmap_world_en.html

3

It appears that there is an “apparent trend in the data that can be eliminated or reversed by splitting the data into natural groups.”

(Reinhart 2015:4)



3

A way to evaluate regressions is to run them a number of times, each time leaving out a different observation and using the results to predict this observation (leave-one-out cross-validation).

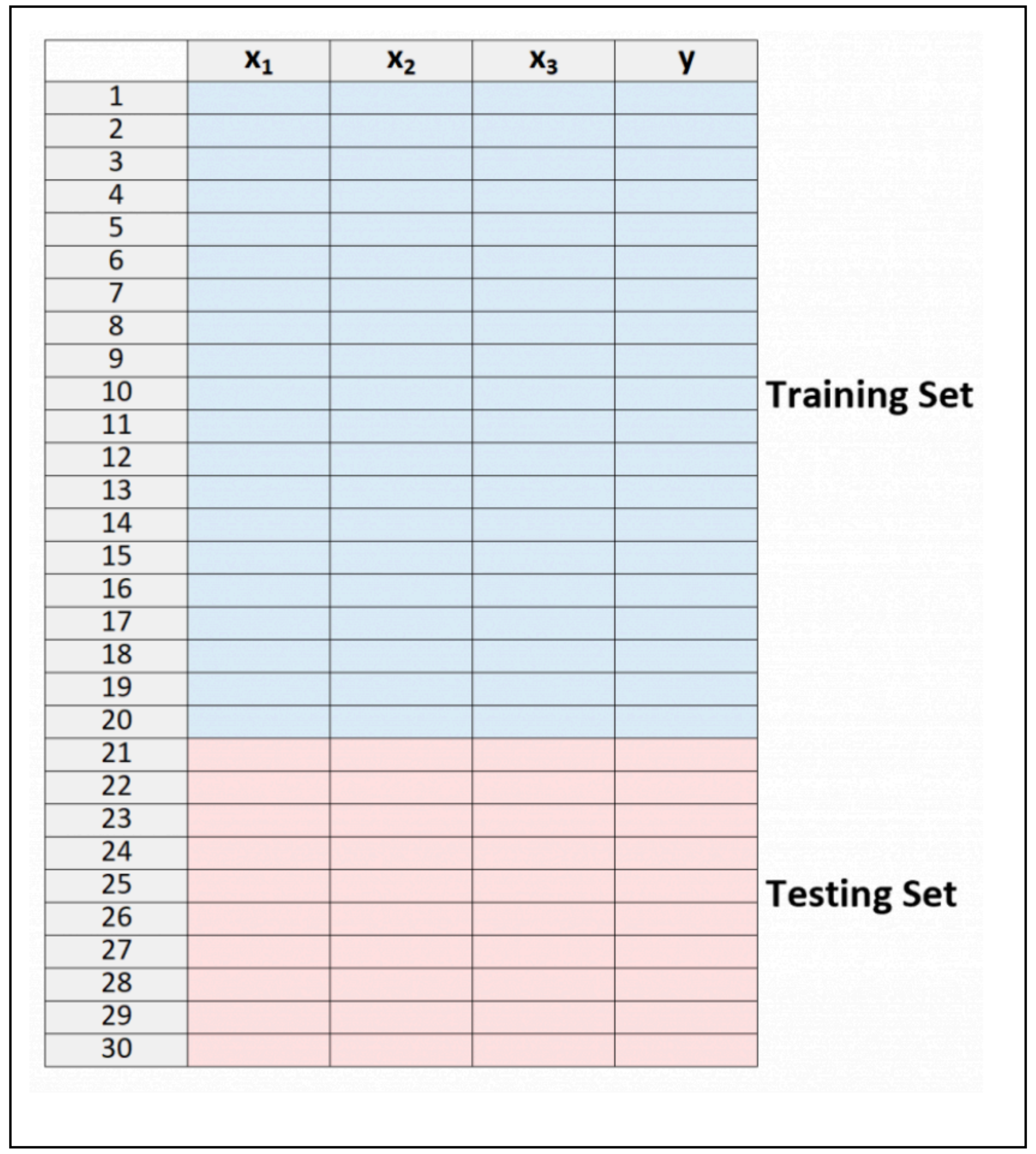
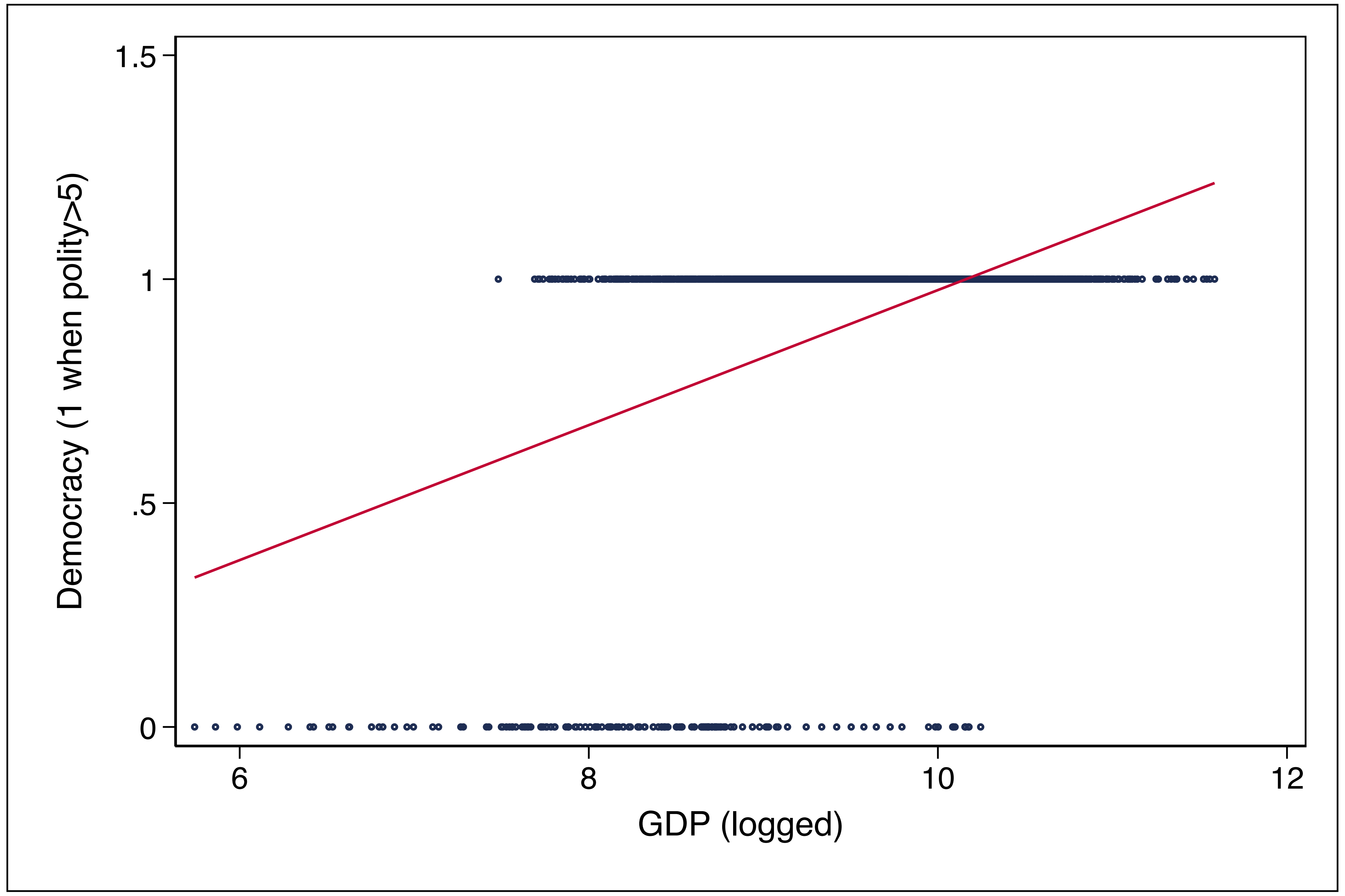


Image source: https://www.statology.org/leave-one-out-cross-validation/

```
.          reg    oecd_unemplrt_t1b  pop_nl gdp_nl

      Source |       SS           df       MS      Number of obs   =       688
-------------+----------------------------------   F(2, 685)       =      0.23
       Model |  7.71111544         2  3.85555772   Prob > F        =    0.7930
    Residual |  11381.2027       685  16.6148945   R-squared       =    0.0007
-------------+----------------------------------   Adj R-squared   =   -0.0022
       Total |  11388.9138       687  16.5777494   Root MSE        =    4.0761

------------------------------------------------------------------------------
oecd_unem~1b | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      pop_nl |  -.0526766   .1053346    -0.50   0.617    -.2594941    .1541409
      gdp_nl |  -.1374431   .2532246    -0.54   0.587    -.6346326    .3597464
       _cons |   8.764164   2.875336     3.05   0.002     3.118634    14.40969
------------------------------------------------------------------------------

.           loocv reg    oecd_unemplrt_t1b  pop_nl gdp_nl


 Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |    4.0865116
Mean Absolute Errors     |    2.9096928
Pseudo-R2                |    .02179182
-----------------------------------------
```

3

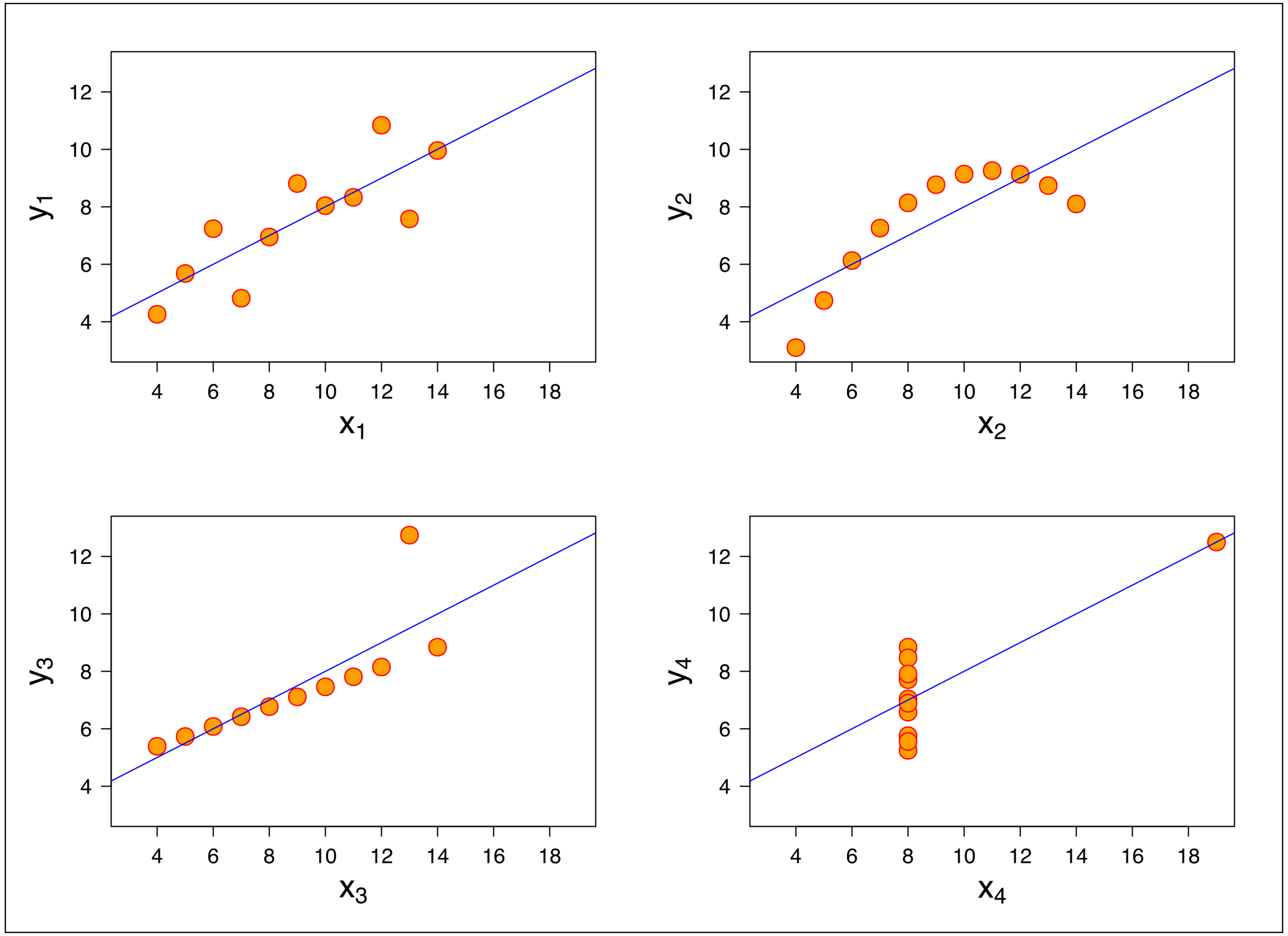
Along a similar vein to Simpson's paradox is the danger of thinking your results apply to a population that may or not be similar to the sample you used.



3

Assuming **linearity** can either lead to null results or understating true relationship (Type 2 errors).

Anscombe's quartet —>



Source: https://en.wikipedia.org/wiki/Anscombe%27s_quartet#/media/File:Anscombe's_quartet_3.svg

3

Table 5
Dyad development, regional dummies, and conflict: multivariate analysis, all dyads, 1880–2001

Variable	Model 9		Model 10		Model 11	
	Parameter estimate	Standard error	Parameter estimate	Standard error	Parameter estimate	Standard error
Shared basin	0.69**	0.28	0.62*	0.25	0.64*	0.26
Dyad development					0.059	0.044
Dyad development squared					−0.028	0.035
Dyad development*shared basin					−0.15***	0.051
Dyad development sq. *shared basin					−0.036	0.047
Middle East and North Africa	0.37	0.20				
MENA*shared basin	0.19	0.28				
Sub-Saharan Africa			−0.76*	0.38		
Sub-Saharan Africa*shared basin			0.63	0.38		
N	107,584		107,584		107,584	
Pseudo-R^2	0.38		0.38		0.38	

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

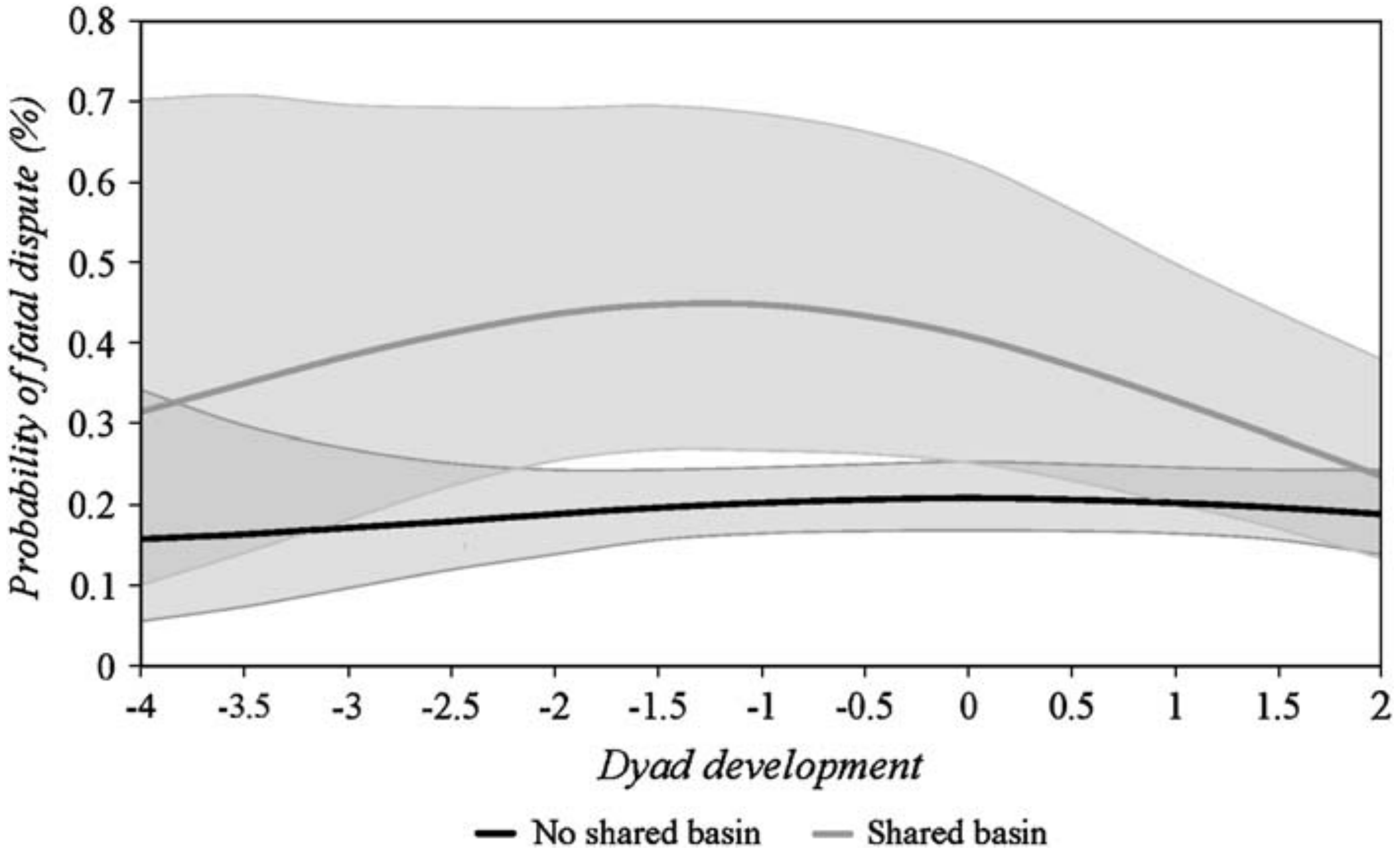


Fig. 3. Estimated probability of fatal dispute by shared basin and dyad development, 1880–2001. The shaded area around each line represents a 90% confidence interval.

Source: Gleditsch, Nils Petter, Kathryn Furlong, Håvard Hegre, Bethany Lacina, and Taylor Owen. 2006. "Conflicts Over Shared Rivers: Resource Scarcity or Fuzzy Boundaries?" *Political Geography* 25: 361-382.

3

It is easy to get results either **counter** to your expectations or **null effects** if your theories are not well matched to your data sample.

Think about whether your theory is more about change **within** units (e.g. countries or people) over time or **between** units.

Think about whether the relationship is linear or **non-linear**.

Make sure to evaluate the **robustness** of your findings.

4


mixpanel
UNDERSTAND EVERY USER'S JOURNEY WITH INSTANT INSIGHTS
PRODUCT SCHOOL
SILICON VALLEY
Product School Guest
Value.

4

Political Analysis, 9:4

Testing for Publication Bias in Political Science

Alan S. Gerber, Donald P. Green, and David Nickerson
Department of Political Science, Yale University,
New Haven, CT 06520-8301
e-mail: alan.gerber@yale.edu

If the publication decisions of journals are a function of the statistical significance of research findings, the published literature may suffer from "publication bias." This paper describes a method for detecting publication bias. We point out that to achieve statistical significance, the effect size must be larger in small samples. If publications tend to be biased against statistically insignificant results, we should observe that the effect size diminishes as sample sizes increase. This proposition is tested and confirmed using the experimental literature on voter mobilization.

1 Introduction

THE DEARTH OF insignificant findings in journals reflects the behavior of both researchers and journal editors. Editors and referees look askance at papers that report insignificant findings (Mahoney 1977), and their reputation for doing so creates the "file-drawer problem"—researchers elect not to submit their findings when their research fails to reject the null hypothesis (Iyengar and Greenhouse 1988; Greenwald 1975).

If articles that do not reject the null hypothesis tend to go unpublished, surveys of published research will create a distorted impression about effect size. To achieve statistical significance, studies with a small sample size require larger estimated effects than those with large samples. Publication bias against statistically insignificant results is therefore a directly testable proposition. One can detect the presence of publication bias by plotting the size of the estimated effect by the sample size (Begg 1985, 1994). For one-tailed tests, the smaller the sample size, the larger the published effect size (Light and Pillemer 1984).

Does publication bias inhabit political science? Although the phenomenon has been well documented in other fields such as psychology (e.g., Coursol and Wagner 1986), medical sciences (e.g., Simes 1986; Begg and Berlin 1988; Dickersin 1990), and economics (e.g., DeLong and Lang 1992), the only extended discussion of publication bias in political science is by Lee Sigelman (1999, p. 206) who argues that small sample size is symptomatic of poor methodology. According to this explanation, what may appear to be bias toward statistical significance may instead be an innocuous process whereby methodologically

Authors' note: We are grateful for the useful comments from the three anonymous referees. We are also grateful to the Smith Richardson Foundation and the Institution for Social and Policy Studies at Yale, which helped fund this research, but bear no responsibility for its content.

Quarterly Journal of Political Science, 2008, 3: 313–326

Research Note

Do Statistical Reporting Standards Affect What Is Published? Publication Bias in Two Leading Political Science Journals

Alan Gerber[1] and Neil Malhotra[2]

[1]*ISPS, Yale University, 77 Prospect Street, New Haven, CT 06520, USA; alan.gerber@yale.edu*
[2]*Graduate School of Business, Stanford University, Stanford, CA 94305-5015, USA; neilm@stanford.edu*

ABSTRACT

We examine the *APSR* and the *AJPS* for the presence of publication bias due to reliance on the 0.05 significance level. Our analysis employs a broad interpretation of publication bias, which we define as the outcome that occurs when, for whatever reason, publication practices lead to bias in the published parameter estimates. We examine the effect of the 0.05 significance level on the pattern of published findings using a "caliper" test, a novel method for comparing studies with heterogeneous effects, and find that we can reject the hypothesis of no publication bias at the 1 in 32 billion level. Our findings therefore raise the possibility that the results reported in the leading political science journals may be misleading due to publication bias. We also discuss some of the reasons for publication bias and propose reforms to reduce its impact on research.

A key objective of political science research is the accurate measurement of causal effects. Methodological advances (such as increased use of natural, laboratory, and field experiments) have made it much more plausible than in earlier decades that the results of individual studies are unbiased estimates. Unfortunately, better research design does not ensure unbiased literatures. For instance, if some results are more likely to be published, then literatures will be biased even if each study is done well.

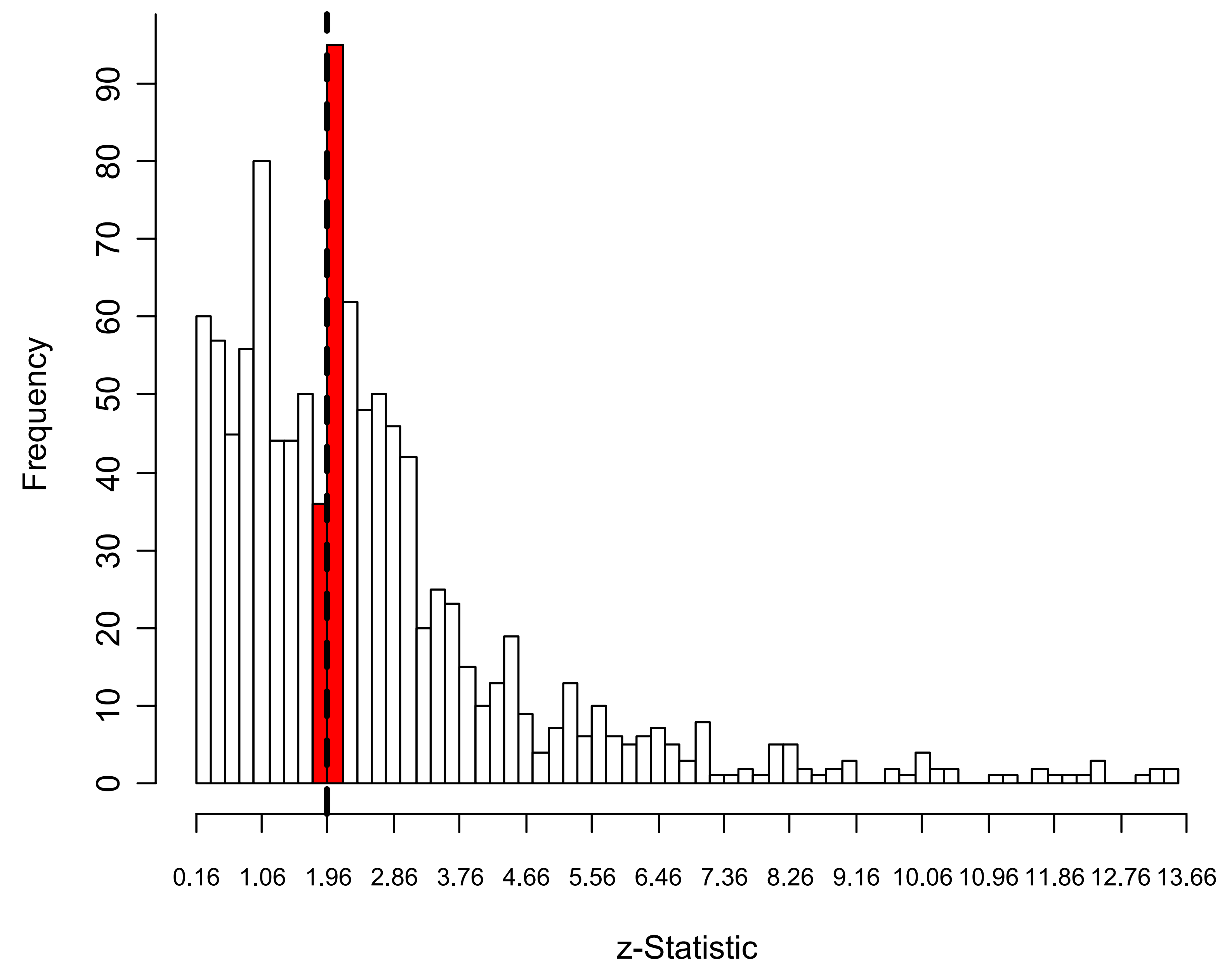


Figure 1(a). Histogram of z-statistics, *APSR & AJPS* (Two-Tailed). Width of bars (0.20) approximately represents 10% caliper. Dotted line represents critical z-statistic (1.96) associated with $p = 0.05$ significance level for one-tailed tests.

4

PNAS RESEARCH ARTICLE | SOCIAL SCIENCES OPEN ACCESS

Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty

Edited by Douglas Massey, Princeton University, Princeton, NJ; received March 6, 2022; accepted August 22, 2022

This study explores how researchers' analytical choices affect the reliability of scientific findings. Most discussions of reliability problems in science focus on systematic biases. We broaden the lens to emphasize the idiosyncrasy of conscious and unconscious decisions that researchers make during data analysis. We coordinated 161 researchers in 73 research teams and observed their research decisions as they used the same data to independently test the same prominent social science hypothesis: that greater immigration reduces support for social policies among the public. In this typical case of social science research, research teams reported both widely diverging numerical findings and substantive conclusions despite identical start conditions. Researchers' expertise, prior beliefs, and expectations barely predict the wide variation in research outcomes. More than 95% of the total variance in numerical results remains unexplained even after qualitative coding of all identifiable decisions in each team's workflow. This reveals a universe of uncertainty that remains hidden when considering a single study in isolation. The idiosyncratic nature of how researchers' results and conclusions varied is a previously underappreciated explanation for why many scientific hypotheses remain contested. These results call for greater epistemic humility and clarity in reporting scientific findings.

metascience | many analysts | researcher degrees of freedom | analytical flexibility | immigration and policy preferences

Significance

Will different researchers converge on similar findings when analyzing the same data? Seventy-three independent research teams used identical cross-country survey data to test a prominent social science hypothesis: that more immigration will reduce public support for government provision of social policies. Instead of convergence, teams' results varied greatly, ranging from large negative to large positive effects of immigration on social policy support. The choices made by the research teams in designing their statistical tests explain very little of this variation; a hidden universe of uncertainty remains. Considering this variation, scientists, especially those working with the complexities of human societies and behavior, should exercise humility and strive to better account for the uncertainty in their work.

Organized scientific knowledge production involves institutionalized checks, such as editorial vetting, peer review, and methodological standards, to ensure that findings are independent of the characteristics or predispositions of any single researcher (1, 2). These procedures should generate interresearcher reliability, offering consumers of scientific findings assurance that they are not arbitrary flukes and that other researchers would generate similar findings given the same data. Recent metascience research challenges this assumption as several attempts to reproduce findings from previous studies failed (3, 4).

In response, scientists have discussed various threats to the reliability of the scientific process with a focus on biases inherent in the production of science. Pointing to both misaligned structural incentives and the cognitive tendencies of researchers (5–7), this bias-focused perspective argues that systematic distortions of the research process push the published literature away from truth seeking and accurate observation. This then reduces the probability that a carefully executed replication will arrive at the same findings.

Here, we argue that some roots of reliability issues in science run deeper than systematically distorted research practices. We propose that to better understand why research is often nonreplicable or lacking interresearcher reliability, we need to account for idiosyncratic variation inherent in the scientific process. Our main argument is that variability in research outcomes between researchers can occur even under rigid adherence to the scientific method, high ethical standards, and state-of-the-art approaches to maximizing reproducibility. As we report below, even well-meaning scientists provided with identical data and freed from pressures to distort results may not reliably converge in their findings because of the complexity and ambiguity inherent to the process of scientific analysis.

Variability in Research Outcomes

The scientific process confronts researchers with a multiplicity of seemingly minor, yet nontrivial, decision points, each of which may introduce variability in research outcomes. An important but underappreciated fact is that this even holds for what is often seen as the most objective step in the research process: working with the data after it has come in. Researchers can take literally millions of different paths in wrangling, analyzing, presenting, and interpreting their data. The number of choices grows exponentially with the number of cases and variables included (8–10).

A bias-focused perspective implicitly assumes that reducing "perverse" incentives to generate surprising and sleek results would instead lead researchers to generate valid

The authors declare no competing interest.

This article is a PNAS Direct Submission.

See online for related content such as Commentaries.

[1]To whom correspondence may be addressed. Email: breznau.nate@gmail.com.

[2]N.B., E.M.R., and A.W. were the Principal Investigators, equally responsible for conceptualization and data collection. Primary meta-analysis of data analysts' results and preparation of metadata for public consumption preformed by N.B., with assistance from H.H.V.N.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2203150119/-/DCSupplemental.

Published October 28, 2022.

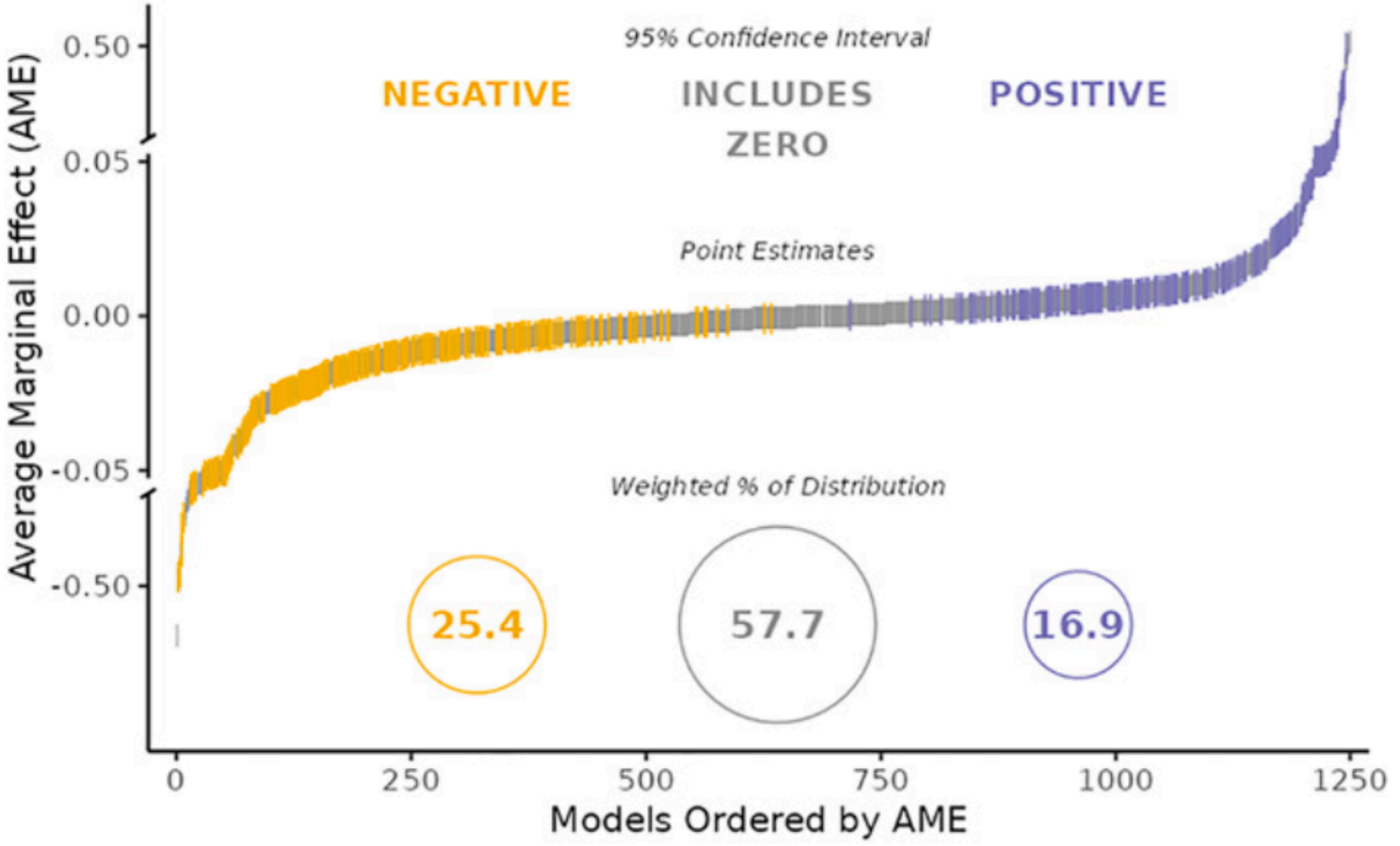


Fig. 1. Broad variation in the findings from 73 teams testing the same hypothesis with the same data. The distribution of estimated AMEs across all converged models (n = 1,253) includes results that are negative (yellow; in the direction predicted by the given hypothesis the teams were testing), not different from zero (gray), or positive (blue) using a 95% CI. AME are xy standardized. The y axis contains two scaling breaks at ±0.05. Numbers inside circles represent the percentages of the distribution of each outcome inversely weighted by the number of models per team.

4

Researchers are human, and they often have a tendency of using a **particular perspective** that favours particular populations, opinions, and research questions.

There are also risks of:

Confirmation bias—interpret incoming information in light of what you already believe

Interpretation bias—e.g., hostile attribution bias

Fundamental attribution error—attribute outcomes as coming more from people's **preferences** rather than the **situation** or the **structural environment**.

4

A Birmingham screwdriver

Researchers have a tendency to use the same:

methods (e.g. OLS or probit),

data (e.g. Polity IV), and

interpretation (coefficient significance)

across papers and (often) research sub-fields.

Source: https://www.toolsdiy.co.uk/media/catalog/product/cache/1/image/3318x/9df78eab33525d08d6e5fb8d27136e95/1/0/10372.jpg


Research is what other people do.
I think I can formulate a falsifiable hypothesis.
All I need to do is run a regression.
The best research involves storytelling & indicative data/case analysis.

How can we minimise the chance of making mistakes when creating our research design?

What theoretical, empirical, and simple human factors should we be aware of?

4