Can Theory-Guided Research Be Improved By Mindless Specification Robustness Algorithms?

Katrin Auspurg & Josef Brüderl
LMU Munich

Seminar “Analytical Sociology”

November 18th 2019, VIU

Background

- We want to identify a (total) causal effect

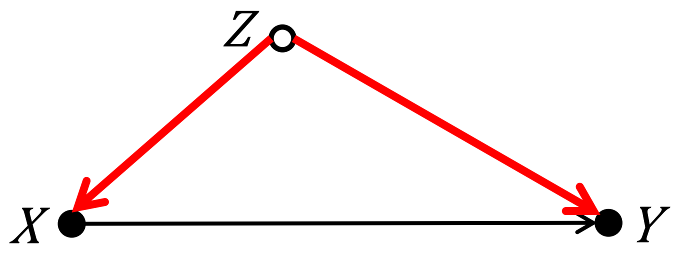
$$X \longrightarrow Y$$

- With observational data
  – Causal inference is threatened by several potential biases

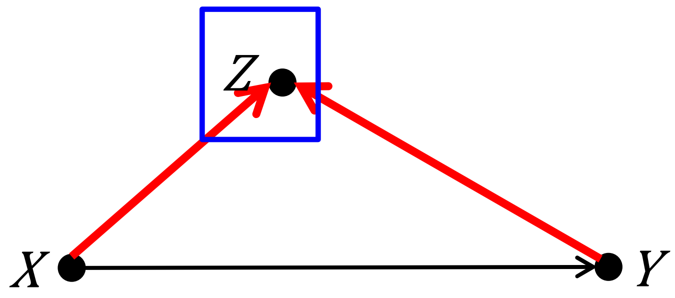
- Assumption: There is one correct model specification that allows for unbiased causal inference
  – Theory tells the researcher which model specification is the correct one (theory-guided research)

Three Fundamental Specification Errors

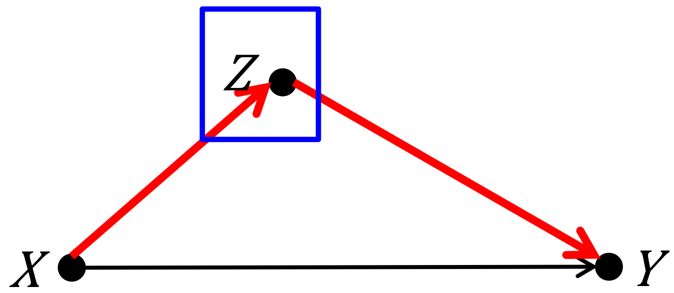
- Not controlling for a confounder
  - **omitted variable bias**



- Controlling for a collider
  - **collider bias**

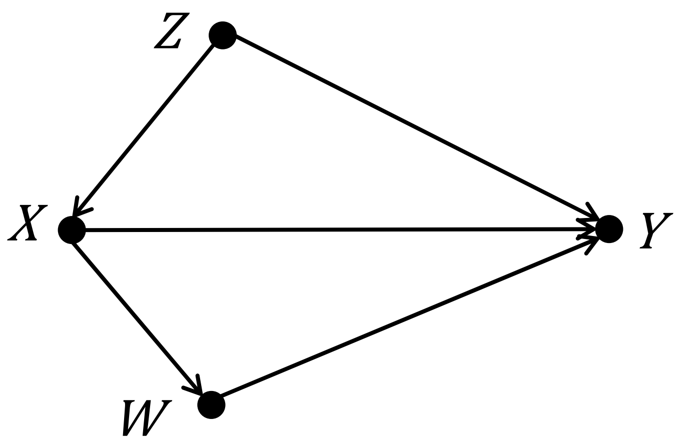


- Controlling for a mediator
  - **overcontrol bias**



Theory-Guided Research

- Theorizing about the causal structure of the research question



Z: confounder
W: mediator

  - Correct model specification

$$Y = \alpha + \beta X + \gamma Z$$

  - Mis-specified models

$$Y = \alpha + \beta X$$
$$Y = \alpha + \beta X \qquad + \delta W$$
$$Y = \alpha + \beta X + \gamma Z + \delta W$$

Does Theory-Guided Research Work?

- Even very competent researchers may fail in finding the correct specification
  - Young (2009) in re-analyzing Barro/McCleary (2003) concludes:
    "… top level competence … is not a solution to the problem of model uncertainty"

- Most social science theories are not informative enough to unambiguously identify the correct specification
  - Statistical models are a "garden of forking paths" and theory does not help (Gelman/Loken 2014)
  - "Analytical flexibility"

Does Theory-Guided Research Work?

- Finally, there are several mechanisms that make social researchers to defy theory-guided specification search:
  - Incentives are such that researchers may strive not for correct but for “significant” results (*p*-hacking, publication bias)
  - Current research practice does not require much effort in getting the correct model specification
    - Kohler et al. (2019) show that only 25 % of all ESR (2016/17) papers justify covariate selection

- **Mis-specified models are widespread in (theory-guided) social research**

Specification Robustness Algorithms

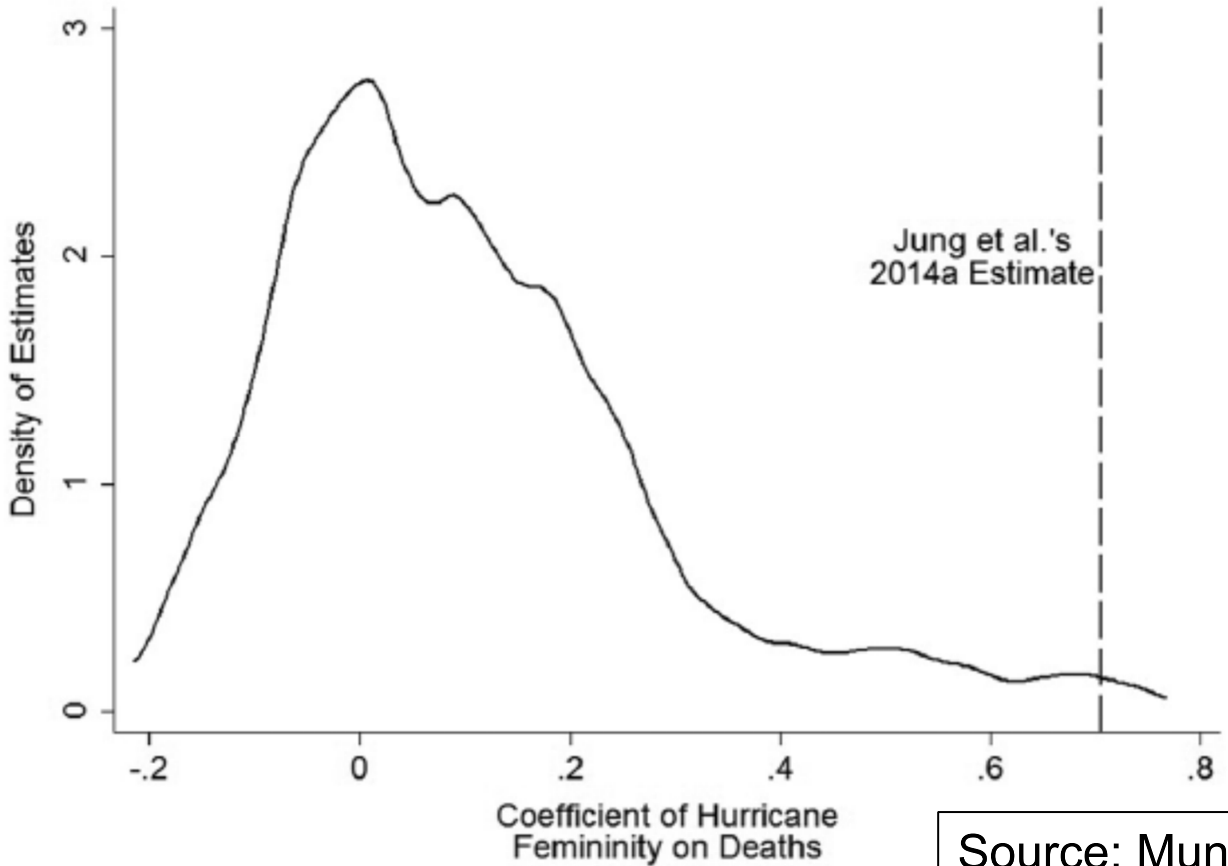
- Recently several specification robustness algorithms have been suggested
  - Specification curve, multiverse analysis, …
- Multimodel analysis (`mrobust`) (Young/Holsteen 2017)
  - Focus on **one** treatment effect
  - Allows for different statistical models / functional forms / operationalizations / controls
  - Runs models with all possible combinations of model ingredients
    - Plots distribution of treatment effect estimates (modeling distribution)
    - Provides influence statistics on treatment effect estimates
  - In the following: robustness to the choice of controls

Can Multimodel Analysis Help?

- Not helpful are
  - Optimal specification search algorithms
  - (Bayesian) model averaging
- Multimodel analysis might be helpful
  - It increases transparency
    - Model robustness analysis: “Are the results robust?”
  - It might stimulate theoretical reflection
    - Model influence analysis: “What modelling decisions are critical for obtaining the result and what is their theoretical justification?”
- Multimodel analysis starts from a given set of controls, thus it checks robustness “**inside”** the model
  - it cannot help identifying omitted variable bias,
  - but it can help identifying collider bias and/or overcontrol bias

Example: Are Female Hurricanes More Deadly?

- Jung et al. (2014) Female Hurricanes Are Deadlier Than Male Hurricanes. PNAS
- Mechanism: Residents tend to dismiss the destructive potential of storms with feminine names and take fewer precautions

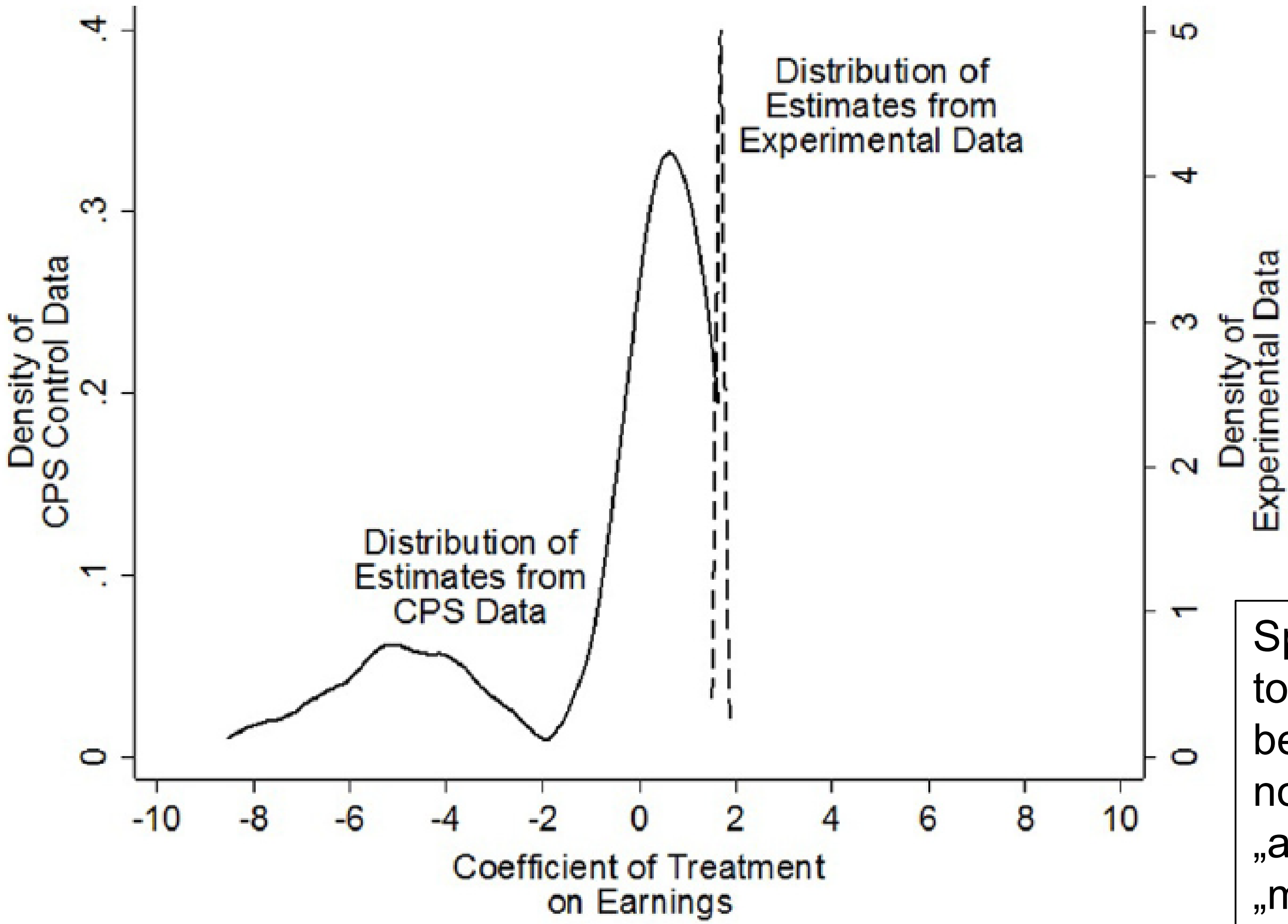


Source: Munoz/Young 2018

Example With an Experimental Benchmark

- Effect of job training on wages (re-employment after unemployment)
  - Field experiment (n=445) (LaLonde 1986)
  - CPS cross-sectional data (n=16,177)
  - Outcome: wage
  - Treatment: program participation dummy
  - Controls: past wages and unemployment status, age, race, marital status, and education
- Results of robustness analysis
  - Experiment: mean 1.69, almost no modeling variation
    - "The conclusions are given by the data, not by the choice of statistical model."
  - CPS: mean -0.82, large modeling variation
    - Depending on model specification one can conclude anything

Example With an Experimental Benchmark

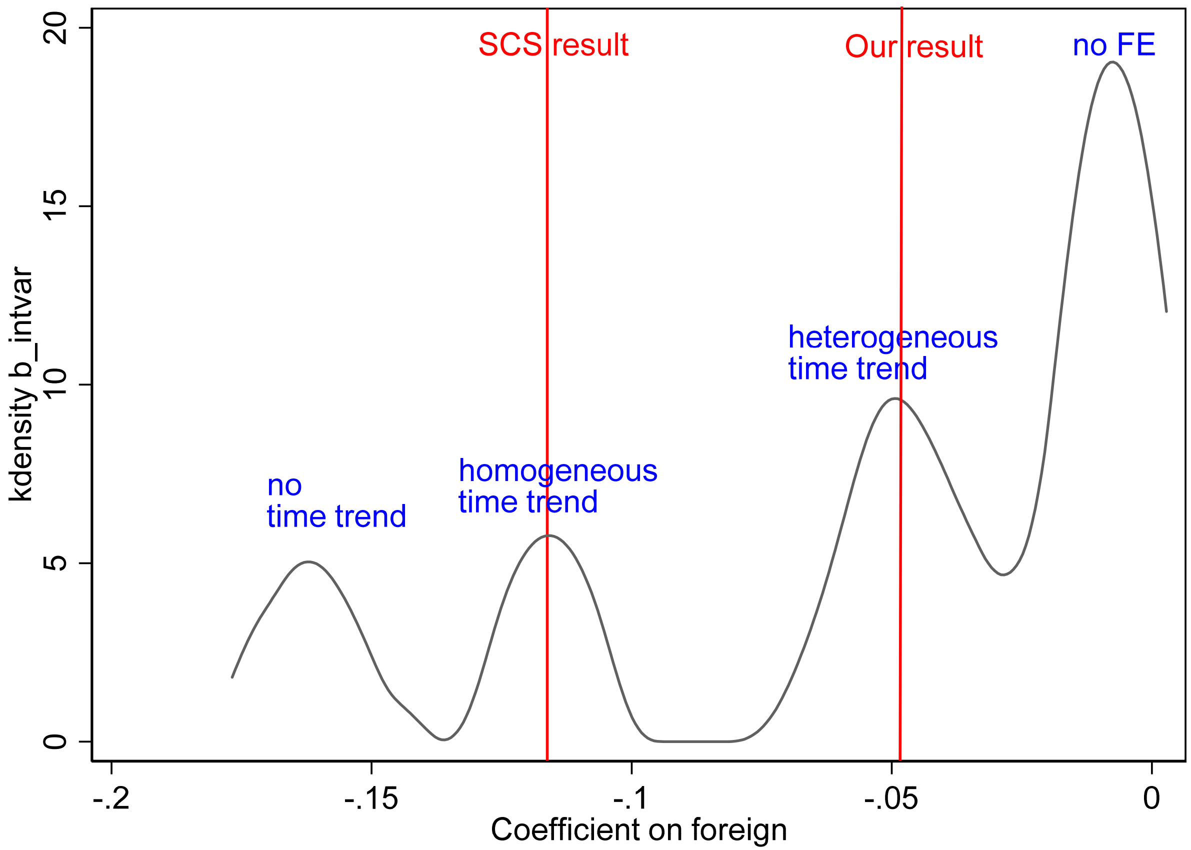


Specifications close to experimental benchmark:
not controlling for „age“ and „marital status“

Source: Munoz/Young 2018

An Example Modeling Distribution

- A recent study reports a strong negative effect of "proportion foreigners" on support for the welfare state
  - Schmidt-Catran/Spies. 2016. *ASR*. (SCS)
- A re-analysis argues that this results from model mis-specification
  - Auspurg/Brüderl/Wöhler. 2019. *ASR.*



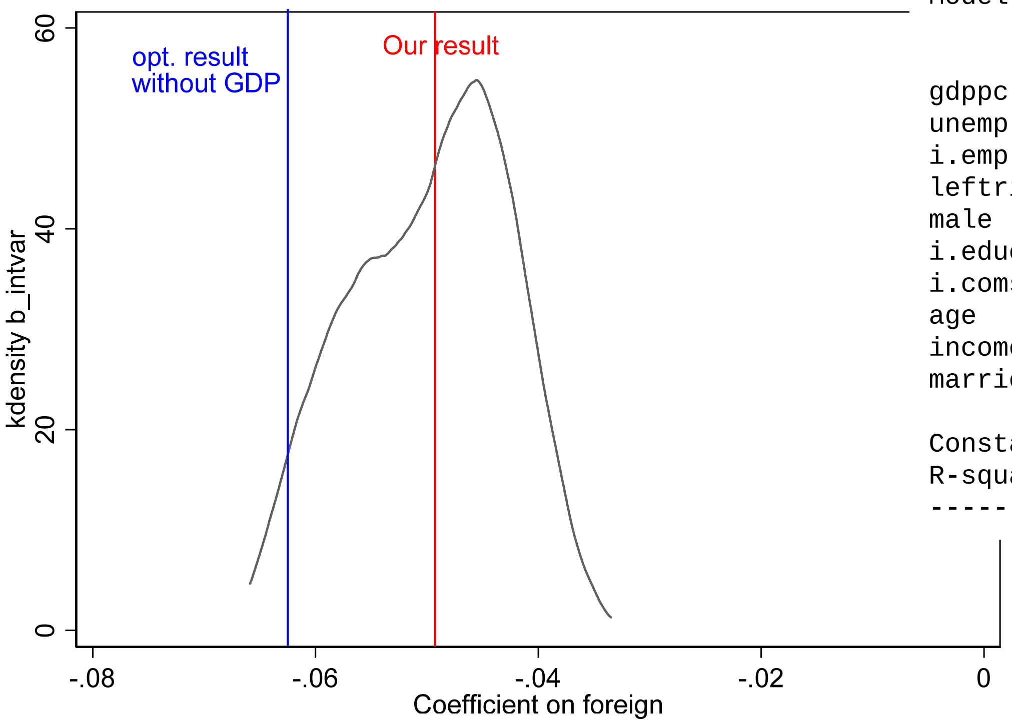
SCS specification:
controls, regional FE

New element added by us:
heterogeneous time trend

Obtained with `mrobust`
(see Young/Holsteen 2017)

Misuse: „Optimizing“ the Model

- Only models with FE + heterogeneous time trend
- Influence statistics say: not controlling for „GDP“ reduces the coefficient strongly
  - Now the effect of „foreign“ is significantly negative!



Model Influence

	Marginal Effect of Variable Inclusion	Percent Change From Mean(b)
gdppc	0.0109	-22.0%
unemplr	-0.0049	9.8%
i.emplstat	-0.0043	8.6%
leftright	-0.0031	6.1%
male	-0.0027	5.3%
i.educ	0.0020	-4.0%
i.comsize	-0.0019	3.8%
age	0.0010	-2.0%
income_hh_equiv	-0.0006	1.2%
married	-0.0006	1.2%
Constant	-0.0477	
R-squared	0.9600	

Conclusion

- Pay attention to model uncertainty (Young 2018)
  - The “footnote approach” to robustness is insufficient
- Use algorithms like `mrobust`
  - This creates transparency
  - This forces researchers to justify their model specification
- Use algorithms and mind!
  - Only mindful specification algorithms are helpful