Pitfalls of Counterfactual Inference

Hannes Keppler
December 5, 2019

Seminar: How do I lie with statistics?
Supervisor: Prof. Dr. Ullrich Köthe

What is Counterfactual Inference?

- **Factual**
  - Facts exist even if they are not known
- **Counterfactual (CF)**
  - e.g. Exam results if the student had studies
  - Used for:
    - "What if?"- questions
    - Forecasts
    - Causal inference, "Causal effect" ≈ "factual" - "counterfactual"

What is Counterfactual Inference?

Main Problem:

How to know, whether a counterfactual can be answered by the given dataset?

More precise:

How model-dependent is the inference?

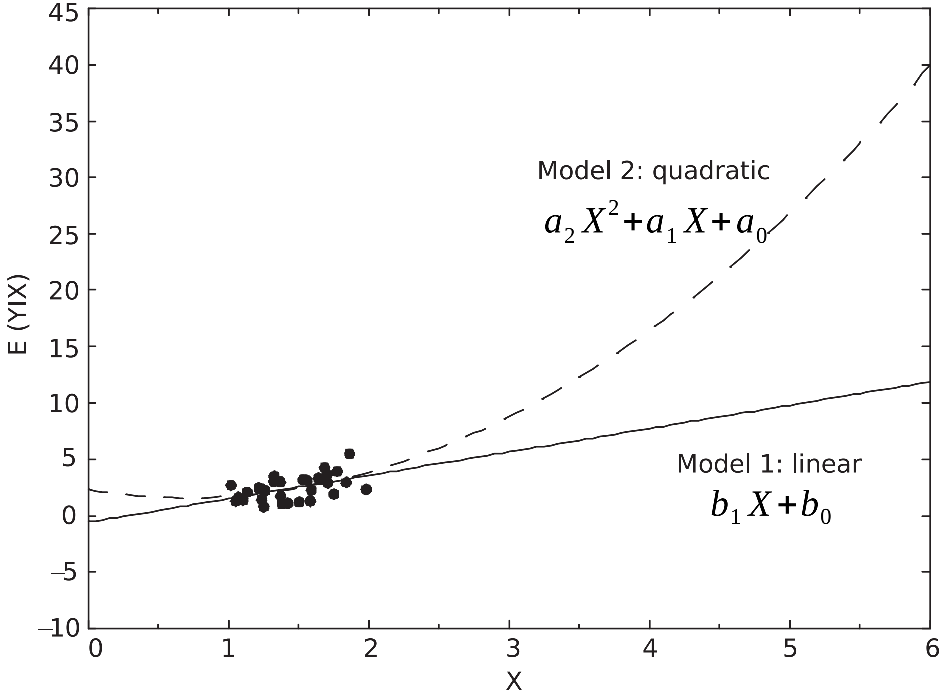
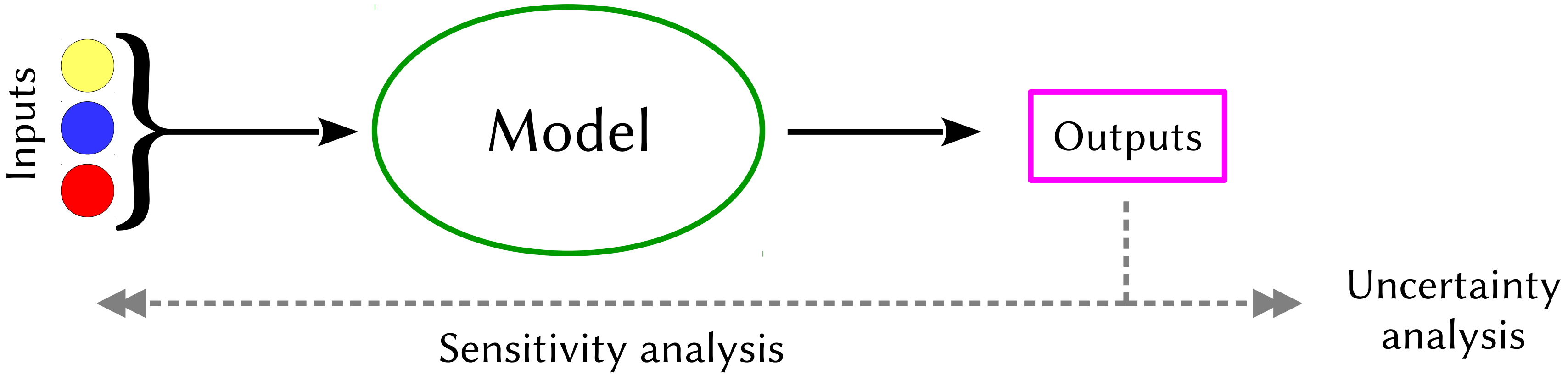


Fig. 1. Linear and quadratic model. Equal fit to data, but different out-of-sample predictions.

Sensitivity Analysis

- Common Method to study model-dependence
- Allocate the uncertainty in the output of a math. Model to its inputs.

e.g.: Sample Input and rerun model, “One At a Time”, Linear regression, Variance based methods, ...



Sensitivity Analysis

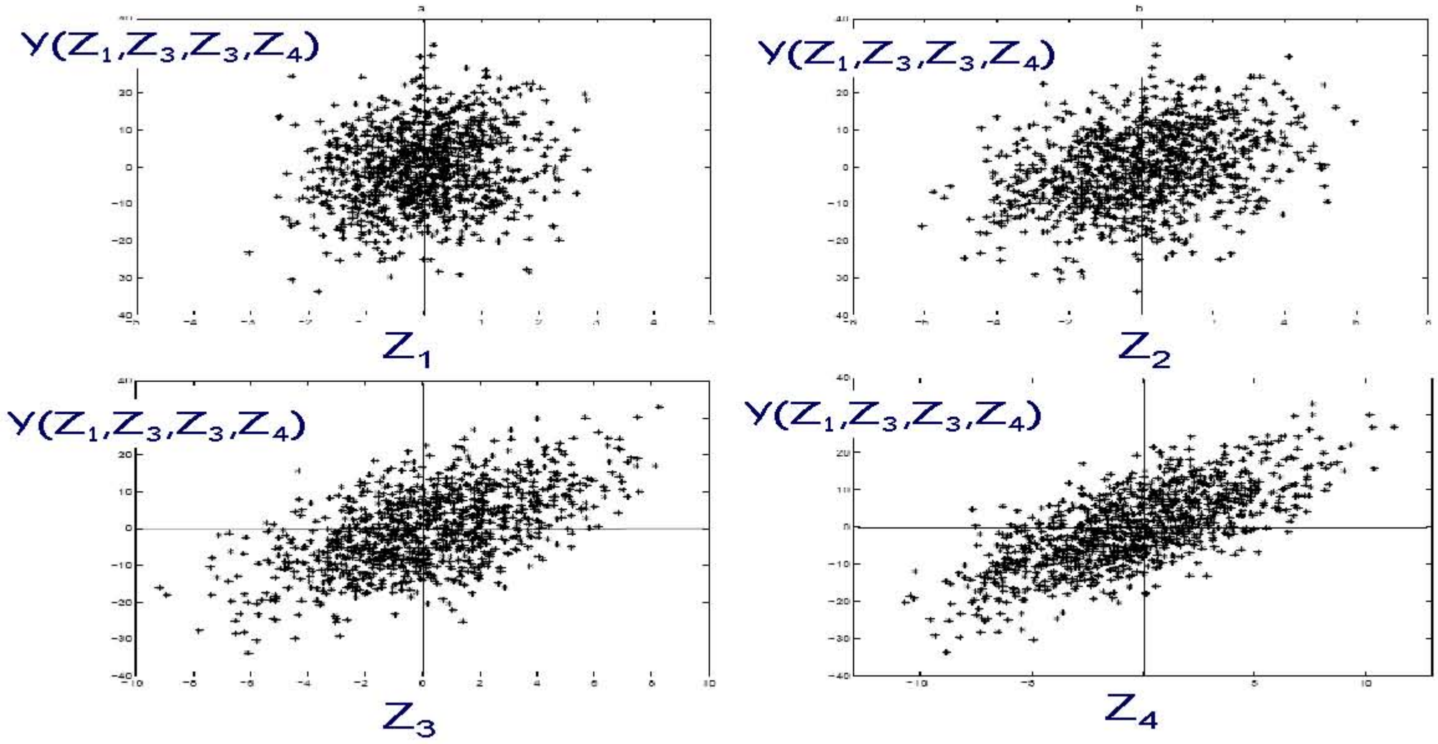


Fig. 2. Sampling-based sensitivity analysis by scatterplots. Z_4 is most important in influencing Y.

Sensitivity Analysis

Problems:

- Only given class of models is tested,
  often convenient ones
- Class of possible models often not easily formalized
- Model influences/determines alternative Hypothesis
- Task of the Analyst: choose model and optimize parameters

Interpolation vs. Extrapolation

Task: Determine how far a CF is from the dataset.

- Extrapolation

  $\longrightarrow$ normally *more* model dependence

- Interpolation

  $\longrightarrow$ normally *less* model dependence

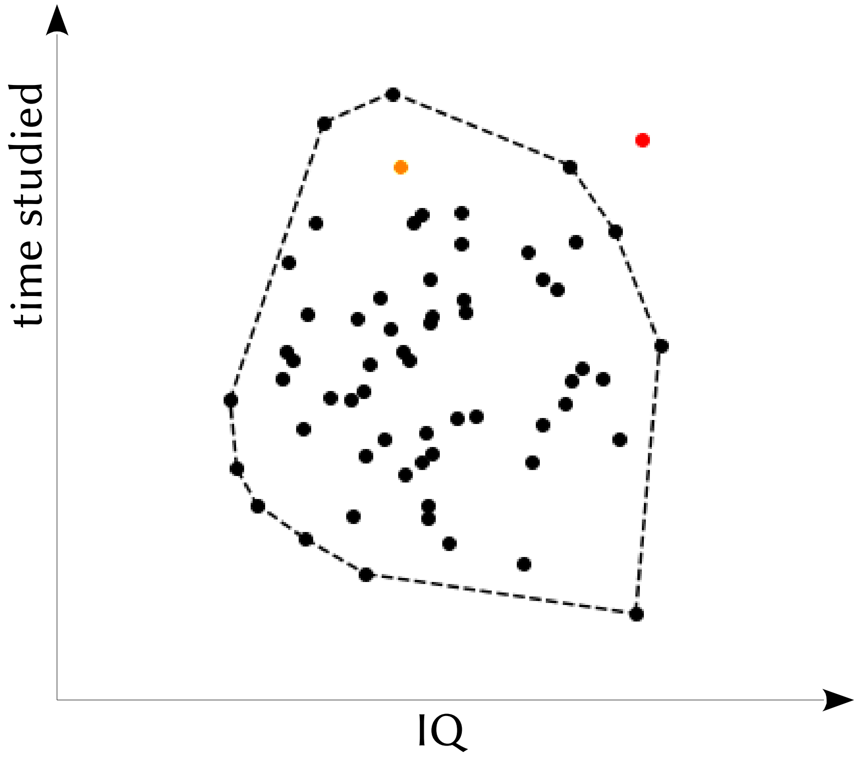


Fig. 3. Dataset with convex hull and two CF (red and orange).

Interpolation vs. Extrapolation

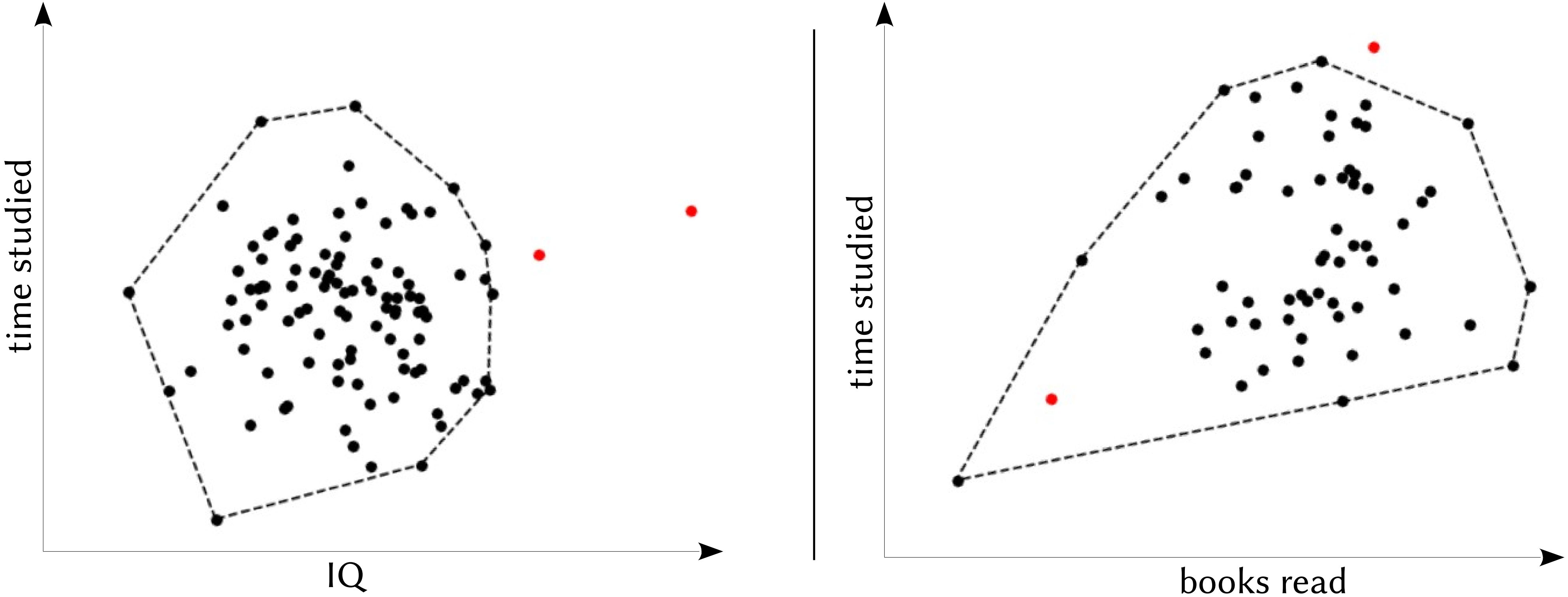


Fig. 4. Possible problems with simple convex hull criterion. Exemplary explanatory variables to predict results in an exam.

Appropriate Metric:

$$G(x,y) = \frac{1}{K} \sum_{k=1}^{K} \frac{|x_k - y_k|}{r_k}$$

$$r_k = max(X_k) - min(X_k)$$

- Distance in relation to range of the data

Geometric variability (GV):

- Standard deviation of distances between data points

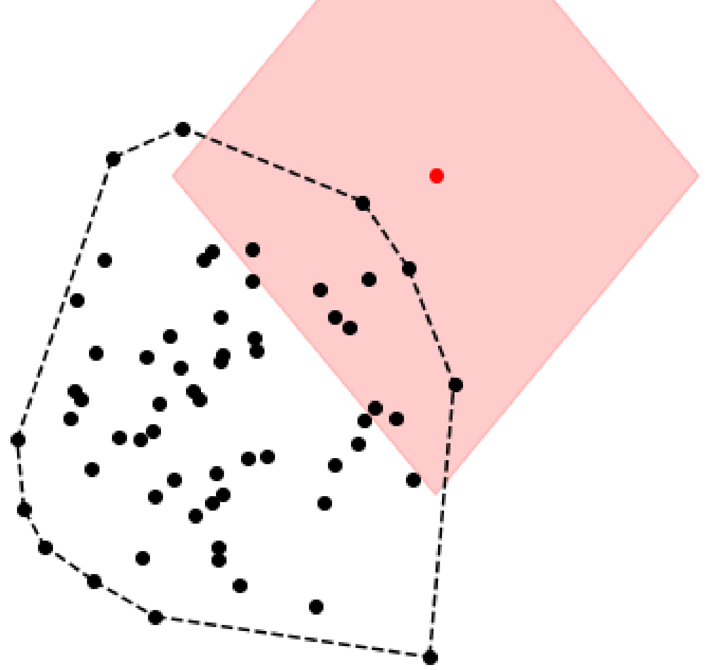


Fig. 5. CF (red) and area within 1GV. 17% of data is in this region.

Example: UN peacebuilding

- *Doyle and Sambanis (2000). International Peacebuilding: A Theoretical and Quantitative Analysis.*
- 124 past World War II civil wars
- Analysis of the correlates of successful peacebuildung and influence of UN operations

- **Dependent variable:** PBS2—peacebuilding success or failure two years after the end of the war

- **Implicit Counterfactual:** 1 – UNOP4
  - Only 1.3% of data inside 1GV range ! ⟶ *expect high model dependence*

Example: UN peacebuilding

	Original Model			Modified Model		
Variables	Coeffcient	Robust SE	p-Value	Coeffcient	Robust SE	p-Value
UNOP4	3.135	1.091	0.004	0.262	1.392	0.851
Wardur x UNOP4	-	-	-	0.037	0.011	0.001

Logistic model / Modified model :
$$P(X) = \frac{1}{1 + \exp\left(\sum_i \beta_i x_i + \alpha \cdot x_{Wardur} x_{UNOP4}\right)}$$

Tab. 1. UN example. Original model and modified model with an interaction between UNOP4 and duration of war.

Example: UN peacebuilding

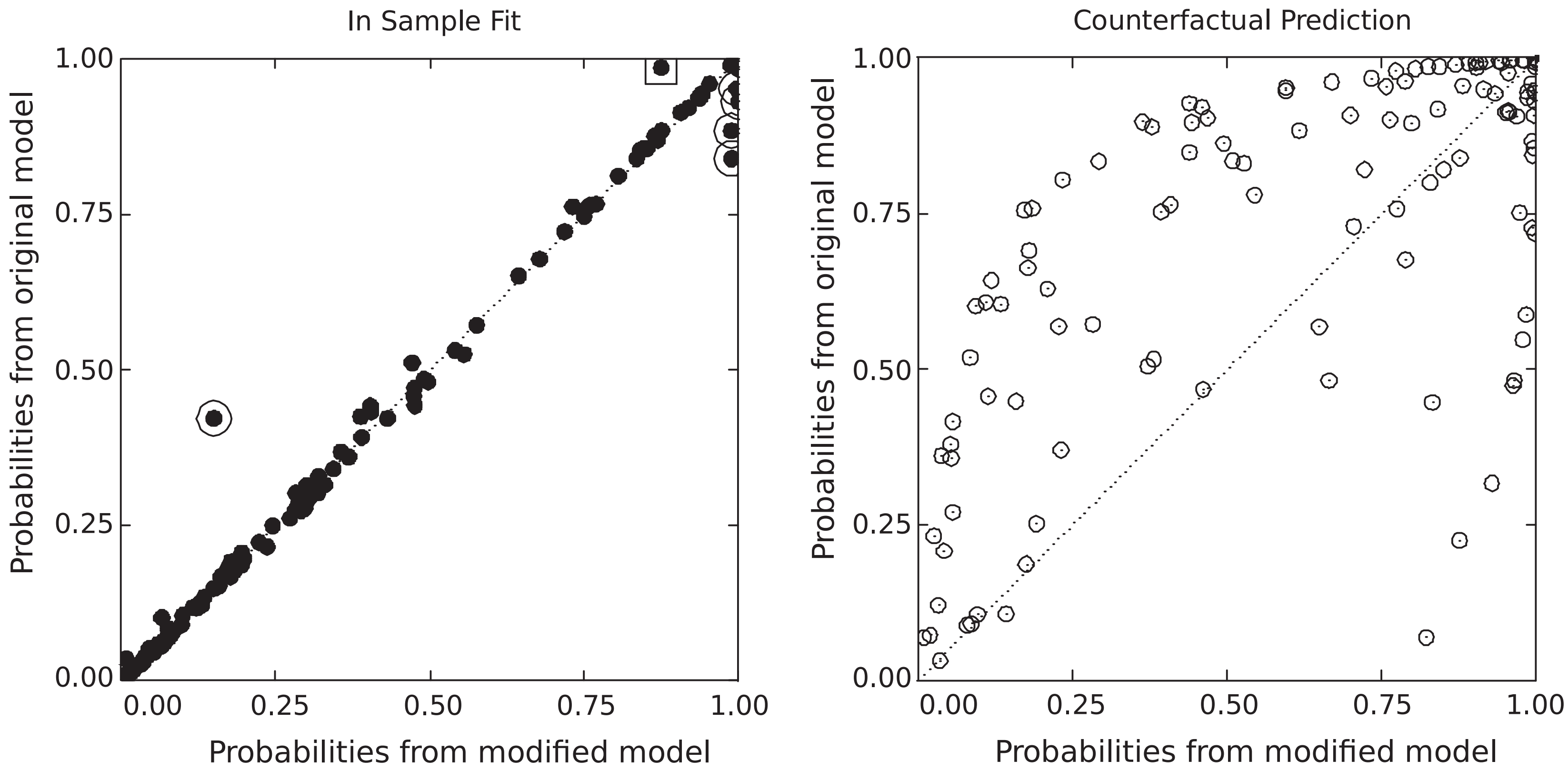


Fig. 6. Comparison of predictions of original and modified model.

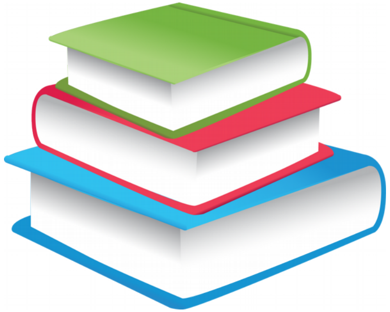
Causal inference

Average causal effect among the treated

$$\theta \ = \ E(Y_1, D{=}1) - E(Y_0, D{=}1)$$

"factual" - "counterfactual"

Example: How does studying effect the result of an exam?



Y_1 ... Result with studying.
Y_0 ... Result without studying.
D ... Has the student studied?

θ is not accessible, but an estimator can be used. **Simple estimator:**

$$d \ = \ mean(Y, D{=}1) - mean(Y, D{=}0)$$

Causal inference – Bias decomposition

- **Bias:** Expected value of results differs from true underlying parameter.

$$Bias = E(d) - \theta = E(Y_0, D=1) - E(Y_0, D=0)$$

- Is it admissible to use $E(Y_0, D=0)$ for $E(Y_0, D=1)$?
- Possible problems: $Bias = \Delta_o + \Delta_p + \Delta_i + \Delta_e$

Causal inference – Bias decomposition

- Omitted variable bias Δ_o
  - Additional control variable Z

- Post treatment bias Δ_p
  - Some Z may be highly correlated to D

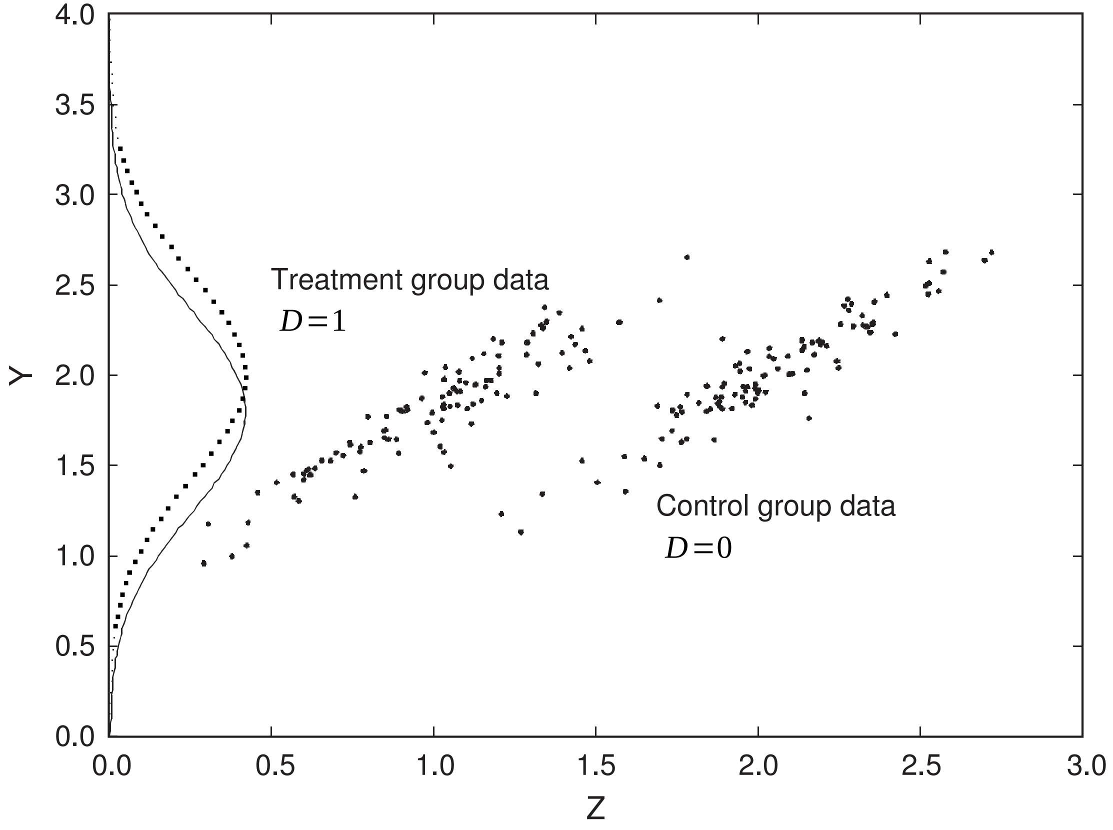


Fig. 7. Omitted variable bias. Influence of additional Z.

Causal inference – Bias decomposition

- Interpolation and Extrapolation bias Δ_i, Δ_e
  - Problems in adjusting for the control variables

    i : in the region of the data

    e : Out of the dataset

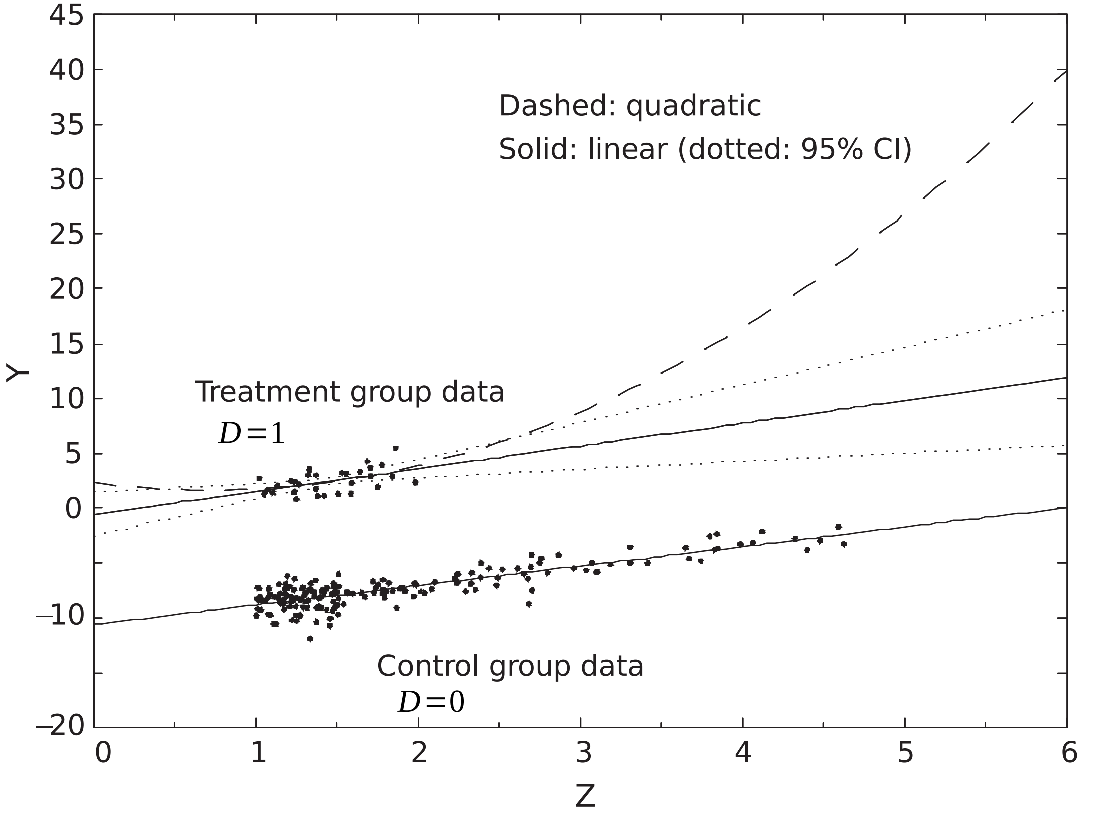


Fig. 8. Extrapolation bias.

Example: UN peacebuilding

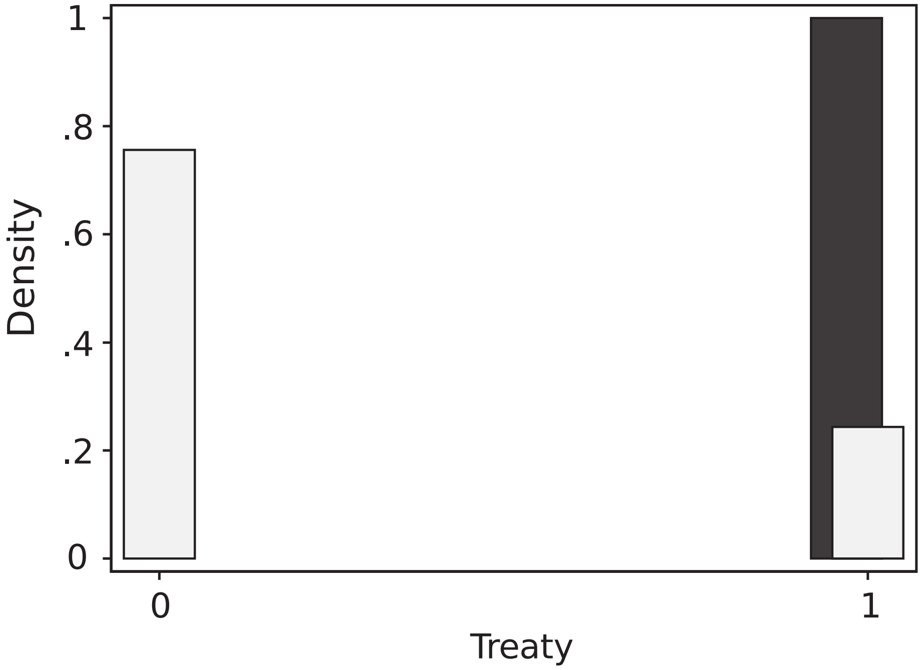


Fig. 9. Wars with and without signed treaty.
Dark: with UN intervention.
Light: without UN intervention

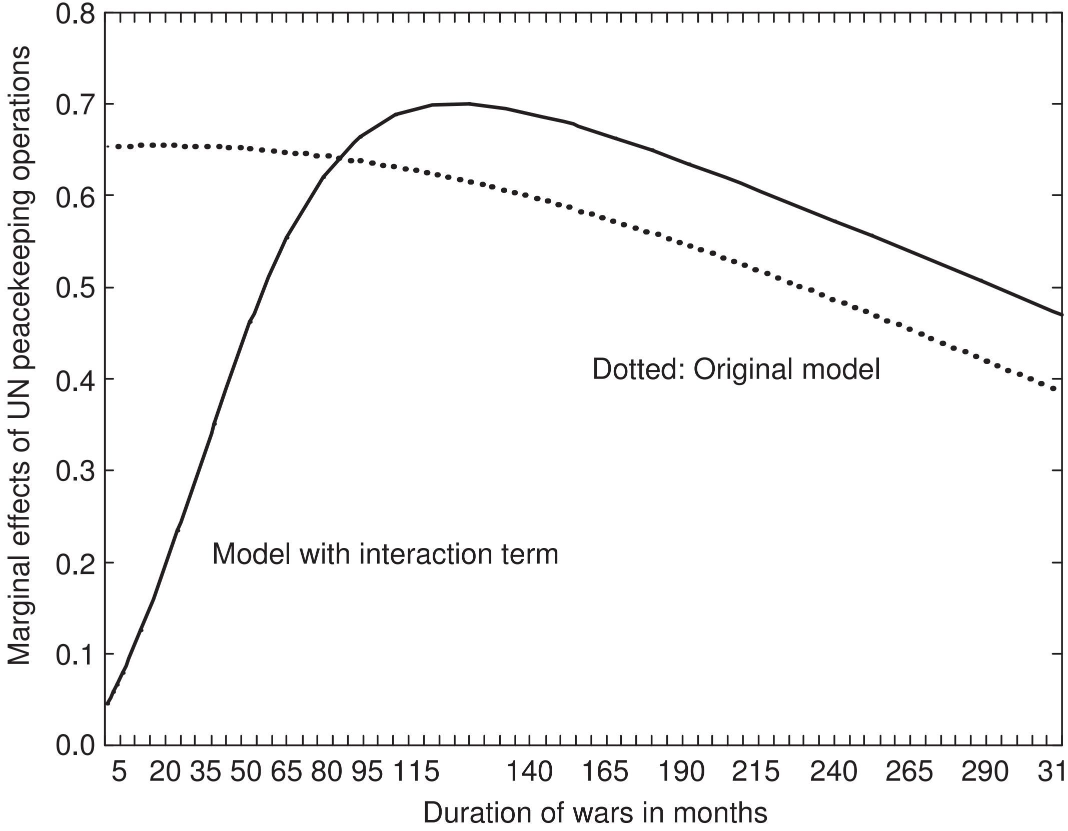


Fig. 10. Estimated causal effect of UN peacekeeping operations.

Summary

- CFs are essential for causal inference
  - Observe effect of a minor change
- Some CFs can't be reliably answered by a given dataset
- Distance of CF to the data $\longrightarrow$ degree of model-dependence
- Classical pre Neural Networks Statistics
  - today even more powerful methods

    e.g. cross validation, typical set, ...

References

- King and Zeng. (2007) When Can History Be Our Guide? The Pitfalls of Counterfactual Inference. *International Studies Quarterly* 51:183-210. Available at: https://gking.harvard.edu/files/gking/files/counterf.pdf
- King and Zeng. (2006) The Dangers of Extreme Counterfactuals. *Political Analysis* 14(2):131–159. Available at: https://gking.harvard.edu/files/gking/files/counterft.pdf
- Doyle and Sambanis. (2000) International Peacebuilding. *American Political Science Review* 94(4):779–801. DOI: https://doi.org/10.2307/2586208
- Pearl. (2009) Causal inference in statistics: An overview. *Statistics Surveys* 3:96–146. Available at: https://ftp.cs.ucla.edu/pub/stat_ser/r350.pdf
- Nalisnick, Matsukawa, Whye Teh, Lakshminarayanan. (2019) Detecting Out-of-Distribution Inputs to Deep Generative Models Using Typicality. In: *arXiv e-prints*. arXiv:1906.02994

- Figures 1, 6-10 and Table 1 taken from King and Zeng (2007)
- Figure 2 from https://en.wikipedia.org/wiki/File:Scatter_plots_for_sensitivity_analysis_bis.jpg

Thanks for your attention.