Applying Informal Benchmarking to the *f*-Sensitivity Model

Author Adrians Slics
A.Slics@student.tudelft.nl
Supervisor Matej Havelka
Responsible Professor Jesse Krijthe

1 Introduction

- Causal inference from observational data typically assumes all confounders are observed, an assumption that rarely holds [4].
- Sensitivity analysis instead quantifies how unobserved confounding could distort causal conclusions [1], via frameworks like the marginal sensitivity model (MSM) and the f-sensitivity model [3].
- Both require a sensitivity parameter to bound the hidden confounding, yet principled guidance for setting it is lacking [1].
- Informal benchmarking uses observed covariates as reference points to suggest plausible values, but has not been studied for the f-sensitivity model and its distributional-robustness perspective.
- Both bound the effect of an unmeasured confounder U on treatment:

  MSM: 1/Γ ≤ OR(X,U) ≤ Γ
  f-sensitivity: E[f(OR(X,U))] ≤ ρ (f = KL)
  Γ, ρ: sensitivity parameters.

2 Research question

Main

Can the idea of informal benchmarking be applied to the f-sensitivity model?

Sub-questions

1.What are the differences between the MSM and the f-sensitivity model that affect the transferability of informal benchmarking?

2. Under which conditions does the adapted benchmarking procedure produce accurate parameter estimates?

3. Under which conditions does the adapted procedure fail or produce misleading results?

3 Methods

For an adapted informal benchmarking procedure, the setup is:

1. For each dropped covariate, calculate the propensity scores.
2. Evaluate the shift in divergence with KL.
3. Choose the largest resulting sensitivity parameter ρ.

The benchmarked ρ is then assessed in two settings:

- **Against the MSM**: placing both on the odds-ratio scale (Figure 1).
- **Against ground truth:** in a logistic DGP with a known confounder U, comparing the benchmarked ρ to the true ρ.

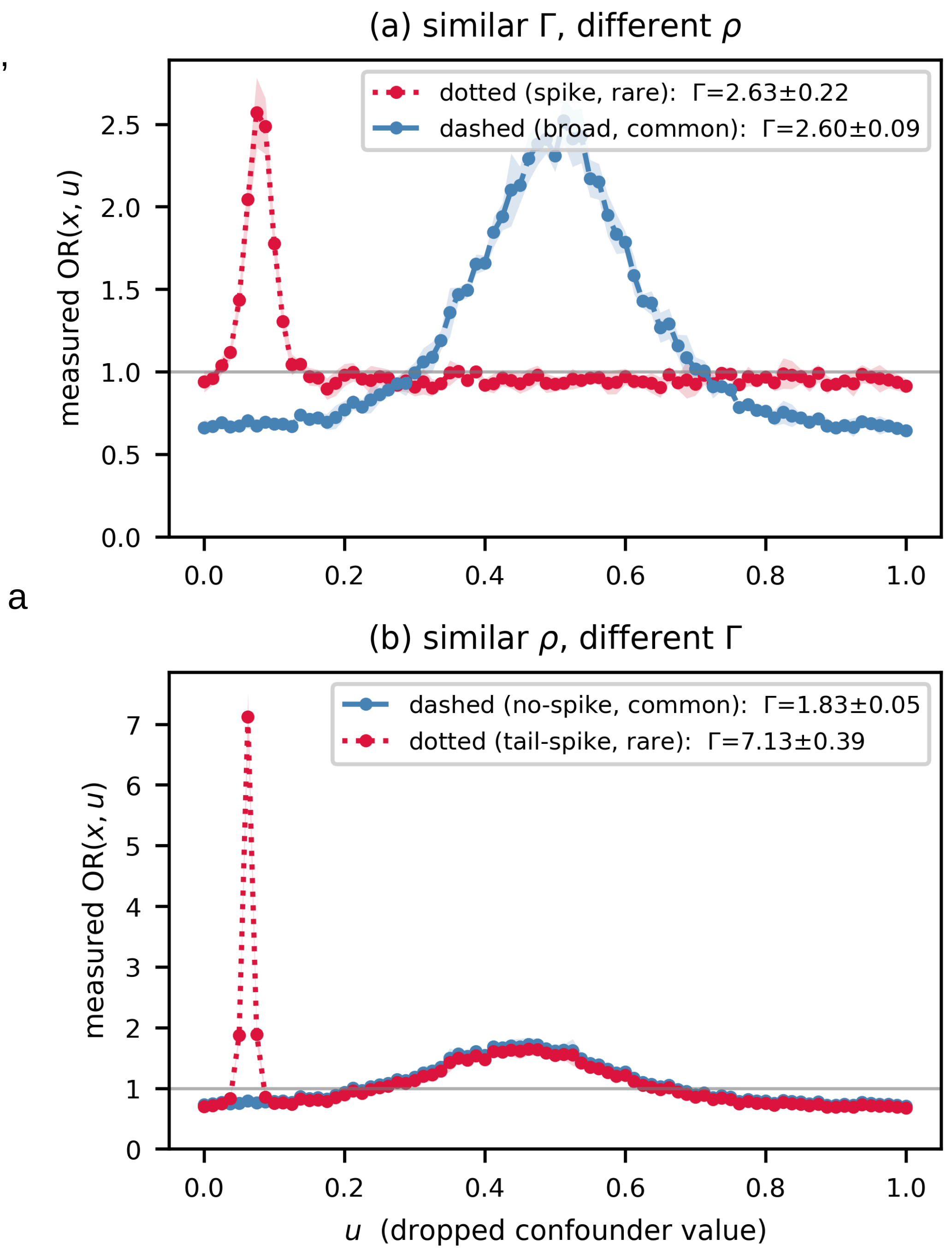

Figure 1. Measured OR under spike vs. broad confounders

4 Results

- **Benchmark recovers an upper bound on the true divergence.** Synthetic DGP with simulated U (true ρ known): ρ_bench covers truth on both scales in all four scenarios, tightest under homogeneous confounding, loosest when one covariate exceeds U (Fig 3).
- **MSM and f-sensitivity respond to different features of the confounding distribution (Fig 2).** A rare, concentrated spike raises Γ but not ρ; heterogeneous confounding of similar strength raises ρ while MSM bounds stay flat.

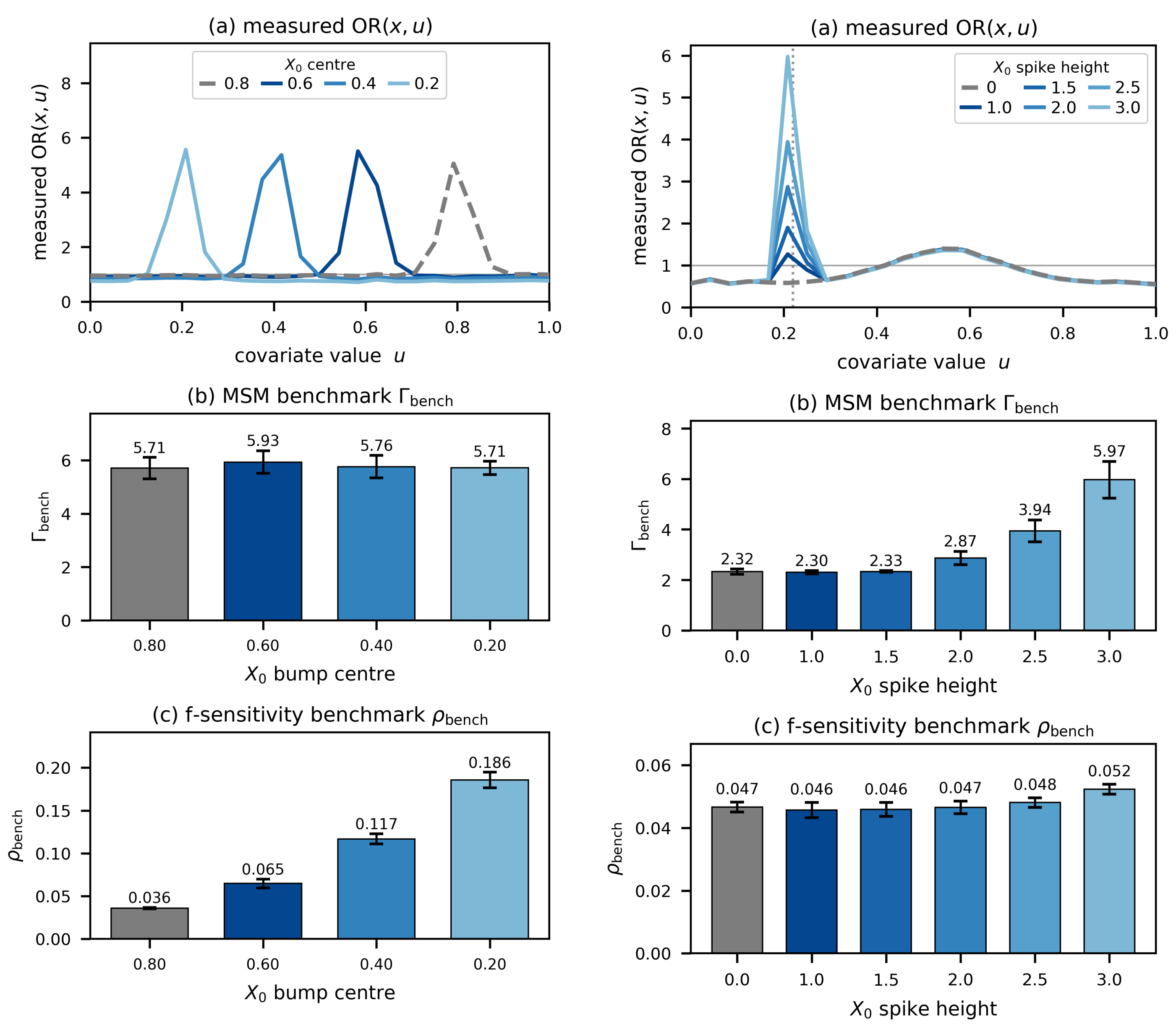

Figure 2. Sensitivity to confounder location and strength

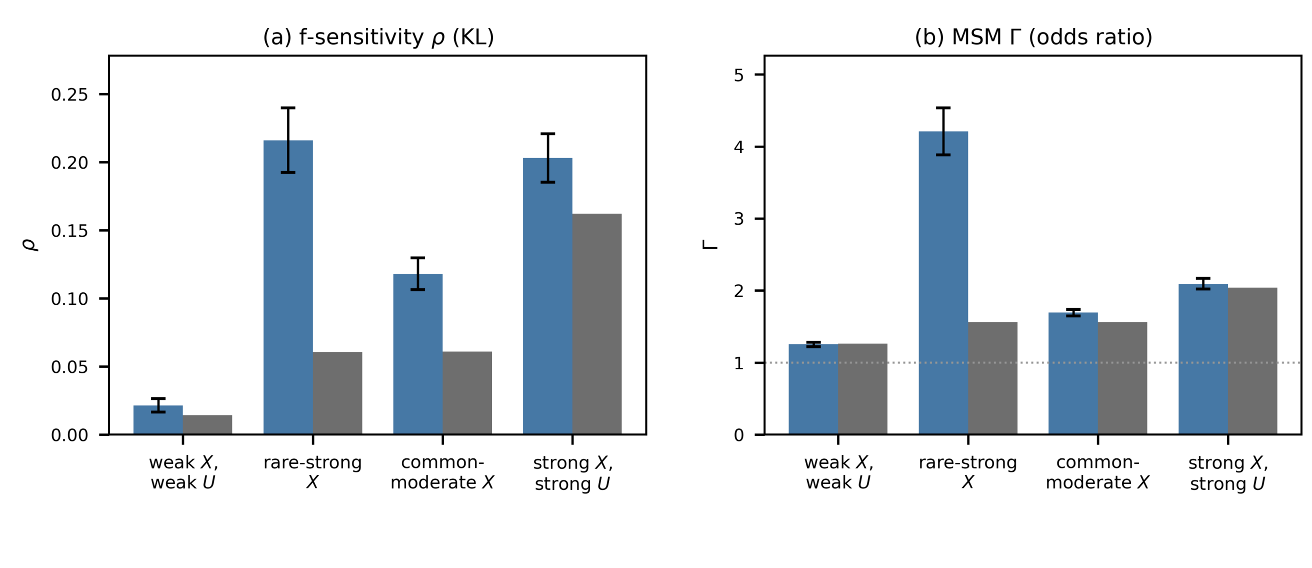

Figure 3. Benchmark vs. true value across confounding forms

5 Conclusions

Informal benchmarking can be applied to the f-sensitivity model. The benchmarked ρ conservatively recovers the ρ of a known confounder, but it is unreliable when confounding hides in rare, low-probability regions.

Future work: extend the procedure to correlated and continuous covariates, and test divergences beyond KL.

References

[1] Baitairian, J.-B., Sebastien, B., Jreich, R., Katsahian, S., & Guilloux, A. (2025). Calibrating confounding strength in sensitivity models for weighting estimators: A comparative review and a new method (arXiv:2510.16560). arXiv. doi.org/10.48550/arXiv.2510.16560

[2] Imbens, G. W. (2003). Sensitivity to exogeneity assumptions in program evaluation. American Economic Review, 93(2), 126–132. doi.org/10.1257/000282803321946921

[3] Jin, Y., Ren, Z., & Zhou, Z. (2022). Sensitivity analysis under the f-sensitivity models: A distributional robustness perspective (arXiv:2203.04373). arXiv. doi.org/10.48550/arXiv.2203.04373

[4] Yao, L., Chu, Z., Li, S., Li, Y., Gao, J., & Zhang, A. (2021). A survey on causal inference. ACM Transactions on Knowledge Discovery from Data, 15(5), 1–46. doi.org/10.1145/3444944

Repo: github.com/ASlics/bsc-informal-benchmarking-f-sensitivity-model