Sensitivity Analysis for Multi-Treatment Causal Inference with Unobserved Confounding

Alexander Franks

October 22, 2021

University of California, Santa Barbara

Multi-Treatment Inference

- Goal: assess the causal effect of multiple treatments applied simultaneously in an observational context
  - Genome Wide Association Studies (GWAS)
  - What is the effect of an actor on movie revenue

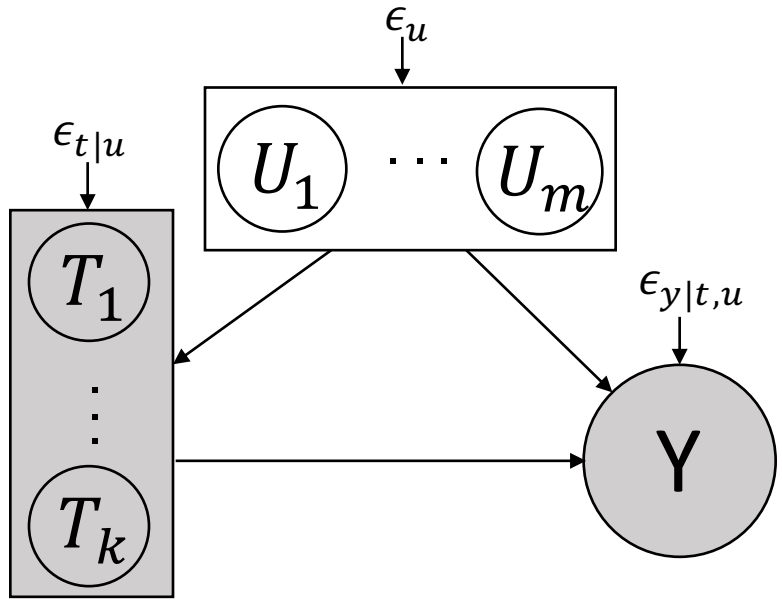
- Can possibly leverage correlation between treatments to control for potential unmeasured confounders.

- Many concurrent estimands (e.g. effect of each gene)

- Renewed interest in causal community due to Wang and Blei, 2019.

Multi-Treatment Inference: Setup

- Outcomes Y (scalar)
- Treatments T (k-vector)
- Unmeasured confounders U (m-vector)



Assumptions

Assumption 1: Latent ignorability

U blocks all backdoor paths between T and Y.

Assumption 2: Positivity

$$f(T = t \mid U = u) > 0 \text{ for all } u.$$

Assumptionm 3: SUTVA

There are no hidden versions of the treatments and there is no interference between units.

“The Deconfounder” Approach (Wang and Blei, 2019)

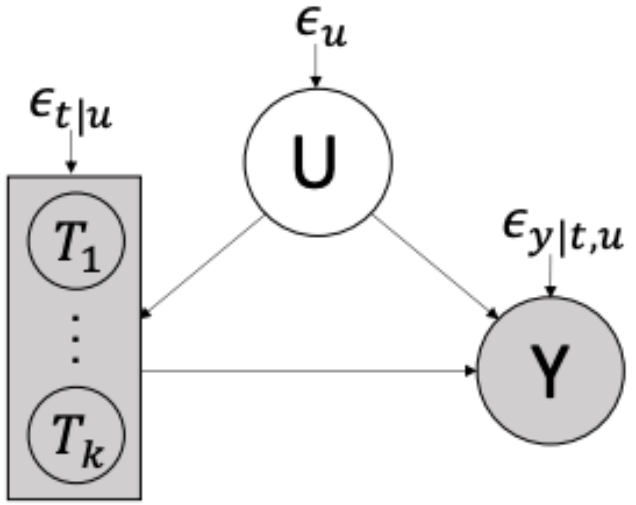
- Fit a factor model to infer substitute confounders:

$$\hat{U} = E[U|T]$$

- “Correct for” bias by including proxy confounder to debias treatment effect estimates:

$$Y \sim \hat{U} + T$$

- Assume U is pinpointed by T as k goes to infinity

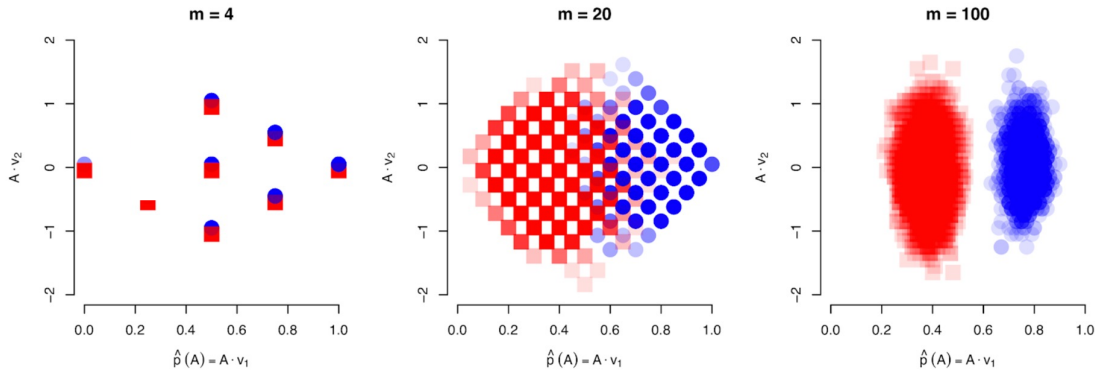


Some problems with “The Deconfounder”

Lack of general nonparametric identification (D’Amour, 2019a,b)

Counterexamples to theoretical results (Ogburn, 2020)

“The Deconfounder” does not outperform naïve regression (Grimmer, 2020)



Methods of this type appear across science, and are standard procedure (e.g., Price et al 2006 in GWAS).

Principal components analysis corrects for stratification in genome-wide association studies

AL **Price**, NJ Patterson, RM Plenge, ME Weinblatt… - Nature …, **2006** - nature.com

Population stratification—allele frequency differences between cases and controls due to systematic ancestry differences—can cause spurious associations in disease studies. We describe a method that enables explicit detection and correction of population stratification on a genome-wide scale. Our method uses principal components analysis to explicitly model ancestry differences between cases and controls. The resulting correction is specific to a candidate marker's variation in frequency across ancestral populations, minimizing …

☆ 99 Cited by 7154 Related articles All 27 versions

This talk

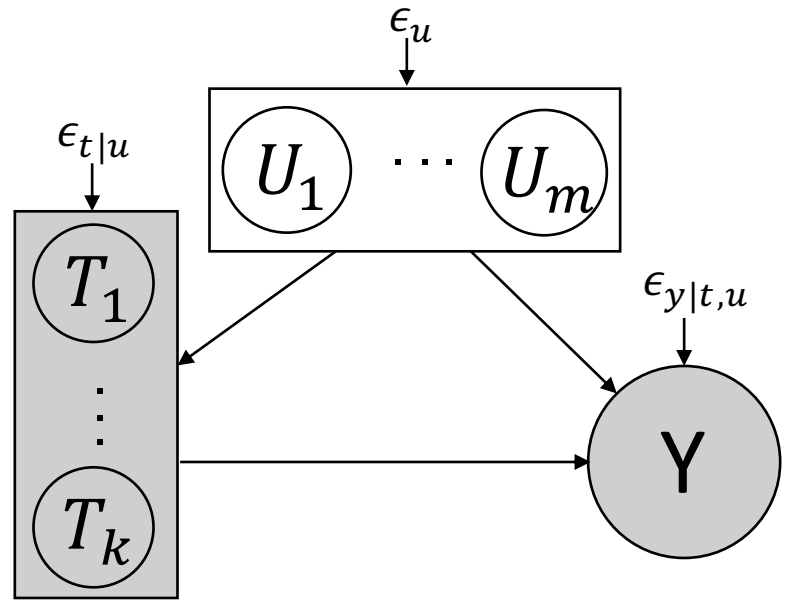
- Reconcile intuition and practical success with negative theoretical results
- Flexible and interpretable sensitivity analysis for multi-treatment inference
- Some theoretical insights about the what might be gained in multi-treatment inference

The Role of Sensitivity Analysis

- Relax unverifiable identifying assumptions
- Readers can assess claims more precisely
- Unique challenges in the multiple treatment setting which require careful consideration

Setup

- Outcomes Y (scalar)
- Treatments T (k-vector),
- Confounders U (m-vector)



- Use the do-calculus framework (Pearl, 2009)
- do operator indicates the density of y in the population in which we **intervened** to assign t

$$\mathrm{PATE}_{t_1,t_2} := E(Y \mid do(t_1)) - E(Y \mid do(t_2))$$

$$\text{for treatment vectors } t_1 \text{ and } t_2$$

Observed and Intervention Densities

Observed: $$f(y \mid t) = \int_{\mathcal{U}} f_{\psi_Y}(y \mid t, u) f_{\psi_T}(u|t) du$$

Intervention: $$f_{\psi}(y \mid do(t)) = \int_{\mathcal{U}} f_{\psi_Y}(y \mid t, u) f_{\psi_T}(u) du$$

ψ are sensitivity parameters

Note: $$f_{\psi_T}(u) du = \int f_{\psi_T}(u \mid \tilde{t}) f(\tilde{t}) d\tilde{t}$$

Copula Approach to Multi-Treatment Sensitivity

$$f_{\psi}(y \mid do(t)) = f(y \mid t) \int c_{\psi_Y}\left(F_{Y\|t}(y), F_{U\|t}^{\psi_T}(u) \mid t\right) f(u)du$$

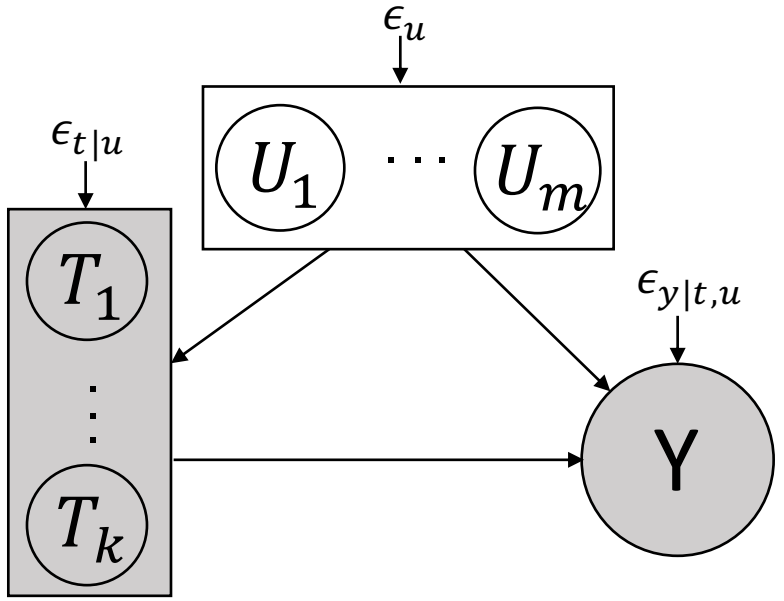
- $c_{\psi}\left(F_{Y|t}(y), F_{U|t}(u) \mid t\right)$ is the conditional copula which characterizes the dependence between Y and U given T.
- ψ_T might be identified (up to an equivalence class) with multiple treatments (e.g. a latent factor model)
- ψ_Y remains unidentified

Motivating Example: Analysis of Mouse Obesity

- Explore the effect of 17 gene expression values on mouse weight (Miao et al 2020)
- Likely confounded due to batch effects and unmeasured phenotypes
- Gene expression are the “treatments” and weight is the outcome
  - Model treatments and outcomes as Gaussian
- Correlation between expression levels (”treatments”) is indicative of potential confounding

Building Intuition: The Linear Gaussian Model

$$U = \epsilon_u$$
$$T = BU + \epsilon_{t|u}$$
$$Y = \tau' T + \gamma' U + \epsilon_{y|t,u}$$



$$\epsilon_u \sim N_m(0, I)$$
$$\epsilon_{t|u} \sim N_k\left(0, \sigma^2_{t|u} I_k\right)$$
$$\epsilon_{y|t,u} \sim N\left(0, \sigma^2_{y|t,u}\right)$$

The Linear Gaussian Model

In this model:

- $f(u \mid t) \sim N(\mu_{u|t}, \Sigma_{u|t})$ is identifiable (up to scale / rotation)
- $$PATE_{t_1,t_2} = PATE_{\Delta t} = \tau' \Delta t$$
  $$\text{Bias}_{\Delta t} = (\hat{\tau} - \tau)' \Delta t$$

Sensitivity analysis:

- Consider confounding bias of naïve estimates as a function of γ
- Bias varies across treatment contrasts, Δt

Worst-case bias of naïve estimators

Theorem

Suppose that the observed data is generated by model 1-3 with $\sigma^2_{t|u} > 0$. Then, $\forall \gamma$ satisfying Assumptions 1 and 2,

$$\gamma^T \Sigma_{u|t} \gamma \leq \sigma^2_{y|t} \tag{4}$$

For any given Δt, we have

$$\text{Bias}^2_{\Delta t} \leq \sigma^2_{y|t} R^2_{Y \sim U|T} \|\Sigma^{-1/2}_{u|t} \mu_{u|\Delta t}\|^2_2 \quad \text{(bounded)}, \tag{5}$$

The bound is achieved when γ is colinear with $\Sigma^{-1}_{u|t} \mu_{u|\Delta t}$.

The omitted variable bias is proportional to the scaled difference in confounder means

Overall worst-case bias

Corollary

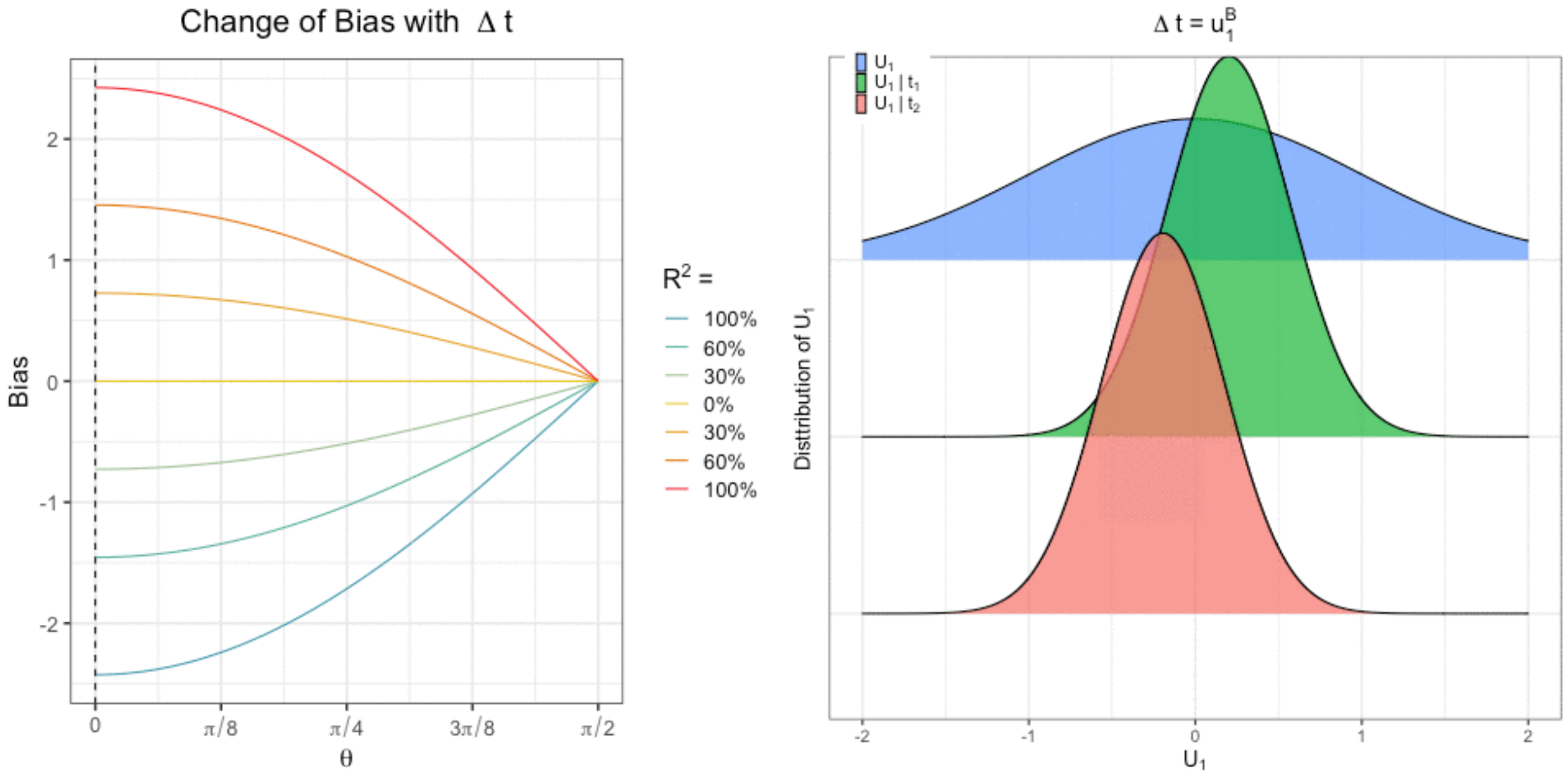
Let d_1 be the largest singular value of B. For all Δt with $\| \Delta t \|_2 = 1$, the squared bias is bounded by

$$\text{Bias}^2_{\Delta t} \leq \frac{d_1^2}{(d_1^2 + \sigma^2_{t|u})} \frac{\sigma^2_{y|t}}{\sigma^2_{t|u}} R^2_{Y \sim U|T}, \tag{6}$$

with equality when $\Delta t = u_1^B$, the first left singular vector of B. When $\Delta t \in Null(B')$, the naive estimate is unbiased, that is, $PATE_{\Delta t} = \tau'_{\text{naive}} \Delta t$.

$\frac{d_1^2}{\left(d_1^2 + \sigma^2_{t|u}\right)}$ is the fraction of variance in the first PC of treatments that can be explained by confounding.

Confounding Bias



Robustness and Calibration

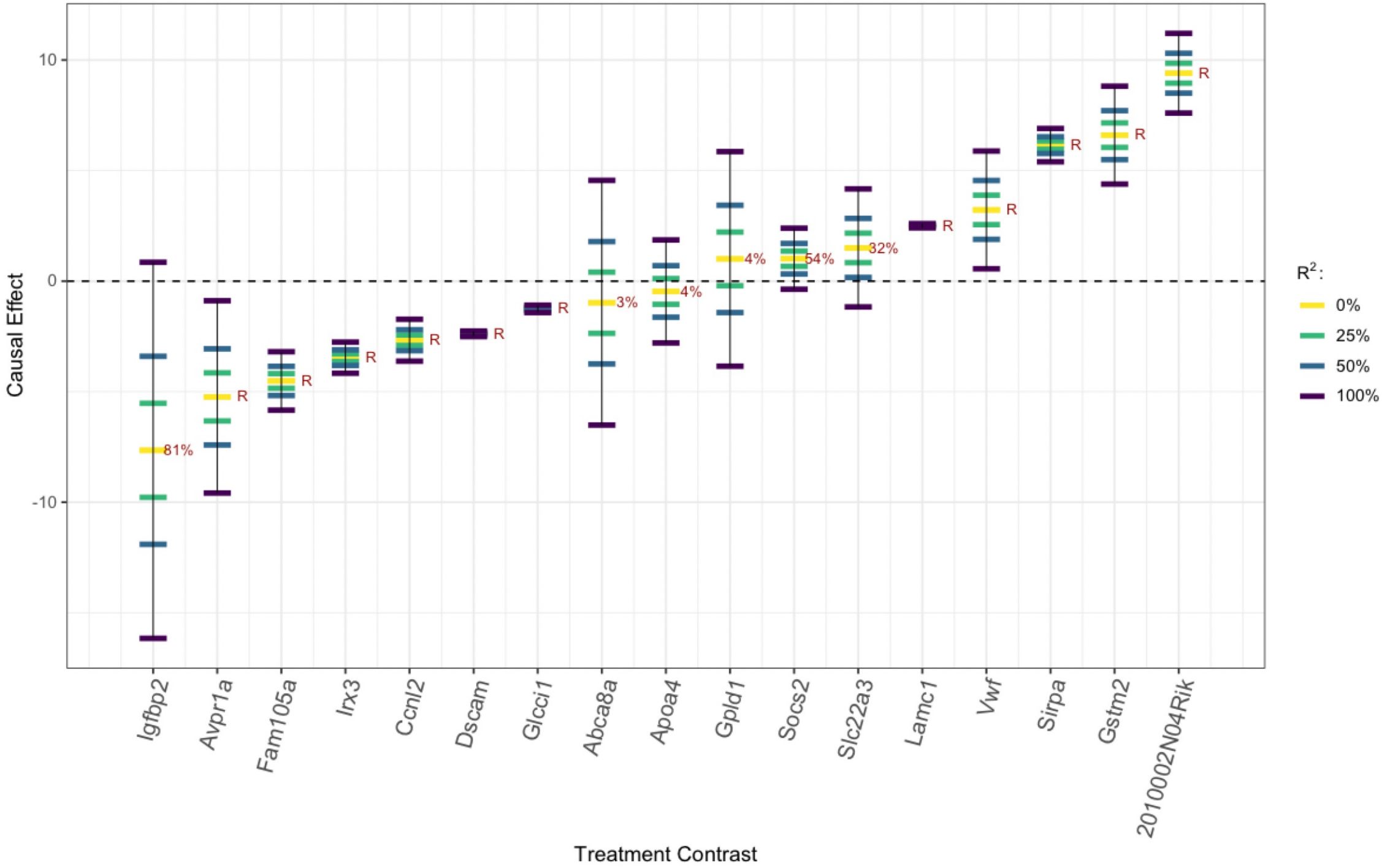
- Sensitivity analysis consists of two parts:
  1. The sensitivity model parameterization
  2. Tools for mapping external assumptions to specific models in the set
- Models parameterized by $\gamma = \sqrt{R^2_{Y \sim U|T}} \Sigma^{-1/2}_{u|t} d$
- Default: choose d to maximize bias and reason about $R^2_{Y \sim U|T}$.

Robustness and Calibration

- The *robustness value* is the smallest value of $R^2_{Y\sim U|T}$ that negates the sign of the treatment effect.
- If no value can change the sign we declare the effect robust to this confounding
- Can reason about by robustness by comparing $R^2_{Y\sim U|T}$ to:
  - $R^2_{Y\sim T}$, the observed fraction of variance explained by treatments
  - $R^2_{Y\sim T_j|T_{-j}}$ the partial fraction of variance explained by some, given others

Analysis of Mouse Obesity

- Explore the effect of 17 gene expression values on mouse weight (Miao et al 2020)
- Likely confounded due to batch effects and unmeasured phenotypes
- Gene expression are the “treatments” and weight is the outcome
  - Model treatments and outcomes as Gaussian
- Correlation between expression levels is indicative of potential confounding


Causal Effect
10
0
-10
81%
R
R
R
R
R
R
3%
4%
4%
54%
32%
R
R
R
R
Igfbp2
Avpr1a
Fam105a
Irx3
Ccnl2
Dscam
Glcci1
Abca8a
Apoa4
Gpld1
Socs2
Slc22a3
Lamc1
Vwf
Sirpa
Gstm2
2010002N04Rik
Treatment Contrast
R²:
0%
25%
50%
100%

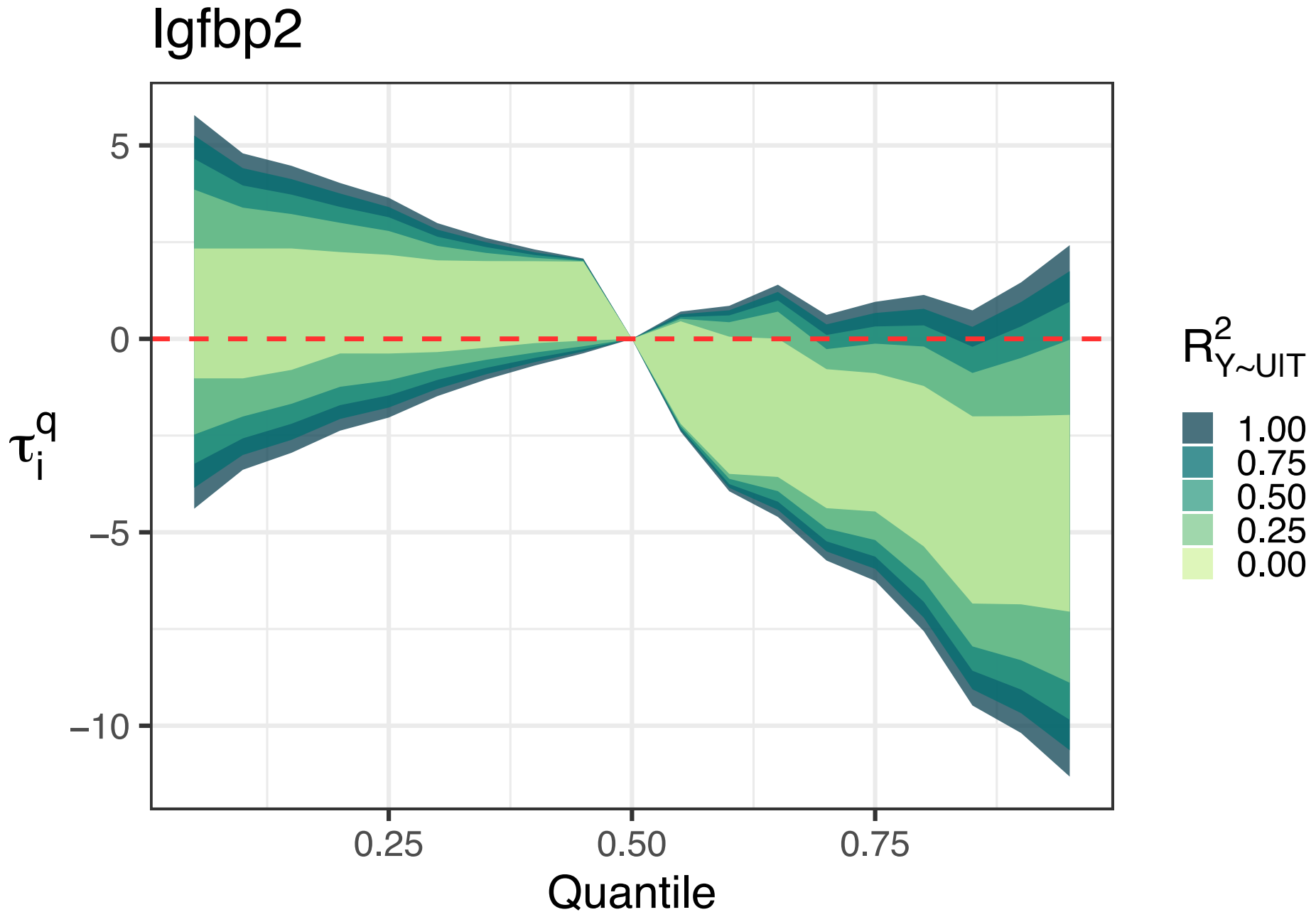
	τ_{mean}^{naive}	τ_{limit}^{naive}	$RV_{limit}(\%)$
2010002N04Rik	11.74	3.63	***
Vwf	10.69	3.87	***
Serpina6	7.52	2.79	16
Mest	5.88	0.78	1
Slc22a3	5.20	0.56	5
Ear11	3.53	0.96	14
Ccnl2	-4.89	-0.76	99
Irx3	-5.11	-0.47	58
Ndrg1	-6.38	-1.56	65
Igfbp2	-6.66	-2.59	31
Kdm4a	-7.54	-0.57	1
Abca8a	-8.01	-1.80	26
Gapdh	-11.67	-4.56	***
Fam105a	-15.60	-6.69	57

Non-linear models

- Model: $Y = g(T) + \gamma' U + \epsilon_{y|t,u}$
- Model the outcome given treatments using Bayesian Additive Regression Trees
- New estimand:

$$\tau_j^q = E\left[Y \mid do\left(t_j^q\right)\right] - E\left[Y \mid do(t) = t_j^{0.5}\right]$$

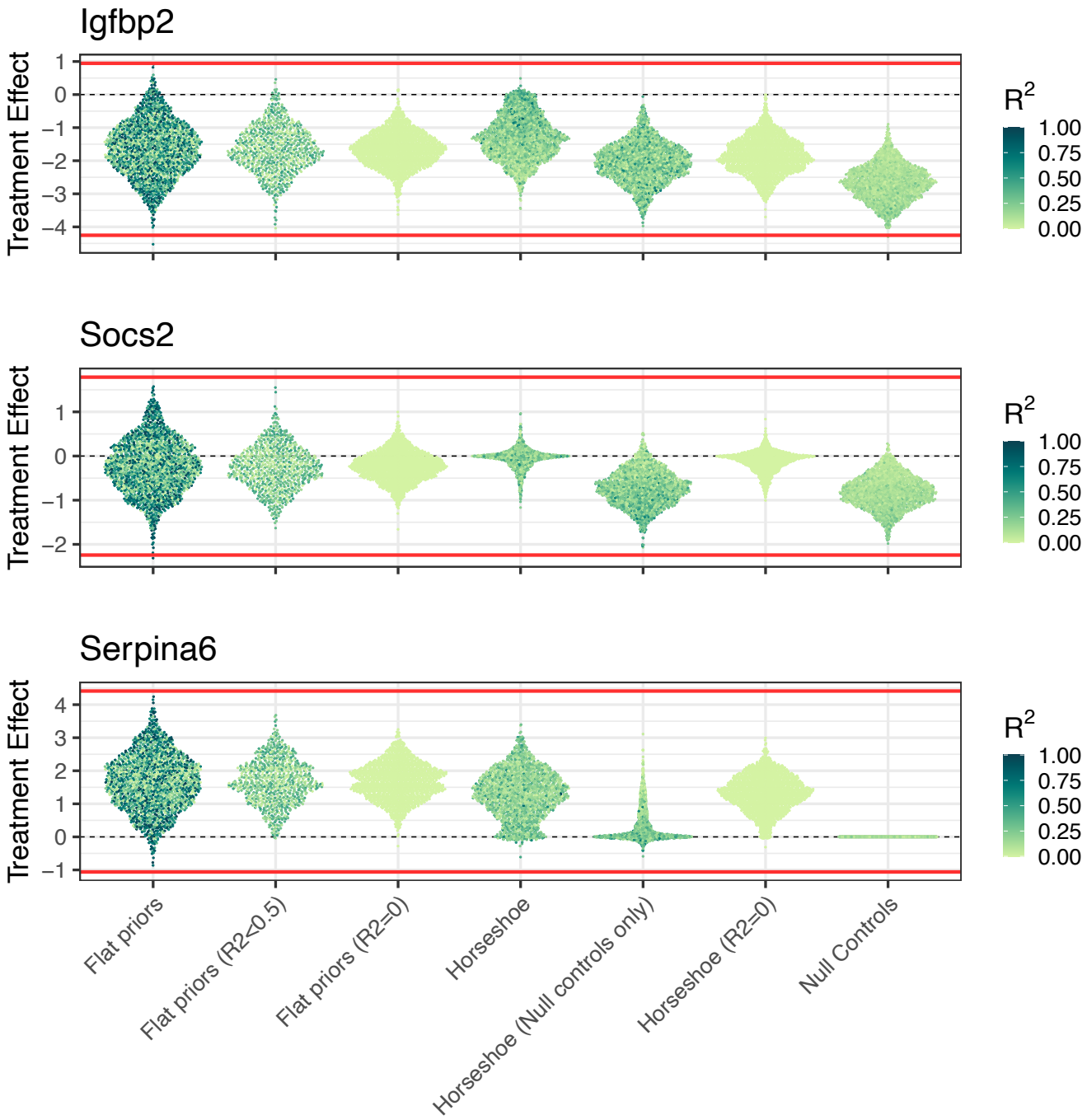
where t_j^q is the treatment vector with all genes at their median level except for the jth gene which has it's level set to the q-th quantile.

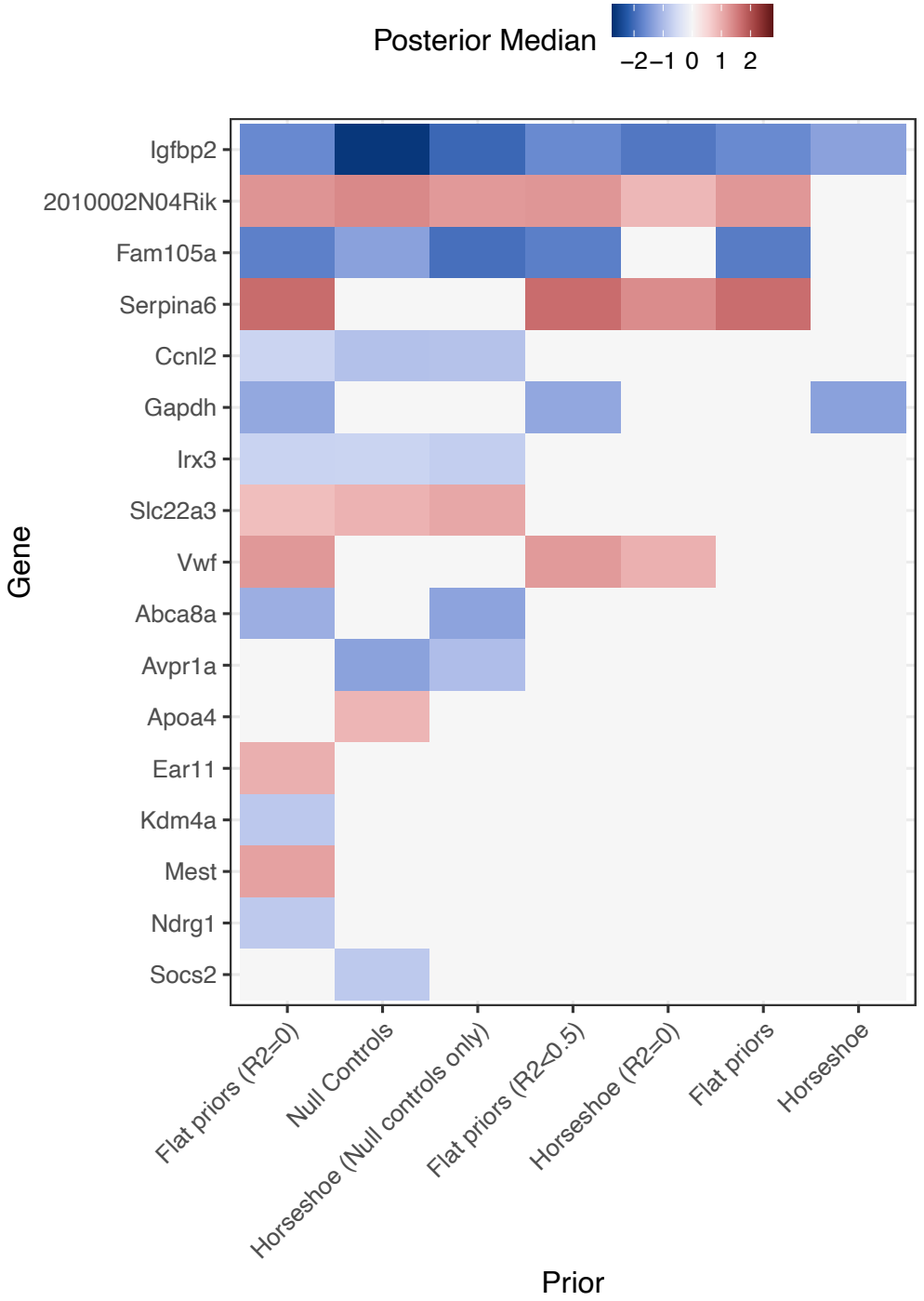

Igfbp2
5
0
−5
−10
τ_i^q
0.25
0.50
0.75
Quantile
$R^2_{Y\sim UIT}$
1.00
0.75
0.50
0.25
0.00

Additional Assumptions for Reducing Ignorance

- **R-squared:** what fraction of the outcome variance is due to confounding?
- **Null controls:** specify which set of treatments are known to have no causal impact on the outcome.
- **Sparsity:** the majority of treatments have no effect
  - Miao et al (2020) propose an identification strategy based on the assumption that at least half of the treatments have no effect (”null treatments”)

Natural to encode these assumptions with prior distributions in a Bayesian framework


Igfbp2
Treatment Effect
1
0
−1
−2
−3
−4
R2
1.00
0.75
0.50
0.25
0.00
Socs2
Treatment Effect
1
0
−1
−2
R2
1.00
0.75
0.50
0.25
0.00
Serpina6
Treatment Effect
4
3
2
1
0
−1
R2
1.00
0.75
0.50
0.25
0.00
Flat priors
Flat priors (R2<0.5)
Flat priors (R2=0)
Horseshoe
Horseshoe (Null controls only)
Horseshoe (R2=0)
Null Controls


Posterior Median
−2 −1 0 1 2
Gene
Igfbp2
2010002N04Rik
Fam105a
Serpina6
Ccnl2
Gapdh
Irx3
Slc22a3
Vwf
Abca8a
Avpr1a
Apoa4
Ear11
Kdm4a
Mest
Ndrg1
Socs2
Flat priors (R2=0)
Null Controls
Horseshoe (Null controls only)
Flat priors (R2<0.5)
Horseshoe (R2=0)
Flat priors
Horseshoe
Prior

Multi-cause sensitivity in general

$$f_\psi(y \mid do(t)) = f(y \mid t) \int c_{\psi_Y}\left(F_{Y|t}(y), F_{U|t}^{\psi_T}(u) \mid t\right) f(u)du$$

- $c_\psi\left(F_{Y|t}(y), F_{U|t}(u) \mid t\right)$ is the conditional copula which characterizes the dependence between Y and U given T.
  - No information in the observed data about this copula
- $F_{Y|t}$ is the CDF of the naïve outcome model
- $F_{U|t}$ is the CDF of U given T=t. Assumed identifiable from the latent variable model.

Additional Assumptions

Asm4: Copula invariance

The conditional copula does not depend on the value of t, that is, the conditional dependence between Y and U is invariant to the level of T.

Asm5: Gaussian copula

The conditional copula between the outcome and m-dimensional latent confounders given treatments, $c_\psi(F_{Y|t}(y), F_{U|t}(u) \mid t)$, is a Gaussian copula.

Multi-cause sensitivity in general

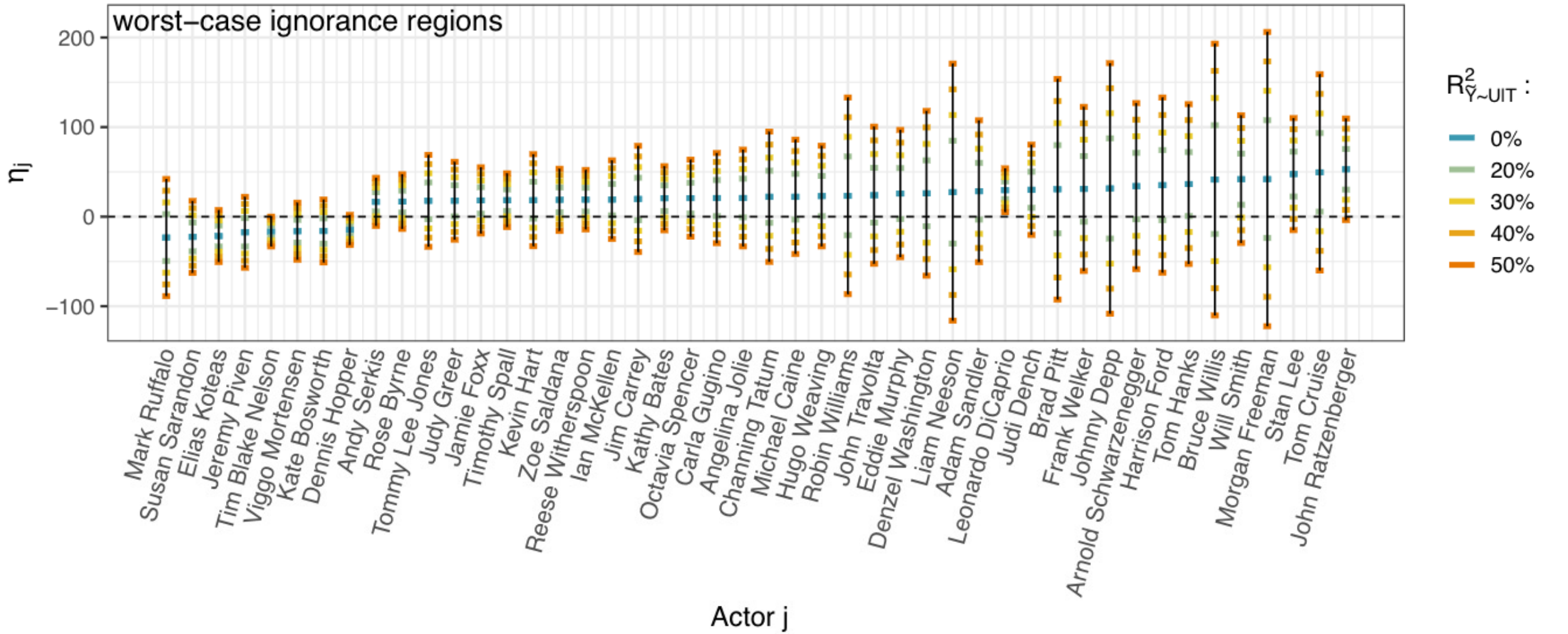
$$
\begin{aligned}
T &\sim F_T \\
f(u \mid t) &\sim N\left(\mu_{u|t}, \Sigma_{u|t}\right) \\
\tilde{Y} &= \gamma'\left(U - \mu_{u|t}\right) + \epsilon_{\tilde{y}|t,u} \\
Y &= F^{-1}_{Y|t}(\Phi(\tilde{Y}))
\end{aligned}
$$

Implied Gaussian copula:

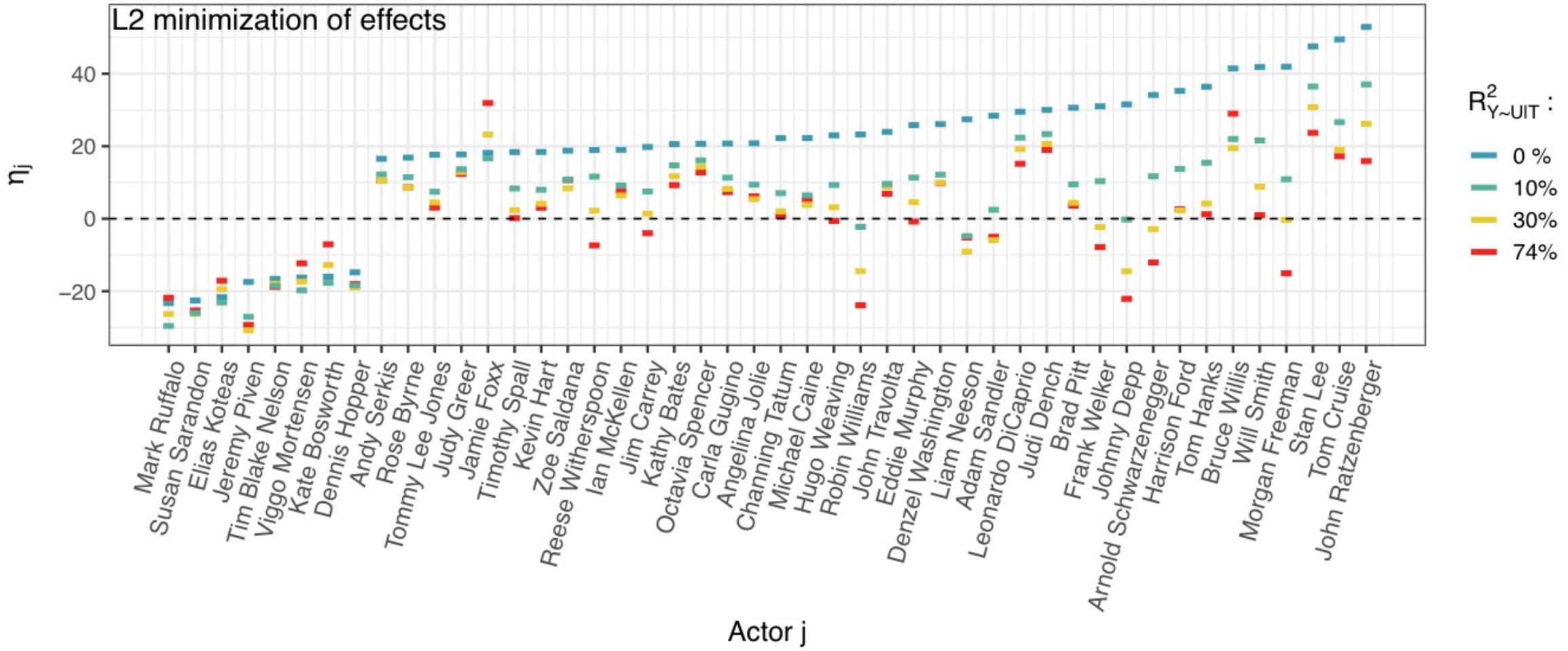
$$
\mathrm{Cov}([\tilde{Y}, U] \mid T = t) = \begin{bmatrix} 1 & \gamma^T \Sigma_{u|t} \\ \Sigma_{u|t}\gamma & \Sigma_{u|t} \end{bmatrix}
$$

Analysis of IMDB movie data

- Analysis of TMDB 5000 Move Dataset
- Estimate the effect of an actor's presence on movie revenue (see Wang and Blei, 2018)
- Regress log revenue of cast indicators
- Explicitly *exclude* observed covariates in order to validate the sensitivity analysis.
    - E.g. budget explains 30% of the variance in revenue


worst−case ignorance regions
200
100
0
−100
η_j
Mark Ruffalo
Susan Sarandon
Elias Koteas
Jeremy Piven
Tim Blake Nelson
Viggo Mortensen
Kate Bosworth
Dennis Hopper
Andy Serkis
Rose Byrne
Tommy Lee Jones
Judy Greer
Jamie Foxx
Timothy Spall
Kevin Hart
Zoe Saldana
Reese Witherspoon
Ian McKellen
Jim Carrey
Kathy Bates
Octavia Spencer
Carla Gugino
Angelina Jolie
Channing Tatum
Michael Caine
Hugo Weaving
Robin Williams
John Travolta
Eddie Murphy
Denzel Washington
Liam Neeson
Adam Sandler
Leonardo DiCaprio
Judi Dench
Brad Pitt
Frank Welker
Johnny Depp
Arnold Schwarzenegger
Harrison Ford
Tom Hanks
Bruce Willis
Will Smith
Morgan Freeman
Stan Lee
Tom Cruise
John Ratzenberger
Actor j
$R^2_{\hat{Y}\sim UIT}$:
0%
20%
30%
40%
50%

Minimum Norm Effects



Conclusions

- If you can identify the distribution of latent confounders given treatments, you get bounded ignorance regions for the causal effects
- Sensitivity analysis allows us to relax strong identifying
- Explore robustness to different kinds of assumptions (R^2, sparsity, null controls, etc)
    - Current work on encoding causal assumptions with Bayesian priors

Software and Future Directions

- Alternative latent variable models
    - Interpretation and calibration is the challenge
- Bayesian inference for encoding (partially) uncheckable causal assumptions.
- Multiple outcomes
- R package available at github.com/JiajingZ/CopSens

Thanks!

Jiajing Zheng
UCSB

Alexander D'Amour
Google Research

Reference: Jiajing Zheng, Alexander D'Amour and Alexander Franks, *Copula-based Sensitivity Analysis for Multi-Treatment Causal Inference with Unobserved Confounding*. **Arxiv: https://arxiv.org/abs/2102.09412**

R package available at github.com/JiajingZ/CopSens

References

D'Amour, A. (2019). On multi-cause approaches to causal inference with unobserved counfounding: Two cautionary failure cases and a promising alternative. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 3478–3486.

Franks, A., A. D'Amour, and A. Feller (2019). Flexible sensitivity analysis for observational studies without observable implications. *Journal of the American Statistical Association* (just-accepted), 1–38.

Grimmer, J., D. Knox, and B. M. Stewart (2020). Naive regression requires weaker assumptions than factor models to adjust for multiple cause confounding. *arXiv preprint arXiv:2007.12702*.

Miao, W., W. Hu, E. L. Ogburn, and X. Zhou (2020). Identifying effects of multiple treatments in the presence of unmeasured confounding.

Ogburn, E. L., I. Shpitser, and E. J. T. Tchetgen (2020). Counterexamples to" the blessings of multiple causes" by wang and blei. *arXiv preprint arXiv:2001.06555*.

Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics 38*(8), 904–909.

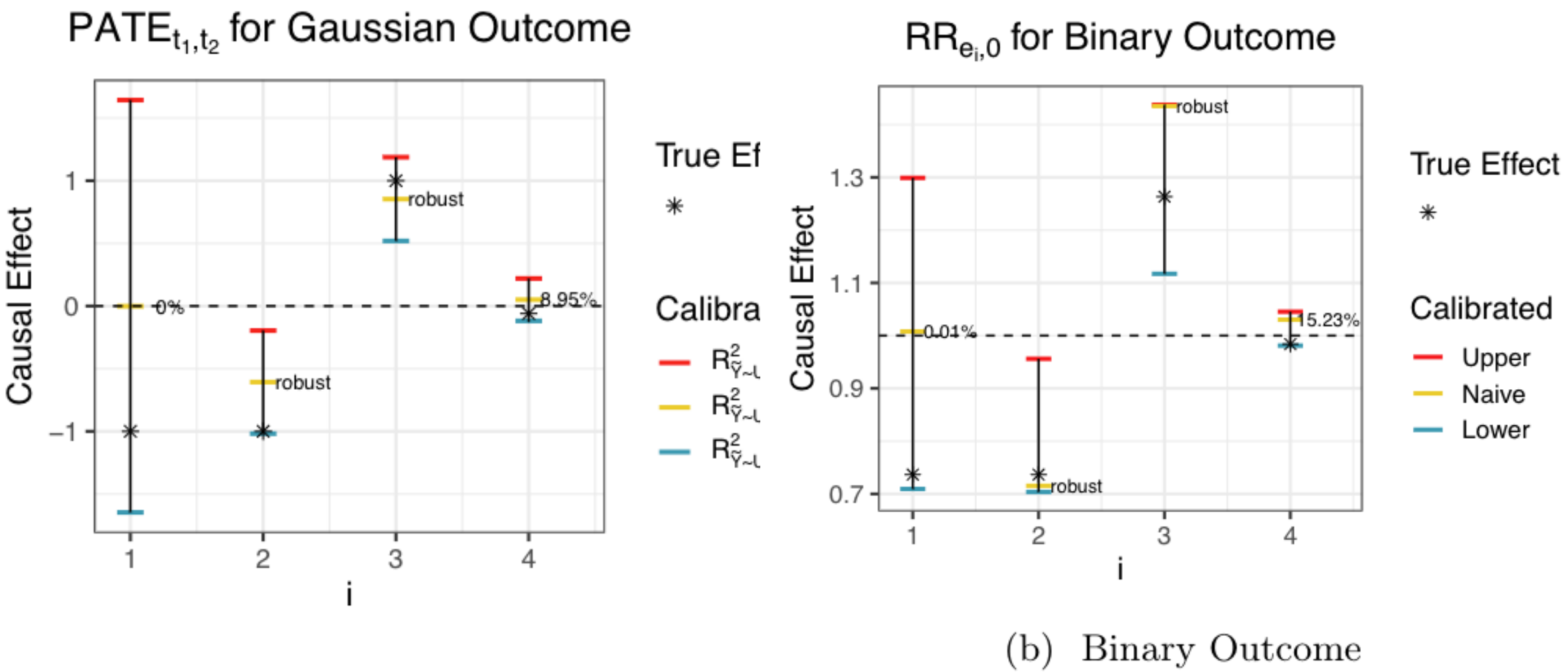
Wang, Y. and D. M. Blei (2018, May). The Blessings of Multiple Causes. *arXiv e-prints*, arXiv:1805.06826.

Binary outcomes and/or binary treatments

- For binary outcomes, focus on the risk ratio:

$$RR_{t_1,t_2} = P\left(Y = 1 \mid do\left(t_1\right)\right) / P\left(Y = 1 \mid do\left(t_2\right)\right)$$

- For non-Gaussian treatments we use variational autoencoder
  - Neural network latent variable model where $f(u \mid t) \sim N(\mu_{u|t}, \Sigma_{u|t})$ holds approximately



(b) Binary Outcome

Observable Partial R-squareds for IMDB example