Flexible Sensitivity Analysis For Observational Studies

Alexander Franks
Department of Statistics, UCSB
May 15, 2019

With Alex D'Amour (Google) and Avi Feller(UC Berkeley)

Observational Causal Inference: Setup

- Use potential outcomes framework (Neyman, 1923; Rubin, 1974).
- Let $Y_i(0)$ and $Y_i(1)$ denote the unit's potential outcomes if assigned to control or treatment, respectively.
- Let T_i denote a binary treatment indicator and X_i denote observed covariates
- For compactness, we often write $Y_i(t), t \in \{0, 1\}$ denote the outcome for treatment level t.

Observational Causal Inference

- Identification of causal effects rests on unconfoundedness:

$$(Y_i(0), Y_i(1) \perp\!\!\!\perp T \mid X$$

- Implies missing equality of observed potential outcome distributions
- Unconfoundedness is unlikely to hold in practice
- Can still make potentially productive statements regarding robustness

The Role of Sensitivity Analysis

The magnitude of the excess lung-cancer risk among cigarette smokers is so great that the results can not be interpreted as arising from an indirect association of cigarette smoking with some other agent or characteristic, since this hypothetical agent would have to be at least as strongly associated with lung cancer as cigarette use, no such agent has been found or suggested.

Cornfield et al (1959)

The Role of Sensitivity Analysis

- Relax unverifiable identifying assumptions
- Readers can assess claims more precisely
- Puts additional focus on the strength of unobserved confounding
- Many approaches to sensitivity since Cornfield*

*Rosenbaum and Rubin (1983); Heckman et al. (1998); Robins (1999); Frank (2000); Rosenbaum (2002); Imbens (2003); Brumback et al. (2004); Hosman et al. (2010); Imai et al. (2010); Vanderweele and Arah (2011); Blackwell (2013); Frank et al. (2013); Dorie et al. (2016); Middleton et al. (2016); VanderWeele and Ding (2017), Oster (2017) and Hazlett et al. (2019) . Thanks to Chad Hazlett for this list.

Model Based Sensitivity Analysis

Assume unconfoundedness conditional on a latent confounder U:

$$(Y_i(0), Y_i(1)) \perp\!\!\!\perp T_i \mid U_i, X_i$$

A common model based approach:

1) Propose a distribution for the latent confounder, U (e.g. binary)
2) Introduce sensitivity parameters for:
    1) The propensity to be treated: T | U, X
    2) The effect of the confounders on the outcome: Y | U, X
3) Assess sensitivity

NHANES data

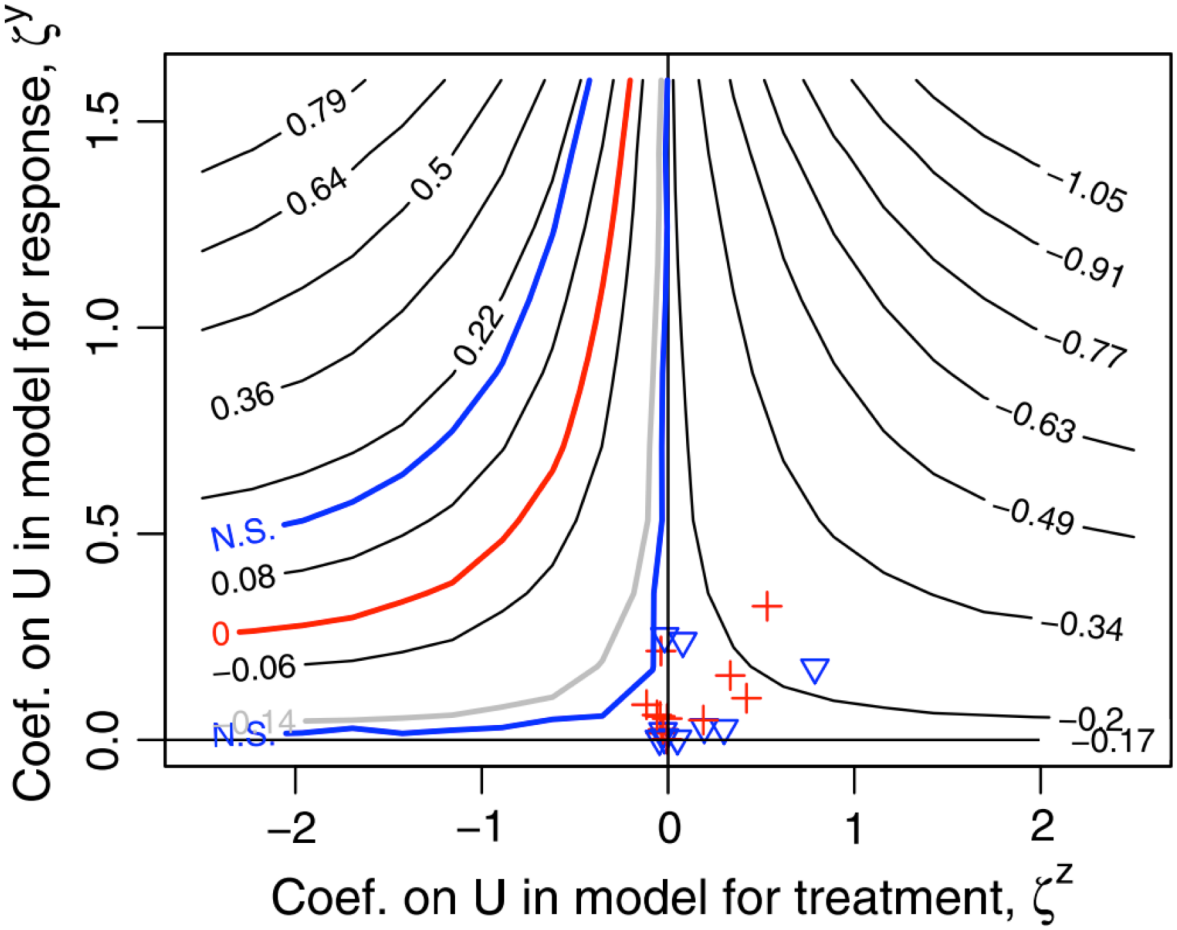
- Analysis of National Health and Nutrition Examination Survey data.
- Explore the effect of taking two or more anti-hypertensives on diastolic blood pressure
- Several pre-treatment covariates:
    - age, race, gender, income and health status

Sensitivity analysis: an example

$$T\,|X, U, \beta \sim \text{Bernoulli}\left(\Phi\left(\beta X + \gamma_t U\right)\right)$$
$$Y|U, \mu_{\text{xt}}, \sigma^2 \sim \text{N}\left(\mu_{\text{xt}} + \gamma_y U, \sigma^2\right)$$

$$\mu_{\text{xt}}, \sigma^2|X, T \sim \text{BART}(X, T)$$
$$U \sim \text{Bernouli}\left(\pi\right)$$

Dorie et al (2016)



Source: Dorie et al (2016)

Implications of latent confounder models:

Complete Data Model:

$$T\,|X, U, \beta \sim \text{Bernoulli}\left(\Phi\left(\beta X + \gamma_t U\right)\right)$$

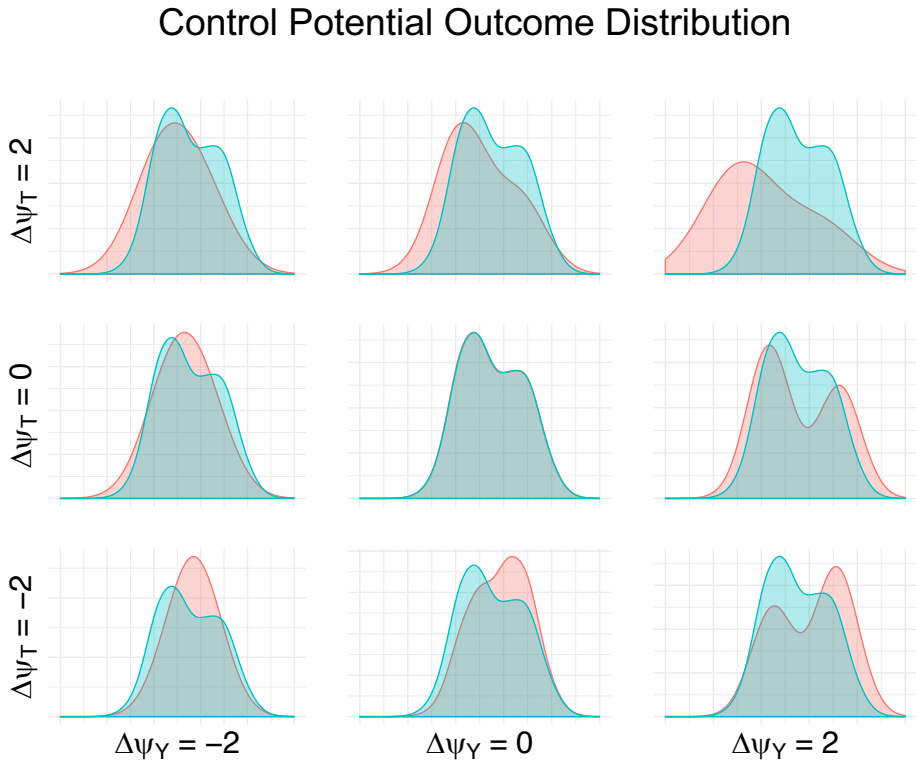
$$Y|U, \mu_{\text{xt}}, \sigma^2 \sim \text{N}\left(\mu_{\text{xt}} + \gamma_y U, \sigma^2\right)$$

Observed Data Density:

$$f^{\text{obs}}(Y(t)|T = t, X) = \int f^{\text{comp}}(Y(t)|T = t, X, U) f(U|T = t) dU$$

$$Y(t)|T = t, X \sim \pi_{\gamma_T} N\left(\mu_{tx} + \gamma_Y, \sigma^2\right) + \left(1 - \pi_{\gamma_T}\right) N\left(\mu_{tx}, \sigma^2\right)$$

Implications of latent confounder models



Implications of latent confounder sensitivity analysis

- Observed data density depends on sensitivity parameters!
- What you do “know” gets combined with what you “don’t know”
- Funny business
- **Our goal:** interpretable sensitivity analysis without observable implications

A practical workflow for sensitivity analysis

- Accommodates a large class of commonly-applied models.
    - Bayesian Additive Regression Trees (BART) for response surface estimation
    - Dirichlet processes mixture (DPM).
    - Semi-continuous data (e.g zero-inflated models)

- Computationally efficient
- Interpretable parameter calibration
- No observable implications

Built on existing methodology for non-ignorable missing data

Non-ignorable Missing Data

- Many ways to factorize a joint distribution
- The selection factorization: $P(Y,T) = P(T|Y)P(Y)$
- The pattern mixture factorization: $P(Y,T) = P(Y|T)P(T)$
- J.W. Tukey, suggested an alternative representation of the joint distribution:

$$P(Y,T) = \frac{P(T|Y)}{P(T=1|Y)} P(Y|T=1)P(T)$$

Tukey’s factorization

- Tukey referred to this as the “simplified selection model” (1986)
- “Exponential tilting” (Scharfstein 1999, Birmingham 2003), non-parametric (just) identified (NPI) models (Robins, 2000), “extrapolation factorization” (Linero, 2017)
- The joint data is expressed as product of the observed data (identified) and the treatment assignment mechanism (unidentified, easy to reason about)

The missing data distribution

- Consider a single scalar potential outcome Y(t)
- Y(t) is observed when T=t and missing when T=1-t
- Tukey's factorization yields the missing potential outcome density as a function of the observed data density and non-ignorable treatment assignment:

$$f^{\text{mis}}_{\gamma,t}(Y(t) \mid T = 1\text{-}t, X) \propto \frac{f_\gamma(T = t \mid Y(t), X)}{f_\gamma(T = 1\text{-}t \mid Y(t), X)} f^{\text{obs}}_t(Y(t) \mid T = t, X).$$

Logistic-Exponential Family Models

- Assume an exponential family model for the observed potential outcomes'

$$f_t^{\text{obs}}(Y(t)|T=t,X) = h(Y(t))g\left(\eta_t(X)\right)e^{s(Y(t))'\eta_t(X)}$$

- Are the sufficient statistics, are the natural parameters and g() is the normalizing constant.

- Assume treatment is logistic in the (potentially unobserved) outcome

$$f_{\gamma_t}(T=1|Y(t),X) = \text{logit}^{-1}\left\{\alpha_t(X) + \gamma_t' s_t(Y(t))\right\}$$

Logistic-Exponential Family Models

- Assume an exponential family model for the observed potential outcomes
- Then, missing outcome distributions have the same exponential family distribution

$$f_t^{\text{mis}}(Y(t)|T = 1\text{-}t, X) = h(Y(t))g\left(\eta_t^*(X)\right)e^{s(Y(t))'\eta_t^*(X)}$$

where $\eta_t^* = \eta_t + \gamma_t$

Extension to Mixtures

- Assume the observed data is well approximated by a mixture of EF distributions (all from the same family)

$$f_t^{\text{obs}}\left(Y(t)|\eta_t, T=t\right) = \sum_k \pi_k h_k(Y(t)) g_k\left(\eta_{tk}\right) e^{s(Y(t))'\eta_{tk}}$$

- Then, missing outcome distributions have the same exponential family distribution

$$f_{\gamma_t,t}^{\text{mis}}\left(Y(t)|\eta_t, \gamma_t, T=1-t\right) = \sum_k \pi_k^* h_k(Y(t)) g_k\left(\eta_{tk}^*\right) e^{s(Y(t))'\eta_{tk}^*}$$

$$\pi_k^* = \frac{\pi_k \frac{g(\eta_k)}{g\left(\eta_k^*\right)}}{\sum_k^K \pi_k \frac{g(\eta_k)}{g\left(\eta_k^*\right)}}$$

Example: Normal Mixtures

- Assume the observed data is well approximated by a mixture of normal:

$$f_{\gamma_t}(T = 1|Y(t), X) = \text{logit}^{-1}\left\{\alpha_t(X) + \gamma_t' s_t(Y(t))\right\}$$

$$f_t^{\text{obs}}(Y(t)|T = t, X) \sim \sum \pi_k \text{N}\left(\mu_{tk}(X), \sigma_{tk}^2\right)$$

- Then, missing outcome distributions have the same exponential family

$$f_{\gamma_t,t}^{\text{mis}}(Y(t)|T = 1 - t, X) \sim \sum \pi_k^* \text{N}\left(\mu_{tk}(X) + \gamma_t\sigma_{tk}^2, \sigma_{tk}^2\right)$$

$$\pi_k^* \propto \pi_k \exp\frac{1}{2}\left(\frac{\mu_{tk}^2(X)}{\sigma_{tk}^2} - \left(\frac{\mu_{tk}(X)}{\sigma_{tk}} - \gamma_t\right)^2\right)$$

Tukey’s factorization for causal inference

- Don’t explicitly introduce latent confounders
- Focus directly on relationship between potential outcomes and treatment
- Apply Tukey’s factorization to the triplet $(T, [Y(0), Y(1)])$
- Leverage logistic-exponential family results

Tukey's factorization in causal inference

$$f(T,[Y(0),Y(1)]) = \frac{f(Y(0),T)}{f(T)} \cdot \frac{f(Y(1),T)}{f(T)} \cdot f(T) \cdot \frac{f(Y(0),Y(1)|T)}{f(Y(0)|T)f(Y(1)|T)}$$

$$= f(Y(0)|T=0)f(T=0) \cdot \frac{f(T|Y(0))}{f(T=0|Y(0))} \cdot$$

$$\cdot f(Y(1)|T=1)f(T=1) \cdot \frac{f(T|Y(1))}{f(T=1|Y(1))} \cdot \frac{1}{f(T)}$$

$$c(F(Y(0)|T), F(Y(1)|T)|T)$$

Applicability of Tukey's factorization

- **Condition 1:** (Integral constraint)

$$\int_{\mathcal{Y}} f^{\text{obs}}(Y(t) \mid T = t, X) \cdot \frac{f_\psi(T = 1 - t \mid Y(t), X)}{f_\psi(T = t \mid Y(t), X)} dY(t) = \frac{f(T = 1 - t \mid X)}{f(T = t \mid X)}.$$

- **Condition 2:** Outcome Overlap

$$P(Y(t) \in A \mid T = 1 - t, X) > 0 \Rightarrow P(Y(t) \in A \mid T = t, X) > 0$$

Marginal Contrast Estimands

$$\tau^{ATE} = E[Y(1) \mid X] - E[Y(0) \mid X]$$ (Average Treatment Effect)

$$\tau^{CATE} = E[Y(1) \mid X] - E[Y(0) \mid X]$$ (Conditional Average Treatment Effect)

$$\tau_q = Q_q(Y(1)) - Q_q(Y(0))$$ (Quantile q Treatment Effect)

- Marginal contrast estimands are uniquely defined by the assignment $f_\gamma(T \mid Y(t), X)$
- Marginal contrast estimands are invariant to the specification of the copula

The treatment assignment function

- Leverage nice properties logistic models

$$f(T = 1 \mid Y(t), X) = \text{logit}^{-1}\{\alpha_t(X) + \gamma_t' s_t(Y(t))\},$$

- Often reasonable to assume monotonicity in Y(t)

- Two sensitivity parameters, γ_0 and γ_1

Interpretation the sign of sensitivity parameters

- Specify how sufficient statistics of the potential outcomes are over- or under-represented among observed control and treated units.

- $\gamma_1 > 0$ implies that units with large $Y(1)$ are over-represented among treated units
    - ATT is positively biased when assuming unconfoundedness

- $\gamma_0 > 0$ implies that units with large $Y(0)$ are over-represented among treated units
    - ATC is negatively biased when assuming unconfoundedness

Interpretation the sign of sensitivity parameters

- Same sign:
  - Treatment and control group means are biased in opposite directions, ATE changes rapidly
- Opposite signs
  - Treatment and control group biases partially cancel out, effect on ATE is gradual
- Single arm confounding
  - Only one of the group means is biased, ATE change is moderate with change in nonzero sensitivity parameter

Calibrating the Magnitude of Sensitivity Parameters

- Collinearity makes it difficult to reason about sensitivity parameters using regression coefficients.

- Calibrate the magnitude of sensitivity parameters to the amount of variation in the treatment assignment T that is explained by Y(t), *above and beyond what is accounted for by X.*

- Use ”implicit R-squared” (Imbens, 2003) for measuring variance explained in logistic regression

Implicit R-squared for Logistic Models

$$Z = m(X) + \epsilon \text{ with } \epsilon \sim \text{Logistic}(0, 1)$$

$$T = \begin{cases} 0 \text{ if } Z < 0 \\ 1 \text{ if } Z \geq 0 \end{cases}$$

$$\rho_X^2 = \frac{\text{Var}(m(X))}{\text{Var}(m(X)) + \pi^2/3}$$

Partial Implicit R-Squared

- Implicit R-squared estimable for observed covariates: $\rho^2_X = \frac{\mathrm{Var}(m(X))}{\mathrm{Var}(m(X)) + \pi^2/3}$

- Partial R-squared:

$$\rho^2_{Y(t)|X} = \frac{\rho^2_{X,Y(t)} - \rho^2_X}{1 - \rho^2_X}$$

- Anchor against partial R-squared for observed covariates:

$$\rho^2_{X_j|X_{-j}} = \frac{\rho^2_X - \rho^2_{X_{-j}}}{1 - \rho^2_{X_{-j}}}$$

Calibration

$$\rho^2_{Y(t)|X} = \frac{\sigma^2_{rt}\gamma^2_t}{\mathrm{Var}(m(X)) + \pi^2/3 + \sigma^2_{rt}\gamma^2_t},$$

- Fix a plausible value partial variance in treatment explained by the outcome
- Solve for the implied sensitivity parameter γ_t

Calibration

$$|\gamma_t| = \frac{1}{\sigma_{rt}} \sqrt{\frac{\rho^2_{Y(t)|X}}{1 - \rho^2_{Y(t)|X}} (\mathrm{Var}(m(X)) + \pi^2/3)}.$$

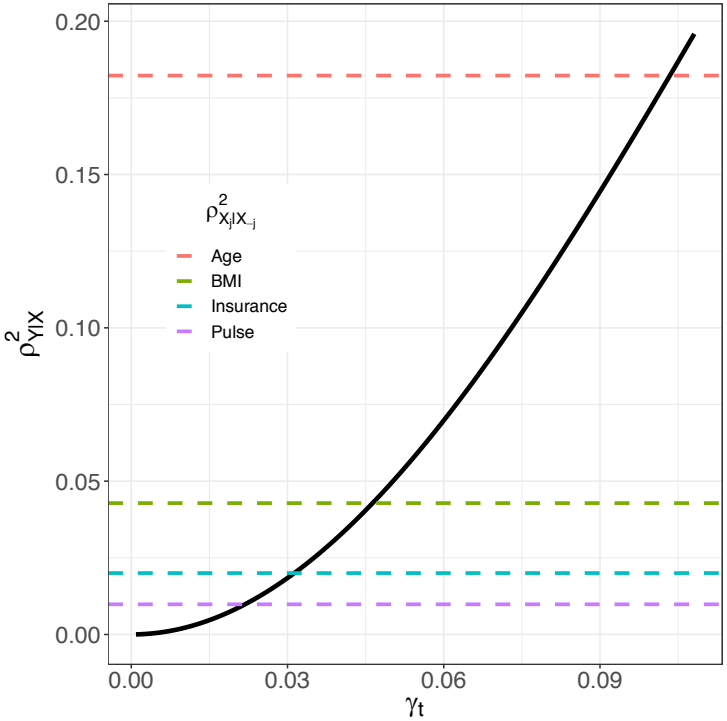
- σ_{rt} is the complete data residual standard deviation

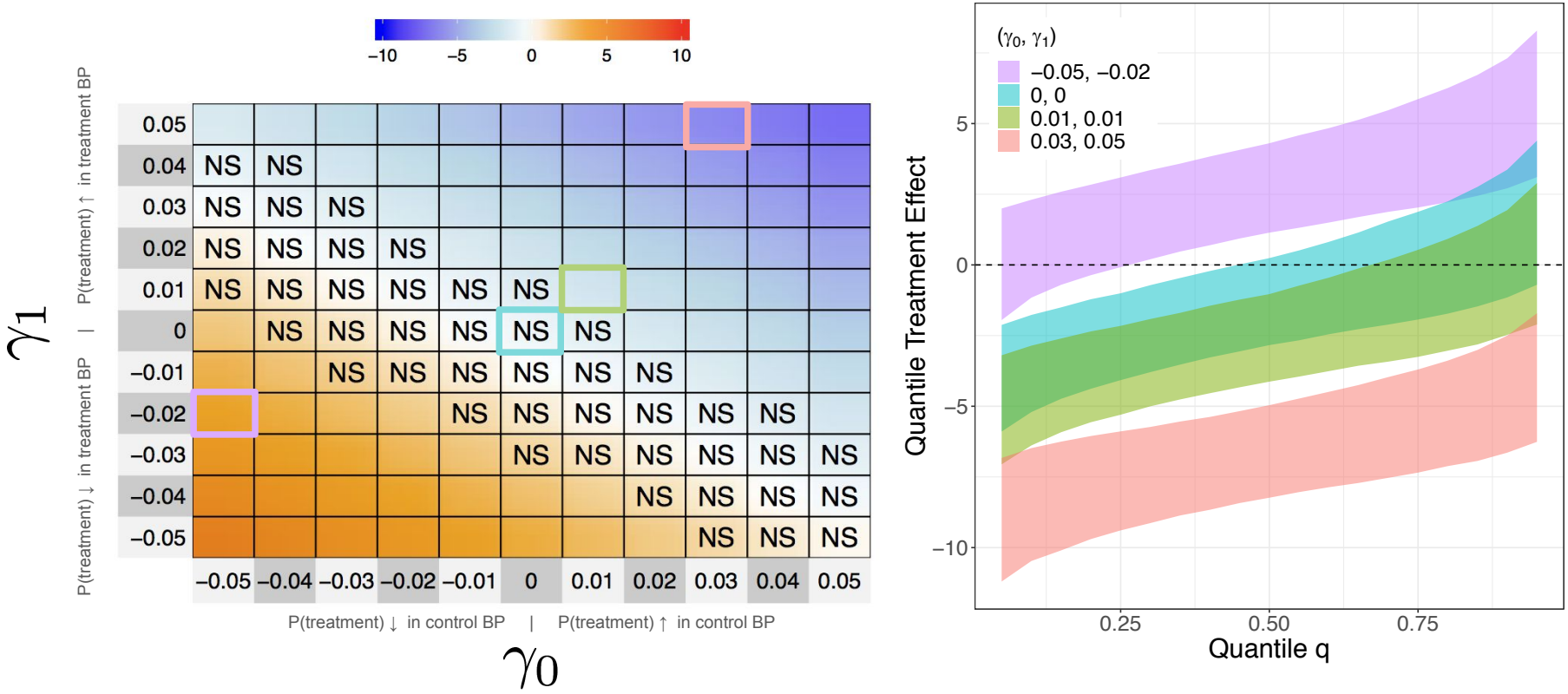
- In logistic-EF models we have analytic solution to this variance as a function of
- Simple to solve a system of equations

Analysis of NHANES data

- Third National Health and Nutrition Examination Survey (NHANES)
- III) a comprehensive survey of Americans' health and nutritional status.
- We follow the same set up as Dorie et al. (2016), and utilize pre-treatment covariates
    - Race, Gender, Age, income BMI, and whether the patient was insured

Calibration in NHANES analysis




−10
−5
0
5
10
0.05
0.04
0.03
0.02
0.01
0
−0.01
−0.02
−0.03
−0.04
−0.05
NS
−0.05
−0.04
−0.03
−0.02
−0.01
0
0.01
0.02
0.03
0.04
0.05
P(treatment) ↓ in treatment BP | P(treatment) ↑ in treatment BP
P(treatment) ↓ in control BP | P(treatment) ↑ in control BP
γ_1
γ_0
(γ_0, γ_1)
−0.05, −0.02
0, 0
0.01, 0.01
0.03, 0.05
Quantile Treatment Effect
Quantile q
5
0
−5
−10
0.25
0.50
0.75

Job Training Partnership Act

- Analysis of Job Training and Partnership Act (JTPA) evaluation,
  - a large randomized trial estimating the impact of workforce development programs on wages.
  - compare outcomes between those who choose to participate and those who did not
- Record employment status, W, and income, Y.
  - Y=0 when W=0
- Estimate quantile treatment effects (see e.g. Xu *et al* 2018)

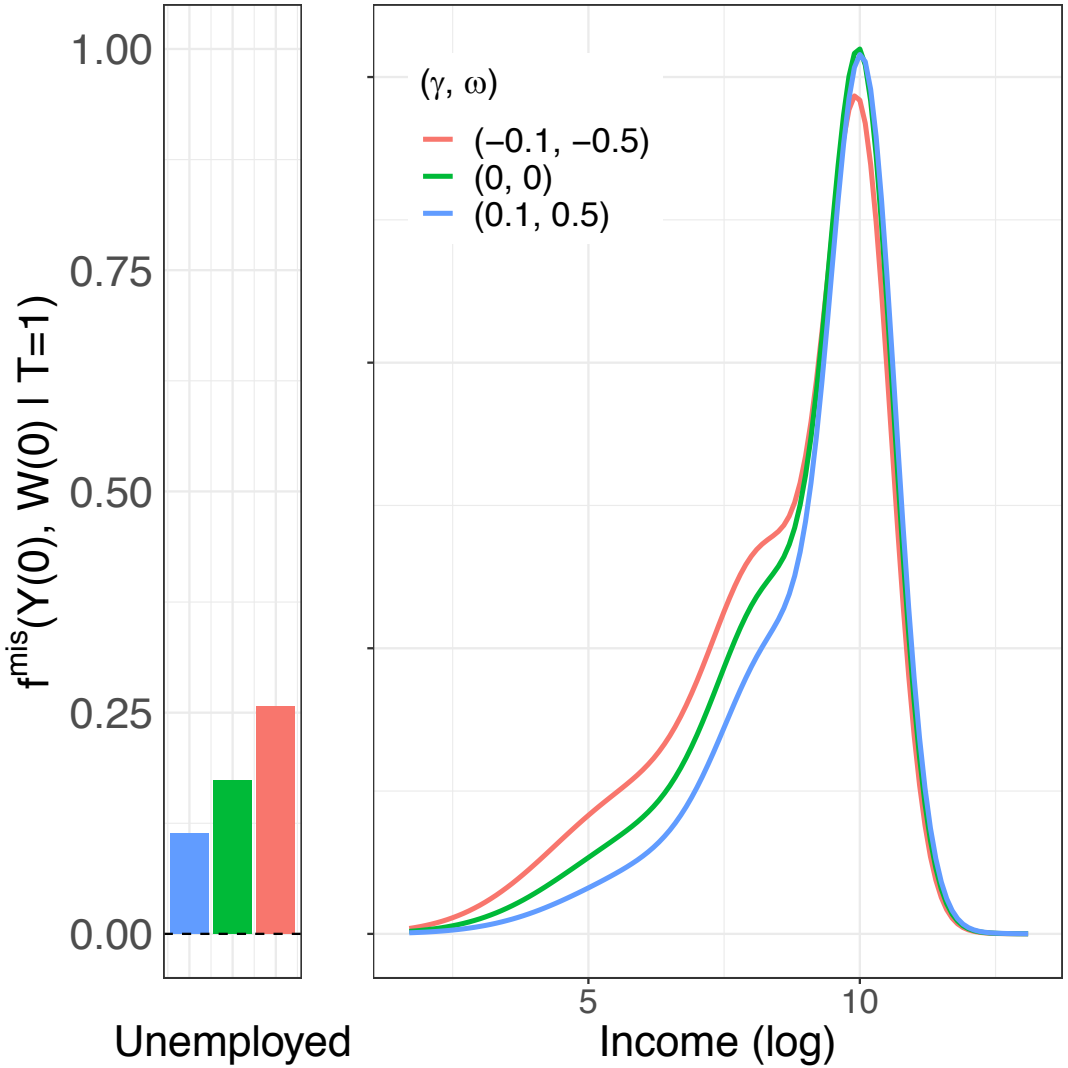
Job Training Partnership Act

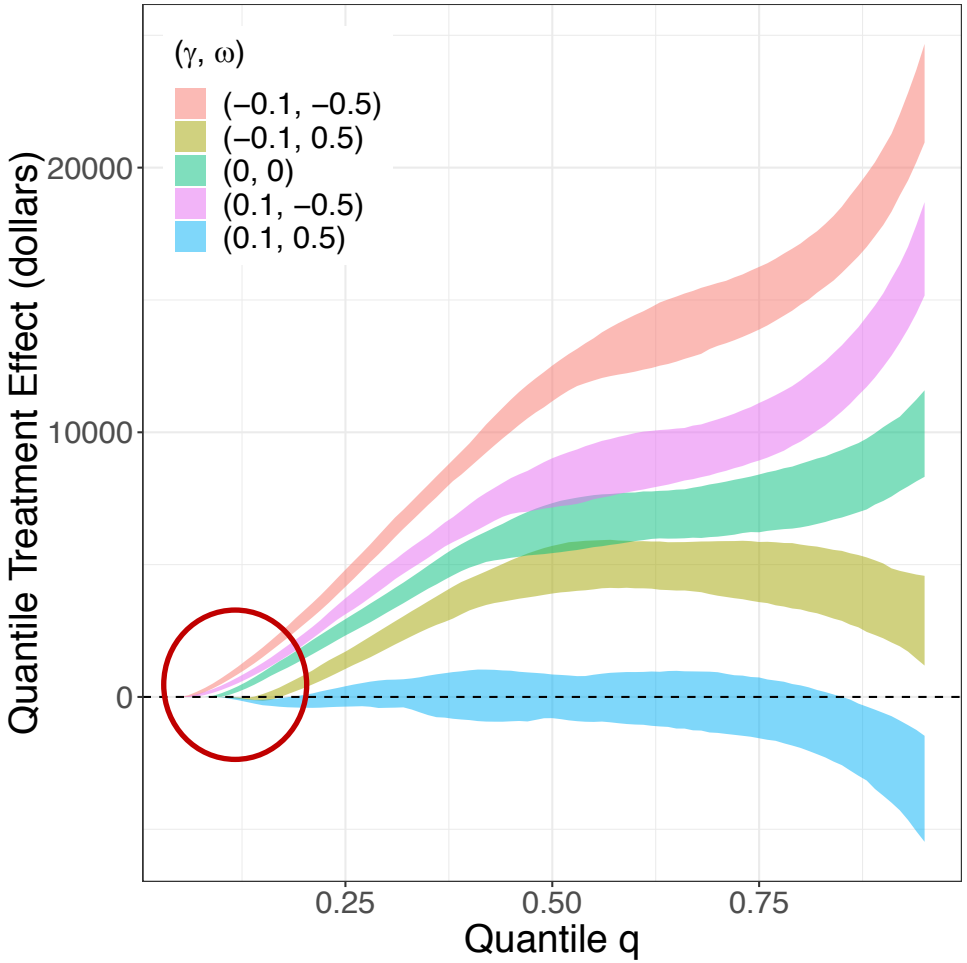
- Model the log-income distribution with a Dirichlet Process Mixture of Normals:

$$f_t^{\text{obs}}\left(\log\left(Y_i(t)\right)|T=t,W=1\right) \sim N\left(\mu_{it},\sigma_{it}^2\right)$$
$$\left(\mu_{it},\sigma_{it}\right) \sim G$$
$$G \sim DP\left(\alpha G_0\right)$$

- Treatment assignment is logistic log-income, log(Y) given employment and logistic in employment indicator, W.

$$f(T=1|\log(Y(t)),W(t)=1) \sim \text{logit}^{-1}\left(\beta_t+\gamma_t\log(Y(t))\right)$$
$$f(T=1|W(t)=0) \sim \text{logit}^{-1}\left(\alpha_t+\omega_t I\{W(t)=0\}\right)$$


1.00
0.75
0.50
0.25
0.00
$f^{mis}(Y(0), W(0) \mid T=1)$
Unemployed
(γ, ω)
(−0.1, −0.5)
(0, 0)
(0.1, 0.5)
5
10
Income (log)


(γ, ω)
(−0.1, −0.5)
(−0.1, 0.5)
(0, 0)
(0.1, −0.5)
(0.1, 0.5)
Quantile Treatment Effect (dollars)
20000
10000
0
0.25
0.50
0.75
Quantile q

Summary

- Flexible and interpretable model based sensitivity analysis
- Can be applied to *post facto* to many models in which unconfoundedness was previously assumed
    - Computationally efficient, don’t necessary need to re-fit
- No observable implications. Independent of model selection.

Ongoing

- R package development in progress: *tukeySens*
- Returning to latent variable models
  - $f^{\text{obs}}(Y(t) \mid T = t, X) = \int f^{\text{comp}}(Y(t) \mid T = t, X, U) f_\gamma(U \mid T) dU$
  - GLMM approach
- Sensitivity analysis in the “multiple treatments” setting
  - GWAS an important application
  - See D’Amour (2019)

Thank you!

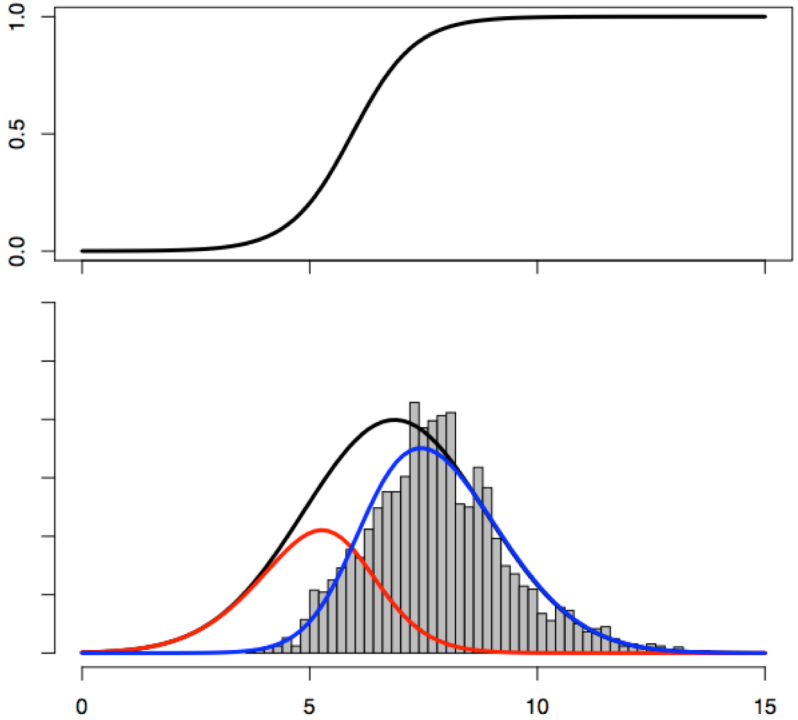
Alex D'Amour
Google Brain

Avi Feller
UC Berkeley

Reference:

Alex Franks, Alex D'Amour & Avi Feller (2019) Flexible sensitivity analysis for observational studies without observable implications*, Journal of the American Statistical Association, DOI: 10.1080/01621459.2019.1604369

The selection factorization



Tukey’s factorization