Sensitivity Analysis

Causal inference in many situations is UNSATISFYING!

- Most of the time we don't have a cool natural experiment (for an instrument or rd for example) or the natural experiment is flawed
- Most of the time ignorability isn't terribly convincing (a referee can always think of a confounder you haven't controlled for)
- How can we make our inferences more convincing???

Sensitivity analysis: If I'm wrong… how far off might I be?

- The problem with almost all causal inference settings is that we never know whether our assumptions are satisfied
- *Sensitivity analysis* is an umbrella term that encapsulates any of a variety of different methods that assess the degree to which our inferences might be altered by changes in structural or parametric assumptions.

Classic example: smoking and cancer

- In the 1950's the empirical *association* between smoking and lung cancer was clear
- However those arguing against a causal interpretation posited that a third variable might be (at least partially) responsible both for the desire to smoke and the probability of getting lung cancer

Cornfield's sensitivity analysis

- Cornfield responded by focusing on this potential omitted variable
- What would that variable have to look like (in particular how *strongly* would it have to be associated with both smoking and lung cancer) in order for it to explain away the association between the two?
- It turns out that it would have to be *very strong*
  - it would have to be an almost perfect predictor of lung cancer
  - it would have to be about 9 times more prevalent in smokers than non-smokers
- This level of association was implausible for any factor that anyone could think of

Sensitivity Analysis: Types of approaches

Classically there have been three main classes of approaches

1. *Primal methods*: Specify relationship between U and Z; assume U and Y are essentially collinear (e.g. Rosenbaum, 1987)
2. *Dual methods*: Specify relationship between U and Y; assume U and Z are essentially collinear (e.g. Manski, 1990, 1995)
3. *Simultaneous methods*: Specify relationship between U and Z; specify relationship between U and Y (e.g. Imbens, 2003, Harada, 2012)

* Labels from Gastwirth (1999)

Sensitivity Analysis: “Primal” approach

"Primal" (one parameter) sensitivity analysis for matched pairs

Rosenbaum motivates his method for matched pairs in the following way:

- If a study is free from hidden bias (that is, if ignorability is satisfied) then if two units have the same values on all their measured confounders ($\mathbf{x}_i=\mathbf{x}_j$) they should also have the same probability of receiving the treatment ($\pi_i=\pi_j$)
- Suppose this is not true at a given level of $\mathbf{x}$… How large would the difference in probabilities have to be to alter the qualitative conclusions of a study?

Sensitivity analysis for matched pairs

- Denote the probability of being assigned treatment for the j^{th} member of the i^{th} pair as π_{ij}.
- Since one member of every pair must be assigned to treatment, but both can't be, $\pi_{i1} + \pi_{i2} = 1$.
- In a matched pairs randomized experiment or matched study with no hidden bias, $\pi_{i1}=\pi_{i2}$.
- In a matched pairs study with hidden bias $\pi_{i1}\neq\pi_{i2}$.
- What if we assume that matched pairs may differ in their chances (odds) of receiving the treatment by at most a factor of Γ as expressed by

$$\frac{1}{\Gamma} \leq \frac{\pi_{i1}/(1-\pi_{i1})}{\pi_{i2}/(1-\pi_{i2})} \leq \Gamma$$

Back to sensitivity analysis for matched pairs

- Denote the probability of being assigned treatment for the j^{th} member of the i^{th} pair as π_{ij}.
- Since one member of every pair must be assigned to treatment, but both can't be, $\pi_{i1} + \pi_{i2} = 1$.
- In a matched pairs randomized experiment or matched study with no hidden bias, $\pi_{i1}=\pi_{i2}$.
- In a matched pairs study with hidden bias $\pi_{i1}\neq\pi_{i2}$.
- What if we assume that matched pairs may differ in their chances (odds) of receiving the treatment by at most a factor of Γ as expressed by

$$\frac{1}{\Gamma} \leq \frac{\pi_{i1}/(1-\pi_{i1})}{\pi_{i2}/(1-\pi_{i2})} \leq \Gamma$$

Note that if Γ=1 this implies that the study is free from hidden bias

Back to sensitivity analysis for matched pairs

- Denote the probability of being assigned treatment for the j^{th} member of the i^{th} pair as π_{ij}.
- Since one member of every pair must be assigned to treatment, but both can't be, $\pi_{i1} + \pi_{i2} = 1$.
- In a matched pairs randomized experiment or matched study with no hidden bias, $\pi_{i1}=\pi_{i2}$.
- In a matched pairs study with hidden bias $\pi_{i1}\neq\pi_{i2}$.
- What if we assume that matched pairs may differ in their chances (odds) of receiving the treatment by at most a factor of Γ as expressed by

$$\frac{1}{\Gamma} \leq \frac{\pi_{i1}/(1-\pi_{i1})}{\pi_{i2}/(1-\pi_{i2})} \leq \Gamma$$

Note that if Γ=1 this implies that the study is free from hidden bias

- We can also think of Γ as the coefficient on the unobserved confounder in a logistic regression of the treatment on the unobserved confounder

RB inference: matched pairs, hidden bias of Γ=2.25, testing τ=0

$$\frac{1}{n}\sum_j \left((Y_{2j} - Y_{1j})Z_{2j} + (Y_{1j} - Y_{2j})Z_{1j}\right) = 4.1$$

Pair	Z_j	(π_{1j},π_{2j})	(Y_{1j},Y_{2j})	$(Y(0)_{1j},Y(0)_{2j})$	$(Y(1)_{1j},Y(1)_{2j})$
1	(0,1)	(.4,.6)	(4.2,8)	(4.2,8)	(4.2,8)
2	(0,1)	(.4,.6)	(5,9.1)	(5,9.1)	(5,9.1)
3	(0,1)	(.4,.6)	(6.2,10)	(6.2,10)	(6.2,10)
4	(0,1)	(.4,.6)	(4,8.3)	(4,8.3)	(4,8.3)
5	(0,1)	(.4,.6)	(6,9.6)	(6,9.6)	(6,9.6)
6	(0,1)	(.4,.6)	(5.6,10)	(5.6,10)	(5.6,10)
7	(0,1)	(.4,.6)	(5,9.2)	(5,9.2)	(5,9.2)
8	(0,1)	(.4,.6)	(4.3,8)	(4.3,8)	(4.3,8)
9	(0,1)	(.4,.6)	(4,8.1)	(4,8.1)	(4,8.1)
10	(0,1)	(.4,.6)	(6,10.2)	(6,10.2)	(6,10.2)

RB inference: matched pairs, hidden bias of Γ=2.25, testing τ=0

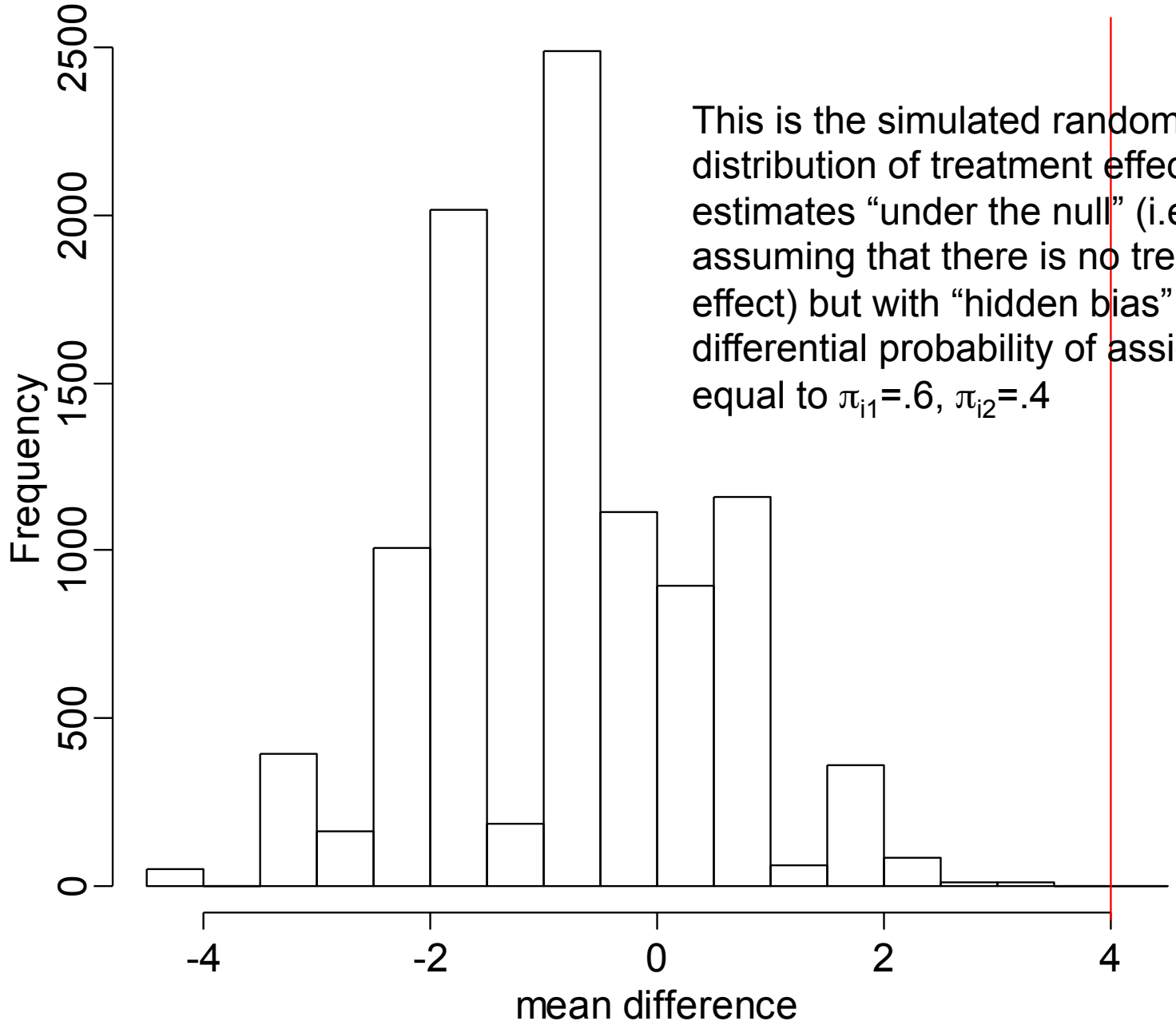
$$\frac{1}{n}\sum_j \left((Y_{2j}-Y_{1j})Z_{2j}+(Y_{1j}-Y_{2j})Z_{1j}\right)=4.1$$

Pair	Z_j	(π_{1j},π_{2j})	(Y_{1j},Y_{2j})	$(Y(0)_{1j},Y(0)_{2j})$	$(Y(1)_{1j},Y(1)_{2j})$
1	(0,1)	(.4,.6)	(4.2,8)	(4.2,8)	(4.2,8)
2	(0,1)	(.4,.6)	(5,9.1)	(5,9.1)	(5,9.1)
3	(0,1)	(.4,.6)	(6.2,10)	(6.2,10)	(6.2,10)
4	(0,1)	(.4,.6)	(4,8.3)	(4,8.3)	(4,8.3)
5	(0,1)	(.4,.6)	(6,9.6)	(6,9.6)	(6,9.6)
6	(0,1)	(.4,.6)	(5.6,10)	(5.6,10)	(5.6,10)
7	(0,1)	(.4,.6)	(5,9.2)	(5,9.2)	(5,9.2)
8	(0,1)	(.4,.6)	(4.3,8)	(4.3,8)	(4.3,8)
9	(0,1)	(.4,.6)	(4,8.1)	(4,8.1)	(4,8.1)
10	(0,1)	(.4,.6)	(6,10.2)	(6,10.2)	(6,10.2)

Now perform a simulation that re-randomizes treatment assignment within pair with the probabilities in the third column. These make it more likely that the second observation in each pair is randomized to the treatment. This simulation creates a "randomization distribution" (similar to a sampling distribution).

Testing $\tau=0$

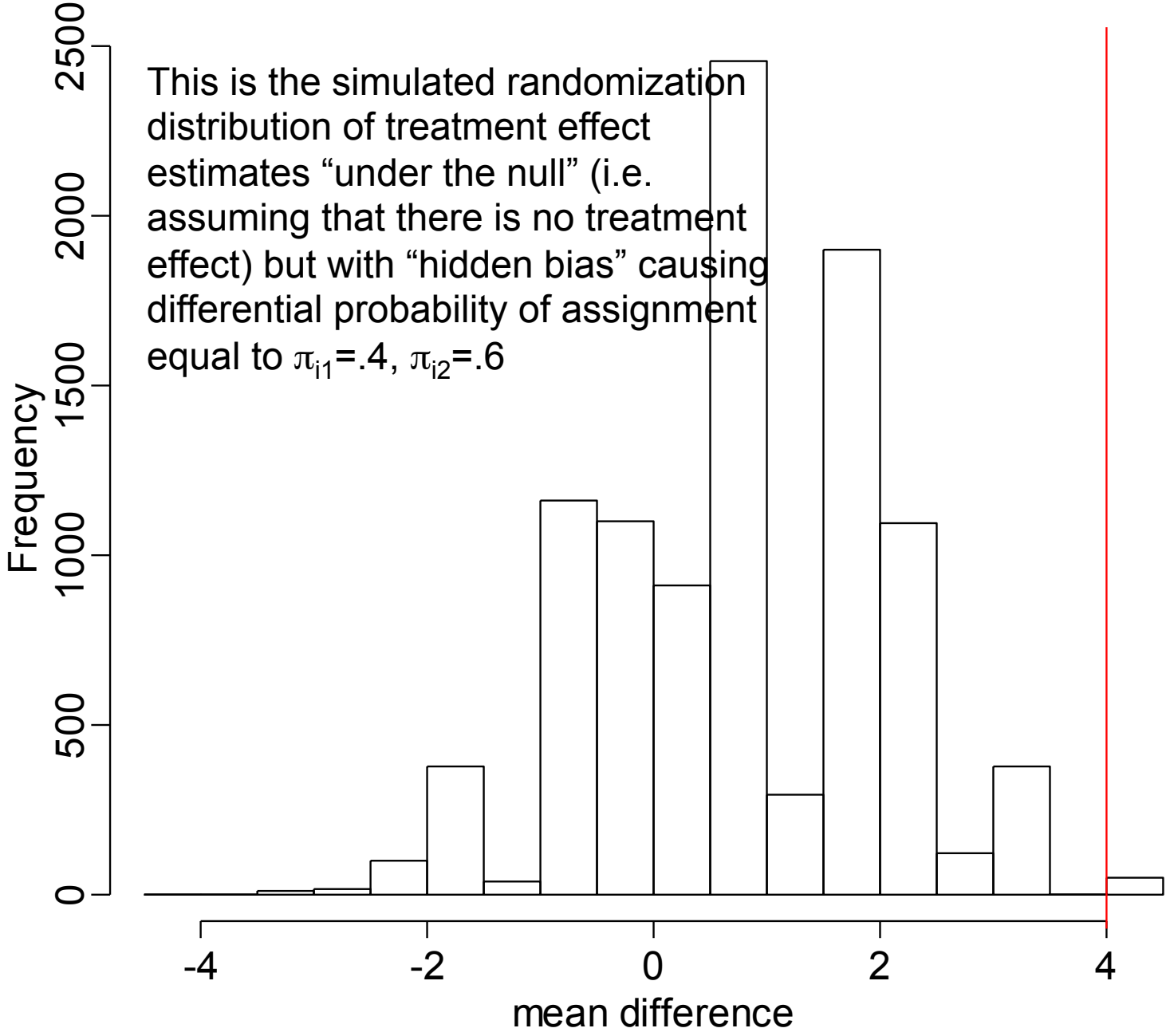
Gamma=2.25, lower bound



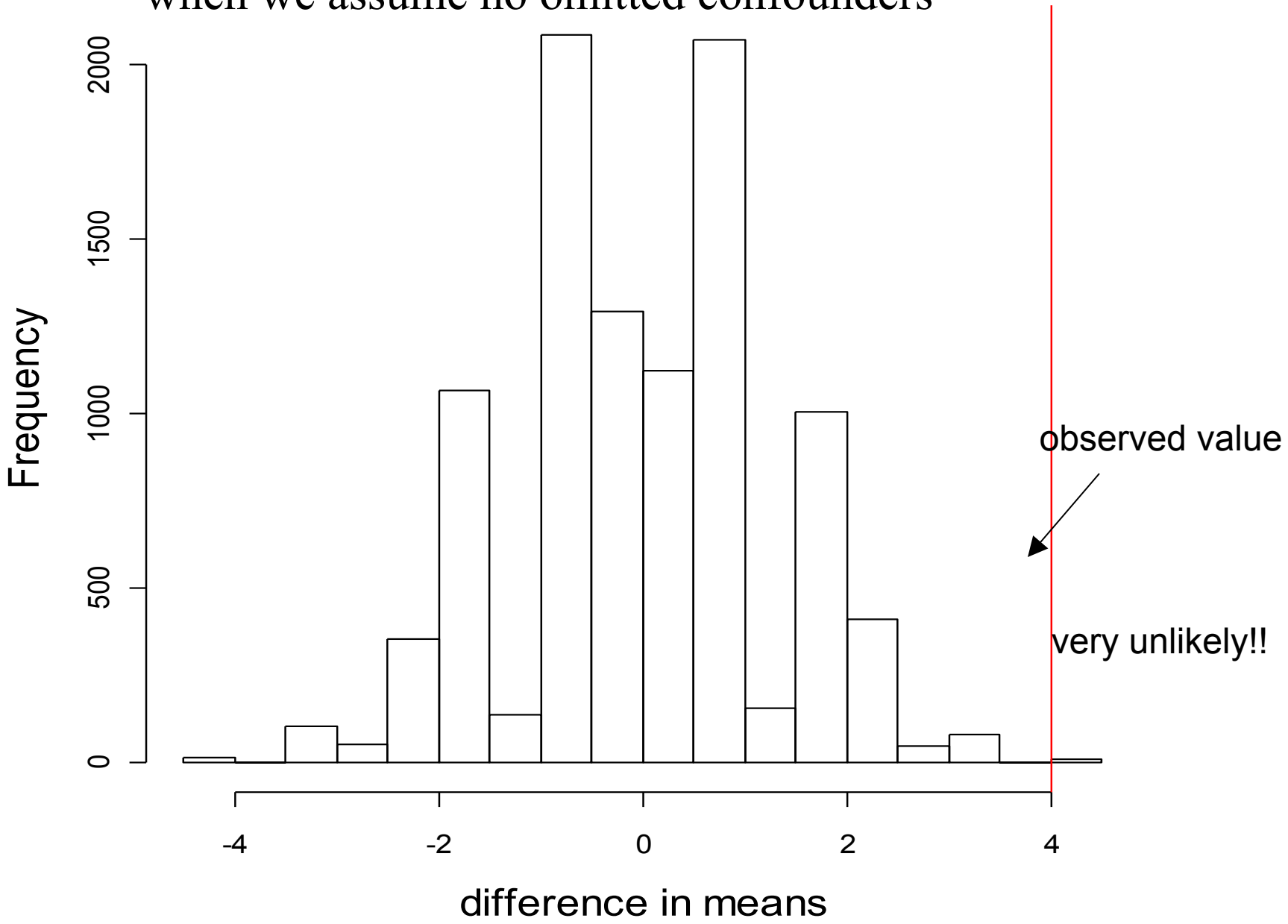
This is the simulated randomization distribution of treatment effect estimates "under the null" (i.e. assuming that there is no treatment effect) but with "hidden bias" causing differential probability of assignment equal to $\pi_{i1}=.6$, $\pi_{i2}=.4$

Testing $\tau=0$

Gamma=2.25, upper bound



This is what the randomization distribution looks like for H_0: $\tau=0$ when we assume no omitted confounders

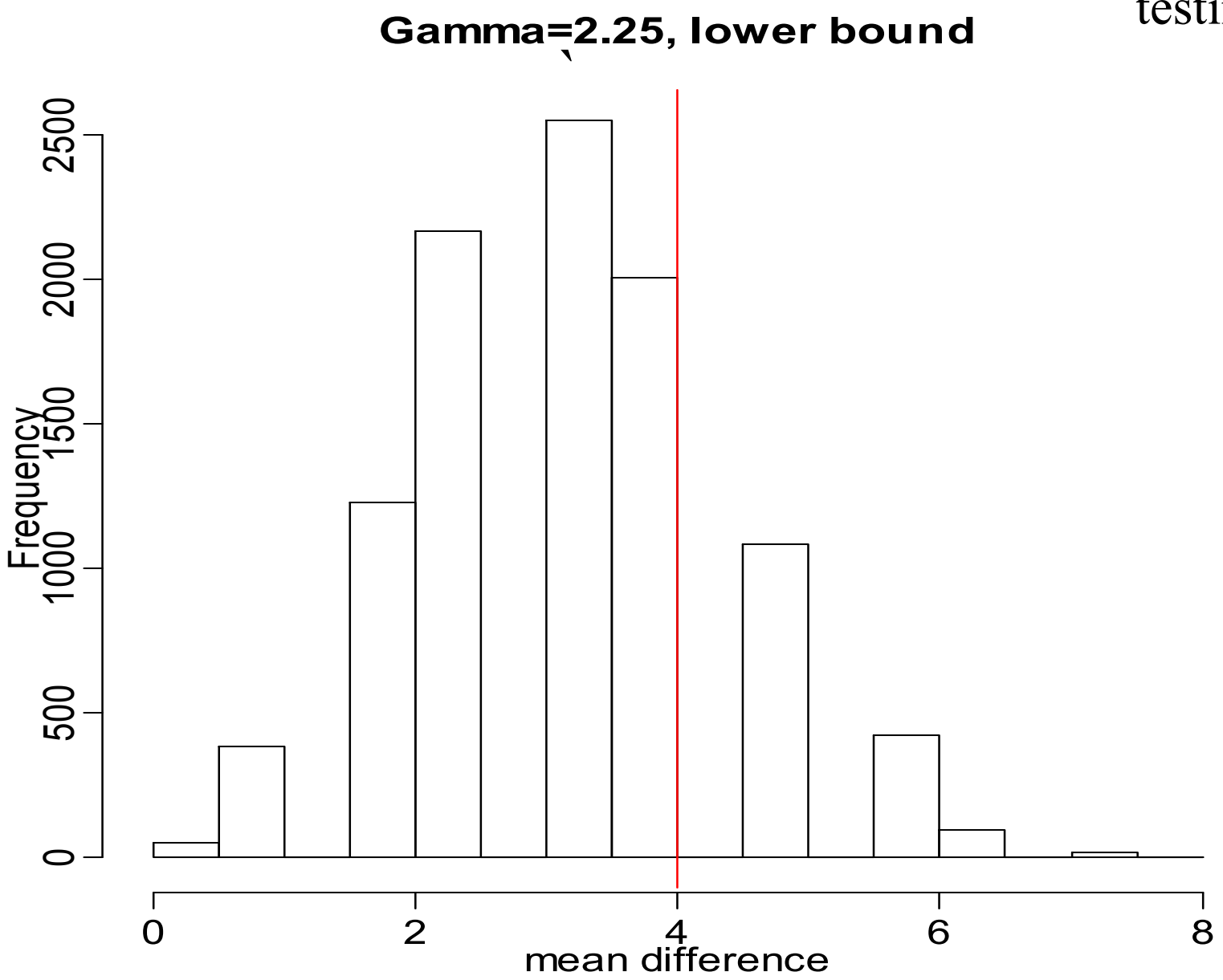


RB inference: matched pairs, hidden bias of Γ=2.25, testing τ=4

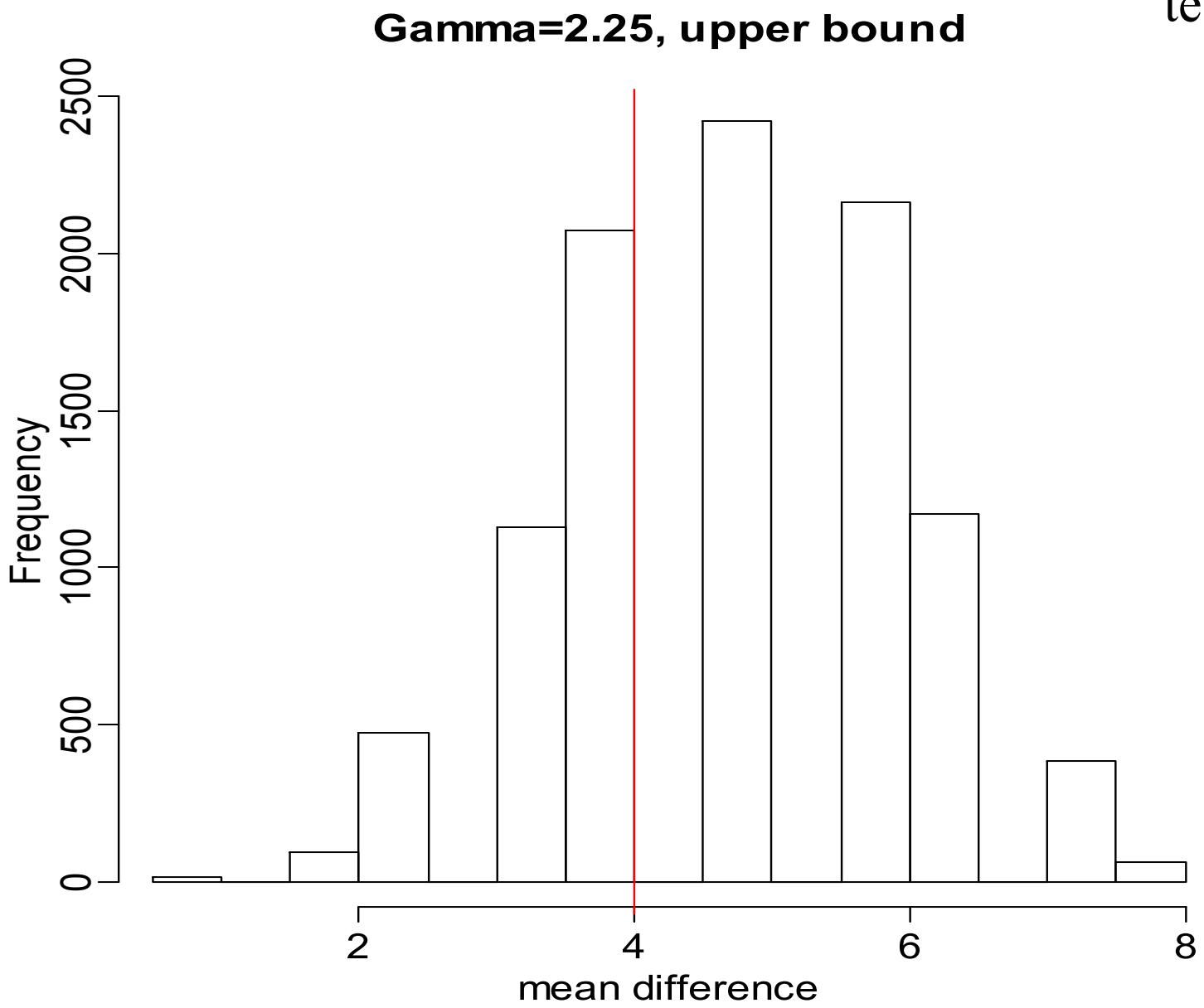
$$\frac{1}{n}\sum_j \Big(((Y_{1j}+\tau)-(Y_{2j}-\tau))Z_{1j}+(Y_{2j}-Y_{1j})Z_{2j}\Big)=4.1$$

Pair	Z_j	(π_{1j},π_{2j})	(Y_{1j},Y_{2j})	$(Y(0)_{1j},Y(0)_{2j})$	$(Y(1)_{1j},Y(1)_{2j})$
1	(0,1)	(.4,.6)	(4.2,8)	(4.2,4)	(8.2,8)
2	(0,1)	(.4,.6)	(5,9.1)	(5,5.1)	(9,9.1)
3	(0,1)	(.4,.6)	(6.2,10)	(6.2,6)	(10.2,10)
4	(0,1)	(.4,.6)	(4,8.3)	(4,4.3)	(8,8.3)
5	(0,1)	(.4,.6)	(6,9.6)	(6,5.6)	(10,9.6)
6	(0,1)	(.4,.6)	(5.6,10)	(5.6,6)	(9.6,10)
7	(0,1)	(.4,.6)	(5,9.2)	(5,5.2)	(9,9.2)
8	(0,1)	(.4,.6)	(4.3,8)	(4.3,4)	(8.3,8)
9	(0,1)	(.4,.6)	(4,8.1)	(4,4.1)	(8,8.1)
10	(0,1)	(.4,.6)	(6,10.2)	(6,6.2)	(10,10.2)

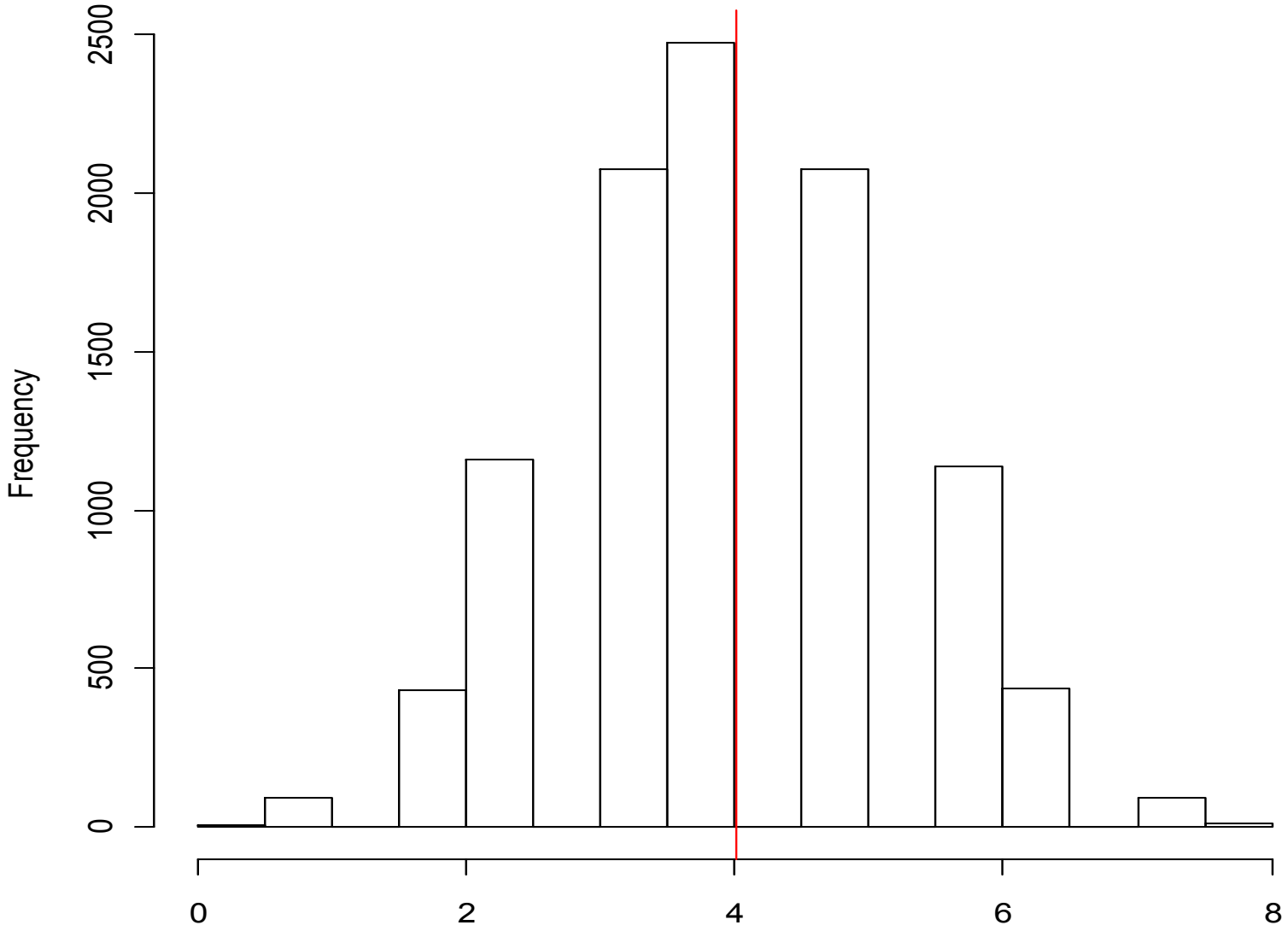
testing τ=4



testing τ=4



This is what the randomization distribution looks like for H_A: τ=4 when we assume no omitted confounders



Summary of this evidence

- When testing $\tau=0$ there is some sensitivity to a moderate to large amount of hidden bias ($\Gamma=2.25$), but not enough to alter our conclusions (either way we reject the null)
- When testing $\tau=4$ there is some sensitivity to this level of hidden bias (either way we do not reject the null)

However…

- We don't want to have to test hypotheses at upper and lower bounds for each Γ for every value of τ…
- So we replace this procedure with a nonparametric rank-based test such as the Wilcoxon signed rank test
- Stata (or R) implements this in the `rbounds` package.

`rbounds` (Stata and R)

- `rbounds` works naturally with the output from `psmatch2`
- I'll show a quick example in Stata using
  - our test data
  - the NSW data used in the pscore lecture

rbounds example test data

```
ssc install rbounds
gen delta=y2-y1
rbounds delta, gamma(1 (1) 3)

Rosenbaum bounds for delta (N = 10 matched pairs)

Gamma           sig+      sig-    t-hat+    t-hat-       CI+       CI-
-------------------------------------------------------------------
    1        .002488   .002488         4         4       3.8      4.25
    2         .02352   .000036      3.95       4.1       3.6       4.4
    3        .052454   5.7e-07       3.9       4.2       -99        99

* gamma  - log odds of differential assignment due to unobserved factors
  sig+   - upper bound significance level
  sig-   - lower bound significance level
  t-hat+ - upper bound Hodges-Lehmann point estimate
  t-hat- - lower bound Hodges-Lehmann point estimate
  CI+    - upper bound confidence interval (a=  .95)
  CI-    - lower bound confidence interval (a=  .95)
```

rbounds: NSW/LaLonde example

```
psmatch2 treat age educ black hisp re74 re75, outc(re78)
gen delta=re78-_re78 if _treat==1
rbounds delta, gamma(1 (.5) 3)

Rosenbaum bounds for delta (N = 185 matched pairs)

Gamma           sig+      sig-    t-hat+    t-hat-       CI+        CI-
-------------------------------------------------------------------------
    1         .220211   .220211   409.916   409.916  -613.693   1701.07
  1.5         .945662   .000739  -875.075   2036.88  -2255.33   3682.27
    2         .999556   3.5e-07  -1989.87   3342.75  -3585.36   4854.15
  2.5         .999999   7.2e-11  -2947.67   4232.05   -4461.7   5759.68
    3               1   9.7e-15  -3697.74   4961.34  -5182.48   6664.83

* gamma  - log odds of differential assignment due to unobserved
   factors
  sig+   - upper bound significance level
  sig-   - lower bound significance level
  t-hat+ - upper bound Hodges-Lehmann point estimate
  t-hat- - lower bound Hodges-Lehmann point estimate
  CI+    - upper bound confidence interval (a=  .95)
  CI-    - lower bound confidence interval (a=  .95)
```

Sensitivity Analysis: simultaneous methods

Sensitivity Analysis: Simultaneous methods

I'll discuss a simultaneous method that I've developed with colleagues (Nicole Carnegie, Masataka Harada).

- It is similar to the models/algorithms developed by others (Imbens (2003), Harada (2012)) with some advantages in terms of
  - flexibility (can estimate ATE, ATT, ATC) and
  - interpretability (sensitivity parameters are expressed as regression coefficients).
- The software (`treatSens`) is available for R (it is on CRAN).

Motivating example:
Effect of breastfeeding on long-term cognitive outcomes

Study with ADDHealth data

- ADDHealth is a nationally representative sample of adolescents in the US in 7th – 12th grade
- *Treatment*. We make comparisons between
  - children who were breastfed for at least six months
  - children who were not breastfed at all
- *Outcome*. Adolescent score on the Peabody Picture Vocabulary Test III
- *Confounders*. Gender, child's age at IQ assessment, ethnicity, birth weight, ELL status, biological mother's and father's level of education (as an IQ proxy), residential mother's level of education, family income, and whether the family was currently on welfare.
- *Unobserved confounder?* Past research has demonstrated the important of mother's IQ as a confounder. Our research may be sensitive to omission of this confounder (or others).

Effect of breastfeeding on age 16 test scores

- Goal: estimate the effect of breastfeeding at least 6 months (compared to not at all) on the test scores of children at age 16
- Estimand: The effect of the treatment on the treated
- Estimation Strategy: Genetic Matching
- Challenge: Determine how problematic the omission of mother IQ might be in terms of identifying the causal effect

Sensitivity to an unobserved confounder, formalization of assumption

- What if our identification strategy (for instance propensity score matching conditioning on all observed covariates) fails to control for one important confounder, ***U***
- That is, what if ***U*** is the final confounder we need to measure in order to satisfy "all confounders measured" (maybe like mom IQ in the breastfeeding example):

$$Y(0),\ Y(1) \perp Z \mid X, \boldsymbol{U}$$

- The problem is of course that we don't know what ***U*** looks like, so we'll need to explore the impact of this potential ***U*** over a range of plausible options

Sensitivity Analysis General Algorithm

Goal is to explore sensitivity of our estimate of τ to a range of potential incarnations of U. To distinguish our guess at U from the true U, we will refer to it as $\ddot{U}$

1. Specify a model for how the data were generated
2. Generate $\ddot{U}$ consistent with this data generating process and accompanying sensitivity parameters (here ζ^z and ζ^y). For this we derive $p(\ddot{U} \mid X, Z, Y)$.
3. Estimate (or draw from a distribution for) the treatment effect, τ, conditional on $\ddot{U}$.
4. Plot contours of estimated τ conditional on the sensitivity parameters (which can also be expressed as partial correlations)

Step 1

What we assume about the world.
Data Generating Process: Binary Z

- We "factor the likelihood" (rewrite the joint model in terms of conditionals) as

$$p(Y, Z, U, X \mid \theta) = p_1(U \mid \theta_1)\, p_2(Z \mid X, U, \theta_2)\, p_3(Y \mid X, U, Z, \theta_3)$$

- We use the following

$$U \sim \text{Bernoulli}(0,1)$$

$$Z \mid X, U \sim \text{Bernoulli}(\Phi^{-1}(\beta_z X + \zeta_z U))$$

$$Y \mid X, U, Z \sim N(\beta_y X + \zeta_y U + \tau Z, \sigma_{y.xuz}^2)$$

Sensitivity parameters

Causal Effect

Step 2

Bernoulli draw of U details

Problem: To draw from this distribution we need to know the parameters below

We derive Pr(U=1 | Y, Z, X) using Bayes rule:

Pr(Y, Z, U=1 | X)/Pr(Y, Z | X)

$$\pi^{y,z,x,u=1} = \left(2\pi\sigma^2_{\text{y·xuz}}\right)^{-1/2} \exp\left(-\frac{(y-\beta^y x-\zeta^y-\tau z)^2}{2\sigma^2_{\text{y·xuz}}}\right)$$

$$\cdot\left(1-\Phi^{-1}(\beta^z x+\zeta^z)\right)^{(1-z)}\left(\Phi^{-1}(\beta^z x+\zeta^z)\right)^{z}\pi^{u}$$

$$\pi^{y,z,x} = \left(2\pi\sigma^2_{\text{y·xuz}}\right)^{-1/2} \exp\left(-\frac{(y-\beta^y x-\tau z)^2}{2\sigma^2_{\text{y·xuz}}}\right)$$

$$\cdot\left(1-\Phi^{-1}(\beta^z x)\right)^{(1-z)}\left(\Phi^{-1}(\beta^z x+\zeta^z)\right)^{z}(1-\pi^{u})+$$

$$\left(2\pi\sigma^2_{\text{y·xuz}}\right)^{-1/2} \exp\left(-\frac{(y-\beta^y x-\zeta^y-\tau z)^2}{2\sigma^2_{\text{y·xuz}}}\right)$$

$$\cdot\left(1-\Phi^{-1}(\beta^z x+\zeta^z)\right)^{(1-z)}\left(\Phi^{-1}(\beta^z x+\zeta^z)\right)^{z}\pi^{u}$$

Step 2

Drawing Ü (Binary Z case)

- Closed form solutions for drawing U are not readily available
- A reasonable iterative solution exists however

0) First we estimate the parameters of $Z \mid X, U \sim$ Bernoulli$(\Phi^{-1}(\beta_z x + \zeta_z u))$ $Y \mid X, U, Z \sim N(\beta_y x + \zeta_y u + \tau z, \sigma_{y.xuz}^2)$ using a candidate value of U, $\ddot{U}$

1) Then we draw Ü conditional on these parameters using

$\ddot{U} \mid Y,Z,X, \sim$ Bernoulli$(\pi_{y,z,u=1} / \pi_{y,z})$

2) Then we get new estimates of the parameters conditional on this new value, $\ddot{U}$

- We iterate between (1) and (2) until convergence (typically obtained in a very small number of steps, i.e. below 10)

Step 3

Using Ü to estimate τ

1) Specify the distributions

2) Draw from p(U | X, Y, Z)

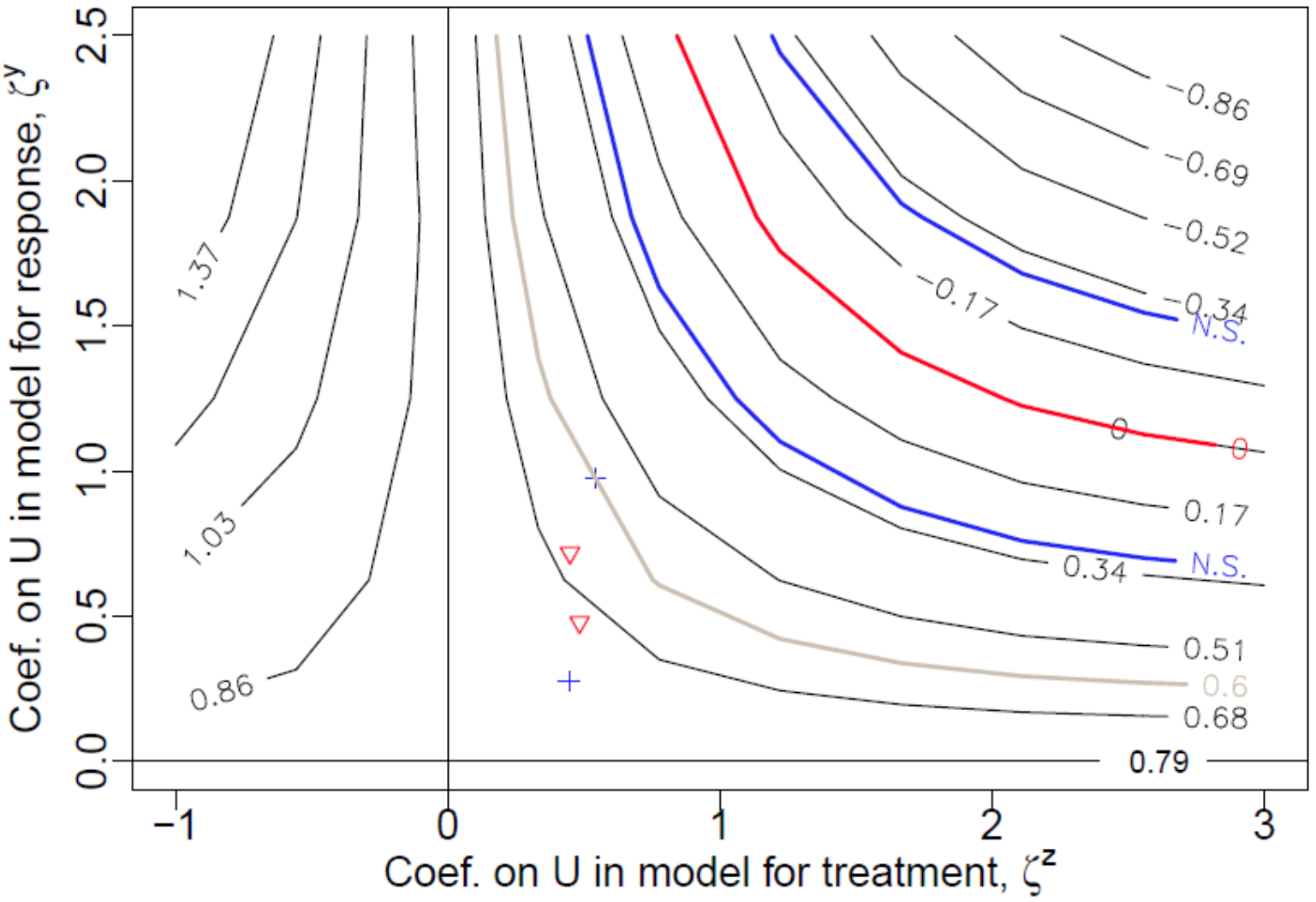
3) Fit E(Y | X, U, Z) using linear regression to estimate τ

$$Y \mid X, U, Z \sim N(\beta\uparrow y\, X + \zeta\uparrow y\, Ü + \tau Z, \sigma\downarrow y.xuz\uparrow 2)$$

4) Draw contours

Step 4

Graphical output from our program



What are we estimating?
Targeting estimands

- When identifying causal effects for observational studies the average treatment effect is often not of primary interest
  - policy, practice considerations
  - problems with lack of sufficient common support
- Existing methods for dual sensitivity analysis implicitly either assume a constant treatment effect or estimate an (variance-weighted) average treatment effect
- We have incorporated weights into our algorithm that allow for estimated of targeted estimands (e.g. effect of the treatment on the treated, effect of the treatment on the controls) assuming that treatment effects vary only by observed X

Targeting estimands: modification weights

- We estimate "modification scores", m(x), by fitting a probit (or similar) model for E[Z | X] and getting estimated probabilities
- Then we use standard IPTW formulations to create weights appropriate for particular estimands
- ATE weights
  - 1/m(x) for the treated
  - 1/(1-m(x)) for the untreated
- ATT weights
  - 1 for the treated
  - m(x)/(1-m(x)) for the untreated
- Weights are incorporated in the steps where we fit models for E[Y | Z, X] and E[Y | Z, X, U]

Breastfeeding example: Analysis and Sensitivity

Study with ADDHealth data

- ADDHealth is a nationally representative sample of adolescents in the US in 7th – 12th grade
- *Treatment*. We make comparisons between
  - children who were breastfed for at least six months
  - children who were not breastfed at all
- *Outcome*. Adolescent score on the Peabody Picture Vocabulary Test III
- *Confounders*. Gender, child's age at IQ assessment, ethnicity, birth weight, ELL status, biological mother's and father's level of education (as an IQ proxy), residential mother's level of education, family income, and whether the family was currently on welfare.
- *Unobserved confounder?* Past research has demonstrated the important of mother's IQ as a confounder. Our research may be sensitive to omission of this confounder (or others).

Effect of breastfeeding on age 16 test scores

- Goal: estimate the effect of breastfeeding at least 6 months (compared to not at all) on the test scores of children at age 16
- Estimand: The effect of the treatment on the treated
- Estimation Strategy: Genetic Matching

Assumptions

- Structural Assumption: All confounders measured (will "relax" by using SA to explore how vulnerable our inferences are to deviations from this assumption)
- Parametric Assumptions? Our matching approach is non-parametric. When coupled with regression it can be considered a semi-parametric approach. This means that our results should be robust to deviations from the strict parametric assumptions (linearity and additivity) of the standard linear regression model

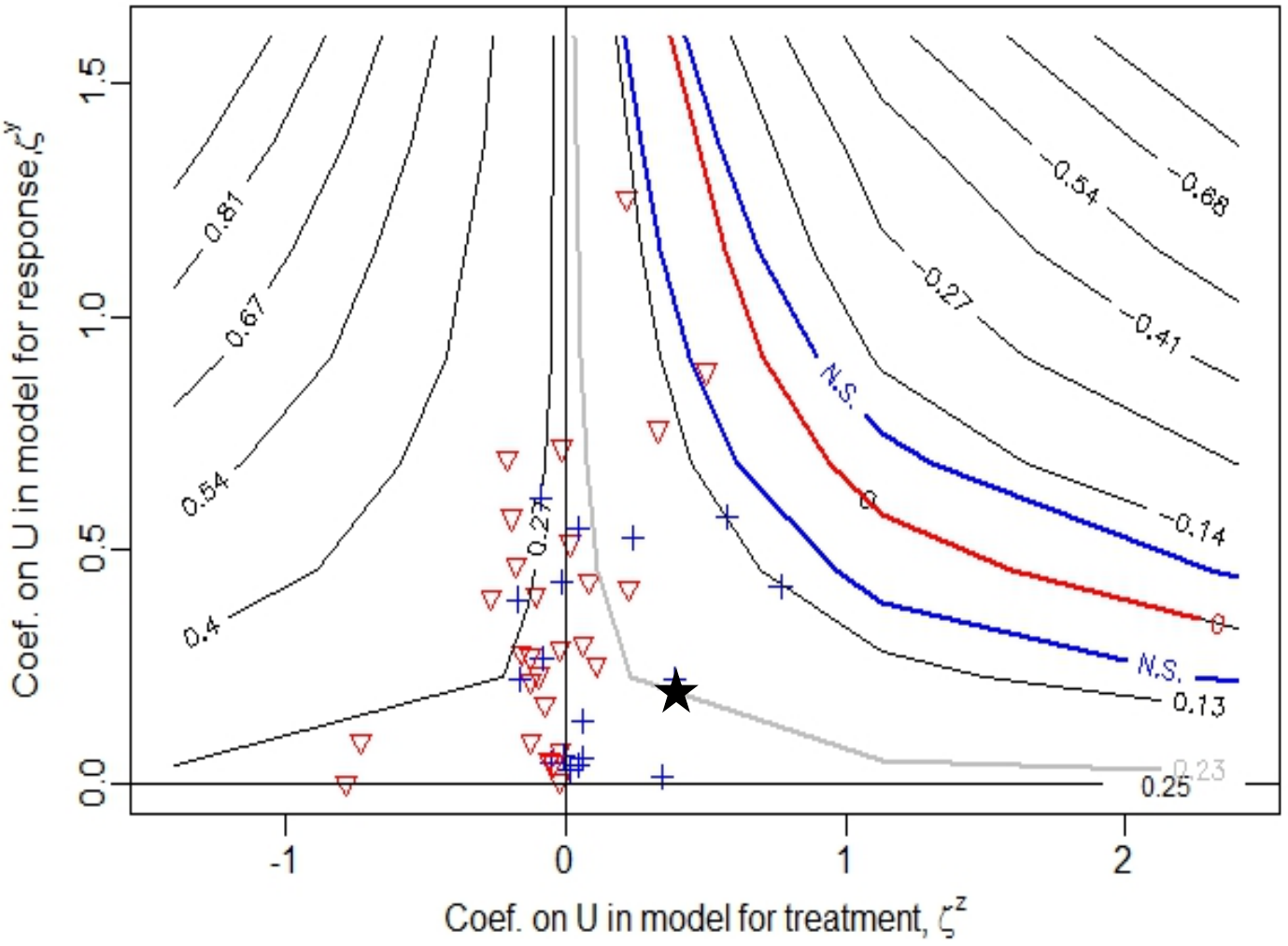
Estimates of the effect of the treatment on the treated

- Difference in means estimate on matched sample is 3.44 (.70)
- Regression estimate on matched sample is 3.52 (.61)

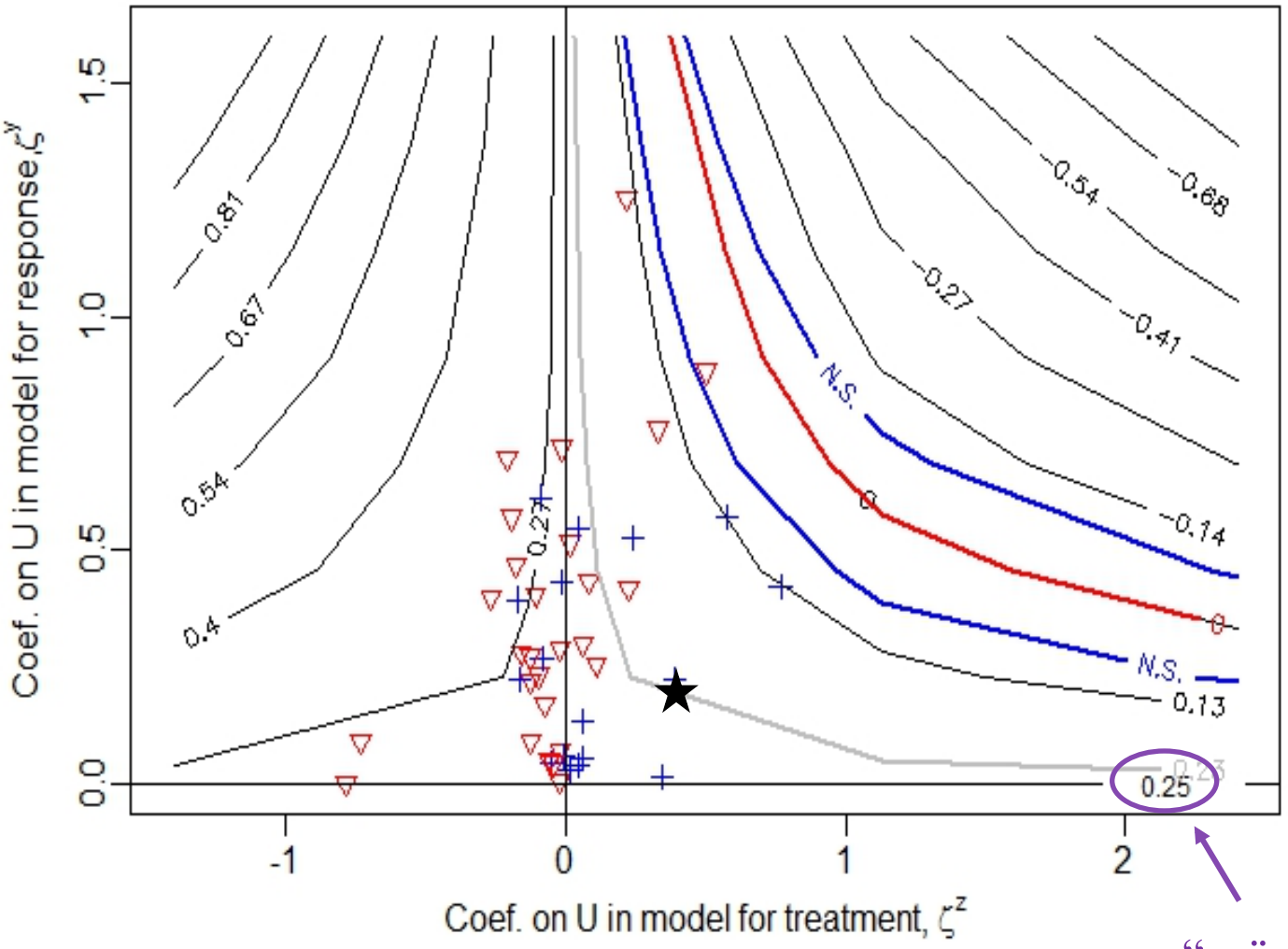
- In standard deviation units this is an effect size of .24 or .25

- (Note that estimates obtained from IPTW using propensity scores yields similar estimates)

Sensitivity of our results



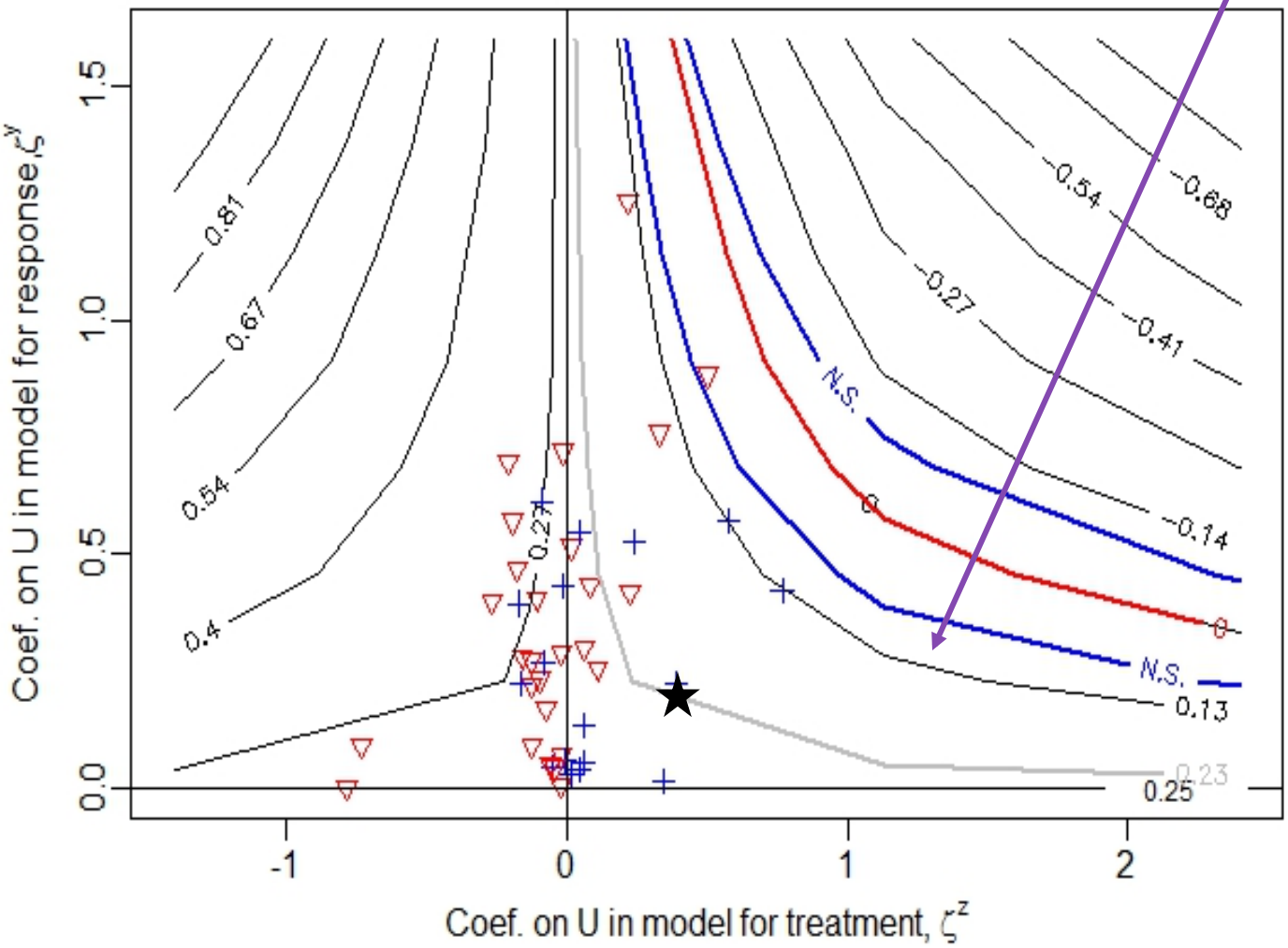
Sensitivity of our results



"naïve" regression estimate

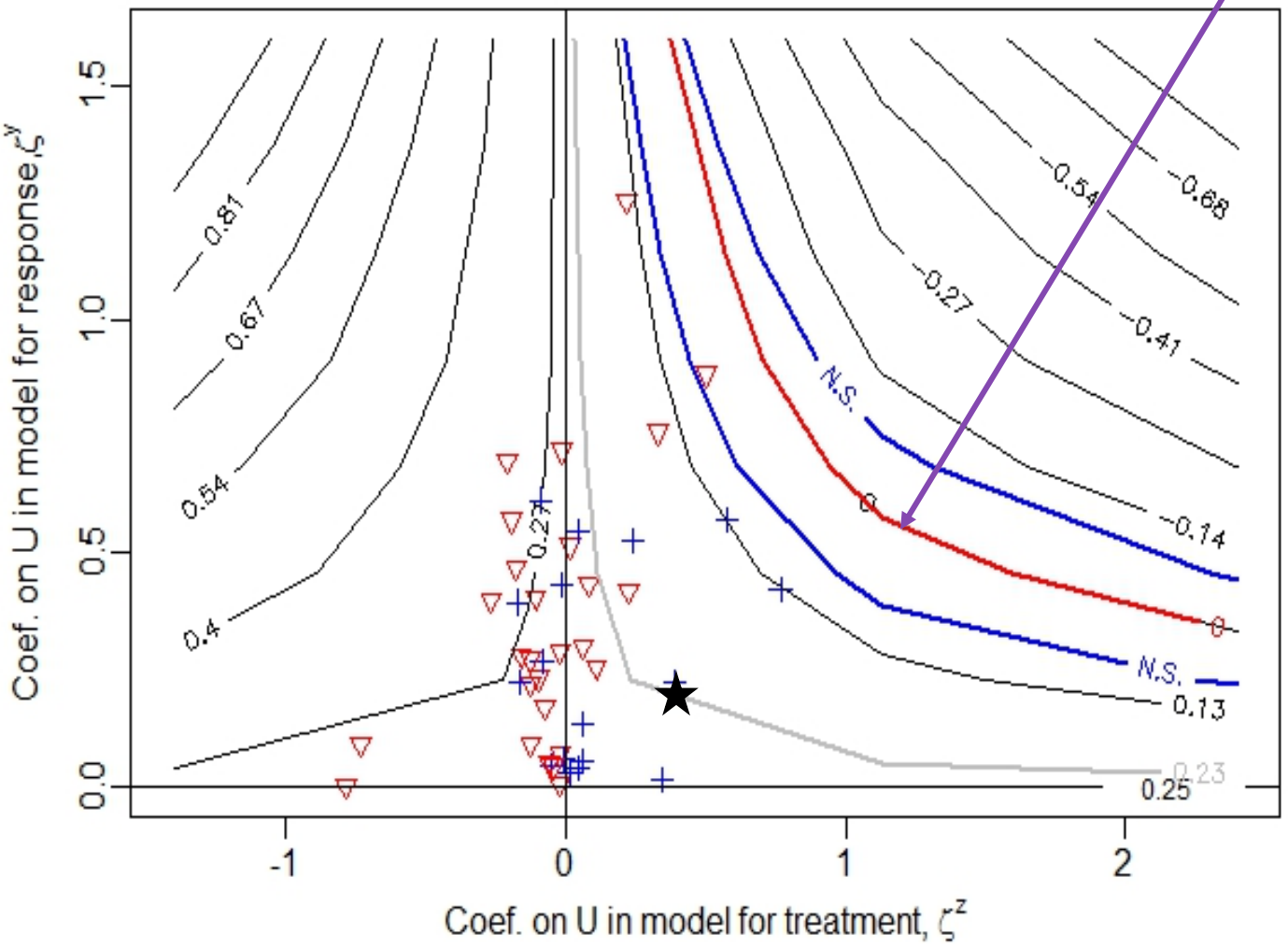
Sensitivity of our results

combinations of ζ^z and ζ^y that would lead to a standardized treatment effect estimate of .13



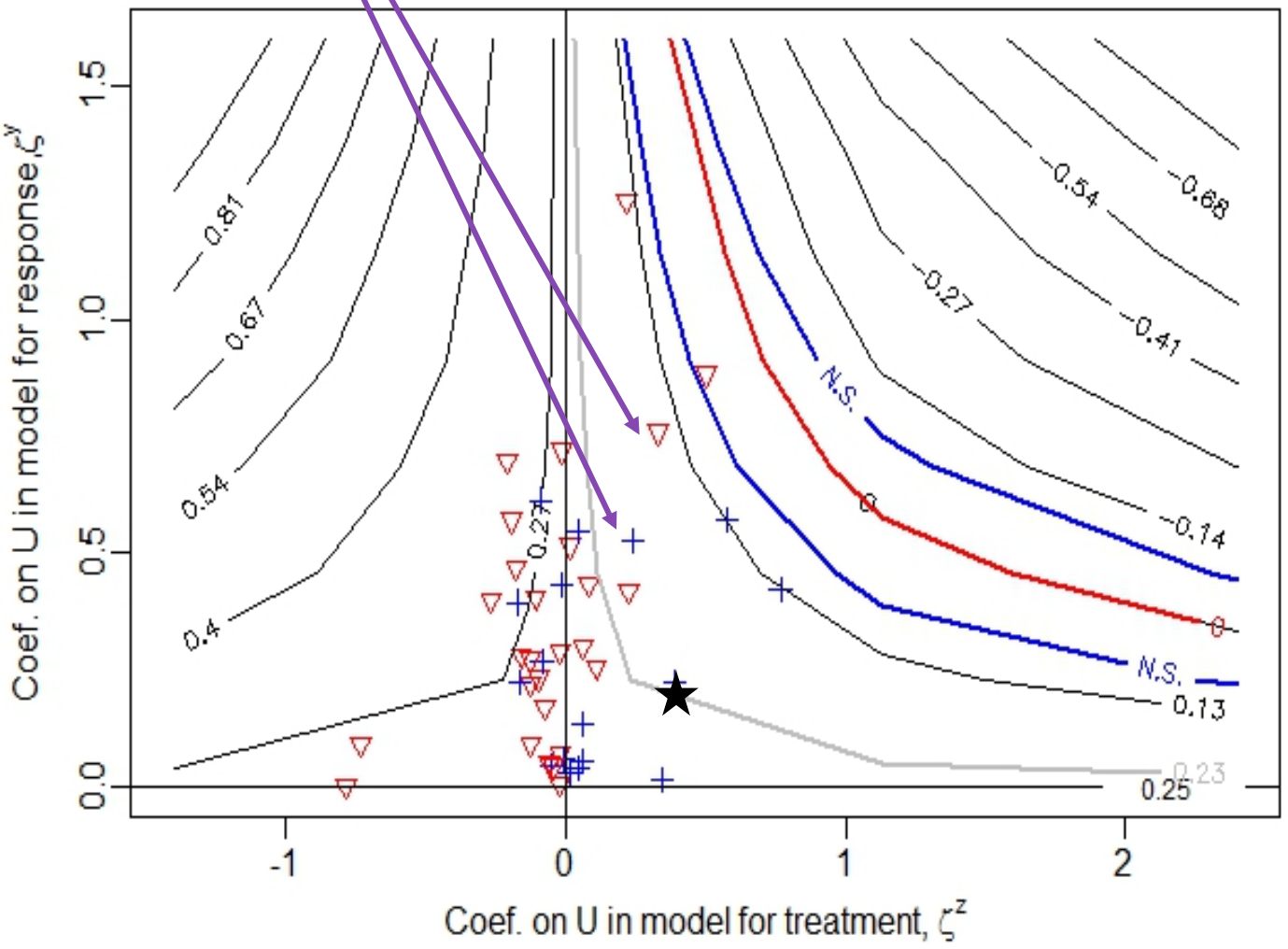
Sensitivity of our results

combinations of ζ^z and ζ^y that would lead to a treatment effect estimate of 0



Sensitivity of our results

(standardized) coefficients from a regression of the outcome on the treatment and observed covariates (∇ are for variables whose sign we reversed)



Calibration to external information about mom IQ

- Let's suppose (somewhat heroically) that mother's intelligence is the only omitted confounder in our analysis
- How can we get a handle on the size of the sensitivity parameters for such a confounder?
- The NLSY child sample has similar measures to those in ADDHealth but, importantly, it also contains a measure of mother's intelligence (the AFQT)
- We ran similar regressions in the NLSY sample to provide an estimate of the size of the coefficients on mom IQ – the result is plotted as a star,★ , on the sensitivity plot (note that this was in a regression predicting 5th grade outcomes so is likely an overestimate of the coefficients that would be observed for longer-term outcomes)
- This suggests that our results would not be very sensitive to inclusion of such a confounder

Discussion

- It is often impossible (for ethical, logistical, financial) reasons to randomize exposure to causal variables of interest
- In such scenarios are only option is often to use observational data to attempt to identify a causal effect
- Such analyses typically rely on strong assumptions such as “all confounders measured”
- Sensitivity analyses present an option that allow the researcher to explore how sensitive her inferences are to deviations from that assumption
- They provide an informal assessment of how confident we can be that our results are identifying a causal estimand
- This software is available for R (sitting on CRAN) as the `treatSens` package