Sensitivity analysis for principal ignorability violation in estimating complier and noncomplier average causal effects

Trang Nguyen

joint work with Liz Stuart, Dan Scharfstein, Betsy Ogburn

arXiv:2303.05032 | trang.nguyen@jhu.edu

CMStatistics 2023-12-17

Sensitivity analysis for principal ignorability violation in estimating complier and noncomplier average causal effects

2023-12-17

Sensitivity analysis for principal ignorability violation in estimating complier and noncomplier average causal effects

Trang Nguyen

joint work with Liz Stuart, Dan Scharfstein, Betsy Ogburn

arXiv:2303.05032 | trang.nguyen@jhu.edu

CMStatistics 2023-12-17

- thank Widemberg da Silva Nobre
- joint work with Liz Stuart, Dan Scharfstein and Betsy Ogburn
- this work is about sens analysis for violation of one of the assumptions used in estimating principal causal effects

Outline

Outline

Background
Principal causal effects
Principal ignorability (PI) – one identification strategy
Sensitivity analysis for PI violation

Mean-based sensitivity analysis: expanding options
Allow different sensitivity parameters
Accommodate different estimation methods

Distribution-based sensitivity analysis: a new thing
Method limitations and information use
A method using full information (ongoing work)

i will give a brief introduction to

- principal causal effects
- the principal ignorability assumption
- and sensitivity analysis for its violation

before going into the work that we do on this topic

The noncompliance problem

People may

not take their pill

not attend the training they are assigned to

volunteer less than they are asked to

The noncompliance problem
People may
not take their pill
not attend the training they are assigned to
volunteer less than they are asked to

the study of treatment effects is often complicated by noncompliance

- some people (in drug treatment trials) might not take the pill they are told to take
- some people might not attend the training program they are assigned to
- some people might volunteer fewer hours than they are asked to

The noncompliance problem

People may

not take their pill

not attend the training they are assigned to

volunteer less than they are asked to

Z	treatment assigned
Y	outcome
$Y(z)$	potential outcomes
X	baseline covariates
S	treatment received

The noncompliance problem
People may
not take their pill
not attend the training they are assigned to
volunteer less than they are asked to
Z treatment assigned
Y outcome
Y(z) potential outcomes
X baseline covariates
S treatment received

so in addition to the usual variables

- treatment assigned Z (here binary 0/1)
- outcome Y
- potential outcome Y(z) where z is either 1 or 0
- and baseline covariates X

we have a post-treatment assignment variable,
generically referred to as treatment received, S

The noncompliance problem

People may
- not take their pill
- not attend the training they are assigned to
- volunteer less than they are asked to

Z	treatment assigned
Y	outcome
$Y(z)$	potential outcomes
X	baseline covariates
S	treatment received

Might be intersted in
- the effect of receiving treatment
- but
- those who received $\neq$ those who did not

The noncompliance problem
People may
not take their pill
not attend the training they are assigned to
volunteer less than they are asked to
Z treatment assigned
Y outcome
Y(z) potential outcomes
X baseline covariates
S treatment received
Might be intersted in
the effect of receiving treatment
but
those who received ≠ those who did not

when not all those assigned to treatment receive the treatment

one might be more interested in
the effect of receiving the treatment than the effect of being assigned to the treatment

but those who received the treatment (or who volunteered) may not be similar to those who didn't

simply comparing their outcomes would be akin to breaking randomization

Principal stratification

People may
- not take their pill
- not attend the training they are assigned to
- volunteer less than they are asked to

Might be intersted in
- the effect of receiving treatment
- but the groups are not comparable

Principal stratification (Frangakis & Rubin 2002)
- avoids this problem
- by creating a new pre-treatment variable based on potential treatment received
- and stratifying on it

Z	treatment assigned
Y	outcome
$Y(z)$	potential outcomes
X	baseline covariates
S	treatment received
C	principal stratum, defined based on potential values $S(1), S(0)$ of S

Principal causal effects:

$$\mathrm{E}[Y(1) - Y(0) \mid C = c]$$

Principal stratification
People may
not take their pill
not attend the training they are assigned to
volunteer less than they are asked to
Might be intersted in
the effect of receiving treatment
but the groups are not comparable
Principal stratification (Frangakis & Rubin 2002)
avoids this problem
by creating a new pre-treatment variable based on potential treatment received
and stratifying on it
Z treatment assigned
Y outcome
Y(z) potential outcomes
X baseline covariates
S treatment received
C principal stratum, defined based on potential values S(1), S(0) of S
Principal causal effects:
E[Y(1) − Y(0) | C = c]

the principal stratification framework avoids this problem

by creating a new pre-treatment-assignment variable called principal stratum
based on potential values of treatment received

[pointing on RHS]
the variable principal stratum (C) here is defined based on S(1) and S(0),
which are potential values of S

now we consider effects
– still of being assigned to treatment vs. control, so still the difference between Y(1) and Y(0) –
but within each principal stratum

this way we don't break randomization
because C, which is about what type of person someone is (in terms of potential behavior),
is a pre-treatment-assignment variable

Principal stratification

People may
- not take their pill
- not attend the training they are assigned to
- volunteer less than they are asked to

Might be intersted in
- the effect of receiving treatment
- but the groups are not comparable

Principal stratification (Frangakis & Rubin 2002)
- avoids this problem
- by creating a new pre-treatment variable based on potential treatment received
- and stratifying on it

Two-sided noncompliance: 4 principal strata

$S(1)$	$S(0)$	C
1	1	always taker
1	0	complier
0	1	defier
0	0	never taker

Z	treatment assigned
Y	outcome
$Y(z)$	potential outcomes
X	baseline covariates
S	treatment received
C	principal stratum, defined based on potential values $S(1), S(0)$ of S

Principal causal effects:

$$\mathrm{E}[Y(1) - Y(0) \mid C = c]$$

Principal stratification
People may
not take their pill
not attend the training they are assigned to
volunteer less than they are asked to
Might be intersted in
the effect of receiving treatment
but the groups are not comparable
Principal stratification (Frangakis & Rubin 2002)
avoids this problem
by creating a new pre-treatment variable based on potential treatment received
and stratifying on it
Two-sided noncompliance: 4 principal strata
Z treatment assigned
Y outcome
Y(z) potential outcomes
X baseline covariates
S treatment received
C principal stratum, defined based on potential values S(1), S(0) of S
Principal causal effects:
E[Y(1) − Y(0) | C = c]

with binary Z and binary S, there are two settings

in the 2-sided noncompliance setting
there are four principal strata based on different combinations of S(1) and S(0) values

Principal stratification

People may
- not take their pill
- not attend the training they are assigned to
- volunteer less than they are asked to

Might be intersted in
- the effect of receiving treatment
- but the groups are not comparable

Principal stratification (Frangakis & Rubin 2002)
- avoids this problem
- by creating a new pre-treatment variable based on potential treatment received
- and stratifying on it

One-sided noncompliance: 2 principal strata

$S(1)$	$S(0)$	C
1	0	complier
0	0	noncomplier

i.e., $C = S(1)$

Z	treatment assigned
Y	outcome
$Y(z)$	potential outcomes
X	baseline covariates
S	treatment received
C	principal stratum, defined based on potential values $S(1), S(0)$ of S

Principal causal effects:

$$E[Y(1) - Y(0) \mid C = c]$$

in the 1-sided noncompliance setting,
where people assigned to control don't have access to the active treatment

S(0) is always zero

so there are two principal strata based on the two values of S(1)

which we refer to simply as compliers and noncompliers

Examples

Our focus is one-sided noncompliance

target: (non)complier average causal effects (CACE and NACE)

JOBS II for unemployed workers

- Z: week-long training on job search and mental health
- S: attending training

Experience Corps for the elderly

- Z: facilitated program for volunteering to help kids in school
- S: volunteering above a certain number of hours

Examples

Our focus is one-sided noncompliance
target: (non)complier average causal effects (CACE and NACE)

JOBS II for unemployed workers
- Z: week-long training on job search and mental health
- S: attending training

Experience Corps for the elderly
- Z: facilitated program for volunteering to help kids in school
- S: volunteering above a certain number of hours

our focus is one-sided noncompliance
where the target estimands are the effects on people of the two compliance types:
the complier average causal effect (CACE) and the noncomplier average causal effect (NACE)

let me give two examples

JOBS II is a study with unemployed workers
where the treatment Z is a week-long training on job search skills and mental health
and S, the treatment received variable, is an indicator attending the training or not

here the CACE and NACE are the effects of being assigned to the training vs. not
on people who would and people who would not attend the training if they were assigned to it

in Experience Corps
the intervention is a volunteering program for elderly folks as a way to improve their health and wellbeing
the investigators define variable S to be volunteering at least a certain number of hours

here the CACE and NACE are the effects of the program vs the control condition
on people who would and who would not volunteer at that level

Identification challenge: C is not observed under control

	compliers	noncompliers
intervention arm		
control arm		

2023-12-17

Identification challenge: C is not observed under control

	compliers	noncompliers
intervention arm		
control arm		

now the challenge is
C (compliance type) is only observed in the intervention arm

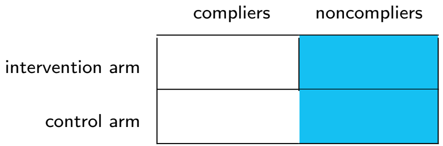
in the control arm we have a mixture two types of people
and do not know who is who

therefore the CACE and NACE are not identified
under standard causal inference assumptions

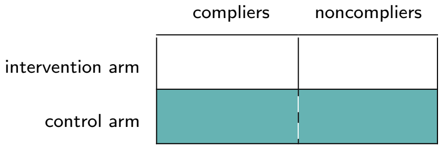
some additional assumption is required

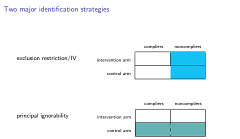
Two major identification strategies

exclusion restriction/IV



principal ignorability





there are two major strategies for identifying these effects

the first one is to assume that
for noncompliers the assigned treatment has no effect
so the outcome is similar in the two conditions

the second one is to assume that
in the control condition, the outcome is similar between compliers and noncompliers,
so principal stratum is ignorable,
hence the name principal ignorability

these are very different assumptions
and each may be appropriate in some situations but not others

our focus in the current work is PI

Principal ignorability (PI)

	compliers	noncompliers
intervention arm		
control arm		

$Y \perp\!\!\!\perp C \mid X, Z = 0$

ie within covariate levels, compliers and noncompliers share the same outcome distribution under control

Principal ignorability (PI)
compliers noncompliers
intervention arm
control arm
$Y \perp\!\!\!\perp C \mid X, Z = 0$
ie within covariate levels, compliers and noncompliers share the same outcome distribution under control

formally, PI is the assumption of this conditional independence
between compliance type and outcome under control

it means that within levels of baseline covariates
compliers and noncompliers share the same outcome distribution under control

this assumption may be plausible
if we have covariates that are predictive
both of compliance and of outcome under control

Principal ignorability (PI)

	compliers	noncompliers
intervention arm	$\mu_{11}(X)$	$\mu_{10}(X)$
control arm	$\mu_{01}(X)$	$\mu_{00}(X)$

$$Y \perp\!\!\!\perp C \mid X, Z = 0$$

or

$$\underbrace{\mathrm{E}[Y \mid X, Z = 0, C = 1]}_{\mu_{01}(X)} = \underbrace{\mathrm{E}[Y \mid X, Z = 0, C = 0]}_{\mu_{00}(X)}$$

Principal ignorability (PI)

the important thing it implies is the equality between these two conditional outcome means of compliers and noncompliers (which we label mu-01 and mu-00)

in this picture in the top right here,
these two conditional means under control are equal under PI

we also label the two means in the intervention arm as mu-11 and mu-10
these are identified from data

Principal ignorability (PI)

	compliers	noncompliers
intervention arm	$\mu_{11}(X)$	$\mu_{10}(X)$
control arm	$\mu_{01}(X)$	$\mu_{00}(X)$

$$Y \perp\!\!\!\perp C \mid X, Z = 0$$

or

$$\underbrace{\mathrm{E}[Y \mid X, Z = 0, C = 1]}_{\mu_{01}(X)} = \underbrace{\mathrm{E}[Y \mid X, Z = 0, C = 0]}_{\mu_{00}(X)}$$

Combined with treatment assignment ignorability, PI identifies CACE, NACE:

$$\mathrm{E}[Y(1) - Y(0) \mid C = c] = \frac{\mathrm{E}\{[\mu_{1c}(X) - \boxed{\kappa_0(X)}\,]\pi_c(X)\}}{\mathrm{E}[\pi_c(X)]}$$

where

$$\begin{aligned} \kappa_0(X) &:= \mathrm{E}[Y \mid X, Z = 0] && \text{(mixture outcome mean)} \\ \pi_c(X) &:= \mathrm{P}(C = c \mid X, Z = 1) && \text{(principal score)} \end{aligned}$$

the combination of PI with treatment assignment ignorability
identifies the two principal causal effects
by this formula

which is a weighted average
of a difference in outcome means between the intervention and control conditions

but here, this piece, which is supposed to be the stratum-specific mean mu-0c,
is replaced with the mixture outcome mean kappa-0 thanks to PI

and the weight, pi-c(x) here,
is the probability of belonging in principal stratum little c given covariate values
this is usually called the principal score

Principal ignorability (PI)

	compliers	noncompliers
intervention arm	$\mu_{11}(X)$	$\mu_{10}(X)$
control arm	$\mu_{01}(X)$	$\mu_{00}(X)$

$$Y \perp\!\!\!\perp C \mid X, Z = 0$$

or

$$\underbrace{\mathrm{E}[Y \mid X, Z = 0, C = 1]}_{\mu_{01}(X)} = \underbrace{\mathrm{E}[Y \mid X, Z = 0, C = 0]}_{\mu_{00}(X)}$$

PI is untestable

Need sensitivity analyses

Principal ignorability (PI)

	compliers	noncompliers
intervention arm	$\mu_{11}(X)$	$\mu_{10}(X)$
control arm	$\mu_{01}(X)$	$\mu_{00}(X)$

$Y \perp\!\!\!\perp C \mid X, Z = 0$

or

$\underbrace{\mathrm{E}[Y \mid X, Z = 0, C = 1]}_{\mu_{01}(X)} = \underbrace{\mathrm{E}[Y \mid X, Z = 0, C = 0]}_{\mu_{00}(X)}$

PI is untestable

Need sensitivity analyses

note that PI is an untestable assumption

so we need sensitivity analysis for its violation

Prior sens analysis method that inspired this work

Ding and Lu (2017) use a mean ratio sensitivity parameter

$$\frac{\mu_{01}(X)}{\mu_{00}(X)} = \rho$$

and modify a PI-based weighting estimator to incorporate ρ

See also Jiang, Yang and Ding (2022)

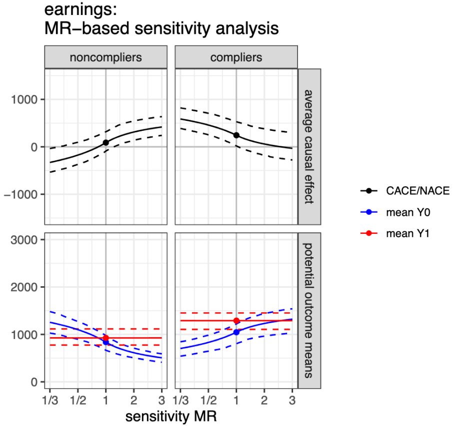
Prior sens analysis method that inspired this work
Ding and Lu (2017) use a mean ratio sensitivity parameter
$$\frac{\mu_{01}(X)}{\mu_{00}(X)} = \rho$$
and modify a PI-based weighting estimator to incorporate ρ
See also Jiang, Yang and Ding (2022)

in this prior work, which inspired ours,
instead of assuming these two conditional outcome means are equal
Ding and Lu used their ratio as a sens param
and consider it over a range for the sens analysis

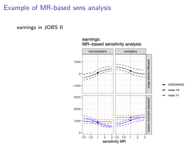
Ding and Lu modified a PI-based weighting estimator to incorporate this param

Example of MR-based sens analysis

earnings in JOBS II



2023-12-17



this is an example of a mean ratio based sens analysis
conducted on the outcome earnings at 6 months
in the JOBS II study

let me take a moment to orient you to this plot
b/c it will appear again

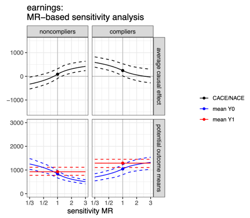
the two columns are for compliers and noncompliers
the top panel shows the effect estimates
the bottom panel shows the potential outcome means
the x-axis shows the sens param

the dots, which are at mean ratio 1, are the PI-based estimates
on the right of 1, compliers are assumed to have higher earnings under control than noncompliers
with the same covariate values
on the left of 1, it's the opposite

we see how the effect estimates change with the sens param

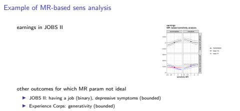
Example of MR-based sens analysis

earnings in JOBS II



other outcomes for which MR param not ideal

- JOBS II: having a job (binary), depressive symptoms (bounded)
- Experience Corps: generativity (bounded)



while the mean ratio param may be appropriate for earnings

there are other outcomes for which this param is not ideal
like binary and bounded outcomes

Outline

Outline

Background
Principal causal effects
Principal ignorability (PI) – one identification strategy
Sensitivity analysis for PI violation

Mean-based sensitivity analysis: expanding options
Allow different sensitivity parameters
Accommodate different estimation methods

Distribution-based sensitivity analysis: a new thing
Method limitations and information use
A method using full information (ongoing work)

that motivates this work on mean-based sensitivity analysis
which aims to allow different sensitivity parameters suitable to different outcome types

also, we want to accommodate different estimation methods
that may be used in practice

A range of sens assumptions with different sens params

PI: $\mu_{01}(X) = \mu_{00}(X)$.

sens-MR: $\dfrac{\mu_{01}(X)}{\mu_{00}(X)} = \rho$,

sens-OR: $\dfrac{\mu_{01}(X)/[1-\mu_{01}(X)]}{\mu_{00}(X)/[1-\mu_{00}(X)]} = \psi$,

sens-GOR: $\dfrac{[\mu_{01}(X)-l]/[h-\mu_{01}(X)]}{[\mu_{00}(X)-l]/[h-\mu_{00}(X)]} = \psi$

where l and h are the outcome lower and upper bounds,

sens-SMD: $\dfrac{\mu_{01}(X)-\mu_{00}(X)}{\sqrt{[\sigma^2_{01}(X)+\sigma^2_{00}(X)]/2}} = \eta$

where $\sigma^2_{0c}(X) := \text{var}(Y \mid X, Z=0, C=c)$,

for some range of ρ, ψ or η that is considered plausible.

A range of sens assumptions with different sens params
PI: $\mu_{01}(X) = \mu_{00}(X)$.
sens-MR: $\dfrac{\mu_{01}(X)}{\mu_{00}(X)} = \rho$,
sens-OR: $\dfrac{\mu_{01}(X)/[1-\mu_{01}(X)]}{\mu_{00}(X)/[1-\mu_{00}(X)]} = \psi$,
sens-GOR: $\dfrac{[\mu_{01}(X)-l]/[h-\mu_{01}(X)]}{[\mu_{00}(X)-l]/[h-\mu_{00}(X)]} = \psi$
where l and h are the outcome lower and upper bounds,
sens-SMD: $\dfrac{\mu_{01}(X)-\mu_{00}(X)}{\sqrt{[\sigma^2_{01}(X)+\sigma^2_{00}(X)]/2}} = \eta$
where $\sigma^2_{0c}(X) := \text{var}(Y \mid X, Z=0, C=c)$,
for some range of ρ, ψ or η that is considered plausible.

this slide shows a range of sens params

in addition to the mean ratio
we consider an odds ratio param, which is suitable for a binary outcome,

a generalized odds ratio, which can be used for an outcome bounded on both ends,

and a standardized mean difference parameter,
where the mean difference is divided by a pooled standard deviation.
this is suitable for a general continuous variable.

Identification

sens-MR and sens-GOR result in point identification of CACE, NACE

because they help solve the mixture equation

$$\pi_1(X)\mu_{01}(X) + \pi_0(X)\mu_{00}(X) = \kappa_0(X).$$

both sens-MR and -GOR result in point identification of the CACE and NACE
b/c they help solve this mixture equation
of the conditional outcome means under control

Identification

sens-MR and sens-GOR result in point identification of CACE, NACE

because they help solve the mixture equation

$$\pi_1(X)\mu_{01}(X) + \pi_0(X)\mu_{00}(X) = \kappa_0(X).$$

Identification

sens-MR and sens-GOR result in point identification of CACE, NACE

because they help solve the mixture equation

$$\pi_1(X)\mu_{01}(X) + \pi_0(X)\mu_{00}(X) = \kappa_0(X).$$

sens-SMD obtains bounds for CACE, NACE

bounds are narrowed if also assume $1/k \leq \frac{\sigma^2_{01}(X)}{\sigma^2_{00}(X)} \leq k$ for some $k > 1$

and reduce to point identification if assume $\sigma^2_{01}(X) = \sigma^2_{00}(X)$ (aka sens-SMDe)

Identification
sens-MR and sens-GOR result in point identification of CACE, NACE
because they help solve the mixture equation
$\pi_1(X)\mu_{01}(X) + \pi_0(X)\mu_{00}(X) = \kappa_0(X).$
sens-SMD obtains bounds for CACE, NACE
bounds are narrowed if also assume $1/k \leq \frac{\sigma^2_{01}(X)}{\sigma^2_{00}(X)} \leq k$ for some $k > 1$
and reduce to point identification if assume $\sigma^2_{01}(X) = \sigma^2_{00}(X)$ (aka sens-SMDe)

sens-SMD obtains bounds for the effects
these bounds are narrowed if we supplement it with an assumption about the conditional variances
and reduce to point identification if we assume equal variance

Identification

sens-MR and sens-GOR result in point identification of CACE, NACE

because they help solve the mixture equation

$$\pi_1(X)\mu_{01}(X) + \pi_0(X)\mu_{00}(X) = \kappa_0(X).$$

sens-SMD obtains bounds for CACE, NACE

bounds are narrowed if also assume $1/k \le \frac{\sigma^2_{01}(X)}{\sigma^2_{00}(X)} \le k$ for some $k > 1$

and reduce to point identification if assume $\sigma^2_{01}(X) = \sigma^2_{00}(X)$ (aka sens-SMDe)

in all cases, effect identification is via identification of $\mu_{0c}(X)$

by a function of sens param, $\pi_c(X)$, $\kappa_0(X)$ (and $\text{var}(Y \mid X, Z = 0)$ w/ sens-SMD)

Identification

sens-MR and sens-GOR result in point identification of CACE, NACE

because they help solve the mixture equation

$$\pi_1(X)\mu_{01}(X) + \pi_0(X)\mu_{00}(X) = \kappa_0(X).$$

sens-SMD obtains bounds for CACE, NACE

bounds are narrowed if also assume $1/k \le \frac{\sigma^2_{01}(X)}{\sigma^2_{00}(X)} \le k$ for some $k > 1$

and reduce to point identification if assume $\sigma^2_{01}(X) = \sigma^2_{00}(X)$ (aka sens-SMDe)

in all cases, effect identification is via identification of $\mu_{0c}(X)$

by a function of sens param, $\pi_c(X)$, $\kappa_0(X)$ (and $\text{var}(Y \mid X, Z = 0)$ w/ sens-SMD)

in all cases
effect identification is obtained via identification of mu-0c(x)

and the result for mu-0c(x) is a function of
the sens param
the principal score and mixture outcome mean
and for the sens-SMD assumption, also the mixture outcome variance

this detail is not so important but will be relevant in the next section

Sens analysis techniques to go with 3 types of PI-based estimators

Sens analysis = a modification of main analysis

Sens analysis techniques to go with 3 types of PI-based estimators
Sens analysis = a modification of main analysis

in practice, a sens analysis often follows and is secondary to a main analysis,

it is thus desirable for the sens analysis to be
a hopefully simple modification of the main analysis

people may use different estimators for the main analysis

here we think of 3 types of PI-based estimators
for which different sens analysis techniques may apply

Sens analysis = a modification of main analysis

- Type A ($\approx$ outcome regression estimators)
  - estimates $\kappa_0(X)$ to first estimate effects conditional on covariates and then aggregates them to estimate CACE/NACE, eg

$$\frac{\sum_{i=1}^{n}\hat{\pi}_c(X_i)[\hat{\mu}_{1c}(X_i)-\hat{\kappa}_0(X_i)]}{\sum_{i=1}^{n}\hat{\pi}_c(X_i)},\quad \frac{\sum_{i=1}^{n}\frac{Z_i}{\hat{e}(X_i,Z_i)}\mathrm{I}(C_i=c)[Y_i-\hat{\kappa}_0(X_i)]}{\sum_{i=1}^{n}\frac{Z_i}{\hat{e}(X_i,Z_i)}\mathrm{I}(C_i=c)}$$

  - sens analysis technique: replace $\kappa_0(X)$ by the identification result of $\mu_{0c}(X)$ under the sens assumption

The first type consists of outcome regression estimators

Roughly speaking, these as estimators that estimate the $\kappa_0(X)$ to first estimate conditional effects then aggregates the conditional effects to estimate the CACE/NACE

An example is the plug-in estimator <<point to it>>.

With this type, the sensitivity analysis technique is to replace $\kappa_0(X)$ with the identification result of $\mu_{0c}(X)$ under the sensitivitiy assumption

Sens analysis techniques to go with 3 types of PI-based estimators

Sens analysis = a modification of main analysis

- Type B ($\approx$ influence function based estimators)
  - write the CACE/NACE as
$$\frac{\nu_{1c} - \nu_{0c}^{\text{PI}}}{\pi_c}$$
    where $\nu_{zc} := \text{E}[\pi_c(X)\mu_{zc}(X)]$, $\nu_{0c}^{\text{PI}} := \text{E}[\pi_c(X)\kappa_0(X)]$, $\pi_c := \text{E}[\pi_c(X)]$
  - a type B estimator can be expressed as combination of IF-based estimators of π_c, ν_{1c} and ν_{0c}^{PI}
$$\frac{\hat{\nu}_{1c,\text{IF}} - \hat{\nu}_{0c,\text{IF}}^{\text{PI}}}{\hat{\delta}_{c,\text{IF}}}$$
  - sens analysis technique: replace $\hat{\nu}_{0c,\text{IF}}^{\text{PI}}$ with an IF-based estimator of ν_{0c} under the sens assumption

Sens analysis techniques to go with 3 types of PI-based estimators

Type B includes some influence function based estimators.

We write the effect under PI as this expression that involves three parameters.

A type B estimator can be expressed as a combination of influence function based estimators of these three parameters.

A relevant sensitivity analysis technique then is to replace this component estimator in red with an influence function based estimator of ν_{0c} under the sensitivity assumption

Sens analysis techniques to go with 3 types of PI-based estimators

Sens analysis = a modification of main analysis

- Type C (≈ other/weighting estimators)
  - an example is the pure weighting estimator

$$\frac{\sum_{i=1}^n \frac{Z_i}{\hat{e}(X_i,Z_i)} \mathrm{I}(C_i=c)Y_i}{\sum_{i=1}^n \frac{Z_i}{\hat{e}(X_i,Z_i)} \mathrm{I}(C_i=c)} - \frac{\sum_{i=1}^n \frac{1-Z_i}{\hat{e}(X_i,Z_i)}\hat{\pi}_c(X_i)Y_i}{\sum_{i=1}^n \frac{1-Z_i}{\hat{e}(X_i,Z_i)}\hat{\pi}_c(X_i)}$$

  - no general sens analysis technique

Sens analysis techniques to go with 3 types of PI-based estimators

Sens analysis = a modification of main analysis

- Type C (≈ other/weighting estimators)
  - an example is the pure weighting estimator
  - no general sens analysis technique

The other estimators are in type C.

They are not IF-based and do not involve estimating $\kappa_0(X)$.

An example is the pure weighting estimator.

For this type, we do not have a general sensitivity analysis technique in mind, and will need to consider them case by case.

Sens analysis techniques to go with 3 types of PI-based estimators

Sens analysis = a modification of main analysis

- Type C (≈ other/weighting estimators)
  - an example is the pure weighting estimator

$$\frac{\sum_{i=1}^{n}\frac{Z_i}{\hat{e}(X_i,Z_i)}\mathrm{I}(C_i=c)Y_i}{\sum_{i=1}^{n}\frac{Z_i}{\hat{e}(X_i,Z_i)}\mathrm{I}(C_i=c)}-\frac{\sum_{i=1}^{n}\frac{1-Z_i}{\hat{e}(X_i,Z_i)}\hat{\pi}_c(X_i)Y_i}{\sum_{i=1}^{n}\frac{1-Z_i}{\hat{e}(X_i,Z_i)}\hat{\pi}_c(X_i)}$$

  - no general sens analysis technique
  - sens-MR: scale Y in control units by a simple function of ρ and $\pi_c(X)$

Sens analysis techniques to go with 3 types of PI-based estimators
Sens analysis = a modification of main analysis
Type C (≈ other/weighting estimators)
an example is the pure weighting estimator
no general sens analysis technique
sens-MR: scale Y in control units by a simple function of ρ and $\pi_c(X)$

for the sens-MR case,
there is a simple technique for type C that involves a scaling of the outcome in control units

Sens analysis techniques to go with 3 types of PI-based estimators

Sens analysis = a modification of main analysis

- Type A ($\approx$ outcome regression estimators)
  - estimates $\kappa_0(X)$ to first estimate effects conditional on covariates and then aggregates them to estimate CACE/NACE
  - sens analysis technique: replace $\kappa_0(X)$ by the identification result of $\mu_{0c}(X)$ under the sens assumption
- Type B ($\approx$ influence function based estimators)
  - can be expressed as combination of IF-based estimators of π_c, ν_{1c} and ν_{0c}^{PI}
  - sens analysis technique: replace $\hat{\nu}_{0c,\text{IF}}^{\text{PI}}$ with an IF-based estimator of ν_{0c} under the sens assumption
- Type C ($\approx$ other/weighting estimators)
  - no general sens analysis technique
  - sens-MR: scale Y in control units by a simple function of ρ and $\pi_c(X)$

Sens analysis techniques to go with 3 types of PI-based estimators

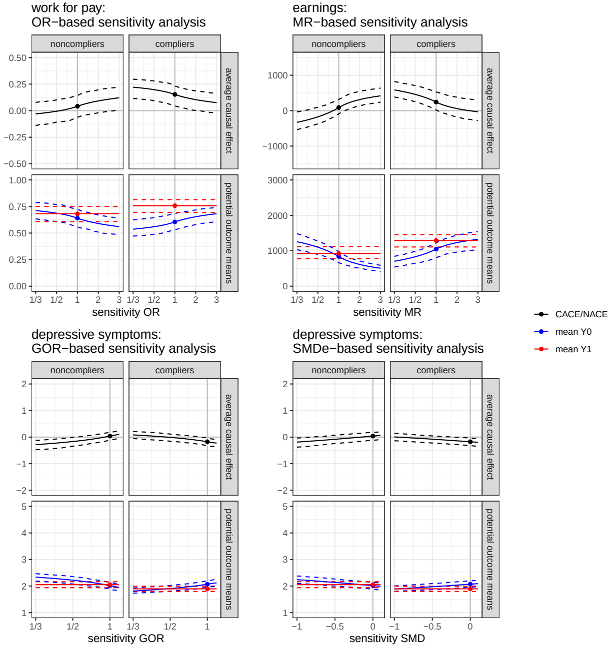
Sens analysis = a modification of main analysis

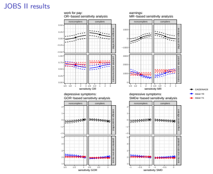
- Type A ($\approx$ outcome regression estimators)
  - estimates $\kappa_0(X)$ to first estimate effects conditional on covariates and then aggregates them to estimate CACE/NACE
  - sens analysis technique: replace $\kappa_0(X)$ by the identification result of $\mu_{0c}(X)$ under the sens assumption
- Type B ($\approx$ influence function based estimators)
  - can be expressed as combination of IF-based estimators of π_c, ν_{1c} and ν_{0c}^{PI}
  - sens analysis technique: replace $\hat{\nu}_{0c,\text{IF}}^{\text{PI}}$ with an IF-based estimator of ν_{0c} under the sens assumption
- Type C ($\approx$ other/weighting estimators)
  - no general sens analysis technique
  - sens-MR: scale Y in control units by a simple function of ρ and $\pi_c(X)$

the paper includes more details about different estimators

but these are basically the key ideas

JOBS II results





we applied the different sens analyses with different outcomes in JOBS II

using the OR param for the binary outcome being employed

using the GOR and SMD params for the continuous outcome depressive symptoms

a lot can be said about this specific example, but we don't have time

Some other things we noticed/figured out

- Partial loss of multiple robustness

  because $\mu_{0c}(X)$ is a function of $\pi_c(X)$ and $\kappa_0(X)$

- A pattern of finite-sample bias for the sens analysis where effect estimates are less extreme than should be

  because $\text{E}[Y_0 \mid C = c] = \dfrac{\text{E}[\pi_c(X)\mu_{0c}(X)]}{\text{E}[\pi_c(X)]}$ is a weighted average where the function being averaged depends on the weight

- If IF-based nonparametric estimation: Rate conditions on nuisance estimation for the estimator to be root-n consistent

there are several things we noticed or figured out

such as partial loss of multiple robustness for IF-based estimation

finite-sample bias in the sens analysis

and rates conditions for root-n consistent semiparamtric estimation

let's skip this slide, the details are in the paper

Some other things we noticed/figured out
Partial loss of multiple robustness
because $\mu_{0c}(X)$ is a function of $\pi_c(X)$ and $\kappa_0(X)$
A pattern of finite-sample bias for the sens analysis where effect estimates are less extreme than should be
because $\text{E}[Y_0 \mid C = c] = \frac{\text{E}[\pi_c(X)\mu_{0c}(X)]}{\text{E}[\pi_c(X)]}$ is a weighted average where the function being averaged depends on the weight
If IF-based nonparametric estimation: Rate conditions on nuisance estimation for the estimator to be root-n consistent

Outline

Outline
Background
Principal causal effects
Principal ignorability (PI) – one identification strategy
Sensitivity analysis for PI violation
Mean-based sensitivity analysis: expanding options
Allow different sensitivity parameters
Accommodate different estimation methods
Distribution-based sensitivity analysis: a new thing
Method limitations and information use
A method using full information (ongoing work)

now I'd like to turn to a different approach

distribution-based sens analysis

which we are working on

A general limitation of above methods

is the risk of contradicting the observed data distribution

Assumption	Risk level	Info used from the observed mixture outcome distribution
sens-MR	greatest risk	mean $\kappa_0(X) = \mathrm{E}[Y \mid X, Z = 0]$ lower bound (0)
sens-GOR (for nonbinary Y)	less risk	mean both bounds
sens-SMD	less risk	mean variance $\mathrm{var}(Y \mid X, Z = 0)$
sens-OR (binary Y)	no risk	full distribution (as mean = probability)

For nonbinary Y, to avoid contradicting the observed data distribution,

sens analysis needs to be fully informed by $\mathrm{P}(Y \mid X, Z = 0)$

A general limitation of above methods
is the risk of contradicting the observed data distribution
For nonbinary Y, to avoid contradicting the observed data distribution,
sens analysis needs to be fully informed by $\mathrm{P}(Y \mid X, Z = 0)$

a general limitation of the mean-based methods
is the risk of contradicting the observed data distribution

which is due to using some but not all the information in the data

for example, the sens-MR assumption can predict outcome means
that exceed the actual outcome range
because it is agnostic of the outcome's upper bound

sens-GOR and sens-SMD have lower risk as they use more information,
one using both bounds
the other using not only the mean but also the variance of the mixture distribution

sens-OR for a binary outcome has no risk, because it actually uses full information

this means it would be ideal for the sens analysis to be fully informed by the observed mixture outcome distribution

Constructing that sens analysis: foundation

PI: $Y \perp\!\!\!\perp C \mid X, Z = 0$

(Everything here conditions on $X, Z = 0$,
so this will be implicit.)

Constructing that sens analysis: foundation
PI: $Y \perp\!\!\!\perp C \mid X, Z = 0$
(Everything here conditions on $X, Z = 0$,
so this will be implicit.)

here's how we construct such a sens analysis

PI says that Y and C are conditionally independent

Constructing that sens analysis: foundation

PI: $Y \perp\!\!\!\perp C \mid X, Z = 0$

(Everything here conditions on $X, Z = 0$,
so this will be implicit.)

Want a sens assumption
that allows Y and C to be dependent AND helps identify $\mu_{0c}(X)$

we want a sens assumption that allows Y and C to be dependent

Constructing that sens analysis: foundation
PI: $Y \perp\!\!\!\perp C \mid X, Z = 0$
(Everything here conditions on $X, Z = 0$,
so this will be implicit.)
Want a sens assumption
that allows Y and C to be dependent AND helps identify $\mu_{0c}(X)$

Constructing that sens analysis: foundation

PI: $Y \perp\!\!\!\perp C \mid X, Z = 0$

(Everything here conditions on $X, Z = 0$,
so this will be implicit.)

Want a sens assumption
that allows Y and C to be dependent AND helps identify $\mu_{0c}(X)$

Principal scores

Recall/recast	$\pi_c(X) = \mathrm{P}(C = c \mid X, Z = 0)$	(outcome-agnostic)
Now define	$\tilde{\pi}_c(X, Y) = \mathrm{P}(C = c \mid X, Z = 0, Y)$	(outcome-specific)

Constructing that sens analysis: foundation
PI: $Y \perp\!\!\!\perp C \mid X, Z = 0$
(Everything here conditions on $X, Z = 0$, so this will be implicit.)
Want a sens assumption
that allows Y and C to be dependent AND helps identify $\mu_{0c}(X)$
Principal scores
Recall/recast $\pi_c(X) = \mathrm{P}(C = c \mid X, Z = 0)$ (outcome-agnostic)
Now define $\tilde{\pi}_c(X, Y) = \mathrm{P}(C = c \mid X, Z = 0, Y)$ (outcome-specific)

the principal score we have, pi-c(x), does not condition on the outcome

now define pi-tilde-c(x), which additionally conditions on Y

we call this the outcome-specific principal score

Constructing that sens analysis: foundation

PI: $Y \perp\!\!\!\perp C \mid X, Z = 0$

(Everything here conditions on $X, Z = 0$,
so this will be implicit.)

Want a sens assumption
that allows Y and C to be dependent AND helps identify $\mu_{0c}(X)$

Principal scores

Recall/recast	$\pi_c(X) = \mathrm{P}(C = c \mid X, Z = 0)$	(outcome-agnostic)
Now define	$\tilde{\pi}_c(X, Y) = \mathrm{P}(C = c \mid X, Z = 0, Y)$	(outcome-specific)

$\tilde{\pi}_c(X, Y)$ is not identified but we know

- $\mathrm{E}[\tilde{\pi}_c(X, Y) \mid X, Z = 0] = \pi_c(X)$
- If Y and C are dependent, $\tilde{\pi}_c(X, Y)$ is a function of Y in addition to X

Constructing that sens analysis: foundation
PI: $Y \perp\!\!\!\perp C \mid X, Z = 0$
(Everything here conditions on $X, Z = 0$, so this will be implicit.)
Want a sens assumption that allows Y and C to be dependent AND helps identify $\mu_{0c}(X)$
Principal scores
Recall/recast $\pi_c(X) = \mathrm{P}(C = c \mid X, Z = 0)$ (outcome-agnostic)
Now define $\tilde{\pi}_c(X, Y) = \mathrm{P}(C = c \mid X, Z = 0, Y)$ (outcome-specific)
$\tilde{\pi}_c(X, Y)$ is not identified but we know
- $\mathrm{E}[\tilde{\pi}_c(X, Y) \mid X, Z = 0] = \pi_c(X)$
- If Y and C are dependent, $\tilde{\pi}_c(X, Y)$ is a function of Y in addition to X

pi-tilde-c is not identified but we know that

its mean conditional on X,Z=0 is pi-c(x)

and that if Y and C are conditionally independent, pi-tilde is a function of Y

Constructing that sens analysis: foundation

PI: $Y \perp\!\!\!\perp C \mid X, Z = 0$

(Everything here conditions on $X, Z = 0$,
so this will be implicit.)

Want a sens assumption
that allows Y and C to be dependent AND helps identify $\mu_{0c}(X)$

Principal scores

Recall/recast	$\pi_c(X) = \mathrm{P}(C = c \mid X, Z = 0)$	(outcome-agnostic)
Now define	$\tilde{\pi}_c(X, Y) = \mathrm{P}(C = c \mid X, Z = 0, Y)$	(outcome-specific)

$\tilde{\pi}_c(X, Y)$ is not identified but we know

- $\mathrm{E}[\tilde{\pi}_c(X, Y) \mid X, Z = 0] = \pi_c(X)$
- If Y and C are dependent, $\tilde{\pi}_c(X, Y)$ is a function of Y in addition to X

(Side note: Very different from the unobserved confounding problem b/c of the observed mixture.
Hence exponential tilting doesn't work, except for a binary outcome, in which case = sens-OR.)

Constructing that sens analysis: foundation
PI: $Y \perp\!\!\!\perp C \mid X, Z = 0$
(Everything here conditions on $X, Z = 0$, so this will be implicit.)
Want a sens assumption that allows Y and C to be dependent AND helps identify $\mu_{0c}(X)$
Principal scores
Recall/recast $\pi_c(X) = \mathrm{P}(C = c \mid X, Z = 0)$ (outcome-agnostic)
Now define $\tilde{\pi}_c(X, Y) = \mathrm{P}(C = c \mid X, Z = 0, Y)$ (outcome-specific)
$\tilde{\pi}_c(X, Y)$ is not identified but we know
- $\mathrm{E}[\tilde{\pi}_c(X, Y) \mid X, Z = 0] = \pi_c(X)$
- If Y and C are dependent, $\tilde{\pi}_c(X, Y)$ is a function of Y in addition to X
(Side note: Very different from the unobserved confounding problem b/c of the observed mixture. Hence exponential tilting doesn't work, except for a binary outcome, in which case = sens-OR.)

let's skip this

Constructing that sens analysis: 2 steps

Use shorthand $\tilde{\pi}_1$ for $\tilde{\pi}_1(X, Y)$

Step 1: Assume a distribution for $\tilde{\pi}_1$ with mean $\pi_1(X)$ that allows $\tilde{\pi}_1$ to vary, indexed by a dispersion param

Step 2: Connect Y to $\tilde{\pi}_1$ to induce Y-C dependence

Constructing that sens analysis: 2 steps
Use shorthand $\tilde{\pi}_1$ for $\tilde{\pi}_1(X, Y)$
Step 1: Assume a distribution for $\tilde{\pi}_1$ with mean $\pi_1(X)$ that allows $\tilde{\pi}_1$ to vary, indexed by a dispersion param
Step 2: Connect Y to $\tilde{\pi}_1$ to induce Y-C dependence

With those two known facts,
we construct the sens analysis in two steps

[read slide]

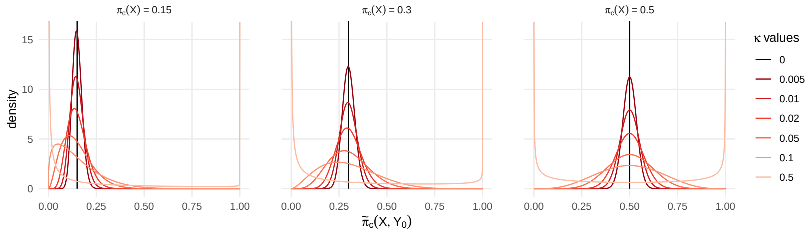
Constructing that sens analysis: 2 steps

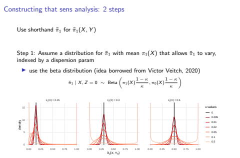
Use shorthand $\tilde{\pi}_1$ for $\tilde{\pi}_1(X, Y)$

Step 1: Assume a distribution for $\tilde{\pi}_1$ with mean $\pi_1(X)$ that allows $\tilde{\pi}_1$ to vary, indexed by a dispersion param

- use the beta distribution (idea borrowed from Victor Veitch, 2020)

$$\tilde{\pi}_1 \mid X, Z = 0 \sim \text{Beta}\left(\pi_1(X)\frac{1-\kappa}{\kappa}, \pi_0(X)\frac{1-\kappa}{\kappa}\right)$$





for step 1,
we assume that conditional on X, Z=0,
pi-tilde follows this beta distribution
with mean pi
and dispersion param kappa

this figure shows
for three values of pi (.15, .3 and .5)
what different kappa values imply
about the distribution of pi-tilde

we see that with kappa as small as 0.02,
there is a lot of variation in the probability

Constructing that sens analysis: 2 steps

Use shorthand $\tilde{\pi}_1$ for $\tilde{\pi}_1(X, Y)$

Step 1: Assume a distribution for $\tilde{\pi}_1$ with mean $\pi_1(X)$ that allows $\tilde{\pi}_1$ to vary, indexed by a dispersion param

- use the beta distribution

$$\tilde{\pi}_1 \mid X, Z = 0 \;\sim\; \text{Beta}\left(\pi_1(X)\frac{1-\kappa}{\kappa}, \pi_0(X)\frac{1-\kappa}{\kappa}\right)$$

Step 2: Connect Y to $\tilde{\pi}_1$ to induce Y-C dependence

- use quantile-to-quantile mapping between distributions of Y and of $\tilde{\pi}_1$ (given $X, Z = 0$)
  - same-quantiles: positive Y-C association
  - opposite-quantiles: negative Y-C association

Constructing that sens analysis: 2 steps

Use shorthand $\tilde{\pi}_1$ for $\tilde{\pi}_1(X, Y)$

Step 1: Assume a distribution for $\tilde{\pi}_1$ with mean $\pi_1(X)$ that allows $\tilde{\pi}_1$ to vary, indexed by a dispersion param

- use the beta distribution

$$\tilde{\pi}_1 \mid X, Z = 0 \;\sim\; \text{Beta}\left(\pi_1(X)\frac{1-\kappa}{\kappa}, \pi_0(X)\frac{1-\kappa}{\kappa}\right)$$

Step 2: Connect Y to $\tilde{\pi}_1$ to induce Y-C dependence

- use quantile-to-quantile mapping between distributions of Y and of $\tilde{\pi}_1$ (given $X, Z = 0$)
  - same-quantiles: positive Y-C association
  - opposite-quantiles: negative Y-C association

for step 2
we use quantile-to-quantile mapping between the conditional distributions of Y and of pi-tilde

where same-quantile mapping
(ie the 80th percentile in the distribution of Y
is connected to the 80th percentile in the distribution of pi-tilde)
means positive Y-C association

and opposite-quantile mapping means negative association

Constructing that sens analysis: 2 steps

Use shorthand $\tilde{\pi}_1$ for $\tilde{\pi}_1(X, Y)$

Step 1: Assume a distribution for $\tilde{\pi}_1$ with mean $\pi_1(X)$ that allows $\tilde{\pi}_1$ to vary, indexed by a dispersion param

- use the beta distribution

$$\tilde{\pi}_1 \mid X, Z = 0 \ \sim \ \text{Beta}\left(\pi_1(X)\frac{1-\kappa}{\kappa}, \pi_0(X)\frac{1-\kappa}{\kappa}\right)$$

Step 2: Connect Y to $\tilde{\pi}_1$ to induce Y-C dependence

- use quantile-to-quantile mapping between distributions of Y and of $\tilde{\pi}_1$ (given $X, Z = 0$)
  - same-quantiles: positive Y-C association
  - opposite-quantiles: negative Y-C association

The sens param: the "signed" κ

Constructing that sens analysis: 2 steps
Use shorthand $\tilde{\pi}_1$ for $\tilde{\pi}_1(X, Y)$
Step 1: Assume a distribution for $\tilde{\pi}_1$ with mean $\pi_1(X)$ that allows $\tilde{\pi}_1$ to vary, indexed by a dispersion param
use the beta distribution
$\tilde{\pi}_1 \mid X, Z = 0 \sim \text{Beta}\left(\pi_1(X)\frac{1-\kappa}{\kappa}, \pi_0(X)\frac{1-\kappa}{\kappa}\right)$
Step 2: Connect Y to $\tilde{\pi}_1$ to induce Y-C dependence
use quantile-to-quantile mapping between distributions of Y and of $\tilde{\pi}_1$ (given $X, Z = 0$)
same-quantiles: positive Y-C association
opposite-quantiles: negative Y-C association
The sens param: the "signed" κ

the sens param is the signed kappa

where the sign is the sign of the Y-C association

Identification and estimation

Identification:

$$\mu_{0c}(X) = \frac{\mathrm{E}[\tilde{\pi}_c Y \mid X, Z = 0]}{\mathrm{E}[\tilde{\pi}_c \mid X, Z = 0]},$$

where $\tilde{\pi}_c$ and Y are quantile-to-quantile connected

Estimation:

- estimate $\mathrm{P}(Y \mid X, Z = 0)$ so can compute quantiles
- then estimate $\mu_{0c}(X)$ using numerical integration

Identification and estimation

Identification: $\mu_{0c}(X) = \frac{\mathrm{E}[\tilde{\pi}_c Y \mid X, Z = 0]}{\mathrm{E}[\tilde{\pi}_c \mid X, Z = 0]}$,

where $\tilde{\pi}_c$ and Y are quantile-to-quantile connected

Estimation:
- estimate $\mathrm{P}(Y \mid X, Z = 0)$ so can compute quantiles
- then estimate $\mu_{0c}(X)$ using numerical integration

under this assumption about Y-C dependence
we have this identification result

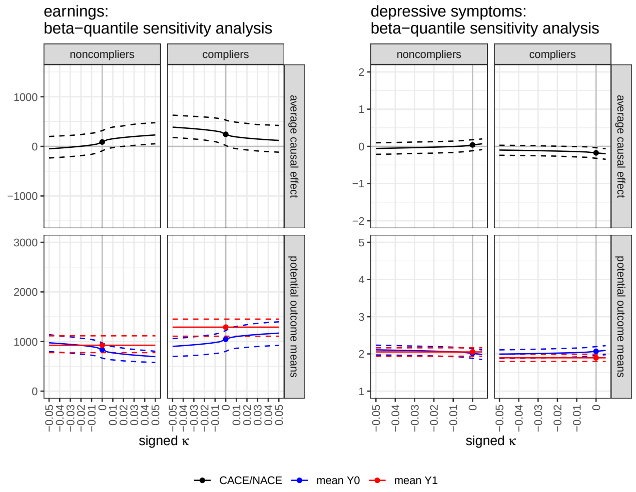
for estimation
we first estimate this mixture outcome distribution
so that we can compute its quantiles

and then we compute mu-0c(x) estimates using numerical integration

this method is harder than our earlier methods
because it requires estimating this distribution (or its quantile)

but as a result we benefit from more information

Example: JOBS II preliminary results



Results are less extreme than mean-based sens analysis

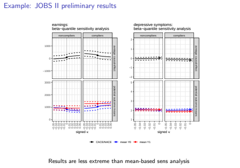
2023-12-17

Sensitivity analysis for principal ignorability violation in estimating complier and noncomplier average causal effects

└─Distribution-based sensitivity analysis: a new thing

  └─A method using full information (ongoing work)

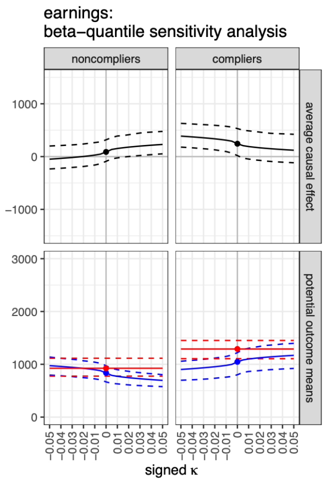
    └─earnings: distribution-based (left), mean-based (right)

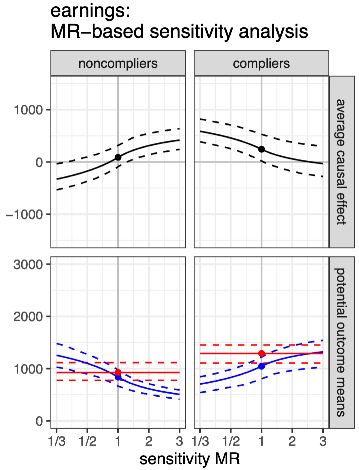


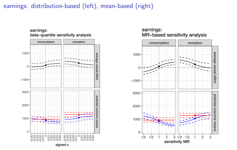
here are results from JOBS II
for the two outcomes earnings and depressive symptoms
for a rather wide range of signed kappa

what we notice is
these results can be less extreme than what we might get from mean-based sens analysis

earnings: distribution-based (left), mean-based (right)

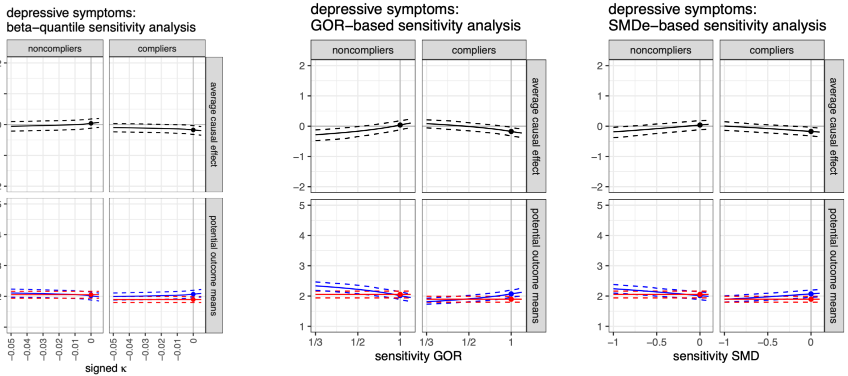




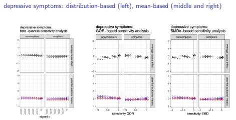


you can see that in this side-by-side shows
of results from the distribution-based and mean-based sens analyses for the outcome earnings

depressive symptoms: distribution-based (left), mean-based (middle and right)



2023-12-17



and here is for depressive symptoms
the first plot is distribution-based
the last two are mean-based

Outline

Outline

Background
Principal causal effects
Principal ignorability (PI) – one identification strategy
Sensitivity analysis for PI violation

Mean-based sensitivity analysis: expanding options
Allow different sensitivity parameters
Accommodate different estimation methods

Distribution-based sensitivity analysis: a new thing
Method limitations and information use
A method using full information (ongoing work)

to sum up
i have presented our work on sens analysis for principal ignorability violation
in the estimation of complier and noncomplier average causal effects

with two approaches
a mean-based approach and a distribution-based approach

THANK YOU

and i am happy to take questions