Does College Selectivity Reduce Obesity? A Partial Identification Approach

Giorgio Brunello
University of Padua

Dimitris Christelis
University of Glasgow

Anna Sanz-de-Galdeano
University of Alicante

Anastasia Terskaya
University of Barcelona

Obesity

- Obesity is a very big problem in the US
  - About 43% of Americans are obese

- Consistent evidence of negative implications not only for health but also for
  - wages
  - productivity
  - skill formation

Why would attending a selective college affect obesity after college

- Selective college attendance implies, on average, a higher human capital
  - Strong negative association between education and obesity, possibly because a higher human capital makes it easier to understand the costs of obesity
  - More likely to exercise more frequently, eat better, drink fewer sugary drinks

- Better education leads, on average, to a higher income, which makes it easier to afford healthier food

Causality

- Causal statements are very difficult to make:
  - Attending a selective college is endogenous – those who attend a more selective college are more likely to be from a more privileged background and less likely to be obese when entering college
  - lack of credible exogenous variation
  - treatment spillovers (violation of the SUTVA) are very likely because being obese can be affected by the behaviours of fellow students – results from standard methods (including RCTs) likely invalid when not accounting for the problem
  - time-invariant treatment, panel data analysis impossible

Causality (cont.)

- Most of the literature relies on methods that use selection on observables (OLS, panel data, matching, etc.). Evidence is quite robust about the negative association between selective college attendance and obesity

Contribution

- We derive informative identification regions for the average treatment of attending a selective college on being obese after college. Evidence at the individual level using partial identification (PI) methods in observational data

- PI results are robust to both endogeneity and violation of the SUTVA

- We rely on weak, and thus credible assumptions

Data

- National Longitudinal Study of Adolescent to Adult Health (Add Health). Longitudinal data starting at 1994-95, following children at 7^{th} to 12^{th} grades then, and till 2016-2018

- In Wave 4 (2008), respondents were 24-34 old and were asked whether they have a college degree, and if so, which college they attended. Colleges were classified as selective or not (primarily based on admission criteria) using Barron’s 2009 college selectivity measure (see also Chetty et al., 2017)

- Respondents were also observed in Wave 5 (2016-2018), when they were 33-44 years old

Data (cont.)

- Obesity was computed using measure height and weight
- Sample selection: interviewees who attended a four-year college in Wave 4, and for whom there is information on the Peabody Picture Vocabulary Test (PPVT), a cognitive test administered in high school
- Sample size: 3,578 individuals in Wave 4, 2,856 individuals in Wave 5
- In Wave 4, 26.2% are obese, and 20% have attended a selective college

Partial identification

- We use partial identification methods (Manski, 1990; Manski, 1997; Manski and Pepper, 2000)

- Starting premise:
  - the treatment is endogenous and there is no exogenous variation that can help identify the ATE
  - SUTVA is violated

- PI puts bounds around counterfactual outcomes, and thus around the average treatment effect (ATE), using credible assumptions

Partial identification

- Average treatment effect (outcome y, observed treatment w, generic treatment value d, controls x) :

$$E[y(d_2)|x]-E[y(d_1)|x]$$

- Average potential outcome:

$$E[y(d)|x]=E[y(d)\,|\,x,w=d]P(w=d|x)$$
$$+E[y(d)\,|\,x,w\neq d]P(w\neq d|x)$$

Simulated, fully specified models

- A simulated model allows one to evaluate the counterfactual $E[y(d)\,|\,x,w\neq d]$, typically in non-closed form as a policy function

- Treatment effect is calculated as the difference in the, typically, average outcomes when policy functions are evaluated at two different (sets of) values of the choice/parameter of interest

- Models make many assumptions (about optimizing behaviour, parameter values, functional forms, extent of heterogeneity, work, networks, exogeneity of errors, global maxima); not clear how each one affects the results

Regression models

- One can also evaluate the counterfactual $E[y(d)\,|\,x, w \neq d]$ using a regression model, by evaluating the estimating equation giving everybody two different values of the treatment, and keeping everything else constant

- In a linear model, the treatment effect is given by the relevant regression coefficient

- Regression models can, in principle, evaluate structural relationships, even as best linear approximations

- Regression requires exogenous variation, choices on which to include, assumptions about unobservables, functional forms, etc.

Partial identification

- PI, instead of evaluating $E[y(d) \mid x, w \neq d]$ using various strong assumptions, put bounds around it

- Bounds on $E[y(d) \mid x, w \neq d]$ imply bounds on average potential outcomes, and bounds on the differences in average potential outcomes, that is, average treatment effects

- PI is non-parametric, rests on few and credible assumptions, and it is absolutely transparent how each assumption affects results.

Partial Identification as a non-RCT (cont.)

- Bounding the counterfactual term implies bounding also $E[y(d)|x]$,

$$LB^M(d|x) \leq E[y(d)|x] \leq UB^M(d|x)$$

- Bounding $E[Y(d)]$ implies bounding also the ATE:

$$LB^M(d_2|x) - UB^M(d_1|x)$$
$$\leq E[y(d_2)|x] - E[y(d_1|x)] \leq$$
$$UB^M(d_2|x) - LB^M(d_1|x)$$

Partial Identification (cont.)

- For example,

$$0.2 \leq E[Y(0)|x] \leq 0.3$$
$$0.1 \leq E[Y(1)|x] \leq 0.15$$

$$0.1 - 0.3 = -0.2$$
$$\leq E[Y(1)|x] - E[Y(0)|x)] \leq$$
$$0.15 - 0.2 = -0.05$$

- To narrow identification regions, we need lower bounds that are as large as possible, and upper bounds that are as small as possible

Partial Identification (cont.)

- Without assumptions, the only credible bounds on the counterfactual terms are the minimum and maximum possible values (i.e., 0 and 1), which typically produce uninformative identification regions

$$E[y(d)\,|\,x,w=d]P(w=d|x)+\ y_{min}P(w\neq d|x)$$
$$\leq E[y(d)\,|\,x,w=d] \leq$$
$$E[y(d)\,|\,x,w=d]P(w=d|x)+\ y_{max}P(w\neq d|x)$$

- Using assumptions, we will replace y_{min} with something larger, and y_{max} with something smaller

Assumption 1: Monotone Treatment Response (Manski, Econometrica, 1997)

- Attending a more selective college does not increase the probability of being obese later on, ***on average***, ***keeping everything else constant, except for intermediate outcomes***
  - weak monotonicity, not sufficient to exclude 0 from the ATE identification region

$$E[y(1)|x, w=d] \leq E[y(0)|x, w=d], \text{ for every } d$$

- Plausible assumption – difficult to think why attending a more selective college would increase obesity on average. Has strong effect on results

Assumption 1: Monotone Treatment Response (Manski, Econometrica, 1997)

- Example of how MTR operates: the counterfactual mean obesity prevalence of those who actually did not attend a selective college, had they attended one. Under MTR, an upper bound for this counterfactual term is the actual obesity prevalence of those who did not attend a selective college

- This MTR upper bound is smaller than the overall maximum, and thus identification regions become narrower and thus more informative

Monotone treatment response (cont.)

- $UB^{MTR}(1) = LB^{MTR}(0) = E(y)$

- Hence, $UB_{ATE}^{MTR}(1-0) = 0$

- After combining with MIV assumption,

$$UB_{ATE}^{MTR+MIV}(1-0) < 0$$

Assumption 2: Monotone Instrumental Variable (Manski and Pepper, Econometrica, 2000)

- MIVs can be weakly monotonically associated with the outcome, and are thus much easier to find

- They do not identify the LATE but the ATE – nothing to do with compliers

- They serve to partition the sample space conditional on their values and search for maximum lower bounds and minimum upper bounds
- Need to narrow the identification region
  - LB example: $LB(d) = \sum_z LB(d|Z = z)P(Z = z)$
  - need to increase $LB(d|Z = z)$, as $P(Z = z)$ is given by the data

Monotone instrumental variable (MIV)

- An instrument z is monotone if

$$z_1 < z_2 \rightarrow E[Y(d)|Z = z_1] \leq E[Y(d)|Z = z_2]$$

- Assumption is about potential outcome $E[Y(d)]$, hence it is untestable. It also reflects a simple correlation

- Our MIV is the PPVT test taken in high school. Higher cognition should be weakly negatively associated with obesity

How do MIVs work? (cont.)

- Typically, the finer the partition (achieved, e.g., by using more than one MIVs) the narrower the identification regions

- Local minima and maxima are perfectly OK – they just lead to wider identification regions

- Different instruments possibly lead to different identification regions – this is fine, as the unobserved average potential outcomes do not change

- Constraint: number of observations

Non-individualistic response (Manski, 2013)

- $E[y(d^J)]$, i.e., average potential outcomes depend on the treatments received by others in a set J (here assumed to be the population)
- Problem: $E[y(d^J)|w = d] \neq E(y|w = d) = E[y(d)|w = d]$
- Reinforcing interactions (RI) assumption: for any two treatment vectors d_1^J and d_2^J such that $d_2^k \geq d_1^k \;\; \forall\, k \in J$,

$$E[y(d_2^J)|w = d] \leq E[y(d_1^J)|w = d]$$

- MTR and RI operate in the same direction: credible in this context

Non-individualistic responses (cont.)

- Just as with MTR, $UB^{RI}(1) = LB^{RI}(0) = E(y)$
- Hence, $UB_{ATE}^{RI}(1-0) = 0$
- After combining with MIV, $UB_{ATE}^{RI+MIV}(1-0) < 0$
- MIV operates with RI exactly as with MTR

PI results for consumption – individualistic response – wave 4

Assumptions	Estimates		Lower Bound Low 95% CI	Lower Bound Upper 95% CI	Upper Bound Low 95% CI	Upper Bound Upper 95% CI
	Lower bound	Upper bound				
Panel A. Wave IV						
Exogenous treatment selection	-0.0781		-0.1233		-0.0297	
No assumptions	-0.3831	0.6169	-0.4223	-0.3490	0.5777	0.6510
MTR	-0.3831	0.0000	-0.4223	-0.3490	0.0000	0.0000
MTR + MIV (bias-corrected)	-0.3088	-0.0196	-0.3175	-0.2247	-0.0688	-0.0112
MTR + MIV (bias-uncorrected)	-0.2485	-0.0556	-0.3175	-0.2247	-0.0688	-0.0112
Number of observations			3,525			

PI results for mechanisms – wave 4

Outcomes	Estimates		Lower Bound Low 95% CI	Lower Bound Upper 95% CI	Lower Bound Low 95% CI	Upper Bound Upper 95% CI
	Lower bound	Upper bound				
Eaten fast food (bias-corrected)	-0.3456	-0.0186	-0.4029	-0.2610	-0.0753	-0.0182
Hours spent watching TV (bias-corrected)	-9.8746	-0.2808	-11.2038	-8.3985	-1.0435	-0.1111
Number of sweet drinks (bias-corrected)	-8.2882	-0.4022	-9.9039	-6.5639	-1.2377	-0.2565
Being physically active (bias-corrected)	-0.1278	-0.0208	-0.1777	-0.0737	-0.0611	-0.0206
Log of income (bias-corrected)	0.0652	1.5779	0.0748	0.2055	1.3438	1.7568

Robustness checks

- Corresponding effect for obesity in Wave 5 (10-12 years after graduation) is at least 1.6 pp. Mechanisms operate also in Wave 5

- Results are similar when using abdominal obesity (waist circumference)

- Results are similar when using alternative MIVs: GPA, socioeconomic status, lagged BMI by one wave

Conclusions

- We find a negative causal effect of attending a selective college in the probability of being obese 2 to 12 years after graduation that is at least 2 pp - non-trivial effect when obesity prevalence is 26%. The corresponding effect 10 to 20 years after graduation is at least 1.6 pp.

- Results obtained using PI methods using weak, and thus credible, assumptions

- Observational data (even cross-sectional ones) can be used to make causal statements, without the use of exogenous variation while taking into account treatment spillovers

- Limitations: increased uncertainty

THANK YOU!

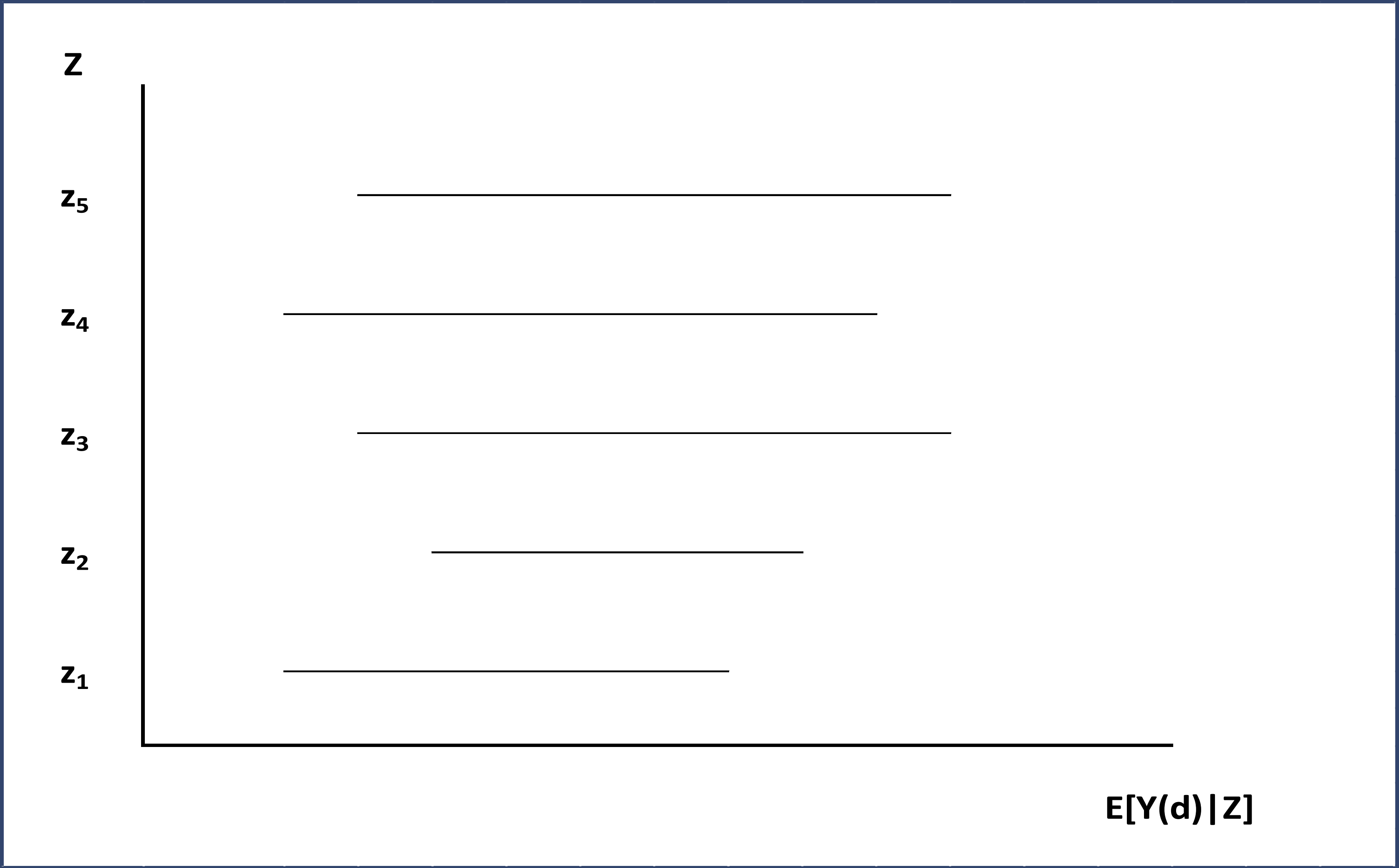
How do MIVs work? - XIVs

- Let us assume for the moment that the instrument is exogenous (XIV)

$$E[Y(d)|Z = z] = E[Y(d)], \forall\, z$$

- Need to narrow the identification region
  - LB example: $LB(d) = \sum_z LB(d|Z = z)P(Z = z)$
  - need to increase $LB(d|Z = z)$, as $P(Z = z)$ is given by the data
  - Correspondingly, need to decrease $UB(d|Z = z)$

How do MIVs work? - XIVs (cont.)



How do MIVs work? - XIVs (cont.)

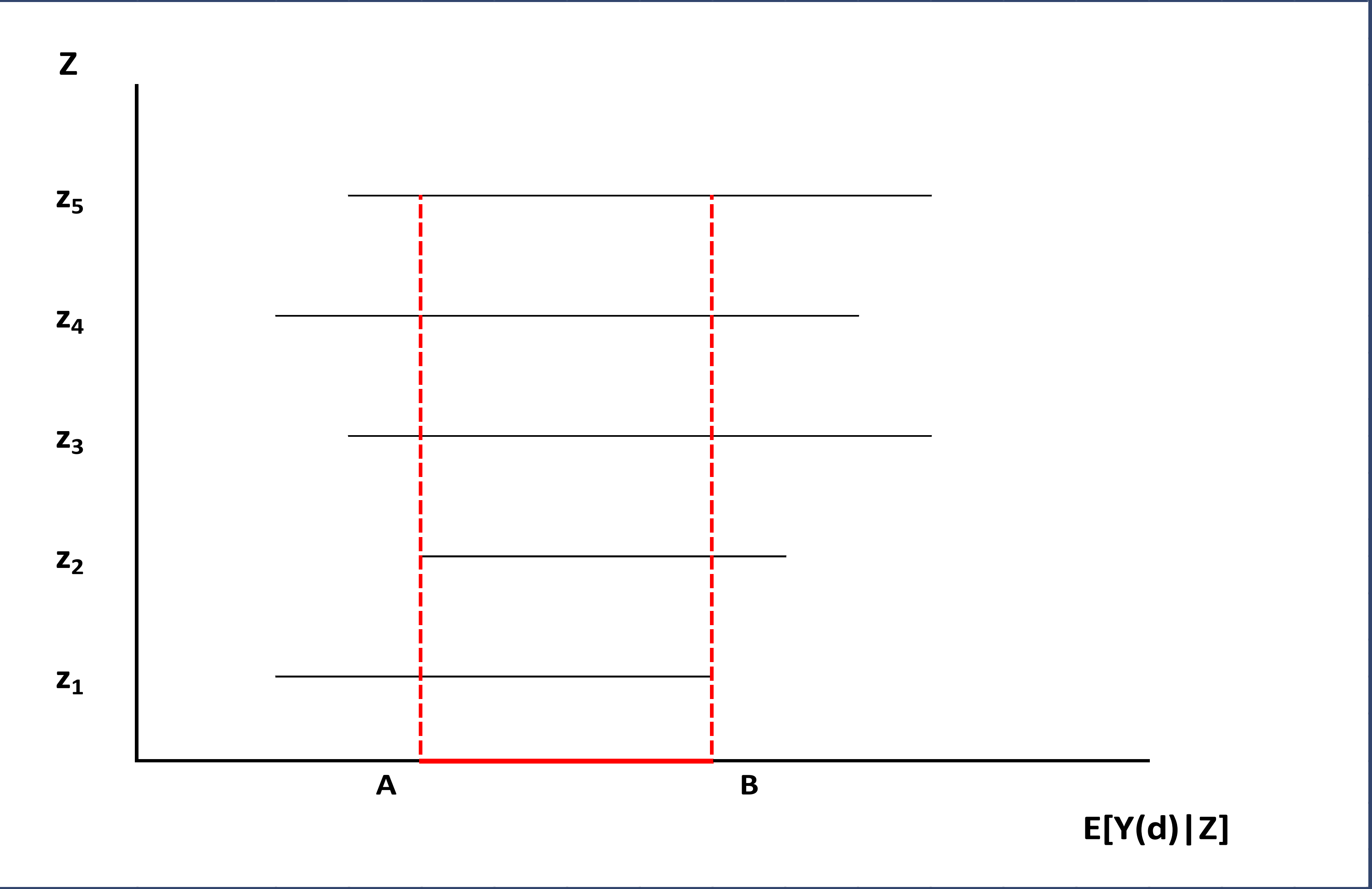
- Since the instrument is exogenous, all conditional identification regions should give the same answer

- Same answer → the intersection of all regions

- The intersection is given by the maximum lower bound and the minimum upper bound

$$\max_{z} LB[d|Z=z] \leq E[Y(d)] \leq \min_{z} UB[d|Z=m]$$

How do MIVs work? - XIVs (cont.)



How do MIVs work?

$$z_1 \leq z_2 \rightarrow E[Y(d)|Z = z_1] \leq E[Y(d)|Z = z_2]$$

- For a given MIV value can examine only what happens to
  - lower bounds at lower instrument values
  - upper bounds of higher instrument values

How do MIVs work? (cont.)

- Hence, the final bounds for the identification region conditional on MIV value z_3 are:
  - the maximum lower bound over z_1, z_2, z_3, which is the one corresponding to MIV value z_2
  - the minimum upper bound over z_3, z_4, z_5, which is the one corresponding to MIV value z_4
- Analogous calculations performed for identification regions conditional on MIV values other than z_3
- Finally, overall bounds are computed by integrating out the MIV

How do MIVs work? (cont.)

- Thus, optimization takes place only for a restricted range of values – identification regions less informative
- To be expected, as the MIV assumption is weaker than the XIV one