Topics in instrumental variable estimation: structural mean models and bounds

Tom Palmer

Division of Health Sciences, Warwick Medical School, University of Warwick, UK

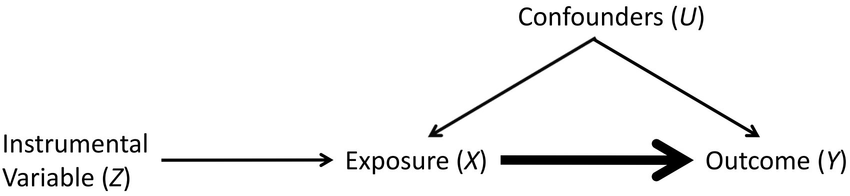
30 May 2013

Outline

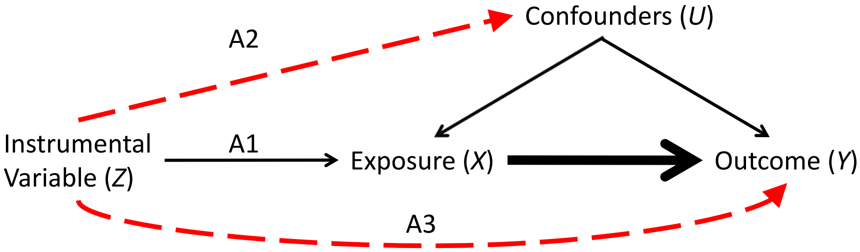
1. Introduction to instrumental variables
   - Assumptions
   - Applications
     - Mendelian randomization
     - Noncompliance in RCTs
   - Test of presence of effect
   - Estimators: ratio, two-stage least squares
2. Structural mean models
   - Potential outcomes and causal parameters
   - Additive SMM: G-estimation example
   - Multiplicative SMM: estimation with multiple instruments
   - (double) Logistic SMM
3. Nonparametric bounds
   - Extensions
   - Limitations
4. Summary

I. Introduction to instrumental variables

Assumptions



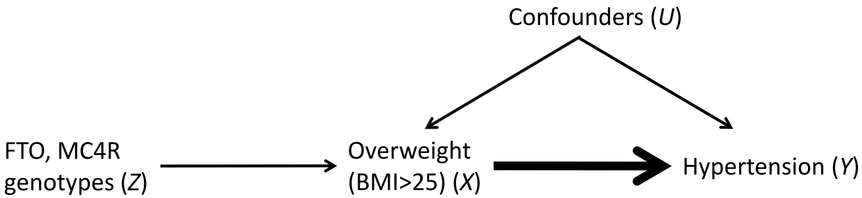
Assumptions



A1. Instrumental variable associated with exposure ($Z \not\perp\!\!\!\perp X$)
A2. Instrumental variable independent of confounders ($Z \perp\!\!\!\perp U$)
A3. No direct effect of instrumental variable on outcome
exclusion restriction ($Y \perp\!\!\!\perp Z \mid (X, U)$)

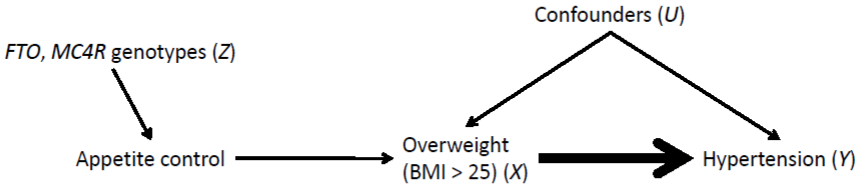
Application: Mendelian randomization

Use of genotypes robustly associated with exposures
(from replicated genome-wide association studies, $P < 5 \times 10^{-8}$)
as instrumental variables (Davey Smith & Ebrahim, 2003)



Application: Mendelian randomization

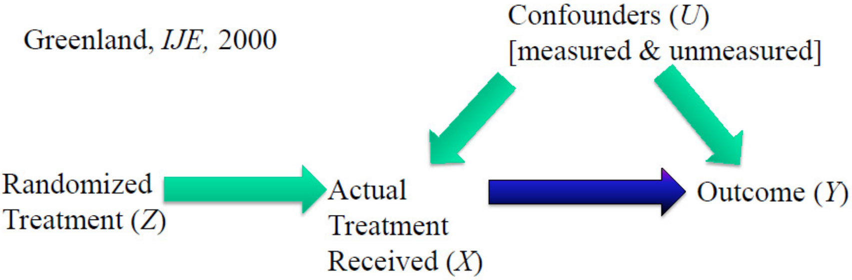
Use of genotypes robustly associated with exposures
(from replicated genome-wide association studies, $P < 5 \times 10^{-8}$)
as instrumental variables (Davey Smith & Ebrahim, 2003)



IV does not have to be causal for exposure

Application: randomized controlled trials

Correcting for noncompliance in randomized controlled trials

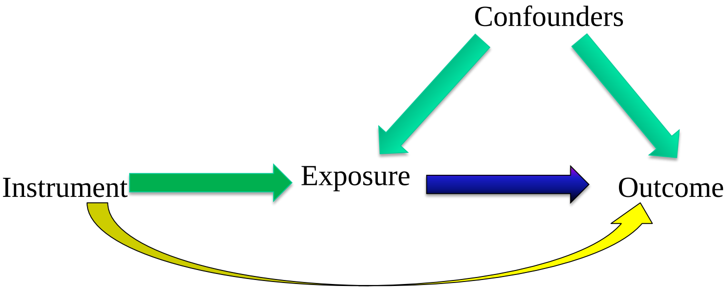


Test of presence of causal effect

If the IV conditions hold, then a test of the

instrument-outcome association

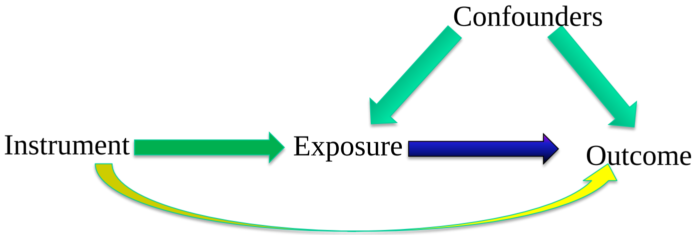
is a test for the presence of a causal effect of exposure on outcome



Ratio estimator

Assumptions: everything linear, binary instrument

IV estimate of Exposure on Outcome = $\dfrac{\text{Instrument-outcome association } (ZY)}{\text{Instrument-exposure association } (ZX)}$



Standard error & confidence interval from delta-method/Fieller's theorem (Thomas et al. *Ann Epi*, 2007)

Two-stage least squares estimator

Estimation with multiple instruments
Two-stage least squares

$$X = \beta_0 + \beta_{z1} Z_1 + \ldots + \beta_{zn} Z_n$$

Stage 1

- Regress Exposure on instrument/s
- Generate predicted values of exposure

$$\hat{X} = \hat{\beta}_0 + \hat{\beta}_Z Z$$

Stage 2

- Regress Outcome on predicted values of exposure
- (***adjust standard errors***)

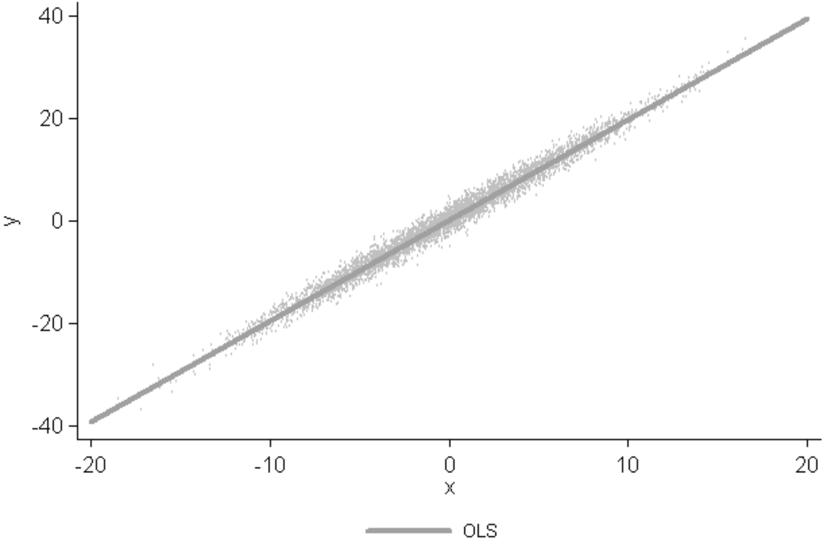
$$Y = \alpha + \beta_{XY} \hat{X}$$

Stata comands: `ivregress`, `ivreg2`

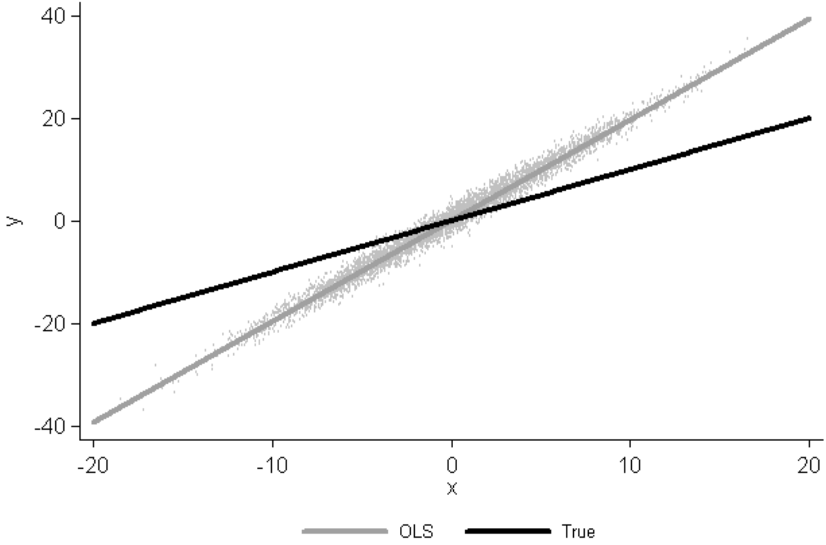
Automatically correct standard errors!!

Binary outcome: parameter = risk difference

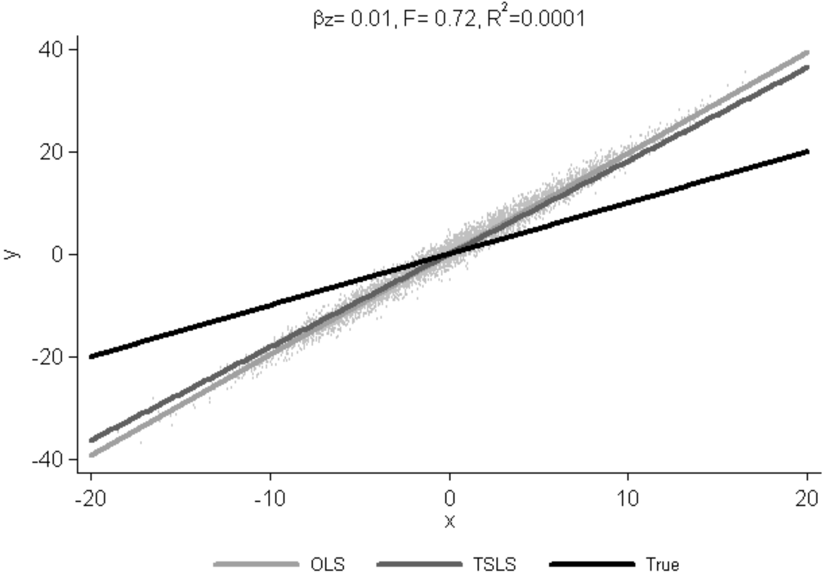
Instrument strength example



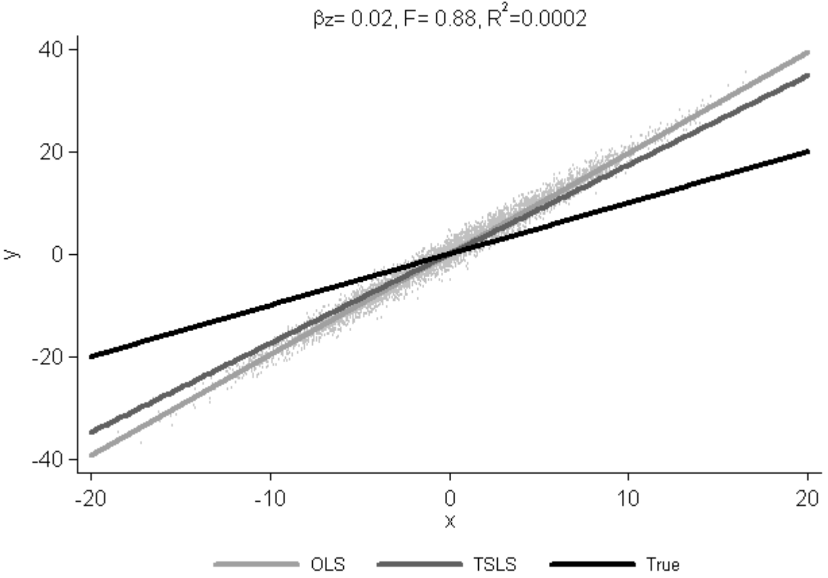
Instrument strength example



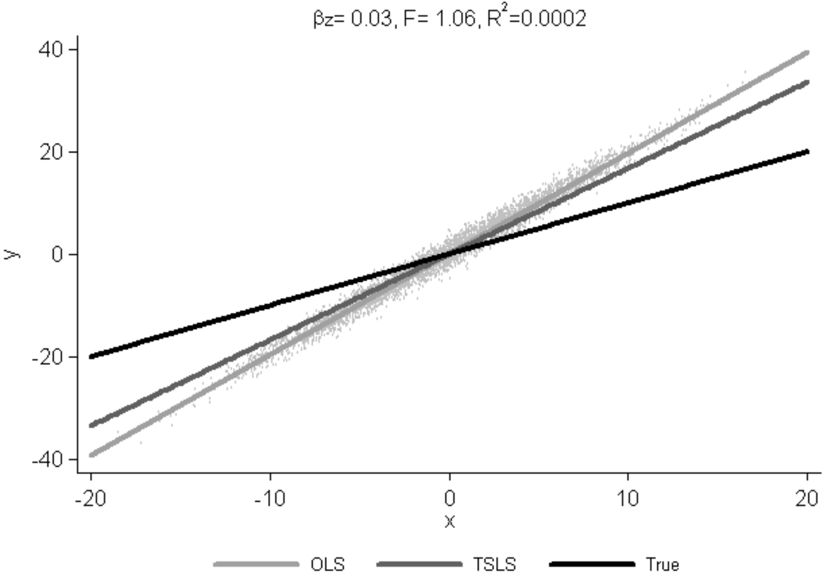
Instrument strength example



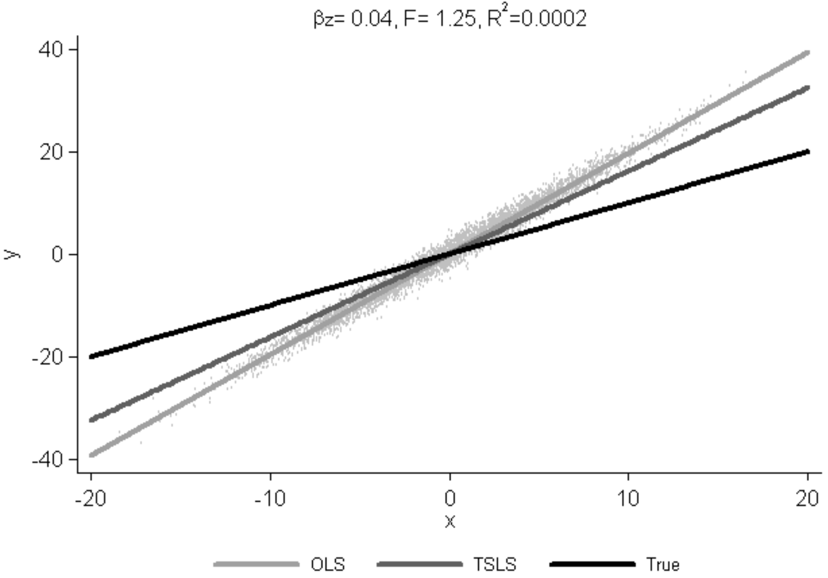
Instrument strength example



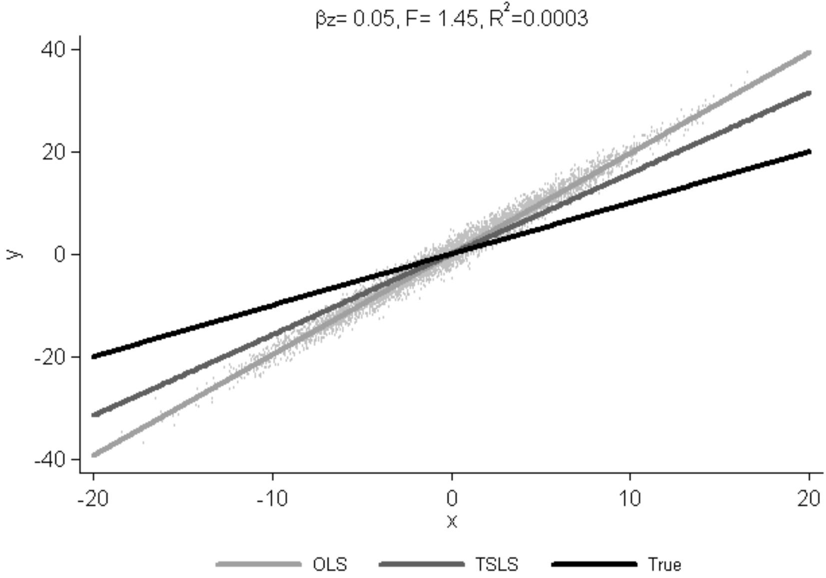
Instrument strength example



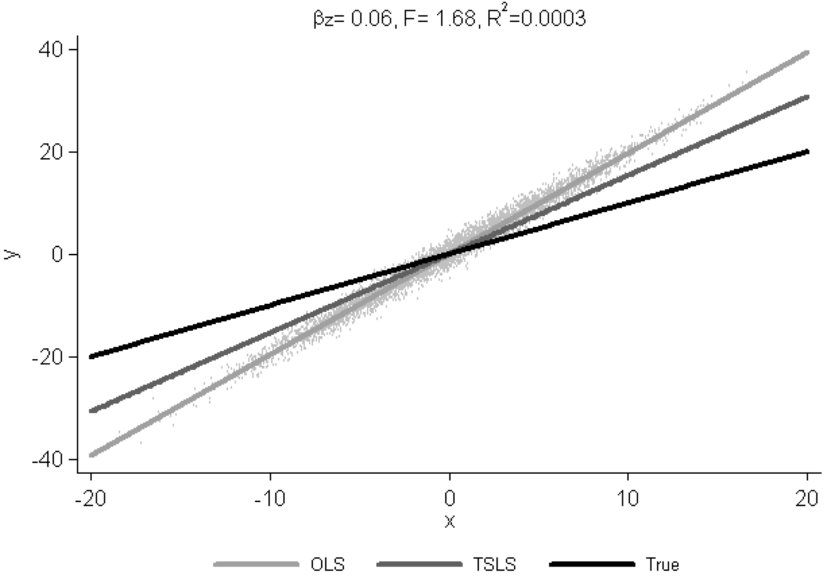
Instrument strength example



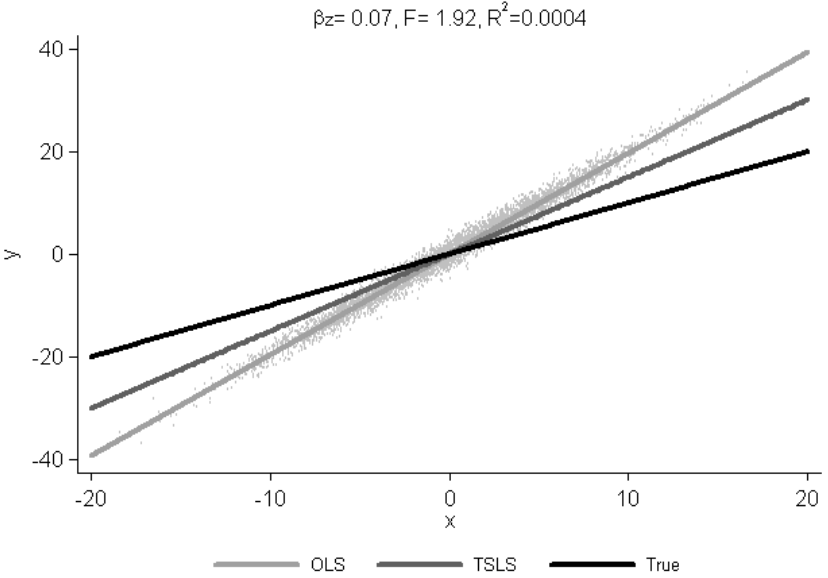
Instrument strength example



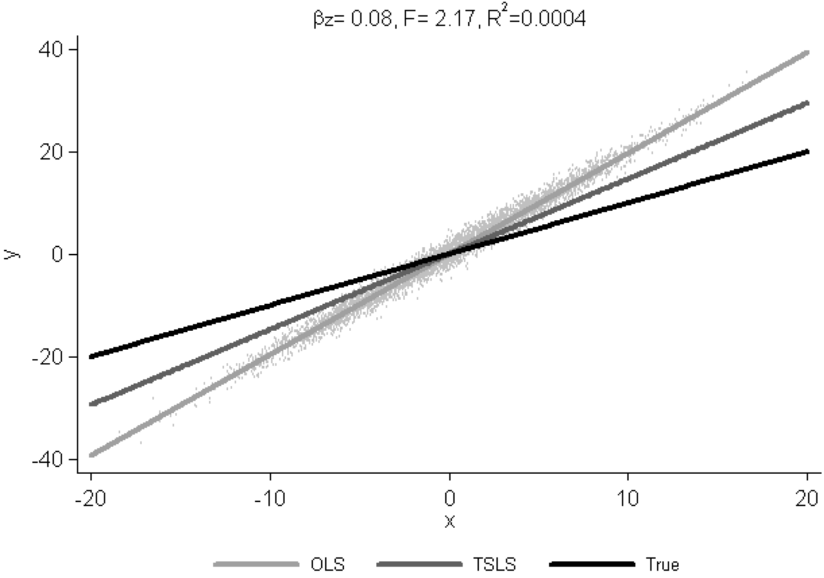
Instrument strength example



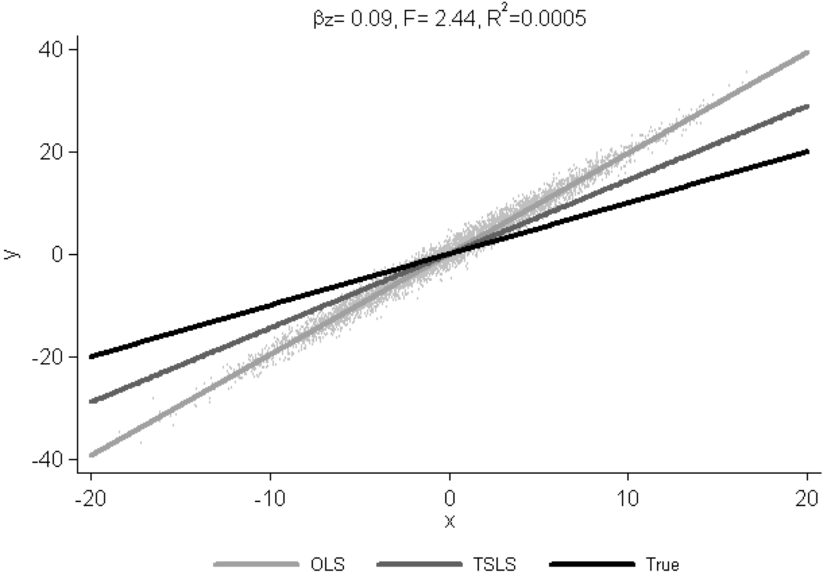
Instrument strength example



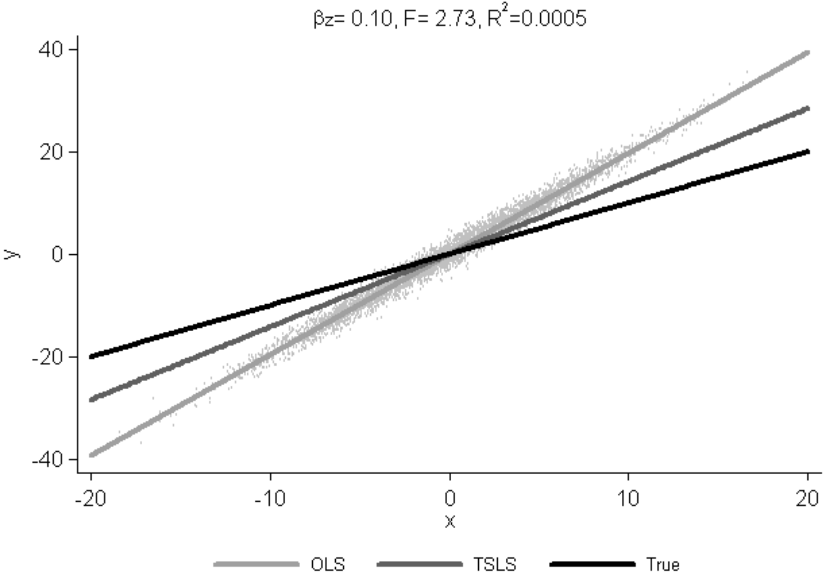
Instrument strength example



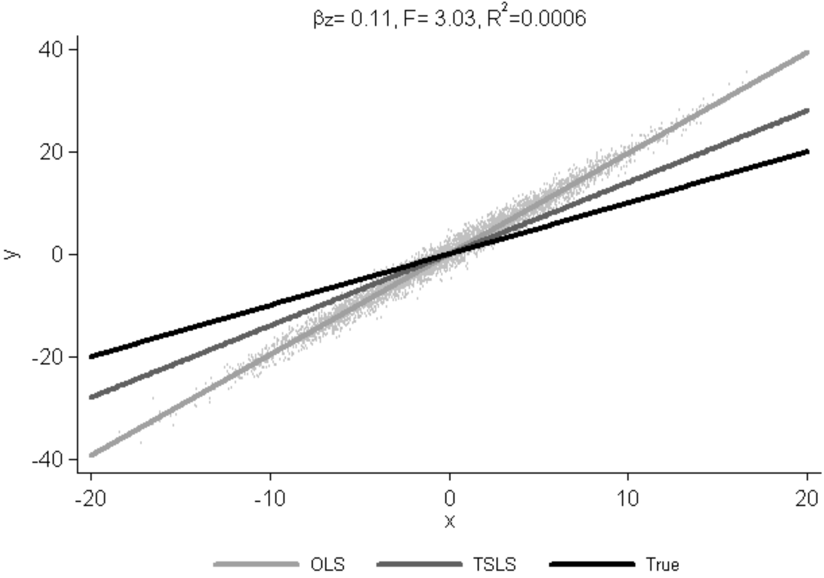
Instrument strength example



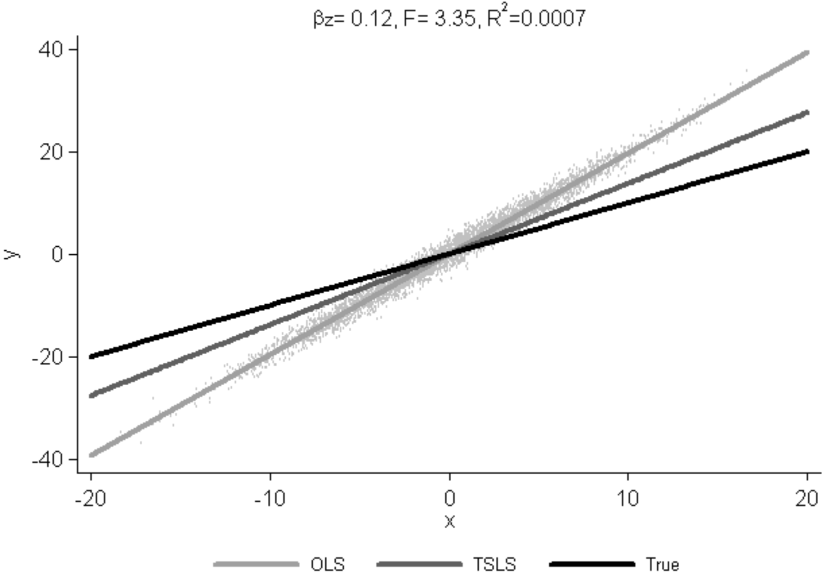
Instrument strength example



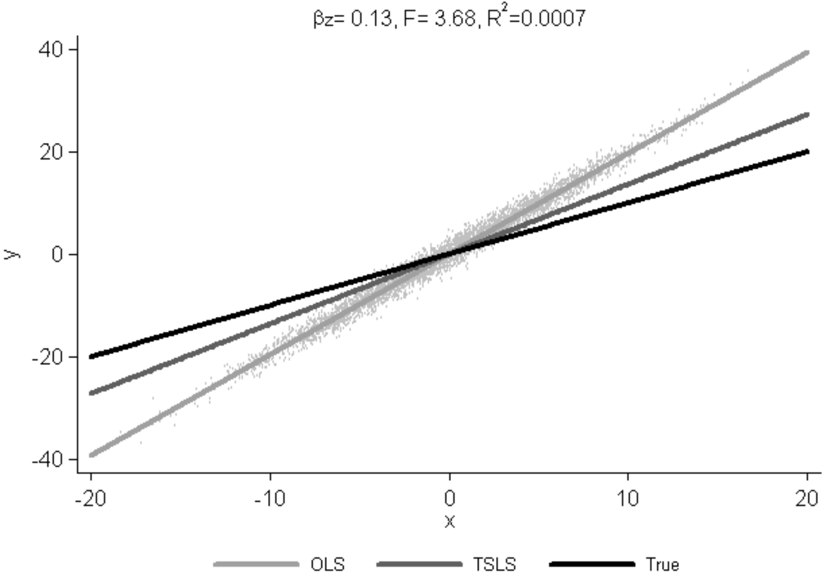
Instrument strength example



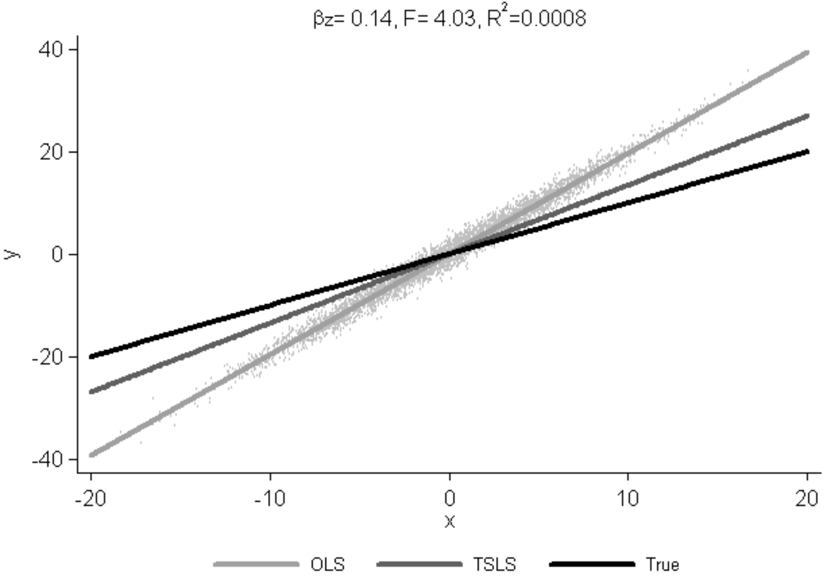
Instrument strength example



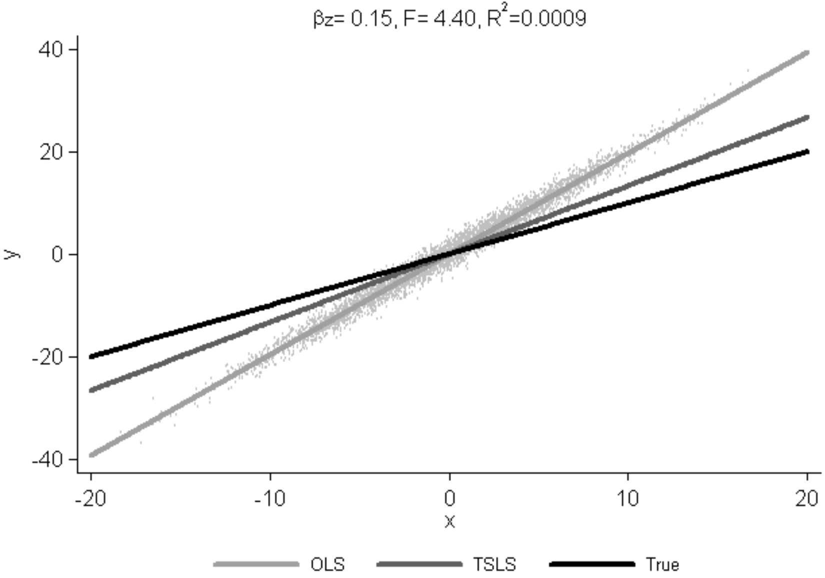
Instrument strength example



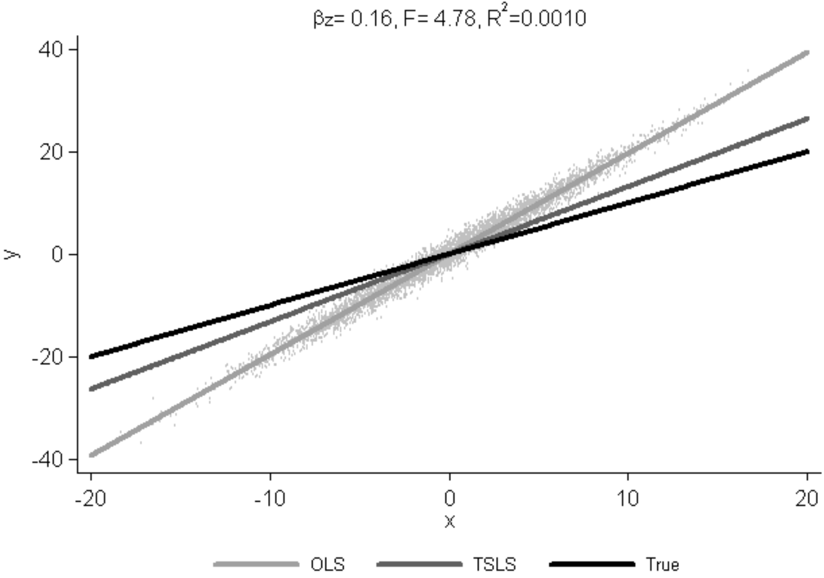
Instrument strength example



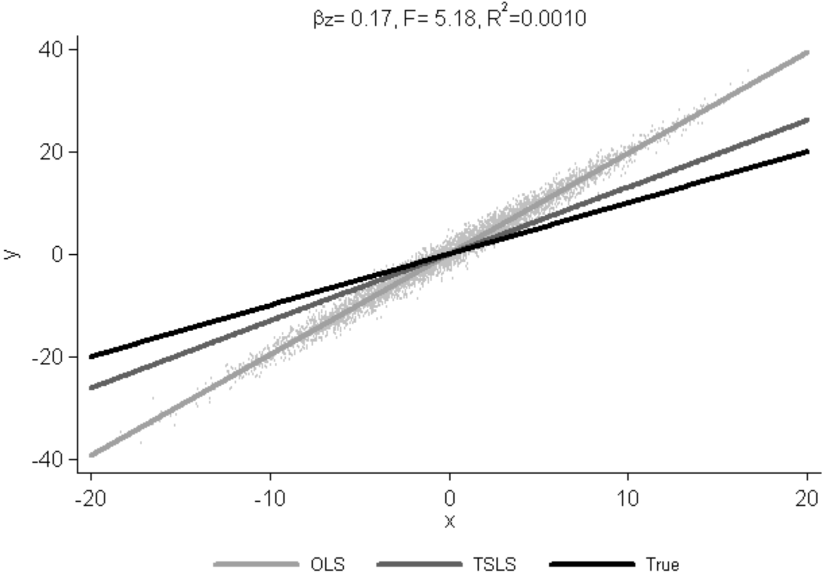
Instrument strength example



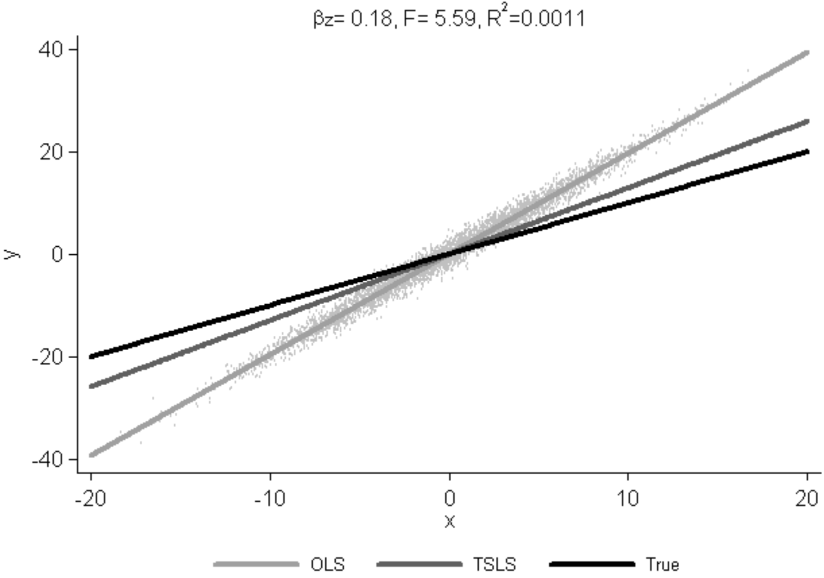
Instrument strength example



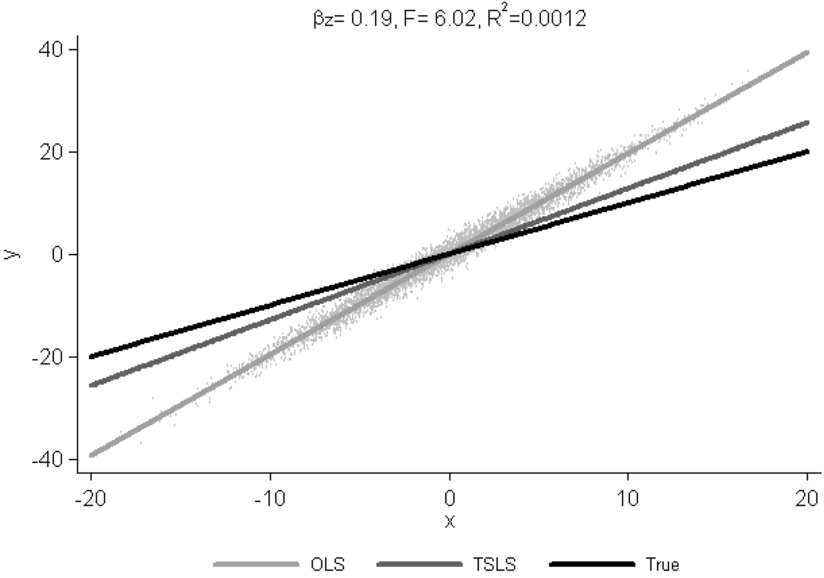
Instrument strength example



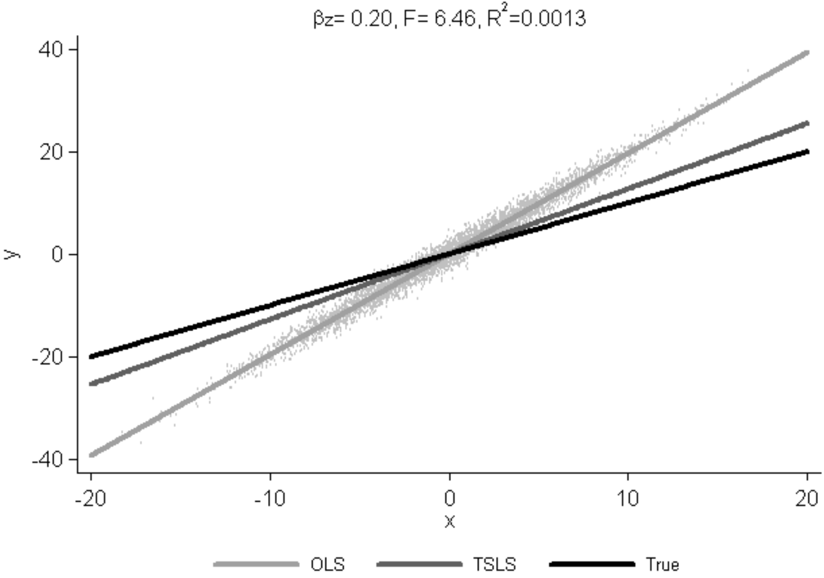
Instrument strength example



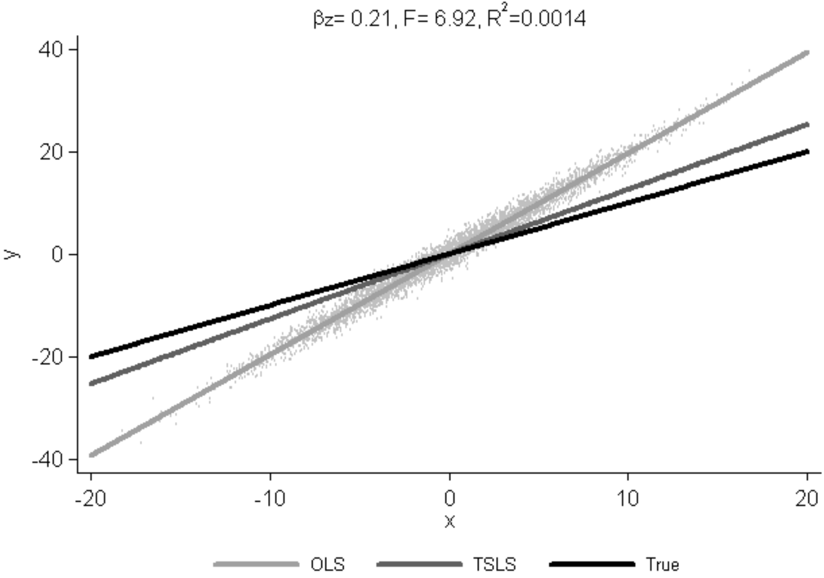
Instrument strength example



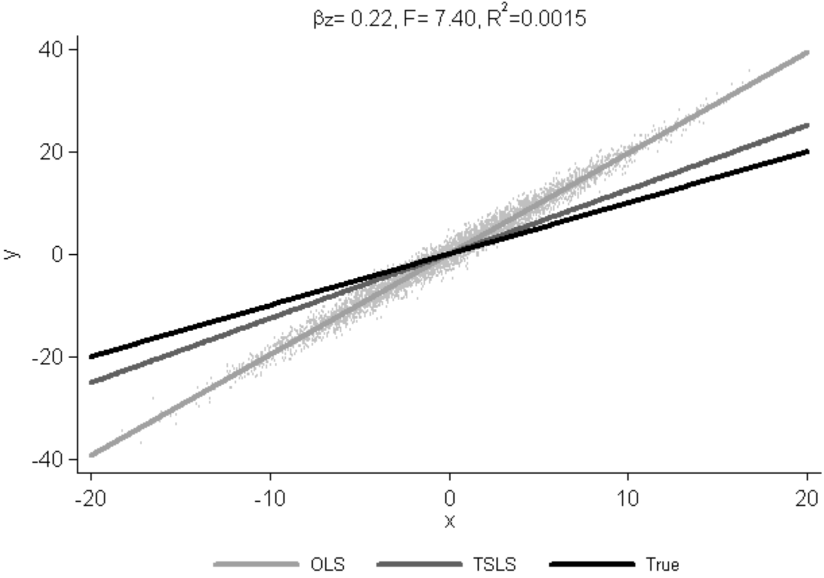
Instrument strength example



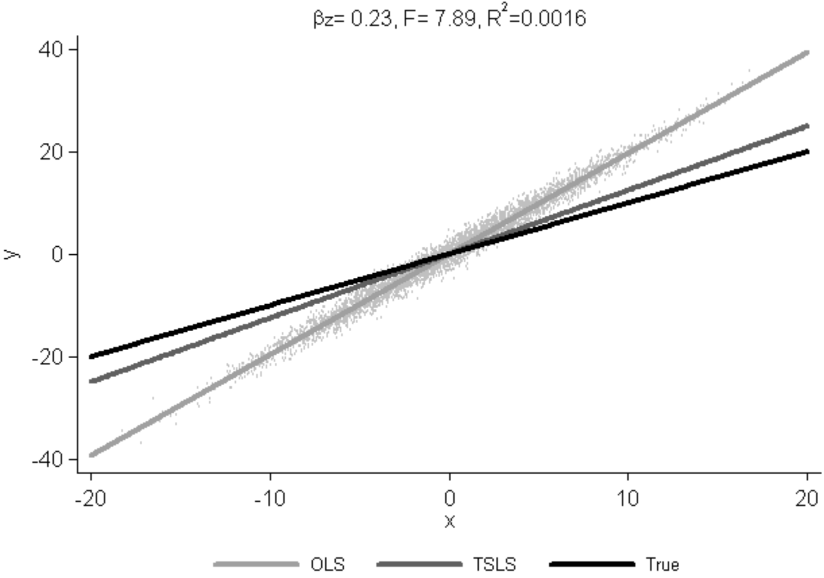
Instrument strength example



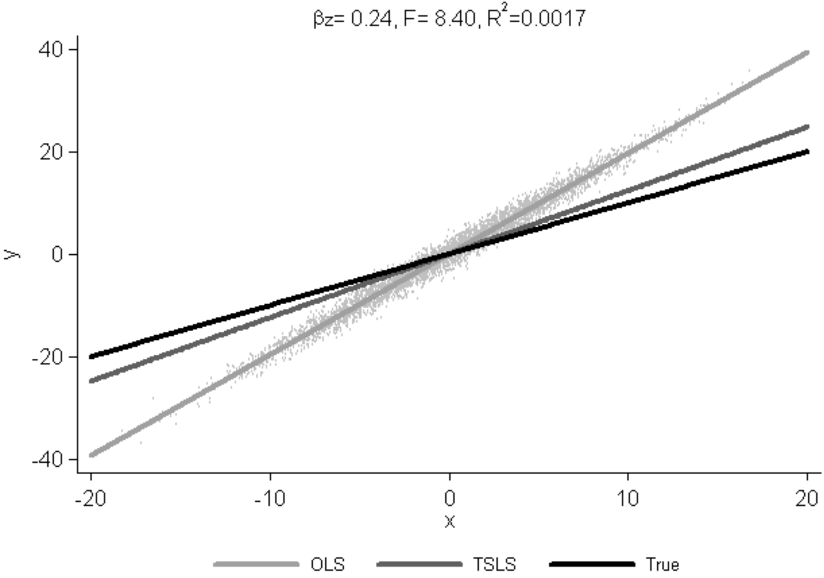
Instrument strength example



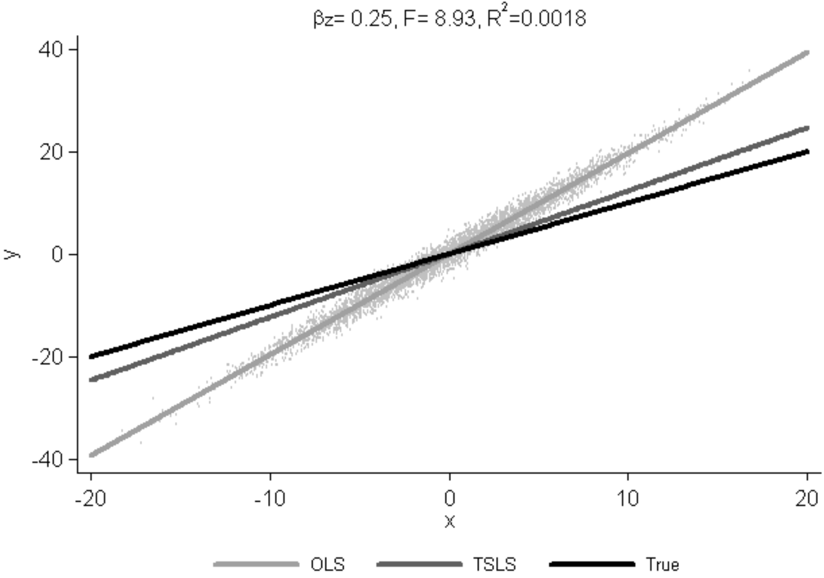
Instrument strength example



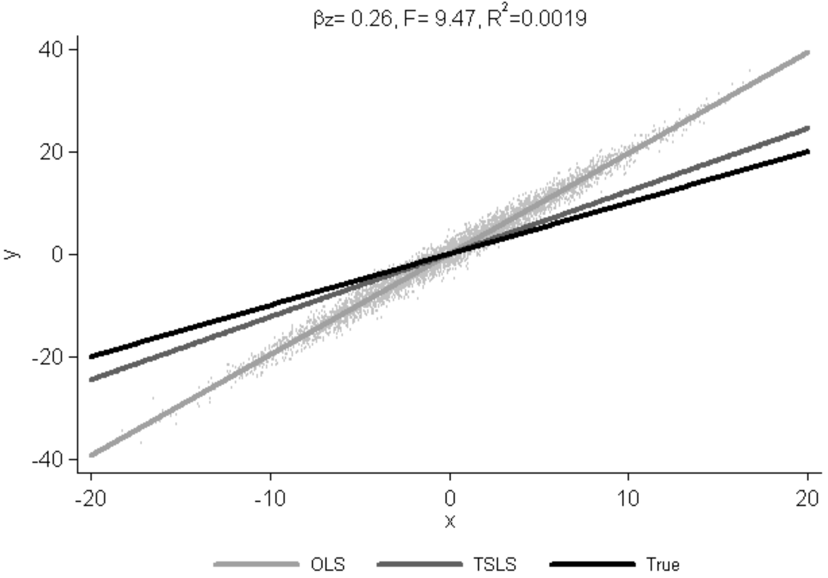
Instrument strength example



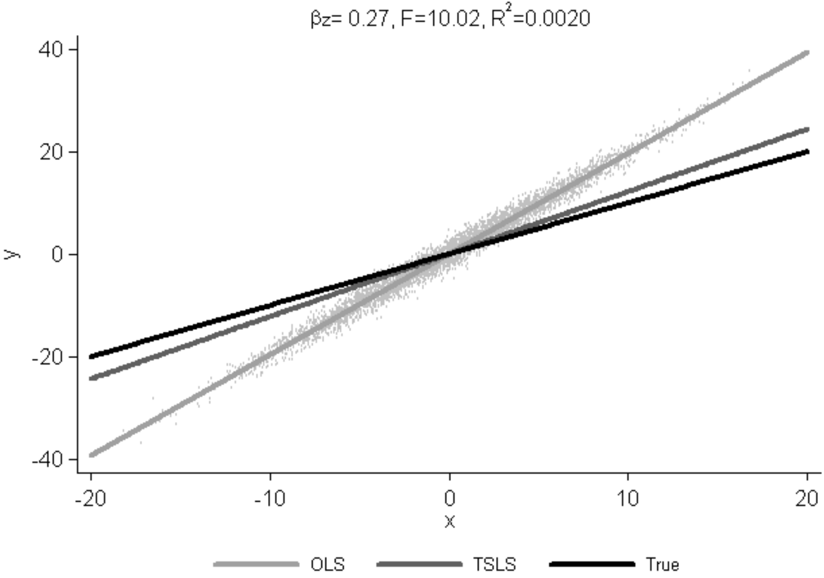
Instrument strength example



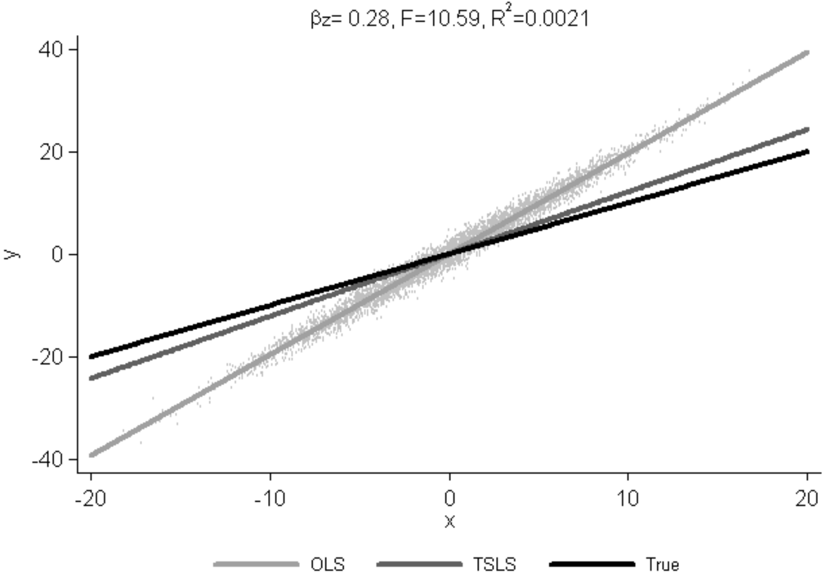
Instrument strength example



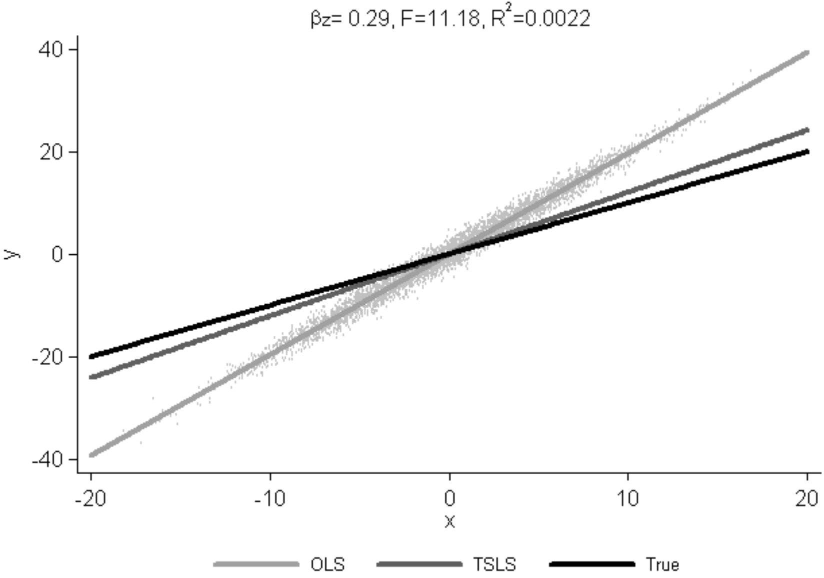
Instrument strength example



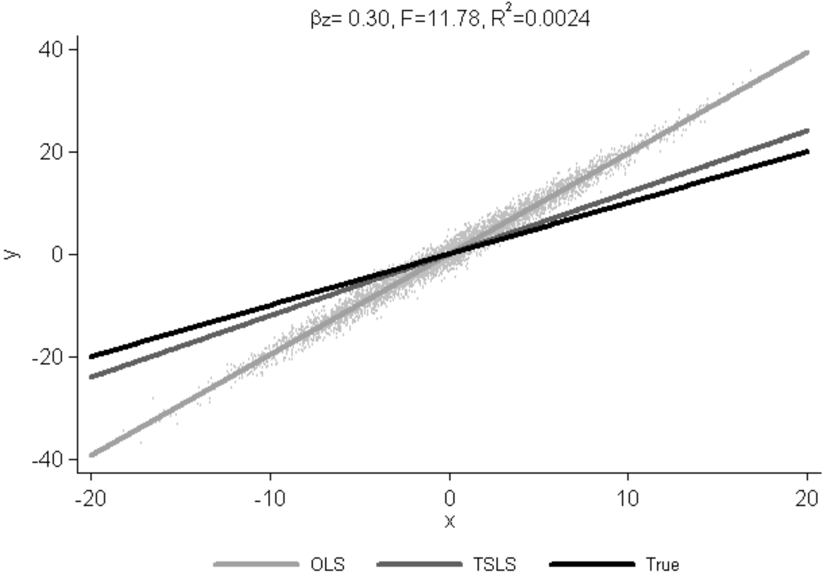
Instrument strength example



Instrument strength example

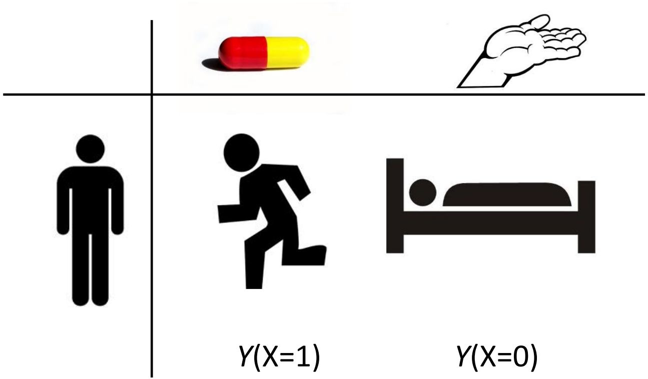


Instrument strength example



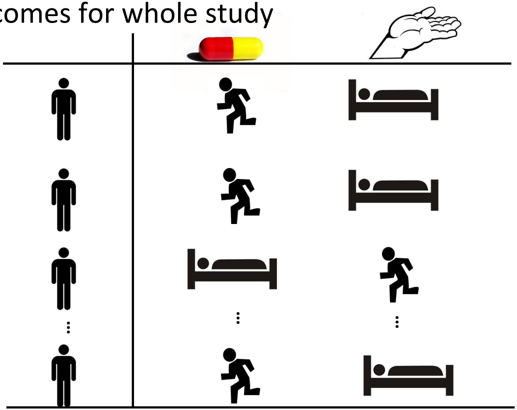
II. Structural mean models (with Frank Windmeijer & Paul Clarke)

Potential outcomes for an individual



Potential outcomes

Potential outcomes for whole study



Average treatment effect = $E[Y(X=1)] - E[Y(X=0)]$

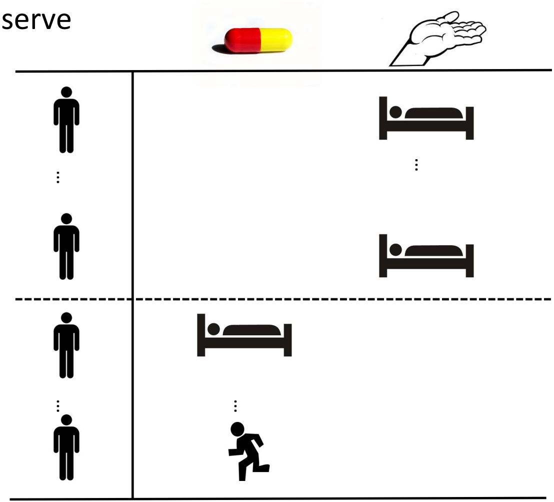
binary outcome: causal risk difference

Causal risk ratio = $E[Y(X=1)] \,/\, E[Y(X=0)]$

Causal odds ratio = $\text{odds}[Y(X=1)] \,/\, \text{odds}[Y(X=0)]$

Potential outcomes

What we observe

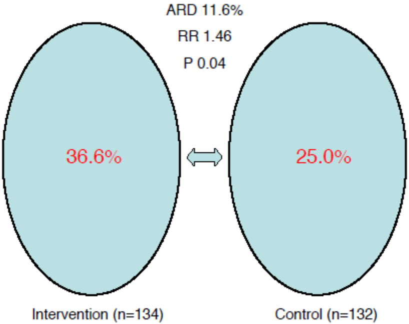


E[Y(1)|X=1] – E[Y(0)|X=1]

SMMs identify effect of treatment of treated

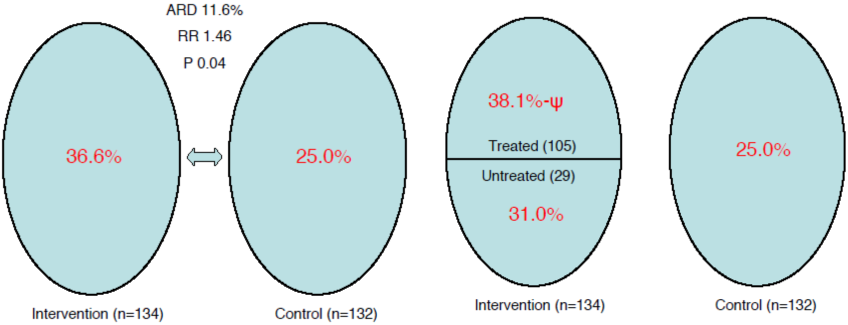
Additive SMM example I

- Ten Have, Elliott, Joffe, Zanutto, & Datto, 2004 266 African American adults with high cholesterol and/or hypertension
- Control: usual care (conventional nutritional info)
- Intervention: usual care + home-based audio tapes
- Outcome: +ve response: beneficial change in cholesterol,
  −ve response: otherwise

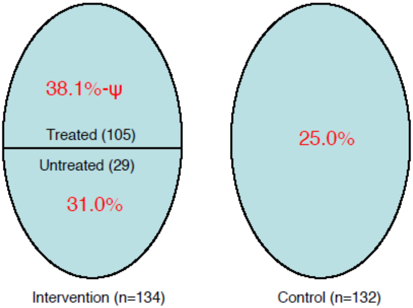


Additive SMM example I

- Ten Have et al., 2004 266 African American adults with high cholesterol and/or hypertension
- Control: usual care (conventional nutritional info)
- Intervention: usual care + home-based audio tapes
- Outcome: +ve response: beneficial change in cholesterol,
  −ve response: otherwise



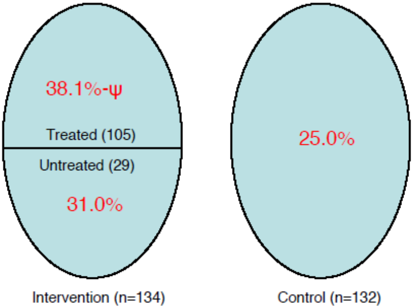
Additive SMM example II



Ratio estimate: $\psi = \dfrac{E(Y|Z=1) - E(Y|Z=0)}{E(X|Z=1) - E(X|Z=0)} = \dfrac{36.6 - 25.0}{105/134 - 0}$

$= 11.6/78.4 = 14.8\%$ (95%CI $0.8\%, 28.7\%$; $P = 0.04$)

Additive SMM example II



$$\text{Ratio estimate: } \psi = \frac{E(Y|Z=1) - E(Y|Z=0)}{E(X|Z=1) - E(X|Z=0)} = \frac{36.6 - 25.0}{105/134 - 0}$$

$$= 11.6/78.4 = 14.8\% \text{ (95\%CI } 0.8\%, 28.7\%; P = 0.04)$$

G-estimation: what would have happened if no-one was treated

ASMM estimate: $(38.1 - \psi)(105/134) + 31.0(29/134) = 25.0$

$$\psi = (38.1 \times 105 + 31.0 \times 29 - 25.0 \times 134)/105 = 14.8\%$$

Multiplicative SMM

- Notation: X exposure/treatment, Y outcome, Z instrument, $Y(X=0)$ exposure/treatment free potential outcome

Robins, Rotnitzky, & Scharfstein, 1999; Hernán & Robins, 2006

$$\log(E[Y|X,Z]) - \log(E[Y(0)|X,Z]) = \psi X$$
$$\frac{E[Y|X,Z]}{E[Y(0)|X,Z]} = \exp(\psi X)$$
$$\psi: \text{ log causal risk ratio}$$
$$\text{Rearrange: } Y(0) = Y\exp(-\psi X)$$

Mutliplicative SMM: estimation with multiple instruments

Under the instrumental variable assumptions (Robins, 1989):

$$Y(0) \perp\!\!\!\perp Z$$
$$Y \exp(-\psi X) \perp\!\!\!\perp Z$$

Mutliplicative SMM: estimation with multiple instruments

Under the instrumental variable assumptions (Robins, 1989):

$$Y(0) \perp\!\!\!\perp Z$$
$$Y \exp(-\psi X) \perp\!\!\!\perp Z$$
$$\text{trick: } Y \exp(-\psi X) - Y(0) \perp\!\!\!\perp Z$$

Mutliplicative SMM: estimation with multiple instruments

Under the instrumental variable assumptions (Robins, 1989):

$$
\begin{aligned}
Y(0) &\perp\!\!\!\perp Z \\
Y\exp(-\psi X) &\perp\!\!\!\perp Z \\
\text{trick: } Y\exp(-\psi X) - Y(0) &\perp\!\!\!\perp Z
\end{aligned}
$$

Moment conditions
Z=0,1

$$
\begin{aligned}
E[(Y\exp(-\psi X) - Y(0))1] &= 0 \\
E[(Y\exp(-\psi X) - Y(0))Z_1] &= 0
\end{aligned}
$$

Mutliplicative SMM: estimation with multiple instruments

Under the instrumental variable assumptions (Robins, 1989):

$$Y(0) \perp\!\!\!\perp Z$$
$$Y \exp(-\psi X) \perp\!\!\!\perp Z$$
$$\text{trick: } Y \exp(-\psi X) - Y(0) \perp\!\!\!\perp Z$$

Moment conditions

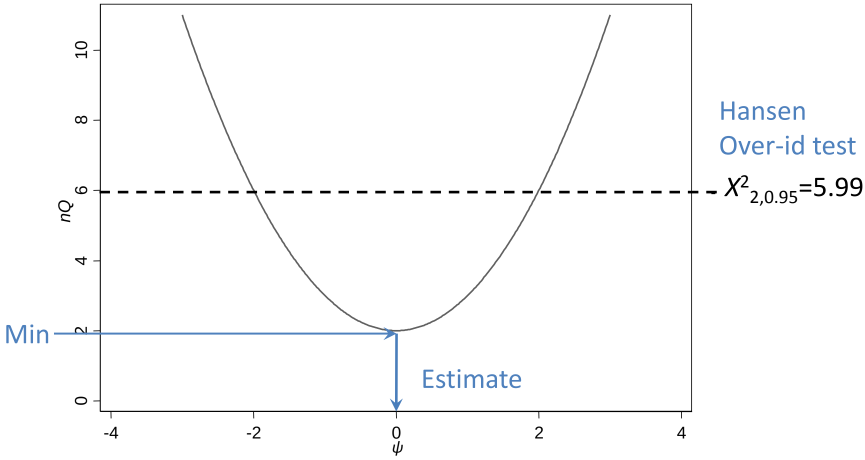
Z=0,1,2,3

Over-identified

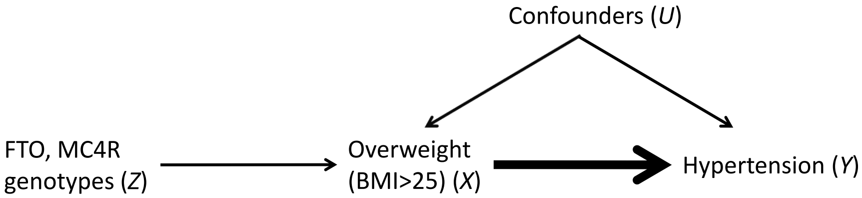
$$E[(Y \exp(-\psi X) - Y(0))1] = 0$$
$$E[(Y \exp(-\psi X) - Y(0))Z_1] = 0$$
$$E[(Y \exp(-\psi X) - Y(0))Z_2] = 0$$
$$E[(Y \exp(-\psi X) - Y(0))Z_3] = 0$$

Generalised Method of Moments

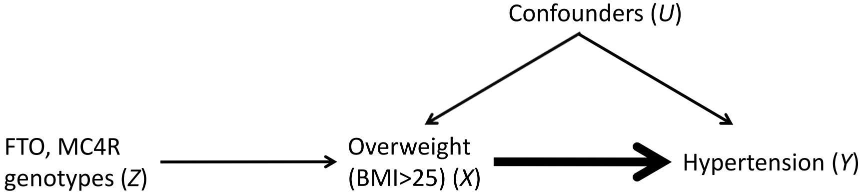
Minimises quadratic form: $Q = m' W^{-1} m$



Copenhagen example descriptive statistics 1



Copenhagen example descriptive statistics 1



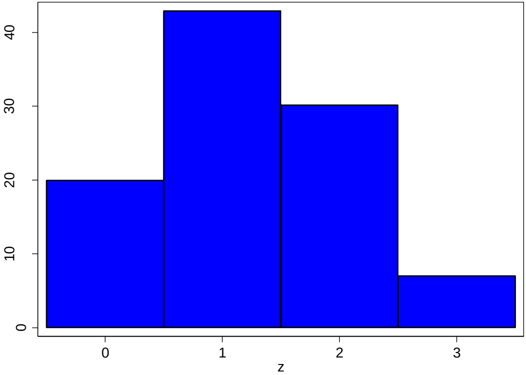
	No Hypertension	Hypertension	Total
Not Overweight	10,066 42%	13,909 58%	23,975
Overweight	6,906 22%	24,642 78%	31,548
Total	16,972 31%	38,551 69%	55,523 χ^2 P<0.001

Risk ratio for hypertension 1.35 (1.32, 1.37)

Copenhagen example descriptive statistics 2

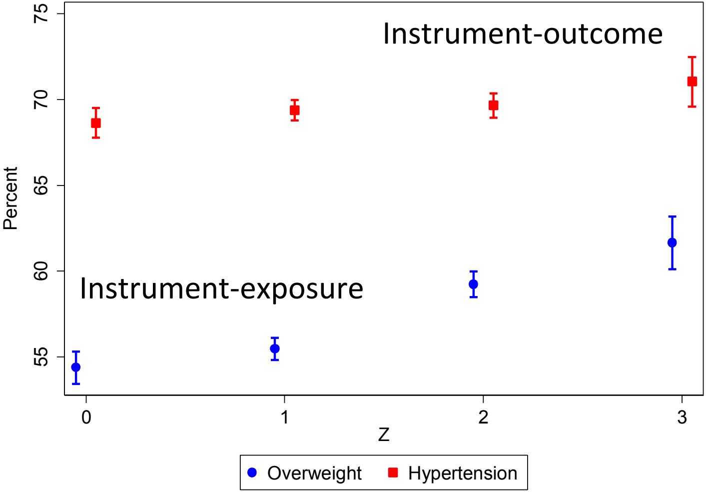
Distribution of instrument (Z)

FTO	*MC4R*	*Z*	Freq
0	0	0	0.20
0	1	1	0.15
1	0	1	0.27
1	1	2	0.21
2	0	2	0.09
2	1	3	0.07



Copenhagen example descriptive statistics 3

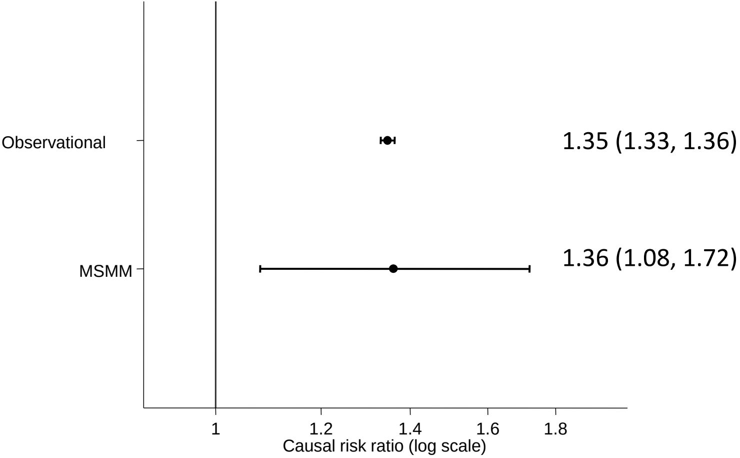
Exposure (over-weight) & outcome (hypertension) by instrument



P=0.007

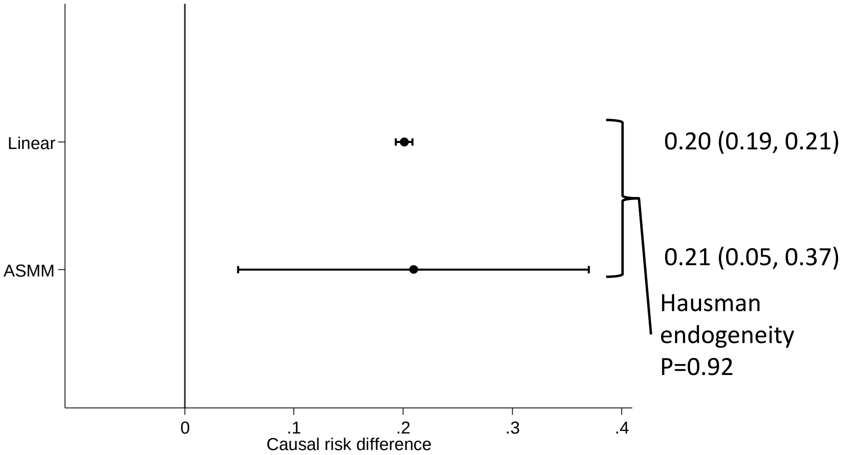
P<0.001
R^2=0.002

Copenhagen example Multiplicative SMM estimates



MSMM: Hansen over-identification test $P = 0.31$
$\mathrm{E}[Y(0)] = 0.58$ (0.50, 0.65)

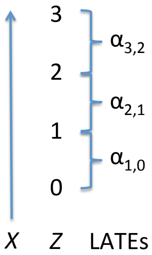
Copenhagen example Additive SMM estimates



MSMM: Hansen over-identification test $P = 0.30$
$E[Y(0)] = 0.58$ (0.48, 0.67)

Local risk ratios for Multiplicative SMM

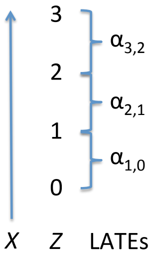
- Identification: NEM by Z ... what if it doesn't hold?
- Alternative assumption of monotonicity: $X(Z_k) \geq X(Z_{k-1})$
- Local Average Treatment Effect (LATE) (Imbens & Angrist, 1994)
  - effect among those whose exposures are changed (upwardly) by changing (counterfactually) the IV from Z_{k-1} to Z_k



$$\alpha_{All} = \lambda_1\alpha_{1,0} + \lambda_2\alpha_{2,1} + \lambda_3\alpha_{3,2}$$

Local risk ratios for Multiplicative SMM

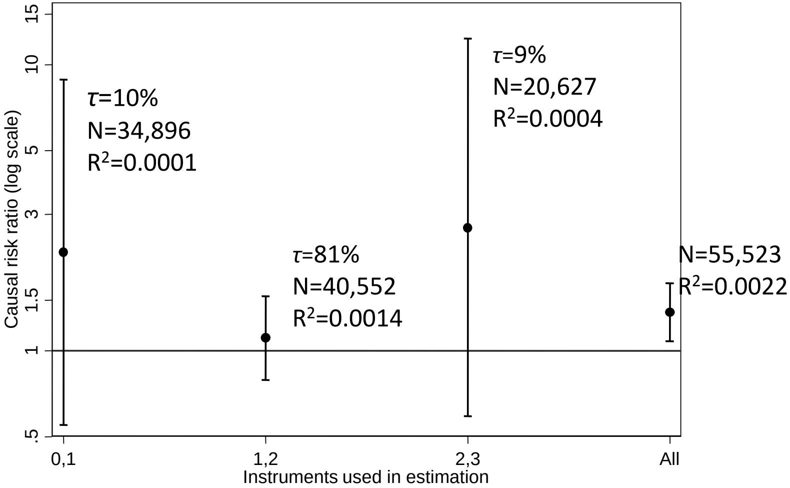
- Identification: NEM by Z ... what if it doesn't hold?
- Alternative assumption of monotonicity: $X(Z_k) \geq X(Z_{k-1})$
- Local Average Treatment Effect (LATE) (Imbens & Angrist, 1994)
  - effect among those whose exposures are changed (upwardly) by changing (counterfactually) the IV from Z_{k-1} to Z_k



$$\alpha_{All} = \lambda_1\alpha_{1,0} + \lambda_2\alpha_{2,1} + \lambda_3\alpha_{3,2}$$

Similar result holds for MSMM: $e^{\psi}_{\text{All}} = \sum_{k=1}^{K} \tau_k e^{\psi}_{k,k-1}$

Copenhagen example local risk ratios



Check: $(0.10 \times 2.21) + (0.81 \times 1.11) + (0.09 \times 2.69) = 1.36$

(double) Logistic SMM

$\text{logit}(p) = \log(p/(1-p)), \text{expit}(x) = e^x/(1+e^x)$

Goetghebeur, 2010

$$\text{logit}(E[Y|X,Z]) - \text{logit}(E[Y(0)|X,Z]) = \psi X$$

ψ : log causal odds ratio

Rearrange for $Y(0) = \text{expit}(\text{logit}(Y) - \psi X)$

(double) Logistic SMM

$\text{logit}(p) = \log(p/(1-p)), \text{expit}(x) = e^x/(1+e^x)$

Goetghebeur, 2010

$$\text{logit}(E[Y|X,Z]) - \text{logit}(E[Y(0)|X,Z]) = \psi X$$

ψ : log causal odds ratio

Rearrange for $Y(0) = \text{expit}(\text{logit}(Y) - \psi X)$

- Can't be estimated in a single step Robins (1999)
- First stage association model Vansteelandt (2003):
  - (i) logistic regression of Y on X & Z & interactions
  - (ii) predict Y, estimate LSMM using predicted Y

(double) Logistic SMM moment conditions

Association model moment conditions

Logistic regression using GMM

$$E[(Y - \text{expit}(\beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ))1] = 0$$
$$E[(Y - \text{expit}(\beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ))X] = 0$$
$$E[(Y - \text{expit}(\beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ))Z] = 0$$
$$E[(Y - \text{expit}(\beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ))XZ] = 0$$

(double) Logistic SMM moment conditions

Association model moment conditions

Logistic regression using GMM

$$E[(Y - \text{expit}(\beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ))1] = 0$$
$$E[(Y - \text{expit}(\beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ))X] = 0$$
$$E[(Y - \text{expit}(\beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ))Z] = 0$$
$$E[(Y - \text{expit}(\beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ))XZ] = 0$$

Causal model moment conditions

$$E[(\text{expit}(\text{logit}(\widehat{p}) - \psi X) - Y(0))1] = 0$$
$$E[(\text{expit}(\text{logit}(\widehat{p}) - \psi X) - Y(0))Z] = 0$$

Problem: SEs incorrect - need association model uncertainty

LSMM joint estimation

Joint estimation = correct SEs Gourieroux (1996)

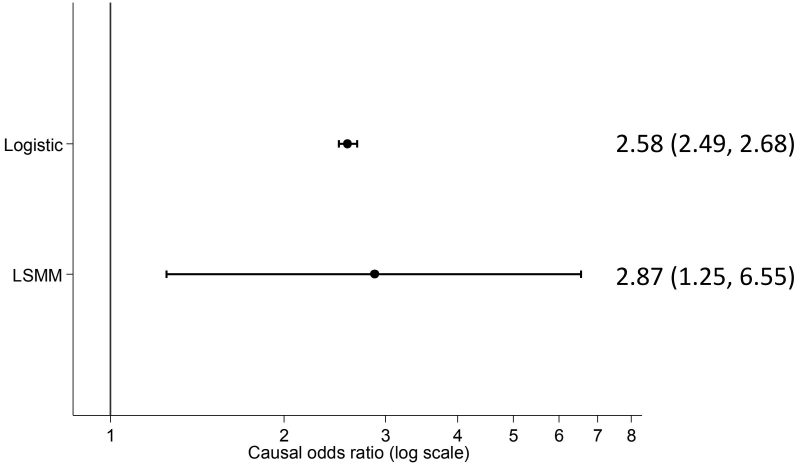
Vanstteelandt & Goetghebeur (2003)

$$
\begin{aligned}
E[(Y - \text{expit}(\beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ))1] &= 0 \\
E[(Y - \text{expit}(\beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ))X] &= 0 \\
E[(Y - \text{expit}(\beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ))Z] &= 0 \\
E[(Y - \text{expit}(\beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ))XZ] &= 0 \\
E[(\text{expit}(\beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ - \psi X) - Y(0))1] &= 0 \\
E[(\text{expit}(\beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ - \psi X) - Y(0))Z] &= 0
\end{aligned}
$$

Stata `gmm` command - allows multiple equations - still 1 line of code

Example: causal model SEs $\times 10$

Copenhagen example LSMM estimates



LSMM: Hansen over-identification test $P = 0.29$
$\mathrm{E}[Y(0)] = 0.57$ (0.45, 0.68)

SMM estimation problems I

- ALSPAC, N = 4647
- Estimate effect of BMI on asthma using FTO genotypes

Table 3. Instrumental Variable Estimates of the Causal Odds Ratio and Causal Risk Ratio for the Effect of Body Mass Index on Asthma Risk, Avon Longitudinal Study of Parents and Children, 1991–1992

	COR or CRR	95% CI
Standard logistic regression analysis		
Unadjusted odds ratio	1.06	1.02, 1.10
Adjusted[a] odds ratio	1.08	1.03, 1.13
Wald/ratio estimator[b]		
CRR	1.37	0.64, 2.96
COR	1.45	0.65, 3.43
2-stage estimator[c]		
CRR	1.37	0.68, 2.78
COR	1.45	0.64, 3.29
Control function[c]		
CRR	1.37	0.68, 2.76
COR	1.44	0.63, 3.28
Logistic structural mean model[d]		
COR	1.64	0.29, 9.31
Multiplicative structural mean model[d]		
CRR	0.81	0.44, 1.48
Multiplicative generalized method of moments[d]		
CRR	0.81	0.44, 1.48

SMM estimation problems II

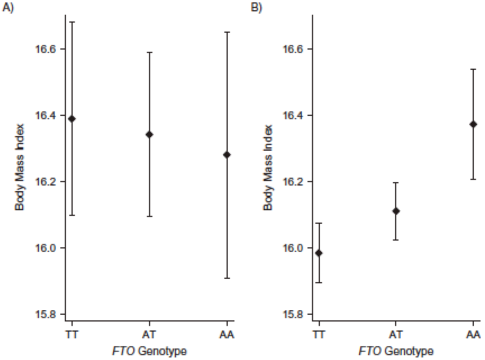


Figure 4. Mean body mass index (weight (kg)/height (m)^2), denoted by diamonds, according to fat mass and obesity-associated (*FTO*) genotype (rs9939609) for A) asthmatic and B) nonasthmatic children aged 7 years, Avon Longitudinal Study of Parents and Children, 1991–1992. Bars, 95% confidence interval.

- explained as interaction between FTO and unmeasured confounder on BMI

SMM estimation problems III

- Provided simulation evidence

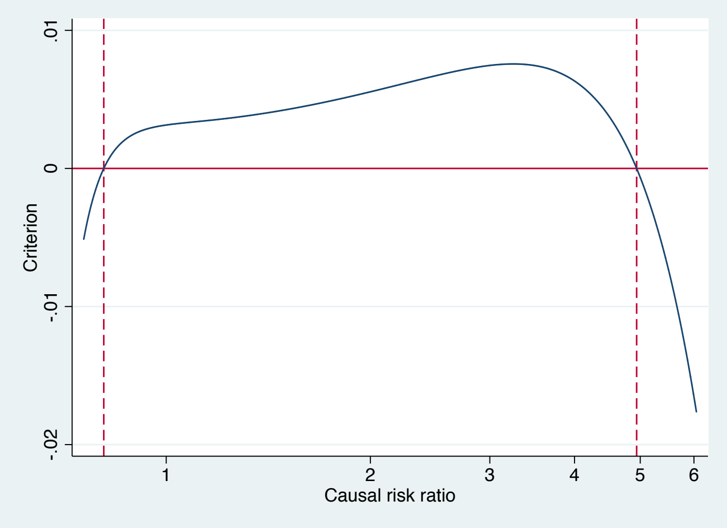
Table 4. Results of Simulations Comparing the Multiplicative Generalized Method of Moments and 2-Stage Estimators of the Causal Risk Ratio

	2-Stage Estimate for Log CRR (MCE)	MGMM Estimate for Log CRR (MCE)
Scenario 1: no causal effect with interaction		
Mean bias	−0.007 (0.0046)	0.009 (0.0094)
MSE	0.021 (0.0010)	0.088 (0.0042)
Coverage	0.952 (0.0068)	0.964 (0.0059)
Correlation between estimates	−0.23	
% of estimates on opposite sides of the CRR of 1	64.1	
Scenario 2: causal effect with interaction		
Mean bias	−0.206 (0.0042)	−0.146 (0.0100)
MSE	0.060 (0.0019)	0.120 (0.0055)
Coverage	0.674 (0.0148)	0.919 (0.0086)
Correlation between estimates	−0.12	
% of estimates on opposite sides of the CRR of 1.2	35.9	
Scenario 3: no causal effect with no interaction		
Mean bias	−0.005 (0.0049)	−0.001 (0.0053)
MSE	0.024 (0.0010)	0.029 (0.0018)
Coverage	0.942 (0.0074)	0.964 (0.0059)
Correlation between estimates	0.88	
% of estimates on opposite sides of the CRR of 1	7.3	
Scenario 4: causal effect with no interaction		
Mean bias	0.003 (0.0043)	0.003 (0.0049)
MSE	0.018 (0.0009)	0.024 (0.0014)
Coverage	0.954 (0.0066)	0.964 (0.0059)
Correlation between estimates	0.82	
% of estimates on opposite sides of the CRR of 1.2	15	

Abbreviations: CRR, causal risk ratio; MCE, Monte Carlo error; MGMM, multiplicative generalized method of moments; MSE, mean squared error.

SMM estimation problems IV

- But really weak identification



III. Nonparametric bounds (with Roland Ramsahai, Nuala Sheehan, Vanessa Didelez)

Nonparametric bounds

X, Y, Z all binary: IV assumptions–instrumental inequality (Bonet, 2001)

Denote $p_{yx.z} = P(Y = y, X = x \mid Z = z)$

$$
\begin{aligned}
p_{00.0} + p_{10.1} &\leq 1 \\
p_{10.0} + p_{00.1} &\leq 1 \\
p_{11.0} + p_{01.1} &\leq 1 \\
p_{01.0} + p_{11.1} &\leq 1
\end{aligned}
$$

- Not a statistical test
- If fail then the IV assumptions must be violated
- But IV assumptions can be violated without failing the inequalities

Nonparametric bounds II

Bounds on intervention probabilities (Balke & Pearl, 1997):
$P(Y(X=0))$, $P(Y(X=1))$

Bounds on the Average Causal Effect (ACE):
$\text{ACE } = P(Y(X=1)) - P(Y(X=0))$

Nonparametric bounds II

Bounds on intervention probabilities (Balke & Pearl, 1997):
$P(Y(X=0))$, $P(Y(X=1))$

Bounds on the Average Causal Effect (ACE):
ACE $= P(Y(X=1)) - P(Y(X=0))$

$$\max \left\{ \begin{array}{c} p_{00.0} + p_{11.1} - 1 \\ p_{00.1} + p_{11.1} - 1 \\ p_{11.0} + p_{00.1} - 1 \\ p_{00.0} + p_{11.0} - 1 \\ 2p_{00.0} + p_{11.0} + p_{10.0} + p_{11.1} - 2 \\ p_{00.0} + 2p_{11.0} + p_{00.1} + p_{01.1} - 2 \\ p_{10.0} + p_{11.0} + 2p_{00.1} + p_{11.1} - 2 \\ p_{00.0} + p_{01.0} + p_{00.1} + 2p_{11.1} - 2 \end{array} \right\} \leq \text{ACE} \leq \min \left\{ \begin{array}{c} 1 - p_{10.0} - p_{01.1} \\ 1 - p_{01.0} - p_{10.1} \\ 1 - p_{01.0} - p_{10.0} \\ 1 - p_{01.1} - p_{10.1} \\ 2 - 2p_{01.1} - p_{10.0} - p_{10.1} - p_{11.1} \\ 2 - p_{01.0} - 2p_{10.0} - p_{00.1} - p_{01.1} \\ 2 - p_{10.0} - p_{11.0} - 2p_{01.1} - p_{10.1} \\ 2 - p_{00.0} - p_{01.0} - p_{01.1} - 2p_{10.1} \end{array} \right\}$$

Nonparametric bounds: Example

	$Z=0$		$Z=1$	
	$Y=0$	$Y=1$	$Y=0$	$Y=1$
$X=0$	74	11514	34	2385
$X=1$	0	0	12	9663

Table : Vitamin A supplementation data (Balke & Pearl, 1997).

Nonparametric bounds: Example

	$Z = 0$		$Z = 1$	
	$Y = 0$	$Y = 1$	$Y = 0$	$Y = 1$
$X = 0$	74	11514	34	2385
$X = 1$	0	0	12	9663

Table : Vitamin A supplementation data (Balke & Pearl, 1997).

	Bounds
Instrumental inequality	`satisfied`
$P(Y(X = 0))$	(0.9936, 0.9936)
$P(Y(X = 1))$	(0.7990, 0.9990)
Average Causal Effect (risk difference)	(-0.1946, 0.0054)
Causal risk ratio	(0.8042, 1.0054)

TSLS: ACE 0.0032 (95%CI 0.0010, 0.0055)

Nonparametric bounds: Interpretation

- Not same as a confidence interval
- Bounds of [-0.1946, 0.0054]:
  there exists some distribution involving U that yields a true ACE as small as -19.46%, and another that gives a true ACE as large as 0.54%, with both distributions satisfying the IV assumptions and having the same observed marginal frequencies on (X, Y, Z)

Nonparametric bounds: Extensions

- Monotonicity assumption: for all values u of U
  $P(X = 1 \mid Z = 1, U = u) \geq P(X = 1 \mid Z = 0, U = u)$
  gives (slightly) tighter inequalities and bounds
- Extensions for a 3rd instrument category
- Data structures: (X, Z) & (Y, Z) in different samples

Nonparametric bounds: Limitation

Simulate data (N=10,000), two outcomes; both do not fulfil IV assumptions

$$
\begin{aligned}
Z &\sim Bern(0.5), \quad U \sim Bern(0.5) \\
p_X &= 0.05 + 0.1Z + 0.1U, \quad X \sim Bern(p_X) \\
p_1 &= 0.1 + 0.2Z + 0.05X + 0.1U, \quad Y_1 \sim Bern(p_1) \\
p_2 &= 0.1 + 0.05Z + 0.05X + 0.1U, \quad Y_2 \sim Bern(p_2)
\end{aligned}
$$

For Y_1 IV instrumental inequality is not satisfied

Nonparametric bounds: Limitation

Simulate data (N=10,000), two outcomes; both do not fulfil IV assumptions

$$
\begin{aligned}
Z &\sim Bern(0.5), \quad U \sim Bern(0.5) \\
p_X &= 0.05 + 0.1Z + 0.1U, \quad X \sim Bern(p_X) \\
p_1 &= 0.1 + 0.2Z + 0.05X + 0.1U, \quad Y_1 \sim Bern(p_1) \\
p_2 &= 0.1 + 0.05Z + 0.05X + 0.1U, \quad Y_2 \sim Bern(p_2)
\end{aligned}
$$

For Y_1 IV instrumental inequality is not satisfied
For Y_2 we get

	Bounds
Instrumental inequality	`satisfied`
$P(Y(X=0))$	(0.1542, 0.2352)
$P(Y(X=1))$	(0.0585, 0.8464)
Average Causal Effect (risk difference)	(-0.1767, 0.6922)

4. Summary

- Instrumental variable assumptions
  - Not fully testable from observational data (as for all causal inf.)
  - Application: Mendelian randomization (IV: genotypes)
  - Application: Correct for noncompliance (IV: randomized treatment)
  - Test for presence of effect
  - Estimators: ratio and two-stage least squares
- Structural Mean Models:
  - G-estimation $Y(0) \perp\!\!\!\perp Z$
  - Additive SMM
  - Multiplicative SMM
  - Review of methods Palmer, Sterne, et al., 2011; Palmer, Lawlor, et al., 2011
- Nonparametric bounds
  - Stata command: `bpbounds` Palmer, Ramsahai, Didelez, & Sheehan, 2011

Acknowledgements

- MRC Collaborative grant G0601625
- MRC CAiTE Centre grant G0600705
- ESRC grant RES-060-23-0011
- SMMs: joint work with Frank Windmeijer, Paul Clarke
- Nonparametric bounds: joint work with Roland Ramsahai, Nuala Sheehan, Vanessa Didelez
- With thanks to George Davey Smith, Debbie Lawlor, Jonathan Sterne, Stijn Vansteelandt, Sha Meng, Neil Davies, Roger Harbord, Nic Timpson, Børge Nordestgaard.

References I

Balke, A., & Pearl, J. (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, *92*(439), 1172–1176.

Bonet, B. (2001). Instrumentality tests revisited. In J. Breese & D. Koller (Eds.), *Proc. 17th conf. on uncertainty in artificial intelligence* (pp. 48–55). Seattle, WA: Morgan Kaufmann.

Chaussé, P. (2010). Computing generalized method of moments and generalized empirical likelihood with R. *Journal of Statistical Software*, *34*(11), 1–35. Retrieved from http://www.jstatsoft.org/v34/i11/

Davey Smith, G., & Ebrahim, S. (2003). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease. *International Journal of Epidemiology*, *32*, 1–22. doi: DOI:10.1093/ije/dyg070

Hernán, M. A., & Robins, J. M. (2006). Instruments for Causal Inference. An Epidemiologist's Dream? *Epidemiology*, *17*, 360–372.

Imbens, G. W., & Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, *62*, 467–467.

Palmer, T. M., Lawlor, D. A., Harbord, R. M., Sheehan, N. A., Tobias, J. H., Timpson, N. J., . . . Sterne, J. A. C. (2011). Using multiple genetic variants as instrumental variables for modifiable risk factors. *Statistical Methods in Medical Research*. (Published online before print January 7) doi: 10.1177/0962280210394459

References II

Palmer, T. M., Ramsahai, R. R., Didelez, V., & Sheehan, N. A. (2011). Nonparametric bounds for the causal effect in a binary instrumental-variable model. *Stata Journal*, *11*(3), 345-367(23).

Palmer, T. M., Sterne, J. A. C., Harbord, R. M., Lawlor, D. A., Sheehan, N. A., Meng, S., . . . Didelez, V. (2011). Instrumental variable estimation of causal risk ratios and causal odds ratios in mendelian randomization analyses. *American Journal of Epidemiology*, *173*, 1392–1403. doi: 10.1093/aje/kwr026

Robins, J. M. (1989). Health services research methodology: A focus on aids. In L. Sechrest, H. Freeman, & A. Mulley (Eds.), (chap. The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies). Washington DC, US: US Public Health Service.

Robins, J. M., Rotnitzky, A., & Scharfstein, D. O. (1999). Statistical models in epidemiology: The environment and clinical trials. In M. E. Halloran & D. Berry (Eds.), (pp. 1–92). New York, US: Springer.

Ten Have, T. R., Elliott, M. R., Joffe, M., Zanutto, E., & Datto, C. (2004). Causal models for randomized physician encouragement trials in treating primary care depression. *Journal of the American Statistical Association*, *99*(465), 16-25. doi: 10.1198/016214504000000034

Copenhagen example comparison estimates

	RR (95% CI)	P over-id
MSMM	1.36 (1.08, 1.72)	0.31
$Y - \exp(\psi X) \perp\!\!\!\perp Z$	1.36 (1.07, 1.75)	0.30
Control function	1.36 (1.08, 1.71)	
	OR (95% CI)	P over-id
LSMM two-stage	1.88 (1.75, 2.02)	
LSMM joint	2.87 (1.25, 6.55)	0.29
$Y - \text{expit}(\psi X) \perp\!\!\!\perp Z$	2.69 (1.23, 5.90)	0.30
Control function	2.69 (1.21, 5.97)	

Including covariates

TSLS: include covariates in both stages
GMM: use covariates as instruments for themselves

Including (pre-exposure) covariates in MSMM

$$Y(0) \perp\!\!\!\perp Z|C$$

$$\log(E[Y|X,Z,C]) - \log(E[Y(0)|X,Z,C]) = \psi X + \psi_c C$$

Including covariates

TSLS: include covariates in both stages
GMM: use covariates as instruments for themselves

Including (pre-exposure) covariates in MSMM

$$Y(0) \perp\!\!\!\perp Z | C$$

$$\log(E[Y|X,Z,C]) - \log(E[Y(0)|X,Z,C]) = \psi X + \psi_c C$$

Copenhagen example estimates

Covariates	RR (95%CI)	Over-id P
	1.36 (1.08, 1.72)	0.31
sex	1.36 (1.07, 1.72)	0.39
sex, age	1.35 (1.07, 1.71)	0.58
sex, age, chol	1.33 (1.05, 1.68)	0.49