Identification- and many instrument-robust inference via invariant moment conditions

Tom Boot Johannes W. Ligtenberg

University of Groningen

August 30, 2023

Introduction

- Focus on the linear IV model.
- Three complications:
  1. Weak or irrelevant instruments;
  2. Heteroskedasticity;
  3. Many instruments (Bekker, 1994).
- With only 1. tests based on the continuous updating objective function are optimal (Andrews et al., 2019).
- What if we add 2. and 3.?

In this presentation

- Discuss the problems with the continuous updating objective function when there are many instruments.
- Show how an invariance assumption can overcome these problems.
- Derive the joint distribution of the AR and score.

Why many instruments?

- Number of instruments is non-negligible relative to the sample size.
- Examples
  - Bartik instrument
    Goldsmith-Pinkham et al. (2020): elasticity of substitution of workers using 124 cities and 38 country groups.
  - Judge design dummies
    Kling (2006): each judge can only handle a limited number of cases.
  - Interactions of instruments and covariates
    Angrist and Krueger (1991): return to education using 329,509 observations and 1530 state-year-quarter dummies.

Why many instruments?

- Previous examples are for independent observations.
- In practice data are often clustered.
- Many and weak instruments are more likely to be problematic under clustered data due to reduced effective sample size.
- Ligtenberg (2023) proposes robust tests for clustered data.

Many instruments and continuous updating

- Linear IV model with heteroskedasticity

$$\begin{aligned} y_i &= \boldsymbol{x}_i' \boldsymbol{\beta}_0 + \varepsilon_i \\ \boldsymbol{x}_i &= \boldsymbol{\Pi}' \boldsymbol{z}_i + \boldsymbol{\eta}_i, \end{aligned}$$

  $\boldsymbol{\beta}_0 \in \mathbb{R}^p$, $\boldsymbol{\Pi} \in \mathbb{R}^{k \times p}$, $i = 1, \ldots, n$.
- Moment conditions $\boldsymbol{g}_i(\boldsymbol{\beta}_0) = \mathrm{E}[\boldsymbol{z}_i(y_i - \boldsymbol{x}_i' \boldsymbol{\beta}_0)] = 0$.
- Continuous updating objective function

$$Q(\boldsymbol{\beta}) = \frac{1}{n} \sum_i \boldsymbol{g}_i(\boldsymbol{\beta})' \left[\sum_i \boldsymbol{g}_i(\boldsymbol{\beta}) \boldsymbol{g}_i(\boldsymbol{\beta})' \right]^{-1} \sum_i \boldsymbol{g}_i(\boldsymbol{\beta})$$

- With many instruments and heteroskedasticity the $k \times k$ weighting matrix is problematic.

How to handle the weighting matrix?

1. Change it for something that does not depend on the errors.
   - Crudu et al. (2021) and Mikusheva and Sun (2022) do this for the objective function.
   - Matsushita and Otsu (2022) do this for the score.
   - Changing the weighting matrix might not weigh the moment conditions optimally.
2. Show that if the moment conditions satisfy an invariance assumption, the weighting matrix does not obstruct finding the limiting distribution.

Invariant moment conditions

- Assume the moment conditions are reflection invariant: $\boldsymbol{g}_i(\boldsymbol{\beta}_0) \overset{(d)}{=} r_i \boldsymbol{g}_i(\boldsymbol{\beta}_0)$ for $r_i = \pm 1$ with probability $1/2$.
- This allows for heteroskedasticity.
- Then under $H_0 \colon \boldsymbol{\beta} = \boldsymbol{\beta}_0$

$$
\begin{aligned}
Q(\boldsymbol{\beta}_0) &\overset{(d)}{=} \frac{1}{n} \sum_i r_i \boldsymbol{g}_i(\boldsymbol{\beta}_0)' [r_i^2 \sum_i \boldsymbol{g}_i(\boldsymbol{\beta}_0) \boldsymbol{g}_i(\boldsymbol{\beta}_0)']^{-1} \sum_i r_i \boldsymbol{g}_i(\boldsymbol{\beta}_0) \\
&= \frac{1}{n} \boldsymbol{r}' \boldsymbol{P}(\boldsymbol{\beta}_0) \boldsymbol{r},
\end{aligned}
$$

with $\boldsymbol{P}(\boldsymbol{\beta})$ a projection matrix of the moment conditions.

Many instrument Anderson-Rubin statistic

Lemma A2 of Chao et al. (2012):

$$\frac{\sum_{i,j} P_{ij}(\beta_0) - k}{\sqrt{\sigma_n^2 k}} \overset{(d)}{=} \frac{\boldsymbol{r}'\boldsymbol{P}(\beta_0)\boldsymbol{r} - k}{\sqrt{\sigma_n^2 k}} \rightsquigarrow N(0,1).$$

Score test

Same approach as for AR:

1. Use invariance to get score with r_i.
2. Condition on $\{\varepsilon_i, \boldsymbol{z}_i\}_{i=1}^n$.
3. Use the randomness in r_i to derive the conditional variance of the score with r_i.
4. Use the randomness in r_i to derive a central limit theorem for the standardised score.
5. Conclude that the conditional distribution implies an unconditional distribution.

Result:

$$\boldsymbol{\Sigma}_n^{-1/2}(\boldsymbol{\beta}_0)\begin{pmatrix}\frac{1}{\sqrt{k}}[\sum_{i,j} P_{ij}(\boldsymbol{\beta}_0) - k] \\ \sqrt{n}\boldsymbol{S}(\boldsymbol{\beta}_0)\end{pmatrix} \rightsquigarrow N(0, \boldsymbol{I}).$$

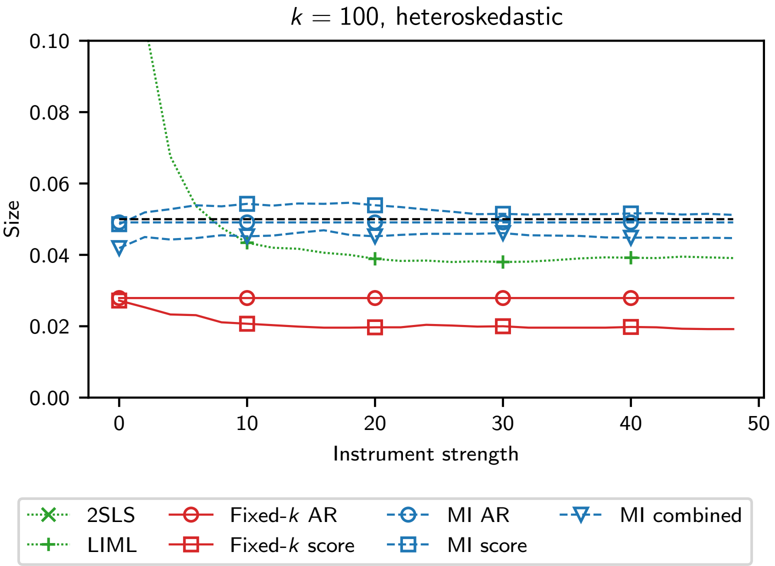
Monte Carlo

$$y_i = x_i\beta_0 + \varepsilon_i$$
$$x_i = \mathbf{\Pi}'\boldsymbol{z}_i + \eta_i.$$

- $n = 800$ observations, $p = 1$ endogenous regressor and $\beta_0 = 0$.
- $k = 100$ instruments of which only one is relevant. We vary its strength.
- There is endogeneity and conditional heteroskedasticity, but the errors are reflection invariant.

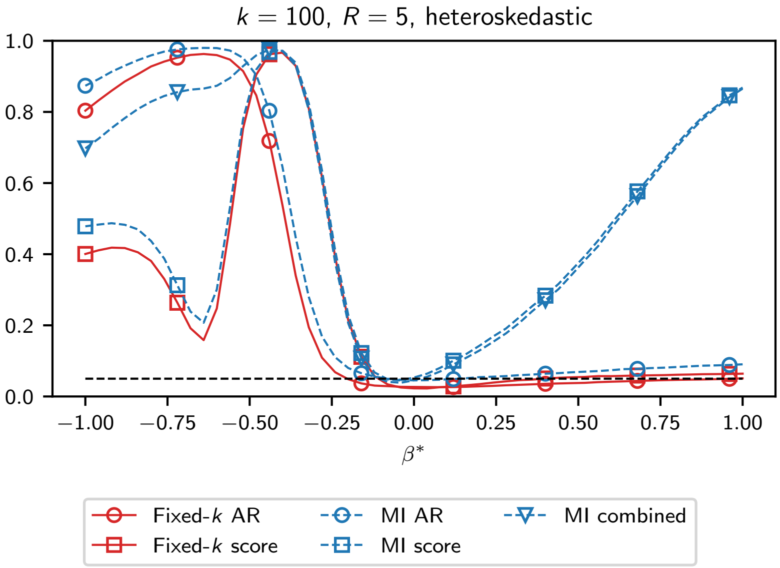
Details

Size

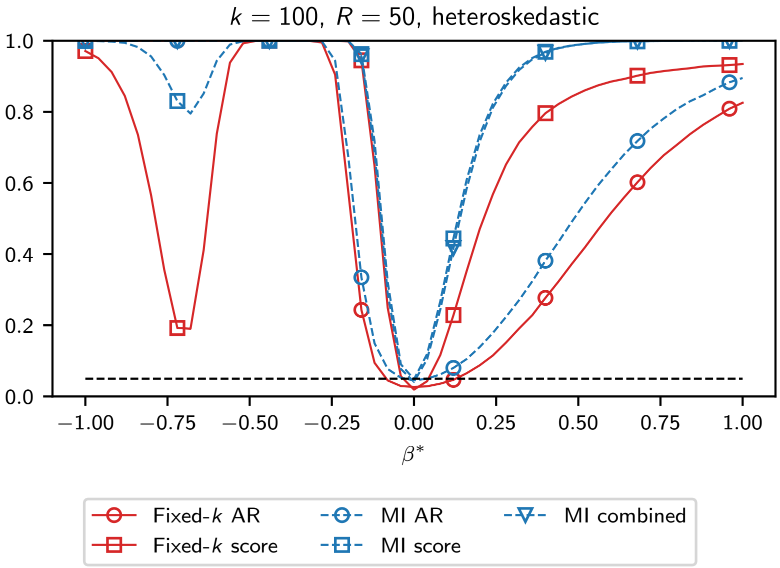


$k = 100$, heteroskedastic
Size
0.10
0.08
0.06
0.04
0.02
0.00
0
10
20
30
40
50
Instrument strength
2SLS
LIML
Fixed-k AR
Fixed-k score
MI AR
MI score
MI combined

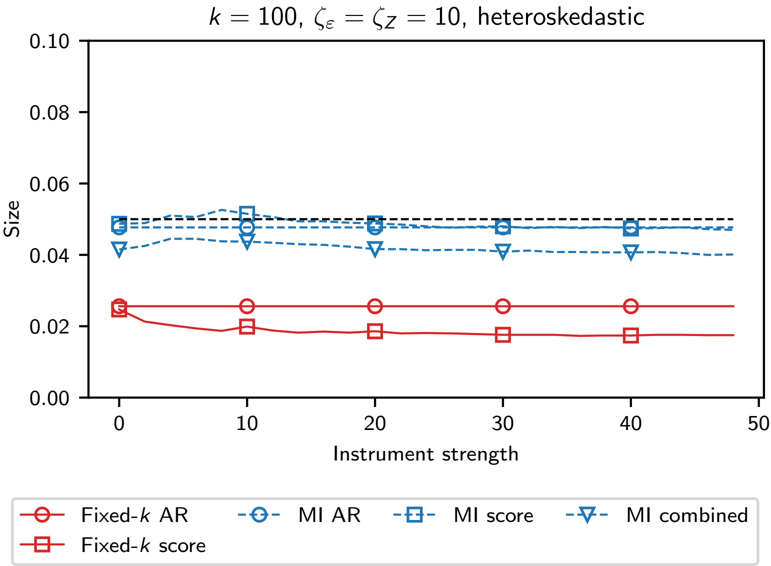
Power weak instruments



Power strong instruments



Size with violated invariance



Conclusion

- Many instruments are empirically relevant.
- The large dimensional weighting matrix of the continuous updating objective is difficult to handle with many instruments and heteroskedasticity.
- These problems can be circumvented when the moment conditions are reflection invariant.
- We can derive an AR and score statistic.
- The AR and score tests are size correct and have good power.

References I

Andrews, D. W. K., V. Marmer, and Z. Yu (2019). On optimal inference in the linear IV model. *Quantitative Economics 10*(2), 457–485.

Angrist, J. D. and A. B. Krueger (1991, November). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics 106*(4), 979–1014.

Bekker, P. A. (1994). Alternative approximations to the distributions of instrumental variable estimators. *Econometrica 62*(3), 657–681.

Chao, J. C., N. R. Swanson, J. A. Hausman, W. K. Newey, and T. Woutersen (2012). Asymptotic distribution of JIVE in a heteroskedastic IV regression with many instruments. *Econometric Theory 28*(1), 42–86.

References II

Crudu, F., G. Mellace, and Z. Sándor (2021). Inference in instrumental variable models with heteroskedasticity and many instruments. *Econometric Theory 37*(2), 281–310.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020). Bartik instruments: What, when, why, and how. *American Economic Review 110*(8), 2586–2624.

Kling, J. R. (2006). Incarceration length, employment, and earnings. *American Economic Review 96*(3), 863–876.

Ligtenberg, J. W. (2023). Inference in IV models with clustered dependence, many instruments and weak identification.

Matsushita, Y. and T. Otsu (2022). Jackknife lagrange multiplier test with many weak instruments.

Mikusheva, A. and L. Sun (2022). Inference with many weak instruments. *Review of Economic Studies 89*(5), 2663–2686.

Many instrument Anderson-Rubin statistic

Assume:

1. $\text{rank}[\boldsymbol{P}(\beta_0)] = k$;
2. $P_{ii}(\beta_0) < 1$ for all i;
3. $\sigma_n^2 = \frac{2}{k}\sum_{i \neq j} P_{ij}(\beta_0)^2 > 0$.

Lemma A2 of Chao et al. (2012):

$$\frac{\sum_{i,j} P_{ij}(\beta_0) - k}{\sqrt{\sigma_n^2 k}} \stackrel{(d)}{=} \frac{\boldsymbol{r}'\boldsymbol{P}(\beta_0)\boldsymbol{r} - k}{\sqrt{\sigma_n^2 k}} \rightsquigarrow N(0,1).$$

Power of many instrument AR

- The many instrument AR has higher power than the AR statistic for a fixed number of instruments.
- Let $\text{AR}(\boldsymbol{\beta}) = nQ(\boldsymbol{\beta})$.
- Let $\phi_1(\boldsymbol{\beta}) = 1$ if $(k\sigma_n^2)^{-1/2}(\text{AR}(\boldsymbol{\beta}) - k) > (2k)^{-1/2}(\chi^2(k)_{1-\alpha} - k)$.
- Let $\phi_2(\boldsymbol{\beta}) = 1$ if $\text{AR}(\boldsymbol{\beta}) > \chi^2(k)_{1-\alpha}$.
- Then if $\alpha < 0.3$, $\text{P}(\phi_1(\boldsymbol{\beta}) = 1) > \text{P}(\phi_2(\boldsymbol{\beta}) = 1)$.

Score test

$$\frac{\partial Q(\boldsymbol{\beta})}{\partial \beta_i} = -\frac{1}{n}\boldsymbol{x}_{(i)}{}'[\boldsymbol{I} - \mathrm{diag}(\boldsymbol{P}(\boldsymbol{\beta})\boldsymbol{\iota})]\boldsymbol{Z}\left[\sum_j \boldsymbol{z}_j\boldsymbol{z}_j'\varepsilon_j^2\right]^{-1}\boldsymbol{Z}'\boldsymbol{\varepsilon}$$

- To apply $\varepsilon_i\boldsymbol{z}_i \overset{(d)}{=} r_i\varepsilon_i\boldsymbol{z}_i$ we need to parameterise dependence of $\boldsymbol{x}_i = \boldsymbol{\Pi}'\boldsymbol{z}_i + \boldsymbol{\eta}_i$ on ε_i.
- $\boldsymbol{\eta}_i = \varepsilon_i\boldsymbol{a}_i + \boldsymbol{u}_i$ with $\boldsymbol{u}_i$ and ε_i independent.

Conditional variance of the score

- The conditional variance of the score is $\Omega_{ij}(\beta_0) = \mathrm{E}\left[n \cdot S_{(i),r}(\beta_0) S_{(j),r}(\beta_0) \middle| \mathcal{J}\right] = \Omega^L_{ij}(\beta_0) + \Omega^H_{ij}(\beta_0)$.
- If $\max_{i=1,\dots,n} P_{ii} \leq 0.9$ then $\Omega^H_{ij}(\beta_0)$ is negative semi-definite.
- Under additional assumptions negative semi-definite changes to negative definite. These additional assumptions are more likely to hold if there are many instruments.

Assumption for joint distribution

1. $\frac{1}{n}\sum_{i=1}^{n}\|\bar{\boldsymbol{z}}_i\|^2 \leq C < \infty$ a.s.n.;
2. $\frac{1}{n}\max_{i=1,\dots,n}\|\bar{\boldsymbol{z}}_i\|^2 \to_{a.s.} 0$;
3. $\frac{1}{n}\max_{i=1,\dots,n}\|\bar{\boldsymbol{Z}}'\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{e}_i\|^2 \to_{a.s.} 0$;
4. $0 < C^{-1} \leq \lambda_{\min}(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{Z}) \leq \lambda_{\max}(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{Z}) \leq C < \infty$ a.s.n., $0 < C^{-1} \leq \lambda_{\min}(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{D}_\varepsilon^2\boldsymbol{Z}) \leq \lambda_{\max}(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{D}_\varepsilon^2\boldsymbol{Z}) \leq C < \infty$ a.s.n.

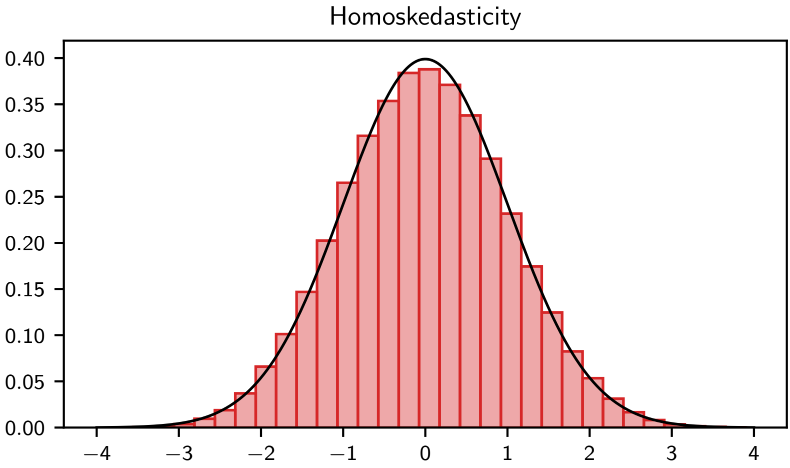
Details of Monte Carlo setup

$$
\begin{aligned}
y_i &= x_i \beta_0 + \varepsilon_i \\
x_i &= \mathbf{\Pi}' \boldsymbol{z}_i + \eta_i .
\end{aligned}
$$

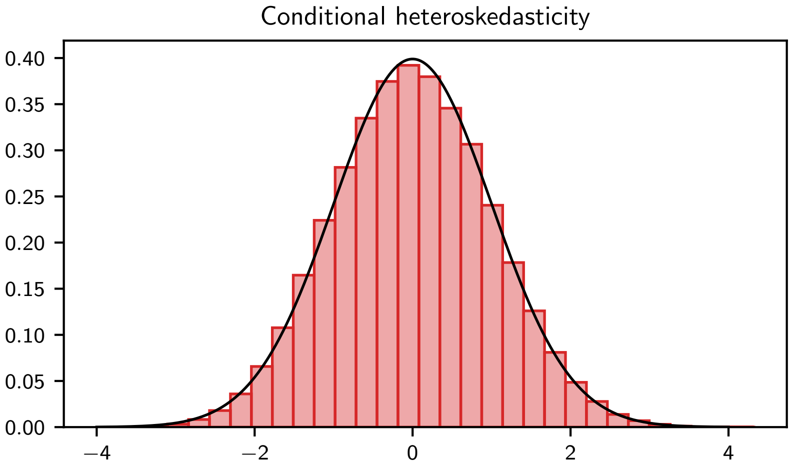
- $n = 800$, $p = 1$, $\beta_0 = 0$.
- $k = 100$ with $Z_{ij} \sim N(0,1)$. $\Pi_1 = \sqrt{R\sqrt{k}/n}$ and zeroes elsewhere. $R = 5$ for weak IV and $R = 50$ for strong IV.
- $\eta_i = |Z_{1i}|\varepsilon_i + w_i/2$ with $w_i \sim N(0,1)$ and $\varepsilon_i \sim N(0,1)$.

Back

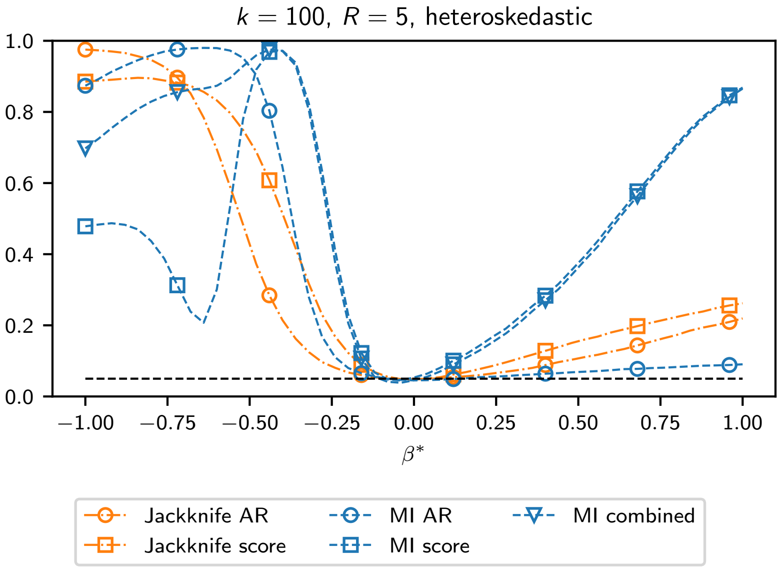
Distribution of the many instrument AR



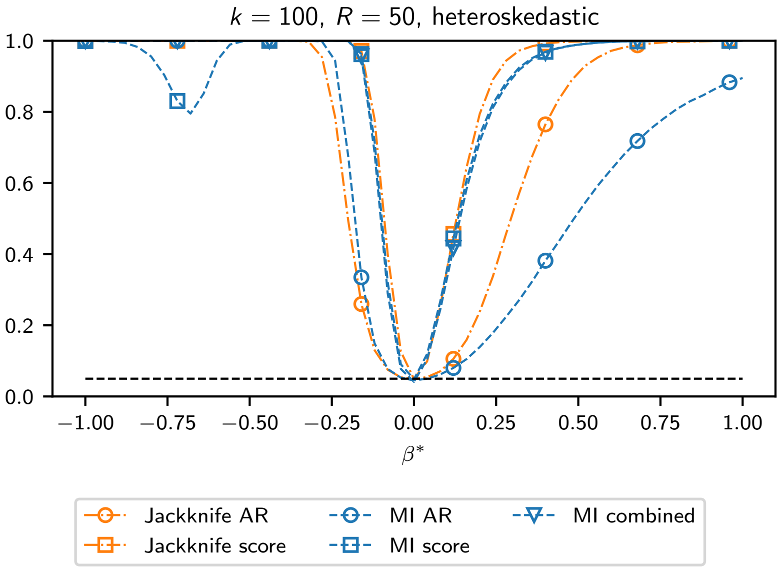
Distribution of the many instrument AR



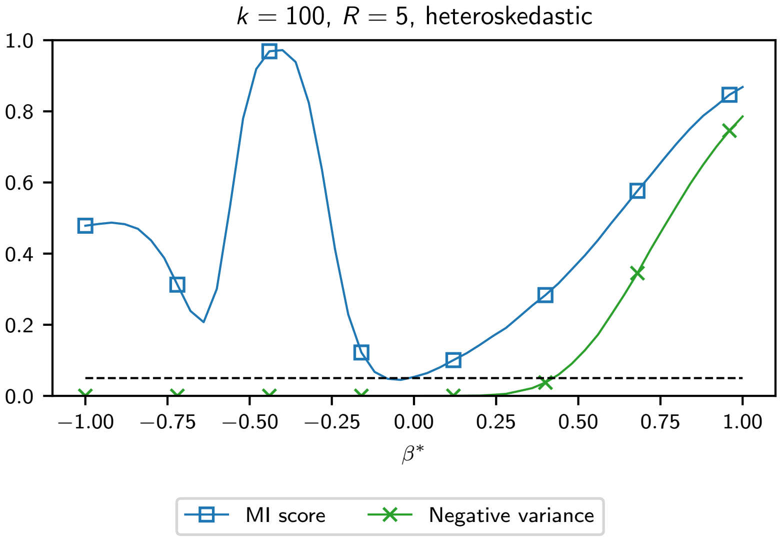
Jackknife weak IV



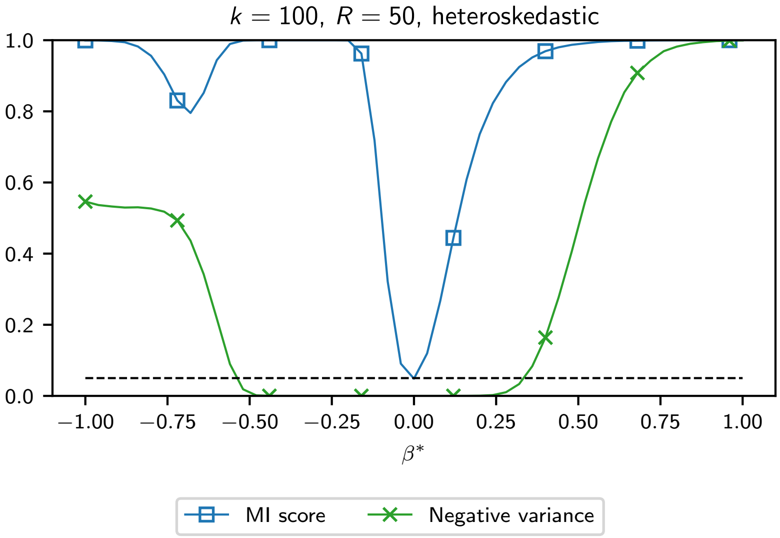
Jackknife strong IV



Negative variance weak IV



Negative variance strong IV

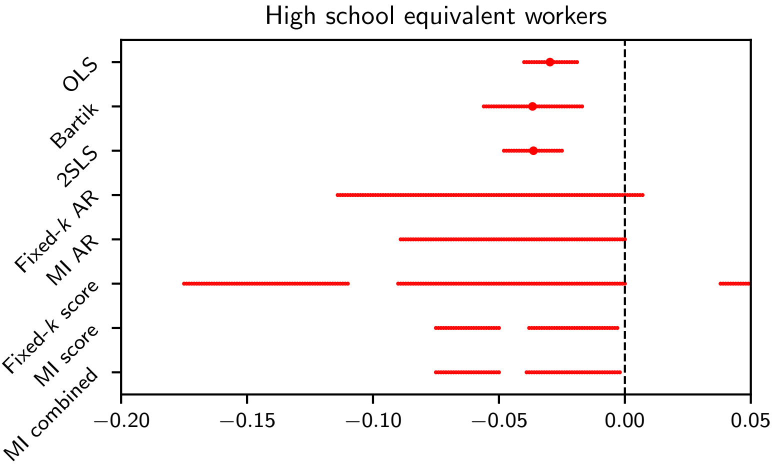


Negative inverse elasticity of substitution

$$y_{lj} = \beta \log x_{lj} + \boldsymbol{\gamma}' \boldsymbol{X}_l + \varepsilon_{lj}$$

- y_{lj} residual log wage gap between immigrant and native men in skill group j and location l, x_{lj} ratio of immigrant to native hours worked in skill group j and location l of both men and women, $\boldsymbol{X}_l$ location specific controls.
- Bartik instrument $B_{lj} = \sum_{k=1}^{38} z_{lk,1980} g_{kj}$, with $z_{lk,1980} = N_{lk,1980}/(N_{k,1980} P_{l,2000}$ for $N_{lk,1980}$ number of immigrants from country group k in 1980 in location l, $N_{k,1980}$ number of immigrants from country group k in 1980 and $P_{l,2000}$ the population in 2000, and g_{kj} the number of immigrants from country group k in skill group j arriving between 1990 and 2000.
- $z_{lk,1980}$ can also be used separately as instruments.

Negative inverse elasticity of substitution



Negative inverse elasticity of substitution