Judging Judges of IV Validity

How to Not Fail Hausman and Omnibus

Hutchings, Nate, Passey, Roundy

Brigham Young University

April 15, 2024

Motivating Question

- How does the presence of multicollinearity in IV affect the output of overidentification tests?

	Hausman Test	**Omnibus Test**
$Z'Z$	?	?
$Z'X$	?	?

Omnibus and Hausman Tests

- Omnibus: All instruments valid $\implies \hat{o} = nR^2$ should not be too different from 0.
- Hausman: All instruments valid $\implies \hat{b}_{\text{sub}}$ and $\hat{b}_{\text{all}}$ should not be too far apart
- If we construct valid instruments, could multicollinearity lead us to rejection more than 5 percent of the time?

$Z'Z$ and $Z'X$ in Over-identification Tests

- Omnibus:

$$\hat{o} = n * (1 - \frac{\hat{u} - Z(Z^T Z)^{-1} Z^T \hat{u}}{\hat{u}^T \hat{u}})$$

$$\hat{u} = y - X(X^T Z(Z^T Z)^{-1} Z^T X)^{-1} X^T Z(Z^T Z)^{-1} Z^T y$$

- Hausman:

$$\hat{h} = \frac{\hat{b}_{\text{sub}} - \hat{b}_{\text{all}}}{\sqrt{\hat{V}_{\text{sub}} - \hat{V}_{\text{all}}}}$$

$$\hat{V} = \frac{\hat{u}^T \hat{u}}{n - p} (X^T Z(Z^T Z)^{-1} Z^T X)^{-1}$$

$Z'Z$ Data Generating Process

- Structure:

$$Y = X\beta + \epsilon$$

$$X = Z\pi + \gamma$$

$$\epsilon = \gamma + \eta$$

- η, γ are drawn randomly.
- $Z \sim N(0, 1)$
- By construction, instruments are valid (relevant and exclusive)

Moment Condition

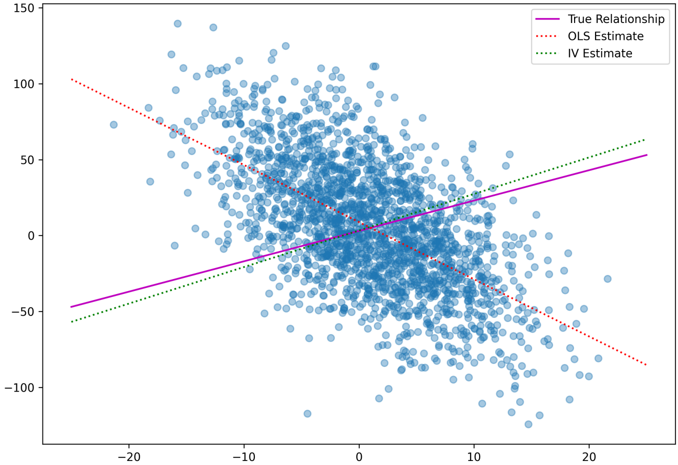
- Moment Condition:

$$m_i(\beta) = Z_i(Y_i - X_i'\beta)$$

$$E[m_i(\beta)] = 0$$

- It holds by construction.

Visualizing the DGP



Constructing Multicollinear Instruments

- $z_2 = z_1 + D$, where D is the mechanism of collinearity:

$$\underset{n\times 1}{D}\, s.t. \sum_{i=1}^{n} |d_i| = C$$

- Iterate Monte-Carlo tests over levels of C

Constructing Multicollinear Instruments

- Compute condition index for each iteration of C

$$C.I. = \frac{\sigma_{Max}}{\sigma_{Min}},$$

  where σ_x are singular values from matrix $Z'Z$
- As C increases, multicollinearity (C.I.) in $Z'Z$ decreases

$Z'Z$ Condition Index

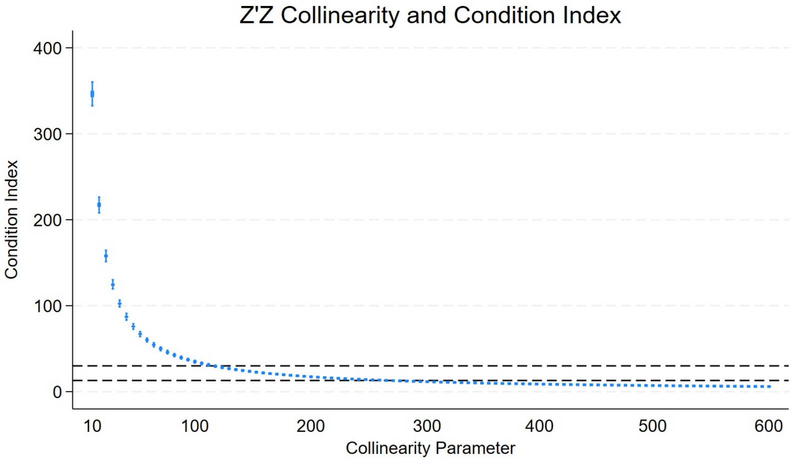
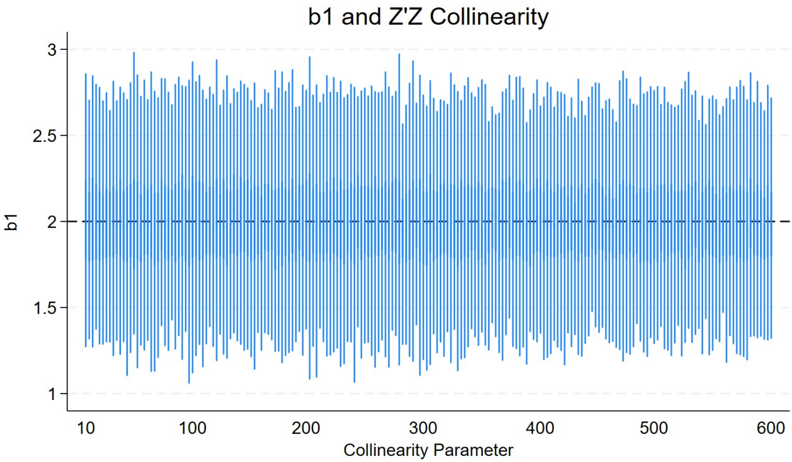


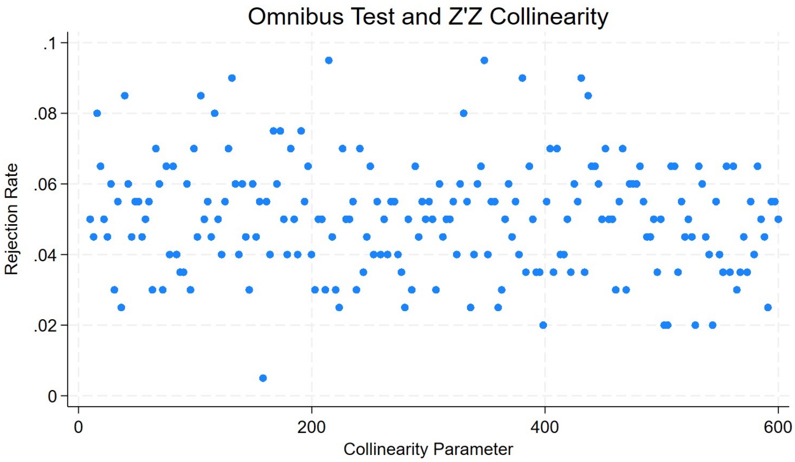
Figure: Lines at 30, 10

$Z'Z$ b_1



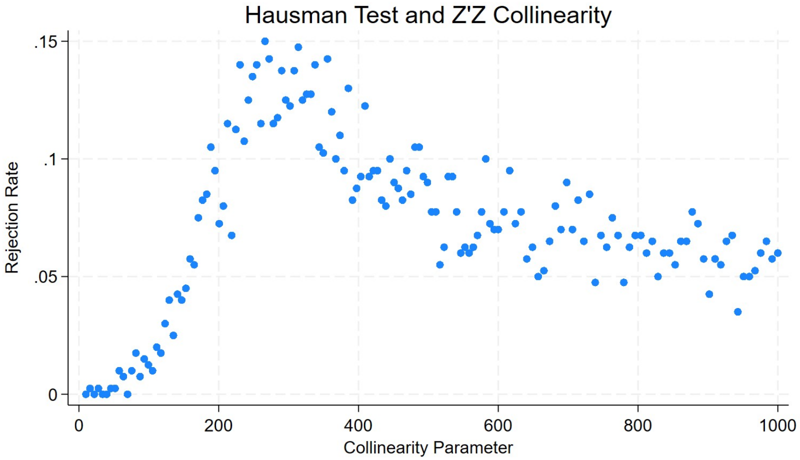
b1 and Z'Z Collinearity
3
2.5
2
1.5
1
b1
10
100
200
300
400
500
600
Collinearity Parameter

$Z'Z$ Results



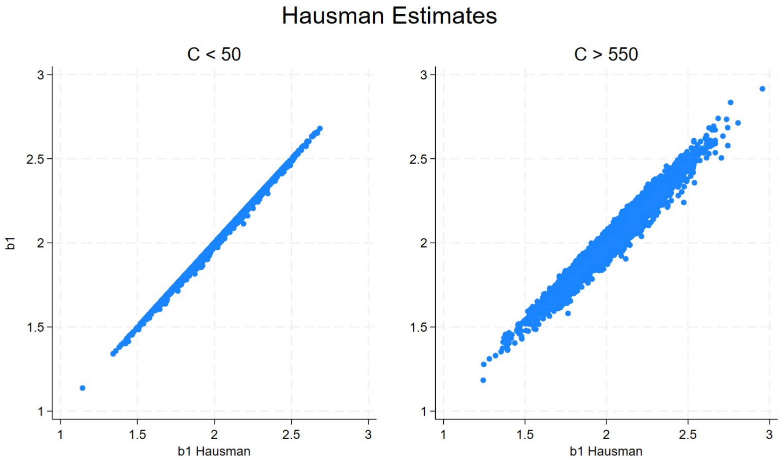
Omnibus Test and Z'Z Collinearity
Rejection Rate
0
.02
.04
.06
.08
.1
0
200
400
600
Collinearity Parameter

$Z'Z$ Results

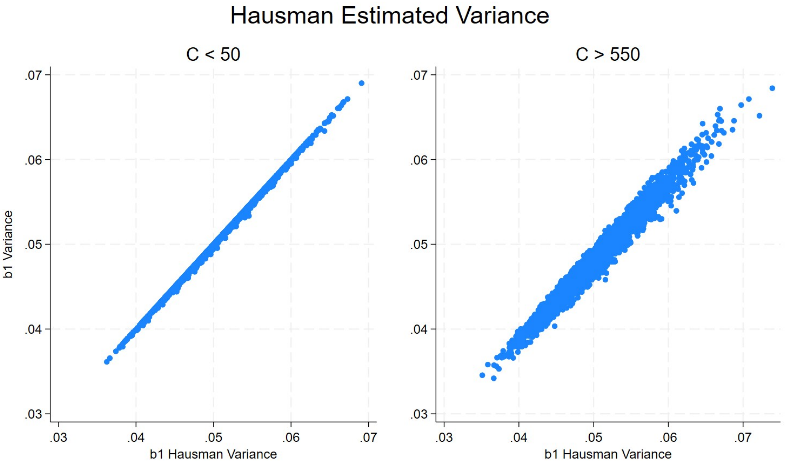


Hausman Test and Z'Z Collinearity
Rejection Rate
.15
.1
.05
0
0
200
400
600
800
1000
Collinearity Parameter

Explaining Hausman Results



Explaining Hausman Results



$Z'X$ Data Generating Process

- Craft $Z'X$ that displays multicollinearity
- Instruments still satisfy relevance and exclusion conditions (classically valid)

$Z'X$ Data Generating Process

- Construct
$$Y = X\beta + \epsilon,$$
where $\epsilon = e_1 + e_2 + \eta$
- Fix Z: $z_1, z_2, z_3 \sim N(0, 1)$
- Two independent variables:
  - $x_1 = Z\alpha + e_1$
  - $x_2 = Z\pi + e_2$
- Note: x_i are endogenous by construction, z_i are not.

Proof of Collinearity in $Z'X$

$$E[z_i x_i^T] = E[z_i[{x_1}_i{}^T \;\; {x_2}_i{}^T]] = E[z_i[(z_i{}^T\alpha + e_1) \;\; (z_i{}^T\pi + e_2)]$$

$$= E[z_i[z_i{}^T\alpha \;\; z_i{}^T\pi]]$$

given that $E[z_i e_1] = E[z_i e_2] = 0$.

Thus, $E[z_i z_i^T\alpha \;\; z_i z_i^T\pi]$ becomes collinear as $\alpha \to \pi$

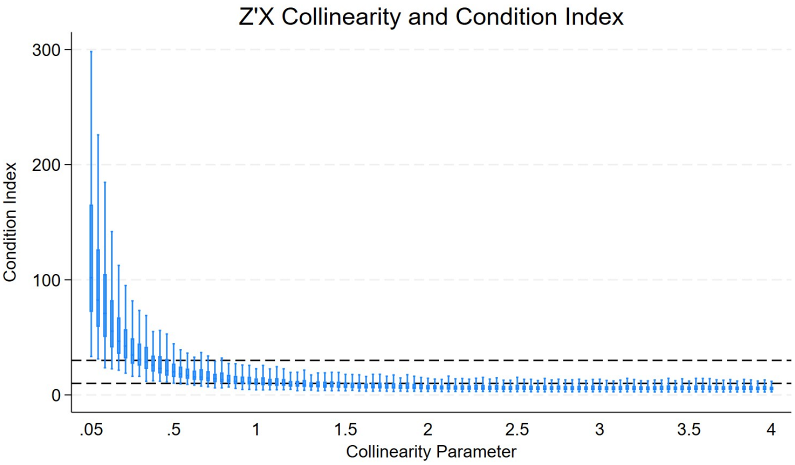
$Z'X$ Data Generating Process

- For independent variables:
  - $x_1 = Z\alpha + e_1$
  - $x_2 = Z\pi + e_2$
- $\alpha = [1, 1, 1, 1]$
- $\pi = \alpha + D$, where

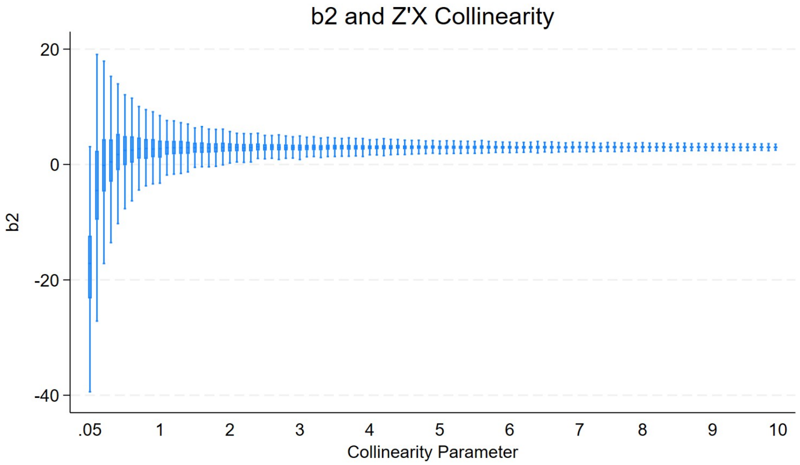
$$\underset{n\times 1}{D}\ s.t. \sum_{i=1}^{n} |d_i| = C$$

- Iterating over C allows us to test different levels of multicollinearity in the *linear combinations of Z that represent x_i*

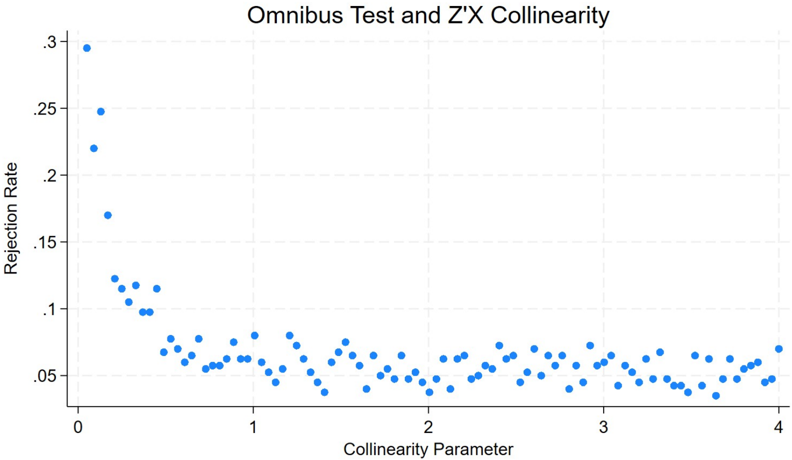
$Z'X$ Condition Index



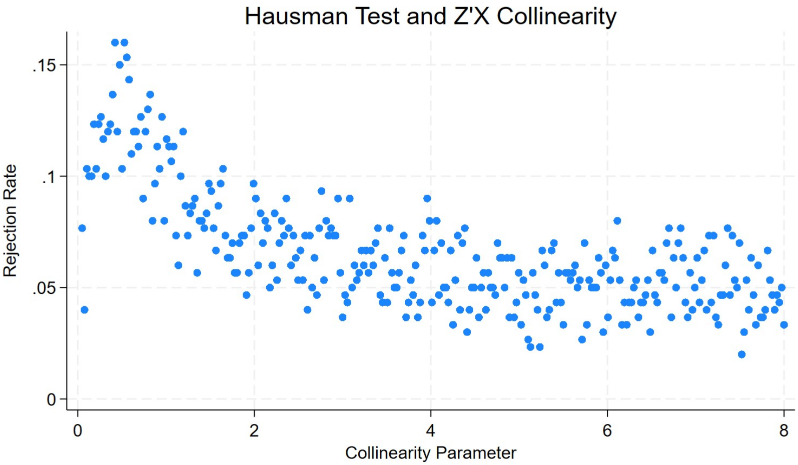
$Z'X$ Instrument Failure



$Z'X$ Results


Omnibus Test and Z'X Collinearity
Rejection Rate
.3
.25
.2
.15
.1
.05
0
1
2
3
4
Collinearity Parameter

$Z'X$ Results



Hausman Test and Z'X Collinearity
Rejection Rate
.15
.1
.05
0
0
2
4
6
8
Collinearity Parameter

Effects on test statistics as collinearity increases

	Hausman Test	**Omnibus Test**
$Z'Z$	?	*No Effect*
$Z'X$	$\uparrow$ *Rej.Rate*	$\uparrow$ *Rej.Rate*

Conclusion

- Classically valid instruments that exhibit high levels of multicollinearity can lead to increased rejection rates in standard overidentification tests