Toulouse School of Economics

M1 INTERMEDIATE ECONOMETRICS

Instrumental-variable estimation

Koen Jochmans   François Poinas

2024 — 2025

This deck of slides goes through instrumental-variable estimation of linear models.

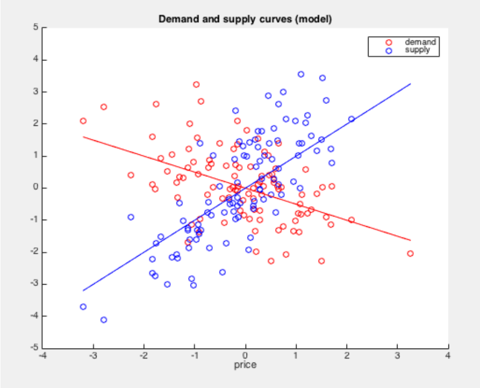
The corresponding chapters in Hansen are 12 and 13.

Example: simultaneity

Temporary deviation from notational conventions to analyze market model

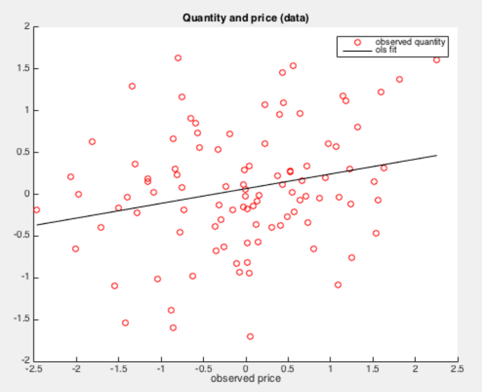
$$d = \alpha_d - \theta_d \, p + u$$
$$s = \alpha_s \; + \theta_s \, p + v$$

where d, s, p are demand, supply, and price, respectively.



We do not observe supply and demand for any given price.

Collected data is on quantity traded and transaction price, (q, p).



Data comes from markets in equilibrium.

So, we solve

$$s = d$$

for the equilibrium price to get

$$p = \frac{\alpha_d - \alpha_s}{\theta_d + \theta_s} + \frac{u - v}{\theta_d + \theta_s}.$$

This gives traded quantity as

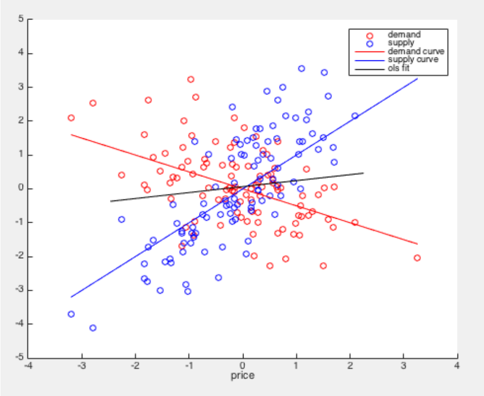
$$q = \frac{\alpha_d \theta_s + \alpha_s \theta_d}{\theta_d + \theta_s} + \frac{\theta_s u + \theta_d v}{\theta_d + \theta_s}.$$

(With $\mathbb{E}(uv) = 0$) the population regression slope of q on p equals

$$\frac{\sigma_u^2}{\sigma_u^2 + \sigma_v^2}\, \theta_s - \frac{\sigma_v^2}{\sigma_u^2 + \sigma_v^2}\, \theta_d,$$

for $\sigma_u^2 = \mathbb{E}(u^2)$ and $\sigma_v^2 = \mathbb{E}(v^2)$.

Least-squares estimates a weighted average of supply and demand elasticities.


demand
supply
demand curve
supply curve
ols fit
price
5
4
3
2
1
0
-1
-2
-3
-4
-5
-4
-3
-2
-1
0
1
2
3
4

Focus on the estimation of the demand curve.

Then, collecting equations from above, we have the triangular system

$$d = \alpha_d - \theta_d p + u, \qquad p = \frac{\alpha_d - \alpha_s}{\theta_d + \theta_s} + \frac{u - v}{\theta_d + \theta_s}.$$

Clearly,

$$\mathbb{E}(pu) = \mathbb{E}\left(u\left(\frac{u - v}{\theta_d + \theta_s}\right)\right) = \frac{\sigma_u^2}{\theta_d + \theta_s} \neq 0,$$

as the errors in both equations are correlated.

The same happens for the supply curve, as

$$s = \alpha_s + \theta_s p + v, \qquad p = \frac{\alpha_d - \alpha_s}{\theta_d + \theta_s} + \frac{u - v}{\theta_d + \theta_s}.$$

and

$$\mathbb{E}(pv) = \mathbb{E}\left(v\left(\frac{u - v}{\theta_d + \theta_s}\right)\right) = -\frac{\sigma_v^2}{\theta_d + \theta_s} \neq 0.$$

Resolving simultaneity with instrumental variables

Now suppose that

$$\begin{aligned} d &= \alpha_d - \theta_d\, p + u \\ s &= \alpha_s + \theta_s\, p + \gamma z + v \end{aligned}.$$

where $\mathbb{E}(zu) = 0$.

Here, z shifts supply (relevance) but not demand (exclusion).

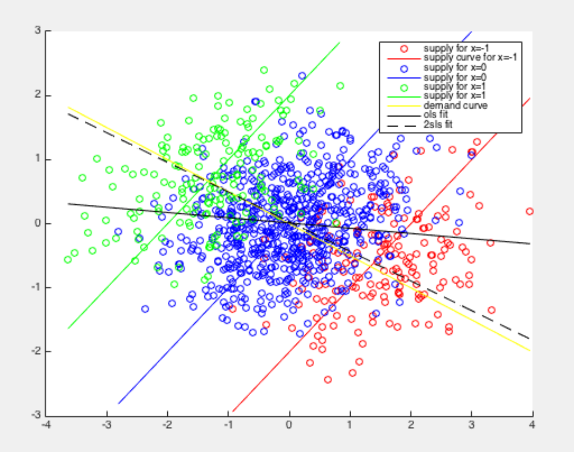
We now have the triangular system of equations

$$\begin{aligned} d &= \alpha_d - \theta_d\, p + u \\ p &= \frac{\alpha_d - \alpha_s}{\theta_d + \theta_s} - \frac{\gamma}{\theta_d + \theta_s}\, z + \frac{u - v}{\theta_d + \theta_s} \end{aligned}.$$

Further, as $\text{cov}(u, z) = 0$,

$$\text{cov}(d, z) = \text{cov}(\alpha_d - \theta_d\, p_i + u, z) = -\theta_d\, \text{cov}(p, z),$$

and so, provided that $\text{cov}(p, z) \neq 0$, $-\theta_d = {}^{\text{cov}(d_i, z_i)}\!/_{\text{cov}(p_i, z_i)}$.


supply for x=-1
supply curve for x=-1
supply for x=0
supply for x=0
supply for x=1
supply for x=1
demand curve
ols fit
2sls fit
3
2
1
0
-1
-2
-3
-4
-3
-2
-1
0
1
2
3
4

Endogeneity and instrumental variables (H12.2, H12.4 and H12.5)

Interest lies in the parameter vector β in the linear model

$$Y = X'\beta + e,$$

when

$$\mathbb{E}(Xe) \neq 0.$$

Hence, β is not a projection coefficient!

Rather see the equation of interest as a structural relationship.

Linearity is an assumption.

We will proceed by using instrumental variables Z, which are variables that satisfy the following two conditions:

Validity: $\mathbb{E}(Ze) = 0$.

Relevance: $\mathbb{E}(ZX')$ has rank k.

We will also maintain that $\mathbb{E}(ZZ')$ is invertible, this simply excludes linearly-dependent instruments.

Note that by setting $Z = X$ this recovers the linear prediction problem that we have studies so far.

It is useful to rechristen $Y_1 = Y$ and to partition

$$Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}, \qquad X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} Z_1 \\ Y_2 \end{pmatrix},$$

and then, because, $\mathbb{E}(Ze) = 0$ but $\mathbb{E}(Xe) \neq 0$,

$$\mathbb{E}(Z_1 e) = 0, \qquad \mathbb{E}(Z_2 e) = 0, \qquad \mathbb{E}(Y_2 e) \neq 0$$

We refer to Z_1 as exogenous regressors and Y_2 as endogenous regressors.

Exogenous regressors Z_1 can serve as instruments.

Endogenous regressors Y_2 cannot, and need to be instrumented for by Z_2.

The relevance condition requires that we need at least one instrument for each regressor.

We let $\ell = \ell_1 + \ell_2$ be the dimension of Z and $k = k_1 + k_2$ be the dimension of X; here $k_1 = \ell_1$.

We need that $\ell_2 \geq k_2$.

IV estimand (H13.4-H13.6)

If

$$Y_1 = X'\beta + e, \qquad \mathbb{E}(Ze) = 0, \qquad \operatorname{rank} \mathbb{E}(ZX') = k,$$

then

$$\mathbb{E}[Z(Y_1 - X'\beta)] = \mathbb{E}(ZY_1) - \mathbb{E}(ZX')\,\beta = 0.$$

When $\ell > k$ we have more equations than unknowns. The problem is overidentified.

For a $k \times \ell$ matrix A with maximal row rank

$$A\,\mathbb{E}(ZY_1) - A\,\mathbb{E}(ZX')\,\beta = 0$$

and so

$$\beta = (A\,\mathbb{E}(ZX'))^{-1}(A\,\mathbb{E}(ZY_1)).$$

When $k = \ell$, the problem is just identified. In this case $(A\,\mathbb{E}(ZX'))^{-1} = \mathbb{E}(ZX')^{-1}A^{-1}$, and so

$$\beta = (\mathbb{E}(ZX'))^{-1}(\mathbb{E}(ZY_1)),$$

independent of A.

Alternatively, when $\ell > k$, can think about doing least squares on the linear relationship

$$\mathbb{E}(ZY_1) = \mathbb{E}(ZX')\beta,$$

i.e.,

$$\beta = \arg\min_b \left(\mathbb{E}(ZY_1) - \mathbb{E}(ZX')b\right)' \left(\mathbb{E}(ZY_1) - \mathbb{E}(ZX')b\right)$$

This has first-order condition

$$\mathbb{E}(XZ')[\mathbb{E}(ZY_1) - \mathbb{E}(ZX')\beta] = 0$$

and solution

$$\beta = (\mathbb{E}(XZ')\,\mathbb{E}(ZX'))^{-1}(\mathbb{E}(XZ')\,\mathbb{E}(ZY_1)).$$

Again, when $\ell = k$ this reduces to $\beta = (\mathbb{E}(ZX'))^{-1}(\mathbb{E}(ZY_1))$.

Could also do generalized least squares.

For $\ell \times \ell$ weight matrix W this is

$$\beta = \arg\min_{b} \left(\mathbb{E}(Y_1 Z') - b'\mathbb{E}(XZ')\right) W \left(\mathbb{E}(ZY_1) - \mathbb{E}(ZX')b\right).$$

The solution is

$$\beta = (\mathbb{E}(XZ')\, W\, \mathbb{E}(ZX'))^{-1}(\mathbb{E}(XZ')\, W\, \mathbb{E}(ZY_1)).$$

This corresponds to $A = \mathbb{E}(XZ')\, W$.

This GLS formulation is the basis for a class of generalized method of moment estimators, as a function of the choice of W.

Simultaneous-equation system and reduced form (H12.7-H12.8)

Let

$$Y_2 = \Gamma' Z + u_2 = \Gamma'_{12} Z_1 + \Gamma'_{22} Z_2 + u_2$$

be defined through $\mathbb{E}(Z u_2) = 0$, i.e.,

$$\Gamma = \mathbb{E}(ZZ')^{-1} \mathbb{E}(Z Y_2').$$

This is essentially a population regression of Y_2 on Z.

Then we have the system of triangular equations

$$\begin{aligned} Y_1 &= Z_1' \beta_1 + Y_2' \beta_2 + e \\ Y_2 &= \Gamma'_{12} Z_1 + \Gamma'_{22} Z_2 + u_2. \end{aligned}$$

Because $\mathbb{E}(Ze) = 0$, we have that $\mathbb{E}(Y_2 e) = \mathbb{E}(u_2 e) (\neq 0)$, so endogeneity flows through the projection error u_2.

We can plug-in the projection of Y_2 on Z into the structural equation for Y_1 to get

$$\begin{aligned} Y_1 &= X'\beta + e \\ &= Z_1'\beta_1 + Y_2'\beta_2 + e \\ &= Z_1'\beta_1 + (\Gamma_{12}'Z_1 + \Gamma_{22}'Z_2 + u_2)'\beta_2 + e \\ &= Z_1'(\beta_1 + \Gamma_{12}\beta_2) + Z_2'\Gamma_{22}\beta_2 + (e + u_2'\beta_2) \\ &= \lambda'Z + u_1 \end{aligned}$$

for

$$\lambda = \begin{pmatrix} \beta_1 + \Gamma_{12}\beta_2 \\ \Gamma_{22}\beta_2 \end{pmatrix}, \qquad u_1 = e + u_2'\beta_2.$$

Can stack the equations

$$Y_1 = \lambda' Z + u_1$$
$$Y_2 = \Gamma' Z + u_2.$$

Here $\mathbb{E}(Zu_1) = 0$ and $\mathbb{E}(Zu_2) = 0$ and, because $\mathbb{E}(ZZ')$ is invertible, we can learn λ and Γ from two separate linear (population) projections.

We have

$$\lambda = \begin{pmatrix} \beta_1 + \Gamma_{12}\beta_2 \\ \Gamma_{22}\beta_2 \end{pmatrix} = \begin{pmatrix} I_{k_1} & \Gamma_{12} \\ 0 & \Gamma_{22} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \bar{\Gamma}\,\beta.$$

This is a set of ℓ equations in k unknowns β. Therefore, provided that $\bar{\Gamma}$ has maximal column rank k,

$$\beta = (\bar{\Gamma}'\bar{\Gamma})^{-1}\bar{\Gamma}'\lambda$$

is recoverable.

Now observe that

$$\bar{\Gamma} = \begin{pmatrix} I_{k_1} & \Gamma_{12} \\ 0 & \Gamma_{22} \end{pmatrix} = \mathbb{E}(ZZ')^{-1}\mathbb{E}(ZX'),$$

recalling that $X' = (Z_1', Y_2')$ and $Z' = (Z_1', Z_2')$.

This matrix has rank k iff (by matrix block inversion formula)

$$\operatorname{rank}\Gamma_{22} = k_2.$$

Alternatively, as $\operatorname{rank}\mathbb{E}(ZZ') = \ell$, we require that

$$\operatorname{rank}\mathbb{E}(ZX') = k,$$

which is our relevance condition.

Assumptions (H Assumption 12.2)

In addition to the relevance and validity conditions on the instrumental variables, assume

1. Random sampling: The variables (Y_i, X_i, Z_i) are i.i.d.

2. Moments: $\mathbb{E}(|Y_1|^4) < \infty$, $\mathbb{E}(\|X\|^4) < \infty$, $\mathbb{E}(\|Z\|^4) < \infty$.

3. Variance: $\Omega = \mathbb{E}(ZZ'e^2)$ is positive definite.

First let $\boldsymbol{W}$ be a fixed non-random matrix.

Our estimator is

$$\hat{\beta}_{\text{gmm}} = ((\boldsymbol{X}'\boldsymbol{Z})\boldsymbol{W}(\boldsymbol{Z}'\boldsymbol{X}))^{-1}((\boldsymbol{X}'\boldsymbol{Z})\boldsymbol{W}(\boldsymbol{Z}'\boldsymbol{Y})).$$

Clearly, as $n \to \infty$,

$$\boldsymbol{Z}'\boldsymbol{X}/n \underset{p}{\to} \mathbb{E}(ZX') = \boldsymbol{Q}_{ZX}, \qquad \boldsymbol{Z}'\boldsymbol{Y}/n \underset{p}{\to} \mathbb{E}(ZY),$$

and so

$$\hat{\beta}_{\text{gmm}} \underset{p}{\to} \beta.$$

Next, because $\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{e}$,

$$\sqrt{n}(\hat{\beta}_{\text{gmm}} - \beta) = \left((\boldsymbol{X}'\boldsymbol{Z}/n)\boldsymbol{W}(\boldsymbol{Z}'\boldsymbol{X}/n)\right)^{-1}\left((\boldsymbol{X}'\boldsymbol{Z}/n)\boldsymbol{W}(\boldsymbol{Z}'\boldsymbol{e}/\sqrt{n})\right).$$

From before, we know that, as $n \to \infty$,

$$\sqrt{n}(\hat{\beta}_{\text{gmm}} - \beta) = \left(\boldsymbol{Q}_{ZX}'\boldsymbol{W}\,\boldsymbol{Q}_{ZX}\right)^{-1}\left(\boldsymbol{Q}_{ZX}'\boldsymbol{W}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i e_i\right)\right) + o_p(1).$$

Also,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i e_i \underset{d}{\to} N(0, \Omega)$$

so that,

$$\sqrt{n}(\hat{\beta}_{\text{gmm}} - \beta) \underset{d}{\to} N(0, \boldsymbol{V}_\beta)$$

for

$$\boldsymbol{V}_\beta = \left(\boldsymbol{Q}_{ZX}'\boldsymbol{W}\,\boldsymbol{Q}_{ZX}\right)^{-1}\left(\boldsymbol{Q}_{ZX}'\boldsymbol{W}\Omega\boldsymbol{W}\boldsymbol{Q}_{ZX}\right)\left(\boldsymbol{Q}_{ZX}'\boldsymbol{W}\,\boldsymbol{Q}_{ZX}\right)^{-1}.$$

Variance estimator (H12.18)

All the limit results go through if we replace $\boldsymbol{W}$ by a $\hat{\boldsymbol{W}}$ that satisfies

$$\hat{\boldsymbol{W}} \underset{p}{\rightarrow} \boldsymbol{W}$$

as $n \to \infty$.

Given residuals $\hat{e}_i = Y_i - X_i' \hat{\beta}_{\text{gmm}}$ we can estimate Ω by

$$\hat{\Omega} = \frac{1}{n} \sum_{i=1}^{n} Z_i Z_i' \hat{e}_i^2.$$

The plug-in estimator of $\boldsymbol{V}$ so constructed is consistent.

Efficient estimator (H13.10)

The asymptotic variance $\boldsymbol{V}_\beta$ is minimized for the choice

$$\hat{\boldsymbol{W}} = \hat{\Omega}^{-1}.$$

In this case,

$$\boldsymbol{V}_\beta = \left(\boldsymbol{Q}'_{ZX}\Omega^{-1}\boldsymbol{Q}_{ZX}\right)^{-1}.$$

The construction of $\hat{\Omega}$ requires residuals, which then require an initial estimator of β.

This leads to a two-step procedure.

When $\mathbb{E}(e|Z=z)=\sigma^2$,

$$\Omega = \mathbb{E}(ZZ'e^2) = \sigma^2\,\mathbb{E}(ZZ') = \sigma^2\,\boldsymbol{Q}_{ZZ}.$$

In this case, the efficient estimator uses $\hat{\Omega} = s^2\,\hat{\boldsymbol{Q}}_{ZZ} \propto \boldsymbol{Z}'\boldsymbol{Z}$, i.e.,

$$\hat{\beta}_{\text{gmm}} = (\hat{\boldsymbol{Q}}'_{ZX}\hat{\boldsymbol{Q}}_{ZZ}^{-1}\hat{\boldsymbol{Q}}_{ZX})^{-1}(\hat{\boldsymbol{Q}}'_{ZX}\hat{\boldsymbol{Q}}_{ZZ}^{-1}\hat{\boldsymbol{Q}}_{ZY}).$$

But note that

$$\hat{\boldsymbol{Q}}_{ZZ}^{-1}\hat{\boldsymbol{Q}}_{ZX} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}(\boldsymbol{Z}'\boldsymbol{X}) = \hat{\bar{\Gamma}},$$

which is an estimator of $\bar{\Gamma}$, and that

$$\hat{\boldsymbol{Q}}'_{ZX}\hat{\boldsymbol{Q}}_{ZZ}^{-1}\hat{\boldsymbol{Q}}_{ZX} = \hat{\boldsymbol{Q}}'_{ZX}\hat{\bar{\Gamma}} = n^{-1}\boldsymbol{X}'\boldsymbol{Z}\hat{\bar{\Gamma}} = n^{-1}\hat{\bar{\Gamma}}'\boldsymbol{Z}'\boldsymbol{Z}\hat{\bar{\Gamma}} = n^{-1}\boldsymbol{X}'\boldsymbol{P_Z}\boldsymbol{X},$$

and so

$$\hat{\beta}_{\text{gmm}} = \hat{\beta}_{\text{2sls}} = (\boldsymbol{X}'\boldsymbol{P_Z}\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{P_Z}\boldsymbol{Y}) = (\hat{\bar{\Gamma}}'\boldsymbol{Z}'\boldsymbol{Z}\hat{\bar{\Gamma}})^{-1}(\hat{\bar{\Gamma}}'\boldsymbol{Z}'\boldsymbol{Y})$$

This is the two-stage least squares estimator.

Its name comes from the observation that

$$Y_1 = \lambda' Z + u_1 = (\beta' \bar{\Gamma}') Z + u_1 = \beta' (\bar{\Gamma}' Z) + u_1 = (Z' \bar{\Gamma})' \beta + u_1$$

so that we could estimate β by OLS from a regression of Y_1 on $Z' \bar{\Gamma}$ if $\bar{\Gamma}$ was known.

$\bar{\Gamma}$ is not known so replaced with its OLS estimator.

We can proceed by following the Wald principle for inference.

For GMM, as

$$\sqrt{n}\hat{\boldsymbol{V}}_{\beta}^{-1/2}(\hat{\beta}_{\text{gmm}} - \beta) \underset{d}{\rightarrow} N(0, I_k)$$

for any continuously-differentiable (vector-valued) function r, $\theta = r(\beta)$ satisfies

$$\sqrt{n}\,(\hat{\boldsymbol{R}}'\hat{\boldsymbol{V}}_{\beta}\hat{\boldsymbol{R}})^{-1/2}(\hat{\theta} - \theta) \underset{d}{\rightarrow} N(0, I_q).$$

So, testing $\mathbb{H}_0 : \theta = \theta_0$ against $\mathbb{H}_1 : \theta \neq \theta_0$ can be done via the Wald statistic

$$n\,(\hat{\theta} - \theta_0)'(\hat{\boldsymbol{R}}'\hat{\boldsymbol{V}}_{\beta}\hat{\boldsymbol{R}})^{-1}(\hat{\theta} - \theta_0)$$

in exactly the same manner as before.

The (infeasible) statistic

$$\frac{\boldsymbol{e}'\boldsymbol{Z}\,\Omega^{-1}\boldsymbol{Z}'\boldsymbol{e}}{n}$$

is asymptotically χ^2_ℓ under the null that $\mathbb{E}(Ze) = 0$.

A feasible version of this statistic is

$$\frac{\hat{\boldsymbol{e}}'\boldsymbol{Z}\,\hat{\Omega}^{-1}\boldsymbol{Z}'\hat{\boldsymbol{e}}}{n} = \frac{(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta}_{\text{gmm}})'\boldsymbol{Z}\,\hat{\Omega}^{-1}\boldsymbol{Z}'(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta}_{\text{gmm}})}{n};$$

this is the efficient-GMM objective function evaluated at its minimizer.

Under the null it is asymptotically $\chi^2_{\ell-k}$.

We loose k degrees of freedom due to the estimation of β.

Feasible statistic is exactly zero in the just-identified case!

Sargan statistic (H12.31)

A special case has homoskedasticity. In this case, we use 2sls and a variance estimator under homoskedasticity.

Then the J-statistic becomes

$$\frac{1}{n}\frac{\hat{\boldsymbol{e}}'\boldsymbol{Z}\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\hat{\boldsymbol{e}}}{s^2} = \frac{\hat{\boldsymbol{e}}'\boldsymbol{P}_{\boldsymbol{Z}}\hat{\boldsymbol{e}}}{s^2} = (n-k)\frac{\hat{\boldsymbol{e}}'\boldsymbol{P}_{\boldsymbol{Z}}\hat{\boldsymbol{e}}}{\hat{\boldsymbol{e}}'\hat{\boldsymbol{e}}} = (n-k)\,R^2,$$

where R^2 refers to the regression of $\hat{e}_i$ on Z_i.

This is Sargan's statistic.

Large values for the J-statistic suggest that the validity of (at least some of) the instruments is in doubt.

Incremental statistic (H13.23)

We can use the same idea to test a subset of the moment conditions.

Prime application is to test whether $\mathbb{E}(Y_2 e) = 0$.

To do this let J_1 be the J-statistic based on $\mathbb{E}(Ze) = 0$. and let J_0 be the J-statistic based on $\mathbb{E}(Ze) = 0$ *and* $\mathbb{E}(Y_2 e) = 0$. Then

$$J_0 - J_1 \underset{d}{\rightarrow} \chi^2_{k_2}$$

under the null that $\mathbb{E}(Y_2 e) = 0$.