M1 INTERMEDIATE ECONOMETRICS

INSTRUMENTAL VARIABLES

Koen Jochmans

November 11, 2025

1. The instrumental-variable problem

The general setting of interest to be discussed here is one where we are interested in the parameter vector β in the equation

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

but we know (or at least suspect) that $\mathbb{E}(Xe) \neq 0$. Thus, β is not the coefficient in a population projection of Y on X. Regressors that correlate with the error term are said to be endogenous. Endogeneity can come from many different sources. A simple one is the omission of relevant explanatory variables.

A solution to the endogeneity problem is to rely on ℓ variables, Z, that satisfy the following two conditions:

C1, $\mathbb{E}(Ze) = 0$ (exclusion).

C2, $\operatorname{rank} \mathbb{E}(ZX') = k$ (relevance).

To see how, note that, from C1, $\mathbb{E}(Z(Y - X'\beta)) = 0$, which is a set of ℓ linear equations. Therefore, for any $\ell \times k$ matrix A with maximal column rank the system of k equations

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

can be solved for β to get

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

A simple choice for A is to set $A = \mathbb{E}(ZX')$. The matrix A is important only when $\ell > k$, that is, when we have more equations than unknown coefficients. We call this the overidentified case. When $\ell = k$ we are in the just-identified case and the matrix A drops out of the above equation, which then reduces to $\beta = \mathbb{E}(ZX')^{-1}\mathbb{E}(ZY)$. The case where $\ell < k$, so we have less instruments than regressors, is not of interest to us here as then β cannot be uniquely determined. This is the underidentified case.

A forms k linear combinations of the ℓ instrumental variables to transform the overidentified system into a just identified system. Define $\check{Z}_A = A'Z$. These are k instrumental variables for which the just-identified system yields

$$\beta = \mathbb{E}(\check{Z}_A X')^{-1}\mathbb{E}(\check{Z}_A Y).$$

Different choices of A give estimators with different asymptotic variances. The optimal choice, in that it makes the asymptotic variance as small as possible, turns out to be $A = \mathbb{E}(XZ')\mathbb{E}(ZZ'e^2)^{-1}$. We will return to this below.

2. A supply and demand example

Let us temporarily deviate from our notational conventions to analyze a market model where demand d and supply s curves are linear in price p, as in

$$d = \alpha_d - \theta_d \, p + u,$$

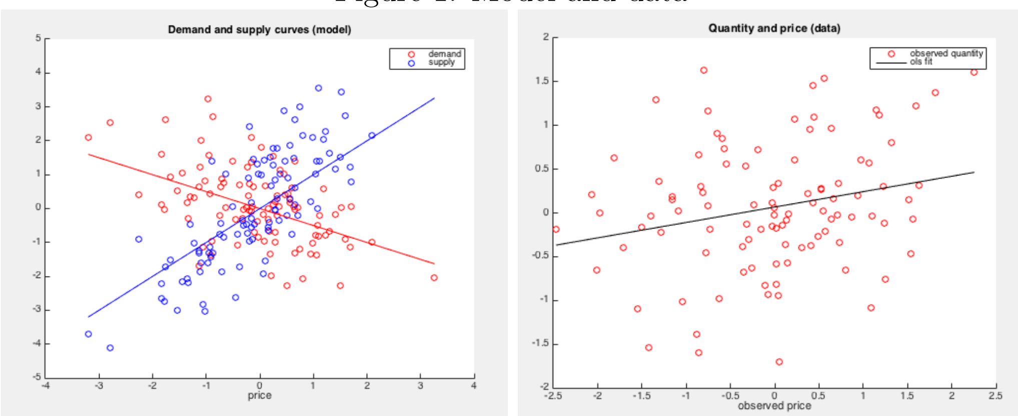
$$s = \alpha_s \; + \theta_s \, p + v,$$

where u, v are unobserved errors. We will take these errors to be independent and homoskedastic, with variances $\sigma_u^2 = \mathbb{E}(u^2)$ and $\sigma_v^2 = \mathbb{E}(v^2)$. Suppose we would like to learn the demand curve. Whether this is an easy task or not depends on the type of data at hand.

2.1. The inconsistency of least squares

The cloud of points in the left plot of Figure 2.1 corresponds to experimental or survey data, where sellers (blue) and buyers (red) are asked how much they would sell or buy for a range of different prices (on the horizontal axis). Such data clearly would allow us to trace out the supply and demand curves. It is, however, not the data an economist would usually have at his or her disposal.

Figure 1: Model and data



The data usually available are based on transactions. Here, we observe only what was traded, q, and at what price, p. Such a scenario is given in the right plot of Figure 2.1. The black curve is a least squares fit through this cloud of points. Although it is upward sloping, it is not a consistent estimator of the supply curve. It is also not a consistent estimator of the demand curve. This is so because, in this specification, the variable price is endogenous.

To see this note that q is obtained by the market clearing condition that $s = d$. This equilibrium is obtained by letting the price p adjust. If we solve

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

for the equilibrium price we get

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

This then also gives traded quantity as

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

The slope on p in a population regression of q on a constant and p equals

$$\frac{\operatorname{cov}(p, q)}{\operatorname{var}(p)} = \frac{\operatorname{cov}(u - v, \theta_s u + \theta_d v)}{\operatorname{var}(u - v)} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_v^2}\,\theta_s - \frac{\sigma_v^2}{\sigma_u^2 + \sigma_v^2}\,\theta_d.$$

This is a weighted average of supply and demand elasticities. The left plot in Figure 2.1 shows this clearly by combining the demand and supply curves from before with the least squares fit.

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,

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

so price is not exogenous but endogenous.

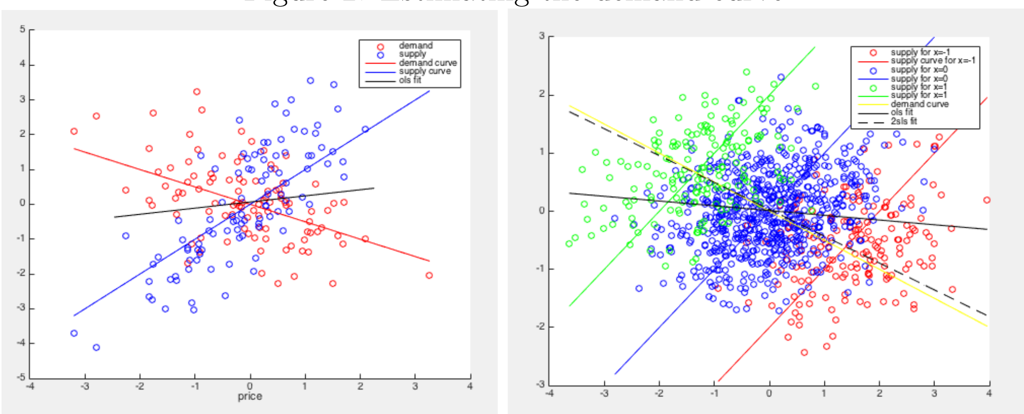
The same happens for the supply curve, as

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

and

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

Figure 2: Estimating the demand curve



2.2. The instrumental-variable solution

To make progress we introduce an instrumental variable z. We thus 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, C2) but not demand (exclusion, C1). A simple example in the market for fish, say, would be a variable related to weather on the ocean. Worse weather makes catching fish more difficult and reduces supply, but does not affect the consumer's appetite for fish.

Proceeding as before shows that 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 $\operatorname{cov}(u, z) = 0$,

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

and so, provided that $\operatorname{cov}(p, z) \neq 0$,

$$-\theta_d = \frac{\operatorname{cov}(d, z)}{\operatorname{cov}(p, z)},$$

the price elasticity of demand. The right plot in Figure 2.1 illustrates what the instrument allows to achieve. The observations in the scatter plot are color-coded according to the value that the instrument can take (here, $z \in \{-1, 0, 1\}$). Changes in z induce a shift in the supply curve. These changes allow to trace out the demand curve, which is immune to changes in the instrument.

To identify the supply curve we would need an instrument that shifts demand but not supply.

3. Estimation by instrumental variables

We will work under the following additional assumptions.

C3. Random sampling

C4. $\mathbb{E}(|Y_1|^4) < +\infty$, $\mathbb{E}(\|X\|^4) < +\infty$, and $\mathbb{E}(\|Z\|^4) < +\infty$.

C5. $\Omega = \mathbb{E}(ZZ'e^2)$ is positive definite.

3.1. A class of estimators

As before, stacking the observations in the sample gives

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

together with an $n \times \ell$ matrix of instrumental variables $\boldsymbol{Z}$. When $\ell > k$ we will generally not be able to find a b for which

$$\boldsymbol{Z}'(\boldsymbol{Y} - \boldsymbol{X}b) = 0$$

holds exactly. The solution is to take k linear combinations of the columns of $\boldsymbol{Z}$. An equivalent insightful way of formulating this is to consider minimizing a distance of the $\ell \times 1$ vector $\boldsymbol{Z}'(\boldsymbol{Y} - \boldsymbol{X}b)$ from zero. This can be written as

$$\min_b (\boldsymbol{Y} - \boldsymbol{X}b)'\boldsymbol{Z}\,W\boldsymbol{Z}'(\boldsymbol{Y} - \boldsymbol{X}b)$$

for a chosen full-rank weight matrix W that defines the distance metric. It is easy to see that this corresponds to setting $A = \boldsymbol{X}'\boldsymbol{Z}\,W$.

The first-order condition is

$$\boldsymbol{X}'\boldsymbol{Z}\,W\boldsymbol{Z}'(\boldsymbol{Y} - \boldsymbol{X}b) = 0$$

and has solution

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

This delivers class of generalized method-of-moment estimators, indexed by the choice of weight matrix W.

3.2. Limit distribution

For the moment suppose that the matrix W is known. Afterwards one can deduce that replacing W in our arguments by a consistent estimator $\hat{W}$ does not alter any of the results. Consistency of the IV estimator follows by the usual arguments, relying on

$$\frac{\boldsymbol{Z}'\boldsymbol{Z}}{n} \underset{p}{\rightarrow} \mathbb{E}(ZX'), \qquad \frac{\boldsymbol{Z}'\boldsymbol{e}}{n} \underset{p}{\rightarrow} \mathbb{E}(Ze) = 0,$$

as $n \to \infty$ under C1, C3, and C4. We focus on deriving the limit distribution here.

To do so we write

$$\hat{\beta} - \beta = \left(\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\, W \,\frac{\boldsymbol{Z}'\boldsymbol{X}}{n}\right)^{-1} \left(\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\, W \,\frac{\boldsymbol{Z}'\boldsymbol{e}}{n}\right)$$

and note that

$$\left(\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\, W \,\frac{\boldsymbol{Z}'\boldsymbol{X}}{n}\right)^{-1} \frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\, W \underset{p}{\rightarrow} (\mathbb{E}(XZ')\, W\, \mathbb{E}(ZX'))^{-1}\, \mathbb{E}(XZ')\, W = Q \text{ (say)}$$

as $n \to \infty$. This follows from the same arguments as above, together with the fact that the rank conditions on W and $\mathbb{E}(XZ')$ ensure that the inverse matrix is well defined.

Therefore,

$$\sqrt{n}(\hat{\beta}-\beta) \underset{p}{\to} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Q(Z_i e_i) \underset{d}{\to} N(0, Q\Omega Q'),$$

where the last transition follows by the central limit theorem, as validated by C1, C3, and C5.

3.3. Optimal (two-step) estimation

The asymptotic variance, written out in full, is the lengthy formula

$$(\mathbb{E}(XZ')W\mathbb{E}(ZX'))^{-1}\mathbb{E}(XZ')\,W\Omega W\,\mathbb{E}(ZX')\;(\mathbb{E}(XZ')W\mathbb{E}(ZX'))^{-1}$$

and shows how the limit distribution depends on the choice of W. This means that some linear combinations of the instrumental variables yield more precise estimators than others.

It can be shown that the asymptotic variance is minimized by the choice

$$W = \mathbb{E}(ZZ'e^2)^{-1} = \Omega^{-1}.$$

In this case the asymptotic variance formula simplifies considerably, and becomes

$$(\mathbb{E}(XZ')\mathbb{E}(ZZ'e^2)^{-1}\mathbb{E}(ZX'))^{-1}.$$

Thus, the optimal instrumental-variable estimator uses the particular linear combination

$$\check{Z}_A = A'Z = \boldsymbol{X}'\boldsymbol{Z}\,\Omega^{-1}Z.$$

Of course, this formula is not feasible in that it depends on $\Omega = \mathbb{E}(ZZ'e^2)$, which is unknown.

This suggests a two-step procedure. In a first step, we obtain a preliminary estimator of β by using a known weight matrix W (which may not be the optimal one). A popular choice is $W = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}$ (see below for more on this). Let $\tilde{\beta}$ be the estimator so constructed. Under our assumptions this estimator is consistent, and so

$$\tilde{e}_i = Y_i - X_i'\tilde{\beta}$$

is a valid estimator for e_i. These, in turn, allow the construction of the estimator

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

which can be shown to be consistent for Ω. In the second step we then compute the optimal estimator as

$$\hat{\beta} = (\boldsymbol{X}'\boldsymbol{Z}\,\hat{\Omega}^{-1}\boldsymbol{Z}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{Z}\,\hat{\Omega}^{-1}\boldsymbol{Z}'\boldsymbol{Y}),$$

which is asymptotically efficient.

3.4. Two-stage least squares

When

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

we have that

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

From the expression of $\hat{\beta}$ it is apparent that the estimator remains unchanged of W is replaced by $c\,W$ for any scalar $c \neq 0$. Thus, an estimator of the optimal weight matrix, up to the scale factor σ^2/n, is given by $\boldsymbol{Z}'\boldsymbol{Z}$. The associated optimal estimator, which here can be computed in a single step, is thus

$$\hat{\beta}_{2SLS} = (\boldsymbol{X}'\boldsymbol{Z}\,(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{Z}\,(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y}).$$

Recalling that $\boldsymbol{Z}\,(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z} = \boldsymbol{P_Z}$ is the projection onto the column space of the instrument matrix $\boldsymbol{Z}$ we see that

$$\hat{\beta}_{2SLS} = (\boldsymbol{X}'\boldsymbol{P_Z}\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{P_Z}\boldsymbol{Y})$$

can be calculated by running two ordinary least-squares regressions. The first is a regression of $\boldsymbol{X}$ on $\boldsymbol{Z}$ to extract the fitted values $\boldsymbol{P_Z}\boldsymbol{X}$. The second is a regression of $\boldsymbol{Y}$ on these fitted values $\boldsymbol{P_Z}\boldsymbol{X}$. This simple interpretation makes 2SLS a popular choice in practice, even when homoskedasticity is not satisfied (in which case robust standard errors should be used for correct inference).

4. Inference

4.1. Hypothesis tests on β

Given that

$$\sqrt{n}(\hat{\beta} - \beta) \underset{d}{\rightarrow} N(0, V_\beta)$$

for $V_\beta = Q\Omega Q'$ we can test hypothesis on β of the form $\mathbb{H}_0 : \theta = r(\beta) = \theta_0$ in the same way as before. Moreover, with $\hat{V}_\theta$ a plug-in estimator of the variance $V_\theta = R'(\beta)V_\beta R(\beta)$ the conventional Wald statistic converges under the null to

$$W = n\,(\hat{\theta} - \theta_0)'\hat{V}_\theta^{-1}(\hat{\theta} - \theta_0) \underset{d}{\rightarrow} \chi^2_m,$$

where m is the number of restrictions being tested. Note that this procedure does not require the use of the optimally-weighted estimator provided, of course, that an estimator of the proper V_β is used in the construction of the test statistic.

4.2. The J-test

The framework allows to test whether the moment conditions are valid. The optimally-weighted objective function was

$$(\boldsymbol{Y} - \boldsymbol{X}b)'\boldsymbol{Z}\hat{\Omega}^{-1}\boldsymbol{Z}'(\boldsymbol{Y} - \boldsymbol{X}b).$$

Evaluating this at $b = \beta$ and re-scaling by n^{-1} gives

$$\frac{\boldsymbol{e}'\boldsymbol{Z}}{\sqrt{n}}\hat{\Omega}^{-1}\frac{\boldsymbol{Z}'\boldsymbol{e}}{\sqrt{n}} = \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i e_i\right)'\hat{\Omega}^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i e_i\right).$$

When the instrumental variables all satisfy C1, the central limit theorem ensures that

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

Thus, in this case,

$$\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i e_i\right)'\hat{\Omega}^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i e_i\right) \underset{d}{\rightarrow} \chi^2_\ell.$$

This implies that a test of the null that C1 holds could be constructed by comparing the scaled objective function at the truth to the quantiles of the relevant chi-squared distribution. Observe how the use of the optimal weight matrix is important in reaching this conclusion.

Of course, the above strategy is not feasible because we do not know β.

However, with some work it can be shown that, under $\mathbb{H}_0 : \mathbb{E}(Ze) = 0$, we have

$$J = \frac{\hat{\boldsymbol{e}}'\boldsymbol{Z}}{\sqrt{n}} \hat{\Omega}^{-1} \frac{\boldsymbol{Z}'\hat{\boldsymbol{e}}}{\sqrt{n}} \underset{d}{\rightarrow} \chi^2_{\ell-k},$$

where we have now used residuals from the optimal estimator to construct a feasible version of the statistic in question. As before, large values for the J-statistic suggest that the validity of (at least some of) the instruments is in doubt. Notice, though, that for the feasible statistic we have reduced the degrees of freedom of the null distribution from ℓ down to $\ell - k$, which is the number of over-identifying restrictions. This means that this approach is not useful when we have as many instruments as regressors in the model (as then $J = 0$). The intuition for this is that we need k instruments to be able to learn β. As a consequence, the power properties of this type of test are complicated.

For the case where the two-stage least squares estimator is the optimally-weighted estimator the J-statistic is

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

where $s^2 = \hat{\boldsymbol{e}}'\hat{\boldsymbol{e}}/(n-k)$ and R^2 refers to the coefficient of variation for the regression of $\hat{e}_i$ on Z_i. This is often called the Sargan statistic. Again, in the just-identified case where $\ell = k$ such an approach would be useless as the residuals are constructed to be exactly uncorrelated to the instruments by construction.

4.3. The incremental J-test

The above approach can help to flag that some instruments are invalid but it is not informative about which ones. An incremental test strategy is helpful here. The strategy is as follows.

Suppose that the validity of a subset of $\ell_0 \geq k$ instruments is not under debate. Collect these instruments in $\boldsymbol{Z}_0$. We can compute the corresponding estimator

$$\hat{\beta}_0 = \arg\min_b (\boldsymbol{Y} - \boldsymbol{X}b)'\boldsymbol{Z}_0 \hat{\Omega}^{-1}\boldsymbol{Z}_0'(\boldsymbol{Y} - \boldsymbol{X}b).$$

This yields residuals $\hat{\boldsymbol{e}}_0 = \boldsymbol{Y} - \boldsymbol{X}\hat{\beta}_0$ and the associated J-statistic

$$J_0 = \frac{\hat{\boldsymbol{e}}_0'\boldsymbol{Z}_0}{\sqrt{n}}\hat{\Omega}^{-1}\frac{\boldsymbol{Z}_0'\hat{\boldsymbol{e}}_0}{\sqrt{n}}.$$

We can also proceed as before and use the full set of instruments $\boldsymbol{Z} = (\boldsymbol{Z}_0, \boldsymbol{Z}_1)$ to compute the estimator $\hat{\beta}$ and its associated J-statistic J from before. It turns out that, under the maintained hypothesis that $\mathbb{E}(Z_0 e) = 0$, we have that, when $\mathbb{H}_0 : \mathbb{E}(Z_1 e) = 0$,

$$J - J_0 \underset{d}{\to} \chi^2_{\ell-\ell_0},$$

which can be used to construct an asymptotically-valid test for the null that the remaining $\ell - \ell_0$ instruments are valid.

There is another variation of the incremental J-statistic idea that is useful. Suppose that the validity of the instruments is not in doubt but we would like to test the null

$$\mathbb{H}_0 : \mathbb{E}(Xe) = 0,$$

that is, that the regressors are, in fact exogenous (the construction for testing the exogeneity of a subset of is analogous). Under the null,

$$J_1 - J \underset{d}{\to} \chi^2_k,$$

where J is the J-statistic from before, based on instrumental variables $\boldsymbol{Z}$ and J_1 is the J-statistic obtained from running our procedure using both $\boldsymbol{Z}$ and $\boldsymbol{X}$ as instruments.