Econometrics
Part II: Further topics. Instrumental variables.
Session 7.

References on Bruce Hansen's textbook will be abbreviated with H. Say, H 1.1 means chapter 1.1 from Bruce Hansen's textbook.

Errors in variables. Endogeneity. (H 12.1-12.4)
Instrumental variables. 2SLS. Control function approach (H 12.5-12.9, 12.12, 12.15-12.18, and 12.28)
Endogeneity tests. Overidentification tests. (H 12.29, 12.31)
Irrelevant instruments. (These notes, H. 12.35)

Classical measurement error.

Measurement errors plague most data sets to one degree or another. Suppose that we observe noisy versions of the variables we would like to observe. We obtain data y_i and x_i for $i = 1, ..., n$, while the true values are y_i^* and x_i^*. Also assume the following

$$\begin{aligned} x_i &= x_i^* + \nu_i \\ y_i &= y_i^* + \eta_i \end{aligned}$$

where the errors in measurement are such that the following hold:

$$\begin{aligned} E(\nu_i) = 0, &\quad E(\eta_i) = 0 \\ E(x_i^*\nu_i) = 0, &\quad E(y_i^*\eta_i) = 0 \\ E(x_i^*\eta_i) = 0, &\quad E(y_i^*\nu_i) = 0 \\ E(\nu_i\eta_i) = 0. & \end{aligned}$$

Given that $E(y_i^*|x_i^*) = x_i^{*\prime}\beta$, if we proceeded as if there were no measurement error, we would estimate the following by OLS

$$\begin{aligned} \hat{\beta}_{OLS} &= \quad (X'X)^{-1} \quad X'Y \\ &= \left(\tfrac{1}{n}\sum x_i x_i'\right)^{-1} \tfrac{1}{n}\sum x_i y_i' \\ &= \quad \downarrow p \quad \downarrow p \\ &= \quad (Ex_i x_i')^{-1} \; E(x_i y_i) \end{aligned}$$

Note that

$$\begin{aligned} E(x_i x_i') &= E\left[(x_i^* + \nu_i)(x_i^* + \nu_i)'\right] \\ &= E[x_i^* x_i^{*\prime}] + E[\nu_i \nu_i'] + E[x_i^* \nu_i'] + E[\nu_i x_i^{*\prime}] \\ &= E[x_i^* x_i^{*\prime}] + E[\nu_i \nu_i'] \end{aligned}$$

$$\begin{aligned} E(x_i y_i) &= E[(x_i^* + \nu_i)(y_i^* + \eta_i)] \\ &= E[x_i^* y_i^*] + E[\nu_i \eta_i] + E[x_i^* \eta_i] + E[\nu_i y_i^*] \\ &= E[x_i^* y_i^*] \\ &= E[x_i^* x_i^{*\prime}]\,\beta \end{aligned}$$

Thus, when measurement error in the independent variables exists, so that $Var(\nu_i) \neq 0$, OLS yields an inconsistent estimator:

$$\hat{\beta}_{OLS} \xrightarrow{p} \left(E[x_i^* x_i^{*\prime}] + E[\nu_i \nu_i']\right)^{-1} E[x_i^* x_i^{*\prime}]\,\beta$$

Note that when we observe x^* directly, so that $E[\nu_i \nu_i'] = 0$, then $\hat{\beta}_{OLS} \xrightarrow{p} \beta$ as expected.

In the univariate case, the above reduces to:

$$\hat{\beta}_{OLS} \xrightarrow{p} \underbrace{\frac{\sigma_{x^*}^2}{\sigma_{x^*}^2 + \sigma_\nu^2}}_{\text{signal/noise ratio}} \beta$$

so the estimate $\hat{\beta}_{OLS}$ is biased towards 0 and inconsistent.

If there were no measurement errors, we would be using

$$y_i^* = x_i^{*\prime}\beta + \varepsilon_i$$

Any measurement error in the dependent variables is subsumed in the error term as follows:

$$y_i = y_i^* + \eta_i = x_i^{*\prime}\beta + (\varepsilon_i + \eta_i)$$

This new error term, $\varepsilon_i + \eta_i$ creates no problem for estimation, as η_i is uncorrelated with x_i^*. However, if there were measurement errors in x_i^* as well, then we have

$$y_i = x_i'\beta + \underbrace{(\varepsilon_i + \eta_i - \nu_i'\beta)}_{u_i}$$

In this case, the new error term, denoted u_i, is correlated with x_i.

$$E[x_i\nu_i'] = E[(x_i^* + \nu_i)\nu_i'] = E[x_i^*\nu_i'] + E[\nu_i\nu_i']$$

Because the error term is correlated with x_i, (OLS2') $Ex_iu_i = 0$ does not hold, and OLS is inconsistent.

There are two potential solutions to the problem of measurement error.

Solution 1. If we can estimate $E[\nu_i\nu_i']$, we can undo the error in estimation by using the following

$$E[x_ix_i'] = E[x_i^*x_i^{*\prime}] + E[\nu_i\nu_i']$$

Unfortunately, this is not usually possible.

Solution 2. Suppose we get another independent measure of x^* such that

$$w_i = x_i^* + \tau_i$$

and τ_i is uncorrelated with any of $y_i^*, x_i^*, \eta_i, \nu_i$.

$$\begin{aligned} E[w_ix_i'] &= E\left[(x_i^* + \tau_i)(x_i^* + \nu_i)'\right] = E[x_i^*x_i^{*\prime}] \\ E[w_iy_i] &= E[(x_i^* + \tau_i)(y_i^* + \eta_i)] = E[x_i^*y_i^*] \end{aligned}$$

Then if $E[w_ix_i']$ is invertible, we have

$$E[w_ix_i']^{-1}E[w_iy_i] = E[x_i^*x_i^{*\prime}]^{-1}E[x_i^*y_i^*] = E[x_i^*x_i^{*\prime}]^{-1}E[x_i^*x_i^{*\prime}]\beta = \beta$$

So $\hat{\beta}_{IV} = (W'X)^{-1}W'Y$ is consistent.

The idea behind this repeated observation estimator extends well beyond the repeated observation case and even beyond the measurement error case. Next we will discuss instrumental variables estimation.

Endogeneity

Consider the following model

$$y_i = x_i'\beta + \varepsilon_i \text{ where } E(x_i\varepsilon_i) \neq 0 \text{ in violation of OLS2 and OLS2'.}$$

This is a core problem in econometrics that distinguishes it from statistics. To distinguish this model from the regression model, we will call the above equation a structural equation, and β a structural parameter. By structural equation we understand one that represents a causal link rather than just an empirical association. When $E(x_i\varepsilon_i) \neq 0$ holds we say that x_i is endogenous for β.

To be clear, usually $E(x_i\varepsilon_i) \neq 0$ is caused by only a few components of x_i being correlated with ε_i. The components causing the problems are then called endogenous and the rest are called exogenous. By rearranging explanatory variables, we may partition x_i into the exogenous part x_{1i} and endogenous part x_{2i}.

The endogeneity may happen in many ways. We have already seen how it arises in the context of measurement error. Another classical example is the Supply and Demand system, where q_i and p_i (quantity and price) are determined jointly by the demand equation

$$q_i = -\beta_d p_i + \varepsilon_{di}$$

and the supply equation

$$q_i = \beta_s p_i + \varepsilon_{si}.$$

In matrix notation,

$$\begin{pmatrix} 1 & \beta_d \\ 1 & -\beta_s \end{pmatrix} \begin{pmatrix} q_i \\ p_i \end{pmatrix} = \begin{pmatrix} \varepsilon_{di} \\ \varepsilon_{si} \end{pmatrix}$$

so that

$$\begin{aligned} \begin{pmatrix} q_i \\ p_i \end{pmatrix} &= \frac{1}{-\beta_s - \beta_d} \begin{pmatrix} -\beta_s & -\beta_d \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{di} \\ \varepsilon_{si} \end{pmatrix} \\ &= \begin{pmatrix} (\beta_s\varepsilon_{di} + \beta_d\varepsilon_{si}) / (\beta_s + \beta_d) \\ (\varepsilon_{di} - \varepsilon_{si}) / (\beta_s + \beta_d) \end{pmatrix} \end{aligned}$$

so neither $E[p_i\varepsilon_{di}] = 0$ nor $E[p_i\varepsilon_{si}] = 0$. If we run an OLS of q_i on p_i we estimating $Cov(q_i, p_i)/Var(p_i)$. Assuming the demand and supply shocks are uncorrelated, this quantity equals

$$\begin{aligned} \frac{Cov(q_i, p_i)}{Var(p_i)} &= \frac{\frac{\beta_s}{(\beta_s+\beta_d)^2} Var(\varepsilon_{di}) - \frac{\beta_d}{(\beta_s+\beta_d)^2} Var(\varepsilon_{si})}{\frac{1}{(\beta_s+\beta_d)^2} Var(\varepsilon_{di}) + \frac{1}{(\beta_s+\beta_d)^2} Var(\varepsilon_{si})} \\ &= \beta_s \frac{Var(\varepsilon_{di})}{Var(\varepsilon_{di}) + Var(\varepsilon_{si})} - \beta_d \frac{Var(\varepsilon_{si})}{Var(\varepsilon_{di}) + Var(\varepsilon_{si})}, \end{aligned}$$

that is, some linear combination of the slopes of the demand and the supply curves.

Our last example of endogeneity would be a structural equation connecting two variables that are both chosen by economic agents, say, wage and education

$$wage_i = \beta_1 + \beta_2 educ_i + \varepsilon_i.$$

Both $wage_i$ and $educ_i$ may be affected by person i's ability or some other factor belonging to ε_i. Here the structural equation can be thought of as reflecting a causal relationship between education and wage, which would be observed if we were able to completely randomly assign education levels to people, independent on their abilities or anything else. In real situation, education is not assigned. It is chosen by people, and this choice may be affected by other factors influencing wage.

Instrumental variables

Consider a linear regression model

$$y_i = x_i'\beta + \varepsilon_i,$$

where x_i is endogenous. Suppose that we have a variable w_i (called instrumental variable) such that

$$\begin{aligned} &E[w_i\varepsilon_i] = 0 \text{ (exogeneity)} \\ &E[w_iw_i'] > 0 \text{ (no redundant instruments)} \\ &E[w_ix_i'] \text{ has full column rank (relevance)} \end{aligned}$$

Then

$$0 = E[w_i\varepsilon_i] = E[w_i(y_i - x_i'\beta)] = E[w_iy_i] - E[w_ix_i']\beta.$$

If $E[w_ix_i']$ is invertible, then

$$\beta = (E[w_ix_i'])^{-1} E[w_iy_i].$$

This motivates the following estimator $\hat{\beta}_{IV}$

$$\begin{aligned} \hat{\beta}_{IV} &= (W'X)^{-1}(W'Y) = \left(\frac{1}{n}\sum w_ix_i'\right)^{-1}\frac{1}{n}\sum w_iy_i \\ &\xrightarrow{p} (E[w_ix_i'])^{-1}(E[w_ix_i']\beta + E[w_i\varepsilon_i]) = \beta. \end{aligned}$$

Note that the Instrumental Variables estimator (IV) reduces to OLS when $W = X$.

Examples.

In the errors in variables examples, the instrumental variable was equal to the independent noisy observation of x_i^*.

In the demand and supply example, instruments are often related to so called demand and supply shifters. Suppose that the market is a local fish market as in Graddy (1995). Then, we may think that the supply would be affected by weather off shore, w_i, so that

$$q_i = \beta_s p_i + \gamma w_i + \varepsilon_{si}$$

whereas demand will not be directly affected by w_i so that

$$q_i = \beta_d p_i + \varepsilon_{di}.$$

Then, w_i can be used as an instrument for p_i in the estimation of the demand (but not supply) equation.

In the wage-education example, many proposals were made. For example, Angrist and Krueger (1991) propose the quarter of birth indicator as the instrument for education. Due to compulsory education laws in the United States, you cannot drop out from school until you are 16, so people who are born in the first quarter of the year, being oldest in their class, may drop out more often than those born in the other quarters. This insures that $E[w_i x_i] \neq 0$, whereas, arguably, the quarter of birth should not be related to any other determinants of your wage, so that $E[w_i \varepsilon_i] = 0$.

Under-, just- and over-identification.

Not all components of x_i have to be endogenous. We may, for example, have

$$y_i = x_{1i}'\beta_1 + x_{2i}'\beta_2 + \varepsilon_i \text{ with } E(x_{1i}\varepsilon_i) = 0 \text{ and } E(x_{2i}\varepsilon_i) \neq 0.$$

Then, we can take

$$w_i = \begin{pmatrix} x_{1i} \\ z_i \end{pmatrix}$$

with x_{1i} being "instruments for themselves" (or included exogenous variables) and z_i (called excluded exogenous variables) instrumenting the endogenous variables

x_{2i}. If w_i is l-dimensional and x_i is k-dimensional, we have

$$\underbrace{E\left[w_i y_i\right]}_{l\times 1} = \underbrace{E\left[w_i x_i'\right]}_{l\times k} \underbrace{\beta}_{k\times 1}$$

If $l < k$, there will exist in general many solutions to β, and we say that this parameter is under-identified. If $l = k$, the number of equations equals the number of unknowns, and we say that β is just-identified. If $l > k$, there are more equations than unknowns, and we say that β is over-identified. $E\left[w_i x_i'\right]$ is no longer invertible. We could throw away the extra variables, but a better option is to use the Two-Stage Least Squares.

Two-Stage Least Squares.

For now, let us assume that $E\left[\varepsilon_i | w_i\right] = 0$

$$\begin{aligned} 0 &= E\left[\varepsilon_i | w_i\right] \\ &= E\left[y_i - x_i'\beta | w_i\right] \\ &= E\left[y_i | w_i\right] - E\left[x_i' | w_i\right]\beta \end{aligned}$$

$$E\left[y_i | w_i\right] = E\left[x_i' | w_i\right]\beta$$

Suppose we also know that

$$E\left[x_i' | w_i\right] = w_i'\pi$$

Then we have

$$E\left[y_i | w_i\right] = \left(w_i'\pi\right)\beta$$

This suggests the following two stage procedure:

Stage1: Regress $X_{n\times k}$ on $W_{n\times l}$ to find $\hat{\pi} = \left(W'W\right)^{-1} W'X$. Use the results to form $\hat{X} = W\hat{\pi}$. note that

$$\hat{X} = W\hat{\pi} = W\left(W'W\right)^{-1} W'X = P_W X$$

Stage2: Regress Y on $\hat{X}$ to find

$$\begin{aligned} \hat{\beta}_{2SLS} &= \left(\hat{X}'\hat{X}\right)^{-1} \hat{X}'Y = \left(X'P_W'P_W X\right)^{-1} X'P_W'Y \\ &= \left(X'P_W X\right)^{-1} X'P_W Y \end{aligned}$$

Consider the following IV assumptions for the model $y_i = x_i'\beta + \varepsilon_i$

(IV0) (y_i, x_i, w_i) is i.i.d. sequence
(IV1) $E[w_i w_i'] < \infty$ non-singular; $E[w_i x_i']$ has full column rank (relevance)
(IV2) $E[\varepsilon_i | w_i] = 0$ or (IV2') $E(w_i \varepsilon_i) = 0$ (exogeneity)
(IV3) $E[\varepsilon_i^2 | w_i] = \sigma^2$ or (IV3') $V = Var(w_i \varepsilon_i)$ is finite nonsingular.

Theorem 1 *Under IV0-1-2', $\hat{\beta}_{2SLS} \xrightarrow{p} \beta$*

Proof:

$$\begin{aligned}
\hat{\beta}_{2SLS} &= \left(\hat{X}'\hat{X}\right)^{-1} \hat{X}Y = \left(X'P_W X\right)^{-1} X'P_W Y \\
&= \beta + \left(X'P_W X\right)^{-1} X'P_W \varepsilon
\end{aligned}$$

$$\begin{array}{cccccccc}
& \hat{\beta}_{2SLS} - \beta = & & & & & \\
= & [X'W & (W'W)^{-1} & W'X]^{-1} & X'W & (W'W)^{-1} & W'\varepsilon \\
= & [\frac{1}{n}\sum x_i w_i' & \left(\frac{1}{n}\sum w_i w_i'\right)^{-1} & \frac{1}{n}\sum w_i x_i']^{-1} & \frac{1}{n}\sum x_i w_i' & \left(\frac{1}{n}\sum w_i w_i'\right)^{-1} & \frac{1}{n}\sum w_i \varepsilon_i \\
& \downarrow p & \downarrow p & \downarrow p & \downarrow p & \downarrow p & \downarrow p \\
\xrightarrow{p} & [E(x_i w_i') & (E(w_i w_i'))^{-1} & E(w_i x_i')]^{-1} & E(x_i w_i') & (E(w_i w_i'))^{-1} & E(w_i \varepsilon_i) = 0 \\
= & 0 & & & & &
\end{array}$$

Note that, in general, $\dim W \neq \dim X$, but our procedure insures that $\dim \hat{X} = \dim X$, so $\hat{\beta}_{2SLS}$ is $\hat{\beta}_{IV}$ using $\hat{X}$ as the instrument,

$$\hat{\beta}_{2SLS} = \left(\hat{X}'X\right)^{-1} \hat{X}'Y.$$

Theorem 2 *Let $C = E[w_i w_i']$ and $D = E[w_i x_i']$. Under IV0-1-2'-3',*

$$\sqrt{n}\left(\hat{\beta}_{2SLS} - \beta\right) \xrightarrow{d} N\left(0, \left(D'C^{-1}D\right)^{-1} D'C^{-1}VC^{-1}D\left(D'C^{-1}D\right)^{-1}\right)$$

Proof:

$$\begin{array}{cccccccc}
& \hat{\beta}_{2SLS} - \beta = & & & & & \\
= & [\frac{1}{n}X'W & \left(\frac{1}{n}W'W\right)^{-1} & \frac{1}{n}W'X]^{-1} & \frac{1}{n}X'W & \left(\frac{1}{n}W'W\right)^{-1} & \frac{1}{\sqrt{n}}W'\varepsilon \\
= & [\frac{1}{n}\sum x_i w_i' & \left(\frac{1}{n}\sum w_i w_i'\right)^{-1} & \frac{1}{n}\sum w_i x_i']^{-1} & \frac{1}{n}\sum x_i w_i' & \left(\frac{1}{n}\sum w_i w_i'\right)^{-1} & \frac{1}{\sqrt{n}}\sum w_i \varepsilon_i \\
& \downarrow p & \downarrow p & \downarrow p & \downarrow p & \downarrow p & \downarrow d \\
\xrightarrow{d} & [E(x_i w_i') & (E(w_i w_i'))^{-1} & E(w_i x_i')]^{-1} & E(x_i w_i') & (E(w_i w_i'))^{-1} & N(0, V) \\
= & [D' & C^{-1} & D]^{-1} & D' & C^{-1} & N(0, V)
\end{array}$$

Under (IV3) (homoskedasticity),

$$V = Var\left(w_i \varepsilon_i\right) = \sigma^2 E\left[w_i w_i'\right] = \sigma^2 C$$

Then, much of the asymptotic variance cancels, leaving

$$\sqrt{n}\left(\hat{\beta}_{2SLS} - \beta\right) \xrightarrow{d} N\left(0, \sigma^2 \left(D' C^{-1} D\right)^{-1}\right)$$

This can be estimated by $\hat{\sigma}^2 \left(\frac{1}{n}\hat{X}'\hat{X}\right)^{-1}$, where $\hat{\sigma}^2 = \hat{\varepsilon}'\hat{\varepsilon}/n$ and $\hat{\varepsilon} = Y - X\hat{\beta}_{2SLS}$. Under (IV3'), or heteroskedasticity, we can use White's estimate as in earlier discussions

$$\hat{V} = \frac{1}{n}\sum\nolimits_{i=1}^{n} \hat{\varepsilon}_i^2 w_i w_i'$$

Homoskedasticity only, as well as robust estimates of the variance of $\hat{\beta}_{2SLS}$ can be used to form F-statistics for testing linear hypotheses in the usual way. Asymptotically, such F-statistics would be distributed as $\chi^2(p)/p$, where p is the number of restrictions. However, the finite sample distribution of the F-statistic would not be $F(p, n-k)$ even if ε_i is normally distributed.

Control function approach

There is an alternative way to derive the 2SLS estimator. Suppose that

$$y_i = x_{1i}'\beta_1 + x_{2i}'\beta_2 + \varepsilon_i$$

with endogenous x_{2i}. The idea of the 2SLS explored above was to extract the exogenous part $w_i'\pi$ of x_i and use it in the second stage regression. An alternative would be to extract the endogenous part of x_i (control function) and add it to the regression as an additional regressor. The two approaches are equivalent. Let us see why.

The exogenous part, $w_i'\pi$, of x_i is nothing else as the best linear predictor of x_i given w_i. The first stage regression(s)

$$x_i' = w_i'\pi + u_i'$$

are called reduced form because they do not have any structural interpretation, we just want to predict x_i by a linear function of w_i in the best possible way. Recall that w_i contains two components: included exogenous variables x_{1i} and excluded exogenous variables z_i. Therefore, the reduced form equations can be partitioned into

$$\begin{aligned} x_{1i}' &= x_{1i}'\pi_{11} + z_i'\pi_{12} + u_{1i}' \\ x_{2i}' &= x_{1i}'\pi_{21} + z_i'\pi_{22} + u_{2i}' \end{aligned}$$

Of course, the BLP of x_{1i} given x_{1i} and z_i is just x_{1i}, so the first of the above two equations is trivial: $x_{1i}' = x_{1i}'$. For the second equation, we will write, slightly abusing notation

$$x_{2i}' = x_{1i}'\pi_1 + z_i'\pi_2 + u_i'$$

In the first stage of the 2SLS procedure, one may run only this regression, obtain $\hat{x}_{2i}$, form $\hat{x}_i$ by combining x_{1i} with $\hat{x}_{2i}$ and proceed to the second stage.

Alternatively, note that x_{2i} can only be endogenous if $E(u_i\varepsilon_i) \neq 0$, that is, the error of the first stage regression, u_i, is correlated with the structural error ε_i. In a sense, the error u_i soaks up the endogeneity in x_{2i}. So, if we could add it to the structural equation as an additional regressor, we would be controlling for the endogeneity and get consistent estimates of the structural parameters.

Consider the BLP of ε_i given u_i :

$$\varepsilon_i = u_i'\alpha + e_i.$$

Substituting this into the structural equation, we obtain

$$y_i = x_{1i}'\beta_1 + x_{2i}'\beta_2 + u_i'\alpha + e_i$$

where

$$\begin{aligned} E(u_ie_i) &= 0, \\ E(x_{1i}e_i) &= E(x_{1i}(\varepsilon_i - u_i'\alpha)) = 0, \text{ and} \\ E(x_{2i}e_i) &= E((\pi_1x_{1i} + \pi_2z_i + u_i)e_i) = E(\pi_2z_ie_i) = \pi_2E(z_i(\varepsilon_i - u_i'\alpha)) = 0 \end{aligned}$$

So OLS2' is satisfied and the OLS estimates of β_1, β_2, and α should be consistent. The only problem is that u_i is not observed. However, one can estimate u_i from the first stage regression, and add $\hat{u}_i$ to the structural equation, and estimate β by OLS, getting $\hat{\beta}_{CF}$ (CF stands for control function). This is exactly what the control function approach is doing.

Let $\hat{U}$ be the matrix with rows $\hat{u}_i'$. Then by the partitioned regression formula

(Frisch-Waugh theorem),

$$\hat{\beta}_{CF} = \left(X' M_{\hat{U}} X\right)^{-1} X' M_{\hat{U}} Y.$$

On the other hand, $\hat{U} = M_W X_2$ so that

$$M_{\hat{U}} = I - M_W X_2 \left(X_2' M_W X_2\right)^{-1} X_2' M_W.$$

Since X_1 is a part of W, $M_W X_1 = 0$ and

$$M_{\hat{U}} X_1 = X_1 = P_W X_1.$$

Further,

$$M_{\hat{U}} X_2 = X_2 - M_W X_2 \left(X_2' M_W X_2\right)^{-1} X_2' M_W X_2 = P_W X_2$$

Therefore,

$$M_{\hat{U}} X = P_W X,$$

which yields

$$\hat{\beta}_{CF} = \left(X' P_W X\right)^{-1} X' P_W Y = \hat{\beta}_{2SLS}.$$

Endogeneity test

If x_{2i} is not endogenous, OLS is efficient (BLUE) and 2SLS is not. How to test whether x_{2i} is exogenous or not?

$$H_0 : E\left(x_{2i}\varepsilon_i\right) = 0 \text{ against } H_1 : E\left(x_{2i}\varepsilon_i\right) \neq 0$$

Recall the control function regression

$$\begin{aligned} y_i &= x_{1i}'\beta_1 + x_{2i}'\beta_2 + u_i'\alpha + e_i, \text{ where} \\ \alpha &= \left[E\left(u_i u_i'\right)\right]^{-1} E\left(u_i \varepsilon_i\right) \text{ (the coefficient of BLP for } \varepsilon_i \text{ given } u_i) \end{aligned}$$

We have $E\left(x_{2i}\varepsilon_i\right) \neq 0$ if and only if $E\left(u_i\varepsilon_i\right) \neq 0$. Therefore, H_0 can be reformulated as

$$H_0 : \alpha = 0 \text{ against } H_1 : \alpha \neq 0.$$

Thus, a natural test statistic will be the Wald statistic for testing linear restrictions $\alpha = 0$ in the control function regression, where u_i is replaced by its estimate $\hat{u}_i$.

It turns out that the asymptotic distribution of such a test statistic under the null hypothesis is unaffected by such a replacement, and remains $\chi^2(k_2)$, where k_2 is the dimensionality of α (the same as the dimensionality of x_{2i}). This follows from a general result on the asymptotic distribution of the OLS estimates of the coefficients of regressions with so-called generated regressors (the ones consistently estimating the true regressors). For details, see H.12-26,12-27.

In STATA, the Wald statistic is reported when you type `estat endogenous` after `ivregress.` If you've used heteroskedasticity robust covariance estimation option when running `ivregress,` then the Wald statistic is reported under the name "Robust regression F", if you've used the default homoskedasticity option in `ivregress`, then the statistic is reported under the name "Wu-Hausman F".

Overidentification tests.

When $l > k$ (there are more instruments than endogenous regressors) we can test the hypothesis that the instruments are exogenous, that is

$$H_0 : E(w_i \varepsilon_i) = 0.$$

Let us assume the homoskedasticity, so that $E(\varepsilon_i^2 | w_i) = \sigma^2$. Consider a reduced form regression

$$\varepsilon_i = w_i' \alpha + e_i,$$

where

$$\alpha = (E(w_i w_i'))^{-1} E(w_i \varepsilon_i).$$

We see that $E(w_i \varepsilon_i) \neq 0$ if and only if $\alpha \neq 0$. Hence, our null and alternative hypotheses are

$$H_0 : \alpha = 0 \text{ against } H_1 : \alpha \neq 0$$

We cannot regress ε_i on w_i because we do not observe ε_i. However, we can try to replace ε_i by $\hat{\varepsilon}_i$, the residuals from the 2SLS. Sargan (1958) proposed to use nR^2 from this regression as the test statistic for testing H_0 against H_1

$$S = nR^2 = n \frac{SSE}{SST} = n \frac{\hat{\varepsilon}' W (W'W)^{-1} W' \hat{\varepsilon}}{\hat{\varepsilon}' \hat{\varepsilon}}.$$

Let us derive the asymptotic distribution of S. Note that S is invariant with respect to transformations $W \mapsto W \times A$, where A is any invertible matrix. Therefore, without loss of generality, we will assume that W is rotated and scaled so that

$E\left(w_i w_i'\right) = I_l$. As $n \to \infty$, we have

$$\frac{1}{\sqrt{n}} W'\varepsilon = \frac{1}{\sqrt{n}} \sum\nolimits_{i=1}^{n} w_i \varepsilon_i \xrightarrow{d} N\left(0, Var\left(w_i \varepsilon_i\right)\right) = N\left(0, \sigma^2 I_l\right) = \sigma N\left(0, I_l\right),$$

$$\frac{1}{n} W'W \xrightarrow{p} I_l,$$

and $\frac{1}{n} W'X$ converges in probability to some full column rank matrix Q. On the other hand, we have

$$\begin{aligned}
\frac{1}{\sqrt{n}} W'\hat{\varepsilon} &= \frac{1}{\sqrt{n}} W'\left(Y - X\hat{\beta}_{2SLS}\right) = \frac{1}{\sqrt{n}} W'\left(Y - X\left(X'P_W X\right)^{-1} X'P_W Y\right) \\
&= \frac{1}{\sqrt{n}} W'\left(\varepsilon - X\left(X'P_W X\right)^{-1} X'P_W \varepsilon\right) \\
&= \left(I - W'X\left(X'P_W X\right)^{-1} X'W\left(W'W\right)^{-1}\right) \frac{1}{\sqrt{n}} W'\varepsilon.
\end{aligned}$$

Combining these results yield

$$\frac{1}{\sqrt{n}} W'\hat{\varepsilon} \xrightarrow{d} \left(I - Q\left(Q'Q\right)^{-1} Q'\right) \sigma N\left(0, I_l\right)$$

and

$$\begin{aligned}
\hat{\varepsilon}'W\left(W'W\right)^{-1} W'\hat{\varepsilon} &= \frac{1}{\sqrt{n}} \hat{\varepsilon}'W \left(\frac{1}{n} W'W\right)^{-1} \frac{1}{\sqrt{n}} W'\hat{\varepsilon} \\
&\xrightarrow{d} \sigma^2 N'\left(I - Q\left(Q'Q\right)^{-1} Q'\right) N,
\end{aligned}$$

where $N \sim N\left(0, I_l\right)$. Similar to the derivation of the $\chi^2\left(n-k\right)$ distribution of $\hat{\sigma}^2$ in normal linear regression that we did in Session 3, we can show that

$$N'\left(I - Q\left(Q'Q\right)^{-1} Q'\right) N \sim \chi^2\left(l-k\right).$$

Hence,

$$\hat{\varepsilon}'W\left(W'W\right)^{-1} W'\hat{\varepsilon} \xrightarrow{d} \sigma^2 \chi^2\left(l-k\right)$$

Finally, $\hat{\varepsilon}'\hat{\varepsilon}/n \xrightarrow{p} \sigma^2$. Therefore,

$$S = n \frac{\hat{\varepsilon}'W\left(W'W\right)^{-1} W'\hat{\varepsilon}}{\hat{\varepsilon}'\hat{\varepsilon}} \xrightarrow{d} \chi^2\left(l-k\right).$$

We reject the null of the instrument exogeneity when S is larger than a critical value of $\chi^2(l-k)$.

The test cannot be performed in the just-identified situation (when $l=k$). In that case, $W'X$ has full rank (both rows and columns) and thus invertible. Therefore,

$$\begin{aligned}
\frac{1}{\sqrt{n}}W'\hat{\varepsilon} &= \left(I - W'X\left(X'P_WX\right)^{-1}X'W\left(W'W\right)^{-1}\right)\frac{1}{\sqrt{n}}W'\varepsilon \\
&= \left(I - W'X\left(X'W\left(W'W\right)^{-1}W'X\right)^{-1}X'W\left(W'W\right)^{-1}\right)\frac{1}{\sqrt{n}}W'\varepsilon \\
&= \left(I - W'X\left(W'X\right)^{-1}\left(\left(W'W\right)^{-1}\right)^{-1}\left(X'W\right)^{-1}X'W\left(W'W\right)^{-1}\right)\frac{1}{\sqrt{n}}W'\varepsilon \\
&= \left(I-I\right)\frac{1}{\sqrt{n}}W'\varepsilon = 0.
\end{aligned}$$

Irrelevant instruments

Some or all instruments are irrelevant if $E\left(w_i x_i'\right)$ is not of full column rank, and IV1 assumption is violated. In this case, parameter β is not identified. For a simple example, consider a model

$$\begin{aligned}
y_i &= x_i\beta + \varepsilon_i \\
x_i &= w_i\gamma + e_i
\end{aligned}$$

with one-dimensional endogenous variable x_i and instrument w_i satisfying $E\left(w_i\varepsilon_i\right) = 0$ (exogeneity), but fails the relevance assumption, so that $\gamma = 0$ and hence $E\left(w_i x_i\right) = 0$ too. The system of equations (moment conditions)

$$\begin{cases} E\left(w_i\varepsilon_i\right) = 0 \\ E\left(w_i x_i\right) = 0 \end{cases}$$

is equivalent to

$$\begin{cases} E\left(w_i\left(y_i - x_i\beta\right)\right) = 0 \\ E\left(w_i x_i\right) = 0 \end{cases} \Longleftrightarrow \begin{cases} E\left(w_i y_i\right) = 0 \\ E\left(w_i x_i\right) = 0 \end{cases}$$

which tells us nothing about β. Parameter β is not identified

Nevertheless, we can compute the IV estimator

$$\hat{\beta}_{IV} = \left(W'X\right)^{-1}W'Y$$

because, in any finite sample, it is unlikely that $W'X$ will be exactly zero. Let us compare such $\hat{\beta}_{IV}$ with $\hat{\beta}_{OLS}$. For simplicity, assume homoskedasticity, and suppose that

$$E(e_i|w_i) = E(\varepsilon_i|w_i) = 0,$$

$$Var(e_i|w_i) = Var(\varepsilon_i|w_i) = 1,\ Cov(e_i, \varepsilon_i|w_i) = \rho \neq 0$$

$$Ew_i = 0 \text{ and } Ew_i^2 = 1$$

By the CLT, we have

$$\frac{1}{\sqrt{n}}\begin{pmatrix} w_i\varepsilon_i \\ w_i e_i \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} \sim N\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

Note that $\xi_0 = \xi_1 - \rho\xi_2$ is a normal random variable, independent from ξ_2.

We have

$$\hat{\beta}_{OLS} - \beta = \frac{\frac{1}{n}\sum x_i\varepsilon_i}{\frac{1}{n}\sum x_i^2} = \frac{\frac{1}{n}\sum e_i\varepsilon_i}{\frac{1}{n}\sum e_i^2} \xrightarrow{p} \rho \neq 0$$

so that the endogeneity of x_i leads to the inconsistency of $\hat{\beta}_{OLS}$. Under the identification failure that we assume is taking place,

$$\hat{\beta}_{IV} - \beta = \frac{\frac{1}{\sqrt{n}}\sum w_i\varepsilon_i}{\frac{1}{\sqrt{n}}\sum w_i x_i} = \frac{\frac{1}{\sqrt{n}}\sum w_i\varepsilon_i}{\frac{1}{\sqrt{n}}\sum w_i e_i} \xrightarrow{d} \frac{\xi_1}{\xi_2} = \rho + \frac{\xi_0}{\xi_2}$$

We see that:

(1) $\hat{\beta}_{IV}$ does not converge in probability to a limit. Instead, it converges in distribution to a random variable. In particular, $\hat{\beta}_{IV}$ is inconsistent.

(2) Since the ratio ξ_0/ξ_2 is distributed symmetrically around zero, the median of the limiting distribution of $\hat{\beta}_{IV}$ is $\beta + \rho$. Hence, $\hat{\beta}_{IV}$ does not have correct centering. In fact, it is centered as the OLS.

(3) The ratio of two independent normals ξ_0/ξ_2 has the Cauchy distribution. Cauchy distribution has such a fat tails that even the first moment (mean) does not exist. We would expect a very wild fluctuations of $\hat{\beta}_{IV}$ from sample to sample.

To summarize, $\hat{\beta}_{IV}$ is inconsistent, median-biased, and asymptotically non-normal with fat tails. What about the t-statistic?

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n}\sum\left(y_i - x_i\hat{\beta}_{IV}\right)^2 = \frac{1}{n}\sum\left(\varepsilon_i - x_i\left(\hat{\beta}_{IV}-\beta\right)\right)^2 \\
&= \frac{1}{n}\sum\left(\varepsilon_i - e_i\left(\hat{\beta}_{IV}-\beta\right)\right)^2 \\
&= \frac{1}{n}\sum\varepsilon_i^2 - 2\frac{1}{n}\sum\varepsilon_i e_i\left(\hat{\beta}_{IV}-\beta\right) + \frac{1}{n}\sum e_i^2\left(\hat{\beta}_{IV}-\beta\right)^2 \\
&\xrightarrow{d} 1 - 2\rho\left(\rho + \frac{\xi_0}{\xi_2}\right) + \left(\rho + \frac{\xi_0}{\xi_2}\right)^2 \\
&= 1 - \rho^2 + \left(\frac{\xi_0}{\xi_2}\right)^2
\end{aligned}$$

Therefore, the t-statistic has the asymptotic distribution

$$t = \frac{\hat{\beta}_{IV}-\beta}{\sqrt{\hat{\sigma}^2\frac{1}{n}\sum w_i^2}/\frac{1}{\sqrt{n}}\left|\sum w_i x_i\right|} = \frac{\hat{\beta}_{IV}-\beta}{\sqrt{\hat{\sigma}^2\frac{1}{n}\sum w_i^2}/\frac{1}{\sqrt{n}}\left|\sum w_i e_i\right|} \xrightarrow{d} \frac{\rho + \frac{\xi_0}{\xi_2}}{\sqrt{1-\rho^2+\left(\frac{\xi_0}{\xi_2}\right)^2}/\left|\xi_2\right|}$$

This distribution is non-normal. Note that, when $\rho \to 1$,

$$Var(\xi_0) = Var\left(\xi_1 - \rho\xi_2\right) = 1 - 2\rho^2 + \rho^2 \to 0$$

so that

$$\xi_0 \xrightarrow{p} 0 \text{ and } 1 - \rho^2 + \left(\frac{\xi_0}{\xi_2}\right)^2 \xrightarrow{p} 0$$

This implies that

$$|t| \xrightarrow{p} \infty.$$

We may decide that our estimates are very statistically significant, when, in fact, they are useless.

References

[1] Angrist, J.D., and Krueger, A.B. (1991) "Does Compulsory Schooling Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics* 106, 976–1014.

[2] Card, D. (1995) "Using geographic variation in college proximity to estimate the return to schooling," in *Aspects of Labor Market Behavior: Essays in Honor*

of John Vanderkamp, L.N. Christofides, E.K. Grant. and R. Swidinsky, eds. Toronto: University of Toronto Press.

[3] Graddy, K. (1995) "Testing for imperfect competition at the Fulton fish market," *RAND Journal of Economics* 26, 75-92.

[4] Sargan (1958) "The estimation of economic relationships using instrumental variables," *Econometrica* 26, 393–415.