Econometrics, Lecture 16, Instrumental Variables.

Ron Smith
EMS, Birkbeck, University of London

Autumn 2020

Last time

- We began the examination of the failure of exogeneity with an examination of simultaneous equations systems.
- We looked at the problem of identification in the context of a demand supply system.
- Showed it has the same form of observational equivalence problem as the examples with IRFs and cointegrating vectors.
- Similar issue with reparameterising a linear regression, e.g. ARDL to ECM: $y = X\beta + u$ is observationally equivalent to $y = (XP)(P^{-1}\beta) + u, \;\; y = X^*\beta^* + u$.
- Showed OLS inconsistent, the consistent estimator was two stage least squares, which is the same as instrumental variables.

Instruments

- Back to the LRM
$$y = X\beta + u$$
with X a $T \times k$ matrix, but not exogenous, so $E(X'u) \neq 0$. The OLS estimates will be biased and inconsistent.
- Suppose there is a $T \times i$, matrix of 'Instruments', W, where $i \geq k$, correlated with X so that $E(W'X) \neq 0$ but not correlated with the disturbances so that $E(W'u) = 0$.
- W will include the elements of X that are exogenous (including the column of ones for the constant), but we need at least one instrument for each endogenous X.
- If $i = k$, the model is just-identified, if $i > k$ it is said to be over-identified. The condition $i \geq k$ is the same order condition we had before.
- The rank condition is $E(W'X) \neq 0$ to make $(W'X)$ full rank and $(W'X)^{-1}$ exist. Notice $(W'X)' = (X'W)$.

Instrumental Variable, IV, estimator

- For the just identified case, IV is a method of moments estimator, using sample equivalent of $E(W'u) = 0$:

$$\begin{aligned} W'\widetilde{u} &= 0 \\ W'(y - X\widetilde{\beta}) &= 0 \\ W'y &= W'X\widetilde{\beta} \\ \left(W'X\right)^{-1} W'y &= \widetilde{\beta}. \end{aligned}$$

- Same form as the ILS estimator in the Keynesian consumption function: $\widetilde{\beta} = \sum i_t c_t / \sum i_t y_t$.
- $$\begin{aligned} \widetilde{\beta} &= \left(W'X\right)^{-1} W'(X\beta + u) \\ &= \left(W'X\right)^{-1} W'X\beta + \left(W'X\right)^{-1} W'u \\ &= \beta + \left(W'X\right)^{-1} W'u \end{aligned}$$

so is consistent, since $E(W'u) = 0$.

IV variance covariance matrix

- Asymptotically

$$
\begin{aligned}
V(\widetilde{\beta}) &= E(\widetilde{\beta}-\beta)(\widetilde{\beta}-\beta)' \\
&= E\left[\left(W'X\right)^{-1} W'u\, u'W\left(W'X\right)^{-1}\right] \\
&= \sigma^2 \left(W'X\right)^{-1} W'W \left(W'X\right)^{-1}
\end{aligned}
$$

- Its efficiency increases (standard errors reduce) with the correlation between W and X.
- With more instruments than required, we want to take the linear combination with maximum correlation to maximise efficiency, that is the GIVE/2SLS estimator.
- Some use IV for all cases, some distinguish IV for just identified and GIVE for over-identified.

Generalised Instrumental Variable estimator

- GIVE is the same as 2SLS. First regress each of the X on the W;
$$X = WB + V$$
to give the $i \times k$ matrix of coefficients $\widehat{B} = (W'W)^{-1}W'X$,
- Then get the predicted values of X as:
$\widehat{X} = W\widehat{B} = W(W'W)^{-1}W'X = P_W X$, with
$P_w = W(W'W)^{-1}W'$ being an idempotent projection matrix.
- Substituting $X = \widehat{X} + \widehat{V}$ into the original regression we get:
$$y = (\widehat{X} + \widehat{V})\beta + u = \widehat{X}\beta + (\widehat{V}\beta + u).$$
- $\widehat{X}$ is uncorrelated with u since it is only a function of the W which are uncorrelated with u, and is uncorrelated with $\widehat{V}$ by construction. Therefore it satisfies our exogeneity conditions.

GIVE Variance

- The GIVE estimator is

$$
\begin{aligned}
\widetilde{\beta} &= (\widehat{X}'\widehat{X})^{-1}\widehat{X}'y \\
&= (X'W(W'W)^{-1}W'X)^{-1}X'W(W'W)^{-1}W'y \\
&= (X'P_wX)^{-1}X'P_wy
\end{aligned}
$$

- **Show that its variance covariance matrix is** $\sigma^2(X'P_wX)^{-1}$
- Estimated residuals use X not $\widehat{X}$:

$$
s_{IV}^2 = (y - X\widetilde{\beta})'(y - X\widetilde{\beta})/(T-k)
$$

- GIVE chooses $\widetilde{\beta}$ to make $X'P_w\widetilde{u} = 0$. Minimises $\widetilde{u}'P_w\widetilde{u}$, the IV minimand, which is zero when model is just identified.
- **Show that IV minimand is zero for the just identified case** where $\widetilde{\beta} = (W'X)^{-1}W'y$ by multiplying out

$$
(y - X\widetilde{\beta})'W(W'W)^{-1}W'(y - X\widetilde{\beta})
$$

Example model

- Suppose that we have two potentially endogenous variables, x_{1t} and x_{2t} and 4 exogenous variables w_{it}. The structural form of the model is

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 w_{1t} + u_t$$

  with w_{2t}, w_{3t}, w_{4t} as excluded exogenous variables and potential instruments.
- $X = [1 \ x_{1t} \ x_{2t} \ w_{1t}]$, and $W = [1 \ w_{1t} \ w_{2t} \ w_{3t} \ w_{4t}]$ so $k = 4$, $i = 5$ and the degree of overidentification is one.

Reduced form

- The two 'reduced form' regressions of $X = WB + V$ are

$$
\begin{aligned}
x_{1t} &= b_{10} + b_{11}w_{1t} + b_{12}w_{2t} + b_{13}w_{3t} + b_{14}w_{4t} + v_{1t} \\
x_{2t} &= b_{20} + b_{21}w_{1t} + b_{22}w_{2t} + b_{23}w_{3t} + b_{24}w_{4t} + v_{2t}
\end{aligned}
$$

  and they are used to get estimates of $\widehat{x}_{1t}, \widehat{x}_{2t}, \widehat{v}_{1t}, \widehat{v}_{2t}$.
- Also use them to test for weak instruments. F stat should be greater than about 10.
- If the instruments are weak, do not explain x_{it} very well, GIVE estimates will be biased and have large variance even in large samples.

GIVE

- GIVE/2SLS is OLS on the structural equation using fitted values of endogenous variables:

$$y_t = \beta_0 + \beta_1 \widehat{x}_{1t} + \beta_2 \widehat{x}_{2t} + \beta_3 w_{1t} + e_t$$

  where $e_t = u_t + \beta_1 \widehat{v}_{1t} + \beta_2 \widehat{v}_{2t}$.
- The residuals are estimated as (not using fitted values of x_{it}):

$$\widetilde{u}_t = y_t - (\widetilde{\beta}_0 + \widetilde{\beta}_1 x_{1t} + \widetilde{\beta}_2 x_{2t} + \widetilde{\beta}_3 w_{1t})$$

- In programs choose IV/2SLS list the model variables, then list the instruments. Include constant and right hand side exogenous variables among the instruments.

Sargan (Hansen) test

- If the over-identifying restrictions which exclude w_{2t}, w_{3t}, w_{4t} from the model are valid, the GIVE residuals should be uncorrelated with the instruments.
- Tested by a Sargan, Bassman, Hansen J test. Regress the GIVE residuals on all the instruments:

$$\widetilde{u}_t = c_0 + c_1 w_{1t} + c_2 w_{2t} + c_3 w_{3t} + c_4 w_{4t} + \varepsilon_t$$

  testing $c_1 = c_2 = c_3 = c_4 = 0$, this is distributed $\chi^2(i-k)$, i.e. with degrees of freedom equal to the number of overidentifying restrictions, one in this case.
- This can also be expressed as the ratio of the IV minimand to the GIVE variance. When the model is just identified, the IV minimand is zero, so the test is not defined.

Wu-Hausman test

.

- This tests whether the x_{it} are in fact exogenous. Include the residuals from the reduced form regressions, $\widehat{v}_{1t}, \widehat{v}_{2t}$, in the model, run by OLS:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 w_{1t} + \delta_1 \widehat{v}_{1t} + \delta_2 \widehat{v}_{2t} + u_t \quad (1)$$

- The null that the x_{it} are exogenous is $H_0 : \delta_1 = \delta_2 = 0$. Rejection indicates that one or both of the x_{it} are endogenous and GIVE should be used.
- This tests whether there is a significant difference between the OLS (where $\delta_1 = \delta_2 = 0$) and GIVE where since $x_{1t} - \widehat{v}_{1t} = \widehat{x}_{1t}$, $x_{2t} - \widehat{v}_{2t} = \widehat{x}_{2t}$, then $\delta_1 = -\beta_1$ and $\delta_2 = -\beta_2$
- (1) is a convenient way to implement GIVE in non-linear models, known as the control function approach.

Hausman test

- The general Hausman test compares two estimators with $k \times 1$ vectors of estimates $\widehat{\delta}$ and $\widetilde{\delta}$
- $\widehat{\delta}$ is efficient under the null, but inconsistent under the alternative,
- $\widetilde{\delta}$ is less efficient under the null but consistent under both the null and alternative.
- The Hausman test statistic is the quadratic form

$$\left(\widehat{\delta}-\widetilde{\delta}\right)'\left[V(\widetilde{\delta})-V(\widehat{\delta})\right]^{-1}\left(\widehat{\delta}-\widetilde{\delta}\right) \sim \chi^2(k)$$

- In the Wu-Hausman case the null is that x_{it} are exogenous, $\widehat{\delta}$ is the OLS estimator and $\widetilde{\delta}$ the GIVE estimate, but degrees of freedom not k but number of potentially endogenous regressors.

Identification through the covariance matrix

- In $\mathbf{By}_t = \mathbf{\Gamma x}_t + \mathbf{u}_t$; $\mathbf{E}(\mathbf{u}_t \mathbf{u}_t') = \mathbf{\Omega}$, we can get identification not just by restrictions on $\mathbf{B}$ and $\mathbf{\Gamma}$ but also on $\mathbf{\Omega}$.
- Assuming $\mathbf{\Omega}$ is diagonal, shocks are uncorrelated, gives $m(m-1)/2$ restrictions, off diagonal elements are zero. Assuming $\mathbf{B}$ is triangular, elements above diagonal are zero, gives $m(m-1)/2$ restrictions. Plus m from normalisation gives m^2 in total.
- Such a "recursive" system is just identified and can be estimated by OLS, as none of the RHS variables are correlated with the errors:

$$\begin{aligned} y_{1t} &= \gamma_1 x_t + u_{1t} \\ y_{2t} &= \beta_{21} y_{1t} + \gamma_2 x_t + u_{2t} \\ y_{3t} &= \beta_{31} y_{1t} + \beta_{32} y_{2t} + \gamma_3 x_t + u_{3t} \end{aligned}$$

- u_{2t} is not correlated with y_{1t} because there is no direct link, y_{2t} does not influence y_{1t}, and no indirect link, u_{2t} is not correlated with u_{1t}.

Recursive system

- The recursive system in the form $\mathbf{B}\mathbf{y}_t = \boldsymbol{\Gamma}\mathbf{x}_t + \mathbf{u}_t$

$$\begin{bmatrix} 1 & 0 & 0 \\ -\beta_{21} & 1 & 0 \\ -\beta_{31} & -\beta_{32} & 1 \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix} x_t + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix}.$$

- $$\boldsymbol{\Omega} = \begin{bmatrix} \omega_{11} & 0 & 0 \\ 0 & \omega_{22} & 0 \\ 0 & 0 & \omega_{33} \end{bmatrix}$$

- There is a "causal ordering".

VAR and SEM

- A VAR is the reduced form of a structural system where $\mathbf{x}_t = \mathbf{y}_{t-1}$:

$$\begin{aligned}
\mathbf{B}\mathbf{y}_t &= \boldsymbol{\Gamma}\mathbf{y}_{t-1} + \mathbf{u}_t; \quad \mathbf{E}(\mathbf{u}_t\mathbf{u}_t') = \boldsymbol{\Omega} \\
\mathbf{y}_t &= \mathbf{B}^{-1}\boldsymbol{\Gamma}\mathbf{y}_{t-1} + \mathbf{B}^{-1}\mathbf{u}_t \\
\mathbf{y}_t &= \boldsymbol{\Pi}\mathbf{y}_{t-1} + \varepsilon_t; \ \mathbf{E}(\varepsilon_\mathbf{t}\varepsilon'_\mathbf{t}) = \boldsymbol{\Sigma} = \mathbf{B}^{-1}\boldsymbol{\Omega}\mathbf{B}^{-1\prime}
\end{aligned}$$

- Impulse response functions, IRFs, are used to analyse the interaction between the variables in a VAR. They measure the time profile of the effect of a one standard error shock on the expected future values of the variables in the system.
- Used the moving average representation

$$y_t = \sum_{i=0}^{\infty} \Phi_i \varepsilon_{t-i}$$

- We looked at them in lecture 13. Next slide is repeated from there.

Generalised and orthogonalised IRFs

- The errors ε_{it} are not orthogonal (uncorrelated) since $E(\varepsilon_t\varepsilon_t') = \Sigma$, not diagonal. Generalised IRFs let the other errors respond to the shock according to the covariance matrix. Some think structural shocks should be orthogonal.
- Orthogonalised IRFs use a causal ordering to make the shock uncorrelated. A Choleski decomposition uses $\Sigma = PP'$, where P is lower triangular and

$$y_t = \sum_{i=0}^{\infty}(\Phi_i P)(P^{-1}\varepsilon_{t-i}) = \sum_{i=0}^{\infty}\Psi_i\eta_i$$

- $E(\eta_i\eta_i') = E\left[(P^{-1}\varepsilon_{t-i})(P^{-1}\varepsilon_{t-i})'\right] = P^{-1}\Sigma P^{-1\prime} = I$. So the η_i are uncorrelated. But P is not unique, it depends on the ordering.
- Orthogonalised IRFs are equivalent to recursive causal orderings. There are other forms of structural VARs.

Identification

- Having an "identification strategy" is central to much empirical economics. It is needed to separate causation from correlation. But identification assumptions are not testable, so source of debate.
- IV and SEM apply to cross-section not just time series. Questions like what is the effect of class size on pupil performance, what is the effect of education on earnings. Angrist & Pischke good on microeconometric issues.
- Big problem is finding valid and relevant instruments. Natural experiments which allow Regression Discontinuity Design are one source.
- Randomised Control Trials are alternatives used in development economics.

Next time

- Up to now we have been using a "frequentist" approach to estimation and inference (testing).
- There is an alternative approach to statistics, the Bayesian approach.
- This treats parameters not as fixed numbers but as random variables and derives a posterior distribution for a parameter from the product of the likelihood and the prior distribution.
- Will introduce it next time.