Treatment Effect Heterogeneity and Weak Instruments

Michal Kolesár

ECO539B, Spring 2025

March 25, 2025

Textbook model

Treatment effect heterogeneity

Uses of IV

Can use instrumental variables (IV) regression to solve a number of issues:

1. Errors-in-variables (e.g., Zellner 1970);
2. Deal with omitted variable bias: we'd like to recover β in the projection $E[Y_i \mid D_i, A_i] = D_i\beta + A_i'\gamma$, but A_i is not observed (e.g., Chamberlain 2007);
3. Estimate a simultaneous equations model, such as a demand-and-supply system (e.g., Angrist, Graddy, and Imbens 2000); or
4. Estimate treatment effects when the unconfoundedness assumption fails.

Focus on last goal, and consider:

1. Implications of treatment effect heterogeneity for estimation and inference; and
2. Weak instrument issues

Setup

- Focus on i.i.d. sampling of $\mathcal{D}_i = (Y_i, D_i, Z_i, W_i)$, with $\dim(Z_i) = k$, $\dim(W_i) = \ell$. Let $X_i = (Z_i', W_i')'$.
- Reduced form and the first stage projections

$$Y_i = Z_i'\delta + W_i'\psi_Y + u_{Y,i}, \tag{1}$$

$$D_i = Z_i'\pi + W_i'\psi_D + u_{D,i}. \tag{2}$$

Normality assumption

$$\sqrt{n}\left(\begin{pmatrix}\hat{\delta}\\ \hat{\pi}\end{pmatrix}-\begin{pmatrix}\delta\\ \pi\end{pmatrix}\right) \Rightarrow \mathcal{N}\left(0, u_i \otimes Q^{-1}\tilde{Z}_i\right), \quad Q = E[\tilde{Z}_i\tilde{Z}_i'], \tag{3}$$

with asymptotic variance consistently estimable

- Fails if
  - k is large relative to n (next lecture)
  - Leverages are high (Young 2022, e.g.). Analogous to ordinary least squares (OLS) diagnostics—these are just OLS regressions!

Valid IV with constant treatment

$\delta = \beta\pi$, with $\beta = E[Y(1) - Y(0)]$.

- Can fail if any of following doesn't hold:

Random assignment Z mean-independent of the potential outcomes given W

Exclusion restriction the potential outcomes $Y_i(d, z)$ in fact only depend on d

Linearity $E[Z \mid W]$ is linear in W (or else $E[Y(0) \mid W]$ linear)

Constant treatment effects $Y(d) = Y(0) + d\beta$.

Estimation: TSLS

- $\pi\beta = \delta$ equivalent to moment condition $E[X_i \epsilon_i] = 0$.
- When $\epsilon_i = u_{Yi} - u_{Di}\beta$ homoskedastic, optimal generalized method of moments (GMM) weighting matrix $\propto E[X_i X_i]^{-1}$, and solving it yields two-stage least squares (TSLS):

$$\hat{\beta}_{\text{TSLS}} = \frac{D' H_{\ddot{Z}} Y}{D' H_{\ddot{Z}} D} = \frac{\hat{\pi} \ddot{Z}' \ddot{Z} \hat{\delta}}{\hat{\pi} \ddot{Z}' \ddot{Z} \hat{\pi}}, \qquad \hat{\gamma}_{\text{TSLS}} = (W'W)^{-1} W'(Y - D\hat{\beta}_{\text{TSLS}}).$$

  If $k = 1$, weighting doesn't matter, $\hat{\beta}_{\text{TSLS}} = \hat{\delta}/\hat{\pi}$.

- Standard GMM (or delta method) arguments deliver

$$\sqrt{n}(\hat{\beta}_{\text{TSLS}} - \beta) \Rightarrow \mathcal{N}(0, \mathcal{V}_1), \qquad \mathcal{V}_1 = \frac{E[\sigma^2(X_i)(\tilde{Z}_i' \pi)^2]}{(\pi' Q \pi)^2}. \tag{4}$$

Estimation: LIML

- Anderson and Rubin (1949): assume $(\epsilon_i, u_{D,i})$ are homoskedastic and jointly normal conditional on X_i, and estimate β by maximum likelihood. This gives:

$$\hat{\beta}_{\text{LIML}} = \operatorname*{argmin}_{\beta} \frac{(1,-\beta)\hat{\Pi}'\ddot{Z}\ddot{Z}'\hat{\Pi}(1,-\beta)}{(1,-\beta)S(1,-\beta)'}, \qquad \hat{\Pi} = (\hat{\delta}, \hat{\pi})$$

- Equivalent to minimum distance estimator minimizing

$$\begin{pmatrix} \hat{\delta} - \pi\beta \\ \hat{\pi} - \pi \end{pmatrix}' W \begin{pmatrix} \hat{\delta} - \pi\beta \\ \hat{\pi} - \pi \end{pmatrix}.$$

  with W optimal weighting matrix under homoskedasticity (Goldberger and Olkin 1971)

- Minimum distance objective doesn't rely on normality or homoskedasticity $\implies$ LIML asymptotically normal and consistent—in fact first-order asymptotically equivalent to TSLS.

What can go wrong

1. One of our two assumptions fails
2. Delta method underlying asymptotic normality of TSLS fails
    - This happens if $\pi = 0$. By continuity, this implies that the delta method will work poorly if π is close to zero. This is a weak instrument problem.

Treatment effect heterogeneity

TO BE CONTINUED

Anderson, Theodore W., and Herman Rubin. 1949. "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations." *The Annals of Mathematical Statistics* 20, no. 1 (March): 46–63. https://doi.org/10.1214/aoms/1177730090.

Angrist, Joshua D., Kathryn Graddy, and Guido W. Imbens. 2000. "The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish." *Review of Economic Studies* 67, no. 3 (July): 499–527. https://doi.org/10.1111/1467-937X.00141.

Chamberlain, Gary. 2007. "Decision Theory Applied to an Instrumental Variables Model." *Econometrica* 75, no. 3 (May): 609–652. https://doi.org/10.1111/j.1468-0262.2007.00764.x.

Goldberger, Arthur S., and Ingram Olkin. 1971. "A Minimum-Distance Interpretation of Limited-Information Estimation." *Econometrica* 39, no. 3 (May): 635–639. https://doi.org/10.2307/1913273.

Young, Alwyn. 2022. "Consistency without Inference: Instrumental Variables in Practical Application." *European Economic Review* 147 (August): 104112. https://doi.org/10.1016/j.euroecorev.2022.104112.

Zellner, Arnold. 1970. "Estimation of Regression Relationships Containing Unobservable Independent Variables." *International Economic Review* 11, no. 3 (October): 441–454. https://doi.org/10.2307/2525323.