MANY INSTRUMENTS AND JUDGES

Michal Kolesár

ECO539B, Fall 2022

April 3, 2024

- Consider linear instrumental variables (IV) model when number of instruments K large. Arises in 2 ways:
  1. Interact low-dim instrument with controls (Angrist and Krueger 1991). Or use approach of Belloni et al. (2012).
  2. Use fixed effects (group indicators) as instruments: "judge" or "examiner" designs. Studies of effects of incarceration on economic outcomes (Kling 2006; Aizer and Doyle 2015), studies that exploit random assignment of judges to bankruptcy cases, criminal cases, or patent cases, doctors to shifts…

Key issues

1. two-stage least squares (TSLS) biased with many instruments. Alternatives?
2. standard errors
3. Interpretation of first-stage F statistic

Setup and Estimation

Inference

First stage F

Summary and illustration

- Same notation as in previous lecture, with reduced form and first stage

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

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

  The parameter of interest is $\beta := \frac{\delta' Q\pi}{\pi' Q\pi}$, where $Q = E[\tilde{Z}_i\tilde{Z}_i']$, and $\tilde{Z}_i = Z_i - E[Z_iW_i']E[W_iW_i']^{-1}W_i$.

- To measure overall strength of instruments, define measure of "effective sample size"

$$r_n = n\pi' Q\pi = nE[(\tilde{Z}_i'\pi)^2].$$

- $X_i = (Z_i', W_i')'$ collects the right-hand side (RHS) (or "exogenous") variables.

- For any matrix A, let $H_A = A(A'A)^{-1}A'$ denote hat matrix (also called a projection matrix), and let $\ddot{A} = A - H_W A$ denote residuals after projecting A onto the covariates.
    - Thus $\tilde{Z}_i$ is the population analog of $\ddot{Z}_i$

Judges design

- Individual i assigned judge Q_i, random conditional on i's geographic location G_i.
- Covariates and instruments both indicators: $W_{i\ell} = \mathbb{1}\{G_i = \ell\}$, and $Z_{ik} = \mathbb{1}\{Q_i = k\}$.
- There are L locations and $K + L$ judges; in each location we drop one judge to avoid collinearity—call this judge reference judge.
- $\psi_{D,\ell}$ is average sentencing rate of the reference judge in location ℓ, and π_k is sentencing rate of judge k relative to the reference judge in the same location.

- Monotonicity assumption requires that judges agree on the ranking of the defendants, they just disagree on the cutoff at which they start sentencing them.
- Can be problematic: Chan, Gentzkow, and Yu (2019) point out that decisions of physicians differ due to both skill and preferences.
- Similarly, Kleinberg et al. (2018) find that the increase in crime associated with judges who are more likely to release defendants on bail is about the same as if these more lenient judges randomly picked the extra defendants to release on bail.
- Failure of monotonicity only affects the interpretation of β; to keep statistical issues separate from identification issues, we put these issues aside here.

Consider asymptotics in which $K = \dim(Z_i)$ and $L = \dim(W_i)$ can to grow with sample size

Lemma

The TSLS estimator suffers from own observation bias towards ordinary least squares (OLS). In particular, suppose $E[u_i \mid X_i] = 0$, with $\Omega(X_i) = E[u_i u_i' \mid X_i]$ bounded, $E[\tilde{Z}_i \mid W_i] = 0$, and $L/n \to 0$.

Then consistency of TSLS is in general requires $K/r_n \to 0$. Under homoskedastic errors,

$$\hat{\beta}_{\mathrm{TSLS}} = \beta + \frac{(\Omega_{YD} - \Omega_{DD}\beta)K}{r_n + \Omega_{DD}K} + o_p(1) = (1-w)\beta + w\beta_{\mathrm{OLS}} + o_p(1), \quad w = \frac{K}{r_n/\Omega_{DD} + K},$$

where $\beta_{\mathrm{OLS}} = (\Omega_{YD} - \Omega_{DD}\beta)/\Omega_{DD} + \beta$ is the probability limit of OLS.

TSLS bias scales with K/r_n, rather than K/n!

- Bias arises because single constructed instrument $\hat{Z}_{\text{TSLS},i} = Z_i'\hat{\pi} + W_i'\hat{\psi}_D$ used by TSLS puts positive weight on own treatment status D_i
- Total net weight, holding overall instrument strength constant, scales linearly with K, so TSLS bias scales with the number of instruments.
- Solution 1 (Bekker 1994): use limited information maximum likelihood (LIML), or variants thereof. Only need $\sqrt{K}/r_n \to 0$ for consistency, substantially weaker requirement. But LIML-like estimators not robust to heterogeneous treatment effects.
- Solution 2: subtract estimate of bias. Easy to do under homoskedastic errors: leads to variants of the bias-corrected TSLS estimator that dates back to Nagar (1959). But consistency does depend on homoskedasticity.

- Bias caused by using D_i in constructing the single instrument $\hat{Z}_{\text{TSLS},i}$: obvious solution is to not use it!
- Option 1: construct predictors $\hat{\pi}_{-i}$ and $\hat{\psi}_{D,-i}$ of π and ψ_D based on a regression of D onto (Z, W), but with the ith observation removed, to construct a single instrument $\hat{Z}_{\text{JIVE1},i} = Z_i'\hat{\pi}_{-i} + W_i'\hat{\psi}_{D,-i}$
  - What is this in judges design?
- Then run an IV regression of Y_i onto D_i and W_i, using $\hat{Z}_{\text{JIVE1},i}$ as an instrument for D_i. Resulting estimator known as jackknife instrumental variables estimator (JIVE1):

$$\hat{\beta}_{\text{JIVE1}} = \frac{\hat{Z}_{\text{JIVE1}}'\ddot{Y}}{\hat{Z}_{\text{JIVE1}}'\ddot{D}}, \quad \hat{Z}_{\text{JIVE1}} = \ddot{H}D, \quad \ddot{H} = (I - \operatorname{diag}(H_X))^{-1}(H_X - \operatorname{diag}(H_X)),$$

- Don't need to run n first-stage regressions
- Problem: when we project out the covariates from $\hat{Z}_i$, we re-introduce the own observation bias
  - In judges design, adjust instrument by average sentencing rate in the location of i
- Generally need $L/r_n \to 0$ for consistency (see Evdokimov and Kolesár (2018)). Under homo, bias goes in opposite direction to TSLS bias:

$$\hat{\beta}_{\text{JIVE1}} = \beta - \frac{(\Omega_{YD} - \Omega_{DD}\beta)L}{r_n - \Omega_{DD}L} + o_p(1) = (1+\lambda)\beta - \lambda\beta_{OLS} + o_p(1), \;\; \lambda = \frac{L}{r_n/\Omega_{DD} - L}.$$

- Leads to option 2: *first* partial out the covariates, and then do the leave-one-out prediction. Called improved improved jackknife instrumental variables estimator (IJIVE1) estimator, proposed in Ackerberg and Devereux (2009).

$$\hat{\beta}_{\text{IJIVE1}} = \frac{\ddot{D}'\ddot{H}'\ddot{Y}}{\ddot{D}'\ddot{H}'\ddot{D}}, \qquad \ddot{H} = (I - \operatorname{diag}(H_{\ddot{Z}}))^{-1}(H_{\ddot{Z}} - \operatorname{diag}(H_{\ddot{Z}})).$$

- Option 3: to exclude own observation both when calculating the severity of the judge assigned to i, and also when calculating the average sentencing rate in the location of i.

Kolesár (2013) calls this estimator unbiased jackknife instrumental variables estimator (UJIVE).

$$\hat{D}_{\text{UJIVE},i} = \hat{Z}_{\text{JIVE1},i} - \frac{\sum_{j \neq i} D_j \mathbb{1}\{G_j = G_i\}}{\sum_{j \neq i} \mathbb{1}\{G_j = G_i\}}$$

That is, (i) run jackknife the regression of D onto (W, Z) to compute $\hat{Z}_{\text{UJIVE},i} = Z_i' \hat{\pi}_{-i} + W_i' \hat{\psi}_{D,-i}$. (ii) adjust $\hat{Z}_{\text{UJIVE},i}$ for covariates by jackknife regression of D onto W. Then use this as a single instrument, $\hat{\beta}_{\text{UJIVE}} = \hat{D}_{\text{UJIVE}}' Y / \hat{D}_{\text{UJIVE}}' D$.

- Both UJIVE and IJIVE1 consistent so long as:
  1. $\sqrt{K}/r_n \to 0$ (without this condition, no estimator can be consistent).
  2. There aren't too many covariates: IJIVE1 requires $LK/r_n n \to 0$, while consistency of UJIVE only requires $\sqrt{L}/r_n \to 0$.

- Define $\pi_\Delta = \delta - \pi\beta$, $u_{\Delta,i} = u_{Yi} - u_{Di}\beta$, $\gamma = E[W_i W_i']^{-1} E[W_i(Y_i - D_i\beta)] = E[W_i W_i']^{-1} E[W_i Z_i]\pi_\Delta + \psi_Y - \psi_D\beta$, and $\epsilon_i = \tilde{Z}_i'\pi_\Delta + u_{\Delta,i}$.
- Write the "structural" equation as

$$Y_i = D_i\beta + W_i'\gamma + \epsilon_i.$$

- saw in the previous lecture that the oracle estimator satisfied

$$\mathcal{V}_{1,n}^{-1/2}(\hat{\beta}^* - \beta) \Rightarrow \mathcal{N}(0,1), \qquad \mathcal{V}_{1,n} = \frac{E[\epsilon_i^2(\tilde{Z}_i'\pi)^2]}{r_n E[(\tilde{Z}_i'\pi)^2]}.$$

  Variance $\mathcal{V}_{1,n}$ is estimated by the conventional robust standard errors, such as those in Stata. Also correct standard errors for TSLS, UJIVE, or IJIVE1 if (i) there is no treatment effect heterogeneity, and (ii) $K/r_n \to 0$ for UJIVE, or IJIVE1, and $K^2/r_n \to 0$ for TSLS.

- If (i) fails, then, as discussed last time, don't achieve oracle variance, but instead the correct asymptotic variance is given by

$$\mathcal{V}_{2,n} = \frac{E[((\tilde{Z}_i'\pi_\Delta)u_{D,i} + \epsilon_i(\tilde{Z}_i'\pi))^2]}{r_n E[(\tilde{Z}_i'\pi)^2]}.$$

- What if (ii) fails? What if $K/r_n \to 0$, but $K^2/r_n \not\to 0$ and we use TSLS?

- If $K/r_n \not\to 0$, and use UJIVE or IJIVE1, we need to account for the presence of many instruments in the asymptotic variance formula (Evdokimov and Kolesár 2018, Theorem 5.4)

$$(\mathcal{V}_{2,n} + \mathcal{V}_{MI,n})^{-1/2}(\hat{\beta}_{\text{UJIVE}} - \beta) \Rightarrow \mathcal{N}(0,1),$$

$$\mathcal{V}_{MI,n} = \frac{1}{r_n^2}\sum_{i \neq j}(H^2_{\tilde{Z},ij}u^2_{\Delta,i}u^2_{2,j} + H^2_{\tilde{Z},ij}u_{\Delta,i}u_{2,i} \cdot u_{\Delta,j}u_{2,j}), \quad (3)$$

- Under homoskedasticity, $\Omega(X_i) = \Omega = E[u_i u_i']$, additional many instrument term simplifies to

$$\mathcal{V}_{MI,n} = \frac{K}{r_n^2}(E[(u_{i1} - u_{2i}\beta)^2] \cdot E[u^2_{2i}] + E[(u_{i1} - u_{2i}\beta)u_{2i}]^2)(1 + o_p(1)).$$

Setup and Estimation

Inference

First stage F

Summary and illustration

- Some papers calculate the instrument $\hat{Z}_i$ manually, often as a leave-one-out prediction, effectively computing JIVE1 by hand. But this does not mean that $K = 1$! That will overstate actual instrument strength.
- When using the (correctly computed) first-stage F for diagnostics, remember that the $F > 10$ rule of thumb tests hypothesis that TSLS bias, relative to the bias of OLS exceeds 0.1.
- But small F statistic not necessarily a concern when UJIVE or IJIVE1 are used. Under homoskedasticity,

$$E[F] = \frac{E[\hat{\pi}_2 \tilde{Z}' \tilde{Z} \hat{\pi}_2]}{KE[u_{2i}^2]} = \frac{\pi_2 E[\tilde{Z}' \tilde{Z}] \pi_2}{KE[u_{2i}^2]} + 1 \approx \frac{r_n}{KE[u_{2i}^2]} + 1.$$

- If $r_n/K \to 0$, TSLS will be inconsistent; but jackknife estimators will remain consistent so long as $r_n/\sqrt{K} \to \infty$

Setup and Estimation

Inference

First stage F

Summary and illustration

- If K non-negligible relative to effective sample size r_n, TSLS biased. Instead, use IJIVE1 or UJIVE.
    - K/r_n may be large even if K/n small!
    - JIVE1 not a good solution
- To ensure reliable inference, standard errors need to account for additional many instrument term in the asymptotic variance
    - Even more important is to avoid downward bias that's present in the default standard errors estimator based on TSLS—see notes.
- Will now illustrate in application to Angrist and Krueger (1991)

Estimator	Estimate	$\hat{\mathcal{V}}_1^{1/2}$	$\hat{\mathcal{V}}_2^{1/2}$	$\sqrt{\hat{\mathcal{V}}_2 + \hat{\mathcal{V}}_{MI}}$	$\hat{r}_n/K$
Panel A: OLS					
OLS	0.0670	0.0004			
Panel B: Instrument is QOB. $F = 34.0$					
TSLS	0.1026	0.0195	0.0198		366.0
JIVE1	0.1039	0.0203	0.0206	0.0209	351.6
UJIVE	0.1036	0.0201	0.0204	0.0207	355.2
Panel C: Instrument is $QOB \times YOB$. $F = 4.9$					
TSLS	0.0891	0.0162	0.0176		52.6
JIVE1	0.0959	0.0224	0.0244	0.0273	38.3
UJIVE	0.0938	0.0204	0.0222	0.0211	41.9
Panel D: Instrument is $QOB \times YOB + QOB \times SOB$, $F = 2.6$					
TSLS	0.0928	0.0097	0.0112		26.2
JIVE1	0.1211	0.0205	0.0243	0.0273	12.7
UJIVE	0.1096	0.0160	0.0187	0.0211	16.1
Panel E: Instrument is $QOB \times YOB \times SOB$, $F = 1.1$					
TSLS	0.0721	0.0049	0.0067		11.6
JIVE1	0.0320	0.0307	0.0425	0.0515	-1.9
UJIVE	0.1110	0.0397	0.0548	0.0663	1.4

Ackerberg, Daniel A., and Paul J. Devereux. 2009. "Improved JIVE Estimators for Overidentified Linear Models with and without Heteroskedasticity." *Review of Economics and Statistics* 91, no. 2 (May): 351–362. https://doi.org/10.1162/rest.91.2.351.

Aizer, Anna, and Joseph J. Doyle Jr. 2015. "Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges." *The Quarterly Journal of Economics* 130, no. 2 (May): 759–803. https://doi.org/10.1093/qje/qjv003.

Angrist, Joshua D., and Alan B. Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics* 106, no. 4 (November): 979–1014. https://doi.org/10.2307/2937954.

Bekker, Paul A. 1994. "Alternative Approximations to the Distributions of Instrumental Variable Estimators." *Econometrica* 62, no. 3 (May): 657–681. https://doi.org/10.2307/2951662.

Belloni, Alexandre, Daniel L. Chen, Victor Chernozhukov, and Christian B. Hansen. 2012. "Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain." *Econometrica* 80, no. 6 (November): 2369–2429. https://doi.org/10.3982/ECTA9626.

Chan, David, Matthew Gentzkow, and Chuan Yu. 2019. *Selection with Variation in Diagnostic Skill: Evidence from Radiologists.* Working Paper 26467. Cambridge, MA: National Bureau of Economic Research, November. https://doi.org/10.3386/w26467.

Evdokimov, Kirill, and Michal Kolesár. 2018. "Inference in Instrumental Variable Regression Analysis with Heterogeneous Treatment Effects." January. https://www.princeton.edu/~mkolesar/papers/het_iv.pdf.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2018. "Human Decisions and Machine Predictions." *The Quarterly Journal of Economics* 133, no. 1 (February): 237–293. https://doi.org/10.1093/qje/qjx032.

Kling, Jeffrey R. 2006. "Incarceration Length, Employment, and Earnings." *American Economic Review* 96, no. 3 (May): 863–876. https://doi.org/10.1257/aer.96.3.863.

Kolesár, Michal. 2013. "Estimation in an Instrumental Variables Model With Treatment Effect Heterogeneity." Working paper, Princeton University, November. https://www.princeton.edu/~mkolesar/papers/late_estimation.pdf.

Nagar, Anirudh Lal. 1959. "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations." *Econometrica* 27, no. 4 (October): 575–595. https://doi.org/10.2307/1909352.