4.6 Further IV Details

鲍余薇

2019.10.24

The Outline

- 4.6.1 2SLS Mistakes
    1. Covariate Ambivalence
    2. Forbidden Regressions
- 4.6.2 Peer Effects
- 4.6.3 Limited Dependent Variables Reprise
- 4.6.4 The Bias of 2SLS

4.6.1 2SLS Mistakes

- In the Manual 2SLS procedure, you estimate the first stage yourself, and plug the fitted values into the second stage equation, which is then estimated by OLS

$$s_i = \pi_{10} X_i' + \pi_{11}' Z_i + \xi_{1i}$$
$$Y_i = a' X_i + \rho \hat{s}_i + [\eta_i + \rho(S_i - \hat{s}_i)]$$

- X_i is a set of covariates, Z_i is a set of excluded instruments, and the first stage fitted values are

$$\hat{s}_i = X_i' \pi_{10} + \pi_{11}' Z_i$$

4.6.1 2SLS Mistakes with Covariate Ambivalence

- **Example:** Constructing 2SLS estimates of a wage equation that treats AFQT scores as an endogenous control variable to be instrumented.
- The instruments for AFQT are early schooling (completed before military service), race, and family background variables.

$$s_i = \pi_{10} X_{0i}' + \pi_{11}' Z_i + \xi_{1i}$$
$$Y_i = \alpha_0' X_{0i} + \alpha_0' X_{1i} + \rho \hat{s}_i + [\eta_i + \rho(S_i - \hat{s}_i)]$$

4.6.1 2SLS Mistakes with Forbidden Regressions

- **Example:**

  The causal model of interest: $Y_i = a'X_i + \rho D_i + \eta_i$

  The usual 2SLS first stage $D_i = \pi'_{10} X_i + \pi'_{11} Z_i + \xi_{1i}$

- Suppose that we use Probit to model $E[D_i|X_i, Z_i]$, The Probit first stage is

$$\widehat{\mathrm{D}}_{pi} = \Phi[X_i'\pi_{p0}+\pi'_{p1}Z_i]$$

- In the second stage, substitute $\widehat{\mathrm{D}}_{pi}$ for D_i

$$Y_i = a'X_i + \rho\widehat{\mathrm{D}}_{pi} + [\eta_i + \rho\big(D_i - \widehat{\mathrm{D}}_{pi}\big)]$$

4.6.1 2SLS Mistakes with Forbidden Regressions

- **A simple alternative:**
- Instead of plugging in nonlinear fitted values, we can **use the nonlinear fitted values as instruments**.
- Advantage: if the nonlinear model gives a better approximation of the first-stage CEF than the linear model, the resulting 2SLS estimates will be more efficient than those using a linear first stage.
- Disadvantage: it implicitly uses nonlinearities in the first stage as a source of **identifying information**..

4.6.1 2SLS Mistakes with Forbidden Regressions

- **Example**: the causal relation between schooling and earnings is approximately quadratic

$$Y_i = a'X_i + \rho_1 s_i + \rho_1 s_i^2 + \eta_i$$

- This model treats both s_i and s_i^2 as endogenous. In this case, there are two first-stage equations, one for s_i and one for s_i^2
- It's natural to use Z_i and its square (unless Z_i is a dummy)

4.6.1 2SLS Mistakes with Forbidden Regressions

- Or maybe we can work with a single first stage
  $Y_i = a'X_i + \rho_1 \hat{s}_i + \rho_2 \hat{s}_i{}^2 + [\eta_i + \rho_1(S_i - \hat{s}_i) + \rho_2(S_i^2 - \hat{s}_i{}^2)]$
- This is a mistake since $\hat{s}_i$ can be correlated with $S_i^2 - \hat{s}_i{}^2$ while $\hat{s}_i{}^2$ can be correlated with both $S_i - \hat{s}_i$ and $S_i^2 - \hat{s}_i{}^2$
- On the other hand, as long as X_i and Z_i are uncorrelated with η_i , and you have enough instruments in Z_i , 2SLS estimation of is straightforward.

4.6.2 Peer Effects

- ***Peer Effects*** means the causal effect of group characteristics on individual outcomes.
- There are two types of peer effects. The first concerns the effect of group characteristics such as the average schooling in a state or city on individually-measured outcome variable.
- The second is the effect of the group average of a variable on the individual level of this same variable.

4.6.2 Peer Effects

- **The first type** links the average of one variable to individual outcomes as described by another variable.
- **Example:** whether a given individual's earnings are affected by the average schooling in his or her state of residence.

$$Y_{ijt} = \delta_j + \lambda_t + \gamma \bar{S}_{jt} + \rho s_i + u_{jt} + \eta_{ijt}$$

4.6.2 Peer Effects

$$Y_{ijt} = \delta_j + \lambda_t + \gamma \bar{S}_{jt} + \rho s_i + u_{jt} + \eta_{ijt}$$

- **A simpler version:**

$$Y_{ij} = \mu + \pi_0 s_i + \pi_1 \bar{S}_j + V_i$$

- Where $E[V_i s_i] = E\left[V_i \bar{S}_j\right] = 0$
- Y_{ij} is the log weekly wage of individual i in state j
- $\bar{S}_j$ is average schooling in the state

4.6.2 Peer Effects

- Let ρ_0 denote the coefficient from a bivariate regression of Y_{ij} on s_i only
- Let ρ_1 denote the coefficient from a bivariate regression of Y_{ij} on $\bar{S}_j$ only

$$\pi_0 = \rho_1 + \phi(\rho_0 - \rho_1)$$
$$\pi_1 = \phi(\rho_1 - \rho_0)$$

- Where $\phi = \frac{1}{1-R^2} > 1$, and R^2 is the first-stage R-squared

4.6.2 Peer Effects

- **The second type** is not really an IV problem; it takes us back to basic regression issues
- **Example:** peer effect in high school graduation rates

$$S_{ij} = \mu + \pi_2 \bar{S}_j + \xi_{ij}$$

- where S_{ij} is individual i's high school graduation status and $\bar{S}_j$ is the average high school graduation rate in school j, which i attends.
- The regression of S_{ij} on $\bar{S}_j$ always has a coefficient of 1

4.6.2 Peer Effects

- A modestly improved version of the bad peer regression

$$S_{ij} = \mu + \pi_4 \bar{S}_{(i)j} + \xi_{ij}$$

- Where $\bar{S}_{(i)j}$ is the mean of S_{ij} in school j, excluding student i
- But it's still problematic because S_{ij} and $\bar{S}_{(i)j}$ are both affected by school-level random shocks

4.6.2 Peer Effects

- **The best shot** focuses on variation in **ex ante** peer characteristics
- **Example:** The link between classmates' family background, as measured by the number of books in their homes, and student achievement in European primary schools.

$$S_{ij} = \mu^* + \pi_4 \bar{B}_{(i)j} + \xi_{ij}$$

- Where $\bar{B}_{(i)j}$ is the average number of books in the home of student i's peers

4.6.2 Peer Effects

- **Another Example:**
- The impact of bused-in low-achieving newcomers on high-achieving residents' test scores

$$S_{ij} = \mu + \pi_3 \overline{m}_j + \xi_{ij}$$

- Where $\overline{m}_j$ is the number of bused-in low-achievers in school j and S_{ij} is resident-student i's test score

4.6.3 Limited Dependent Variables Reprise

- 2SLS is not the only way to go
- **An alternative** more elaborate approach tries to build up a causal story by describing the process generating LDVs in detail
- **Example:** Bivariate Probit
- Suppose that a woman decides to have a third child by comparing costs and benefits using a net benefit function or latent index that is linear in covariates and excluded instruments, with a random component or error term, v_i

4.6.3 Limited Dependent Variables Reprise

- The bivariate Probit first stage can be written
$$D_i = 1[X_i'\gamma_0 + \gamma_1 z_i > v_i]$$
- Where z_i is an instrumental variable that increases the benefit of a third child, conditional on covariates, X_i
- An outcome of primary interest in this context is employment status, a Bernoulli random variable with a conditional mean between zero and one.
$$Y_i = 1[X_i'\beta_0 + \beta_1 D_i > \varepsilon_i]$$
- Where ε_i is a second random component or error term

4.6.3 Limited Dependent Variables Reprise

- The source of omitted variables bias in the bivariate Probit setup is correlation between v_i and ε_i
- So the parameters can be estimated by maximum likelihood

$$\sum Y_i \ln \Phi_b \left(\frac{X_i'\beta_0 + \beta_1 D_i}{\sigma_\varepsilon}, \frac{X_i'\gamma_0 + \gamma_1 Z_i}{\sigma_v}; \rho_{\varepsilon v}\right) + (1 - Y_i)\ln[1 - \Phi_b\left(\frac{X_i'\beta_0 + \beta_1 D_i}{\sigma_\varepsilon}, \frac{X_i'\gamma_0 + \gamma_1 Z_i}{\sigma_v};\ \rho_{\varepsilon v}\right)]$$

4.6.3 Limited Dependent Variables Reprise

- The potential outcomes defined by the bivariate Probit model are
- $Y_{0i} = 1[X_i'\beta_0 > \varepsilon_i]$ and $Y_{1i} = 1[X_i'\beta_0 + \beta_1 > \varepsilon_i]$

- While potential treatment assignments are
- $D_{0i} = 1[X_i'\gamma_0 > v_i]$ and $D_{1i} = 1[X_i'\gamma_0 + \gamma_1 > v_i]$

4.6.3 Limited Dependent Variables Reprise

- The average causal effect of childbearing is
- $E[Y_{1i} - Y_{0i}] = E\{1[X_i'\beta_0 + \beta_1 > \varepsilon_i] - 1[X_i'\beta_0 > \varepsilon_i]\}$
- While the average effect on the treated is
- $E[Y_{1i} - Y_{0i} | D_i = 1] = E\{1[X_i'\beta_0 + \beta_1 > \varepsilon_i] - 1[X_i'\beta_0 > \varepsilon_i] \,|\, X_i'\gamma_0 + \gamma_1 > v_i\}$

4.6.3 Limited Dependent Variables Reprise

- Under normality, the average causal effect is
$$E\{1[X_i'\beta_0+\beta_1>\varepsilon_i]-1[X_i'\beta_0>\varepsilon_i]\}$$
$$=E\left\{\Phi\left[\frac{X_i'\beta_0+\beta_1}{\sigma}\right]-\Phi\left[\frac{X_i'\beta_0}{\sigma}\right]\right\}$$
- The effect on the treated is a little more complicated since it involves the bivariate normal CDF
- $E[Y_{1i}-Y_{0i}|D_i=1]=$
$$E\left\{\frac{\Phi_b\left(\frac{X_i'\beta_0+\beta_1}{\sigma_\varepsilon},\frac{X_i'\gamma_0+\gamma_1Z_i}{\sigma_v};\rho_{\varepsilon v}\right)-\Phi_b\left(\frac{X_i'\beta_0}{\sigma_\varepsilon},\frac{X_i'\gamma_0+\gamma_1Z_i}{\sigma_v};\rho_{\varepsilon v}\right)}{\Phi_b\left(\frac{X_i'\gamma_0+\gamma_1Z_i}{\sigma_v}\right)}\right\}$$

4.6.4 The Bias of 2SLS

- Now we are estimating the effect of a single endogenous regressor, stored in a vector x, on a dependent variable, stored in the vector y, with no other covariates.

$$y = \beta x + \eta$$

- The associated first-stage equation is:

$$x = Z\pi + \xi$$

- OLS estimates are biased because η_i is correlated with ξ_i. The instruments, Z_i are uncorrelated with ξ_i by construction and uncorrelated with η_i by assumption.

4.6.4 The Bias of 2SLS

- After tedious calculation, we can get that

$$E\left[\hat{\beta}_{2SLS} - \beta\right] \approx \frac{\sigma_{\eta\xi}}{\sigma_\xi^2} \frac{1}{F+1}$$

- From this we see that as the first stage F-statistic gets small, the bias of 2SLS approaches $\frac{\sigma_{\eta\xi}}{\sigma_\xi^2}$. The bias of the OLS estimator is $\frac{\sigma_{\eta\xi}}{\sigma_x^2}$, which also equals $\frac{\sigma_{\eta\xi}}{\sigma_\xi^2}$ if $\pi = 0$
- On the other hand, the bias of 2SLS vanishes when **F gets large**, as it should happen in large samples when $\pi \neq 0$

4.6.4 The Bias of 2SLS

- **The LIML estimator** is approximately median-unbiased for over-identified constant-effects models, and therefore provides an attractive alternative to just-identified estimation using one instrument at a time.
- It has two advantages:

1. Having the same large-sample distribution as 2SLS while providing finite-sample bias reduction.
2. Many statistical packages compute it while other estimators typically require some programming.

4.6.4 The Bias of 2SLS

- ***So what should we handle this problem in practice?***

1. Report the first stage and think about whether it makes sense.
2. Report the F-statistic on the excluded instruments. The bigger this is, the better.
3. Pick your best single instrument and report just-identified estimates using this one only.
4. Check over-identified 2SLS estimates with LIML.
5. Look at the coefficients, t-statistics, and F-statistics for excluded instruments in the reduced-form regression of dependent variables on instruments.

4.6.4 The Bias of 2SLS

Table 4.6.2: Alternative IV estimates of the economic returns to schooling

	(1)	(2)	(3)	(4)	(5)	(6)
2SLS	0.105	0.435	0.089	0.076	0.093	0.091
	(0.020)	(0.450)	(0.016)	(0.029)	(0.009)	(0.011)
LIML	0.106	0.539	0.093	0.081	0.106	0.110
	(0.020)	(0.627)	(0.018)	(0.041)	(0.012)	(0.015)
F-statistic (excluded instruments)	32.27	0.42	4.91	1.61	2.58	1.97
	Controls					
Year of birth	✓	✓	✓	✓	✓	✓
State of birth					✓	✓
Age, Age squared		✓		✓		✓
	Excluded Instruments					
Quarter of birth	✓	✓				
Quarter of birth*year of birth			✓	✓	✓	✓
Quarter of birth*state of birth					✓	✓
Number of excluded instruments	3	2	30	28	180	178

Notes: The table compares 2SLS and LIML estimates using alternative sets of instruments and controls. The OLS estimate corresponding to the models reported in columns 1-4 is .071; the OLS estimate corresponding to the models reported in columns 5-6 is .067. Data are from the Angrist and Krueger (1991) 1980 Census sample. The sample size is 329,509. Standard errors are reported in parentheses.

4.6.4 The Bias of 2SLS

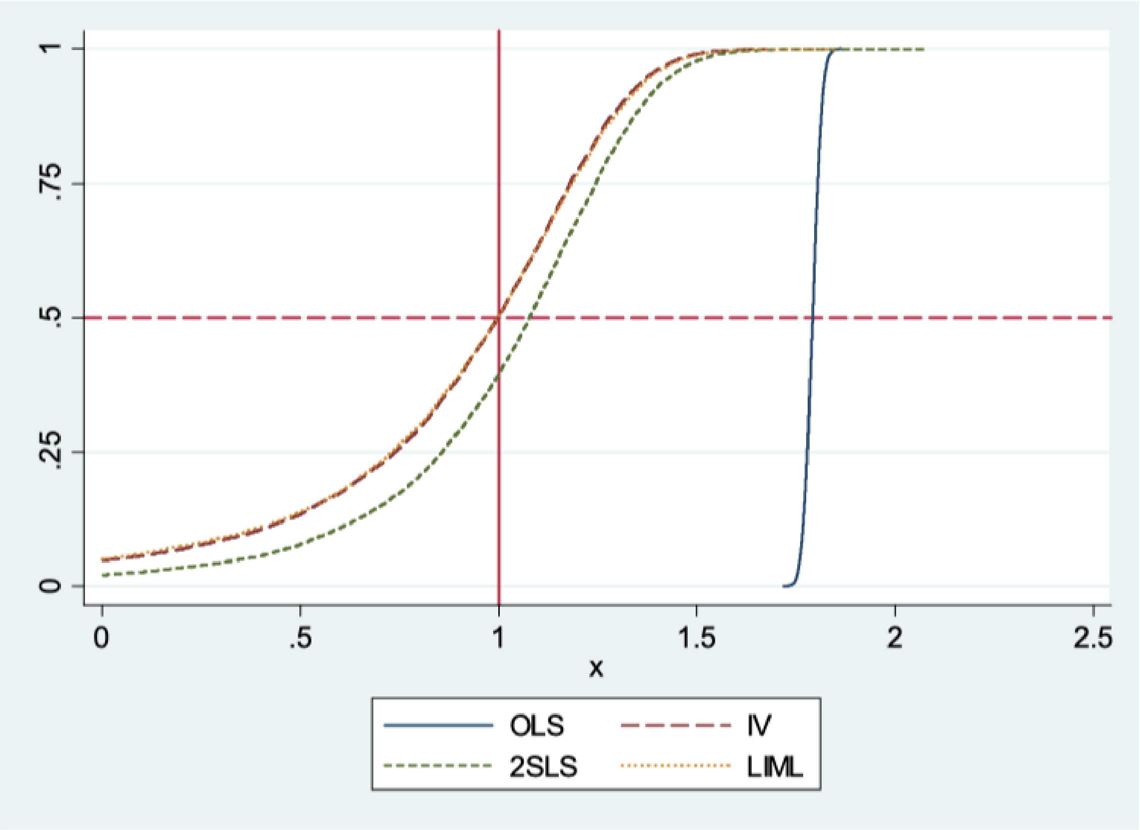

Figure 4.6.1: Distribution of the OLS, IV, 2SLS, and LIML estimators. IV uses one instrument, while 2SLS and LIML use two instruments.

4.6.4 The Bias of 2SLS

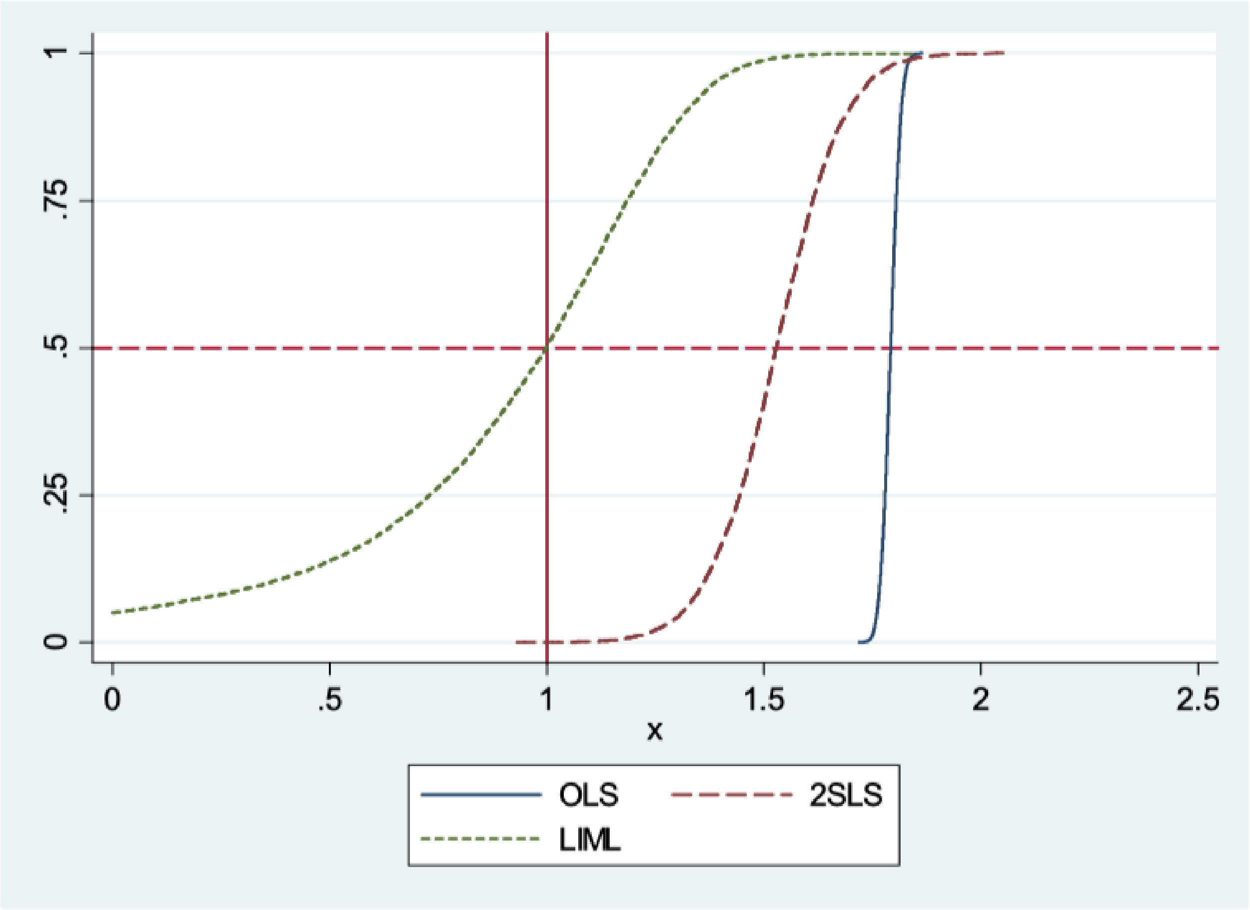

Figure 4.6.2: Distribution of the OLS, 2SLS, and LIML estimators with 20 instruments

4.6.4 The Bias of 2SLS

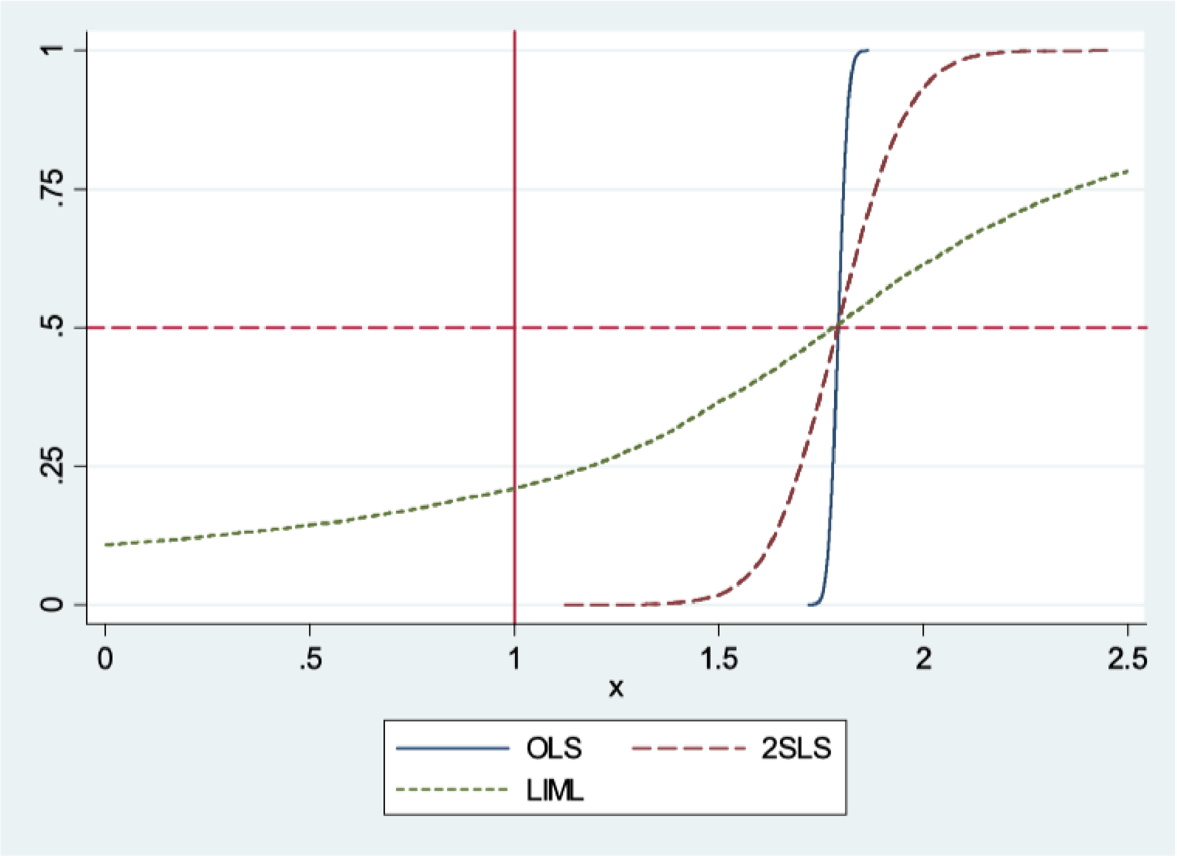

Figure 4.6.3: Distribution of the OLS, 2SLS, and LIML estimators with 20 worthless instruments