1 Pitfalls in 2SLS Regressions

1.1 Computing Standard Errors

Consider the following regression:

$$
\begin{aligned}
Y &= X'\alpha + X^*\beta + \eta \\
X^* &= X'\pi_{10} + \pi_{11} Z + \epsilon \\
&\mathrm{Cov}(\eta, \epsilon) \neq 0
\end{aligned}
$$

Then

$$
\begin{pmatrix} \hat{\pi}_{10} \\ \hat{\pi}_{11} \end{pmatrix} = \left(W^\top W\right)^{-1} W^\top X^*, \qquad W = (X, Z)
$$

so

$$
\begin{aligned}
Y &= X'\alpha + \beta X^* + \eta \\
&= X'\alpha + \beta\left(\hat{X}^*\right) + \beta\left(X^* - \hat{X}^*\right) + \eta
\end{aligned}
$$

The takeaway is that when you compute the standard errors, you have to use X^* not $\hat{X}^*$:

$$
\hat{u} = Y - X'\hat{\alpha}_{2SLS} - \hat{\beta}_{2SLS} X^*
$$

1.2 Concocting Endogenous Variables

Consider only using X_1 in the first-stage regression but both X_1 and X_2 in the second-stage regression:

$$
\begin{aligned}
X^* &= \alpha X_1 + \pi Z + \epsilon \\
Y &= (X_1, X_2)^\top \gamma + \beta X^* + \eta \\
&= (X_1, X_2)^\top \gamma + \beta \hat{X}^* + \underbrace{\beta\left(X^* - \hat{X}^*\right) + \eta}_{\text{error terms}}
\end{aligned}
$$

so the error features variance explained by X_2. Congratulations! – You have successfully made X_2 endogenous.

1.3 Loss of Efficiency

Consider a converse situation:

$$
\begin{aligned}
X^* &= (X_1, X_2)^\top \gamma + \pi Z + \epsilon \\
Y &= \alpha X_1 + \beta \hat{X}^* + \beta\left(X^* - \hat{X}^*\right) + \eta
\end{aligned}
$$

You will have a consistent estimate but higher standard errors:

1.4 Quadratic Endogeneity

Do we need to use two instruments or one in this case?

$$
Y = \alpha X + \beta_1 X^* + \beta_2 (X^*)^2 + \eta
$$

The answer is that we should use two. To see this:

$$\begin{aligned} X^* &= \gamma' X + \pi Z + \epsilon \\ \Rightarrow \hat{X}^* &= \hat{\gamma}' X + \hat{\pi} Z \end{aligned}$$

and substituting into the equation:

$$Y = \alpha X + \beta_1 \hat{X}^* + \beta_2 \left(\hat{X}^*\right)^2 + \beta_1 \left(X^* - \hat{X}^*\right) + \beta_2 \left(\underbrace{(X^*)^2 - \left(\hat{X}^*\right)^2}\right) + \eta$$

but the underlined term is equal to $\epsilon^2 + 2\epsilon\left(\gamma' X + \pi Z\right)$. Both terms are problematic.

1.5 Finite Sample Bias

Consider the following:

$$\begin{aligned} Y &= \beta X + \eta \\ X &= Z\pi + \epsilon \end{aligned}$$

Writing out the fitted values:

$$Y = \beta \hat{X} + \beta \left(X - \hat{X}\right) + \eta$$

where

$$\hat{X} = Z\hat{\pi} = \underbrace{Z' \left(Z'Z\right)^{-1} Z}_{P} X$$

which means

$$\begin{aligned} \hat{\beta} &= \left(\hat{X}'\hat{X}\right)^{-1} \hat{X}' Y \\ &= \left(X'PX\right)^{-1} X'PY \\ &= \left(X'PX\right)^{-1} \left(Z\pi + \epsilon\right)' PY \\ &= \beta + \left(X'PX\right)^{-1} \pi' Z' \eta + \left(X'PX\right)^{-1} \epsilon' P \eta \end{aligned}$$

2 RDD

2.1 Expression of LATE for RDD

We have

$$\mathbb{E}\left[Y_1 - Y_0 | R = c,\ \text{compliers}\right] = \frac{\lim_{\tau \downarrow c} \mathbb{E}\left[Y | R = \tau\right] - \lim_{\tau \uparrow c} \mathbb{E}\left[Y | R = \tau\right]}{\lim_{\tau \downarrow c} \mathbb{E}\left[D | R = \tau\right] - \lim_{\tau \uparrow c} \mathbb{E}\left[D | R = \tau\right]}$$

Write:

$$Z = \mathbb{I}\left\{D \geq c\right\}$$

and

$$D = D_0 + Z\left(D_1 - D_0 So\right) the$$

Thus the numerator can be rewritten as:

$$
\begin{aligned}
& \lim_{\tau \downarrow c} \mathbb{E}\left[Y|R=\tau\right] - \lim_{\tau \uparrow c} \mathbb{E}\left[Y|R=\tau\right] \\
=& \lim_{\tau \downarrow c} \mathbb{E}\left[Y_0 + D\left(Y_1 - Y_0\right)|R=\tau\right] - \lim_{\tau \uparrow c} \mathbb{E}\left[Y_0 + D\left(Y_1 - Y_0\right)|R=\tau\right] \\
=& \mathbb{E}\left[Y_1 - Y_0|R=c, T=cp\right] P\left(T=cp|R=c\right) + \mathbb{E}\left[Y_1 - Y_0|R=c, T=at\right] P\left(T=at|R=c\right) \\
& - \mathbb{E}\left[Y_1 - Y_0|R=\tau, T=at\right] P\left(T=at|R=\tau\right) \\
=& \mathbb{E}\left[Y_1 - Y_0|R=c, T=cp\right] P\left(T=cp|R=c\right)
\end{aligned}
$$

The denominator can be rewritten as:

$$P\left(T=cp|R=c\right)$$

And thus we arrive at our desired result.