Econometrics II

Fabian Waldinger (LMU Munich)

Lecture Structure

1. Recap from last lecture
2. Continuation Instrumental Variables

Recap from Last Lecture I

- IV allows us to obtain unbiased estimates (overcome violations of GM3)
- Two key assumptions:
  1. $Cov(S, Z) \neq 0$ (first stage exists)
  2. $Cov(Z, \varepsilon) = 0$ (exclusion restriction: Z is uncorrelated with any other determinants of the dependent variable)
- IV is not unbiased but consistent if $N \to \infty$

Recap from Last Lecture II

- Some IV jargon:
  - Causal relationship of interest:
    $$y = \beta_1 + \beta_2 x + \varepsilon$$
  - First-Stage regression:
    $$x = \gamma_1 + \gamma_2 z + \mu$$
  - Second-Stage regression:
    $$y = \beta_1 + \beta_2 \hat{x} + \upsilon$$
  - Reduced form:
    $$y = \lambda_1 + \lambda_2 z + \upsilon$$

IV with Matrix Algebra

- As for OLS it is often useful to write the IV estimator using matrix algebra
- The model of interest is:

$$y_i = \beta_1 + \beta_2 x_{i2} + ... + \beta_K x_{iK} + \varepsilon$$

- We assume that all x_{ik} apart from x_{iK} are exogenous. x_{iK}, however, violates GM3 and is correlated with the error term: $E(x_K \varepsilon) \neq 0$
- Suppose we have an IV for x_{iK} which we call z_{i1}

Data Structure Matrix Notation

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_{12} & x_{13} & \cdots & x_{1K} \\ 1 & x_{22} & x_{23} & \cdots & x_{2K} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N2} & x_{N3} & \cdots & x_{NK} \end{pmatrix}_{N \times k}$$

$$\boldsymbol{Z} = \begin{pmatrix} 1 & z_{12} & z_{13} & \cdots & z_{11} \\ 1 & z_{22} & z_{23} & \cdots & z_{21} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_{N2} & z_{N3} & \cdots & z_{N1} \end{pmatrix}_{N \times k}$$

Deriving the IV Estimator - Special Case

- We first consider the special case where the number of IV equals the number of endogenous variables: $\boldsymbol{X'Z}$ has dimensions $k \times k$ and is nonsingular
- In matrix notation the model can be written as:

$$y = \boldsymbol{X}\beta + \varepsilon$$

- If the IV is exogenous (exclusion restriction is satisfied):

$$E(\boldsymbol{Z}'\varepsilon) = 0$$

- Plugging in the model:

$$E(\boldsymbol{Z}'(y - \boldsymbol{X}\beta)) = 0$$

- From this it follows that:

$$E(\boldsymbol{Z'X})\beta = E(\boldsymbol{Z}'y)$$

Deriving the IV Estimator

- If the matrix $E(\boldsymbol{Z'X})$ has full rank i.e. (this is the equivalence of the first stage assumption in last week's lecture)

$$rank\ E(\boldsymbol{Z'X}) = k$$

- we can invert $E(\boldsymbol{Z'X})$ and solve for β:

$$\beta = [E(\boldsymbol{Z'X})]^{-1} E(\boldsymbol{Z}'y)$$

- We can use sample moments to consistently estimate β:

$$\hat{\beta}^{IV} = (\boldsymbol{Z'X})^{-1} \boldsymbol{Z}'y$$

2SLS

- With more than one IV we can estimate 2SLS:
  1. Estimate the first stage and get the predicted values $\hat{\boldsymbol{X}}$
  2. Use the $\hat{\boldsymbol{X}}$ in a second stage that you estimate using OLS
- The second stage regression would be as follows:

$$\hat{\beta}^{2SLS} = (\hat{\boldsymbol{X}}'\hat{\boldsymbol{X}})^{-1}\hat{\boldsymbol{X}}'y$$

2SLS

- How does the 2SLS estimator look like if we express it in terms of the raw data?

$$\hat{\mathbf{X}} = \mathbf{Z}\hat{\gamma} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}$$

- Plugging this into the equation above:

$$\hat{\beta}^{2SLS} = ([\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]'[\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}])^{-1}[\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]'y$$

$$= (\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'y$$

$$= (\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'y$$