

Endogeneity and Instrumental Variables: Lecture XXI

Charles B. Moss ¹

¹University of Florida

March 9, 2012

- 1 Readings
- 2 Instrumental Variables
 - Indiana Land Values
 - OLS and IV-Exact Estimation
 - Overidentified Instrumental Variable
- 3 Two Stage Least Squares
- 4 Readings – Next Time

Readings

- * Bowden, R.J. and D.A. Turkington. 1984. *Instrumental Variables*. Cambridge University Press. (Chapters 1 and 2 pp. 1-65)
- * Cameron, A. C. and P. K. Trivedi. 2010. *Microeconometrics Using Stata Revised Edition*. Stata Press. (Chapter 6).
- * Greene, W.H. 2012. *Econometric Analysis* Seventh Edition (Chapter 8 Section 8.1 8.3 pp. 219 230)
- Radner, R. and F. Bobkoski. 1992. Some Recent Work of H. Theil on Estimation in Systems of Simultaneous Equations. In *Henri Theils Contributions to Economics and Econometrics: Volume I* (eds) Baldev,R. and J. Koerts. Kluwer Academic Publishers, 117-125.

Reading

- * Theil, H.1992. Estimation and Simultaneous Correlation in Incomplete Equations Systems. In *Henri Theils Contributions to Economics and Econometrics: Volume I* (eds) Baldev,R. and J. Koerts. Kluwer Academic Publishers, 65-107.

Instrumental Variables - A Concept

The method of instrumental variables is developed around the following strategy: Suppose that in the model

$$y_i = x_i' \beta + \epsilon_i, \quad (1)$$

the K variables x_i may be correlated with ϵ_i . Suppose as well that there exists a set of L variables z_i , such that z_i is correlated with x_i , but not with ϵ_i . We cannot estimate β consistently by using the familiar least squares estimator. But, the assumed lack of correlation between z_i and ϵ_i implies a set of relationships that may allow us to construct a consistent estimator of β by using the relationship among z_i , x_i , and ϵ_i .

Indiana Land Value Data

Year	Prices					Interest Rate	Land Value	Income per Acre	Cash Rent	Debt to Asset
	Corn	Seed	Fert.	Fuel	Labor					
2010	4.94	311	286	309	192	0.054	4419	114.79	161	0.108
2009	3.62	304	230	264	189	0.060	4367	101.25	158	0.111
2008	3.55	275	396	207	185	0.069	4240	147.52	157	0.100
2007	4.04	211	253	308	179	0.075	3688	112.07	139	0.091
2006	3.68	191	174	227	172	0.070	3162	46.96	127	0.109
2005	3.01	171	185	221	166	0.062	2945	80.90	126	0.120
2004	1.71	158	153	167	161	0.063	2693	83.29	122	0.120
2003	1.82	157	128	130	156	0.064	2509	50.06	120	0.118

2SLS and IV

Land Value Model - Continued

- There are actually two formulation:
 - Exact number of instruments – In this case three instruments (we will use the corn price, the fertilizer price, and the debt-to-asset ratio).
 - More instruments than independent variables – In this case we will use five instruments (corn price, seed price, fertilizer price, fuel price, and debt-to-asset ratio).
- The exact IV estimator is

$$\beta_{IV} = (Z'X)^{-1} (Z'y). \quad (4)$$

The variance for this estimator is then

$$V_{IV} = \hat{\sigma}^2 (Z'X)^{-1} (Z'Z) (X'Z)^{-1}. \quad (5)$$

OLS and IV-Exact Estimation

Parameter	Ordinary Least Squares			Instrumental Variable		
	Estimate	Std. Dev.	t-Ratio	Estimate	Std. Dev.	t-Ratio
α_0	5.550	1.802	3.080	-16.684	78.337	-0.213
α_1	0.416	0.163	2.548	0.922	1.778	0.519
α_2	-0.268	0.634	-0.423	-7.540	26.178	-0.288

$$F_{5,2}^{OLS} = \frac{(r - R\beta_{OLS})' [R\hat{V}_{OLS}R']^{-1} (r - R\beta_{OLS})}{\hat{s}_{OLS}^2} = 7.678 \quad (6)$$

$$F_{5,2}^{IV} = \frac{(r - R\beta_{IV})' [R\hat{V}_{IV}R']^{-1} (r - R\beta_{IV})}{\hat{s}_{IV}^2} = 0.109$$

Overidentified Instrumental Variable

- The Overidentified Instrumental Variable formulation allows a simple transition to the Two-Stage Least Squares formulation.
- Returning to the spanning-space concept, consider the projection of the X variable into a space spanned by a set of instrumental variables

$$X^* = \left[Z (Z'Z)^{-1} Z' \right] X \quad (7)$$

The transformed regression is then

$$\beta_{IV}^* = ([X^*]' X^*)^{-1} [X^*]' y^* \quad (8)$$

Overidentified Instrumental Variable - Continued

- In this case, we can label $P_Z = Z(Z'Z)^{-1}Z'$ as a projection matrix. Note that P_Z is idempotent

$$P_Z P_Z = \left(Z(Z'Z)^{-1}Z' \right) \left(Z(Z'Z)^{-1}Z' \right) = Z(Z'Z)^{-1}Z' \quad (9)$$

- Thus, the Overidentified Instrumental Variable estimator can be expressed as

$$\beta_{IV}^* = (X'P_ZX)^{-1}X'P_Zy \quad (10)$$

with the variance matrix

$$V_{IV}^* = \hat{\sigma}^2 (X'P_ZX)^{-1} \quad (11)$$

OLS and Overidentified IV

Parameter	Ordinary Least Squares			Instrumental Variable		
	Estimate	Std. Dev.	t-Ratio	Estimate	Std. Dev.	t-Ratio
α_0	5.550	1.802	3.080	2.347	3.545	0.662
α_1	0.416	0.163	2.548	0.509	0.224	2.274
α_2	-0.268	0.634	-0.423	-1.282	1.174	-1.092

$$F_{5,2}^{IV*} = \frac{(r - R\beta_{IV}^*)' [R\hat{V}_{IV}^*R']^{-1} (r - R\beta_{IV}^*)}{\hat{S}_{IV*}^2} = 2.611 \quad (12)$$

Two Stage Least Squares

- Next, change the formulation a little bit.
- For example, assume that the supply and demand curves for a given market can be written as

$$q^s = -3 + 4p_1 - p_2 + \epsilon_1 \quad (13)$$

$$q_d = 10 - p_1 + 2y + \epsilon_2$$

Solving this two equation systems yields

$$p_1 = \frac{13}{5} + \frac{1}{5}p_2 + \frac{2}{5}y - \epsilon_1 + \epsilon_2 \quad (14)$$

Two Stage Least Squares - Continued

- Ignoring the problem of simultaneity, the supply equation can be estimated as

$$\begin{aligned} q^s &= \alpha_0 + \alpha_1 p_1 + \alpha_2 p_2 + \tilde{\epsilon}_1 \\ q^d &= \beta_0 + \beta_1 p_1 + \beta_2 y + \tilde{\epsilon}_2 \end{aligned} \quad (15)$$

Table: Ordinary Least Squares Estimates

Parameter	Estimate	Parameter	Estimate
α_0	-1.957	β_0	9.564
α_1	3.688	β_1	0.736
α_2	-0.515	β_2	-0.688

Two Stage Least Squares - Continued

- The point of estimating a Two-Stage Least Squares is to first estimate the value of the endogenous variable in terms of the exogenous information (i.e., the instruments).
- In this case, we regress p_1 on p_2 , y , and for a little pizzazz $p_2 \times y$ yielding

$$\beta_{1SLS} = \begin{bmatrix} 3.4906 \\ -0.2232 \\ 0.3285 \\ 0.0347 \end{bmatrix} \quad (16)$$

These values are then used to generate $\hat{p}_1 = Z\hat{\beta}_{1SLS}$.

Two Stage Least Squares - Continued

- The "new" X matrix then becomes $\begin{pmatrix} 1 & \hat{p}_1 & p_2 & y \end{pmatrix}$. The simple regression results then become

$$\hat{\alpha}_{2SLS} = \begin{bmatrix} -2.835 \\ 3.843 \\ -0.599 \end{bmatrix} \quad \hat{\beta}_{2SLS} = \begin{bmatrix} 11.319 \\ -1.507 \\ 2.227 \end{bmatrix} \quad (17)$$

- Of course, the same results can be generated by applying Overidentified Instrumental Variables to the original problem.

Readings – Next Time

- * Greene, W.H. 2012. *Econometric Analysis* Seventh Edition (Chapter 10 Section 10.6 10.7 pp. 314 335)
- * Theil, H. 1971. *Principles of Econometrics*. John Wiley & Sons. (Chapters 9 and 10 pp. 429 439)
- Theil, H. 1992. Estimation of Parameters of Econometric Model. In *Henri Theils Contributions to Economics and Econometrics: Volume I* (eds) Baldev,R. and J. Koerts. Kluwer Academic Publishers, 109-116.
- Theil, H. 1992. The Aggregate Implications of Identifiable Structural Macrorelations. In *Henri Theils Contributions to Economics and Econometrics: Volume I* (eds) Baldev,R. and J. Koerts. Kluwer Academic Publishers, 127-145.

Readings – Next Time

- * Zellner, A. and H. Theil. 1992. Three-Stage Least Squares: Simultaneous Estimation of Simultaneous Equations. In *Henri Theils Contributions to Economics and Econometrics: Volume I* (eds) Baldev, R. and J. Koerts. Kluwer Academic Publishers, 148-178.