Benjamin Faanunu

**Instrumental Variables**

The Problem: Arises when a regressor is correlated with the error term or in other words x is not independent of the error. When this happens, the OLS estimator β will not be consistent.

Main Causes:

1st: When an omitted variable (now in the error term) is correlated with an included variable

 2nd: X contains measurement errors

 3rd: If X contains endogenous variables jointly determined with y

Underlying considerations in choosing an Instrumental Variable:

 1st: Uncorrelated with the error

 2nd: Highly correlated with the regressor for which it is to serve as an instrument

 3rd: \*No direct effect on the regressand

How to choose an appropriate Instrumental Variable:

 1st: Consider context

 2nd: Consider theory

 3rd: Literature (Kennedy has a list of techniques)

Draw Back:

Variance of the Instrumental Variable estimator is larger than that of the OLS estimator.

Instrumental Variable Estimation:

$$Consider the OLS equation y=Xβ+ε where X contians K\_{1} columns of observations on exogenous $$

$$variables that are uncorrelated with ε, and K\_{2} columns of observations on variables correlated with ε.$$

$$Choose K\_{3} columns of instruments \left(at least one for each troublesome variable\right), combine it with K\_{1} $$

$$\left(since each of these can act as their own instrument\right) and call it Z. Regressing X on Z, we get $$

$$\hat{X}=Z(Z^{'}Z)^{-1}Z^{'}X, then our β^{IV}=[X^{'}Z\left(Z^{'}Z\right)^{-1}Z^{'}X]^{-1}X'Z(Z^{'}Z)^{-1}Z^{'}y=(\hat{X}'\hat{X})^{-1}\hat{X}^{'}y $$

$$if Z has the same dimensions as X then β^{IV}=\left(Z'X\right)^{-1}Z'y$$

The variance-covariance matrix of $β^{IV}$ can be estimated using$ \hat{σ}^{2}[X^{'}Z\left(Z^{'}Z\right)^{-1}Z^{'}X]^{-1}=\hat{σ}^{2}(\hat{X}'\hat{X})^{-1}$. Actual variance is too difficult to calculate so it is asymptotically estimated.

**Two-Stage Least Squares**

A special case of the IV sometimes called “Best Instrument” because it combines all the exogenous variables to create a combined variable.

Stage 1: Regress each endogenous variable acting as a regressor in the equation being estimated on all the exogenous variables in the system of simultaneous equation, and calculate the estimated values of these endogenous variables.

Stage 2: Use the estimated values as IVs for the endogenous variables or simply use these estimated values and the included exogenous variables as regressors in OLS regression.

2SLS Example

Demand: $p\_{t}=f\left(q\_{t}^{d},ps\_{t},di\_{t}\right)=β\_{11}+γ\_{11}q\_{t}^{d}+β\_{12}ps\_{t}+β\_{13}di\_{t}+ε\_{1t}$

Supply: $q\_{t}^{s}=f\left(p\_{t},pf\_{t}\right)=β\_{21}+γ\_{21}p\_{t}+β\_{24}pf\_{t}+ε\_{2t}$

Where $p and q are endogenous variables jointly determined and ps\_{t},di\_{t}, and pf\_{t} are all exogenous$

Price is a function of quantity demanded, price of subsidies, and disposable income; quantity is a function of price and price factor (the cost of producing the good).

For convenience, I have rewritten the equations with y representing the endogenous variables and x representing the exogenous variables.

$$y\_{1}=\left[\begin{matrix}y\_{2}&x\_{1}&x\_{2}&x\_{3}\end{matrix}\right]\left[\begin{matrix}γ\_{11}\\β\_{11}\\β\_{12}\\β\_{13}\end{matrix}\right]+e\_{1}= Z\_{1}λ\_{1}+e\_{1}$$

$$y\_{2}=\left[\begin{matrix}y\_{1}&x\_{1}&x\_{4}\end{matrix}\right]\left[\begin{matrix}γ\_{21}\\β\_{21}\\β\_{24}\end{matrix}\right]+e\_{2}= Z\_{2}λ\_{2}+e\_{2}$$

Let$ X=\left[\begin{matrix}x\_{1}&x\_{2}&x\_{3}&x\_{4}\end{matrix}\right]$, the set of all exogenous variables which are assumed to be uncorrelated with the error. Then regressing the endogenous variables on the exogenous variables we have

$$y\_{1}=\left[\begin{matrix}x\_{1}&x\_{2}&x\_{3}&x\_{4}\end{matrix}\right]\left[\begin{matrix}π\_{11}\\π\_{12}\\π\_{13}\\π\_{14}\end{matrix}\right]+v\_{1}= X\hat{π}\_{1}+\hat{v}\_{1}=\hat{y}\_{1}+\hat{v}\_{1}$$

$$y\_{2}=\left[\begin{matrix}y\_{1}&x\_{1}&x\_{4}\end{matrix}\right]\left[\begin{matrix}π\_{21}\\π\_{22}\\π\_{23}\end{matrix}\right]+v\_{2}= X\hat{π}\_{2}+\hat{v}\_{2}=\hat{y}\_{2}+\hat{v}\_{2}$$

Where $\hat{π}\_{1}$ is the least squares estimator of $π\_{1}$ and $\hat{v}\_{1}$ are the least squares residuals. Again substituting we obtain

$$y\_{2}=\left[\begin{matrix}\hat{y}\_{1}+\hat{v}\_{1}&x\_{1}&x\_{4}\end{matrix}\right]\left[\begin{matrix}γ\_{21}\\β\_{21}\\β\_{24}\end{matrix}\right]+e\_{2}=\left[\begin{matrix}\hat{y}\_{1}&x\_{1}&x\_{4}\end{matrix}\right]\left[\begin{matrix}γ\_{21}\\β\_{21}\\β\_{24}\end{matrix}\right]+\hat{v}\_{1}γ\_{21}+e\_{2}=\hat{Z}\_{2}λ\_{2}+\overbar{e}\_{2}$$

$$y\_{1}=\left[\begin{matrix}\hat{y}\_{2}+\hat{v}\_{2}&x\_{1}&x\_{2}&x\_{3}\end{matrix}\right]\left[\begin{matrix}γ\_{11}\\β\_{11}\\β\_{12}\\β\_{13}\end{matrix}\right]+e\_{1}=\left[\begin{matrix}\hat{y}\_{2}&x\_{1}&x\_{2}&x\_{3}\end{matrix}\right]\left[\begin{matrix}γ\_{11}\\β\_{11}\\β\_{12}\\β\_{13}\end{matrix}\right]+\hat{v}\_{2}γ\_{11}+e\_{1}=\hat{Z}\_{1}λ\_{1}+\overbar{e}\_{1}$$

These equations derive consistent estimators $\hat{λ}\_{1}=(\hat{Z}\_{1}'\hat{Z}\_{1})^{-1}\hat{Z}\_{1}'y\_{1} $ and $\hat{λ}\_{2}=(\hat{Z}\_{2}'\hat{Z}\_{2})^{-1}\hat{Z}\_{2}'y\_{2}$

Definitions:

Endogenous variable: a variable whose variation an econometric model is designed to explain

Exogenous variable: a variable whose variation is externally determined

Sources:

Kennedy, A Guide to Econometrics 6E

Wooldridge, Introductory Econometrics: A Modern Approach, 4th edition

<http://elsa.berkeley.edu/~mcfadden/e240b_f01/ch4.pdf>