Simultaneous Equations

Math 463, Spring 2017, University of Oregon

David A. Levin

University of Oregon

May 31, 2017

Supply and Demand

- Response schedules: F price of bread, O price of olive oil (determinants of demand), W wage rate, H price of hay (determinants of supply). These are *exogenous*; P price and Q quantity are endogenous.

$$Q = a_0 + a_1 P + a_2 W + a_3 H + \delta_t \qquad \text{supply} \qquad (1a)$$

$$Q = b_0 + b_1 P + b_2 F + b_3 O + \varepsilon_t \qquad \text{demand} \qquad (1b)$$

  Experiment: Set P, W, H, F, O and observe how much butter is brought to market (supply); Set P, W, H, F, O, and observe how much butter consumers will buy. Disturbance terms and parameters are invariant under interventions. The assumptions are that $\{(\delta_t, \varepsilon_t)\}$ are i.i.d., but δ_t and ε_t are correlated.
- The variables (W_t, H_t, F_t, O_t) are drawn independently of δ's and ε's, and then placed into the above equations:

$$Q = a_0 + a_1 P + a_2 W_t + a_3 H_t + \delta_t \qquad \text{supply} \qquad (2a)$$

$$Q = b_0 + b_1 P + b_2 F_t + b_3 O_t + \varepsilon_t \qquad \text{demand} \qquad (2b)$$

- Then (P_t, Q_t) is obtained by solving the two equations above for two unknowns P and Q. Note that P_t and Q_t depend on the errors.
- You can write down formulas for P_t and Q_t in terms of all the quantities above, note that P_t will depend on ε_t and δ_t, and in general is correlated with them.

Instrumental Variables

- Model is that
$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\delta}.$$
Assume $\boldsymbol{\delta}$ has mean zero and variance $\sigma^2 I$, *not* assumed independent of $\boldsymbol{X}$.
-
    - $\boldsymbol{X}$ is $n \times p$ and $\boldsymbol{Z}$ is $n \times q$ with $n > q \geq p$.
    - $\boldsymbol{Z}'\boldsymbol{X}$ and $\boldsymbol{Z}'\boldsymbol{Z}$ have full rank
    - $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\delta}$
    - $\Sigma_{\boldsymbol{\delta}} = \sigma^2 I$
    - $\boldsymbol{Z}$ is independent of $\boldsymbol{\delta}$.
- Perform OLS on
$$(\boldsymbol{Z}'\boldsymbol{Z})^{-1/2}\boldsymbol{Z}'\boldsymbol{Y} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1/2}\boldsymbol{Z}'\boldsymbol{X}\boldsymbol{\beta} + \underbrace{(\boldsymbol{Z}'\boldsymbol{Z})^{-1/2}\boldsymbol{Z}'\boldsymbol{\delta}}_{\boldsymbol{\eta}}$$
Then $\mathrm{Cov}(\boldsymbol{\eta} \mid \boldsymbol{Z}) = \sigma^2 I$.
- Are the assumptions of OLS satisfied, so that we obtain unbiased estimates?

Simulated Example

Model:

$$Y_i = 1.2X_i + \varepsilon_i$$

where $\mathrm{Cov}(X_i, \varepsilon_i) = 0.4$. There are two exogenous "instruments" $Z_{i,1}, Z_{i,2}$. Thus the covariance matrix of $(X_1, Z_{1,1}, Z_{1,2}, \varepsilon_1)$ is

$$\Sigma = \begin{bmatrix} 1 & 0.3 & 0.4 & 0.4 \\ 0.3 & 1 & 0 & 0 \\ 0.4 & 0 & 1 & 0 \\ 0.4 & 0 & 0 & 1 \end{bmatrix}$$

We have $N = 100000$ to eliminate "small-sample bias".
We compare OLS with IVLS.

```
> library(expm)
> set.seed(101)
> Sig = matrix(c(1,0.3,0.4,0.4,0.3,1,0,0,0.4,0,1,0,0.4,0,0,1),ncol=4,by
> Sig
     [,1] [,2] [,3] [,4]
[1,]  1.0  0.3  0.4  0.4
[2,]  0.3  1.0  0.0  0.0
[3,]  0.4  0.0  1.0  0.0
[4,]  0.4  0.0  0.0  1.0
> SigSD = eigen(Sig)
> SigSR = SigSD$vectors%*%diag(sqrt(SigSD$values))%*%t(SigSD$vectors)
> xzzd = t(SigSR%*%matrix(rnorm(I(100000*4)),ncol=100000))
> y = 1.2*xzzd[,1]+6*xzzd[,4]
> M = sqrtm(solve(t(xzzd[,2:3])%*%xzzd[,2:3]))%*%t(xzzd[,2:3])
> my = M%*%y
> mx = M%*%xzzd[,1]
> fit = lm(my~mx-1)
> fit2 = lm(y~xzzd[,1]-1)
> b=coef(fit);bb=coef(fit2)
> b;bb
      mx
1.179119
xzzd[, 1]
 3.606102
```

Note that

$$\mathrm{Cov}(\hat{\beta} \mid \boldsymbol{Z}) = \hat{\sigma^2}[\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-\mathbf{1}}\boldsymbol{Z}'\boldsymbol{X}]^{-1}$$

```
> x = xzzd[,1];z=xzzd[,2:3]
> fitval = x*b; sighat = sqrt(sum((y-fitval)^2)/(100000-1))
> sighat

[1] 5.988697

> unscalecov = solve(t(x)%*%z%*%solve(t(z)%*%z)%*%t(z)%*%x)
> sighat*sqrt(unscalecov)

          [,1]
[1,] 0.0381629
```

Compare with

Example: Education and fertility

Table 1. Variables in the model (Rindfuss et al 1980).

The endogenous variables	
ED	Respondent's education (Years of schooling completed at first marriage)
AGE	Respondent's age at first birth
The exogenous variables	
OCC	Respondent's father's occupation
RACE	Race of respondent (Black = 1, other = 0)
NOSIB	Respondent's number of siblings
FARM	Farm background (coded 1 if respondent grew up on a farm, else coded 0)
REGN	Region where respondent grew up (South = 1, other = 0)
ADOLF	Broken family (coded 0 if both parents present when respondent was 14, else coded 1)
REL	Religion (Catholic = 1, other = 0)
YCIG	Smoking (coded 1 if respondent smoked before age 16, else coded 0)
FEC	Fecundability (coded 1 if respondent had a miscarriage before first birth; else coded 0)

Notes: The data are from a probability sample of 1766 women 35–44 years of age residing in the continental United States. The sample was restricted to ever-married women with at least one child. OCC was measured on Duncan's scale (section 6.1), combining information on education and income. Notation differs from Rindfuss et al.

Two equations:

$$\begin{aligned} \mathrm{ED} &= a_0 + a_1\mathrm{AGE} + a_2\mathrm{OCC}_i + \cdots + a_{10}\mathrm{YCIG}_i + \delta_i \\ \mathrm{AGE} &= b_0 + b_1\mathrm{ED} + a_3\mathrm{FEC}_i + \cdots + a_{10}\mathrm{YCIG}_i + \varepsilon_i \end{aligned}$$

- Experiments in which OCC and FEC and other exogenous variables are assigned at random.
- Experiment 1: Daughters are assigned to various levels of AGE, ED observed.
- Experiment 2: Daughers are assigned to various levels of ED, then AGE is observed.
- Nature randomizes exogenous variables, then women choose ED and AGE as if solving two simultaneous equations.
- The equations estimated from survey data should also apply to experimental situations where ED and AGE are manipulated. These constancy assumptions are basis for causal inference from the non-experimental data.
- "Without the response schedules that embody the constancy assumptions, it is hard to see what "effects" migh mean."

Questions

- Assumptions about errors.
- Omitted variables.
- Why additive linear effects?
- Constant coefficients.
- Are exogenous variables really exogenous?
- Identifying restrictions?
- Are equations structural?