UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE DE RIBEIRÃO PRETO
DEPARTAMENTO DE ECONOMIA

Resumo: Equações Simultâneas

Felipe S. Costa

7 de maio de 2018

1 PROBLEMA

Para exemplificar a questão central do problema de EQUAÇÕES SIMULTÂNEAS podemos pensar em um problema de identificação de relações de oferta e demanda por algum bem. Veja que quando estamos modelando uma relação de preço e quantidade não temos a certeza, nem mesmo na teoria, de qual das duas variáveis é determinada primeiro (um problema de *causalidade*). Neste caso, o estimador de Mínimos Quadrados Ordinários (MQO) deixa de ser o estimador mais eficiente e terá, possivelmente, três fontes de viés:

1 viés de variável omitida;

2 viés causado por erros de medida;

3 viés de simultaneidade.

2 EXEMPLO

Vamos começar pensando em um problema de oferta de horas de trabalho. É razoável afirmar que um indivíduo ofertará mais horas de trabalho quanto maior for o salário/hora que enxerga no mercado de trabalho. Logo, se queremos medir a elasticidade do trabalho com relação ao trabalho, podemos propor a relação abaixo:

$$\ln h_i^S = \alpha_1 \ln w_i + \beta_1 z_{1i} + u_{1i}$$

em que h_i^S é o número de horas de trabalho ofertadas, w_i é o salário por hora trabalhada e $z_{1,i}$ é uma variável exógena hipotética que afeta apenas a relação de oferta (*observed supply shifter*). Essa é a nossa *equação estrutural de oferta* do trabalho.

Todavia, para descrever como ocorre o equilíbrio entre horas e salários, é preciso ter, além do lado da oferta, também o lado da demanda. Seja a *equação estrutural de demanda* do trabalho dada por

$$\ln h_i^D = \alpha_2 \ln w_i + \beta_2 z_{2i} + u_{2i}$$

sendo h_i^D quantidade de horas demandadas, z_2 uma variável exógena hipotética que afeta apenas de demanda por trabalho (*observed demand shifter*). u_1 e u_2 são os erros estruturais das nossas equações, que são denominados também como *unobservable supply shifter* e *unobservable demand shifter*, respectivamente. A condição de equilíbrio desse mercado é então dada por

$$h_i^S = h_i^D$$

mas como só observamos no mundo a quantidade de horas de equilíbrio, passamos a denotar as horas observadas como h_i. Assim, nossas equações estruturais passam a ser:

$$\begin{cases} \ln h_i = \alpha_1 \ln w_i + \beta_1 z_{1i} + u_{1i} \\ \ln h_i = \alpha_2 \ln w_i + \beta_2 z_{2i} + u_{2i} \end{cases}$$

O sistema de equações acima é o que chamamos de *Simultaneous Equations Model* (SEM). Como para dados z_{1i}, z_{2i}, u_{1i} e u_{2i}, as duas equações determinam h_i e w_i, estas duas variáveis são as variáveis endógenas do SEM.

3 Viés de Simultaneidade de MQO

Suponha um sistema que é definido por duas equações e que y_1 e y_2 são definidos simultaneamente por uma relação de equilíbrio econômico[1]:

$$y_{1i} = \alpha_1 y_{2i} + \beta_1 z_{1i} + u_{1i},\ u_{1i} \sim N(0, \sigma_{u_1}^2) \tag{3.1}$$

$$y_{2i} = \alpha_2 y_{1i} + \beta_2 z_{2i} + u_{2i},\ u_{2i} \sim N(0, \sigma_{u_2}^2) \tag{3.2}$$

Temos então um problema de causalidade reversa, sendo que y_1 e y_2 são variáveis endógenas e z_1 e z_2, exógenas. Substitua (3.1) em (3.2):

$$y_{2i} = \alpha_2(\alpha_1 y_{2i} + \beta_1 z_{1i} + u_{1i}) + \beta_2 z_{2i} + u_{2i} \tag{3.3}$$

$$y_{2i} - \alpha_1\alpha_2 y_{2i} = \alpha_2\beta_1 z_{1i} + \alpha_2 u_{1i} + \beta_2 z_{2i} + u_{2i} \tag{3.4}$$

$$y_{2i} = \frac{\alpha_2\beta_1 z_{1i} + \alpha_2 u_{1i} + \beta_2 z_{2i} + u_{2i}}{(1 - \alpha_1\alpha_2)} \tag{3.5}$$

[1] As equações estruturais não contêm intercepto para facilitar as manipulações.

$$y_{2i} = \frac{\alpha_2\beta_1}{1-\alpha_1\alpha_2} z_{1i} + \frac{\beta_2}{1-\alpha_1\alpha_2} z_{2i} + \underbrace{\frac{\alpha_2 u_{1i} + u_{2i}}{1-\alpha_1\alpha_2}}_{=v_{2i}} \tag{3.6}$$

Por hipótese, $\alpha_1\alpha_2 \neq 1$, o que garante que a equação reduzida acima é bem definida. Além disso, supõe-se que os erros das equações estruturais (u_1 e u_2) são não correlacionados. Assim, pode-se provar que, se $\alpha_2 \neq 0$ então $\mathrm{E}[u_1|y_2, z_1] \neq 0$, causado pelo chamado *viés de simultaneidade*.

Demonstração. Sendo z_1 e z_2 variáveis exógenas, por definição, estas não serão correlacionadas com os erros estruturais. Adicionalmente, por hipótese, os erros estruturais não são correlacionados entre si. Deste modo,

$$\begin{aligned}
\mathrm{cov}(y_2, u_1) &= \mathrm{cov}\left(\frac{\alpha_2\beta_1}{1-\alpha_1\alpha_2} z_1 + \frac{\beta_2}{1-\alpha_1\alpha_2} z_2 + \frac{\alpha_2 u_1 + u_2}{1-\alpha_1\alpha_2}, u_1\right) = \\
&= \mathrm{cov}\left(\frac{\alpha_2\beta_1}{1-\alpha_1\alpha_2} z_1, u_1\right) + \mathrm{cov}\left(\frac{\beta_2}{1-\alpha_1\alpha_2} z_2, u_1\right) + \mathrm{cov}\left(\frac{\alpha_2}{1-\alpha_1\alpha_2} u_1, u_1\right) + +\mathrm{cov}\left(\frac{1}{1-\alpha_1\alpha_2} u_2, u_1\right) \\
&= \frac{\alpha_2\beta_1}{1-\alpha_1\alpha_2} \mathrm{cov}(z_1, u_1) + \frac{\beta_2}{1-\alpha_1\alpha_2} \mathrm{cov}(z_2, u_1) + + + \frac{1}{1-\alpha_1\alpha_2} \mathrm{cov}(u_2, u_1) \\
&= \frac{\alpha_2}{1-\alpha_1\alpha_2} \mathrm{var}(u_1) = \frac{\alpha_2}{1-\alpha_1\alpha_2} \sigma^2_{u_1} \neq 0
\end{aligned} \tag{3.7}$$

Logo, generalizando, as hipóteses de $\mathrm{E}[u_1|y_2, z_1] = 0$ e $\mathrm{E}[u_2|y_1, z_2] = 0$ não se verificam. Então, MQO será viesado e inconsistente. □

4 Identificação das Equações Estruturais

Como demonstrado, o estimador de MQO será viesado e inconsistente para estimadar as equações estruturais do modelo hipotético de equações simultâneas. Então, como é possível estimar um modelo como esse? No exemplo usual de oferta e demanda, apenas é possível identificar a demanda usando deslocamentos da oferta, para assim obter sua inclinação a partir dos diversos pontos de equilíbrio em que a demanda não foi deslocada (sendo que é possível identificar a oferta, mantendo-a estática enquanto desloca-se a curva de demanda).

Seja então o sistema abaixo[2]:

$$\begin{cases} \mathbf{y}_1 = \beta_{10} + \alpha_1 \mathbf{y}_2 + \mathbf{Z}_1 \boldsymbol{\beta}_1 + \mathbf{u}_1 \text{ (I)} \\ \mathbf{y}_2 = \beta_{20} + \alpha_2 \mathbf{y}_1 + \mathbf{Z}_2 \boldsymbol{\beta}_2 + \mathbf{u}_2 \text{ (II)} \end{cases}$$

[2] $\mathbf{y}_1$, $\mathbf{y}_2$, $\mathbf{u}_1$ e $\mathbf{u}_2$ são vetores $(n \times 1)$; β_{10}, β_{20}, α_1 e α_2 são escalares; $\mathbf{Z}_1$ é uma matriz $(n \times k_1)$; e $\mathbf{Z}_2$ é uma matriz $(n \times k_2)$.

sendo que $\mathbf{Z}_1$ é conjunto de k_1 variáveis exógenas que explicam y_1, e $\mathbf{Z}_2$ é conjunto de k_2 variáveis exógenas que explicam y_2, sendo que podem ser reescritos como:

$$\begin{cases} \mathbf{Z}_1\boldsymbol{\beta}_1 = \beta_{11}\mathbf{z}_{11} + \beta_{12}\mathbf{z}_{12} + \cdots + \beta_{1k_1}\mathbf{z}_{1k_1} \\ \mathbf{Z}_2\boldsymbol{\beta}_2 = \beta_{21}\mathbf{z}_{21} + \beta_{22}\mathbf{z}_{22} + \cdots + \beta_{2k_2}\mathbf{z}_{2k_2} \end{cases}$$

Para a identificação das equações por meio de deslocamentos de uma das relações, mantendo a outra inalterada, é necessário satisfazer o que se chama de *restrição de exclusão.*

RESTRIÇÃO DE EXCLUSÃO: $\mathbf{Z}_1$ e $\mathbf{Z}_2$ precisam ter ao menos uma variável diferente (ou seja, ao menos uma das exógenas deve aparecer em apenas uma das equações).

CONDIÇÃO DE CLASSIFICAÇÃO PARA IDENTIFICAÇÃO DE UMA EQUAÇÃO ESTRUTURAL:

Identificação de (I): A equação (I) será identificada se, e somente se, a equação (II) contiver ao menos uma variável exógena excluída de (I), com coeficiente diferente de zero.

Identificação de (II): A equação (II) será identificada se, e somente se, a equação (I) contiver ao menos uma variável exógena excluída de (II), com coeficiente diferente de zero.

Note: É possível que se tenha em mãos casos em que apenas uma das equações seja identificada, como, por exemplo,

$$\begin{cases} \mathbf{y}_1 = \beta_{10} + \alpha_1\mathbf{y}_2 + \beta_{12}\mathbf{z}_2 + \beta_{13}\mathbf{z}_3 + \mathbf{u}_1 \text{ (I')} \\ \mathbf{y}_2 = \beta_{20} + \alpha_2\mathbf{y}_1 + \beta_{22}\mathbf{z}_2 + \mathbf{u}_2 \text{ (II')} \end{cases}$$

Nesse caso, apenas a equação (II') é identificada, pois $\mathbf{z}_3$ é uma variável exógena excluída a (II') e presente em (I'), sendo que como (I') não possui uma exógena excluída a ela, esta será não identificada.

5 ESTIMAÇÃO DE EQUAÇÕES SIMULTÂNEAS

A solução para esse tipo de problema é bastante simples e já familiar. Uma vez que o problema se dá pela quebra da hipótese de exogeneidade dos modelos de regressão linear múltipla, uma vez que $\text{cov}(y_2, u_1) \neq 0$, o estimador consistente para esse tipo de caso é o estimador de *Variáveis Instrumentais* (VI) ou de *Mínimos Quadrados em Dois Estágios* (MQ2E), sendo que os instrumentos w que são variáveis que, para a equação (I), atendem as restrições

$$\begin{cases} \text{cov}(y_2, u_1) = 0 \\ \text{cov}(y_2, w) \neq 0, \end{cases}$$

já possuem candidatos naturais, que são as variáveis presentes em $\mathbf{Z}_2$ e que não pertencem a $\mathbf{Z}_1$. Sendo que o paralelo é facilmente traçado para (II).

5.1 USANDO O STATA®

Vamos usar um exemplo do livro-texto para mostrar como estimar um regressões que sofrem do problema de simultaneidade.

EXEMPLO 16.5: Neste exemplo do livro-texto, são utilizados dados sobre mulheres casadas que trabalham fora, buscando estimar sua oferta de horas por MQ2E. As duas equações estruturais de oferta e demanda por trabalho são descritas no Exemplo 16.3, as quais são:

$$\text{hours} = \alpha_1 \ln(\text{wage}) + \beta_{10} + \beta_{11}\text{educ} + \beta_{12}\text{age} + \beta_{13}\text{kidslt6} + \beta_{14}\text{nwifeinc} + u_1 \tag{5.1}$$

e

$$\ln(\text{wage}) = \alpha_2\text{hours} + \beta_{20} + \beta_{21}\text{educ} + \beta_{22}\text{exper} + \beta_{23}\text{exper}^2 + u_2 \tag{5.2}$$

sendo que “kidslt6” é o número de filhos com menos de seis anos e “nwifeinc” são os rendimentos de outra pessoa família que não a mulher (inclui os ganhos do marido). Assim, pode-se obter a equação reduzida abaixo:

$$\ln(\text{wage}) = \pi_{20} + \pi_{21}\text{educ} + \pi_{22}\text{age} + \pi_{23}\text{kidslt6} + \pi_{24}\text{nwifeinc} + \pi_{25}\text{exper} + \pi_{26}\text{exper}^2 + \ + u_2 \tag{5.3}$$

Para a identificação de (5.1), é necessário que $\pi_{25} \neq 0$ e $\pi_{26} \neq 0$, o que já discutimos que pode ser testado usando uma estatística F padrão.

No Stata,

```
use http://fmwww.bc.edu/ec-p/data/wooldridge/mroz, clear

reg lwage educ age kidslt6 nwifeinc exper expersq

test exper expersq
```

O resultado do teste é então

```
.  test exper expersq
( 1) exper = 0
( 2) expersq = 0
F( 2, 421) = 9.33
Prob > F = 0.0001
```

Sendo possível rejeitar a hipótese nula do teste F, usaremos as variáveis exógenas excluídas à equação estrutural de oferta por trabalho das mulheres casadas para estimar esta. Assim, vamos para o Stata:

```
ivregress 2sls hours (lwage = exper expersq ) educ age
kidslt6 nwifeinc
```

E o resultado é dado abaixo:

	hours
lwage	1639.6***
	(3.51)
educ	-183.8**
	(-3.13)
age	-7.806
	(-0.84)
kidslt6	-198.2
	(-1.09)
nwifeinc	-10.17
	(-1.55)
Constant	2225.7***
	(3.90)
Observations	428

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

6 Observações Finais

Considerando o problema de estimação de equações simultâneas, qualquer equação que seja identificada pode ser estimada por MQ2E, sendo que as variáveis instrumentais são as variáveis exógenas excluídas que aparecem em qualquer outra equação do sistema. Além disso, dado que uma equação é identificada, pode-se realizar testes como os de endogeneidade, heterocedasticidade e de restrições de sobreidentificação (tópicos cujas aplicações no Stata foram abordados anteriormente).

Referências

[1] WOOLDRIDGE, J. M. *Introductory Econometrics: A modern approach*, 5^{th} Edition (2013). South Western, Cengage Learning. Chapter 16.