Aula 2.3: Equações Simultâneas

Econometria I - IBMEC

Marcus L. Nascimento

31 de outubro de 2025

Introdução

Introdução

- Ao estudarmos os modelos de regressão linear (simples e múltipla) e os modelos probit, tobit e logit, analisamos casos com uma única equação, ou seja, com uma única variável dependente.
  - Tais modelos descrevem uma relação unidirecional entre as variáveis (das variáveis explicativas para a variável resposta).
- Em muitos problemas, porém, a variável resposta é determinada por um conjunto de variáveis explicativas e algumas destas (endógenas) são determinadas pela variável resposta.
  - Há uma simultaneidade entre a variável resposta e alguns regressores endógenos;
  - Distinção entre variáveis dependentes e independentes torna-se complicada.

Introdução

- **Modelos de equações simultâneas** agrupam um conjunto de variáveis que podem ser determinadas simultaneamente por um conjunto restante de variáveis.
  - Múltiplas equações, uma para cada variável endógena;
  - A estimação dos parâmetros de uma equação deve considerar as informações presentes nas demais equações do sistema.

Observações:

- Na especificação do sistema, cada equação deve ter uma interpretação isolada.
- As variáveis endógenas devem ser definidas por agentes econômicos distintos.
- O fato de duas variáveis serem determinadas simultaneamente não implica que um modelo de equações simultâneas seja adequado.

Exemplo 1

- Podemos pensar em um problema no qual órgãos de governo querem investigar quanto um aparato adicional de aplicação da lei reduzirá a taxa de homicídios. Neste caso, é possível especificar o seguinte modelo:

$$murdpc = \alpha_1 polpc + \beta_{10} + \beta_{11} incpc + \varepsilon_1, \tag{1}$$

onde

  - *murdpc*: homicídios *per capita*;
  - *polpc*: número de policiais *per capita*;
  - *incpc*: renda *per capita*.

Exemplo 1

- Faz sentido pensar que o tamanho da força policial é determinado de maneira exógena?

Exemplo 1

- Faz sentido pensar que o tamanho da força policial é determinado de maneira exógena?
    - Provavelmente não. A despesa orçamentária destinada ao aparato de aplicação da lei é, ao menos parcialmente, determinada pela taxa de homicídio esperada.

Exemplo 1

- Faz sentido pensar que o tamanho da força policial é determinado de maneira exógena?
  - Provavelmente não. A despesa orçamentária destinada ao aparato de aplicação da lei é, ao menos parcialmente, determinada pela taxa de homicídio esperada.
- Para refletir tal endogeneidade, podemos especificar a seguinte relação:

$$polpc = \alpha_2 murdpc + \beta_{20} + outros\ fatores, \tag{2}$$

onde uma vez especificados os outros fatores, temos um modelo com duas equações simultâneas.

Exemplo 1

Observações:

- Note que não estamos interessados na Equação (2), mas é necessário especificá-la para estimar a Equação (1);
- A Equação (2) descreve o comportamento dos órgãos governamentais, enquanto a Equação (1) as ações de potenciais criminosos;
- Cada equação possui uma interpretação isolada.

Exemplo 2

- Suponha que para um domicílio aleatório na população, assumimos que as despesas anuais com habitação e poupança são determinadas conjuntamente da seguinte forma:

$$\begin{aligned} housing &= \alpha_1 saving + \beta_{10} + \beta_{11} inc + \beta_{12} educ + \beta_{13} age + \varepsilon_1 \\ saving &= \alpha_2 housing + \beta_{20} + \beta_{21} inc + \beta_{22} educ + \beta_{23} age + \varepsilon_2, \end{aligned}$$

  onde *inc* é a renda anual e *educ* e *age* são medidas em anos.

- À primeira vista, a especificação acima parece ser uma forma de entender como *housing* e *saving* são determinados. Porém ...
  - *housing* e *saving* não podem ser interpretadas isoladamente - ambas são definidas pelo mesmo agente;
  - A seguinte pergunta não faz sentido: Se a renda anual aumentar em R$ 10.000,00, como a despesa seria alterada mantendo a poupança fixa?

Viés de Simultaneidade

Viés de Simultaneidade (MQO)

- A partir de um exemplo simples, verificaremos que uma variável explicativa determinada simultaneamente com a variável dependente geralmente é correlacionada com o termo de erro.
  - Viés e inconsistência nos estimadores de mínimos quadrados ordinários.
- Considere o sistema de equações a seguir:

$$
\begin{aligned}
y_1 &= \alpha_1 y_2 + \beta_1 z_1 + \varepsilon_1 \\
y_2 &= \alpha_2 y_1 + \beta_2 z_2 + \varepsilon_2,
\end{aligned}
$$

onde o foco é estimar a primeira equação.

Viés de Simultaneidade (MQO)

Para mostrar que y_2 geralmente é correlacionado com ε_1, resolvemos as duas equações para y_2 em termos da variável exógena e do erro:

$$y_2 = \alpha_2(\alpha_1 y_2 + \beta_1 z_1 + \varepsilon_1) + \beta_2 z_2 + \varepsilon_2.$$

Logo,

$$(1 - \alpha_1\alpha_2)y_2 = \alpha_2\beta_1 z_1 + \beta_2 z_2 + \alpha_2\varepsilon_1 + \varepsilon_2,$$

que para $\alpha_1\alpha_2 \neq 0$ pode ser reescrita da seguinte forma:

$$y_2 = \frac{\alpha_2\beta_1}{1 - \alpha_1\alpha_2} z_1 + \frac{\beta_2}{1 - \alpha_1\alpha_2} z_2 + \frac{\alpha_2\varepsilon_1 + \varepsilon_2}{1 - \alpha_1\alpha_2}. \tag{3}$$

Viés de Simultaneidade (MQO)

A partir da Equação (3), temos

$$
\begin{aligned}
\operatorname{Cov}(y_2, \varepsilon_1) &= \operatorname{Cov}\left(\frac{\alpha_2\beta_1}{1-\alpha_1\alpha_2} z_1 + \frac{\beta_2}{1-\alpha_1\alpha_2} z_2 + \frac{\alpha_2\varepsilon_1 + \varepsilon_2}{1-\alpha_1\alpha_2}, \varepsilon_1\right) \\
&= \operatorname{Cov}\left(\frac{\alpha_2\varepsilon_1}{1-\alpha_1\alpha_2} + \frac{\varepsilon_2}{1-\alpha_1\alpha_2}, \varepsilon_1\right) \\
&= \frac{\alpha_2}{1-\alpha_1\alpha_2} \operatorname{Cov}(\varepsilon_1, \varepsilon_1) \\
&= \frac{\alpha_2}{1-\alpha_1\alpha_2} \operatorname{Var}(\varepsilon_1) \\
&\neq 0
\end{aligned}
$$

Identificação

Identificação

Suponha o seguinte sistema com duas equações:

$$\begin{aligned} y_1 &= \alpha_1 y_2 + \boldsymbol{z}_1' \boldsymbol{\beta}_1 + \varepsilon_1 && (4) \\ y_2 &= \alpha_2 y_1 + \boldsymbol{z}_2' \boldsymbol{\beta}_2 + \varepsilon_2, && (5) \end{aligned}$$

onde y_1 e y_2 são variáveis endógenas, ε_1 e ε_2 são termos de erro e $\boldsymbol{z}_1$ $\boldsymbol{z}_2$ são variáveis exógenas tais que

$$\begin{aligned} \boldsymbol{z}_1' \boldsymbol{\beta}_1 &= \beta_{11} z_{11} + \beta_{12} z_{12} + \ldots + \beta_{1k_1} z_{1k_1} \\ \boldsymbol{z}_2' \boldsymbol{\beta}_2 &= \beta_{21} z_{21} + \beta_{22} z_{22} + \ldots + \beta_{2k_2} z_{2k_2}. \end{aligned}$$

OBSERVAÇÃO: o fato de que $\boldsymbol{z}_1$ e $\boldsymbol{z}_2$ geralmente contêm conjuntos de variáveis exógenas diferentes significa que estamos impondo **restrições de exclusão** no modelo.

Identificação

Quando podemos escrever as Equações (4) e (5) para y_1 e y_2 como funções lineares de todas as variáveis exógenas e dos erros ε_1 e ε_2 (**forma reduzida**)?

A condição é a mesma que vimos anteriormente, ou seja, $\alpha_1\alpha_2 \neq 0$.

A partir da forma reduzida, sob quais condições, conseguimos recuperar os parâmetros do modelo original?

Esta é a pergunta-chave sobre identificação.

Quando tal recuperação não é possível, dizemos que o sistema em pauta é **identificado**. Caso a recuperação não seja possível, dizemos que o sistema é **não identificado** (ou **subidentificado**).

Condição de ordem

- Em um modelo de equações simultâneas com duas equações, a **identificação** é obtida através das seguintes condições:
  - A primeira equação é identificada se ...
    - ... pelo menos uma das variáveis exógenas é excluída da equação;
    - ... dentre as variáveis excluídas, pelo menos uma tem coeficiente diferente de zero.
  - A segunda equação será identificada sob as mesmas condições acima, porém considerando variáveis exógenas excluídas desta.

Exemplo

Como exemplo, considere a oferta de mão de obra para mulheres casadas que já estão no mercado de trabalho. Suponha as seguintes equações:

$$\begin{aligned} hours &= \alpha_1 \ln(wage) + \beta_{10} + \beta_{11} educ + \beta_{12} age + \beta_{13} kidslt6 + \beta_{14} nwifeinc + \varepsilon_1 \\ \ln(wage) &= \alpha_2 hours + \beta_{20} + \beta_{21} educ + \beta_{22} exper + \beta_{23} exper^2 + \varepsilon_2, \end{aligned}$$

onde a primeira equação é a função de demanda e a segunda, a função de oferta. Além disso,

educ: anos de escolaridade;

age: idade em anos;

kidslt6: número de filhos com menos de 6 anos de idade;

nwifeinc: renda da mulher não oriunda do trabalho;

exper: anos de experiência prévia.

Exemplo

- Considerando a primeira equação, as variáveis *exper* e $exper^2$ não aparecem na mesma, então, se β_{22} ou β_{23} forem diferentes de zero, a mesma será identificada.

- A segunda equação será identificada se pelo menos um dos coeficientes relativos às variáveis exógenas *age*, *kidslt*6 e *nwifeinc* for diferente de zero.