Aula – Econometria

Cláudio R. Lucinda

FEA/USP

2020

Objetivos da Aula

1. Estimação de Sistemas de Equações
   - OLS
   - SUR
   - Mínimos Quadrados Generalizados Factíveis
   - Mínimos Quadrados a Dois Estágios
   - Mínimos Quadrados a Três Estágios
   - GMM em Sistema

Estimação de Sistemas de Equações

- A estimação de sistemas de equações de demanda, qualquer que seja a especificação, possui uma estrutura comum:

$$\begin{aligned}
\mathbf{y}_1 &= \mathbf{X}_1(\beta_1) + \epsilon_1 \\
\mathbf{y}_2 &= \mathbf{X}_2(\beta_2) + \epsilon_2 \\
&\vdots \\
\mathbf{y}_M &= \mathbf{X}_M(\beta_M) + \epsilon_M
\end{aligned}$$

- Existem M equações, em que as variáveis em cada uma delas possuem T observações.
- Escolhemos por denotar a parte independente das equações de demanda por $\mathbf{X}(\beta)$ para enfatizar o fato que, na maior parte das vezes, estaremos lidando com uma função não linear envolvendo os regressores $\mathbf{X}$ e os coeficientes β.
    - Além disso, esta função é diferente – envolvendo diferentes coeficientes, por exemplo – em cada uma da equações.

Sistemas de Equação

- Iremos, neste texto, enfatizar a similaridade das técnicas de estimação e a sua base comum no Método Generalizado dos Momentos.
- Para começarmos, vamos introduzir uma notação adicional.
  - Seja $\mathbf{y}_\bullet$ o vetor de dimensão $MT \times 1$ obtido a partir da "vetorização" dos pequenos vetores $\mathbf{y}_i$.
  - Da mesma forma, iremos denotar $\mathbf{X}_\bullet(\beta)$ o vetor obtido a partir do "empilhamento" dos lados direitos da igualdade,
  - $\epsilon_\bullet$ o vetor obtido a partir do "empilhamento" dos resíduos de cada uma das regressões.
- Desta forma, o que vimos no slide anterior fica sendo igual a:

$$\mathbf{y}_\bullet = \mathbf{X}_\bullet(\beta) + \epsilon_\bullet$$

OLS

- Neste caso, podemos escrever a i-ésima equação do sistema da seguinte forma:

$$\begin{aligned} \mathbf{y}_i &= \mathbf{X}_i(\beta) + \epsilon_i \\ E(\epsilon_i \epsilon_i^T) &= \sigma_{ii} \mathbf{I}_T \end{aligned}$$

- Em que $\mathbf{I}_T$ denota a matriz identidade de dimensão $T \times T$. Além disso, precisamos da seguinte condição:

$$E(\epsilon_i \epsilon_j^T) = \mathbf{0}, \forall j \neq i$$

- Em que $\mathbf{0}$ denota a matriz zero de dimensão $T \times T$. Neste caso, podemos construir a matriz variância-covariância do sistema:

$$E(\epsilon_\bullet \epsilon_\bullet^T) = \mathbf{V} = \begin{bmatrix} \sigma_{11}\mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \sigma_{22}\mathbf{I} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \sigma_{MM}\mathbf{I} \end{bmatrix}$$

OLS (cont.):

- Neste caso, podemos estimar os coeficientes do sistema se estimarmos equação por equação por OLS, caso não tivéssemos restrições dos coeficientes entre as equações.
- A estimação se daria por meio da minimização da função critério – que no caso seria a soma dos quadrados dos resíduos, ponderada pela inversa da matriz de variância-covariância dos resíduos:

$$SSR_{\bullet} = (\mathbf{y}_{\bullet} - \mathbf{X}_{\bullet}(\beta))^{T}(\mathbf{\Sigma}^{-1} \otimes \mathbf{I})(\mathbf{y}_{\bullet} - \mathbf{X}_{\bullet}(\beta))$$

- Em que $\mathbf{\Sigma}$ representa a matriz de variância-covariância dos resíduos, que no caso em questão possuem a seguinte forma:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1M} = \mathbf{0} \\ \vdots & \ddots & \vdots \\ \sigma_{M1} = \mathbf{0} & \cdots & \sigma_{MM} \end{bmatrix}$$

OLS (III):

- Geralmente são estimados os membros desta matriz com os resíduos da regressão, ou seja - $\sigma_{11} = \epsilon_{\mathbf{1}}^{\mathbf{T}} \epsilon_1$Com isto, temos como estimativa de variância de β:

$$\hat{Var}(\beta) = (\mathbf{Xd^T}_{\bullet}(\beta)(\mathbf{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{Xd}_{\bullet}(\beta))^{-1}$$

- Em que $\mathbf{Xd}_{\bullet}(\beta)$ é uma matriz de derivadas parciais de $\mathbf{X}_{\bullet}(\beta)$ com relação a cada um dos coeficientes, e $\otimes$ denota o Produto Kronecker.
    - Se estivéssemos com equações lineares nos coeficientes, $\mathbf{Xd}_{\bullet}(\beta) = \mathbf{X}_{\bullet}$.

SUR

- A diferença entre este método e o anterior reside na estrutura de correlação entre os erros das equações.
- A idéia desta técnica é que, mesmo que tenhamos as variáveis independentes e as variáveis dependentes sendo completamente diferentes entre as equações, a correlação entre os resíduos da regressão pode ser diferente, de forma que estimar isto por Mínimos Quadrados Ordinários não dá certo.
    - Daí o nome de *Seemingly Unrelated Regressions* (Regressões Aparentemente Não Relacionadas)
- Neste caso, temos que a matriz $\boldsymbol{\Sigma}$ tem a seguinte estrutura:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1M} \\ \vdots & \ddots & \vdots \\ \sigma_{M1} & \cdots & \sigma_{MM} \end{bmatrix}$$

SUR (Cont.)

- O que nos dá a seguinte matriz de variância-covariância dos resíduos empilhados:

$$E(\epsilon_{\bullet}\epsilon_{\bullet}^{T}) = \mathbf{V} = \mathbf{\Sigma} \otimes \mathbf{I}_T$$

Temos, neste caso, a mesma função critério a ser minimizada:

$$SSR_{\bullet} = (\mathbf{y}_{\bullet} - \mathbf{X}_{\bullet}(\beta))^{T}(\mathbf{\Sigma}^{-1} \otimes \mathbf{I})(\mathbf{y}_{\bullet} - \mathbf{X}_{\bullet}(\beta))$$

- A matriz de variância-covariância dos cieficientes neste caso é similar também:

$$Va\hat{r}(\beta) = (\mathbf{Xd^{T}}_{\bullet}(\beta)(\mathbf{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{Xd}_{\bullet}(\beta))^{-1}$$

FGLS System

- Neste caso, a estrutura é bastante similar, o que mudaremos é a forma pela qual a nossa já queria matriz $\mathbf{\Sigma}$ é calculada.
- Neste caso, teríamos uma situação em que cada um dos elementos da matriz $\mathbf{\Sigma}$ seria calculada por um método como o de Newey-West, em que σ_{ii} seria uma soma ponderada das variâncias, das covariâncias entre os resíduos entre t e $t-1$, t e $t-2$, e assim por diante.
- Mais especificamente:

$$
\begin{aligned}
\sigma_{ii} &= \sum_{j=-n+1}^{n-1} k(x) \times \hat{\mathbf{\Gamma}}_{ij} \\
k(x) &= \begin{cases} 1-|x| & para\, |x| \leq 1 \\ 0 & c.c. \end{cases} \\
\hat{\mathbf{\Gamma}}_{ij} &= \epsilon_i \epsilon_{i(-j)} \\
x &= \frac{j}{q}
\end{aligned}
$$

FGLS System

- Neste caso, a estrutura é bastante similar, o que mudaremos é a forma pela qual a nossa já queria matriz $\mathbf{\Sigma}$ é calculada.
- Neste caso, teríamos uma situação em que cada um dos elementos da matriz $\mathbf{\Sigma}$ seria calculada por um método como o de Newey-West, em que σ_{ii} seria uma soma ponderada das variâncias, das covariâncias entre os resíduos entre t e $t-1$, t e $t-2$, e assim por diante.
- Mais especificamente:

$$
\begin{aligned}
\sigma_{ii} &= \sum_{j=-n+1}^{n-1} k(x) \times \hat{\mathbf{\Gamma}}_{ij} \\
k(x) &= \begin{cases} 1-|x| & para\,|x| \leq 1 \\ 0 & c.c. \end{cases} \\
\hat{\mathbf{\Gamma}}_{ij} &= \epsilon_i \epsilon_{i(-j)} \\
x &= \frac{j}{q}
\end{aligned}
$$

1. Sendo que q é a chamada largura de banda *bandwidth*

TSLS

- Neste caso, temos uma diferença em relação aos casos anteriores, em que temos entre os regressores uma – ou mais de uma – variável que se considera como sendo determinada conjuntamente com a variável dependente.
- Vamos começar supondo um modelo linear, em que a i -ésima equação possui a seguinte forma:

$$\mathbf{y}_i = \mathbf{z}_i \delta + \epsilon_i$$

- Em que $\mathbf{z}_i = [\mathbf{Y_i} | \mathbf{x_i}]$. Ou seja, juntamos as variáveis que são consideradas como endógenas ($\mathbf{Y_i}$) e as que são consideradas como exógenas ($\mathbf{x_i}$).

TSLS (III):

- Supondo que $\mathbf{x}_i$ inclua todas as variáveis exógenas e que estas variáveis exógenas sejam suficientes para o atendimento das condições de posto e condições de ordem do sistema de equações, temos que podemos obter a estimativa dos coeficientes do sistema por meio das chamadas "condições de ortogonalidade":

$$\mathbf{m}(\delta) = \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t(\mathbf{y}_i - \mathbf{z}_i\delta)$$

- A idéia é escolher os coeficientes de forma a minimizar este valor para torná-lo o mais próximo de zero possível.
- Supondo que tenhamos mais instrumentos do que variáveis endógenas no sistema – ou seja, ele é sobre-identificado – este sistema nunca dá exatamente zero.

TSLS (Cont.):

- Para resolver isso, geralmente se escolhe o vetor de coeficientes que minimiza a seguinte função objetivo:

$$q = \mathbf{m}(\delta)^T W^{-1} \mathbf{m}(\delta)$$

Na verdade isto aqui é uma forma quadrática, com uma ponderação dada pela variância dos resíduos. Se os resíduos da equação são homoscedásticos e não autocorrelacionados, a matriz $W = (\mathbf{X}^T\mathbf{X})$nos dá os resultados do Método dos Mínimos Quadrados a Dois Estágios:

$$\begin{aligned} \delta_{TSLS} &= (\mathbf{X}^T(\mathbf{I} - \mathbf{M_X})\mathbf{Z})^{-1}\mathbf{X}^T(\mathbf{I} - \mathbf{M_X})\mathbf{y} \\ (\mathbf{I} - \mathbf{M_X}) &= \mathbf{X^T}(\mathbf{X^T}\mathbf{X})^{-1}\mathbf{X} \end{aligned}$$

MQ3E

- Vamos agora começar a discutir os métodos de estimação que consideram a informação de todos os resíduos do sistema. O primeiro destes modelos é o método dos Mínimos Quadrados a Três Estágios, em que supomos que a estrtura de covariância entre os resíduos das equações é a seguinte:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1M} \\ \vdots & \ddots & \vdots \\ \sigma_{M1} & \cdots & \sigma_{MM} \end{bmatrix}$$

- O que nos dá a seguinte matriz de variância-covariância dos resíduos empilhados:

$$E(\epsilon_{\bullet}\epsilon_{\bullet}^{T}) = \mathbf{V} = \mathbf{\Sigma} \otimes \mathbf{I}_T$$

MQ3E (Cont.)

- Com esta estrutura de resíduos, podemos calcular os estimadores de Mínimos Quadrados a Três Estágios da seguinte forma:

$$\begin{aligned} \beta_{3SLS} &= (\hat{\mathbf{Z}}_{\bullet}(\mathbf{\Sigma}^{-1} \otimes \mathbf{I})\hat{\mathbf{Z}}_{\bullet})^{-1}\hat{\mathbf{Z}}_{\bullet}(\mathbf{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{y} \\ \hat{\mathbf{Z}} &= \mathbf{X}^{\mathbf{T}}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}\mathbf{Z} \end{aligned}$$

- Neste caso, a matriz $\mathbf{\Sigma}$ é calculada com os resíduos de Mínimos Quadrados a Dois Estágios calculados equação por equação. Na verdade esta é uma aplicação do caso mais geral do Método Generalizado dos Momentos com equações não-lineares, que é o que veremos mais adiante.

GMM em Sistema

- Finalmente, passamos ao método mais amplo de todos, o Método Generalizado dos Momentos para o sistema de equações. a idéia neste caso começa com o seguinte conjunto de condições de ortogonalidade para cada uma das equações do sistema:
$$\mathbf{m}_i(\delta) = \mathbf{Z}^T(\mathbf{Y}_i - \mathbf{X}_i(\beta))$$
- Em que $\mathbf{Z}$ denota o conjunto de instrumentos utilizado. A partir destas condições de ortogonalidade, podemos definir a seguinte função a ser minimizada:

$$q = \sum_{i=1}^{M}\sum_{j=1}^{M}((\mathbf{Y_i} - \mathbf{X_i}(\beta))^T\mathbf{Z})(\mathbf{W})^{ij}(\mathbf{Z}^T(\mathbf{Y_i} - \mathbf{X_i}(\beta)))$$

- Nesta equação, o único termo ainda não visto é o $(\mathbf{W})^{ij}$, que seria a inversa da covariância entre os momentos da equação i e os momentos da equação j.

GMM em Sistema (Cont.)

- Ou ainda, podemos abrir o sistema da seguinte forma:

$$
q = \begin{bmatrix} (\mathbf{y}_1 - \mathbf{X}_1(\beta))^T\mathbf{Z} \\ \vdots \\ (\mathbf{y}_M - \mathbf{X}_M(\beta))^T\mathbf{Z} \end{bmatrix}^T \begin{bmatrix} \sigma_{11}(\mathbf{Z}^T\mathbf{Z}) & \cdots & \sigma_{1M}(\mathbf{Z}^T\mathbf{Z}) \\ \vdots & \ddots & \vdots \\ \sigma_{M1}(\mathbf{Z}^T\mathbf{Z}) & \cdots & \sigma_{MM}(\mathbf{Z}^T\mathbf{Z}) \end{bmatrix}
\begin{bmatrix} \mathbf{Z}^T(\mathbf{y}_1 - \mathbf{X}_1(\beta)) \\ \vdots \\ \mathbf{Z}^T(\mathbf{y}_M - \mathbf{X}_M(\beta)) \end{bmatrix}
$$

- Estamos supondo aqui que os resíduos em cada uma das equações são homocedásticos e não autocorrelacionados.

GMM em Sistema (Cont.)

- Ou ainda, podemos abrir o sistema da seguinte forma:

$$q = \begin{bmatrix} (\mathbf{y}_1 - \mathbf{X}_1(\beta))^T\mathbf{Z} \\ \vdots \\ (\mathbf{y}_M - \mathbf{X}_M(\beta))^T\mathbf{Z} \end{bmatrix}^T \begin{bmatrix} \sigma_{11}(\mathbf{Z}^T\mathbf{Z}) & \cdots & \sigma_{1M}(\mathbf{Z}^T\mathbf{Z}) \\ \vdots & \ddots & \vdots \\ \sigma_{M1}(\mathbf{Z}^T\mathbf{Z}) & \cdots & \sigma_{MM}(\mathbf{Z}^T\mathbf{Z}) \end{bmatrix} \begin{bmatrix} \mathbf{Z}^T(\mathbf{y}_1 - \mathbf{X}_1(\beta)) \\ \vdots \\ \mathbf{Z}^T(\mathbf{y}_M - \mathbf{X}_M(\beta)) \end{bmatrix}$$

- Estamos supondo aqui que os resíduos em cada uma das equações são homocedásticos e não autocorrelacionados.

1. Em geral, ele é mais vantajoso do que o Método dos Mínimos Quadrados a Três Estágios porque tem ganhos de eficiência na presença de correlação e heterocedasticidade (a matriz do meio pode ser ajeitada com o uso do estimador de Newey-West para dar estimativas robustas).