GMM e Análise Longitudinal

(rascunho de notas de aula)

Victor Gomes

Universidade de Brasilia

15/04/2019

GMM

- Hipótese mais importante de OLS: ortogonalidade entre termo de erro e regressores. Sem essa hipótese OLS não é nem mesmo consistente.

- Em diversas aplicações em economia não garantimos a condição de ortogonalidade. Estimação por GMM é fundamental e (quase) obrigatória.

- GMM inclui como casos especiais OLS, IV, regressão multivariada e 2SLS.

Viés de Endogeneidade

- Clássico exemplo de Working (1927)

- Exemplo para mercado intermediário de café - commodity

- Modelo simples de oferta e demanda

$$\begin{aligned} q_i^d &= \alpha_0 + \alpha_1 p_i + u_i && (1) \\ q_i^s &= \beta_0 + \beta_1 p_i + v_i && (2) \\ q_i^d &= q_i^s && (3) \end{aligned}$$

- q = quantidade, p = preço. Indicadoras são: tempo i, demanda d, e oferta s. u é o termo de erro da demanda e v o termo de erro da oferta.

Viés de Endogeneidade

- Vamos assumir $E(u_i) = 0$ e $E(v_i) = 0$, e por simplicidade, $Cov(u_i, v_i) = 0$.

- Fazendo $q_i^d = q_i^s = q_i$, temos:

$$q_i = \alpha_0 + \alpha_1 p_i + u_i \quad (4)$$

$$q_i = \beta_0 + \beta_1 p_i + v_i \quad (5)$$

Viés de Endogeneidade

- Dissemos que um *regressor é endógeno* se ele não é predeterminado, i.e. não é ortogonal ao erro.

- Em nosso caso, p_i é endógeno em ambas as equações:

$$p_i = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{v_i - u_i}{\alpha_1 - \beta_1} \quad (6)$$

$$q_i = \frac{\alpha_1 \beta_0 - \alpha_0 \beta_1}{\alpha_1 - \beta_1} + \frac{\alpha_1 v_i - \beta_1 u_i}{\alpha_1 - \beta_1} \quad (7)$$

Viés de Endogeneidade

- Desse sistema de equações podemos calcular a covariância de p_i:

$$\mathsf{Cov}(p_i, u_i) = \frac{-\mathsf{Var}(u_i)}{\alpha_1 - \beta_1},\ \mathsf{Cov}(p_i, u_i) = \frac{-\mathsf{Var}(v_i)}{\alpha_1 - \beta_1} \qquad (8)$$

- Observe que as covariâncias não são zero. Nesse caso a endogeneidade é resultado do equilíbrio de mercado.

Viés de Endogeneidade

- Quando rodados uma regressão de quantidade contra preço e uma constante, estamos estimando a oferta ou demanda? Como preço é endógeno em ambas equações a resposta não é clara.

- Da projeção do OLS sabemos que:

$$\text{plim da estimativa de OLS de } p_i = \frac{\text{Cov}(p_i, q_i)}{\text{Var}(p_i)} \tag{9}$$

- Para explicitar o efeito-preço na curva de demanda use $q_i = \alpha_0 + \alpha_1 p_i + u_i$ e faça:

$$\text{Cov}(p_i, q_i) = \alpha_1 \text{Var}(p_i) + \text{Cov}(p_i, u_i) \tag{10}$$

Viés de Endogeneidade

- Substituindo (10) em (9), temos o **viés assintótico** para α_1:

$$\text{plim da estimativa de OLS de } p_i - \alpha_1 = \frac{\text{Cov}(p_i, u_i)}{\text{Var}(p_i)} \tag{11}$$

- Similarmente o **viés assintótico** para β_1 é dado por:

$$\text{plim da estimativa de OLS de } p_i - \beta_1 = \frac{\text{Cov}(p_i, v_i)}{\text{Var}(p_i)} \tag{12}$$

- Então como a covariância não é zero, a estimativa OLS não é consistente para α_1 e para β_1.

Viés de Endogeneidade

- Isto é conhecido como **viés de endogeneidade**.

- Porque o erro e o regressor são relacionados por meio de um sistema de equações simultâneas, esse viés também é conhecido como **viés de equações simultâneas** ou **viés de simultaniedade**.

- Sem viés o plim da estimativa OLS do coeficiente de preço seria:

$$= \frac{\alpha_1 \text{Var}(v_i) + \beta_1 \text{Var}(u_i)}{\text{Var}(v_i) + \text{Var}(u_i)}$$

Mudanças na oferta

- Como identificar se as mudanças nos preços são devido a alterações na curva de oferta ou na demanda?

- Suponha que uma mudança na oferta v_i possa ser divida em um fator observável x_i e em um fator não observável ζ_i não-correlacionado com x_i. Assim:

$$q_i = \beta_0 + \beta_1 p_i + \beta_2 x_i + \zeta_i \text{ com } \beta_2 \neq 0 \tag{13}$$

- Imagine que esta variável x_i que provoca mudanças na curva de oferta é predeterminada (i.e. não-correlacionada com o termo do erro)

Mudanças na oferta

- Na equação em questão, uma variável predeterminada que é correlacionada com o regressor endógeno é chamada de *variável instrumental* ou um *instrumento*. Em nosso exemplo, o deslocador da curva de oferta x_i pode servir como um instrumento para a equação de demanda.

- Usando equação (13) e (7) temos:

$$p_i = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{\beta_2}{\alpha_1 - \beta_1} x_i + \frac{\zeta_i - u_i}{\alpha_1 - \beta_1} \tag{14}$$

$$q_i = \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1} + \frac{\alpha_1\beta_2}{\alpha_1 - \beta_1} x_i + \frac{\alpha_1\zeta_i - \beta_1 u_i}{\alpha_1 - \beta_1} \tag{15}$$

Mudanças na oferta: condições do instrumento

- Como $\text{Cov}(x_i, \zeta_i) = 0$ por construção e $\text{Cov}(x_i, u_i) = 0$ por hipótese (no momento).

- A partir da equação $p_i = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{\beta_2}{\alpha_1 - \beta_1} x_i + \frac{\zeta_i - u_i}{\alpha_1 - \beta_1}$

$$\text{Cov}(x_i, p_i) = \frac{\beta_2}{\alpha_1 - \beta_1} \text{Var}(x_i) \neq 0 \tag{16}$$

- Estas são as condições de validade de um instrumento

Mudanças na oferta

- Com um instrumento válido podemos estimar α_1 consistentemente.

- Usando a curva de demanda podemos calcular $Cov(x_i, q_i)$:

$$Cov(x_i, q_i) = \alpha_1 Cov(x_i, p_i) + Cov(x_i, u_i) = \alpha_1 Cov(x_i, p_i)$$

  dado que $Cov(x_i, u_i) = 0$.

Mudanças na oferta

- Dividindo ambos os lados por $Cov(x_i, p_i)$, temos

$$\alpha_1 = \frac{Cov(x_i, q_i)}{Cov(x_i, p_i)}$$

  Assim um estimador natural de α_1 é:

$$\widehat{\alpha}_{1,IV} = \frac{\text{covariancia amostral entre } x_i \text{ e } q_i}{\text{covariancia amostral entre } x_i \text{ e } p_i}$$

- Este é o estimador de variável instrumental (IV) com x_i como instrumento.

2SLS

- Outro estimador popular é o de mínimos quadrados de dois estágios (2SLS).

- (Estágio 1) regressão de p_i contra x_i

- (Estágio 2) regressão de q_i contra $\widehat{p}_i$ (onde $\widehat{p}_i$ é o valor predito no primeiro estágio).

- Então o estimador 2SLS pode ser escrito da seguinte forma:

$$\widehat{\alpha}_{1,2SLS} = \frac{\text{covariancia amostral entre } \widehat{p}_i \text{ e } q_i}{\text{variância amostral de } \widehat{p}_i}$$

- Para relacionar a equação de demanda com o 2SLS:

$$q_i = \alpha_0 + \alpha_1 \widehat{p}_i + [u_i + \alpha_1(p_i - \widehat{p}_i)]$$

  A estimativa de α_1 é consistente pela seguinte razão: se o valor previsto $\widehat{p}$ é exatamente igual a projeção de mínimos quadrados $\widehat{E}^*(p|1, x)$, então nem u_i nem $(p_i - \widehat{p}_i)$ devem ser correlacionados com $\widehat{p}_i$. u_i não é correlacionado porque não é correlacionado com x_i e $\widehat{p}_i$ é uma função linear de x_i e $(p_i - \widehat{p}_i)$ é não correlacionado porque é uma projeção de mínimos quadrados do erro. Esses resultados valem quando $N \to \infty$

- No caso destes exemplos os estimadores IV e 2SLS são os numericamente os mesmos. O estimador IV é um caso de GMM.

Exemplos: Modelo Macro

- Modelo de consumo de Haavelmo (1943)

$$C_i = \alpha_o + \alpha_1 Y_i + u_i \tag{17}$$

$$Y_i = C_i + I_i \tag{18}$$

- Variáveis: C_i consumo agregado; Y_i é PNB; α_i é a propensão marginal a consumir ($0 < \alpha_1 < 1$).

- GNP em equilíbrio é:

$$Y_i = \frac{\alpha_0}{1-\alpha_1} + \frac{I_i}{1-\alpha_i} + \frac{u_i}{1-\alpha_i} \tag{19}$$

Exemplos: Modelo Macro

- Se o investimento é predeterminado, i.e., Cov$(I_i, u_i) = 0$ segue que

$$\text{Cov}(Y_i, u_i) = \frac{Var(u_i)}{1 - \alpha_i} > 0$$

$$\text{Cov}(Y_i, I_i) = \frac{Var(I_i)}{1 - \alpha_i} > 0$$

- Nesse caso, Y_i é endógena por meio da função consumo. O investimento é um instrumento válido para o regressor endógeno.

- O plim da estimativa de OLS da equação de consumo é assintoticamente viesado:

$$\text{plim}\widehat{\alpha}_{1,OLS} - \alpha_1 = \frac{\text{Cov}(Y_i, u_i)}{\text{Var}(Y_i)} = \frac{1-\alpha_1}{1 + \frac{Var(I_i)}{Var(u_i)}} > 0$$

- O viés assintótico pode ser corrigido pelo uso do investimento como instrumento para a renda. Aqui se usa o mesmo papel do descolador de oferta no exemplo de Working.

Exemplos: Função de Produção

- Em muitas aplicações é comum se aplicar uma hipótese sobre a estrutura de informação. Neste exemplo, a idéia fundamental é que a empresa possui mais informações do que o economista. Portanto, o que não é conhecido quando estimamos um modelo pode ser conhecido no mercado.

- A endogeneidade surge quando regressores são decisões realizadas pelo agente com base em fatores que não são observáveis por quem estima.

Exemplos: Função de Produção

- Suponha uma amostra cross-section onde firmas escolhem o insumo trabalho para maximizar seus lucros. A função de produção para a firm i é:

$$Q_i = A_i(L_i)^{\phi_1} \exp(v_i), 0 < \phi_1 < 1 \qquad (20)$$

- Tal que Q_i é o produto, L_i é o trabalho contratado, A_i é o nível de eficiência da conhecido pela firma, e v_i é o choque tecnológico. Em contrate a A_i, v_i não é observado pela firma quando ela escolhe L_i. Nem A_i nem v_i é observável pelo econometrista.

- Assuma que $B = E[\exp(v_i)]$ é o mesmo para todas as firmas (ver hipóteses no texto). O nível de produto que as firmas esperam quando escolhem L_i é

$$A_i(L_i)^{\phi_1}B$$

- Assuma indústria competitiva. Então p (preço de Q_i) e w (salário) são constantes entre as firmas i. O objetivo da firma é maximizar lucros

$$\max_{L_i}\left\{pA_i(L_i)^{\phi_1}B - wL_i\right\}$$

solucionando para L_i temos:

$$L_i = \left(\frac{w}{p}\right)^{\frac{1}{\phi_1 - 1}}(A_i B\phi_1)^{\frac{1}{1-\phi_1}}$$

Exemplos: Função de Produção - Viés

- Faça u_i ser o desvio da firma i da média do log da eficiência: $u_i = log(A_i) - E[\log(A_i)]$. Faça $\phi_0 = E[\log(A_i)]$ $(E(u_i) = 0 e A_i = \exp(\phi_0 + u_i))$. As equações para estimação são as seguintes:

$$\log(Q_i) = \phi_0 + \phi_1 \log(L_i) + (v_i + u_i) \tag{21}$$

$$\log(L_i) = \beta_0 + \frac{1}{1 - \phi_1} u_i \tag{22}$$

tal que $\beta_0 = \frac{1}{\phi_1 - 1}[\log(w/p) - \phi_0 - \log(\phi_1 B)]$

- A estimativa OLS de ϕ_1 é viesada. Isto ocorre porque L_i é um regressor endógeno que é relacionado com o termo de erro

$(v_i + u_i)$ por meio de u_i. Este exemplo ilustra outra fonte de endogeneidade: *a variável escolhida pelo agente possui um termo de erro que não é conhecido pelo econometrista*.

Formulação Geral

Hip. 3.1 (linearidade): $y_i = \boldsymbol{z}_i'\boldsymbol{\delta} + \varepsilon_i \qquad$ (i = 1,2,...). $\boldsymbol{z}$ é o vetor de regressores, $\boldsymbol{\delta}$ é o vetor dos coeficientes $(L \times 1)$.

Hip. 3.2 (estacionaridade ergódica): Seja $\boldsymbol{x}_i$ o vetor dos instrumentos (dimensão K), e faça $\boldsymbol{w}_i$ serem os elementos únicos e não-constantes de $(y_i, \boldsymbol{z}_i, \boldsymbol{x}_i)$. $\{\boldsymbol{w}_i\}$ é conjuntamente estacionária e ergódica.

Hip. 3.3 (regressores pré-determinados): Todas as variáveis em $\boldsymbol{x}_i$ são pré-determinadas no sentido de que todas são ortogonais ao erro: $\mathrm{E}(x_{ik}\varepsilon_i) = 0$ para todo i e $k(= 1, 2, ..., K)$.

Ou seja: $\mathrm{E}[\boldsymbol{x}_i(y_i - \boldsymbol{z}_i'\boldsymbol{\delta})] = 0$ ou $\mathrm{E}(\boldsymbol{g}_i) = 0$, tal que $\boldsymbol{g}_i \equiv \boldsymbol{x}_i\varepsilon_i$.

Formulação Geral

Exemplo 3.2 (equação de salário): Considere a seguinte equação de salário:

$$\log W_i = \delta_1 + \delta_2 S_i + \delta_3 EXPR_i + \delta_4 IQ_i + \varepsilon_i,$$

Então

$$y_i = \log W_i, L = 4,$$

$$\boldsymbol{z}_i = \begin{bmatrix} 1 \\ S_i \\ EXPR_i \\ IQ_i \end{bmatrix}, K = 5, \boldsymbol{x}_i = \begin{bmatrix} 1 \\ S_i \\ EXPR_i \\ AGE_i \\ MED_i \end{bmatrix}, \boldsymbol{w}_i = \begin{bmatrix} \log W_i \\ S_i \\ EXPR_i \\ IQ_i \\ AGE_i \\ MED_i \end{bmatrix}$$

Identificação

Um instrumento deve ser predeterminado e correlacionado com os regressores.

Hipótese 3.4 (condição de posto para identificação): A matriz $\mathsf{E}(\boldsymbol{x}_i \boldsymbol{z}_i')$ é posto-completo. Representamos por $\boldsymbol{\Sigma}_{\boldsymbol{xz}}$

Para enfatizar a dependência de $\boldsymbol{g}_i \equiv \boldsymbol{x}_i \varepsilon_i$ sobre o vetor de parâmetros e os dados escrevemos $\boldsymbol{g}_i$ como $\boldsymbol{g}_i(\boldsymbol{w}_i; \boldsymbol{\delta}) = \mathsf{E}[\boldsymbol{x}_i(y_i - \boldsymbol{z}_i'\boldsymbol{\delta})] = 0$. Deste modo a *condição de ortogonalidade* é rescrita como:

$$\boldsymbol{g}_i(\boldsymbol{w}_i; \boldsymbol{\delta}) = \mathbf{0} \tag{23}$$

Agora faça $\tilde{\delta}(L \times 1)$ ser um valor hipotético de $\boldsymbol{\delta}$ e considere o sistema de $\boldsymbol{K}$ equações simultâneas em L incognitas (os elementos do valor hipotético):

$$\boldsymbol{g}_i(\boldsymbol{w}_i; \tilde{\boldsymbol{\delta}}) = \mathbf{0} \tag{24}$$

As condições de ortogonalidade (23) implicam que os valores do vetor de coeficientes $\boldsymbol{\delta}$ são uma solução para o sistema (24) de $\boldsymbol{K}$ equações simultâneas. As quatro hipóteses GMM garatem que existe solução para o sistema de equações simultâneas.

Dizemos que o vetor de coeficientes é *identificado* se $\tilde{\boldsymbol{\delta}} = \boldsymbol{\delta}$ é *a única solução*.

(24) é um sistema de K equações lineares:

$$\mathsf{E}(\boldsymbol{x}_i y_i) - \mathsf{E}(\boldsymbol{x}_i \boldsymbol{z}_i')\tilde{\boldsymbol{\delta}} = \mathbf{0} \tag{25}$$

ou

$$\underset{(K\times L)}{\boldsymbol{\Sigma}_{\boldsymbol{xz}}}\underset{(L\times 1)}{\tilde{\boldsymbol{\delta}}} = \underset{(K\times 1)}{\boldsymbol{\sigma}_{\boldsymbol{xy}}} \tag{26}$$

tal que

$$\boldsymbol{\Sigma}_{\boldsymbol{xz}} = \mathsf{E}(\boldsymbol{x}_i\boldsymbol{z}_i'); \qquad \boldsymbol{\sigma}_{\boldsymbol{xy}} = \mathsf{E}(\boldsymbol{x}_i y_i)$$

$\tilde{\boldsymbol{\delta}} = \boldsymbol{\delta}$ é a única solução para o sistema se e somente se $\boldsymbol{\Sigma}_{\boldsymbol{xz}}$ é posto completo (que é a condição da hipótese 3.4).

Condição de Ordem para Identificação

Como o posto de $\Sigma_{xz} < L$ se $K < L$, uma condição para identificação é que

$$K(\#\text{ variáveis predeterminadas}) \geq L(\#\text{ regressores}) \quad (27)$$

Esta condição possui como equivalentes:

- # condições de ortogonalidade $\geq$ # parametros
- # variáveis predeterminadas excluídas da equação $\geq$ # regressores endógenos

Taxonomia para Identificação

A equação pode ser:

Sobreidentificada condição de ordem é satisfeita com $K > L$

Exatamente identificada quando a condição de ordem é satisfeita com $K = L$

Subidentificada a condição de ordem não é satisfeita, $K < L$

Normalidade Assintótica

Hipótese 3.5 (g_i é um mds com segundos momentos finitos): Faça $\boldsymbol{g}_i = \boldsymbol{x}_i \varepsilon_i$. $\{\boldsymbol{g}_i\}$ é uma sequência martingale diferença (E($\boldsymbol{g}_i$) = 0). A matriz dos momentos cruzados $(K \times K)$, E($\boldsymbol{g}_i \boldsymbol{g}_i'$), é não-singular. Usamos $\boldsymbol{S}$ para Avar($\bar{\boldsymbol{g}}$). Pela Hip. 3.2 e pelo teorema do limite central (ergódico diferença martingale estacionário), $\boldsymbol{S} = \mathsf{E}(\boldsymbol{g}_i \boldsymbol{g}_i')$.

Normalidade Assintótica

- Se os instrumentos incluem uma constante, então esta hipótese implica que o termo de erro é uma sequência martingale em diferenças.

- Uma condição suficiente:

$$\mathrm{E}(\varepsilon_i \mid \varepsilon_{i-1}, ..., \varepsilon_1, \boldsymbol{x}_i, \boldsymbol{x}_{i-1}, ..., \boldsymbol{x}_1) = 0$$

  o que significa que o termo de erro é ortogonal não apenas ao instrumento contemporaneamente mas a todos os instrumentos passados.

- Como $\boldsymbol{g}_i\boldsymbol{g}_i' = \varepsilon_i^2\boldsymbol{x}_i\boldsymbol{x}_i'$, $\boldsymbol{S}$ é uma matriz de quarto momentos. A estimativa consistente de $\boldsymbol{S}$ irá requerer uma hipótese de quarto momento (3.6).

- Se $\{\boldsymbol{g}_i\}$ é seriamente correlacionado, então $\boldsymbol{S}$ não é igual a $\mathsf{E}(\boldsymbol{g}_i\boldsymbol{g}_i')$ e será necessária uma forma mais complicada para estimação.

Definição GMM

- As condições de ortogonalidade afirmam que um conjunto de momentos da população são todos iguais a zero – $\mathrm{E}[\boldsymbol{g}(\boldsymbol{w}_i;\boldsymbol{\delta})]$.

- O princípio básico do método dos momentos é escolher uma estimativa dos parâmetros tal que corresponda aos momentos amostrais que também são iguais a zero.

$$\boldsymbol{g}_n(\tilde{\boldsymbol{\delta}}) \equiv \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{g}(\boldsymbol{w}_i;\tilde{\boldsymbol{\delta}})$$

Definição GMM

- Aplicando o princípio do método dos momentos, devemos escolher o $\tilde{\boldsymbol{\delta}}$ que soluciona o sistema de K equações simultâneas e L incógnitas: $\boldsymbol{g}_n(\tilde{\boldsymbol{\delta}}) = 0$.

- Como o sistema é linear podemos escrever:

$$\boldsymbol{g}_n(\tilde{\boldsymbol{\delta}}) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i(y_i - \boldsymbol{z}_i'\tilde{\boldsymbol{\delta}}) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i y_i - \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{z}_i'\right)\tilde{\boldsymbol{\delta}} \quad (28)$$

$$\boldsymbol{g}_n(\tilde{\boldsymbol{\delta}}) = \boldsymbol{s}_{\boldsymbol{xy}} - \boldsymbol{S}_{\boldsymbol{xz}}\tilde{\boldsymbol{\delta}} \quad (29)$$

$$\boldsymbol{s}_{\boldsymbol{xy}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i y_i \quad \text{e} \quad \boldsymbol{S}_{\boldsymbol{xz}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{z}_i'$$

- O análogo amostral de $\boldsymbol{g}_n(\tilde{\boldsymbol{\delta}}) = 0$ é um sistema de K equações lineares com L incógnitas:

$$\boldsymbol{S_{xz}}\tilde{\boldsymbol{\delta}} = \boldsymbol{s_{xy}} \tag{30}$$

- Se $K > L$ então o sistema pode não ter uma solução. A extensão do método dos momentos que lida com este caso é o GMM.

Método dos Momentos

- Pela hip. 3.2, $\boldsymbol{S}_{xz}$ converge para $\boldsymbol{\Sigma_{xz}} - \boldsymbol{S_{xz}}$ é inversível em amostras grandes.

- Então, quando possuímos amostras grandes o sistema tem solução:
$$\tilde{\boldsymbol{\delta}}_{IV} = \boldsymbol{S_{xz}^{-1} s_{xy}} \tag{31}$$
Este é o estimador de variáveis intrumentais (IV) para $\boldsymbol{x}_i$ servindo como instrumento.

Método dos Momentos

- Se $\boldsymbol{z}_i = \boldsymbol{x}_i$ então todos os regressores são pré-determinados, então $\tilde{\boldsymbol{\delta}}_{IV}$ se reduz ao estimador OLS. Portanto, a estimação OLS é um estimador do método dos momentos.

Método dos Momentos Generalizado

- Se a equação é sobre-identificada ($K > L$) então não podemos escolher em geral um $\tilde{\boldsymbol{\delta}}$ que satisfaça as K equações do sistema em questão.

- Podemos então escolher um $\tilde{\boldsymbol{\delta}}$ tal que $\boldsymbol{g}_n(\tilde{\boldsymbol{\delta}})$ seja o mais perto de 0.

- Para definir "perto", determinamos a **distância** entre dois vetores quaisquer (dimensão K) ξ e η pela forma quadrática
$$(\xi - \eta)'\widehat{\boldsymbol{W}}(\xi - \eta)$$
$\widehat{\boldsymbol{W}}$ é uma matriz simétrica e positiva definida que define a distância.

Método dos Momentos Generalizado

Definição 3.1 (Estimador GMM) Faça $\widehat{\boldsymbol{W}}$ ser uma matriz simétrica e positiva definida, dependente da amostra, tal que $\widehat{\boldsymbol{W}} \to_p \boldsymbol{W}$ a medida que $n \to \infty$, com $\boldsymbol{W}$ simétrica e positiva definida. O **estimador GMM** de $\boldsymbol{\delta}$, representado por $\hat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}})$, é

$$\hat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}) \equiv \underset{\tilde{\boldsymbol{\delta}}}{\text{argmin}} J(\tilde{\boldsymbol{\delta}}, \widehat{\boldsymbol{W}})$$

onde

$$J(\tilde{\boldsymbol{\delta}}, \widehat{\boldsymbol{W}}) \equiv n \left[\boldsymbol{g}_n(\tilde{\boldsymbol{\delta}})' \widehat{\boldsymbol{W}} \boldsymbol{g}_n(\tilde{\boldsymbol{\delta}}) \right].$$

Método dos Momentos Generalizado

- Como $\boldsymbol{g}_n(\tilde{\boldsymbol{\delta}})$ é linear em $\tilde{\boldsymbol{\delta}}$, a função objetivo é quadrática em $\tilde{\boldsymbol{\delta}}$ quando a equação é linear:
$$J(\tilde{\boldsymbol{\delta}}, \widehat{\boldsymbol{W}}) \equiv n(\boldsymbol{s_{xy}} - \boldsymbol{S_{xz}}\tilde{\boldsymbol{\delta}})'\widehat{\boldsymbol{W}}(\boldsymbol{s_{xy}} - \boldsymbol{S_{xz}}\tilde{\boldsymbol{\delta}}) \tag{32}$$
A FOC com respeito a $\tilde{\boldsymbol{\delta}}$ é:
$$\boldsymbol{S'_{xz}}\widehat{\boldsymbol{W}}\boldsymbol{s_{xy}} = \boldsymbol{S'_{xz}}\widehat{\boldsymbol{W}}\boldsymbol{S_{xz}}\tilde{\boldsymbol{\delta}}$$

- Como $\widehat{\boldsymbol{W}}$ é positiva semidefinida a matriz $(L \times L)$ $\boldsymbol{S'_{xz}}\widehat{\boldsymbol{W}}\boldsymbol{s_{xy}}$ é não singular. A única solução pode ser obtida multiplicando ambos os lados por pela inversa de $\boldsymbol{S'_{xz}}\widehat{\boldsymbol{W}}\boldsymbol{s_{xy}}$. A única solução forma o estimador GMM:
$$\textbf{estimador GMM: } \hat{\delta}(\widehat{\boldsymbol{W}}) = (\boldsymbol{S'_{xz}}\widehat{\boldsymbol{W}}\boldsymbol{S_{xz}})^{-1}\boldsymbol{S'_{xz}}\widehat{\boldsymbol{W}}\boldsymbol{s_{xy}} \tag{33}$$

- Se $K = L$, então $\boldsymbol{S_{xz}}$ é uma matriz quadrática e o estimador GMM se reduz ao estimador IV.

GMM: Sampling Error

Multiple ambos os lados por da equação de estimação $y_i = \boldsymbol{z}_i'\boldsymbol{\delta} + \varepsilon_i$ por e tomando as médias se tem:

$$\boldsymbol{s_{xy}} = \boldsymbol{S_{xz}}\boldsymbol{\delta} + \widehat{\boldsymbol{g}} \tag{34}$$

tal que

$$\widehat{\boldsymbol{g}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i\varepsilon_i = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{g}(\boldsymbol{w}_i;\boldsymbol{\delta}) = \boldsymbol{g}_n(\boldsymbol{\delta})$$

Substituindo (34) na equação do estimador GMM (33):

$$\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}) - \boldsymbol{\delta} = (\boldsymbol{S'_{xz}}\widehat{\boldsymbol{W}}\boldsymbol{S_{xz}})^{-1}\boldsymbol{S'_{xz}}\widehat{\boldsymbol{W}}\widehat{\boldsymbol{g}} \tag{35}$$

Distribuição assintótica do estimador GMM

Proposição 3.1 (distribuição assintótica do estimador GMM):

(a) (Consistência) Sob as hipóteses 3.1-3.4, $\text{plim}_{n\to\infty}\hat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}) = \boldsymbol{\delta}$.

(b) (Normalidade assintótica) Se a hipótese 3.3 é expandida como a hipótese 3.5, então

$$\sqrt{n}(\hat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}) - \boldsymbol{\delta}) \to_d N(0, \text{Avar}(\hat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}))) \quad \text{a medida que } n \to \infty,$$

tal que

$$\text{Avar}(\hat{\boldsymbol{\delta}}(\widehat{\text{W}})) = (\boldsymbol{\Sigma}'_{\boldsymbol{xz}}\boldsymbol{W}\boldsymbol{\Sigma}_{\boldsymbol{xz}})^{-1}\boldsymbol{\Sigma}'_{\boldsymbol{xz}}\boldsymbol{W}\boldsymbol{S}\boldsymbol{W}\boldsymbol{\Sigma}_{\boldsymbol{xz}}(\boldsymbol{\Sigma}'_{\boldsymbol{xz}}\boldsymbol{W}\boldsymbol{\Sigma}_{\boldsymbol{xz}})^{-1}$$

sendo que $\boldsymbol{\Sigma}_{\boldsymbol{xz}} \equiv \text{E}(\boldsymbol{x}_i\boldsymbol{z}'_i), \boldsymbol{S} = \text{E}(\boldsymbol{g}_i\boldsymbol{g}'_i), \boldsymbol{W} \equiv \text{plim}\hat{\boldsymbol{W}}$.

Distribuição assintótica do estimador GMM

Proposição 3.1 (distribuição assintótica do estimador GMM):

(c) (Consistência da estimativa de Avar($\hat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}})$)) Suponha que é disponível um estimador consistente, $\widehat{\boldsymbol{S}}$, de $\boldsymbol{S}(K \times K)$. Então, sob a hipótese 3.2, Avar($\hat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}})$) é consistentemente estimada por

$$\widehat{\text{Avar}(\hat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}))} = (\boldsymbol{S}'_{\boldsymbol{xz}}\hat{\boldsymbol{W}}\boldsymbol{S}_{\boldsymbol{xz}})^{-1}\boldsymbol{S}'_{\boldsymbol{xz}}\hat{\boldsymbol{W}}\hat{\boldsymbol{S}}\hat{\boldsymbol{W}}\boldsymbol{S}_{\boldsymbol{xz}}(\boldsymbol{S}'_{\boldsymbol{xz}}\hat{\boldsymbol{W}}\boldsymbol{S}_{\boldsymbol{xz}})^{-1},$$

tal que $\boldsymbol{S}_{\boldsymbol{xx}}$ é a média amostral de $\boldsymbol{x}_i\boldsymbol{x}'_i$:

$$\boldsymbol{S}_{\boldsymbol{xx}} \equiv \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}'_i = \frac{1}{n}\boldsymbol{X}\boldsymbol{X}'$$

Distribuição assintótica do estimador GMM

Proposição 3.2 (estimação consistente da variância do erro): Para qualquer estimador consistente, $\widehat{\boldsymbol{\delta}}$, de $\boldsymbol{\delta}$, defina $\widehat{\varepsilon}_i \equiv y_i - \boldsymbol{z}_i'\widehat{\boldsymbol{\delta}}$. Sob as hipóteses 3.1, 3.2, mais a hipótese que $\mathrm{E}(\boldsymbol{z}_i\boldsymbol{z}_i')$ existe e é finita,

$$\frac{1}{n}\sum_{i=1}^{n}\widehat{\varepsilon}_i^2 \to_p \mathrm{E}(\widehat{\varepsilon}_i^2)$$

dado que $\mathrm{E}(\widehat{\varepsilon}_i^2)$ existe e é finito.

Teste de Hipótese

Proposição 3.3 (razão t robusta e estatística de Wald): Suponha que as hipóteses 3.1-3.5 valem, e suponha que é disponível uma estimativa consistente $\widehat{\boldsymbol{S}}$ de $\boldsymbol{S}$. Faça

$$\widehat{\text{Avar}(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}))} = (\boldsymbol{S'_{xz}}\hat{\boldsymbol{W}}\boldsymbol{S_{xz}})^{-1}\boldsymbol{S'_{xz}}\hat{\boldsymbol{W}}\hat{\boldsymbol{S}}\hat{\boldsymbol{W}}\boldsymbol{S_{xz}}(\boldsymbol{S'_{xz}}\hat{\boldsymbol{W}}\boldsymbol{S_{xz}})^{-1},$$

Então

(a) Sob a hipótese nula $H_0 : \delta_\ell = \bar{\delta}_\ell$,

$$t_\ell \equiv \frac{\sqrt{n}(\hat{\delta}_\ell(\widehat{\boldsymbol{W}}) - \bar{\delta}_\ell)}{\sqrt{(\widehat{\text{Avar}(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}))})_{\ell\ell}}} = \frac{\hat{\delta}_\ell(\widehat{\boldsymbol{W}}) - \bar{\delta}_\ell}{SE^*_\ell} \to_d N(0,1).$$

tal que $\left(\widehat{\text{Avar}(\hat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}))}\right)_{\ell\ell}$ é o elemento $(\ell\ell)$ da $\widehat{\text{Avar}(\hat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}))})$ e

$$SE_\ell^* \equiv \sqrt{\frac{1}{n}(\widehat{\text{Avar}(\hat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}))})_{\ell\ell}}$$

(b) Sob a hipótese nula $H_0 : \boldsymbol{R\delta} = \boldsymbol{r}$ tal que $\#\boldsymbol{r}$ é o número de restrições e $\boldsymbol{R}(\#\boldsymbol{r} \times L)$ é posto completo em linha,

$$W \equiv n.(\boldsymbol{R}\hat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}) - \boldsymbol{r})'\{\boldsymbol{R}[\widehat{\text{Avar}(\hat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}))}]\boldsymbol{R}'\}^{-1}(\boldsymbol{R}\hat{\boldsymbol{\delta}}\widehat{(\boldsymbol{W})} - \boldsymbol{r}) \to_d \chi^2(\#r$$

(c) Sob a hipótese nula $H_0 : \boldsymbol{a}(\boldsymbol{\delta}) = \boldsymbol{0}$ tal que $\boldsymbol{A}(\boldsymbol{\delta})$, a matriz $(\#\boldsymbol{a} \times L)$ de primeira derivada de $\boldsymbol{a}(\boldsymbol{\delta})$ é contínua e de posto completo,

$$W \equiv n.a(\hat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}))'\{\boldsymbol{A}(\hat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}))[\widehat{\text{Avar}(\hat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}))}]\boldsymbol{A}(\hat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}))'\}^{-1}(\boldsymbol{a}\hat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}})) \to$$

Restrições: H_0 : $\boldsymbol{R\delta} = \boldsymbol{r}$, exemplo: $\boldsymbol{R} = \left[\boldsymbol{I}_q \mathbf{0}_{q\times(k+1-q)}\right]$ e $\boldsymbol{r} = \mathbf{0}_q$ (veja Stock e Watson, p. 760).

Estimativa de S

- Vimos anteriormente que uma estimativa consistente era

$$\widehat{\boldsymbol{S}} \equiv \frac{1}{n} \sum_{i=1}^{n} \widehat{\varepsilon}_i^2 \boldsymbol{x}_i \boldsymbol{x}_i'$$

tal que $\widehat{\varepsilon}_i = y_i - \boldsymbol{z}'\widehat{\boldsymbol{\delta}}$.

Estimando S consistentemente

Generalização da hip. 2.6.

Hipótese 3.6 (quarto momento finito): $\mathrm{E}[(x_{ik}z_{i\ell})^2]$ existe e é finito para todo k, $j = (1, 2, ..., K)$, e $\ell(= 1, ..., L)$.

Estimando S consistentemente

Proposição 3.4 (estimativa consistente de S): Suponha o coeficiente estimado $\widehat{\boldsymbol{\delta}}$ usado para calcular o resíduo $\widehat{\varepsilon}_i$ para $\widehat{\boldsymbol{S}}$ de $(1/n)\sum_{i=1}^{n}\widehat{\varepsilon}_i^2\boldsymbol{x}_i\boldsymbol{x}_i'$, e suponha que $\boldsymbol{S} = \mathsf{E}(g_i g_i')$ exista e é finito. Então, sob as hips 3.1, 3.2 e 3.6, $\widehat{\boldsymbol{S}}$ é consistente para $\boldsymbol{S}$.

Estimador eficiente GMM

Proposição 3.5 (escolha ótima da matriz $\widehat{\boldsymbol{W}}$): Um limite inferior para a variância assintótica dos estimadores GMM, indexados por $\widehat{\boldsymbol{W}}$, é dado por $(\boldsymbol{\Sigma}'_{\boldsymbol{xz}}\boldsymbol{S}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{xz}})^{-1}$. Esse limite inferior é alcançado se $\widehat{\boldsymbol{W}}$ é tal que $\boldsymbol{W}(\equiv \text{plim}\widehat{\boldsymbol{W}}) = \boldsymbol{S}^{-1}$.

Portanto, o **estimador GMM eficiente** é aquele que satisfaz $\text{plim}\widehat{\boldsymbol{W}} = \boldsymbol{S}^{-1}$.

Simplesmente substituindo $\widehat{\boldsymbol{W}}$ por $\widehat{\boldsymbol{S}^{-1}}$ (que é consistenta para $\boldsymbol{S}^{-1}$) nas fórmulas da Proposição 3.1 obtemos:

$$\textbf{estimador GMM: } \hat{\delta}(\widehat{\boldsymbol{S}}^{-1}) = (\boldsymbol{S}'_{\boldsymbol{xz}}\widehat{\boldsymbol{S}}^{-1}\boldsymbol{S}_{\boldsymbol{xz}})^{-1}\boldsymbol{S}'_{\boldsymbol{xz}}\widehat{\boldsymbol{S}}^{-1}\boldsymbol{s}_{\boldsymbol{xy}} \quad (36)$$

$$\text{Avar}(\hat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1})) = (\boldsymbol{\Sigma}'_{\boldsymbol{xz}}\boldsymbol{S}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{xz}})^{-1} \tag{37}$$

$$\widehat{\text{Avar}(\hat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}))} = (\boldsymbol{S}'_{\boldsymbol{xz}}\widehat{\boldsymbol{S}}^{-1}\boldsymbol{S}_{\boldsymbol{xz}})^{-1} \tag{38}$$

Com $\widehat{\boldsymbol{W}} = \boldsymbol{S}^{-1}$, as fórmulas para o teste t robusto e estatística de Wald da Proposição 3.3 passa a ser:

$$t_\ell = \frac{\hat{\delta}_\ell(\widehat{\boldsymbol{S}}^{-1}) - \bar{\delta}_\ell)}{SE^*_\ell} \tag{39}$$

o SE^*_ℓ é o erro-robusto dado por

$$SE^*_\ell = \sqrt{\frac{1}{n}((\boldsymbol{S}'_{\boldsymbol{xz}}\widehat{\boldsymbol{S}}^{-1}\boldsymbol{S}_{\boldsymbol{xz}})^{-1})_{\ell\ell}}$$

$$W = n.\boldsymbol{a}(\hat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}))'\{\boldsymbol{A}(\hat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}))(\boldsymbol{S}'_{\boldsymbol{xz}}\widehat{\boldsymbol{S}}^{-1}\boldsymbol{S}_{\boldsymbol{xz}})^{-1}\boldsymbol{A}(\hat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}))'\}^{-1}(\boldsymbol{a}\hat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1})) \tag{40}$$

Estimador GMM eficiente

- Para se calcular um estimador GMM eficiente precisamos de uma estimativa consistente de $\widehat{\boldsymbol{S}}$.

- A proposição 3.4 garante que $\widehat{\boldsymbol{S}}$ baseado em qualquer estimador consistente de δ é consistente para $\boldsymbol{S}$.

Estimador GMM eficiente

- Procedimento de dois estágios de estimador GMM eficiente.

  1. Escolha uma matriz $\widehat{\boldsymbol{W}}$ que converge em probabilidade para uma matriz simétrica positiva definda, e minimize $J(\tilde{\delta}, \widehat{\boldsymbol{W}})$ sobre $\tilde{\delta}$ para obter $\tilde{\delta}(\widehat{\boldsymbol{W}})$. Usualmente fazemos $\widehat{\boldsymbol{W}} = \boldsymbol{S}_{xx}^{-1}$. Esse estimador é o conhecido método de mínimos quadrados de dois estágios (2SLS). Use isto para calcular o resíduo $\widehat{\varepsilon}$ e obter um estimador consistente $\widehat{\boldsymbol{S}}$ de $\boldsymbol{S}$.

  2. Minimize $J(\tilde{\delta}, \widehat{\boldsymbol{S}}^{-1})$ sobre $\tilde{\delta}$. O minimizador é o estimador GMM eficiente.

Teste de Hansen de Restrições Sobre-identificadoras (Hansen, 1982)

Quando a equação é sobre-identificada então a distância $J(\tilde{\boldsymbol{\delta}}, \widehat{\boldsymbol{W}}) \equiv n\left[\boldsymbol{g}_n(\tilde{\boldsymbol{\delta}})'\widehat{\boldsymbol{W}}\boldsymbol{g}_n(\tilde{\boldsymbol{\delta}})\right]$ (ver Definição 3.1) não é exatamente zero. Se a matriz de pesos $\widehat{\boldsymbol{W}}$ é escolhida otimamente então plim$\widehat{\boldsymbol{W}} = \boldsymbol{S}^{-1}$, assim a distância minimizada é assintóticamente qui-quadrada.

Proposição 3.6 (Teste de Hansen de sobre-identificação de restrições:) Suponha que disponível um estimador consistente de $\boldsymbol{S}$, $\widehat{\boldsymbol{S}}(= \mathsf{E}(\boldsymbol{g}_i\boldsymbol{g}_i'))$. Sob as hipóteses 3.1-3.5,

$$J(\hat{\delta}(\widehat{\boldsymbol{S}}^{-1}), \widehat{\boldsymbol{S}}^{-1}) = n.\boldsymbol{g}_n(\hat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}))'\widehat{\boldsymbol{S}}^{-1}\boldsymbol{g}_n(\hat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1})) \rightarrow_d \chi^2(K-L). \tag{41}$$

Comentários:

- Este é um teste de especificação: se todas as restrições do modelo são satisfeitas. Se a estatística J é grande significa que tanto as condições de ortogonalidade e/ou outras hipóteses são falsas. Apenas quando temos confiança nas outras hipóteses podemos interpretar a estatística J grande como evidência para endogeneidade de algum dos K instrumento incluidos em $\boldsymbol{x}_i$.

- O teste não é consistente contra algumas falhas das condições de ortogonalidade.

- Existem preocupações quanto a aplicação do teste para amostras pequenas.

Testando subconjuntos para condições de ortogonalidade

Suponha que podemos dividir os K instrumentos em dois grupos:

1. o vetor $\boldsymbol{x}_{i1}$ de K_1 instrumentos confiáveis e,

2. o vetor $\boldsymbol{x}_i2$ dos demais instrumentos $K - K_1$ que podem ser suspeitos de violar a hipótese de ortogonalidade.

Como a ordenação dos instrumentos não muda os valores da estimação e das estatísticas de teste, podemos assumir que os últimos K elementos de $\boldsymbol{x}_i$ são os suspeitos:

$$\boldsymbol{x}_i = \begin{bmatrix} \boldsymbol{x}_{i1} \\ \boldsymbol{x}_{i2} \end{bmatrix} \begin{matrix} \} & K_1 \text{ linhas} \\ \} & K - K_1 \text{ linhas} \end{matrix}$$

A parte do modelo que queremos testar é:

$$\mathrm{E}(\boldsymbol{x}_{i2}\varepsilon_i) = \mathbf{0} \tag{42}$$

Esta restrição é testável se existe ao menos instrumentos não suspeitos quanto os coeficientes: $K_1 \geq L$.

A idéia é comparar duas estatísticas J de duas estimativas GMM do mesmo coeficiente $\boldsymbol{\delta}$: uma usando o subconjunto de instrumentos $\boldsymbol{x}_{i1}$ e outra utilizando todos os instrumentos $\boldsymbol{x}_i$. Se a inclusão dos instrumentos eleva a estatística J então existe uma boa razão para duvidar da predeterminação dos $\boldsymbol{x}_{i2}$.

De acordo com as particições de $\boldsymbol{x}_{i1}$ as condições de ortogonalidade podem ser escritas assim:

$$\underset{(K\times 1)}{\boldsymbol{g}_n(\tilde{\boldsymbol{\delta}})} = \begin{bmatrix} \underset{(K_1\times 1)}{\boldsymbol{g}_{1n}(\tilde{\boldsymbol{\delta}})} \\ \underset{((K-K_1)\times 1)}{\boldsymbol{g}_{2n}(\tilde{\boldsymbol{\delta}})} \end{bmatrix}, \quad \underset{(K\times K)}{\boldsymbol{S}} = \begin{bmatrix} \boldsymbol{S}_{11} & \boldsymbol{S}_{12} \\ \boldsymbol{S}_{21} & \boldsymbol{S}_{22} \end{bmatrix} \tag{43}$$

tal que $\boldsymbol{S}_{11} = \mathsf{E}(\varepsilon_i^2 \boldsymbol{x}_{i1}\boldsymbol{x}'_{i1})$, etc.

Proposição 3.7 (Testando um subconjunto de condições de ortogonalidade:) Suponha que as hipóteses 3.1-3.5 valem. Faça $\boldsymbol{x}_i$ ser um subvetor de $\boldsymbol{x}_i$, e expanda a hipótese 3.4 pelo requerimento da condição do posto para identificação seja satisfeita para $\boldsymbol{x}_{i1}$. Então para qualquer estimador consistente $\widehat{\boldsymbol{S}}$ de $\boldsymbol{S}$ e $\widehat{\boldsymbol{S}}_{11}$ de $\boldsymbol{S}_{11}$,

$$C \equiv J - J_1 \rightarrow_d \chi^2(K - K_1).$$

Exemplo 3.3 (testando o quanto escolaridade é predeterminado na equação salarial) Na equação salarial do exemplo 3.2, suponha que escolaridade S_i é suspeita de ser endógena. Para testar esta situação faça:

$$\boldsymbol{x}_{i1} = \begin{bmatrix} 1 \\ EXPR_i \\ AGE_i \\ MED_i \end{bmatrix}, \ \boldsymbol{x}_{i2} = [S_i]$$

O vetor de regressores $\boldsymbol{z}_i$ é o mesmo. O primeiro passo é estimar $\boldsymbol{\delta}$ pelo GMM eficientes de duas etapas com $\boldsymbol{x}_i = [\boldsymbol{x}_{i1}\boldsymbol{x}_{i2}]'$ como instrumentos. Isto produz J e a matriz (5×5) $\widehat{\boldsymbol{S}}$. Utilize a matriz $\widehat{\boldsymbol{S}}$ da primeira estimação para formas a matriz $\boldsymbol{S}_{11}$ e estime os mesmos coeficientes $\boldsymbol{\delta}$ por GMM (usando $\boldsymbol{x}_{i1}$ e $\widehat{\boldsymbol{S}}_{11}$). A diferença das estatísticas J deve ser assintóticamente $\chi^2(1)$.

Teste de Hipótese por Princípio Razão da Verosimilhança (LR)

Na seção 3.5 foi derivada as estatísticas de teste chi-quadrada para as hipóteses nula $\mathrm{H}_0 : \boldsymbol{a}(\boldsymbol{\delta}) = \mathbf{0}$ pelo princípio de Wald. Aqui é feito o mesmo pelo princípio LR: examina a diferença na função objetivo com e sem a imposição da hipótese nula.

Na estimativa GMM eficiente a função objetivo é $J(\tilde{\boldsymbol{\delta}}, \widehat{\boldsymbol{S}}^{-1})$ para uma estimativa consistente $\hat{\boldsymbol{S}}$ de $\boldsymbol{S}$.

Estimador GMM eficiente restrito

$$\bar{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}) \equiv \underset{\tilde{\boldsymbol{\delta}}}{\text{argmin}}\ J(\tilde{\boldsymbol{\delta}}, \widehat{\boldsymbol{S}}^{-1}) \text{ s.t. } \mathrm{H}_0 \tag{44}$$

O princípio LR sugere que

$$LR \equiv J(\bar{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}), \widehat{\boldsymbol{S}}^{-1}) - J(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}), \widehat{\boldsymbol{S}}^{-1}) \tag{45}$$

deve ser assintoticamente χ-quadrado.

Proposição 3.8 (estatística de teste pelo princípio LR): Suponha que as Hipóteses 3.1-3.5 valem e suponha que seja disponível um estimador consistente $\widehat{\boldsymbol{S}}$ de $\boldsymbol{S}(= \mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_i'))$. Considere a hipótese nula de $\#\boldsymbol{a}$ restrições $\mathrm{H}_0 : \boldsymbol{a}(\boldsymbol{\delta}) = \boldsymbol{0}$ tal que $\boldsymbol{A}(\boldsymbol{\delta}) = \boldsymbol{0}$, a matriz $(\#\boldsymbol{a} \times L)$ das primeiras derivadas, é contínua e de posto completo. Defina duas estatísticas, W e LR, por (40) e (45) , respectivamente. Então, sob a hipótese nula, o seguinte vale:

(a) As duas estatisticas são assintoticamente equivalentes – as distribuições assintóticas são as mesmas $(\chi^2(\#\boldsymbol{a}))$.

(b) As duas estatísticas são assintoticamente equivalentes no sentido mais forte que suas diferenças numéricas convergem em probabilidade para zero: $LR - W \to_p 0$.

(c) Além disso, se a hipótese é linear tal que as restrições podem ser escritas como $\boldsymbol{R\delta} = \boldsymbol{r}$, então as duas estatísticas são numericamente iguais.

Teste de Hipótese por Princípio Razão da Verosimilhança (LR)

- A vantagem da LR sobre W é invariante: o valor numérico da LR não depende de como as restrições são representadas por $\boldsymbol{a}(\cdot)$. Quando a hipótese é não linear é necessário escrever um programa para o teste.

- Proposição 3.8 requer que a matriz de distância $\hat{\boldsymbol{W}}$ satisfaça a condição plim$\hat{\boldsymbol{W}} = \boldsymbol{S}^{-1}$. Caso contrário LR não é estatisticamente χ-quadrado. (A estatistica de Wald é assintoticamente χ-quadrado sem satisfazer esta condição de eficiência).

- A mesma estimativa de $\boldsymbol{S}$ deve ser usada para calcular a LR (os dois J).

- Para existir equivalência numérica, $LR = W$, no caso linear então a mesma estimativa $\widehat{\boldsymbol{S}}$ deve ser a mesma para ambos. Caso contrário, LR e W são apenas assintoticamente equivalentes.

Teste de Hipótese por Princípio Razão da Verosimilhança (LR)

Estatística LR para OLS: OLS é caso especial de GMM. A estatística LR pode ser escrita da seguinte forma: em OLS fazemos $\boldsymbol{x}_i = \boldsymbol{z}_i$, então o modelo GMM eficiente irrestrito é OLS e $J(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}), \widehat{\boldsymbol{S}}^{-1}) = 0$. Portanto,

$$LR = J(\bar{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}), \widehat{\boldsymbol{S}}^{-1}) \tag{46}$$

Aqui $\bar{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1})$ é o estimador GMM eficiente restrito.

Pela Proposição 3.8 esta estatística é assintoticamente χ-quadrada e é numericamente igual a estatística de Wald se a hipótese nula é linear. (Assumindo homocedasticidade condicional a LR é a diferença na soma do quadrado dos resíduos normalizado pela variância do erro).

Implicações da Homocedasticidade Condicional

A análise anterior não assume homocedasticidade. Esta análise é apresentada na seção 3.8 do livro de Hayashi como caso particular de GMM.

Hipótese 3.7 (homocedasticidade condicional:)

$$\mathsf{E}(\varepsilon_i^2 \mid \boldsymbol{x}_i) = \sigma^2$$

- Nesse caso, a matriz do quarto momento é $\boldsymbol{S} = \mathsf{E}(\boldsymbol{g}_i \boldsymbol{g}_i') = \mathsf{E}(\varepsilon_i^2 \boldsymbol{x}_i \boldsymbol{x}_i')$ pode ser escrito como o produto de segundos momentos:

$$\boldsymbol{S} = \sigma^2 \boldsymbol{\Sigma}_{\boldsymbol{xx}} \tag{47}$$

tal que $\boldsymbol{\Sigma}_{\boldsymbol{xx}} = \mathsf{E}(\boldsymbol{x}_i \boldsymbol{x}_i')$.

Implicações da Homocedasticidade Condicional

- Como $\boldsymbol{S}$ é não singular pela hipótese 3.5 isso implica que $\sigma^2 > 0$ e $\boldsymbol{\Sigma}_{\boldsymbol{xx}}$ é não-singular.

- Para este caso, o estimador de $\boldsymbol{S}$ é:

$$\widehat{\boldsymbol{S}} = \widehat{\sigma}^2 \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' = \widehat{\sigma}^2 \boldsymbol{S}_{\boldsymbol{xx}} \tag{48}$$

  Aqui $\widehat{\sigma}^2$ é um estimador consistente a ser especificado depois. Por estacionariedade ergódica $\boldsymbol{S}_{\boldsymbol{xx}} \rightarrow_{a.s.} \boldsymbol{\Sigma}_{\boldsymbol{xx}}$. Dado um $\widehat{\sigma}^2$, não precisamos da Hip. 3.6 do quarto momento para $\widehat{\boldsymbol{S}}$ ser consistente.

Implicações da Homocedasticidade Condicional Com homcedasticidade condicional simplesmente vários resultados podem ser simplificados com a substituição de $\boldsymbol{S}$ por $\hat{\sigma}^2\boldsymbol{S_{xx}}$ e a expressão $\hat{\boldsymbol{S}}$ de $(1/n)\sum_{i=1}^{n}\hat{\varepsilon}_i^2\boldsymbol{x}_i\boldsymbol{x}_i'$ por $\hat{\boldsymbol{S}}$ em (48). As simplificações são as seguintes:

- O estimador GMM eficiente se torna 2SLS:

$$\begin{aligned}\tilde{\delta}(\hat{\boldsymbol{S}}^{-1}) &= [\boldsymbol{S}_{xz}'(\hat{\sigma}^2\boldsymbol{S}_{xx})^{-1}\boldsymbol{S}_{xz}]\boldsymbol{S}_{xz}'(\hat{\sigma}^2\boldsymbol{S}_{xx})^{-1}\boldsymbol{s}_{xy}\\ &= [\boldsymbol{S}_{xz}'\boldsymbol{S}_{xx}^{-1}\boldsymbol{S}_{xz}]\boldsymbol{S}_{xz}'\boldsymbol{S}_{xx}^{-1}\boldsymbol{s}_{xy}\\ &= \tilde{\delta}(\boldsymbol{S}_{xx}^{-1}) \equiv \tilde{\delta}_{2SLS} \qquad (49)\end{aligned}$$

- Sob homocedasticidade condicional, não há necessidade do primeiro passo do estimador GMM porque o segundo passo se reduz ao estimador GMM com $\boldsymbol{S}_{xx}$ sendo usado para $\widehat{\boldsymbol{S}}$. Este estimador é chamado de Mínimo Quadrado de Dois-Estágios (2SLS) – Proposto por Theil (1953).

Implicações da Homocedasticidade Condicional – Avar($\tilde{\delta}_{2SLS}$)
Podemos calcular a Avar para o caso do 2SLS:

$$\text{Avar}(\tilde{\delta}_{2SLS}) = \sigma^2(\boldsymbol{\Sigma'_{xz}}\boldsymbol{\Sigma^{-1}_{xx}}\boldsymbol{\Sigma_{xz}})^{-1}.$$

Um estimador natural para a Avar é:

$$\widehat{\text{Avar}(\tilde{\delta}_{2SLS})} = \hat{\sigma}^2(\boldsymbol{S'_{xz}}\boldsymbol{S^{-1}_{xx}}\boldsymbol{S_{xz}})^{-1}.$$

Para $\hat{\sigma}^2$, considere a variância amostral do resíduo do 2SLS:

$$\hat{\sigma}^2 \equiv \frac{1}{n}\sum_{i=1}^{n}(y_i - \boldsymbol{z}'_i\hat{\boldsymbol{\delta}}_{2SLS})^2.$$

Pela proposição 3.2, $\hat{\sigma}^2 \to_p \sigma^2$ se E($\boldsymbol{z}_i\boldsymbol{z}'_i$) existe é finito. Portanto, se $\hat{\boldsymbol{S}}$ é definida como (48) com este $\hat{\sigma}^2$ acima.

Substituindo (48) nas equações da razão t e da estatística de Wald, se tem:

$$t_\ell = \frac{\hat{\delta}_{2SLS,\ell} - \bar{\delta}_\ell}{SE_\ell} \tag{50}$$

com

$$SE_\ell \equiv \sqrt{\frac{\hat{\sigma}^2}{n}\left((\boldsymbol{S'_{xz}}\boldsymbol{S_{xx}^{-1}}\boldsymbol{S_{xz}})^{-1}\right)_{\ell\ell}}$$

$$W = n.\frac{\boldsymbol{a}(\hat{\boldsymbol{\delta}}_{2SLS})'\left[\boldsymbol{A}(\hat{\boldsymbol{\delta}}_{2SLS})'(\boldsymbol{S'_{xz}}\boldsymbol{S_{xx}^{-1}}\boldsymbol{S_{xz}})^{-1}\boldsymbol{A}(\hat{\boldsymbol{\delta}}_{2SLS})'\right]^{-1}\boldsymbol{a}(\hat{\boldsymbol{\delta}}_{2SLS})}{\hat{\sigma}^2} \tag{51}$$

Implicações da Homocedasticidade Condicional

- Quando $\hat{\boldsymbol{W}} = (\hat{\sigma}^2 \boldsymbol{S}_{xx})^{-1}$:

$$J(\tilde{\boldsymbol{\delta}}, (\hat{\sigma}^2 \boldsymbol{S}_{xx})^{-1}) = n \frac{(\boldsymbol{s_{xy}} - \boldsymbol{S_{xz}}\tilde{\delta})' \boldsymbol{S}_{xx}^{-1} (\boldsymbol{s_{xy}} - \boldsymbol{S_{xz}}\tilde{\delta})}{\hat{\sigma}^2}.$$

  Para o estimador $\tilde{\delta}_{2SLS}$, a distância J é chamada de estatística de Sargan (1958):

$$\text{Estatística de Sargan} = n \frac{(\boldsymbol{s_{xy}} - \boldsymbol{S_{xz}}\tilde{\delta}_{2SLS})' \boldsymbol{S}_{xx}^{-1} (\boldsymbol{s_{xy}} - \boldsymbol{S_{xz}}\tilde{\delta}_{2SLS})}{\hat{\sigma}^2}.$$

2SLS

Proposição 3.9 (Propriedades assintóticas do 2SLS):

(a) Sob as hipóteses 3.1-3.4, o estimador 2SLS é consistente. Se adicionamos a hipótese 3.5, o estimador é assintoticamente normal com a variância assintótica dada por Avar($\widehat{\delta}_{2SLS}$) com $\boldsymbol{W} = (\sigma^2 \Sigma_{xx})^{-1}$. Se a hip. 3.7 é adicionada às hips. 3.1-3.5, então o estimador GMM é eficiente.

Além disso, se E($\boldsymbol{z}_i \boldsymbol{z}_i'$), então:

(b) a variância assintótica é consistentemente estimada por Avar($\tilde{\delta}_{2SLS}$).

(c) $t_\ell = \frac{\hat{\delta}_{2SLS,\ell} - \bar{\delta}_\ell}{SE_\ell} \to_d N(0,1)$,

$W = n \frac{a(\hat{\delta}_{2SLS})' \{ \boldsymbol{A}(\hat{\delta}_{2SLS}) (\boldsymbol{S}'_{\boldsymbol{xz}} \boldsymbol{S}^{-1}_{\boldsymbol{xx}} \boldsymbol{S}_{\boldsymbol{xz}})^{-1} \boldsymbol{A}(\hat{\delta}_{2SLS})' \}^{-1} \boldsymbol{a}(\hat{\delta}_{2SLS})}{\hat{\sigma}^2} \to_d \chi^2(\#\boldsymbol{r})$, e

(d) a estatística de Sargan $\to_d \chi^2(K - L)$.

GMM com múltiplas equações

Estimar mais de uma equação conjutamente por GMM.

O "payoff" de dominar a estimação de múltiplas esquações é considerável. Quando se assume homocedasticidade condicional, o GMM se reduz ao "full-information instrumental variable efficient (FIVE) estimator", que se reduz no 3SLS se o conjunto de variáveis instrumentais é comum para todas as equações. Se assumirmos que todos os regressores são predeterminados, 3SLS se reduz no estimador SUR (seemingly unrelated regressions), que por sua vez se reduz na regressão multivariada quando todas as equações possuem os mesmos regressores.

Será mostrado que o sistema de múltiplas equações pode ser escrito como um sistema de equação com os coeficientes restritos a serem os mesmos entre equações. O estimador GMM para este sistema é de novo um caso especial do GMM de equação única. O estimador GMM quando todos os regressores são pre-determinados e os erros são condicionalmente homocedásticos é chamado do estimador de efeitos aleatórios (RE). Assim sendo, SUR e RE são estimadores equivalentes.

GMM com múltiplas equações

Hip. 4.1 (linearidade): Existem M equações lineares:

$$y_{im} = \boldsymbol{z}'_{im}\boldsymbol{\delta}_m + \varepsilon_{im} \quad (i = 1, 2, ...)(m = 1, ..., M).$$

$\boldsymbol{z}$ é o vetor de regressores, $\boldsymbol{\delta}$ é o vetor dos coeficientes ($L \times 1$). O modelo não faz hipóteses sobre a correlação entre os erros das equações. Além disso, não há restrições sobre os coeficientes de equações diferentes.

Hip. 4.2 (estacionaridade ergódica): Faça $\boldsymbol{w}_i$ serem os elementos únicos e não-constantes de $(y_{i1}, ..., y_{iM}, \boldsymbol{z}_{i1}, ..., \boldsymbol{z}_{iM}, \boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{iM}$ $\{\boldsymbol{w}_i\}$ é conjuntamente estacionária e ergódica.

Condições de ortogonalidade é o conjunto das condições de ortogonalidade das condições individuais.

Hip. 4.3 (condições de ortogonalidade): Para cada equação m, as K_m variáveis em $\boldsymbol{x}_{im}$ são pré-determinadas no sentido de que todas são ortogonais ao erro: $\mathrm{E}(x_{im}\varepsilon_{im}) = \mathbf{0}$ para todo i e $m(= 1, 2, ..., M)$. Ou seja, existem $\sum_m K_m$ condições de ortogonalidade no total. Se nós definirmos

$$\underset{(\sum_{m=1}^{M} K_m \times 1)}{\boldsymbol{g}_i} \equiv \begin{bmatrix} \boldsymbol{x}_{i1}.\varepsilon_{i1} \\ \vdots \\ \boldsymbol{x}_{iM}.\varepsilon_{iM} \end{bmatrix}. \tag{52}$$

então todas as condições de ortogonalidade podem ser escritas compactamente como

$$\mathrm{E}(g_i) = \mathbf{0} \tag{53}$$

Não é assumido *ortogonalidade cruzada*. Por exemplo, $\boldsymbol{x}_{i1}$ não precisa ser ortogonal a ε_{i2}, embora tenha que ser a ε_{i1}. Todavia, se a variável for incluída em ambos os vetores, então a hipótese implica que a variável é ortogonal a ambos ε_{i1} e ε_{i2}.

Exemplo. Assuma $\sum_m K_m$ como $8(= 4 + 4)$ e $\boldsymbol{g}_i$ é

$$\boldsymbol{g}_i = \begin{bmatrix} \varepsilon_{i1} \\ S_i\varepsilon_{i1} \\ EXPR_i\varepsilon_{i1} \\ MED_i\varepsilon_{i1} \\ \varepsilon_{i2} \\ S_i\varepsilon_{i2} \\ EXPR_i\varepsilon_{i2} \\ MED_i\varepsilon_{i2} \end{bmatrix}$$

Este é o caso pois $\boldsymbol{x}_{i1}$ e $\boldsymbol{x}_{i2}$ possuem o mesmo conjunto de instrumentos, cada instrumento é ortogonal a ε_{i1} e ε_{i2}.

Identificação

Com as condições de ortogonalidade podemos derivar as condições de indentificação.

- Na versão de equação múltipla:

$$\boldsymbol{g}(\boldsymbol{w}_i;\boldsymbol{\delta}) \equiv \begin{bmatrix} \boldsymbol{x}_{i1}.(y_{i1} - \boldsymbol{z}'_{i1}\delta_1) \\ \vdots \\ \boldsymbol{x}_{iM}.(y_{iM} - \boldsymbol{z}'_{iM}\delta_M) \end{bmatrix} \tag{54}$$

  e $\boldsymbol{\delta}$ sem o subescrito é o vetor "empilhado" dos coeficientes:

$$\boldsymbol{\delta} = [\boldsymbol{\delta}_1 \ldots \boldsymbol{\delta}_M]'.$$

Identificação

- O vetor de coeficientes é identificado se $\tilde{\boldsymbol{\delta}} = \boldsymbol{\delta}$ é a única solução para o sistema de equações:

$$\mathsf{E}[\boldsymbol{g}(\boldsymbol{w}_i; \tilde{\boldsymbol{\delta}})] = \mathbf{0}. \tag{55}$$

Nesse caso,

$$\mathsf{E}[\boldsymbol{g}(\boldsymbol{w}_i; \tilde{\boldsymbol{\delta}})] =$$

$$\begin{bmatrix} \mathsf{E}(\boldsymbol{x}_{i1}.y_{i1}) \\ \vdots \\ \mathsf{E}(\boldsymbol{x}_{iM}.y_{iM}) \end{bmatrix} - \begin{bmatrix} \mathsf{E}(\boldsymbol{x}_{i1}.\boldsymbol{z}_{i1}') & & \\ & \ddots & \\ & & \mathsf{E}(\boldsymbol{x}_{iM}.\boldsymbol{z}_{iM}') \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{\delta}}_1 \\ \vdots \\ \tilde{\boldsymbol{\delta}}_M \end{bmatrix}$$

$$\mathsf{E}[\boldsymbol{g}(\boldsymbol{w}_i; \tilde{\boldsymbol{\delta}})] \equiv \boldsymbol{\sigma}_{\boldsymbol{xy}} - \boldsymbol{\Sigma}_{\boldsymbol{xz}} \tilde{\boldsymbol{\delta}}.$$

Logo, o sistema de equações determinando $\tilde{\boldsymbol{\delta}}$ pode ser escrito como:

$$\boldsymbol{\Sigma}_{\boldsymbol{xz}}\tilde{\boldsymbol{\delta}} = \boldsymbol{\sigma}_{\boldsymbol{xy}}. \tag{56}$$

Identificação

- Uma condição para identificação é que $\boldsymbol{\Sigma}_{\boldsymbol{xz}}$ tenha posto completo. Mas como esta matriz é bloco-diagonal, esta condição é equivalente a:

Hipótese 4.4 (condição de posto para identificação): Para cada $m = 1, ..., M$, a matriz $\mathrm{E}(\boldsymbol{x}_{im}\boldsymbol{z}'_{im})$ é posto-completo (de ordem $K_m \times L_m$).

Normalidade Assintótica

Hipótese 4.5 (g_i é um mds com segundos momentos finitos): $\{\boldsymbol{g}_i\}$ é uma sequência diferença martingale conjunta. $\boldsymbol{S} = \mathsf{E}(\boldsymbol{g}_i\boldsymbol{g}_i')$ é não-singular.

- A matriz tem a seguinte estrutura:

$$\boldsymbol{S} = \underset{(\sum_{m=1}^{M} K_m \times \sum_{m=1}^{M} K_m)}{\mathsf{E}(\boldsymbol{g}_i\boldsymbol{g}_i')} = \tag{57}$$

$$= \begin{bmatrix} \mathsf{E}(\varepsilon_{i1}\varepsilon_{i1}\boldsymbol{x}_{i1}\boldsymbol{x}_{i1}') & \dots & \mathsf{E}(\varepsilon_{i1}\varepsilon_{iM}\boldsymbol{x}_{i1}\boldsymbol{x}_{iM}') \\ \vdots & & \vdots \\ \mathsf{E}(\varepsilon_{iM}\varepsilon_{i1}\boldsymbol{x}_{iM}\boldsymbol{x}_{i1}') & \dots & \mathsf{E}(\varepsilon_{iM}\varepsilon_{iM}\boldsymbol{x}_{iM}\boldsymbol{x}_{iM}') \end{bmatrix}$$

- Em suma, o modelo de múltiplas equações é um sistema onde se aplicam todas as hipóteses que fizemos para o modelo de equação simples, com a adição da estacionariedade conjunta.

Definição GMM

Definição como em única equação. Faça $\tilde{\boldsymbol{\delta}}$ ser um valor hipotético dos parâmetros verdadeiros $\boldsymbol{\delta}$ e defina $\boldsymbol{g}_n(\tilde{\boldsymbol{\delta}})$. A definição GMM é a mesma, mas a matriz de pesos $\widehat{\boldsymbol{W}}$ é agora $(\sum_{m=1}^{M} K_m \times \sum_{m=1}^{M} K_m)$. Agora podemos escrever a mesma de definição de $\boldsymbol{g}$ para o análogo amostral:

$$
\begin{aligned}
\underset{(\sum_{m=1}^{M} K_m \times 1}{\boldsymbol{g}_n \quad (\tilde{\boldsymbol{\delta}})} &= \qquad (58) \\
&= \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_{i1}(y_{i1} - \boldsymbol{z}'_{i1}\tilde{\boldsymbol{\delta}}_1) \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_{iM}(y_{iM} - \boldsymbol{z}'_{iM}\tilde{\boldsymbol{\delta}}_M) \end{bmatrix} \\
&= \boldsymbol{s}_{\boldsymbol{xy}} - \boldsymbol{S}_{\boldsymbol{xz}}\tilde{\boldsymbol{\delta}}
\end{aligned}
$$

Teoria para Amostras-Grandes

A teoria de amostras grandes é a mesma de GMM de única equação. Apenas é realizada a substituição dos indicadores.

- (Teste de hipóteses) No caso presente de equações múltiplas, $\boldsymbol{\delta}$ é um vetor empilhado composto por coeficientes de equações diferentes. As proposições 3.3 e 3.8 (testes t, W e LR) permitem testar restrições entre equações.

- (Teste de restrições sobre-identificadas) O número de condições de ortogonalidade é $\sum_{m=1}^{M} K_m$ e o número de coeficientes é $\sum_{m=1}^{M} L_m$. Os graus de liberdade para a estatística J na Proposição 3.6 são $\sum_{m=1}^{M} K_m - \sum_{m=1}^{M} L_m$ e para a estatística C

na Proposição 3.7 são o total de instrumentos suspeitos de diferentes equações.

Proposição 4.1 (estimação consistente do erro contemporâneo dos momentos entre equações): Faça $\widehat{\boldsymbol{\delta}}_m$ ser um estimador consistente de $\boldsymbol{\delta}_m$ e faça $\widehat{\varepsilon}_{im} \equiv y_{im} - \boldsymbol{z}'_{im}\widehat{\boldsymbol{\delta}}_m$ ser o resíduo de $m = 1, 2, \ldots, M$. Sob as hips 4.1 e 4.2, mais a hipótese de que $\mathsf{E}(\boldsymbol{z}_{im}\boldsymbol{z}'_{ih})$ existe e é finito para todo m, e $h (= 1, \ldots, M)$,

$$\widehat{\sigma}_{mh} \rightarrow_p \sigma_{mh}$$

tal que

$$\widehat{\sigma}_{mh} = \frac{1}{n}\sum_{i=1}^{n} \widehat{\varepsilon}_{im}\widehat{\varepsilon}'_{ih} \text{ e } \sigma_{mh} \equiv \mathsf{E}(\boldsymbol{\varepsilon}_{im}\boldsymbol{\varepsilon}'_{ih})$$

dado que $\mathrm{E}(\boldsymbol{\varepsilon}_{im}\boldsymbol{\varepsilon}'_{ih})$ existe e é finito.

$$\widehat{\sigma}_{mh} \to_p \sigma_{mh} \tag{59}$$

tal que

$$\widehat{\sigma}_{mh} \equiv \frac{1}{n}\sum_{i=1}^{n} \widehat{\varepsilon}_{im}\widehat{\varepsilon}_{ih} \text{ e } \sigma_{mh} \equiv \mathrm{E}(\varepsilon_{im}\varepsilon_{ih})$$

dado que $\mathrm{E}(\varepsilon_{im}\varepsilon_{ih})$ existe e é finito.

Estimando S consistentemente

Hipótese 4.6 (quarto momento finito): $\mathrm{E}[(x_{imk}z_{ihj})^2]$ existe e é finito para todo k, $j = (1, 2, ..., K)$, m e $h(= 1, ..., M)$, tal que x_{imk} é o k-ésimo elemento de $\boldsymbol{x}_{im}$ e z_{ihj} é o j-ésimo elemento de $\boldsymbol{z}_{ih}$.

Proposição 4.2 (estimação consistente de S, a variância assintót
Faça $\widehat{\delta}_m$ ser um estimador consistente de δ_m, e faça $\widehat{\varepsilon}_{im} \equiv y_{im} - \boldsymbol{z}'_{im}\widehat{\delta}_m$ ser o resíduo para $m = 1, \ldots, M$. Sob as hipóteses 4.1, 4.2 e 4.6, $\widehat{\boldsymbol{S}}$ é consistente para $\boldsymbol{S}$. Tal que:

$$\widehat{\boldsymbol{S}} = \frac{1}{n}\sum_{i=1}^{n} \widehat{\varepsilon}_{im}\widehat{\varepsilon}_{im}\boldsymbol{x}_{im}\boldsymbol{x}'_{im}$$

Table 4.1: Multiple-Equation GMM in the Single-Equation Format

Sample Analogue of Orthogonality Conditions:	$\mathbf{g}_n(\tilde{\boldsymbol{\delta}}) = \mathbf{s}_{\mathbf{xy}} - \mathbf{S}_{\mathbf{xz}}\tilde{\boldsymbol{\delta}} = \mathbf{0}$
GMM Estimator:	$\hat{\boldsymbol{\delta}}(\widehat{\mathbf{W}}) = (\mathbf{S}'_{\mathbf{xz}}\widehat{\mathbf{W}}\mathbf{S}_{\mathbf{xz}})^{-1}\mathbf{S}'_{\mathbf{xz}}\widehat{\mathbf{W}}\mathbf{s}_{\mathbf{xy}}$
Its Sampling Error:	$\hat{\boldsymbol{\delta}}(\widehat{\mathbf{W}}) - \boldsymbol{\delta} = (\mathbf{S}'_{\mathbf{xz}}\widehat{\mathbf{W}}\mathbf{S}_{\mathbf{xz}})^{-1}\mathbf{S}'_{\mathbf{xz}}\widehat{\mathbf{W}}\bar{\mathbf{g}}$
Asymptotic Variance of Optimal GMM:	$\text{Avar}(\hat{\boldsymbol{\delta}}(\widehat{\mathbf{S}}^{-1})) = (\boldsymbol{\Sigma}'_{\mathbf{xz}}\mathbf{S}^{-1}\boldsymbol{\Sigma}_{\mathbf{xz}})^{-1}$
Its Estimator:	$\widehat{\text{Avar}(\hat{\boldsymbol{\delta}}(\widehat{\mathbf{S}}^{-1}))} = (\mathbf{S}'_{\mathbf{xz}}\widehat{\mathbf{S}}^{-1}\mathbf{S}_{\mathbf{xz}})^{-1}$
J Statistic:	$J(\hat{\boldsymbol{\delta}}(\widehat{\mathbf{S}}^{-1}), \widehat{\mathbf{S}}^{-1}) = n \cdot \mathbf{g}_n(\hat{\boldsymbol{\delta}}(\widehat{\mathbf{S}}^{-1}))'\widehat{\mathbf{S}}^{-1}\mathbf{g}_n(\hat{\boldsymbol{\delta}}(\widehat{\mathbf{S}}^{-1}))$

	Single-Equation GMM applied to the equation in question	Multiple-Equation GMM
$\mathbf{g}_i$	$\mathbf{x}_i \cdot \varepsilon_i$	(4.1.4)
$\boldsymbol{\delta}$	$\boldsymbol{\delta}$	(4.1.6)
$\mathbf{s}_{\mathbf{xy}}$	$\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i \cdot y_i$	(4.2.2)
$\mathbf{S}_{\mathbf{xz}}$	$\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{z}'_i$	(4.2.2)
Size of $\mathbf{W}$	$K \times K$	$\sum_m K_m \times \sum_m K_m$
$\boldsymbol{\Sigma}_{\mathbf{xz}}$	$\text{E}(\mathbf{x}_i\mathbf{z}'_i)$	(4.1.9)
$\mathbf{S}$ (= Avar($\bar{\mathbf{g}}$))	$\text{E}(\varepsilon_i^2\mathbf{x}_i\mathbf{x}'_i)$	(4.1.11)
$\widehat{\mathbf{S}}$	$\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2\mathbf{x}_i\mathbf{x}'_i$	(4.3.2)
Estimator consistent under which assumptions?	3.1–3.4	4.1–4.4
Estimator asymptotic normal under which assumptions?	3.1–3.5	4.1–4.5
$\widehat{\mathbf{S}} \to_p \mathbf{S}$ under which assumptions?	3.1, 3.2, 3.6, $\text{E}(\mathbf{g}_i\mathbf{g}'_i)$ finite	4.1, 4.2, 4.6, $\text{E}(\mathbf{g}_i\mathbf{g}'_i)$ finite
d.f. of J	$K - L$	$\sum_m (K_m - L_m)$

Se pelo menos uma das M equações é sobreidentificada então a escolha de $\hat{\boldsymbol{W}}$ afeta o valor do estimador GMM, quando comparado com a alterativa de estimar cada M equação em separado.

Proposição 4.3 (equivalência entre estimação GMM simples e po

Se todas as equações são exatamente identificadas, então a estimação GMM equação por equação e por múltiplas equações são numericamente os mesmos e igual ao estimador IV.

- Se ao menos uma equação é sobreidentificada mas a equação e "não-relacionada" no sentindo de $\mathrm{E}(\varepsilon_{im}\varepsilon_{ih})\boldsymbol{x}_{im}\boldsymbol{x}'_{ih} = 0$, então a estimativa eficiente equação por equação e por equação múltipla são assintoticamente equivalentes dado que $\sqrt{n}$ vezes a diferença converge em probabilidade para zero.

Homecedasticidade Condicional

Assumindo homocedasticidade condicional:

Hipótese 4.7 (homocedasticidade condicional:)

$$\mathsf{E}(\varepsilon_{im}\varepsilon_{ih} \mid \boldsymbol{x}_{im}, \boldsymbol{x}_{ih}) = \sigma_{mh}$$

para todo m, $h = 1, 2, ..., M$.

Portanto, os momentos cruzados condicionais E($\varepsilon_{im}\varepsilon_{ih}$ é igual a σ_{mh} pelo lei das expectativas totais.

Portanto a $\boldsymbol{S}$ em (4.1.11) pode ser escrita como:

$$\boldsymbol{S} = \begin{bmatrix} \sigma_{11}\mathsf{E}\left(\boldsymbol{x}_{i1}\boldsymbol{x}'_{i1}\right) & \dots & \sigma_{1M}\left(\mathsf{E}\boldsymbol{x}_{i1}\boldsymbol{x}'_{iM}\right) \\ \vdots & & \vdots \\ \sigma_{M1}\mathsf{E}\left(\boldsymbol{x}_{iM}\boldsymbol{x}'_{i1}\right) & \dots & \sigma_{MM}\left(\mathsf{E}\boldsymbol{x}_{iM}\boldsymbol{x}'_{iM}\right) \end{bmatrix} \tag{60}$$

Como pela Hipótese 4.5 $\boldsymbol{S}$ é finita, esta decomposição implica que $\mathrm{E}(\boldsymbol{x}_{im}\boldsymbol{x}'_{ih})$ existe e é finito para todo m, $h(= 1, 2, ..., M)$.

FIVE: Full-Information Instrumental Variables Efficient

Um estimador de $\boldsymbol{S}$ explorando a estrutura do quarto momento mostrado em (60)

$$\widehat{\boldsymbol{S}} = \begin{bmatrix} \widehat{\sigma}_{11}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{i1}\boldsymbol{x}_{i1}'\right) & \cdots & \widehat{\sigma}_{1M}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{i1}\boldsymbol{x}_{iM}'\right) \\ \vdots & & \vdots \\ \widehat{\sigma}_{M1}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{iM}\boldsymbol{x}_{i1}'\right) & \cdots & \widehat{\sigma}_{MM}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{iM}\boldsymbol{x}_{iM}'\right) \end{bmatrix} \quad (61)$$

Neste caso, para algum estimador consistente $\widehat{\delta}_m$ de $\boldsymbol{\delta}$,

$$\widehat{\sigma}_{mh} \equiv \frac{1}{n}\sum_{i=1}^{n}\widehat{\varepsilon}_{im}\widehat{\varepsilon}_{ih}, \text{ com } \widehat{\varepsilon}_{ih} \equiv y_{im} - \boldsymbol{z}'\widehat{\boldsymbol{\delta}}_m \quad (62)$$

FIVE: Full-Information Instrumental Variables Efficient

Estimador FIVE $\widehat{\boldsymbol{\delta}}_{FIVE} \equiv \widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1})$, com $\widehat{\boldsymbol{S}}^{-1}$ dada por (61).

Proposição 4.4 (propriedades de amostras grandes para FIVE): Suponha que as Hipóteses 4.1-4.5 e 4.7 valem. Suponha também que E($\boldsymbol{z}_{im}\boldsymbol{z}'_{ih}$) existe para todo m e h. Faça $\boldsymbol{S}$ e $\widehat{\boldsymbol{S}}$ como definido em (60) e (61), respectivamente. Então:

- $\widehat{\boldsymbol{S}} \to_p \boldsymbol{S}$;

- $\widehat{\boldsymbol{\delta}}_{FIVE} \equiv \widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1})$ é consistente, assintoticamente normal e eficiente com Avar($\widehat{\boldsymbol{\delta}}_{FIVE}$) dado por (37) (ver GMM cap. 3);

- A variância assintótica estimada dada por (38) é consistente para Avar($\widehat{\boldsymbol{\delta}}_{FIVE}$);

- Estatística Sargan dada por

$$J(\widehat{\boldsymbol{\delta}}_{FIVE}, \widehat{\boldsymbol{S}}^{-1}) = n.\boldsymbol{g}_n(\widehat{\boldsymbol{\delta}}_{FIVE})'\widehat{\boldsymbol{S}}^{-1}\boldsymbol{g}_n(\widehat{\boldsymbol{\delta}}_{FIVE}) \rightarrow_d \chi^2(K_m - L_m). \tag{63}$$

tal que $\boldsymbol{g}_n(\cdot)$ é dada por (58).

3SLS

Quando os instrumentos do modelo FIVE são os mesmos entre equações então temos o estimador 3SLS: $\hat{\boldsymbol{\delta}}_{3SLS} \equiv \hat{\boldsymbol{\delta}}(\hat{\boldsymbol{S}}^{-1})$.

Defina:

$$\underset{(M \times 1)}{\boldsymbol{\varepsilon}_i} = \begin{bmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iM} \end{bmatrix}$$

A matriz dos momentos cruzados de $\boldsymbol{\varepsilon}_i$ é:

$$\underset{(M \times M)}{\boldsymbol{\Sigma}} = \mathsf{E}(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i') = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1M} \\ \vdots & & \vdots \\ \sigma_{M1} & \cdots & \sigma_{MM} \end{bmatrix} \tag{64}$$

Para estimar $\boldsymbol{\Sigma}$ precisamos de um estimador inicial consisntente de $\boldsymbol{\delta}_m$ para calcular $\widehat{\varepsilon}_{im}$. Um estimador 2SLS é usado para calcular esta estimativa inicial. Dados estes resíduos um estimador natural de

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n} \widehat{\boldsymbol{\varepsilon}}_i \widehat{\boldsymbol{\varepsilon}}_i' \tag{65}$$

Se $\boldsymbol{x}_i = \boldsymbol{x}_{i1} = \boldsymbol{x}_{i2} = ...$ é o conjunto comum de instrumentos (dimensão K), então $\boldsymbol{g}_i$ dado por (52), $\boldsymbol{S}$ por (60) e $\widehat{\boldsymbol{S}}$ por (61) pode ser escrito como:

$$\underset{(MK\times 1)}{\boldsymbol{g}_i} = \boldsymbol{\varepsilon}_i \otimes \boldsymbol{x}_i \tag{66}$$

$$\underset{(MK\times MK)}{\boldsymbol{S}} = \underset{(M\times M)}{\boldsymbol{\Sigma}} \otimes \underset{(K\times K)}{\mathsf{E}(\boldsymbol{x}_i \boldsymbol{x}_i')} \tag{67}$$

com

$$S^{-1} = \Sigma^{-1} \otimes [\mathsf{E}(x_i x_i')]^{-1}$$

$$\widehat{S} = \widehat{\Sigma} \otimes \left(\frac{1}{n} \sum_{i=1}^{n} x_i x_i' \right) \tag{68}$$

então

$$\widehat{S}^{-1} = \widehat{\Sigma}^{-1} \otimes \left(\frac{1}{n} \sum_{i=1}^{n} x_i x_i' \right)^{-1}$$

(Σ e $\mathsf{E}(x_i x_i')$ não são singulares.) A matriz e pesos é:

$$\widehat{W}_{mh} = \widehat{\sigma}^{mh} \left(\frac{1}{n} \sum_{i=1}^{n} x_i x_i' \right)^{-1} \tag{69}$$

aqui $\widehat{\sigma}^{mh}$ é o elemento (m, h) de $\widehat{\Sigma}^{-1}$. Substitua $\widehat{W}$ na equação

$\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}})$ (4.2.6):

$$\widehat{\boldsymbol{\delta}}_{3SLS} = \left[\begin{array}{ccc} \widehat{\sigma}^{11}\widehat{\boldsymbol{A}}_{11} & \ldots & \widehat{\sigma}^{1M}\widehat{\boldsymbol{A}}_{1M} \\ \vdots & & \vdots \\ \widehat{\sigma}^{M1}\widehat{\boldsymbol{A}}_{M1} & \ldots & \widehat{\sigma}^{MM}\widehat{\boldsymbol{A}}_{MM} \end{array} \right]^{-1} \left[\begin{array}{ccc} \widehat{\sigma}^{11}\widehat{\boldsymbol{c}}_{11} & \ldots & \widehat{\sigma}^{1M}\widehat{\boldsymbol{c}}_{1M} \\ \vdots & & \vdots \\ \widehat{\sigma}^{M1}\widehat{\boldsymbol{c}}_{M1} & \ldots & \widehat{\sigma}^{MM}\widehat{\boldsymbol{c}}_{MM} \end{array} \right] \tag{70}$$

tal que

$$\widehat{\boldsymbol{A}}_{mh} = \left(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{z}_{im}\boldsymbol{x}_i' \right) \left(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i' \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{z}_{ih}' \right) \tag{71}$$

$$\widehat{\boldsymbol{c}}_{mh} = \left(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{z}_{im}\boldsymbol{x}_i' \right) \left(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i' \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i y_{ih} \right) \tag{72}$$

A expressão para a variância é:

$$\text{Avar}(\hat{\boldsymbol{\delta}}_{3SLS}) = \begin{bmatrix} \sigma^{11}\boldsymbol{A}_{11} & \dots & \sigma^{1M}\boldsymbol{A}_{1M} \\ \vdots & & \vdots \\ \sigma^{M1}\boldsymbol{A}_{M1} & \dots & \sigma^{MM}\boldsymbol{A}_{MM} \end{bmatrix}^{-1} \tag{73}$$

para $\hat{\boldsymbol{A}}_{mh} = \mathsf{E}\left(\boldsymbol{z}_{im}\boldsymbol{x}_i'\right)\mathsf{E}\left(\boldsymbol{x}_i\boldsymbol{x}_i'\right)^{-1}\mathsf{E}\left(\boldsymbol{x}_i\boldsymbol{z}_{ih}'\right)$ A matriz da variância é consistentemente estimada por:

$$\widehat{\text{Avar}(\hat{\boldsymbol{\delta}}_{3SLS})} = \begin{bmatrix} \hat{\sigma}^{11}\hat{\boldsymbol{A}}_{11} & \dots & \hat{\sigma}^{1M}\hat{\boldsymbol{A}}_{1M} \\ \vdots & & \vdots \\ \hat{\sigma}^{M1}\hat{\boldsymbol{A}}_{M1} & \dots & \hat{\sigma}^{MM}\hat{\boldsymbol{A}}_{MM} \end{bmatrix}^{-1} \tag{74}$$

Proposição 4.5 (propriedades de amostras grandes para 3SLS): Suponha que as Hipóteses 4.1-4.5 e 4.7 valem e que os instrumentos são comuns: $\boldsymbol{x}_{im} = \boldsymbol{x}_i$. Suponha também que

E$(\boldsymbol{z}_{im}\boldsymbol{z}'_{ih})$ existe para todo m e h. Faça $\widehat{\boldsymbol{\Sigma}}$ ser a matriz de momentos cruzados do termo de erro calculada por (65) usando os resíduos de 2SLS. Então:

- $\widehat{\boldsymbol{\delta}}_{3SLS}$ dado por (70) é consistente, assintoticamente normal e eficiente com Avar$(\widehat{\boldsymbol{\delta}}_{3SLS})$ dada por (73);

- A variância assintótica estimada dada por (74) é consistente para Avar$(\widehat{\boldsymbol{\delta}}_{3SLS})$;

- Estatística Sargan dada por

$$J(\widehat{\boldsymbol{\delta}}_{3SLS}, \widehat{\boldsymbol{S}}^{-1}) = n.\boldsymbol{g}_n(\widehat{\boldsymbol{\delta}}_{3SLS})'\widehat{\boldsymbol{S}}^{-1}\boldsymbol{g}_n(\widehat{\boldsymbol{\delta}}_{3SLS}) \rightarrow_d \chi^2(MK - L_m). \tag{75}$$

  tal que $\boldsymbol{g}_n(\cdot)$ é dada por (58), K é o número de instrumentos comuns e $\widehat{\boldsymbol{S}} = \widehat{\boldsymbol{\Sigma}} \otimes \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}'_i\right)$.

SUR – Seemingly Unrelated Regression

3SLS pode ser simplificado se

$$\boldsymbol{x}_i = \text{união de } (\boldsymbol{z}_{i1}, ..., \boldsymbol{z}_{iM}) \tag{76}$$

Isto é equivalente a condição

$$\mathrm{E}(\boldsymbol{z}_{im}\varepsilon_{ih}) = 0 \tag{77}$$

Isto significa que os regressores predeterminados satisfazem as ortogonalidades cruzadas. Aqui eles não são apenas determinados em cada equação, $\mathrm{E}(\boldsymbol{z}_{im}\varepsilon_{im}) = 0$, mas também predeterminados entre equações, $\mathrm{E}(\boldsymbol{z}_{im}\varepsilon_{ih}) = 0$, para $h \neq m$. Esta forma simplificada é chamada de estimativa SUR, $\widehat{\boldsymbol{\delta}}_{SUR}$.

Como SUR é um caso especial do 3SLS as fórmulas que se aplicam para ambos os casos. A implicação do estimador SUR é que nas expressões para $\widehat{\boldsymbol{A}}_{mh}$, $\widehat{\boldsymbol{c}}_{mh}$ e $\boldsymbol{A}_{mh}$ o $\boldsymbol{x}_i$ desaparece.

$$\widehat{\boldsymbol{A}}_{mh} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{z}_{im} \boldsymbol{z}'_{ih} \tag{78}$$

$$\widehat{\boldsymbol{c}}_{mh} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{z}_{im} y_{ih} \tag{79}$$

$$\widehat{\boldsymbol{A}}_{mh} = \mathsf{E}(\boldsymbol{z}_{im} \boldsymbol{z}'_{ih}) \tag{80}$$

Proposição 4.6 (propriedades de amostras grandes para SUR): Suponha que as Hipóteses 4.1-4.5 e 4.7 valem e que os instrumentos são: $\boldsymbol{x}_i =$ união de $(\boldsymbol{z}_{i1}, ..., \boldsymbol{z}_{iM})$. Faça $\widehat{\boldsymbol{\Sigma}}$ ser a matriz de momentos cruzados do termo de erro calculada por (65) usando os resíduos OLS. Então:

- $\widehat{\boldsymbol{\delta}}_{SUR}$ com $\widehat{\boldsymbol{A}}_{mh}$ dada por (78), $\widehat{\boldsymbol{c}}_{mh}$ dado por (79) é consistente, assintoticamente normal e eficiente com Avar($\widehat{\boldsymbol{\delta}}_{SUR}$) dada por (73) com $\widehat{\boldsymbol{A}}_{mh}$ dado por (80).

- A variância assintótica estimada dada por (74), $\widehat{\boldsymbol{A}}_{mh}$ dada por (78), é consistente para Avar($\widehat{\boldsymbol{\delta}}_{SUR}$);

- Estatística Sargan dada por

$$J(\widehat{\boldsymbol{\delta}}_{SUR}, \widehat{\boldsymbol{S}}^{-1}) = n.\boldsymbol{g}_n(\widehat{\boldsymbol{\delta}}_{SUR})'\widehat{\boldsymbol{S}}^{-1}\boldsymbol{g}_n(\widehat{\boldsymbol{\delta}}_{SUR}) \rightarrow_d \chi^2(MK - \sum_m L_m). \tag{81}$$

  tal que $\boldsymbol{g}_n(\cdot)$ é dada por (58), K é o número de instrumentos comuns e $\widehat{\boldsymbol{S}} = \widehat{\boldsymbol{\Sigma}} \otimes \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i'\right)$.

Eficiência do SUR: exemplo modelo de regressão multivariada com restrições de exclusão – SUR mais eficiente do que OLS equação por equação. Suponha o sistema:

$$LW_i = \phi_1 + \beta_1 S_i + \gamma_1 IQ_i + \pi_1 EXPR_i + \varepsilon_{i1}$$

$$KWW_i = \phi_2 + \beta_2 S_i + \gamma_2 IQ_i + \pi_2 EXPR_i + \varepsilon_{i2}$$

O conjunto de instrumentos comuns é $(1, S_i, IQ_i, EXPR_i)$. Este sistema é um modelo de regressão multivariada com o mesmo conjunto de regressores. Mas se π_2 é a priori restrito para ser 0, significando que $EXPR_i$ é excluida da segunda equação, então o modelo se torna o SUR. O SUR é mais eficiente do que o modelo multivariado pois explora a restrição de exclusão.

Coeficiente Comum

Panel data: caso particular do modelo GMM de múltiplas equações impondo restrição de regressores comuns.

O sistema agora pode ser escrito como:

Hip. 4.1 (linearidade): Existem M equações lineares:

$$y_{im} = \boldsymbol{z}'_{im}\boldsymbol{\delta} + \varepsilon_{im} \quad (i = 1, 2, ...)(m = 1, ..., M).$$

$\boldsymbol{z}$ é o vetor de regressores, $\boldsymbol{\delta}$ é o vetor dos coeficientes $(L \times 1)$ comum a todas as equações. O modelo não faz hipóteses sobre a correlação entre os erros das equações. Além disso, não há restrições sobre os coeficientes de equações diferentes.

Apenas é necessário modificar a Hipótese 4.4 (identificação). A versão de $\boldsymbol{g}(\cdot)$ é a seguinte:

$$\boldsymbol{g}(\boldsymbol{w}_i;\tilde{\boldsymbol{\delta}}) \equiv \begin{bmatrix} \boldsymbol{x}_{i1}\cdot(y_{i1} - \boldsymbol{z}'_{i1}\tilde{\boldsymbol{\delta}}) \\ \vdots \\ \boldsymbol{x}_{iM}\cdot(y_{iM} - \boldsymbol{z}'_{iM}\tilde{\boldsymbol{\delta}}) \end{bmatrix} \tag{82}$$

neste caso temos

$$\mathsf{E}[\boldsymbol{g}(\boldsymbol{w}_i;\tilde{\boldsymbol{\delta}})] = \begin{bmatrix} \mathsf{E}(\boldsymbol{x}_{i1}\cdot y_{i1}) \\ \vdots \\ \mathsf{E}(\boldsymbol{x}_{iM}\cdot y_{iM}) \end{bmatrix} - \begin{bmatrix} \mathsf{E}(\boldsymbol{x}_{i1}\cdot\boldsymbol{z}'_{i1})\tilde{\boldsymbol{\delta}}) \\ \vdots \\ \mathsf{E}(\boldsymbol{x}_{iM}\cdot\boldsymbol{z}'_{iM})\tilde{\boldsymbol{\delta}}) \end{bmatrix} \tag{83}$$

$$\mathsf{E}[\boldsymbol{g}(\boldsymbol{w}_i;\tilde{\boldsymbol{\delta}})] = \underset{(\sum_{m=1}^{M} K_m\times 1)}{\boldsymbol{\sigma}_{\boldsymbol{xy}}} - \underset{(\sum_{m=1}^{M} K_m\times L)}{\boldsymbol{\Sigma}_{\boldsymbol{xz}}} \underset{(L\times 1)}{\tilde{\boldsymbol{\delta}}}$$

A implicação de coeficientes comuns significa que a matriz $\boldsymbol{\Sigma}_{\boldsymbol{xz}}$ é agora empilhada, não uma matriz bloco diagonal.

A condição para identificação é:

Hipótese 4.4' (condição de posto): A matriz $\underset{(\sum_{m=1}^{M} K_m \times L)}{\boldsymbol{\Sigma_{xz}}}$ definida em (83) é posto completo.

Essa condição é mais fraca do que a Hipótese 4.4, que requer que cada equação do sistema seja identificada. Portanto, uma condição suficiente para identificação é que E$(\boldsymbol{x}_{im}.\boldsymbol{z}'_{im})$ seja posto completo para algum m. É possível que o sistema seja identificado mesmo se nenhuma das equações seja identificada individualmente.

GMM – Coeficiente Comum

O estimador GMM é:

$$
\begin{aligned}
\hat{\boldsymbol{\delta}}(\hat{\boldsymbol{W}}) = & \\
= & \left[\sum_{m=1}^{M}\sum_{h=1}^{M}\left\{\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_{im}.\boldsymbol{x}'_{im}\right)\hat{\boldsymbol{W}}_{mh}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{ih}.\boldsymbol{z}'_{ih}\right)\right\}\right]^{-1} \\
& \left[\sum_{m=1}^{M}\sum_{h=1}^{M}\left\{\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_{im}.\boldsymbol{x}'_{im}\right)\hat{\boldsymbol{W}}_{mh}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{ih}.y_{ih}\right)\right\}\right] \quad (84)
\end{aligned}
$$

O estimador eficiente é obtido com $\hat{\boldsymbol{W}}$ sendo substituido pela inversa de $\hat{\boldsymbol{S}}$.

Impondo Homocedasticidade Condicional

Como antes se pode impor homocedasticidade condicional. O estimador FIVE é obtido se é aplicada a correspondente matriz $\widehat{\boldsymbol{S}}$.

Assumindo que os conjuntos de instrumentos são os mesmos entre as equações, então, se tem a estimativa de $\widehat{\boldsymbol{S}}$:

$$\widehat{\boldsymbol{S}} = \widehat{\boldsymbol{\Sigma}} \otimes \left(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' \right) \tag{85}$$

data a estimativa de $\widehat{\boldsymbol{W}}_{mh}$ se chega no estimador 3SLS com coeficientes comuns:

$$\widehat{\boldsymbol{\delta}}_{3SLS} =$$
$$= \left[\sum_{m=1}^{M}\sum_{h=1}^{M}\left\{\widehat{\sigma}^{mh}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_{im}.\boldsymbol{x}_i'\right)\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i.\boldsymbol{x}_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i.\boldsymbol{z}_{ih}'\right)\right\}\right]^{-1}$$
$$\left[\sum_{m=1}^{M}\sum_{h=1}^{M}\left\{\widehat{\sigma}^{mh}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_{im}.\boldsymbol{x}_i'\right)\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i.\boldsymbol{x}_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i.y_{ih}\right)\right\}\right] \tag{86}$$

tal que $\widehat{\sigma}^{mh}$ é o elemento (m, h) de $\widehat{\boldsymbol{\Sigma}}^{-1}$.

Estimador de Efeitos Aleatórios

Assumindo as condições do estimador SUR, "desaparecimento de $\boldsymbol{x}$," o estimador GMM passa a ser o estimador de *efeitos-aleatórios*:

$$\hat{\boldsymbol{\delta}}_{RE} = \left[\sum_{m=1}^{M}\sum_{h=1}^{M}\hat{\sigma}^{mh}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_{im}.\boldsymbol{z}'_{ih}\right)\right]^{-1} \left[\sum_{m=1}^{M}\sum_{h=1}^{M}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_{im}.y_{ih}\right)\right] \quad (87)$$

A variância assintótica é

$$\text{Avar}(\hat{\boldsymbol{\delta}}_{RE}) = \left[\sum_{m=1}^{M}\sum_{h=1}^{M}\sigma^{mh}\mathsf{E}\left(\boldsymbol{z}_{im}.\boldsymbol{z}'_{ih}\right)\right]^{-1} \quad (88)$$

$$\widehat{\mathsf{Avar}(\widehat{\boldsymbol{\delta}}_{RE})} = \left[\sum_{m=1}^{M} \sum_{h=1}^{M} \widehat{\sigma}^{mh} \left(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{z}_{im} \cdot \boldsymbol{z}_{ih}' \right) \right]^{-1} \tag{89}$$

Estimador de Efeitos Aleatórios

Proposição 4.7 (propriedades de amostras grandes para RE): Suponha que as Hipóteses 4.1', 4.2, 4.3, 4.4', 4.5 e 4.7 valem e que os instrumentos são: $\boldsymbol{x}_i =$ união de $(\boldsymbol{z}_{i1}, ..., \boldsymbol{z}_{iM})$. Faça $\hat{\boldsymbol{\Sigma}}$ ser a matriz de momentos cruzados do termo de erro calculada por (65) usando os resíduos OLS e $\hat{\boldsymbol{\Sigma}}$ uma estimativa consistente de $\boldsymbol{\Sigma}$. Então:

- $\widehat{\boldsymbol{\delta}}_{RE}$ é consistente, assintoticamente normal e eficiente com Avar($\widehat{\boldsymbol{\delta}}_{RE}$) dada por (88).

- A variância assintótica estimada dada por (89) é consistente para Avar($\widehat{\boldsymbol{\delta}}_{RE}$);

- Estatística Sargan dada por

$$J(\widehat{\boldsymbol{\delta}}_{RE}, \widehat{\boldsymbol{S}}^{-1}) = n.\boldsymbol{g}_n(\widehat{\boldsymbol{\delta}}_{RE})'\widehat{\boldsymbol{S}}^{-1}\boldsymbol{g}_n(\widehat{\boldsymbol{\delta}}_{RE}) \to_d \chi^2(MK - L). \tag{90}$$

tal que $\boldsymbol{g}_n(\cdot)$ é dada por (58), K é o número de instrumentos comuns e $\widehat{\boldsymbol{S}} = \widehat{\boldsymbol{\Sigma}} \otimes \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i'\right)$.

Pooled OLS

Para o modelo SUR foi estimado $\hat{\boldsymbol{\Sigma}}$ a partir dos resíduos OLS (equação por equação). Vamos assumir que os coeficientes a serem estimados são os mesmos para todas as equações na estimação de $\hat{\boldsymbol{\Sigma}}$. Então considere determinar $\hat{\boldsymbol{W}}$ como

$$\boldsymbol{I}_M \otimes \left(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' \right)^{-1} \tag{91}$$

ao invés de

$$\hat{\boldsymbol{\Sigma}}^{-1} \otimes \left(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' \right)^{-1} \tag{92}$$

no primeiro estágio da estimativa GMM para obter a estimativa consistente inicial de $\boldsymbol{\delta}$. Este estimador é o RE (87) com $\widehat{\sigma}^{mh} = 1$

para $m = h$ e 0 para $m \neq h$, que pode ser escrito como:

$$\begin{aligned} \hat{\boldsymbol{\delta}}_{POLS} &= \left[\sum_{m=1}^{M}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_{im}.\boldsymbol{z}'_{im}\right)\right]^{-1}\left[\sum_{m=1}^{M}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_{im}.y_{im}\right)\right] \\ &= \left(\sum_{i=1}^{n}\sum_{m=1}^{M}\boldsymbol{z}_{im}.\boldsymbol{z}'_{im}\right)^{-1}\left(\sum_{i=1}^{n}\sum_{m=1}^{M}\boldsymbol{z}_{im}.y_{im}\right) \end{aligned} \quad (93)$$

Este é o modelo OLS com amostra de tamanho nM onde as observações são empinhadas nas equações. Por esta razão o estimador é chamado de *Pooled OLS*.

A variância assintótica para o POLS é

$$\text{Avar}(\hat{\boldsymbol{\delta}}_{POLS}) = \left(\sum_{m=1}^{M}\text{E}(\boldsymbol{z}_{im}.\boldsymbol{z}'_{im})\right)^{-1}\sum_{m=1}^{M}\sum_{h=1}^{M}\sigma_{mh}\text{E}(\boldsymbol{z}_{im}.\boldsymbol{z}'_{ih})\left(\sum_{m=1}^{M}\text{E}(\boldsymbol{z}_{im} \quad (94)$$

$$\widehat{\text{Avar}(\hat{\boldsymbol{\delta}}_{POLS})} = \left(\sum_{m=1}^{M} \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{z}_{im}.\boldsymbol{z}'_{im}\right)^{-1} \sum_{m=1}^{M} \sum_{h=1}^{M} \hat{\sigma}_{mh} \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{z}_{im}.\boldsymbol{z}'_{ih} \left(\sum_{m=1}^{M} \frac{1}{n} \right. \tag{95}$$

Como o estimador é consistente, então o resíduo pode ser usado para calcular $\hat{\sigma}_{mh}$ nesta expressão. Os erros-padrão corretos do Pooled OLS são a raiz quadrada de $(1/n\times)$ os elementos da diagonal desta matriz.

Flexibilidade definições

Um modelo completo de duas equações se encaixa na Definição 4.1'. Suponha o exemplo 4.1 e defina as matrizes da seguinte forma:

$$\boldsymbol{z}_{i1} = \begin{bmatrix} 1 \\ S_i \\ IQ_i \\ EXPR_i \\ 0 \\ 0 \\ 0 \end{bmatrix}, \ \boldsymbol{z}_{i1} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ S_i \\ EXPR_i \end{bmatrix}, \ \boldsymbol{z}_{i1} = \begin{bmatrix} \phi_1 \\ \beta_1 \\ \gamma_1 \\ \pi \\ \phi_2 \\ \beta_2 \\ \gamma_2 \end{bmatrix}$$

Um sistema com apenas um subconjunto de variáveis comuns também pode ser escrito na forma da Hipótese 4.1. Considere o Exemplo 4.2 assumindo que IQ e S são constantes no tempo. Então:

$$\boldsymbol{z}_{i1} = \begin{bmatrix} 1 \\ 0 \\ S69_i \\ IQ_i \\ EXPR69_i \end{bmatrix}, \ \boldsymbol{z}_{i1} = \begin{bmatrix} 1 \\ 0 \\ S80_i \\ IQ_i \\ EXPR80_i \end{bmatrix}, \ \boldsymbol{z}_{i1} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \beta \\ \gamma \\ \pi \end{bmatrix}$$

Assim a restrição de coeficientes comuns não é tão restritiva.

Modelo de Componentes de Erro (Cap.5)

Modelo de múltiplas equações com coeficientes comuns (Proposição 4.7).

Considere um sistema de M equações lineares, amostras aleatórias e as hipóteses "SUR":

$$\boldsymbol{y}_i = \boldsymbol{Z}_i \boldsymbol{\delta} + \boldsymbol{\varepsilon}_i \tag{96}$$

$$\{\boldsymbol{y}_i, \boldsymbol{Z}_i\} \text{ são i.i.d.} \tag{97}$$

$$\mathrm{E}(\boldsymbol{z}_{im} \varepsilon_{ih}) = 0 \tag{98}$$

neste caso $\mathrm{E}(\boldsymbol{\varepsilon}_i \otimes \boldsymbol{x}_i) = 0$ tal que $\boldsymbol{x}$ é a união de todos os $\boldsymbol{z}$.

Identificação: $\mathrm{E}(\boldsymbol{Z}_i \otimes \boldsymbol{x}_i)$ é posto completo.

Homocedasticidade condicional:

$$\mathsf{E}(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i' \mid \boldsymbol{x}_i) = \mathsf{E}(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i') = \boldsymbol{\Sigma} \tag{99}$$

$$\mathsf{E}(\boldsymbol{g}_i \boldsymbol{g}_i') \text{ é não singular, com } \boldsymbol{g}_i = \boldsymbol{\varepsilon}_i \otimes \boldsymbol{x}_i \tag{100}$$

Como notado na seção 4.5, $\mathsf{E}(\boldsymbol{g}_i \boldsymbol{g}_i') = \boldsymbol{\Sigma} \otimes \mathsf{E}(\boldsymbol{x}_i \boldsymbol{x}_i')$ sob homocedasticidade condicional, (5.1.6) é equivalente a condição

$$\boldsymbol{\Sigma} \text{ e } \mathsf{E}(\boldsymbol{x}_i \boldsymbol{x}_i') \text{ são não singulares.} \tag{101}$$

Componente de Erros: Se assume que o termo de erro possa ser decomposto como:

$$\varepsilon_{im} = \alpha_i + \eta_{im} \tag{102}$$

O termo α_i é chamado de *efeito individual, heterogeneidade individual, ou o efeito fixo*.

Defina o modelo em forma matricial:

$$\boldsymbol{y}_i = \boldsymbol{Z}_i \boldsymbol{\delta} + \mathbf{1}_M . \alpha_i + \boldsymbol{\eta}_i \tag{103}$$

tal que:

$$\begin{bmatrix} y_{i1} \\ \vdots \\ y_{iM} \end{bmatrix} = \begin{bmatrix} \boldsymbol{z}_{i1}\boldsymbol{\delta} \\ \vdots \\ \boldsymbol{z}_{iM}\boldsymbol{\delta} \end{bmatrix} + \begin{bmatrix} \alpha_i \\ \vdots \\ \alpha_i \end{bmatrix} + \begin{bmatrix} \boldsymbol{\eta}_{i1} \\ \vdots \\ \boldsymbol{\eta}_{iM} \end{bmatrix} \tag{104}$$

A condições de ortogonalidade (98) são satisfeitas se os regressores do sistema são ortogonais a ambos componentes de erro, isto é, se

$$\mathsf{E}(\boldsymbol{z}_{im}\alpha_i) = \mathbf{0} \tag{105}$$

e

$$\mathsf{E}(\boldsymbol{z}_{im}\eta_i) = \mathbf{0} \tag{106}$$

Todavia em muitas aplicações não é uma hipótese razoável. Isto ocorre porque o efeito fixo representa algumas características permanentes da unidade economica individual. Ver exemplo.

Modelo de Componentes de Erro

Função de produção com heterogeneidade Continuação exemplo da sec. 3.2.

Usando o exemplo da função de produção log-linear, suponha que a eficiência da firma, u_i, permaneça constante ao longo do tempo. Então a equação para o ano m é:

$$log(Q_{im}) = \phi_0 + \phi_1 log(L_{im}) + u_i + v_{im}$$

onde Q_{im} é a produção da firma i no tempo m, L_{im} é o fator trabalho e v_{im} é o choque tecnológico. Esta equação pode ser escirta na forma matricial (vetorial) fazendo:

$$y_{im} = log(Q_{im}); \boldsymbol{z}_{im} = (1, log(L_{im}))'; \boldsymbol{\delta} = (\phi_0, \phi_1)';$$

$$\alpha_i = u_i; \eta_{im} = v_{im}$$

Além disso,

$$\boldsymbol{x}_i = (1, log(L_{im}), ..., log(L_{iM}))'$$

sob concorrência perfeita, o efeito fixo u_i deveria ser positivamente correlacionado com o insumo trabalho porque firmas eficientes, dado que u_i é maior do que a média, deveriam contratar mais trabalho para expandir suas atividades. Se v_{im} representa choques não esperados pela firma *quando* ela escolhe insumos, é razoável supor que v_{im} não é relacionado com os regressores.

Modelo de Componentes de Erro

equação salarial Continuação exemplo 4.2.

Suponha o exemplo sem experiência ($EXPR$), mas com $m = 1969, 1980, 1982$. Suponha coeficientes para escolaridade S e IQ constantes ao longo do tempo, mas o intercepto pode variar. Com a decomposição do termo de erro o sistema é

$$LW69_i = \phi_1 + \beta S69_i + \gamma IQ_i + \alpha_i + \eta_{i1}$$
$$LW80_i = \phi_1 + \beta S80_i + \gamma IQ_i + \alpha_i + \eta_{i1}$$
$$LW82_i = \phi_1 + \beta S82_i + \gamma IQ_i + \alpha_i + \eta_{i1}$$

Escrevendo usando a definição de *coeficientes comuns*:

$$\boldsymbol{Z}_i = \begin{bmatrix} \boldsymbol{z}'_{i1} \\ \boldsymbol{z}'_{i2} \\ \boldsymbol{z}'_{i3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & S69_i & IQ_i \\ 0 & 1 & 0 & S80_i & IQ_i \\ 0 & 0 & 1 & S82_i & IQ_i \end{bmatrix}, \ \boldsymbol{\delta}' = (\phi_1, \phi 2, \phi_3, \beta, \gamma) \tag{107}$$

$$\boldsymbol{x}_i = (1, S69_i, S80_i, S82_i, IQ_i)'$$

O termo de erro inclui os determinantes que não estão na equação salarial. Pode ser razoável assumir que eles são dividos entre o que é permanente para o indivíduo (afetando escolha de educação, por exemplo) e o que não é relacionado com determinantes do salário (erro de medida do salário, por exemplo).

Média de Grupos

Método de efeitos fixos resolve o problema de (105). O estimador é aplicado a um sistema de M-equações transformada do sistema original (103). A matriz usada é a aniquiladora associada com $\mathbf{1}_M$:

$$\underset{(M\times M)}{\boldsymbol{Q}} = \boldsymbol{I}_M - \mathbf{1}_M(\mathbf{1}_M'\mathbf{1}_M)^{-1}\mathbf{1}_M' \tag{108}$$

$$= \boldsymbol{I}_M - \frac{1}{M}\mathbf{1}_M'\mathbf{1}_M$$

$$= \boldsymbol{I}_M - \begin{bmatrix} \frac{1}{M} & \cdots & \frac{1}{M} \\ \vdots & & \vdots \\ \frac{1}{M} & \cdots & \frac{1}{M} \end{bmatrix}$$

O que essa matriz faz é extrair desvios da média dos grupos. Por exemplo, multimplicando pela esquerda o vetor $\boldsymbol{y}_i$ de dimensão M por $\boldsymbol{Q}$, temos:

$$\tilde{\boldsymbol{y}}_i = \boldsymbol{Q}\boldsymbol{y}_i = \begin{bmatrix} y_{i1} - \bar{y}_i \\ \vdots \\ y_{im} - \bar{y}_M \end{bmatrix} = \boldsymbol{y}_i - \mathbf{1}_M \bar{y}_i \tag{109}$$

tal que a média do grupo da variável dependente é:

$$\bar{y}_i = \frac{1}{M}\mathbf{1}'_M \boldsymbol{y}_i = \frac{1}{M}\sum_{m=1}^{M} y_{im}$$

Essa transformação é similar para os regressores: $\boldsymbol{Q}\boldsymbol{Z}_i$ é o vetor de desvios de $\boldsymbol{Z}_i$.

Uma Reparemetrização

A robustez em relação a (105) possui um preço: alguns parâmetros do modelo podem não se identificados após a transformação por $\boldsymbol{Q}$. Dois exemplos sobre o tema:

1. O caso óbvio é o coeficiente comum para todas as equações. Por exemplo, IQ é um coeficiente comum em todas as equações no exemplo 5.2. Então a coluna de $\boldsymbol{Z}_i$ correspondente a IQ é $\mathbf{1}_M.IQ_i$ e a quinta coluna de $\boldsymbol{QZ}_i$ é de zeros. Consequentemente o coeficiente de IQ, γ, não pode ser identificado após as transformações. Para separar os coeficientes comuns dos demais faça $\boldsymbol{b}_i$ ser o vetor de regressores comuns e escreva a matriz $M \times L$ de regressores $\boldsymbol{Z_i}$ como:

$$\boldsymbol{Z}_i = (\boldsymbol{F} \vdots \mathbf{1}_M \boldsymbol{b}_i') \tag{110}$$

o vetor de coeficientes pode ser particionado como:

$$\boldsymbol{\delta} = \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{bmatrix} \tag{111}$$

O vetor de coeficientes $\boldsymbol{\gamma}$ não pode ser identificado após a transformação.

2. Também pode ser o caso, em adição a $\boldsymbol{\gamma}$, que algum coeficiente de $\boldsymbol{\beta}$ seja não identificado após a reparametrização.

A condição geral de identificação da estimação FE é que $\boldsymbol{F}_i$ e $\boldsymbol{b}_i$ são definidos tal que

$$\text{E}(\boldsymbol{QF}_i \otimes \boldsymbol{x}_i)\text{é (coluna) posto completo} \tag{112}$$

tal que $\boldsymbol{x}_i$ é a união de $(\boldsymbol{z}_{i1}, ..., \boldsymbol{z}_{iM})$. Por que esta é uma condição de identificação? Porque o estimador de efeitos fixos é uma especialização do RE aplicado a regressão transformada de $\boldsymbol{Qy}$ em $\boldsymbol{QF}$. (Esta condição é apenas uma adptação da anterior.)

Com $\boldsymbol{Z}_i$ divido entre $\boldsymbol{F}_i$ e $\boldsymbol{b}_i$, o sistema (103) pode ser rescrito como:

$$\boldsymbol{y}_i = \boldsymbol{F}_i\boldsymbol{\beta} + \mathbf{1}_M\boldsymbol{b}_i\gamma + \mathbf{1}_M\alpha_i + \boldsymbol{\eta}_i, \text{ ou} \tag{113}$$

$$y_{im} = \boldsymbol{f}'_{im}\boldsymbol{\beta} + \boldsymbol{b}_i\gamma + \alpha_i + \eta_i \tag{114}$$

tal que $\boldsymbol{f}'_{im}$ é a linha m de $\boldsymbol{F}_i$. Os estimadores RE e FE são bem definidos para o mesmo modelo.

Estimador de Efeitos Fixos

Estimador de efeitos fixos definido como o sistema transformado (dado que $\boldsymbol{QI}_M = \mathbf{0}$):

$$\boldsymbol{Q}\boldsymbol{y}_i = \boldsymbol{Q}\boldsymbol{F}_i\boldsymbol{\beta} + \boldsymbol{Q}\boldsymbol{\eta}_i \tag{115}$$

ou

$$\tilde{\boldsymbol{y}}_i = \tilde{\boldsymbol{F}}_i\boldsymbol{\beta} + \tilde{\boldsymbol{\eta}}_i \tag{116}$$

tal que:

$$\underset{(M\times 1)}{\tilde{\boldsymbol{y}}_i} = \boldsymbol{Q}\boldsymbol{y}_i; \ \underset{(M\times \#\beta)}{\tilde{\boldsymbol{F}}_i} = \boldsymbol{Q}\boldsymbol{F}_i; \ \underset{M\times 1}{\tilde{\boldsymbol{\eta}}_i} = \boldsymbol{Q}\boldsymbol{\eta}_i$$

Forme uma amostra "pooled" das transformadas $\tilde{\boldsymbol{y}}_i$ e $\tilde{\boldsymbol{F}}_i$ como

$$\underset{(nM\times 1)}{\tilde{\boldsymbol{y}}} = \begin{pmatrix} \tilde{\boldsymbol{y}}_1 \\ \vdots \\ \tilde{\boldsymbol{y}}_n \end{pmatrix}, \underset{(nM\times \#\beta)}{\tilde{\boldsymbol{F}}} = \begin{pmatrix} \tilde{\boldsymbol{F}}_1 \\ \vdots \\ \tilde{\boldsymbol{F}}_n \end{pmatrix}. \tag{117}$$

O estimador de efeitos fixos de $\boldsymbol{\beta}$, representado por $\widehat{\boldsymbol{\beta}}_{FE}$, é o estimador "Pooled OLS," i.e., o estimador OLS aplicado a amostra "pooled" $(\tilde{\boldsymbol{y}}, \tilde{\boldsymbol{F}})$ de tamanho nM. Assim,

$$\hat{\boldsymbol{\beta}}_{FE} = (\tilde{\boldsymbol{F}}_i'\tilde{\boldsymbol{F}}_i)^{-1}\tilde{\boldsymbol{F}}_i'\tilde{\boldsymbol{y}}_i$$

$$= (\frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{F}}_i'\tilde{\boldsymbol{F}}_i)^{-1}\frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{F}}_i'\tilde{\boldsymbol{y}}_i$$

$$= \left[(\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{Q}\boldsymbol{F}_i)'(\boldsymbol{Q}\boldsymbol{F}_i)\right]^{-1}\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{Q}\boldsymbol{F}_i)'(\boldsymbol{Q}\boldsymbol{y}_i)$$

$$= \left(\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{F}_i'\boldsymbol{Q}\boldsymbol{Q}\boldsymbol{F}_i\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{F}_i'\boldsymbol{Q}\boldsymbol{Q}\boldsymbol{y}_i$$

$$= \left(\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{F}_i'\boldsymbol{Q}\boldsymbol{Q}\boldsymbol{F}_i\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{F}_i'\boldsymbol{Q}\boldsymbol{Q}\boldsymbol{y}_i \text{ (uma vez que } \boldsymbol{Q} \text{ é simétrica e iden}$$

Substituindo (116) em (118) se pode obter o sampling error:

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{FE}-\boldsymbol{\beta} &= (\tilde{\boldsymbol{F}}_i'\tilde{\boldsymbol{F}}_i)^{-1}\tilde{\boldsymbol{F}}_i'\tilde{\boldsymbol{\eta}}_i \\ &= \left(\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{F}_i'\boldsymbol{Q}\boldsymbol{Q}\boldsymbol{F}_i\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{F}_i'\boldsymbol{Q}\boldsymbol{Q}\boldsymbol{\eta}_i \quad (119)\end{aligned}$$

Como o estimador de efeitos fixos é desvio da médio de grupos, ele também é conhecido como *within estimator* ou estimador covariância. Outra aplicação comum é o LSDV (Least Square Dummy Variable): estimar o modelo em nível por OLS adicionando dummies de grupo (i).

Proposição 5.1 (propriedades de amostras grandes para FE): Suponha as Hipóteses do modelo de componentes de erro, mas relaxe as hipóteses SUR requerendo apenas que

$$\mathrm{E}(\boldsymbol{f}_{im}'\eta_{ih})=\mathbf{0}$$

onde $\boldsymbol{f}'_{im}$ é a linha m de $\boldsymbol{F}_i$. Defina $\tilde{\boldsymbol{F}}_i$, $\tilde{\boldsymbol{y}}_i$ e $\tilde{\boldsymbol{\eta}}_i$ por (116). Então

- O estimador de efeitos fixos (118) é consistente e assintoticamente normal com

$$\mathsf{Avar}(\widehat{\boldsymbol{\beta}}_{FE}) = [\mathsf{E}(\tilde{\boldsymbol{F}}_i'\tilde{\boldsymbol{F}}_i)]^{-1}\mathsf{E}[\tilde{\boldsymbol{F}}_i'\mathsf{E}(\tilde{\boldsymbol{\eta}}_i\tilde{\boldsymbol{\eta}}_i')\tilde{\boldsymbol{F}}_i][\mathsf{E}(\tilde{\boldsymbol{F}}_i'\tilde{\boldsymbol{F}}_i)]^{-1}$$

- A variância assintótica é consistentemente estimada dada por

$$\widehat{\mathsf{Avar}(\widehat{\boldsymbol{\beta}}_{FE})} = [\frac{1}{n}\sum \tilde{\boldsymbol{F}}_i'\tilde{\boldsymbol{F}}_i]^{-1}[\frac{1}{n}\sum \tilde{\boldsymbol{F}}_i'\tilde{\boldsymbol{V}}\tilde{\boldsymbol{F}}_i][\frac{1}{n}\sum \tilde{\boldsymbol{F}}_i'\tilde{\boldsymbol{F}}_i]^{-1}$$

  aqui $\tilde{\boldsymbol{V}}$ é a matriz dos momentos cruzados dos resíduos transformados associados com o estimador FE:

$$\underset{(M\times M)}{\tilde{\boldsymbol{V}}} = \frac{1}{n}\sum(\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{F}}_i\widehat{\boldsymbol{\beta}}_{FE})(\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{F}}_i\widehat{\boldsymbol{\beta}}_{FE})' \qquad (120)$$

Função de Custo Translog (sec. 4.7)

Função de custo translog é uma generalização da função de custo Cobb-Douglas. Uma boa característica da translog é que as equações de demanda otimizada por insumo se forem transformadas em participação no custo total são lineares no (log) do produto e preço dos fatores. Os coeficientes inerentes a este sistema são um subconjunto de parâmetros da função custo que descrevem a tecnologia.

Será assumida a hipótese de homocedasticidade condicional. Assumir também que os regressores no sistema são predeterminados. Como foi argumentado no cap. 1, esta é uma hipótese que pode considerada razoável para o setor de geração de energial elétrica antes da desregulamentação.

Função de Custo Translog (sec. 4.7)

A função custo translog: assumindo termos quadráticos e cruzados no log de todos os argumentos, a função custo log-linear com três insumos pode ser escrita como:

$$\log(C) = \alpha_0 + \sum_{j=1}^{3} \alpha_j \log(p_j) + \frac{1}{2}\sum_{j=1}^{3}\sum_{k=1}^{3} \gamma_{jk} \log(p_j)\log(p_k)$$
$$+ \alpha_Q \log(Q) + \frac{1}{2}\gamma_{QQ}(\log(Q))^2 + \sum_{j=1}^{3} \gamma_{jQ} \log(p_j)\log(Q) \quad (121)$$

Nessa expressão o termo $\frac{1}{2}\sum_{j=1}^{3}\sum_{k=1}^{3}\gamma_{jk}\log(p_j)\log(p_k)$ é uma forma representando o efeito de segunda ordem dos preços dos fatores. Se pode assumir que a matriz 3×3 de forma quadradática coeficientes é simétrica:

$$\gamma_{jk} = \gamma_{kj},\ (j, k = 1, 2, 3) \quad (122)$$

O grau de retorno de escala pode ser calculado como:

$$\frac{1}{\partial \log(C)/\partial \log(Q)} = \frac{1}{\alpha_Q + \gamma_{QQ} \log(Q) + \sum_{j=1}^{3} \gamma_{jQ} \log(p_j)} \quad (123)$$

Participação dos Fatores

A conexão entre os parâmetros da função custo e a demanda dos fatores é dado pelo Lemma de Shepard. Faça x_j ser a demanda minimizadora de custos para o insumo j dado o preço dos fatores (p_1, p_2, p_3) e produto Q. Então $\sum_{j=1}^{3} p_j x_j = C$. O lema afirma que

$$\frac{\partial C}{\partial p_j} = x_j \tag{124}$$

Notando que:

$$\frac{\partial \log(C)}{\partial \log(p_j)} = \frac{p_j}{C}\frac{\partial C}{\partial p_j} \tag{125}$$

o lema também afirma que o log das derivas parciais da função custo é igual a participação dos fatores:

$$\frac{\partial \log(C)}{\partial \log(p_j)} = \frac{p_j x_j}{C} \tag{126}$$

Para o caso da função de custo translog a derivada parcial é:

$$\frac{\partial \log(C)}{\partial \log(p_j)} = \alpha_j + \sum_{j=1}^{3} \gamma_{jk} \log(p_k) + \gamma_{jQ} \log(Q) \tag{127}$$

Combinando (127) e (126) e definindo o "cost share" como $s_j = p_j x_j / C$, temos o seguinte sistema de "cost share":

$$s_j = \alpha_j + \sum_{k=1}^{3} \gamma_{jk} \log(p_k) + \gamma_{jQ} \log(Q) \tag{128}$$

Restrições de semietria: o sistema s_j é sujeito as retrições cruzadas entre equações: o coeficiente de $\log(p_k)$ em s_j é igual ao coeficiente de $\log(p_j)$ em s_k

Elasticidades de Substituição

A elasticidade de substituição entre insumos j e k, representado por η_{jk} é relacionado com a função custo C pela fórmula:

$$\eta_{jk} = \frac{C.\frac{\partial^2 C}{\partial p_j \partial p_k}}{\frac{\partial C}{\partial p_j}.\frac{\partial C}{\partial p_k}} \tag{129}$$

Para a função de custo translog temos:

$$\eta_{jk} = \begin{cases} \frac{\gamma_{jk}+s_j s_k}{s_j s_k} & \text{para } j \neq k \\ \frac{\gamma_{jj}+s_j^2-s_j}{s_j^2} & \text{para } j = k \end{cases} \tag{130}$$

Propriedades da Função Custo

Homogeneidade:

$$\begin{pmatrix} \alpha_1 + \alpha_2 + \alpha_3 = 1 \\ \gamma_{11} + \gamma_{12} + \gamma_{13} = 0 \\ \gamma_{21} + \gamma_{22} + \gamma_{23} = 0 \\ \gamma_{31} + \gamma_{32} + \gamma_{33} = 0 \\ \gamma_{1Q} + \gamma_{2Q} + \gamma_{3Q} = 0 \end{pmatrix}$$

Monotonicidade: Participações (shares) não podem ser negativas. O lado direito das equações de shares não podem ser negativos para quaisquer combinações de preços de fator e produto.

Concavidade: esta condição requer que a matriz de elasticidade de substituição seja dada por (130) seja negativa semidefinda para qualquer combinação de s_j. É possível mostrar que uma condição necessária e suficiente é que a matriz γ_{jk} seja negativa semidefinida. Os elementos da diagonal principal não podem ser positivos.

Especificação estocástica: adicionar termos de erro nas equações de share. Entretanto, justicativas diversas não suficientes para adicionar erros estocásticos nas equações de participação. Estimativa apenas das equações de shares.

A natureza das restrições

Com a adições dos erros, o sistema é escrito como:

$$
\begin{aligned}
s_1 &= \alpha_1 + \gamma_{11}\log(p_1) + \gamma_{12}\log(p_2) + \gamma_{13}\log(p_3) + \gamma_{1Q}\log(Q) + \varepsilon_1 \\
s_2 &= \alpha_2 + \gamma_{21}\log(p_1) + \gamma_{22}\log(p_2) + \gamma_{23}\log(p_3) + \gamma_{2Q}\log(Q) + \varepsilon_2 \\
s_1 &= \alpha_3 + \gamma_{31}\log(p_1) + \gamma_{32}\log(p_2) + \gamma_{33}\log(p_3) + \gamma_{1Q}\log(Q) + \varepsilon_3
\end{aligned}
\tag{131}
$$

O sistema possui 15 coeficientes. As restrições cruzadas e as de homogeneidade formam um conjunto de 8 restrições. Isso significa que os 15 parâmetros podem ser descritos por 7 parâmetros.

$$\begin{pmatrix} \alpha_1 + \alpha_2 + \alpha_3 = 1 \\ \gamma_{11} + \gamma_{12} + \gamma_{13} = 0 \\ \gamma_{21} + \gamma_{22} + \gamma_{23} = 0 \\ \gamma_{31} + \gamma_{32} + \gamma_{33} = 0 \\ \gamma_{1Q} + \gamma_{2Q} + \gamma_{3Q} = 0 \end{pmatrix}$$

Homogeneidade:

$$\begin{pmatrix} \gamma_{11} + \gamma_{12} + \gamma_{13} = 0 \\ \gamma_{21} + \gamma_{22} + \gamma_{23} = 0 \\ \gamma_{31} + \gamma_{32} + \gamma_{33} = 0 \end{pmatrix}$$

Simetria:

$$\begin{pmatrix} \gamma_{12} = \gamma_{21} \\ \gamma_{13} = \gamma_{31} \\ \gamma_{23} = \gamma_{32} \end{pmatrix}$$

Impor a restrição de homogeneidade é direta, eliminando os parâmetros $\gamma_{.3}$:

$$\begin{aligned} s_1 = \alpha_1 + \gamma_{11} \log(p_1/p_3) + \gamma_{12} \log(p_2/p_3) + \gamma_{1Q} \log(Q) + \varepsilon_1 \\ s_2 = \alpha_2 + \gamma_{21} \log(p_1/p_3) + \gamma_{22} \log(p_2/p_3) + \gamma_{2Q} \log(Q) + \varepsilon_2 \\ s_3 = \alpha_3 + \gamma_{31} \log(p_1/p_3) + \gamma_{32} \log(p_2/p_3) + \gamma_{1Q} \log(Q) + \varepsilon_3 \end{aligned} \tag{132}$$

Os shares soman 1 para todas as unidades. A soma dos termos de erro é sempre zero, logo a matriz de covariância é $\Sigma = \mathrm{Var}(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ é singular. Solução: se a terceira equação for eliminada para incorporar restrições de soma temos:

$$\begin{aligned} s_1 &= \alpha_1 + \gamma_{11} \log(p_1/p_3) + \gamma_{12} \log(p_2/p_3) + \gamma_{1Q} \log(Q) + \varepsilon_1 \\ s_2 &= \alpha_2 + \gamma_{21} \log(p_1/p_3) + \gamma_{22} \log(p_2/p_3) + \gamma_{2Q} \log(Q) + \varepsilon_2 \end{aligned} \tag{133}$$

Como os regressores são predeterminados o sistema com regressores comuns pode ser estimado por regressão multivariada s.a. restrição de simetria entre equações.

Esta estimação com restrição pode ser transformada em um estimação irrestrita. I.e., o sistema pode ser escrito no formato de coeficientes comuns.

$$\boldsymbol{y}_i = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$

$$
\boldsymbol{Z}_i = \begin{bmatrix} 1 & 0 & \log(p_1/p_3) & \log(p_2/p_3) & 0 & \log(Q) & 0 \\ 0 & 1 & 0 & \log(p_1/p_3) & \log(p_2/p_3) & 0 & \log(Q) \end{bmatrix}
$$

$$
\boldsymbol{\delta}' = [\alpha_1, \alpha 2, \gamma_{11}, \gamma_{12}, \gamma_{22}, \gamma_{1Q}, \gamma_{2Q}] \tag{134}
$$

Os demais parâmetros podem ser calculados como resíduo. Por exemplo,

$$
\widehat{\gamma}_{33} = \widehat{\gamma}_{11} + 2\widehat{\gamma}_{12} + \widehat{\gamma}_{22}
$$

Translog

$\widehat{\Sigma}^*$ calculado com OLS equação por equação. Estimativa do modelo por RE.

Table 4.3: Simple Statistics (Sample Size = 99)

	Output in kilowatt hours	Labor share	Capital share	Fuel share
Mean	9.0	0.141	0.227	0.631
Std. deviation	10.3	0.059	0.062	0.095

Table 4.4: Random-Effects Estimates

Parameter	Point estimate	Standard error	t-value
α_1	−0.132	0.106	−1.25
α_2	0.318	0.085	3.75
α_3	0.813	0.094	8.69
γ_{11}	0.084	0.020	4.19
γ_{12}	−0.023	0.016	−1.46
γ_{13}	−0.060	0.015	−3.92
γ_{22}	0.122	0.020	6.19
γ_{23}	−0.099	0.017	−5.75
γ_{33}	0.159	0.023	6.90
γ_{1Q}	−0.0211	0.0025	−8.55
γ_{2Q}	−0.0086	0.0030	−2.87
γ_{3Q}	0.0297	0.0037	7.98

$$\widehat{\boldsymbol{\Sigma}} \text{ by pooled OLS} = \begin{bmatrix} 0.00173 & -0.000171 & -0.00156 \\ -0.000171 & 0.00253 & -0.00236 \\ -0.00156 & -0.00236 & 0.00391 \end{bmatrix}$$

Table 4.5: Substitution Elasticities

Labor-Capital	Capital-Fuel	Labor-Fuel
0.17	0.27	0.29

Referências

Hansen, Bruce *Econometrics*. 2019.

Hansen, Lars Peter "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50 (4), 1982.

Hayashi, Fumio *Econometrics*. Princeton University Press, 2000.

Stock, James H. e Mark W. Watson *Introduction to Econometrics*. 3a ed. Pearson, 2015.