GMM - Parte I

Bibliografia: Hayashi, cap. 3

Claudio Lucinda

FEA/USP

Overview

Método Generalizado dos Momentos

- A base do GMM é a seguinte. Com amostragem aleatória, e sob hipóteses razoáveis, uma estatística amostral irá convergir em probabilidade para alguma constante.
- Por exemplo, com amostragem aleatória i.i.d., $\bar{m}_2 = (1/n)\sum_{i=1}^{n} y_i^2$ irá convergir em mean square à variância mais o quadrado da média da variável aleatória y_i.
- Esta constante será uma função dos parâmetros – desconhecidos – da distribuição.
- Se tivermos K estatísticas amostrais, $\bar{m}_1, \cdots, \bar{m}_K$, podemos estimar K parâmetros dos quais as estatísticas amostrais dependem.

GMM II

- Estes momentos serão consistentes (ou seja, convergem para as constantes relevantes), por conta de alguma versão da Lei dos Grandes Números
- Eles serão normalmente distribuídos pelo Teorema Central do Limite de Lindberg-Lévy.
- Podemos extrair as estimativas dos parâmetros vão ser consistentes pelo teorema de Slutsky, e as estimativas deles serão normalmente distribuídas pelo Método Delta.

Método Generalizado dos Momentos (Continuação)

- Etapas do GMM:
  1. Especificar o modelo econômico e definir os parâmetros a serem estimados.
  2. Identificar os momentos empíricos relevantes relacionados aos parâmetros.
  3. Escolher um estimador de momentos adequado.
  4. Resolver as equações de momento para estimar os parâmetros.
  5. Verificar a validade das estimativas e realizar testes de qualidade do ajuste.

GMM- Revisitando IV

O modelo de Variáveis Instrumentais é um tipo de estimador de GMM. Em muitos casos, temos que a premissa de

$$E\left(\mathbf{Z}^{\mathbf{T}}\mathbf{u}\right) = 0$$

Vamos supor que, dada uma amostra grande o suficiente, mesmo assim, a covariância entre os regressores e o termo erro do modelo não converge para zero plim $_{N\rightarrow\infty}\mathbf{Z}^{\mathbf{T}}\mathbf{u} = \gamma$. Isso faz com que o estimador de XQO seja viesado e inconsistente.

Revisitando IV

No entanto, vamos supor que exista um ou um conjunto de variáveis, denotadas por **X** que, ao mesmo tempo:

$$E\left(\mathbf{X}^{\mathrm{T}}\mathbf{Z}\right) \neq 0$$
$$E\left(\mathbf{X}^{\mathrm{T}}\mathbf{u}\right) = 0$$

Esta variável **X** é chamada de instrumento para **Z** (**essa notação é do Hayashi**). Podemos derivar o estimador de variáveis instrumentais de uma forma bem simples, desde que o número de variáveis seja igual ao número de Instrumentos:

$$E\left(\mathbf{X}^{\mathbf{T}}\mathbf{u}\right) = 0$$
$$E\left(\mathbf{X}^{\mathbf{T}}(\mathbf{Y} - \mathbf{Z}\beta)\right) = 0$$

IV Revisitado

Supondo os regressores fixos em amostras repetidas, temos que:

$$\mathbf{X}^{\mathbf{T}}(\mathbf{Y} - \mathbf{Z}\beta) = 0$$
$$\mathbf{X}^{\mathrm{T}}\mathbf{Y} - \mathbf{X}^{\mathrm{T}}\mathbf{Z}\beta = 0$$
$$\hat{\beta}_{iv} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{Z}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}$$

Para provarmos a ausência de viés:

$$\hat{\beta}_{iv} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{Z}\right)^{-1}\mathbf{X}^{\mathbf{T}}(\mathbf{Z}\beta + \mathbf{u})$$
$$= \beta + \left(\mathbf{X}^{\mathrm{T}}\mathbf{Z}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{u}$$

IV Revisitado

Ou seja, dada a hipótese de $E\left(\mathbf{X}^{\mathrm{T}}\mathbf{u}\right) = 0$, temos que o estimador é não viesado. Do ponto de vista dos erros-padrão:

$$\mathrm{Var}\left(\hat{\beta}_{iv}\right) = E\left[\left(\hat{\beta}_{iv} - \beta\right)\left(\hat{\beta}_{iv} - \beta\right)^{T}\right] = \\ = E\left[\left(\mathbf{X}^{\mathbf{T}}\mathbf{Z}\right)^{\mathbf{-1}}\mathbf{X}^{\mathbf{T}}\mathbf{u}\mathbf{u}^{\mathrm{T}}\mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{Z}\right)^{\mathbf{-1}}\right]$$

Ou, alternativamente,

$$\mathrm{Var}\left(\hat{\beta}_{iv}\right) = \sigma^2\left(\mathbf{X}^{\mathbf{T}}\mathbf{Z}\right)^{\mathbf{-1}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{Z}\right)^{\mathbf{-1}}$$

IV Revisitado

O problema é que temos que ter a dimensão de $\mathbf{X}$ igual à dimensão de $\mathbf{Z}$. E se tivermos mais instrumentos do que variáveis endógenas? a idéia é que o sistema subjacente é sobre-identificado. O que fazer neste caso? Uma sugestão seria apenas usar os instrumentos na quantidade certinha, mas aí estaríamos desperdiçando informações. Outra forma interessante seria usar, no lugar dos $\mathbf{X}$ diretamente, a projeção dos $\mathbf{X}$ no espaço determinado pelos $\mathbf{Z}$:

$$\hat{\mathbf{Z}} = \mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Z} = \mathbf{P}_{\mathbf{X}}\mathbf{Z}$$

Neste caso, as fórmulas ficam:

$$\hat{\beta}_{iv} = \left(\mathbf{Z}^{\mathbf{T}}\mathbf{P}_{\mathbf{z}}\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathbf{T}}\mathbf{P}_{\mathbf{X}}\mathbf{Y}$$

A Variância, por outro lado, fica sendo:

$$\mathrm{Var}\left(\hat{\beta}_{iv}\right) = \sigma^2\left(\mathbf{Z}^{\mathbf{T}}\mathbf{P}_{\mathbf{X}}\mathbf{Z}\right)^{-1}$$

Condição de Posto

Suposição 3.4 (condição de posto para identificação): A $K \times L$ matriz $\mathrm{E}(\mathbf{x}_i \mathrm{z}_i{}')$ é de posto completo na dimensão coluna (ou seja, seu posto é igual a L, o número de suas colunas). Denotamos esta matriz por $\mathbf{\Sigma_{xz}}$.
Intuição, para um modelo linear nos parâmetros. Podemos escrevê-lo como $\mathbf{x}_i \cdot y_i - \mathbf{x}_i \mathbf{z}_i' \tilde{\delta}$. Ou um sistema de K equações lineares:

$$\mathrm{E}(\mathbf{x}_i \cdot y_i) - \mathrm{E}\left(\mathbf{x}_i \mathbf{z}_i'\right) \tilde{\delta} = \mathbf{0} \quad \text{or} \quad \underset{(K\times L)}{\mathbf{\Sigma_{xz}}} \underset{(L\times 1)}{\tilde{\delta}} {}^{(K\times 1)} = \underset{(Kx}{\boldsymbol{\sigma}_{\mathbf{xy}}},$$

Em que

$$\sigma_{\mathbf{xy}} \equiv \mathrm{E}(\mathbf{x}_i \cdot y_i), \quad \mathbf{\Sigma_{xz}} \equiv \mathrm{E}\left(\mathbf{x}_i \mathbf{z}_i'\right)$$

Condição de Ordem

Como rank $(\mathbf{\Sigma_{xz}}) < L$ se $K < L$, uma condição necessária para identificação é que

$$K \ (= \text{nro. variáveis predeterminadas}) \ \geq L \ (= \text{nro. regressores}).$$

Isso é chamado de condição de ordem para identificação. Pode ser afirmado de diferentes maneiras. Visto que K é também o número de condições de ortogonalidade e L é o número de parâmetros, a condição de ordem pode ser exposta de forma equivalente como

$$\text{nro. condições de ortogonalidade} \ \geq \ \text{nro. parâmetros}.$$

Ao subtrair o número de regressores predeterminados de ambos os lados da desigualdade, obtemos outra declaração equivalente:

$$\#\text{variáveis predeterminadas excluídas da equação} \geq \ \#\text{regressores endógenos}.$$

Condições de Ordem II

Dependendo se a condição de ordem é satisfeita, dizemos que a equação é

- Sobreidentificada se a condição de posto for satisfeita e $K > L$,
- Exatamente identificada ou apenas identificada se a condição de posto for satisfeita e $K = L$
- Subidentificada (ou não identificada) se a condição do posto não for satisfeita (ou seja, se $K < L$).

Como a condição de ordem é uma condição necessária para identificação, uma falha na condição de ordem significa que a equação não é identificada.

Condição Necessária para Normalidade Assintótica

g_i é uma sequência de diferenças em martingale (MDS) com segundos momentos finitos:
Seja $\mathbf{g}_i \equiv \mathbf{x}_i \cdot \varepsilon_i$. Temos que $\mathbf{g}_i$ é uma MDS (e $\mathrm{E}\left(\mathbf{g}_i\right) = \mathbf{0}$).
A matriz $K \times K$ de momentos cruzados, $\mathrm{E}\left(\mathbf{g}_i \mathbf{g}_i'\right)$, é não singular. Definindo $S = Avar(\bar{g})$ (a variância da distribuição limite de $\sqrt{n}\overline{\mathbf{g}}$, where $\overline{\mathbf{g}} \equiv \frac{1}{n} \sum_{i=1}^{n} \mathbf{g}_i$). Pela Hipótese 3.2 do Hayashi e o TCL de MDS ergódico estacionário, $\mathbf{S} = \mathrm{E}\left(\mathbf{g}_i \mathbf{g}_i'\right)$.

Intuitivamente...

Uma forma mais intuitiva de entender essa coisa do MDS é:

$$\mathrm{E}\left(\varepsilon_i \mid \varepsilon_{i-1}, \ldots, \varepsilon_1, \mathbf{x}_i, \mathbf{x}_{i-1}, \ldots, \mathbf{x}_1\right) = 0$$

Que em palavras quer dizer que, além de ser um MDS, o termo erro tem que ser ortogonal não apenas ao instrumentos contemporâneos mas aos defasados também. Depois vamos ver isso com mais cuidado.

GMM - Definição

Seja $\widehat{\mathbf{W}}$ uma matriz definida positiva simétrica $K \times K$, possivelmente dependente da amostra, tal que $\widehat{W} \to_{mathrmp} \mathbf{W}$ com o tamanho da amostra n tendendo a infinito, e $\mathbf{W}$ simétrico e positivo definido. O estimador GMM de $\boldsymbol{\delta}$, denotado $\hat{\boldsymbol{\delta}}(\widehat{\mathbf{W}})$, é

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

where

$$J(\tilde{\boldsymbol{\delta}}, \widehat{\mathbf{W}}) \equiv n \cdot \mathbf{g}_{\boldsymbol{n}}(\tilde{\boldsymbol{\delta}})' \widehat{\mathbf{W}} \mathbf{g}_n(\tilde{\boldsymbol{\delta}})$$

GMM – Definição II

Quando $\mathbf{g}_n(\tilde{\delta}\ )$ é linear nos coeficientes $\tilde{\delta}$ a função objetivo é quadrática no $\tilde{\delta}$ quando a equação é linear:

$$J(\tilde{\delta}, \widehat{\mathbf{W}}) = n \cdot \left(\mathbf{s_{xy}} - \mathbf{S_{xz}}\tilde{\delta}\right)' \widehat{\mathbf{W}} \left(\mathbf{S_{xy}} - \mathbf{S_{xz}}\tilde{\delta}\right)$$

Podemos resolver e chegar a uma solução analítica nesse caso:

$$\text{Estimador GMM: } \hat{\delta}(\widehat{\mathbf{W}}) = \left(\mathbf{S}'_{\mathbf{xz}}\widehat{\mathbf{W}}\mathbf{S_{xz}}\right)^{-1} \mathbf{S}'_{\mathbf{xz}}\widehat{\mathbf{W}}\mathbf{s_{xy}}$$

Propriedades Assintóticas do GMM

Resultados assintóticos para $\hat{\delta}(\widehat{\mathbf{W}})$ válidos para qualquer escolha de $\widehat{\mathbf{W}}$ são
(a) (Consistência) $\text{plim}_{n\to\infty}\, \hat{\delta}(\widehat{\mathbf{W}}) = \boldsymbol{\delta}$.
(b) (Normalidade Assintótica) Dada a condição de alguns slides pra trás

$$\sqrt{n}(\hat{\boldsymbol{\delta}}(\widehat{\mathbf{W}}) - \delta) \underset{\mathrm{d}}{\to} N(\mathbf{0}, \text{Avar}(\hat{\boldsymbol{\delta}}(\widehat{\mathbf{W}}))) \quad \text{as } n \to \infty$$

em que

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

Lembrando: $\left(\boldsymbol{\Sigma}_{\mathbf{xz}} \equiv \mathrm{E}\left(\mathbf{x}_i\mathbf{z}_i'\right), \mathbf{S} = \mathrm{E}\left(\mathbf{g}_i\mathbf{g}_i'\right) = \mathrm{E}\left(\varepsilon_i^2\mathbf{x}_i\mathbf{x}_i'\right), \mathbf{W} \equiv \text{plim}\,\widehat{\mathbf{W}}.\right)$

Propriedades Assintóticas II

(c) (Estimador Consistente de Avar $(\hat{\delta}(\widehat{\mathbf{W}}))$) Suponha um estimador consistente de, $\widehat{\mathbf{S}}$, de $\mathbf{S}(K \times K)$. Podemos mostrar que Avar$(\hat{\delta}(\widehat{\mathbf{W}}))$ é estimado de forma consistente por

$$\text{Avar}(\hat{\delta}(\widehat{\mathbf{W}})) \equiv \left(\mathbf{S}'_{\mathbf{xz}}\widehat{\mathbf{W}}\mathbf{S}_{\mathbf{xz}}\right)^{-1} \mathbf{S}'_{\mathbf{xz}}\widehat{\mathbf{W}}\widehat{\mathbf{S}}\widehat{\mathbf{W}}\mathbf{S}_{\mathbf{xz}} \left(\mathbf{S}'_{\mathbf{xz}}\widehat{\mathbf{W}}\mathbf{S}_{\mathbf{xz}}\right)^{-1}$$

Em que $\mathbf{S}_{\mathbf{xz}}$ é a média amostral de $\mathbf{x}_i\mathbf{z}_i'$:

$$\underset{(K\times L)}{\mathbf{S}_{\mathbf{xz}}} \equiv \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i\mathbf{z}_i'$$

Teste t

Supondo a existência de um estimador consisrente $\widehat{\mathbf{S}}$ de $\mathbf{S}(\equiv \text{Avar}(\overline{\mathbf{g}}) = \mathrm{E}\left(\mathbf{g}_i\mathbf{g}_i'\right))$. Seja

$$\text{Avar}(\hat{\delta}(\widehat{\mathbf{W}})) \equiv \left(\mathbf{S}_{\mathbf{xz}}'\widehat{\mathbf{W}}\mathbf{S}_{\mathbf{xz}}\right)^{-1}\mathbf{S}_{\mathbf{xz}}'\widehat{\mathbf{W}}\widehat{\mathbf{S}}\widehat{\mathbf{W}}\mathbf{S}_{\mathbf{xz}}\left(\mathbf{S}_{\mathbf{xz}}'\widehat{\mathbf{W}}\mathbf{S}_{\mathbf{xz}}\right)^{-1}$$

Então:

(a) Sob $\mathrm{H}_0 : \delta_\ell = \bar{\delta}_\ell$,

$$t_\ell \equiv \frac{\sqrt{n}\left(\hat{\delta}_\ell(\widehat{\mathbf{W}}) - \bar{\delta}_\ell\right)}{\sqrt{(\text{Avar}(\hat{\delta}(\widehat{\mathbf{W}})))_{\ell\ell}}} = \frac{\hat{\delta}_\ell(\widehat{\mathbf{W}}) - \bar{\delta}_\ell}{SE_\ell^*} \to\underset{\mathrm{d}}{\to} N(0,1),$$

$$SE_\ell^*(\text{ robust standard error }) \equiv \sqrt{\frac{1}{n} \cdot (\text{Avar}(\hat{\delta}(\widehat{\mathbf{W}})))_{\ell\ell}}$$

Teste Wald

(b) (H0 linear) $\mathrm{H}_0 : \mathbf{R}\delta = \mathbf{r}$ em que r é a dimensão de $\mathbf{r}$ e $\mathbf{R}(\#\mathbf{r} \times L)$ é posto cheio,

$$W \equiv n \cdot (\mathbf{R}\hat{\delta}(\widehat{\mathbf{W}}) - \mathbf{r})' \left\{ \mathbf{R}[\widehat{\mathrm{Avar}(\hat{\delta}(\widehat{\mathbf{W}}))}]\mathbf{R}' \right\}^{-1} (\mathbf{R}\hat{\delta}(\widehat{\mathbf{W}}) - \mathbf{r}) \underset{\mathrm{d}}{\rightarrow} \chi^2(\#\mathbf{r})$$

(c) (H0 não linear) $\mathrm{H}_0 : \mathbf{a}(\boldsymbol{\delta}) = \mathbf{0}$ tal que $\mathbf{A}(\boldsymbol{\delta})$, a $\#\mathrm{a} \times L$ matriz de primeiras derivadas de $\mathbf{a}(\boldsymbol{\delta})$ (em que $\#\mathrm{a}$ é a dimensão de $\mathbf{a}$), é contínua de posto linha cheio,

$$W \equiv n \cdot \mathbf{a}(\hat{\delta}(\widehat{\mathbf{W}}))' \left\{ \mathbf{A}(\hat{\delta}(\widehat{\mathbf{W}}))[\mathrm{Avar}(\hat{\delta}(\widehat{\mathbf{W}}))]\mathbf{A}(\hat{\delta}(\widehat{\mathbf{W}}))' \right\}^{-1} \mathbf{a}(\hat{\delta}(\widehat{\mathbf{W}})) \underset{\mathrm{d}}{\rightarrow} \chi^2(\#\mathbf{a})$$

LR Test - Versão GMM

A função objetivo é $J\left(\tilde{\delta}, \widehat{\mathbf{S}}^{-1}\right)$ para uma estimativa consistente $\widehat{\mathbf{S}}$ de $\mathbf{S}$. O estimador restrito é definido $\bar{\delta}\left(\widehat{\mathbf{S}}^{-1}\right) \equiv \underset{\tilde{\delta}}{\operatorname{argmin}} J\left(\tilde{\delta}, \widehat{\mathbf{S}}^{-1}\right)$ sujeito a H_0.

O princípio da LR sugere que

$$LR \equiv J\left(\bar{\delta}\left(\widehat{\mathbf{S}}^{-1}\right), \widehat{\mathbf{S}}^{-1}\right) - J\left(\hat{\delta}\left(\widehat{\mathbf{S}}^{-1}\right), \widehat{\mathbf{S}}^{-1}\right)$$

Deve ser distribuído qui-quadrado com g.l. igual ao número de restrições. E é.