GMM - Parte II

Bibliografia: Hayashi, cap. 3

Claudio Lucinda

FEA/USP

Overview

GMM em 2 estágios

Naturalmente, desejamos escolher entre os estimadores GMM indexados por $\widehat{\mathbf{W}}$ aquele que tiver a menor variância assintótica. A próxima proposição fornece uma escolha de $\mathbf{W}$ que minimiza a variância assintótica.

Proposition

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

Como a variância assintótica para qualquer $\widehat{\mathbf{W}}$ é (3.5.1), essa proposição é provada se pudermos mostrar que

$$(\mathbf{\Sigma}'_{\mathbf{xz}}\mathbf{W}\mathbf{\Sigma}_{\mathbf{xz}})^{-1}\mathbf{\Sigma}'_{\mathbf{xz}}\mathbf{WSW}\mathbf{\Sigma}_{\mathbf{xz}}(\mathbf{\Sigma}'_{\mathbf{xz}}\mathbf{W}\mathbf{\Sigma}_{\mathbf{xz}})^{-1} \geq (\mathbf{\Sigma}'_{\mathbf{xz}}\mathbf{S}^{-1}\mathbf{\Sigma}_{\mathbf{xz}})^{-1}$$

Estimador GMM Eficiente

$$\hat{\delta}\left(\widehat{\mathbf{S}}^{-1}\right) = \left(\mathbf{S}_{\mathbf{xz}}'\widehat{\mathbf{S}}^{-1}\mathbf{S}_{\mathbf{xz}}\right)^{-1}\mathbf{S}_{\mathbf{xz}}'\widehat{\mathbf{S}}^{-1}\mathbf{S}_{\mathbf{xy}},$$

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

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

Teste t e Wald GMM Eficiente

Com $\widehat{\mathbf{W}} = \widehat{\mathbf{S}}^{-1}$, as fórmulas pra estatística t e teste Wald ficam

$$t_\ell = \frac{\hat{\delta}_\ell\left(\widehat{\mathbf{S}}^{-1}\right) - \bar{\delta}_\ell}{SE^*_\ell}$$

em que SE^*_ℓ é o erro-padrão robusto dado por

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

e

$$W = n \cdot \mathbf{a}\left(\hat{\delta}\left(\widehat{\mathbf{S}}^{-1}\right)\right)' \left\{\mathbf{A}\left(\hat{\delta}\left(\widehat{\mathbf{S}}^{-1}\right)\right)\left(\mathbf{S}'_{\mathbf{xz}}\widehat{\mathbf{S}}^{-1}\mathbf{S}_{\mathbf{xz}}\right)^{-1}\mathbf{A}\left(\hat{\delta}\left(\widehat{\mathbf{S}}^{-1}\right)\right)'\right\}^{-1} \mathbf{a}\left(\hat{\delta}\left(\widehat{\mathbf{S}}^{-1}\right)\right)$$

GMM em duas etapas

Para calcular o estimador GMM eficiente, precisamos do estimador consistente $\widehat{\mathbf{S}}$. Mas a Proposição 3.4 do livro nos garante que o $\widehat{\mathbf{S}}$ baseado em qualquer estimador consistente de δ é consistente para S. Isso nos leva ao seguinte procedimento GMM eficiente em duas etapas:

Passo 1: Escolha uma matriz $\widehat{W}$ que converja em probabilidade para uma matriz definida simétrica e positiva, e minimize $J(\tilde{\delta}, \widehat{\mathbf{W}})$ sobre $\tilde{\delta}$ para obter $\hat{\delta}(\widehat{\mathbf{W}})$. Não faltam tais matrizes $\widehat{\mathbf{W}}$ (por exemplo, $\widehat{\mathbf{W}} = \mathbf{I}$), mas geralmente definimos $\widehat{\mathbf{W}} = \mathbf{S}_{\mathbf{xx}}^{-1}$. O estimador resultante $\hat{\boldsymbol{\delta}}\left(\mathbf{S}_{\mathbf{xx}}^{-1}\right)$ é o celebrado mínimo quadrado de dois estágios. Use isso para calcular o resíduo $\hat{\varepsilon}_i \equiv y_i - \mathbf{z}_i'\hat{\boldsymbol{\delta}}(\ widehat\mathbf{W})$ e obter um estimador consistente $\widehat{\mathbf{S}}$ de $\mathbf{S}$ por (3.5.10),

Passo 2: Minimize $J\left(\tilde{\delta}, \widehat{\mathbf{S}}^{-1}\right)$ em termos de $\tilde{\delta}$. O valor que minimiza é o estimador GMM eficiente.

Amostra Pequena

- Essas propriedades assintóticas desejáveis do estimador GMM eficiente e estatísticas de teste associadas são transportadas para suas distribuições de amostras finitas? O estimador GMM eficiente usa $\widehat{\mathbf{S}}^{-1}$, uma função dos quartos momentos estimados, para $\widehat{\mathbf{W}}$.
- Esperaríamos que o estimador GMM eficiente tivesse propriedades de amostra pequena mais pobres do que os estimadores GMM que não usam quartos momentos para $\widehat{\mathbf{W}}$.
- A edição de julho de 1996 do JBES tem vários artigos examinando a distribuição de pequenas amostras de estimadores GMM e estatísticas de teste associadas para vários DGPs. Sua conclusão geral é que o estimador GMM igualmente ponderado com $\widehat{\mathbf{W}} = \mathbf{I}$ geralmente supera o GMM eficiente em termos de viés e variância em amostras finitas.
- Se α é o nível de significância assumido e c_α é o valor crítico associado para que $\text{Prob}\left(\chi^2 > c_\alpha\right) = \alpha$, a probabilidade em amostras finitas de que a estatística de Wald seja maior que c_α excede em muito α; o teste rejeita o nulo com muita frequência.

Testando Restrições Sobre-Identificadoras

Se a equação for exatamente identificada, então é possível escolher $\tilde{\delta}$ para que todos os elementos dos momentos amostrais $\mathbf{g}_n(\tilde{\boldsymbol{\delta}})$ são zero e a distância

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

é zero. (O $\tilde{\boldsymbol{\delta}}$ que faz isso é o estimador IV.) Se a equação for sobreidentificada, então a distância não pode ser definida exatamente como zero, mas esperamos que a distância minimizada seja próxima a zero. Acontece que, se a matriz de ponderação $\widehat{\mathbf{W}}$ for escolhida de forma ótima para que plim $\widehat{\mathbf{W}} = \mathbf{S}^{-1}$, então a distância minimizada é assintoticamente qui-quadrada.

Testando Restrições Sobre-Identificadoras

Seja $\widehat{\mathbf{S}}$ um estimador consistente de $\mathbf{S}$, e considere primeiro o caso em que a distância é avaliada no verdadeiro valor do parâmetro $\delta, J\left(\delta, \widehat{\mathbf{S}}^{-1}\right)$. Já que por definição $\mathbf{g}_n(\tilde{\boldsymbol{\delta}}) = \overline{\mathbf{g}}\left(\equiv \frac{1}{n}\Sigma_i \mathbf{g}_i\right)$ para $\tilde{\boldsymbol{\delta}} = \delta$, a distância é igual a

$$J\left(\delta, \widehat{\mathbf{S}}^{-1}\right) = n \cdot \overline{\mathbf{g}}' \widehat{\mathbf{S}}^{-1} \overline{\mathbf{g}} = (\sqrt{n}\overline{\mathbf{g}})' \widehat{\mathbf{S}}^{-1} (\sqrt{n}\overline{\mathbf{g}})$$

Proposition

Proposition 3.6 (Teste de Hansen sobre restrições sobreidentificadoras (Hansen, 1982)): Suponha um estimador consistente, $\widehat{\mathbf{S}}$, de $\mathbf{S}\left(= \mathrm{E}\left(\mathbf{g}_i \mathbf{g}_i'\right)\right)$. Sob algumas premissas,

$$J\left(\hat{\delta}\left(\widehat{\mathbf{S}}^{-1}\right), \widehat{\mathbf{S}}^{-1}\right)\left(= n \cdot \mathbf{g}_n\left(\hat{\delta}\left(\widehat{\mathbf{S}}^{-1}\right)\right)' \widehat{\mathbf{S}}^{-1} \mathbf{g}_n\left(\hat{\delta}\left(\widehat{\mathbf{S}}^{-1}\right)\right)\right) \underset{\mathrm{d}}{\rightarrow} \chi^2(K-L)$$

Testando Condições de Ortogonalidade

Proposition

Proposição 3.7 (testando um subconjunto de condições de ortogonalidade): Seja $\mathbf{x}_{i1}$ *um subvetor de* $\mathbf{x}_i$ *e fortaleça a Suposição 3.4 exigindo que a condição de posto para identificação seja satisfeita para* $\mathbf{x}_{i1}$ *(então* $\mathrm{E}\left(\mathbf{x}_{i1}\mathbf{z}_i'\right)$ *é de posto coluna cheio). Então, para quaisquer estimadores consistentes* $\widehat{\mathbf{S}}$ *de* $\mathbf{S}$ *e* $\widehat{\mathbf{S}}_{11}$ *de* $\mathbf{S}_{11}$*,*

$$C \equiv J - J_1 \underset{\mathrm{d}}{\rightarrow} \chi^2\left(K - K_1\right)$$

onde $K = \#\mathbf{x}_i$ *(dimensão de* $\mathbf{x}_i$ *),* $K_1 = \#\mathbf{x}_{i1}$ *(dimensão de* $\mathbf{x}_{i1}$ *), e* J *e* J_1 *são as estatísticas de Hansen para o modelo com todas e o modelo só com um subconjunto de restrições.*

Homocedasticidade

Assumindo:

$$\mathrm{E}\left(\varepsilon_i^2 \mid \mathbf{x}_i\right) = \sigma^2$$

Sob homocedasticidade condicional, a matriz de quartos momentos $\mathbf{S}$ $(= \mathrm{E}\left(\mathbf{g}_i\mathbf{g}_i'\right) = \mathrm{E}\left(\varepsilon_i^2\mathbf{x}_i\mathbf{x}_i'\right))$ pode ser escrito como um produto de segundos momentos:

$$\mathbf{S} = \sigma^2\mathbf{\Sigma}_{\mathbf{xx}}$$

Em que $\mathbf{\Sigma}_{\mathbf{xx}} = \mathrm{E}\left(\mathbf{x}_i\mathbf{x}_i'\right)$.

Homocedasticidade

O estimador de $\mathbf{S}$ sob esta premissa é

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

onde $\hat{\sigma}^2$ é algum estimador consistente a ser especificado abaixo. Pela estacionariedade ergódica, $\mathbf{S}_{\mathbf{xx}} \to_{\text{a.s.}} \mathbf{\Sigma}_{\mathbf{xx}}$. Assim, desde que $\hat{\sigma}^2$ seja consistente, não precisamos premissas sobre os quartos momentos para que $\widehat{\mathbf{S}}$ seja consistente.

GMM Eficiente vira MQ2E (e IV)

Como vimos na estimação eficiente de GMM, a matriz de ponderação é $\widehat{\mathbf{S}}^{-1}$. Usando o estimador do slide anterior $\widehat{\mathbf{S}}$ isso fica

$$\begin{aligned}\hat{\delta}\left(\widehat{\mathbf{S}}^{-1}\right) &= \left[\mathbf{S}_{\mathbf{xz}}'\left(\hat{\sigma}^2\mathbf{S}_{\mathbf{xx}}\right)^{-1}\mathbf{S}_{\mathbf{xz}}\right]^{-1}\mathbf{S}_{\mathbf{xz}}'\left(\hat{\sigma}^2\mathbf{S}_{\mathbf{xx}}\right)^{-1}\mathbf{s}_{\mathbf{xy}} \\ &= \left(\mathbf{S}_{\mathbf{xz}}'\mathbf{S}_{\mathbf{xx}}^{-1}\mathbf{S}_{\mathbf{xz}}\right)^{-1}\mathbf{S}_{\mathbf{xz}}'\mathbf{S}_{\mathbf{xx}}^{-1}\mathbf{S}_{\mathbf{xy}} \\ &= \hat{\boldsymbol{\delta}}\left(\mathbf{S}_{\mathbf{xx}}^{-1}\right) \equiv \hat{\boldsymbol{\delta}}_{\mathrm{2SLS}}\end{aligned}$$

A expressão para Avar $\left(\hat{\delta}_{2SLS}\right)$ é:

$$\text{Avar}\left(\hat{\boldsymbol{\delta}}_{\mathrm{2SLS}}\right) = \sigma^2 \cdot \left(\boldsymbol{\Sigma}_{\mathbf{xz}}'\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\Sigma}_{\mathbf{xz}}\right)^{-1}$$

Com o estimador:

$$\text{Avar}\left(\hat{\boldsymbol{\delta}}_{\mathrm{2SLS}}\right) = \hat{\sigma}^2 \cdot \left(\mathbf{S}_{\mathbf{xz}}'\mathbf{S}_{\mathbf{xx}}^{-1}\mathbf{S}_{\mathbf{xz}}\right)^{-1}$$

Sargan e Hansen

Um outro teste de restrições sobre-identificadoras é o Teste de Sargan - basicamente é n multiplicado pelo R^2 dos resíduos da regressão estimada por IV contra as variáveis exógenas. Sob homocedasticidade, a gente mostra que isso é igual ao nosso teste de Hansen:
Quando $\widehat{\mathbf{W}}$ é definido como sendo $\left(\hat{\sigma}^2 \mathbf{S_{xx}}\right)^{-1}$, a função critério GMM se torna:

$$J\left(\tilde{\delta}, \left(\hat{\sigma}^2 \cdot \mathbf{S_{xx}}\right)^{-1}\right) = n \cdot \frac{\left(\mathbf{S_{xy}} - \mathbf{S_{xz}}\tilde{\delta}\right)' \mathbf{S_{xx}}^{-1} \left(\mathbf{S_{xy}} - \mathbf{S_{xz}}\tilde{\delta}\right)}{\hat{\sigma}^2}$$

Que é a estatística de Sargan

HAC GMM

- E quando não temos motivo para assumir a homocedasticidade?
- Aí precisamos construir a matriz de pesos que seja sobusta a heterocedasticidade - e autocorrelação
- Se sabemos a priori que $\boldsymbol{\Gamma}_j = \mathbf{0}$ para $j > q$ em que q é conhecido e finito, então claramente $\mathbf{S}$ pode ser estimado por

$$\widehat{\mathbf{S}} = \widehat{\boldsymbol{\Gamma}}_0 + \sum_{j=1}^{q} \left(\widehat{\boldsymbol{\Gamma}}_j + \widehat{\boldsymbol{\Gamma}}_j' \right) = \sum_{j=-q}^{q} \widehat{\boldsymbol{\Gamma}}_j \quad \left(\text{assumindo: } \widehat{\boldsymbol{\Gamma}}_{-j} = \widehat{\boldsymbol{\Gamma}}_j' \right).$$

$$\widehat{\boldsymbol{\Gamma}}_j = \frac{1}{n} \sum_{t=j+1}^{n} \hat{\mathbf{g}}_t \hat{\mathbf{g}}_{t-j}' \quad (j = 0, 1, \ldots, n-1)$$

e

$$\hat{\mathbf{g}}_t \equiv \mathbf{x}_t \cdot \hat{\varepsilon}_t, \quad \hat{\varepsilon}_t = y_t - \mathbf{z}_t' \hat{\boldsymbol{\delta}}, \quad \hat{\boldsymbol{\delta}} \text{ consistente para } \boldsymbol{\delta}.$$

Usando Kernels para estimar S

Uma classe de estimadores, chamados de baseados em kernels (ou "não paramétricos") podem ser expressos como uma média ponderada das autocovariâncias estimadas:

$$\widehat{\mathbf{S}} = \sum_{j=-n+1}^{n-1} k\left(\frac{j}{q(n)}\right) \cdot \widehat{\mathbf{\Gamma}}_j$$

A função $k(\cdot)$ é chamada de kernel, e $q(n)$, bandwidth. O do slide anterior é um exemplo de kernel truncado, com $q(n) = q$

$$k(x) = \begin{cases} 1 & \text{for } |x| \leq 1 \\ 0 & \text{for } |x| > 1 \end{cases}$$

Kernel de Bartlett e Spectral Quadrático

Newey e West (1987) observaram que o estimador baseado em kernel pode ser definido como não negativo em amostras finitas se o kernel for o kernel de Bartlett:

$$k(x) = \begin{cases} 1 - |x| & \text{for } |x| \leq 1 \\ 0 & \text{for } |x| > 1 \end{cases}$$

O estimador baseado no kernel de Bartlett de S é chamado (em econometria) de estimador Newey-West. Por exemplo, para $q(n) = 3$, o estimador baseado em kernel inclui autocovariâncias de até dois (não três) lags:

$$\widehat{\mathbf{S}} = \widehat{\mathbf{\Gamma}}_0 + \left(\frac{2}{3}\right)\left(\widehat{\mathbf{\Gamma}}_1 + \widehat{\mathbf{\Gamma}}_1'\right) + \left(\frac{1}{3}\right)\left(\widehat{\mathbf{\Gamma}}_2 + \widehat{\mathbf{\Gamma}}_2'\right)$$

Outro exemplo é o Spectral Quadrático

$$k(x) = \frac{25}{12\pi^2 x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right).$$

Uma vez que $k(x) \neq 0$ for $|x| > 1$ aqui, todas as autocovariâncias estimadas $\widehat{\Gamma}_j$ $(j = 0, 1, \ldots, n-1)$ entram no cálculo de $\widehat{\mathbf{S}}$ mesmo se $q(n) < n - 1$.