GMM - Parte III

Bibliografia: Hayashi, cap. 3

Claudio Lucinda

FEA/USP

Overview

GMM e ML

- Seja $f\left(y_i \mid \mathbf{x}_i, \beta\right)$ a verdadeira densidade de probabilidade para uma variável aleatória y_i dado um conjunto de covariadas $\mathbf{x}_i$ e vetor de parâmetros β.
- A função de log-verossimilhança é $(1/n) \log L(\beta \mid \mathbf{y}, \mathbf{X}) = (1/n) \sum_{i=1}^{n} \log f\left(y_i \mid \mathbf{x}_i, \beta\right)$.
- O MLE, $\hat{\beta}_{\mathrm{ML}}$, é a estatística que maximiza esta função. (A divisão de log L por n não afeta a solução.)
- Maximizamos a função de verossimilhança de log igualando suas derivadas a zero, então o MLE é obtido resolvendo o conjunto de equações dos momento empíricos

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\partial \log f\left(y_i \mid \mathbf{x}_i, \hat{\beta}_{\mathrm{ML}}\right)}{\partial \hat{\beta}_{\mathrm{ML}}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{d}_i\left(\hat{\beta}_{\mathrm{ML}}\right) = \overline{\mathbf{d}}\left(\hat{\beta}_{\mathrm{ML}}\right) = \mathbf{0}$$

GMM e ML– II

A contrapartida populacional é

$$E\left[\frac{1}{n}\frac{\partial \log L}{\partial \beta}\right] = E\left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{d}_i(\beta)\right] = E[\bar{\mathbf{d}}(\beta)] = \mathbf{0}$$

Usando o que sabemos sobre estimadores GMM, se $E[\bar{\mathbf{d}}(\beta)] = \mathbf{0}$, então $\hat{\beta}_{\mathrm{ML}}$ é consistente e assintoticamente distribuído normalmente, com matriz de covariância assintótica igual a

$$\mathbf{V}_{\mathrm{ML}} = \left[\mathbf{G}(\beta)'\mathbf{G}(\beta)\right]^{-1}\mathbf{G}(\beta)'\{\mathrm{Var}[\bar{\mathbf{d}}(\beta)]\}\mathbf{G}(\beta)\left[\mathbf{G}(\beta)'\mathbf{G}(\beta)\right]^{-1},$$

Como visto antes, $\mathrm{Var}[\partial \log L/\partial \beta] = -E[\mathbf{H}(\beta)]$. Cortando um monte de coisa $(1/n)E[\mathbf{H}(\beta)]$, obtemos o resultado conhecido $\mathbf{V}_{\mathrm{ML}} = \{-E[\mathbf{H}(\beta)]\}^{-1}$.

Modelos de Escolha Discreta com Dados Agregados:

- Agora iremos discutir um pouco mais como podemos utilizar os modelos de escolha discreta, quando apenas temos dados agregados de mercado – ou seja, não temos microdados com as escolhas dos indivíduos.
- Supondo que os indivíduos ajam de acordo com os modelos de escolha discreta colocados anteriormente, temos que a quantidade vendida em um determinado mercado é dada por:

$$q_j = M \times s_j$$

Escolha Discreta com Dados Agregados: O bem externo

- Além disso, tem um ponto adicional, que é a existência de um bem “externo”, cujo preço não é estabelecido em resposta aos preços dos J produtos.
- Caso não fizéssemos isto, os consumidores seriam forçados a escolher apenas entre os bens existentes e a demanda dependeria somente das diferenças de preços. Portanto, um aumento generalizado dos preços não reduzirá a produção agregada.
- A colocação deste produto externo torna necessário que estimemos o tamanho do mercado, uma vez que $s_0 = M \times (1 - \sum^J s_j)$

Derivando as Especificações – Modelo Logit

- Supondo que os consumidores funcionem como no modelo Logit Multinomial tradicional, temos que

$$s_j = M \times \frac{e^{V_j+\xi_j}}{1+\sum_k e^{V_k+\xi_k}}$$

- Passando o Log dos dois lados:

$$\ln s_j = \ln M + V_j + \xi_j + \ln(1 + \sum_k e^{V_k+\xi_k})$$

- O bem externo, por sua vez, tem a sua participação de mercado dada por:

$$s_0 = M \times \frac{1}{1+\sum_k e^{V_k+\xi_k}}$$

- Passando o log também, temos que:

$$\ln s_0 = \ln M + \ln(1 + \sum_k e^{V_k+\xi_k})$$

Especificações – Modelo Logit (II):

- Tirando a diferença das duas equações, temos que:

$$\ln s_j - \ln s_0 = V_j + +\xi_j$$

- Em que ξ_j seria o termo erro econométrico.
- Esta especificação poderia ser estimada – mas existem alguns problemas que iremos discutir mais adiante:
  - IIA
  - Características não observadas dos consumidores
  - Endogeneidade - Vamos resolver por GMM
- Com relação ao IIA, o slide seguinte mostra uma possível solução – o modelo logit aninhado

Especificações – Modelo Nested Logit

- Seguindo Berry(1994, RAND), temos que a especificação de teste é a seguinte:

$$\ln s_j - \ln s_0 = V_j + (1-\sigma)\ln s_{j|K} + \xi_j$$

- Mais uma vez, o termo ξ_j representa o termo erro econométrico.
- CUIDADO!!! Na demonstração do Berry, o σ dele é exatamente igual ao meu $1-\sigma$.
- Caso $\sigma \to 1$, temos que esta especificação colapsa para o Logit Multinomial tradicional.
- Note que o "aninhamento" das alternativas é algo intrinsecamente determinado pelo analista; é ele quem coloca os produtos nos diferentes ninhos.

Estimação de Funções de Produção

- A estimação das funções de produção começou com o trabnalho de Cobb e Douglas (1928), que buscavam testar as implicações da teoria da distribuição baseada na produtividade marginal dos fatores.
- A principal crítica deste tipo de literatura é que os dados sobre fatores de produção, quando estamos falando em quantidades agregadas, são determinados simultaneamente aos valores do produto;
- desta forma, a função de produção não seria identificável.
- Vamos ilustrar este ponto mais detalhadamente, considerando a seguinte equação:,

$$q = a + \alpha z + \beta x + u$$

- Em que q é o log da quantidade produzida, z é o log do capital (ou qualquer outra quantidade de fatores "fixos" de produção), e x o log de todos os insumos variáveis.
  - O problema, levantado já em 1944 por Marshak e Andrews (1944), é que não podemos tratar x e z como verdadeiramente exógenos e estimar este negócio por OLS.

Estimação de Funções de Produção (II):

- A demanda pelo insumo variável, supondo que as empresas escolham as quantidades de x ao observar a realização de u, é dada por:

$$X = \left[\frac{p}{w}\beta e^{a+u} Z^{\alpha}\right]^{\frac{1}{1-\beta}}$$

- Uma vez que a escolha de X depende de u, temos problemas de endogeneidade.
- Um segundo problema é o da seleção de amostra. Um exemplo clássico é o de Dunne, Roberts e Samuelson (1988) encontrou taxas de saída maiores do que 30% entre intervalos de 5 e 5 anos.
- É de se supor que o principal determinante deste padrão de saída não é o componente aleatório ortogonal à escolha das variáveis.
  - Pelo contrário! É de se supor que as decisões da empresa tenham papel preponderante nas decisões de saída (i.e., falência) das empresas.

Endogeneidade da Função de Produção

- Dois exemplos de endogeneidade como a mencionada no slide anterior:
  1. Vamos supor que observemos um *cross section* de empresas. Algumas delas são mais produtivas e têm melhores gestores.E por isso, elas podem precisar de menos trabalho para produzir a mesma quantidade. Ou seja, estas empresas vão produzir mais com menos trabalho e por isso OLS vai subestimar β_l
  2. Suponha que, agora observamos um painel e, em cada período a empresa tem um choque de produtividade – positivo por ela observado e com este valor vai contratar mais. Ou seja, no final o aumento de produção com o choque de produtividade vai ser devido às duas coisas mas OLS vai atribuir TODO o aumento de produção ao aumento de trabalho, sobrestimando β_l
- Ou seja, pode ir para qualquer direção.
- Usualmente, assumimos que o problema da endogeneidade é mais presente no trabalho.

Atrição da Amostra em Funções de Produção

- Pra ilustrar melhor este ponto, suponha que as empresas sejam monpólios que são dotados exogenamente de diferentes quantidades de capital, como mencionado por Griliches e Mairesse (1995)
- Desta forma, dependendo do valor de u, elas podem decidir sair ou não.
  - Ou seja, se u for “muito ruim”, pode ser melhor vender o valor residual da empresa.
- Isto pode ser racionalizado com a seguinte regra de saída:

$$\chi(u, Z, p, w, a, \beta, \alpha) = 0 \, se \, \Pi(u, Z, p, w, a, \beta, \alpha) < \Psi$$

- Em que Π é a parte variável dos lucros e Ψo valor residual da empresa.

Atrição de Amostra em Funções de Produção

- O ponto aqui é que esta condição gerará uma correlação entre u e Z condicional à empresa estar no mercado.
- Isto ocorre porque as empresas com maiores estoques de capital devem ter maiores lucros variáveis e, portanto, podem suportar piores choques u sem sair do mercado.
  - Ou seja, devemos observar apenas aquelas empresas em que o Z é relativamente grande e/ou u relativamente pequeno.
  - Isso implica que as empresas menores devem sair da amostra

Soluções Tradicionais para o Problema:

- Existem duas formas de lidar com alguns dos problemas mencionados aqui:
  - Aproveitamento de amostra de dados em painel
  - Utilização de Variáveis Instrumentais
- Vamos representar nosso modelo da seguinte forma:

$$y_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + \omega_{it} + \eta_{it}$$

- Em que ω_{it} representa a parte de informação não observada pelo econometrista que é observada pela empresa na tomada de suas decisões, e η_{it} representa a parte da informação não observada pelo econometrista que também não é observada pela empresa.
  - ω_{it}: capacidade gerencial
  - η_{it}: comportamento anômalo.

Solução I – Dados em Painel

- Uma solução interessante para o problema da endogeneidade é utilizar a informação da estrutura em painel dos dados.
- Aqui estamos considerando que a parte ω_{it} é constante ao longo do tempo
- Neste caso, podemos usar os diferentes estimadores mencionados em Wooldridge (2002), e que alguns de vocês viram no curso de Econometria com Dados em Painel:
  - Primeiras Diferenças: $(y_{it} - y_{it-1}) = \beta_k(k_{it} - k_{it-1}) + \beta_l(l_{it} - l_{it-1}) + (\eta_{it} - \eta_{it-1})$
  - Efeitos Fixos: $(y_{it} - \bar{y}_i) = \beta_k(k_{it} - \bar{k}_i) + \beta_l(l_{it} - \bar{l}_i) + \eta_{it}$

Dados em Painel:

- Dada a hipótese que η_{it} são independentes das escolhas de insumos em qualquer instante do tempo, podemos estimar as duas equações por OLS.
  - Esta hipótese é a chamada "exogeneidade estrita". Em alguns casos, podemos estimar este modelo de efeitos fixos sob a premissa de "exogeneidade seqüencial", em que η_{it} não é correlacionado com a escolha de insumos nos instantes anteriores à t.
- Esta premissa de ω constante ao longo do tempo também resolveria o problema da atrição da amostra, caso a regra de saída dependa somente de ω, e não de η_{it}.
- No entanto, existem algumas limitações da abordagem com dados em painel.

Dados em Painel – Limitações

- É uma premissa complicada assumir que os ω sejam constantes ao longo do tempo, especialmente quando bases de microdados mais longas estão disponíveis.
- Além disso, pode haver interesse nas mudanças em ω propriamente dito.
- Outro problema é que, quando há erros de medida nos insumos, os estimadores de dados em painel podem gerar estimativas piores que OLS - em especial, β_k muito baixos
  - Griliches e Hausman (1986) mostram que quando os insumos são mais correlacionados que os erros de medida, pode se reduzir a razão sinal/ruído nas variáveis independentes (a parcela da variabilidade mais devida a alterações na variável mesmo do que nos erros de medida).
- Um terceiro problema é que, em geral, efeitos fixos dão estimativas muito baixas para os coeficientes de retornos de escala.
  - Além disso, os dados mudam muito se pegamos o painel inteiro ou apenas a parte balanceada do mesmo.

Solução II – Variáveis Instrumentais:

- As abordagens de variáveis instrumentais se baseiam na premissa que é possível encontrar instrumentos adequados.
- Alguns instrumentos “naturais”
  - Preços dos fatores de produção: se eles forem independentes de ω, tudo bem
- Estamos, neste caso, assumindo que não existe poder de mercado por parte das empresas na aquisição de insumos.
- No entanto, existem problemas com esta abordagem:
  - Preços pagos por insumos não são reportados pelas empresas
  - Nem sempre há variação econometricamente “boa” nestas variáveis
  - É difícil imaginar que ω não seja afetado pelos preços dos insumos
  - Não resolve a questão da saída

Solução III – Painéis Dinâmicos

- Uma linha de ataque aos problemas mencionados anteriormente envolve a estimação de modelos de painel dinâmico.
- Vamos começar supondo o seguinte modelo:

$$\begin{aligned} y_{it} &= \gamma_t + \beta_k k_{it} + \beta_l l_{it} + f_i + \eta_{it} \\ \eta_{it} &= \rho \eta_{it-1} + \epsilon_{it} \\ \epsilon_{it} &\sim MA(0) \end{aligned}$$

- Assume-se que a parte da produtividade tenha um componente aleatório e um componente persistente – para refletir o fato que a produtividade apresenta forte persistência ao longo do tempo.
- Este modelo tem uma representação dinâmica da seguinte forma:

$$\begin{aligned} y_{it} &= \beta_l l_{it} - \rho \beta_l l_{it-1} + \beta_k k_{it} - \rho \beta_k k_{it-1} + \rho y_{it-1} + \\ &+ (\gamma_t - \rho \gamma_{t-1}) + (f_i(1-\rho) + \epsilon_{it}) \end{aligned}$$

Painéis Dinâmicos

- Podemos reescrever esta equação como:

$$y_{it} = \pi_1 l_{it} + \pi_2 l_{it-1} + \pi_3 k_{it} + \pi_4 k_{it-1} + \pi_5 y_{it-1} + \gamma_t^* + (f_i^* + \epsilon_{it})$$

- Sujeita a duas restrições:
  - $\pi_2 = -\pi_1 \pi_5$
  - $\pi_4 = -\pi_3 \pi_5$
- Arellano e Bond (1991) supõem as seguintes premissas sobre as condições iniciais:
  - $E(\mathbf{x_{i1}} \epsilon_{it}) = 0$, sendo que $\mathbf{x_{it}} = (y_{it}, l_{it}, k_{it})$
- Podemos utilizar as seguintes condições de momento:

$$m(\theta) = E(\mathbf{x_{it-s}} \Delta \epsilon_{it}) = 0$$

- Em que $s \geq 2$ caso não tenhamos erros de medida.

Painéis Dinâmicos (II):

- O problema é que a estimação tem propriedades ruins quando os níveis defasados da série, os $\mathbf{x_{it-s}}$ são pouco correlacionados com as primeiras diferenças subseqüentes $\Delta\epsilon_{it}$.
- Causas possíveis para isso:
  - Processo marginal de determinação de l_{it} e k_{it} são muito persistentes, próximos a ter uma raiz unitária.
- Neste casos, os $\mathbf{x_{it-s}}$ são instrumentos fracos

Problemas de GMM-Diff

- Esta abordagem também tem suas limitações. Em algumas aplicações, é comum encontrar estimativas muito baixas de β_l e β_k e grandes erros-padrão.
- Geralmente, a validade das restrições sobre-identificadoras é rejeitada. Além disso, a hipótese que o processo dos η seja exatamente AR(1) pode ser rejeitada, o que implica que os $\mathbf{x_{it-2}}$ não seriam instrumentos válidos.
- Além disso, a transformação em primeira diferença pode levar ao mesmo problema no caso de erros de medida nas variáveis

GMM – Sistema

- Supondo adicionalmente que $E(\Delta l_{it} f_i^*) = E(\Delta k_{it} f_i^*) = 0$, e que as condições iniciais incluam $E(\Delta y_{i2} f_i^*) = 0$, podemos incluir as seguintes condições de momento na estimação:

$$m^2(\theta) = E(\Delta \mathbf{x_{it-s}}(f_i^* + \epsilon_{it})) = 0$$

- Com $s = 1$ caso não haja erros de medida.
- Este é o chamado estimador GMM em Sistema de Blundell e Bond (1998).
- Podemos testar a adequação das restrições adicionais por meio de um teste de diferença de Sargan:
  - Calcular a diferença entre os valores da função objetivo e comparar com o valor crítico de uma distribuição χ^2, com número de graus de liberdade igual à diferença de condições de ortogonalidade nos dois casos.