ACH3657

Métodos Quantitativos para Avaliação de Políticas Públicas

Aula Teórica 09

A Variância dos Estimadores de MQO

Alexandre Ribeiro Leichsenring
alexandre.leichsenring@usp.br

Organização

1. A Variância dos Estimadores de MQO
   - Variâncias amostrais dos Estimadores de Inclinação de MQO
2. Componentes das Variâncias de MQO: Multicolinearidade
   - Variância do erro: σ^2
   - Variação amostral total em x_j: SQT$_j$
   - As relações lineares entre as variáveis independentes: R_j^2
   - Multicolinearidade
3. Estimação de σ^2: Os Erros-Padrão dos Estimadores de MQO

A Variância dos Estimadores de MQO

- Além de conhecermos as tendências centrais dos $\hat{\beta}_j$, é importante termos uma medida da dispersão de sua distribuição amostral
- Antes de encontrarmos as variâncias, vamos adicionar uma hipótese de homoscedasticidade, como no caso da RLS. Por duas razões:
  - Ao impor a hipótese de variância constante do erro, as fórmulas são simplificadas
  - MQO tem uma propriedade importante de eficiência se acrescentamos a hipótese de homoscedasticidade

Hipótese RLM.5 (Homocedasticidade)

$$\text{Var}(u|x_1, x_2, \ldots, x_k) = \sigma^2$$

- A hipótese RLM.5 significa que a variância do termo erro, u, condicionada às variáveis explicativas, é a mesma para todas as combinações de $x_1, x_2, \ldots, x_n$
- Se essa hipótese é violada, o modelo exibe heteroscedasticidade, (exatamente como no caso de duas variáveis)

Na equação

$$salarioh = \beta_0 + \beta_1\ educ + \beta_2\ exper + \beta_3\ perm + u$$

a homoscedasticidade requer que a variância de u não dependa dos níveis de educação, experiência ou permanência. Isto é:

$$\text{Var}(u|educ,\ exper,\ perm) = \sigma^2$$

Se a variância varia com qualquer uma das três variáveis explicativas, então há heteroscedasticidade

$\Rightarrow$ As hipóteses RLM. 1 a RLM.5 são, em conjunto, conhecidas como as hipóteses de Gauss-Markov (para a regressão de corte transversal)

Teorema

Sob as hipóteses RLM.1 a RLM.5, condicionadas aos valores amostrais das variáveis independentes

$$\text{Var}(\hat{\beta}_j) = \frac{\sigma^2}{\text{SQT}_j(1 - R_j^2)}$$

para $j = 1, 2, \ldots, k$.

- $\text{SQT}_j = \sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$ é a variação total em x_j
- R_j^2 é o R^2 da regressão de x_j sobre todas as outras variáveis independentes (incluindo um intercepto)

Obs.

- Todas as hipóteses de Gauss-Markov são usadas na obtenção dessa fórmula
- A magnitude de $\text{Var}(\hat{\beta}_j)$ é importante na prática: uma variância maior significa um estimador menos preciso, e isso se traduz em intervalos de confiança maiores e testes de hipóteses menos acurados

A variância de $\hat{\beta}_j$ depende de 3 fatores:

- σ^2
- SQT_j
- R_j^2

Vamos considerar cada um desses fatores.

Variância do erro: σ^2

$$\text{Var}(\hat{\beta}_j) = \frac{\sigma^2}{\text{SQT}_j(1 - R_j^2)}$$

- Quanto maior σ^2 maior a variância de $\hat{\beta}_j$
  - Quanto mais "ruído" na equação, mais difícil estimar o efeito parcial de qualquer uma das variáveis independentes sobre y
  - σ^2 é uma característica da população (não tem nada a ver com o tamanho da amostra)
- Componente desconhecido (veremos como estimar)
- Só uma maneira de reduzir a variância do erro: adicionar mais variáveis explicativas à equação (retirar alguns fatores do termo erro) – nem sempre é possível ou desejável

Variação amostral total em x_j: SQT$_j$

$$\text{Var}(\hat{\beta}_j) = \frac{\sigma^2}{\text{SQT}_j(1 - R_j^2)}$$

$$\text{SQT}_j = \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2$$

- Quanto maior a variação total em x_j menor é Var$(\hat{\beta}_j)$
- Tudo o mais sendo igual, preferimos ter tanta variação amostral em x_j quanto possível
- Aumentar tamanho da amostra faz aumentar variação nas variáveis explicativas x_j
- Componente da variância que depende sistematicamente do tamanho da amostra
- Quando SQT$_j$ é pequeno, Var$(\hat{\beta}_j)$ pode ser muito grande

$$\text{SQT}_j \to 0 \Rightarrow \text{Var}(\hat{\beta}_j) \to \infty$$

O caso extremo de nenhuma variação amostral em x_j não é permitido pela hipótese RLM.4

As relações lineares entre as variáveis independentes: R_j^2

$$\text{Var}(\hat{\beta}_j) = \frac{\sigma^2}{\text{SQT}_j(1 - R_j^2)}$$

- R_j^2 é obtido de uma regressão que envolve somente as variáveis independentes do modelo original – que x_j desempenha o papel de uma variável dependente
- R_j^2 é a proporção da variação total de x_j, que pode ser explicada pelas outras variáveis independentes
- Para dados σ^2 e SQT_j, a menor $\text{Var}(\hat{\beta}_j)$ é obtida quando $R_j^2 = 0$
  - Só ocorre quando x_j tem correlação amostral zero com cada uma das outras variáveis independentes
  - Melhor caso, mas raramente encontrado
- Outro caso extremo: $R_j^2 = 1$
  - Excluído pela hipótese RLM.4, pois $R_j^2 = 1$ significa que, na amostra, x_j, é uma combinação linear perfeita de outras variáveis independentes da regressão
- Caso mais relevante: R_j^2 "próximo" de 1
  - Pode fazer com que $\text{Var}(\hat{\beta}_j)$ seja grande:

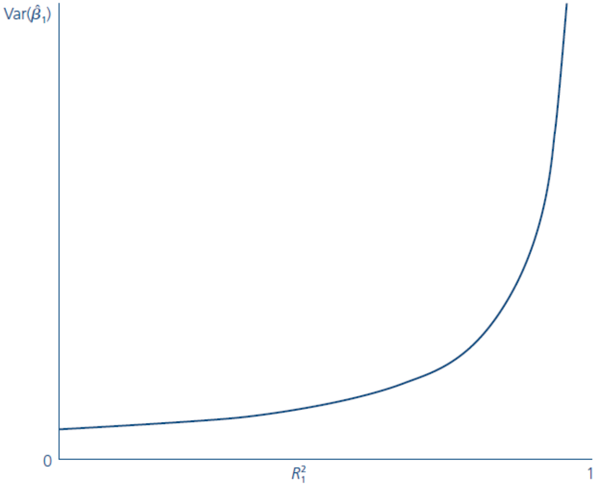
    $$R_j^2 \to 1 \Rightarrow \text{Var}(\hat{\beta}_j) \to \infty$$

  - Correlação alta (mas não perfeita) entre duas ou mais variáveis independentes é chamada **multicolinearidade**

Multicolinearidade

- R^2 próximo de um não viola hipótese RLM.4
- O “problema” da multicolinearidade não é, de fato, bem definido.
- Não há um número absoluto que podemos citar para concluir que a multicolinearidade é um problema
- E.g.: $R_j^2 = 0,9$ significa que 90% da variação amostral em x_j pode ser explicada pelas outras variáveis independentes do modelo
- x_j tem uma forte relação linear com as outras variáveis independentes
- No entanto, se isso se traduz em uma Var$(\hat{\beta}_j)$ grande demais depende da magnitude de σ^2 e de SQT$_j$
- Essencialmente o que importa é a relação entre Var$(\hat{\beta}_j)$ e a magnitude de β_j

Var$(\hat{\beta}_j)$ como uma função de R_j^2



- Assim como um valor grande de R_j^2 pode causar uma Var$(\hat{\beta}_j)$ grande, um valor pequeno de SQT$_j$ também pode fazer com que Var$(\hat{\beta}_j)$ seja grande
- Portanto, um tamanho pequeno da amostra pode também levar a variâncias amostrais grandes
- Preocupar-se com graus elevados de correlação entre variáveis independentes da amostra não é, de fato, diferente de se preocupar com um tamanho pequeno da amostra: ambos funcionam para aumentar Var$(\hat{\beta}_j)$
- Reagindo à obsessão dos econometristas pela multicolinearidade, o famoso econometrista da Universidade de Wisconsin Arthur Goldberger, criou (jocosamente) o termo *micronumerosidade*, que ele define como o "problema do tamanho pequeno da amostra"
- Uma coisa é clara: tudo mais sendo igual, para estimar Var$(\hat{\beta}_j)$ é melhor ter menos correlação entre x_j e as outras variáveis independentes
- Nas ciências sociais somos geralmente coletores passivos de dados: não há uma boa maneira de reduzir as variâncias dos estimadores não-viesados que não seja coletar mais dados
- Podemos tentar suprimir outras variáveis independentes do modelo – infelizmente, suprimir uma variável que pertence ao modelo populacional pode levar viés, como vimos

Exemplo

Suponha que estamos interessados em estimar o efeito de várias categorias de despesas de escolas sobre o desempenho de estudantes.

- Provável que as despesas com salários de professores, materiais institucionais, atletismo etc. estejam altamente correlacionadas:
  - Escolas mais ricas tendem a gastar mais com tudo
  - Escolas mais pobres gastam menos com tudo
- Pode ser difícil estimar o efeito de qualquer categoria de despesa específica sobre o desempenho dos estudantes quando há pouca variação em uma categoria que não pode ser, em grande medida, explicada por variações das outras categorias de despesas (isso leva a um R_j^2 alto para cada uma das variáveis de despesas)
- Em certo sentido, nós mesmos nos impusemos o problema: estamos formulando questões que podem ser sutis demais para que os dados disponíveis as respondam com alguma precisão
- Provavelmente, podemos fazer algo muito melhor mudando o escopo da análise e agrupando todas as categorias de despesa em uma única categoria, desde que não mais estivéssemos tentando estimar o efeito parcial de cada categoria separadamente

Ponto importante

- Um elevado grau de correlação entre certas variáveis independentes pode ser irrelevante no que diz respeito a quão bem podemos estimar outros parâmetros do modelo
- Por exemplo, considere um modelo com três variáveis independentes:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

  com x_2 e x_3 altamente correlacionados.

  $\Rightarrow$ Var$(\hat{\beta}_2)$ e Var$(\hat{\beta}_2)$ podem ser grandes.
- Mas o valor da correlação entre x_2 e x_3 não tem efeito sobre Var$(\hat{\beta}_1)$
- Se x_1 é não correlacionado com x_2 e x_3, então:
  - $R_1^2 = 0$
  - Var$(\hat{\beta}_1) = \frac{\sigma^2}{\text{SQT}_1}$
- Se β_1 é o parâmetro de interesse, não devemos nos preocupar com o valor da correlação entre x_2 e x_3.

Exercício

Suponha que você postula um modelo que explica a nota do exame final em termos da frequência às aulas. Assim, a variável dependente é a nota do exame final, e a principal variável explicativa é o número de aulas frequentadas. A fim de controlar as aptidões dos estudantes e pelos esforços fora da sala de aula, você inclui entre as variáveis explicativas a nota acumulada durante todo o curso, a nota do teste de avaliação de conhecimentos para ingresso em curso superior e as medidas do desempenho do estudante no ensino médio. Alguém diz: "Você não pode esperar aprender nada com esse exercício, pois todas essas variáveis são, provavelmente, altamente colineares". Qual seria sua resposta?

Estimação de σ^2: Os Erros-Padrão dos Estimadores de MQO

- Lembremos que $\sigma^2 = \text{Var}(u)$
- É necessário de um estimador não-viesado de σ^2 para obter estimadores não-viesados de $\text{Var}(\hat{\beta}_j)$
- Estimativas são obtidas através dos termos de erro estimados

$$\hat{u}_i = y_i - \hat{y}_i$$

com

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k$$

- O estimador não-viesado de σ^2 no caso geral da regressão múltipla é:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(\hat{u}_i^2)}{n-k+1} = \frac{\text{SQR}}{n-k+1}$$

Teorema (Estimador não viesado de σ^2

Sob as hipóteses de Gauss-Markov RLM.1 a RLM.5:

$$\mathbf{E}(\hat{\sigma}^2) = \sigma^2$$

Erro-Padrão da Regressão

$\hat{\sigma} = \sqrt{\hat{\sigma}^2}$ é denominado **Erro-Padrão da Regressão** (EPR)

- O EPR é um estimador do desvio-padrão do termo de erro
- Essa estimativa é usualmente informada pelos programas de regressão, embora ela seja chamada de nomes diferentes pelos diferentes programas.
- Para construir intervalos de confiança e conduzir testes sobre os parâmetros da equação de regressão, precisaremos estimar o desvio-padrão de $\hat{\beta}_j$:

$$\text{dp}(\hat{\beta}_j) = \frac{\sigma}{\left[\text{SQT}_j(1 - R_j^2)\right]^{1/2}}$$

Como σ é desconhecido, ele é substituído pelo seu estimador $\hat{\sigma}$. Isso nos dá o erro-padrão de $\hat{\beta}_j$:

$$\text{ep}(\hat{\beta}_j) = \frac{\hat{\sigma}}{\left[\text{SQT}_j(1 - R_j^2)\right]^{1/2}}$$

Observações

- Se os erros exibem heteroscedasticidade, fórmula do erro-padrão não é um estimador válido de dp$(\hat{\beta}_j)$
- A presença de heteroscedasticidade não causa viés em $\hat{\beta}_j$, mas leva viés à fórmula usual da Var$(\hat{\beta}_j)$

  $\Rightarrow$ Invalida a estimativa do erro padrão!
- Se suspeitarmos de heteroscedasticidade, então os erros-padrão de MQO "habituais" não são válidos (alguma ação corretiva deve ser tomada)

Eficiência de MQO

Teorema de Gauss-Markov

Sob as hipóteses RLM.1 a RLM.5:

$$\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$$

são os melhores estimadores lineares não-viesados (BLUEs) de

$$\beta_0, \beta_1, \ldots, \beta_k$$

- Por causa desse teorema as hipóteses RLM.I a RLM.5 são conhecidas como as hipóteses de Gauss-Markov
- Nenhum outro estimador será melhor que MQO
- Se qualquer uma das hipóteses de Gauss-Markov for violada, o teorema não é mais válido

Resumo

1. O modelo de regressão múltipla nos permite, efetivamente, manter os outros fatores fixos ao examinarmos os efeitos de uma variável independente particular sobre a variável dependente. Ele permite, explicitamente, que as variáveis sejam correlacionadas
2. Embora o modelo seja linear em seus *parâmetros*, ele pode ser usado para modelar relações não-lineares ao se escolher, apropriadamente, as variáveis dependente e independente
3. O método de mínimos quadrados ordinários é facilmente aplicado para estimar o modelo de regressão múltipla. Cada estimativa de inclinação mede o efeito parcial da variável independente correspondente sobre a variável dependente, mantendo todas as outras variáveis independentes fixas.
4. R^2 é a proporção da variação amostral da variável dependente explicada pelas variáveis independentes, e é utilizado como uma medida do grau de ajuste. E importante não dar importância demais ao valor do R^2 na avaliação de modelos econométricos

Resumo

5. Sob RLM.1 a RLM.4, os estimadores de MQO são não-viesados
   - incluir uma variável irrelevante em um modelo não tem nenhum efeito sobre a inexistência de viés dos estimadores
   - omitir uma variável importante faz com que MQO seja viesado (em muitas circunstâncias, a direção do viés pode ser determinada)
6. Sob RLM.1 a RLM.5, a variância de um estimador de inclinação de MQO é dada por

$$\text{Var}(\hat{\beta}_j) = \frac{\sigma^2}{[\text{SQT}_j(1 - R_j^2)]}$$

   - quando a variância do erro σ^2 cresce, $\text{Var}(\hat{\beta}_j)$ cresce
   - quando a variação amostral em x_j, SQT_j, aumenta, $\text{Var}(\hat{\beta}_j)$ diminui
   - o termo R_j^2 mede a magnitude da colinearidade entre x_j e as outras variáveis explicativas
   - Quando R_j^2 se aproxima de um, $\text{Var}(\hat{\beta}_j)$ cresce ilimitadamente

7. Adicionar uma variável irrelevante a uma equação geralmente aumenta as variâncias dos demais estimadores de MQO, por causa da multicolinearidade
8. Sob as hipóteses de Gauss-Markov (RLM.I a RLM.5), os estimadores de MQO são os melhores estimadores lineares não-viesados (BLUE).