AULA 12 - Normalidade e Inferência em Regressão Múltipla - Parte 2

Econometria I

Área quantitativa - IE/UFRJ

Testes de hipóteses sobre combinação linear dos parâmetros

- Na aula passada testamos hipóteses sobre um único β_j de cada vez.
- Em algumas aplicações queremos testar mais de um dos parâmetros da população.
- Vamos começar a análise com um exemplo para comparar os retornos da educação de cursos superiores profissionalizantes de dois anos (cp) e de cursos superiores de quatro anos (univ).

$$\log(\text{salário}) = \alpha + \beta_1 \text{cp} + \beta_2 \text{univ} + \beta_3 \text{exper} + u \qquad (1)$$

- A hipótese de interesse é se um ano no curso profissionalizante é equivalente a um ano na universidade

$$H_0 : \beta_1 = \beta_2$$

sob H_0 um ano a mais no curso profissionalizante e um ano a mais na universidade levam ao mesmo aumento percentual em salário.

Testes de hipóteses sobre combinação linear dos parâmetros

- A hipótese alternativa de interesse é unilateral: um ano no curso profissionalizante é menos valioso do que um ano na universidade

$$H_1 : \beta_1 < \beta_2$$

- Não podemos usar simplemente as estatísticas t individuais de β_1 e β_2 para testar H_0
- Para fazer o teste reescrevemos

$$H_0 : \beta_1 - \beta_2 = 0$$

$$H_1 : \beta_1 - \beta_2 < 0$$

- A estatística t é baseada em se a diferença estimada $\hat{\beta}_1 - \hat{\beta}_2$ é estatísticamente zero.

Testes de hipóteses sobre combinação linear dos parâmetros

- Para considerar o erro de nossos estimadores, padronizamos essa diferença ao dividí-la pelo erro-padrão:

$$t = \frac{\hat{\beta}_1 - \hat{\beta}_2}{ep(\hat{\beta}_1 - \hat{\beta}_2)}$$

  Assim o teste segue o procedimento anterior, escolhendo um nível de significância.

- Nossa única dificuldade neste teste é calcular $ep(\hat{\beta}_1 - \hat{\beta}_2)$, que não é fornecido automaticamente nos programas econométricos. Temos duas formas de calcular:
  - $ep(\hat{\beta}_1 - \hat{\beta}_2) = \{[ep(\hat{\beta}_1)]^2 + [ep(\hat{\beta}_2)]^2 - 2Cov(\hat{\beta}_1, \hat{\beta}_2)\}^{1/2}$
  - Estimar um modelo diferente que produz diretamente o erro-padrão de interesse.

Teste F

- Agora queremos testar se um grupo de variáveis não tem efeito sobre a variável dependente.
- Suponha que queremos testar $q < k + 1$ variáveis, assuma que sejam as q últimas variáveis: $x_{k-q+1}, ..., x_k$.
- A hipótese nula testada, será

$$H_0 : \beta_{k-q+1} = 0, ..., \beta_k = 0$$

- A hipótese alternativa é

$$H_1 : H_0 \text{ não é verdadeira}$$

- Em princípio poderíamos fazer um teste t para cada restrição individualmente, mas isso, pode ser enganoso.
- Podemos ter resultados em que as estísticas t não podem rejeitar H_0 para as q variáveis, mas quando testamos em conjunto não necessariamente isso pode ser verdade.

Teste F

- O que devemos observar para o grupo de variáveis que serão ou não retiradas do modelo será a relação custo-benefício da soma dos resíduos ao quadrado (SQR).
- Sabemos que o SQR sempre aumenta quando variáveis são retiradas do modelo.
- A questão é saber se esse aumento é suficientemente grande, relativamente ao SQR do modelo com todas as varáveis.
- Para fazer o teste F considere dois modelos:
  - Modelo irrestrito com k variáveis:

$$y = \alpha + \beta_1 x_1 + ... + \beta_k x_k + u$$

  - Modelo restrito com $q < k$ variáveis:

$$y = \alpha + \beta_1 x_1 + ... + \beta_{k-q} x_{k-q} + u$$

Teste F

- Como dizemos anteriormente olharemos para o aumento relativo em SQR quando nos movemos do modelo irrestrito para o restrito.
- Neste sentido a estatística F é definida como:

$$F = \frac{(SQR_r - SQR_{ir})/q}{SQR_{ir}/(n-k-1)}$$

  em que SQR_r é a soma dos resíduos quadrados do modelo restrito e SQR_{ir} é do modelo irrestrito.
- Como SQR_r não pode ser menor que o SQR_{ir} a estatística F é sempre não negativa.
- No numerador de F temos a divisão por q, que o número de restrições impostas:

$$q = \text{graus de liberdade do numerador} = gl_r - gl_{ir}$$

- No denominador do F temos a divisão pelos graus de liberdade do modelo irrestrito

$$n - k - 1 = \text{graus de liberdade do denominador} = gl_{ir}$$

Teste F

- Para usar a estatística F, devemos conhecer sua distribuição amostral sob a hipótese nula para escolher os valores críticos e as regras de rejeição.
- Temos que sob H_0 (e assumindo que as hipóteses se mantêm), F é distribuído como uma variável aleatória F com $(q, n-k-1)$ graus de liberdade.

$$F \sim F_{q,n-k-1}$$

- A distribuição $F_{q,n-k-1}$ está tabelada.
- Da definição de F, é claro que rejeitaremos H_0 em favor de H_1 quando F for suficientemente " grande" $F > c$
- Se H_0 é rejeitada, dizemos que $x_{k-q+1}, ..., x_k$ são estatisticamente significantes conjuntamente.

EXERCÍCIO: Usando os dados Wooldridge MLB1.RAW do exemplo do capítulo 4.5 estimar o modelo irrestrito e restrito e fazer o teste F para as três últimas variáveis do modelo irrestrito.