E

O Modelo de Regressão Linear em Forma Matricial

Este apêndice deriva vários resultados da estimação do modelo de regressão linear múltipla usando notação matricial e álgebra de matrizes (veja um resumo no Apêndice D). O material apresentado aqui é muito mais avançado do que aparece no texto deste livro.

E.1 O MODELO E A ESTIMAÇÃO DE MÍNIMOS QUADRADOS ORDINÁRIOS

Em todo este Apêndice, usamos o subscrito t para indexar observações e n para representar o tamanho da amostra. É útil escrevermos o modelo de regressão linear múltipla com k parâmetros, como segue:

$$y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \ldots + \beta_k x_{tk} + u_t,\ t = 1, 2, \ldots, n, \tag{E.1}$$

onde y_i é a variável dependente para a observação t, e x_{ij}, $j = 2, 3, \ldots, k$ são as variáveis independentes. Observe como nossa convenção notacional neste apêndice difere do texto do livro: chamamos o intercepto de β_1 e fazemos $\beta_2, \ldots, \beta_k$ representar os parâmetros de inclinação. Essa terminologia não é importante, mas simplifica a notação do método matricial na regressão múltipla.

Para cada t, defina um vetor $1 \times k$, $\boldsymbol{x}_t = (1, x_{t2}, \ldots, x_{tk})$ e seja $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_k)'$ o vetor $k \times 1$ de todos os parâmetros. Então, podemos escrever (E.1) como

$$y_t = \boldsymbol{x}_t\boldsymbol{\beta} + u_t,\ t = 1, 2, \ldots, n. \tag{E.2}$$

[Alguns autores preferem definir $\boldsymbol{x}_t$ como um vetor coluna, caso em que $\boldsymbol{x}_t$ é substituído por $\boldsymbol{x}_t'$ em (E.2). Matematicamente, faz mais sentido defini-lo como um vetor linha]. Podemos escrever (E.2) em notação matricial completa, definindo apropriadamente os dados dos vetores e matrizes. Seja $\boldsymbol{y}$ o vetor $n \times 1$ das observações de y: o $t^{\text{-ésimo}}$ elemento de $\boldsymbol{y}$ é y_t. Seja $\mathbf{X}$ o vetor $n \times k$ das observações das variáveis explicativas. Em outras palavras, a $t^{\text{-ésima}}$ linha de $\mathbf{X}$ consistirá do vetor x_t. De forma equivalente, o $(t, j)^{\text{-ésimo}}$ elemento de $\mathbf{X}$ é simplesmente x_{ij}:

$$\underset{n \times k}{\mathbf{X}} \equiv \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_n \end{bmatrix} = \begin{bmatrix} 1 & x_{12} & x_{13} & \cdots & x_{1k} \\ 1 & x_{22} & x_{23} & \cdots & x_{2k} \\ \vdots & & & & \\ 1 & x_{n2} & x_{n3} & \cdots & x_{nk} \end{bmatrix}.$$

Finalmente, seja $\boldsymbol{u}$ o vetor $n \times 1$ dos distúrbios não observados. Então, podemos escrever (E.2) de todas as n observações em **notação matricial**:

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{u}. \tag{E.3}$$

Lembre-se, como $\mathbf{X}$ é $n \times k$ e $\boldsymbol{\beta}$ é $k \times 1$, $\mathbf{X}\boldsymbol{\beta}$ será $n \times 1$.

A estimação de $\boldsymbol{\beta}$ prossegue minimizando a soma dos resíduos quadrados, como na Seção 3.2. Defina a função da soma dos resíduos quadrados para qualquer possível vetor de parâmetros $\boldsymbol{b}$ $k \times 1$ como

$$\text{SQR}(\boldsymbol{b}) \equiv \sum_{t=1}^{n} (y_t - \boldsymbol{x}_t\boldsymbol{b})^2.$$

O vetor $k \times 1$ das estimativas de mínimos quadrados ordinários, $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_k)'$, minimiza SQR($\boldsymbol{b}$) de todos os possíveis vetores $\boldsymbol{b}$ $k \times 1$. Isso é um problema no cálculo multivariado. Para que $\hat{\boldsymbol{\beta}}$ minimize a soma dos resíduos quadrados, ele deve satisfazer a **condição de primeira ordem**

$$\partial\text{SQR}(\hat{\boldsymbol{\beta}})/\partial\boldsymbol{b} \equiv 0. \tag{E.4}$$

Usando o fato que a derivada de $(y_1 - \boldsymbol{x}_t\boldsymbol{b})^2$ em relação a $\boldsymbol{b}$ é o vetor $1 \times k$ $-2(y_t - \boldsymbol{x}_t\boldsymbol{b})\boldsymbol{x}_t$, (E.4) é equivalente a

$$\sum_{t=1}^{n} \boldsymbol{x}_t'(y_t - \boldsymbol{x}_t\hat{\boldsymbol{\beta}}) \equiv \mathbf{0}. \tag{E.5}$$

(Dividimos por -2 e tomamos a transposta.) Podemos escrever essa condição de primeira ordem como

$$\begin{aligned}
&\sum_{t=1}^{n} (y_t - \hat{\beta}_1 - \hat{\beta}_2 x_{t2} - ... - \hat{\beta}_k x_{tk}) = 0 \\
&\sum_{t=1}^{n} x_{t2}(y_t - \hat{\beta}_1 - \hat{\beta}_2 x_{t2} - ... - \hat{\beta}_k x_{tk}) = 0 \\
&\vdots \\
&\sum_{t=1}^{n} x_{tk}(y_t - \hat{\beta}_1 - \hat{\beta}_2 x_{t2} - ... - \hat{\beta}_k x_{tk}) = 0,
\end{aligned}$$

que, exceto pela nova convenção notacional, é idêntica às condições de primeira ordem da equação (3.13). Queremos escrevê-las na forma matricial para torná-las mais úteis. Usando a fórmula da multiplicação particionada do Apêndice D, vemos que (E.5) é equivalente a

$$\mathbf{X}'(\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{0} \tag{E.6}$$

ou

$$(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}'\boldsymbol{y}. \tag{E.7}$$

É possível mostrar que (E.7) sempre tem pelo menos uma solução. Soluções múltiplas não nos ajudam, já que estamos querendo um conjunto único de estimativas MQO, dado nosso conjunto de dados. Assumindo que a matriz simétrica $k \times k$ $\mathbf{X}'\mathbf{X}$ é não singular, podemos premultiplicar ambos os lados de (E.7) por $(\mathbf{X}'\mathbf{X})^{-1}$ para encontrar o estimador MQO $\hat{\boldsymbol{\beta}}$:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \tag{E.8}$$

Essa é a fórmula básica para a análise matricial do modelo de regressão linear múltipla. A hipótese de que $\mathbf{X}'\mathbf{X}$ é inversível é equivalente à hipótese de que posto($\mathbf{X}$) $= k$, o que significa que as colunas de $\mathbf{X}$ devem ser linearmente independentes. Essa é a versão matricial do RLM.4 do Capítulo 3.

Antes de continuar, precisamos fazer um alerta sobre (E.8). É tentador simplificar a fórmula de $\hat{\boldsymbol{\beta}}$ como segue:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\boldsymbol{y} = \mathbf{X}^{-1}(\mathbf{X}')^{-1}\mathbf{X}'\boldsymbol{y} = \mathbf{X}^{-1}\boldsymbol{y}.$$

A falha desse raciocínio é que em geral $\mathbf{X}$ não é uma matriz quadrada e, portanto, ela não poderá ser invertida. Em outras palavras, não podemos escrever $(\mathbf{X}'\mathbf{X})^{-1} = \mathbf{X}^{-1}(\mathbf{X}')^{-1}$ a menos que $n = k$, um caso que na prática virtualmente nunca aparece.

Os vetores $n \times 1$ dos valores ajustados dos MQO e os resíduos são dados por

$$\hat{\boldsymbol{y}} = \mathbf{X}\hat{\boldsymbol{\beta}},\ \hat{\boldsymbol{u}} = \boldsymbol{y} - \hat{\boldsymbol{y}} = \boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{\beta}}.$$

A partir de (E.6) e da definição de $\hat{\boldsymbol{u}}$, podemos ver que a condição de primeira ordem de $\hat{\boldsymbol{\beta}}$ é o mesmo que

$$\mathbf{X}'\hat{\boldsymbol{u}} = \mathbf{0}. \tag{E.9}$$

Como a primeira coluna de $\mathbf{X}$ consiste inteiramente de unidades, (E.9) implica que os resíduos MQO sempre somarão zero quando um intercepto for incluído na equação e que a covariância amostral entre cada variável independente e os resíduos MQO será zero. (Examinamos essas duas propriedades no Capítulo 3.)

A soma dos resíduos quadrados pode ser escrita como

$$\mathrm{SQR} = \sum_{t=1}^{n} \hat{u}_t^2 = \hat{\boldsymbol{u}}'\hat{\boldsymbol{u}} = (\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\,(\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \tag{E.10}$$

Todas as propriedades algébricas do Capítulo 3 podem ser derivadas usando álgebra matricial. Por exemplo, podemos mostrar que a soma dos quadrados total é igual à soma dos quadrados explicada mais a soma dos resíduos quadrados [veja (3.27)]. O uso de matrizes não oferece uma prova mais simples que a notação somatória, de modo que não apresentamos outra derivação.

O método matricial da regressão múltipla pode ser usado como a base para uma interpretação geométrica da regressão. Isso envolve conceitos matemáticos ainda mais avançados dos que os abordados no Apêndice D [Veja Goldberger (1991) ou Greene (1997)].

E.2 PROPRIEDADES DE AMOSTRA FINITA DO MQO

A derivação do valor esperado e da variância do estimador MQO $\hat{\boldsymbol{\beta}}$ é facilitada pela álgebra matricial, mas devemos ter algum cuidado na apresentação das hipóteses.

HIPÓTESE E.1 (LINEAR NOS PARÂMETROS)
O modelo poderá ser escrito como em (E.3), quando $\boldsymbol{y}$ for um vetor observado $n \times 1$, $\mathbf{X}$ for uma matriz observada $n \times k$, e $\boldsymbol{u}$ for um vetor $n \times 1$ de erros ou distúrbios não observados.

HIPÓTESE E.2 (MÉDIA CONDICIONAL ZERO)
Condicional em relação à totalidade da matriz $\mathbf{X}$, cada erro u_t terá média zero: $\mathrm{E}(u_t|\mathbf{X}) = 0$, $t = 1, 2, ..., n$. Na forma vetorial,

$$\mathrm{E}(\boldsymbol{u}|\mathbf{X}) = \mathbf{0}. \quad \text{(E.11)}$$

Essa hipótese é sugerida por RLM.3 sob a hipótese de amostragem aleatória, RLM.2. Em aplicações de séries temporais, a hipótese E.2 impõe exogeneidade estrita nas variáveis explicativas, algo que explicamos em profundidade no Capítulo 10. Ela exclui variáveis explicativas cujos valores futuros sejam correlacionados com u_t; em particular, ela elimina variáveis dependentes defasadas. Sob a hipótese E.2, podemos condicionar a x_{tj} quando calculamos o valor esperado de $\hat{\boldsymbol{\beta}}$.

HIPÓTESE E.3 (INEXISTÊNCIA DE COLINEARIDADE PERFEITA)
A matriz $\mathbf{X}$ tem posto k.

Essa é uma declaração cuidadosa da hipótese que elimina dependências lineares entre as variáveis explicativas. Sob a Hipótese E.3, $\mathbf{X}'\mathbf{X}$ é não singular e, assim, $\hat{\boldsymbol{\beta}}$ será única e poderá ser escrita como em (E.8).

TEOREMA E.1 (INEXISTÊNCIA DE VIÉS DO MQO)
Sob as Hipóteses E.1, E.2 e E.3, o estimador MQO $\hat{\boldsymbol{\beta}}$ será não-viesado para $\boldsymbol{\beta}$.

PROVA: Use as Hipóteses E.1 e E.3 e álgebra simples para escrever

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{u}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{u} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{u},\end{aligned} \quad \text{(E.12)}$$

(Continua...)

(... continuação)

onde usamos o fato de que $(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X}) = \mathbf{I}_k$. Tomando a expectativa condicional em relação a $\mathbf{X}$, teremos

$$\begin{aligned} \mathrm{E}(\hat{\boldsymbol{\beta}}|\mathbf{X}) &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathrm{E}(\boldsymbol{u}|\mathbf{X}) \\ &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{0} = \boldsymbol{\beta}, \end{aligned}$$

porque $\mathrm{E}(\boldsymbol{u}|\mathbf{X}) = \mathbf{0}$ sob a Hipótese E.2. Esse argumento claramente não depende do valor de $\boldsymbol{\beta}$, de modo que mostramos que $\hat{\boldsymbol{\beta}}$ é não-viesado.

Para obter a forma mais simples da matriz de variâncias-covariâncias de $\hat{\boldsymbol{\beta}}$, impomos as hipóteses de homoscedasticidade e correlação não-serial.

HIPÓTESE E.4 (HOMOSCEDASTICIDADE E AUSÊNCIA DE CORRELAÇÃO SERIAL)
(i) $\mathrm{Var}(u_t|\mathbf{X}) = \sigma^2$, $t = 1, 2, ..., n$. (ii) $\mathrm{Cov}(u_t, u_s|\mathbf{X}) = 0$, para todo $t \neq s$. Em forma matricial, podemos escrever essas duas hipóteses como

$$\mathrm{Var}(\boldsymbol{u}|\mathbf{X}) = \sigma^2\mathbf{I}_n, \tag{E.13}$$

onde $\mathbf{I}_n$ é a matriz identidade $n \times n$.

A Parte (i) da Hipótese E.4 é a hipótese de homoscedasticidade: a variância de u_t não pode depender de qualquer elemento de **X**, e a variância deve ser constante ao longo das observações, *t*. A Parte (ii) é a hipótese de correlação não-serial: os erros não podem ser correlacionados ao longo das observações. Sob amostragem aleatória e em qualquer outro esquema de amostragem de corte transversal com observações independentes, a parte (ii) da Hipótese E.4 é mantida automaticamente. Em aplicações de séries temporais, a parte(ii) elimina a correlação nos erros ao longo do tempo (tanto condicional em relação a **X** como incondicionalmente).

Por causa de (E.13), freqüentemente dizemos que $\boldsymbol{u}$ tem **matriz de variâncias-covariâncias escalar** quando a hipótese E.4 for válida. Agora podemos derivar a **matriz de variâncias-covariâncias do estimador MQO**.

TEOREMA E.2 (MATRIZ DE VARIÂNCIA-COVARIÂNCIA DO ESTIMADOR MQO)
Sob as Hipóteses E.1 a E.4,

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}. \tag{E.14}$$

PROVA: Da última fórmula na equação (E.12), temos

(Continua...)

(... continuação)

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) = \mathrm{Var}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{u}|\mathbf{X}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[\mathrm{Var}(\boldsymbol{u}|\mathbf{X})]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}.$$

Agora, usamos a Hipótese E.4 para obter

$$\begin{aligned}\mathrm{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\sigma^2\mathbf{I}_n)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}.\end{aligned}$$

A fórmula (E.14) significa que a variância de $\hat{\beta}_j$ (condicional em relação a $\mathbf{X}$) é obtida pela multiplicação de σ^2 pelo $j^{\text{-ésimo}}$ elemento de $(\mathbf{X}'\mathbf{X})^{-1}$. Para os coeficientes de inclinação, demos uma fórmula interpretável na equação (3.51). A equação (E.14) também nos informa como obter a covariância entre quaisquer duas estimativas MQO: multiplicar σ^2 pelo elemento fora da diagonal apropriado de $(\mathbf{X}'\mathbf{X})^{-1}$. No Capítulo 4, mostramos como evitar explicitamente, encontrar covariâncias para obter intervalos de confiança e testes de hipóteses pela reformulação apropriada do modelo.

O Teorema de Gauss-Markov, em sua generalidade total, pode ser provado.

TEOREMA E.3 (TEOREMA DE GAUSS-MARKOV)

Sob as hipóteses E.1 a E.4, $\hat{\boldsymbol{\beta}}$ é o melhor estimador linear não-viesado.

PROVA: Qualquer outro estimador linear de $\boldsymbol{\beta}$ pode ser escrito como

$$\tilde{\boldsymbol{\beta}} = \boldsymbol{A}'\boldsymbol{y}, \tag{E.15}$$

onde **A** é uma matriz $n \times k$. Para que $\tilde{\boldsymbol{\beta}}$ seja não-viesado condicional em **X**, **A** pode consistir de números e funções não aleatórias de **X**. (Por exemplo, **A** não poderá ser uma função de ($\boldsymbol{y}$). Para verificar que outras restrições sobre **A** são necessárias, escreva

$$\tilde{\boldsymbol{\beta}} = \mathbf{A}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{u}) = (\mathbf{A}'\mathbf{X})\boldsymbol{\beta} + \mathbf{A}'\mathbf{u}. \tag{E.16}$$

Então,

$$\begin{aligned}\mathrm{E}(\tilde{\boldsymbol{\beta}}|\mathbf{X}) &= \mathbf{A}'\mathbf{X}\boldsymbol{\beta} + \mathrm{E}(\mathbf{A}'\boldsymbol{u}|\mathbf{X}) \\ &= \mathbf{A}'\mathbf{X}\boldsymbol{\beta} + \mathbf{A}'\mathrm{E}(\boldsymbol{u}|\mathbf{X}), \text{ pois } \mathbf{A} \text{ é uma função de } \mathbf{X} \\ &= \mathbf{A}'\mathbf{X}\boldsymbol{\beta}, \text{ pois } \mathrm{E}(\boldsymbol{u}|\mathbf{X}) = \mathbf{0}.\end{aligned}$$

Para que $\tilde{\boldsymbol{\beta}}$ seja um estimador não-viesado de $\boldsymbol{\beta}$, dever ser verdade que $\mathrm{E}(\tilde{\boldsymbol{\beta}}|\mathbf{X}) = \boldsymbol{\beta}$ para todos os vetores $k \times 1$ $\boldsymbol{\beta}$, isto é,

$$\mathbf{A}'\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta} \text{ para todos os vetores } k \times 1\ \boldsymbol{\beta}. \tag{E.17}$$

(Continua...)

(... continuação)

Como $\mathbf{A}'\mathbf{X}$ é uma matriz $k \times k$, (E.17) será válida se, e somente se, $\mathbf{A}'\mathbf{X} = \mathbf{I}_k$. As equações (E.15) e (E.17) caracterizam a classe de estimadores lineares, não-viesados de $\boldsymbol{\beta}$.

Em seguida, a partir de (E.16), temos

$$\mathrm{Var}(\tilde{\boldsymbol{\beta}}|\mathbf{X}) = \mathbf{A}'[\mathrm{Var}(\boldsymbol{u}|\mathbf{X})]\mathbf{A} = \sigma^2\mathbf{A}'\mathbf{A},$$

pela Hipótese E.4. Portanto,

$$\begin{aligned}\mathrm{Var}(\tilde{\boldsymbol{\beta}}|\mathbf{X}) - \mathrm{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) &= [\sigma^2\mathbf{A}'\mathbf{A} - (\mathbf{X}'\mathbf{X})^{-1}] \\ &= \sigma^2[\mathbf{A}'\mathbf{A} - \mathbf{A}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}], \text{ pois } \mathbf{A}'\mathbf{X} = \mathbf{I}_k \\ &= \sigma^2\mathbf{A}'[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{A} \\ &= \sigma^2\mathbf{A}'\mathbf{M}\mathbf{A},\end{aligned}$$

onde $\mathbf{M} \equiv \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Como $\mathbf{M}$ é simétrica e idempotente, $\mathbf{A}'\mathbf{M}\mathbf{A}$ é positiva definida para qualquer matriz $\mathbf{A}$ $n \times k$. Isso estabelece que o estimador MQO $\hat{\boldsymbol{\beta}}$ é o melhor estimador linear não-viesado. O quanto ele é significante? Seja $\boldsymbol{c}$ qualquer vetor $k \times 1$ e considere a combinação linear $\boldsymbol{c}'\boldsymbol{\beta} = c_1\beta_1 + c_2\beta_2 + ... + c_k\beta_k$, que é um escalar. Os estimadores não-viesados de $\mathbf{c}'\boldsymbol{\beta}$ são $\mathbf{c}'\hat{\boldsymbol{\beta}}$ e $\boldsymbol{c}'\tilde{\boldsymbol{\beta}}$. Entretanto,

$$\mathrm{Var}(\boldsymbol{c}'\tilde{\boldsymbol{\beta}}|\mathbf{X}) - \mathrm{Var}(\boldsymbol{c}'\hat{\boldsymbol{\beta}}|\mathbf{X}) = \boldsymbol{c}'[\mathrm{Var}(\tilde{\boldsymbol{\beta}}|\mathbf{X}) - \mathrm{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X})]\boldsymbol{c} \geq 0,$$

pois $[\mathrm{Var}(\tilde{\boldsymbol{\beta}}|\mathbf{X}) - \mathrm{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X})]$ é p.s.d. Portanto, quando usado para estimar qualquer combinação de $\boldsymbol{\beta}$, os MQO produzem a menor variância. Particularmente, $\mathrm{Var}(\hat{\beta}_j|\mathbf{X}) \leq \mathrm{Var}(\tilde{\beta}_j|\mathbf{X})$ para qualquer outro estimador linear, não-viesado de β_j.

O estimador não-viesado da variância do erro σ^2 pode ser escrito como

$$\hat{\sigma}^2 = \hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}/(n - k),$$

onde neste apêndice denominamos as variáveis explicativas de forma que há um total de k parâmetros, incluindo o intercepto.

TEOREMA E.4 (INEXISTÊNCIA DE VIÉS EM $\hat{\sigma}^2$)

Sob as hipóteses E.1 a E.4, $\hat{\sigma}^2$ é não-viesado para σ^2: $\mathrm{E}(\hat{\sigma}^2|\mathbf{X}) = \sigma^2$ para todo $\sigma^2 > 0$.

PROVA: Escreva $\hat{\mathbf{u}} = \boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \boldsymbol{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{M}\boldsymbol{y} = \mathbf{M}\boldsymbol{u}$, onde $\mathbf{M} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, e a última igualdade segue-se automaticamente, pois $\mathbf{M}\mathbf{X} = \mathbf{0}$. Como $\mathbf{M}$ é simétrica e idempotente,

$$\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}} = \boldsymbol{u}'\mathbf{M}'\mathbf{M}\boldsymbol{u} = \boldsymbol{u}'\mathbf{M}\boldsymbol{u}.$$

Como $\boldsymbol{u}'\mathbf{M}\boldsymbol{u}$ é uma escalar, ele é igual a seu traço. Portanto,

(Continua...)

(... continuação)

$$
\begin{aligned}
\mathrm{E}(\boldsymbol{u}'\mathbf{M}\boldsymbol{u}|\mathbf{X}) &= \mathrm{E}[\mathrm{tr}(\boldsymbol{u}'\mathbf{M}\boldsymbol{u})|\mathbf{X}] = \mathrm{E}[\mathrm{tr}(\mathbf{M}\boldsymbol{u}\boldsymbol{u}')|\mathbf{X}] \\
&= \mathrm{tr}[\mathrm{E}(\mathbf{M}\boldsymbol{u}\boldsymbol{u}'|\mathbf{X}] = \mathrm{tr}[\mathbf{M}\mathrm{E}(\boldsymbol{u}\boldsymbol{u}'|\mathbf{X})] \\
&= \mathrm{tr}(\mathbf{M}\sigma^2\mathbf{I}_n) = \sigma^2\mathrm{tr}(\mathbf{M}) = \sigma^2(n-k).
\end{aligned}
$$

A última igualdade resulta de $\mathrm{tr}(\mathbf{M}) = \mathrm{tr}(\mathbf{I}_n) - \mathrm{tr}[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] = n - \mathrm{tr}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}] = n - \mathrm{tr}(\mathbf{I}_k) = n - k$. Portanto,

$$
\mathrm{E}(\sigma^2|\mathbf{X}) = \mathrm{E}(\boldsymbol{u}'\mathbf{M}\boldsymbol{u}|\mathbf{X})/(n-k) = \sigma^2.
$$

E.3 INFERÊNCIA ESTATÍSTICA

Quando adicionamos a hipótese final do modelo linear clássico, $\hat{\boldsymbol{\beta}}$ terá uma distribuição normal multivariada, que leva as distribuições t e F para os testes estatísticos padrões tratados no Capítulo 4.

HIPÓTESE E.5 (NORMALIDADE DOS ERROS)

Condicional em relação a $\mathbf{X}$, os u_t serão independente e identicamente distribuídos como Normal$(0,\sigma^2)$. De forma equivalente, $\boldsymbol{u}$ dado $\mathbf{X}$ terá distribuição normal multivariada com média zero e matriz de variâncias-covariâncias $\sigma^2\mathbf{I}_n$: $\boldsymbol{u} \sim \mathrm{Normal}(\mathbf{0},\sigma^2\mathbf{I}_n)$.

Sob a hipótese E.5, cada u_t será independente das variáveis explicativas para todas as t. Em um cenário de séries temporais, essa é basicamente a hipótese de exogeneidade estrita.

TEOREMA E.5 (NORMALIDADE DE $\hat{\beta}$)

Sob as hipóteses E.1 a E.5 do modelo linear clássico, $\hat{\boldsymbol{\beta}}$ condicional em relação a $\mathbf{X}$ terá distribuição normal multivariada com média $\boldsymbol{\beta}$ e matriz de variâncias-covariâncias $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.

O teorema E.5 é a base da inferência estatística envolvendo $\boldsymbol{\beta}$. De fato, com as propriedades das distribuições t, F e qui-quadrado que resumimos no Apêndice D, podemos usar o Teorema E.5 para estabelecer que as estatísticas t têm uma distribuição t sob as Hipóteses E.1 a E.5 (sob as hipóteses nulas) e igualmente para as estatísticas F. Ilustramos a questão com uma prova para a estatística t.

TEOREMA E.6

Sob as hipóteses E.1 até a E.5,

$$(\hat{\beta}_j - \beta_j)/\text{ep}(\hat{\beta}_j) \sim t_{n-k}, j = 1, 2, ..., k.$$

PROVA: A prova exige várias etapas; as seguintes afirmações são, inicialmente, condicionais em relação a **X**. Primeira, pelo Teorema E.5, $(\hat{\beta}_j - \beta_j)/\text{dp}(\hat{\beta}) \sim \text{Normal}(0,1)$, onde $\text{dp}(\hat{\beta}_j) = \sigma\sqrt{c_{jj}}$, e c_{jj} é o $j^{\text{ésimo}}$ elemento de $(\mathbf{X}'\mathbf{X})^{-1}$. Em seguida, sob as hipóteses E.1 a E.5, condicional em relação a **X**,

$$(n - k)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{n-k}. \quad \textbf{(E.18)}$$

Isso é assim porque $(n - k)\hat{\sigma}^2/\sigma^2 = (\boldsymbol{u}/\sigma)'\mathbf{M}(\boldsymbol{u}/\sigma)$, onde **M** é a matriz simétrica, idempotente $n \times n$ definida no Teorema E.4. Mas $\boldsymbol{u}/\sigma \sim \text{Normal}(\mathbf{0},\mathbf{I}_n)$ pela hipótese E.5. Resulta, pela Propriedade 1 da distribuição qui-quadrado no Apêndice D, que $(\boldsymbol{u}/\sigma)'\mathbf{M}(\boldsymbol{u}/\sigma) \sim \chi^2{}_{n-k}$ (pois **M** tem posto $n - k$).

Também precisamos mostrar que $\hat{\boldsymbol{\beta}}$ e $\hat{\sigma}^2$ são independentes. Mas $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{u}$, e $\hat{\sigma}^2 = \boldsymbol{u}'\mathbf{M}\boldsymbol{u}/(n - k)$. Agora, $[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{M} = \mathbf{0}$, pois $\mathbf{X}'\mathbf{M} = \mathbf{0}$. Resulta, pela Propriedade 5 da distribuição normal multivariada do Apêndice D, que $\hat{\boldsymbol{\beta}}$ e $\mathbf{M}\boldsymbol{u}$ são independentes. Como $\hat{\sigma}^2$ é uma função de $\mathbf{M}\boldsymbol{u}$, $\hat{\boldsymbol{\beta}}$ e $\hat{\sigma}^2$ também são independentes.

Finalmente, podemos escrever

$$(\hat{\beta}_j - \beta_j)/\text{ep}(\hat{\beta}_j) = [(\hat{\beta}_j - \beta_j)/\text{dp}(\hat{\beta}_j)]/(\hat{\sigma}^2/\sigma^2)^{1/2},$$

que é a razão entre uma variável aleatória normal padrão e a raiz quadrada de uma variável aleatória $\chi^2{}_{n-k}/(n - k)$. Acabamos de mostrar que elas são independentes e, assim, pela definição de uma variável aleatória t, $(\hat{\beta}_j - \beta_j)/\text{ep}(\hat{\beta}_j)$ terá distribuição t_{n-k}. Como essa distribuição não depende de **X**, ela também é a distribuição incondicional de $(\hat{\beta}_j - \beta_j)/\text{ep}(\hat{\beta}_j)$.

A partir desse teorema, podemos agregar qualquer valor hipotético de β_j e usar a estatística t para testar hipóteses, como sempre.

Sob as hipóteses E.1 a E.5, podemos calcular o que é conhecido como o limite inferior de *Cramer-Rao* para a matriz de variâncias-covariâncias dos estimadores não-viesados de $\boldsymbol{\beta}$ (novamente condicional em **X**) [veja Greene (1997, Capítulo 4)]. É possível mostrar que isso é $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$, que é exatamente a matriz de variâncias-covariâncias dos estimadores MQO. Isso implica que $\hat{\boldsymbol{\beta}}$ é o estimador de **variância mínima não-viesado** de $\boldsymbol{\beta}$ (condicional em **X**): $\text{Var}(\tilde{\boldsymbol{\beta}}|\mathbf{X}) - \text{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X})$ é positiva semidefinida para qualquer outro estimador não-viesado $\tilde{\boldsymbol{\beta}}$; não mais precisamos restringir nossa atenção sobre os estimadores lineares em $\boldsymbol{y}$.

É fácil mostrar que o estimador MQO é na verdade o estimador de máxima verossimilhança de $\boldsymbol{\beta}$ sob a hipótese E.5. Para cada t, a distribuição de y_t dado **X** é $\text{Normal}(\boldsymbol{x}_t\boldsymbol{\beta},\sigma^2)$. Como os y_t são independentes condicionais em **X**, a função de verossimilhança para a amostra é obtida do produto das densidades:

$$\prod_{t=1}^{n}(2\pi\sigma^2)^{-1/2}\exp[-(y_t - \boldsymbol{x}_t\boldsymbol{\beta})^2/(2\sigma^2)],$$

onde $\prod$ representa o produto. Maximizar essa função em relação a $\boldsymbol{\beta}$ e σ^2 é o mesmo que maximizar seu logaritmo natural:

$$\sum_{t=1}^{n}[-(1/2)\log(2\pi\sigma^2) - (y_t - \boldsymbol{x}_t\boldsymbol{\beta})^2/(2\sigma^2)].$$

Para a obtenção de $\hat{\boldsymbol{\beta}}$, isso é o mesmo que minimizar $\sum_{t=1}^{n}(y_t - \boldsymbol{x}_t\boldsymbol{\beta})^2$ — a divisão por $2\sigma^2$ não afeta a otimização —, que é simplesmente o problema que o MQO soluciona. O estimador de σ^2 que temos usado, SQR/$(n - k)$, revela-se como não sendo o EMV de σ^2; o EMV é SQR/n, que é um estimador viesado. Como o estimador não-viesado de σ^2 resulta nas estatísticas t e F com distribuições t e F exatas sob a hipótese nula, ele sempre é usado em lugar do EMV.

E.4 UM POUCO SOBRE ANÁLISE ASSIMPTÓTICA

O método matricial do modelo de regressão múltipla também pode tornar as derivações das propriedades assimptóticas mais concisas. De fato, podemos apresentar provas generalizadas das afirmações do Capítulo 11, usando as mesmas notações e hipóteses deste capítulo, com a pequena exceção que fazemos k representar o número total de parâmetros no modelo, como viemos fazendo ao longo de todo este apêndice.

Começamos provando o resultado da consistência do Teorema 11.1. Recorde que essas hipóteses contêm, como um caso especial, as hipóteses da análise de corte transversal sob a amostragem aleatória.

PROVA DO TEOREMA 11.1: Como no problema E.1 e usando a hipótese ST.1′, escrevemos o estimador MQO como

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \left(\sum_{t=1}^{n} x_t'\boldsymbol{x}_t\right)^{-1}\left(\sum_{t=1}^{n}\boldsymbol{x}_t' y_t\right) = \left(\sum_{t=1}^{n}\boldsymbol{x}_t'\boldsymbol{x}_t\right)^{-1}\left(\sum_{t=1}^{n}\boldsymbol{x}_t'(\boldsymbol{x}_t\boldsymbol{\beta} + u_t\right) \\ &= \boldsymbol{\beta} + \left(\sum_{t=1}^{n}\boldsymbol{x}_t'\boldsymbol{x}_t\right)^{-1}\left(\sum_{t=1}^{n}\boldsymbol{x}_t' u_t\right) \\ &= \boldsymbol{\beta} + \left(n^{-1}\sum_{t=1}^{n}\boldsymbol{x}'_t\boldsymbol{x}_t\right)^{-1}\left(n^{-1}\sum_{t=1}^{n}\boldsymbol{x}_t' u_t\right).\end{aligned} \tag{E.19}$$

Agora, de conformidade com a lei dos grandes números,

$$n^{-1}\sum_{t=1}^{n}\boldsymbol{x}_t'\boldsymbol{x}_t \xrightarrow{p} \mathbf{A} \text{ e } n^{-1}\sum_{t=1}^{n}\boldsymbol{x}_t' u_t \xrightarrow{p} \mathbf{0}, \tag{E.20}$$

onde $\mathbf{A} = \mathrm{E}(\boldsymbol{x}_t'\boldsymbol{x}_t)$ é a matriz não singular $k \times k$ sob a hipótese ST.3′ e onde usamos o fato de que $\mathrm{E}(\boldsymbol{x}_t' u_t) = \mathbf{0}$ sob a hipótese ST.2′. Agora, devemos usar uma versão matricial da Propriedade PLIM.1 do Apêndice C. Ou seja, como $\mathbf{A}$ é não singular,

$$\left(\sum_{t=1}^{n} \boldsymbol{x}'_t \boldsymbol{x}_t\right)^{-1} \xrightarrow{p} \mathbf{A}^{-1}. \qquad \textbf{(E.21)}$$

[Wooldridge (2002, Capítulo 3) contém uma discussão sobre esses tipos de resultados de convergência]. Agora segue de (E.19), (E.20) e (E.21) que

$$\text{plim}\,(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} + \mathbf{A}^{-1} \cdot \mathbf{0} = \boldsymbol{\beta}.$$

Isso completa a prova. □

A seguir, delineamos uma prova do resultado da normalidade assimptótica do Teorema 11.2.

PROVA DO TEOREMA 11.2: Pela equação (E.19), podemos escrever

$$\begin{aligned}\sqrt{n}\,(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \left(n^{-1}\sum_{t=1}^{n} \boldsymbol{x}'_t \boldsymbol{x}_t\right)^{-1} \left(n^{-1/2}\sum_{t=1}^{n} \boldsymbol{x}'_t u_t\right) \\ &= \mathbf{A}^{-1}\left(n^{-1/2}\sum_{t=1}^{n} \boldsymbol{x}'_t u_t\right) + o_p(1),\end{aligned} \qquad \textbf{(E.22)}$$

onde o termo "$o_p(1)$" é o termo residual que converge em probabilidade para zero. Esse termo é igual a $\left[\left(n^{-1}\sum_{t=1}^{n} \boldsymbol{x}'_t \boldsymbol{x}_t\right)^{-1} - \mathbf{A}^{-1}\right]\left(n^{-1/2}\sum_{t=1}^{n} \boldsymbol{x}'_t u_t\right)$. O termo entre colchetes converge em probabilidade para zero (pelo mesmo argumento usado na prova do Teorema 11.1), enquanto $\left(n^{-1/2}\sum_{t=1}^{n} \boldsymbol{x}'_t u_t\right)$ é limitado em probabilidade porque ele converge para uma distribuição normal multivariada pelo teorema do limite central. Um resultado bastante conhecido na teoria assimptótica é que o produto de tais termos converge em probabilidade para zero. Além disso, $\sqrt{n}\,(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ herda sua distribuição assimptótica de $\mathbf{A}^{-1}\left(n^{-1/2}\sum_{t=1}^{n} \boldsymbol{x}'_t u_t\right)$. Veja Wooldridge (2002, Capítulo 3) para mais detalhes sobre os resultados de convergência usados nesta prova.

De acordo com o teorema do limite central, $n^{-1/2}\sum_{t=1}^{n} \boldsymbol{x}'_t u_t$ tem uma distribuição normal assimptótica com média zero e, digamos, matriz de variâncias-covariâncias $k \times k$ $\mathbf{B}$. Então, $\sqrt{n}\,(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ tem uma distribuição normal multivariada assimptótica com média zero e matriz de variâncias-covariâncias $\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$. Agora mostramos que, sob as hipóteses ST.4′ e ST.5′, $\mathbf{B} = \sigma^2\mathbf{A}$. (A expressão geral é útil, pois ela sustenta os erros-padrão robustos em relação à heteroscedasticidade e à correlação serial do MQO, de forma semelhante à examinada no Capítulo 12.) Primeiro, sob a hipótese ST.5′, $\boldsymbol{x}'_t u_t$ e $\boldsymbol{x}'_s u_s$ são não-correlacionados quando $t \neq s$. Por quê? Concretamente, suponha que $s < t$. Então, pela lei das expectativas iteradas, $\mathrm{E}(\boldsymbol{x}'_t u_t u_s \boldsymbol{x}_s) = \mathrm{E}[\mathrm{E}(u_t u_s \boldsymbol{x}'_t \boldsymbol{x}_s) | \boldsymbol{x}'_t \boldsymbol{x}_s)] = \mathrm{E}[\mathrm{E}(u_t u_s | \boldsymbol{x}'_t \boldsymbol{x}_s) \boldsymbol{x}'_t \boldsymbol{x}_s)] = \mathrm{E}[0 \cdot \boldsymbol{x}'_t \boldsymbol{x}_s] = \mathbf{0}$. As covariâncias zero implicam que a variância da soma é a soma das variâncias. Entretanto, $\mathrm{Var}(\boldsymbol{x}'_t u_t) = \mathrm{E}(\boldsymbol{x}'_t u_t u_t \boldsymbol{x}_t) = \mathrm{E}(u_t^2 \boldsymbol{x}'_t \boldsymbol{x}_t)$. Pela lei das expectativas iteradas, $\mathrm{E}(u_t^2 \boldsymbol{x}'_t \boldsymbol{x}_t) = \mathrm{E}[\mathrm{E}(u_t^2 \boldsymbol{x}'_t \boldsymbol{x}_t | \boldsymbol{x}_t)] =$

$\mathrm{E}[\mathrm{E}(u_t^2 | \boldsymbol{x}_t)\boldsymbol{x}'_t\boldsymbol{x}_t] = \mathrm{E}(\sigma^2\boldsymbol{x}'_t\boldsymbol{x}_t) = \sigma^2\mathrm{E}(\boldsymbol{x}'_t\boldsymbol{x}_t) = \sigma^2\mathbf{A}$, onde usamos $\mathrm{E}(u_t^2|\boldsymbol{x}_t) = \sigma^2$ sob as hipóteses ST.2′ e ST.4′. Isso mostra que $\mathbf{B} = \sigma^2\mathbf{A}$ e, assim, sob as hipóteses ST.1′ a ST.5′, temos

$$\sqrt{n}\,(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \overset{a}{\sim} \text{Normal}\,(\mathbf{0},\sigma^2\mathbf{A}^{-1}). \qquad \textbf{(E.23)}$$

Isso completa a prova. □

A partir da equação (E.23), tratamos $\hat{\boldsymbol{\beta}}$ como se fosse, de maneira aproximada, normalmente distribuído com média $\boldsymbol{\beta}$ e matriz de variâncias-covariâncias $\sigma^2\mathbf{A}^{-1}/n$. A divisão pelo tamanho da amostra, n, nesse caso é esperada: a aproximação à matriz de variâncias-covariâncias de $\hat{\boldsymbol{\beta}}$ se contrai para zero à taxa $1/n$. Quando substituímos σ^2 por seu estimador consistente, $\hat{\sigma}^2 = \mathrm{SQR}/(n - k)$, e substituímos $\mathbf{A}$ por seu estimador consistente $n^{-1}\sum_{t=1}^{n} \boldsymbol{x}'_t\boldsymbol{x}_t = \mathbf{X}'\mathbf{X}/n$, obtemos um estimador para a variância assimptótica de $\hat{\boldsymbol{\beta}}$:

$$\mathrm{A}\hat{v}\mathrm{ar}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}. \qquad \textbf{(E.24)}$$

Observe como as duas divisões por n se anulam, e o lado direito de (E.24) é a maneira habitual como estimamos a matriz de variâncias do estimador MQO sob as hipóteses de Gauss-Markov. Para resumir, mostramos que, sob as hipóteses ST.1′ a ST.5′ — que contêm RLM.1 a RLM.5 como casos especiais —, os erros-padrão e as estatísticas t habituais são assimptoticamente válidas. É perfeitamente legítimo usar a distribuição t habitual para obter valores críticos e p-valores para se testar uma hipótese única. Interessantemente, na organização geral do Capítulo 11, assumir a normalidade dos erros — digamos u_t dados $\mathbf{x}_t, u_{t-1}, \boldsymbol{x}_{t-1}, ..., u_1, \boldsymbol{x}_1$ é distribuída como Normal$(0,\sigma^2)$ — não ajuda necessariamente, pois as estatísticas t não serão de forma geral estatísticas t exatas sob esse tipo de hipótese de normalidade. Quando não assumimos a exogeneidade estrita das variáveis explicativas, resultados distribucionais exatos são difíceis, se não impossíveis, de serem obtidos.

Se modificarmos o argumento acima, poderemos derivar uma matriz de variâncias-covariâncias robustas em relação à heteroscedasticidade. O importante é que deveremos estimar $\mathrm{E}(u_t^2\mathbf{x}'_t\mathbf{x}_t)$ separadamente, pois essa matriz não mais iguala $\sigma^2\mathrm{E}(\boldsymbol{x}'_t\boldsymbol{x}_t)$. Entretanto, se os $\hat{\boldsymbol{u}}_t$ forem os resíduos MQO, um estimador consistente será

$$(n - k)^{-1} \sum_{t=1}^{n} \hat{u}_t^2\boldsymbol{x}'_t\boldsymbol{x}_t, \qquad \textbf{(E.25)}$$

onde a divisão por $n - k$ em vez de por n é um ajuste dos graus de liberdade que hipoteticamente auxilia nas propriedades de amostra finita do estimador. Quando usamos a expressão na equação (E.25), obtemos

$$\mathrm{A}\hat{v}\mathrm{ar}(\hat{\boldsymbol{\beta}}) = [n/(n - k]\mathbf{X}'\mathbf{X})^{-1}\left(\sum_{t=1}^{n} \hat{u}_t^2\boldsymbol{x}'_t\boldsymbol{x}_t\right)(\mathbf{X}'\mathbf{X})^{-1}. \qquad \textbf{(E.26)}$$

As raízes quadradas dos elementos diagonais dessa matriz são os mesmos erros-padrão robustos em relação à heteroscedasticidade que obtivemos na Seção 8.2, no caso de corte transversal puro. Uma extensão matricial de erros-padrão robustos em relação à correlação serial (e à heteroscedasticidade) que obtivemos na Seção 12.5 também está disponível, mas a matriz que deve substituir (E.25) é complicada devido à correlação serial. Veja, por exemplo, Hamilton (1994, Seção 10.5).

Estatística de Wald para Testes de Múltiplas Hipóteses

Argumentos semelhantes podem ser usados para a obtenção da distribuição assimptótica da **estatística de Wald** para testar múltiplas hipóteses. Seja $\mathbf{R}$ uma matriz $q \times k$, com $q \leq k$. Assuma que as q restrições sobre o vetor $q \times 1$ de parâmetros, $\boldsymbol{\beta}$, possam ser expressas como H_0: $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, onde $\mathbf{r}$ é um vetor $q \times 1$ de constantes conhecidas. Sob as hipóteses ST.1′ a ST.5′, é possível mostrar que, sob H_0,

$$[\sqrt{n}(\mathbf{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})]'(\sigma^2\mathbf{R}\mathbf{A}^{-1}\mathbf{R}')^{-1}[\sqrt{n}(\mathbf{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})] \sim \chi^2_q, \qquad \textbf{(E.27)}$$

onde $\mathbf{A} = \mathrm{E}(\boldsymbol{x}_t'\boldsymbol{x}_t)$, como nas provas dos teoremas 11.1 e 11.2. A intuição por trás da equação (E.25) é simples. Como $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ é aproximadamente distribuída como Normal$(\mathbf{0},\sigma^2\mathbf{A}^{-1})$, $\mathbf{R}[\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})] = \sqrt{n}\mathbf{R}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ será aproximadamente Normal$(\mathbf{0},\sigma^2\mathbf{R}\mathbf{A}^{-1}\mathbf{R}')$, pela Propriedade 3 da distribuição normal multivariada do Apêndice D. Sob H_0, $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ e, assim, $\sqrt{n}(\mathbf{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}) \sim$ Normal$(\mathbf{0},\sigma^2\mathbf{R}\mathbf{A}^{-1}\mathbf{R}')$ sob a H_0. Para obter formalmente o resultado final, precisaremos usar uma versão assimptótica desta propriedade, que pode ser encontrada em Wooldridge (2002, Capítulo 3).

Conhecido o resultado em (E.25), obtemos uma estatística computável, fazendo a substituição de $\mathbf{A}$ e σ^2 por seus estimadores consistentes; isso não alterará a distribuição assimptótica. O resultado será a chamada estatística de Wald, a qual, após o cancelamento dos tamanhos das amostras e o uso de um pouco de álgebra, poderá ser escrita como

$$W = (\mathbf{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathrm{R}']^{-1}(\mathrm{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})/\hat{\sigma}^2. \qquad \textbf{(E.28)}$$

Sob H_0, $W \sim X^2_q$, onde recordamos que q é o número de restrições sendo testadas. Se $\hat{\sigma}^2 = \mathrm{SQR}/(n - k)$, é possível mostrar que W/q será exatamente a estatística F que obtivemos no Capítulo 4 no teste de restrições lineares múltiplas. [Veja, por exemplo, Greene (1997, Capítulo 7).] Portanto, sob as hipóteses do modelo linear clássico, ST.1 a ST.6 do Capítulo 10, W/q terá uma distribuição $F_{q,n-k}$ exata. Sob as hipóteses ST.1′ a ST.5′, somente teremos o resultado assimptótico em (E.26). No entanto, é apropriado, e comum, tratar a estatística F usual como tendo uma distribuição $F_{q,n-k}$ aproximada.

Uma estatística de Wald robusta quanto à heteroscedasticidade de forma desconhecida pode ser obtida com o uso da matriz em (E.26) em lugar de $\hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$, e de forma semelhante para um teste estatístico robusto tanto em relação à heteroscedasticidade como quanto à correlação serial. As versões robustas da estatística dos testes não podem ser computadas via somas dos resíduos quadrados ou dos R-quadrados das regressões restritas e irrestritas.

RESUMO

Este apêndice apresentou uma breve explicação do modelo de regressão linear usando a notação matricial. Este material foi incluído para aulas mais avançadas que usam álgebra matricial, mas não é essencial para a leitura do livro. Na verdade, este apêndice faz prova de alguns dos resultados que, ou afirmamos sem provas, provamos somente em casos especiais, ou provamos por meio de métodos excessivamente complicados. Outros tópicos — tais como, propriedades assimptóticas, estimação de variáveis instrumentais e modelos de dados de painel — poderão ser tratados de forma concisa com o uso de matrizes. Textos avançados sobre econometria, entre os quais incluímos Davidson e MacKinnon (1993), Greene (1997) e Wooldridge (2002), podem ser consultados para detalhes.

PROBLEMAS

E.1 Seja $\boldsymbol{x}_t$ o vetor $1 \times k$ das variáveis explicativas da observação t. Mostre que o estimador MQO $\hat{\boldsymbol{\beta}}$ pode ser escrito como

$$\hat{\boldsymbol{\beta}} = \left(\sum_{t=1}^{n} \boldsymbol{x}'_t \boldsymbol{x}_t \right)^{-1} \left(\sum_{t=1}^{n} \boldsymbol{x}'_t y_t \right).$$

A divisão de cada somatório por n mostrará que $\hat{\boldsymbol{\beta}}$ é uma função de médias amostrais.

E.2 Seja $\hat{\boldsymbol{\beta}}$ o vetor $k \times 1$ das estimativas MQO.

(i) Mostre que para qualquer vetor $\boldsymbol{b}$ $k \times 1$, podemos escrever a soma dos resíduos quadrados como

$$\text{SQR}(\boldsymbol{b}) = \hat{\boldsymbol{u}}'\hat{\boldsymbol{u}} + (\hat{\boldsymbol{\beta}} - \boldsymbol{b})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{b}).$$

{*Sugestão*: Escreva $(\boldsymbol{y} - \mathbf{X}\boldsymbol{b})'(\boldsymbol{y} - \mathbf{X}\boldsymbol{b}) = [\hat{\boldsymbol{u}} + \mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{b})]'[\hat{\boldsymbol{u}} + \mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{b})]$ e use o fato que $\mathbf{X}'\hat{\boldsymbol{u}} = \mathbf{0}$.}

(ii) Explique como a expressão de SQR($\boldsymbol{b}$) na parte (i) prova que $\hat{\boldsymbol{\beta}}$ minimiza de forma única a SQR($\boldsymbol{b}$) para quaisquer possíveis valores de $\boldsymbol{b}$, assumindo que $\mathbf{X}$ tem posto k.

E.3 Seja $\hat{\boldsymbol{\beta}}$ a estimativa pelos MQO da regressão de y sobre $\mathbf{X}$. Seja $\mathbf{A}$ uma matriz não singular $k \times k$ e defina $z_t \equiv \boldsymbol{x}_t\mathbf{A}$, $t = 1, ..., n$. Portanto, z_t é $1 \times k$ e é uma combinação linear não singular de $\boldsymbol{x}_t$. Seja $\mathbf{Z}$ a matriz $n \times k$ com linhas z_t. Seja $\tilde{\boldsymbol{\beta}}$ a estimativa MQO de uma regressão de $\boldsymbol{y}$ sobre $\mathbf{Z}$.

(i) Mostre que $\tilde{\boldsymbol{\beta}} = \mathbf{A}^{-1}\hat{\boldsymbol{\beta}}$.

(ii) Sejam $\hat{y}_t$ os valores ajustados da regressão original e $\tilde{y}_t$ os valores ajustados da regressão de $\boldsymbol{y}$ sobre $\mathbf{Z}$. Mostre que $\tilde{y}_t = \hat{y}_t$ para todo $t = 1, 2, ..., n$. Como se comparam os resíduos das duas regressões?

(iii) Mostre que a matriz de variâncias estimada de $\tilde{\boldsymbol{\beta}}$ é $\hat{\sigma}^2\mathbf{A}^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}^{-1\prime}$, onde $\hat{\sigma}^2$ é a estimativa usual da variância da regressão de $\boldsymbol{y}$ sobre $\mathbf{X}$.

(iv) Sejam $\hat{\beta}_j$ as estimativas MQO da regressão de y_t sobre $1, x_{t2}, ..., x_{tk}$, e $\tilde{\beta}_j$ as estimativas MQO da regressão de y_t sobre $1, a_2x_{t2}, ..., a_kx_{tk}$, onde $a_j \neq 0, j = 2, ..., k$. Use os resultados da parte (i) para encontrar a relação entre $\tilde{\beta}_j$ e $\hat{\beta}_j$.

(v) Assumindo a estrutura da parte (iv), use a parte (iii) para mostrar que ep($\tilde{\beta}_j$) = ep($\hat{\beta}_j$)/$|a_j|$.

(vi) Assumindo a estrutura da parte (iv), mostre que os valores absolutos das estatísticas t de $\tilde{\beta}_j$ e $\hat{\beta}_j$ são idênticos.

E.4 Assuma que o modelo $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{u}$ satisfaz as hipóteses de Gauss-Markov, defina $\mathbf{G}$ como uma matriz não aleatória, não singular $k \times k$, e defina $\boldsymbol{\delta} = \mathbf{G}\boldsymbol{\beta}$, de forma que $\boldsymbol{\delta}$ também seja um vetor $k \times 1$. Seja $\hat{\boldsymbol{\beta}}$ o vetor $k \times 1$ dos estimadores MQO e defina $\hat{\boldsymbol{\delta}} = \mathbf{G}\hat{\boldsymbol{\beta}}$ como o estimador MQO de $\boldsymbol{\delta}$.

(i) Mostre que $\mathrm{E}(\hat{\boldsymbol{\delta}}|\mathbf{X}) = \boldsymbol{\delta}$.

(ii) Encontre $\mathrm{Var}(\hat{\boldsymbol{\delta}}|\mathbf{X})$ em termos de σ^2, $\mathbf{X}$ e $\mathbf{G}$.

(iii) Use o Problema E.3 para verificar que $\hat{\boldsymbol{\delta}}$ e a estimativa apropriada de $\mathrm{Var}(\hat{\boldsymbol{\delta}}|\mathbf{X})$ são obtidas a partir da regressão de $\boldsymbol{y}$ sobre $\mathbf{XG}^{-1}$.

(iv) Agora, seja $\mathbf{c}$ um vetor $k \times 1$ com pelo menos uma entrada diferente de zero. Para maior clareza, assuma que $c_k \neq 0$. Defina $\theta = \boldsymbol{c}'\boldsymbol{\beta}$, de forma que θ seja um escalar. Defina $\delta_j = \boldsymbol{\beta}_j$, $j = 1, 2, ..., k-1$ e $\delta_k = \theta$. Mostre como definir uma matriz $\mathbf{G}$ não singular $k \times k$, de forma que $\boldsymbol{\delta} = \mathbf{G}\boldsymbol{\beta}$. (*Sugestão*: Cada uma das primeiras $k - 1$ linhas de $\mathbf{G}$ deve conter $k - 1$ zeros e um 1. Qual será a última linha?)

(v) Mostre que para a escolha de $\mathbf{G}$ na parte (iv),

$$\mathbf{G}^{-1} = \begin{pmatrix} 1 & 0 & 0\,.\,. & . & 0 \\ 0 & 1 & 0\,.\,. & . & 0 \\ \vdots & & & & \\ 0 & 0 & \ldots & 1 & 0 \\ -c_1/c_k & -c_2/c_k & \ldots & -c_{k-1}/c_k & 1/c_k \end{pmatrix}.$$

Use essa expressão para $\mathbf{G}^{-1}$ e a parte (iii) para concluir que $\hat{\theta}$ e seus erros-padrão são obtidos como o coeficiente de x_{tk}/c_k na regressão de y_t sobre

$$[x_{t1} - (c_1/c_k)x_{tk}],\ [x_{t2} - (c_2/c_k)x_{tk}],\ ...,\ [x_{t,\,k-1} - (c_{k-1}/c_k)x_{tk}],\ x_{tk}/c_k,\ t = 1, ..., n.$$

Essa regressão é exatamente a que foi obtida escrevendo $\boldsymbol{\beta}_k$ em termos de θ e $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, ..., $\boldsymbol{\beta}_{k-1}$, agregando o resultado no modelo original e reorganizando. Portanto, podemos justificar formalmente o truque que usamos em todo o livro para obter o erro-padrão de uma combinação linear de parâmetros.

E.5 Assuma que o modelo $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{u}$ satisfaz as hipóteses de Gauss-Markov e que $\hat{\boldsymbol{\beta}}$ seja o estimador MQO de $\boldsymbol{\beta}$. Seja $\mathbf{Z} = \mathbf{G}(\mathbf{X})$ uma função matricial $n \times k$ de $\mathbf{X}$ e assuma que $\mathbf{Z}'\mathbf{X}$ (uma matriz $k \times k$) é não singular. Defina um novo estimador de $\boldsymbol{\beta}$ por $\tilde{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}$.

(i) Mostre que $\mathrm{E}(\tilde{\boldsymbol{\beta}}|\mathbf{X}) = \boldsymbol{\beta}$, de forma que $\tilde{\boldsymbol{\beta}}$ também é não-viesado condicional em relação a $\mathbf{X}$.

(ii) Encontre $\mathrm{Var}(\tilde{\boldsymbol{\beta}}|\mathbf{X})$. Certifique-se de que ela é uma matriz $k \times k$ simétrica que depende de $\mathbf{Z}$, $\mathbf{X}$ e σ^2.

(iii) Qual estimador você prefere, $\hat{\boldsymbol{\beta}}$ ou $\tilde{\boldsymbol{\beta}}$? Explique.