Como o exemplo anterior sugere, com uma amostra grande raramente vemos discrepâncias importantes entre os resultados dos testes *LM* e *F*. Usaremos a estatística *F* para a maior parte dos problemas porque ela é rotineiramente calculada pela maioria dos programas de regressão. No entanto, você deve estar consciente sobre a estatística *LM* para reconhecê-la, pois ela é usada em trabalhos aplicados.

Um último comentário sobre a estatística *LM*. Assim como com a estatística *F*, devemos estar seguros de usar as mesmas observações nos passos (i) e (ii). Se faltarem dados para algumas das variáveis independentes excluídas sob a hipótese nula, os resíduos do passo (i) devem ser obtidos de uma regressão sobre o conjunto de dados reduzido.

5.3 EFICIÊNCIA ASSIMPTÓTICA DE MQO

Sabemos que, sob as hipóteses de Gauss-Markov, os estimadores de MQO são os melhores estimadores não viesados lineares. MQO também é, sob as hipóteses de Gauss-Markov, **assimptoticamente eficiente** dentro uma classe de estimadores. Um tratamento geral requer álgebra matricial e análise assimptótica avançada. Em primeiro lugar, vamos descrever o resultado para o caso da regressão simples.

No modelo

$$y = \beta_0 + \beta_1 x + u, \tag{5.16}$$

u tem média condicional zero sob RLM.4: $E(u|x) = 0$. Isso dá lugar a uma variedade de estimadores consistentes de β_0 e β_1; como habitual, vamos nos concentrar no parâmetro de inclinação, β_1. Seja $g(x)$ qualquer função de x; por exemplo, $g(x) = x^2$ ou $g(x) = 1/(1 + |x|)$. Então u é não correlacionado com $g(x)$ (veja a Propriedade CE.5 no Apêndice B, disponível no site da Cengage). Seja $z_i = g(x_i)$ para todas as observações i. Então, o estimador

$$\tilde{\beta}_1 = \left(\sum_{i=1}^{n}(z_i - \bar{z})y_i\right) \Big/ \left(\sum_{i=1}^{n}(z_i - \bar{z})x_i\right) \tag{5.17}$$

é consistente para β_1, desde que $g(x)$ e x sejam correlacionados. [Lembre-se: é possível que $g(x)$ e x sejam não correlacionados porque a correlação mensura a dependência *linear*.] Para ver isso, podemos colocar $y_i = \beta_0 + \beta_1 x_i + u_i$ e escrever $\tilde{\beta}_1$ como

$$\tilde{\beta}_1 = \beta_1 + \left(n^{-1}\sum_{i=1}^{n}(z_i - \bar{z})u_i\right) \Big/ \left(n^{-1}\sum_{i=1}^{n}(z_i - \bar{z})x_i\right). \tag{5.18}$$

Agora, podemos aplicar a lei dos grandes números ao numerador e denominador, os quais convergem em probabilidade para $\mathrm{Cov}(z,u)$ e $\mathrm{Cov}(z,x)$, respectivamente. Na condição de que $\mathrm{Cov}(z,x) \neq 0$ — de modo que z e x sejam correlacionados —, temos

$$\operatorname{plim} \tilde{\beta}_1 = \beta_1 + \mathrm{Cov}(z,u)/\mathrm{Cov}(z,x) = \beta_1,$$

porque $\mathrm{Cov}(z,u) = 0$ sob RLM.4.

É mais difícil mostrar que $\tilde{\beta}_1$ é assimptoticamente normal. No entanto, usando argumentos semelhantes àqueles do apêndice deste capítulo, pode ser mostrado que $\sqrt{n}\,(\tilde{\beta}_1 - \beta_1)$ é assimptoticamente normal com média zero e variância assimptótica $\sigma^2\,\mathrm{Var}(z)/[\mathrm{Cov}(z,x)]^2$. A variância assimptótica do estimador de MQO é obtida quando $z = x$, caso em que $\mathrm{Cov}(z,x) = \mathrm{Cov}(x,x) = \mathrm{Var}(x)$. Portanto, a variância assimptótica de $\sqrt{n}\,(\hat{\beta}_1 - \beta_1)$, em que $\hat{\beta}_1$ é o estimador de MQO, é $\sigma^2\,\mathrm{Var}(x)/[\mathrm{Var}(x)]^2 = \sigma^2/\mathrm{Var}(x)$. Agora, a desigualdade de Cauchy-Schwartz (veja Apêndice B.4, disponível no site da Cengage) implica $[\mathrm{Cov}(z,x)]^2 \leq \mathrm{Var}(z)\mathrm{Var}(x)$, o que implica que a variância assimptótica de $\sqrt{n}\,(\hat{\beta}_1 - \beta_1)$ não é maior do que a de $\sqrt{n}\,(\tilde{\beta}_1 - \beta_1)$. Assim, para o caso da regressão simples, mostramos que, sob as hipóteses de Gauss-Markov, o estimador de MQO tem uma variância assimptótica menor do que qualquer outro estimador da forma (5.17). [O estimador em (5.17) exemplifica um *estimador de variáveis instrumentais*, que estudaremos extensivamente no Capítulo 15.] Se a hipótese de homoscedasticidade não for válida, então há estimadores da forma (5.17) que têm uma variância assimptótica menor do que a de MQO. Veremos isso no Capítulo 8.

O caso geral é semelhante, mas matematicamente muito mais difícil. No caso de k regressores, a classe de estimadores consistentes é obtida ao generalizar as condições de primeira ordem de MQO:

$$\sum_{i=1}^{n} g_j(\mathbf{x}_i)(y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_{i1} - \ldots - \tilde{\beta}_k x_{ik}) = 0, j = 0, 1, \ldots, k, \tag{5.19}$$

em que $g_j(\mathbf{x}_i)$ representa qualquer função de todas as variáveis explicativas para a observação i. Como pode ser visto ao comparar (5.19) com as condições de primeira ordem de MQO em (3.13), obtemos os estimadores de MQO quando $g_0(\mathbf{x}_i) = 1$ e $g_j(\mathbf{x}_i) = x_{ij}$, para $j = 1, 2, \ldots, k$. A classe dos estimadores em (5.19) é infinita, pois podemos usar qualquer função de x_{ij} que quisermos.

TEOREMA 5.3 (EFICIÊNCIA ASSIMPTÓTICA DE MQO)

Sob as hipóteses de Gauss-Markov, sejam $\tilde{\beta}_j$ os estimadores que solucionam as equações da forma (5.19) e $\hat{\beta}_j$ os estimadores de MQO. Então, para $j = 0, 1, 2, \ldots, k$, os estimadores de MQO têm as menores variâncias assimptóticas: $\mathrm{Avar}\,\sqrt{n}\,(\hat{\beta}_j - \beta_j) \leq \mathrm{Avar}\,\sqrt{n}\,(\tilde{\beta}_j - \beta_j)$.

Provar a consistência dos estimadores em (5.19), sem mencionar que eles são assimptoticamente normais, é matematicamente difícil. [Veja Wooldridge (2002, Capítulo 5).]

RESUMO

As afirmações subjacentes ao material deste capítulo são razoavelmente técnicas, mas suas implicações práticas são diretas. Mostramos que as quatro primeiras hipóteses de Gauss-Markov implicam que MQO é consistente. Além disso, todos os métodos de testar e construir intervalos de confiança que aprendemos no Capítulo 4 são aproximadamente válidos, sem presumir que os erros são extraídos de uma distribuição normal (equivalentemente, a distribuição de y, dadas as variáveis explicativas, não é normal). Isso significa que podemos aplicar MQO e usar os métodos anteriores para um conjunto de aplicações em que a variável dependente não é de fato aproximada e normalmente distribuída. Também mostramos que, em vez da estatística F, a estatística LM pode ser usada para testar restrições de exclusão.

Antes de deixarmos este capítulo, devemos observar que situações como o Exemplo 5.3 podem muito bem apresentar problemas que, *de fato*, exigem atenção especial. Para uma variável como *npre86*, que é zero ou um para a maioria dos homens na população, um modelo linear pode não ser capaz de capturar adequadamente a relação funcional entre *npre86* e as variáveis explicativas. Além do mais, mesmo se um modelo linear descreve o valor esperado das prisões, a heteroscedasticidade poderia ser um problema. Problemas como esses não são mitigados quando o tamanho da amostra aumenta e, portanto, retornaremos a eles em capítulos posteriores.

PROBLEMAS

5.1 No modelo de regressão simples (5.16), sob as primeiras quatro hipóteses de Gauss-Markov, mostramos que o estimador da forma (5.17) é consistente com a inclinação β_1. Dado tal tipo de estimador, defina um estimador de β_0 por $\tilde{\beta}_0 = \bar{y} - \tilde{\beta}_1 \bar{x}$. Mostre que plim $\tilde{\beta}_0 = \beta_0$.

5.2 O conjunto de dados do arquivo SMOKE.RAW contém informações sobre o comportamento tabagista e outras variáveis para uma amostra aleatória de adultos solteiros dos Estados Unidos. A variável *cigs* é o número (médio) de cigarros fumados por dia. Você acha que *cigs* tem uma distribuição normal na população adulta dos Estados Unidos? Explique.

5.3 Suponha que o modelo

$$pctação = \beta_0 + \beta_1 funds + \beta_2 risctol + u$$

satisfaça as quatro primeiras hipóteses de Gauss-Markov, em que *pctação* é a percentagem da pensão de um trabalhador investida no mercado de ações, *funds* é o número de fundos mútuos que o trabalhador pode escolher e *risctol* é alguma medida de tolerância de risco (*risctol* maior significa que a pessoa tem uma tolerância maior ao risco). Se *funds* e *risctol* são positivamente correlacionados, qual é a inconsistência em $\tilde{\beta}_1$, o coeficiente de inclinação da regressão de *pctação* sobre *funds*?

5.4 No modelo de regressão simples sob os RLM.1 até RLM.4, afirmamos que o estimador de inclinação, $\hat{\beta}_1$, é consistente com β_1. Usando $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1$, demonstre que plim $\hat{\beta}_0 = \beta_0$. [Você precisará usar a consistência de $\hat{\beta}_1$ e a lei dos grandes números, juntamente com o fato de que $\beta_0 = \mathrm{E}(y) - \beta_1 \mathrm{E}(x_1)$.]

APÊNDICE 5A

Normalidade Assimptótica dos MQO

Vamos delinear uma prova da normalidade assimptótica de MQO [Teorema 5.2(i)] no caso da regressão simples. Escreva o modelo de regressão simples como na equação (5.16). Em seguida, por meio da álgebra usual da regressão simples, podemos escrever

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) = (1/s_x^2)\left[n^{-1/2}\sum_{i=1}^{n}(x_i - \bar{x})u_i\right],$$

em que usamos s_x^2 para representar a variância amostral de $\{x_i: i = 1, 2, \ldots, n\}$. Pela lei dos grandes números (veja o Apêndice C, disponível no site da Cengage), $s_x^2 \xrightarrow{p} \sigma_x^2 = \mathrm{Var}(x)$. A Hipótese RLM.3 exclui a perfeita colinearidade, o que significa que $\mathrm{Var}(x) > 0$ (x_i varia na amostra, e portanto x não é constante na população). Em seguida, $n^{-1/2}\sum_{i=1}^{n}(x_i - \bar{x})u_i = n^{-1/2}\sum_{i=1}^{n}(x_i - \mu)u_i + (\mu - \bar{x})[n^{-1/2}\sum_{i=1}^{n}u_i]$, em que $\mu = \mathrm{E}(x)$ é a média populacional de x. Agora $\{u_i\}$ é a sequência de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) com média zero e variância σ^2, e portanto $n^{-1/2}\sum_{i=1}^{n}u_i$ converge para a distribuição Normal(0, σ^2) quando $n \to \infty$; isso é exatamente o teorema do limite central do Apêndice C, disponível no site da Cengage. Pela lei dos grandes números, $\mathrm{plim}(\mu - \bar{x}) = 0$. Um resultado padrão da teoria assimptótica é que se $\mathrm{plim}(w_n) = 0$ e z_n tem uma distribuição normal assimptótica, então $\mathrm{plim}(w_n z_n) = 0$. [Veja Wooldridge (2002, Capítulo 3) para mais discussão.] Isso implica $(\mu - \bar{x})[n^{-1/2}\sum_{i=1}^{n}u_i]$ plim zero. Em seguida $\{(x_i - \mu)u_i: i = 1, 2, \ldots\}$ é uma sequência indefinida de variáveis aleatórias i.i.d. com média zero — porque u e x são não correlacionados sob RLM.4 — e variância $\sigma^2\sigma_x^2$, pela hipótese de homoscedasticidade RLM.5. Portanto, $n^{-1/2}\sum_{i=1}^{n}(x_i - \mu)u_i$ tem uma distribuição Normal($0, \sigma^2\sigma_x^2$) assimptótica. Acabamos de mostrar que a diferença entre $n^{-1/2}\sum_{i=1}^{n}(x_i - \bar{x})u_i$ e $n^{-1/2}\sum_{i=1}^{n}(x_i - \mu)u_i$ tem plim zero. Um resultado da teoria assimptótica é que se z_n tem uma distribuição normal e $\mathrm{plim}(v_n - z_n) = 0$, então v_n tem a mesma distribuição normal assimptótica que z_n. Em decorrência $n^{-1/2}\sum_{i=1}^{n}(x_i - \bar{x})u$ também tem uma distribuição Normal($0, \sigma^2\sigma_x^2$) assimptótica. Colocando todas essas peças juntas, temos

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) = (1/\sigma_x^2)\left[n^{-1/2}\sum_{i=1}^{n}(x_i - \bar{x})u_i\right]$$

$$+ [(1/s_x^2) - (1/\sigma_x^2)]\left[n^{-1/2}\sum_{i=1}^{n}(x_i - \bar{x})u_i\right],$$

e como $\mathrm{plim}(1/s_x^2) = 1/\sigma_x^2$, o segundo termo tem plim zero. Portanto, a distribuição assimptótica de $\sqrt{n}(\hat{\beta}_1 - \beta_1)$ é Normal($0, \{\sigma^2\sigma_x^2\}/\{\sigma_x^2\}^2$) = Normal ($0, \sigma^2/\sigma_x^2$). Isso completa a prova para o caso da regressão simples, quando $a_1^2 = \sigma_x^2$ neste caso. Veja Wooldridge (2002, Capítulo 4) para o caso geral.