Notas de aula para o curso de Econometria I

Nota 5: modelo clássico de regressão linear (MCRL) e suas propriedades estatísticas incluindo o teorema de Gauss-Markov

Thiago Fonseca Morello

fonseca.morello@ufabc.edu.br

sala 301, Bloco Delta, SBC

1 O modelo clássico de regressão linear e a hipótese de Gauss-Markov

As propriedades estatísticas do modelo de regressão simples foram obtidas tomando-se por base algumas hipóteses, como, por exemplo, a ausência de covariância entre o termo de perturbação, u, e a variável independente, X. A maneira de proceder é equivalente para o modelo de regressão linear. Antes, porém, de passar à discussão das propriedades estatísticas, cabe detalhar em que consiste o modelo de regressão linear múltipla padrão, o qual, a rigor, recebe a denominação de "Modelo Clássico de Regressão Linear (MCRL)". Tal modelo não deve ser entendido apenas como um procedimento para obter estimadores, ou seja, ele não compreende apenas o método de mínimos quadrados ordinários (MQO). Este, de fato, é apenas um dos três componentes do MCRL. Cabe atentar para os outros dois componentes.

O segundo componente corresponde a um critério de inferência. Uma vez que geralmente os dados disponíveis consistem em uma amostra da população de interesse, os momentos populacionais que compõem o estimador de MQO, $E[x_i.x_i']^{-1}$ e $E[x_i.y_i]$ não são observados. A única saída viável, conforme colocado ao se tratar da estimação, é tomar os momentos amostrais correspondentes, solução esta sugerida pelo princípio da analogia.

O ponto crucial a ser ressaltado é que sempre que a análise empírica dispõe apenas de uma amostra, a regressão linear é uma ferramenta para inferir a expectativa da variável dependente condicional às variáveis explicativas.

O terceiro componente, o qual ainda não foi apresentado de maneira devidamente organizada corresponde a um "conjunto de hipóteses sobre como os dados são produzidos por um 'processo gerador de dados' subjacente" (Greene, 2005, p.10). O objetivo do conjunto de hipóteses é duplo. Trata-se de garantir que (i) o método de estimação, combinado com o critério de inferência, possa ser colocado em prática, (ii) que os estimadores de MQO tenham três propriedades estatísticas desejáveis, quais sejam, ausência de viés, eficiência e consistência.

Cabe assinalar que é apenas sob a validade das hipóteses que constituem o MCRL que essas três propriedades são garantidas. Tais hipóteses não são fatos e, pois, podem se mostrar equivocadas a depender da estrutura dos dados disponíveis. No curso de Econometria I, porém, é assumido que tais hipóteses são válidas, deixando-se para Econometria II as consequências e o tratamento de eventuais violações.

A atenção se volta agora para as hipóteses do MCRL, este é o objetivo da presente seção. Há sete hipóteses, as quais são explicadas nas subseções a seguir e resumidas no quadro ao final desta seção. O número e a ordem das hipóteses, bem como a notação nelas empregada, tende a variar ao longo dos livros-texto, mas o essencial pode ser encontrado no que segue. Alguns livros, como o de Wooldridge, denominam as hipóteses a seguir de hipóteses de Gauss-Markov – o motivo ficará claro na seção 2.

1.2 Linearidade e aleatoriedade

(MCRL 1) A aproximação linear (nos parâmetros) para a FRP é adequada, i.e., não representa um erro considerável.

A primeira hipótese diz respeito à aproximação linear da FEC. Caso esta aproximação seja ruim, i.e., gere um modelo linear que cometa erros de previsão numerosos e consideráveis, as estimativas pontuais para os parâmetros populacionais, bem como os demais resultados da análise empírica, não são confiáveis. Daí a necessidade da hipótese.

Alguns autores, como Wooldridge e Greene[1], enunciam a hipótese MCRL1 assumindo que a FEC é efetivamente linear ou então que a variável dependente pode ser descrita por uma forma linear nos parâmetros envolvendo as explicativas. Opta-se, aqui, seguindo Angrist & Pischke[2] e também Gujarati, por outra abordagem, que consiste em considerar que a expectativa condicional da variável dependente em relação às explicativas não necessariamente é linear nos parâmetros. A aproximação linear é utilizada para tornar tratável o problema de inferir tal expectativa condicional.

Vale a pena recordar que o termo “linear” se refere à maneira como os parâmetros entram no modelo. São excluídas, pois, especificações que contém funções não-lineares dos parâmetros (intercepto e coeficientes), mas não são excluídas especificações que contêm funções não-lineares das explicativas.

(MCRL 2) A amostra disponível é aleatória, i.e., todas as unidades da população têm a mesma probabilidade de serem selecionadas para participar da amostra. É equivalente a afirmar que não há um mecanismo particular que determina quais observações são selecionadas (favorecendo a seleção de algumas observações em detrimento de outras).

A principal implicação da hipótese MCRL2 está em garantir que as observações são independentes e identicamente distribuídas, o que é crucial para obter os estimadores e também para garantir algumas de suas propriedades, especialmente as propriedades de ausência de viés e consistência.

[1] Greene, W. H., Econometric analysis, 2003, quinta edição em inglês.
[2] Angrist, J.D., Pischke, J-F., Mostly harmless econometrics, an empiricist’s companion, 2009, Princeton University Press.

1.3 Posto completo de X

Na nota de aula anterior, seção 4 (estimação com notação matricial), foi assumido que a matriz inversa $E[X_i X_i']^{-1}$ existe. De fato, se a matriz $E[X_i X_i']$ não for invertível, é impossível resolver o sistema de equações composto pelas condições de primeira ordem e, portanto, os estimadores de MQO não podem ser obtidos. Uma matriz quadrada não é invertível quando há uma relação linear perfeita entre pelo menos duas de suas colunas. Ou, para utilizar um termo de álgebra linear, há pelo menos duas colunas linearmente dependentes. É possível demonstrar que a matriz $E[X_i X_i']$ possui colunas linearmente dependentes se e somente se a matriz de explicativas, X, possui colunas linearmente dependentes. Conclusivamente, se e somente se há pelo menos duas explicativas entre as quais há uma relação linear perfeita (determinística) a matriz $E[X_i X_i']$ é não invertível. É preciso assinalar que este princípio incorpora relações lineares entre grupos de explicativas, i.e., os casos em que uma das explicativas é uma função linear de mais de uma das demais explicativas.

Cabe recordar o conceito de dependência linear. Há dependência linear entre as explicativas sempre que elas puderem ser expressas a partir da relação abaixo, na qual $a_1,...,a_K$ são constantes tais que pelo menos uma delas não seja nula.

$$a_1x_1 + a_2x_2 + ... + a_Kx_K = 0$$

Neste caso é sempre possível, pois, escrever uma das explicativas, a j-ésima, por exemplo, como uma combinação linear[3] das demais, ou seja:

$$x_j = -(a_1/a_jx_2 + ... + a_{j-1}/a_jx_{j-1} + a_{j+1}/a_jx_{k+1} + ... + a_K/a_jx_K)$$

Ou, ainda, de maneira mais clara:

$$x_j = b_1x_2 + ... + b_{j-1}x_{j-1} + b_{j+1}x_{k+1} + ... + b_Kx_K$$

Em que $b_s = - a_s/a_j$, s = 1,...,K.

Mas como verificar se há relações lineares perfeitas entre as explicativas? É claro que não se trata de olhar as colunas da matriz X, em uma planilha, por exemplo, comparando-as, em busca de relações lineares. Uma melhor maneira de proceder consiste em examinar a definição das explicativas incorporadas ao modelo. Vejamos dois exemplos.

Retomando o problema da desnutrição infantil, a principal explicativa é a renda. Seja assumido que a teoria ou algum estudo prévio sugere a incorporação da despesa familiar em alimentação infantil. Porém, por alguma imprecisão nos dados disponíveis, é impossível, com base neles, chegar ao valor da despesa em alimentação infantil. Mesmo

[3] O conceito de combinação linear é oriundo da teoria de álgebra linear. Ele define a combinação linear de K variáveis como sendo equivalente à expressão $a_1x_1 + a_2x_2 + a_3x_3 + ... + a_Kx_K$, em que $a_1, a_2, ..., a_K$ são escalares constantes. Para mais detalhes, ver o capítulo 3 da apostila de Sérgio Luis Zani disponível no link http://www.icmc.usp.br/~szani/alglin.pdf.

assim, o pesquisador encontrou em um estudo prévio, não necessariamente sobre o mesmo tema, a informação de que as famílias brasileiras com crianças de zero a cinco anos gastam em torno de 30% de sua renda em alimentação infantil. Ocorre ao pesquisador incorporar, como explicativa, a seguinte proxy para a despesa em alimentação infantil: desp_ali_fam = 0,3renda_fam, em que renda_fam é a renda familiar. Não há problema algum em fazer isso - exceto pelas imprecisões introduzidas pela baixa qualidade da proxy - caso a variável renda_fam não seja também incluída. Porém, a inclusão de ambas, desp_ali_fam e renda_fam daria origem a um conjunto de explicativas linearmente dependente, dado que a primeira é uma combinação linear da segunda. O que talvez seja mais claramente percebido a partir da seguinte manipulação da definição da proxy: desp_ali_fam - 0,3renda_fam = 0.

O segundo exemplo é o da armadilha das variáveis binárias (dummies). A região brasileira de localização do domicílio pode ser incorporada à equação de desnutrição infantil por meio de quatro binárias, cada uma indicando com valor unitário a localização em uma das cinco regiões e com zero a localização em qualquer outra região. Como visto na nota de aula 8, com cinco binárias, uma para cada uma das regiões, tem-se redundância, a qual pode ser apreendida a partir do fato de que:

$$d\_N + d\_NE + d\_CO + d\_SE + d\_S = 1 \quad (1)$$

A redundância, pois, se manifesta como dependência linear entre as cinco variáveis binárias e o intercepto. Isso pois o intercepto também pode ser compreendido como um coeficiente multiplicado por uma variável não-aleatória constante, a qual assume valor unitário para todas as observações. É exatamente o que se leva em conta ao ampliar a matriz X, introduzindo uma coluna de valores unitários. De maneira mais clara, e utilizando a notação matricial, a matriz X pode ser expressa como um conjunto de vetores, cada um deles representando os valores de uma dada explicativa para todas as observações.

$$X = \begin{bmatrix} 1 & Z & d_N & d_{NE} & d_{CO} & d_{SE} & d_S \end{bmatrix}$$

Em que Z é uma matriz com todas as explicativas quantitativas.

Uma combinação linear de Z e dos vetores que compõem X pode ser escrita como:

$$a_1 1 + a_2 Z + a_3 d\_N + a_4 d\_NE + a_5 d\_CO + a_6 d\_SE + a_7 d\_S$$

Seja assumido que $a_2 = 0$, $a_1 = -1$ e, $a_3 = a_4 = ... = a_7 = 1$, o que é plenamente compatível com a definição de combinação linear e não implica em perda de generalidade para os propósitos em questão. Com isso obtém-se:

$$-1 + d\_N + d\_NE + d\_CO + d\_SE + d\_S \quad (2)$$

Esta expressão (2) é claramente nula à luz da expressão (1). É por isso que a inclusão de um número de dummies equivalente ao número de categorias de uma variável qualitativa introduz dependência linear na matriz de explicativas.

Conclusivamente, se todas as dummies possíveis forem incorporadas ao modelo de regressão linear, será impossível, pois, obter os estimadores. A insistência, pois, em manter uma variável binária redundante repercute em um gravíssimo resultado. Este problema é o que se entende por “armadilha das variáveis dummy”. A solução geral é, porém, simples: as M categorias de uma variável qualitativa devem ser incorporadas a partir de M – 1 variáveis binárias. Consequentemente, evitando-se a redundância, evita-se o problema causado por uma matriz de explicativas com dependência linear.

Uma vez entendida a importância da independência linear entre as explicativas, pode-se enunciar a hipótese a seguir, referente à ausência de independência linear na população.

(MCRL 3) Não há dependência linear entre as variáveis explicativas, o que garante que a matriz $E[x_i x_i']$ seja invertível. É equivalente afirmar que X tem posto completo, entendendo-se por posto o número de colunas linearmente independentes de X (K +1 colunas, no caso).

1.4 Exogeneidade

(MCRL 4) A expectativa condicional do termo de perturbação é nula, i.e., $E[u_i|x_i] = 0$, $i=1,...,N$.

A hipótese MCRL 4 tem grande importância, uma vez que é apenas sob sua validade que os estimadores de MQO são não-viesados. Este fato já foi estabelecido para a regressão linear, e, para a regressão múltipla, a exposição pode ser encontrada na próxima seção. A propriedade de consistência, pelo menos tão importante quanto a de ausência de viés, a ser discutida mais à frente no curso (parte IV), também depende da hipótese MCRL 4.

Uma implicação da hipótese MCRL 4 é a ausência de covariância entre o termo de perturbação e as explicativas, conforme estabelecido na nota de aula 5, com um argumento válido também para regressões múltiplas (ver seção 3.2.2).

Uma variável explicativa é dita “exógena” quando não-correlacionada com o termo de perturbação e, em caso contrário, é dita “endógena”. Como ausência de covariância é equivalente à ausência de correlação, a hipótese MCRL 4 implica que as variáveis explicativas são exógenas.

1.5 Perturbações esféricas

As duas hipóteses a seguir dizem respeito ao comportamento estatístico do termo de perturbação. Quando elas são válidas, as variâncias dos estimadores são consideravelmente mais simples e seguem a fórmula geral a ser derivada no estudo das propriedades estatísticas.

(MCRL 5) As perturbações são homocedásticas, i.e., $V[u_i|\, x_i] = E[(u_i - E[u_i])^2|\, x_i] = E[u_i^2|\, x_i] = \sigma^2$.

(MCRL 6) As perturbações são não-autocorrelacionadas, i.e., $corr[u_i, u_j|\, x_i, x_j] = \frac{cov[u_i, u_j \vee x i, x j]}{\sqrt{V[u_i \vee x i, x j] V[u_j \vee x i, x j]}} = 0 \rightarrow cov[u_i, u_j|\, x_i, x_j] = E[u_i u_j|\, x_i, x_j] - E[u_i|\, x_i, x_j] E[u_j|\, x_i, x_j] = 0 \rightarrow E[u_i u_j|\, x_i, x_j] = 0, \forall\, i \neq j$.

A primeira hipótese, de homocedasticidade, requer que o valor da variância das perturbações não difira ao longo das observações. Já a segunda hipótese, de ausência de autocorrelação, exige que não exista correlação entre perturbações referentes a observações distintas.

1.6 Normalidade dos erros

(MCRL 7) A função de distribuição de probabilidades (FD) condicional do termo de perturbação é normal, com média zero e variância σ^2 *(hipótese de Gauss-Makov), i.e,* $u_i|X \sim N(0, \sigma^2), i=1,...,N$.

A última hipótese diz respeito à função de distribuição de probabilidades (FD) do termo de perturbação. É requerido que esta seja normal. Como a expectativa condicional do termo de perturbação é nula, segundo MCRL 4 e sua variância é constante e equivalente a σ^2 de acordo com MCRL 5, a distribuição é obrigatoriamente a normal com média e variância populacionais equivalentes, respectivamente, a 0 e σ^2.

A normalidade das perturbações é crucial para a inferência, especificamente para a estimação em intervalos e para a realização de testes de hipóteses. A razão está em que ambos os procedimentos requerem o conhecimento da FD dos estimadores de MQO, a qual apenas pode ser obtida quando a distribuição de probabilidades do termo de perturbação é conhecida.

Quadro Hipóteses do MCRL

(MCRL 1) A aproximação linear (nos parâmetros) para a FRP é adequada, i.e., não representa um erro considerável.

(MCRL 2) A amostra disponível é aleatória, i.e., todas as unidades da população têm a mesma probabilidade de serem selecionadas para participar da amostra. É equivalente a afirmar que não há um mecanismo particular que determina quais observações são selecionadas.

(MCRL 3) Não há dependência linear entre duas variáveis explicativas, o que garante que a matriz $E[x_i x_i']$ seja invertível. É equivalente afirmar que X tem posto completo, entendendo-se por posto o número de colunas linearmente independentes de X (K +1 colunas, no caso).

(MCRL 4) A expectativa condicional do termo de perturbação é nula, i.e., $E[u_i|x_i] = 0$, $i=1,\ldots,N$.

(MCRL 5) As perturbações são homocedásticas, i.e., $V[u_i| x_i] = E[(u_i - E[u_i])^2| x_i] = E[u_i^2| x_i] = \sigma^2$.

(MCRL 6) As perturbações são não-autocorrelacionadas, i.e., $corr[u_i, u_j| x_i, x_j] = \frac{cov[u_i, u_j \vee x\,i, x\,j]}{\sqrt{V[u_i \vee x\,i, x\,j]\,V[u_j \vee x\,i, x\,j]}} = 0 \rightarrow cov[u_i, u_j| x_i, x_j] = E[u_i u_j| x_i, x_j] - E[u_i| x_i, x_j]E[u_j| x_i, x_j] = 0 \rightarrow E[u_i u_j| x_i, x_j] = 0, \forall\, i \neq j$.

(MCRL 7) A função de distribuição condicional do termo de perturbação é normal, com média zero e variância σ^2 (hipótese de Gauss-Makov), i.e, $u_i|X \sim N(0, \sigma^2)$, $i=1,\ldots,N$.

2 Comentário sobre as propriedades estatísticas, variância dos estimadores e teorema de Gauss-Markov

2.1 Propriedades estatísticas

É possível demonstrar, como consta no apêndice A ao final desta nota, que, sob as hipóteses MCRL1 a MCRL4 os estimadores de MQO para os K+1 parâmetros da FRP são calculáveis e, cabe frisar, não-viesados. As hipóteses MCRL 5 e 6 têm duas finalidades. A primeira é a de permitir fórmulas simplificadas para as variâncias dos estimadores, da mesma maneira como foi demonstrado para a regressão simples. A segunda finalidade é a de garantir, em conjunto com as hipóteses MCRL1 a MCRL4, a validade do importante resultado em que consiste o teorema de Gauss-Markov (seção 2.4 abaixo).

2 Variância dos estimadores de mínimos quadrados ordinários

2.2 Componentes da variância dos estimadores

É sabido que a qualidade de um estimador é função não apenas da distância entre sua expectativa e o valor populacional-alvo mas também de sua variância. Um estimador com alta variância tende a gerar valores que variam consideravelmente ao longo das amostras (retomando, neste raciocínio, o experimento mental com base no qual as propriedades desejáveis dos estimadores são definidas). O que não é adequado, uma vez que geralmente apenas uma amostra está disponível. De fato, um estimador com alta variância é impreciso e, pois, pouco confiável.

Quais fatores determinam o valor da variância dos estimadores de MQO para os parâmetros da FRP? Sejam considerados apenas os estimadores para os coeficientes. Sob as hipóteses de homocedasticidade e ausência de autocorrelação (MCRL 5 e MCRL

6), a fórmula do estimador de MQO para o k-ésimo coeficiente da FRP é tal como segue.

$$V\left[\hat{\beta}_{MQO\,k}\middle|X\right]=\frac{\sigma^2}{\sum_{i=1}^{N}\left(x_{ik}-\bar{x}_k\right)^2\left(1-R_k^2\right)}$$

A fórmula deixa claro que há apenas três componentes que determinam o tamanho da variância do estimador para o coeficiente. Cabe compreender o que cada componente representa e como contribui.

Componente 1, σ^2, variância das perturbações. É esclarecedor estabelecer a relação entre a variância das perturbações e a variância da variável explicada, Y, como os passos a seguir mostram.

(Passo 1) É fato que $V[y_i|X] = E[(y_i - E[y_i|X])^2|X] = E[u_i^2|X] = \sigma^2$, pois, de acordo com a propriedade de decomposição de y_i que fundamenta a FRP, $y_i = E[y_i |X] + u_i$ (1). Este resultado implica que $E_x[V[y_i|X]] = E_x[\sigma^2] = \sigma^2$, uma vez que σ^2 é uma constante;

(Passo 2) Do teorema ANOVA (vide gabarito da lista 1, exercício 3), sabemos que $V[y_i] = E_x[V[y_i|X]] + V_x[E[y_i|X]]$.

(Passo 3) Incorporando o resultado do passo 2 ao passo 3, temos: $V[y_i] = \sigma^2 + V_x[E[y_i|X]]$, i.e., $\sigma^2 = V[y_i] - V_x[E[y_i|X]]$.

Conclusivamente, pois, a variância do termo de perturbação é equivalente ao cômputo da variabilidade de Y e da variabilidade da média condicional de X. Portanto, mantidos constantes os demais componentes da variância do estimador de MQO e também a variância das médias condicionais, quanto maior a variabilidade de y_i na população, maior a variância do estimador do k-ésimo coeficiente.

Componente 2, SQT da k-ésima variável, $\sum_{i=1}^{N}\left(x_{ik}-\bar{x}_k\right)^2$. Quanto maior a variabilidade da k-ésima explicativa, mantidos os demais componentes fixos, menor a variância do estimador para o k-ésimo coeficiente, portanto, mais preciso é este estimador.

Componente 3, R_k^2. Uma alta fração da variabilidade do k-ésimo regressor explicada pelos demais regressores, esta a essência de R_k^2, indica que o primeiro é redundante, i.e., a informação nele contida já está contida nos demais regressores. Há, pois, correlação considerável (estatisticamente significativa) entre o k-ésimo regressor e pelo menos um dos outros regressores, um problema denominado por multicolinearidade – tema tratado em econometria 2[4]. Com R_k^2 próximo de 100%, i.e. com a variação de x_k na amostra sendo quase totalmente explicada pelas demais variáveis independentes, a inclusão da k-ésima explicativa no modelo pouco acrescenta, em termos informacionais. Decorre que, no limite, $1-R_k^2$ se aproxima de zero e, pois, a variação da k-ésima explicativa se

[4] Uma referência interessante para multicolinearidade é o capítulo 23 do livro de Goldberger, A.S., A course in econometrics, 1991, Harvard College.

aproxima do infinito. É desejável, pois, que cada explicativa tenha um componente de sua variação que seja específico, não correlacionado com as demais explicativas.

A principal conclusão a ser retirada da discussão desta seção é a que, para que o estimador do k-ésimo coeficiente tenha um nível minimamente aceitável de precisão (variância tolerável), é preciso que a explicativa a ele associada (i) tenha variabilidade relevante e (ii) não seja redundante com as demais explicativas. Também é válido observar, ainda que não totalmente correto (dada a influência da variância das expectativas condicionais de Y), que a variabilidade de Y contribui negativamente para a precisão do estimador.

2.3 Estimador para a variância dos estimadores de MQO

Assim como para a regressão simples, o cálculo da variância dos estimadores da regressão múltipla pressupõe o conhecimento do valor de σ^2, a variância das perturbações. Como as últimas nunca são observadas é impossível conhecer σ^2. É preciso recorrer à estimação. O estimador não-viesado neste caso é:

$$\hat{\sigma}^2 = \frac{1}{N-(K+1)} \sum_{i=1}^{N} \hat{u}_i^2$$

Em que $\sum_{i=1}^{N} \hat{u}_i^2 = \sum_{i=1}^{N} ¿¿$, a soma dos quadrados dos resíduos da regressão. É necessário dividir esta quantidade por N - (K+1) pois este é o número de graus de liberdade remanescentes após a estimação dos K+1 parâmetros populacionais da FRP pressuposta pelo cálculo dos resíduos. É com base, pois, nesta quantidade de "informação bruta" que a SQR ($\sum_{i=1}^{N} \hat{u}_i^2$¿ é calculada.

A necessidade de estimar σ^2 implica que a variância dos estimadores de MQO também tem de ser estimada, sendo o estimador dado por:

$$\widehat{V\left[\hat{\beta}_{MQOk} \middle| X\right]} = \frac{\hat{\sigma}^2}{\sum_{i=1}^{N} \left(x_{ik} - \bar{x}_k\right)^2 \left(1 - R_k^2\right)}, k = 1, \ldots, N$$

É preciso assinalar que o sinal "^" indica que a expressão acima se refere ao estimador para a variância do estimador de MQO para o k-ésimo coeficiente.

3 Teorema de Gauss-Markov: eficiência dos estimadores de MQO

A ausência de viés garante que se o estimador de MQO para o vetor β for calculado, repetidas vezes, em todas as amostras que podem ser retiradas da população, então o valor médio dele é equivalente ao valor efetivo de β. Esta propriedade é de grande importância.

É, porém, igualmente importante saber quanto à volatilidade das estimativas pontuais. A razão para isso está em que um estimador pode ser não viesado e, mesmo assim, gerar estimativas pontuais consideravelmente discrepantes entre si. De fato, viés e variância são medidas distintas e, portanto, complementares para o desempenho de um estimador. Enquanto o viés compara as estimativas pontuais com o valor "verdadeiro" do parâmetro, a variância compara as estimativas pontuais com a própria média delas. Os gráficos abaixo ilustram a diferença entre o viés e os desvios da média captado pelas variâncias.

Figura 1 Viés de um estimador

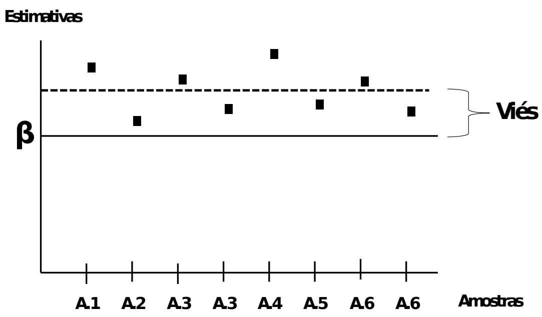
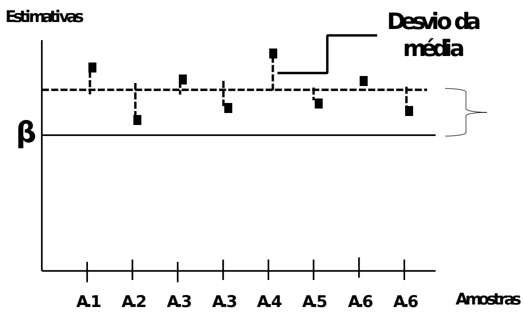


Figura 2 Variância de um estimador



O teorema de Gauss Markov, enunciado a seguir, é um dos principais resultados que garante a qualidade dos estimadores de MQO. Trata-se de um resultado forte pois assevera que o estimador de MQO é o melhor estimador entre todos os estimadores lineares não viesados, de maneira que não é necessário procurar outros estimadores para os parâmetros populacionais. A generalidade do resultado explica porque a regressão linear é uma das principais ferramentas de análise empírica utilizadas por economistas.

Teorema de Gauss-Markov: sob as hipóteses MCRL1 a MCLR6, o estimador $\hat{\beta}_{MQO}$ possui a menor variância entre todos os estimadores lineares não viesados. O que o atribui a propriedade de "melhor estimador linear não viesado" (MELNV ou, na sigla em inglês, BLUE, Best Linear Unbiased Estimator).

A demonstração do teorema acima pode ser encontrada no apêndice B ao final desta nota de aula. Por hora, basta assinalar que o teorema de Gauss-Markov é válido apenas se todas as hipóteses MCRL1 a MCRL6 são verificadas.

Apêndice A Propriedades estatísticas (com notação matricial)

A.1 Ausência de viés

O estudo da propriedade de ausência de viés segue a mesma estrutura adotada para a regressão simples, ou seja, a de uma demonstração. O resultado que se deseja demonstrar é:

$$B¿$$

É preciso assinalar que tanto $\hat{\beta}_{MQO}$ como β são vetores de dimensão (K+1)x1. A demonstração, portanto, tem de recorrer a operações matemáticas próprias ao "mundo das matrizes".

O primeiro passo consiste em relembrar a definição de $\hat{\beta}_{MQO}$:

$$\hat{\beta}_{MQO}=\left(\frac{1}{N}\sum_{i=1}^{N}x_i x_i{'}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}x_i y_i\right)$$

Ou, de modo equivalente:

$$\hat{\beta}_{MQO}=N\left(\sum_{i=1}^{N}x_i x_i{'}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}x_i y_i\right)$$

Esta passagem considera que $\left(\frac{1}{N}\right)^{-1}=N$.

Simplificando mais ainda:

$$\hat{\beta}_{MQO}=N\frac{1}{N}\left(\sum_{i=1}^{N}x_i x_i{'}\right)^{-1}\left(\sum_{i=1}^{N}x_i y_i\right)=\left(\sum_{i=1}^{N}x_i x_i{'}\right)^{-1}\left(\sum_{i=1}^{N}x_i y_i\right)$$

A definição de FRP postula:

$y_i = x_i'\beta + u_i$, i=1,...,N (2)

Aplicando (2) em (1):

$$\hat{\beta}_{MQO}=\left(\sum_{i=1}^{N}x_i x_i{'}\right)^{-1}\left(\sum_{i=1}^{N}x_i(x_i{'}\beta+u_i)\right)$$

Manipulando:

$$\hat{\beta}_{MQO}=\left(\sum_{i=1}^{N}x_i x_i{'}\right)^{-1}\left(\sum_{i=1}^{N}x_i x_i^{'}\beta+x_i u_i\right)=\left(\sum_{i=1}^{N}x_i x_i{'}\right)^{-1}\left(\sum_{i=1}^{N}x_i x_i{'}\beta+\sum_{i=1}^{N}x_i u_i\right)\leftrightarrow$$

$$\hat{\beta}_{MQO}=\left(\sum_{i=1}^{N}x_i x_i{'}\right)^{-1}\left(\sum_{i=1}^{N}x_i x_i^{'}\beta\right)+\left(\sum_{i=1}^{N}x_i x_i{'}\right)^{-1}\left(\sum_{i=1}^{N}x_i u_i\right)(3)$$

Para prosseguir, é preciso perceber que:

$$\sum_{i=1}^{N}x_i x_i^{'}\beta=\left(x_1 x_1^{'}\beta+x_2 x_2^{'}\beta+\ldots+x_K x_K^{'}\beta\right)=\left(x_1 x_1^{'}+x_2 x_2^{'}+\ldots+x_K x_K^{'}\right)\beta$$

De modo que:

$$\sum_{i=1}^{N} x_i x_i' \beta = \left(\sum_{i=1}^{N} x_i x_i'\right) \beta \, (4)$$

Levando (4) em (3):

$$\hat{\beta}_{MQO} = \left(\sum_{i=1}^{N} x_i x_i'\right)^{-1} \left(\sum_{i=1}^{N} x_i x_i'\right) \beta + \left(\sum_{i=1}^{N} x_i x_i'\right)^{-1} \sum_{i=1}^{N} x_i u_i \, (3')$$

Neste ponto, deve-se atentar para o fato de que $\left(\sum_{i=1}^{N} x_i x_i'\right)^{-1}$ é a inversa da matriz $\sum_{i=1}^{N} x_i x_i'$. Consequentemente, $\left(\sum_{i=1}^{N} x_i x_i'\right)^{-1} \left(\sum_{i=1}^{N} x_i x_i'\right) = I_{K+1}$ em que I_{K+1} é a matriz identidade de dimensão K+1.

O primeiro termo do lado direito de (3’) pode ser consideravelmente simplificado:

$$\hat{\beta}_{MQO} = I_{K+1} \beta + \left(\sum_{i=1}^{N} x_i x_i'\right)^{-1} \sum_{i=1}^{N} x_i u_i = \beta + \left(\sum_{i=1}^{N} x_i x_i'\right)^{-1} \sum_{i=1}^{N} x_i u_i \, (3'')$$

Chega-se, pois, à seguinte fórmula simplificada para o estimador do vetor de parâmetros:

$$\hat{\beta}_{MQO} = \beta + \left(\sum_{i=1}^{N} x_i x_i'\right)^{-1} \sum_{i=1}^{N} x_i u_i \, (3'')$$

Aplicando esta fórmula na definição de viés do estimador $\hat{\beta}_{MQO}$, tem-se:

$$B\left[\hat{\beta}_{MQO} \middle| X\right] = E\left[\hat{\beta}_{MQO} \middle| X\right] - \beta = E\left[\beta + \left(\sum_{i=1}^{N} x_i x_i'\right)^{-1} \sum_{i=1}^{N} x_i u_i \middle| X\right] - \beta \leftrightarrow$$

$$B\left[\hat{\beta}_{MQO} \middle| X\right] = E\left[\left(\sum_{i=1}^{N} x_i x_i'\right)^{-1} \sum_{i=1}^{N} x_i u_i \middle| X\right]$$

Utilizando o fato de que a expectativa é condicional em X:

$$B\left[\hat{\beta}_{MQO} \middle| X\right] = \left(\sum_{i=1}^{N} x_i x_i'\right)^{-1} E\left[\sum_{i=1}^{N} x_i u_i \middle| X\right]$$

A expectativa de uma soma é sempre equivalente à soma das expectativas, portanto:

$$B\left[\hat{\beta}_{MQO} \middle| X\right] = \left(\sum_{i=1}^{N} x_i x_i'\right)^{-1} \sum_{i=1}^{N} E[x_i u_i \vee X]$$

Empregando mais uma vez o fato de que a expectativa é condicional a X:

$$B\left[\hat{\beta}_{MQO} \middle| X\right] = \left(\sum_{i=1}^{N} x_i x_i'\right)^{-1} \sum_{i=1}^{N} x_i E[u_i \vee X]$$

É neste estágio que a hipótese MCRL 4 se mostra necessária. Ela estabelece que $E\left[u_i \middle| X\right] = 0$, i=1,...,N. Sendo ela válida, portanto, tem-se que:

$$B\left[\hat{\beta}_{MQO} \middle| X\right] = \left(\sum_{i=1}^{N} x_i x_i'\right)^{-1} \sum_{i=1}^{N} x_i 0 = 0$$

Ou seja:

$$B\left[\hat{\beta}_{MQO} \middle| X\right] = 0$$

Fica, pois, demonstrado que, sob MCRL 4, o estimador de MQO para o vetor de parâmetros β é não-viesado.

A.2 Variância do estimador de MQO

$$V\left[\hat{\beta}_{MQO}\middle|X\right]=E\left[\left(\hat{\beta}_{MQO}-E\left[\hat{\beta}_{MQO}\middle|X\right]\right)\left(\hat{\beta}_{MQO}-E\left[\hat{\beta}_{MQO}\middle|X\right]\right)'\middle|X\right]=E\left[\left(\hat{\beta}_{MQO}-\beta\right)\left(\hat{\beta}_{MQO}-\beta\right)'\middle|X\right](i)$$

A última passagem decorre do fato de que $\hat{\beta}_{MQO}$ é não-viesado. Incorporando a definição segundo a qual $\hat{\beta}_{MQO}=\beta+\left(\sum_{i=1}^{N}x_i x_i'\right)^{-1}\sum_{i=1}^{N}x_i u_i$ e, portanto, $\hat{\beta}_{MQO}-\beta=\left(\sum_{i=1}^{N}x_i x_i'\right)^{-1}\sum_{i=1}^{N}x_i u_i$, a expressão (i) acima passa a:

$$V\left[\hat{\beta}_{MQO}\middle|X\right]=E\left[\left(\left(\sum_{i=1}^{N}x_i x_i'\right)^{-1}\sum_{i=1}^{N}x_i u_i\right)\left(\left(\sum_{i=1}^{N}x_i x_i'\right)^{-1}\sum_{i=1}^{N}x_i u_i\right)'\middle|X\right]=¿$$

$$E\left[\left(\sum_{i=1}^{N}x_i x_i'\right)^{-1}\left(\sum_{i=1}^{N}x_i u_i\right)\left(\sum_{i=1}^{N}x_i u_i\right)'\left(\sum_{i=1}^{N}x_i x_i'\right)^{-1}\middle|X\right]=¿$$

$$\left(\sum_{i=1}^{N}x_i x_i'\right)^{-1}E\left[\left(\sum_{i=1}^{N}x_i u_i\right)\left(\sum_{i=1}^{N}x_i u_i\right)'\middle|X\right]\left(\sum_{i=1}^{N}x_i x_i'\right)^{-1}(ii)$$

Neste ponto é preciso relembrar uma propriedade do operador de transposição, segundo a qual [A+B]' = A' + B', de modo que

$$\left(\sum_{i=1}^{N}x_i u_i\right)'=\sum_{i=1}^{N}\left(x_i u_i\right)'(iii)$$

E, de acordo com outra propriedade do operador de transposição,

$$\sum_{i=1}^{N}\left(x_i u_i\right)'=\sum_{i=1}^{N}u_i' x_i'(iv)$$

Neste ponto, é importante lembrar que u_i é escalar, i.e., tem dimensão 1 x 1. Com isso, $u_i' = u_i$.

Incorporando (iii) e (iv) em (ii):

$$V\left[\hat{\beta}_{MQO}\middle|X\right]=\left(\sum_{i=1}^{N}x_i x_i'\right)^{-1}E\left[\left(\sum_{i=1}^{N}x_i u_i\right)\sum_{i=1}^{N}u_i x_i'\middle|X\right]\left(\sum_{i=1}^{N}x_i x_i'\right)^{-1}(v)$$

Um fato crucial a considerar diz respeito à maneira de representar $\left(\sum_{i=1}^{N}x_i u_i\right)\sum_{i=1}^{N}u_i x_i'$.

Quanto a isso:

$$\left(\sum_{i=1}^{N}x_i u_i\right)\sum_{i=1}^{N}u_i x_i'=\left(x_1 u_1+x_2 u_2+\ldots+x_N u_N\right)\left(u_1 x_1'+u_2 x_2'+\ldots+u_N x_N'\right)$$

Esta expressão pode ser representada como $\left(a_1+a_2+\ldots+a_N\right)\left(b_1+b_2+\ldots+b_N\right)$, esta última sendo equivalente a $\sum_{i=1}^{N}a_i b_i+\sum_{j\neq 1}^{N}\sum_{i=1}^{N}a_i b_j$. Reincorporando as definições de a_i e b_i, tem-se, finalmente, que:

$$\left(\sum_{i=1}^{N} x_i u_i\right)\sum_{i=1}^{N} u_i' x_i' = \sum_{i=1}^{N} x_i u_i u_i x_i' + \sum_{j\neq 1}^{N}\sum_{i=1}^{N} x_i u_i u_j x_j' = \sum_{i=1}^{N} x_i u_i^2 x_i' + \sum_{j\neq 1}^{N}\sum_{i=1}^{N} x_i u_i u_j x_j'$$

Levando (vi) a (v), tem-se:

$$V\left[\hat{\beta}_{MQO}\middle|X\right]=\left(\sum_{i=1}^{N} x_i x_i'\right)^{-1} E\left[\sum_{i=1}^{N} x_i u_i^2 x_i' + \sum_{j\neq 1}^{N}\sum_{i=1}^{N} x_i u_i u_j x_j' \middle| X\right]\left(\sum_{i=1}^{N} x_i x_i'\right)^{-1}$$

Aplicando a propriedade de que a expectativa da soma é a soma das expectativas:

$$V\left[\hat{\beta}_{MQO}\middle|X\right]=\left(\sum_{i=1}^{N} x_i x_i'\right)^{-1}\left(\sum_{i=1}^{N} x_i E\left[u_i^2\right] x_i' + \sum_{j\neq 1}^{N}\sum_{i=1}^{N} x_i E\left[u_i u_j\right] x_j'\right)\left(\sum_{i=1}^{N} x_i x_i'\right)^{-1} (vi)$$

Supondo válidas as hipóteses de homocedasticidade e de ausência de autocorrelação, MCRL5 e MCRL6, i.e., $E\left[u_i^2\right]=\sigma^2$ e $E\left[u_i u_j\right]=0, i, j=1, \ldots, N$,a equação (vi) se converte em:

$$V\left[\hat{\beta}_{MQO}\middle|X\right]=\left(\sum_{i=1}^{N} x_i x_i'\right)^{-1}\left(\sum_{i=1}^{N} \sigma^2 x_i x_i'\right)\left(\sum_{i=1}^{N} x_i x_i'\right)^{-1}=\sigma^2\left(\sum_{i=1}^{N} x_i x_i'\right)^{-1}\left(\sum_{i=1}^{N} x_i x_i'\right)\left(\sum_{i=1}^{N} x_i x_i'\right)^{-1} \rightarrow$$

$$V\left[\hat{\beta}_{MQO}\middle|X\right]=\sigma^2\left(\sum_{i=1}^{N} x_i x_i'\right)^{-1}$$

Chega-se, pois, ao formato geral da matriz que contém em sua diagonal principal as variâncias dos K+1 estimadores para os K+1 parâmetros populacionais. Os elementos que não pertencem à diagonal principal correspondem a covariâncias entre os estimadores. Deste formato geral decorre que a variância de um estimador para o k-ésimo coeficiente é:

$$V\left[\hat{\beta}_{MQO\,k}\middle|X\right]=\frac{\sigma^2}{\sum_{i=1}^{N}\left(x_{ik}-\bar{x}_k\right)^2\left(1-R_k^2\right)}$$

Em que R^2_k é o coeficiente de determinação da regressão da k-ésima explicativa contra todas as demais explicativas. A estatística $\sum_{i=1}^{N}\left(x_{ik}-\bar{x}_k\right)^2$ mede a variação total da k-ésima explicativa na amostra (ela é praticamente igual à variância de x_k), sendo denotada também por SQT_k. Cabe recordar que σ^2 é a variância do termo de perturbação, $E\left[u_i^2\right]=\sigma^2$.

Apêndice B, teorema de Gauss-Markov

Em primeiro lugar é preciso deixar claro que “estimador linear” é um objeto mais geral do que meramente um estimador obtido a partir de um modelo de regressão linear. De fato, por estimador linear se entende um estimador em cuja fórmula a variável dependente (Y) entra de forma linear, i.e., trata-se de um estimador que é uma função linear da variável dependente. Retomando a fórmula do estimador de MQO, pode-se ver que ele é linear:

$$\hat{\beta}_{MQO}=\left(\sum_{i=1}^{N} x_i x_i'\right)^{-1}\sum_{i=1}^{N} x_i y_i$$

Basta perceber que a variável dependente aparece em nível e não na forma de uma função não-linear. Talvez, com a notação matricial abaixo, a natureza linear do estimador de MQO fique mais clara:

$$\hat{\beta}_{MQO} = (X'X)^{-1} X'Y$$

Em que X é a matriz Nx(K+1) de variáveis explicativas e Y o vetor Nx1 com os valores da variável dependente.

Uma forma genérica para um estimador linear para o vetor β, pois, é a seguinte:

$$\tilde{\beta} = H(X)'Y$$

Em que H(X) é uma função não necessariamente linear de X – a inversa de X'X, por exemplo, não é uma função linear. Partindo-se da premissa de que $\tilde{\beta}$ é não viesado, pode-se demonstrar que sua variância será obrigatoriamente maior do que a variância de $\hat{\beta}_{MQO}$.

Passo 1, ausência de viés

Tomando a forma genérica acima, pois, $B¿$, ou seja, $B¿$. Considerando que a expectativa é condicional em X, esta condição passa a $H(X)'E¿$. A FRP pode ser expressa, em notação matricial, como Y = Xβ + u (1c). Aplicando (1c) em (1b), tem-se: $H(X)'E¿$ = 0, o que é equivalente a $H(X)'X\beta + H(X)'E¿$, ou, reorganizando, $H(X)'X\beta - \beta + H(X)'E¿$. Assumindo que MCRL 4 vale, a condição é reduzida a $H(X)'X\beta - \beta = 0$ ou, fatorando, $[H(X)'X - I_{K+1}¿\beta = 0$ (1d). A última passagem toma por base que uma matriz identidade é sempre quadrada, de maneira que, como a dimensão de X é Nx(K+1),então a fatoração apenas é coerente se a matriz identidade tiver dimensão (K+1)x(K+1). A equação (1d), em que se desdobra o requerimento de ausência de viés, apenas tem solução quando o termo entre colchetes é nulo, dado que β é um vetor constante e, não necessariamente nulo em todos seus componentes. Portanto, para que o estimador linear genérico seja não viesado, é preciso ter $H(X)'X = I_{K+1}$. Esta condição será assumida, uma vez que o resultado que se procura estabelecer é o de que o estimador de MQO é o de menor variância dentro da classe de estimadores lineares não viesados. Não se tem por objetivo estabelecer que o estimador de MQO é o de menor variância na classe de todos os estimadores possíveis para os parâmetros da FRP.

Passo 2, variância de $\tilde{\beta}$

É necessário obter a variância do estimador genérico, no que será adotada a notação matricial. É sabido (vide subseção anterior) que V[$\tilde{\beta}$|X] = E[($\tilde{\beta}$-E[$\tilde{\beta}$|X])($\tilde{\beta}$-E[$\tilde{\beta}$|X])'|X]. Como se está assumindo ausência de viés para o estimador genérico tem-se: V[$\tilde{\beta}$|X] = E[($\tilde{\beta}$ -β) ($\tilde{\beta}$- β)'|X] (2a). Neste ponto, cabe escrever o estimador genérico de maneira a explicitar o termo de perturbação. Para isso, basta considerar que $\tilde{\beta} = H(X)'(X\beta + u) = H(X)'X\beta + H(X)'u(2b)$. Aplicando a condição $H(X)'X = I_{K+1}$ em (2b), a qual é equivalente a assumir ausência de viés para $\tilde{\beta}$, tem-se:$\tilde{\beta} = \beta + H(X)'u(2c)$. Combinando (2c) e (2a), tem-se: V[$\tilde{\beta}$|X] = E[$H(X)'u$ ($H(X)'u$)' |X] = E[$H(X)'uu'H(X)$|X]. A última passagem empregou a propriedade distributiva do operador de transposição, segundo a qual (AB)' = B'A', em que A e B são duas matrizes genéricas. Utilizando o fato de que a expectativa é condicional em X, V[$\tilde{\beta}$|X] = $H(X)$'E[uu'|X]$H(X)$ (1d). Aplicando as hipóteses MCRL5 e MCRL6, tem-se que E[uu'|X] = σ^2 o que, incorporado a (1d), leva a V[$\tilde{\beta}$|X] = σ^2H(x)'H(x) (1.e);

<u>Passo 3, comparando V[$\tilde{\beta}$|X] e V[$\hat{\beta}_{MQO}$|X]</u>

Como a variância do estimador de MQO é σ^2(X'X)$^{-1}$, tem-se V[$\tilde{\beta}$|X] ≥ V[$\hat{\beta}_{MQO}$|X], i.e., V[$\tilde{\beta}$|X] - V[$\hat{\beta}_{MQO}$|X] ≥ 0 se e somente se σ^2H(x)'H(x) - σ^2(X'X)$^{-1}$ ≥ 0, ou, de maneira equivalente, H(x)'H(x) - (X'X)$^{-1}$ ≥ 0, pois σ^2 é uma constante. Cabe, portanto, focar na expressão H(x)'H(x) - (X'X)$^{-1}$, procurando determinar seu sinal. Seguindo o apêndice E de Wooldridge (versão em inglês), recorre-se à manipulação a seguir.

3.i) H(x)'H(x) - (X'X)$^{-1}$ = H(x)'H(x) - I_{K+1}(X'X)$^{-1}$$I_{K+1}$ (4.a), dado que multiplicar uma matriz pela identidade não a altera.

3.ii)Sabemos, porém, que $H(X)'X=I_{K+1}$, o que é equivalente a $X'H(X)=I_{K+1}$ fato que, incorporado a 4.a conduz a: H(x)'H(x) - H(x)'X(X'X)$^{-1}$X'H(x), expressão esta que também pode ser escrita como H(x)'[I_N- X(X'X)$^{-1}$ X']H(x)

A matriz M = I_N- X(X'X)$^{-1}$ X' é denominada por "matriz geradora de resíduos" uma vez que [I_N- X(X'X)$^{-1}$ X']Y = Y - X(X'X)$^{-1}$X'Y = Y - X$\hat{\beta}_{MQO}$ = $\hat{u}$. Essa matriz tem duas propriedades importantes. Em primeiro lugar, ela é simétrica, ou seja, o elemento na i-ésima linha e na j-ésima coluna é equivalente ao elemento na j-ésima coluna e na i-ésima linha. O que pode ser representando formalmente como M = M'. Em segundo lugar, M é idempotente, o que significa que seu quadrado é equivalente à própria matriz, ou seja, M'M = M. De acordo com a seção A.7.2 de Greene (2003), uma matriz M com essas duas propriedades é tal que z'Mz = $\sum_{i=1}^{N} q_i^2$, sendo z uma matriz genérica e q uma variável genérica. O importante é que a forma quadrática z'Mz, composta com base em M sempre é equivalente a uma soma de quadrados, esta última sempre não-negativa. Dado que H(x)'[I_{K+1}- X(X'X)$^{-1}$ X']H(x) = H(x)'MH(x) é uma forma quadrática em M, pode-se concluir que H(x)MH(x)' ≥ 0. Recordando V[$\tilde{\beta}$|X] - V[$\hat{\beta}_{MQO}$|X] = σ^2H(x)MH(x)' e σ^2 é sempre positivo, fica, pois, demonstrado que V[$\tilde{\beta}$|X] ≥ V[$\hat{\beta}_{MQO}$|X].

Uma conclusão importante a ser retirada desta demonstração é a de que a propriedade de eficiência dos estimadores de MQO pressupõe a validade das hipóteses de homocedasticidade (MCRL 5) e de ausência de autocorrelação (MCRL 6), ambas referentes ao termo de perturbação. A violação de qualquer uma destas duas hipóteses (ou de ambas) implica na perda da propriedade de eficiência por parte dos estimadores de MQO – um tema tratado no curso de econometria II.