Esta aula

- Plano
  - Inclusão de Variáveis Irrelevantes
  - Omissão de Variáveis Relevantes
  - Revisão: Formas Funcionais
  - Testes de Especificação Funcional
  - Variáveis Proxy

- Bibliografia
  - Wooldridge, J. M. Introductory Econometrics: A modern Approach, 6th Ed.

Inclusão de Variáveis Irrelevantes

Variáveis Irrelevantes

- Considere o seguinte modelo populacional:

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

Com $\beta_3 = 0$

- Se estimarmos os parâmetros, obtemos:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$$

- Essas são estimativas não-enviesadas dos verdadeiros parâmetros populacionais.

Variáveis Irrelevantes

- Porém lembre-se que:

$$Var\left(\hat{\beta}_j\right) = \frac{\sigma^2}{SST_j\left(1 - R_j^2\right)}$$

- Dessa forma, se a variável irrelevante for correlacionada com as demais variáveis explicativas do modelo, aumentará a variância ou imprecisão na estimação dos parâmetros. Os testes de hipótese tenderão a rejeitar a significância de parâmetros relevantes com maior probabilidade.

Omissão de Variáveis Relevantes

Omissão de Variáveis Relevantes

- Considere o seguinte modelo populacional:

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

- Porém, por falta de conhecimento ou de dados, estima-se:

$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$$

- Obtemos uma estimativa não-enviesada do impacto de x1 em y?

Omissão de Variáveis Relevantes

- A seguinte relação se verifica:

$$\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \tilde{\delta}_1$$

- Em que delta1 é o impacto de x1 em x2 em uma regressão de x2 contra x1

- Portanto, tem-se:

$$\begin{aligned} \mathrm{E}(\tilde{\beta}_1) &= \mathrm{E}(\hat{\beta}_1 + \hat{\beta}_2 \tilde{\delta}_1) = \mathrm{E}(\hat{\beta}_1) + \mathrm{E}(\hat{\beta}_2)\tilde{\delta}_1 \\ &= \beta_1 + \beta_2 \tilde{\delta}_1, \end{aligned}$$

- E o Viés é dado por: $\mathrm{Bias}(\tilde{\beta}_1) = \mathrm{E}(\tilde{\beta}_1) - \beta_1 = \beta_2 \tilde{\delta}_1.$

Omissão de Variáveis Relevantes

TABLE 3.2 Summary of Bias in $\tilde{\beta}_1$ when x_2 Is Omitted in Estimating Eqution (3.40)

	$\mathbf{Corr}(x_1, x_2) > 0$	$\mathbf{Corr}(x_1, x_2) < 0$
$\beta_2 > 0$	Positive bias	Negative bias
$\beta_2 < 0$	Negative bias	Positive bias

Omissão de Variáveis Relevantes

- Caso Geral - Considere o seguinte modelo populacional:

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

- Porém, por falta de conhecimento ou de dados, estima-se:

$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + \tilde{\beta}_2 x_2.$$

- Obtemos uma estimativa não-enviesada do impacto de x1 em y?

Omissão de Variáveis Relevantes

- Se x1 e x2 não forem correlacionadas, tem-se:

$$E(\tilde{\beta}_1) = \beta_1 + \beta_3 \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)x_{i3}}{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2}$$

Revisão: Formas Funcionais

Linearidade

- Até agora, consideramos o modelo linear:

$$y = \beta_0 + \beta_1 x + u$$

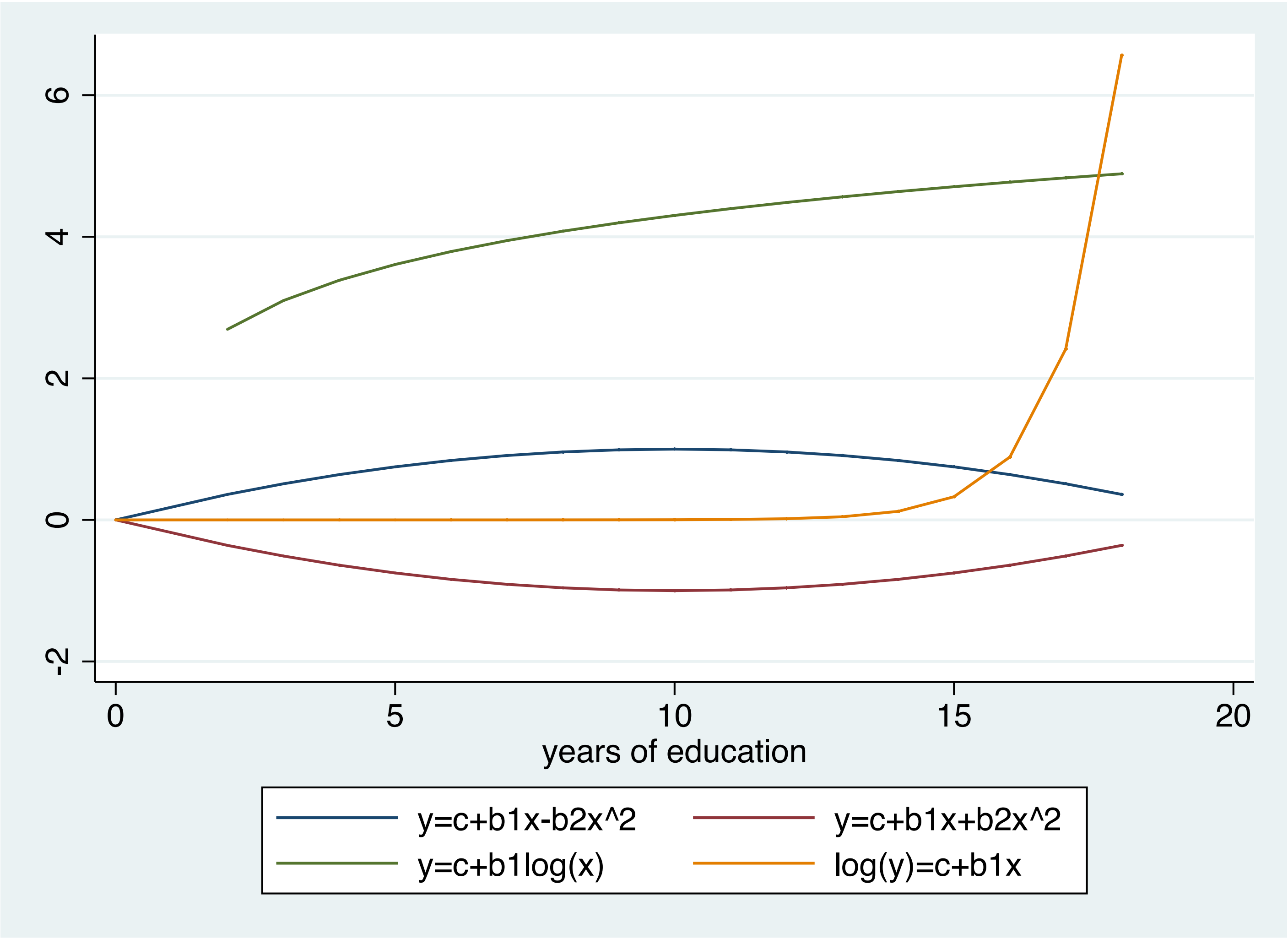
- Esse modelo preve um impacto constante β_1, também chamado de efeito marginal de x em y, que independe do valor inicial de x: $\Delta y = \beta_1 \Delta x$

- Em: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2.$

$$\Delta y = \beta_1 \Delta x_1 \text{ if } \Delta x_2 = 0$$

- β_1 é chamado de efeito parcial de x_1 em y

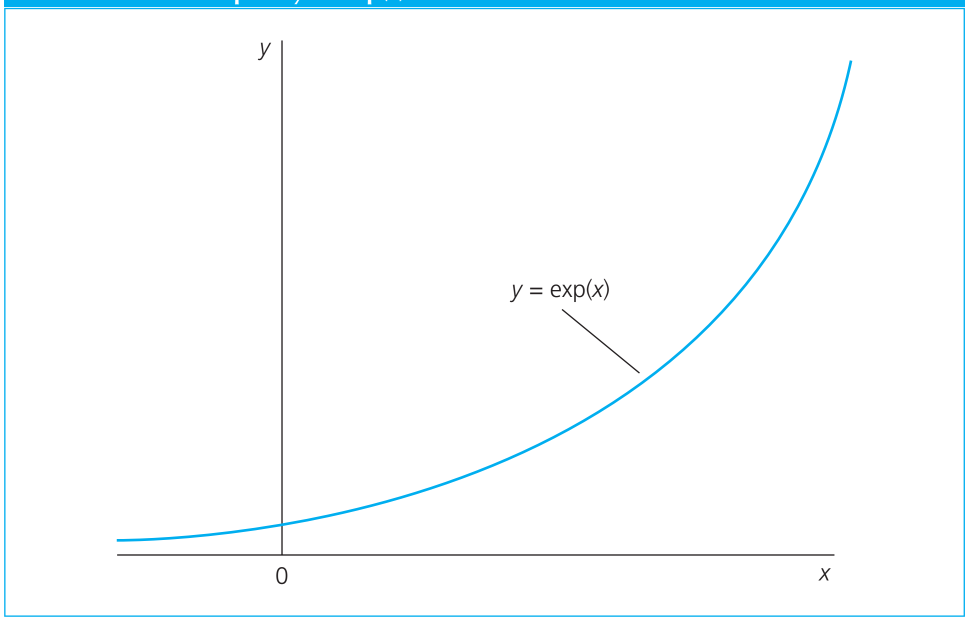
Linearidade



Linearidade

- É comum fazer-se a regressão de log(y) contra x. A interpretação do modelo faz mais sentido para muitos problemas em ciências sociais. Em nosso exemplo anterior, teríamos: $\log(wage) = \beta_0 + \beta_1 educ + u,$

FIGURE A.5 Graph of $y = \exp(x)$.



Linearidade

- Utilizando cálculo, é possível demonstrar que:

$$\log(x_1) - \log(x_0) \approx (x_1 - x_0)/x_0 = \Delta x/x_0$$

- Em $\log(wage) = \beta_0 + \beta_1 educ + u,$

  , multiplicando-se por 100 tem-se:

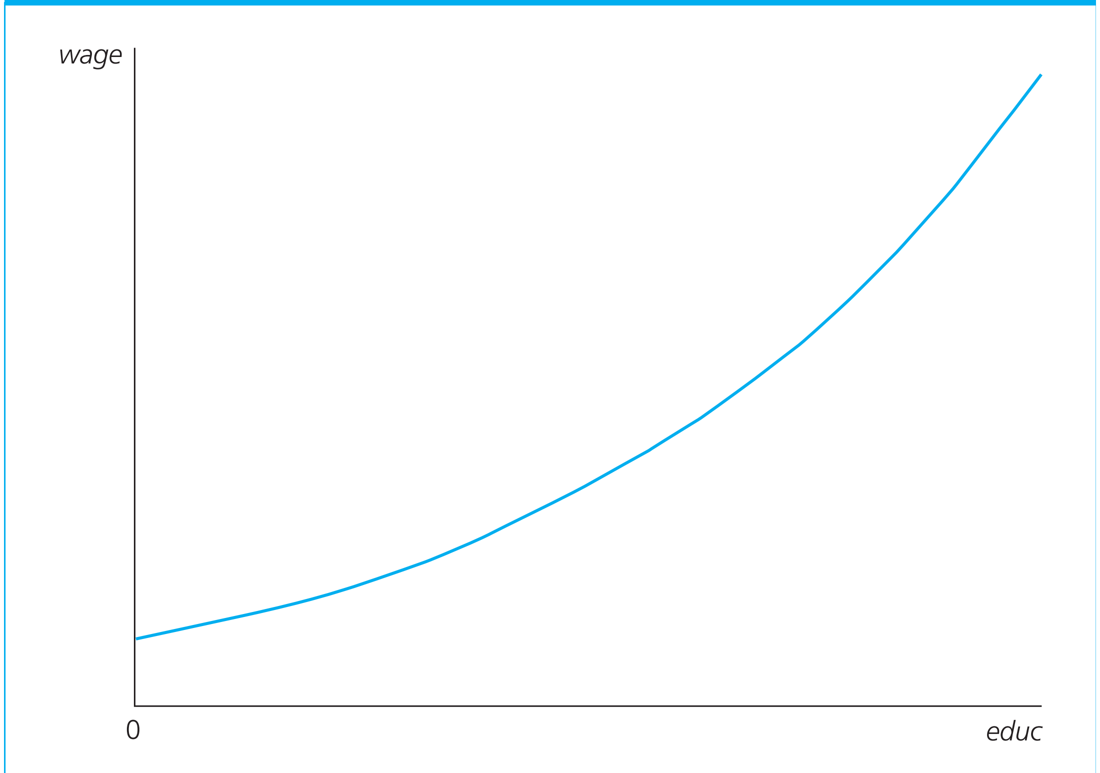
$$\%\Delta wage \approx (100 \cdot \beta_1)\Delta educ.$$

Linearidade

- Nesse caso, tem-se:

$$\%\Delta wage \approx (100 \cdot \beta_1)\Delta educ.$$

FIGURE 2.6 $wage = \exp(\beta_0 + \beta_1 educ)$, with $\beta_1 > 0$.

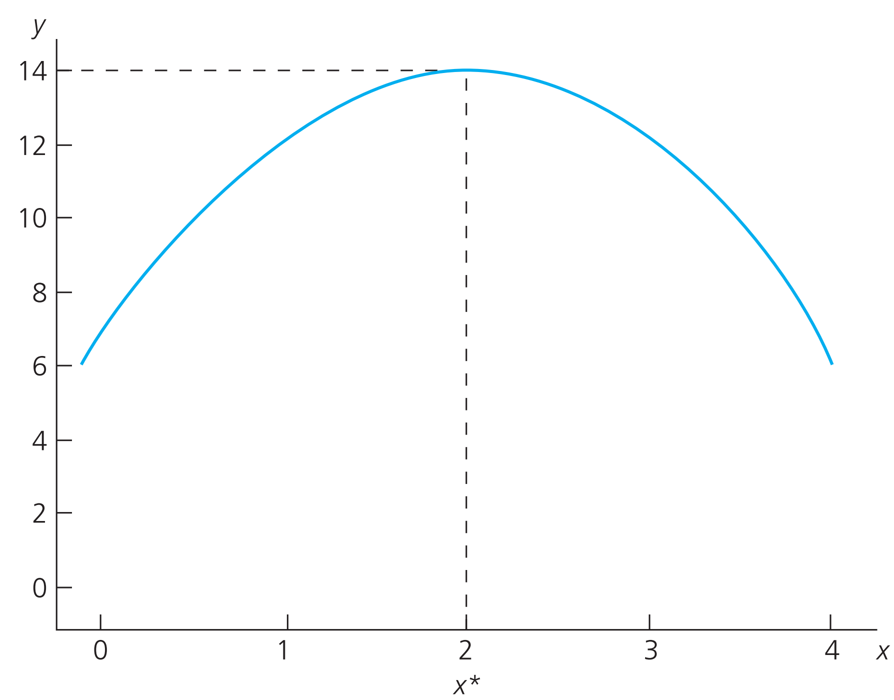


Linearidade

- Função quadrática:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2,$$

FIGURE A.3 Graph of $y = 6 + 8x - 2x^2$.



Linearidade

- Portanto, em

$$\log(y) = \beta_0 + \beta_1 \log(x)$$

- $\beta_1 = \partial \log(y) / \partial \log(x)$, sendo aproximadamente a elasticidade de y com relação x

Linearidade

- Elasticidade de y com relação a x é definida pela variação percentual de y dividida pela variação percentual de x:

$$\frac{\Delta y}{\Delta x} \cdot \frac{x}{y} = \frac{\% \Delta y}{\% \Delta x}$$

- (Cálculo) A elasticidade pode ser aproximada por:

$$\Delta \log(y) / \Delta \log(x)$$

Linearidade

- Nesse caso, tem-se:

TABLE 2.3 Summary of Functional Forms Involving Logarithms

Model	Dependent Variable	Independent Variable	Interpretation of β_1
Level-level	y	x	$\Delta y = \beta_1 \Delta x$
Level-log	y	$\log(x)$	$\Delta y = (\beta_1/100)\%\Delta x$
Log-level	$\log(y)$	x	$\%\Delta y = (100\beta_1)\Delta x$
Log-log	$\log(y)$	$\log(x)$	$\%\Delta y = \beta_1 \%\Delta x$

Testes de Especificação Funcional

Testes de Especificação Funcional

- Regression Especification Error Test (RESET) – Ramsey's (1969)
- Considere o seguinte modelo sob H0:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u$$

- Queremos testar se o quadrado das variáveis explicativas e produtos cruzados deveriam ser incluídos no modelo.
- Considere o seguinte modelo alternativo, sob H1:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \delta_1 \hat{y}^2 + \delta_2 \hat{y}^3 + error.$$

Testes de Especificação Funcional

- Ramsey's (1969) RESET consiste no teste F de H1 contra H0:

H_0: $\delta_1 = \delta_2 = 0$

H_1: H_0 *não é verdadeira*

Considerar o valor critíco considerando-se a distribuição

$F_{2,n-k-1-2}$

Teste de Hipótese

- As etapas do teste são:

1. Escrever as hipóteses alternativas e nulas
2. Escolher o nível de significância do teste α
3. Calcular a estatística F, conhecida como a **estatística do teste**
4. Encontrar o **valor crítico** do teste F^*,
5. Decidir: Se o valor de F for maior do que o de F^*, rejeitar H_0 com um nível de confiança de $1-\alpha$

Testes de Especificação Funcional

- Testes de Nonnested Models
- Considere o seguinte modelo sob H0:

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

- Considere agora o seguinte modelo alternativo, sob H1:

$$y = \beta_0 + \beta_1 \log(x_1) + \beta_2 \log(x_2) + u,$$

- Uma possibilidade seria agrugar todas as variáveis em um único modelo:

$$y = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 \log(x_1) + \gamma_4 \log(x_2) + u.$$

Testes de Especificação Funcional

- Mizon e Richard (1968) Test consite no teste F:

  H_0: $\gamma_3 = \gamma_4 = 0$

  H_1: *H_0 não é verdadeira*

  Se trocarmos o modelo sob H0 pelo modelo sob H1, poderíamos teríamos as hipóteses:

  H_0: $\gamma_1 = \gamma_2 = 0$

  H_1: *H_0 não é verdadeira*

Testes de Especificação Funcional

- Considere novamente o seguinte modelo sob H0:

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

E o seguinte modelo alternativo, sob H1:

$$y = \beta_0 + \beta_1 \log(x_1) + \beta_2 \log(x_2) + u,$$

- Tome os valores estimados $\hat{\hat{y}}$ sob H1. Davidson e MacKinnon propuseram o seguinte teste t nessa variável:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \theta_1 \hat{\hat{y}} + error.$$

H_0: $\theta_1 = 0$

H_1: *H_0 não é verdadeira*

Variável Proxy

- Suponha que o modelo populacional seja:

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 abil + u.$$

Sendo ability sendo não-observável

- De uma forma geral, considere o seguinte modelo populacional:

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

Sendo x_3^* não-observável.

Variável Proxy

- Se x_3^* puder ser escrita como uma função da variável x_3 :

$$x_3^* = \delta_0 + \delta_3 x_3 + v_3,$$

E substituindo-se dentro do modelo populacional, obtem-se:

$$y = (\beta_0 + \beta_3 \delta_0) + \beta_1 x_1 + \beta_2 x_2 + \beta_3 \delta_3 x_3 + u + \beta_3 v_3.$$

Ou, reescrevendo-se:

$$y = \alpha_0 + \beta_1 x_1 + \beta_2 x_2 + \alpha_3 x_3 + e,$$

Variável Proxy

- Se x_3^* for uma função de todas as variáveis explicativas, tem-se:

$$x_3^* = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \delta_3 x_3 + v_3.$$

Nesse caso, substituindo-se dentro do modelo populacional, obtém-se:

$$\begin{aligned} y = {} & (\beta_0 + \beta_3\delta_0) + (\beta_1 + \beta_3\delta_1)x_1 + (\beta_2 + \beta_3\delta_2)x_2 \\ & + \beta_3\delta_3 x_3 + u + \beta_3 v_3, \end{aligned}$$

Variável Proxy

- Em trabalhos empíricos, é comum utilizar-se a primeira defasagem da variável dependente como variável proxy:

$$crime = \beta_0 + \beta_1 unem + \beta_2 expend + \beta_3 crime_{-1} + u,$$

Obrigada!