Econometria Aplicada

Correlação Serial

João Ricardo Costa Filho

“The most important questions of life are, for the most part, really only problems in probability."

Laplace (1812)

“In God we trust. All others must bring data."

William Edwards Deming

A regressão linear

Motivação (tudo começa com uma pergunta)

Qual é a relação entre capital humano e o desenvolvimento econômico?

Referências

- Como responder a questão que motivou a nossa análise?

Referências

- Como responder a questão que motivou a nossa análise?
  - Desenvolvimento econômico, capital humano e instituições: Acemoglu, Gallego, and Robinson (2014).
  - Econometria: capítulo 6 de Wooldridge (2006).

Variância do Estimador $\hat{\beta}_j$

Regressão múltipla

Regressão múltipla

Seja $\widehat{\boldsymbol{\beta}}$ um vetor de dimensão $k \times 1$ tal que:

Regressão múltipla

Seja $\widehat{\boldsymbol{\beta}}$ um vetor de dimensão $k \times 1$ tal que:

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

onde $\mathbf{X}$ é uma matriz de dimensão $n \times k$ e $\mathbf{y}$ um vetor de dimensão $n \times 1$.

Regressão múltipla

Seja $\widehat{\boldsymbol{\beta}}$ um vetor de dimensão $k \times 1$ tal que:

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

onde $\mathbf{X}$ é uma matriz de dimensão $n \times k$ e $\mathbf{y}$ um vetor de dimensão $n \times 1$. A variância do vetor de estimadores é dada por:

Regressão múltipla

Seja $\widehat{\boldsymbol{\beta}}$ um vetor de dimensão $k \times 1$ tal que:

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

onde $\mathbf{X}$ é uma matriz de dimensão $n \times k$ e $\mathbf{y}$ um vetor de dimensão $n \times 1$. A variância do vetor de estimadores é dada por:

$$\text{Var}\left[\widehat{\boldsymbol{\beta}}\right] = \text{Var}\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}\right].$$

Variância do Estimador $\hat{\beta}_j$

Podemos substituir a $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ no resultado anterior para obtermos:

Variância do Estimador $\hat{\beta}_j$

Podemos substituir a $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ no resultado anterior para obtermos:

$$\text{Var}\left[\widehat{\boldsymbol{\beta}}\right] = \text{Var}\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})\right].$$

Variância do Estimador $\hat{\beta}_j$

Podemos substituir a $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ no resultado anterior para obtermos:

$$\text{Var}\left[\widehat{\boldsymbol{\beta}}\right] = \text{Var}\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})\right].$$

Expandindo, temos:

$$\text{Var}\left[\widehat{\boldsymbol{\beta}}\right] = \text{Var}\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\right]$$

Variância do Estimador $\hat{\beta}_j$

Podemos substituir a $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ no resultado anterior para obtermos:

$$\text{Var}\left[\widehat{\boldsymbol{\beta}}\right] = \text{Var}\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})\right].$$

Expandindo, temos:

$$\text{Var}\left[\widehat{\boldsymbol{\beta}}\right] = \text{Var}\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\right] \iff$$

$$\text{Var}\left[\widehat{\boldsymbol{\beta}}\right] = \text{Var}\left[\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\right].$$

Variância do Estimador $\hat{\beta}_j$

Como $\boldsymbol{\beta}$ é um vetor de constantes, temos:

Variância do Estimador $\hat{\beta}_j$

Como $\boldsymbol{\beta}$ é um vetor de constantes, temos:

$$\text{Var}\left[\widehat{\boldsymbol{\beta}}\right] = \text{Var}\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\right].$$

Variância do Estimador $\hat{\beta}_j$

Como $\boldsymbol{\beta}$ é um vetor de constantes, temos:

$$\mathsf{Var}\left[\widehat{\boldsymbol{\beta}}\right] = \mathsf{Var}\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\right].$$

Agora, lembre-se que $\mathsf{Var}(\mathbf{AX}) = \mathbf{A}\mathsf{Var}(\mathbf{X})\mathbf{A}'$

Variância do Estimador $\hat{\beta}_j$

Como $\boldsymbol{\beta}$ é um vetor de constantes, temos:

$$\text{Var}\left[\widehat{\boldsymbol{\beta}}\right] = \text{Var}\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\right].$$

Agora, lembre-se que $\text{Var}(\mathbf{AX}) = \mathbf{A}\text{Var}(\mathbf{X})\mathbf{A}'$ e como $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ é uma matriz de constantes, o resultado anterior torna-se "sanduíche":

Como $\boldsymbol{\beta}$ é um vetor de constantes, temos:

$$\text{Var}\left[\widehat{\boldsymbol{\beta}}\right] = \text{Var}\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\right].$$

Agora, lembre-se que $\text{Var}(\mathbf{AX}) = \mathbf{A}\text{Var}(\mathbf{X})\mathbf{A}'$ e como $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ é uma matriz de constantes, o resultado anterior torna-se "sanduíche":

$$\text{Var}\left[\widehat{\boldsymbol{\beta}}\right] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{Var}(\boldsymbol{\varepsilon})\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]',$$

Finalmente, dado que $(AB)' = B'A'$, podemos reescrever $\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]'$ como $\mathbf{X}\left[(\mathbf{X}'\mathbf{X})^{-1}\right]'$, e obtemos:

$$\text{Var}\left[\widehat{\boldsymbol{\beta}}\right] = (\mathbf{X}'\mathbf{X})^{-1}\text{Var}(\boldsymbol{\varepsilon})$$

Variância do Estimador sem correlação serial

Assuma que $(\mathbf{X}'\mathbf{X})^{-1}$ seja dada pela seguinte matriz:

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix},$$

e que a variância do vetor de erros $\boldsymbol{\varepsilon}$ é:

$$\mathsf{Var}\left[\boldsymbol{\varepsilon}\right] = \sigma^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Variância do Estimador sem correlação serial

Substituindo as definições anteriores na forma geral, temos:

$$\text{Var}\left[\widehat{\boldsymbol{\beta}}\right] = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}.$$

Variância do Estimador com correlação serial

Agora, trabalhemos com o caso com correlação serial.

Variância do Estimador com correlação serial

Agora, trabalhemos com o caso com correlação serial. A matriz de covariância do termo de erro, $\boldsymbol{\varepsilon}$, deixa de ser diagonal:

$$\mathrm{Cov}\left[\boldsymbol{\varepsilon}\right] = \begin{bmatrix} \sigma^2 & \mathrm{Cov}(\varepsilon_1, \varepsilon_2) \\ \mathrm{Cov}(\varepsilon_1, \varepsilon_2) & \sigma^2 \end{bmatrix}.$$

Variância do Estimador com correlação serial

Agora, trabalhemos com o caso com correlação serial. A matriz de covariância do termo de erro, ε, deixa de ser diagonal:

$$\text{Cov}\left[\boldsymbol{\varepsilon}\right] = \begin{bmatrix} \sigma^2 & \text{Cov}(\varepsilon_1, \varepsilon_2) \\ \text{Cov}(\varepsilon_1, \varepsilon_2) & \sigma^2 \end{bmatrix}.$$

Por definição, o coeficiente de correlação $\rho = \frac{\text{Cov}[\varepsilon_1, \varepsilon_2])}{\sqrt{\text{Var}[\varepsilon_1]\text{Var}[\varepsilon_2]}}$

Variância do Estimador com correlação serial

Agora, trabalhemos com o caso com correlação serial. A matriz de covariância do termo de erro, $\boldsymbol{\varepsilon}$, deixa de ser diagonal:

$$\text{Cov}\left[\boldsymbol{\varepsilon}\right] = \begin{bmatrix} \sigma^2 & \text{Cov}(\varepsilon_1, \varepsilon_2) \\ \text{Cov}(\varepsilon_1, \varepsilon_2) & \sigma^2 \end{bmatrix}.$$

Por definição, o coeficiente de correlação $\rho = \frac{\text{Cov}[\varepsilon_1, \varepsilon_2])}{\sqrt{\text{Var}[\varepsilon_1]\text{Var}[\varepsilon_2]}}$ e assumindo que $\text{Var}\left[\varepsilon_1\right] = \text{Var}\left[\varepsilon_2\right] = \sigma^2$,

Variância do Estimador com correlação serial

Agora, trabalhemos com o caso com correlação serial. A matriz de covariância do termo de erro, $\boldsymbol{\varepsilon}$, deixa de ser diagonal:

$$\text{Cov}\left[\boldsymbol{\varepsilon}\right] = \begin{bmatrix} \sigma^2 & \text{Cov}(\varepsilon_1, \varepsilon_2) \\ \text{Cov}(\varepsilon_1, \varepsilon_2) & \sigma^2 \end{bmatrix}.$$

Por definição, o coeficiente de correlação $\rho = \frac{\text{Cov}[\varepsilon_1,\varepsilon_2])}{\sqrt{\text{Var}[\varepsilon_1]\text{Var}[\varepsilon_2]}}$ e assumindo que $\text{Var}\left[\varepsilon_1\right] = \text{Var}\left[\varepsilon_2\right] = \sigma^2$, a matriz de covariância pode ser reescrita como:

$$\text{Cov}(\boldsymbol{\varepsilon}) = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

Variância do Estimador com correlação serial

Finalmente, temos:

$$\text{Var}\left[\widehat{\boldsymbol{\beta}}\right] = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix},$$

Variância do Estimador com correlação serial

Finalmente, temos:

$$\operatorname{Var}\left[\widehat{\boldsymbol{\beta}}\right] = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix},$$

que é uma matriz 2×2 em que a variância do estimador depende explicitamente de ρ.

Variância do Estimador com correlação serial

Finalmente, temos:

$$\text{Var}\left[\widehat{\boldsymbol{\beta}}\right] = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix},$$

que é uma matriz 2×2 em que a variância do estimador depende explicitamente de ρ. Em geral, a correlação serial aumenta a variância do estimador,

Variância do Estimador com correlação serial

Finalmente, temos:

$$\text{Var}\left[\widehat{\boldsymbol{\beta}}\right] = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix},$$

que é uma matriz 2×2 em que a variância do estimador depende explicitamente de ρ. Em geral, a correlação serial aumenta a variância do estimador, mas o impacto exato depende da magnitude e do sinal de ρ.

Referências i

Acemoglu, Daron, Francisco A Gallego, and James A Robinson. 2014. "Institutions, Human Capital, and Development." *Annual Review of Economics* 6 (1): 875–912.

Wooldridge, Jeffrey M. 2006. *Introdução à Econometria: Uma Abordagem Moderna*. Cengage Learning.