Econometria I

Exame 1 (Gabarito)

26/10/2023

1. a. O estimador de OLS de β, b, é

$$\begin{aligned} b &= (\mathbf{y_2'y_2})^{-1}\mathbf{y_2'y_1} \\ &= (\mathbf{y_2'y_2})^{-1}\mathbf{y_2'}(\mathbf{y_2}\beta + \varepsilon_1) \\ &= \beta + \left(\frac{\mathbf{y_2'y_2}}{n}\right)^{-1}\left(\frac{\mathbf{y_2'y_1}}{n}\right) \end{aligned}$$

Dado que y_{2i} são *iid*, podemos aplicar a lei dos grandes números e encontrar: $\mathbf{y_2'y_2}/n \to E(y_{2i}^2)$. Ademais, temos:

$$\begin{aligned} \frac{\mathbf{y_2'}\boldsymbol{\varepsilon_1}}{n} &= \frac{(\alpha\mathbf{x}' + \varepsilon_2')\boldsymbol{\varepsilon_1}}{n} \\ &= \alpha\frac{\mathbf{x}'\boldsymbol{\varepsilon_1}}{n} + \frac{\boldsymbol{\varepsilon_2'}\boldsymbol{\varepsilon_1}}{n} \\ &\xrightarrow{p} \alpha E(x_i\varepsilon_{1i}) + E(\varepsilon_{2i}\varepsilon_{1i}) \end{aligned}$$

Por suposição, x_i é uma variável exógena, o que significa $E(x_i\varepsilon_{1i}) = 0$. Por suposição, $E(\varepsilon_{2i}\varepsilon_{1i}) = \sigma_{12}$. Portanto, em geral, este estimador não é consistente.

b. Para recuperar a consistência, precisamos impor a restrição $\sigma_{12} = 0$, ou seja, não há correlação entre os termos de erro em ambas as equações: tal restrição também garantiria que não há problemas de endogeneidade, o que tornaria o OLS o estimador preferido (e consistente).

c. O estimador de OLS de α, $\hat{\alpha}$, é

$$\begin{aligned} \hat{\alpha} &= (\mathbf{x'x})^{-1}\mathbf{x'y_2} \\ &= (\mathbf{x'x})^{-1}\mathbf{x}'(\mathbf{x}\alpha + \varepsilon_2) \\ &= \alpha + \left(\frac{\mathbf{x'x}}{n}\right)^{-1}\left(\frac{\mathbf{x}'\boldsymbol{\varepsilon_2}}{n}\right) \end{aligned}$$

Dado que x_i é *iid*, podemos aplicar a lei dos grandes números e encontrar: $\mathbf{x'x}/n \xrightarrow{p} E(x_i^2)$. Além disso, também temos:$\mathbf{x}'\boldsymbol{\varepsilon_2}/n \xrightarrow{p} E(x_i\varepsilon_{2i}) = 0$ dado que x_i é exógeno. Concluímos que o estimador $\hat{\alpha}$ é consistente.

d. O estimador é consistente.

e. O estimador de IV de β usando $\mathbf{x}$ como instrumento é:

$$\begin{aligned} b_{IV} &= (\mathbf{x'y_2})^{-1}\mathbf{x'y_1} \\ &= (\mathbf{x'y_2})^{-1}\mathbf{x'}(\mathbf{y_2}\beta + \varepsilon_1) \\ &= \beta + \left(\frac{\mathbf{x'y_2}}{n}\right)^{-1}\left(\frac{\mathbf{x'}\varepsilon_\mathbf{1}}{n}\right) \end{aligned}$$

Note que $\mathbf{x'y_2}/n \xrightarrow{p} E(x_i y_{i2}) = \alpha$ (conforme mostrado anteriormente) e $\mathbf{x'}\varepsilon_\mathbf{1}/n \xrightarrow{p} E(x_i\varepsilon_1) = 0$ pela exogeneidade de x. Portanto, o estimador IV é consistente. Ao comparar o estimador de OLS e o estimador de IV, o estimador de IV deve ser usado quando há endogeneidade para fornecer um estimador consistente; caso contrário, o estimador de OLS deve ser usado.

2. a. A matriz de covariância convencionalmente estimada para o estimador LS é $\sigma^2(\mathbf{X'X})^{-1}$, o que é inadequado, pois a matriz adequada é $\sigma^2(\mathbf{X'X})^{-1}(\mathbf{X'\Omega X})(\mathbf{X'X})^{-1}$. Note que podemos escrever a diferença como

$$\mathbf{D} = \frac{\sigma^2}{n}\left(\frac{\mathbf{X'X}}{n}\right)^{-1}\left[\frac{\mathbf{X'X}}{n} - \frac{\mathbf{X'\Omega X}}{n}\right]\left(\frac{\mathbf{X'X}}{n}\right)^{-1}$$

a qual é positiva, indicando que o estimador de OLS é ineficiente. Entretanto, caso a hipótes de exogeneidade $E[\varepsilon|\mathbf{X}] = 0$ seja satisfeita, o estimador de OLS ainda é não-viesado e consistente.

b. Como $\mathbf{\Omega}$ é uma matriz simétrica definida positiva, se pode ser fatorada em $\mathbf{\Omega} = \mathbf{C\Lambda C'}$, onde as colunas de $\mathbf{C}$ são os vetores característicos de $\mathbf{\Omega}$ e as raízes características de $\mathbf{\Omega}$ são dispostas na matriz diagonal Λ.Defina $\mathbf{\Lambda}^{1/2}$ como a matriz diagonal com o i-ésimo elemento diagonal $\sqrt{\lambda_i}$ e seja $\mathbf{T} = \mathbf{C\Lambda^{1/2}}$. Então, $\mathbf{TT'} = \mathbf{\Omega}$. Também defina $\mathbf{P'} = \mathbf{C\Lambda^{-1/2}}$, então $\mathbf{\Omega}^{-1} = \mathbf{P'P}$. Se pré-multiplicarmos $\mathbf{y} = \mathbf{X}\beta + \varepsilon$ por P, obtemos

$$\mathbf{Py} = \mathbf{PX}\beta + \mathbf{P}\varepsilon$$

Como assumimos que $\mathbf{\Omega}$ é conhecida, podemos definir

$$\begin{aligned} \hat{\beta} &= (\mathbf{X'P'PX})^{-1}\mathbf{X'P'Py} \\ &= (\mathbf{X'\Omega^{-1}X})^{-1}\mathbf{X'\Omega^{-1}y} \end{aligned}$$

como o estimador GLS.

c. Sabemos que o estimador de OLS é consistente mas ineficiente. Vejamos a consistência do estimador de GLS: Exigimos que os dados transformados $\mathbf{PX} = \mathbf{X}_*$, e não os dados originais $\mathbf{X}$, sejam bem comportados. O estimador GLS é consistente se plim $(1/n)\mathbf{X}_*\mathbf{X}_* = \mathbf{Q}_*$, onde $\mathbf{Q}_*$ é uma matriz definida positiva finita. Fazendo a substituição, vemos que isso implica

$$\text{plim } [(1/n)\mathbf{X'\Omega}^{-1}\mathbf{X}] = \mathbf{Q}_*^{-1}$$

Em relação à eficiência, note que a covariancia conditional do termo de erro transformado $\mathbf{P}\varepsilon = \varepsilon_*$ é dada por

$$E[\varepsilon_*\varepsilon_*'|\mathbf{X}] = E[\mathbf{P}\varepsilon\varepsilon'\mathbf{P'}|\mathbf{PX}] = \mathbf{P}\sigma^2\mathbf{\Omega P'} = \sigma^2\mathbf{I}$$

Portanto, aplicando a transformação, recuperamos a hipótese de homoscedasticidade e eficiência segue do teorema de Gauss-Markov.

d. Podemos escrever o estimador de LS como

$$\mathbf{b} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\sum_{i=1}^{n}\mathbf{x}_i\varepsilon_i$$

Então, temos

$$Var[\mathbf{b}|\mathbf{X}] = \frac{1}{n}\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}\left(\frac{\mathbf{X}'(\sigma^2\boldsymbol{\Omega})\mathbf{X}}{n}\right)\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}$$

Portanto, a variância assintótica é dada por:

$$Asy.Var[\mathbf{b}] = \frac{1}{n}\mathbf{Q}^{-1}\left[\text{plim }\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2\mathbf{x}_i\mathbf{x}_i'\right]\mathbf{Q}^{-1} = \frac{1}{n}\mathbf{Q}^{-1}\mathbf{Q}^*\mathbf{Q}^{-1}$$

onde assumimos que $\mathbf{Q}^* = \text{plim }\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2\mathbf{x}_i\mathbf{x}_i'$. O estimador de $\mathbf{Q}^*$ robusto à Heteroscedasticidade proposto por White (1980) é dado por

$$\mathbf{W}_{het} = \frac{1}{n}\sum_{i=1}^{n}e_i^2\mathbf{x}_i\mathbf{x}_i',$$

onde e_i é o resíduo de LS, $y_i - \mathbf{x}_i'\mathbf{b}$. Portanto, temos

$$Est.Asy.Var[\mathbf{b}] = n(\mathbf{X}'\mathbf{X})^{-1}\mathbf{W}_{het}(\mathbf{X}'\mathbf{X})^{-1}$$

e. Considere o seguinte teste:

$$\begin{aligned} H_0 :& \sigma_i^2 = E[\varepsilon_i^2|\mathbf{X}] = \sigma^2 \text{ for all } i \\ H_1 :& \text{Not } H_0 \end{aligned}$$

O Teste Breusch-Pagan considera $E[\varepsilon^2|\mathbf{X}]$ uma função linear das variáveis explicativas:

$$E[\varepsilon^2|\mathbf{X}] = \delta_0 + \delta_1 x_1 + \ldots + \delta_k x_k + \nu$$

Portanto, a hipótese nula de homocedasticidade é

$$\begin{aligned} H_0 :& \delta_1 = \delta_2 = \ldots = \delta_k = 0 \\ H_1 :& \text{Pelo menos de } \delta_j \text{ não é igual a } 0 \end{aligned}$$

Assumindo as suposições clássicas, H_0 pode ser testado usando o teste usual F para a significância conjunta de todas as variáveis explicativas. A equação que estimamos é

$$e_i^2 = \delta_0 + \delta_1 x_{1i} + \ldots + \delta_k x_{ki} + \varepsilon_i'$$

As etapas para realizar o teste são as seguintes:
1. Estime $\mathbf{y} = \mathbf{x}\beta + \varepsilon$ e encontre e
2. Estime $\mathbf{e}'\mathbf{e} = \mathbf{x}\delta + \varepsilon'$
3. Execute o teste F para $\delta = 0$

3. a. A resposta a esta questão deveria conter duas partes: a primeira mostrando (ou explicando) que o estimador de desvio das médias

$$\mathbf{b}^{within} = [S_{xx}^{within}]^{-1}S_{xy}^{within}$$

onde $S_{xx}^{within} = \sum_{i=1}^{n}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'$ e $S_{xy}^{within} = \sum_{i=1}^{n}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{y}_{it} - \bar{\mathbf{y}}_i)'$ controla para os efeitos heterogêneos fixos no tempo. Segundo, assumindo exogeneidade estrita

$$E[\varepsilon_{it}|\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, c_i] = E[\varepsilon_{it}|\mathbf{X}_i, c_i] = 0$$

e dados bem-comportados

$$\text{plim}\ \frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\mathbf{X_i'X_i} = \bar{\mathbf{Q}}_n,$$

mostrar que o estimador é não-viesado e consistente segue o mesmo raciocínio básico do estimador de OLS visto anteriormente.

b. Neste caso, o modelo de efeitos aleatórios é o mais apropriado, dado que o pressuposto $E[\alpha_i \mathbf{x}_i] = 0$ é satisfeito. Dessa forma, o modelo de efeitos aleatórios torna-se mais atraente, dado que o número de parâmetros a serem estimados é consideravelmente menor, apesar de ambos serem estimadores consistentes.

4. Essa questão permite vários caminhos para a resposta. O ponto importante é o aluno explicitar de forma clara e precisa a(s) equação(ões) regressão proposta, além dos pressupostos do modelo de regressão clássico e possíveis violações. Um exemplo de estratégia a ser seguido é o modelo de variáveis instrumentais, onde uma variável dummy indicando qualificação para o programa de crédito poderia ser usada como instrumento para a variável que indica a participação efetiva, a qual é provavelmente endógena devido à viés de seleção.