lista teorica 2

Miguel Sallum

06/07/2021

Questão 1

Considere o modelo de regressão linear simples, $y_t = \alpha + \beta x_t + u_t$ na presença de endogeneidade, isto é, $E(ui|xi) \neq 0$.Responda os itens a seguir:

a Apresente as condições que uma variável instrumental, z_i, deve satisfazer e derive o estimador de Variáveis Instrumentais β^{IV}.

b Apresente o procedimento necessario para a obtenção do estimador de Mínimos Quadrados em Dois Estágios.

c Verifique que todos os estimadores apresentados nos itens anteriores sao equivalentes. Mostre que o estimador β^{IV} (e, obviamente, os demais) é consistente. Derive a distribuição assintótica de β^{IV}.

d O que acontece com o estimador β^{IV} quando $cov(z1, x1) \to 0$? E com a sua variância? Interprete esse resultado.

Questão 2

Derive o estimador de Variaveis Instrumentais e de Mínimos Quadrados em Dois Estágios para o caso de regressão linear múltipla. Mostre a propriedade de consistência e normalidade assintótica desses estimadores.

Questão 3

Com base no modelo de regressão com dados em painel responda:

a Compare uma base de dados em painel com uma base de dados em cross-section. Quais as vantagens em utilizar painel?

b Escreva um modelo de regressao linear simples com dados em painel, onde y_{it} é a variável dependente e x_{it} a independente. Estenda para o caso de regressão múltipla.

c Escreva o modelo de regressao com efeitos fixos, em que $\alpha_1, \cdots \alpha_n$ sao tratados como interceptos. Associe o termo α_i com problema de variável omitida (ou não observada). Apresente também uma intuicão gráfica para esse termo.

d Utilizando variáveis binárias, apresente uma forma equivalente para o modelo de regressão com efeitos fixos visto no item anterior. Nesse caso, os parâmetros poderiam ser estimados por OLS?

e Defina as seguintes médias (no tempo): $\bar{y}_t = \frac{1}{T}\sum_t y_{it}$ e $\bar{x}_t = \frac{1}{T}\sum_t x_{it}$. Transforme o modelo da letra (b) em termos de desvios de y_{it} e x_{it} das respectivas médias (no tempo). Argumente a favor da estimação via OLS nos dados assim transformados.

f Generalize o modelo de regressão com efeitos fixos, apresentado no item (c), para incluir efeito fixo temporal

Questão 3.5

Respondam as questões a seguir: (a) Um instrumento fraco tera boas propriedades assintóticas. **falso** (b) O uso de uma variável proxy e o método de variável instrumental são duas possiveis formas de lidar o vies de variável omitida. **verdadeiro** (c) O estimador de variável instrumental será não viesado sob as hipóteses usuais associadas a esse estimador. **falso** (d) Uma boa variável instrumental deve ser uma variável exógena excluída da equação estrutural que tenha alguma correlação com a explicativa endógena. **verdadeiro** (e) O problema de endogeneidade surge quando as variáveis explicativas são correlacionadas entre si. **falso** (f) Sob a hipótese de ausência de correlação entre o efeito fixo não observado e as variaveis independentes, o estimador de efeitos aleatórios será o mais adequado relativamente ao estimador de efeitos fixos. **verdadeiro** (g) Para dois períodos de tempo, os estimadores de efeitos fixos, primeiras diferenças e efeitos aleatórios são idênticos. **falso** (h) O estimador de primeiras diferenças exige que as variáveis de controle utilizadas na análise variem ao longo do tempo. **verdadeiro** (i) A existência de 2 grupos e períodos de tempo distintos é suficiente para a utilização do método de diferenças-em-diferenças. **falso**

Questão 4

Suponha que voce possua os seguintes dados anuais sobre o número médio de filhos por domicílio em duas localidades, onde cada localidade é composta por 1000 domicílios:

```r
data <- tibble(
  ano = c(1980, 1985, 1990, 1995, 2000),
  A = c(3.7, 4.4, 3.8, 5.1, 5.8),
  B = c(2.3, 3.3, 3.2, 3.1, 3.1)
)
```

Em 1992 o prefeito de B implementou amplo programa de planejamento familiar e pediu que voce estimasse o impacto do programa. a Calcule o estimador de diferenças-em-diferenças do impacto do programa, definindo $t_0 = 1985$ e $t_1 = 1995$.

b Quais os aspectos relevantes para escolher o período pós-intervenção(t_1)?

c Em um gráfico com o número de filhos no eixo vertical (ordenadas) e o tempo no eixo horizontal (abcissas), desenhe um gráfico que inclua (i) as curvas de tendência do número de filhos por domicílio em cada localidade; (ii) o instante da intervenção; (iii) a magnitude do efeito estimado por diferenças-em-diferenças em 1995.

d Quais as condições que deveriam ser satisfeitas para o estimador de diferenças-em-diferenças ser uma boa estimativa do efeito causal da intervenção? Você considera que neste caso voce possui uma boa resposta para seu efeito causal? Justifique.

Questão 5

The following exercises guide you through the process of estimating and interpreting a simple regression discontinuity.

$$Y_i = \begin{cases} \alpha_0 + \alpha_1 X_i + \epsilon_i & for\ X_i < c \\ \beta_0 + \beta_1 X_i + \epsilon_i & for\ X_i \geq c \end{cases}$$

a In terms of $\alpha_0, \alpha_1, \beta_0, \beta_1$ e c, what is the the value of the RD causal effect?

$$LATE_{RD} = (\beta_0 - \alpha_0) + (\beta_1 - \alpha_1) \cdot c$$

b Show how to write the two "separate" linear regressions from above as a single "joint" regression:

$$Y_i = \gamma_0 + \gamma_1 D_i + \gamma_2 X_i + \gamma_3 D_i X_i + \epsilon_i$$

where D_i is a dummy variable that equals one if $Xi \geq c$. What is the relationship between the coefficients $(\gamma_0, \gamma_1, \gamma_2, \gamma_3)$ of the joint regression, and the coefficients (α_0, α_1) and (β_0, β_1) of the "separate" regressions?

$$\begin{aligned} \gamma_0 &= \alpha_0 \\ \gamma_1 &= \beta_0 - \alpha_0 \\ \gamma_2 &= \alpha_1 \\ \gamma_3 &= \beta_1 - \alpha_1 \end{aligned}$$

c Combine your answers to parts 1 and 2, to write the RD causal effect in terms of $(\gamma_0, \gamma_1, \gamma_2, \gamma_3)$.

$$LATE_{RD} = (\beta_0 - \alpha_0) + (\beta_1 - \alpha_1) \cdot c = \gamma_1 + \gamma_3 \cdot c$$

d To make the regression from part 3 easier to interpret, define $\bar{x}_i = x_i - c$ and and substitute $(X - c + c)$ in place of X. Using this substituting, show that we can re-write the joint regression as

$$Y_i = \bar{\gamma}_0 + \bar{\gamma}_1 D_i + \bar{\gamma}_2 \bar{X}_i + \bar{\gamma}_3 D_i \bar{X}_i + \epsilon_i$$

where $\bar{\gamma}_1$ equals the RD causal effect. How do the the other parameters of this "modified" regression relate to the original coefficients $(\gamma_0, \gamma_1, \gamma_2, \gamma_3)$?

$$\bar{\gamma}_0 = \gamma_0 + \gamma_2 \cdot c$$
$$\bar{\gamma}_1 = \gamma_1 + \gamma_3 \cdot c$$
$$\bar{\gamma}_2 = \gamma_2$$
$$\bar{\gamma}_3 = \gamma_3$$