

Lista Prática 4

Exercício 1. Neste exercício, usaremos a base de dados *meap00_01*, do pacote *wooldridge*, cujas observações são escolas de Michigan. Queremos estimar o seguinte modelo:

$$\text{math4} = \beta_0 + \beta_1 \text{lunch} + \beta_2 \text{lenroll} + \beta_3 \text{lexppp} + \varepsilon \quad (1.1)$$

em que

- *math4*: percental de alunos de 4^o ano com nível de matemática satisfatório
 - *lunch*: percentual de estudantes elegíveis ao programa de auxílio para almoço
 - *lenroll*: log de matrículas escolares
 - *lexppp*: log de gastos por estudante
- a) Estime por MQO com erros padrão usuais e robustos à heterocedasticidade por White (1980). Obtenha a matriz de variâncias-covariâncias do estimador robusta “na mão”, ou seja, não use funções que façam os cálculos para você (como `vcovHC(..., "HC0")` ou outras)¹. As estimativas e os erros padrão são distintos? Justifique.
- b) Faça o teste de White para heterocedasticidade. Quais são os modelos restrito e irrestrito utilizados no teste *F*? Analise o resultado do teste a partir estatística obtida.
- c) Faça a regressão de $g_i \equiv \log(\hat{\varepsilon}_i^2)$ em função dos valores preditos $\widehat{\text{math4}}$ e $\widehat{\text{math4}}^2$, obtidos pelo modelo do item (a). Calcule os pesos $w_i = 1/\exp(\hat{g}_i)$ para estimar o modelo (1.1) por MQGF. Há diferenças em relação aos resultados obtidos por MQO?
- d) Davidson e MacKinnon (1999, pág. 264) afirmam que é possível iterar o procedimento de MQGF. Para isto usam-se as estimativas obtidas no item (c) para calcular novos pesos $w_i = 1/\exp(\hat{g}_i)$ e novas estimativas, e isso pode ser repetido iteradamente. Faça iterações do procedimento de MQGF enquanto a seguinte condição for verdadeira:

$$\max\{\hat{\beta}^{it+1} - \hat{\beta}^{it}\} > 1 \times 10^{-15},$$

em que *it* é o número de uma iteração.

Informe quantas iterações foram necessárias para a convergência e compare os quatro modelos estimados (MQO, MQO com erros padrão robustos, MQGF, e MQGF com iteração) usando a função *stargazer()* do pacote de mesmo nome. Há algum modelo melhor para analisar o efeito de gastos (*lexppp*) em *math4*? Explique.

¹Além de obter os resíduos por meio da função `residuals()` sobre o objeto de regressão gerado por `lm()`, é possível obter a matriz *X* usando `model.matrix()` no mesmo objeto.

Exercício 2. Neste exercício, usaremos a base de dados de Angrist e Evans (1998) que contém quase 32 mil observações de mulheres negras ou hispânicas, todas casadas e com pelo menos dois filhos. Para carregá-la no R, use:

```
1 data(labsup, package="wooldridge")
```

Queremos saber como o número de crianças (*kids*) afeta as horas trabalhadas por semana (*hours*). Há suspeita de que *kids* seja um regressor endógeno e, portanto, utilizaremos variáveis instrumentais para gerar variação exógena que permita identificar o efeito do *nº* de crianças sobre *hours*.

a) Estime a equação por MQO com erros padrão robustos para heterocedasticidade:

$$hours = \beta_0 + \beta_1 kids + \beta_2 nonmomi + \beta_3 educ + \beta_4 age + \beta_5 agesq + \beta_6 black + \beta_7 hispan + \varepsilon$$

- *kids*: nº de crianças
- *nonmomi*: renda familiar que não é da mãe
- *educ*: anos de educação
- *age*: idade
- *agesq*: idade ao quadrado
- *black*: dummy negra
- *hispan*: dummy hispânica

b) Um dos instrumentos propostos por Angrist e Evans é a variável dummy *samesex*, que possui valor 1 quando os dois primeiros filhos são do mesmo sexo, e 0 caso contrário. Exponha argumentos (favoráveis ou contrários) para que *samesex* esteja relacionado a *kids*, e que não afete diretamente as horas trabalhadas (*hours*).

c) Faça o teste de instrumento fraco no primeiro estágio da regressão. O quão estatisticamente significativa é a variável *samesex* em relação a *kids*?

d) Use *samesex* como instrumento de *kids*. Estime o modelo estrutural do item (a) pelo estimador VI por meio da **função** *ivreg()* e **analiticamente** ("na mão"). Compare os resultados MQO e VI usando a função *stargazer()* do pacote de mesmo nome.

e) Avalie a potencial endogeneidade de *kids* por meio do Teste de (Durbin-Wu-)Hausman. Analise o seu resultado, explicitando a hipótese nula testada.

c') Deste item em diante, considere como um segundo instrumento para *kids* a variável *multi2nd*, que é uma dummy indicando que os dois primeiros filhos são gêmeos. Faça o teste de instrumentos fracos no primeiro estágio da regressão. O quão conjuntamente significantes são as variáveis *samesex* e *multi2nd* em relação a *kids*?

d') Use *samesex* e *multi2nd* como instrumentos de *kids* e estime o modelo estrutural do item (a) pelo estimador MQ2E por meio da **função** *ivreg()*, da **função** *lm()* e **analiticamente** ("na mão"). Compare os resultados MQO, VI e MQ2E usando a função *stargazer()*.

- e') Avalie a potencial endogeneidade de *kids* por meio do Teste de (Durbin-Wu-)Hausman, porém agora comparando $\hat{\beta}^{MQO}$ e $\hat{\beta}^{MQ2E}$. Analise o seu resultado, explicitando a hipótese nula testada.
- f) Faça o teste de sobreidentificação de Sargan e analise o resultado, explicitando sua hipótese nula.