**Econometria II
Exercícios para revisão e autoteste**
“Introdução à Econometria”, Jefrey M. Wooldridge
**VARIÁVEIS INSTRUMENTAIS E MÍNIMOS QUADRADOS DE DOIS ESTÁGIOS**

Obs.: os exercícios que indicam ‘arquivos’ para serem resolvidos são do livro do Wooldridge. Os arquivos necessários estão na pasta “Banco de dados Wooldridge”, na área “Programação em R”.

**1.** Suponha que queiramos estimar o efeito do uso de maconha na nota média de graduação. Em uma população de alunos veteranos de uma universidade, defina *diasusados* como o número de dias do mês anterior que um aluno fumou maconha e considere a equação estrutural
 *supGPA* = β0 + β1*diasusados* + β2*SAT* + *u*.

Defina *percento* como a porcentagem de uma sala do ensino médio que informou uso regular de maconha. Se ela for uma candidata a VI de *diasusados*, escreva a forma reduzida de *diasusados*. Você acha que *percento* é verdadeiramente exógena na equação estrutural? Que problemas podem surgir nesse caso?

**2.** O modelo seguinte explica a taxa de crimes violentos, ao nível de cidades, em termos de uma variável binária que indica se existem leis de controle de armas e outros controles:

*violento* = β0+ β1*controledearmas* + β2*desemp +* β3*popul* + β4*porcnegro +* β5*idade18\_21*  + ...

Alguns pesquisadores estimaram equações semelhantes usando variáveis como o número de membros da National Rifle Association na cidade e o número de assinantes de revistas sobre armas como variáveis instrumentais de *controledearmas*. Essas variáveis instrumentais são convincentes?

**3.** Um modelo para testar o efeito do crescimento dos gastos governamentais sobre o crescimento da produção é

cPIBt = β0+ β1*cGOVt* + β2*RAZINVt +* β3*cTRABt* + *ut* ,

em que *c* indica crescimento, *PIB* é o produto interno bruto real, *GOV* é o gasto governamental real, *RAZINV* é a razão do investimento interno sobre o PIB e *TRAB* é o tamanho da força de trabalho. Sob quais hipóteses uma variável dummy indicando se o presidente no ano *t-1* era um republicano seria uma VI adequada de cGOV?

**4.** Considere um modelo simples para estimar o efeito da propriedade de um computador pessoal (PC) na nota média de graduação de formandos de uma grande universidade pública:

*supGPA = β0+ β1PC + u*,

em que *PC* é uma variável binária que indica a propriedade de um PC.
a) Por que a propriedade de um PC pode estar correlacionada com u?
b) Explique por que *PC* possivelmente está relacionado a renda anual dos pais. Isso significa que a renda dos pais será uma boa VI de *PC*? Por quê?
c) Suponha que, quatro anos atrás, a universidade tenha concedido subsídios para a compra de computadores a aproximadamente metade dos alunos novos, e que os alunos que receberam esses subsídios tenham sido escolhidos aleatoriamente. Explique cuidadosamente como você usaria essa informação para construir uma variável instrumental de *PC*.

**5.**Suponha que você queira usar dados de séries temporais de determinado estado dos EUA, para estimar o efeitodo salário mínimo em nível estadual sobre o emprego de pessoas entre 18 e 25 anos de idade (*EMP*). Um modelo simples é

 *cEMPt* = *β0+ β1cMINt*  + *β2cPOPt  + β3cPEBt* + *β4cPIBt + ut ,*

em que MINt é o salário mínimo, em dólares reais, POPt é a população com idade entre 18 e 25 anos, PEBt é o produto estadual bruto e PIBt é o produto interno bruto norte-americano. O prefixo c indica a taxa de crescimento do ano t-1 ao ano 1, que em geral será aproximada pela diferença dos logs.
a) Se estivermos preocupados que o estado escolha seu salário mínimo parcialmente baseado em fatores não observados (por nós) que afetem o emprego dos jovens, qual será o problema da estimação por MQO?
b) Defina SMAmt como o salário mínimo dos EUA, que também é indicado em termo de dólares reais. Você acha que cSMAmt é não correlacionado com ut?
c) Por lei, qualquer salário mínimo estadual deve ser pelo menos igual ao salário mínimo nacional. Explique por que isso torna cSMAmt um candidato em potencial para ser uma VI de cMINt­.

**6.** O que segue é um modelo simples para medir o efeito de um programa de escolha de escola sobre o desempenho de um teste padronizado:

 *nota =* *β0+ β1escolha + β2rendfam + u1,*

em que *nota* é a nota em um teste de âmbito estadual, *escolha* é uma variável binária indicando se o aluno frequentou uma escola de sua escolha no último ano e *rendfam* é a renda familiar. A VI de *escolha* é *conc*, o montante em dólares concedido aos alunos para ser usado como pagamento da anuidade da escola particular de sua escolha. O montante da concessão difere conforme o nível de renda familiar, razão pela qual controlamos *rendfam* na equação.
a) Mesmo com *rendfam* na equação por que *escolha* pode ser correlacionada com *ui*?
b) Se no interior de cada classe de rendimento os montantes de concessão fossem atribuídos aleatoriamente, *conc* seria não correlacionada com *ut*?
c) Escreva a forma reduzida da equação de *escolha*. O que é necessário para *conc* ser parcialmente correlacionado com *escolha?*
d) Escreva a equação na forma reduzida de *nota*. Explique por que isso é importante.

**7.** Suponha que você queira testar se meninas que frequentam uma escola de ensino médio só para meninas se saem melhor em matemática do que as que frequentam escolas mistas. Você tem uma amostra aleatória de meninas veteranas de escolas de ensino médio de um estado dos EUA, e *nota* é a nota de um teste padronizado em matemática. Defina *meninaem* como uma variável dummy indicando se uma aluna frequenta uma escola de ensino médio só para meninas.
a) Que outros fatores você controlaria na equação? (Você deve ter condições razoáveis de coletar dados sobre esses fatores.)
b) Escreva uma equação que relaciona *nota* com *meninaem* e os outros fatores que você listou na parte a).
c) Suponha que o suporte e o incentivo dos pais sejam fatores não indicados no termo de erro na parte b). É possível que eles sejam correlacionados com *meninaem*? Explique.
d) Discuta as hipóteses necessárias para que o número de escolas do ensino médio só para meninas situadas em um raio de 20 milhas (aproximadamente 32km) da residência de uma menina seja uma VI valida de *meninaem.*

**8.** Considere o problema de estimar o efeito causal de faltar às aulas sobre as notas do exame final. Em uma estrutura de regressão simples, temos:

 *nota =* *β0+ β1faltas + u,*

em que *nota* é a nota no exame final e *faltas* é o número total de faltas às aulas durante o semestre. Suponha que você não tenha uma boa candidata a variável instrumental de *faltas*. Entretanto, você tem duas outras informações sobre os alunos: a nota média ponderada de matemática e habilidade verbal do estudante para ingresso em curso superior (SAT) e a nota média acumulada anterior ao semestre (GPA). O que você faria em vez da estimação de VI?

**9.** Em um artigo recente, Evans e Schwab (1995) estudaram os efeitos que frequentar uma escola católica do ensino médio teriam sobre a probabilidade de cursar uma faculdade. Concretamente, defina *faculdade* como uma variável binária igual a um se o aluno estiver na faculdade, e zero caso contrário. Defina *EMCat* como uma variável binária igual a um se o aluno frequenta uma escola católica do ensino médio. Um modelo de probabilidade linear é

 *faculdade =* *β0+ β1EMCat + outros fatores + u,*

dm que, entre outros fatores, estão sexo, raça, renda familiar e instrução do país.
a) Por que *EMCat* pode ser correlacionada com *u*?
b) Evan e Schwab tinham dados sobre a nota de um teste padronizado feito quando cada estudante era aluno do 2º ano. O que pode ser feito com essa variável para melhorar a estimativa *ceteris paribus* de frequentar uma escola católica no ensino médio?
c) Defina *RelCat* como uma variável binária igual a um se o estudante for da religião católica. Detalhe os dois requisitos necessários para que essa seja uma VI válida de *EMCat* na equação precedente. Qual deles pode ser testado?
d) Não surpreendentemente, o fato de ser católico tem um efeito significante sobre frequentar uma escola católica no ensino médio. Você julga que *RelCat* é uma variável instrumental convincente de *EMCat*?

**10.** Considere um modelo simples de séries temporais no qual a variável explicativa tem erro clássico de medida:

 *yt =* *β0+ β1xt\* + ut,
 xt= xt\* + et,*

em que *ut* tem media zero e é não correlacionado com *xt\** e *et.* Observamos somente *yt* e *xt*. Suponha que *et* tem média zero e é não correlacionado com *xt\** e que *xt\** também tem média zero ( esta última hipótese é só para simplificar a álgebra).
a) Escreva *xt\* = xt – et* e insira essa expressão na equação acima. Mostre que o termo de erro da nova equação, digamos *vt*, é negativamente correlacionado com *xt* se *β1 > 0.* O que isso sugere sobre o estimador MQO de *β1* da regressão de *yt* sobre *xt*?
b) Além das hipóteses anteriores, presuma que *ut* e *et* são não correlacionados com todos os valores passados de *xt\** e *et* ; em particular com *xt-1\** e *et-1* . Mostre que *E(xt-1,vt)=0,* em que *vt* é o termo de erro no modelo da parte a).
c) É possível que *xt* e *xt-1* sejam correlacionados? Explique.
d) O que as partes b) e c) sugerem como uma estratégia vantajosa para estimarmos consistentemente *β0* e *β1* ?

**SOLUÇÕES**

1. Temos que assumir que a porcentagem de estudantes que fumava maconha em uma escola de ensino médio não está correlacionada com fatores não observados que afetam a nota média no curso superior. Embora estejamos, até certo ponto, controlando a qualidade do ensino médio ao incluirmos *sat* na equação, isso pode não ser suficiente. Talvez as escolas de ensino médio que tenham feito um melhor trabalho na preparação dos estudantes para a universidade também tenham tido um menor número de alunos que fumavam maconha. Ou, o uso de maconha poderia estar correlacionado com os níveis médios de renda. Naturalmente, essas são questões empíricas que podem ou não ser respondidas.

2. Embora a predominância da National Riffle Association e os assinantes de revistas sobre armas estejam provavelmente correlacionados com a presença de leis de controle de armas, não é óbvio que eles sejam não-correlacionados com outros fatores não observados que afetem a taxa de crimes violentos. De fato, podemos argumentar que uma população interessada em armas seja um reflexo das elevadas taxas de criminalidade, e que o controle de variáveis econômicas e demográficas não seja suficiente para capturar esse aspecto. Seria difícil sustentar de forma persuasiva que elas são realmente exógenas na equação de crimes violentos.

3. Como sempre, existem dois requisitos. Primeiro, poderia ser o caso de que o crescimento dos gastos governamentais esteja sistematicamente relacionado ao partido do presidente, após a remoção da taxa de investimento e o crescimento da força de trabalho. Em outras palavras, a variável instrumental deve ser parcialmente correlacionada com a variável explicativa endógena. Embora possamos imaginar que os gastos governamentais cresçam mais lentamente em governos com presidentes Republicanos, isso certamente não tem sido sempre verdade nos Estados Unidos e teria que ser testado usando a estatística t de *REPt-1* na forma reduzida *cGOVt = π0 + π1 REPt-1 + π2RAZINVt + π3TRABt + vt*. Devemos assumir que o partido do presidente não tem efeito separado sobre *cPIB*. Isso seria violado se, por exemplo, a política monetária diferisse sistematicamente por partido presidencial e tivesse um efeito separado sobre o crescimento do PIB.