**Nomes**: João Rodrigues, Marcos de Campos, Vinícius Ferreira

## **Exercício 1**

**Questão (i)**

A posse ou não de um computador pode estar bastante positivamente correlacionada com variáveis omitidas, como renda pessoal ou familiar.

**Questão (ii)**

Sendo PC positivamente correlacionado com a renda anual dos pais – já que famílias mais ricas tem mais condições de fornecer esses equipamentos aos filhos -, ele satisfaz a condição de relevância de uma VI. Contudo, a renda pode estar correlacionada com outros fatores omitidos que afetam a nota – como aptidão -, de modo que a condição de exogeneidade dos instrumentos será violada. Assim, é melhor explicitá-la na equação (se os dados estiverem disponíveis) e buscar uma outra VI ou *proxy* para PC.

**Questão (iii)**

Caso essa doação de subsídios pela universidade tenha sido feita de forma aleatória, configura-se um experimento natural e *grant* – uma dummy que indica o recebimento do subsídio - estaria não correlacionada com o termo de erro e certamente positivamente correlacionada com *PC* – já que essa era a finalidade da doação. Assim, seguindo essas hipóteses, *grant* poderia ser usada como uma VI de *PC* já que é tanto exógena como relevante.

## **Exercício 2**

**Questão (i)**

Estudantes de maior renda tem mais condição de morar perto do campus; como renda certamente afeta a performance, *dist* estaria negativamente correlacionada com ela.

**Questão (ii)**

Assumindo a não-correlação entre *dist* e *u* (exogeneidade do instrumento), *dist* ainda teria que satisfazer a condição de relevância, ou seja, deve ser correlacionada com *atndrte*. Espera-se que esse seja o caso, já que, quanto maior a distância à sala de aula, menos um aluno tende a ir (ou maior a chance de chegar atrasado), de forma que *dist* e *atndrte* são, provavelmente, negativamente correlacionados.

**Questão (iii)**

Sendo *atndrte* correlacionado com *u,* surge a necessidade de Vis tanto para *atndrte* como para o termo de interação proposto. Assumindo que *dist* e *u* são não-correlacionados, *dist* pode ser usada como VI de *atndrte* (como discutido no item anterior).

Sendo *dist* não-correlacionada com *u*, o termo de interação *priGPAdist* também será, de forma que o modelo pode ser estimado usando Mínimos Quadrados em Dois Estágios (MQ2E).

## **Exercício 3**

A equação (15.10) é

, onde *z* é uma variável *dummy*. Pelas propriedades do somatório:

Desenvolvendo o numerador e usando o fato de que é o número de observações em que *zi* = 1 (denotado por *z1*):

Como a média de uma variável *w* qualquer é , pode-se alterar a expressão acima para:

Como pode ser visto como uma média ponderada entre os casos em que *zi* = 0 e *zi* = 1, o numerador se torna:

O procedimento análogo pode ser feito para o denominador, de forma que

## **Exercício 4**

**Questão (i)**

Se os estados escolhem o salário-mínimo com base nos fatores não-observáveis em *u*, haverá a violação da RLM.4, o que tornará os estimadores viesados e inconsistentes.

**Questão (ii)**

Como o salário-mínimo agora é nacional, é muito pouco provável que ele esteja correlacionado com características individuais de cada estado, o que não pode ser dito para o salário-mínimo local, o qual pode estar correlacionado com o partido dominante na região, por exemplo (estados democratas tendem a ter salários-mínimos maiores do que aqueles de maioria republicana).

**Questão (iii)**

Como o mínimo para um estado é o salário-mínimo nacional, aumentos do salário-mínimo nacional provocarão aumentos naqueles estados cuja lei ditava um montante mais escasso. Assim, *gUSMIN* estará ao menos parcialmente correlacionado com *gMIN* - ou seja, é relevante - e, como argumentado na questão (ii), é não correlacionado com *u* – ou seja, é exógeno. Assim, *gUSMIN* pode ser uma boa variável instrumental para *gMIN*, especialmente para aqueles estados cuja legislação não estabelece maiores salários.

## **Exercício 7**

**Questão (i)**

Pode haver um problema de autosseleção: estudantes mais aptos – cujo *score* é maior – podem decidir ir para uma “choice school”.

**Questão (ii)**

Se dentro de cada classe de renda os subsídios forem dados aleatoriamente, *grant* será não-correlacionada com *u*, já que ele não foi afetado por nenhuma característica individualmente relevante para o modelo.

**Questão (iii)**

A equação reduzida (ou de primeiro estágio) de *choice* é

Para que *grant* seja parcialmente correlacionado com *choice*, de modo que a condição de relevância seja satisfeita, necessita-se que .

**Questão (iv)**

A equação reduzida de *score* é

, o que pode ser útil a fim de analisar os efeitos parciais da política de subsídios dados aos alunos sobre a nota final do teste estadual.

## **Exercício C1**

**Questão (i)**

A regressão simples de *lwage* sobre *sibs* produz os resultados abaixo; enquanto a estimativa de *educ* com *sibs* como VI é 0,122, a estimativa do efeito parcial percentual (e significante) do número de irmãos é ≈ -2,8%. De fato, um número maior de irmãos está correlacionado com uma menor educação, o que, tudo o mais constante, diminui os salários futuros.

A correlação negativa entre *educ* e *sibs* pode ser vista pelos resultados da regressão simples envolvendo essas duas variáveis: um irmão adicional está relacionado a uma queda média de quase 3 meses na educação:

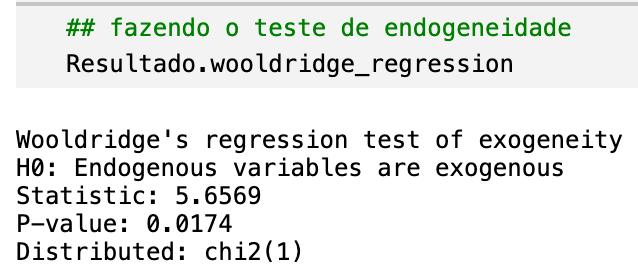
**Questão (ii)**

Nos EUA, dado que faculdades são privadas e caras, muitas famílias só têm dinheiro para mandar o primeiro filho a essas instituições, de modo que se espera que *educ* e *brthord* estejam negativamente correlacionados. De fato, a regressão de *educ* sobre *brthord* mostra que há uma correlação negativamente e estatisticamente significante, sendo o p-valor da variável igual a 0 até quatro casas decimais:

**Questão (iii)**

Usando *brthord* como uma VI de educ, o retorno estimado da educação é aproximadamente 13% e possui p-valor 0 até quatro casas decimais:

Essa estimativa é mais que o dobro daquela obtida por MQO (≈ 6%), o que evidencia que de fato *educ* é endógena. Realizando um teste de endogeneidade, vê-se que seu p-valor é 0,0174. Assim, rejeita-se a hipótese nula de que *educ* é exógena.



**Questão (iv)**

Controlando explicitamente para o número de irmãos e usando *brthord* como VI de *educ*, vê-se que a condição de relevância de *brthord* é satisfeita: na regressão de 1º estágio, a estimativa é -0,1527 e sua estatística t é -2,68.

Tabela

Descrição gerada automaticamente

**Questão (v)**

O resultado da equação estrutural do modelo proposto no item anterior é mostrado abaixo. A significância estatística do coeficiente de *educ* agora só é significante a 5% contra uma alternativa unilateral; além disso, o coeficiente de *sibs* tem um p-valor de 0,9, sendo bastante estatisticamente insignificante.

**Questão (vi)**

Usando os valores de *educ* obtidos pela regressão de (iv), vê-se que a correlação entre eles e *sibs* é igual a -0,9294. Essa alta correlação indica problemas de multicolinearidade, o que aumenta os erros padrão dos estimadores do modelo em (v).

Tabela

Descrição gerada automaticamente

## **Exercício C2**

**Questão (i)**

Usando o modelo proposto, o efeito parcial marginal de um ano de educação é de -0,09 filhos; caso 100 mulheres tenham um ano a mais de educação, espera-se que nasçam 9 crianças a menos, tudo o mais constante.

**Questão (ii)**

Sendo *frsthalf* exógena, restar testar se ela satisfaz a condição de relevância. Para isso, é necessário regredir *educ* sobre todas as variáveis exógenas do modelo e verificar a significância estatística de *frsthalf*. Ao se fazer isso, vê-se que a estimativa de *frsthalf* tem p-valor 0 até 4 casas decimais, indicando um forte efeito negativo de quase um ano sobre a educação para aquelas mulheres que nasceram na primeira metade do ano.

**Questão (iii)**

Usando *frsthalf* como VI de educ no modelo da questão (i), vê-se que o efeito parcial da educação sobre o número de nascimentos quase dobra, passando de -9 para -17 a cada 100 mulheres. O erro padrão aumenta em quase 10 vezes, mas o p-valor de *educ* ainda é 0,001, ou seja, a estimativa ainda é estatisticamente significante, mas com um intervalo de confiança mais amplo.

**Questão (iv)**

Adicionando as *dummies* exógenas de eletricidade, TV (as quais possibilitam mais formas de passar o tempo) e bicicleta (que pode dar mais lazer a possíveis crianças), o coeficiente *educ* reduz um pouco em ambos os modelos frente as suas estimativas anteriores.

No modelo de MQ2E, apenas *bicycle* é estatisticamente significante; já no modelo de MQO, todas as *dummies* são estatisticamente significantes. Em ambos os casos, os sinais estão de acordo com o que se espera.

Abaixo estão os resultados da estimação por MQ2E e MQO, respectivamente;

Texto preto sobre fundo branco

Descrição gerada automaticamente

Tela de celular com texto preto sobre fundo branco

Descrição gerada automaticamente

## **Exercício C3**

**Questão (i)**

A proximidade do domicílio de uma faculdade que oferece cursos de quatro anos pode estar correlacionada com vários outros fatores, como renda. Pessoas de maior renda costumam ter melhor nutrição e mais atenção dos pais (ou babás), o que tende a levar a maior aptidão. Assim, a condição de exogeneidade estaria sendo violada.

**Questão (ii)**

Regredindo *IQ* sobre *nearc4*, vê-se que há uma correlação positiva muito forte e estatisticamente significante entre essas variáveis, contribuindo para a hipótese desenvolvida no item anterior:

**Questão (iii)**

Controlando para diferenças geográficas, o coeficiente de *nearc4* diminui para 0,3479 e tem um p-valor de 0,6684, sendo estatisticamente não-significante. Assim, controlando para o ambiente em que o homem cresceu, não há diferença nas aptidões entre aqueles que cresceram perto de uma faculdade e aqueles que não:

**Questão (iv)**

Dada a diferença entre a exogeneidade ou não de *nearc4* quando os controles regionais são adicionados, é importante adicioná-los ao modelo estrutural a fim de garantir que *nearc4* seja uma boa VI para *educ*.

## **Exercício C5**

**Questão (i)**

Regredindo *educ* sobre os controles da tabela 15.1, obtemos os resíduos . Colocando esses resíduos como uma variável explicativa no modelo estrutural – no qual *educ* é instrumentado por *nearc4* -, obtém-se:

Como o p-valor de é 0,2785, não se rejeita H0: *educ* é exógeno, de modo que a instrumentação dessa variável pode ser vista como desnecessária.

**Questão (ii)**

Adicionando *nearc2* como instrumento ao lado de *nearc4*, o retorno da educação aumenta de aproximadamente 13,2% para 15,71% e seu erro-padrão aumenta ligeiramente, mas seu p-valor ainda é inferior a 0,01. Contudo, o intervalo de confiança de *educ* passa a ter um raio de 10%, algo bastante significativo e semelhante à estimativa usando apenas *nearc4* como instrumento:

Tela de celular com texto preto sobre fundo branco

Descrição gerada automaticamente

**Questão (iii)**

O teste de restrições sobreidentificadoras produz um p-valor de 0,26, de modo que não se rejeita H0: os instrumentos são exógenos.

## **Exercício C9**

**Questão (i)**

Tabela

Descrição gerada automaticamente

Como podemos ver no relatório acima, a instrumentalização da variável *sibs* no lugar de *educ* foi bem sucedida, incorporando as outras variáveis (*black*, *exper* e *tenure*) também como instrumentos ao lado de *sibs*. Também podemos observar que todas as variáveis são relevantes estatisticamente, exceto por *exper* que é marginalmente relevante ao nível de 1%.

**Questão (ii)**

Tabela

Descrição gerada automaticamente



Tabela

Descrição gerada automaticamente

Com os dois passos da MQ2E demonstrados acima, podemos ver que os resultados para os coeficientes são iguais aos do item **(i)**, diferenciando-se apenas pelos erros-padrão ligeiramente maiores.

**Questão (iii)**

Texto, Tabela

Descrição gerada automaticamente

Tabela

Descrição gerada automaticamente

Nesse caso, podemos ver que, além dos erros-padrão inconsistentes, os coeficientes estimados também são incongruentes com os resultados dos itens **(i)** e **(ii)**. Como fizemos o primeiro passo ea MQ2E de maneira incorreta, o fato de os resultados estimados serem incoerentes não nos surpreende.

## **Exercício C11**

**Questão (i)**

Uma imagem contendo Padrão do plano de fundo

Descrição gerada automaticamente Dos 990 alunos da amostra, 522 atenderam à escola escolhida. O número de alunos que tiveram a vaga disponível por 4 anos foi de 108 e os que frequentaram a escola escolhida por 4 anos foram 56.

**Questão (ii)**

Texto, Tabela

Descrição gerada automaticamente

Na regressão simples acima podemos observar que a correlação entre *choiceyrs* e *selectyrs* é grande e ocorre no sentido esperado. Isso se dá pois a correlação é positiva, indo de encontro com a intuição de que quanto mais anos (ou mais cedo nos anos observados) o candidato foi escolhido para a vaga, maior o número de anos esperados em que o aluno estudou de fato em tal instituição.

**Questão (iii)**

Tabela

Descrição gerada automaticamente

Como pode ser observado no relatório acima, o coeficiente estimado para *choiceyrs* não é o esperado, pois indica que a nota no teste de matemática é consideravelmente pior nos alunos que mais frequentaram a escola de escolha.

Tabela

Descrição gerada automaticamente

Quando adicionamos as variáveis *black*, *hispanic* e *female* ao modelo, podemos observar uma redução nesse impacto negativo inesperado de *choiceyrs*. Porém, seu sentido de influência em *mnce* ainda é contra intuitivo e sua magnitude ainda é consideravelmente alta. Além disso, podemos ver que *choiceyrs* perde sua significância estatística nesse modelo; enquanto *black* e *hispanic* são relevantes estatisticamente, *female* também se mostra como insignificante.

**Questão (iv)**

Pois, se considerarmos que o sorteio para ingressar na escola só é realizado em alunos que se candidataram às vagas, podemos traçar uma relação intuitiva entre a nota do teste *mnce* os alunos que se candidataram e um fator de aptidão não mensurado incorporado pelo termo de erro.

Isso se dá pois podemos deduzir que alunos com maior aptidão tem uma maior probabilidade de se candidatar à vaga e uma maior probabilidade de permanecer na escola escolhida por mais anos (fator medido pela variável *choiceyrs*). Assim a variável *choiceyrs*, o termo de erro e a nota *mnce* estariam correlacionados, enviesando todo o modelo.

**Questão (v)**

Tabela

Descrição gerada automaticamente

No modelo acima, ao usarmos a Variável Instrumental em *choiceyrs* podemos ver uma atenuação no efeito negativo da variável, porém, ainda assim, o efeito é não esperado pois é negativo. Quanto aos coeficientes das outras variáveis, nenhuma sofre alterações de magnitude considerável.

**Questão (vi)**

Por MQO:

Tabela

Descrição gerada automaticamente

Por VI:

Tabela

Descrição gerada automaticamente

Agora, ao incluirmos a variável *mnce90* e controlarmos para as conquistas anteriores do aluno, o coeficiente estimado para a variável *choiceyrs* passa a assumir o sentido positivo esperado. Além disso, no modelo de VI o coeficiente de *choiceyrs* passa a ser mais impactante, representando 1.8% a mais de nota no teste a cada ano estudado na escola de preferência.

**Questão (vii)**

Apesar de os resultados das duas regressões estimadas no item anterior **(vi)** irem de encontro com a intuição lógica, eles não são completamente confiáveis pois o número de observações decai muito quando incluímos a variável *mnce90*. Isso se dá, pois, o conjunto de observações não é 100% balanceado em todas as variáveis e quando reduzimos o número amostral de observações perdermos significância nos resultados do modelo.

**Questão (viii)**

Tabela

Descrição gerada automaticamente

Apesar de as variáveis *dummy* não serem individualmente estatisticamente significantes (o que é esperado), elas trazem ao modelo resultados que fazem muito mais sentido com a intuição lógica. Pode-se observar claramente no modelo acima que há a tendência dos alunos que ficaram mais anos na escola escolhida terem maior sucesso no teste *mnce*. O modelo, por incluir a variável *mnce90* ainda contém a suspeita de inviesamento por baixa quantidade de observações, mas no geral parece ser um modelo explicativo infinitamente mais completo e robusto.