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

## **Exercício 1**

Assumindo as hipóteses da questão, a Correlação de Pearson entre *∆uit* e *∆uit+1* é:

Como, por hipótese, os resíduos são não correlacionados temporalmente e a variância dos erros idiossincráticos são constantes,

A covariância, por sua vez, será:

Como cada erro tem média 0, a covariância será reduzida à esperança da multiplicação entre os termos acima, de modo que:

Como a esperança da soma é a soma das esperanças – e levando em conta que os erros são não correlacionados dois a dois[[1]](#footnote-1):

, onde a última igualdade da esperança de *uit*2 ser igual à sua variância decorre do fato de que a média de *uit* é, por hipótese, 0. Por fim, a Correlação de Pearson será:

Portanto, vê-se que, mesmo que os *uit*individuais tenham média 0, sejam não-correlacionados adjuntamente e que tenham variância constante – ou seja, sob as hipóteses ideais de Efeitos Fixos (EF) -, os resíduos de Primeiras Diferenças (PD) terão correlação serial negativa e serão, portanto, ineficientes relativamente aos estimadores obtidos por meio de EF.

## **Exercício 3**

**Questão (i)**

Como o modelo sempre possui um intercepto, podemos garantir que a esperança de *vit* será igual 0, o que faz com que a esperança de sua média também seja 0. Assim, mostra-se que a esperança do erro do modelo de efeitos aleatórios também será 0.

**Questão (ii)**

**Questão (iii)**

## **Exercício 5**

**Questão (i)**

Os dados obtidos seriam de agrupamento porque cada aluno possuirá o número de observações igual ao número de matérias feitas por ele. Nesse sentido, cada matéria possuirá um grupo de estudantes com efeitos individuais *ai* diferentes entre si.

**Questão (ii)**

Como não são relacionadas com cada classe, só dependem de cada estudante as variáveis *ira*, *vestibular* e suas características não observadas, *as*. De forma similar à um agrupamento temporal, cada matéria *i* pode receber um intercepto diferente, sendo *materiai* uma série de dummies para cada matéria.

**Questão (iii)**

Ao usar MQO agrupado, assume-se que *as* não está correlacionado com nenhuma outra das variáveis explicativas que dependa do agrupamento (da matéria) – o que pode ser bastante falso no caso de *pct,* por exemplo, uma vez que alunos com maior aptidão tem maior tendência a comparecerem mais às aulas.

Como *ira* e *vestibular* são proxies de aptidão, sua adição ao modelo pode fazer com que os efeitos não-observados já sejam capturados por elas, permitindo assim uma estimação não-viesada dos parâmetros do modelo.

**Questão (iv)**

Caso *ira* e *vestibular* não captem bem o efeito não-observado, ele seria, de fato, correlacionado com as variáveis explicativas que dependem do agrupamento. Assim, teríamos que eliminar *as*– e, consequentemente, as variáveis constantes para cada aluno *ira* e *vestibular* - via estimação de EF ou PD, a depender se há ou não correlação serial (no caso, entre diferentes matérias) entre os resíduos *usc*. Se houver correlação serial, prefere-se PD por ser mais eficiente.

## **Exercício C1**

**Questão (i)**

Usando MQO agrupado, encontram-se as estimativas abaixo. O coeficiente da dummy *y90* é significativo tanto economicamente quanto estatisticamente, indicando que há fatores estruturais não captados por outras variáveis que fizeram o preço do aluguel aumentar em aproximadamente 26,22% entre 1980 e 1990. O coeficiente de *pctstu*, por sua vez, é estatisticamente significante, mas bastante baixo: um aumento de 1% na porcentagem de estudante aumenta em 0,5% o preço dos aluguéis.

Tela de celular com texto preto sobre fundo branco

Descrição gerada automaticamente

**Questão (ii)**

O uso de MQO agrupado produz erros padrões incorretos, uma vez que, a menos que a variância das características não-observadas sejam seja igual a 0 (ou seja, que todas as cidades sejam iguais entre si, o que é impossível), haverá correlação serial entre os termos de erro.

**Questão (iii)**

Estimando a equação por PD, o impacto de uma mudança no percentual da população que é estudante dobra, chegando a um impacto percentual aproximado de +1,12% no preço dos aluguéis. O impacto ainda é estatisticamente significante mesmo ao nível de 1%, já que o p-valor da estimativa é 0,009.

Tabela

Descrição gerada automaticamente

**Questão (iv)**

Como esperado, o resultado das estimações por EF é idêntico à de PD, com a exceção de que EF produz uma constante e, por isso, possui um diferente R2:

Tela de celular com texto preto sobre fundo branco

Descrição gerada automaticamente

## **Exercício C2**

**Questão (i)**

A imagem abaixo compara o resultado das estimações de PD e EF para o modelo proposto, sendo a primeira coluna os resultados da estimação por PD, sendo os valores entre parênteses:

Tabela

Descrição gerada automaticamente

A tabela comparativa mostra que não há mudança significativa no valor das variáveis e na sua significância estatística, com exceção de *lavgsen* – que possui um valor maior (em módulo) na estimação de PD, mas, apesar de uma estatística *t* maior que o de EF, ainda é insignificante.

**Questão (ii)**

Controlando para todas as variações possíveis de remuneração, obtemos as estimativas descritas na segunda coluna da tabela abaixo. Vemos que a maioria das variáveis de justiça criminal tem magnitude e significância muito parecida em ambos os modelos, indicando que são, em grande parte, não correlacionadas com as remunerações.

Tabela

Descrição gerada automaticamente

Tabela

Descrição gerada automaticamente

**Questão (iii)**

Esperava-se que os coeficientes de *wage* tivessem sinal negativo: uma maior remuneração desestimularia a prática de crimes, o que não é visto em todas as variáveis, apesar de todas terem efeitos próximos de 0.

Um teste F de significância conjunta das variáveis produz um *p-valor* de 0,0065 usando a distribuição *qui-quadrado (9)* – apesar de sinais contra-intuitivos, as variáveis de remuneração são conjuntamente estatisticamente significantes até a 1%.

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

## **Exercício C3**

**Questão (i)**

A regressão estimada por EF é feita com 135 empresas e 390 observações; se todas as firmas possuíssem todas as informações, o modelo de EF seria estimado com 135\*3 = 405 empresas.

Tela de celular com texto preto sobre fundo branco

Descrição gerada automaticamente

**Questão (ii)**

O coeficiente de *grant* é igual a 34,23, ou seja, uma empresa que recebeu subsídio treinou os funcionários 34,23 horas a mais, em média. A estimativa também é estatisticamente significante, com uma estatística *t* igual a 11,974. Isso é esperado, já que o recebimento de subsídio permite às empresas investirem mais em seus colaboradores.

**Questão (iii)**

O coeficiente de *grant\_1* é igual a 0,5 e possui um p-valor de 0,9, sendo estatisticamente não-significante. Isso é esperado, já que o subsídio provavelmente já foi gasto no anterior e indica que não há inércia do efeito do recebimento de verbas.

**Questão (iv)**

O coeficiente de *lemploy* é negativo e estatisticamente não-significante, mostrando que o tamanho da firma não tem efeito sobre as horas de treinamento.

## **Exercício C5**

**Questão (i)**

## Diferentes ocupações podem estar correlacionadas com variáveis explicativas que estão no modelo, como *educ* e *married*. Como algumas pessoas mudam de ocupação ao longo do período 1980-87, a inclusão dessas *dummies* não seria possível via PD ou EF.

**Questão (ii)**

Como explicado no item anterior, caso opte-se pela estimativa de EF ou PD, a inclusão das *dummies* só é possível caso haja alguma variação de ocupação ao longo do período analisado, já que qualquer efeito fixo ao longo do tempo é eliminado com as transformações dessas metodologias. Caso o modelo fosse estimado por MQO Agrupado, Efeitos Aleatórios ou Efeitos Aleatórios Não-Correlacionados, a inclusão das *dummies* poderia ser feita independentemente da existência de variação.

**Questão (iii)**

Colocando *occ1* como o grupo-base, o modelo proposto estimado por EF produz os resultados dispostos abaixo. Mesmo controlando para as diferentes ocupações, o coeficiente de *union* é estatisticamente significante (com um p-valor de 0,0000) e indica que pessoas sindicalizadas tem, em média e aproximadamente, um salário 8,04% maior.

Tabela

Descrição gerada automaticamente

## **Exercício C7**

**Questão (i)**

Caso execuções tenham um efeito redutor na taxa de homicídios, espera-se que seu coeficiente seja negativo. Quanto ao coeficiente de desemprego, espera-se que estados com mais desempregados tenham mais crime, o que provavelmente aumenta a taxa de homicídios do local. Assim, espera-se que seu coeficiente seja maior que 0.

**Questão (ii)**

Estimando o modelo proposto por MQO Agrupado, vê-se que o coeficiente de *exec* é estatisticamente insignificante, possuindo um p-valor de 0,34. Assim, as estimativas de MQOA indicam que o número de execuções não parece ter efeito sobre a taxa de homicídios.

Recibo com texto preto sobre fundo branco

Descrição gerada automaticamente

**Questão (iii)**

Estimando o modelo proposto por Efeitos Fixos, vê-se que o coeficiente de *exec* é negativo e estatisticamente significante: mais execuções penais diminuem a taxa de mortalidade do estado. A diferença frente a estimativa de MQOA se deve ao fato de ela ser viesada, já que, muito provavelmente, *cov(ai,xit) ≠ 0* – a cultura estadual ou até mesmo a liberação do porte de armas podem afetar as execuções e a taxa de desemprego.

**Tabela

Descrição gerada automaticamente**

**Questão (v)**

O estado com o maior número de execuções entre 1991 e 1993 foi o Texas, com 34 execuções – mais que o triplo observado no segundo colocado (Virginia, 11).

**Gráfico, Gráfico de barras, Histograma

Descrição gerada automaticamente**

**Questão (vi)**

Retirando Texas dos dados, obtém-se as seguintes estimativas, as quais mostram que o efeito de *exec* passa a ser estatisticamente insignificante após a retirada do estado com a maioria das execuções.

Contudo, ao contrário de regressões em cortes transversais, não é possível replicar o resultado adicionando uma dummy para o estado do Texas, uma vez que a variável é constante ao longo do tempo.

Tela de celular com texto preto sobre fundo branco

Descrição gerada automaticamente

**Questão (vii)**

Usando os dados de 1987 a 1993 e incluindo o estado do Texas, vê-se que *exec* possui um coeficiente estatisticamente não-significante, ao contrário do encontrado usando apenas os dados de 1990 a 1993. Apesar disso, a estimativa do parâmetro de *exec* aumenta (em módulo), passando de -0,10 para -0,1383.

Tabela

Descrição gerada automaticamente

## **Exercício C9**

**Questão (i)**

Tabela

Descrição gerada automaticamente

Ao fazermos a regressão requerida pelo modelo de MQO, temos que o efeito da variável de maior interesse (*choice*) é de 11,74 pontos em *pctstck*. Além disso, ela é relativamente significante estatisticamente, sendo relevante à 10%.

**Questão (ii)**

Aparentemente não, pois como pode ser visto na regressão acima, nenhuma dessas variáveis de controle é estatisticamente significativa. Conjuntamente, esses resultados de insignificância estatística se mantem, tendo um *p* valor de 0,417.

**Questão (iii)**

Tabela

Descrição gerada automaticamente

Como pode ser visto no mecanismo feito acima, apesar de o número total de observações ser 194, as observações únicas ou diferentes entre si são apenas 171.

**Questão (iv)**

Tabela

Descrição gerada automaticamente

Comparando esse modelo de desvios-padrão robustos agrupados com a regressão inicial, pode-se perceber que há uma alteração muito leve nas variáveis, especialmente na variável de interesse (*choice*) que passa de 6,232 para 6,198. Isso não é surpreendente pois 171 das 194 são independentes umas das outras.

**Questão (v)**

Primeiro nós geramos as variáveis necessárias para fazer a regressão pelo método das primeiras diferenças nesse painel que não possui a variável tempo. Em seguida, também eliminamos as observações que não representam casais.

Interface gráfica do usuário, Texto

Descrição gerada automaticamente

Então, podemos fazer a regressão:

Texto

Descrição gerada automaticamente

Uma imagem contendo Texto

Descrição gerada automaticamente

As variáveis sobre renda e outros fatores foram notavelmente abandonadas no modelo pelo fato de serem correspondentes ao casal, ou seja, não possui variação quando fazem a primeira diferença entre um cônjuge e outro.

**Questão (vi)**

Nenhuma das outras variáveis restantes é estatisticamente significante. Isso não é surpreendente pois é muito pequena a amostra de observações que se encaixam na restrição de cônjuges posta.

## **Exercício C12**

**Questão (i)**

Padrão do plano de fundo

Descrição gerada automaticamente com confiança baixa

Interface gráfica do usuário, Texto

Descrição gerada automaticamente

Como pode ser visto na função acima, o menor número de escolas em um distrito é de 1 escola e o maior número é 162 escolas, com sua média sendo 22,34.

**Questão (ii)**

Regressão por MQO:

Tabela

Descrição gerada automaticamente

Regressão por MQO agrupado entre as 1848 escolas: Tabela

Descrição gerada automaticamente

Como é esperado em ambos os casos o coeficiente de *bs* é -0,516, porém no MQO comum seu erro padrão é 0,101, mas para o MQO agrupado é 0,283.

**Questão (iii)**

Tabela

Descrição gerada automaticamente O erro padrão nessas condições para a variável *bs* é de 0,253, e seu *t* valor cai (em módulo) para –2,04.

**Questão (iv)**



Tabela

Descrição gerada automaticamente

Para o caso desse tratamento na variável *bs* há uma queda significativa no coeficiente, de -0,561 para -0,1864, em contrapartida seu erro-padrão robusto quase não se altera, de 0,253 para 0,273. Dessa forma, vemos muito pouco *tradeoff* entre os benefícios e o salário.

**Questão (v)**

Primeiro, definimos a variáveis de identificação e o equivalente à variável de tempo nesse painel sem que não possui observações temporais.



Assim, podemos fazer a regressão pelo método das primeiras diferenças:

Tabela

Descrição gerada automaticamente

Nesse caso, vemos que o coeficiente de *bs* volta a patamares mais altos (-0,523) indicando um *tradeoff* mais relevante entre benefícios e salário.

**Questão (vi)**

Ao tratarmos o painel para as observações menores do que 0,5 em *bs*, temos a ausência de evidências significativas para apontar um *tradeoff* entre os benefícios e o salário recebido pelos professores observados. Porém, ao tratarmos para os efeitos fixos dentro de cada distrito podemos ver que há pelo menos alguma evidência da existência desse *tradeoff.*

## **Exercício C13**

**Questão (i)**

*Totfatrte* é a taxa de fatalidade por 100.000 pessoas. Sua média em 1980 era 25,49; em 1992, 17,15; e em 2004, 16,72.

Ao regredirmos *totfatrte* sobre as dummies de cada ano, vê-se que cada dummy possui um coeficiente menor que a do ano anterior, indicando que a mortalidade no trânsito diminuiu ao longo do tempo.

Tabela

Descrição gerada automaticamente

**Questão (ii)**

Ao adicionarmos outros controles, os coeficientes de *bac08* e *bac10* indicam que estados com limites de álcool no sangue possuem menores taxas de fatalidade no trânsito. Além disso, um limite menor (*bac08*) possui um efeito redutor maior do que o de *bac10*.

Além disso, a estimativa do coeficiente que capta a existência de leis que permitam a cassação da licença de motorista (*perse*) tem um sinal negativo e estatisticamente significante, com um p-valor (usando erros padrões robustos para corrigir para correlação serial inerente a MQO Agrupado) de 0,0223.

Por fim, a estimativa do uso de um cinto de segurança primário (*sbprim*) é estatisticamente não-significante, com um p-valor (com erros padrões robustos) de 0,869.

Texto preto sobre fundo branco

Descrição gerada automaticamente

Tabela

Descrição gerada automaticamente

**Questão (iii)**

Ao usarmos um modelo de efeitos fixos, todas as variáveis descritas na questão anterior possuem sinal negativo e estatisticamente significante, o que vai de acordo com o que era esperado: limites de álcool no sangue, uso de cinto de segurança e leis mais restritivas limitam a mortalidade no trânsito. Ao contrário do modelo anterior, o uso do cinto de segurança primário passa a reduzir significantemente a taxa de mortalidade.

Como a estimativa de MQO Agrupado será viesada se qualquer uma das variáveis explicativas for correlacionada com características não-observadas da cidade – como a qualidade das estradas -, prefere-se o modelo de Efeitos Fixos. Ademais, as estimativas de EF estão de acordo com o que se espera pela realidade empírica.

Tabela

Descrição gerada automaticamente

**Questão (iv)**

Ao assumirmos que o total de milhas dirigidas per capita aumenta em 1000x, vemos que, para cada nova milha dirigida, a taxa de fatalidade no trânsito/100.000 pessoas diminui em 0,000009 – um efeito quase nulo, mas ainda estatisticamente diferente de zero.

A estimativa de *vehicmilepcs\_1000* é igual a de *vehicmiles*/1000, já que, apenas alteramos a unidade de medida da variável. Os coeficientes de regressão têm uma interpretação ceteris paribus (ou seja, de derivadas parciais); assim, o efeito de dirigir uma a milha a mais no ano continua o mesmo independentemente da unidade de medida.

**Questão (v)**

Estimando o modelo por EF e considerando erros padrões robustos para os agrupamentos (ou seja, para os estados), as estimativas do modelo não se alteram, mas as variáveis de política (*bac08, bac10, perse, sbprim, sbsecon, sl70plus* e *gdl*) têm sua significância estatística diminuída; *bac10*, por exemplo, não é mais significante a 5% contra uma alternativa bilateral.

Tabela

Descrição gerada automaticamente

## **Exercício C14**

**Questão (i)**

Há, no máximo, 1149 diferentes médias temporais (números distintos de *ids*). A maior média temporal de *concen* é 0,9997 (id 768) e a menor é 0,1861 (id 174).

Texto

Descrição gerada automaticamente

**Questão (ii)**

O modelo proposto na questão equivale a estimação por Efeitos Aleatórios Correlacionados e, como esperado, a estimativa de *concen* é igual àquela obtida por Efeitos Fixos (0,1689):

Recibo com texto preto sobre fundo branco

Descrição gerada automaticamente

**Questão (iii)**

Ao retirarmos *ldist* e *ldistsq* da estimação de EA, o coeficiente de *concen* fica inalterado, enquanto o de *concenbar* se torna negativo:

Tabela

Descrição gerada automaticamente

**Questão (iv)**

Os resultados do modelo de EA da questão (ii) mostram que o estimador de (o parâmetro de *concenbar*) tem um p-valor de 0,0017, ou seja, rejeita-se H0: até a 1% de significância.

Como esse coeficiente mede, em última instância, a correlação entre *ai* e as variáveis explicativas, vê-se que há, de fato, correlação entre os efeitos não-observados e as variáveis dependentes. Assim, EA se mostra viesado e prefere-se o estimador de EF.

**Questão (v)**

Usando um erro robusta com relação a heteroscedasticidade e à autocorrelação serial, o coeficiente de *concenbar* passa a ter um p-valor de 0,48, de modo que não se rejeita H0:.

Assim, como ele mede a correlação entre os efeitos não-observados e as variáveis explicativas, pode-se argumentar que não há essa correlação. Portanto, apesar de ambos os estimadores serem não-viesados e consistentes, prefere-se EA por ser mais eficiente.

Texto preto sobre fundo branco

Descrição gerada automaticamente

1. Lembrando que, assim como feito acima, a média 0 dos erros idiossincráticos reduz a covariância à esperança da multiplicação das variáveis. [↑](#footnote-ref-1)