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

**Observação:** parte da lista foi realizada usando um notebook em Python. Ao final do documento há o código com as definições das funções utilizadas. O notebook com todos os códigos e cálculos foi enviado juntamente com esse PDF.

## **Exercício 2**

A estimativa de *a81proxincin* na equação (13.9) é igual a -0,063, sendo bem menor que as apresentadas nas equações do exercício. Isso ocorre porque (13.9) controla pra *proxincin* e para *a81* simultaneamente: enquanto o coeficiente de *proxincin* captura diferentes já pré-existentes entre os grupos, o de *a81* capta efeitos estruturais do tempo sobre os preços das residências.

No primeiro modelo, que omite *a81*, todo o efeito natural de valorização de preços nominais das casas é captado pelo coeficiente de *a81proxincin*, o que o torna bastante positivo. Já no segundo – que omite *proxincin* -, não é possível perceber o fato de que casas perto do local do incinerador já possuíam menores preços antes mesmo da construção. Em ambos os casos, há omissão de variáveis importantes e que estão, por natureza do termo de interação, correlacionadas com *a81proxincin*, de modo que os estimadores se tornam viesados.

## **Exercício 3**

O uso de primeiras diferenças não é possível em agrupamentos independentes de cortes transversais justamente porque o efeito não observado – *ai* – não será o mesmo para as observações, já que muito provavelmente irão conter indivíduos diferentes nos dois anos analisados.

## **Exercício 6**

**Questão (i)**

A situação da questão ilustra um exemplo típico de experimento natural, de modo que é possível usar o seguinte modelo simples de diferença em diferenças:

No modelo acima, *d90* é uma dummy para as observações de 1990 e *dFL* é uma dummy para as observações do estado da Florida que, por ter passado a lei proibindo recipientes abertos de bebidas alcoólicas nos veículos, será tratado como o grupo do tratamento. Portanto, as observações da Georgia serão o grupo controle.

Quanto aos coeficientes, estima os efeitos estruturais da passagem do tempo entre 1985 na probabilidade de uma pessoa ser presa por dirigir embriagada. , por sua vez, indica se há diferenças pré-existentes entre os estados nas probabilidades de prisão por direção sob embriaguez. Portanto, têm-se que, em 1985:

Já em 1990:

Assim, o Efeito do Tratamento nos Não-Tratados é , enquanto o Efeito do Tratamento nos Tratados é . O Efeito do Tratamento – ou seja, da implementação da lei - é, portanto, .

**Questão (ii)**

Seria interessante controlar para outros fatores em virtude da existência de outros pontos que podem alterar a probabilidade de prisão por embriaguez, como a magnitude do consumo de bebida alcóolica em 1985 e sua mudança entre 1985 e 1990. Caso esse fator esteja correlacionado com algum parâmetro do modelo anterior, o viés das estimações poderá ser reduzido; caso não, a adição de mais variáveis aumenta o R2, diminuindo o SQR e aumentando as precisões das estimações.

**Questão (iii)**

Usando os dados a nível municipal, pode-se interpretar a fração dos motoristas habilitados presos por dirigem embriagados como o *número de prisões por direção sob embriaguez per capita*. Assim, o modelo deixaria de ser de probabilidade linear para ser uma estimação de uma variável dependente “comum”, mas medida a níveis per capita.

Contudo, é preciso atentar para o que é dito na seção (8.4) do livro do Wooldridge: assumindo que os dados a níveis individuais satisfazem as hipóteses de Gauss-Markov, dados per capita sofreram, naturalmente, de heteroscedasticidade.

Isso ocorre por conta da natureza de um dado per capita, ou seja, a soma do total dos valores individuais divididos pelo tamanho da população. Assim, se *um,i*é o resíduo da observação *i* de um munícipio *m* com população *p*:

Ou seja, o erro médio do munícipio *m* terá uma variância proporcional ao inverso do tamanho da população, de modo que o modelo pode ser feito usando Mínimos Quadrados Ponderados com pesos (ou caso a equação tomada para a definição seja a de transformação das variáveis explicativas e não a dos quadrados dos resíduos).

## **Exercício 7**

**Questão (i)**

A estimativa de *apmud\*altrenda* na questão é próxima à obtida com *apmud* no modelo em virtude da baixa significância econômica e estatística do efeito da ampliação de limite para as pessoas cuja renda não é tão alta.

**Questão (ii)**

A retirada de *altrend* faz com que o coeficiente do termo de interação seja maior porque, sem ele, pressupõe-se que não há diferenças nos tempos de afastamento entre pessoas com baixa e alta renda antes da mudança, o que não é verdade – praticamente o inverso do que ocorre com a retirada de *apmud* proposta na questão (i). Portanto, o coeficiente do termo de interação tem um valor alto porque está capturando diferenças já pré-existentes nos tempos de afastamento remunerado entre diferentes classes sociais.

## **Exercício C1**

**Questão (i)**

O modelo do exemplo 13.1 está na imagem abaixo. Aplicando a estatística F para testar *H0:ßfarm= ßothrural =ßtown= ßsmcity=0*, vê-se que o p-valor do teste de significância conjunta é 0,327. Portanto, não se rejeita H0, ou seja, as variáveis ambientais não são conjuntamente estatisticamente significantes para determinar a fertilidade aos 16 anos.

Tabela

Descrição gerada automaticamente

**Questão (ii)**

Fazendo um teste F para testar a significância conjunta das variáveis regionais, vê-se que seu p-valor é 0,029. Assim, se rejeita a hipótese *H0:ßeast= ßnorthcenl =ßwest=0* a 5% de significância, ou seja, fatores regionais são importantes para determinar a fertilidade aos 16 anos.

Interface gráfica do usuário, Texto

Descrição gerada automaticamente

**Questão (iii)**

A regressão dos quadrados dos resíduos do modelo original sobre o conjunto de dummies temporais indica se há variação intertemporal da variância do erro ou, em outras palavras, se o erro é heteroscedástico no tempo. Assim, a hipótese nula de homoscedasticidade intertemporal é *H0:ßj=0* *para todo j*, onde o conjunto *ßj­* representa os parâmetros das dummies. Os resultados podem ser vistos abaixo:

Tabela

Descrição gerada automaticamente

Como se pode ver na tabela acima, o teste F de significância conjunta de todas as dummies fornece um p-valor de 0,0082, ou seja, há fortes indícios de heteroscedasticidade no modelo até ao nível de 1%.

**Questão (iv)**

Os termos de interação entre os anos de educação e as dummies temporais indicam a **mudança** na influência da educação sobre a fertilidade ao longo do período analisado. Assim, a soma dos coeficientes de *y84* e *y84educ* indica o retorno da educação em 1984.

O teste F dos termos de interação resulta em um p-valor 0,18, ou seja, os termos de interaçãosão conjuntamente não significantes até ao nível de 15%: isso indica que não há mudanças no impacto da educação sobre a fertilidade ao longo do tempo.

Tabela

Descrição gerada automaticamente

## **Exercício C2**

**Questão (i)**

O coeficiente de *y85* mostra o percentual de mudança no salário, entre 1978 e 1985, decorrentes de efeitos temporais agregados. Contudo, como há um termo de interação *y85educ,* o coeficiente de *y85* mede a mudança intertemporal no salário para aqueles homens cuja escolaridade é de 0 anos, algo muito atípico e pouco interessante para a análise.

**Questão (ii)**

Para um homem com 12 anos de educação, a mudança no retorno da educação entre 1978 e 1985 é

Assim, a mudança seria igual a aproximadamente 33,98%. Para obter os intervalos de confiança, estamos interessados em estimar o parâmetro:

De forma que o modelo resultante é

Portanto, estamos interessados nos valores relacionados ao coeficiente da dummy temporal, . As estimativas do modelo são:

Tabela

Descrição gerada automaticamente

O intervalo de confiança a 95% de significância de é [0,273, 0,406], de modo que se pode dizer, com grande significância estatística, que houve um aumento percentual aproximado que varia de 27,3% a 40,6% para os homens com 12 anos de educação entre 1978 e 1985.

**Questão (iii)**

Como a variável dependente está em forma de logaritmo, podemos encontrar o salário real em 1985 pela expressão , onde log é o logaritmo natural.

Ao regredir o logaritmo do salário real sobre as variáveis dependentes do modelo anterior, vê-se que só há mudança no intercepto de 1985, o que era esperado, já que:

Tabela

Descrição gerada automaticamente

**Questão (iv)**

Como o SQR é o mesmo, a única outra explicação para a diminuição do R2 de 0,426 para 0,356 é um aumento do SQTy, o que ocorre justamente porque houve uma mudança nas variáveis dependentes: o logaritmo do salário real possui, portanto, uma variação maior do que o logaritmo do salário nominal.

**Questão (v)**

De 1978 para 1985, a participação sindical caiu cerca de 12,5% entre 1978 (30,5%) e 1985 (17,98%).

Interface gráfica do usuário, Texto

Descrição gerada automaticamente

**Questão (vi)**

Para verificar se o efeito dos sindicatos mudou ao longo do tempo, pode-se adicionar o termo de interação *d85union* ao modelo original. Cabe dizer que em 1978 o retorno parcial da sindicalização era ≈20,23% e estatisticamente muito significante.

Ao analisar o coeficiente do termo de interação, vê-se que ele é tanto economicamente quanto estatisticamente insignificante, com um p-valor de 0,995, ou seja, não há indícios que o retorno da sindicalização tenha mudado ao longo do tempo.

Tabela

Descrição gerada automaticamente

**Questão (vii)**

A análise mostra que os achados são, em primeira análise, conflituosos, uma vez que, mesmo que o efeito da sindicalização não tenha aumentado entre 1978 e 1985, ele continua contribuindo em aproximadamente 20,23% para um salário maior. Assim, não há incentivos para os agentes deixarem os sindicatos, o que mostra que há outras pressões ambientais contribuindo para a diminuição da adesão aos sindicatos.

Uma possível explicação é a aceleração da globalização e a enxurrada de mão-de-obra barata proveniente das cadeias globais de valor, que pressionou os custos locais e pode ter forçado trabalhadores a saírem de sindicatos a fim de diminuir os salários e, em última instância, conseguir manter o seu emprego.

Isso fica evidente no documentário “[**American Factory**](https://www.netflix.com/title/81090071)” que, apesar de retratar eventos do século XXI, mostra a pressão dos empregadores pela não-adesão dos empregados aos sindicatos trabalhistas locais.

## **Exercício C3**

**Questão (i)**

Se a construção do incinerador efetivamente reduziu o preço das casas da redondeza, espera-se que seja positivo: residências com menor distância ao local do incinerador têm valores menores do que aquelas mais distantes.

Caso seja positivo, isso indica que casas mais afastadas do incinerador já possuíam preços maiores, ou seja, que a construção foi feita em áreas cujas casas já possuíam um valor de mercado menor.

**Questão (ii)**

O resultado do modelo proposto no exercício (i) é:

Tabela

Descrição gerada automaticamente

O coeficiente de mede a mudança no impacto da distância de uma casa do local do incinerador sobre o preço após a sua construção. O coeficiente, contudo, é nada significante, com um p-valor de 0,55. Portanto, conclui-se que a construção do incinerador não afetou o preço das casas na região (o que é corroborado também pelo coeficiente também não significante de *y81*).

**Questão (iii)**

O resultado do modelo proposto é:

Tabela

Descrição gerada automaticamente

O coeficiente de continua positivo e, apesar de não-significante até mesmo ao nível de 20%, sua estimação é feita com maior precisão. Contudo, os resultados ainda indicam que a instalação do incinerador não provocou mudanças no preço das casas da região.

**Questão (iv)**

O coeficiente de *ldist* deixa de ser significante no modelo (iii) justamente porque se adicionou outros controles ao modelo. Assim, casas mais distantes do incinerador tem preços maiores não porque estão mais distantes, mas sim por terem terrenos maiores, um maior número de quartos e banheiros, serem mais novas, etc.

## **Exercício C4**

**Questão (i)**

Com um modelo mais completo – levando em conta status civil, gênero e dummies de tipo de lesão e indústria de origem -, o coeficiente de *afhigh* aumenta com relação ao modelo da equação (13.2), sendo significante até mesmo ao nível de 1%. Assim, pessoas de maior renda que sofreram algum tipo de lesão ficaram afastadas por um período de tempo de, em média, aproximadamente 23,09% maior do que antes do aumento do limite da indenização.

Tabela

Descrição gerada automaticamente

**Questão (ii)**

O pequeno R2 (0,041) mostra que há muitos outros fatores – como a severidade da lesão, por exemplo – que influenciam a duração do afastamento coberto por indenização governamental. Assim, a previsão do tempo de duração para um indivíduo seria bastante difícil.

Contudo, como a amostra tem mais de 5.000 observações, isso permite ainda encontrar estimações bastante precisas para os efeitos causais parciais de cada uma das variáveis (ao menos aquelas estatisticamente significativas, como é o caso de *afhigh*). Isso depende, claro, da hipótese de que, mesmo com tantos fatores delegados ao termo de erro, E (u|x) = 0.

**Questão (iii)**

Usando o modelo da equação (13.12), a estimativa de *afhigh* (0,1962) para os 1.475 dados do estado de Michigan acaba sendo maior que a obtida usando as 5.626 observações de Kentucky (0,191), mas não é significante contra uma alternativa unilateral nem ao nível de 10%, tendo um p-valor bilateral igual a 0,208.

Tabela

Descrição gerada automaticamente

## **Exercício C5**

**Questão (i)**

Usando MQO agrupado, os resultados do modelo são:

Tabela

Descrição gerada automaticamente

O coeficiente de *y90* é estatisticamente significante e indica, tudo o mais constante, que houve um aumento aproximado de 26,22% nos preços médios dos aluguéis entre 1980 e 1990. Já o coeficiente de *pctstu* mostra que, tudo o mais constante, um aumento de 1% no percentual de estudantes frente a população total quase não tem efeito sobre os aluguéis, aumentando-os em aproximadamente 0,5%.

**Questão (ii)**

Os erros padrão da equação anterior não serão válidos se características não-observadas da cidade – como seu clima – estiverem correlacionadas com qualquer uma das variáveis explicativas usadas.

**Questão (iii)**

Usando o método de Primeiras Diferenças, o coeficiente de ∆*pctstu* indica que a mudança no percentual estudantil da população tem um impacto mais significativo sobre o preço dos aluguéis em 1990, sendo que uma mudança de 1% implica, tudo o mais constante e supondo que as hipóteses RLM.1-4 sejam satisfeitas, em um aumento de 1,12% nos aluguéis médios.

Tabela

Descrição gerada automaticamente

**Questão (iv)**

Aplicando erros robustos com relação à heteroscedasticidade, vê-se que o erro padrão de ∆*pctstu* diminui de 0,004132 para 0,002936, tornando seu coeficiente ainda mais estatisticamente significante.

Tabela

Descrição gerada automaticamente

## **Exercício** **C7**:

**Questão (i)**

Usando MQO agrupado, os resultados do modelo são:

Tabela

Descrição gerada automaticamente

O modelo mostra que o coeficiente da variável explicativa *season* é de –0,0272. Como essa variável indica a mudança na performance dos atletas (medida pelo GPA) durante a temporada de jogos, o resultado do coeficiente associado a tal variável mostra que durante essa temporada o desemprenho dos atletas diminui nessa taxa, tudo o mais constante.

**Questão (ii)**

Se, de fato a habilidade dos jogadores de futebol americano não é bem capturada por essas outras variáveis, os efeitos desse fator estarão correlacionados com a variável *season*. Assim, com o fato de essa habilidade específica que é absorvida pelo termo de erro estar relacionada com uma variável explicativa tornará de fato as variáveis do modelo viesadas e inconsistentes.

**Questão (iii)**

Uma imagem contendo Tabela

Descrição gerada automaticamente

Como o sumario da regressão acima mostra, as varáveis que não possuem variação com o tempo (mudança de estação), e, portanto, são retiradas do modelo, são: *spring*, *sat*, *hsperc*, *female*, *black* e *white*.

Texto

Descrição gerada automaticamente com confiança média

O teste F realizado acima mostra que, apesar de a impacto do termo *season* ter aumentado, com seu coeficiente passando de –0,0272 para –0,0645, essa variável não possui uma grande relevância estatística, sendo relevante apenas a 15%.

**Questão (iv)**

Pode-se argumentar que a ocorrência de lesões em atletas é um fator a ser considerado, especialmente se tratando de esportes de impacto, como o próprio futebol americano. Essa variável pode influenciar fortemente o desempenho de um atleta de um período para o outro se esse estiver em um período de recuperação.

Por um olhar, se a ocorrência de lesões for vista como aleatória entre os atletas esse fator poderia ser negligenciado à variável de erro. Entretanto, se considerado que há uma maior chance de ocorrer lesões em atletas que já praticam o esporte há mais de um semestre poderá haver, então, uma correlação do termo de erro com a variável explicativa *frstsem* que indica se o estudando está no primeiro semestre ou não.

## **Exercício** **C10**

**Questão (i)**

Estimando a equação usando o método de Primeiras Diferenças (que pode ser calculada usando um intercepto considerando uma dummy apenas para o terceiro ano), encontramos os seguintes resultados:

Tabela

Descrição gerada automaticamente com confiança média

O número de observações usadas na regressão é 251, mas para encontrar o número de firmas cujos dados são não-nulos pode-se contar o número de *fcodes* únicos:



Caso o painel fosse equilibrado, todas as firmas possuiriam dados para os três anos. Como são 451 observações, o número de firmas cujos dados poderiam ser utilizados seriam 451/3 = 127. Como o método de Primeiras Diferenças “descarta” o primeiro ano de observações, a regressão seria feita utilizando 314 observações.

**Questão (ii)**

O coeficiente de ∆*grant* é tanto estatisticamente quanto economicamente significante (≈32,6), indicando que o recebimento de uma concessão governamental aumenta as horas médias de treinamento de uma firma em cerca de 32 horas a cada ano.

**Questão (iii)**

Como a verba de treinamento de um ano é majoritariamente usada no ano de seu recebimento, não é surpreendente que o coeficiente *∆grant-1* é estatisticamente não-significante, possuindo um alto p-valor (0,72).

**Questão (iv)**

O uso de *lemploy* pode ser visto como uma proxy para o tamanho das firmas; o coeficiente tem um valor muito baixo (0,744) e um p-valor bastante alto (0,879), indicando que ele não é estatisticamente significante. Assim, conclui-se que não há diferença significativa nas horas de treinamento entre empresas grandes e pequenas.

## **Exercício** **C13**

**Questão (i)**

Tabela

Descrição gerada automaticamente

Todos os parâmetros podem ser estimados nesse modelo (*d81, d82, d83, d84, d85, d86, d87* e suas respectivas interações com *educ*, assim como *union*), exceto pelo intercepto, pois esse é posto para fora no modelo na subtração, em que ele é constante e acaba anulado.

**Questão (ii)**

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Assim como demonstrado no teste F acima, não há evidência significante de que o retorno da educação tenha aumentado no período observado.

**Questão (iii)**

Tabela

Descrição gerada automaticamente

Texto

Descrição gerada automaticamente

Rodando o modelo de forma robusta e fazendo os mesmos testes do item (ii), temos resultados diferentes para o teste F, ainda assim, a interpretação não muda, sendo insignificante a evidência de que o retorno da educação tenha aumentado no período observado.

**Questão (iv)**

Tabela

Descrição gerada automaticamente Tabela

Descrição gerada automaticamente

Como o extenso modelo acime demonstra, o diferencial por fazer parte de uma união calculado para 1980 é o parâmetro base de *union* sem nenhuma interação de ano, ou seja, 0,105. E o mesmo diferencial para o ano de 1987 é dado pela soma do diferencial base de 1980 (0,105) com o parâmetro da variável de interação de *union* com o ano de 1987 (-0,147), ou seja, a subtração resulta num diferencial negativo de aproximadamente –0,041.

**Questão (v)**

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Como o teste acima aponta, não há evidências relevantes de que o retorne de pertencer a uma união no modelo apresentado muda conforme os anos observados, pois seu p-valor (0,3346) é alto e só significaria relevância à 35%.

## **Anexo: definições das funções utilizadas**

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente