

Exercício 3 - Escolher qualquer base de dados desde que ela esteja no formato painel.

- a) Explique as variáveis importantes utilizando algumas estatísticas descritivas.

A base de dados utilizada neste exercício ("Grunfeld") pertence do pacote ERA e é formada pela variável dependente nível de investimento ("invest") de diversas empresas ("firm"), bem como pelas variáveis explicativas valor de mercado ("value") e valor do estoque de capital ("capital") durante o período de 1935-1954 (20 anos).

Foram selecionadas quatro empresas ("General Electric", "General Motors", "US Steel" e "Westinghouse") para fins de análise dos dados. Utiliza-se o pacote plm e a função pdata.frame para alocar a base de dados para a análise de regressão de dados em painel, uma vez que é necessário definir o atributo individual ("firm") e temporal ("year") das observações. Para isso utiliza-se o argumento index.

Posteriormente, extraiu-se as estatísticas básicas das séries para fins de explicação do comportamento dessas. Os resultados são apresentados a seguir:

invest	value	capital	firm	year
Min. : 12.93	Min. : 191.5	Min. : 0.8	General Motors : 20	1935:04
1st Qu.: 55.27	1st Qu.: 1192.3	1st Qu.: 118.1	US Steel : 20	1936:04
Median : 199.75	Median : 1971.2	Median : 254.7	General Electric : 20	1937:04
Mean : 290.92	Mean : 2229.5	Mean : 357.3	Westinghouse : 20	1938:04
3rd Qu.: 459.77	3rd Qu.: 2795.0	3rd Qu.: 368.9		1939:04
Max. : 1486.70	Max. : 6241.7	Max. : 2226.3		1940:04

Nota-se que nesta base que será trabalhada os dados das empresas aparecem "empilhados", uma vez que a variável referente ao ano da observação ("year") está repetida para cada observação da referida empresa (corte transversal repetido em diversos períodos de tempo). Desta forma a nossa base de dados possui igualmente 20 informações para cada empresa, se constituindo em um painel equilibrado. Caso o número de informações para cada empresa fosse desiguais, teríamos um painel desequilibrado.

- b) Estime uma regressão por MQO e execute alguns testes para ver a significância conjunta de alguns parâmetros do modelo.

Modelo Pooled

Este modelo trata de "empilhar" todas as observações da base de dados, ignorando a estrutura de dados em painel. Desta forma, todas as observações são tratadas como não correlacionadas para os indivíduos, com erros homoscedásticos. Trata-se, portanto, da forma mais simplista e ingênua pois desconsidera as dimensões de tempo e espaço combinados, ao mesmo tempo que estima a regressão pelo método dos Mínimos Quadrados Ordinários (MQO). Assim:

$$Y_{it} = \beta_1 + \beta_2 X_{2it} + \beta_3 X_{3it} + u_{it}$$
$$i = 1, 2, 3, 4$$
$$t = 1, 2, \dots, 20$$

em que i corresponde à i -ésima unidade de corte transversal e t o t -ésimo período de tempo.

Para executar este modelo de regressão utiliza-se a função plm, juntamente com as variáveis dependente e independentes, indicando a base de dados (data) e o tipo da regressão ("pooling"). Os resultados são apresentados a seguir:

Pooling Model

Call:

```
plm(formula = invest ~ value + capital, data = Grunfeld, model =  
"pooling")
```

Balanced Panel: n = 4, T = 20, N = 80

Residuals:

Min.	1st Qu.	Median	3rd Qu.	Max.
-319.6766	-99.9523	1.9647	65.9905	336.2072

Coefficients:

	Estimate	Std. Error	t-value	Pr(> t)	
(Intercept)	-62.8318	29.72539	-2.1137	0.03778	*
value	0.110521	0.013776	8.023	9.19E-12	***
capital	0.300463	0.049399	6.0823	4.27E-08	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 6410400

Residual Sum of Squares: 1572700

R-Squared: 0.75466

Adj. R-Squared: 0.74829

F-statistic: 118.424 on 2 and 77 DF, p-value: < 2.22e-16

A estimação da regressão *pooled* apresentou alta significância estatística nas variáveis dependentes ("value" e "capital"), indicando sinal positivo para os coeficientes, conforme esperado, bem como um valor de R2 elevado. Este tipo de modelo não diferencia entre a influência/diferença das empresas na variável investimento e nem se a resposta do investimento às variáveis explicativas é a mesma ao longo do tempo. Isto faz com que não se saiba se existe heterogeneidade entre as empresas. A comparação do modelo *pooled* com as regressões de efeitos fixos e efeitos aleatórios, que serão estimados na sequência, servirá mostrar ao pesquisador qual é o melhor modelo dentre eles.

- c) Estime esta regressão com efeitos fixos. Todos os coeficientes do modelo anterior podem ser estimados? Explica e mostra por quê?

O modelo de regressão com efeitos fixos considera, como visto anteriormente, que os valores dos interceptos para cada regressão (α_i) variam de acordo com o efeito de cada indivíduo ("firma") e que os coeficientes de declividade (das variáveis independentes "value" e "capital") para cada equação são os mesmos para cada empresa, conforme equação abaixo:

$$invest_{it} = value_{it} + capital_{it} + \alpha_i + \varepsilon_{it}$$

em que $i = 1, \dots, 4$, $t = 1, \dots, 20$ (painel balanceado).

Desta forma, o intercepto da equação é diferente para cada empresa, mas o efeito das variáveis independentes é o mesmo sobre a variável dependente. Isto indica que existe características

especiais em cada empresa influenciando o investimento, como por exemplo o estilo de gestão. A seguir apresenta-se os resultados da regressão com efeitos fixos:

Oneway (individual) effect Within Model

Call:

plm(formula = invest ~ value + capital, data = Grunfeld, model = "within")

Balanced Panel: n = 4, T = 20, N = 80

Residuals:

Min. 1st Qu. Median 3rd Qu. Max.
-184.6581 -48.2612 9.3252 40.5471 197.6681

Coefficients:

	Estimate	Std.	Error	t-value	Pr(> t)
value	0.1084	0.017566	6.1711	0.00	***
capital	0.345058	0.026708	12.9195	0.00	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 2171500

Residual Sum of Squares: 422220

R-Squared: 0.80556

Adj. R-Squared: 0.79242

F-statistic: 153.291 on 2 and 74 DF, p-value: < 2.22e-16

Nota-se que o impacto do valor da empresa ("value") e do capital ("capital") é positivo sobre o investimento ("invest"), para todas as empresas como visto acima. Inclusive, há significância estatística para estas variáveis. No entanto, ainda resta definir o efeito dos interceptos de cada empresa, como segue:

Firmas	Estimate	Std.	Error	t-value	Pr(> t)
General Motors	-85.515	73.49	-1.1636	0.24831	
US Steel	94.988	36.664	2.5907	0.01153	*
General Electric	-246.228	35.938	-6.8515	1.86E-09	***
Westinghouse	-59.386	20.233	-2.9351	0.004439	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Com este resultado é possível observar que o efeito das firmas sobre o investimento parece ser diferente para cada indivíduo. Desta forma, somente a empresa US Steel consta com efeito positivo sobre o investimento. Por outro lado, a fórmula da regressão é apresentada de maneira diversa, por exemplo:

$invest = -85,515 + 0,108400value + 0,345058capital$, para a regressão considerando a General Motors;

$invest = 94.988 + 0,108400value + 0,345058capital$, considerando a US Steel e assim por diante.

- d) Compare a estimação acima com a estimação com efeitos aleatórios. Quais são as diferenças mais importantes que você observa? O que poderiam indicar estas diferenças?

No modelo de regressão com efeitos aleatórios, os efeitos individuais das firmas ("firms") são considerados variáveis aleatórias, ao contrário do modelo visto anteriormente. Desta forma:

$$Y_{it} = \beta_{1i} + \beta_2 X_{2it} + \beta_3 X_{3it} + u_{it}$$

em que β_{1i} é variável aleatória com valor médio β_1 , e o intercepto para a empresa individual é dado por:

$$\beta_{1i} = \beta_1 + \varepsilon_i \quad i = 1, 2, \dots, N$$

- e) Qual é o resultado do teste de Hausman?

O teste de Hausmann (Hausman 1978) efetua a especificação dos modelos de Efeito Fixo e de Efeitos Aleatórios, sendo que se o teste rejeitar a hipótese nula, o modelo de Efeitos Fixos é o mais adequado.

H0: α_i não são correlacionados com X_{it}

H1: α_i são correlacionados com X_{it}

A função utilizada para este teste foi `phptest`. O resultado é apresentado a seguir:

Hausman Test

data: invest ~ value + capital

chisq = 0.074639, df = 2, p-value = 0.9634

alternative hypothesis: one model is inconsistent

Pelo resultado do teste o modelo de efeitos aleatórios é mais adequado.

- f) Replique a estimação do modelo de efeitos fixos usando *dummies*.

Outra forma de visualizar a equação de efeitos fixos é utilizando a função `lm` para definir a regressão, definindo a variável "firm" como um fator:

Call:

`lm(formula = invest ~ value + capital + as.factor(firm), data = Grunfeld)`

Residuals:

Min	1Q	Median	3Q	Max
-184.658	-48.261	9.325	40.547	197.668

Coefficients:

	Estimate	Std. Error	t	Pr(> t)
(Intercept)	-85.5153	73.48978	-1.164	0.24831
value	0.1084	0.01757	6.171	0.000 ***
capital	0.34506	0.02671	12.919	0.000 ***
as.factor(firm)US Steel	180.503	45.71679	3.948	0.000 ***
as.factor(firm)General Electric	-160.712	46.62236	-3.447	0.001 ***
as.factor(firm)Westinghouse	26.12959	64.94348	0.402	0.689

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 75.54 on 74 degrees of freedom

Multiple R-squared: 0.9341, Adjusted R-squared: 0.9297

F-statistic: 209.9 on 5 and 74 DF, p-value: < 2.2e-16

Note que o intercepto definido (-85,51533) refere-se à presença da empresa General Motors. Caso seja evidenciada a presença da empresa US Steel, o valor do intercepto passa para 94,988 (= -85,51533 + 180,50295), a mesma lógica vale para as demais empresas.

g) Compare as estimações com efeitos fixos e aleatórios com a estimação em primeiras diferenças. Quais são as alterações mais importantes que você observa? O que poderiam indicar estas diferenças?

Oneway (individual) effect First-Difference Model

plm(formula = invest ~ value + capital, data = Grunfeld, model = "fd")

Balanced Panel: n = 4, T = 20, N = 80

Observations used in estimation: 76

Residuals:

Min.	1st Qu.	Median	3rd Qu.	Max.
-202.4563	-34.5701	-8.5034	35.1266	189.7779

Coefficients:

	Estimate	Std. Error	t-value	Pr(> t)
(Intercept)	-0.824600	9.078483	-0.0908	0.927876
value	0.090577	0.013417	6.7508	3.009e-09 ***
capital	0.302697	0.089806	3.3706	0.001202 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 550550

Residual Sum of Squares: 326830

R-Squared: 0.40635

Adj. R-Squared: 0.39009

Além de uma pequena mudança no valor dos coeficientes, observamos uma redução significativa no valor de R^2 que cai para 0,40.

h) É possível aplicar o teste de Hausman com o modelo de efeitos aleatórios e em primeira diferença?

O teste de Hausman é aplicado para escolha entre modelos de efeitos fixos e aleatórios e, no caso em questão, como $T > 2$ os modelos de EF e primeira diferença não são comparáveis.

i) Consegue imaginar um instrumento externo para alguma variável de interesse da sua estimação? Discute a adequação do instrumento proposto.

O variável valor de mercado talvez possa ser instrumentalizada por outra medida do tamanho do negócio da firma como o volume total de vendas ou faturamento. É esperado, por exemplo, que firmas maiores (i.e. com maior valor de mercado) possuam maior faturamento. Este, todavia, pode estar correlacionado com o estoque de capital.

j) Estime a equação novamente em primeiras diferenças com uso de uma ou mais variáveis instrumental

Modelo com defasagem na variável capital.

Oneway (individual) effect First-Difference Model

Call:

```
plm(formula = invest ~ value + lag(capital, 1L), data = Grunfeld,
     model = "fd")
```

Balanced Panel: n = 4, T = 19, N = 76

Observations used in estimation: 72

Residuals:

Min.	1st Qu.	Median	3rd Qu.	Max.
-226.356	-31.885	-12.734	27.773	250.112

Coefficients:

	Estimate	Std. Error	t-value	Pr(> t)
(Intercept)	8.897589	9.774321	0.9103	0.3658
value	0.090561	0.015015	6.0314	7.185e-08 ***
lag(capital, 1)	0.156819	0.111107	1.4114	0.1626

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 531890

Residual Sum of Squares: 346930

R-Squared: 0.34774

Adj. R-Squared: 0.32884

F-statistic: 18.3933 on 2 and 69 DF, p-value: 3.9577e-07