

State University of Campinas - UNICAMP

*CRESCIMENTO ECONÔMICO E O USO DE
ENERGIA RENOVÁVEL: UMA AVALIAÇÃO
COM DADOS EM PAINEL*

Trabalho Final HO-235

Supervisor: Alexandre Gori Maia

João Paulo F. Fenelon (211467) e Henri Makika (211042)

Julho 10, 2019



1 INTRODUÇÃO

Este trabalho analisa o impacto do crescimento sobre o uso de energia de fontes renováveis mediante o uso de técnicas de regressão. Supõe-se que é medida que a renda das economias aumenta ocorre um processo de substituição das fontes de energias fósseis pelas mais limpas (renováveis). Utiliza-se na primeira parte um modelo clássico baseado no método de mínimos quadrados ordinários (OLS), de modo a estimar os resultados para o ano de 2010; em seguida, usa-se o modelo de dados em painel para avaliar um período mais longo, de 1960 a 2018. Os dados foram extraídos do world development indicators do Banco Mundial. Finalmente, as conclusões são descritas na última seção.

2 ESTIMAÇÃO E DISCUSSÃO DOS RESULTADOS

São utilizados os dados do World Development Indicators (WDI) do Banco Mundial. Para análise de corte transversal utilizou-se o ano de 2010 como referência. Das 217 observações, 80 foram eliminadas devido à presença de missingness. As seguintes variáveis são consideradas ao longo do trabalho:

renovavel = produção de energia de fontes renováveis em kWh, excluindo hidroelétricas;

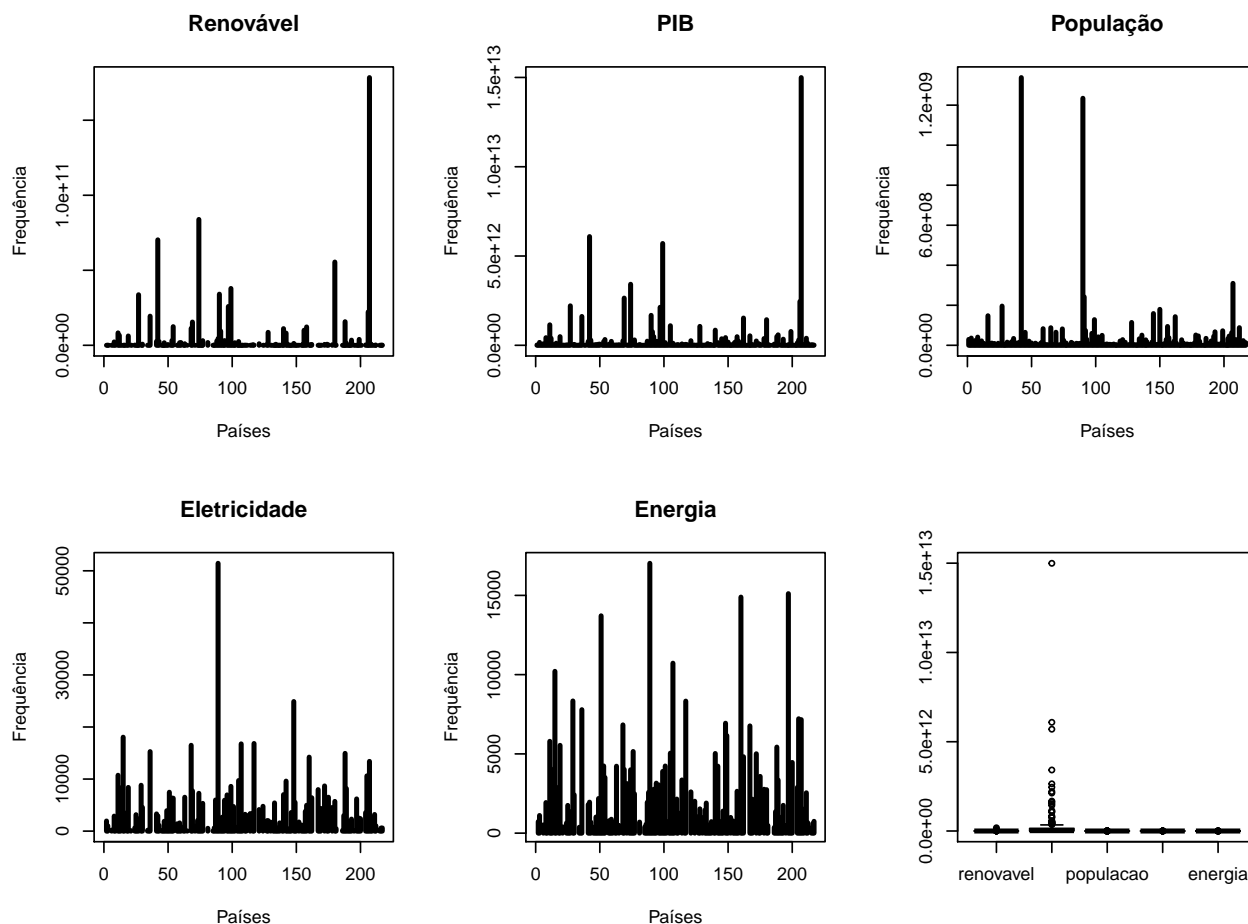
pib = produto interno bruto em US\$ a preços de 2010;

populacao = total da população;

eletricidade = consumo de energia elétrica em kWh per capita;

energia = uso de energia em kg per capita.

##	renovavel	pib	populacao
##	Min. :0.000e+00	Min. :3.182e+07	Min. :1.000e+04
##	1st Qu.:0.000e+00	1st Qu.:4.895e+09	1st Qu.:6.897e+05
##	Median :1.040e+08	Median :2.027e+10	Median :5.824e+06
##	Mean :5.445e+09	Mean :3.194e+11	Mean :3.180e+07
##	3rd Qu.:1.927e+09	3rd Qu.:1.466e+11	3rd Qu.:2.053e+07
##	Max. :1.784e+11	Max. :1.499e+13	Max. :1.338e+09
##	NA's :77	NA's :12	
##	eletricidade	energia	
##	Min. : 24.52	Min. : 135.4	
##	1st Qu.: 775.95	1st Qu.: 687.3	
##	Median : 2491.63	Median : 1474.7	
##	Mean : 4259.60	Mean : 2641.6	
##	3rd Qu.: 5700.86	3rd Qu.: 3347.6	
##	Max. :51439.91	Max. :17023.2	
##	NA's :76	NA's :76	



Supõe-se que a produção de energia renovável (*renovavel*) depende positivamente das variáveis produto interno bruto (*pib*) e negativamente da população (*populacao*). Argumenta-se que países com maior PIB tendem, ao longo do tempo, substituir as fontes de energia fósseis por fontes mais limpas; países que apresentam maior população tendem a consumir maior quantidade energia fóssel. O que foi dito pode ser estimado por meio de um modelo de regressão linear baseado no método de mínimos quadrados ordinários (OLS) na forma:

$$renovavel_i = \hat{\alpha} + \hat{\beta}_1 pib_i + \hat{\beta}_2 populacao_i + \hat{e}_i$$

Como forma de encontrar o melhor ajuste na forma funcional do modelo, deve-se testar algumas especificações. Os resultados do modelo na forma linear, linear-log, log-linear e log-log são descritos a seguir:

```
##
## =====
##                               Dependent variable:
##                               -----
##                               renovavel      log(renovavel + 1)
##                               Linear      Lin-Log      Log-Lin      Log-Log
##                               (1)        (2)        (3)        (4)
## -----
## pib          0.0115***                0.0000***
##              (0.0004)                (0.0000)
##
## populacao    4.7875                    0.0000
##              (4.0306)                (0.0000)
##
```

```

## log(pib)                    5,188,506,697.0000***          3.0299***
##                             (1,001,320,220.0000)          (0.5329)
##
## log(populacao)              919,936,393.0000             -0.7444
##                             (1,171,945,667.0000)          (0.6237)
##
## Constant                    -135,113,932.0000  -139,634,102,491.0000***  12.7654*** -50.1408***
##                             (605,477,858.0000)  (19,017,008,819.0000)    (0.8557)   (10.1206)
##
## -----
## Observations                137                137                137                137
## R2                          0.8801                0.3044                0.0860                0.2479
## Adjusted R2                 0.8783                0.2940                0.0724                0.2367
## F Statistic (df = 2; 134)   491.9049***           29.3220***           6.3050***   22.0888***
## =====
## Note:                                     *p<0.1; **p<0.05; ***p<0.01

```

Escolhe-se o modelo log-log, uma vez que foi aquele que apresentou melhor ajuste de acordo com os critérios de informação Akaique (AIC) e Schwarz (BIC)¹. Ademais, o teste de White não verificou suspeitas de heterocedasticidade em nenhum dos modelos; a hipótese de normalidade dos resíduos foi rejeitada em todos os modelos, conforme os teste Jarque-Bera e Shapiro-Wilk; e o teste geral de especificação *Ramsey-Reset* rejeitou a hipótese de correta especificação para todos os modelos, exceto o modelo log-log.

As variáveis relacionadas à hipótese teórica mostram que no modelo linear apenas o PIB é estatisticamente significativo a 5%. O sinal positivo do coeficiente associado ao PIB está de acordo com a hipótese levantada. O aumento de um por cento no PIB gera um aumento de 3% no uso de energia renovável, ceteris paribus. A variável população, neste caso, não é estatisticamente significativa na determinação do maior uso da energia renovável. Quanto ao modelo, pode-se concluir pelo teste F que o mesmo é estatisticamente significativo a qualquer nível de significância.

Pode-se testar a hipótese de maior uso de energia renovável no Grupo das sete economias mais avançadas (G7) em relação ao resto do mundo. Para isto, inclui-se uma variável binária em que 1 representa o uso de energia renovável dos países do G7 e 0 para o caso contrário.

```

##
## =====
##                               Dependent variable:
##                               -----
##                               log(renovavel + 1)
##                               Log-Log          Log-Log-Binária
##                               (1)              (2)
## -----
## log(pib)                    3.0299***          2.9841***
##                             (0.5329)          (0.5855)
##
## log(populacao)              -0.7444           -0.7381
##                             (0.6237)          (0.6268)
##
## binaria_G7                  0.7299
##                             (3.7988)
##
## Constant                    -50.1408***       -49.1305***
##                             (10.1206)        (11.4374)
##

```

¹981.4439 (AIC) e 993.1238 (BIC)

```
## -----
## Observations          137          137
## R2                    0.2479        0.2482
## Adjusted R2           0.2367        0.2312
## F Statistic    22.0888*** (df = 2; 134) 14.6324*** (df = 3; 133)
## =====
## Note:                                *p<0.1; **p<0.05; ***p<0.01
```

Os países do G7 usam, em média, 107.5% de energia renovável a mais que os demais países. No entanto, a hipótese não é validada, pois a nula não é rejeitada; isto é, a variável binária é estatisticamente insignificante. Logo, mantém-se com a especificação padrão do modelo log-log.

Feito isso, verifica-se a possibilidade de endogeneidade na variável PIB. O problema de endogeneidade pode ocorrer devido a três fatores: (1) omissão de variáveis relevantes correlacionadas ao PIB; (2) erros de medição com o PIB (e.g. uma proxy mal especificada); (3) simultaneidade entre y (renovavel) e uma ou mais variáveis explicativas (pib e populacao). Para este trabalho, testa-se o caso (1). Considera-se existir uma variável de controle associada ao PIB e que não deve ser omitida do modelo, uma vez que se teria um modelo com viés de especificação. Supõe-se dois modelos: o primeiro possui, supostamente, viés ao omitir variáveis; o segundo considera variáveis (proxies) de controle associadas à demanda por energia, as quais ajudam a explicar o uso de energia renovável e estão associadas ao PIB. Estas variáveis são o uso de energia per capita e o consumo de eletricidade per capita.

```
##
## =====
##                               Dependent variable:
##                               -----
##                               log(renovavel + 1)
##                               Linear      Linear com Controles
##                               (1)         (2)
## -----
## log(pib)                3.0299***      5.4435***
##                          (0.5329)      (1.2355)
##
## log(populacao)          -0.7444         -3.2398**
##                          (0.6237)      (1.3280)
##
## log(energia)                        -4.5628***
##                                      (1.7394)
##
## log(eletricidade)                        0.5725
##                                      (1.2735)
##
## Constant                -50.1408***     -40.8971***
##                          (10.1206)     (10.4703)
## -----
## Observations            137            137
## R2                      0.2479         0.2918
## Adjusted R2             0.2367         0.2703
## Residual Std. Error    8.5407 (df = 134) 8.3507 (df = 132)
## F Statistic            22.0888*** (df = 2; 134) 13.5951*** (df = 4; 132)
## =====
## Note:                                *p<0.1; **p<0.05; ***p<0.01
```

A inclusão das variáveis explicativas de controle melhorou o ajuste do modelo. Neste caso, o impacto positivo do crescimento econômico sobre o uso de energia renovável fica mais nítido: os sinais dos coeficientes do

PIB e da população estão de acordo com o que se supões. Apenas a variável de consumo de energia elétrica per capita não se mostrou estatisticamente significativa a 5%. De fato, pelo teste de restrição de Wald, a contribuição da variável eletricidade é nula para modelo. Diante disso, opta-se pelo uso do modelo com controles mas sem a variável eletricidade.

Uma desvantagem da análise com dados em corte transversal são os limitados graus de liberdade. Isso significa que é possível incluir apenas um conjunto restrito de variáveis no modelo de regressão; caso contrário, reduz-se sua eficiência: a variância tende a se tornar muito elevada e os testes de hipótese inválidos. Por outro lado, omitir variáveis pode gerar estimativas tendenciosas. A principal forma de enfrentar este trade-off entre eficiência e viés é aumentar o número de observações (n). Para um mesmo número de indivíduos (países) a solução passa por incluir mais períodos. Para o caso analisado, é possível estender o período da amostra para um intervalo maior que 2010. A base do WDI do Banco Mundial permite analisar os períodos de 1960 a 2018. Para lidar com esses dados alguns modelos são considerados.

Uma alternativa é usar o método OLS com dados empilhados. Os resultados são descritos a seguir comparando-os ao modelo com dados em corte transversal.

```
##
## =====
##                               Dependent variable:
##                               -----
##                               log(renovavel + 1)
##                               Transversal      Empilhado(i)
##                               (1)              (2)
## -----
## log(pib)                5.7105***          4.3798***
##                          (1.0800)          (0.1637)
##
## log(populacao)          -3.5304***          -2.4689***
##                          (1.1568)          (0.1735)
##
## log(energia)            -4.1433***
##                          (1.4635)
##
## log(energia + 1)                -2.0356***
##                                  (0.2246)
##
## Constant                -41.6222***          -43.7718***
##                          (10.3143)          (1.6471)
##
## -----
## Observations                137              5,340
## R2                        0.2907              0.2453
## Adjusted R2                0.2747              0.2449
## F Statistic      18.1684*** (df = 3; 133) 578.2465*** (df = 3; 5336)
## =====
## Note:                                *p<0.1; **p<0.05; ***p<0.01
```

Pode-se ter três especificações para modelos com dados empilhados: (i) intercepto e coeficientes angulares constantes; (ii) intercepto diferente e coeficientes angulares constantes; e (iii) intercepto e coeficientes angulares diferentes. No caso descrito na tabela anterior, observa-se que o modelo com dados empilhados (i) gera resultados mais robustos. Neste caso, todas as variáveis são estatisticamente significantes a 1%. A seguir são apresentados os resultados usando as especificações (ii) e (iii).

```
##
## =====
```

```

##                                     Dependent variable:
##                                     -----
##                                     log(renovavel + 1)
##                                     Empilhado(i)      Empilhado(ii)      Empilhado(iii)
##                                     (1)              (2)              (3)
## -----
## log(pib)                4.3798***                4.6664***                4.5013***
##                          (0.1637)                (0.1619)                (0.1637)
##
## log(populacao)          -2.4689***                -2.7400***                -2.8266***
##                          (0.1735)                (0.1714)                (0.1714)
##
## log(energia + 1)        -2.0356***                -2.3264***                -2.4206***
##                          (0.2246)                (0.2214)                (0.2213)
##
## periodo                 6.4026***                6.0595***
##                          (0.4490)                (0.4512)
##
## int_pibperiodo          0.0000***
##                          (0.0000)
##
## Constant                -43.7718***                -50.3335***                -44.0816***
##                          (1.6471)                (1.6810)                (1.9780)
## -----
## Observations            5,340                    5,340                    5,340
## R2                      0.2453                    0.2730                    0.2778
## Adjusted R2             0.2449                    0.2725                    0.2772
## Residual Std. Error      8.6666 (df = 5336)        8.5068 (df = 5335)        8.4796 (df = 5334)
## F Statistic              578.2465*** (df = 3; 5336)  500.9686*** (df = 4; 5335)  410.4312*** (df = 5; 5334)
## =====
## Note:                                                            *p<0.1; **p<0.05; ***p<0.01

```

Mais uma vez os resultados indicam que o modelo com dados empilhados é estatisticamente mais eficiente que o modelo com dados em corte transversal.

Alternativamente, pode-se estar interessado em controlar os efeitos de variáveis omitidas no modelo e que se supõe relevantes. A análise de dados em painel permite usar estimadores que controlam os efeitos dessas variáveis omitidas sobre a variável dependente. Uma possibilidade de fazer isso é usar o estimador de efeitos fixos. O controle pode ser feito por meio da criação de dummies (c_i) para cada país da amostra menos um. Dessa forma, o componente c_i mantém constante o efeito das variáveis omitidas que impactam o uso de energia renovável. O problema disso é que se gastam muitos graus de liberdade e, consequentemente, reduz-se a eficiência do modelo. Outro método é tratar os efeitos das variáveis omitidas como valores aleatórios. Neste caso, usa-se o estimador de efeitos aleatórios. Incorpora-se o efeito das variáveis omitidas (as quais são tratadas com uso de binárias no estimador de efeitos fixos) dentro do erro de previsão. É uma alternativa válida quando não se deseja reduzir a eficiência do modelo. No entanto, o uso do estimador de efeitos aleatórios requer que a correlação entre as variáveis explicativas e o erro de previsão seja igual ou próxima a zero.

Os resultados estimados para o modelo com dados empilhados (i) e os modelos com estimadores fixos e aleatórios são descritos a seguir.

```

##
## =====
##                                     Dependent variable:
##                                     -----
##                                     log(renovavel + 1)

```

```

##                               OLS                               panel
##                               Empilhado(i)                       linear
##                               (1)                               Fixos-Oneway
##                               (2)                               Fixos-Twoway
##                               (3)
## -----
## log(pib)                    4.3798***                    8.9181***                    3.7101***
##                               (0.1637)                    (0.2701)                    (0.3559)
##
## log(populacao)              -2.4689***                    -3.8842***                    -6.6756***
##                               (0.1735)                    (0.4576)                    (0.5170)
##
## log(energia + 1)            -2.0356***                    0.6711**                     0.5578*
##                               (0.2246)                    (0.3056)                    (0.3028)
##
## Constant                    -43.7718***
##                               (1.6471)
## -----
## Observations                5,340                        5,340                        5,340
## R2                          0.2453                        0.3533                        0.0512
## Adjusted R2                 0.2449                        0.3359                        0.0153
## F Statistic                 578.2465*** (df = 3; 5336) 946.9478*** (df = 3; 5199) 92.5988*** (df = 3; 5144)
## =====
## Note:                                                                *p<0.1; **p<0.05; ***p<0.01

```

O modelo com estimador de efeitos fixos por país (one-way) indica que, na média, o uso de energia de fontes renováveis aumenta por país 8.9% quando o PIB aumenta em um por cento, ceteris paribus. O aumento em um por cento na população leva a uma queda, em média por país, de 3.9%, ceteris paribus. No modelo two-way controla-se, além dos efeitos entre os países, o efeito entre os períodos (anos). Assim, o aumento em um por cento do PIB gera um aumento médio anual para cada país de 3.7% no uso de energia renovável, ceteris paribus. O aumento de um por cento da população reduz anualmente por país em média o uso de energia renovável em 6.7%, ceteris paribus. O consumo de energia elétrica é estatisticamente significativa a 10% em ambos os modelos, embora o sinal esteja invertido quando comparado ao modelo empilhado(i) e haja perda de eficiência devido às restrições que se colocam no estimador de efeitos fixos.

```

##
## =====
##                               Dependent variable:
##                               -----
##                               log(renovavel + 1)
##                               OLS                               panel
##                               Empilhado(i) Aleatórios-Oneway Aleatórios-Twoway
##                               (1)           (2)           (3)
## -----
## log(pib)                    4.3798***                    7.9164***                    5.5149***
##                               (0.1637)                    (0.2538)                    (0.0559)
##
## log(populacao)              -2.4689***                    -3.1193***                    -3.3513***
##                               (0.1735)                    (0.3533)                    (0.0707)
##
## log(energia + 1)            -2.0356***                    0.2911                       0.0206
##                               (0.2246)                    (0.2938)                    (0.0573)
##
## Constant                    -43.7718***                    -136.9199***                    -73.6536***

```



```

##              (1.6471)          (3.9357)          (1.0963)
##
## -----
## Observations          5,340          5,340          5,340
## R2                    0.2453          0.3206          0.2213
## Adjusted R2           0.2449          0.3202          0.2209
## F Statistic (df = 3; 5336) 578.2465***    839.2010***    185.8546***
## =====
## Note:                  *p<0.1; **p<0.05; ***p<0.01

```

O modelo com estimador de efeitos aleatórios apresenta interpretação similar ao caso anterior quando se analisa o valor dos coeficientes. No entanto, por ser mais parcimonioso, consome-se menos graus de liberdade quando comparado ao estimador de efeitos fixos. A implicação disso são testes de hipótese mais robustos (ganho de eficiência). Contudo, as estimativas são confiáveis apenas se sustenta a condição $cor(X, c_i) = 0$. Caso contrário, deve-se utilizar o estimador de efeitos fixos.

Uma forma de avaliar qual estimador é o mais adequado é feita pelo teste de Hausman. Os parâmetros dos estimadores com efeitos fixos e aleatórios são comparados. Se ambos são estatisticamente iguais, opta-se pelo estimador de efeitos aleatórios, uma vez que este é mais parcimonioso (mais eficiente).

No caso do modelo one-way (i.e., em que se controla os efeitos de variáveis omitidas entre os países), o modelo mais adequado, segundo o teste de Hausman é o estimador de efeitos fixos, uma vez que se rejeitou a hipótese nula de coeficientes estatisticamente iguais. Da mesma forma, no modelo two-way (i.e., em que se controla o efeitos das variáveis omitidas entre países e períodos) o estimador de efeitos fixos é apontado como o mais adequado.

3 CONCLUSÃO

A análise de dados em painel permite gerar estimativas que ajudam a tratar o problema da endogeneidade na econometria. De um lado, a inclusão de um número maior de observações aumenta a variabilidade da amostra e a eficiência do modelo estimado; por outro, é possível controlar variáveis importantes omitidas, de modo a gerar estimativas não tendenciosas.

Dito isso, os resultados para o caso analisado entre crescimento econômico (PIB e população) e uso de energia renovável sustentam da hipótese inicial. Os sinais em todos os modelos analisado indicam que o crescimento econômico, *ceteris paribus*, eleva o uso de energia renovável.

REFERÊNCIAS

- ANGRIST, J.; PISCHKE, J. Mostly harmless econometrics: An empiricist's Companion. *Princeton University Press*, 2009.
- Baltagi, B. Econometric analysis of panel data. Third Edition. *John Wiley & Sons*. 2005, Chapters 1-4.
- BALTAGI, B. H. Econometric analysis of Panel Data. 2nd edition, *John Wiley & Sons*. 2002.
- MAIA, A. G. Econometria: conceitos e aplicações. *Editores Saint Paul*, 2017.
- WOOLDRIDGE, J. M. Econometric analysis of Cross Section and Panel Data. 2nd edition. Massachusetts. *The MIT Press*. 2007.
- Wooldridge, J. M. 2001. Econometric analysis of cross section and panel data. Cap. 10.