

Impact de la croissance économique sur l'utilisation des énergies renouvelables

Henri Makika*

Julho 15, 2019

Introduction

Ce travail présente et discute des résultats d'un modèle de régression linéaire analysant l'impact de la croissance économique sur l'utilisation de l'énergie produite à partir de sources renouvelables. On suppose que lorsque les revenus des économies augmentent, un processus de remplacement des sources d'énergie fossiles par des sources plus propres (renouvelables) se produit. Dans la première partie, un modèle classique basé sur la méthode des moindres carrés ordinaires (MCO) est utilisé afin d'estimer les résultats pour l'année 2010; le modèle de données de panel est ensuite utilisé pour évaluer une période plus longue de 1960 à 2018. Les données ont été extraites des indicateurs de développement dans le monde (*world development indicators*) de la Banque mondiale.

Estimation et Discussion des résultats

Les données suivantes extraites des indicateurs de développement dans le monde de la Banque mondiale sont utilisées:

```
library(readxl)      # Pour lire les fichiers Excel
library(dplyr)        # Pour la manipulation des données
library(stargazer)    # Pour construire des tableaux
library(lmtest)       # Pour réaliser les tests d'autocorrélation sérielle
library(normtest)     # Pour réaliser les tests de normalité des résidus
library(car)          # Pour exécuter le test de Hausman
library(plm)          # Pour estimer le modèle en panel
library(knitr)        # Pour Kable
```

Importation et Exploration des données

```
Data_Renewable <- read_excel("~/Videos/Inverno 2019/Ensaio/Data_Renewable.xlsx",
                             na = ".")
dados_transversal <- Data_Renewable %>% filter(ano==2010)
summary(dados_transversal[,5:9]) %>% kable(digits = 0)
```

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

*State University of Campinas, São Paulo. E-mail : hd.makika@gmail.com

Les données utilisées se rapportent à l'année 2010 contenant 217 observations (pays). Les variables suivantes sont prises en compte tout au long du travail:

renouvelable est la production d'énergie à partir de sources renouvelables en kWh, à l'exclusion de l'hydroélectricité;

PIB représente le produit intérieur brut en USD aux prix de 2010;

population est la population totale;

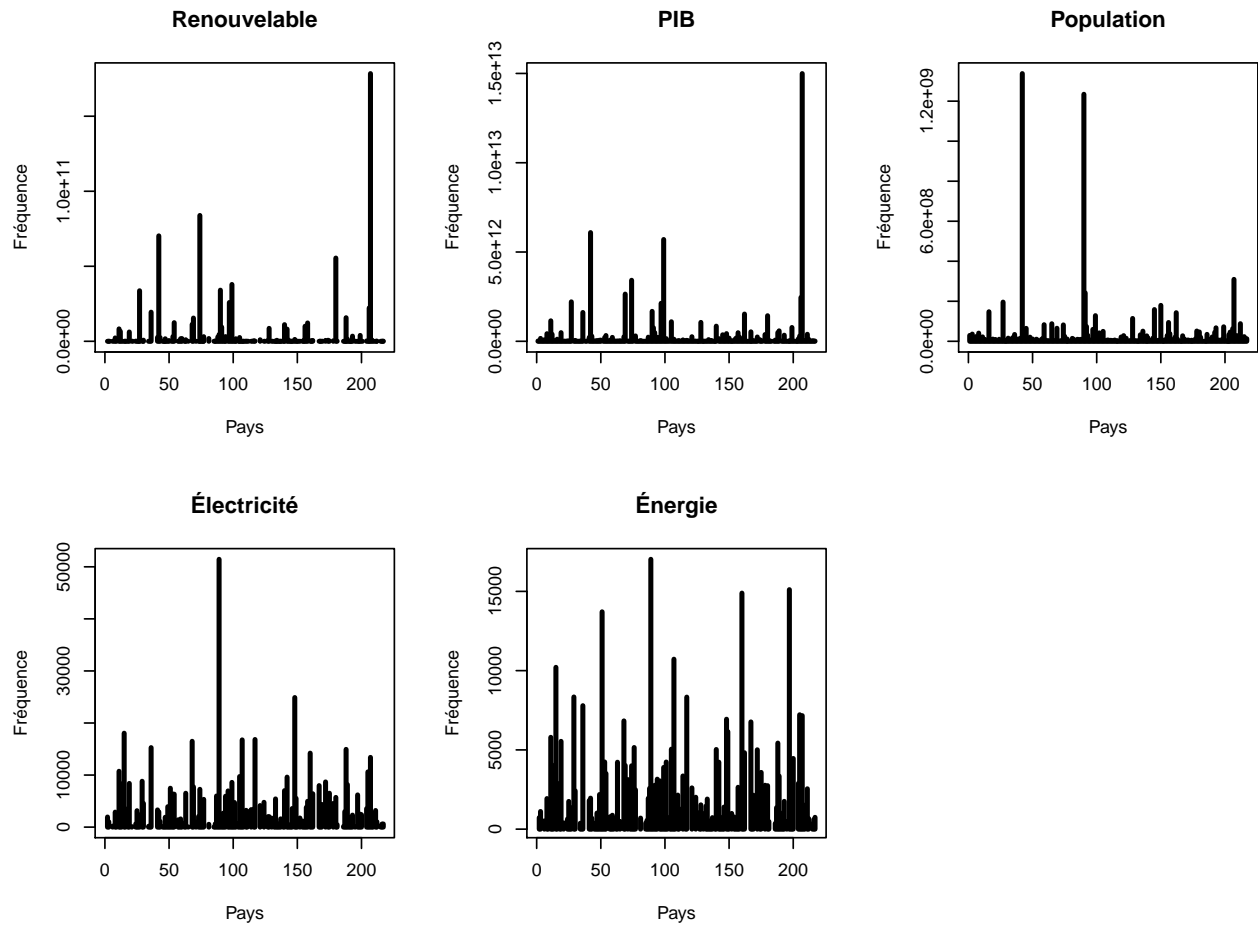
électricité est la consommation d'électricité en kWh par habitant;

énergie représente la consommation d'énergie en kWh par habitant.

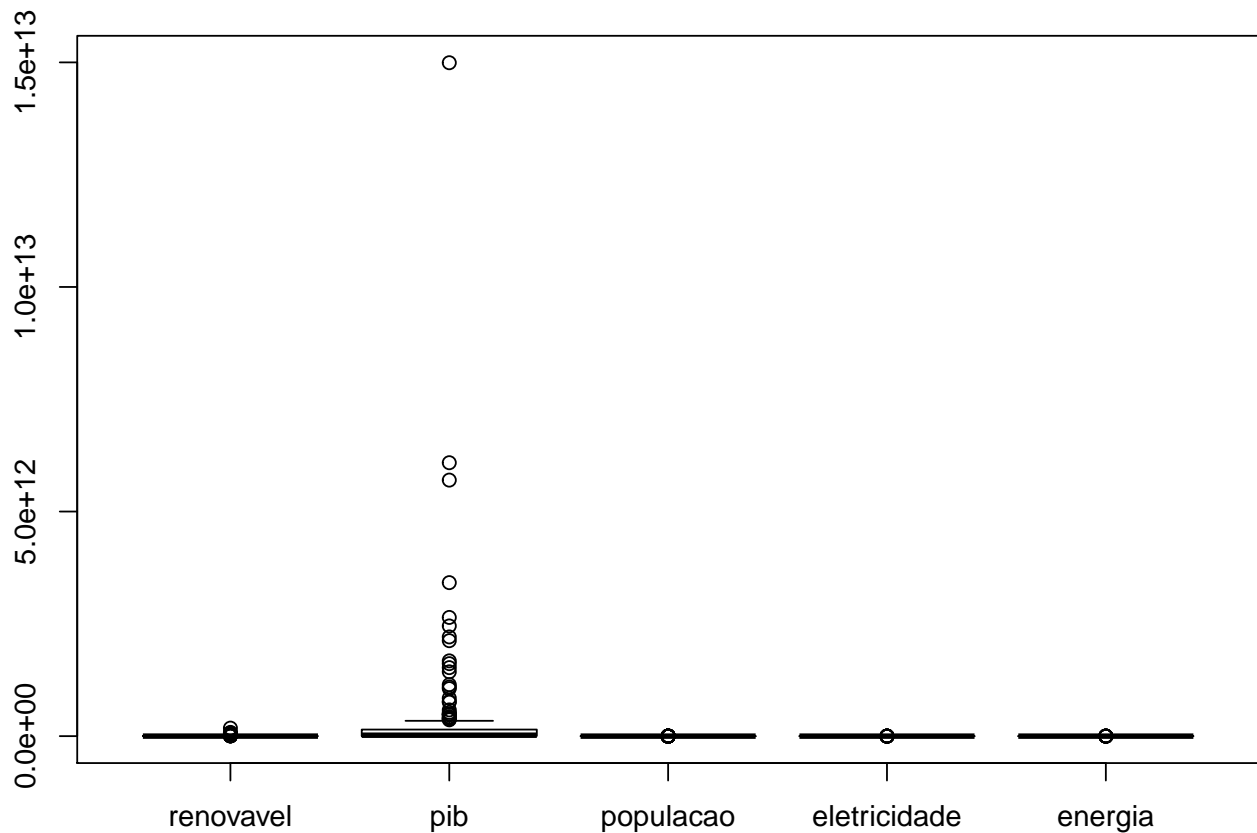
Représentation graphique

```
par(mfrow = c(2,3))
plot(dados_transversal$renovavel, type = "h", lwd = 3, main = "Renouvelable",
     xlab = "Pays", ylab = "Fréquence")
plot(dados_transversal$pib, type = "h", lwd = 3, main = "PIB", xlab = "Pays",
     ylab = "Fréquence")
plot(dados_transversal$populacao, type = "h", lwd = 3, main = "Population",
     xlab = "Pays", ylab = "Fréquence")
plot(dados_transversal$eletricidade, type = "h", lwd = 3, main = "Électricité",
     xlab = "Pays", ylab = "Fréquence")
plot(dados_transversal$energia, type = "h", lwd = 3, main = "Énergie",
     xlab = "Pays", ylab = "Fréquence")

par(mfrow = c(1, 1))
```



```
boxplot(dados_transversal[,5:9])
```



On suppose que la production d'énergie renouvelable (renouvelable) dépend positivement du produit intérieur brut (PIB) et négativement de la population (population). Il est avancé que les pays ayant un PIB plus élevé ont tendance, au fil du temps, à remplacer les sources d'énergie fossiles par des sources plus propres; Les pays plus peuplés ont tendance à consommer plus d'énergie fossile. Ce qui a été dit peut être estimé au moyen d'un modèle de régression linéaire basé sur la méthode MCO sous la forme:

$$renouvelable_i = \hat{\beta}_0 + \hat{\beta}_1 pib_i + \hat{\beta}_2 population_i + \hat{\mu}_i$$

Afin de trouver le meilleur ajustement dans la forme fonctionnelle du modèle ci-dessus, il convient de tester certaines spécifications. Les spécifications du modèle en linéaire, log-linéaire, linéaire-log et log-log sont décrites ci-dessous:

```
linear <- lm(renovavel ~ pib + populacao, data = dados_transversal)

linlog <- lm(renovavel ~ log(pib) + log(populacao), data = dados_transversal)

loglin <- lm(log(renovavel+1) ~ pib + populacao, data = dados_transversal)

loglog <- lm(log(renovavel+1) ~ log(pib) + log(populacao), data = dados_transversal)

stargazer(linear, linlog, loglin, loglog, digits = 4, type = 'text',
           column.labels = c("", "", "", ""),
           keep.stat = c('n', 'rsq', 'adj.rsq', 'f'))
```

```
##
## =====
##                               Dependent variable:
## -----
```

```
##                                renovavel                                log(renovavel + 1)
##
##                                (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
```

L'utilisation de (+ 1) dans la variable *renouvelable* est appliquée car il y a beaucoup de valeurs 0 dans le vecteur renouvelable.

Diagnostic de tests

i. Test de normalité

L'hypothèse nulle H_0 du test est que les résidus sont normalement distribués. Il sied de souligner que le rejet de l'hypothèse nulle ne signifie pas l'interprétation et l'analyse des résultats, juste pour suggérer seulement de la prudence. La non-normalité des résidus dans les analyses de séries macroéconomiques est courante dans les études conduisant au test de Jarque-Bera.

```
jb.norm.test(linear$residuals)
```

```
##
## Jarque-Bera test for normality
##
## data: linear$residuals
## JB = 3498, p-value < 2.2e-16
```

```
shapiro.test(linear$residuals)
```

```
##
## Shapiro-Wilk normality test
##
## data: linear$residuals
## W = 0.55193, p-value < 2.2e-16
```

```

jb.norm.test(linlog$residuals)

##
##  Jarque-Bera test for normality
##
## data:  linlog$residuals
## JB = 13019, p-value < 2.2e-16
shapiro.test(linlog$residuals)

##
##  Shapiro-Wilk normality test
##
## data:  linlog$residuals
## W = 0.56979, p-value < 2.2e-16
jb.norm.test(loglin$residuals)

##
##  Jarque-Bera test for normality
##
## data:  loglin$residuals
## JB = 20.799, p-value < 2.2e-16
shapiro.test(loglin$residuals)

##
##  Shapiro-Wilk normality test
##
## data:  loglin$residuals
## W = 0.74816, p-value = 4.805e-14
jb.norm.test(loglog$residuals)

##
##  Jarque-Bera test for normality
##
## data:  loglog$residuals
## JB = 14.421, p-value = 0.007
shapiro.test(loglog$residuals)

##
##  Shapiro-Wilk normality test
##
## data:  loglog$residuals
## W = 0.89399, p-value = 1.966e-08

```

L'hypothèse nulle peut être rejetée, les résidus ne sont pas normalement distribués pour nos quatre régressions. L'hypothèse nulle du test est que les résidus sont normalement distribués. Il sied de souligner que le rejet de l'hypothèse nulle ne signifie pas l'interprétation et l'analyse des résultats, juste pour suggérer seulement de la prudence. La non-normalité des résidus dans les analyses de séries macroéconomiques est courante dans les études conduisant au test de Jarque-Bera.

ii. Test d'homoscédasticité

En statistique, l'homoscédasticité est une propriété fondamentale du modèle de la régression linéaire générale et fait partie de ses hypothèses de base. On parle d'homoscédasticité lorsque la variance des erreurs stochastiques de la régression est la même pour chaque observation i (de 1 à n observations). La notion d'homoscédasticité

s'oppose à celle d'hétéroscédasticité, qui correspond au cas où la variance de l'erreur des variables est différente. L'hypothèse nulle H_0 est la présence d'homoscédasticité et l'hypothèse alternative H_1 est l'hétéroscédasticité.

```
bptest(linear) # rejet de l'hypothèse nulle
```

```
##
## studentized Breusch-Pagan test
##
## data: linear
## BP = 11.479, df = 2, p-value = 0.003216
```

```
bptest(linear, ~fitted(linear) + fitted(linear)^2) # non rejet la nulle
```

```
##
## studentized Breusch-Pagan test
##
## data: linear
## BP = 0.12266, df = 1, p-value = 0.7262
```

```
bptest(linlog) # rejet de l'hypothèse nulle
```

```
##
## studentized Breusch-Pagan test
##
## data: linlog
## BP = 12.845, df = 2, p-value = 0.001625
```

```
bptest(linlog, ~fitted(linlog) + fitted(linlog)^2) # non rejet la nulle
```

```
##
## studentized Breusch-Pagan test
##
## data: linlog
## BP = 0.018464, df = 1, p-value = 0.8919
```

```
bptest(loglin) # non rejet de l'hypothèse nulle
```

```
##
## studentized Breusch-Pagan test
##
## data: loglin
## BP = 2.4055, df = 2, p-value = 0.3004
```

```
bptest(loglin, ~fitted(loglin) + fitted(loglin)^2) # non rejet la nulle
```

```
##
## studentized Breusch-Pagan test
##
## data: loglin
## BP = 0.7205, df = 1, p-value = 0.396
```

```
bptest(loglog) # rejet de l'hypothèse nulle
```

```
##
## studentized Breusch-Pagan test
##
## data: loglog
## BP = 7.8889, df = 2, p-value = 0.01936
```

```
bptest(loglog, ~fitted(loglog) + fitted(loglog)^2) # non rejet la nulle
```

```
##  
## studentized Breusch-Pagan test  
##  
## data: loglog  
## BP = 0.32147, df = 1, p-value = 0.5707
```

iii. Test de multicolinéarité

La multicolinéarité est le fait qu'une variable indépendante est prédictible par (ou partage sa variance avec) une combinaison linéaire des autres variables indépendantes du modèle estimé. Pour faire simple, disons qu'une combinaison linéaire est une variable que l'on obtient en faisant la somme pondérée de plusieurs autres variables. La fonction *VIF* (pour Variance Inflation Factor), donne la valeur réciproque de la tolérance.

Des valeurs élevées de *VIF* indiquent donc la présence de multicolinéarité. L'utilité diagnostique de cet indice résulte du fait que la variance du coefficient de régression augmente de la même façon que le facteur d'inflation de la variance. Pour nos régressions nous remarquons aucun suspect de la multicolinéarité.

```
vif(linear)
```

```
##      pib populacao  
## 1.253587 1.253587
```

```
vif(linlog)
```

```
##      log(pib) log(populacao)  
##      1.784668      1.784668
```

```
vif(loglin)
```

```
##      pib populacao  
## 1.253587 1.253587
```

```
vif(loglog)
```

```
##      log(pib) log(populacao)  
##      1.784668      1.784668
```

iv. Test de spécification du modèle

En statistique, le test d'erreur de spécification de régression de Ramsey (RESET) est un test de spécification générale pour le modèle de régression linéaire. Plus spécifiquement, il teste si les combinaisons non linéaires des valeurs ajustées aident à expliquer la variable de réponse. L'intuition qui sous-tend le test est que si des combinaisons non linéaires des variables explicatives ont le pouvoir d'expliquer la variable de réponse. Ainsi, le modèle est mal spécifié en ce sens que le processus de génération de données pourrait être mieux approché par un polynôme ou un autre fonctionnel de la forme non linéaire. Pour notre cas, le modèle lolog qui est bien spécifié.

```
resettest(linear)
```

```
##  
## RESET test  
##  
## data: linear  
## RESET = 4.1673, df1 = 2, df2 = 132, p-value = 0.01758
```

```
resettest(linlog)
```

```
##  
## RESET test
```



```
##
## data:  linlog
## RESET = 178.25, df1 = 2, df2 = 132, p-value < 2.2e-16
resettest(loglin)

##
## RESET test
##
## data:  loglin
## RESET = 7.6217, df1 = 2, df2 = 132, p-value = 0.0007371
resettest(loglog)

##
## RESET test
##
## data:  loglog
## RESET = 0.9187, df1 = 2, df2 = 132, p-value = 0.4016
```

v. Critère d'informations

Lorsque l'on estime un modèle statistique, il est possible d'augmenter la vraisemblance du modèle en ajoutant un paramètre par exemple. Le critère d'information d'Akaike (AIC), tout comme le critère d'information bayésien (BIC), permet de pénaliser les modèles en fonction du nombre de paramètres afin de satisfaire le critère de parcimonie. On choisit alors le modèle avec le critère d'information d'Akaike le plus faible.

```
AIC(linear, linlog, loglin, loglog)
```

```
##          df          AIC
## linear   4 6591.5635
## linlog   4 6832.4482
## loglin   4 1008.1595
## loglog   4  981.4439
```

```
BIC(linear, linlog, loglin, loglog)
```

```
##          df          BIC
## linear   4 6603.2434
## linlog   4 6844.1282
## loglin   4 1019.8395
## loglog   4  993.1238
```

Le modèle linéaire est choisi car c'est celui qui présente le meilleur ajustement en fonction du R^2 (voir le tableau ci-haut). De plus, le test de White n'a pas permis de vérifier les suspicions d'hétéroscédasticité et cela dans aucun des modèles; l'hypothèse de la normalité des résidus a été rejetée dans tous les modèles, selon les tests de Jarque-Bera et Shapiro-Wilk; et le test général de Ramsey-Reset on a rejeté l'hypothèse de spécification correcte pour tous les modèles, à l'exception du modèle loglog.

L'hypothèse théorique montre que, dans le modèle linéaire, seul le PIB est statistiquement significatif à 5%. Le signe positif du coefficient associé au PIB conforme l'hypothèse émise. L'augmentation de 100 dollars du PIB génère une augmentation de 1.16 kWh de l'utilisation des énergies renouvelables, toutes choses étant égales par ailleurs. La variable population, dans ce cas, n'est pas statistiquement significative pour déterminer la plus grande utilisation des énergies renouvelables. En ce qui concerne le modèle, nous pouvons conclure du test F qu'il est statistiquement significatif à n'importe quel niveau de signification.

On peut tester l'hypothèse d'une utilisation accrue des énergies renouvelables dans le groupe des sept économies les plus avancées (G7) par rapport au reste du monde. Pour cela, une variable binaire est incluse où 1 représente l'utilisation de l'énergie renouvelable des pays du G7 et 0 dans le cas contraire.

```

sigla = dados_transversal$codigo

d1=as.numeric(sigla=='DEU')    # Allemagne
d2=as.numeric(sigla=='CAN')    # Canada
d3=as.numeric(sigla=='USA')    # Etats-Unis d'Amérique
d4=as.numeric(sigla=='ITA')    # Italie
d5=as.numeric(sigla=='FRA')    # France
d6=as.numeric(sigla=='JPN')    # Japon
d7=as.numeric(sigla=='GBR')    # Grande Bretagne

binaria_G7 = d1 + d2 + d3 + d4 + d5 + d6 + d7

linear.bin <- lm(renovavel ~ pib + populacao + binaria_G7, data = dados_transversal)

stargazer(linear, linear.bin, digits = 4, type = 'text', column.labels = c("", ""),
  keep.stat = c('n', 'rsq', 'adj.rsq', 'f'), out = "mdw.txt")

```

```

##
## =====
##                               Dependent variable:
##                               -----
##                               renovavel
##
##                               (1)                (2)
## -----
## pib                0.0115***                0.0115***
##                   (0.0004)                (0.0006)
##
## populacao          4.7875                    4.9181
##                   (4.0306)                (4.2438)
##
## binaria_G7                                363,258,057.0000
##                                       (3,564,800,222.0000)
##
## Constant           -135,113,932.0000         -140,929,245.0000
##                   (605,477,858.0000)         (610,399,688.0000)
##
## -----
## Observations              137                137
## R2                        0.8801                0.8801
## Adjusted R2               0.8783                0.8774
## F Statistic  491.9049*** (df = 2; 134) 325.5182*** (df = 3; 133)
## =====
## Note:                      *p<0.1; **p<0.05; ***p<0.01

```

Les pays du G7 utilisent en moyenne plus de 3,5 millions de kwh d'énergie renouvelable de plus que les autres pays. Cependant, l'hypothèse n'est pas validée, puisque l'hypothèse nulle de la signficance de la variable *binaria_G7* n'est pas rejetée. c'est-à-dire que la variable binaire est statistiquement non significative. Par conséquent, il est maintenu avec la spécification du modèle linéaire.

Une fois que cela est fait, la possibilité d'endogénéité dans la variable pib est vérifiée. Le problème d'endogénéité peut survenir en raison de trois facteurs: (1) l'omission de variables pertinentes corrélées au PIB; (2) des erreurs de mesure avec la variable PIB (par exemple un proxy mal spécifié); (3) la simultanéité entre y (renouvelable) et une ou plusieurs variables explicatives (PIB et population). Pour ce travail, nous testons le cas (1), où elle est considérée comme une variable de contrôle associée au PIB et ne devrait pas être

omise du modèle, car il y aurait un modèle avec biais de spécification. Deux modèles sont supposés: le premier a, supposément, un biais lorsqu'on omet des variables; la seconde considère les variables de contrôle associées (variables de remplacement) est la demande en énergie, ce qui aide à expliquer l'utilisation des énergies renouvelables et est associée au PIB. Ces variables sont la consommation d'énergie par habitant et la consommation d'électricité par habitant.

```
modelo_vies <- lm(renovavel ~ pib + populacao, data = dados_transversal)

modelo_sem_vies <- lm(renovavel ~ pib + populacao + energia + eletricidade,
                      data = dados_transversal)

stargazer(modelo_vies, modelo_sem_vies, digits = 4, type = 'text',
           column.labels = c("", ""), keep.stat = NULL, out = "mdw.txt")
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               renovavel
##
##                               (1)                (2)
## -----
## pib                        0.0115***          0.0115***
##                          (0.0004)            (0.0004)
##
## populacao                  4.7875              5.0456
##                          (4.0306)            (4.1097)
##
## energia                    -287,172.8000
##                          (356,869.4000)
##
## eletricidade               178,950.9000
##                          (175,439.3000)
##
## Constant                   -135,113,932.0000    -154,382,459.0000
##                          (605,477,858.0000)    (788,217,963.0000)
##
## -----
## Observations                137                137
## R2                         0.8801              0.8811
## Adjusted R2                 0.8783              0.8775
## Residual Std. Error 6,662,336,361.0000 (df = 134) 6,686,198,842.0000 (df = 132)
## F Statistic                491.9049*** (df = 2; 134) 244.4613*** (df = 4; 132)
## =====
## Note:                        *p<0.1; **p<0.05; ***p<0.01
```

Puisque l'inclusion des variables de contrôle (instrument) n'a pas amélioré la signification des estimateurs ni de la valeur de R^2 , le modèle d'origine est conservé.

L'analyse transversale des données présente l'inconvénient de limiter les degrés de liberté. Cela signifie qu'il est possible d'inclure uniquement un ensemble restreint de variables dans le modèle de régression; sinon, son efficacité est réduite: la variance a tendance à devenir très élevée et les tests d'hypothèses sont souvent invalides. En revanche, l'omission de variables peut générer des estimations biaisées. Le principal moyen de remédier à ce compromis entre efficacité et biais consiste à augmenter le nombre d'observations (n). Pour le même nombre d'individus (pays), la solution consiste à inclure plus de périodes. Pour le cas étudié, il est possible d'étendre la période d'échantillonnage à un intervalle supérieur à 2010. La base de données WDI de

la Banque mondiale permet d'analyser les périodes de 1960 à 2018. Certains modèles sont considérés pour traiter ces données.

Une alternative consiste à utiliser la méthode OLS avec des données empilées. Les résultats sont décrits ci-dessous en les comparant au modèle avec les données transversales.

```
dados_painel <- Data_Renewable
```

```
summary(dados_painel[,1:4]) %>% kable(digits = 0)
```

ano	tempoano	pais	codigo
Min. :1960	Length:12803	Length:12803	Length:12803
1st Qu.:1974	Class :character	Class :character	Class :character
Median :1989	Mode :character	Mode :character	Mode :character
Mean :1989	NA	NA	NA
3rd Qu.:2004	NA	NA	NA
Max. :2018	NA	NA	NA

```
dados_painel <- Data_Renewable
```

```
summary(dados_painel[,5:9]) %>% kable(digits = 0)
```

renovavel	pib	populacao	eletricidade	energia
Min. :0.000e+00	Min. :2.144e+07	Min. :3.893e+03	Min. : 0.0	Min. : 0.0
1st Qu.:0.000e+00	1st Qu.:3.515e+09	1st Qu.:4.628e+05	1st Qu.: 399.5	1st Qu.: 512.7
Median :0.000e+00	Median :1.495e+10	Median :4.100e+06	Median : 1574.4	Median : 1212.1
Mean :2.181e+09	Mean :2.449e+11	Mean :2.415e+07	Mean : 3173.1	Mean : 2348.6
3rd Qu.:3.200e+08	3rd Qu.:1.100e+11	3rd Qu.:1.312e+07	3rd Qu.: 4305.1	3rd Qu.: 3067.7
Max. :3.174e+11	Max. :1.784e+13	Max. :1.393e+09	Max. :54799.2	Max. :40710.1
NA's :6789	NA's :3554	NA's :108	NA's :6896	NA's :6721

```
lpib = log(dados_painel$pib)
```

```
lpop = log(dados_painel$populacao)
```

```
lrenov = log(dados_painel$renovavel + 1)
```

```
lener = log(dados_painel$energia + 1)
```

```
transversal <- lm(lrenov ~ lpib + lpop + lener, data = dados_transversal)
```

```
empilhado1 <- lm(lrenov ~ lpib + lpop + lener, data = dados_painel)
```

```
stargazer(transversal, empilhado1, type = 'text', digits = 4, column.labels = c("", "1"),
  keep.stat = c('n', 'rsq', 'adj.rsq', 'f'))
```

```
##
## =====
##               Dependent variable:
##               -----
##               lrenov
##               1
##               (1)      (2)
## -----
```

```
## lpib          4.3798***      4.3798***
##              (0.1637)      (0.1637)
##
## lpop          -2.4689***      -2.4689***
##              (0.1735)      (0.1735)
##
## lener         -2.0356***      -2.0356***
##              (0.2246)      (0.2246)
##
## Constant     -43.7718***      -43.7718***
##              (1.6471)      (1.6471)
##
## -----
## Observations      5,340      5,340
## R2                0.2453      0.2453
## Adjusted R2       0.2449      0.2449
## F Statistic (df = 3; 5336) 578.2465*** 578.2465***
## =====
## Note:              *p<0.1; **p<0.05; ***p<0.01
```

```
dados_painel <- Data_Renewable
ano = dados_painel$ano
n = length(ano)
periodo = rep(0,n)
for(i in 1:n){if(ano[i]>=1973)periodo[i]=1}

empilhado2 <- lm(lrenov ~ lpib + lpop + lener + periodo, data = dados_painel)

int_pibperiodo <- dados_painel$pib*periodo

empilhado3 <- lm(lrenov ~ lpib + lpop + lener + periodo + int_pibperiodo,
               data = dados_painel)

stargazer(empilhado1, empilhado2, empilhado3, type = 'text', digits = 3,
          column.labels = c("1", "2", "3"), keep.stat = NULL, out = "mdr.txt")
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               lrenov
##                               1          2          3
##                               (1)        (2)        (3)
## -----
## lpib          4.380***      4.666***      4.501***
##              (0.164)      (0.162)      (0.164)
##
## lpop          -2.469***      -2.740***      -2.827***
##              (0.173)      (0.171)      (0.171)
##
## lener         -2.036***      -2.326***      -2.421***
##              (0.225)      (0.221)      (0.221)
##
## periodo              6.403***      6.060***
##                  (0.449)      (0.451)
```

```
##
## int_pibperiodo                                0.000***
##                                                (0.000)
##
## Constant                -43.772***            -50.334***            -44.082***
##                        (1.647)                (1.681)                (1.978)
##
## -----
## Observations                5,340                5,340                5,340
## R2                        0.245                0.273                0.278
## Adjusted R2                0.245                0.273                0.277
## Residual Std. Error      8.667 (df = 5336)      8.507 (df = 5335)      8.480 (df = 5334)
## F Statistic              578.247*** (df = 3; 5336) 500.969*** (df = 4; 5335) 410.431*** (df = 5; 5334)
## =====
## Note:                                                    *p<0.1; **p<0.05; ***p<0.01
```

Il convient de noter que trois spécifications sont décrites pour le modèle avec des données empilées: (1) des coefficients angulaires d'interception et constants; (2) différents coefficients angulaires d'interception et constants; et (3) intercept et différents coefficients angulaires. Ainsi, il est possible d'identifier des changements dans la relation entre croissance économique et production d'énergie renouvelable au fil du temps. Bien que le R^2 n'augmente pas dans les modèles avec des données empilées, la variable population devient statistiquement significative.

Une autre solution consiste à contrôler les effets des variables omises dans le modèle et supposés pertinents. Ainsi, il est possible d'utiliser des estimateurs permettant de contrôler les effets de ces variables omises sur les autres. Ce contrôle peut être effectué en créant des binaires pour chaque pays de l'échantillon moins un. Cette méthode est réalisée via l'estimateur à effets fixes. Le problème, c'est que de nombreux degrés de liberté sont utilisés et que, par conséquent, l'efficacité du modèle est réduite. Une autre méthode consiste à traiter les effets des variables omises comme des valeurs aléatoires, c'est-à-dire que, par le biais de l'estimateur à effets aléatoires, l'effet des variables omises (traitées avec une utilisation binaire dans l'estimateur à effets fixes) dans l'erreur de prédiction est incorporé. C'est une alternative valable quand on ne veut pas réduire l'efficacité du modèle. Cependant, l'utilisation de l'estimateur à effets aléatoires exige que la corrélation entre les variables explicatives et l'erreur de prédiction soit égale ou proche de zéro.

Les résultats estimés pour le modèle linéaire en coupe transversale, le modèle de données empilées et les modèles avec des estimateurs fixes et aléatoires sont décrits ci-dessous.

i. Modèle à effets fixes (within and twoway)

```
dados_painel <- plm.data(dados_painel, index=c("pais","ano"))

## Warning: use of 'plm.data' is discouraged, better use 'pdata.frame' instead

lpib = log(dados_painel$pib)
lpop = log(dados_painel$populacao)
lrenov = log(dados_painel$renovavel + 1)
lener = log(dados_painel$energia + 1)

painel.fe1 <- plm(lrenov ~ lpib + lpop + lener, data = dados_painel,
                 model = "within")

painel.fe2 <- plm(lrenov ~ lpib + lpop + lener, data = dados_painel,
                 effect = "twoway", model = "within")

stargazer(empilhado1, painel.fe1, painel.fe2, type = 'text', digits = 3,
          column.labels = c("Empile1", "Fixe-Oneway", "Fixe-Twoway"),
          keep.stat = c('n', 'rsq', 'adj.rsq', 'f'), out = "inesh.txt")
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               lrenov
##                               OLS          panel
##                               Empile1      linear
##                               (1)          (2)          (3)
## -----
## lpib          4.380***          8.918***          3.710***
##                (0.164)          (0.270)          (0.356)
##
## lpop          -2.469***          -3.884***          -6.676***
##                (0.173)          (0.458)          (0.517)
##
## lener          -2.036***          0.671**          0.558*
##                (0.225)          (0.306)          (0.303)
##
## Constant      -43.772***
##                (1.647)
## -----
## Observations    5,340          5,340          5,340
## R2              0.245          0.353          0.051
## Adjusted R2     0.245          0.336          0.015
## F Statistic  578.247*** (df = 3; 5336) 946.948*** (df = 3; 5199) 92.599*** (df = 3; 5144)
## =====
## Note:                                     *p<0.1; **p<0.05; ***p<0.01
```

Le modèle d'estimation à effets fixes pour pays (unidirectionnels) indique que la consommation d'énergie provenant de sources renouvelables augmente en moyenne de 8.9% par pays lorsque le PIB augmente de 1%, toutes choses étant égales par ailleurs. L'augmentation de 1% de la population entraîne une chute de la consommation d'énergie renouvelable par pays de 3.9% en moyenne, *ceteris paribus*. Dans le modèle *two-way*, en plus des effets entre pays, l'effet entre périodes (années) est contrôlé. Ainsi, l'augmentation de 1% du PIB génère pour chaque pays une augmentation moyenne de 3.7% de l'utilisation des énergies renouvelables, **ceteris paribus*. L'augmentation annuelle de la population de 1% réduit la consommation d'énergie renouvelable par pays de 6.7% en moyenne, *ceteris paribus*. La consommation d'électricité est statistiquement significative à 10% dans les deux modèles, bien que le signal soit inversé par rapport au modèle empilé et qu'il existe une perte d'efficacité due aux restrictions imposées à l'estimateur à effets fixes.

ii. Modèle à effets aléatoires

```
painel.re1 <- plm(lrenov ~ lpib + lpop + lener, data = dados_painel,
                  model = "random")

painel.re2 <- plm(lrenov ~ lpib + lpop + lener, data = dados_painel,
                  effect = "twoways", model = "random")

stargazer(empilhado1, painel.re1, painel.re2, type = 'text', digits = 4,
          column.labels = c("Empile1", "Oneway", "Twoway"),
          keep.stat = c('n', 'rsq', 'adj.rsq', 'f'), out = "inesd.txt")
```

```
##
## =====
##                               Dependent variable:
```

```
## -----
##               lrenov
##               OLS      panel
##               Empile1  linear
##               (1)      (2)      (3)
## -----
## lpib          4.3798***  7.9164***  5.5149***
##               (0.1637)  (0.2538)  (0.0559)
##
## lpop          -2.4689*** -3.1193*** -3.3513***
##               (0.1735)  (0.3533)  (0.0707)
##
## lener         -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
```

Le modèle avec estimateur à effets aléatoires présente une interprétation similaire à celle du cas précédent lors de l'analyse de la valeur des coefficients. Cependant, comme il est plus parcimonieux, moins de degrés de liberté sont consommés par rapport à l'estimateur à effets fixes. Ceci implique des tests d'hypothèses plus robustes (gain d'efficacité). Cependant, les estimations ne sont fiables que si la condition $cor(X, c_i) = 0$ est vérifiée. Sinon, l'estimateur à effets fixes devrait être utilisé.

Le test de Hausman est un moyen d'estimer quel estimateur est le plus approprié. Les paramètres des estimateurs à effets fixes et aléatoires sont comparés. Si les deux sont statistiquement égaux, l'estimateur à effets aléatoires est choisi, car il est plus parcimonieux (plus efficace).

```
phtest(painel.fe1, painel.re1)
```

```
##
## Hausman Test
##
## data: lrenov ~ lpib + lpop + lener
## chisq = 334.34, df = 3, p-value < 2.2e-16
## alternative hypothesis: one model is inconsistent
```

```
phtest(painel.fe2, painel.re2)
```

```
##
## Hausman Test
##
## data: lrenov ~ lpib + lpop + lener
## chisq = 84.078, df = 3, p-value < 2.2e-16
## alternative hypothesis: one model is inconsistent
```

Dans le modèle unidirectionnel (c'est-à-dire où les effets des variables omises sont contrôlés entre pays), le modèle le plus approprié, selon le test de Hausman, est l'estimateur à effets fixes, l'hypothèse nulle des

coefficients statistiquement égaux est rejetée. De même, dans le modèle *two-way* (c'est-à-dire lorsque les effets des variables omises entre pays et périodes sont contrôlés), l'estimateur à effets fixes est indiqué comme étant le plus approprié.

Conclusion

L'analyse des données de panel permet de générer des estimations permettant de résoudre le problème de l'endogénéité en économétrie. D'une part, l'inclusion d'un plus grand nombre d'observations augmente la variabilité de l'échantillon et l'efficacité estimée du modèle; d'autre part, il est possible de contrôler les variables importantes omises afin de générer des estimations non biaisées.

Cela dit, les résultats du cas analysés entre la croissance économique (PIB et population) et l'utilisation des énergies renouvelables confirment l'hypothèse initiale. Les signaux de tous les modèles analysés indiquent que la croissance économique, toutes choses étant égales par ailleurs, augmente l'utilisation des énergies renouvelables. L'estimateur à effets fixes, parmi les modèles analysés, s'est révélé le plus approprié pour cette évaluation.

Textes sources

ANGRIST, J.; PISCHKE, J. **Mostly harmless econometrics: An empiricist's Companion**. Princeton University Press, 2009.

BALTAGI, B.H. **Econometric analysis of Panel Data**. 2nd edition, John Wiley & Sons. 2002.

MAIA, A.G. **Econometria: conceitos e aplicações**. Editora Saint Paul, 2017.

WOOLDRIDGE, J.M. **Econometric analysis of Cross Section and Panel Data**. 2nd edition. Massachusetts. The MIT Press. 2007.

Appendice

Test de dépendance sérielle

```
pcdtest(painel.fe1, test = c("lm"))
```

```
## Warning in pcdres(tres = tres, n = n, w = w, form =  
## paste(deparse(x$formula)), : Some pairs of individuals (0.021 percent) do  
## not have any or just one time period in common and have been omitted from  
## calculation
```

```
##  
## Breusch-Pagan LM test for cross-sectional dependence in panels  
##  
## data: lrenov ~ lpib + lpop + lener  
## chisq = 88626, df = 9452, p-value < 2.2e-16  
## alternative hypothesis: cross-sectional dependence
```

```
pcdtest(painel.fe1, test = c("cd"))
```

```
## Warning in pcdres(tres = tres, n = n, w = w, form =  
## paste(deparse(x$formula)), : Some pairs of individuals (0.021 percent) do  
## not have any or just one time period in common and have been omitted from  
## calculation
```

```
##
## Pesaran CD test for cross-sectional dependence in panels
##
## data:  lrenov ~ lpib + lpop + lener
## z = 13.406, p-value < 2.2e-16
## alternative hypothesis: cross-sectional dependence
```

Test de corrélation sérielle

```
bptest(renovavel ~ pib + populacao + factor(pais), data = dados_painel, studentize = F)
```

```
##
## Breusch-Pagan test
##
## data:  renovavel ~ pib + populacao + factor(pais)
## BP = 111230, df = 139, p-value < 2.2e-16
```

Test d'hétéroscédasticité

```
bptest(renovavel ~ pib + populacao + factor(pais), data = dados_painel,
       studentize = FALSE)
```

```
##
## Breusch-Pagan test
##
## data:  renovavel ~ pib + populacao + factor(pais)
## BP = 111230, df = 139, p-value < 2.2e-16
```

Contrôle d'hétéroscédasticité

```
coeftest(painel.fe1)
```

```
##
## t test of coefficients:
##
##      Estimate Std. Error t value Pr(>|t|)
## lpib    8.91813    0.27006 33.0225 < 2e-16 ***
## lpop   -3.88416    0.45760 -8.4881 < 2e-16 ***
## lener    0.67112    0.30557  2.1963  0.02811 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
coeftest(painel.fe1, vcovHC(painel.fe1, method = "arellano"))
```

```
##
## t test of coefficients:
##
##      Estimate Std. Error t value Pr(>|t|)
## lpib    8.91813    1.14755  7.7715 9.286e-15 ***
## lpop   -3.88416    1.89532 -2.0493  0.04048 *
## lener    0.67112    1.26659  0.5299  0.59623
## ---
```

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

paine1.fe1.consistent <- coeftest(paine1.fe1, vcovHC(paine1.fe1, method = "arellano"))

stargazer(paine1.fe1, paine1.fe1.consistent, type = 'text', digits = 4,
  column.labels = c("P.f1", "P.f1consistent"), keep.stat = NULL,
  out = "mdf.txt")
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               lrenov
##                               panel      coefficient
##                               linear      test
##                               P.f1      P.f1consistent
##                               (1)      (2)
## -----
## lpib          8.9181***          8.9181***
##                (0.2701)          (1.1475)
##
## lpop          -3.8842***          -3.8842**
##                (0.4576)          (1.8953)
##
## lener          0.6711**           0.6711
##                (0.3056)          (1.2666)
## -----
## Observations      5,340
## R2                 0.3533
## Adjusted R2       0.3359
## F Statistic  946.9478*** (df = 3; 5199)
## =====
## Note:              *p<0.1; **p<0.05; ***p<0.01
```

```
pool <- plm(renovavel ~ pib + populacao, data = dados_paine1, model = "pooling")

plmtest(pool, type=c("bp"))
```

```
##
## Lagrange Multiplier Test - (Breusch-Pagan) for unbalanced panels
##
## data:  renovavel ~ pib + populacao
## chisq = 2027.9, df = 1, p-value < 2.2e-16
## alternative hypothesis: significant effects
```