

# Analyse de l'impact de la croissance économique sur les émissions de CO2 dans le monde en 2017

Bill Yehouenou

2022-11-21

## 1. Problématique et équation à estimer

### 1.1. Problématique

Historiquement, les émissions de CO2 ont toujours été fortement corrélées à la richesse d'un pays. Cependant, les récentes tendances du développement des énergies renouvelables permettent d'entretenir l'espoir d'une croissance verte. À l'échelle mondiale, des dizaines de pays ont maintenu une croissance économique tout en diminuant leur empreinte carbone depuis les années 1990. Ainsi, une question se pose **“la croissance fait-elle l'effet de serre ?”**. Telle est la problématique de notre projet où nous nous focalisons sur l'année **2017**.

### 1.2. Equation du modèle

Dans l'optique de construire notre modèle, nous avons fait recours à plusieurs revues scientifiques afin de comprendre les variables pouvant expliquer une quelconque augmentation des émissions des gaz à effets de serre. Une équation dite de *Kaya* à particulièrement retenu notre attention. Développée par l'économiste japonais **Yoichi Kaya** en 1993 dans son ouvrage *Environment, Energy, and Economy : strategies for sustainability*, l'équation de Kaya démontre qu'il y aurait un lien entre les émissions de gaz à effet de serre, la croissance économique, l'intensité énergétique et la population. Nous avons ajouté une variable indicatrice *Pays moins avancées* pour étudier les différences entre ces différents groupes. Ainsi, l'équation de notre modèle devrait se présenter comme suit :

$$CO_2/habitant = b_0 + b_1(PIB/habitant) + b_2(PIB/habitant)^2 + b_3(EnergiesRenouvelables) + b_4(IntensiteEnergétique_economie) + b_5(PaysMoinsAvances) + e$$

## 2. Collecte des données

Notre échantillon est constitué de données prises en coupes instantanées en 2017 pour **217** états dans le monde. Les données ont été prises sur le site <https://data.worldbank.org> de la Banque Mondiale. Nous avons combiné cette base avec la liste des pays moins avancées pour créer la variable *pma* (variable indicatrice) qui prend les valeurs :

- 0 si le pays n'est pas un pays moins avancée ;
- 1 si le pays l'est

En ce qui concerne le traitement des valeurs manquantes, notre base de données contenait une variable utile (d'après la revue de littérature) **pourcentage d'énergie fossile** mais elle avait moins de 4% d'informations. On a jugé utile de ne pas la considérer. Pour finir, nous avons donné aux variables le type adéquat (“character”, “numeric”) pour pouvoir faciliter l'étude des phénomènes. Le dictionnaire des données ci-dessous résume toutes nos variables :

Table 1: Dictionnaire des variables

Code	Définition	Nature	Signe attendu
pays	Nom des pays	Qualitative	
pop_active	Population active en 2017	Quantitative	+
pop	Population totale estimé en 2017	Quantitative	+
pib_hbt	PIB par habitant en 2017 (prix constants)	Quantitative	+
int_energy	Intensité énergétique de l'économie en tonnes/PIB	Quantitative	+
energy_ren	Pourcentage d'énergie renouvelable par pays	Quantitative	-
co2_hbt	Emissions de CO2 par habitant en tonnes	Quantitative	
pma	Pays moins avancés	Indicatrice	+/- selon le cas

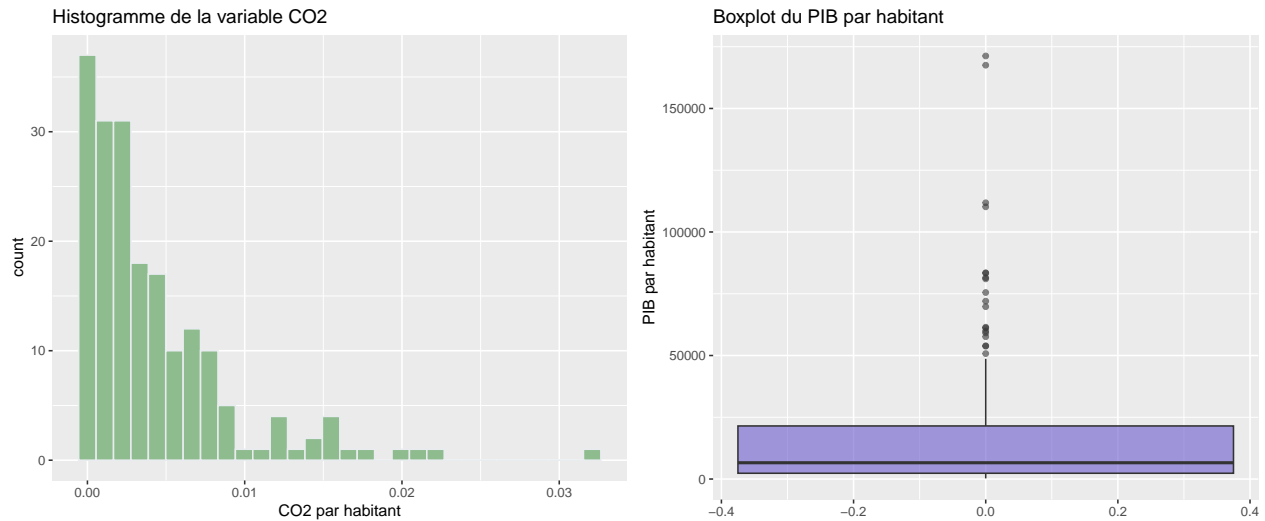
Table 2: Aperçu de la base de données

pays	pop_active	pop	pib_hbt	int_energy	energy_ren	co2_hbt	pma
Afghanistan	9 670 662	36 296 111	516.7	2.30	19.21	0.0001317	1
Albania	1 367 557	2 873 457	4 531.0	2.69	37.07	0.0017888	0
Algeria	12 184 474	41 389 174	4 109.7	4.93	0.14	0.0038256	0

### 3. Analyse descriptive des variables

#### 3.1. Analyse univariée

Nous avons décidé d'axer notre analyse univariée autour de trois variables intéressantes.



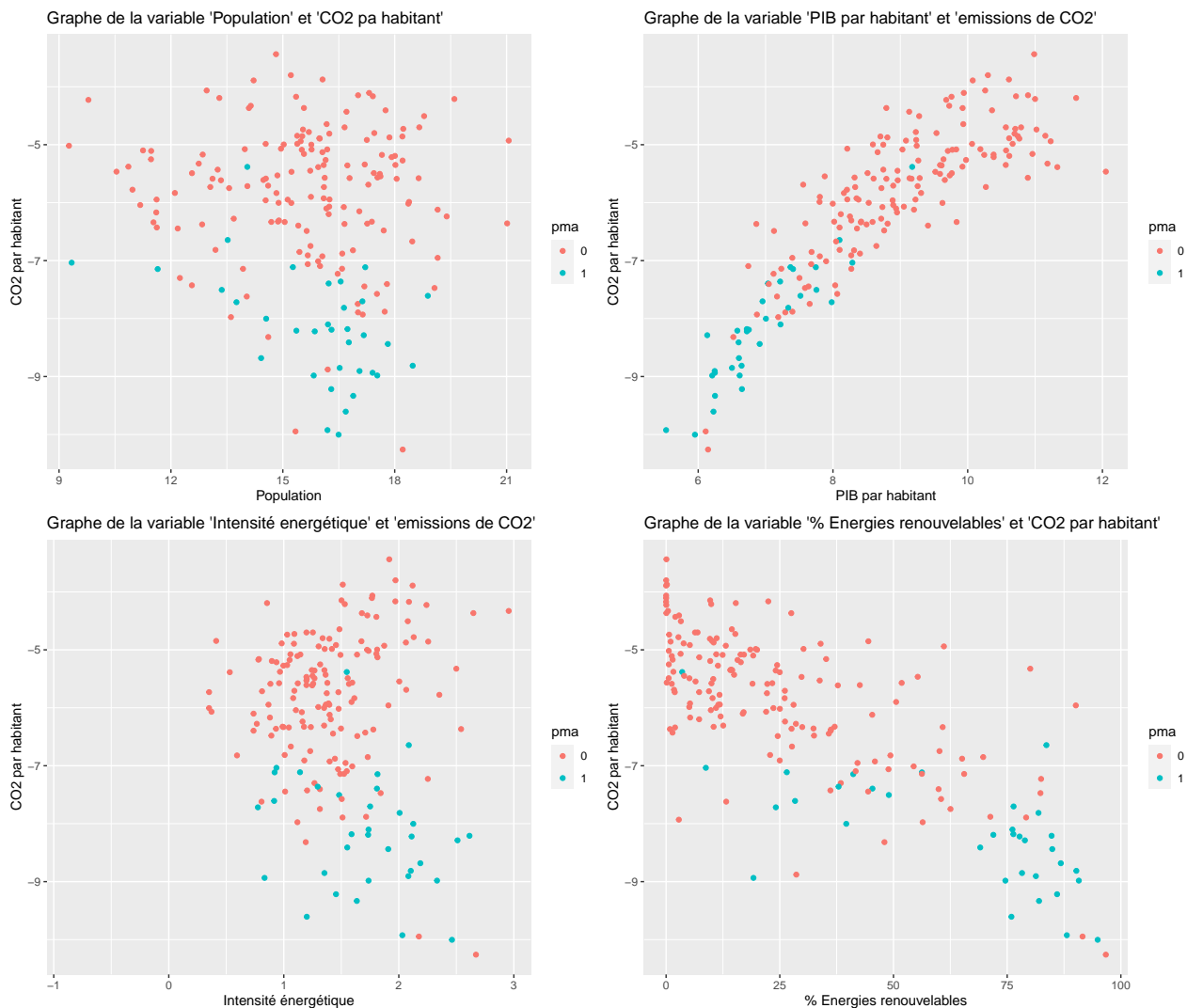
L'analyse de l'histogramme des émissions de CO2 montre que les observations sont assez concentrées entre 0 et 0,015. Toutefois, on note quelques valeurs atypiques comme la Chine. La raison sera qu'en tant que gros producteur pour la plupart des pays, il pollue beaucoup plus et couvre les tonnes d'émissions de gaz à effet de serre des pays vers lesquelles il exporte. En ce qui concerne la population active, elle est assez regroupée autour d'une valeur, mais possède quelques outliers qu'il faudra surveiller.

```
summary(ges_dt$energy_ren)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
0.00	5.26	19.15	28.62	45.33	96.70	4

On remarque que la plupart des valeurs sont situées entre le 1er et le 3ème quartile donc les valeurs sont bien regroupées autour de la moyenne. Ainsi, la plupart des pays utilisent en moyenne 30% d'énergies renouvelables pour leur consommation énergétique annuelle.

### 3.2. Analyse bivariée



Afin de mieux comprendre les relations entre nos variables, nous avons mis certaines variables en logarithme pour nous affranchir des problèmes d'échelles de mesure. L'analyse du nuage de points entre les *émissions de CO2 par habitant* et le *PIB par habitant* nous permet de constater l'existence d'une relation linéaire entre ces deux variables. De plus, on pourrait noter qu'il n'y a pas de réelles différences entre la manière dont les PMA et les non PMA sont dispersées. On peut donc conclure que plus le PIB par habitant est grand, plus on a d'émissions de CO2. La tendance s'inverse dans le cas des *énergies renouvelables*. En ce qui concerne les 02 autres graphes, on a des variables beaucoup plus dispersées, cela ne permet pas de statuer sur l'existence d'une relation linéaire claire entre chaque variable explicative et les émissions de CO2 par habitant. Par ailleurs, il est nécessaire de mettre en relief quelques particularités :

- au niveau de *l'intensité énergétique de l'économie*, on constate que les PMA sont plus regroupées (variance intra-groupe faible) tandis que les non PMA sont plus dispersées (variance intra-groupe forte) en fonction des émissions de CO2. Les pays moins avancés ont une intensité énergétique très **faible** ;
- même constat pour *la population* et les émissions de CO2. Plus un pays a une population élevée alors, chaque habitant a un impact environnemental plus important en matière d'émission de CO2.

Table 3: Matrice de corrélation

	pop_active	pop	pib_hbt	int_energy	energy_ren	co2_hbt
pop_active	1.0000000	0.9728681	-0.0214087	0.0283818	-0.0578384	0.0375479
pop	0.9728681	1.0000000	-0.0415370	0.0258385	-0.0423792	0.0113427
pib_hbt	-0.0214087	-0.0415370	1.0000000	-0.1644705	-0.3006836	0.6100919
int_energy	0.0283818	0.0258385	-0.1644705	1.0000000	0.2814451	0.1335438
energy_ren	-0.0578384	-0.0423792	-0.3006836	0.2814451	1.0000000	-0.5631909
co2_hbt	0.0375479	0.0113427	0.6100919	0.1335438	-0.5631909	1.0000000

### 3.3. Matrice de corrélation

```
corr_mat <- cor(na.omit(ges_dt[,2:7]))
kable(corr_mat,format="latex", booktabs = T, caption = "Matrice de corrélation")>%
  kable_styling(latex_options = "striped")
```

On constate sur la matrice de corrélation ci-dessus, une corrélation positive entre les émissions de CO2 et le PIB par habitant et inversement pour le cas du taux d'énergies renouvelables. Par ailleurs, ce taux est corrélé négativement avec le PIB, mais positivement avec l'intensité énergétique. Il faut aussi noter une faible corrélation entre l'intensité énergétique et les émissions de CO2.

## 4. Modèle

### 4.1. Spécification du modèle

Après avoir mis en route et comparé différents modèles, nous avons retenu un modèle en logarithme (3) avec des effets croisés sur le PIB et la variable PMA.

Dependent variable:			
	(1)	co2_hbt (2)	(3)
pop_active	0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)
pib_hbt	0.00000*** (0.000)		
int_energy	0.001*** (0.0001)		
log(pib_hbt)		0.002*** (0.0002)	
log(int_energy)		0.005*** (0.0005)	0.005*** (0.0005)
energy_ren	-0.0001*** (0.00001)	-0.0001*** (0.00001)	-0.0001*** (0.00001)

log(pib_hbt):pma0			0.002*** (0.0002)
log(pib_hbt):pma1			0.003*** (0.0003)
Constant	0.002*** (0.001)	-0.021*** (0.002)	-0.022*** (0.002)
-----			
Observations	173	173	173
R2	0.651	0.696	0.703
Adjusted R2	0.642	0.689	0.694
Residual Std. Error	0.003 (df = 168)	0.003 (df = 168)	0.003 (df = 167)
F Statistic	78.186*** (df = 4; 168)	96.308*** (df = 4; 168)	78.968*** (df = 5; 167)
=====			
Note:	*p<0.1; **p<0.05; ***p<0.01		

En effet, il s'agit du modèle avec le meilleur R2 ajusté. Il faudra faire un test de Ramsey pour vérifier la spécification du modèle.

## 4.2. Test de Ramsey

```
resettest(model3)
```

RESET test

```
data: model3
RESET = 53.08, df1 = 2, df2 = 165, p-value < 2.2e-16
```

La p-value du test de Ramsey est inférieure au seuil de 5% alors, le modèle est bien spécifié.

## 4.3. Interprétation des coefficients et qualité du modèle

- **Qualité d'ajustement du modèle**

Environ 69,39% de la variance des émissions en CO2 est expliquée par la population active, l'intensité énergétique et les effets croisés entre le PIB par habitant et le fait que le pays soit moins avancé ou non.

- **Test de Fisher global (test de significativité globale du modèle)**

Le modèle est **globalement satisfaisant** car la p-value de la statistique de Fisher est inférieure à 5%

- **Significativité des coefficients**

Le coefficient de la variable *pop\_active* n'est pas significatif au seuil de 5% car sa *p-value* est inférieure à 5%. En revanche, les variables sont significatives *energy\_ren* et *log(int\_energy)* au seuil de 5%. De plus, les effets croisés des variables *log(pib\_hbt)* et *pma* sont significatives au seuil de 5% et de 1%.

Lorsque l'intensité énergétique augmente d'un pourcent, les émissions de CO2 augmentent de 0,00005 tonne, soit 0,05 kg par habitant, toute chose égale par ailleurs. De plus, si le pourcentage d'énergie renouvelable utilisée augmente d'une unité de pourcentage alors les émissions de CO2 augmentent de 0,0001 tonne, soit 0,1 kg par habitant. En ce qui concerne les effets croisés :

- si le pays est un pma, lorsque le PIB par habitant augmente d'1% alors les émissions de CO2 augmentent 0,003 tonne soit 3kg par habitant

- sinon, lorsque le PIB par habitant augmente d'1% alors les émissions de CO2 augmente 0,002 tonne soit 2kg par habitant

#### 4.4. Analyse des résidus du modèle

```
model <- lm(co2_hbt~log(pib_hbt):pma+log(int_energy)+energy_ren,data=ges_dt)
```

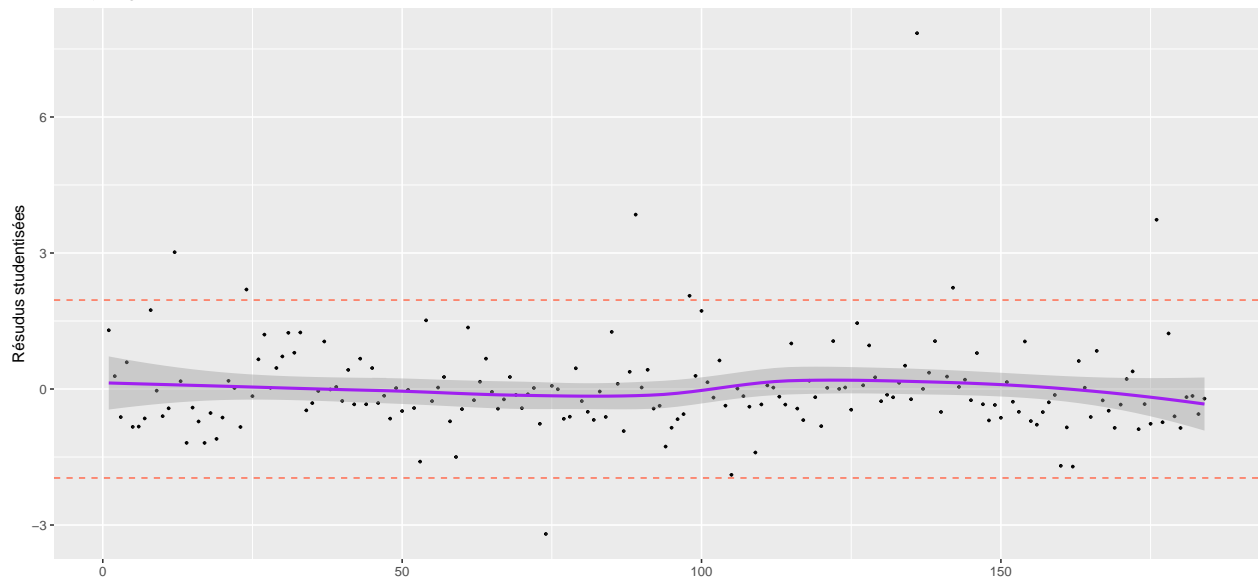
Avant d'analyser le résidu nous avons retiré la variable *pop\_active* qui n'était pas significative dans notre modèle.

##### Résidus studentisés

Les résidus studentisés obtenus par validation croisée nous montrent que les pays cités ci-dessous sont des variables atypiques dans notre modèle.

```
[1] "Austria"      "Bolivia"      "Germany"      "Iceland"
[5] "Jamaica"      "Myanmar"      "New Zealand"  "Solomon Islands"
```

Analyse graphique des résidus studentisés



On a procédé aussi à un lissage de nos résidus et ce lissage est assez proche de 0. Donc, on soupçonne que nos résidus ont une espérance nulle et de variance constante.

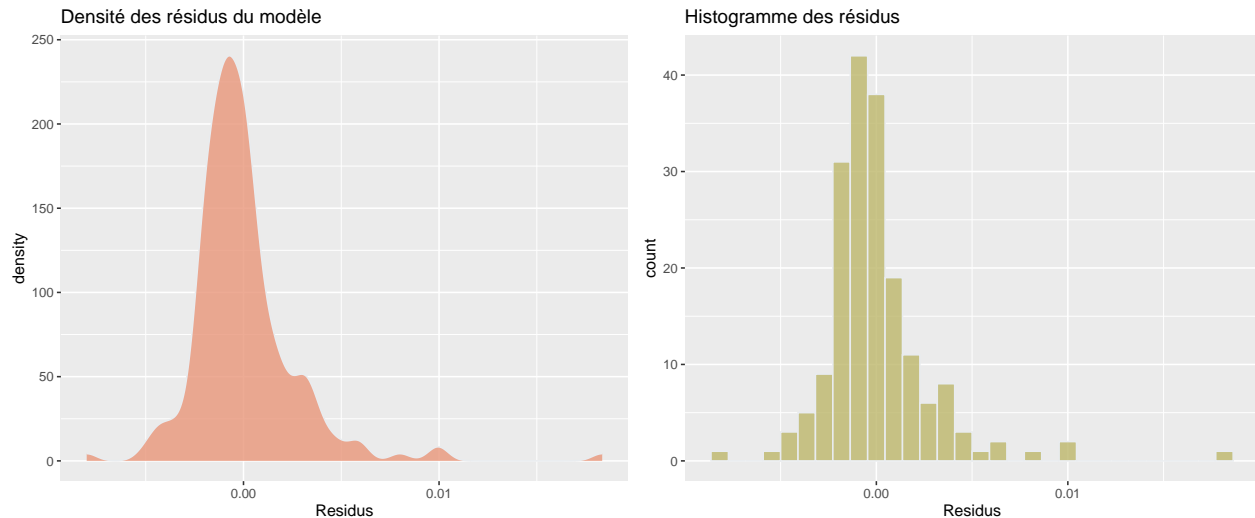
##### Test de normalité des résidus de Shapiro-Wilk

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

Shapiro-Wilk normality test

```
data: model$residuals
W = 0.82576, p-value = 1.458e-13
```

La p-value du test de normalité des résidus est inférieure au seuil de 5% alors on rejette l'hypothèse de normalité des résidus. Les résidus ne suivent pas une loi normale. Cela est dû à l'existence de valeurs atypiques qui étendent la distribution.



### Test d'hétéroscédasticité

```
bptest(model)
```

studentized Breusch-Pagan test

data: model

BP = 21.843, df = 4, p-value = 0.0002154

La p-value est inférieure au seuil de 5% alors on rejette l'hypothèse nulle d'homoscédasticité. Donc, on est en présence d'un problème d'hétéroscédasticité des résidus. Il faut donc appliquer la correction de White.

### Correction de White

```
coeftest(model,vcovHC(model))
```

t test of coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-2.3041e-02	3.5954e-03	-6.4083	1.263e-09	***
log(int_energy)	4.6931e-03	6.5888e-04	7.1229	2.485e-11	***
energy_ren	-5.1718e-05	1.1463e-05	-4.5117	1.162e-05	***
log(pib_hbt):pma0	2.5332e-03	3.2893e-04	7.7013	8.832e-13	***
log(pib_hbt):pma1	2.6930e-03	3.6891e-04	7.2998	9.076e-12	***

---

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

Après la mise en oeuvre de la méthode, on remarque que le niveau de significativité des coefficients a diminué car le modèle a été corrigée de l'hétéroscédasticité. Toutefois, les variables significatives demeurent toujours significatives au seuil de 5%.