

UNIVERSITÉ DE RENNES 1

ÉCONOMÉTRIE

Courbe de Kuznets environnementale pour les émissions de CO₂

Numéro étudiant :
22010709

Responsable :
Véronique THELEN

Auteur :
Romane LE GOFF

Janvier 2021



Table des matières

1	Introduction et présentation de l'article étudié	2
2	Présentation des données	3
3	Statistiques descriptives	4
4	Modélisation économétrique	9
4.1	Premier modèle : courbe de Kuznets environnementale classique	9
4.2	Second et troisième modèle : courbe de Kuznets environnementale augmentée de plusieurs variables	13
4.3	Modèle final : courbe de Kuznets environnementale augmentée du taux d'urbanisation	16
4.4	Pour aller plus loin : introduction de variables instrumentales	19
5	Conclusion	21
6	Bibliographie	21

Résumé

Le sujet de ce rapport porte sur la courbe de Kuznets et son application à une analyse environnementale. Le but de l'étude était d'évaluer la relation entre les émissions de CO₂ par habitant d'un pays et le PIB par habitant, puis d'essayer de l'enrichir de variables additionnelles. C'est une étude économétrique en coupe transversale portant sur les pays ayant des données sur leurs émissions en CO₂. 180 pays ainsi que 3 indicateurs supplémentaires ont été sélectionnés, sur la période de 2014. Les résultats ont démontré l'existence d'une relation entre la pollution émise à l'échelle d'un individu et le PIB par habitant. Le taux d'urbanisation joue a priori également un rôle dans la détermination des émissions de CO₂ par habitant. Les résultats suivent les affirmations de T. Jobert et F. Karanfil dans leur article sur la formation et la déformation de la courbe de Kuznets environnementale, laissant le doute sur la forme quadratique de la relation et l'existence même de la courbe de Kuznets environnementale. Les tests de Sargan et Wu-Hausman suggèrent une modélisation avec des variables instrumentales.

1 Introduction et présentation de l'article étudié

L'étude économétrique au centre de ce document s'appuie sur la modélisation économétrique des émissions de CO₂ au niveau mondial en 2014. Nous nous appuyerons sur un article de T. Jobert et F. Karanfil qui traite de la formation et déformation de la courbe de Kuznets environnementale. La courbe de Kuznets est connue pour décrire une relation en forme de U inversé entre le niveau de développement d'un pays et les inégalités de revenus. Il s'avère qu'elle peut également s'appliquer à une analyse environnementale qui décrirait une relation entre le développement économique (PIB par habitant) et la pollution (émissions de CO₂ par habitant). L'article, publié en 2012, s'intéresse à cette relation et plus particulièrement à la nature de l'apparition de cette relation depuis les 1970. Existe-t-il réellement une relation en forme de U inversé et un point de retournement suggérant une diminution de la pollution d'un État lorsqu'il atteint un stade de développement économique avancé ?

Jobert et Karanfil soulignent que les études empiriques ne permettent pas de trancher sur la nature de la relation. Certains, comme Halkos et Tsionas, parlent de relation monotone. D'autres suggèrent l'existence d'un « havre de pollution » traduisant la délocalisation de la pollution des pays riches vers les pays en voie de développement, qui serait déterminé par l'ouverture commerciale. Souvent, les économistes s'intéressent à l'ajout de variables additionnelles au modèle pour déterminer de nouvelles tendances, parfois autres que quadratiques. Jobert et Karanfil se sont intéressés à une modélisation économétrique avec la méthode des Moindres Carrés Ordinaires comprenant 51 pays et incluant la consommation d'énergie primaire. Les 51 pays qu'ils ont sélectionnés recouvrent plus de 80% des émissions de CO₂. Même si leurs résultats laissent penser à l'apparition d'une

forme en U inversé dans la relation entre les émissions de CO₂ et la richesse à partir des années 80, ils ne confirment malheureusement pas l'idée optimiste de l'existence d'une CKE dans le temps. La tendance d'amélioration d'efficacité observée ne concernerait que les pays les plus riches et serait combinée avec une diminution des résultats dans les pays pauvres. Pire, le point de retournement semble se décaler vers la droite au fil des années. Ainsi, nous allons comparer ces résultats avec ceux de notre jeu de données, tout en essayant de dégager d'autres déterminants des émissions de CO₂ per capita à l'échelle d'un pays. Le taux d'urbanisation, en plus du PIB par habitant, fera notamment l'objet de notre analyse. Avant de commencer, soulignons tout de même que la dimension temporelle présente dans l'article ne fait pas partie de notre analyse, nous nous intéressons à un point (une année) fixe dans le temps. L'analyse faite dans les sections suivantes sera donc en quelque sorte une "mise à jour" des observations pessimistes faites par Jobert et Karanfil, les données étudiées étant plus récentes (2014), et concernant un plus grand nombre de pays.

2 Présentation des données

Le jeu de données sur lequel nous nous appuyerons principalement est issu de World Development Indicators (WDI), une banque mondiale de données développant des indicateurs à la portée de tous. Les indicateurs du jeu de données sont fournis par différents organismes, principalement la Banque Mondiale. On y retrouve les émissions de CO₂ par habitant en 2014 de 180 pays dans le monde ainsi que 4 variables principales :

- Le PIB à parité du pouvoir d'achat par habitant (gdp)
- Les émissions de CO₂ en tonnes par habitant (CO2emissions)
- Le total d'importations et d'exportations de biens et services en % de PIB, autrement dit le ratio Commerce/PIB. Cette variable mesure l'ouverture commerciale des pays (trade)
- Le pourcentage de population vivant en zone urbaine (urban_pop), autrement dit le taux d'urbanisation

Nous y avons également ajouté une variable représentant la production de bœuf par habitant (*new*), qui est fournie par le site web OurWorldInData dont la source est l'Organisation des Nations Unies pour l'alimentation et l'agriculture (FAO). L'ajout de cette variable réduit notre jeu de données à 173 pays. Il nous semblait intéressant d'établir un modèle prenant en compte la production de viande bovine, qui est très polluante, de chaque pays, réduite par habitant. Néanmoins, nous verrons plus tard qu'elle ne sera pas retenue dans le modèle.

Ont donc été fusionnées avec le jeu de données principal la production de bœuf en kilogrammes par pays et la population de chaque pays en 2014 (source : Nations Unies). Cette fusion de jeu

de données a été réalisée dans l'objectif de créer la variable "new" qui représente la production de bœuf en kilogrammes par habitant, plus pertinente dans notre analyse.

Dans un objectif de mise en place d'un modèle avec variables instrumentales, nous avons également inclus des variables décrivant l'espérance de vie (*life*) et de densité de population (*density*) des États présents dans le jeu de données. Les données correspondantes sont issues une nouvelle fois de WDI, dont les sources sont la Banque Mondiale et l'Organisation de l'Alimentation et de l'Agriculture.

Finalement, une variable indicatrice segmentant les pays faisant partis de l'OCDE, selon la Banque Mondiale, des autres a été créée dans le but de comparer les émissions de CO₂ par habitant de ces deux catégories de pays.

3 Statistiques descriptives

Nous commençons cette analyse par une brève description des données et des variables que nous avons à disposition. La table 1 résume nos variables.

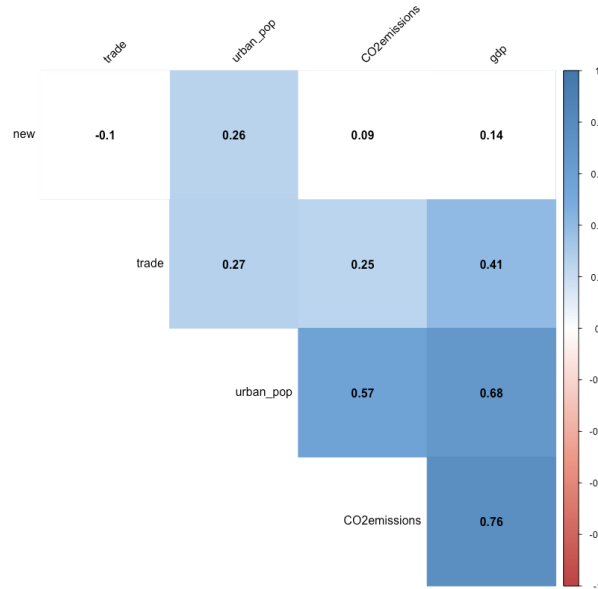
TABLE 1 – Résumé des variables du jeu de données

Statistic	N	Mean	St. Dev.	Min	Max
CO2emissions	173	4.514	5.870	0.044	45.423
trade	173	93.712	56.032	19.459	425.976
gdp	173	18,711.410	21,101.570	597.135	130,755.200
urban_pop	173	57.340	23.396	11.776	100.000
new	173	11.961	21.212	0.009	151.336
life	171	71.486	8.277	49.891	83.980
density	171	339.759	1,676.863	1.892	19,478.810
OECD	173	0.121	0.328	0	1

Le pays moyen de ce jeu de données émet 4.5 tonnes de CO₂ par habitant, a une richesse par habitant de 18711\$, et 57% de sa population vit en zone urbaine. L'espérance de vie de sa population est de 71 ans, et il y a en moyenne 340 individus vivant sur un km² de surface. Le rapport moyen commerce/PIB tourne autour de 93%. Finalement, la consommation de viande bovine annuelle par habitant atteint 11.776 kg.

Il manque deux observations d'espérance de vie et de densité de population. Ces variables étant utilisées uniquement pour la modélisation des variables instrumentales, il n'a pas été jugé nécessaire de les supprimer pour la première modélisation économétrique.

FIGURE 1 – Corrélations entre les variables



La matrice des corrélations est représentée graphiquement par la figure 1 ci-dessus. Plus la couleur est foncée, plus les deux variables sont corrélées entre elles. On remarque notamment que toutes les corrélations sont positives, sauf une qui est néanmoins négligeable, et que les coefficients associés à la variable *new* sont proches de 0. La production de viande bovine annuelle par habitant n'a donc pas de lien direct avec le reste des variables, et encore moins avec les émissions de CO_2 . En ce qui concerne le PIB par habitant, il est fortement corrélée positivement avec les émissions de CO_2 par habitant, à hauteur de 0.76. Cela signifie qu'un pays avec une richesse par habitant élevée engendre davantage de pollution à l'échelle de l'habitant. Le PIB par habitant et le taux d'urbanisation sont également corrélés positivement (coefficient égal à 0.68). Par conséquent, la population d'un pays s'urbanise à mesure qu'il s'enrichit.

Vous trouverez également les p-valeurs associées à chaque coefficient de corrélation dans la table 2 page suivante, révélant que certains coefficients ne sont pas statistiquement significatifs (corrélations entre *new* et *CO2emissions* ; *trade* et *new*). Ces coefficients ne sont donc pas à prendre comme

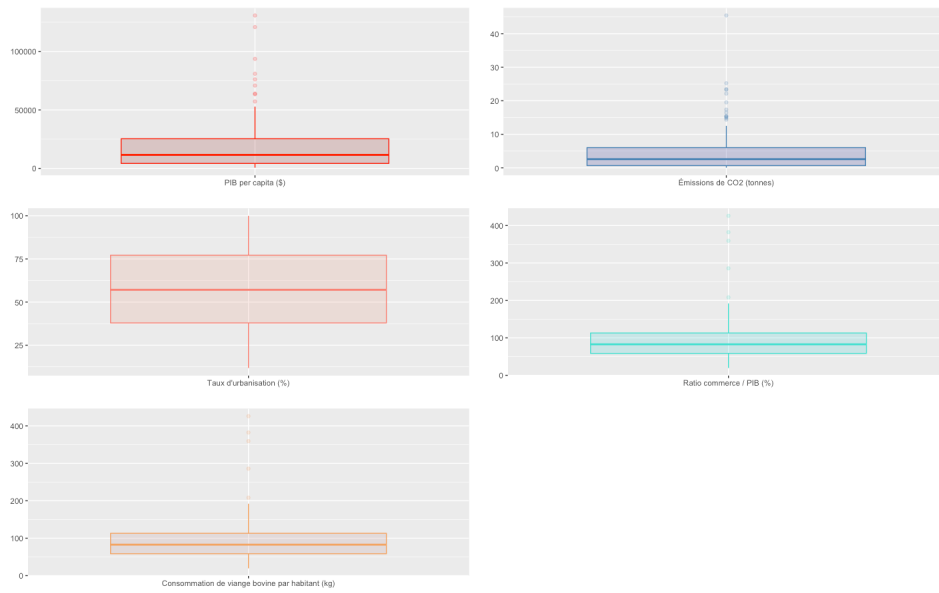
vérifiables et représentatifs de la réelle relation entre la production de viande bovine (*new*) et les émissions de CO₂ ou celle entre *new* et l'ouverture commerciale.

TABLE 2 – p-valeurs de la matrice des corrélations

	CO2emissions	urban_pop	gdp	new	trade
CO2emissions	0	0	0	0.24	0.001
urban_pop	0	0	0	0.001	0.0004
gdp	0	0	0	0.07	0.0000
new	0.24	0.001	0.07	0	0.19
trade	0.001	0.0004	0.0000	0.19	0

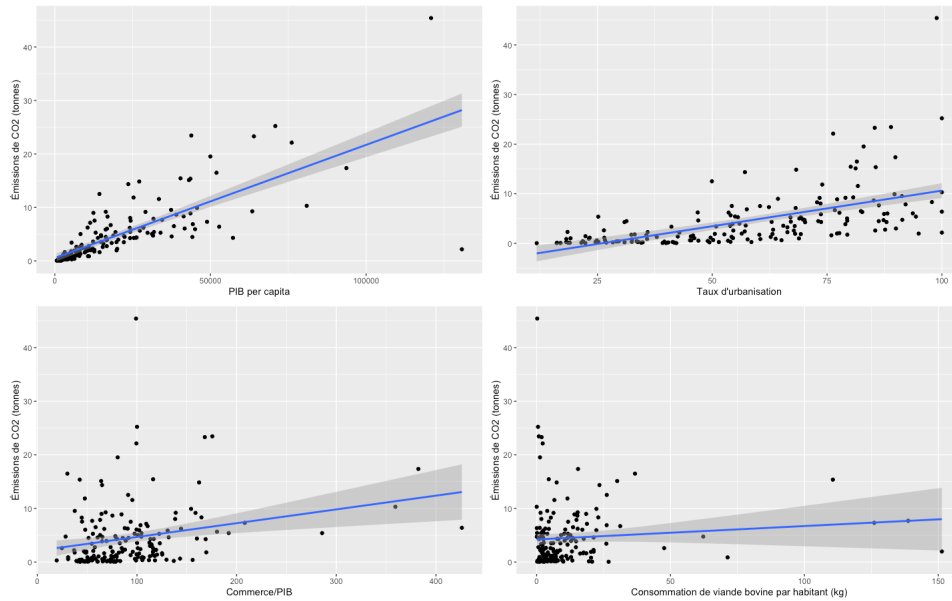
La disparité des variables ainsi que les éventuels individus atypiques sont représentés par les boîtes à moustache ci-dessous. La variable la plus étendue est le taux d'urbanisation tandis que le PIB par habitant est celle la plus concernée par les individus atypiques. On remarque également de nombreux outliers pour les émissions de CO₂ par habitant. La Chine fait notamment partie de ces individus. Elle connaît effectivement une période de croissance disproportionnée par rapport aux autres pays. Le Qatar en fait également partie.

FIGURE 2 – Boîtes à moustache des variables



Le modèle que nous voulons produire expliquera les émissions de CO_2 par habitant en fonction d'une ou plusieurs variables explicatives. Voyons tout d'abord par le biais d'une représentation graphique s'il existe une relation linéaire entre les variables concernées et la pollution émise.

FIGURE 3 – Émissions de CO₂ par habitant en fonction des autres variables explicatives



On remarque qu'il existe des cas isolés s'éloignant considérablement de la droite de régression générée par la fonction *ggplot* sur R, ils correspondent à des pays évoluant anormalement par rapport aux autres. Le reste de chacun des nuages de points est regroupé dans un petit intervalle, excepté encore une fois pour la consommation de viande bovine, et également pour le ratio commerce/PIB. Cela sous-entend donc qu'il n'existe pas réellement de relation linéaire entre les émissions de CO₂ en tonnes par habitant et ces deux variables.

Pour les deux variables restantes, nous pouvons dégager une tendance croissante en fonction de notre variable à expliquer. Ainsi, d'un point de vue économique, plus la richesse par habitant augmente et plus le taux d'urbanisation augmente, plus les émissions de CO₂ par habitant seront également importantes.

Au final, ces statistiques descriptives nous donnent déjà une idée de quelles variables seront conservées plus tard dans notre modèle économétrique final (*urban_pop*, *gdp*).

4 Modélisation économétrique

4.1 Premier modèle : courbe de Kuznets environnementale classique

Dans un premier temps, nous allons modéliser la courbe de Kuznets environnementale classique, c'est-à-dire uniquement avec le PIB par habitant passé en logarithme et en logarithme carré.

Voyons tout d'abord si la relation linéaire suggérée par le graphique n°1 de la figure 3 page précédente est envisageable. Nous avons appliqué le test de Ramsey sur le modèle suivant : $\log(CO2emissions) = \beta_0 + \beta_1 \log(gdp)_i + \epsilon_i$. Celui-ci nous renvoie une p-valeur nettement inférieure à 0.05, nous confirmant la pertinence d'ajouter un terme quadratique à notre régression.

Une modélisation de notre jeu de données avec la méthode MCO nous donne le modèle de régression quadratique suivant :

$$\log(CO2emissions) = \beta_0 + \beta_1 \log(gdp)_i + \beta_2 \log(gdp)_i^2 + \epsilon_i$$

avec $\beta_0 = -19.70$, la constante, $\beta_1 = 3.34$ et $\beta_2 = -0.12$ les coefficients des variables explicatives et ϵ_i les aléas. Ce modèle est soumis aux hypothèses suivantes, le validant :

- $E[\epsilon] = 0$
- $V[\epsilon] = \sigma^2$: les aléas sont constants (homoscédasticité des résidus)
- $Cov(\epsilon_i, \epsilon_j) = 0 \quad \forall i \neq j$ (indépendance des résidus)

Le résultat de la régression est visible dans la table 2 page 6. On remarque tout d'abord que la qualité d'ajustement est déjà très bonne, avec un coefficient de détermination égal à 0.8567 et un coefficient de détermination ajusté égal à 0.855. Ainsi, environ 86% de la variation des émissions de CO₂ par habitant est expliquée par le modèle. La table nous donne également une analyse supplémentaire, qui est le test de Fisher, nous informant de la validité du modèle. Ce test est composé de plusieurs contraintes avec les hypothèses suivantes :

$$\begin{cases} H_0 : \beta_0 = \beta_1 = \beta_2 = 0 \\ H_1 : \exists k \in \{0, 1, 2\} \text{ tel que } \beta_k \neq 0 \end{cases}$$

TABLE 3 – Résultat MCO n°1

émissions de CO ₂ en log	
log(CO ₂ emissions)	
modèle Kuznets environnementale	
log(gdp)	3.34*** (0.51)
I(log(gdp) ²)	−0.12*** (0.03)
Constant	−19.70*** (2.29)
Observations	173
R ²	0.86
Adjusted R ²	0.86
F Statistic	508.16*** (df = 2 ; 170)
<i>Note :</i>	*p<0.1 ; **p<0.05 ; ***p<0.01

La statistique de Fisher étant de 508.16 et la p-valeur associée à cette statistique étant nettement inférieure au seuil de 5%, nous rejetons l'hypothèse nulle que le modèle n'est pas valide.

En outre, R a effectué un test de Student à chaque variable nous indiquant si oui ou non les variables sont significatives et doivent être gardées dans le modèle. Le test comprend les hypothèses ci-dessous :

$$\begin{cases} H_0 : \beta_k = 0 \ \forall k \in \{0, 1, 2\} \\ H_1 : \beta_k \neq 0 \ \forall k \in \{0, 1, 2\} \end{cases}$$

L'hypothèse nulle est rejetée pour la constante ainsi que pour les deux autres variables du modèle au seuil de 5 comme de 1%, nous confirmant la significativité de chacune d'entre elles dans le modèle.

Le modèle est donc globalement satisfaisant, mais nous allons tenter plus loin dans le document de l'améliorer et/ou de le mettre à jour en y incorporant de nouvelles variables susceptibles d'influer sur la variable dépendante.

D'après la table 3 page précédente, β_2 est négatif, ce qui signifie que la courbe est concave et donc en forme de U inversé, en accord avec la théorie initiale de la courbe de Kuznets. Le point de retournement, c'est-à-dire le point à partir duquel la relation devient décroissante est donné par :

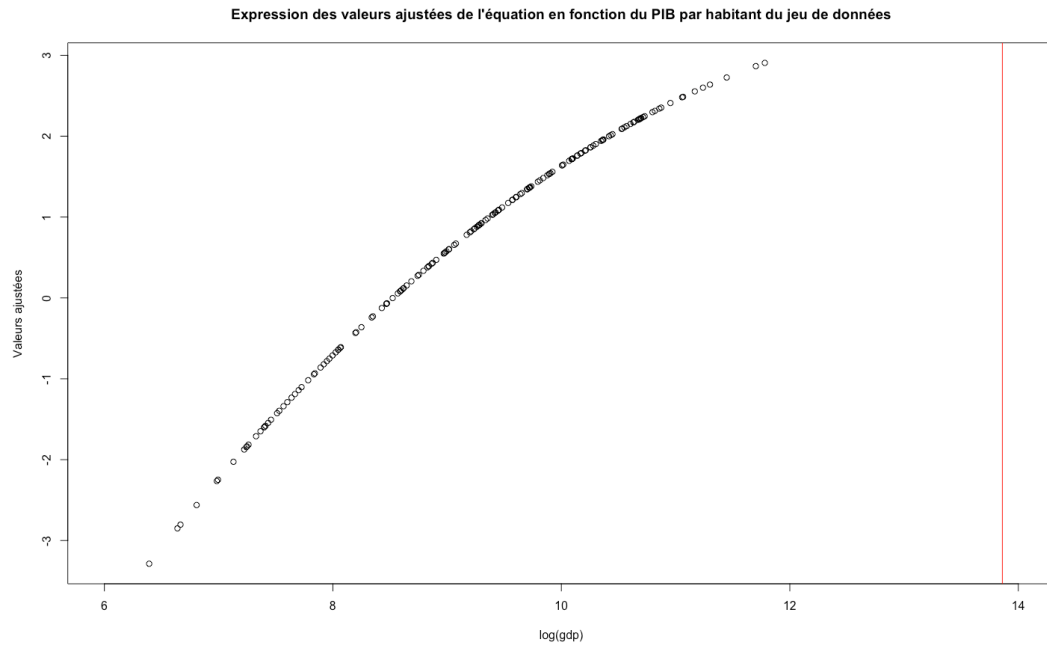
$$X = -\frac{\beta_1}{2\beta_2} = 13.86$$

Il représente donc le niveau de revenu permettant de pourvoir les besoins primaires, correspondant à un "niveau-seuil", où le souci pour l'environnement est censé s'accroître et la tendance s'inverser. Ce seuil marque le début d'une croissance économique s'accompagnant d'une amélioration des conditions environnementales et d'une diminution des émissions de CO₂ par habitant, donc de la pollution.

Si l'on compare ce nouveau point de retournement avec ceux calculés par Jobert et Karanfil, il s'avère être nettement inférieur aux derniers calculs du début des années 2000, se rapprochant de 20. Néanmoins, cette différence est à nuancer étant donné que notre jeu de données regroupe 173 pays contre 51 des plus polluants pour leurs résultats.

Ainsi, en repassant en niveau, le point de retournement de notre modèle correspond à un niveau de revenu de $\exp(13.92) = 1,045,653\$$ (PIB /habitant), qui ne concerne aucun pays de notre jeu de données (la moyenne étant à 18 711\$ de PIB par habitant, et le max étant de 130 755\$). En appliquant le modèle à nos données par une représentation graphique à la page 11, nous remarquons le début d'une courbe en forme de U inversée : une future stagnation commence à apparaître.

Néanmoins, nos données sont encore loin de l'utopique point de retournement (ligne rouge), mettant le doute sur la forme de la courbe. On se demande si la tendance va effectivement s'inverser, ou si la future stagnation de la pollution émise se transformera en une nouvelle tendance croissante à mesure que les pays s'enrichissent. Dans le deuxième cas, cela traduirait une monotonie de la relation.



Finalement, nous pouvons interpréter ce premier modèle de la façon suivante : si le PIB par habitant augmente de 1%, alors les émissions de CO_2 par habitant augmentent de $3.34 - 0.12x1^2 = 3.22\%$. Un test de Ramsey sur ce modèle nous donne une p-valeur de $0.1916 > 0.05$ nous confirmant la bonne spécification du modèle. Néanmoins, même si le modèle est satisfaisant et confirmé par la littérature, nous aimerions en faire une mise à jour et expliquer les émissions de CO_2 par d'autres variables que le PIB par habitant.

4.2 Second et troisième modèle : courbe de Kuznets environnementale augmentée de plusieurs variables

Dans cette partie, nous tentons d'ajouter à la courbe de Kuznets environnementale de nouvelles variables explicatives qui pourraient influencer la pollution en dioxyde de carbone d'un pays. Au nouveau modèle s'ajoutent les variables *trade*, *urban_pop*, *new* et *OECD*, introduites dans la section 3.

$$\begin{aligned}\log(CO2emissions) = & \beta_0 + \beta_1 \log(gdp)_i + \beta_2 \log(gdp)_i^2 + \beta_3 trade_i \\ & + \beta_4 urban_pop_i + \beta_5 new_i + \beta_6 OECD_i + \epsilon_i\end{aligned}$$

Avec $\beta_1 = 3.29$, $\beta_2 = -0.12$, $\beta_3 = -0.0002$, $\beta_4 = 0.01$, $\beta_5 = -0.25$ et $\beta_6 = -19.49$. On remarque que les coefficients β_1 et β_2 restent quasiment inchangés, de même pour la constante. Visibles dans la table 3, première colonne, les coefficients de détermination et de détermination ajusté restent presque les mêmes également, mais la statistique de Fisher a diminué ($174.50 < 508.16$) suggérant une baisse de la qualité du modèle. Certaines variables explicatives ne sont pas statistiquement significatives individuellement (*trade*, *new* et *OECD*).

Cela signifie donc que l'ouverture commerciale n'a statistiquement pas d'effet significatif sur la pollution émise par habitant en matière de dioxyde de carbone. Nous ne pourrions alors pas vérifier l'hypothèse du « havre de pollution », suggérant la délocalisation de la pollution des pays riches vers les pays en voie de développement par l'ouverture commerciale, vue dans l'introduction. De même, la pollution générée par la production de viande bovine, réduite en kilogrammes par habitant, en 2014 ne semble pas être suffisante pour avoir un effet significatif sur les émissions totales de CO₂.

Nous n'avons donc aucun intérêt à conserver ce modèle.

TABLE 4 – Résultats MCO n°2

	émissions de CO ₂ par habitant en log	
	log(CO2emissions)	
	MCO modèle initial	MCO modèle intermédiaire
	(1)	(2)
log(gdp)	3.29*** (0.52)	3.29*** (0.51)
I(log(gdp)^2)	-0.12*** (0.03)	-0.12*** (0.03)
trade	-0.0002 (0.001)	
urban_pop	0.01** (0.003)	0.01* (0.003)
new	-0.003 (0.002)	
OECD	-0.25 (0.16)	-0.23 (0.15)
Constant	-19.49*** (2.33)	-19.51*** (2.28)
Observations	173	173
R ²	0.86	0.86
Adjusted R ²	0.86	0.86
F Statistic	174.50*** (df = 6 ; 166)	261.03*** (df = 4 ; 168)

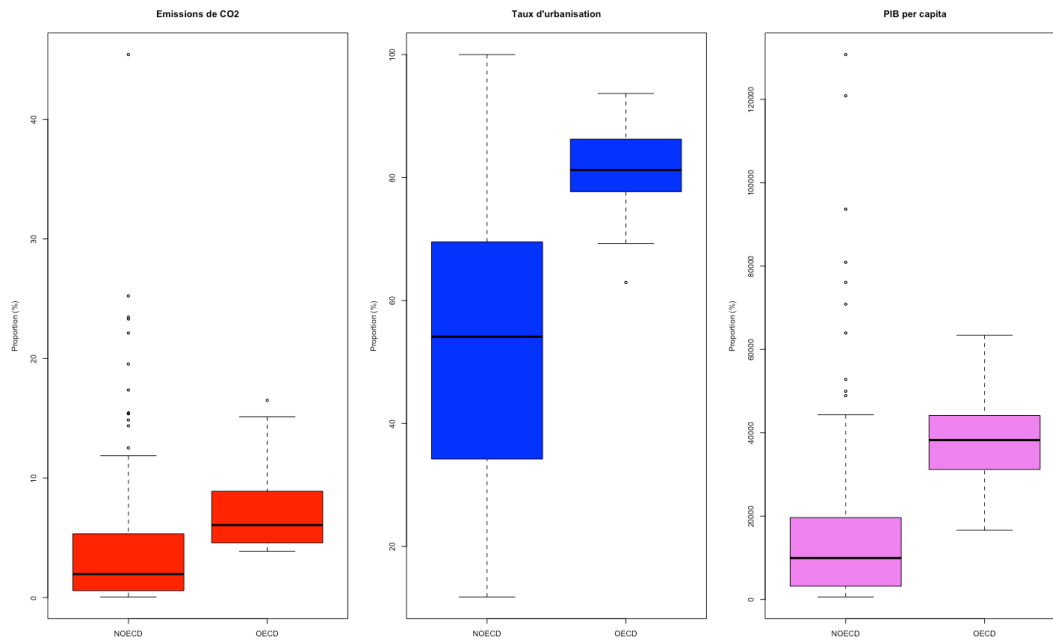
Note :

*p<0.1 ; **p<0.05 ; ***p<0.01

Après avoir retiré *trade*, puis *new*, qui s'avéraient n'apporter aucune information au modèle, nous obtenons le modèle visible sur la deuxième colonne de la table 4 page précédente, où il ne reste que les variables additionnelles *OECD* et *urban_pop*. *urban_pop* est statistiquement significative à un niveau de 10% tandis que la variable indicatrice *OECD* ne l'est pas.

Effectuons un test de Chow sur la variable indicatrice, afin d'être certain qu'il n'y ait pas de changement structurel dans notre modèle. Ci-dessous trois boîtes à moustaches décrivant les différences de valeurs entre les pays faisant partie de l'OCDE (*OECD*) et ceux qui ne le font pas (*NOECD*) pour les variables que l'on a gardées dans le troisième modèle. Il semblerait qu'il y ait des différences au niveau du PIB et de l'urbanisation, bien que la variance des pays hors OCDE soit très dispersée pour l'urbanisation. La différence entre ces deux catégories de pays pour les émissions de CO₂ par habitant est moins marquée.

Vérifions cela avec les résultats du test de Chow. Ce dernier (cf code R en annexe) nous renvoie une p-valeur de $0.58 > 0.05$ nous confirmant qu'il n'y a pas de différence significative entre ces deux catégories de pays, et donc pas de changement structurel à prévoir. Ainsi, nous ne gardons pas la variable *OECD* dans notre modèle.



4.3 Modèle final : courbe de Kuznets environnementale augmentée du taux d'urbanisation

Finalement, après avoir retiré toutes les variables n'apportant rien à la modélisation, nous obtenons le modèle suivant :

$$\log(CO2emissions) = -19.89 + 3.40 \log(gdp)_i - 0.12 \log(gdp)_i^2 + 0.01 urban_pop_i + \epsilon_i$$

TABLE 5 – Résultat MCO final

émissions de CO ₂ par habitant en log	
	log(CO2emissions)
	courbe Kuznets augmenté
log(gdp)	3.40*** (0.51)
I(log(gdp)^2)	-0.13*** (0.03)
urban_pop	0.01* (0.003)
Constant	-19.89*** (2.28)
Observations	173
R ²	0.86
Adjusted R ²	0.86
F Statistic	344.02*** (df = 3 ; 169)
Note :	*p<0.1 ; **p<0.05 ; ***p<0.01

Toutes les variables sont significatives (cf table 5) à hauteur d'au moins 10%, le modèle a perdu en significativité globale par rapport au premier (statistique de Fisher = 355.02 < 508.16) mais

reste tout de même très satisfaisant. Le modèle explique toujours 86% de la variance de la pollution en dioxyde de carbone par habitant et nous apporte désormais une information sur l'impact du taux d'urbanisation d'un pays sur ce type de pollution. Une augmentation de 1% du taux d'urbanisation implique ainsi une hausse de 1% des émissions de CO₂ par habitant.

Le test de Ramsey appliqué à ce dernier modèle nous renvoie une p-valeur de 0.429, confirmant une bonne spécification. Néanmoins, nous avons vu dans la partie de statistiques descriptives que les variables *gdp* et *urban_pop* étaient très corrélées positivement. C'est pourquoi un test d'analyse de la variance a également été réalisé afin de déterminer s'il était pertinent de préférer ce modèle à la courbe de Kuznets environnementale classique vue en 5.1. Les hypothèses suivantes sont testées :

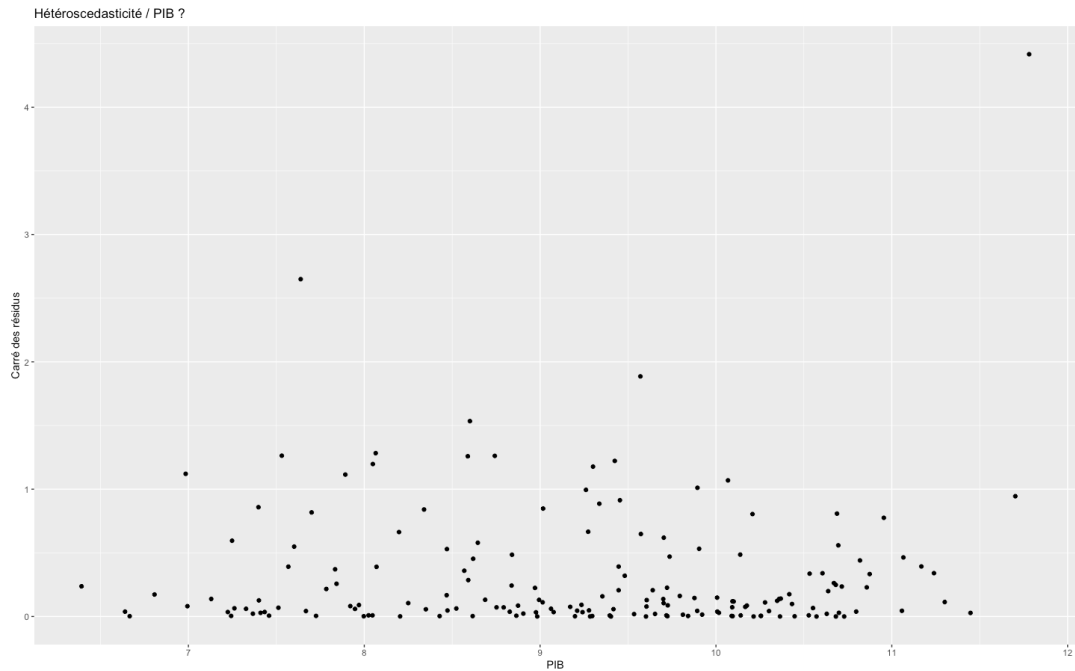
$$\begin{cases} H_0 : \log(CO2emissions) = \beta_0 + \beta_1 \log(gdp)_i + \beta_2 \log(gdp)_i^2 + \epsilon_i \\ H_1 : \log(CO2emissions) = \beta_0 + \beta_1 \log(gdp)_i + \beta_2 \log(gdp)_i^2 + \beta_3 urban_pop_i + \epsilon_i \end{cases}$$

TABLE 6 – ANOVA

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Res.Df	2	169.500	0.707	169	169.2	169.8	170
RSS	2	55.245	0.713	54.740	54.992	55.497	55.749
Df	1	1.000		1.000	1.000	1.000	1.000
Sum of Sq	1	1.009		1.009	1.009	1.009	1.009
F	1	3.115		3.115	3.115	3.115	3.115
Pr(>F)	1	0.079		0.079	0.079	0.079	0.079

La table 6 ci-dessus nous indique une p-valeur de $0.079 < 0.10$ au résultat du test. On peut donc conclure que l'ajout de la variable *urban_pop* apporte effectivement de l'information supplémentaire au modèle, à un niveau de 10%. Il est donc préférable de conserver la variable *urban_pop*.

Analysons désormais les résidus de notre modèle final. À la page 17, vous trouverez le graphique représentant les résidus au carré fonction du PIB par habitant transformé en logarithme.



Le nuage de points semble plus ou moins homogène, à l'exception de quelques individus atypiques. Ces individus peuvent ainsi être la cause d'une hétéroscédasticité, ne respectant pas l'hypothèse d'homogénéité de la variance décrite plus haut. Cependant, les tests de Breusch-Pagan et de Goldfeld-Quandt nous renvoient des p-valeurs de 0.1312 et 0.6999 respectivement (cf code R en annexe), suggérant que le caractère homoscedastique des résidus est respecté. En effet, lors de ces tests, les hypothèses suivantes sont testées :

H_0 : Homoscedasticité de la variance des résidus

H_1 : Hétéroscédasticité de la variance des résidus

pour le test de Breusch-Pagan. Et, pour le test de Goldfeld-Quandt :

$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma$

H_1 : Variance est fonction de la variable $\log(gdp)$

Les deux hypothèses nulles étant acceptées, nous pouvons conclure que les individus atypiques n'ont pas une influence suffisamment importante sur les données pour avoir un effet hétéroscédastique sur la variance des résidus.

4.4 Pour aller plus loin : introduction de variables instrumentales

Nous avons choisi l'espérance de vie et la densité de population comme variables pouvant expliquer du PIB par habitant que nous considérons comme endogène dans notre modèle. En effet, l'espérance de vie d'un pays est directement liée au PIB par habitant de celui-ci. Une hausse de l'espérance de vie mène à une baisse de la fertilité et à de plus grands investissements en capital humain. Par conséquent, le revenu par habitant augmente. Ce phénomène ne concerne néanmoins que les pays ayant déjà connu une transition démographique, selon le travail de M. Cervallati en 2009 (Life Expectancy and Economic Growth). D'autre part, la taille de la population, donc la densité, conduit à un changement technologique, résultant de la croissance de la production et donc du PIB par habitant. C'est pourquoi il a été jugé pertinent d'utiliser ces variables comme instruments du PIB par habitant.

Pour réaliser ce dernier modèle avec les variables *life* et *density* introduites au début de ce document, il a été nécessaire de supprimer les valeurs manquantes qui leur étaient associées. Nous avons donc supprimé 4 pays du jeu de données.

La fonction *ivreg* du package *AER* permet d'effectuer les deux étapes nécessaires à la modélisation d'un modèle avec variables instrumentales. La première étape consiste à déterminer s'il existe bien une relation caractérisant l'espérance de vie et la densité de population comme fonction du PIB par habitant, ici transformé en log. La seconde étape réalisée par la fonction consiste à remplacer $\log(gdp)$ dans la régression par la relation établie dans l'étape 1. Cela permet ainsi de comparer les MCO et la régression instrumentée, en effectuant différents tests tels que le test de Sargan, de Wu-Hausman et des instruments faibles. Ces derniers déterminent si les instruments sont valides, c'est-à-dire exogènes, si la régression instrumentée est préférable aux MCO et si les instruments sont faibles. On obtient les p-valeurs suivantes (cf code R en annexe) :

- Sargan : $0.104023 > 0.05$
- Wu-Hausman : $0.000243 < 0.05$
- Faibles instruments : $0.00000204 < 0.05$ et statistique de Fisher = 14

Ainsi, le test de Sargan accepte l'hypothèse nulle de l'exogénéité des instruments et le test de Wu-Hausman n'accepte pas l'hypothèse nulle que la méthode des MCO est préférable à la méthode des variables instrumentales. Quant au dernier test, l'hypothèse nulle que les instruments sont faibles n'est pas acceptée (statistique de Fisher = 14 > 10, le choix des instruments est alors pertinent pour l'analyse des émissions de CO₂ par habitant).

De plus, le modèle généré par *ivreg*, visible dans la table 7 page suivante, nous indique que toutes variables explicatives passent le test individuel de significativité de Student. Ils sont statistiquement

significatifs au niveau de 5%. Les coefficients de détermination et de détermination ajusté perdent 10 points de pourcentage par rapport au modèle de la partie 5.3, mais restent toujours très satisfaisants.

TABLE 7 – Modèle généré par ivreg

	<i>Variable dépendante :</i>
	log(CO2emissions)
log(gdp)	8.899*** (1.838)
I(log(gdp)^2)	−0.439*** (0.103)
urban_pop	0.013** (0.006)
Constant	−44.147*** (8.162)
Observations	169
R ²	0.754
Adjusted R ²	0.750
Residual Std. Error	0.750 (df = 165)

Note : *p<0.1 ; **p<0.05 ; ***p<0.01

Nous pouvons alors conclure qu’il serait pertinent d’utiliser ce dernier modèle pour modéliser les émissions de CO₂ par habitant.

5 Conclusion

In fine, l'analyse de la courbe de Kuznets environnementale est d'autant plus importante que les questions sur l'environnement croissent dans notre société. L'analyse chronologique apportée par Jobert et Karanfil en 2012 liée particulièrement à l'évolution des émissions de CO₂ des 51 pays les plus pollueurs remet en cause l'existence même de la CKE aujourd'hui. À travers notre étude économétrique, nous avons pu faire apparaître le début d'une telle courbe par la modélisation des émissions de CO₂ enregistrées en 2014. Néanmoins, ce début de courbe ne confirme pas l'existence d'une CKE. En effet, le point de retournement n'est atteint par aucun des pays de notre jeu de données. Ainsi, la relation pourrait, au fur et à mesure que les pays s'enrichissent, dévier vers une relation monotone plutôt que quadratique.

Nous avons ensuite décidé d'enrichir l'équation classique de la CKE de variables additionnelles, pour finalement conclure de la pertinence du modèle suivant :

$$\log(CO2emissions) = \beta_0 + \beta_1 \log(gdp)_i + \beta_2 \log(gdp)_i^2 + \beta_3 urban_pop_i + \epsilon_i$$

Les émissions de CO₂ d'un pays seraient liées à la fois à son PIB par habitant et à son taux d'urbanisation. Il n'y aurait également pas de différence d'émissions entre les pays de l'OCDE et les pays hors OCDE, ni de particularité significative suggérant une régression différenciée pour les individus atypiques.

Une dernière approche incluant des variables instrumentales s'est avérée plus pertinente que la méthode de Moindres Carrés Ordinaires . L'endogénéité observée du PIB par habitant peut être expliquée par la densité de la population et l'espérance de vie. Ces derniers indicateurs ont ainsi été choisis comme variables instrumentales afin d'améliorer la régression.

6 Bibliographie

- Formation et déformation de la courbe de Kuznets environnementale pour les émissions de CO₂, T. JOBERT et F. KARANFIL, 2012
- Cours d'économétrie, V. THELEN, 2020
- Life Expectancy and Economic Growth : The Role of the Demographic Transition, M. CERVELLATI, 2009 : <http://ftp.iza.org/dp4160.pdf>
- Population density and Economic Growth, R. K. OWUSUS, 2012 : <https://core.ac.uk/download/pdf/12521417.pdf>

Les données sont majoritairement issues de la Banque Mondiale. Voici quelques informations supplémentaires en ce concernant les sources des données :

- *Jeu de données principal provenant de WDI (sources : Banque Mondiale, OCDE, Carbon Dioxide Information Analysis Center (USA), Division de la population des Nations Unies)*
- *Pays membres de l'OCDE Source Banque Mondiale*
- *Meat Production, OurWorldInData (source : FAO) (last update 2019) : <https://ourworldindata.org/meat-production>*
- *Life expectancy, WDI (source : Division de la population des Nations Unies) : <https://data.worldbank.org/indicator/EN.POP.DNST> (last update 2019)*
- *Density Population, WDI (sources : FAO, Banque Mondiale) : <https://data.worldbank.org/indicator/SP.DYN.LE00.IN?locations=1W> (last update 2019)*
- *Population by country, DataHub (source : Nations Unies) : <https://datahub.io/JohnSnowLabs/population-figures-by-country#resource-population-figures-by-country-csv>*