

# Sensibilité environnementale et dépenses pour l'environnement

Quentin Merrien et Catherine Berleur

13/04/2021

## Contents

I. Mise en perspective introductive dans le contexte économique et social. . . . .	1
II. Formulation d'un problème économétrique. . . . .	1
III. Présentation des bases de données et exploration des données. . . . .	2
1. Analyse en composantes principales (ACP). . . . .	3
2. Analyse univariée. . . . .	3
3. Analyse bivariée. . . . .	4
IV. Proposition et justification de différents modèles linéaires. . . . .	4
V. Interprétation des modèles retenus et proposition de pistes de réflexion complémentaires (limites blabla). . . . .	9
VII. Conclusion. . . . .	15

## I. Mise en perspective introductive dans le contexte économique et social.

Le 9 mai dernier, des marches pour le climat eurent lieu dans toute la France, afin d'infléchir la loi « Climat et résilience » qui vient d'être votée par les députés. Ces mobilisations en faveur du climat ont-elles une influence sur la politique écologique et environnementale des gouvernements ?

Notre sujet de recherche vise à contribuer à cette vaste question, en se focalisant sur le lien entre ce que l'on a appelé sensibilité environnementale et les dépenses publiques de protection de l'environnement. Les difficultés de mesure statistique concernant les mobilisations environnementales nous ont en effet poussé à choisir comme proxy la sensibilité environnementale.

Cette contribution s'inscrit dans lignée de la sociologie de l'action publique. En particulier, l'article fondateur de Felstiner & alli (1981) sur l'émergence des problèmes publics peut servir de cadre analytique. Les auteurs y définissent trois conditions d'apparition d'un problème public : *naming*, *blaming*, *claiming*. L'étape du *naming* conduit à une transformation de l'ordre symbolique, et impose une nouvelle manière de voir les choses : la sensibilité publique envers certaines questions participe à l'établissement de mise à l'agenda de problèmes publics. Ainsi, une sensibilité environnementale plus importante conduirait à un cadrage différent des problèmes environnementaux, et serait nécessaire aux autres conditions de l'émergence de l'action environnementale et écologique comme problème public.

## II. Formulation d'un problème économétrique.

Le premier enjeu de notre démarche économétrique se situe dans le choix des variables expliquées et explicatives. Pour étudier l'effet de la sensibilité environnementale sur les dépenses de protection de l'environnement, la variable dépendante retenue est la *NEEP* (dépenses nationales pour la protection de l'environnement) dans l'UE et ses États membres. Cette mesure, définie par Eurostat, permet d'apprécier les ressources que consacrent les unités résidentes à la protection de l'environnement naturel au cours d'une période donnée. Elles sont calculées comme étant la somme des dépenses courantes consacrées aux activités de protection de l'environnement et des investissements destinés aux activités de protection de l'environnement, y compris les transferts nets vers le reste du monde.

Il s'agit ensuite de définir ce qu'est la sensibilité environnementale afin de construire un jeu de variables à même de la refléter. Dès lors, on retiendra l'acception suivante : la sensibilité environnementale correspond à l'importance attachée aux problèmes environnementaux et aux enjeux écologiques.

Elle comprend donc la perception de la gravité des problèmes environnementaux, le degré de connaissance des problèmes environnementaux, et le soutien aux dépenses gouvernementales pour la protection de l'environnement [Thiery-Seror, 1996]<sup>1</sup> La difficulté réside alors en la traduction de ces éléments en une variable économétrique. Pour construire cette première variable explicative, nous avons croisé plusieurs bases de données : l'Environmental Awareness Index (EAI)<sup>2</sup> et l'European Values Study (EVS) de 2017. L'*Environmental Awareness Index* est en effet une variable permettant d'apprécier la sensibilisation des citoyens aux questions écologiques, que l'on peut assimiler au degré de connaissance des problèmes environnementaux, et que l'on appellera **conn\_env**. L'EVS permet elle de couvrir les deux autres dimensions de la sensibilité environnementale. La variable **perc\_env** rend compte de la perception des problèmes environnementaux des personnes sondées, tandis que la variable **soutien\_pol\_env** témoigne du soutien aux mesures gouvernementales pour la défense de l'environnement. Afin de construire une variable explicative agrégée pour la sensibilité environnementale (*sensi\_env*), on pondère les variables des ces trois facteurs par 1/3.

Les autres variables explicatives, que l'on pourra appeler variables de contrôle, sont au nombre de 4. Il paraît raisonnable de penser que les dépenses environnementales d'un pays sont liées à son niveau de développement économique, conformément à la thèse avancée par R. Inglehart dans *La Révolution silencieuse* (1977), qui considère qu'à partir d'un certain niveau de développement économique, des valeurs post-matérialistes (et notamment écologiques) ont plus de chance de se diffuser au sein de la société. Pour exprimer l'influence du niveau de développement économique d'un pays sur sa NEEP, nous avons retenu le revenu médian (*rev\_med*) établi par Eurostat dans l'enquête EU-SILC. Il est aussi possible de considérer que les inégalités intra-pays influencent ces dépenses. D'une part, une inégalité, lorsqu'elle est rendue publique, réduit la contribution à un bien public de tous les participants [Anderson et al., 2008]. Puisqu'il est possible de considérer l'environnement comme un bien public (non-rival et non-exclusif), il apparaît pertinent de prendre en compte les inégalités dans notre modèle économétrique. De plus, les sociétés inégalitaires auraient plus de mal à mener les réformes nécessaires pour surmonter un choc externe [Rodrik, 1999]. Nous avons donc retenu le coefficient de Gini du revenu en euros par ménage (indicateur synthétique de mesure des inégalités et de la redistribution très utilisé au niveau international) comme variable explicative. Les deux dernières variables de contrôle sont l'Environmental Performance Index (*EPI*) qui permet d'évaluer de manière synthétique la performance environnementale d'un pays ainsi que le pourcentage de personnes de 15 à 64 ans disposant d'un diplôme d'enseignement supérieur pour rendre compte du niveau d'éducation.

## III. Présentation des bases de données et exploration des données.

Nous avons ainsi créé une base de données à partir de plusieurs sources différentes. Le processus complet de création de la base est disponible dans le fichier .R "création base de données".

<sup>1</sup>Thiery-Seror, P. (1996). La solidarité à travers les produits: l'achat-geste écologique. Actes du XIIème Congrès de l'AFM, Poitiers, 22-23 mai.

<sup>2</sup><http://jultika.oulu.fi/files/nbnfioulu-201312142043.pdf>

Pour l'indicateur de sensibilité environnementale, les données viennent de l'*European Values Survey*, un sondage mené à l'échelle européenne, pour avoir un état des lieux des valeurs des européens. Nous avons extrait de ces données 6 variables sur 56491 observations, et que nous avons résumé en moyenne par pays. Nous avons pris pour la variable **perc\_env** la moyenne des indicateurs associés aux questions suivantes :

- Je donnerais une partie de mes revenus si j'étais sûr que l'argent soit utilisé pour éviter la pollution de l'environnement.
- C'est juste trop difficile pour les gens comme moi de vraiment pour l'environnement.
- Il y a plus important à faire dans la vie que de protéger l'environnement.
- Ça ne sert à rien de faire ce que je peux pour l'environnement si les autres ne font pas la même chose.
- Beaucoup des affirmations sur les menaces environnementales sont exagérées.

La première des questions allant dans un sens opposé aux suivantes, nous avons inversé l'indicateur pour celle-ci par rapport aux autres.

Pour la variable **soutien\_pol\_env**, nous avons pris la moyenne par pays de la réponse :

“On devrait donner la priorité à la protection de l'environnement, même si cela ralentit la croissance économique et si certains perdent leur emploi”,

donnée plutôt que :

“On devrait donner la priorité à la croissance économique et à la création d'emplois, même si l'environnement en souffre d'une manière ou d'une autre”.

Pour la variable **conn\_env**, nous avons pris l'indicateur de conscience environnementale développé par Eevi Kokkinen dans son mémoire de master, “Measuring environmental awareness in the world”. Celui-ci est construit à partir de questionnaires, afin d'estimer la connaissance et les compétences en matière environnementale et écologique des personnes sondées.

Les deux potentielles variables dépendantes (**part\_depenses\_env** et **depenses\_brutes\_env**) proviennent de bases de données établies par Eurostat. Les variables de contrôle **EDUC\_SUP\_2017** et **REV\_MED\_EURO\_2017** proviennent elles aussi d'Eurostat, et sont issues de la base de données EU-SILC (European Union Statistics and Living Conditions).

L'exploration des données préalablement à la réalisation de régressions linéaires est primordiale. Pour ce faire, nous mobilisons à la fois des éléments de statistiques descriptives et l'analyse en composantes principales.

## 1. Analyse en composantes principales (ACP).

Afin d'observer les relations entre variables de manière synthétique, on peut réaliser une ACP, dont les résultats sont disponibles en **annexe 7**.

Le premier graphique représente la contribution des différents axes à la variance totale, et on voit que le premier axe contribue à la moitié de la variance totale, soulignant qu'il existe des facteurs à l'oeuvre rendant le nuage de points non aléatoire. Les deuxièmes et troisièmes axes ont une valeur propre supérieure à 1, sachant que l'analyse est normée. En comparant les cercles de variables pour les espaces des axes 1 et 2, et 1 et 3, le cercle des axes 1 et 3 semble plus intéressant, puisque trois directions se distinguent clairement :

- La première correspond à l'indice de gini. Celui-ci est dans la direction opposée à beaucoup de variable différentes, en particulier le niveau d'éducation supérieure, et les variables relatives à la sensibilité environnementale.
- La deuxième correspond au montant des dépenses pour l'environnement, et correspond à la principale variable participant à la création de l'axe 3.

- La troisième rassemble toutes les variables restantes, c'est-à-dire la sensibilité environnementale (et ses composantes), l'indice de performance environnementale, le revenu médian, le niveau d'éducation supérieure, et les dépenses environnementales en points de PIB. Ainsi, on devrait observer une corrélation positive entre toutes ces variables, ainsi qu'une corrélation négative de toutes ces variables avec l'indice de gini.

Le graphique des individus représente les pays sur l'espace des variables, ainsi que la sensibilité environnementale et le pourcentage du PIB dédié aux dépenses environnementales. En regardant les pays les plus à l'extrême, on voit que la Suède est située la plus à droite, et a l'indice de sensibilité environnementale le plus élevé, tandis que les pays les plus à droite ont un cercle plus petit, c'est-à-dire des dépenses environnementales en points de PIB plus faibles, comme la Roumanie.

Ainsi, plus de la moitié des variables sont corrélées positivement entre elles, et s'opposent à l'indice gini, tandis que le montant des dépenses environnementales se distingue, et que la part des dépenses environnementales dans le PIB et le revenu médian sont corrélées positivement, mais faiblement.

## 2. Analyse univariée.

Nous procédons à une première analyse des variables présentes dans notre jeu de données. Pour ce faire, il s'agit de dresser un tableau des principaux indicateurs statistiques (**Annexe 1**).

La visualisation des données grâce aux graphiques facilite grandement l'analyse. Les **annexes 2 et 3** retranscrivent les disparités importantes entre les pays européens en terme de dépenses pour la protection de l'environnement brutes, ces disparités semblant moins importantes une fois la NEEP rapportée en points de PIB.

## 3. Analyse bivariée.

Il s'agit désormais d'étudier les liens entre nos différentes variables explicatives. Dans un premier temps, on s'intéresse aux trois variables qui ont permis de construire notre indicateur de sensibilité environnementale **sensi\_env** : **conn\_env**, **perc\_env** et **soutien\_pol\_env**. Graphiquement, on visualise la relation de dépendance entre ces trois variables, deux-à-deux (**Annexe 4**). On observe ainsi une corrélation positive entre ces différentes variables, plus ou moins prononcée. En effet, les corrélations linéaires (méthode Pearson) entre ces variables sont : - Entre la connaissance et la perception des problèmes environnementaux : 0.1135332; - Entre la connaissance des problèmes environnementaux et le soutien aux politiques de protection de l'environnement : 0.6305951; - Entre la perception des problèmes environnementaux et le soutien aux mesures de protection de l'environnement : 0.4566094.

On s'intéresse ensuite aux corrélations entre toutes les variables. Pour ce faire, nous dressons la matrice des corrélations entre les différentes variables (**Annexe 5**). Une représentation graphique de ces coefficients facilite leur interprétation (**Annexe 6**).

## IV. Proposition et justification de différents modèles linéaires.

Le faible nombre d'observations a limité nos possibilités d'analyse, mais nous avons tenté de réaliser des modèles avec un minimum de significativité, dans une démarche d'apprentissage, plutôt que d'administration de la preuve.

Tout d'abord, nous avons régressé nos variables dépendantes (part du PIB dédié aux dépenses environnementales, et montant des dépenses environnementales) selon nos variables explicatives centrales (l'indice de sensibilité environnementale, ainsi que ses composantes). Nous comparons pour cela plusieurs régressions linéaires :

- reg1, dont la formule est simplement " $\text{part\_depenses\_env} \sim \text{sensi\_env}$ ",

- reg1bis, dont la formule est “part\_depenses\_env ~ conn\_env + perc\_env + soutien\_pol\_env”, permettant d’identifier le rôle différent des trois composantes de l’indice construit,
- reg1ter, dont la formule est “depenses\_brutes\_env ~ sensi\_env”, pour voir quelle mesure des dépenses environnementales semble la plus appropriée,
- reg1quater, dont la formule est “depenses\_brutes\_env ~ conn\_env + perc\_env + soutien\_pol\_env”.

Table 1: Modèles simples sans variables de contrôle.

	<i>Dependent variable:</i>			
	part_depenses_env		depenses_brutes_env	
	reg1	reg1bis	reg1ter	reg1quater
	(1)	(2)	(3)	(4)
sensi_env	0.018 (0.012)		306.553 (474.735)	
conn_env		0.034*** (0.012)		295.435 (543.034)
perc_env		0.003 (0.010)		57.415 (486.889)
soutien_pol_env		−0.012 (0.011)		−8.982 (527.144)
Constant	0.666 (0.797)	0.454 (0.786)	−7,957.756 (31,156.420)	−8,653.055 (36,564.380)
Observations	20	20	20	20
R <sup>2</sup>	0.114	0.381	0.023	0.033
Adjusted R <sup>2</sup>	0.065	0.266	−0.032	−0.149
Residual Std. Error	0.440 (df = 18)	0.390 (df = 16)	17,212.820 (df = 18)	18,162.970 (df = 16)
F Statistic	2.317 (df = 1; 18)	3.289** (df = 3; 16)	0.417 (df = 1; 18)	0.180 (df = 3; 16)

Note:

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

On observe ainsi dans le tableau résumant ces quatre modèles que seul le modèle (2) (reg1bis) possède un tant soit peu de significativité. D’une part, c’est le seul à posséder une significativité totale correcte, puisque le  $R^2$  correspondant est égal à 0.381, c’est-à-dire que les variables expliquent 38.1% de la variance totale de la variable dépendante, ici, la part du PIB des dépenses environnementales. Les autres modèles ont un  $R^2$  au mieux trois fois plus faible, une différence encore plus flagrante si l’on considère le  $R^2$  ajusté, qui prend en compte le nombre de variables, permettant donc de comparer les modèles alors qu’ils ont un nombre différent de variables.

En approfondissant l’analyse des résultats obtenus, on observe que seule une variable au sein du modèle (2) est significative : **conn\_env**, avec une  $p$ -value inférieure à 0.01. On se propose donc de se focaliser en partie sur cette variable, en l’isolant dans des modèles différents :

- reg2, dont la formule est “part\_depenses\_env ~ conn\_env”,
- reg2bis, dont la formule est “depenses\_brutes\_env ~ conn\_env”.

Pour le modèle reg2, la significativité globale augmente (le  $R^2$  ajusté est plus élevé), ce qui est logique puisque nous venons de ponctionner le modèle de variables non significatives (perc\_env et soutien\_pol\_env). On

Table 2: Modèles sans variables de contrôle, avec une variable repérée comme significative.

	<i>Dependent variable:</i>	
	part_depenses_env reg2	depenses_brutes_env reg2bis
	(1)	(2)
conn_env	0.026*** (0.009)	295.261 (384.663)
Constant	0.373 (0.505)	-5,141.841 (22,667.330)
Observations	20	20
R <sup>2</sup>	0.334	0.032
Adjusted R <sup>2</sup>	0.297	-0.022
Residual Std. Error (df = 18)	0.382	17,132.900
F Statistic (df = 1; 18)	9.033***	0.589

Note:

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

peut donc partir de ce modèle simple et ajouter des variables de contrôle. Etant donné le faible nombre d'observations, et la valeur de l'exercice, nous proposons aussi un certain nombre de modèles reposant sur **sensi\_env**, afin d'analyser les modèles linéaires obtenus. N'ayant pas à l'esprit de variable de contrôle ayant *a priori* une influence plus grande que les autres, nous décidons de tester les modèles suivants, intégrant toutes les variables de contrôle :

- reg3, avec la formule “part\_depenses\_env = sensi\_env + EPI.new + REV\_MED\_EURO\_2017 + EDUC\_SUP\_2017 + GINI\_2017”,
- reg4, avec la formule “part\_depenses\_env ~ conn\_env + EPI.new + REV\_MED\_EURO\_2017 + EDUC\_SUP\_2017 + GINI\_2017”,
- reg3bis, avec la formule “depenses\_brutes\_env ~ sensi\_env + EPI.new + REV\_MED\_EURO\_2017 + EDUC\_SUP\_2017 + GINI\_2017”,
- reg4bis, avec la formule “depenses\_brutes\_env ~ conn\_env + EPI.new + REV\_MED\_EURO\_2017 + EDUC\_SUP\_2017 + GINI\_2017”.

Ce tableau nous montre que ces modèles sont assez peu significatifs : le  $R^2$  ajusté max est égal à 0.15, et la seule variable significative individuellement est l'indice de performance individuelle, lorsque régressé avec le montant des dépenses pour l'environnement. On observe que le revenu médian, variable proxy utilisée pour mesurer le niveau de vie des pays, n'est pas du tout significatif, on peut donc abandonner cette variable, et refaire des modèles sans :

- reg5, avec la formule “part\_depenses\_env = sensi\_env + EPI.new + EDUC\_SUP\_2017 + GINI\_2017”,
- reg6, avec la formule “part\_depenses\_env ~ conn\_env + EPI.new + EDUC\_SUP\_2017 + GINI\_2017”,
- reg5bis, avec la formule “depenses\_brutes\_env ~ sensi\_env + EPI.new + EDUC\_SUP\_2017 + GINI\_2017”,
- reg6bis, avec la formule “depenses\_brutes\_env ~ conn\_env + EPI.new + EDUC\_SUP\_2017 + GINI\_2017”.

Ainsi, le modèle reg5bis semble intéressant, et à retravailler. En explorant un peu plus les modèles, on trouve que le modèle suivant ont un  $R^2$  ajusté correct, c'est-à-dire au-dessus de 0.2 :

Table 3: Modèles avec variables de contrôle, pour sensi.env et conn.env.

	<i>Dependent variable:</i>			
	part_depenses_env		depenses_brutes_env	
	reg3	reg4	reg3bis	reg4bis
	(1)	(2)	(3)	(4)
sensi_env	−0.006 (0.019)		−336.305 (692.576)	
conn_env		0.022 (0.018)		108.729 (703.563)
EPI.new	0.024 (0.030)	0.010 (0.030)	1,969.223* (1,102.069)	1,802.809 (1,150.375)
REV_MED_EURO_2017	−0.00000 (0.00002)	−0.00001 (0.00002)	−0.161 (0.750)	−0.253 (0.736)
EDUC_SUP_2017	0.002 (0.019)	−0.005 (0.019)	−891.047 (690.563)	−897.014 (728.499)
GINI_2017	−0.044 (0.030)	−0.020 (0.032)	798.316 (1,114.194)	1,096.733 (1,230.626)
Constant	1.890 (2.678)	0.701 (2.426)	−104,242.900 (98,231.170)	−128,027.900 (94,125.220)
Observations	20	20	20	20
R <sup>2</sup>	0.313	0.377	0.332	0.322
Adjusted R <sup>2</sup>	0.068	0.154	0.094	0.080
Residual Std. Error (df = 14)	0.440	0.419	16,129.970	16,251.380
F Statistic (df = 5; 14)	1.276	1.693	1.395	1.332

*Note:*

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

Table 4: Modèles avec variables de contrôle, sans educ.sup.

	<i>Dependent variable:</i>			
	part_depenses_env		depenses_brutes_env	
	reg5	reg6	reg5bis	reg6bis
	(1)	(2)	(3)	(4)
sensi_env	−0.007 (0.018)		−522.244 (606.115)	
conn_env		0.022 (0.018)		93.865 (681.265)
EPI.new	0.019 (0.020)	0.002 (0.019)	1,726.900** (705.111)	1,504.822* (732.184)
EDUC_SUP_2017	−0.0002 (0.017)	−0.008 (0.017)	−972.413 (595.914)	−990.684 (655.283)
GINI_2017	−0.045 (0.029)	−0.022 (0.030)		1,020.994 (1,174.538)
Constant	2.332 (1.748)	1.377 (1.657)	−50,304.560 (33,727.920)	−105,054.700 (64,202.110)
Observations	20	20	20	20
R <sup>2</sup>	0.311	0.370	0.308	0.317
Adjusted R <sup>2</sup>	0.127	0.202	0.178	0.134
Residual Std. Error	0.426 (df = 15)	0.407 (df = 15)	15,363.330 (df = 16)	15,766.240 (df = 15)
F Statistic	1.689 (df = 4; 15)	2.202 (df = 4; 15)	2.373 (df = 3; 16)	1.738 (df = 4; 15)

*Note:*

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01



- reg7, avec la formule “part\_depenses\_env ~ conn\_env + soutien\_pol\_env + GINI\_2017 + EPI.new”, avec un  $R^2$  ajusté de 0.25,

On peut donc retenir pour une analyse approfondie les modèles reg2, reg5bis, reg7, et reg7bis :

- reg2, dont la formule est “part\_depenses\_env ~ conn\_env”,
- reg5bis, avec la formule “depenses\_brutes\_env ~ sensi\_env + EPI.new + EDUC\_SUP\_2017 + GINI\_2017”,
- reg7, avec la formule “part\_depenses\_env ~ conn\_env + soutien\_pol\_env + GINI\_2017 + EPI.new”,

## V. Interprétation des modèles retenus et proposition de pistes de réflexion complémentaires (limites blabla).

On peut donc reprendre un tableau récapitulatif des modèles retenus.

Table 5: Modèles retenus.

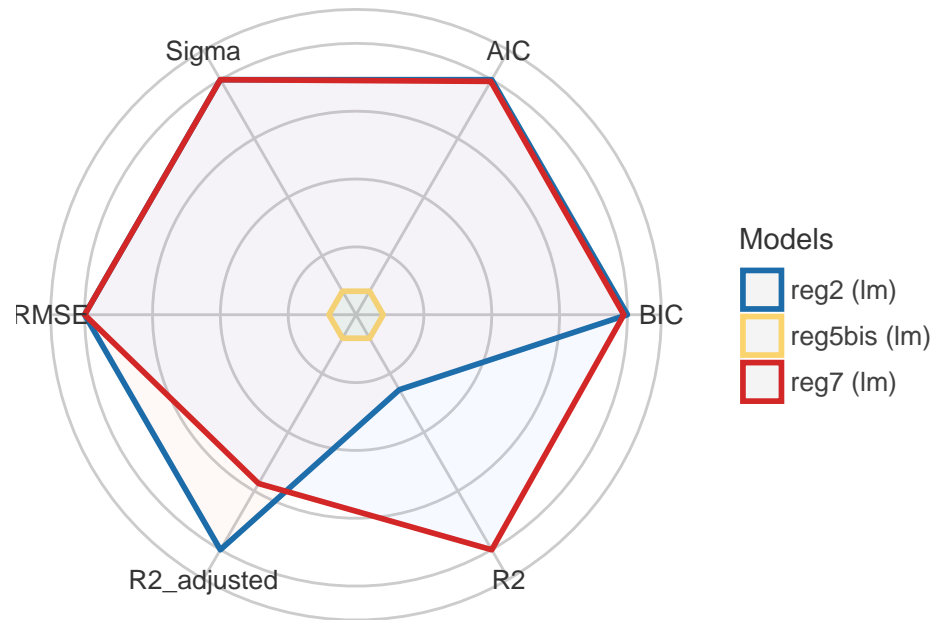
	<i>Dependent variable:</i>		
	part_depenses_env		depenses_brutes_env
	reg2	reg7	reg5bis
	(1)	(2)	(3)
conn_env	0.026*** (0.009)	0.024 (0.016)	
soutien_pol_env		-0.012 (0.010)	
sensi_env			-522.244 (606.115)
EPI.new		0.006 (0.019)	1,726.900** (705.111)
GINI_2017		-0.028 (0.029)	
EDUC_SUP_2017			-972.413 (595.914)
Constant	0.373 (0.505)	1.714 (1.595)	-50,304.560 (33,727.920)
Observations	20	20	20
R <sup>2</sup>	0.334	0.416	0.308
Adjusted R <sup>2</sup>	0.297	0.260	0.178
Residual Std. Error	0.382 (df = 18)	0.392 (df = 15)	15,363.330 (df = 16)
F Statistic	9.033*** (df = 1; 18)	2.668* (df = 4; 15)	2.373 (df = 3; 16)

*Note:*

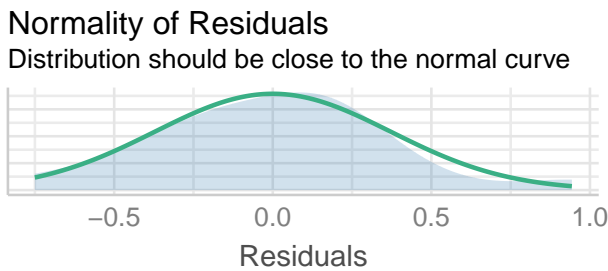
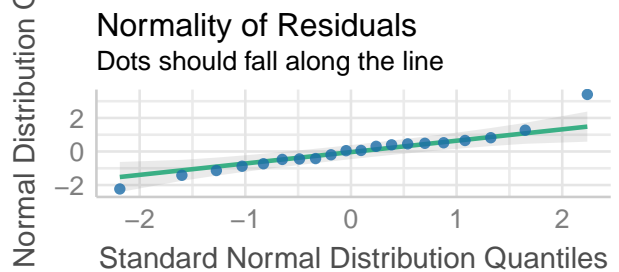
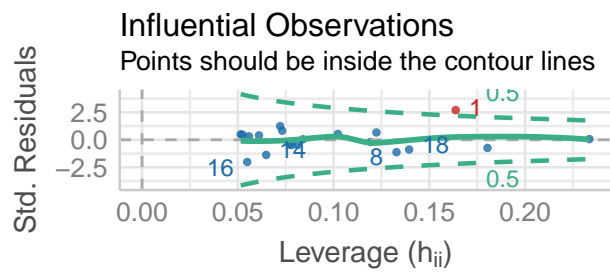
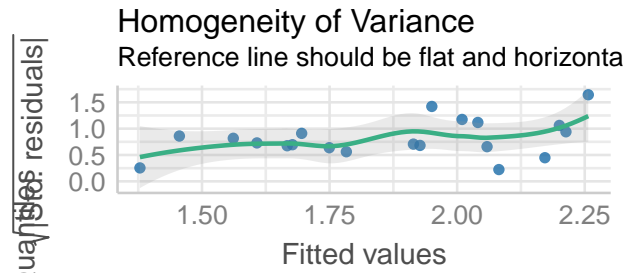
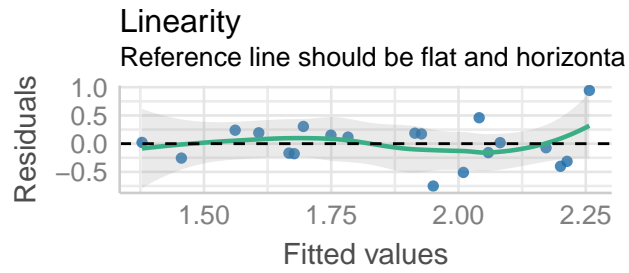
\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Ainsi, on observe que les différents modèles ont un  $R^2$  ajusté assez bon, il est plus de 0.2 pour 3 des modèles, et que 2 des tests ont une variable significative à 1 ou 5%. On peut comparer différentes statistiques de significativité dans un graphique. On observe ainsi que je n'y connais pas grand chose à ces indicateurs.

## Comparison of Model Indices



Avant d'interpréter les coefficients, il convient de vérifier les hypothèses de Gauss-Markov. Pour cela, une représentation graphique s'avère utile.



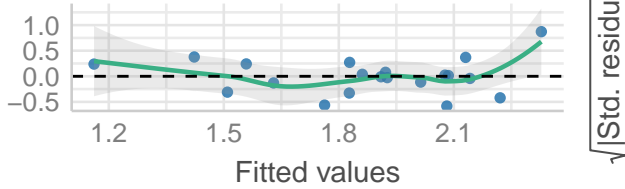
- Normal Distribution Quantiles

Quantile Inflation Factor (VIF)

Residuals

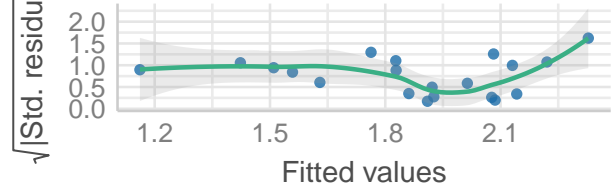
### Linearity

Reference line should be flat and horizontal



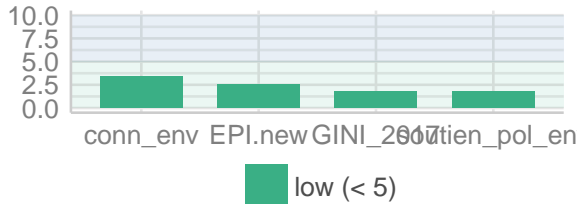
### Homogeneity of Variance

Reference line should be flat and horizontal



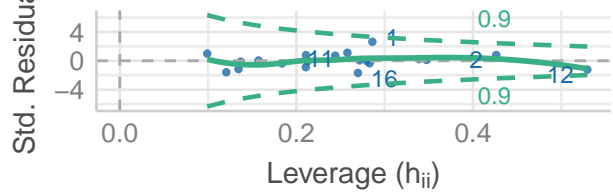
### Collinearity

Higher bars (>5) indicate potential collinear



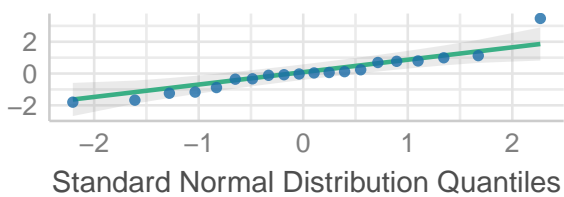
### Influential Observations

Points should be inside the contour lines



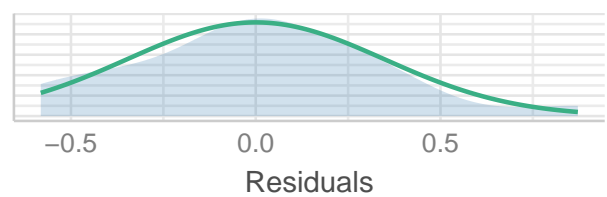
### Normality of Residuals

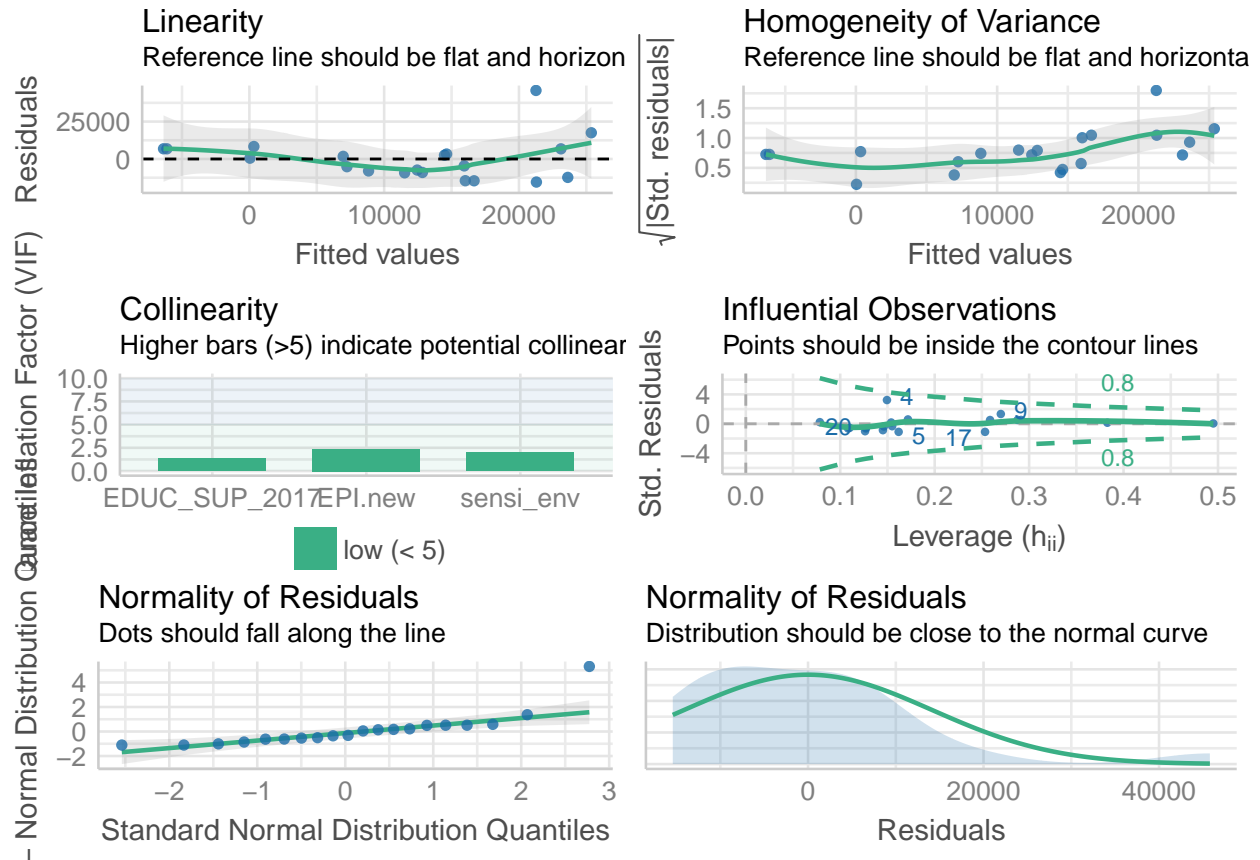
Dots should fall along the line



### Normality of Residuals

Distribution should be close to the normal curve





Ainsi, on voit que les deux modèles linéaires multiples ont des variables explicatives non parfaitement colinéaires, et que les modèles respectent les hypothèses d'homoscédasticité et de normalité des résidus. De plus, on observe que les modèles ne sont pas déformés par des observations influentes.

On peut donc passer à l'interprétation des coefficients des différents modèles, en analysant à la fois le signe et la valeur des coefficients.

Dans le premier modèle, le modèle linéaire simple, le coefficient associé à **conn\_env** est significatif, positif, et égal à 0.026. Ainsi, lorsque l'indice de connaissance environnementale augmente de 10 points (sur un maximum théorique de 100 points), le pourcentage du PIB dédié aux dépenses environnementales augmente de 0.26 point de pourcentage. D'un point de vue théorique, il paraît logique que l'indice de connaissance environnementale fasse augmenter les dépenses d'environnement : il s'agit d'une relation similaire à celle que nous voulions montrer. Cependant, il faut rester prudent sur l'analyse de cette corrélation : une part du PIB plus élevée dédiée aux dépenses d'environnement peut conduire à une plus grande connaissance environnementale de la population, en finançant des infrastructures ou programmes dédiées à la découverte de l'environnement. Il faudrait donc disposer d'outils statistiques et économétriques plus précis pour tester si cette corrélation est une causalité, et dans quel sens elle agit, peut-être à partir d'une variable instrumentale, causée par la variable explicative, et non causée par la variable expliquée. De plus, la constante est égale à 0.373, mais n'est pas significative. On peut toutefois penser que son signe est fiable : y compris avec un indice de connaissance environnementale égal à 0, on aurait 0.37 point de PIB dédié aux dépenses environnementales.

Dans le deuxième modèle, le modèle linéaire multiple vise à expliquer la part du PIB des dépenses environnementales à partir des indices de connaissance environnementale, de soutien aux politiques environnementales, de performance individuelle, et de gini. Aucun coefficient n'est significatif, et la plupart sont assez faibles, même si **conn\_env** est associé à un coefficient proche de celui dans le premier modèle. Le deuxième coefficient, associé à **soutien\_pol\_env**, n'est pas significatif, mais il est de signe négatif, ce qui peut être interprété comme le signe qu'il s'agit d'une variable redondante par rapport à **conn\_env**. Ce

signe négatif pourrait aussi venir du fait que dans les pays où une part plus importante du PIB est allouée aux dépenses pour l'environnement, les populations connaissent les conséquences, peut-être négatives, sur leur vie quotidienne de dépenses pour l'environnement. L'indice de performance environnementale possède un coefficient proche de 0 pour ce modèle, ce qui tendrait à montrer que cet indice, construit à 40% à partir de la santé de l'environnement (qualité de l'air, qualité de l'eau, risques sanitaires) et à 60% à partir de la vitalité des écosystèmes (ressources en eau, forêt, poissons, biodiversité, etc. ), n'a aucune influence sur la part du PIB dédiée aux dépenses pour l'environnement. Le coefficient associé à l'indice de Gini est lui négatif, et non nul, même s'il n'est pas statistiquement significatif. Ainsi, une augmentation des inégalités réduirait la part du PIB dédiée à l'environnement, ce qui va dans le sens de notre analyse *a priori*, décrite en partie 2. La constante associée à ce modèle linéaire est plus élevée que dans le premier modèle, ce qui vient potentiellement de l'introduction de deux variables avec un coefficient négatif.

Dans le dernier modèle, le modèle linéaire teste une relation entre le montant des dépenses pour l'environnement et la somme de notre indice composit de sensibilité environnementale, de l'indice de performance environnementale, et du niveau d'éducation supérieure. La constante et les coefficients sont négatifs, à l'exception du seul coefficient statistiquement significatif : l'indice de performance environnementale. La comparaison avec le modèle précédent n'est pas aisée, compte tenu de variables dépendante, explicative et de contrôle différentes, mais il est toutefois frappant de voir que cet indice devient significatif alors qu'il était presque nul dans l'autre modèle.

On peut souligner qu'il est plus logique que le degré de performance environnementale, une mesure d'états de faits physiques, soit corrélée avec le montant des dépenses pour l'environnement, qui quantifie l'investissement matériel réalisé en faveur de l'environnement. Au contraire, la part du PIB dédié à l'environnement est un indicateur relatif, qui mesure plutôt l'importance accordée à l'environnement, et donc le pourcentage du total des ressources monétaires qu'on lui alloue, et il semble donc plus plausible de le relier à des indicateurs de mesure des préférences et points de vue subjectifs des populations.

Le signe négatif des coefficients associés aux variables **sensi\_env** et **educ\_sup** est cependant plus difficile à interpréter, puisque cela va à l'encontre d'hypothèses de travail *a priori* : on pourrait penser que la sensibilité à l'environnement et le niveau d'éducation supérieure conduise à un niveau plus élevé de dépenses pour l'environnement. Toutefois, on peut penser que ces variables sont plus négativement corrélées avec le montant du PIB total, qu'avec le montant des dépenses pour l'environnement. En effet, au vu de la variation du montant des dépenses comparée à celle des dépenses en points de PIB, on peut penser que le montant des dépenses pour l'environnement est positivement corrélé avec le montant du PIB, qui est lui-même négativement corrélé avec l'indice de sensibilité environnementale. Mais c'est totalement incongru et difficile à expliquer que je ne sais pas pourquoi j'ai dit ça.

Il faut ensuite aborder les pistes de réflexion complémentaires.

Tout d'abord, la distinction entre variable de mesure physique, et variable de mesure psychologique d'importance, semble bien s'appliquer à celle entre respectivement le montant des dépenses pour l'environnement, et la part du PIB qui leur est dédiée.

De plus, il semble que les indicateurs permettant de discriminer les pays selon leur niveau de richesse/développement (ici, le revenu médian, et le niveau d'éducation supérieure) ne semblent pas jouer leur rôle, et devraient peut-être être remplacés par un autre indicateur, comme le PIB. Nous avons décidé de pas intégrer le PIB dans nos analyses pour éviter une endogénéité (est-ce bien ça le terme pour parler de l'influence de la variable expliquée **sur** la variable explicative) de la valeur expliquée, mais il pourrait être intéressant de néanmoins en tester le rôle.

La construction de l'indicateur de sensibilité environnementale gagnerait à être approfondie et précisée, afin de mieux comprendre le rôle des différentes composantes dans sa construction.

Toutefois, la limite principale de nos analyses est le nombre trop faible d'observations, qui vient du fait que les données proviennent de bases de données différentes ne comptant pas toutes exactement les mêmes pays : nous avons des données pour toutes les variables sélectionnées concernant uniquement 20 pays, notamment en raison de la difficulté à obtenir des variables concernant les dépenses dédiées à l'environnement (disponibles sur eurostat.eu, donc pour les pays européens uniquement). De plus, la fragmentation territoriale des pays européens, notamment d'Europe de l'Est, qui provient des guerres successives (guerres mondiales, guerre de Yougoslavie) contribue à un décalage entre des pays de taille relativement importante et des pays de taille

très réduite, ce qui fausse la comparaison des variables absolues, mais aussi relatives, puisque pour une taille plus faible, il est plus probable d'obtenir des statistiques plus élevées. Par exemple, le niveau d'éducation supérieure se calcule comme un ratio de la population diplômée sur la population totale, pour une tranche d'âge précise, ce qui souligne le rôle de la taille totale dans la mesure de statistiques relatives.

Enfin, les données issues de l'enquête EVS ne sont pas très solides, puisque le questionnaire EVS ne propose pas beaucoup de questions relatives à l'environnement, et celles-ci sont formulées de manière non neutre, ce qui peut induire un biais dans les questions. A signaler : Parler de deux des problèmes (environnement v. emploi et "m'en balec") L'International Social Survey Programme conduit une enquête par questionnaire parmi les 42 pays membres du programme, sur le thème de l'environnement. Le questionnaire final est disponible ici<sup>3</sup> mais les résultats ne sont pas encore publiés. Au vu des questions posées par le questionnaire, introduire des variables qui en proviennent permettrait de limiter les biais relatifs à la formulation des questions (cf. la question 6 : how concerned are you about environmental issues?) et d'affiner l'analyse.

## VII. Conclusion.

---

<sup>3</sup>[http://w.issp.org/fileadmin/user\\_upload/Module\\_development/Module\\_2020/Final\\_Source\\_Questionnaire/ISSP2020\\_final\\_sourcequestionnaire\\_corrected.pdf](http://w.issp.org/fileadmin/user_upload/Module_development/Module_2020/Final_Source_Questionnaire/ISSP2020_final_sourcequestionnaire_corrected.pdf)