

PROJET

ECONOMETRIE DES DONNEES DE PANEL

Sihame JDID

2024-2025

M1 SE

Dans quelle mesure les conditions climatiques influencent-elles la production économique régionale à travers le monde entre 1970 et 1974 ?

PREAMBULE

Nous déclarons sur l'honneur que ce projet a été écrit de notre main, sans aide extérieure non autorisée, qu'il n'a pas été présenté auparavant pour évaluation et qu'il n'a jamais été publié, dans sa totalité ou en partie.

Toutes parties, groupes de mots ou idées, aussi limités soient-ils, y compris des tableaux, graphiques, cartes etc. qui sont empruntés ou qui font référence à d'autres sources bibliographiques sont présentés comme tels, sans exception aucune.

TABLE DES MATIERES

Préambule.....	1
Introduction	1
I. Cadre Statistique	2
A. Statistiques descriptives	2
B. Estimations et tests	2
II. Commentaires	4
Conclusion.....	5
Bibliographie.....	6
Annexes.....	7

La question des conséquences du changement climatique sur l'activité économique fait de plus en plus débat au sein des économistes compte tenu de l'impact environnemental que celle-ci engendre. La littérature empirique met en lumière une relation significative entre la productivité économique d'une part et les conditions climatiques d'autre part. L'article de **Kalkuhl et Wenz (2020)** intitulé *The Impact of Climate on Economic Growth: Evidence from a Global Panel of Regions* et publiée dans le *Journal of Environmental Economics and Management*, analyse l'effet des températures et des précipitations sur la croissance économique grâce à un ensemble de données de panel, couvrant plus de 1500 régions dans 77 pays différents dans le monde et remonte jusqu'au début des années 1900. Les auteurs montrent que les chocs climatiques n'affectent pas les économies régionales de la même manière à travers le monde selon leur capacité d'adaptation à ces chocs et à leur niveau de développement.

Dans le prolongement de cette étude, ce projet vise à nous concentrer sur un échantillon restreint de la base initiale, mais homogène puisque les données ont été sélectionnées sur une même période de 5 ans pour toutes les régions que nous avons retenu. De cette manière, il sera possible de comparer plus facilement les données sans avoir à garder des données Non Applicables (NA). De ce fait, nous chercherons à quantifier l'impact de variables climatiques comme la température ou encore les précipitations, sur la production économique régionale.

Nous pouvons nous demander alors, *dans quelle mesure les conditions climatiques influencent-elles la production économique régionale sur une période donnée ?* Pour répondre à cette question, nous utiliserons donc un modèle économétrique en panel permettant d'analyser simultanément les évolutions temporelles et régionales que la variation du climat a sur la production économique. Comme évoqués précédemment, nous utiliserons un panel équilibré de 280 régions (issues de 12 pays) de 1970 à 1974.

Le modèle se définit de la manière suivante :

$$\log_gdp_pc_usd_{i,t} = \beta_0 + \beta_1 T_{i,t} + \beta_2 P_{i,t} + \beta_3 dT_{i,t} + \beta_4 dP_{i,t} + \beta_5 gdpres_{i,t} + \beta_6 PPP_{i,t} + \gamma_i + \delta_t + \epsilon_{i,t}$$

Avec :

- $\log_gdp_pc_usd_{i,t}$: logarithme (que nous avons décidé d'appliquer pour lisser les données) du PIB par habitant pour chaque région et année (en USD)
- $T_{i,t}$: la température moyenne (en °C)
- $P_{i,t}$: les précipitations (en m)
- $dT_{i,t}$: la variation annuelle de température (en °C)
- $dP_{i,t}$: la variation annuelle de précipitations (en m)
- $gdpres_{i,t}$: variable qui capte la capacité d'une région à absorber les chocs économiques et à maintenir un niveau stable de croissance du PIB (indicateur standardisé, plus il augmente, plus la région a une économie moins vulnérable aux crises)
- $PPP_{i,t}$: le taux de parité de pouvoir d'achat (indice de conversion monétaire, plus il est bas, plus le coût de la vie est élevé et le pouvoir d'achat de la monnaie locale est faible)
- γ_i : les effets fixes régionaux, capturant les caractéristiques inobservables spécifiques à chaque région
- δ_t : les effets fixes temporels, contrôlant les chocs globaux qui affectent toutes les régions en même temps
- $\epsilon_{i,t}$: le terme d'erreur capturant les variations idiosyncratiques non expliquées par les variables du modèle

Afin d'estimer ce modèle, nous présenterons les données et les statistiques descriptives avant de nous attarder sur différents types d'estimations et de tests. Nous commenterons les résultats de ces estimations et analyserons les conséquences économiques des résultats obtenus ainsi que la validité statistique et économique du modèle. Enfin, nous conclurons en essayant d'apporter des perspectives aux politiques économiques liées à l'environnement et au climat.

I. CADRE STATISTIQUE

A. STATISTIQUES DESCRIPTIVES

Statistiques descriptives des variables

Statistiques	log_gdp_pc_usd	T	P	dT	dP	gdpres	ppp_rate
Minimum	0.000000	-3.038446	0.0168025	-2.3394070	-1.3850024	-4.7216930	2.951929
Médiane	6.899209	13.240351	0.9708659	-0.0196705	0.0048648	0.1044052	5.730871
Moyenne	6.562209	13.876685	1.0367394	-0.0183252	0.0049546	-0.1324723	5.990111
Maximum	10.164658	28.362186	3.3635504	2.5924609	1.0314436	2.4968486	11.445905
Écart-Type	1.893369	6.079473	0.5717427	0.5481204	0.2543518	0.9662178	2.058412

- $\log_gdp_pc_usd_{i,t}$:

Le logarithme du PIB par habitant, exprimé en USD, varie entre 0 (PIB par habitant de 1 USD à Siaya au Kenya en 1970, donc donnée logique) et 10.164658, avec une moyenne de 6.562209 et un écart-type de 1.893369, mettant en lumière une dispersion modeste du niveau de vie entre les régions des pays de notre échantillon.

- $T_{i,t}$:

La température moyenne, exprimée en °C, prend des valeurs entre -3.03°C et 28.36°C, avec une moyenne de 13.88°C et un écart-type de 6.08°C, montrant une grande variabilité de climat entre les régions.

- $P_{i,t}$:

Les précipitations annuelles, exprimées en m, fluctuent entre 0.0168 m et 3.36 m, avec une moyenne de 1.04 m et un écart-type de 0.57 m. Ces valeurs soulignent des différences importantes en termes de précipitations régionales.

- $dT_{i,t}$:

La variation annuelle de température, exprimée en °C, oscille entre -2.34°C et 2.59°C, avec une moyenne de -0.018°C et un écart-type de 0.55°C, montrant une tendance globale stable avec des écarts limités.

- $dP_{i,t}$:

La variation annuelle des précipitations, exprimée en m, s'étend de -1.39 m à 1.03 m, avec une moyenne égale à 0.005, ce qui est proche de zéro et un écart-type de 0.25 m, reflétant des fluctuations climatiques annuelles modérées.

- $gdpres_{i,t}$:

L'indice de résilience économique des régions, sans unité car standardisé, oscille entre -4.72 et 2.50, avec une moyenne de -0.13 et un écart-type de 0.97, indiquant que certaines régions sont plus vulnérables à l'impact d'un choc économique que d'autres.

- $PPP_{i,t}$:

L'indice du taux de parité de pouvoir d'achat est compris entre 2.95 et 11.45, avec une moyenne de 5.99 et un écart-type de 2.06, reflétant des disparités dans le coût de la vie et le pouvoir d'achat entre les régions.

B. ESTIMATIONS ET TESTS

Dans cette sous-partie, il sera question d'effectuer des comparaisons entre les estimations des différents modèles possibles (OLS, Within et GLS) et de réaliser une série de tests qui nous permettra d'identifier l'estimateur le plus

pertinent pour notre modèle. Ainsi, on a regroupé les résultats des estimations des trois modèles possibles dans le tableau ci-dessous :

Comparaison des Estimations des Modèles (OLS, Within, GLS)

	OLS	Within (Effets Fixes)	GLS (Effets Aléatoires)
(Intercept)	10.3031 (0.0718) ***	NA	8.7278 (0.0937) ***
T	-0.0112 (0.0034) ***	-0.013 (0.0101)	-0.0522 (0.0057) ***
P	0.2034 (0.0329) ***	0.068 (0.023) ***	0.1171 (0.0265) ***
dT	0.1517 (0.0322) ***	0.0112 (0.0066) *	0.0326 (0.0058) ***
dP	-0.0044 (0.0718)	-0.0603 (0.0145) ***	-0.0822 (0.0172) ***
gdpres	1.3749 (0.0207) ***	1.1137 (0.0171) ***	1.1281 (0.019) ***
ppp_rate	-0.6029 (0.0092) ***	-0.2252 (0.0025) ***	-0.2356 (0.003) ***
R ²	0.8792	0.9326	0.8936
R ² ajusté	0.8787	0.9154	0.8932
F-test (Effets Fixes)	NA	329.776 ***	NA
Hausman	NA	12.651 **	NA
LM-test (Effets Aléatoires)	NA	NA	1142.773 ***
Nombre d'observations	1400	1400	1400
Valeur de la p-value	*** p<0.01, ** p<0.05, * p<0.1		

En plus du fait que le tableau regroupe les estimations des échantillons, il présente également les statistiques de test des tests d'Effets Fixes, d'Effets Aléatoires et celui d'Hausman.

Au vu des résultats, l'estimation par les MCO (OLS) suppose évidemment l'absence d'hétérogénéité individuelle non observée et de ce fait elle fournit forcément des estimations biaisées si de tels effets existent. On remarque que tous les coefficients sont significatifs au seuil $\alpha = 1\%$ compte tenu de leur p-value inférieur à ce seuil, seul la variable explicatives représentant la variation annuelle des précipitation ne l'est pas. Ce modèle présente tout de même une grande capacité explicative $R^2 = 0.8792$ (dont un R^2 ajusté quasiment identique).

Pour le modèle à Effets Fixes, nous avons fait le choix d'utiliser l'estimateur Within avec Effets Fixes Individuels puisque notre échantillon couvre un nombre significatif de régions i mais peu de périodes t (5 ans). Contrairement aux autres modèles, celui-ci contrôle pour les effets spécifiques aux individus de notre échantillon en supprimant la constante. A l'inverse du modèle OLS, certains coefficients deviennent moins significatifs, comme la variation de la température annuelle, et d'autres ne le sont plus du tout, comme la température moyenne. En revanche, les autres variables sont toujours significatives au seuil $\alpha = 1\%$. On remarque que l'introduction des effets fixes individuels améliore la valeur de notre R^2 qui passe à 0.9326 (toujours avec un R^2 ajusté quasiment identique). En effet, la statistique de test des effets fixes (F-test) confirme que les effets fixes régionaux sont significatifs et le test d'Hausman montre même que ce modèle est préférable à celui donné par l'estimateur GLS.

Concernant l'estimation GLS, celle-ci fait l'hypothèse que les effets spécifiques sont aléatoires et non corrélés avec nos variables explicatives. Dans ce modèle, tous les coefficients sont fortement significatifs avec des p-value nettement inférieur au seuil $\alpha = 1\%$. Par ailleurs, le test d'existence d'effets aléatoires (LM-test) nous donne une statistique de test qui justifie l'utilisation d'un modèle de panel au lieu d'une régression par les MCO. En revanche, comme cité précédemment, c'est le modèle à Effets Fixes Individuels qui sera retenu compte tenu du test de Hausman.

II. COMMENTAIRES

Dans cette partie, il sera question de commenter en détail la manière dont nous avons prouvé que le modèle à Effets Fixes Individuels était le meilleur, ainsi que ces implications.

Avec le test des Effets Fixes (F-test), on a cherché à vérifier s'il est intéressant ou non d'introduire des effets fixes régionaux (γ_i) dans notre modèle. Voici comment s'écrit notre modèle à Effets Fixes Individuels :

$$\log_gdp_pc_usd_{i,t} = \beta_1 T_{i,t} + \beta_2 P_{i,t} + \beta_3 dT_{i,t} + \beta_4 dP_{i,t} + \beta_5 gdpres_{i,t} + \beta_6 PPP_{i,t} + \gamma_i + \epsilon_{i,t}$$

Pour notre test, on a posé les hypothèses suivantes : $H_0: \gamma_i = 0$ et $H_1: \gamma_i \neq 0$. Au vu des résultats obtenus [1], il est clair que nous rejetons l'hypothèse nulle au seuil $\alpha = 1\%$ puisque la statistique F est de 329.776 ce qui est supérieur à la valeur critique $F^\alpha(279,1114)$ et la p-value est inférieure à $2.2e^{-16}$ ce qui est inférieur à notre seuil. Ainsi, les effets fixes régionaux sont significatifs $\gamma_i \neq 0$ et le modèle des MCO est clairement inadapté face à un modèle à Effets Fixes.

Par la suite, nous réalisons le test des Effets Aléatoires (LM-test) avec comme nouveau modèle à Effets Aléatoires Individuels avec $u_{i,t} = \mu_i + v_{i,t}$ et $\mu_i \sim \mathcal{N}(0, \sigma_\mu^2)$, $v_{i,t} \sim \mathcal{N}(0, \sigma_v^2)$:

$$\log_gdp_pc_usd_{i,t} = \beta_0 + \beta_1 T_{i,t} + \beta_2 P_{i,t} + \beta_3 dT_{i,t} + \beta_4 dP_{i,t} + \beta_5 gdpres_{i,t} + \beta_6 PPP_{i,t} + u_{i,t}$$

On a testé les hypothèses suivantes : $H_0: \sigma_\mu^2 = 0$ et $H_1: \sigma_\mu^2 \neq 0$. En analysant les résultats [2], on remarque que $LM = 1142.773 > \chi^2(1) = 3.841$ de plus que la p-value est largement inférieure au seuil $\alpha = 1\%$ ce qui confirme le fait que nous rejetons l'hypothèse nulle à ce seuil. En d'autres termes, il existe des effets aléatoires significatifs dans notre modèle.

Ainsi, quel modèle choisir ? C'est le test de Hausman qui nous a permis de départager entre ces deux modèles. Il a pour but de déterminer si les effets individuels μ_i sont corrélés ou non avec nos variables explicatives. On peut écrire nos hypothèses sous cette forme : $H_0: E(\mu_i | X_{i,t}) = 0$ et $H_1: E(\mu_i | X_{i,t}) \neq 0$. Au seuil $\alpha = 5\%$, le test de Hausman [3] nous donne une statistique de $12.651 > \chi^2(6) = 12.592$. Ainsi, on rejette l'hypothèse H_0 à ce seuil et la p-value nous confirme cela. On note tout de même le fait qu'on rejette l'hypothèse nulle à quelques chiffres après la virgule donc on préfère tout de même le modèle à Effets Fixes.

L'estimation de notre modèle à Effets Fixes Individuels [4] montre que les variables $gdpres$, P , dP , et ppp_{rate} ont toutes des effets significatifs au seuil $\alpha = 1\%$ puisque leur p-value est inférieure à ce seuil. D'une part, compte tenu des coefficients estimés, les précipitations et la capacité d'une région à absorber les chocs économiques ont un impact positif sur la croissance régionale par habitant. D'autre part, un coût de la vie plus élevée et une hausse de la variation des précipitations impactent quant à eux négativement le PIB par habitant régional.

Cependant, compte tenu de leur p-value, les variables T et dT , respectivement la température moyenne et sa variation annuelle, n'ont pas de véritables effets sur la croissance d'une région par habitant.

Notre modèle final obtient le R^2 et R^2 ajusté les plus élevés par rapport aux autres modèles (0.9326 et 0.9154) ce qui représente une excellente capacité explicative. Dit autrement, plus de 91% (si on prend le R^2 ajusté) de la variabilité du PIB par habitant régional est expliquée par nos variables explicatives.

Pour finir, il convient d'effectuer des tests de robustesse pour tester si notre modèle souffre d'Hétéroscédasticité ou d'Autocorrélation. On cherche à tester les hypothèses suivantes $\forall i, t$: **Hétéroscédasticité** : $H_0: \text{Var}(\epsilon_{i,t}) = \sigma^2$ et $H_1: \text{Var}(\epsilon_{i,t}) \neq \sigma^2$; **Autocorrélation** : $H_0: E(\epsilon_{i,t} \epsilon_{i,t-1}) = 0$ et $H_1: E(\epsilon_{i,t} \epsilon_{i,t-1}) \neq 0$

Suite aux tests de robustesse de notre modèle [5], on constate qu'il souffre tout de même de problème d'hétéroscédasticité et d'autocorrélation (les statistiques de tests sont nettement supérieures aux valeurs critiques et les p-value sont nettement inférieures au seuil $\alpha = 1\%$). Ainsi on rejette les hypothèses nulles d'homoscédasticité et d'absence d'autocorrélation. Il faudrait donc corriger ces biais à l'aide d'estimateurs robustes à l'hétéroscédasticité et à l'autocorrélation.

CONCLUSION

En somme, ce projet avait pour but d'analyser l'impact de l'environnement climatique sur la production régionale de différents pays dans le monde entre 1970 et 1974 grâce à des modèles économétrique en panel. Nous avons donc comparé trois modèles basés sur des estimateurs différents avec pour objectif de déterminer celui qui appréhende les effets du climat sur l'économie de la façon la plus optimale. Cette analyse s'inscrit dans la continuité de l'article dont nous avons puisé nos données et donc des travaux de **Kalkuhl et al. (2020)** qui mettent en lumière l'effet des conditions climatiques sur la croissance économique.

Les différents modèles que nous avons construits ont conduit à retenir celui qui est le plus pertinent : le modèle à Effets Fixes Individuels. C'est en effet grâce au test de Hausman que nous avons déterminé cette conclusion bien que la décision ait été serrée entre ce modèle et celui à Effets Aléatoires Individuels (à quelques décimales près et au seuil $\alpha = 5\%$), c'est tout de même le modèle déterminé par l'estimateur Within qui l'emporte.

De plus, afin de garantir la robustesse de nos résultats, nous avons tout de même décidé de réaliser des tests spécifiques aux données de panel. D'une part, le test de Wald d'hétéroscédasticité a mis en avant son existence et donc des différences de dispersion des erreurs en fonction des régions. D'autre part, le test de Wooldridge a également révélé la présence d'autocorrélation et donc d'une dépendance temporelle des erreurs montrant que les chocs climatiques pouvaient en réalité avoir des effets persistants sur le PIB par habitant régional. Idéalement, il faudrait essayer de corriger ces problèmes pour obtenir un modèle véritablement optimal.

Ainsi, le modèle retenu présente évidemment des limites. Le fait que la constante n'existe pas dans ce modèle nous empêche d'estimer un effet moyen du climat sur l'ensemble de nos régions. De même, bien que les effets fixes contrôlent les biais en relation avec les caractéristiques des régions, ils ne suffisent pas pour éliminer les problèmes d'endogénéité potentielle. On note également le fait que notre modèle se base sur cinq périodes ce qui peut également être assez restreint lorsqu'on travaille avec des données de panel.

Si on souhaite faire le parallèle avec l'article de **Kalkuhl et al. (2020)**, nos conclusions rejoignent leurs idées puisqu'effectivement, les conditions climatiques influencent significativement la production économique régionale. Une des améliorations possibles serait de considérer plus de périodes et plus de régions à notre modèle mais cela devra se faire sous la contrainte de trouver des données pour ces nouvelles périodes et ces nouvelles régions. Introduire d'autres variables explicatives serait également un moyen de peaufiner cette analyse. Ainsi, ces conclusions mettent en lumière l'important d'intégrer la question du climat dans les politiques économiques dans le but de faciliter la croissance économique des régions du monde, et de façon générale, des pays du monde.

BIBLIOGRAPHIE

Kalkuhl, M., Wenz, L., Fang, S., Levermann, A., & Otto, C. (2020). *The Impact of Climate on Economic Growth: Evidence from a Global Panel of Regions*. *Journal of Environmental Economics and Management*.

El Ouardighi, J. (2024-2025). *Master 1ère année – Analyse et Politiques Économiques, UE – Économétrie appliquée II*. Document de cours officiel, Université de Strasbourg

ANNEXES

1 : Résultat du F-test des effets fixes

Résultats du F-test des effets fixes

	Statistique.F	Ddl.1	Ddl.2	p.value
F	329.776	279	1114	< 2.2e-16

2 : Résultats du test de Breusch-Pagan pour les effets aléatoires individuels

Résultats du test de Breusch-Pagan pour
les effets aléatoires (Individuels)

	Statistique.LM	Ddl	p.value
chisq	1142.773	1	< 2.2e-16

3 : Résultats du test de Hausman

Résultats du test de Hausman

	Statistique.Chi.	Ddl	p.value
chisq	12.651	6	0.0489

4 : Résultats de l'estimation finale du modèle à effets fixes individuels

Résultats de l'estimation finale (Effets Fixes Individuels - Within)

	Estimate	Standard.Error	t.value	p.value
T	-0.0130	0.0101	-1.2904	0.197170
P	0.0680	0.023	2.9605	0.003137
dT	0.0112	0.0066	1.6938	0.090585
dP	-0.0603	0.0145	-4.1532	3.529e-05
gdpres	1.1137	0.0171	65.152	< 2.2e-16
ppp_rate	-0.2252	0.0025	-88.7652	< 2.2e-16
R ²	0.9326			
R ² ajusté	0.9154			

5 : Résultats des tests de robustesse

Résultats des tests de robustesse (Hétéroscédasticité et Autocorrélation)

	Test	Statistique	Ddl	p.value
Chisq	Wald (Hétéroscédasticité)	15424.726	6	<2e-16
F	Wooldridge (Autocorrélation)	247.686	-	<2e-16