



Econometrie des données de panel

Rapport de travaux pratiques

Revisiter la relation entre le revenu et la démocratie : Une analyse basée sur l'article d'Acemoglu, Johnson, Robinson et Yared (2008)

Auteur:

Moulaye Ahmed BRAHIM
Zeinebou TAKI

Sous la supervision de :

Ilyes BOUMAHDI

January 1, 2025

Contents

1	Introduction	2
2	Présentation des données	2
3	Modèle économétrique	3
4	Résultats et Discussions	3
5	Conclusion	11

1 Introduction

L'une des régularités empiriques les plus intrigantes en économie politique est l'association positive entre le revenu par habitant et la démocratie. Ce lien est intuitivement observable : les pays de l'Organisation de Coopération et de Développement Économiques (OCDE) sont largement démocratiques, tandis que de nombreux régimes non démocratiques se concentrent dans les régions économiquement défavorisées, telles que l'Afrique subsaharienne et l'Asie du Sud-Est. Historiquement, la transition vers la démocratie semble souvent accompagner le processus de croissance économique moderne, amorcé au début du XIX^e siècle.

Cependant, cette corrélation entre le revenu par habitant et la démocratie ne suffit pas pour conclure à une relation causale. Les travaux antérieurs, bien que démontrant une relation positive dans les analyses transversales, négligent fréquemment deux problèmes majeurs : la causalité inverse (la démocratie pourrait influencer le revenu) et le biais des variables omises (des facteurs historiques ou institutionnels affectant simultanément le revenu et la démocratie). L'étude d'Acemoglu, Johnson, Robinson et Yared (2008) revisite cette problématique en intégrant des effets fixes par pays et des stratégies d'instrumentation afin de neutraliser ces biais.

Dans ce projet, nous reproduirons les résultats de cet article en nous concentrant sur les estimations à effets fixes. En nous appuyant sur deux ensembles de données de panel, nous étudierons la relation dynamique entre le revenu par habitant et le niveau de démocratie, tout en testant l'effet causal hypothétique du revenu sur la démocratie.

2 Présentation des données

Pour évaluer la relation entre le revenu par habitant et la démocratie, nous utilisons un ensemble de données de panel couvrant la période 1950-2000, contenant des observations quinquennales pour 211 pays.

La principale mesure de la démocratie est l'indice des droits politiques de Freedom House, qui évalue des critères tels que la liberté des élections, le rôle des partis politiques, la participation des groupes minoritaires et l'existence d'une opposition effective. Cet indice est normalisé entre 0 (absence totale de démocratie) et 1 (démocratie parfaite).

Les données sur le PIB par habitant proviennent de deux sources : les Tables de Maddison (2003) pour les séries longues, et les estimations ajustées en parité de pouvoir d'achat (PPA) pour les séries plus récentes. Ces deux mesures permettent de capturer la dynamique économique sur des périodes et des échantillons différents.

Pour garantir une analyse robuste, les données sont transformées en panels quinquennaux. De plus, des effets fixes par pays et des effets temporels sont inclus pour contrôler les chocs spécifiques à un pays et les tendances globales, respectivement.

3 Modèle économétrique

La relation entre la démocratie et le revenu par habitant est modélisée à l'aide de l'équation suivante :

$$d_{it} = \alpha d_{i,t-1} + \gamma y_{i,t-1} + \mathbf{X}'_{i,t-1} \beta + \mu_t + \delta_i + u_{it}$$

- d_{it} : Niveau de démocratie dans le pays i à la période t .
- $d_{i,t-1}$: Valeur retardée de la démocratie, capturant la persistance et la dynamique de retour à un équilibre spécifique.
- $y_{i,t-1}$: Logarithme du revenu par habitant retardé, principale variable d'intérêt.
- $\mathbf{X}_{i,t-1}$: Vecteur de covariables, incluant d'autres facteurs économiques ou sociaux influençant la démocratie.
- μ_t : Effets temporels, contrôlant les tendances globales communes à tous les pays.
- δ_i : Effets fixes spécifiques aux pays, capturant les facteurs historiques invariants.
- u_{it} : Terme d'erreur, supposé non corrélé avec les variables explicatives.

L'estimation à effets fixes est utilisée pour neutraliser les biais liés aux facteurs inobservables constants dans le temps, tels que les héritages historiques ou les spécificités institutionnelles d'un pays. En complément, des stratégies d'instrumentation sont employées pour pallier d'éventuels problèmes d'endogénéité, notamment par l'utilisation de variables retardées en différence ou en niveau.

4 Résultats et Discussions

Comme mentionné précédemment, nous utilisons l'article d'Acemoglu, Johnson, Robinson et Yared (2008) pour répliquer et illustrer les résultats. Cette étude examine la relation causale entre le niveau de richesse et celui de la démocratie dans un pays. Les auteurs s'appuient sur différents ensembles de données de panel, parmi lesquels nous avons considéré :

- Un premier ensemble composé de données observées tous les 5 ans, avec 11 observations couvrant la période 1950-2000 pour 211 pays.

Dans l'analyse transversale, une relation positive entre le degré de démocratie et le revenu par habitant est apparente. Cela est illustré dans la Figure 1, qui utilise les données d'Acemoglu et al. (2008) pour l'année 1999. Cependant, cette corrélation contemporaine n'implique pas nécessairement une relation de cause à effet entre les deux variables.

L'utilisation de données de panel permet d'étudier la causalité en spécifiant une relation dynamique. Cette approche aide à dépasser les limites des analyses transversales en intégrant les dimensions temporelles et structurelles des données.

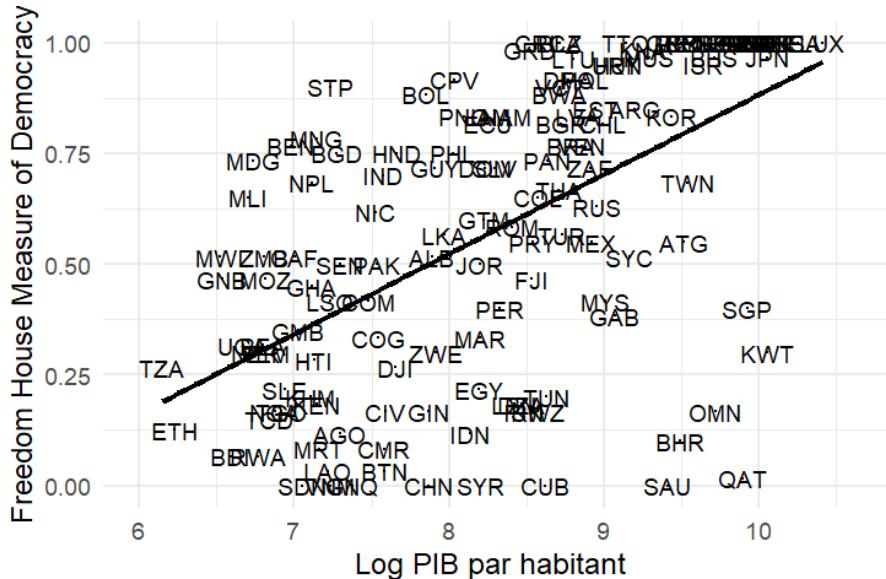


Figure 1: Relation entre le revenu par habitant et la démocratie en 1999 (Acemoglu et al., 2008).

Dans le modèle d'Acemoglu et al. (2008), la variable dépendante est l'indice de démocratie, et les régresseurs sont les retards d'une période de l'indice de démocratie lui-même et du revenu par habitant. L'estimation des moindres carrés ordinaires (OLS) par la fonction `lm()` de R est biaisée en raison de la présence de valeurs retardées. En effet, la méthode `lag()` utilisée par le programme est conçue pour les séries temporelles et non pour les données de panel.

Pour cette raison, la fonction `plm()` du paquetage `plm` sera utilisée à la place, avec l'argument `model` fixé à '`'pooling'`', ce qui permet de conserver les données non transformées. Le `-1` dans la formule indique que nous ne souhaitons pas estimer une constante générale, mais un coefficient pour toutes les instances de la variable année, ce qui n'affecte pas l'estimation.

```
ols <- plm(democracy ~ lag(democracy) + lag(income),
             DemocracyIncome, index = c("country", "year"),
             model = "within", effect = "time",
             subset = sample == 1)
summary(ols)
```

Le même modèle peut être estimé en réglant le modèle sur « `within` » et l'effet sur « `time` » :

```
ols <- plm(democracy ~ lag(democracy) + lag(income) + year - 1,
             DemocracyIncome, index = c("country", "year"),
             model = "pooling", subset = sample == 1)
summary(ols)
```

L'estimation initiale du modèle révèle deux résultats principaux. Tout d'abord, la variable retardée de la démocratie montre une forte persistance, avec un coefficient de

0,71, confirmant que la démocratie tend à maintenir une certaine inertie au fil du temps. Cette persistance est également illustrée par la significativité élevée du décalage de la démocratie (de cinq ans) dans les régressions MCO groupées, comme le montre le tableau 1. Ce résultat indique que la démocratie présente une dynamique de retour vers un équilibre.

Table 1: Résultats du Modèle Pooling

Variable	Estimate	Std. Error	t-value	Pr(> t)
<i>lag(democracy)</i>	0.7064	0.0243	29.08	$< 2.2 \times 10^{-16} ***$
<i>lag(income)</i>	0.0723	0.0083	8.67	$< 2.2 \times 10^{-16} ***$
<i>year1960-1964</i>	-0.3469	0.0639	-5.42	$7.39 \times 10^{-8} ***$
<i>year1965-1969</i>	-0.4297	0.0603	-7.13	$2.05 \times 10^{-12} ***$
<i>year1970-1974</i>	-0.5462	0.0607	-9.00	$< 2.2 \times 10^{-16} ***$
<i>year1975-1979</i>	-0.4586	0.0630	-7.28	$6.92 \times 10^{-13} ***$
<i>year1980-1984</i>	-0.3970	0.0634	-6.26	$5.78 \times 10^{-10} ***$
<i>year1985-1989</i>	-0.4195	0.0633	-6.63	$5.75 \times 10^{-11} ***$
<i>year1990-1994</i>	-0.3995	0.0632	-6.32	$4.04 \times 10^{-10} ***$
<i>year1995-1999</i>	-0.3791	0.0633	-5.99	$2.93 \times 10^{-9} ***$
<i>year2000-2004</i>	-0.4031	0.0627	-6.43	$2.03 \times 10^{-10} ***$

Statistiques du Modèle	
<i>R-Squared</i>	0.7252
<i>Adj. R-Squared</i>	0.7222
<i>F-statistic</i>	$979.81 (\text{df} = 11, 934), p < 2.2 \times 10^{-16}$
<i>Total SS</i>	124.67
<i>Residual SS</i>	34.27

Ensuite, le logarithme du revenu par habitant, en tant que variable retardée, exerce une influence positive et statistiquement significative sur l'indice de démocratie. Ce résultat corrobore la relation positive bien documentée entre le revenu et la démocratie. Toutefois, bien que significatif, l'effet du revenu demeure quantitativement limité : un coefficient de 0,072 (erreur standard de 0,010) implique qu'une augmentation de 10 % du PIB par habitant entraînerait une augmentation de moins de 0,007 dans le score Freedom House, ce qui reste modeste. Si cette estimation transversale identifiait l'effet causal, l'effet à long terme serait amplifié par la dynamique temporelle, car l'influence du retard de la démocratie continuerait à croître dans le temps, induisant des augmentations successives du score de démocratie.

Cependant, l'estimateur MCO souffre d'un biais positif en raison de la corrélation entre le terme d'erreur et la variable endogène retardée, causée par la présence d'effets individuels. On peut donc penser à résoudre ce problème par une transformation qui élimine l'effet individuel. Le choix le plus évident est l'estimateur *within*.

Modèle Within

Le modèle *within* est obtenu avec `plm` en fixant les arguments `model` et `effect` à "within" et "twoways", puisque nous voulons introduire des effets individuels et temporels. Le modèle peut être simplement estimé en mettant à jour le modèle OLS précédent avec le code suivant :

```
within <- update(ols, model = "within", effect = "twoways")
summary(within)
```

Les résultats dans le tableau 2 montrent que la relation entre le revenu et la démocratie disparaît une fois que les effets fixes sont inclus. Par exemple, dans le tableau 2 avec les données de Freedom House, l'estimation de σ est de 0,010 avec une erreur standard de 0,035, ce qui la rend très peu significative. Par rapport au modèle OLS, le coefficient autorégressif est plus faible (0,38 contre 0,71), ce qui était prévisible puisque l'estimateur *within* est biaisé à la baisse, alors que l'OLS est biaisé à la hausse. Il convient également de noter qu'après l'introduction des effets individuels, le coefficient du revenu est très proche de 0 et n'est plus significatif.

Table 2: Résultats du Modèle *Twoways Effects Within*

Variable	Estimate	Std. Error	t-value	Pr(> t)
<i>lag(democracy)</i>	0.3786	0.0334	11.32	< 2e-16***
<i>lag(income)</i>	0.0104	0.0264	0.39	0.6933
Statistiques Résiduelles				
Min.		-0.7159		
1er Quartile		-0.0747		
Médiane		-0.0006		
3e Quartile		0.0845		
Max.		0.6075		
Statistiques Globales				
R-Squared		0.1436		
Adj. R-Squared		-0.0299		
F-statistic	65.82 on 2 and 785 DF, p-value < 2.22e-16			
Total Sum of Squares		29.713		
Residual Sum of Squares		25.446		

Ces premiers résultats montrent qu'une fois les effets fixes pris en compte, le revenu par habitant n'est pas un déterminant majeur de la démocratie. Les effets individuels (et donc le biais) ont disparu de l'estimateur OLS, mais une deuxième source de biais a été introduite. En effet, le régresseur *lag(democracy)* est mécaniquement corrélé avec l'erreur u_{it} pour $s < t$ (biais introduits par la présence de la variable dépendante retardée). Ce biais diffère de celui des OLS à deux égards. D'une part, il est négatif et d'autre part, il tend

vers 0 lorsque T tend vers l'infini. Ce biais ne peut donc pas être ignoré dans le cas des micro-panel, où la dimension temporelle est courte.

La section suivante examine des stratégies d'estimation alternatives pour traiter les biais potentiels introduits par la présence de la variable dépendante retardée.

Méthodes d'Estimation Consistantes pour les Modèles Dynamiques

Les méthodes d'estimation les plus courantes utilisées pour les modèles statiques étant inappropriées, diverses stratégies d'estimation peuvent être adoptées pour atteindre la cohérence. L'une d'entre elles consiste à utiliser la méthode des variables instrumentales, les instruments étant des niveaux ou des différences retardés de la variable dépendante. La méthode généralisée des moments, qui est une extension de la méthode des variables instrumentales, est devenue de plus en plus populaire à cette fin.

L'estimateur proposé par Anderson et Hsiao (1982) utilise le modèle écrit en première différence afin d'éliminer les effets individuels.

Pour calculer l'estimateur d'Anderson et Hsiao (1982), il faut spécifier que le régresseur et les régresseurs sont différenciés et que la variable endogène retardée en différences est instrumentée avec la variable endogène en niveaux retardée de deux périodes. Le modèle est simplement décrit à l'aide d'une formule en deux parties, la première indiquant les variables explicatives et la seconde les instruments, les deux parties étant séparées par le signe |.

```
ahsiao <- plm(diff(democracy) ~ lag(diff(democracy)) +
  lag(diff(income)) + year - 1 |
  lag(democracy, 2) + lag(income, 2) + year - 1,
  DemocracyIncome, index = c("country", "year"),
  model = "pooling", subset = sample == 1)
summary(ahsiao)
```

Le modèle d'Anderson et Hsiao (1982) étant consistant, on s'attend à ce que le coefficient autorégressif estimé soit compris entre celui du modèle within (biaisé à la baisse) et celui du modèle OLS (biaisé à la hausse). C'est effectivement le cas ici, la valeur obtenue de 0,47 se situe entre 0,38 et 0,71 (Tableau 3).

Estimation GMM du Modèle Différencié

L'estimateur des variables instrumentales présenté dans la section précédente est inefficace pour deux raisons :

- Premièrement, il ne tient pas compte de la corrélation induite dans les erreurs par la première différenciation.
- Deuxièmement, il existe d'autres instruments valides.

Variable	Estimate	Std. Error	z-value	Pr(> z)
lag(diff(democracy))	0.4686593	0.1181956	3.9651	7.336e-05***
lag(diff(income))	-0.1035793	0.3048546	-0.3398	0.73403
year1960-1964	0.0617385	0.0614301	1.0050	0.31489
year1965-1969	-0.0269418	0.0542965	-0.4962	0.61975
year1970-1974	-0.0901173	0.0520273	-1.7321	0.08325 .
year1975-1979	0.0593314	0.0473566	1.2529	0.21025
year1980-1984	0.0903238	0.0397899	2.2700	0.02321 *
year1985-1989	0.0032207	0.0398274	0.0809	0.93555
year1990-1994	0.0353261	0.0261877	1.3490	0.17735
year1995-1999	0.0437321	0.0333553	1.3111	0.18982
year2000-2004	0.0029709	0.0284488	0.1044	0.91683
Total Sum of Squares			39.296	
Residual Sum of Squares			53.376	
R-Squared			0.0053628	
Adj. R-Squared			-0.0066643	
Chi-squared			64.8996 on 11 DF	
p-value			1.1255e-09	

Table 3: Estimation des coefficients du modèle avec l'estimateur d'Anderson et Hsiao

Variables Instrumentales et Méthode des Moments Généralisée

Bien que l'estimateur à variable instrumentale d'Anderson et Hsiao (1982) conduise à des estimations consistantes, il n'est pas efficace, sous l'hypothèse d'absence de corrélation sérielle. Arellano et Bond (1991) développent un estimateur par la méthode généralisée des moments (GMM) qui utilise toutes ces conditions de moment. Lorsque ces conditions sont valides, cet estimateur GMM est plus efficace que l'estimateur d'Anderson et Hsiao (1982).

Cet estimateur tient compte du fait que le nombre d'instruments valides croît avec t . Le caractère dynamique du modèle rend la première observation inutilisable et la première différenciation devient la seconde.

L'estimation GMM d'un modèle de panel est réalisée par la fonction `pgmm` de la bibliothèque `plm`. Les arguments de cette fonction sont les mêmes que pour la fonction `plm`, avec quelques arguments spécifiques :

- **formula** : la formule est particulière, car elle comporte trois parties : la première, comme d'habitude, contient les variables explicatives, la deuxième les instruments « gmm », et la troisième les instruments « normaux ».
- **modèle** : le modèle à estimer, soit en une étape : "onestep", soit en deux étapes : "twosteps".
- **effet** : les effets sont soit « individuels » (ils sont alors éliminés par différenciation),

soit « twoways », auquel cas des variables indicatrices pour chaque période sont ajoutées en tant qu'instruments « normaux ».

Nous calculons l'estimateur à une étape avec le code suivant :

```
diff1 <- pgmm(democracy ~ lag(democracy) + lag(income) |
                 lag(democracy, 2:99) | lag(income, 2),
                 DemocracyIncome, index=c("country", "year"),
                 model="onestep", effect="twoways", subset = sample == 1)
summary(diff1)
```

Nous utilisons cet estimateur GMM dans le tableau 4. Les coefficients sont maintenant encore plus négatifs et estimés plus précisément, par exemple $-0,129$ (erreur standard = $0,076$).

Table 4: Résultats de l'estimation GMM

Balanced Panel:				
n	T	N	Number of Observations Used: 838	
211	11	2321		
Residuals:				
Min.	1st Qu.	Median	Mean	3rd Qu.
-1.2491	0.0000	0.0000	0.0006	0.0000
Max.	1.0133			
Coefficients:	Estimate	Std. Error	z-value	Pr(> z)
lag(democracy)	0.504994	0.090490	5.5806	$2.396 \times 10^{-8} ***$
lag(income)	-0.090108	0.080291	-1.1223	0.2617
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1				

La méthode généralisée des moments n'est pas consistante si les conditions des moments ne sont pas vérifiées, ce qui implique notamment que les erreurs ne sont pas corrélées en série. Arellano et Bond (1991) ont proposé un test approprié.

Le test d'Arellano-Bond vérifie si les erreurs différencierées d'ordre supérieur sont autocorrélées. En particulier, il s'intéresse à :

- **Autocorrélation d'ordre 1 (AR(1))** : Normalement, on s'attend à une autocorrélation dans les erreurs différencierées d'ordre 1 à cause de la manière dont elles sont construites (différenciation des données).
- **Autocorrélation d'ordre 2 (AR(2))** : C'est cette autocorrélation qu'on cherche à éviter, car une autocorrélation d'ordre 2 indiquerait que les erreurs en niveaux ne sont pas indépendantes, ce qui rendrait les instruments invalides.

Le test repose sur les résidus estimés du modèle GMM. Les hypothèses du test sont les suivantes :

- **Hypothèse nulle (H_0) :** Les erreurs différenciées ne présentent pas d'autocorrélation d'ordre 2.
- **Hypothèse alternative (H_1) :** Les erreurs différenciées présentent une autocorrélation d'ordre 2.

Si la statistique du test d'AR(2) est significative, cela indique une autocorrélation des erreurs et remet en question la validité des instruments utilisés.

Ce test est obtenu par la fonction suivante :

```
mtest(diff1, order = 2)
```

Test	Valeurs
Données	democracy ~ lag(democracy) + lag(income) lag(democracy, 2:99) ...
Statistique normale	0.84374
<i>p-value</i>	0.3988
Hypothèse alternative	Autocorrélation présente

Table 5: Test d'autocorrélation Arellano-Bond de degré 2 (AR(2))

Le test AR(2) indique qu'il n'y a plus d'autocorrélation.

En outre, dans un modèle avec des instruments, les instruments doivent être exogènes, c'est-à-dire non corrélés avec l'erreur résiduelle. Le test de Hansen-J évalue si cette hypothèse est valide lorsque le nombre d'instruments est supérieur au nombre de variables instrumentées (situation de sur-identification).

- **Hypothèse nulle (H_0) :** Les instruments sont exogènes, c'est-à-dire qu'ils ne sont pas corrélés avec les erreurs résiduelles.
- **Hypothèse alternative (H_1) :** Au moins un des instruments est endogène, c'est-à-dire corrélé avec les erreurs résiduelles.

Un *p-value* faible (par exemple, $< 0,05$) suggère que H_0 est rejetée, ce qui indique que les instruments pourraient être invalides.

Le test de Sargan-Hansen peut être effectué à l'aide de la fonction `sargan`. Par exemple, pour le modèle de différence en une étape, on utilise la commande suivante :

```
sargan(diff1)
```

Le test J de Hansen (Tableau 6) indique que les restrictions de suridentification ne sont pas rejetées.

Les résultats présentés jusqu'à présent ne montrent pas d'effet causal du revenu sur la démocratie. Bien que le revenu et la démocratie soient positivement corrélés, il n'y a aucune preuve d'un effet causal. Au lieu de cela, des facteurs omis – très probablement historiques – semblent avoir façonné les trajectoires divergentes de développement

Test	Valeurs
Données	<code>democracy ~ lag(democracy) + lag(income) lag(democracy, 2:99) ...</code>
Statistique χ^2	70.745
Degrés de liberté (df)	44
$p\text{-value}$	0.006447
Hypothèse alternative	Les restrictions de sur-identification ne sont pas valides

Table 6: Résultats du test de Sargan-Hansen

politique et économique de diverses sociétés, conduisant à l’association positive entre la démocratie et la performance économique. Par conséquent, les régressions qui incluent des effets fixes par pays et/ou des régressions de variables instrumentales ne montrent aucune preuve d’un effet causal du revenu sur la démocratie.

5 Conclusion

Ce projet a permis d’explorer la relation complexe entre le revenu par habitant et la démocratie en reproduisant les travaux d’Acemoglu, Johnson, Robinson et Yared (2008). En utilisant des données de panel et des méthodologies économétriques avancées, nous avons démontré que, bien que la démocratie et le revenu soient positivement corrélés dans les analyses transversales, cette relation ne peut être interprétée comme un lien de causalité direct.

Les estimations à effets fixes ont révélé que la relation positive entre le revenu et la démocratie disparaît une fois que les facteurs spécifiques aux pays, tels que les héritages historiques et les caractéristiques institutionnelles, sont contrôlés. De plus, les modèles dynamiques utilisant des méthodes comme GMM ou l’estimateur d’Anderson-Hsiao ont montré que la variable de revenu retardé n’a pas d’effet significatif sur le niveau de démocratie. Ces résultats confirment que des biais potentiels, tels que la causalité inverse ou l’omission de variables pertinentes, peuvent expliquer les corrélations observées dans les études antérieures.

L’absence d’effet causal du revenu sur la démocratie met en lumière l’importance des facteurs historiques et institutionnels dans le développement politique et économique des nations. Cette conclusion incite à approfondir l’étude des déterminants historiques et culturels qui façonnent simultanément les trajectoires démocratiques et économiques.

En reproduisant ces résultats, nous avons souligné la robustesse des méthodologies économétriques modernes pour évaluer les hypothèses causales et avons confirmé que les résultats des analyses transversales doivent être interprétés avec prudence.