
PIB et Démocratie : une relation causale ?

Édith DARIN
Noémie HAOUZI

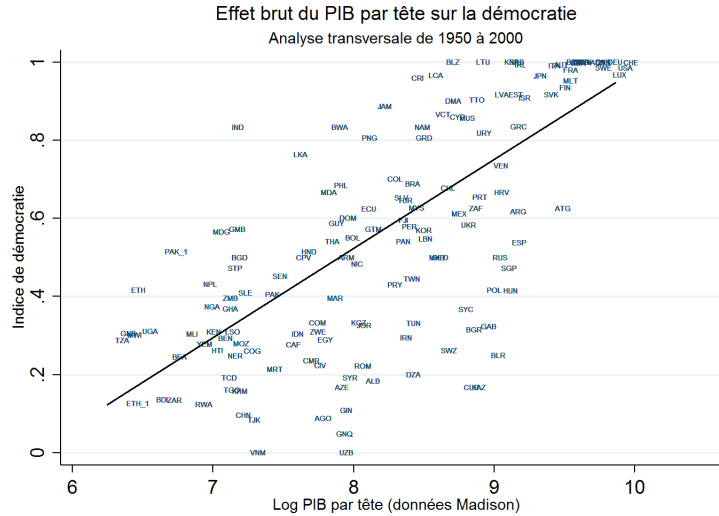
Sous la direction de Jérémy L'Hour

ENSAE ParisTech
Projet d'Econométrie 2

1 Introduction

1.1 Problématique

On se propose dans ce projet de discuter les résultats obtenus par Acemoglu *et alli* dans leur article de 2008¹ sur les relations entre PIB et démocratie. Leur idée est la suivante : il semblerait qu'il y ait de nos jours une relation causale entre la richesse d'un pays, qui se manifeste par un PIB par tête élevé, et son niveau de démocratie, que l'on peut mesurer à travers le *Freedom House Political Rights Index*. L'existence d'un lien entre ces deux variables est clairement représentée par le graphique suivant qui montre une relation croissante entre ces deux indicateurs.²



Sont représentés la moyenne de l'indice Freedom House sur les années 1950 à 2000 de chaque région géographique, en fonction du PIB par tête, en PPA, moyen sur cette même période. La droite de régression a pour pente 0.23 (écart-type = 0.005), $N=211$ et $R^2 = 0.51$.

FIGURE 1 – Démocratie et Revenu

Cependant, et c'est l'objet de l'article que nous discutons, cette relation n'avait été analysée jusqu'ici³ que sous l'angle transversale, en comparant les pays, sans considérer notamment la présence d'effets fixes par pays à comprendre comme des facteurs historiques structurels spécifiques à chaque pays qui influencent à la fois son développement économique et son développement politique sans varier dans le temps. Ainsi, grâce aux données de panel rassemblées par Acemoglu *et alli* on peut chercher à mesurer également les variations ayant lieu à l'intérieur d'un même pays, une fois contrôlées les caractéristiques invariantes au cours du temps. Se pose à ce moment la question de l'existence d'un biais de simultanéité : peut-on mesurer l'effet pur causal de la richesse d'un pays sur sa démocratie ou est-il forcément entâché d'un phénomène de causalité inverse, le contexte politique d'un pays ayant un impact sur sa croissance économique ? Pour répondre à ces deux problèmes, Acemoglu *et alli* ont choisi de mener une analyse en régression à effets fixes en y ajoutant des variables instrumentales pour pouvoir conclure sur un effet causal.

1.2 Démarche méthodologique

Toutefois, si nous nous rangeons à leur étude des biais résultant de l'approche transversale, nous souhaitons émettre ici les doutes quant à l'implémentation de la résolution faite par les auteurs de l'article.⁴ En effet, ils présentent dans un premier temps un modèle de régression à effets fixes sur données de panel. Or, en incluant des

1. Daron Acemoglu, Simon Johnson, James Robinson et Pierre Yared. 2008. "Income and Democracy." *American Economic Review* 98 (3) : 808-842.

2. Interprétation du coefficient de la régression : On a donc un impact positif et significatif du PIB sur l'indice de démocratie qui signifie que si on double le PIB moyen d'un pays sur cette période (hausse de 100%), cela augmente de 0.22 l'indice moyen du *Freedom House* sur cette même période.

3. Entre autres : Robert Barro. 1997. *Determinants of Economic Growth : A Cross-Country Empirical Study*. Cambridge, MA : MIT Press

4. Ceci nous est permis dans la mesure où ces auteurs ont mis à disposition de tous leurs données et leurs codes.

variables retardées (voir l'équation (3)), ils construisent un modèle dynamique qui nie, de fait, l'hypothèse d'exogénéité stricte, nécessaire à une estimation *Within* ou en différence première. Leur estimateur est par conséquent biaisé. Par ailleurs, ils négligent la forte présence de données manquantes qui, selon nous, est liée à un biais de sélection.

Nous proposons donc une autre approche sur ce même sujet. Nous commençons par mettre en évidence que l'absence d'information sur l'indice de démocratie concerne fréquemment des pays anciennement colonisés, avant leur date d'indépendance. Ceci nous porte donc vers un modèle de sélection endogène (Section 2) que l'on applique à l'année 1960 afin de voir l'impact de cette absence de données sur une année particulièrement touchée par ce problème.

Ensuite, une fois ce biais écarté (mais non résolu), nous estimons un modèle de panel à exogénéité faible (Section 3) sur l'ensemble de nos données - ce qui signifie que les variables explicatives sont certes corrélées aux chocs passés, mais on postule l'absence de corrélation avec les chocs futurs. Pour cela, nous revenons sur l'instrument utilisé dans l'article, nous en ajoutons un nouveau, et, contrairement à la méthode utilisée par Acemoglu *et alii*, nous appliquons la régression avec variable instrumentale sur l'ensemble des données de panel et non uniquement sur des moyennes. Précisons qu'outre ce problème d'exogénéité, les variables instrumentales permettront également de fournir une explication à la relation entre démocratie et richesse, en supprimant la possibilité d'une relation de causalité inverse entre ces deux variables.

1.3 Présentation des données

Les données rassemblées par Acemoglu *et alii* sont des données de panel couvrant la période 1950-2000, pour 7 variables, reportées tous les cinq ans (et non moyennisées sur cinq ans) pour 211 zones géographiques. La mesure de l'état de démocratie dans un pays nous est donné par le *Freedom House Political Rights Index* normalisé par Acemoglu *et alii* entre 0 et 1 où 1 correspond aux pays les plus démocratiques⁵.

Nous disposons également du logarithme du PIB par habitant en Parité de Pouvoir d'Achat pour les années susdites. Par ailleurs s'y ajoute un ensemble de variables qu'on l'on utilisera en variables de contrôle : la population (en log), l'éducation (représentée par la moyenne du nombre d'années de scolarisation), la structure par âge de la population, le taux d'épargne et un indice d'ouverture commerciale.

5. Les données proviennent de l'ONG "Freedom House" et de Bollen(2001) "Cross-National Indicators of Liberal Democracy 1950-1990" disponible auprès de l'ICPSR

2 Modèle de Sélection

2.1 Mise en lumière d'un biais de sélection

Lorsqu'on observe de plus près notre base, on remarque qu'il n'y a que 1514 observations qui disposent de l'indice de démocratie sur les 2321 présentes dans la base ($2321 = 211$ zones géographiques \times 11 dates, une date tous les 5 ans entre 1950 et 2000). Il y a plusieurs raisons à cette proportion de données manquantes qui sont plus ou moins dommageables pour notre réflexion. Voici la répartition des explications en table (annexe).

Après cette analyse, on remarque que 420 données manquantes ont été créées artificiellement. Par contre les autres pourraient biaiser l'analyse. En effet, lors des régressions où pour une observation certaines caractéristiques sont manquantes, Stata supprime la ligne. Or nous considérons que supprimer les observations des pays lorsqu'ils sont sous un régime de colonisation, a un impact direct sur la relation que nous cherchons à mettre en évidence.

Il faut noter tout d'abord que souvent les pays colonisés sont des pays au PIB proportionnellement plus faible. En effet, les pays colonisateurs n'ont pas forcément comme mission première d'apporter l'industrialisation et le progrès technique au pays colonisé alors que ce sont les bases de la croissance économique. Par ailleurs, ces pays étant soumis à un régime de colonisation, leur niveau de démocratie est nécessairement plus faible. Ainsi, il semble qu'un effet de sélection apparaisse.

2.2 Le modèle économétrique

Nos données se prêtent donc particulièrement bien à une analyse selon un modèle de sélection. Nous avons décidé de l'appliquer à l'année 1960 qui présente le plus de données manquantes liées aux pays colonisés à cette date. Nous avons aussi supprimé les zones géographiques ne faisant pas sens (les régions) et les données manquantes liées aux tout petit pays. Nous obtenons donc une base avec 109 pays et 83 pour lesquels nous avons l'indice de démocratie.

Précisons le modèle économétrique sous-jacent. L'équation que nous cherchons à estimer est la suivante :

$$dem_i = \beta_0 + \beta_1 y_i + \beta_2 lpop + \epsilon_i \quad (1)$$

avec dem_i l'indice du *Freedom House* pour le pays i en 1960, y_i le log du PIB par tête pour le pays i en 1960 et ϵ_i , le terme d'erreur qui capture le résidu non explicable par le modèle. Nous utilisons $lpop$, le logarithme de la population, comme variable de contrôle⁶.

En appliquant une estimation par MCO (donc sur 83 observations non censurées), on obtient les coefficients présentés dans la Table 1. Ils indiquent que lorsqu'on double le PIB, soit une hausse de 100%, on observe un impact positif et significatif au seuil de 1% (test de Student) sur l'indice de démocratie. Celui-ci augmenterait de 0,19 environ. La relation causale n'est donc pas négligeable ici.

TABLE 1 – Estimation par MCO (1960)

Variable	Coefficient	(Ecart type)
y	0.186**	(0.025)
$lpop$	-0.008	(0.015)
<hr/>		
N	83	
R ²	0.415	
F (2,80)	28.383	
<hr/>		
Niveau de significativité :	† : 10%	* : 5% ** : 1%

Cependant, comme souligné auparavant, nous observons dem_i uniquement pour les pays indépendants en 1960. Or, les pays colonisés sont structurellement et historiquement les pays plus pauvres et les moins développés. Nous avons donc une corrélation entre l'absence de données et les caractéristiques du pays.

Autrement dit, on a

$$\begin{cases} dem^* = D \times dem \\ dem^* \text{ qui dépend de } dem|y \end{cases}$$

6. Nous aurions souhaité en utiliser plus mais à partir du moment où la variable n'est pas disponible pour tous les pays, cela supprime l'observation entière. Sur 109 observations au total, on ne peut pas se le permettre.

avec dem^* l'indice de démocratie observé et D l'indicatrice d'être un pays libre en 1960.

Pour prendre en compte cet effet de sélection endogène, nous cherchons un instrument corrélé au fait d'être indépendant et qui n'affecterait pas la démocratie. Nous avons choisi d'utiliser la composition par âge de la population selon le raisonnement qui suit : les pays colonisés présentent en proportion une population plus jeune⁷ puisqu'on peut considérer que les colonies n'ont pas encore vécu leur transition démographique, la métropole ne cherchant pas à réduire la mortalité en développant les infrastructures sanitaires. Par ailleurs, on peut penser que la composition démographique n'a pas d'influence particulière sur le contexte de démocratie. Cette variable semble donc répondre aux critères nécessaires à un bon instrument.

Nous obtenons donc comme modèle économétrique :

$$\begin{cases} dem_i^* = \beta_0 + \beta_1 y_i + \beta_2 lpop_i + \epsilon_i \\ D = \mathbb{1}\{\alpha_0 + \alpha_1 age5_i + \alpha_2 y_i + \alpha_3 lpop_i + \nu_i \geq 0\} \end{cases} \quad (2)$$

avec $age5$ la proportion de la population âgée de moins de 15 ans. Pour que ce modèle soit identifiable, il nous faut supposer que $(\epsilon_i)_i$ et $(\nu_i)_i$ soient exogènes dans leur équation respective. Par ailleurs, il faut que ν suive une $\mathcal{N}(0,1)$ et que $\mathbb{E}(\epsilon|\nu) = \sigma_0^2 \nu$. Nous allons utiliser l'estimation proposée par Heckmann qui consiste en un probit sur D puis à insérer l'estimation de $\hat{\alpha}_1$ dans la première régression.

Les résultats obtenus sont les suivants :

TABLE 2 – Estimation par un modèle de sélection			
	Variable	Coefficient	(Ecart-type)
dem	y	0.187**	(0.035)
	$lpop$	-0.007	(0.022)
D	$age5$	20.5*	(10.3)
	y	0.354	(0.268)
	$lpop$	0.509**	0.133
	λ	0.0096	(0.143)
N		106	
Obs censurées		23	
Obs non censurées		83	
Wald Test		45.07**	
Niveau de significativité : † : 10% * : 5% ** : 1%			

Ainsi, même si l'instrument a un effet positif significatif sur la variable observée (test de Student avec un seuil de 5%), le modèle de sélection n'est pas validée. En effet, le λ , qui traduit l'effet de sélection, soit la corrélation entre ν et ϵ , indique que cette dernière n'est pas significative (la p-valeur du test de Student est de 0.947). Par ailleurs, on obtient le même effet positif significatif de y sur dem que dans la première estimation par MCO (cf. figure 1).

Il semblerait donc que la forte présence de données manquantes dans notre base n'a pas d'impact sur l'estimation de l'effet de la richesse sur la démocratie via un modèle de sélection. Nous pouvons donc nous atteler à présent à utiliser la profondeur temporelle de nos données afin de raisonner sur la présence de caractéristiques structurelles propres à chaque pays qui influencerait la relation.

7. 43% de la population des pays colonisés a moins de 50 ans en moyenne, alors que seuls 38% ont moins de 50 ans moyenne dans les pays libres.

3 Estimation d'un modèle de panel dynamique

3.1 Modèle économétrique

Voici le modèle économétrique de référence concernant notre analyse en termes de panel. Au cours de l'étude, nous reviendrons sur différentes hypothèses qui ont trait directement à cette équation.

$$dem_{it} = \gamma_0 + \gamma_1 y_{it-1} + \gamma_2 dem_{it-1} + W'_{it} \beta + \alpha_i + \epsilon_{it} \quad (3)$$

où dem_{it} est notre variable d'intérêt, soit la valeur du *Freedom House Political Rights Index* du pays i à la date t . Pour chercher à l'expliquer, nous incluons sa valeur retardée d'une période afin de capturer la persistance d'un effet de la démocratie ainsi que -selon Acemoglu *et alii*- la tendance de cet indice à revenir à sa moyenne, telle une valeur d'équilibre. Le modèle étudié est donc un modèle dynamique.

La variable explicative sur laquelle nous nous concentrons est y_{it-1} , la variable retardée d'une période du log du revenu par tête, dans une idée d'effet retardé du PIB sur la démocratie. Nous voulons donc que γ_1 reflète l'effet causal de l'économie sur la démocratie.

$(\alpha_i)_i$ est l'effet fixe lié au pays à comprendre comme des facteurs historiques structurels spécifiques à chaque pays qui influencent à la fois son développement économique et son développement politique sans varier dans le temps. (ϵ_{it}) est le terme d'erreur qui contient toutes les variables inobservées comme par exemple la culture pour la démocratie d'un pays, la structure des institutions politiques ou la nature des conflits politiques dans une société qui évoluent dans le temps.

Concernant W_{it} , il s'agit d'un vecteur-colonne de taille 16 contenant les variables de contrôle inclues dans l'équation dans le but de limiter au plus le biais de variables omises et d'avoir la meilleure spécification du modèle possible.

$$W_{it} = \begin{pmatrix} age_1 \\ | \\ age_4 \\ educ \\ pop \\ year_{1955} \\ | \\ year_{2000} \end{pmatrix}$$

Les quatre premières variables sont des variables indicatrices de la structure d'âge (avec age_1 la proportion de la population appartenant au cinquième de la population la plus jeune, age_5 étant la valeur de référence). Les deux suivantes sont des variables quantitatives concernant la taille et l'éducation de la population. Les 10 dernières sont des indicatrices concernant l'année : elles cherchent à isoler un effet pur de l'année pour capturer des chocs symétriques sur tous les pays s'étant produit à cette date. Dans la suite, nous noterons $X_{it} = (y_{i,t-1}, dem_{i,t-1}, W_{it})'$.

3.2 Modèle naïf : MCO empilés

La régression standard, menée dans les travaux de Barro par exemple, consiste en des Moindres Carrés Ordinaires (MCO) empilés. Elle s'appuie sur la même équation qu'en (3) sauf en ce qui concerne l'effet fixe α_i , qui est ignoré. Ayant mené cette opération, nous obtenons sur notre base, les résultats présentés dans la Table 3. Ainsi, ils indiquent que le revenu a un effet significatif à 5% sur la démocratie : si le revenu double (augmentation de 100%), l'indice de démocratie augmenterait de 0.047.

TABLE 3 – Estimation par MCO empilés

Variable	Coefficient	(Ecart-type)
$dem_{i,t-1}$	0.597**	(0.049)
$y_{i,t-1}$	0.047*	(0.020)
<hr/>		
N	644	
R ²	0.699	
F _(15,94)	192.677	
<hr/>		
Niveau de significativité :	† : 10%	* : 5% ** : 1%

Cependant, à partir du moment où l'on considère que la relation entre richesse et démocratie est régie par le modèle 3, on assiste à la faillite des MCO empilés. En effet, ces derniers ne prennent pas en compte l'existence des α_i qui sont de facto inclus dans le résidu $\epsilon_{i,t}$. Ceci entraîne donc un biais de variable omise puisqu'il y a de forte chance (cf. équation 3) que $\text{Cov}(y_{it-1}, \alpha_i) \neq 0$ et $\text{Cov}(dem_{it}, \alpha_i) \neq 0$, la présence d'effets fixes liés aux structures des pays influençant à la fois le PIB et la démocratie. C'est pour cela que nous décidons d'inclure dans notre estimation une réflexion sur les effets fixes.

3.3 Estimation sous hypothèse d'exogénéité faible

Le modèle étudié est un modèle dynamique, c'est à dire que la variable dem_{it} est expliquée par cette même variable retardée d'une période, dem_{it-1} . Il permet de distinguer l'hétérogénéité inobservée α_i de la dépendance d'état, effet de dem_{it-1} . Ainsi, par construction, l'exogénéité stricte n'est pas respectée (hypothèse classique non remise en cause chez Acemoglu *et alii*). En effet, on peut montrer par récurrence que, $\forall t \geq 2$:

$$\begin{aligned} dem_{i,t-1} = & \gamma_0 \left(\frac{1 - \gamma_2^{t-1}}{1 - \gamma_2} \right) + \gamma_1 \sum_{k=0}^{t-3} \gamma_2^k y_{i,t-2-k} + \gamma_2^{t-1} dem_{i,0} + \beta \sum_{k=0}^{t-2} \gamma_2^k W_{i,t-1-k} \\ & + \alpha_i \left(\frac{1 - \gamma_2^{t-1}}{1 - \gamma_2} \right) + \sum_{k=0}^{t-2} \gamma_2^k \epsilon_{i,t-1-k} \end{aligned}$$

Par conséquent, on a que $\forall t' \leq t$, $\mathbb{E}(dem_{it-1} \epsilon_{it'}) \neq 0$.

Ainsi, dans cette partie, nous supposons que seule l'hypothèse d'exogénéité faible est vérifiée ie :

$$\forall t' \geq t, \mathbb{E}(X_{it} \epsilon_{it'}) = 0.$$

Ceci sous-entend que les $(W_{it}, y_{it-1})_t$ sont faiblement exogènes et que les $(\epsilon_{it})_t$ ne sont pas corrélés entre eux ni à (α_i, dem_{i0}) . Il apparaît que sous cette hypothèse, les estimateurs *Within* et des différences premières sont biaisés. L'utilisation de variables instrumentales sur données de panel est donc nécessaire.

Premièrement, pour prendre en compte la présence d'effets fixes α_i et dans le but de simplifier les estimations en retirant cet effet fixe, nous utilisons la méthode des différences premières. Nous considérons donc le modèle suivant :

$$\Delta dem_{i,t} = \gamma_1 \Delta y_{i,t-1} + \gamma_2 \Delta dem_{i,t-1} + \Delta W'_{i,t} \beta + \Delta \epsilon_{i,t} \quad (4)$$

où, pour une variable $x_{i,t}$ quelconque, $\Delta x_{i,t} = x_{i,t} - x_{i,t-1}$.

Nous cherchons donc des variables instrumentales pour $\Delta y_{i,t-1}$. Il faut noter que pour qu'une variable z_{it} soit un bon instrument de $\Delta y_{i,t-1}$, elle doit vérifier les conditions suivantes :

- la condition de rang : $\mathbb{E}(\Delta y_{i,t-1} z_{it}) \neq 0$. Cette condition indique la pertinence de la variable instrumentale (*Relevance condition*).
- l'hypothèse d'exogénéité (*Exclusion restriction*) : $\mathbb{E}(z_{it} \Delta \epsilon_{it}) = 0$.

Nous disposons de deux instruments potentiels pour $\Delta y_{i,t-1}$.

Premièrement, le taux d'épargne nominal, *nsave*. Dans cette base, il est défini comme le revenu nominal auquel la consommation et les dépenses du gouvernement sont retirés et le tout est divisé par le revenu nominal. Il semble naturel que le taux d'épargne de la période précédente influence la variation du revenu global et ainsi la condition de rang serait vérifiée. De plus, il semble plausible que les changements de taux d'épargne n'aient pas d'effet direct sur la culture de la démocratie d'un pays, la structure des institutions politiques ou la nature des conflits politiques dans une société, variables qui pourraient être contenues dans le bruit $(\Delta \epsilon_{it})_{i,t}$. Ainsi, l'hypothèse d'exogénéité tiendrait.

Par ailleurs, le log du PIB par tête à la date $t-2$, y_{it-2} pourrait également vérifier les deux hypothèses mentionnées précédemment. En effet, il est clair que y_{it-2} influence la variation du PIB entre $t-1$ et $t-2$, ainsi la condition de rang serait respectée. De plus, l'hypothèse d'exogénéité semble possible, même si nous n'avons pas de théorie précise pouvant expliquer ce phénomène : y_{it-2} n'a pas d'effet direct sur les changements de conflits politiques ou la qualité des institutions du pays.

Les régressions avec chacun des instruments ont été effectuées grâce à une estimation en deux étapes et les résultats apparaissent dans la Table 4.

TABLE 4 – Comparaison des variables instrumentales

	VI y_{t-2}	VI $nsave_{t-2}$	2 VI
	(1)	(2)	(3)
y_{t-1} (2 ^{ème} étape)	0.029 (0.386)	1.847 (6.56)	-0.019 (0.372)
y_{t-2} (1 ^{ère} étape)	-0.05*** (0.011)	- -	-0.053*** (0.012)
$nsave_{t-2}$ (1 ^{ère} étape)	- -	-0.023 (0.661)	0.061 (0.05)
F-Test (1 ^{ère} étape)	10.93***	9.29***	10.29***
N	549	549	549

Sont présentés ici les résultats des régressions avec l'instrument y_{t-2} pour la première colonne (VI y_{t-2}), $nsave_{t-2}$ pour la seconde (VI $nsave_{t-2}$) et les deux instruments simultanément pour la 3^{ème} (2 VI). La 1^{ère} ligne correspond au coefficient associé à y_{t-1} dans la régression de seconde étape pour chacun des modèles. Les 2^{ème} et 3^{ème} lignes présentent les coefficients estimés associés à chacun des instruments dans la régression de première étape.

Les statistiques de Fisher, qui sont les statistiques de test de nullité simultanée de tous les coefficients, indiquent que les variables instrumentales présentent dans les régressions (1) et (3) satisfont bien la règle du pouce (F-test > 10). Ainsi, il semble que y_{t-2} ou la combinaison de y_{t-2} et de $nsave_{t-2}$ soient des bons instruments. De plus, le test de Student de nullité du coefficient associé à $nsave_{t-2}$ dans la régression de première étape montre qu'on ne peut pas rejeter l'hypothèse nulle. $nsave_{t-2}$ ne serait pas, à lui seul, un bon instrument pour Δy_{t-1} . Le choix entre les instruments présents dans les régressions (1) et (3) va finalement se porter sur les deux instruments combinés car les statistiques du F-test sont très proches et il est toujours préférable d'avoir plusieurs instruments plutôt qu'un afin de pouvoir tester l'exogénéité de ceux-ci.

Ainsi, nous allons procéder à un test de Sargan de suridentification du modèle. Il a pour hypothèse nulle :

$$\mathbf{H}_0 : \mathbb{E}(y_{t-2}\Delta\epsilon_t) = \mathbb{E}(nsave_{t-2}\Delta\epsilon_t) = 0$$

ie qu'il teste l'exogénéité des instruments. L'implémentation grâce à une estimation par Moindres Carrés en deux étapes (*Two-Stage Least Squares (2SLS)*) indique une p-valeur ≈ 0.57 c'est-à-dire que nous ne rejetons pas l'hypothèse nulle à 5% ou à 1%. Ainsi, les deux instruments semblent valides et notre modèle correctement spécifié.

3.4 Résultats et limites

Les instruments choisis pour Δy_{t-1} dans le modèle (4) sont donc $nsave_{t-2}$ et y_{t-2} . Les estimations de ce modèle sont présentées de manière plus précise dans la Table 5.

Les résultats indiquent que le PIB à la date $t-1$ n'a pas d'effet significatif sur le niveau de l'indice de démocratie, *Freedom House Political Rights Index*. Ainsi, il semble que prouver l'existence d'une relation causale entre le PIB retardé d'une date et le niveau de démocratie d'un pays entre les années 1950 et 2000 soit impossible.

Néanmoins, il faut noter que l'utilisation de l'instrument $nsave_{t-2}$ pourrait être remis en cause. En effet, il y a de nombreux canaux par lesquels le taux d'épargne nominal pourrait être corrélé au terme d'erreur de la régression de seconde étape.

TABLE 5 – Comparaison des estimation avec instruments

Variable	Coefficient	(Ecart-type)
<i>1^{ère} étape</i>		
Δdem_{t-1}	0.019	(0.0027)
$nsave_{t-2}$	0.061	(0.055)
y_{t-2}	-0.053***	(0.012)
N		549
R ²		0.225
F (15,533)		10.29
<i>2^{ème} étape</i>		
Δy_{t-1}	-0.019	(0.372)
Δdem_{t-1}	-0.250***	(0.053)
N		549
R ²		0.225
F (14,534)		4.63***
Niveau de significativité : † : 10% * : 5% ** : 1%		

Tout d'abord, comme le soulignent Acemoglu et alii, le taux d'épargne, lui-même pourrait être influencé par le type de régime politique du pays, plutôt libéral ou socialiste, ce qui romprait la condition d'exogénéité de cet instrument.

Par ailleurs, le taux d'épargne pourrait être corrélé avec les changements de distribution du revenu ou la composition des portefeuilles, ce qui pourrait avoir un effet direct sur l'équilibre politique. Et, à nouveau, la condition d'exogénéité serait violée.

Les résultats sont donc à nuancer et peuvent être biaisés par ce type d'argument.

4 Conclusions

Cette étude a tenté de mettre en lumière une relation causale entre le logarithme du PIB par tête d'un pays et son niveau de démocratie entre les années 1950 et 2000 via l'indice *Freedom House Political Rights Index*. Mais elle a montré que, bien que le revenu et la démocratie soient corrélés, il n'y a pas de véritable preuve de relation causale entre les deux dans l'époque après-guerre.

L'étude du grand nombre de données manquantes a d'abord permis de mettre en évidence un potentiel effet de sélection lié à la colonisation du pays. Néanmoins, les estimations n'ont pas confirmé l'hypothèse du biais de sélection et ont, par ailleurs, montré un effet non significatif du PIB par tête sur la démocratie des pays en 1960. Cette approche est d'ailleurs totalement novatrice par rapport à l'article de Acemoglu *et alii* qui ne se soucie pas de ce grand nombre de données manquantes dans la base, ce qui biaise, en partie, leurs résultats. De plus, l'utilisation des données de manière temporelle, grâce à une étude en panel, a permis de confirmer que l'effet du PIB par tête sur le niveau de démocratie n'est guère significatif entre 1950 et 2000. La relation causale est, à nouveau, inexistante.

Néanmoins, des précautions sont à prendre dans l'interprétation de nos résultats. Tout d'abord, il est clair que ces résultats peuvent manquer de robustesse face au grand nombre de valeurs manquantes, au potentiel biais de variables omises dû au faible nombre de variables disponibles ou au problème de la simultanéité, malgré l'utilisation de variables instrumentales.

De plus, bien que nos résultats ne fournissent pas de preuve quant à l'existence d'une relation causale entre PIB par tête et démocratie, un tel effet peut exister mais à des fréquences plus faibles, par exemple sur un horizon de 100 ans ou plus (c'est d'ailleurs l'approche qu'ajoutent Acemoglu *et alii* dans leur article) ou bien cet effet peut être lié à d'autres caractéristiques qui n'ont pas été mises en évidence ici.

Enfin, il faut bien noter que nos résultats ne prouvent pas que la démocratie n'a pas d'effet sur la croissance économique (cf. Torsten Persson et Guido Tabellini, 2007). Ceci ferait l'objet d'une nouvelle étude.

Bibliographie

- [1] Daron ACEMOGLU, Simon JOHNSON, James A. ROBINSON et Pierre YARED, *Income and Democracy*, American Economic Review 2008, 98 :3, 808-842
- [2] Michael VISSER, *Econometrics 1*, ENSAE ParisTech, 2015
- [3] Xavier D'HAULTEFŒUILLE, *Econométrie 2*, 2015

Annexe : Distribution des données manquantes

Causes des NA	Régions ¹	Petit pays ²	Pays colonisés ² (date d'indépendance)	Partition ³	Inconnu ⁴
Pays concernés	Modena Etat libre d'Orange Parme Sardaigne Saxe Two Sicilies Toscane United Province CA Wurttemberg	Liechtenstein Etat papal Antigua Andorra	Grenade(1974) Guyane(1966) Indonésie(1945) Inde(1947) Israel(1948) Jamaïque (1962) Kenya, Cambodge (1953), Kiribati (1979) St. Kitts and Nevis(1983) Koweït (1961) Sainte Lucie (1979) Lesotho (1966) Maroc (1955) Madagascar (1960) Maldives (1965) Mali(1960) Malte (1964) Mozambique (1975),Mauritanie (1960) Maurice (1968) Malawi (1964) Malaisie (1957) Namibie (1990), Niger (1960) Nigeria (1960), Oman (1971) Papua New Guinea(1975) Qatar (1971) Rwanda Soudan Sénégal(1960) Singapour (1965) Salomon (1978) Sierra Leone, Somalia,Sao Tome and Principe, Suriname (1975) Swaziland (1968) Séchelles (1976) Chad, Togo Tonga Trinidad and Tobago Tunisie (1957) Tanzania Uganda St. Vincent and the Grenadines(1976), Vanuatu Western Samoa, Congo (1958), Dem. Rep. Zambia, Zimbabwe (1965), Angola, Burundi (1950) Benin,Burkina Faso, Botswana Central African Republic, Cote d'Ivoire, Cameroun, Congo, Rep. Belize(indep 1981), Brunei(1984) Comoros (1975)Cape Verde(1975), United Arab, Emirates (1971), Bangladesh(1971) Bahrain (1971) Bahamas (1971), Barbados(1966) Bhoutan(1949) ,Dominica (1978) Algérie(1961), Eritrea(1993) East Timor (2002) Fiji(1970), Gabon,Ghana Guinée, (1958), Gambie(1965) Guinée-Bissau, Guinée équatoriale (1958)	Corée Pakistan Slovaquie Vietnam Yemen République tchèque Allemagne Ethiopie Kazakhstan Kirghizstan Lituanie Lettonie Moldavie Russie Tadjikistan Turkmenistan Ukraine URSS Ouzbékistan Azerbaïdjan Estonie Arménie Belarus Géorgie Croatie Macédoine Serbie Slovénie Bosnie and Herzégovine	Japon : 1950 Laos : 1950-1955 Libye : " Sri Lanka : 1950 Mexique : 1955 Myanmar : 1950 & 1965 Corée du Nord : 1950-1955 Sénégal : 1965 Thaïlande : 1950 Taïwan: 1950-1955 Vietnam : 1950-1955 Dominican République : 1950 Égypte: 1950 Grèce Colombie Colombie : 1955 Chine: 1950
Nombre d'observations concernées (Total : 805)	99	34	364	287	21
<p>1 Ce sont des régions ajoutées à la base dont il manque les observations</p> <p>2 Ce sont des données manquantes associées à des micro-états.</p> <p>3 Ce sont des données manquantes liées au fait que les États étaient sous domination coloniale aux dates concernées.</p> <p>4 Ce sont des données manquantes créées artificiellement du fait de la partition pendant la période.</p> <p>5 Ce sont des données manquantes sans raison évidentes.</p>					

Annexe : Codes Stata

```
ssc install outtex, replace
ssc install estout, replace

clear
import excel using /Users/noemiehaouzi/Desktop/ENSAE2A/S2/Econometrie2/
    ProjetEcono2/Base_econo.xlsx, sheet(Obs) firstrow

/*-----GENERATION DES VARIABLES-----*/
///Gnration des variables principales
tsset code_numeric year_numeric
sort code_numeric year_numeric
tab year, gen (yr)
tab code, gen(cd)

gen dem = fhpolrigaug
gen y = lrgdpch
destring dem y, dpcomma replace

// Variables de contrles
gen age1 = age_veryyoung
gen age2 = age_young
gen age3 = age_midage
gen age4 = age_old
gen age5 = age_veryold
gen educ = education

gen l_dem = L.dem
gen l_y =L.y
foreach var of varlist l_dem l_y dem yr* {
    bysort code : egen mx = mean ('var')
    gen w'var' = 'var' - mx
    drop mx
}

sort code_numeric year_numeric
gen d_ldem=d.l_dem
gen d_dem=d.dem
gen d_ly=d.l_y

sort code_numeric year_numeric
gen l2_y =l2.y
gen dl2_y = d.l2_y
foreach var of varlist age1 age2 age3 age4 age5 educ lpop yr*{
    gen d'var' = d.'var'
}

gen l2_nsave = l2.nsave
gen dl2_nsave = d.l2_nsave

gen lworldinc = l.worldincome
gen l2worldinc = l2.worldincome
gen dlworldinc =d.lworldinc
gen dl2worldinc =d.l2worldinc
```

```

xtset code_numeric year_numeric

/*-----GRAPHES-----*/
//Graph de l'effet brut
egen moy_dem = mean(dem), by(code)
egen moy_y = mean(y), by(code)

reg moy_dem moy_y
tway (scatter moy_dem moy_y, msymbol(none)
mlabel(code) mlabsize(tiny))
(lfit moy_dem moy_y, clcolor(black)), ytitle("Indice de dmocratie")
xttitle("Log PIB par tte (donnes Madison)")
title("Effet brut du PIB par tte sur la dmocratie")
subtitle("Analyse transversale (cross-section) de 1950 2000")
legend(off) xscale(r(6 10.6))

reg dem y
tway (scatter dem y, msymbol(none) mlabel(code) mlabsize(tiny)) (lfit dem y,
clcolor(black)),
ytitle("Indice de dmocratie") xttitle("Log PIB par tte (donnes Madison)")
title("Effet brut du PIB par tte sur la dmocratie")
subtitle("Analyse transversale de 1950 2000") legend(off) xscale(r(6 10.6))

histogram dem, title("Distribution du niveau de dmocratie") ytitle("")
xttitle("")
subtitle("Entre 1950 et 2000") color(gray)

/*-----POOLED OLS-----*/
eststo: reg dem l_dem l_y yr* age1 age2 age3 age4 age5 educ lpop,
cluster(code)
label variable dem "démocratie(t)"
label variable l_dem "démocratie(t-1)"
label variable l_y "pib(t-1)"
outttx, labels level plain detail legend title("Estimation par MCO empils")

/*----- PANER AVEC VARIABLES INSTRUMENTALES POUR Y(t-1)-----*/

//////////Instrument : Yt
//Instrument = y(t-2) F=10,93
eststo: ivreg d_dem (d_ly = l2_y) d_ldem yr* age1 age2 age3 age4 age5 deduc
dlpop, first robust

//////////Instrument : savings
//instrument nsave(t-2) F=9,29
eststo: ivreg d_dem (d_ly = l2_nsave) d_ldem yr* age1 age2 age3 age4 age5
deduc dlpop, first robust

////////// Plusieurs instruments
//Instruments : nsave(t-2) et y(t-2) F=10,29
eststo : ivreg d_dem (d_ly = l2_nsave l2_y) d_ldem yr* age1 age2 age3 age4
age5 deduc dlpop, first robust

```

```

esttab          using "/Users/noemiehaouzi/Desktop/ENSAE2A/S2/Econometrie2/ProjetEcono2/
tableau.tex" , se r2 title(Comparaison des modeles\label{tabcomp})
nonumbers  nostar replace booktabs mtitle("VI 1" "VI 2" "VI 3")

//SARGAN TEST
ivregress 2sls d_dem (d_ly = l2_nsave l2_y) d_ldem yr*      age1 age2 age3 age4
              age5 deduc dpop, robust
estat overid

/***** MODELE DE SELECTION *****/
clear
import excel using /Users/noemiehaouzi/Desktop/ENSAE2A/S2/Econometrie2/
              ProjetEcono2/Base_econoV2.xlsx, sheet(1960) firstrow

gen dem = fhpolrigaug
gen y = lrgdpch
destring dem y, dpcomma replace

// Variables de contrles
gen age1 = age_veryyoung
gen age2 = age_young
gen age3 = age_midage
gen age4 = age_old
gen age5 = age_veryold
gen educ = education

//regression de comparaison
reg dem y lpop
outtex, labels level plain detail legend title("Estimation par MCO")

//>> biai du 2SLS car tous les pays ne sont pas pris en compte
// Donc modle de slection

xi : heckman dem y lpop, select(obs_dem = age5 lpop y) twostep
outtex, labels level plain detail legend title("Estimation par Heckit")

/* instrument moins significatif*/
xi : heckman dem y, select(obs_dem = age1 y) twostep
xi : heckman dem y, select(obs_dem = age1 age5 y) twostep

```