

DONNEES DE PANEL

Le taux de pauvreté des pays d'Amérique latine

BARRY Mamadou Yaya

Réalisé par : TANNANI Manal
ZYATE Yassine

Professeur : COMPAIRE Philippe

SOMMAIRE

Introduction	1
I. Étude Économique	2
II. Étude Économétrique	4
II.1. Présentation des données	4
II.2. Présentation des variables	5
II.3. Spécification du modèle	6
II.4. Statistiques descriptives	6
II.5. Modèle à conserver (sous TSP).....	8
II.6. Estimation par la méthode <i>within</i> (sous STATA)	12
II.6.1. Modèle à effet fixe individuel (One-way FE).....	12
II.6.2. Modèle within avec effet temporel	13
II.6.3. Autocorrélation intra-individus (Test de Wald modifié)	14
II.6.4. Test d'hétéroscédasticité inter-individus (Test de Wald modifié)..	15
II.7. Interprétations économiques du modèle final	17
Conclusion.....	19
Bibliographie	20

Introduction

Les économètres, par le biais des méthodes économétriques, sont toujours à la recherche de relations de causes à effets entre différents phénomènes (économiques, sociales, démographiques...). Et notamment pour tester la pertinence empirique d'un modèle.

C'est dans cette optique que nous introduirons avec vous "les modèles de panel" pour une étude de cas concernant "le taux de pauvreté des pays d'Amérique latine".

Les données de panel sont constituées d'observations périodiques pour un même *individu statistique* (populations, informations économiques et sociales relatives aux ménages, entreprises, pays, régions, etc.).

Elles sont issues d'enquêtes ou de fichiers administratifs ou de gestion actualisés périodiquement, de données agrégées régionales ou nationales. Elles visent à quantifier les comportements socioéconomiques tant dans leurs différences individuelles que dans leurs propriétés dynamiques tout en apportant une vision chronologique des trajectoires de chaque unité d'observation.

Les modèles économétriques destinés à l'analyse des données de panel intègrent une composante permanente spécifique aux individus observés, qui capture l'hétérogénéité des situations particulières de chaque unité.

Dans un premier temps, nous présenterons l'étude économique, avant d'aborder dans un second temps le modèle économétrique de choix que nous aurons au préalable défini et estimé à travers différentes méthodes de régression qui nous permettront donc d'analyser l'existence d'un ou de plusieurs liens entre nos différentes variables. Et en dernier lieu, dans une conclusion, nous effectuerons un rapprochement entre l'étude réalisée et les littératures existantes dans le domaine.

I. Étude Économique

Les pays d'Amérique latine sont historiquement frappés par la pauvreté, le chômage et le sous-emploi, et par des inégalités considérables dans la distribution des revenus et du patrimoine des personnes. Cependant, la situation socioéconomique de la région s'est améliorée dans la foulée de l'expansion économique des années 2004-2007, mais on ignore quelle a été son évolution pendant le ralentissement économique de 2008 et la grave récession de 2009.

De là, est née l'idée d'étudier le taux de pauvreté pour la proportion d'individus dont la parité de pouvoir d'achat (PPP) au quotidien est de 3,20\$.

La pauvreté est un concept, dépendant non seulement du revenu monétaire, mais également des indicateurs d'accès à plusieurs services de base traduisant le bien-être de la population. La pauvreté est donc un état de vie multidimensionnel, relatif à l'accès à l'éducation, eau potable, services de santé basiques etc.

Dans ce cadre, nous discutons essentiellement le sujet de développement économique et social qui détermine les seuils de pauvreté de chaque pays. La réalisation d'un projet économique cohérent, adéquat à la situation économique et sociale du pays est le moyen le plus optimal pour atteindre un scénario de lutte contre pauvreté et redistribution équitable.

Cette composition riche amplifie également la grandeur de la problématique de pauvreté vers un autre aspect inégalitaire, qui caractérise la région latine par suite des revenus monétaires majoritairement. L'indice de développement humain des Nations Unies précise plusieurs facteurs à considérer.

La région d'Amérique Latine a profité de Momentum intéressant durant la première décennie du 21^{ème} siècle. Une translation économique a été en ordre suivant des crises politiques et des offres de coopération internationales pour les dépasser. Ces circonstances ont aidé à réaliser un profit en termes de PIB important et surpassant la norme, grâce à la richesse en matière première de la région. Cependant, cette situation a créé une corrélation de l'économie aux prix dans les marchés internationaux.

Les travaux de ces pays concernant l'économie ont été focalisé sur les régions rurales avec une variété industrielle (Hydrocarbures, métallurgie, agroalimentaire,). Le renforcement du marché de travail rural a pris un tournant positif, malheureusement non déclaré statistiquement. Toutefois, cela n'ignore point l'avancement de la région urbaine qui aussi évolué positivement,

surtout que nous parlons d'une réduction de plus du quart en cinq ans. Ce changement n'a pas contribué à la diminution de pauvreté, qui est dépendante de l'augmentation des revenus. Chose non réalisée à cause du manque d'intervention administrative, le facteur le plus important pour l'achever, du moins plus que la régulation du marché de travail.

Les revenus bas d'une population fortement active, caractéristique de la région qui attire les Investissements directs étrangers, et source de pauvreté, sont également le noyau d'inégalité surtout en accès aux services tels que la santé.

L'essai de positionnement international a également créé une issue d'inflation grave qui a été réglée par des politiques industrielles sélectives de crédits et subventions et des nouvelles formes d'intervention du côté social. Le protectionnisme a été donc adapté pour garder le pouvoir compétitif et une appréciation acceptable de la monnaie.

L'objectif principal est de voir quels sont les facteurs qui contribuent à l'expansion économique tout en réduisant le niveau de pauvreté que connaissent ces populations. L'heure est au questionnement, notamment avec la transformation, ces dernières années, du scénario politique latino-américain par un double et très important phénomène : l'essor de régimes politiques « de gauche » ou « progressistes », et la multiplication de mouvements populaires contestataires (des régimes de « gauche » s'opposant au néolibéralisme ambiant, voire prônant des options plus radicales) avec une base environnementale, ethnique et sociale¹.

Nous verrons également, à travers notre analyse, les éléments de contradictions qui confirment ou infirment certaines réalités telles que la gabegie financière entretenue par une corruption qui gangrène ces états, comme par ailleurs le cas des pays africains qui vivent le même cauchemar.

¹ Cocaleros (Bolivie et Pérou), collectifs antimondialisation et le libre commerce, communautés victimes de l'exploitation minière et d'hydrocarbures (Pérou), piqueteros (Argentine), paysans sans terre (Brésil), zapatistes (Mexique). Très souvent ce sont des mouvements citoyens, en quête de droits, où l'on trouve des militants écologistes, minorités (ou majoritaires, selon les pays) ethniques, étudiants, syndicalistes, défenseurs du droit à la différence, la mouvance de la théologie de la libération et anciens militants de la gauche classique désireux de se recycler.

II. Étude Économétrique

Un modèle économique est une représentation simplifiée de la réalité économique ou d'une partie de celle-ci : par exemple la croissance, le commerce international, la monnaie, une entreprise ou un ménage.

Afin de mener à bien notre étude, nous pouvons établir notre modèle économique (qui nous permettra de décrire les liens divers entre les variables) comme suit :

$$Y = f(X_1, X_2, X_3, \dots, X_n)$$

Notre variable d'intérêt (le taux de pauvreté) sera fonction de plusieurs variables explicatives issues d'une base de données établie à partir des données de la banque mondiale :

$$\text{POV_RATE} = \beta_1 X_{1,i} + \beta_2 X_{2,i} + \dots + \beta_{P-1} X_{P-1,i} + \beta_P + u_i, i \in \{1, 2, \dots, N\}$$

Avec :

- POV_RATE = Le taux de pauvreté (à 3,20\$ PPP) qui est la variable expliquée.
- $\beta_1, \beta_2, \dots, \beta_{P-1}, \beta_P$ les paramètres à estimer.
- $X_{1,i} ; X_{2,i} ; \dots ; X_{P-1,i}$ = nombres de variables explicatives.

II.1. Présentation des données

Les données utilisées dans cette étude proviennent du site de la Banque mondiale (World Bank Data²). Données à partir desquelles, nous avons réalisé une base de données de quatre cent quatre-vingt-huit (480) observations après traitement des données brutes.

Elle est constituée de 23 variables toutes quantitatives (24 pays d'Amérique latine sur 20 années de 2000 à 2019) que nous avons réorganisées, regroupées puis transformées en fonction des pays afin d'avoir une base de données complète sur le taux de pauvreté.

Cependant, suite à la matrice V de cramer pour observer les relations de corrélations entre les variables, nous avons retenu onze (11) variables dont une expliquée et dix (10) autres qui sont les explicatives.

² <https://donnees.banquemondiale.org/>

II.2. Présentation des variables

Dans le cadre de notre étude, le choix des variables repose sur les différentes revues empiriques portant sur les déterminants de la pauvreté, surtout dans les pays d'Amérique latine, que nous avons pu consulter durant la réalisation de ce travail.

Ces déterminants sont divers et variés. Entre autres, nous pouvons citer les variables que nous avons retenues pour cette étude dans le tableau suivant :

Noms	Description	Sigles
Taux de pauvreté	La proportion des populations du seuil de 3,20\$ par jour.	<i>Pov_rate</i>
Taux d'emploi	Taux d'insertion dans l'activité économique.	<i>emp_rate</i>
Indice GINI	Une mesure statistique permettant de mesurer le niveau d'inégalité de la répartition d'une variable dans la population.	<i>gini</i>
Dépenses courantes de santé	Niveau des dépenses de santé courantes exprimé en pourcentage du PIB.	<i>curr_health_exp</i>
Crédit intérieur fourni au secteur privé	Désigne les ressources financières fournies au secteur privé par les sociétés financières.	<i>crdpriv_gdp</i>
Industrie, valeur ajoutée	La valeur ajoutée des activités industrielles.	<i>indusva_gdp</i>
Agriculture, valeur ajoutée	La valeur ajoutée des activités agricoles.	<i>agrva_gdp</i>
Prévalence de la sous-alimentation	Indique le pourcentage de la population dont l'apport alimentaire est insuffisant pour satisfaire les besoins en énergie alimentaire de façon continue.	<i>undernour_prev</i>
Services, valeur ajoutée	La valeur ajoutée de secteur de services	<i>servva_gdp</i>
Chômage	La population active sans emploi	<i>unemprate_tot</i>
Taux de participation à la population active	Toutes les personnes qui fournissent du travail pour la production de biens et de services au cours d'une période donnée.	<i>Labf_tot3</i>

II.3. Spécification du modèle

Après établissement du modèle économique, il conviendra de le transformer en modèle économétrique. Un modèle économétrique est un ensemble d'équations permettant de déterminer les valeurs prises par un groupe de variables (variables-résultats ou endogènes) en fonction des valeurs attribuées à un groupe de variables (variables-causes ou exogènes).

Cette forme nous permettra non seulement d'étudier les variables qui nous ont été imposées par la théorie pour cette étude, mais également d'observer les variables qui n'ont pas été prises en compte ; c'est-à-dire qu'il y a de nombreuses autres variables qui peuvent avoir un effet sur le taux de pauvreté et que l'on ne peut pas toutes les lister ni observer.

Notre modèle économétrique se présente comme suit :

$$\text{POV_RATE} = \text{Cste} + \beta_1 \text{EMP_RATE} + \beta_2 \text{GINI} + \dots + \beta_{10} \text{LABF_TOT3} + \varepsilon_t$$

Le but ici, est de faire l'analyse statistique tout en décrivant la structure des données, c'est-à-dire définir la composition des variables, la significativité des variables, les nuages de points et les corrélations. Dans un premier temps nous ferons une analyse des statistiques descriptives de l'ensemble des variables (moyenne, médiane, maximum, minimum, nombre d'observations), ensuite nous ferons les tests nécessaires pour notre étude afin de déterminer le modèle correspondant.

II.4. Statistiques descriptives

Les statistiques descriptives des données obtenues avec les logiciels SAS, R et celles obtenues avec les logiciels TSP, STATA et Gretl sont identiques. Ce pendant avec STATA, on a plus de modèles par rapport aux autres.

Variable		Mean	Std. Dev.	Min	Max	Observations	
Id	overall	12.5	6.929408	1	24	N =	480
	between		7.071068	1	24	n =	24
	within		0	12.5	12.5	T =	20
year	overall	2009.5	5.772297	2000	2019	N =	480
	between		0	2009.5	2009.5	n =	24
	within		5.772297	2000	2019	T =	20
Pov_rate	overall	14.37957	1.428051	11.51293	17.56971	N =	262
	between		1.436163	11.65155	17.03058	n =	20
	within		.4276385	12.75786	15.80401	T =	13.1
emp_rate	overall	15.23363	1.308522	12.3737	18.35489	N =	480
	between		1.329402	12.44373	18.26	n =	24
	within		.1225633	14.93277	15.56689	T =	20
gini	overall	49.369	4.718388	38	61.6	N =	271
	between		3.646712	41.1	54.63529	n =	20
	within		3.240602	40.47489	59.87489	T =	13.55
unempr~t	overall	7.093125	3.939476	1.6	20.5	N =	480
	between		3.559041	2.535	13.655	n =	24
	within		1.831704	1.638125	16.34813	T =	20
labf_t~3	overall	67.28069	5.58749	49.55	81.712	N =	480
	between		5.40748	52.1666	78.43065	n =	24
	within		1.771744	58.19903	72.80954	T =	20
undern~v	overall	12.77488	11.05998	2.5	57.6	N =	414
	between		10.77466	2.583333	52.69444	n =	23
	within		3.317913	3.152657	23.64155	T =	18
curr_h~p	overall	6.675482	1.693149	1.18121	12.81401	N =	414
	between		1.432149	4.795879	9.823821	n =	23
	within		.9487507	1.499051	9.775282	T =	18
indusv~p	overall	30.49864	9.698249	16.66919	66.20369	N =	452
	between		9.105276	20.04705	50.21059	n =	24
	within		4.01663	11.32299	50.12985	T =	18.8333
crdpri~p	overall	37.84285	21.4383	8.771811	116.6492	N =	410
	between		19.99196	14.24588	93.40944	n =	22
	within		8.757754	8.131751	77.38166	T =	18.6364
servva~p	overall	53.98002	10.17113	13.49093	73.4483	N =	452
	between		9.444012	29.12491	71.39218	n =	24
	within		4.28304	38.34604	87.60525	T =	18.8333
agrva~p	overall	8.149419	5.610129	.3655144	32.11281	N =	452
	between		5.363923	.6901588	21.65952	n =	24
	within		1.927954	-3.269959	18.9026	T =	18.8333

En réalisant une statistique descriptive sur les variables, nous pouvons observer que nous avons effectivement 480 observations avec 24 pays sur une période de 20 ans (de 2000 à 2019). Cependant, suite à la présence des données manquantes (d'autres variables sont mieux renseignées que d'autres) dans notre base, *le panel est dit non cylindré (unbalanced)* : donc présence de bruit blanc. Malgré cela, le nombre total de nos données reste supérieur à 3000 ($480 \times 11 = 5280 - n \times NA > 3000$).

II.5. Modèle à conserver (sous TSP)

A l'issue de cette procédure (utilisant la commande *panel* de TSP), on dispose des réalisations des estimateurs *Pooled*, *Between*, des estimateurs du modèle à effets individuels fixes (*Within*), du modèle à effets individuels aléatoires (*Error Component Model*), des résultats de trois tests de Fischer, d'un estimateur de la variance des effets individuels, d'un estimateur de la variance totale, de l'estimateur d'un paramètre de pondération et de la statistique du test d'Hausman. Voilà ainsi résumés tous les éléments que nous nous proposons d'étudier tout au long de notre étude.

Méthode Pooled ou des Moindres Carrés Ordinaires (MCO)

```
PANEL DATA ESTIMATION
=====

Unbalanced data: N=    20, TMIN=    1 TMAX=   18, NOB=   249
Ahrens-Pincus Unbalancedness measure APU1 = .48988

Plain OLS (TOTAL)
=====

Dependent variable: POV_RATE

      Mean of dep. var. = 14.3945
      Std. dev. of dep. var. = 1.42804
      Sum of squared residuals = 47.2212
      Variance of residuals = .198409
      Std. error of regression = .445431
      R-squared = .906630
      Adjusted R-squared = .902707
      LM het. test = 23.5556 ** [.000]
      Durbin-Watson = .210538 ** [.000,.000]
      Schwarz B.I.C. = 176.667
      Akaike Information Crit. = 157.321
      Log likelihood = -146.321

Variable      Estimated      Standard
Coefficient      Error      t-statistic      P-value
EMP_RATE      1.12430      .033785      33.2783      ** [.000]
GINI           .082052      .814728E-02      10.0711      ** [.000]
CURR_HEALTH_EXP .063027      .027213      2.31608      * [.021]
CRDPRIV_GDP    -.963418E-02 .180787E-02      -5.32901      ** [.000]
INDUSVA_GDP    .048853      .823142E-02      5.93494      ** [.000]
AGRVA_GDP      .056031      .014846      3.77412      ** [.000]
UNDERNOUR_PREV .051113      .537711E-02      9.50574      ** [.000]
SERVVA_GDP     .038904      .868194E-02      4.48108      ** [.000]
UNEMP_RATE_TOT -.491394E-02 .987531E-02      -.497599      [.619]
LABF_TOT3      -.022301      .686893E-02      -3.24665      ** [.001]
C              -10.2888      1.17705      -8.74114      ** [.000]

F test of A,B=Ai,Bi: F(165,73) = 21.759, P-value = [.0000]
Critical F value for diffuse prior (Leamer, p.114) = 16.685
```

Le OLS nous indique que toutes les variables à l'exception de *unemprate_tot* sont significatives à 5%.

Toutes les variables significatives, sauf ces deux (*labf_tot3* et *crdpriv_gdp*) présentent *ceteris paribus* (toutes choses étant égales par ailleurs) des effets marginaux positifs sur le taux de pauvreté.

Le R^2 et le R^2 -ajusté sont à plus de 90% : ce qui signifie que plus de 90% de la distribution (variation du taux de pauvreté) est pris en charge par le modèle. Cependant le modèle rencontre un problème d'hétéroscédasticité des coefficients.

La *p-value* associée à ce test est très largement inférieure au seuil de 5%. Donc pour ce test, on rejette l'hypothèse nulle Ho^1 d'homogénéité globale : il convient alors de tester l'hypothèse Ho^2 des coefficients (associés aux variables explicatives) entre les pays.

Estimation par le modèle Between

OLS on individual means (BETWEEN)
=====

Dependent variable: POV_RATE

Mean of dep. var. = 14.3925 Std. error of regression = .466710
Std. dev. of dep. var. = 1.43948 R-squared = .950207
Sum of squared residuals = 1.96037 Adjusted R-squared = .894881
Variance of residuals = .217818 LM het. test = 2.30402 [.129]

Variable	Estimated Coefficient	Standard Error	t-statistic		P-value
EMP_RATE	1.04596	.141422	7.39602	**	[.000]
GINI	.106148	.043184	2.45803	*	[.036]
CURR_HEALTH_EXP	.089785	.113210	.793079		[.448]
CRDPRIV_GDP	-.015931	.723717E-02	-2.20121		[.055]
INDUSVA_GDP	.032300	.034305	.941544		[.371]
AGRVA_GDP	.350864E-02	.062483	.056153		[.956]
UNDERNOUR_PREV	.059510	.020448	2.91033	*	[.017]
SERVVA_GDP	.031280	.041262	.758076		[.468]
UNEMPRATE_TOT	-.019522	.046429	-.420483		[.684]
LABF_TOT3	-.014732	.029676	-.496428		[.631]
C	-9.40995	5.58220	-1.68571		[.126]

Le modèle *between*, malgré une bonne variabilité du modèle, reste un mauvais modèle car la plupart des variables ne sont pas représentatives. Il s'emble s'éloigné de l'objectif souligné par le OLS.

Estimation par le modèle à effets fixes - Within

```

Fixed Effects - Individual (WITHIN)
=====

Dependent variable: POV_RATE

      Mean of dep. var. = 14.3945
      Std. dev. of dep. var. = 1.42804
      Sum of squared residuals = 9.38640
      Variance of residuals = .042860
      Std. error of regression = .207027
      R-squared = .981440
      Adjusted R-squared = .978983
      LM het. test = 9.62964 ** [.002]
      Durbin-Watson = .913840 ** [.000,.000]
      Schwarz B.I.C. = 27.9427
      Akaike Information Crit. = -24.8191
      Log likelihood = 54.8191

      Variable      Estimated      Standard
      Coefficient      Error      t-statistic      P-value
EMP_RATE      .386718      .287742      1.34397      [.180]
GINI      .054479      .791465E-02      6.88328      ** [.000]
CURR_HEALTH_EXP      -.064081      .024071      -2.66218      ** [.008]
CRDPRIV_GDP      -.011756      .204876E-02      -5.73818      ** [.000]
INDUSVA_GDP      .046540      .010443      4.45656      ** [.000]
AGRVA_GDP      .049272      .013988      3.52241      ** [.001]
UNDERNOUR_PREV      .012557      .590140E-02      2.12779      * [.034]
SERVVA_GDP      .024663      .928471E-02      2.65628      ** [.008]
UNEMP_RATE_TOT      .057424      .934146E-02      6.14722      ** [.000]
LABF_TOT3      -.042699      .010110      -4.22360      ** [.000]

      F test of Ai,B=Ai,Bi: F(146,73) = 4.4873, P-value = [.0000]
      Critical F value for diffuse prior (Leamer, p.114) = 12.205

      F test of A,B=Ai,B: F(19,219) = 46.460, P-value = [.0000]
      Critical F value for diffuse prior (Leamer, p.114) = 6.0340
  
```

Les p -value du H_o^2 et H_o^3 (F test of $A_i, B = A_i, B_i$ et F test of $A, B = A_i, B$) d'égalité des coefficients et de constantes individuelles sont très largement inférieurs à 5%. Les hypothèses sont rejetées. Il est donc nécessaire d'introduire des effets individuels.

Nos variables sont toutes significatives à 5%. Seule exception pour le taux d'emploi (*emp_rate*).

NB : le modèle *within* ne comporte pas de constante (à supprimer sur les autres logiciels tels que STATA).

Estimation par le modèle à effets aléatoires – MCG

```

Random Effects - Individual - GLS
=====

VWITH (variance of Uit)   =  0.42860E-01
VBET  (variance of Ai)    =  0.15555
(computed from small sample formula)
THETA (0=WITHIN, 1=TOTAL) =  0.15077E-01
(evaluated at TMAX =   18)

Dependent variable: POV_RATE

      Mean of dep. var. = 14.3945
      Std. dev. of dep. var. = 1.42804
      Sum of squared residuals = 81.6595
      Variance of residuals = .343107
      Std. error of regression = .585753
      R-squared = .840399
      Adjusted R-squared = .833694
      LM het. test = 43.0301 ** [.000]
      Durbin-Watson = .113284 ** [.000,.000]

Variable      Estimated      Standard      t-statistic      P-value
Coefficient      Error
EMP_RATE      .995020      .077751      12.7976      ** [.000]
GINI           .066147      .615422E-02      10.7483      ** [.000]
CURR_HEALTH_EXP -.069485      .021780      -3.19037      ** [.001]
CRDPRIV_GDP    -.012303      .179949E-02      -6.83692      ** [.000]
INDUSVA_GDP    .045564      .925173E-02      4.92488      ** [.000]
AGRVA_GDP      .057933      .012661      4.57559      ** [.000]
UNDERNOUR_PREV .017484      .480865E-02      3.63588      ** [.000]
SERVVA_GDP     .025186      .861084E-02      2.92491      ** [.003]
UNEMPRATE_TOT .056369      .808534E-02      6.97174      ** [.000]
LABF_TOT3     -.051667      .790212E-02      -6.53841      ** [.000]
C              -3.67894      1.49884      -2.45453      *  [.014]

Hausman test of H0:RE vs. FE:  CHISQ(10) = 23.616,  P-value = [.0087]

Summary of estimated Panel models (* = best SBIC)
LOGL      SBIC      Model
-146.32    176.67      Plain OLS (TOTAL)
341.17     144.36      Varying slopes and intercepts (BYID)
54.819     27.943      * Fixed Effects - Individual (WITHIN)

```

Ce modèle à effet aléatoire présente des variables toutes significatives à 5%, avec une très bonne variabilité expliquée du modèle. Cependant, un test de spécification d'Hausman permettant de déterminer si les coefficients des deux estimations (fixe et aléatoire) sont statistiquement différents, est nécessaire. L'idée de ce test est que, sous l'hypothèse nulle d'indépendance entre les erreurs et les variables explicatives, les deux estimateurs sont non biaisés, donc les coefficients estimés devraient peu différer. Il résulte donc de nos résultats TSP que le modèle à estimer sera un modèle à effets fixes individuels avec présence d'hétéroscédasticité.

II.6. Estimation par la méthode *within* (sous STATA)

II.6.1. Modèle à effet fixe individuel (*One-way FE*)

```

Fixed-effects (within) regression              Number of obs   =       249
Group variable: Id                          Number of groups =       20

R-sq:                                       Obs per group:
    within = 0.7813                        min =           1
    between = 0.6874                       avg =          12.4
    overall = 0.6521                      max =           18

corr(u_i, Xb) = 0.4388                     F(10,219)       =       78.24
                                           Prob > F        =       0.0000

```

Pov_rate	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
emp_rate	.3867173	.2877424	1.34	0.180	-.1803813	.953816
gini	.0544787	.0079146	6.88	0.000	.0388801	.0700773
unemprate_tot	.0574239	.0093415	6.15	0.000	.0390133	.0758346
labf_tot3	-.0426986	.0101095	-4.22	0.000	-.062623	-.0227742
undernour_prev	.0125569	.0059014	2.13	0.034	.0009261	.0241878
curr_health_exp	-.0640805	.0240707	-2.66	0.008	-.1115204	-.0166406
indusva_gdp	.0465399	.010443	4.46	0.000	.0259582	.0671215
crdpriv_gdp	-.0117562	.0020488	-5.74	0.000	-.015794	-.0077184
servva_gdp	.0246628	.0092847	2.66	0.008	.0063639	.0429616
agrva_gdp	.0492717	.0139881	3.52	0.001	.0217032	.0768402
sigma_u	.9213096					
sigma_e	.20702723					
rho	.95193267	(fraction of variance due to u_i)				

```

F test that all u_i=0: F(19, 219) = 46.46                Prob > F = 0.0000

```

Le test reste le même que sur le logiciel TSP. Cependant, comme le soulignait le test d'Hausman, il y a bel et bien une forte corrélation positive entre les effets individuels et les variables explicatives de notre modèle.

II.6.2. Modèle *within* avec effet temporel

Fixed-effects (within) regression			Number of obs = 249			
Group variable: Id			Number of groups = 20			
R-sq:			Obs per group:			
within = 0.8391			min = 1			
between = 0.7620			avg = 12.4			
overall = 0.7719			max = 18			
corr(u_i, Xb) = -0.9728			F(27,202) = 39.02			
			Prob > F = 0.0000			

Pov_rate	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	

emp_rate	3.437294	.4994675	6.88	0.000	2.452455	4.422132
gini	.0378574	.0076385	4.96	0.000	.0227959	.0529188
unemprate_tot	.0699631	.0089831	7.79	0.000	.0522505	.0876757
labf_tot3	-.0746795	.0101247	-7.38	0.000	-.0946431	-.0547159
undernour_prev	.0030417	.0055013	0.55	0.581	-.0078056	.013889
curr_health_exp	-.1055527	.0227524	-4.64	0.000	-.1504153	-.06069
indusva_gdp	.0363273	.0096882	3.75	0.000	.0172244	.0554303
crdpriv_gdp	-.0086993	.0019451	-4.47	0.000	-.0125347	-.0048639
servva_gdp	.0274116	.008467	3.24	0.001	.0107165	.0441067
agrva_gdp	.0461468	.0129079	3.58	0.000	.0206953	.0715982
year						
2001	-.0290645	.0748561	-0.39	0.698	-.176664	.118535
2002	-.0675614	.0741937	-0.91	0.364	-.2138547	.078732
2003	-.1228717	.0777498	-1.58	0.116	-.2761769	.0304336
2004	-.1649226	.0822148	-2.01	0.046	-.3270319	-.0028134
2005	-.2697455	.0889254	-3.03	0.003	-.4450865	-.0944045
2006	-.3982804	.0955368	-4.17	0.000	-.5866576	-.2099031
2007	-.4948558	.10739	-4.61	0.000	-.706605	-.2831065
2008	-.5515313	.1117233	-4.94	0.000	-.7718249	-.3312378
2009	-.5869412	.1170161	-5.02	0.000	-.817671	-.3562114
2010	-.6554457	.1272687	-5.15	0.000	-.9063913	-.4045001
2011	-.8542228	.133471	-6.40	0.000	-1.117398	-.5910476
2012	-.8351724	.1436962	-5.81	0.000	-1.118509	-.5518354
2013	-1.049455	.152373	-6.89	0.000	-1.3499	-.7490093
2014	-1.003729	.1608779	-6.24	0.000	-1.320945	-.6865138
2015	-1.084294	.1708124	-6.35	0.000	-1.421098	-.7474896
2016	-1.125615	.1746803	-6.44	0.000	-1.470046	-.7811846
2017	-1.341923	.1842964	-7.28	0.000	-1.705315	-.9785318
_cons	-37.7045	7.284546	-5.18	0.000	-52.068	-23.341

sigma_u	2.8426827					
sigma_e	.1848913					
rho	.99578747	(fraction of variance due to u_i)				

F test that all u_i=0: F(19, 202) = 57.03				Prob > F = 0.0000		

Notre test présente un effet temporel (la p-value de Fisher reste très largement inférieure à 5%. Donc le model est globalement significatif). Dans ce cas, il y a une variabilité dans le temps pour chaque individu (pays) et le problème de colinéarité reste persistant.

II.6.3. Autocorrélation intra-individus (Test de *Wald* modifié)

Cross-sectional time-series FGLS regression

Coefficients: generalized least squares
Panels: heteroskedastic
Correlation: no autocorrelation

Estimated covariances	=	20	Number of obs	=	249
Estimated autocorrelations	=	0	Number of groups	=	20
Estimated coefficients	=	45	Obs per group:		
			min	=	1
			avg	=	12.45
			max	=	18
			Wald chi2(44)	=	43967.16
			Prob > chi2	=	0.0000

Pov_rate	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
emp_rate	3.262649	.2843533	11.47	0.000	2.705326	3.819971
gini	.0337955	.0045987	7.35	0.000	.0247822	.0428088
unemprate_tot	.0776053	.005833	13.30	0.000	.0661729	.0890378
labf_tot3	-.0630147	.0059325	-10.62	0.000	-.0746422	-.0513872
curr_health_exp	-.114358	.0141939	-8.06	0.000	-.1421776	-.0865384
indusva_gdp	.029184	.0063116	4.62	0.000	.0168134	.0415546
crdpriv_gdp	-.0078558	.0014222	-5.52	0.000	-.0106433	-.0050683
servva_gdp	.0263398	.0054552	4.83	0.000	.0156477	.0370318
agrva_gdp	.0439408	.0096237	4.57	0.000	.0250788	.0628029
Id						
2	4.777783	.4064058	11.76	0.000	3.981242	5.574323
3	-2.655431	.4725479	-5.62	0.000	-3.581608	-1.729254
4	1.632302	.3015666	5.41	0.000	1.041242	2.223362
5	.5985271	.0949772	6.30	0.000	.4123752	.7846791
6	4.867905	.6061929	8.03	0.000	3.679789	6.056022
8	3.568181	.4066831	8.77	0.000	2.771096	4.365265
9	3.20472	.2913529	11.00	0.000	2.633679	3.775761
10	3.730709	.3748519	9.95	0.000	2.996013	4.465405
12	5.569266	.4899519	11.37	0.000	4.608978	6.529554
13	0 (omitted)					
14	5.731129	.7276772	7.88	0.000	4.304908	7.15735
15	-1.720604	.3306714	-5.20	0.000	-2.368708	-1.072499
16	5.699117	.5744653	9.92	0.000	4.573185	6.825048
17	6.579769	.6772795	9.71	0.000	5.252325	7.907212
18	1.880491	.1094907	17.17	0.000	1.665893	2.095089
20	4.75494	.5413889	8.78	0.000	3.693838	5.816043
21	5.787402	.5405432	10.71	0.000	4.727957	6.846847
23	5.141346	.6988875	7.36	0.000	3.771551	6.51114
24	1.822837	.1672094	10.90	0.000	1.495113	2.150562

Dans ce test de Wald modifié (sous STATA), on ne peut rejeter l'hypothèse nulle d'absence d'autocorrélation des erreurs. Le modèle est globalement significatif, c'est donc un modèle à effets fixes (individuel et temporel) hétéroscédastique.

year						
2001	-.069303	.0486566	-1.42	0.154	-.1646682	.0260623
2002	-.0638864	.0459994	-1.39	0.165	-.1540435	.0262708
2003	-.1371511	.0515322	-2.66	0.008	-.2381523	-.0361499
2004	-.1855649	.0490496	-3.78	0.000	-.2817003	-.0894294
2005	-.2732289	.0536343	-5.09	0.000	-.3783501	-.1681077
2006	-.3873064	.0593597	-6.52	0.000	-.5036493	-.2709636
2007	-.4708998	.0659026	-7.15	0.000	-.6000665	-.3417331
2008	-.5422304	.0675931	-8.02	0.000	-.6747104	-.4097504
2009	-.5790244	.0753984	-7.68	0.000	-.7268026	-.4312463
2010	-.6542127	.0781768	-8.37	0.000	-.8074364	-.500989
2011	-.8094283	.0821426	-9.85	0.000	-.9704249	-.6484317
2012	-.8221148	.0859395	-9.57	0.000	-.9905531	-.6536765
2013	-.9683619	.0939447	-10.31	0.000	-1.15249	-.7842338
2014	-.9749882	.0964876	-10.10	0.000	-1.1641	-.785876
2015	-1.054865	.1055684	-9.99	0.000	-1.261775	-.8479545
2016	-1.126111	.1064657	-10.58	0.000	-1.33478	-.9174417
2017	-1.259182	.1149135	-10.96	0.000	-1.484408	-1.033956

II.6.4. Test d'hétéroscédasticité inter-individus (Test de *Wald* modifié)

Cross-sectional time-series FGLS regression

Coefficients: generalized least squares
Panels: homoskedastic
Correlation: no autocorrelation

Estimated covariances	=	1	Number of obs	=	249
Estimated autocorrelations	=	0	Number of groups	=	20
Estimated coefficients	=	46	Obs per group:		
			min	=	1
			avg	=	12.45
			max	=	18
			Wald chi2(45)	=	17960.08
Log likelihood	=	92.84839	Prob > chi2	=	0.0000

Pov_rate	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
emp_rate	3.440904	.4501679	7.64	0.000	2.558591	4.323217
gini	.0389086	.0066684	5.83	0.000	.0258388	.0519785
unemprate_tot	.0692875	.0080218	8.64	0.000	.053565	.08501
labf_tot3	-.0757218	.0089665	-8.44	0.000	-.0932959	-.0581477
curr_health_exp	-.1082991	.0200137	-5.41	0.000	-.1475252	-.0690731
indusva_gdp	.0348644	.0084007	4.15	0.000	.0183994	.0513294
crdpriv_gdp	-.0083713	.0016698	-5.01	0.000	-.011644	-.0050986
servva_gdp	.0269575	.007596	3.55	0.000	.0120697	.0418453
agrva_gdp	.0461215	.0116347	3.96	0.000	.0233178	.0689252

Après avoir corrigé le problème d'autocorrélation, c'est autour de l'hétéroscédasticité.

Le modèle reste très largement significatif au seuil de 5%. On ne peut rejeter l'hypothèse nulle qui stipule que la variance des erreurs est la même pour tous les individus. Le modèle demeure un *within* à effets fixes individuel et temporel.

Id						
2	4.991221	.6310916	7.91	0.000	3.754304	6.228138
3	-2.954363	.7384091	-4.00	0.000	-4.401618	-1.507107
4	1.73981	.409714	4.25	0.000	.9367848	2.542834
5	.589371	.1096639	5.37	0.000	.3744336	.8043084
6	5.213274	.9606885	5.43	0.000	3.330359	7.096189
8	3.707505	.6448292	5.75	0.000	2.443663	4.971347
9	3.273714	.4517467	7.25	0.000	2.388307	4.159122
10	3.793724	.5365537	7.07	0.000	2.742098	4.84535
12	5.737975	.7636167	7.51	0.000	4.241314	7.234637
13	5.462259	.7707611	7.09	0.000	3.951595	6.972923
14	6.267365	1.170479	5.35	0.000	3.973269	8.561462
15	-2.055764	.4969967	-4.14	0.000	-3.029859	-1.081668
16	5.970074	.8841256	6.75	0.000	4.23722	7.702929
17	6.963501	1.083459	6.43	0.000	4.839961	9.087041
18	1.950846	.1660048	11.75	0.000	1.625483	2.27621
20	5.019892	.8469277	5.93	0.000	3.359944	6.67984
21	6.050506	.8603737	7.03	0.000	4.364205	7.736808
23	5.661667	1.102763	5.13	0.000	3.500291	7.823043
24	1.762434	.2447812	7.20	0.000	1.282672	2.242196
year						
2001	-.0262304	.0673148	-0.39	0.697	-.158165	.1057043
2002	-.0659776	.0668263	-0.99	0.323	-.1969547	.0649995
2003	-.1207785	.0699984	-1.73	0.084	-.2579728	.0164158
2004	-.1621765	.0739708	-2.19	0.028	-.3071566	-.0171964
2005	-.2704245	.0801473	-3.37	0.001	-.4275102	-.1133388
2006	-.4002292	.0860556	-4.65	0.000	-.5688951	-.2315633
2007	-.4989278	.0965705	-5.17	0.000	-.6882026	-.309653
2008	-.5554768	.1004987	-5.53	0.000	-.7524507	-.3585028
2009	-.5900122	.1053563	-5.60	0.000	-.7965067	-.3835177
2010	-.6616614	.1142681	-5.79	0.000	-.8856228	-.4376999
2011	-.8619029	.1196539	-7.20	0.000	-1.09642	-.6273856
2012	-.8435019	.1288101	-6.55	0.000	-1.095965	-.5910388
2013	-1.06011	.136242	-7.78	0.000	-1.327139	-.7930805
2014	-1.015626	.1437081	-7.07	0.000	-1.297289	-.7339638
2015	-1.098906	.1521117	-7.22	0.000	-1.397039	-.8007722
2016	-1.14047	.1555787	-7.33	0.000	-1.445398	-.835541
2017	-1.355917	.1645459	-8.24	0.000	-1.678421	-1.033413

Pour la composante temporelle : tous les individus sont significatifs à l'exception des années 2001-2003. Sans doute dû à des des périodes de crises.

II.7. Interprétations économiques du modèle final

Dans cette partie, nous allons exposer ce qui ressort de notre étude (cf. II.6.4) et expliquer les différentes relations de causes à effets vis-à-vis des différentes littératures sur le sujet.

Les effets marginaux négatifs :

- « labf_tot3 », la proportion de la population active avec une élasticité d'environ 0,08% ;
- « curr_health_exp », avec une élasticité d'environ 0,11% pour les dépenses de santé publique (en % PIB) ;
- « crdpriv_gdp », et enfin une élasticité de 0,008% pour le crédit intérieur contracté par le secteur privé.

Une augmentation de 1% des ces différentes variables, toutes choses étant égales par ailleurs (*ceteris paribus*) entraîne un effet marginal négatif sur le taux de pauvreté. C'est-à-dire que toutes les politiques publiques visant à améliorer le niveau de santé et de relancer l'activité économique, baissent le niveau de pauvreté des populations latino-américaines ayant un PPP (*Purchasing Power Parity : Parité de Pouvoir d'Achat*) de 3,20\$ par jour.

Les variables telles que l'indice de gini et le taux de chômage, présentent *ceteris paribus*, des effets marginaux positifs sur le taux de pauvreté. Leur augmentation de 1%, entraîne respectivement une accentuation des inégalités entre pauvres et riches. L'écart se creuse d'environ 0,039% entre les différentes classes sociales et de 0,069% entre individus de la population active.

Par contre, le reste des variables (le taux d'emploi et les différentes valeurs ajoutées dans le secteur primaire et tertiaire) présentent également des effets marginaux positifs. Cela ne semble-t-il pas contradictoire ? La logique nous aurait laissé penser ainsi. Mais non, car ces élasticités renferment une subtilité qui s'apparente à deux cas de phénomènes majeurs que rencontrent ces pays :

- Premièrement les pays d'Amérique latine tels que le Mexique, le Brésil et l'Argentine, ont libéralisé leurs économies : c'est une nouvelle Amérique latine qui émerge dans les années "2000".
- Deuxièmement, les économies latino-américaines sont beaucoup influencées par les différents courants politiques notamment avec l'avènement des régimes populistes d'un côté ; et de l'autre, une faible croissance de l'emploi et des sous-emplois malgré l'essor des secteurs miniers et agro-industriels qui favorisent la force de travail rurale.

C'est suite à cela, que malgré leur faible taux d'augmentation, le taux de pauvreté s'accroît. Il y a une mauvaise redistribution des richesses. Seules les grandes firmes trans et multinationales en tirent profit. A cela, s'ajoute les problèmes de corruption et d'insécurité.

Quant aux composantes temporelles (période avant 2004) : la non significativité de ces 3 variables se traduit par le fait qu'en 2004 (notamment la période de 2004-2008 où les économies latino-américaines connaissent un essor) on sortait d'une période de récession mondiale.

En palliant, résultats économétriques et littératures économiques, nous avons réussi à s'assurer de la véracité du modèle : but ultime de l'économètre.

Conclusion

En somme, nous retenons que l'aboutissement à un bon modèle se trouve au préalable dans la construction et la gestion des données. Cette étape nous permet d'avoir, notamment avec un nuage de point ou une matrice V de cramer, une idée sur le choix des variables visant à expliquer le phénomène d'étude.

En effet, la méthodologie de travail utilisée dans ce projet, visait au-delà d'une simple analyse économétrique. L'objectif était de réaliser des relations de causes à effets afin de confirmer ou infirmer certaines littératures qui le plus souvent des cas, sont basées sur des cas factuels et non sur la recherche de causalités via des méthodes qui pourraient exposer des phénomènes non apparents.

Les niveaux de pauvreté et de chômage galopants sont, de nos jours, des problèmes de sociétés majeurs devant impliquer toutes les différentes couches sociales.

Des politiques de relance économiques et budgétaires sont vivement recommandées, notamment en cette période de crise sanitaire afin d'éviter une implosion des classes sociales déficitaires qui se soldera par des troubles d'ordre sociaux où personne ne sera épargné.

Bibliographie

BOSSIO ROTONDO, J. C. « Amérique latine: pauvreté, inégalités, chômage et orientations politiques. », *Vie économique* 1.3 (2010).

ÉPAULARD, Anne. « Croissance et réduction de la pauvreté dans les pays en développement et les pays en transition », *Reflets et perspectives de la vie économique*, vol. tome xlii, no. 2, 2003, pp. 9-20.

VALDES, A., and MISTIAEN J. A. « Pauvreté rurale en Amérique Latine: tendances récentes et nouveaux enjeux. », (2001).

Webographie

<https://www.banquemondiale.org/fr/region/lac/overview>, consulté le 12/01/2021

https://www.lepoint.fr/monde/l-extreme-pauvrete-touche-plus-10-de-la-population-en-amerique-latine-15-01-2019-2286104_24.php, consulté le 12/01/2021