

Les inégalités et la croissance : une relation introuvable ?

Denis COGNEAU
Charlotte GUENARD

LES INEGALITES ET LA CROISSANCE : UNE RELATION INTROUVABLE¹ ?

Denis Cogneau
(IRD et DIAL)
Charlotte Guénard
(IEP Paris et DIAL)

e-mail : cogneau@dial.prd.fr
guenard@dial.prd.fr

Document de travail DIAL / Unité de Recherche CIPRE
Janvier 2002

RESUME

Ce travail revient sur l'économétrie de la relation entre inégalités et croissance, à l'échelle macro-économique des pays. Il s'intéresse particulièrement aux conséquences des choix de spécification, de la méthode d'estimation et de la sélection de l'échantillon. Il conclut à l'absence d'une relation robuste allant des inégalités de revenu vers la croissance du produit par tête ou vers l'investissement physique et humain. Une relation de causalité inverse, conforme à la philosophie de la courbe de Kuznets, paraît mieux assise. En tout état de cause, les stratégies de développement nationales et les configurations structurelles et historiques de chaque société conservent une place très large pour déterminer les évolutions conjointes du revenu et de sa répartition.

ABSTRACT

This work takes a new look at the econometrics of the national-level macroeconomic relationship between inequality and growth. It focuses, in particular, on the effects of specification choices, estimation method and sample selection. It concludes that there is no robust relation running from income inequality to per capita product growth or physical and human investment. There would seem to be more grounds for an inverse causal relation in keeping with the Kuznets curve philosophy. In any case, the national development strategies and structural and historical configurations of each society still play a major role in determining concurrent changes in income and its distribution.

¹ Nous remercions les participants du séminaire d'économie du développement DIAL-Paris I, et particulièrement Jean-Claude Berthélemy, ainsi que les membres du groupe de travail du Laboratoire de macro-économie du CREST, et particulièrement Stéphane Grégoir. Notre responsabilité reste entière.

Table des matières

INTRODUCTION	4
1. LES SPECIFICATIONS	7
2. L'ESTIMATION	8
3. LES DONNEES	10
4. DES INEGALITES NUISIBLES OU FAVORABLES A LA CROISSANCE ?	13
5. UN MODELE KALDORIEN D'INVESTISSEMENT ?	15
6. DES INEGALITES FAVORABLES A LA SCOLARISATION ET A LA BAISSSE DE LA FECONDITE ?	16
7. UNE CROISSANCE EGALITAIRE ?	17
CONCLUSION	20
REFERENCES BIBLIOGRAPHIQUES	21
ANNEXES	24

Liste des tableaux

<i>Tableau n° 4-1 : Equations de croissance</i>	<i>13</i>
<i>Tableau n° 4-2 : Test des spécifications non-linéaires de Banerjee et Duflo</i>	<i>14</i>
<i>Tableau n° 4-3 : Test de spécifications non-linéaires pour l'indice de Gini dans l'équation de croissance</i>	<i>14</i>
<i>Tableau n° 5-1 : Equation d'investissement</i>	<i>16</i>
<i>Tableau n° 6-1 : Equation de scolarisation secondaire</i>	<i>16</i>
<i>Tableau n° 6-2 : Equation du taux de fécondité</i>	<i>17</i>
<i>Tableau n° 7-1 : Estimations d'une courbe de Kuznets</i>	<i>18</i>
<i>Tableau n° 7-2 : Estimations d'une courbe de Kuznets augmentée</i>	<i>19</i>

Liste des annexes

<i>Annexe 1 : Dictionnaire des variables</i>	<i>24</i>
<i>Annexe 2 : Statistiques descriptives</i>	<i>24</i>
<i>Annexe 3 : Analyse des correspondances multiples</i>	<i>25</i>

INTRODUCTION

L'économie s'interroge depuis longtemps sur le lien entre la répartition et la production, les grandes théories historiques de la production et de l'allocation des ressources étant également des théories de la répartition et de la formation du prix des ressources². Durant les années 1990, une abondante littérature est revenue sur la question, dans la foulée des "nouvelles" théories de la croissance. La disponibilité d'une base de données internationale à couverture large sur les inégalités de revenu, réunie par Deininger et Squire (1996) à la Banque Mondiale, a contribué à relancer une série de travaux économétriques³. Le présent papier s'inscrit dans cette série, en procédant à un réexamen de plusieurs faits stylisés avancés par la littérature théorique. Il s'intéresse particulièrement aux conséquences des choix de spécification, de la méthode d'estimation et de la sélection de l'échantillon. Il conclut à l'absence d'une relation robuste allant des inégalités de revenu vers la croissance du produit par tête ou vers l'investissement, physique et humain, une relation de causalité inverse paraissant mieux assise.

Depuis les années 1960 jusqu'au début des années 1990, de nombreux articles ont été consacrés à cette relation de causalité allant de la croissance vers les inégalités, dans la lignée des travaux de Kaldor (1956) et de Kuznets (1955). Par exemple les modèles de Stiglitz (1969), puis de Bourguignon (1981 et 1990), partant du cadre solowien, discutaient les conditions d'apparition d'une courbe de Kuznets, et plus généralement les hypothèses sous lesquelles la croissance pouvait engendrer ou au contraire résorber les inégalités de revenu⁴. Dans les années 1970 et 1980, les données transversales réunies sur les pays en développement semblaient corroborer⁵ la courbe de Kuznets (Adelman et Morris, 1973 ; Ahluwalia, 1976 ; Papanek et Kyn 1986). Pourtant, la sensibilité des estimations aux formes fonctionnelles testées et à la composition des échantillons sont progressivement apparues (Anand et Kanbur, 1993a et 1993b). Sur données transversales, la courbe témoigne toutefois d'une certaine résistance, une fois contrôlé l'effet d'autres variables démographiques, de capital humain ou de dualisme (Barro, 1999 ; Higgins et Williamson, 1999 ; Bourguignon et Morrisson, 1999). Elle passe toutefois plus difficilement l'épreuve des données longitudinales (Fields et Jakobson, 1994 ; Li, Squire et Zou, 1998). Des modèles théoriques continuent de s'inspirer de cette courbe (par exemple Galor et Tsiddon, 1997 ; Dahan et Tsiddon, 1998).

A partir des années 1990, l'intérêt des économistes s'est tourné vers la relation de causalité inverse, allant des inégalités vers la croissance. C'est en effet en 1991 qu'apparaît un nouveau fait stylisé issu de l'économétrie de la croissance : une relation négative entre les inégalités « initiales » et la croissance de long terme (Person et Tabellini, 1994 ; Alesina et Rodrik 1994 ; Bourguignon, 1993 ; Birsdall et Sabbot, 1995 ; Clarke, 1995 ; Perotti, 1996 ; Galor et Zang, 1997). De nombreux modèles théoriques ont été élaborés pour rendre compte de cette régularité, dont la robustesse fut un court moment considérée comme éprouvée (cf. en particulier Clarke, 1995). Pourtant, comme pour la courbe de Kuznets, la « nouvelle courbe » révéla bien vite sa fragilité, même sur données transversales (Fishlow, 1996 ; Barro, 1999). Chez Forbes (1999), l'analyse longitudinale semble mettre à jour une relation opposée, c'est-à-dire des inégalités favorables à la croissance. Chez

² Cf. Schumpeter J. A. (1954), notamment t.III, pp.220-233 à propos de l'unification théorique (au sens de la physique) des deux problèmes par les auteurs considérés comme les premiers "néo"-classiques (Menger, Wieser, Jevons, Walras).

³ La brève revue de la littérature qui suit se focalise sur *les tests économétriques* de la relation entre les inégalités de revenu et la croissance du revenu. De surcroît, elle ne prétend à l'exhaustivité en ce domaine que pour les dix dernières années de publication. On pourra par exemple se référer à Bourguignon (1998) pour une synthèse sur le plan théorique.

⁴ On se rappellera notamment que Stiglitz (1969) adapte le modèle de Solow à une population d'individus dont les dotations initiales en capital sont inégales. Dans un cadre similaire, Bourguignon (1981) explore les conséquences d'un taux d'épargne microéconomique fonction convexe du niveau de revenu.

⁵ Dans ce texte, le terme "corroboration" est employé dans le sens que lui donne l'épistémologie popperienne : lorsqu'une hypothèse "falsifiable" passe l'épreuve d'un test de réfutation, ni plus ni moins. Il n'est pas synonyme de "vérification". Cf. Popper K. R. (1959), pp.256-287.

Banerjee et Duflo (2000), c'est la variation des inégalités, quel que soit son signe, qui semble avoir un impact négatif sur la croissance.

Une première série de modèles théoriques repose sur une hypothèse centrale d'imperfection des marchés du crédit, que ce soit pour le financement des projets d'investissement matériel ou de la scolarisation (Loury, 1981 ; Banerjee et Newman, 1993 ; Galor et Zeira, 1993 ; Aghion et Bolton, 1997 ; Piketty, 1997 ; Galor et Zang, 1997). Toujours sur données transversales, le lien négatif entre les inégalités et la scolarisation fut corroboré par Bourguignon (1993 et 1998) et par Perotti (1998). Le lien négatif entre les inégalités et l'investissement physique apparaît chez Person et Tabellini (1994) et Perotti (1994). Il est en revanche inversé chez Bourguignon (1993), et invalidé chez Barro (1999).

Une seconde série de modèles est centrée sur l'économie politique. Ceci apparaissait d'autant plus justifié que l'effet négatif des inégalités sur la croissance semblait persister malgré l'inclusion de l'investissement physique et humain dans les équations économétriques testées. Les deux premiers sont ceux de Person et Tabellini (1994) et Alesina et Rodrik (1994), qui postulent un mécanisme par lequel des inégalités trop élevées induisent une demande de redistribution génératrice de distorsions fiscales. Ce mécanisme est contesté empiriquement par Bénabou (1996) et Perotti (1993 et 1996) qui ne trouvent pas de lien significatif entre les inégalités et les transferts aux fins de redistribution. Un second mécanisme plus flou postule qu'un niveau élevé d'inégalité favorise des situations d'instabilité politique. Cette corrélation est testée par Perotti (1994 et 1996). Banerjee et Duflo (2000) proposent un modèle très simple de "hold-up" où les opportunités de croissance sont plus ou moins efficacement exploitées, selon l'intensité des conflits distributifs qu'elles soulèvent. Économétriquement, ils concluent à l'existence d'une relation non-linéaire entre la variation des inégalités et la croissance postérieure. Enfin des modèles combinent des mécanismes d'économie politique - selon différentes variantes de l'électeur médian - et les imperfections du marché du crédit (Perotti, 1993 ; Saint-Paul et Verdier, 1993 ; Bourguignon et Verdier, 2000 ; Lee et Roemer, 1997). Les résultats de ces modèles sont plus complexes et n'ont pas connu jusqu'à présent de corroboration empirique.

D'autres arguments ont été avancés pour rendre compte du lien négatif entre inégalités et croissance. Les mécanismes d'externalité locale et de distribution de la population ont été particulièrement explorés par Bénabou (1994 et 1996) et Durlauf (1994 et 1996). Murphy, Schleifer et Vishny (1989) proposent une explication originale par des effets de taille de marché en présence de rendements d'échelle croissants. L'impact des inégalités sur la fécondité a été également considéré, dans la suite du modèle de Becker, Murphy et Tamura (1990). Perotti (1996) constate un lien positif entre les inégalités et la fécondité, qui pourrait expliquer comment les inégalités pèsent sur la croissance, en retardant la transition démographique.

On voit ainsi que la littérature théorique propose une grande variété de spécifications testables, et que la littérature empirique correspondante a obtenu des résultats pour le moins contrastés. D'un strict point de vue économétrique, on peut constater que les techniques utilisées jusqu'à récemment étaient assez frustes. Les analyses empiriques sont donc sujettes aux deux types de biais classiques de l'économétrie : des biais de sélection d'une part, et des biais d'endogénéité d'autre part. Ceci s'explique par les limitations inhérentes aux données disponibles, même pour la base la plus complète à ce jour, à savoir celle de Deininger et Squire (1996). En effet, les données internationales concernant les inégalités de revenu sont encore aujourd'hui particulièrement lacunaires et de qualité diverse⁶. Cela conduit à une sélection drastique des échantillons : typiquement, les analyses portent sur une cinquantaine de pays, parmi lesquels les pays de l'OCDE

⁶ Les données considérées comme de haute qualité par Deininger et Squire ne sont pas exemptes de critiques, même pour les pays de l'OCDE. La série de la France connaît par exemple une rupture en 1975.

sont sur-représentés. Par ailleurs, la dimension longitudinale des données est fortement limitée. Typiquement là encore, on aboutit à un panel non-cylindré comprenant seulement une centaine de points pour une quarantaine de pays, dont les pays d'Afrique sub-saharienne et les pays les plus pauvres sont quasiment absents. La variance inter-temporelle exploitable est de surcroît très inférieure à la variance inter-individuelle.

Concernant les biais de sélection, certains papiers montrent une sensibilité des résultats au type d'échantillon retenu, même dans la dimension transversale. Par exemple, chez Barro (1999), l'impact des inégalités sur la croissance varie selon le niveau de revenu initial: il apparaît plutôt négatif pour les pays pauvres et plutôt positif chez les pays riches. Comme l'exploitation de la dimension longitudinale implique une sélection encore plus forte des échantillons, les tests de robustesse sont limités à la prise en compte de spécificités régionales (Amérique Latine notamment) ou de deux niveaux de développement (pays riches et pays pauvres). Or, l'analyse descriptive des données suggère que les biais de sélection risquent bien de ne pas être inoffensifs (voir ci-dessous). Dès lors, il est apparu préférable de conserver le maximum de pays dans les analyses. Pour cela, nous avons restreint le nombre de variables explicatives en privilégiant des spécifications simples, et introduit des indicatrices pour les données manquantes sur les inégalités et leur variation temporelle.

Concernant les biais dus à la présence d'effets fixes individuels endogènes, les techniques correctrices sont sensibles aux erreurs de mesure, et privilégient les relations de court et de moyen terme. Ces réserves amènent Barro (1999) à rejeter leur utilisation dans le cadre de l'économétrie de la croissance. Il n'en demeure pas moins que l'exploitation de la dimension inter-individuelle néglige totalement la corrélation des variables explicatives avec un effet individuel non observé, et ne permet pas un test satisfaisant des différentes hypothèses théoriques. Au-delà de cela, les estimateurs en écart à la moyenne individuelle ("*within*") ou en différences premières ne suffisent pas, même corrigés de l'hétéroscédasticité, souvent importante. Car la corrélation des variables avec l'effet de conjoncture individuel est en fait très probable. Premièrement, la plupart des modèles proposés impliquent une *codétermination* de la croissance (ou du niveau du PIB) et des inégalités sous la forme d'un système d'équation simultanées. Deuxièmement, les variables des bases de données internationales sont sujettes à des erreurs de mesure très fortes, surtout pour les pays en développement. Troisièmement et enfin, les spécifications de la plupart des équations sont auto-régressives, à commencer par les équations de croissance qui incluent systématiquement le niveau du PIB retardé – l'effet de convergence. Ces trois problèmes ont commencé à être traités en économétrie de la croissance (cf. par exemple, Caselli, Esquivel et Lefort, 1996), avec la mise en œuvre d'estimateurs à variables instrumentales, dans le cadre de la méthode des moments généralisés (GMM). Les estimateurs GMM pour les modèles dynamiques sur données de panel ont été particulièrement étudiés et vulgarisés par les travaux d'Arellano et Bond⁷ (1991).

Enfin, au-delà de la méthode économétrique utilisée, il convient de ne pas perdre de vue les problèmes liés à la spécification des modèles. La littérature empirique répertoriée ci-dessus présente d'ailleurs une grande diversité sous ce rapport.

La suite de ce papier est donc organisée comme suit : une première section présente les spécifications retenues pour les tests économétriques, une seconde section les techniques d'estimation et le traitement des biais d'endogénéité et de sélection, une troisième section la construction des données. Les sections 4 à 7 présentent les résultats des estimations économétriques concernant successivement la croissance, l'investissement physique, l'investissement humain et la fécondité, et enfin la courbe de Kuznets.

⁷ Les estimations effectuées ici utilisent la dernière version du logiciel "DPD" (1998) proposée par ces auteurs.

1. LES SPECIFICATIONS

Comme Bourguignon (1993), nous avons choisi de tester un ensemble de relations théoriquement envisageables entre les inégalités et la croissance, en examinant quelques uns des canaux par lesquels cette relation pourrait transiter.

Nous nous sommes tout d'abord attachés à éprouver à nouveau l'effet direct des inégalités sur la croissance, comme dans la plupart des papiers antérieurs (section 4). La question se pose néanmoins du choix d'une équation de croissance. L'équation retenue s'inspire de la spécification de Mankiw, Romer et Weil (1992). Elle fait donc intervenir, outre l'indice de Gini retardé d'une période (GINI(-1)) :

- (i) l'effet de convergence sous la forme du niveau du PIB antérieur, LGDP(-1) ;
- (ii) l'effet de dilution démographique sous la forme du logarithme du taux de croissance annuel moyen de la population incrémenté d'un taux de progrès technique identique et égal à 5 % pour tous les pays, LGPOPC ;
- (iii) l'effet du logarithme du taux d'investissement courant, LINV ;
- (iv) le flux de main-d'œuvre qualifiée sous la forme du logarithme taux de scolarisation secondaire retardé de deux périodes, SECR(-2) ;⁸.

Ces trois dernières variables nous paraissent constituer les facteurs fondamentaux de la croissance de moyenne période. En outre, elles sont chacune susceptibles d'être affectées directement par les inégalités. Enfin, plusieurs spécifications non-linéaires ont été testées pour l'indice de Gini, avec les mêmes variables de contrôle, afin de comparer nos résultats avec ceux de Banerjee et Duflo (2000).

Les modèles théoriques évoqués précédemment, qui font intervenir les imperfections sur les marchés du crédit, impliquent une relation négative entre les inégalités et l'investissement, physique ou humain selon les cas. *A contrario*, un modèle kaldorien, tirant la conséquence d'une relation positive entre le taux d'épargne des ménages et leur revenu, prévoit une relation positive entre les inégalités et l'accumulation. Nous chercherons donc à tester l'impact des inégalités sur l'investissement physique et humain. Les deux spécifications retenues sont également auto-régressives. En ce qui concerne l'investissement physique, nous ajoutons un effet accélérateur-demande, considéré comme l'un des plus robustes par de nombreux macro-économistes (section 5). En ce qui concerne l'investissement humain, nous nous intéresserons aux taux de scolarisation primaire et secondaire, que nous relierons au niveau de revenu par habitant (section 6).

Dans le cadre d'un modèle de choix de fécondité, l'arbitrage entre le nombre d'enfants et l'éducation de ces enfants fait intervenir le rendement de la scolarité, et ainsi, indirectement, les inégalités de revenu entre qualifiés et non-qualifiés, comme chez Becker, Murphy et Tamura (1990). On a donc des raisons de penser que la fécondité dépend négativement des inégalités. Nous chercherons donc à tester l'impact des inégalités sur les transitions de fécondité, dans le cadre d'un modèle là encore auto-régressif introduisant par ailleurs le niveau de revenu par habitant et le niveau d'éducation de la population (section 6).

Enfin, même si la littérature empirique récente a plutôt délaissé la courbe de Kuznets, nous présentons un test longitudinal de la relation de causalité allant de la croissance et des variables de développement vers les inégalités (section 7).

⁸ Il s'agit d'une spécification différente de celles de Barro, Forbes et Banerjee-Duflo, op.cit., qui introduisent une externalité "à la Arrow-Lucas" liée au *stock* de capital humain. Néanmoins, l'utilisation d'une variable comme la proportion d'individus de 15 ans et plus ayant fait des études secondaires (Barro et Lee, 1993) ne modifie pas les principales conclusions quant à l'impact des inégalités de revenu sur la croissance. Voir section 4 pour le commentaire des résultats sur les variables de capital humain.

2. L'ESTIMATION

Trois types d'estimateurs sont considérés (voir ci-dessous). Ils sont chacun corrigés de l'hétéroscédasticité selon la méthode proposée par White (1982). Les variables sont chacune exprimées en déviation par rapport à la moyenne de la période – ce qui correspond à la prise en compte d'un effet fixe temporel considéré comme certain. Le caractère non-cylindré du panel, source supplémentaire d'hétéroscédasticité, est lui aussi pris en compte (Arellano et Bond, 1991).

De façon générale, les modèles estimés se présentent sous la forme suivante :

$$y_{i,t} = \mathbf{I} y_{i,t-1} + x_{i,t} \mathbf{b} + u_i + v_{i,t} \quad (1)$$

L'estimateur des moindres carrés ordinaires dans la dimension totale des données (OLS levels) fournit un point de référence par rapport aux résultats des travaux précédents sur données transversales. En effet, il privilégie largement la dimension inter-individuelle du panel, et fournit des estimations proches de celles que donnerait un estimateur "between".

De toute manière, ces deux estimateurs, de même que l'estimateur des moindres carrés quasi-généralisés du modèle à erreurs composées⁹, sont sans ambiguïté non convergents, compte tenu du caractère auto-régressif du modèle :

$$E(y_{i,t-1} u_i) \neq 0 \quad (2)$$

Par ailleurs, la corrélation de l'effet fixe individuel avec les autres variables explicatives n'est pas à exclure :

$$E(x_{i,t} u_i) \neq 0 \quad (3)$$

Ainsi, pour les équations considérées, des tests d'Hausman comparant l'estimateur "within" et l'estimateur des moindres carrés quasi-généralisés, conduisent systématiquement à rejeter l'exogénéité de l'effet individuel.

L'estimateur des moindres carrés ordinaires pour les différences premières (OLS FD) supprime l'effet fixe individuel :

$$y_{i,t} - y_{i,t-1} = \Delta y_{i,t} = \mathbf{I}(y_{i,t-1} - y_{i,t-2}) + \Delta x_{i,t} \mathbf{b} + \Delta v_{i,t} \quad (4)$$

Il a été préféré à l'estimateur "within" dans la mesure où il fournit un point de référence pour l'estimateur GMM d'Arellano et Bond, qui est un estimateur IV (variables instrumentales) des différences premières. L'estimateur FD est sensible à trois types de biais: (i) biais d'auto-régressivité, (ii) biais de simultanéité, (iii) biais d'erreurs de mesure sur les variables explicatives.

Le caractère auto-régressif des modèles implique immédiatement une corrélation entre la variable endogène retardée et les différences premières des erreurs :

$$E(y_{i,t-1} v_{i,t-1}) \neq 0 \Rightarrow E(\Delta y_{i,t-1} \Delta v_{i,t}) \neq 0 \quad (5a)$$

⁹ Cet estimateur, correspondant à des effets individuels aléatoires ("random effect model"), est une combinaison linéaire de l'estimateur "between" et de l'estimateur "within". Il se situerait probablement dans notre cas "entre" l'estimateur "OLS levels" et l'estimateur "OLS FD". C'est le principal estimateur mobilisé par Banerjee et Duflo, op.cit., pour le test de leur hypothèse non-linéaire (voir section 4).

Par ailleurs, la détermination simultanée de certaines variables, le taux de croissance et le taux d'investissement par exemple, par des éléments de conjoncture idiosyncrasique et non-observés, peut entraîner une corrélation entre les valeurs courantes des variables explicatives et les différences premières des erreurs :

$$\begin{aligned} x_{i,t} &= z_{i,t} \mathbf{q} + \mathbf{a}_i + \mathbf{h}_{i,t} \\ E(v_{i,t} \mathbf{h}_{i,t}) &\neq 0 \Rightarrow E(x_{i,t} v_{i,t}) \neq 0 \Rightarrow E(\Delta x_{i,t} \Delta v_{i,t}) \neq 0 \end{aligned} \quad (5b)$$

Enfin la présence d'erreurs de mesure additives sur les variables, particulièrement difficile à écarter dans le cas des indices d'inégalité, a tendance à biaiser vers zéro le coefficient de ces variables :

$$\begin{aligned} x_{i,t} &= x_{i,t}^* + \mathbf{e}_{i,t} \\ y_{i,t} &= \mathbf{l} y_{i,t-1} + x_{i,t} \mathbf{b} + u_i + v_{i,t}^* \\ v_{i,t}^* &= v_{i,t} - \mathbf{e}_{i,t} \mathbf{b} \\ E(x_{i,t} v_{i,t}^*) &\neq 0 \\ E(\Delta x_{i,t} \Delta v_{i,t}^*) &= -\mathbf{b} \text{Var}(\mathbf{e}_{i,t}) \leq 0 \end{aligned} \quad (5c)$$

Afin de traiter ces trois sources de biais, l'estimateur GMM(n) proposé par Arellano et Bond (1991) consiste à instrumenter les différences premières des variables par les niveaux de ces variables retardés de 2 à (n) périodes. Cet estimateur fait les deux hypothèses (6a) et (6b).

Premièrement, les variables sont prédéterminées d'au moins une période; elles ne sont pas corrélées avec les résidus futurs :

$$\begin{aligned} \forall h &\geq 1 \\ E(x_{i,t} v_{i,t+h}) &\neq 0 \end{aligned} \quad (6a)$$

Cette première hypothèse est la plus problématique, surtout dans le cadre d'une détermination simultanée des variables: par exemple les non observables de la croissance en t peuvent être corrélés avec les non-observables des inégalités en $t-1$. Pour cette raison, nous avons utilisé autant que possible des valeurs retardées des variables explicatives.

Deuxièmement, les termes d'erreur ne sont pas auto-corrélés :

$$E(v_{i,t} v_{i,t-1}) = 0 \Rightarrow E(\Delta v_{i,t} \Delta v_{i,t-2}) = 0 \quad (6b)$$

Arellano et Bond (1991) ont élaboré un test de cette seconde hypothèse, soit un test de corrélation du second ordre entre les différences premières des termes d'erreur individuels-temporels.

Sous ces deux hypothèses, il est légitime d'instrumenter les différences premières des variables par les niveaux retardés d'au moins deux périodes :

$$\begin{aligned} \forall h &\geq 2, \\ E(x_{i,t-h} \Delta v_{i,t}) &= E(y_{i,t-h} \Delta v_{i,t}) = 0 \end{aligned} \quad (7)$$

Le nombre des conditions d'orthogonalité correspondant aux hypothèses étant supérieur au nombre de paramètres à estimer, plusieurs estimateurs GMM sont possibles. On utilise donc un test de sur-identification de Sargan, compatible avec l'hétéroscédasticité des termes d'erreur (cf. Arellano et Bond, 1991).

L'efficacité de l'estimateur GMM de seconde étape a été considérée avec circonspection. En effet, l'étude par simulation d'Arellano et Bond montre la possibilité d'une sous-estimation de 20 à 30 % des écarts-types des coefficients, à distance finie. Ce biais serait plus prononcé en cas d'hétéroscédasticité. Pour pallier ce risque d'inférence trop optimiste¹⁰, nous avons signalé par des étoiles les coefficients significatifs *selon l'estimateur de première étape*. Par ailleurs, nous avons fait varier le nombre (n) de retards des instruments, afin de tester la robustesse des estimateurs¹¹.

Enfin, nous avons cherché à traiter partiellement le problème de sélection de l'échantillon, par une méthode fruste de variables indicatrices. En effet, la construction d'un estimateur en deux étapes "à la Heckman" nous est apparu trop complexe. Nous nous sommes bornés à introduire dans les estimations des indicatrices valant 1 si l'indice d'inégalité g est manquant, pour la période t , et la même chose pour la période $t-1$ dans le cas des estimations en différences premières:

$$y_{i,t} = \mathbf{I} y_{i,t-1} + x_{i,t} \mathbf{b} + \mathbf{I}\{\exists g_{i,t}\} + u_i + v_{i,t} \quad (8a)$$

$$y_{i,t} - y_{i,t-1} = \mathbf{I}(y_{i,t-1} - y_{i,t-2}) + \Delta x_{i,t} \mathbf{b} + \mathbf{I}\{\exists g_{i,t}\} + \mathbf{I}\{\exists g_{i,t-1}\} + \Delta v_{i,t} \quad (8b)$$

Cette méthode de variables indicatrices donc a été appliquée à un échantillon étendu, contenant les données manquantes pour l'indice de Gini (échantillon n°1, voir section 3 ci-dessous). Ces indicatrices sont elles-mêmes instrumentées par leurs valeurs retardées dans les estimations GMM. Les estimations ont également été menées sur un échantillon réduit ne contenant aucune donnée manquante, et donc sans ces variables indicatrices (échantillon n°2).

3. LES DONNEES

La liste, les sigles et les sources des variables utilisées sont récapitulés en annexe 1. La variable de PIB par tête en dollars internationaux, et donc la variable de croissance (GROWTH), proviennent de la base "Chelem" du CEPII. Les six autres variables d'analyse proviennent des bases de données de la Banque Mondiale : le taux de croissance de la population (GPOP) et le taux de fécondité (FERT), les taux de scolarisation primaire (PRIMR) et secondaire (SECR), le taux d'investissement (INV), enfin l'indice de Gini (GINI). Cette dernière variable provient de la base de données de Deininger et Squire. Nous n'avons conservé que les données d'inégalité labellisées comme de "haute qualité" : couverture nationale, source statistique clairement identifiée, méthode de calcul connue. Après analyse économétrique, l'indice de Gini a été corrigé de la manière suivante : nous avons ajouté 6.5 points dans le cas où il s'agissait de données de dépenses de consommation et non de revenu¹².

Deux échantillons ont été construits, dont les statistiques descriptives figurent en annexe 2. Tout d'abord, un panel de base a été constitué en regroupant les données en sept périodes quinquennales¹³ de 1960 à 1995. Les niveaux du PIB et des taux de scolarisation secondaire sont des niveaux de début de période (1960, 1965,...). Les niveaux des autres variables sont des moyennes quinquennales (1960-64, 1965-69,...).

¹⁰ qui semble caractériser les conclusions obtenues par Forbes, op.cit., avec le même type d'estimateur.

¹¹ Nous avons évité de nous trouver dans le cas où le nombre d'instruments dépasse le nombre d'unités individuelles (nombre de pays). Une pseudo-inverse est alors utilisée par le logiciel "DPD" pour calculer la matrice des poids à partir des conditions d'orthogonalité, et fournit des résultats outrageusement efficaces.

¹² La même correction est préconisée par Deininger et Squire et appliquée par Forbes et Banerjee et Duflo, op.cit. D'après nos estimations, l'unité statistique de calcul de l'indice de Gini, selon qu'il s'agit de ménages ou de personnes, n'influence pas significativement le niveau et la variation de l'indice.

¹³ Les deux premières périodes n'interviennent pas directement dans l'analyse économétrique qui utilise les niveaux des variables retardés de deux périodes.

Ensuite, l'échantillon n°1 a été extrait de ce panel de base de telle sorte qu'il ne comprend aucune donnée manquante pour les variables d'analyse, à l'exception de l'indice de Gini¹⁴. Il comprend finalement 90 pays dont 67 sont des pays en développement et 23 des pays développés ou en transition. Pour les cinq périodes d'analyse, ce panel de base est non-cylindré et inclut 387 points, relativement bien répartis dans le temps (voir annexe 2). Sur ces 387 points, 252 correspondent à une donnée manquante pour la variation de l'indice de Gini, et 135 points à une donnée exploitable qui permet l'identification de l'effet des inégalités dans les estimations en différences premières. Ces 135 points sont fournis par 52 pays sur les 90 du panel de base, 32 étant des pays en développement et 20 des pays développés ou en transition.

Si l'on ne retient que les pays pour lesquels on dispose des variations de l'indice de Gini pour deux périodes quinquennales successives, soit trois indices de Gini successifs, on obtient l'échantillon n°2. Celui-ci se réduit à 92 points et 37 pays, parmi lesquels 21 sont des pays en développement et 16 des pays développés ou en transition. Par sa taille et sa composition, cet échantillon est très proche de ceux des auteurs qui ont travaillé sur la dimension longitudinale de la base de Deininger et Squire (Forbes, et Banerjee et Duflo, à notre connaissance). Par exemple, les échantillons utilisés par Banerjee et Duflo comprennent entre 98 et 128 points et concernent au maximum 45 pays, parmi lesquels 35 pays sont en intersection avec notre échantillon n°2. Les statistiques descriptives (annexe 2) confirment que le profil des deux échantillons est très proche. En comparaison de l'échantillon n°1, les pays les plus pauvres et à faible croissance sont très peu représentés, ainsi que la décennie 1970¹⁵. Au-delà du caractère asymptotique des estimateurs, et quels que soient les tests de robustesse interne que l'on peut effectuer, il n'est pas improbable que certains résultats soient sensibles à la sélection de l'échantillon et aux périodes disponibles. Cette sélection de l'échantillon par des variables endogènes constitue de notre point de vue un problème crucial.

Afin d'approfondir ce problème de sélection, nous avons effectué une analyse des correspondances multiples¹⁶ sur chacun des deux échantillons, à partir des variables appartenant à la spécification de notre équation de croissance, exprimées en différences premières (voir sections 2 et 3 ci-dessus). L'analyse des correspondances multiples est sans doute la méthode d'analyse descriptive des données multivariées la plus adaptée à notre objet, dans la mesure où elle ne fait pas l'hypothèse de liaisons linéaires entre les variables, et permet aussi un traitement satisfaisant des données manquantes¹⁷. Un découpage des variables en quatre classes de taille approximativement égale¹⁸ (quatre quartiles) a été retenu. Cette "discrétisation" offre également l'avantage d'une moindre sensibilité aux erreurs de mesure et aux points aberrants, surtout en différences premières. Les premiers plans factoriels de chacune des deux analyses sont présentés sur les graphes de l'annexe 3. Les graphes font uniquement figurer les modalités des variables actives de l'analyse, les bornes des classes étant indiquées explicitement. Des indicatrices de pays et de période ont été projetées en tant que variables supplémentaires pour faciliter l'interprétation des graphes, mais ne sont pas reportées.

Les deux premiers axes de l'analyse rendent compte d'une proportion sensiblement égale de l'inertie¹⁹ du nuage de points : respectivement 9 % et 7.9 % pour l'échantillon n°1, 10.9 % et 10.3 % pour l'échantillon n°2. Ce premier constat souligne que les deux nuages sont relativement

¹⁴ Le passage aux moyennes quinquennales, pratiqué par la plupart des auteurs sur ce thème, se justifie dans notre cas par les données manquantes des variables autres que l'indice de Gini, certaines variables comme les taux de scolarisation n'étant pas enregistrées annuellement dans la plupart des pays.

¹⁵ Le seul pays africain représenté dans l'échantillon n°2 est l'île Maurice. De même, la liste de 45 pays donnée par Banerjee et Duflo ne comprend aucun pays d'Afrique sub-saharienne. Cf. Banerjee et Duflo, op.cit., table 1.

¹⁶ Nous remercions Eric Maurin de cette suggestion.

¹⁷ Contrairement à l'analyse en composantes principales qui se résume à l'analyse de la matrice de corrélations linéaires. Cf. Benzécri, 1973 et Volle, 1981.

¹⁸ A l'exception de la modalité "donnée manquante pour la variation de l'indice de Gini", $DGINI(-)=?$, dans le cas de l'échantillon n°1, qui regroupe un plus grand nombre de points.

¹⁹ Les histogrammes des valeurs propres de la matrice d'inertie sont par ailleurs assez réguliers. Il faut noter que dans le cas de l'analyse des correspondances multiples, le niveau absolu de ces pourcentages sous-estime beaucoup la quantité d'information statistique résumée par chaque axe, contrairement à l'analyse en composantes principales pour laquelle la somme des pourcentages est égale à 100 %. Cf. Benzécri, 1973 et Volle, op. cit.

sphériques, et se livrent donc difficilement à l'extraction de corrélations ; cet élément devrait handicaper les analyses économétriques. Au-delà de ce point commun, la structure d'information des deux échantillons se révèle assez différente, telle qu'elle apparaît à la lecture des contributions et des coordonnées factorielles des variables. Cette différence laisse augurer que les équations de croissance correspondantes seront distinctes, notamment pour les effets de l'indice de Gini, entre les estimations avec indicatrices de données manquantes (échantillon n°1) et les estimations apurées (échantillon n°2). Bien entendu, il n'est pas immédiat de formuler des prédictions concernant l'analyse économétrique, qui, contrairement à l'analyse descriptive, procède du principe "*ceteris paribus*"²⁰. On verra néanmoins apparaître, dans les sections suivantes, certaines des régularités identifiées par l'analyse des correspondances multiples.

Dans le cas de l'échantillon n°1, le premier plan factoriel est structuré par un fort "effet Gutmann" (cf. Volle, 1981) associé à la croissance du PIB par habitant : la corrélation négative entre l'accélération contemporaine du PIB et la croissance de la période antérieure produit le premier axe factoriel, le second axe opposant les valeurs intermédiaires de ces deux variables aux valeurs extrêmes. Il est important de se souvenir qu'il s'agit d'une analyse de périodes de croissance, chaque pays ayant pu connaître au cours des trente dernières années des phases de croissance forte ou faible, stable ou instable. Les pays de l'OCDE sont toutefois plutôt caractérisés par une croissance modérée et relativement stable, ils ont donc tendance à se situer dans le quadrant "nord-est" du graphe. En revanche, le groupe des pays en développement est plus hétérogène et se situe en moyenne au centre du premier plan. Au sein de ce groupe, les pays d'Afrique sub-saharienne occupent en moyenne une position "ouest-sud-ouest", associée à un niveau faible de croissance économique accompagné de "booms" transitoires, à une faible progression de la scolarisation et à la première phase de la transition démographique. Les pays asiatiques à forte croissance comme la Corée du Sud, la Chine, la Malaisie et Singapour, se situent inversement en position "sud-est", associés à une croissance forte, une forte progression du taux de scolarisation secondaire, et surtout de fortes hausses du taux d'investissement. En ce qui concerne la variation de l'indice de Gini, il est intéressant de noter que la modalité de baisse accentuée (plus de deux points) contribue à la formation du premier axe factoriel, associée à une croissance soutenue et à une forte progression de la scolarisation. Toutefois, les autres modalités de la variation de l'indice de Gini ne s'ordonnent pas régulièrement le long de ce premier axe, à cause de l'autre modalité extrême (hausse de plus de 2 points) qui revient fortement vers le centre du graphe et le quadrant "sud-est" (pays asiatiques à forte croissance). De même, les données manquantes se caractérisent plutôt par un profil moyen, même si elles sont plutôt associées au quadrant "sud-ouest", comme les pays d'Afrique sub-saharienne. En définitive, s'il devait exister une relation de causalité allant des inégalités vers la croissance, l'analyse descriptive de l'échantillon n°1 nous suggère qu'elle serait plutôt négative, une baisse des inégalités étant associée à une plus forte croissance, mais également dissymétrique, une baisse des inégalités étant plus favorable à la croissance qu'une hausse lui est nuisible. Cette hypothèse est testée dans la section 4.

La structure de corrélations de l'échantillon n°2 est assez différente. Ici les variations de l'indice de Gini ne semblent associées à aucun schéma de croissance particulier. On retrouve en revanche une forte association entre les progrès de la scolarisation secondaire et la baisse des inégalités, déjà notable dans l'échantillon n°1. Ce lien de causalité²¹ sera testé dans les sections 6 et 7.

²⁰ De surcroît, les méthodes de variables instrumentales appliquées dans la suite prétendent se rapprocher de l'analyse de causalité.

²¹ Plus précisément, même si l'ordre des retards des deux variables DGINI(-1) et DSECR(-2) suggère une causalité allant de la scolarisation vers les inégalités, le lien causal inverse est néanmoins considéré dans la section 6, parce qu'il correspond à une hypothèse théorique d'imperfection des marchés du crédit qui a pris beaucoup d'importance dans la littérature économique récente.

4. DES INEGALITES NUISIBLES OU FAVORABLES A LA CROISSANCE ?

Concernant l'impact des inégalités sur la croissance, comme chez Fishlow (1996) et Forbes (1999), le lien de causalité apparaît positif sur l'échantillon n°2, contrairement à la plupart des résultats antérieurs, obtenus sur d'autres bases de données. Ce lien apparaît en différences premières et non en niveau, et il est renforcé par l'instrumentation. L'ordre de grandeur de l'effet est assez conséquent : 1 point supplémentaire d'indice de Gini engendre en moyenne entre 0.20 et 0.3 points de croissance supplémentaire du produit par tête²². Une variation d'un écart-type intertemporel (égal à 3.0) augmente donc la croissance de la période suivante de 0.6 à 0.9 points, toutes choses égales par ailleurs. Il convient toutefois de rester prudent sur le résultat obtenu. Il peut provenir d'une variable omise, comme par exemple une variable de termes de l'échange qui affecterait simultanément le revenu national et sa distribution²³, mais aussi de la spécificité de l'échantillon.

Tableau n° 4-1 : Equations de croissance

Variable dépendante :	Echantillon 1			Echantillon 2		
	OLS levels	OLS FD (1 st diff.)	GMM(5)	OLS levels	OLS FD (1 st diff.)	GMM(3)
GROWTH						
Constante	-0.0600	+0.0222**	+0.0152*	-0.1063	+0.0078	+0.0034
LGDP(-1)	-0.0042	-0.1410**	-0.0858** (0.0347)	-0.0084	-0.0947**	-0.0837** (0.0094)
LGPOPC	-0.0111	-0.0007	+0.0194 (0.0068)	-0.0201	-0.0728	-0.1696** (0.0186)
LSECR(-2)	+0.0007	-0.0057	-0.0112 (0.0021)	-0.0131	+0.0003	+0.0020 (0.0057)
LINV	+0.0145**	+0.0349**	+0.0233* (0.0027)	+0.0269**	+0.0323*	+0.0280* (0.0077)
GINI(-1)	-0.0146	-0.0291	-0.0338 (0.0064)	+0.0139	+0.1993*	+0.2919* (0.0542)
Wald test	51.7 (.000)	106 (.000)	1474(.000)	13.5(.019)	45.3(.000)	260(.000)
Sargan test	-	-	80.3 (.374)	-	-	30.0 (.749)
2 nd order correlation test	-	1.74 (.082)	.815 (.415)	-	-.91 (.362)	-.53 (.598)
N		387 (90)			92 (37)	

*: 10 % ; ** : 1 % (voir note GMM)

Estimateurs et tests robustes à l'hétéroscédasticité

Variables indicatrices de périodes (éch. 1 et 2) et de données manquantes pour l'indice de Gini (éch.1) incluses mais non reportées

GMM(n) : instrumentation des différences premières par les niveaux des variables retardés de 2 à (n) périodes. Pour cet estimateur, les étoiles indiquent la significativité des estimateurs de première étape, les écartstypes de seconde étape étant potentiellement sous-estimés à distance finie, comme le montrent Arellano et Bond dans leur étude par simulation. Les écartstypes de seconde étape sont indiqués entre parenthèses.

²² Remplacer l'indice de Gini par son logarithme donne le même type de résultats.

²³ L'introduction du niveau du prix du PIB en dollars internationaux, reflétant le taux de change réel, n'a donné aucun résultat. De la même façon, le taux d'inflation n'est pas apparu comme un facteur explicatif des performances de croissance.

Tableau n° 4-2 : Test des spécifications non-linéaires de Banerjee et Duflo

Variable dépendante : GROWTH		OLS FD (1 st diff.)		
GINI(-1)	+0.3764 (0.3429)		+0.2710* (0.1128)	+0.2689* (0.1122)
GINI(-1) ²	-0.2224 (0.4173)			
GINI(-1)-GINI(-2)		+0.0901 (0.0680)	+0.0899 (0.0680)	-0.0667 (0.0835)
(GINI(-1)-GINI(-2)) ²			-0.0631 (1.6723)	+0.0802 (1.6091)
(GINI(-1)-GINI(-2))* (GINI(-1)-GINI(-2))>0				+0.0524 (0.1311)
(GINI(-1)-GINI(-2))* (GINI(-1)-GINI(-2))<0				-0.0370 (0.1414)
				+0.1266 (0.1266)
				-0.0952 (0.1305)
N			92 (37)	

*: 10 % ; ** : 1 % (voir note GMM)

Estimateurs et tests robustes à l'hétéroscédasticité

Variables de contrôle identiques à celles du tableau 1: LGDP(-1), LSECR(-2), LINV et LGPOPC, non reportées

Variables indicatrices de périodes incluses mais non reportées

GMM(n) : voir note du tableau 4.1.

Tableau n° 4-3 : Test de spécifications non-linéaires pour l'indice de Gini dans l'équation de croissance

Variable dépendante : GROWTH		GMM (5)		
GINI(-1)	-0.6041 (0.0824)		-0.0413 (0.0073)	-0.6165 (0.0917)
GINI(-1) ²	+0.6610 (0.0958)			+0.6686 (0.1105)
GINI(-1)* (GINI(-1)-0.457)>0		+0.0598 (0.0127)		
GINI(-1)* (GINI(-1)-0.457)<0		-0.0474 (0.0085)		
GINI(-1)-GINI(-2)		-0.0026 (0.0266)	+0.0434 (0.0261)	+0.0300 (0.0294)
(GINI(-1)-GINI(-2)) ²		+1.6486 (0.6212)	+1.4999 (0.7211)	+1.9868 (0.8148)
N			387 (90)	

*: 10 % ; ** : 1 % (voir note GMM)

Estimateurs et tests robustes à l'hétéroscédasticité

Variables de contrôle identiques à celles du tableau 1: LGDP(-1), LSECR(-2), LINV et LGPOPC, non reportées

Variables indicatrices de périodes (éch. 1 et 2) et de données manquantes pour l'indice de Gini (éch.1) incluses mais non reportées

GMM(n) : voir note du tableau 4.1. La batterie d'instruments utilisée dans ce cas ne comprend que les niveaux de l'indice de Gini retardés de deux à cinq périodes et les niveaux retardés de l'indicatrice de donnée manquante, mais aucun terme quadratique ou tronqué, ni aucun terme en différence première.

En effet, le signe de la causalité est renversé lorsqu'on considère l'échantillon n°1: les inégalités apparaissent nuisibles à la croissance (moitié gauche du tableau 4.1). Cette fois le paramètre n'est significatif que dans le cas de l'estimateur GMM de seconde étape²⁴, or nous avons déjà signalé le risque d'inférence optimiste associé à cet estimateur, ressortant des simulations d'Arellano et Bond. De surcroît, l'équation de croissance est de mauvaise qualité, et l'effet de la scolarisation

²⁴ Une instrumentation plus faible avec des niveaux retardés de deux périodes seulement fait disparaître la significativité et le signe de ce paramètre.

secondaire²⁵ et du taux de croissance de la population ont des signes contraires aux signes attendus. Cette différence entre les deux échantillons était prévisible, sur la base des analyses des correspondances multiples exposées dans la section précédente.

Banerjee et Duflo contestent le résultat de Forbes à partir d'autres arguments. Ils mettent en avant un effet non-linéaire des variations de l'indice de Gini. Nous avons voulu tester leur spécification, en commençant avec l'échantillon n°2 qui ressemble fortement au leur. Le nombre de retards élevés et la faible taille de l'échantillon ne nous autorise pas à mettre en œuvre l'estimateur d'Arellano et Bond. Nous nous sommes donc restreints à l'estimateur des moindres carrés ordinaires en différences premières²⁶. Celui-ci rejette la significativité des spécifications non-linéaires, en maintenant le résultat de Forbes (tableau 4.2). Sur l'échantillon n°1, l'instrumentation GMM est possible. Aucune des spécifications de Banerjee et Duflo ne passe les tests de significativité. Un effet non-linéaire du niveau de l'indice de Gini est en revanche accepté (tableau 4.3). Il renvoie au résultat de l'analyse des correspondances multiples sur l'échantillon n°1. Cette spécification quadratique est significative à 15 % en retenant les écarts-types de l'estimateur de première étape. Elle suggère que les inégalités seraient nuisibles à la croissance seulement jusqu'à un certain niveau de l'indice de Gini, estimé autour de 0.457. Notons que cette causalité n'est pas maintenue sur l'échantillon n°2, en différences premières (tableau 4.2, première colonne).

5. UN MODELE KALDORIEN D'INVESTISSEMENT ?

Comme pour la croissance, sur l'échantillon réduit, la dimension temporelle des données révèle un effet positif des inégalités sur l'investissement en capital physique. Ce résultat s'inscrit en contradiction des modèles de marché du crédit imparfait évoqués en introduction. Il est en revanche cohérent avec un modèle kaldorien d'investissement fondé sur des différentiels de taux d'épargne entre revenus du capital et du travail, ou sur un taux d'épargne des ménages croissant avec le revenu. Là encore, les biais corrigés par l'estimateur GMM apparaissent très importants, et s'appliquent à toutes les variables. Cet estimateur conduit à une estimation raisonnable de l'hystérésis du taux d'investissement, et atténue fortement l'ampleur de l'effet accélérateur-demande, tout en augmentant l'effet des inégalités. L'ampleur de ce dernier effet est néanmoins modeste, puisque, selon le coefficient de la dernière colonne de droite du tableau 5.1, une variation d'un écart-type intertemporel de l'indice de Gini entraîne, *ceteris paribus*, une hausse de 9% du taux d'investissement.

Par ailleurs, cet effet disparaît complètement lorsqu'on effectue l'estimation sur l'échantillon n°1, même si le coefficient de l'indice de Gini demeure positif. Les tests sont donc plutôt favorables au modèle kaldorien, mais sans le corroborer de façon probante.

²⁵ A l'instar de Barro, Forbes et Banerjee et Duflo utilisent comme variables de contrôle deux variables de stock capital humain différenciées par sexe. Ces deux variables sont toutefois très fortement corrélées, dans la dimension transversale comme dans la dimension longitudinale. Cette collinéarité peut d'ailleurs expliquer pourquoi les coefficients de ces deux variables sont toujours très significatifs et affectés de signes opposés.

²⁶ Banerjee et Duflo proposent des estimations avec effets individuels aléatoires (moindres carrés quasi-généralisés) dont nous contestons la pertinence. Mais ils fournissent également des estimations en différences premières et "within" ("fixed effects").

Tableau n° 5-1 : Equation d'investissement

Variable dépendante :	Echantillon 1		Echantillon 2	
	OLS levels	GMM(6)	OLS levels	GMM(2)
LINV				
Constante	+0.3234**	+0.0562	+0.2226	+0.0524
LINV(-1)	+0.9015**	+0.5679** (0.0478)	+0.9290**	+0.2185 (0.0780)
GROWTH(-1)	+2.1955**	+1.4892 (0.2964)	+3.2830**	+1.1604 (0.8571)
GINI(-1)	-0.0931	+0.1007 (0.1725)	-0.0822	+2.9614** (0.6538)
Wald test	1869(.000)	240(.000)	339(.000)	78.4 (.000)
Sargan test	-	61.5 (.253)	-	14.9 (.244)
2 nd order correlation test	-	-1.3 (.194)	-	1.20 (.232)
N	387 (90)		92 (37)	

*: 10 % ; ** : 1 % (voir note GMM)

Estimateurs et tests robustes à l'hétéroscédasticité

Variables indicatrices de périodes (éch. 1 et 2) et de données manquantes pour l'indice de Gini (éch.1) incluses mais non reportées

GMM(n) : voir note du tableau 4.1

6. DES INEGALITES FAVORABLES A LA SCOLARISATION ET A LA BAISSSE DE LA FECONDITE ?

Dans le cas de la scolarisation, les modèles d'imperfection du crédit sont également contredits par les résultats du tableau 6.1. Ces modèles semblaient pourtant particulièrement susceptibles de faire surgir une relation négative entre les inégalités et les investissements humains, lesquels sont sujets à des coûts fixes importants et à des asymétries d'information élevées. Ainsi dans la dimension transversale, les inégalités apparaissent effectivement corrélées négativement aux taux de scolarisation : à niveau de PIB par habitant égal, les pays les plus inégalitaires sont également ceux où les taux de scolarisation secondaire sont les moins élevés²⁷.

Tableau n° 6-1 : Equation de scolarisation secondaire

Variable dépendante :	Echantillon 1		Echantillon 2	
	OLS levels	GMM(6)	OLS levels	GMM(2)
LSECR				
Constante	+2.3823**	+0.0441	+1.4214*	-0.0290
LPRIMR(-1)	+0.8894**	+0.4683** (0.0486)	+1.0356**	+0.2860 (0.1118)
LGDP(-1)	+0.4256**	+0.5777* (0.0588)	+0.2800**	+0.6442** (0.2091)
GINI(-1)	-1.1410**	+0.6516 (0.1255)	-1.9043**	+1.2734 (0.6611)
Wald test	1197(.000)	854(.000)	192 (.000)	52.7(.000)
Sargan test	-	49.5(.684)	-	13.1 (.364)
2 nd order correlation test	-	.347 (.728)	-	-.95(.343)
N	387 (90)		92 (37)	

*: 10 % ; ** : 1 % (voir note GMM)

Estimateurs et tests robustes à l'hétéroscédasticité

Variables indicatrices de périodes (éch. 1 et 2) et de données manquantes pour l'indice de Gini (éch.1) incluses mais non reportées

GMM(n) : voir note du tableau 4.1

²⁷ Cette corrélation conditionnelle se vérifie pas pour les taux de scolarisation primaire.

Dans la dimension longitudinale, les estimateurs GMM font au contraire apparaître un effet positif et significatif des inégalités de revenu sur la scolarisation. Là encore, ils vont plutôt dans le sens d'un modèle kaldorien de taux d'épargne. L'effet estimé est toutefois de faible ampleur, se situant entre 1.5 et 6 % de croissance du taux de scolarisation pour une variation de l'indice de de Gini d'un écart type inter-temporel (3 points). Et sur l'échantillon n°1, il perd sa significativité lorsqu'on retient l'écart-type de première étape du coefficient²⁸. Des estimations similaires ont été effectuées pour la scolarisation primaire. L'effet des inégalités est apparu à chaque fois positif mais non significatif.

Tableau n° 6-2 : Equation du taux de fécondité

Variable dépendante : LFERT	Echantillon 1		Echantillon 2	
	OLS levels	GMM(6)	OLS levels	GMM(2)
Constante	+0.2864**	-0.0672	+0.2196	-0.0166
LFERT(-1)	+0.9202**	+0.8298**	+0.9618**	+0.9181**
LGDP(-1)	-0.0036	+0.1311*	+0.0324	+0.0226
LSECR(-2)	-0.0638**	+0.0321	-0.0108	-0.0741
GINI(-1)	-0.1928*	-0.2940	-0.1499	-0.4224
		(0.0165)		(0.0759)
		(0.0144)		(0.1112)
		(0.0073)		(0.0414)
		(0.0296)		(0.4358)
Wald test	13371(.00)	5758(.000)	1043(.000)	1314(.000)
Sargan test	-	80.4(.164)	-	19.5 (.242)
2 nd order correlation test	-	0.07 (.948)	-	0.16 (.876)
N	387 (90)		92 (37)	

*: 10 % ; ** : 1 % (voir note GMM)

Estimateurs et tests robustes à l'hétéroscédasticité

Variables indicatrices de périodes (éch. 1 et 2) et de données manquantes pour l'indice de Gini (éch.1) incluses mais non reportées

GMM(n) : voir note du tableau 4.1

Pour finir, le tableau 6 présente une équation pour le taux de fécondité²⁹. Cette variable manifeste l'inertie la plus forte, comme le montre le coefficient GMM de l'effet d'hystérésis, égal à 0.83 sur l'échantillon n°1. Les résultats concernant les autres variables sont assez décevants car peu robustes. Le signe des inégalités de revenu est assez surprenant, en comparaison des résultats antérieurs de Perotti (1996). Toutefois le coefficient est d'amplitude très modeste, et se révèle la plupart du temps non significatif³⁰.

7. UNE CROISSANCE EGALITAIRE ?

Nous terminons l'exposé de nos résultats par l'examen d'une causalité inverse, allant de la croissance ou des variables de développement humain vers les inégalités de revenu. Pour traiter le cas des données manquantes de la variable d'inégalités, devenue cette fois la variable à expliquer, il aurait fallu concevoir une procédure spécifique de redressement du biais de sélection. Quel que soit le type d'estimateur retenu, paramétrique "à la Heckman" (1976), ou semi-paramétrique comme

²⁸ Par ailleurs, on notera encore une fois que l'estimateur GMM corrige fortement, dans ces équations agrégées, l'effet d'hystérésis et l'effet du PIB. Dans nos estimations, le niveau de revenu semble mieux expliquer la scolarisation que l'inverse (voir sur ce thème Bils et Klenow, 1998).

²⁹ Les mêmes estimations ont été effectuées pour la variable LGPOPC et fournissent des résultats qualitativement similaires.

³⁰ Les signes observés pour l'effet des inégalités sur la scolarisation et la fécondité sont cohérents avec le modèle de Becker, Murphy et Tamura (1990), où l'accumulation scolaire, c'est-à-dire le choix de la "qualité" des enfants au détriment de leur quantité, vont de pair avec la croissance des rendements de l'éducation, et donc des inégalités plus fortes, pour une répartition donnée du capital humain. Pour tirer ces conclusions, il faudrait toutefois pouvoir considérer que la variation temporelle des rendements de l'éducation explique une part importante de la variation des inégalités totales. Or, cette dernière hypothèse est plus que douteuse.

chez Lewbel (2000), le problème provient principalement du choix d'une variable identifiante, intervenant dans l'équation de sélection (expliquant l'absence de données pour l'indice de Gini) sans intervenir dans l'équation expliquant les inégalités. Aucune variable disponible ne s'étant imposée à nous, nous ne travaillons qu'avec l'échantillon n°2.

Le premier fondement théorique de la causalité inverse est celui de la courbe de Kuznets (1955). Dans les régressions transversales, l'effet d'hystérésis introduit écrase la significativité de l'effet quadratique du PIB par habitant, même si le signe des coefficients confirme la forme en U inversé (tableau 7.1).

En revanche, la dimension temporelle des données ne corrobore que la seconde partie de la courbe de Kuznets, c'est-à-dire une baisse des inégalités accompagnant la croissance économique. Ce résultat est en cohérence avec le celui obtenu par Fields et Jakubson (1994) sur un autre échantillon et avec un estimateur "*within*". L'indice de Gini étant borné entre 0 et 1, nous avons vérifié la robustesse du résultat en effectuant la même estimation sur une transformation de cette variable (moitié droite du tableau 6.2). Au point moyen de l'échantillon, les coefficients de la parabole sont identiques et le point de retournement très proche. Toutes les estimations introduisent un effet fixe temporel, ce qui fait qu'il n'est pas nécessaire de "détrender" la variable de PIB par habitant. La courbe estimée se déplace en fait d'une constante sur l'axe des abscisses, avec le trend de croissance mondial³¹.

Tableau n° 7-1 : Estimations d'une courbe de Kuznets

Variable dépendante :	GINI		f(GINI)= ln(GINI)-ln(1-GINI)	
	OLS levels	GMM(3)	OLS levels	GMM (3)
Constante	+0.0354*	+0.0211*	-0.8516	+0.0816*
GINI(-1) ou f(GINI(-1))	+0.8852**	+0.2159 (0.0868)	+0.8504**	+0.1116 (0.1466)
GDP(-1)	+0.0231	-0.0292** (0.0063)	+0.0060	-0.1077** (0.0242)
GDPSQ(-1)/100	-0.0190	+0.0660* (0.0210)	-0.0530*	+0.2542* (0.0800)
PIB / hab. de retournement	-	22 121	-	21 184
Wald test	907 (.000)	29.8 (.000)	9.5 (.024)	28.9 (.000)
Sargan test	-	19.5 (.191)	-	13.5 (.563)
2 nd order correlation test	-	-.13(.896)	-	-.56 (.588)
N		92 (37)		

* : 10 % ; ** : 1 % (voir note GMM)

Estimateurs et tests robustes à l'hétéroscédasticité

Variables indicatrices de périodes incluses mais non reportées

GMM(n) : voir note du tableau 4.1

Les estimations signalent même l'existence d'une demie courbe en U ("non-inversée"), le point de retournement de la parabole se situant au bout de l'échantillon, c'est-à-dire autour du niveau du PIB par habitant des Etats-Unis d'Amérique au début des années 1990. L'absence de pays très pauvres dans l'échantillon pourrait expliquer que la première branche de la véritable courbe de Kuznets en U inversé ne puisse être identifiée. Dès lors, rien n'interdit de penser que les inégalités connaissent des cycles longs, rythmés par les révolutions technologiques, à la manière des cycles de Kondratieff. En

³¹ Un coup d'œil sur les effets-fixes temporels révèle également une augmentation du niveau moyen des inégalités au cours du temps, avec un premier saut au cours de la période 1976-80 et un second saut au cours de la période 1986-90.

effet, même si la différenciation temporelle des variables privilégie des relations de moyen terme, les estimateurs GMM conduisent à rejeter la significativité de l'effet d'hystérésis, si bien que les élasticités de moyen terme et de long terme apparaissent identiques. Ces élasticités sont toutefois assez modestes : au point moyen de l'échantillon (9000 \$ par habitant, voir annexe 2), une hausse de 10 % du PIB diminue l'indice de Gini de 0.2 points. Au point extrême de l'échantillon (Etats-Unis), comme on l'a déjà noté, l'élasticité est nulle.

Enfin, un autre modèle où la variation des inégalités dans le temps serait spécifique à chaque pays serait peut-être plus conforme à l'article de Kuznets de 1955 sur les données anglaises et allemandes³². Anand et Kanbur (1993a) envisagent cette hypothèse. Un tel modèle n'est évidemment pas testable ici, étant donné le faible nombre d'observations par pays. De toutes façons, la question de l'interprétation de la causalité mise à jour reste entière.

Pour essayer de progresser dans cette direction, nous avons ajouté à la courbe de Kuznets simple le taux de scolarisation secondaire, le taux de fécondité, et l'indicateur de dualisme de Bourguignon et Morisson (1999), c'est-à-dire le rapport de productivité du travail entre l'agriculture et les autres secteurs (tableau 7.2). Comme chez Higgins et Williamson (1999), la forme en U inversé de la courbe de Kuznets réapparaît nettement, dans les régressions transversales augmentées comprenant la fécondité³³. Cette dernière variable est d'ailleurs fortement corrélée à l'indice de Gini pour la variance inter-individuelle, avec un coefficient de corrélation de plus de 0.6, comme le montre le coefficient de l'équation en niveau sans effet d'hystérésis.

Tableau n° 7-2 : Estimations d'une courbe de Kuznets augmentée

Variable dépendante :	GINI	
	OLS levels	GMM (2)
Constante	-0.1323	+0.0284
GDP(-1)	+0.0305**	-0.0184 (0.0067)
GDPSQ(-1)/100	-0.1157**	+0.0367 (0.0275)
LSECR(-2)	-0.0125	-0.0798 (0.0151)
LFERT(-1)	+0.2167**	-0.0140 (0.0296)
LRLP	+0.0567**	+0.0554* (0.0153)
PIB / hab. de retournement	13 200	25 100
Wald test	113 (.000)	186 (.000)
Sargan test	-	11.4 (.651)
2 nd order correlation test	-	.081 (.936)
N	92 (37)	

* : 10 % ; ** : 1 % (voir note GMM)

Estimateurs et tests robustes à l'hétéroscédasticité

Variables indicatrices de périodes et de données manquantes pour l'indicateur de dualisme (LRLP) incluses mais non reportées

GMM(n) : voir note du tableau 4.1

³² De même, des données historiques confirment la baisse continue des inégalités en Italie depuis la fin du XIX^{ème} siècle (Rossi, Toniolo et Vecchi, 1999).

³³ Higgins et Williamson considèrent en fait la structure de la population par âge, en montrant qu'un fort degré d'inégalités est associé à un poids élevé des classes d'âge jeunes. Cette variable est néanmoins très corrélée à la fécondité ; il n'est donc pas possible de faire la différence entre les deux variables avec les données disponibles.

Dans la dimension longitudinale des données, la significativité du PIB par habitant et de la fécondité disparaissent, tandis que l'effet de la variable de dualisme subsiste³⁴. Le coefficient de la scolarisation secondaire est fortement ré-estimé à la hausse par rapport aux régressions transversales. La dimension de ces deux effets n'est pas négligeable. Une baisse de 10 % du "dualisme agricole" entraîne, toutes choses égales par ailleurs, une baisse de l'indice de Gini d'un demi-point. Une hausse du taux de scolarisation de 10 % entraîne une baisse de l'indice de Gini de près d'un point. Au niveau macro-économique les politiques publiques de développement de la scolarisation universelle contribuent sans doute à la réduction des inégalités de revenu et de la pauvreté monétaire. Par ailleurs, même s'il ne fait pas de doute que la croissance est "bonne pour les pauvres" (Dollar et Kraay, 2000), elle n'est probablement pas suffisante. Au mieux, elle diminue légèrement les inégalités (voir commentaires du tableau 6.2) et de manière décroissante au fur et à mesure du processus de développement.

CONCLUSION

Le lecteur retirera sans doute une impression dominante de perplexité. Comme le concluent Hoff et Stiglitz (1999) à propos de la théorie économique du développement, la certitude socratique de notre ignorance est pour le moins établie. Cependant, cette "ignorance avertie" ne débouche pas sur un scepticisme généralisé. En effet, les voies de la recherche émergent assez clairement.

Comme nous l'avons rappelé en introduction, la littérature théorique récente s'attache à examiner la détermination concomitante de la plupart des variables fondamentales du développement : niveau de revenu, investissement physique et humain, démographie, mais également répartition du revenu et des ressources.

Les données empiriques disponibles ne rejettent pas l'idée générale d'une codétermination ; toutefois, les tests économétriques ne permettent pas de corroborer une influence décisive des inégalités de revenu sur les autres variables. A ce propos, la sélection des échantillons, relativement négligée par les travaux empiriques antérieurs, s'avère avoir des conséquences problématiques.

Il semble un peu mieux assuré que le processus de développement pris dans son ensemble favorise une réduction des inégalités de revenu, sans qu'il soit là encore possible de démêler, au niveau macro-économique, les différents canaux par lesquels cette causalité s'organise – transition démographique, développement humain et institutionnel, croissance... Et là encore, la question de l'universalité de cette relation demeure posée, à cause de la sélection des échantillons.

A l'heure actuelle et en ce domaine, on doit admettre que les travaux sur données macro-économiques atteignent leurs limites. Premièrement, la qualité des données continue de poser problème. Dans le cas des inégalités, n'importe quel statisticien d'enquête sait à quel point les calculs d'indices d'inégalité sont sensibles aux données de base et aux méthodes d'estimation. Mais le comptable national ou le spécialiste des données sociales sait aussi les incertitudes qui pèsent sur les données plus traditionnelles concernant la croissance ou les autres variables du développement socio-économique. Deuxièmement, les estimations menées sur des données transversales sont particulièrement sujettes à caution, quant aux erreurs de spécification qu'elles peuvent recouvrir, et aux problèmes de biais. Cependant, même si une analyse raisonnée des données longitudinales conduit à remettre en cause la plupart des constats obtenus par l'analyse transversale, celle-ci s'avère également limitée par la qualité et la sélection des données. Enfin, quelles que soient les "régularités planétaires" qui pourraient être découvertes, elles laisseraient encore la place à de nombreuses originalités locales. La relation inégalités-croissance demeure conditionnée par les

³⁴ Malgré la sélection de l'échantillon, qui élimine un grand nombre de pays à spécialisation agricole.

stratégies de développement nationales et les configurations structurelles et historiques de chaque société.

REFERENCES BIBLIOGRAPHIQUES

- Adelman I., Morris C.T. (1973), "Economic Growth and Social Equity in Developing Countries", *Stanford University Press*, California.
- Aghion P., Bolton P. (1997), "A Theory of Trickle-Down Growth and Development", *Review of Economic Studies*, 64, 151-172.
- Ahluwalia M. (1976), "Inequality, Poverty and Development", *Journal of Development Economics*, 6, 307-342.
- Alesina A., Rodrik D. (1994), "Distributive Politics and Economic Growth", *The Quarterly Journal of Economics*, 104, 465-490.
- Anand S., Kanbur S.M.R. (1993a), "The Kuznets Process and the Inequality-Development Relationship", *Journal of Development Economics*, 40, 25-40.
- Anand S., Kanbur S.M.R. (1993b), "Inequality and development, A critique", *Journal of Development Economics*, 41, 25-40.
- Arellano M., Bond S. (1991), "Some Tests of Specification for Panel Data : Monte Carlo Evidence and an Application to Employment Equations", *Review of Economic Studies*, 58, 277-297.
- Arellano M., Bond S. (1998), *Dynamic Panel Data Estimation Using DPD98 for Gauss : A Guide for Users*, mimeo, 27 pp.
- Banerjee A.V., Newman A.F. (1991), "Risk-bearing and the Theory of Income Distribution", *Review of Economic Studies*, 58, 211-235.
- Banerjee A.V., Newman A.F. (1993), "Occupational Choice and the Process of Development", *Journal of Political Economy*, 101, 274-299.
- Banerjee A.V., Duflo E. (2000), "Inequality and Growth: What Can the Data Say?", *NBER WP 7793*.
- Barro R. (1999), "Inequality, growth and investment", *NBER WP 7038*.
- Barro R.J., Lee J.W. (1993), "International Comparisons of Educational Attainment", *Journal of Monetary Economics*, 32, 363-394.
- Becker G.S., Murphy K.M., Tamura R. (1990), "Human capital, Fertility and Economic Growth", *Journal of Political Economy*, 98, n°5, 12-37.
- Bénabou R. (1994), "Human capital, inequality and growth: a local perspective", *European Economic Review*, 38, 817-826.
- Bénabou R. (1996), "Inequality and Growth", *NBER WP 5658*.
- Benzécri J.-P. (1973), *L'analyse de données*, t.1 et t.2, Dunod, Paris
- Bils M., P. J. Klenow (1998), "Does Schooling Cause Growth or the Other Way Around ?" *NBER Working Paper 6393*, 41 pp.
- Birdsall N., Ross D., Sabot R. (1995), "Inequality and Growth Reconsidered: Lessons From East Asia", *The World Bank Economic Review*, 9(3), 347-508.
- Bourguignon F. (1981), "Pareto Superiority of Unegalitarian Equilibria in Stiglitz' Model of Wealth Distribution with Convex Saving Function", *Econometrica*, 49(6), 1469-1475
- Bourguignon F. (1990), "Growth and inequality in the dual model of development : the role of demand factors", *Review of Economic Studies*, 57, 215-228.
- Bourguignon F. (1993), « Croissance, distribution et ressources humaines: comparaison internationale et spécificités régionales », *Revue d'économie du développement*, 4, 3-35.
- Bourguignon F. (1996), "Equity and Economic Growth: Permanent questions and changing

- answers?», *Document de travail DELTA-EHESS* n°96/15, August.
- Bourguignon F. (1998), « Equité et croissance économique : une nouvelle analyse ? », *Revue Française d'Economie*, 3, 25-84.
- Bourguignon F., Morrisson C. (1990), "Income distribution, Development and Foreign Trade: a Cross-Sectional Analysis", *European Economic Review*, 34, 1113-1132.
- Bourguignon F., Morrisson C. (1998), "Inequality and Development: The Role of Dualism", *Journal of Development Economics*, 57(2), 233-258.
- Bourguignon F., Verdier T. (2000), "Oligarchy, Democracy, Inequality and Growth", *Journal of Development Economics*, 62(2), 285-314
- Caselli F., Esquivel G., Lefort F. (1996), "Reopening the Convergence Debate : A New Look at Cross-Country Growth Empirics", *Journal of Economic Growth*, 1, 363-389.
- Clarke G.R.G. (1995), "More Evidence On Income Distribution and Growth", *Journal of Development Economics*, 47(2), 403-427.
- Dahan M., D. Tsiddon (1998), "Demographic Transition, Income Distribution, and Economic Growth", *Journal of Economic Growth*, 3(1), 29-52
- Deininger K., Squire L. (1996), "A New Data Set Measuring Income Inequality", *The World Bank Economic Review*, 10(3), 565-591.
- Deininger K., Squire L. (1998), "New Ways of Looking at Old Issues: Inequality and Growth", *Journal of Development Economics*, 57, 259-287.
- Dollar D., Kraay A. (2000), *Growth Is Good for the Poor*, mimeo, The World Bank, 43 pp.
- Durlauf S.N. (1996), "A theory of persistent inequality", *Journal of Economic Growth*, 1(1), 75-93.
- Durlauf S.N. (1994), "Spillovers, stratification and inequality", *European Economic Review*, 38, 836-845
- Fields G.S., Jakubson (1994), "New evidence on the Kuznets Curve", *Cornell University Working Paper*, mimeo.
- Fishlow A. (1996), "Inequality, Poverty and Growth: Where Do We Stand?, in Bruno M., B. Pleskovic (eds)", *Annual World Bank Conference on Development Economics 1995*, The World Bank.
- Forbes K.J. (1999), "A Reassessment of the Relationship Between Inequality and Growth", *MIT Working Paper*, mimeo, 26 pp.
- Galor O., Tsiddon D. (1997), "The Distribution of Human Capital and Economic Growth", *Journal of Economic Growth*, 2(1), 93-124
- Galor O., Zeira J. (1993), "Income distribution and macroeconomics", *Review of Economic Studies*, 60, 35-52.
- Galor O., Zang H. (1997), "Fertility, income distribution, and economic growth: Theory and cross country evidence", *Japan and the World Economy*, 9(1997), 197-229.
- Heckman J. (1979), *Sample Selection Bias as a Specification Error*, *Econometrica*, 47, 153-161.
- Higgins M., Williamson J.G. (1999), "Explaining Inequality the World Around : Cohort Size, Kuznets Curves and Openness", mimeo, 44 pp.
- Hoff K., J. Stiglitz (1999), *Modern Economic Theory and Development*, mimeo, The World Bank, november, 70 pp.
- Kaldor N. (1956), "Alternative Theories of Distribution", *Review of Economic Studies*, 23(2), 94-100.
- Kuznets S. (1955), "Economic Growth and Income Inequality", *American Economic Review*, 45(1), 1-28.
- Lee W., J. Roemer (1997), "Income Distribution, Redistributive Politics and Economic Growth", *Journal of Economic Growth*, 3(3), 217-240

- Lewbel A. (2000), "Two Stage Least Squares Estimation of Endogeneous Sample Selection Threshold, or Censoring Model", mimeo, Boston College, 32 pp.
- Li H., Squire L., Zou H. (1998), "Explaining International and Intertemporal Variations in Income Inequality", *Economic Journal*, 108, 26-43.
- Loury G. (1981), *Intergenerational Transfers and the Distribution of Earnings*, *Econometrica*, 49, 843-867
- Mankiw N.G., Romer D., Weil D.N. (1992), "A contribution to the empirics of economic growth", *Quarterly Journal of Economics*, 107(2), 407-437
- Murphy K.M., Shleifer A., Vishny R. (1989), "Income Distribution, Market Size and Industrialization", *Quarterly Journal of Economics*, 104, 537-564.
- Pananek G., Kyn O. (1986), "The effect on income distribution of development, the growth rate and economic strategy", *Journal of Development Economics*, 23, 55-65.
- Perotti R. (1993), "Political equilibrium, income distribution and growth", *Review of Economics Studies*, 60, 755-776.
- Perotti R. (1994), "Income Distribution and Investment", *European Economic Review*, 38, 827-835.
- Perotti R. (1996), "Growth, Income Distribution and Democracy: What the Data Say", *Journal of Economic Growth*, 1(1), 149-187.
- Persson T., Tabellini G. (1994), "Is Inequality Harmful for Growth?", *The American Economic Review*, 84(3), 600-620.
- Piketty T. (1997), "The Dynamics of Wealth Distribution and the Interest Rate with Credit Rationing", *Review of Economics Studies*, 64, 173-189.
- Popper K. R. (1959), *La logique de la découverte scientifique*, Payot, Paris, trad. fr. 1973
- Rossi N., G. Toniolo, G. Vecchi (1999), "Is the Kuznets Curve Still Alive ? Evidence from Italy's Household Budgets, 1881-1961", *CEPR Discussion Paper N°2140*, 57 pp.
- Saint-Paul G., Verdier T. (1993), "Education, democracy and growth", *Journal of Development Economics*, 42, 399-407.
- Stiglitz J. (1969), *Distribution of Income and Wealth Among Individuals*, *Econometrica*, 37(3), 382-397.
- Schumpeter J. A. (1954), *Histoire de l'analyse économique*, tome III, Gallimard, Paris, trad. fr. 1983.
- Volle M. (1981), *L'analyse des données*, Economica, Paris.
- White H. (1982), *Instrumental Variables Regression with Independent Observations*, *Econometrica*, 50, 483-499.
- World Bank (1998), *World Development Indicators*, CD ROM.

ANNEXES

Annexe 1 : Dictionnaire des variables

Variable	Définition	Source
GDP(-1)	PIB par hab. en \$ internationaux en début de période (ex. : pour 1980-84, valeur de 1980)	CEPII (base CHELEM 1998)
GDPSQ(-1)	GDP(-1) au carré	idem
LGDP(-1)	Logarithme de GDP(-1)	idem
GROWTH	Taux de croissance de GDP sur la période =LGDP-LGDP(-1)	idem
GINI	Indice de Gini	Deininger et Squire (1996)
[L]SECR(-2)	[Log. du] Taux brut de scolarisation secondaire de début de période, retardé une fois (ex.: pour 1980-84, valeur de 1975)	World Development Indicators (1998)
[L]PRIMR	[Log. du] Taux brut de scolarisation primaire	idem
[L]FERT	[Log. du] Taux de fécondité	idem
[L]GPOPC	[Log. du] Taux de croissance de la population +0.05	idem
[L]INV	[Log. du] Taux d'investissement	idem
[L]RLP	[Log. du] Rapport de productivité du travail entre secteur agricole et autres secteurs	

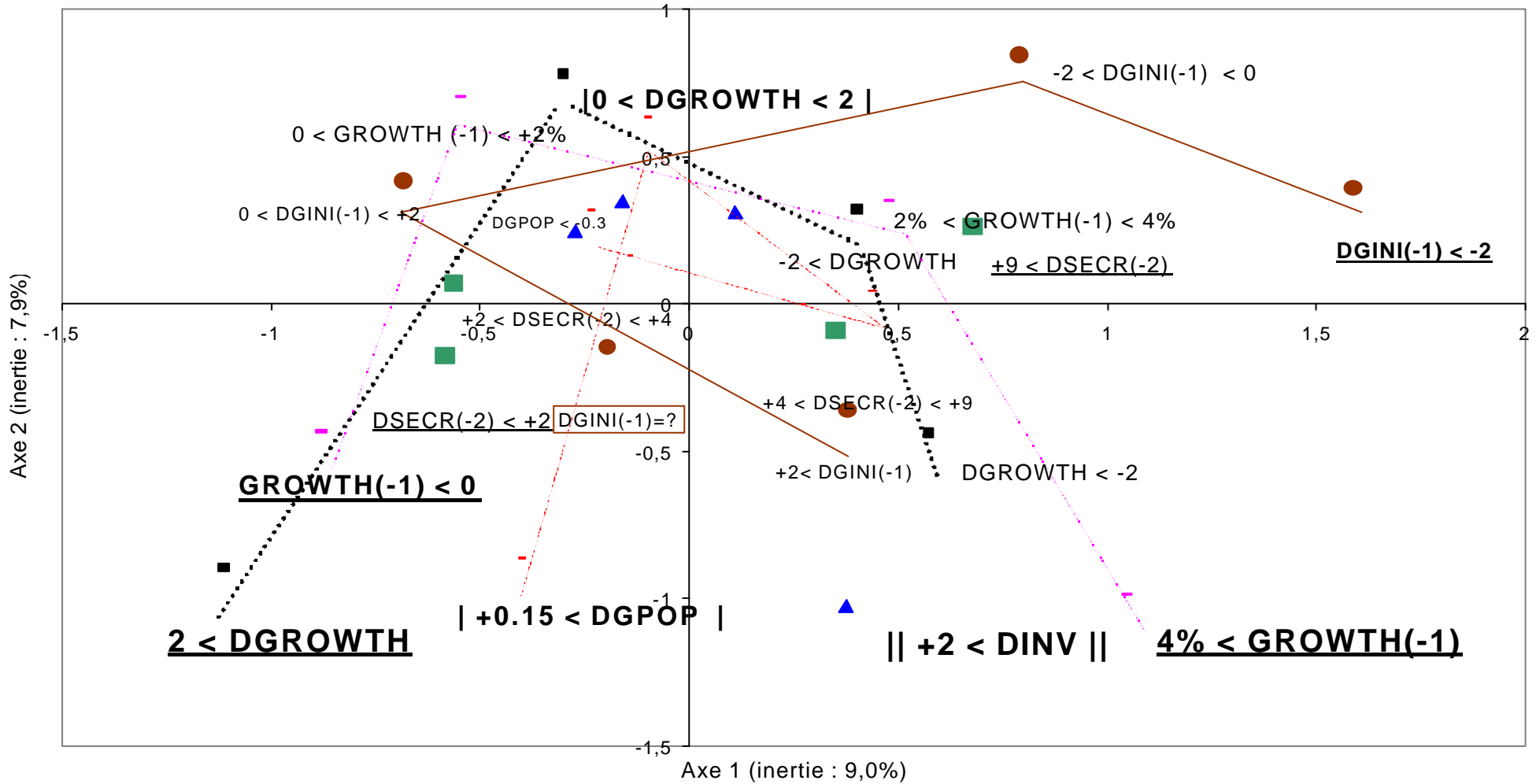
Annexe 2 : Statistiques descriptives

	Echantillon n°1	Echantillon n°2
A. Nombre de périodes		
1970-74	77	9
1975-79	81	10
1980-84	82	20
1985-89	78	27
1990-94	69	26
Total	387	92
B. Moyennes		
PIB par hab. en dollars internationaux 1985	5185.6	9020.2
Taux de croissance du PIB par hab. (%)	1.4	2.2
Indice de Gini	0.415	0.377
Valeurs manquantes de l'indice de Gini (%)	45.2	0.0
Taux d'investissement (%)	17.2	22.1
Taux brut de scolarisation secondaire (%)	45.0	67.4
Taux brut de scolarisation primaire (%)	89.5	100.7
Taux de croissance de la population (%)	1.9	1.2
Taux de fécondité	4.3	2.6
C. Nombre de pays par région		
Afrique sub-saharienne	29	1
Amérique Latine et Caraïbes	18	8
Asie du Sud et de l'Est	12	10
Pays de l'OCDE	19	13
Autres	12	5
Total	90	37

Annexe 3 : Analyse des correspondances multiples

**Analyse des correspondances multiples
Echantillon n°1**

387 pays-périodes, différences premières des variables (D)
contributions à l'axe 1 | contributions à l'axe 2 |



Annexe 3 (suite et fin)
Analyse des Correspondances Multiples
Echantillon n°2
 92 points pays-périodes, différences premières des variables
~~contributions à l'axe 1~~, | contributions à l'axe 2 |

