# Manuel sur la **Mesure des Inégalités pour les Études Pays**

*Muna Shifa et Vimal Ranchhod*

AFD
AGENCE FRANÇAISE DE DÉVELOPPEMENT

ACEIR
AFRICAN CENTRE OF EXCELLENCE FOR INEQUALITY RESEARCH

# Manuel sur la **mesure des inégalités pour les études pays**

*Muna Shifa[1] et Vimal Ranchhod[2]*

## 1. INTRODUCTION

Rédiger une étude sur les inégalités au niveau national peut être une tâche intimidante pour les chercheurs. Comment aborder la conceptualisation d'une telle étude ? Quels éléments inclure ? Quels types d'analyses utiliser, et comment faire pour générer les résultats requis ? Ce Manuel est avant tout un document de référence destiné à servir de guide aux chercheurs qui s'apprêtent à rédiger un rapport de recherche sur les inégalités dans un contexte national spécifique.[3]

Nous avons rédigé ce Manuel en partant du principe que les chercheurs qui seraient amenés à l'utiliser n'auraient qu'une expérience limitée de ce type de tâche, mais auraient toutefois quelques notions en gestion des données et seraient capables de travailler avec le logiciel Stata. Dans la mesure où chaque contexte est unique, nous partons également du principe que l'équipe de recherche concernée possède des connaissances approfondies du contexte local et de ses institutions. Les chercheurs doivent notamment bénéficier de connaissances dans deux domaines en particulier.

Premièrement, ils doivent connaître l'environnement socio-économique global dans lequel se déroule leur étude. C'est une condition indispensable pour garantir que les principaux facteurs d'inégalité soient bien inclus dans l'étude, et que les facteurs plus négligeables en soient exclus. Par exemple, il est possible que dans certains contextes, les écarts de rémunération entre populations des zones rurales et urbaines expliquent une grande partie des inégalités au niveau national, là où dans un autre pays, ces écarts ne sont que marginalement responsables des inégalités globales. Il en va de même pour les inégalités découlant du rendement des actifs financiers. Les chercheurs sont donc invités à s'appuyer sur leur expertise locale pour déterminer quel sous-ensemble de facteurs d'inégalité est le plus déterminant dans le contexte qui les intéresse.

Deuxièmement, les chercheurs doivent connaître les différentes enquêtes et sources de données alternatives susceptibles d'être utilisées dans le cadre de leur étude. Les différents types de données requises seront abordés plus en détail dans la section Données, mais de manière générale, deux situations potentielles nécessitent de bien connaître les données

1 Chargé de recherche principal au SALDRU, à l'Université du Cap.

2 Professeur d'économie au SALDRU, à l'Université du Cap.

3 Bien que nous ayons rédigé ce document dans le but spécifique de servir de guide intermédiaire aux membres du projet ACEIR, il peut également être utile aux chercheurs qui envisagent une étude de ce type dans différents contextes. Il peut notamment s'agir d'une étude à portée plus restreinte ou plus large, comme par exemple une analyse régionale ou transnationale.





disponibles. Premièrement, il arrive que plusieurs enquêtes puissent être utilisées pour analyser les inégalités. Dans ce cas, les chercheurs vont devoir choisir quelles sources de données inclure dans leur étude en se basant sur leur connaissance de l'ensemble des jeux de données disponibles. Deuxièmement, il arrive que les différents éléments d'information pertinents figurent dans des ensembles de données distincts.[4] Par exemple, la meilleure source de données sur l'éducation peut être le ministère de l'éducation, alors que les données sur la fécondité sont collectées par le ministère de la santé. Pour réaliser une étude sur les liens entre réussite scolaire et fécondité, la meilleure solution serait alors de fusionner ces deux bases de données institutionnelles. Pour prendre en toute connaissance de cause la décision d'inclure ou non ce type d'analyse dans une étude, il faut savoir que les informations dont on a besoin figurent dans ces deux ensembles de données, que ceux-ci sont tous deux accessibles aux chercheurs et qu'il est possible de combiner ces informations de manière utile.

En plus de donner les moyens aux chercheurs de réaliser une étude sur les inégalités dans un pays donné, ce Manuel a également été rédigé en tenant compte d'un méta-objectif, à savoir faciliter la comparabilité des résultats et des conclusions entre pays. Cette externalité représente l'une des principales motivations d'une collaboration multi-pays comme l'African Center of Excellence for Inequality Research (ACEIR). Afin de maximiser la comparabilité des résultats, il est utile de coordonner délibérément nos approches en amont. En respectant les méthodes et les principes d'interprétation décrits dans ce Manuel, les chercheurs s'assureront que les données sont préparées et analysées de façon homogène, et donc que les résultats obtenus dans les différents pays peuvent être comparés.

Il est évident que deux pays ne sont jamais totalement comparables. Comme nous l'avons vu plus haut, l'histoire, le contexte et les institutions ont tous une importance particulière, de sorte que chaque pays demeure toujours unique dès lors que nos analyses intègrent des niveaux de détail de plus en plus précis. Il est donc probable qu'il y ait un arbitrage à faire entre la comparabilité des études entre pays, d'une part, et la spécificité et l'exhaustivité de l'étude dans un pays donné, d'autre part. Pour concilier ces objectifs quelque peu contradictoires, nous avons adopté ce que nous considérons comme une approche pragmatique. Tout aspect des inégalités qu'un chercheur juge important pour parvenir à une compréhension exhaustive des inégalités dans un pays donné doit être inclus dans l'étude sur ce pays. Par ailleurs, la section 5.2 fournit une liste des résultats que toute étude nationale devrait idéalement inclure. Cet ensemble minimal de résultats requis, sous réserve qu'il soit possible de les mettre en œuvre dans chaque pays, constituera la base d'une méta-étude portant sur la comparaison des inégalités entre les différents pays.

Le reste de ce Manuel est structuré comme suit : dans la section 2, nous abordons quelques paramètres que toute étude empirique sur les inégalités doit définir en amont. La section 3 traite des exigences en matière de données et des solutions possibles aux problèmes les plus courants. Dans la section 4, nous expliquons comment mettre en œuvre les différents estimateurs dans le contexte plus spécifique d'une étude sur les inégalités de revenus. La section 5 présente la structure de base des rapports pays tels que nous les envisageons. Enfin, la section 6 conclut par un bref résumé de ce Manuel.

4 On notera que ces deux possibilités ne sont pas mutuellement exclusives. En effet, il est probable que ces deux considérations relatives aux données s'appliquent simultanément.

# 2. PARAMÈTRES DES ÉTUDES SUR LES INÉGALITÉS

Le corpus des recherches que l'on peut considérer comme ayant trait aux inégalités est très vaste. Tout sujet qui touche à la structure de la société est susceptible d'inclure des notions d'inégalité. Ainsi, les disciplines qui ont contribué à la compréhension holistique des inégalités sont notamment la sociologie, l'histoire, la politique, l'économie, la santé, la littérature, les statistiques, la géographie, la philosophie morale et la psychologie ; et il est tout à fait possible que cette liste pourtant exceptionnellement longue ne soit pas complète. Au sein de chaque discipline, il faut ensuite déterminer quelles méthodes utiliser pour étudier le sujet. Ces méthodes peuvent aller de la théorie purement abstraite à des analyses quantitatives à grande échelle, en passant par des études qualitatives sur de petits échantillons. Une fois le champ de l'étude suffisamment circonscrit, il faut encore déterminer quelles sources d'information vont être mobilisées. Par exemple, les données peuvent provenir d'enquêtes, de bases de données administratives, de registres d'entreprises, d'archives historiques, de cartes, de systèmes juridiques de droits et de registres de propriété, ou même indirectement de données sur les prix et de systèmes de comptabilité. Tout chercheur travaillant sur les inégalités doit donc définir les paramètres de son étude de façon à être en mesure de répondre à la question de recherche.

Au sein de la vaste catégorie des études empiriques sur les inégalités en économie, il nous faut encore répondre au moins à trois questions pour définir la portée de notre étude : quel type d'inégalités, au sein de quelle population, et sur quelle période ?

Le contexte de l'étude pays suppose, presque par définition, que le groupe sur lequel se concentre notre analyse est composé des individus qui résident à l'intérieur des frontières du pays qui nous intéresse. On notera que cette définition inclut les immigrants, quel que soit leur statut juridique, et exclut les individus ayant émigré, qu'il s'agisse d'une situation temporaire ou permanente. Bien qu'un tel choix puisse sembler anodin, il peut avoir un impact significatif sur la mesure des inégalités. Sans entrer dans le détail des avantages ou des inconvénients de ce choix, notons simplement qu'il reflète la manière dont les recensements sont généralement menés. De ce fait, il est probable que la disponibilité des données rende cette décision inutile en pratique.

En ce qui concerne l'horizon temporel, on pourrait très bien envisager de mesurer les inégalités sur une longue période, ou à une autre époque politique ou historique. Dans ce Manuel, cependant, nous nous intéresserons plutôt aux inégalités contemporaines et aux tendances relativement récentes dans ce domaine. De façon plus précise, nous entendons par là la période correspondant aux données existantes les plus récentes qui conviennent à l'analyse que l'on souhaite réaliser.

Enfin, une autre décision peut avoir une influence considérable sur la mise en œuvre, et donc les résultats, d'une étude sur les inégalités : le choix des aspects spécifiques des inégalités que l'on souhaite étudier. Le présent projet s'intéresse plus précisément aux inégalités économiques, bien que cette catégorie d'études reste elle-même assez vaste. On peut s'interroger sur les inégalités en matière de pouvoir de marché, d'accès au crédit ou de revenus locatifs. On peut mesurer la discrimination sur le marché du travail en fonction de l'origine ethnique ou du sexe, ou encore étudier la répartition des professions et la stratification professionnelle. Une décision s'impose et, pour cette série d'articles, nous nous concentrons principalement sur les inégalités contemporaines en matière de revenu global ou de consommation au sein de la population dans son ensemble. Néanmoins, comme nous l'avons indiqué en introduction, les chercheurs sont invités à approfondir cette question en procédant à une décomposition des inégalités en sous-groupes définis sur la base de leurs connaissances préalables et de leur expertise locale.

La section suivante présente les données nécessaires à la mise en œuvre d'une étude pays, ainsi que quelques problèmes courants et leurs solutions possibles.

# 3. DONNÉES ET MESURE DU BIEN-ÊTRE

Dans la section précédente, nous avons évoqué les différents paramètres à définir pour étudier les inégalités. Dans cette section, nous allons nous concentrer sur les problèmes de données et de mesure liés à l'analyse des inégalités économiques. Bien que le nombre et la fréquence des enquêtes ménages collectant des informations sur les revenus et la consommation aient augmenté récemment, la qualité des données reste un problème majeur dans la plupart des pays en développement. Examinons quelques problèmes dont il faut tenir compte lorsqu'on cherche à mesurer les inégalités.

## *Les principaux indicateurs de bien-être*

Les indicateurs classiques utilisés dans la littérature pour mesurer le bien-être individuel et analyser les inégalités économiques sont les revenus ou la consommation. Dans le contexte des pays en développement, cependant, on utilise plus souvent les données de consommation pour estimer aussi bien la pauvreté que les inégalités. C'est notamment lié au fait que les données de revenu n'y sont pas facilement accessibles. La plupart des pays en développement et des pays émergents possèdent un secteur informel important, et il est difficile de collecter des informations sur les revenus du travail indépendant et de l'agriculture de subsistance. En outre, dans la mesure où les ménages lissent leur consommation (via l'épargne et l'emprunt), on a tendance à privilégier la consommation comme indicateur du bien-être actuel. Ainsi, si le revenu peut être envisagé comme un moyen de parvenir au bien-être, la consommation est un indicateur plus direct du bien-être individuel.

Néanmoins, mesurer la consommation en s'appuyant sur des enquêtes ménages pose problème à plusieurs titres. Par exemple, il est souvent difficile d'attribuer une valeur monétaire aux biens et services consommés dans le cadre d'une production personnelle (par exemple, l'agriculture de subsistance) ou fournis par le secteur public (par exemple, l'accès à l'éducation et aux services de santé gratuits). Ces problèmes de mesure peuvent fausser nos estimations des inégalités, et ils rendent également difficile la comparaison des inégalités entre différents pays.

## *Absence de réponse et sous-déclaration*

Dans la plupart des enquêtes sur les ménages, les foyers situés dans la tranche supérieure de la distribution des revenus sont sous-représentés en raison du taux élevé de non-réponse qui les caractérise. En outre, ces ménages ont aussi tendance à sous-déclarer leurs revenus. Ces problèmes peuvent donner lieu à une sous-estimation du niveau d'inégalité. On utilise alors des pondérations (stratification a posteriori) pour corriger les problèmes liés à l'absence de réponse chez les plus riches. Même si on utilise des données de consommation pour mesurer les inégalités, on risque tout de même de sous-estimer les inégalités économiques dans la mesure où les plus aisés ont tendance à épargner davantage que les plus modestes. Il arrive aussi qu'on utilise les données fiscales pour estimer les inégalités. Toutefois, les données sur le revenu imposable ne sont généralement disponibles que pour les personnes dont le revenu dépasse un certain seuil (Wittenberg, 2017).

On peut également trouver dans les données des cas de non-réponse partielle et des revenus indiqués par tranches. À moins que certaines valeurs ne soient manquantes de manière totalement aléatoire (*missing completely at random*, ou MCAR), ignorer les valeurs manquantes peut biaiser l'estimation des inégalités. Si les données manquantes ne sont pas de type MCAR, on peut utiliser une méthode d'imputation pour attribuer les valeurs manquantes. Quant à la déclaration des revenus par tranches, elle est courante dans les enquêtes ménages, et certaines études n'ont pas d'autre choix que d'utiliser ce type de données. Par exemple, dans les recensements sud-africains, les valeurs de revenus ne sont indiquées que par tranches. Dans ce cas, la plupart des études utilisent des techniques d'imputation pour convertir les valeurs rapportées sous forme de tranches en estimations ponctuelles. Cependant, ces techniques présentent elles aussi le risque de sous-estimer les inégalités si chaque individu d'une tranche donnée se voit attribuer une seule et même valeur de revenu, comme c'est souvent le cas.

### *Comparabilité des enquêtes*

Un autre défi majeur auquel sont confrontés les chercheurs qui souhaitent mesurer les tendances en matière d'inégalités est la question de l'homogénéité dans le temps ou entre pays/régions. Les variations des indices d'inégalité peuvent être dues à un changement réel des niveaux de vie, à une modification de la méthode de collecte des données, ou encore à une combinaison de ces deux effets. Les changements de méthodologie incluent notamment la modification des modalités de collecte des données (conception de l'enquête et instruments utilisés) et de la mesure des variables, la fluctuation des prix et les ajustements liés à la saisonnalité. Par exemple, la façon dont on mesure les revenus ou la consommation doit être homogène d'une année à l'autre. Les changements apportés aux catégories de revenus ou de consommation (par exemple en raison d'une mise à jour de la liste des articles de consommation, ou du revenu net par rapport au revenu brut), les modifications de la période couverte[5] et le caractère saisonnier des activités économiques sont susceptibles de générer des incohérences dans la mesure des données de revenu ou de consommation au fil du temps. Il faut donc être attentif à ces problèmes quand on utilise des données d'enquête pour comparer les inégalités *dans le temps, entre régions ou entre pays*.

### *Échelles d'équivalence*

Les données de revenu ou de consommation sont souvent collectées au niveau des ménages. La National Income Dynamics Study (NIDS), une enquête par panel représentative au niveau national menée en Afrique du Sud, est l'une des rares à collecter des données de revenu à la fois au niveau individuel et au niveau des ménages. Pour étudier la pauvreté ou les inégalités, il faut disposer d'informations sur le bien-être au niveau individuel. Il peut donc être nécessaire de formuler des hypothèses concernant la répartition des revenus ou de la consommation au sein des ménages afin de convertir les données au niveau des ménages en données au niveau individuel.

5 Par exemple, la période de référence pour la collecte des données de consommation ou de revenu peut être annuelle ou mensuelle. Cela pose un problème de comparabilité dans la mesure où les données mensuelles peuvent inclure des fluctuations transitoires, ce qui ne serait peut-être pas le cas en se basant sur une période de référence annuelle. On s'attend donc à ce que la mesure des inégalités soit plus élevée si on utilise une période de référence mensuelle plutôt qu'annuelle.



Même dans les cas où l'on dispose d'informations individuelles sur les revenus, il faut d'abord les agréger sous forme de données au niveau des ménages, car les familles mettent en commun leurs revenus et autres ressources. Pour ce faire, l'une des approches consiste à utiliser une échelle par habitant, ce qui implique de diviser la consommation ou le revenu total du ménage par sa taille et d'attribuer cette valeur moyenne à tous les individus du ménage. Dans ce cas, on part du principe que les revenus ou la consommation du ménage sont répartis de manière égale entre tous ses membres, et on ne tient pas compte des économies d'échelle.[6] Une autre solution consiste à utiliser une échelle d'équivalence adulte. L'approche dite de l'équivalent-adulte tient compte à la fois des économies d'échelle et du coût des enfants (en supposant que les enfants consomment moins que les adultes).[7] Notons que dans les deux cas, l'allocation des ressources au sein du foyer n'est pas prise en compte, car elle nécessite des informations détaillées sur la consommation de chaque membre du ménage. Il est donc impossible de ventiler les inégalités par groupes comme le sexe si les membres de ces groupes sont généralement co-résidents au sein des mêmes ménages.

# 4. MÉTHODES DE MESURE DES INÉGALITÉS ÉCONOMIQUES

## 4.1 Sélectionner un indicateur d'inégalité

Dans la littérature, différents outils (indices d'inégalité) sont utilisés pour mesurer les inégalités. Le choix de l'indice d'inégalité dépend en partie du type de question que l'on souhaite étudier. Imaginons par exemple que l'on cherche à déterminer dans quelle mesure les inégalités en Afrique du Sud sont dues à une répartition inégale des revenus au sein des groupes ethniques, plutôt qu'à une distribution inégale des revenus entre ces différents groupes ethniques. Répondre à ce type de questions requiert un indicateur d'inégalité permettant de décomposer l'inégalité globale en différents groupes.

Pour choisir parmi les différents indicateurs d'inégalité (indices), on peut notamment adopter l'approche axiomatique (Cowell, 1985). Cette approche consiste tout d'abord à spécifier un ensemble de propriétés minimales qu'un indicateur d'inégalité doit satisfaire. On utilise ensuite ces axiomes pour faire un choix parmi les différents indices possibles. Vous trouverez ci-dessous quelques-unes des principales propriétés (axiomes) qu'un indice d'inégalité doit satisfaire. Cette section est largement inspirée de l'ouvrage de Foster et al. (2013) intitulé *A unified approach to measuring poverty and inequality: Theory and practice*.

*Axiome 1 : Anonymat (symétrie) : cet axiome stipule que la mesure des inégalités doit rester inchangée en cas de permutation, c'est-à-dire que l'identité des individus ne doit pas être prise en compte dans l'analyse des inégalités.*

6 Par exemple, le coût de la vie par habitant peut être plus élevé pour un ménage composé d'une seule famille que pour un ménage composé de deux familles, car les membres du ménage bifamilial peuvent partager le coût du loyer et autres coûts communs du ménage. Si l'on ne tient pas compte de ce type d'économies d'échelle, on aura tendance à sous-estimer le bien-être des ménages les plus nombreux.

7 Voici la formule souvent utilisée pour calculer un équivalent-adulte : Équivalents-adultes = $(adultes+a\times enfants)^q$ où a est le paramètre enfant, souvent compris entre 0,5 et 0,75, et q est le paramètre correspondant aux économies d'échelle. Par exemple, dans le cas de l'Afrique du Sud, q= 0,9.

Prenons l'exemple d'un groupe de quatre individus nommés A, B, C et D, dont les revenus sont respectivement de 10, 20, 30 et 40.

$$Y_{1} = (10, 20, 30, 40)$$

A, B, C, D

Considérons maintenant un second groupe avec le même ensemble de revenus, mais dont les bénéficiaires sont différents.

$$Y_{2} = (10, 20, 30, 40)$$

D, C, B, A

En vertu de l'axiome d'anonymat/symétrie, les distributions Y1 et Y2 sont tout aussi inégales l'une que l'autre.

*Axiome 2 : Principe de population :* cette propriété implique que le niveau d'inégalité au sein d'un groupe ne varie pas en fonction de la taille de sa population. Par exemple, si on obtient $Y_3$= (10, 10, 20, 20, 30, 30, 40, 40) à partir de $Y_1$ (en doublant le nombre d'individus sans changer la distribution des revenus), alors selon le principe de population, ces deux distributions sont aussi inégales l'une que l'autre.

*Axiome 3 : Normalisation :* cette propriété stipule que l'indice d'inégalité doit être égal à zéro lorsque tous les revenus sont distribués de manière égale.

*Axiome 4 : Invariance à l'échelle :* cet axiome suppose qu'augmenter les revenus de tous les individus dans la même proportion ne modifie pas les inégalités. Par exemple, en multipliant par deux le revenu de chaque membre de $Y_1$, on obtient $Y_4$ = (20, 40, 60, 80). On constate que le niveau de revenu de la personne la plus riche est 4 fois supérieur à celui de la personne la plus pauvre aussi bien dans le cas de $Y_1$ que de $Y_4$. L'invariance à l'échelle implique que le niveau d'inégalité est identique dans $Y_1$ et $Y_4$, et que la notion d'inégalité est purement relative. Ainsi, le niveau de revenu n'a pas d'importance. Il convient de noter que la pertinence de cette propriété varie selon que l'on s'intéresse à des mesures d'inégalités absolues ou relatives. Si l'on décide de mesurer les inégalités absolues, l'écart absolu entre les individus les plus riches et les plus pauvres est de 30 dans $Y_1$ contre 60 dans $Y_4$. Dans ce cas, on peut donc dire que le niveau d'inégalité est plus élevé dans $Y_4$ que dans $Y_1$.

En fonction de leur propriété d'invariance, on peut donc classer les mesures d'inégalités en deux catégories : les mesures d'inégalités absolues et les mesures d'inégalités relatives. Les mesures absolues sont invariantes par translation : la mesure des inégalités de revenus n'est pas modifiée si l'on ajoute ou si l'on retranche un montant absolu au revenu de tous les individus. Les mesures d'inégalités relatives, quant à elles, sont invariantes à l'échelle : elles ne sont pas modifiées si l'on multiplie tous les revenus par une valeur scalaire positive. Les mesures d'inégalités relatives ne sont pas invariantes par translation, tandis que les mesures d'inégalités absolues ne sont pas invariantes à l'échelle.

D'un point de vue analytique, la propriété d'invariance à l'échelle est intéressante car elle garantit que la valeur d'une mesure d'inégalité ne varie pas selon l'unité utilisée pour mesurer les revenus, contrairement aux mesures invariantes par translation qui violent cette propriété.





Par exemple, la variance est l'une des mesures absolues les plus simples, mais sa valeur dépend de l'unité de mesure. De ce fait, les mesures d'inégalités relatives sont préférables aux mesures absolues, et c'est donc sur elles que se concentre le présent document.

*Axiome 5 : Principes de transfert :* Que se passe-t-il lorsqu'on transfère le revenu d'une personne relativement aisée à une personne relativement pauvre, tout en maintenant constant leur rang respectif dans la distribution des revenus ? Par exemple, supposons que la personne la plus riche de $Y_4$ transfère dix rands de son revenu à la personne la plus pauvre, générant ainsi une nouvelle distribution des revenus $Y_5$ = (30, 40, 60, 70). Le raisonnement selon lequel un tel transfert réduit les inégalités, le niveau d'inégalité de $Y_5$ étant donc inférieur à celui de $Y_4$, fait l'objet d'un large consensus. Il existe deux versions du principe de transfert : (i) le principe de transfert faible, selon lequel le niveau d'inégalités est censé diminuer ou rester stable après un transfert de revenus d'un individu relativement riche vers un individu relativement pauvre, et (ii) le principe de transfert fort, selon lequel le niveau d'inégalités est censé diminuer strictement après un transfert de revenus d'un individu relativement riche vers un individu relativement pauvre.

*Axiome 6 : Sensibilité au transfert :* Cette propriété stipule qu'un indice d'inégalité doit être plus sensible aux transferts réalisés dans la tranche la plus basse de la distribution (c'est-à-dire entre deux individus pauvres plutôt qu'entre deux individus aisés). Prenons l'exemple d'une distribution des revenus $Y_6$ = (30, 30, 60, 80), obtenue en transférant 10 rands du deuxième individu le plus pauvre de $Y_4$ = (20, 40, 60, 80) à l'individu le plus pauvre de cette distribution. Comparons cela avec la distribution $Y_7$ = (20, 40, 70, 70), obtenue en transférant le même montant de l'individu le plus riche de $Y_4$ = (20, 40, 60, 80) au deuxième individu le plus riche. Selon le principe de sensibilité au transfert, notre indicateur d'inégalité doit être plus sensible aux transferts qui ont généré $Y_6$ qu'à ceux qui ont généré $Y_7$.

*Axiome 7 : Décomposabilité :* Si l'on veut pouvoir décomposer notre mesure des inégalités de façon à rendre compte de la contribution de différents sous-groupes, notre indicateur doit être décomposable. Les sous-groupes en question peuvent être des sources de revenus (revenus du travail vs revenus hors travail) ou relever d'autres dimensions telles que l'origine ethnique, le sexe ou le lieu de résidence. Une mesure des inégalités décomposable doit satisfaire deux propriétés souhaitables :

i. <u>Décomposabilité additive :</u> L'inégalité totale est la somme de toutes les inégalités intra-groupes et inter-groupes. L'inégalité intra-groupe est la somme pondérée des inégalités au sein des sous-groupes (les poids peuvent refléter les parts de population concernées ou les revenus relatifs), tandis que l'inégalité inter-groupe correspond aux inégalités entre les différents sous-groupes (le revenu moyen du groupe est attribué à chaque individu de chaque groupe).

ii. <u>Décomposition cohérente en sous-groupes :</u> Ce concept fait référence à la réactivité de la mesure de l'inégalité totale aux variations du niveau d'inégalité au sein des différents sous-groupes. Si les inégalités augmentent au sein d'un sous-groupe donné de la population et qu'elles ne diminuent pas dans le reste des sous-groupes, notre mesure de l'inégalité totale est censée augmenter.

## 4.2 Mesures d’inégalité courantes

Dans cette section, nous allons passer en revue quelques-uns des indices d’inégalité les plus répandus. De manière générale, les mesures d’inégalité se répartissent en deux grandes catégories selon l’approche utilisée pour les calculer : les mesures descriptives et les mesures normatives (Sen, 1973). Les mesures d’inégalité descriptives sont généralement des formules mathématiques ou statistiques. Par conséquent, les caractéristiques de ces indices sont fonction de leurs propriétés mathématiques ou statistiques, respectivement. La plupart des indices d’inégalité sont de nature descriptive. Les indices d’inégalité normatifs sont dérivés d’une fonction de bien-être social fondée sur un jugement de valeur préalable concernant les effets des inégalités sur le bien-être social. Ces mesures associent l’indice d’inégalité à une évaluation sociale et précisent si les inégalités sont néfastes ou non, ainsi que le degré de bien-être qu’une société perd ou gagne du fait de ces inégalités. Les indices d’inégalité d’Atkinson comptent parmi les mesures normatives les plus souvent citées. Notons que les indices d’inégalité examinés ici ne vérifient pas nécessairement tous les axiomes abordés dans la sous-section précédente. Par exemple, l’indice d’Atkinson satisfait presque tous les axiomes, mais il n’est pas additivement décomposable. Ainsi, si notre objectif est de décomposer les inégalités totales par sous-groupes de population, on utilisera plutôt les mesures d’inégalité issues de l'entropie.

Dans le cadre de notre analyse des différentes mesures d’inégalité, nous utilisons des données provenant de l’enquête NIDS (vagues 1 et 4), des enquêtes sur la démographie et la santé (DHS) sud-africaines de 1998, qui sont des enquêtes ménages représentatives au niveau national, et du recensement sud-africain de 2011. Nous utilisons le revenu par habitant comme mesure du bien-être individuel (c’est-à-dire le revenu total du ménage divisé par la taille du ménage). Nos variables de revenu prennent en compte les revenus issus de toutes les sources (c’est-à-dire aussi bien les revenus du travail que les revenus autres que ceux du travail, comme les aides sociales). Nous utilisons des poids d’échantillonnage pour toutes nos estimations des inégalités de revenus.

Nous utilisons les données des enquêtes DHS pour analyser les approches visant à mesurer les aspects des inégalités qui ne sont pas liés au revenu, à savoir les inégalités d’équipement des ménages. Nous n’utilisons les variables de nos ensembles de données que pour illustrer par des exemples pratiques comment estimer et interpréter les différentes mesures d’inégalité. Les résultats de cet exercice ne doivent donc pas être utilisés à d'autres fins.

Nous utilisons le logiciel Stata et le package DASP pour estimer la plupart des indices d’inégalité. Les instructions relatives à l’installation du package DASP figurent en annexe (voir Araar & Duclos, 2013 pour plus de détails).

### 4.2.1 Rapports interquantiles / interdéciles / intercentiles

La façon la plus simple d’étudier les inégalités de revenus consiste à diviser la population en quantiles ou en déciles après en avoir classé les membres du plus pauvre au plus riche. On peut ainsi calculer le niveau ou la proportion de revenus qui échoit à chaque quantile ou décile. Le Tableau 1 ci-dessous représente la part de chaque centile dans le revenu total par décile en Afrique du Sud en 2008. Pour estimer la part de chaque centile, on utilise la commande Stata suivante :[8]

*pshare estimate pcminc [w=wgt],nquantiles(10) percent*

Tableau 1 : part des différents centiles, 2008

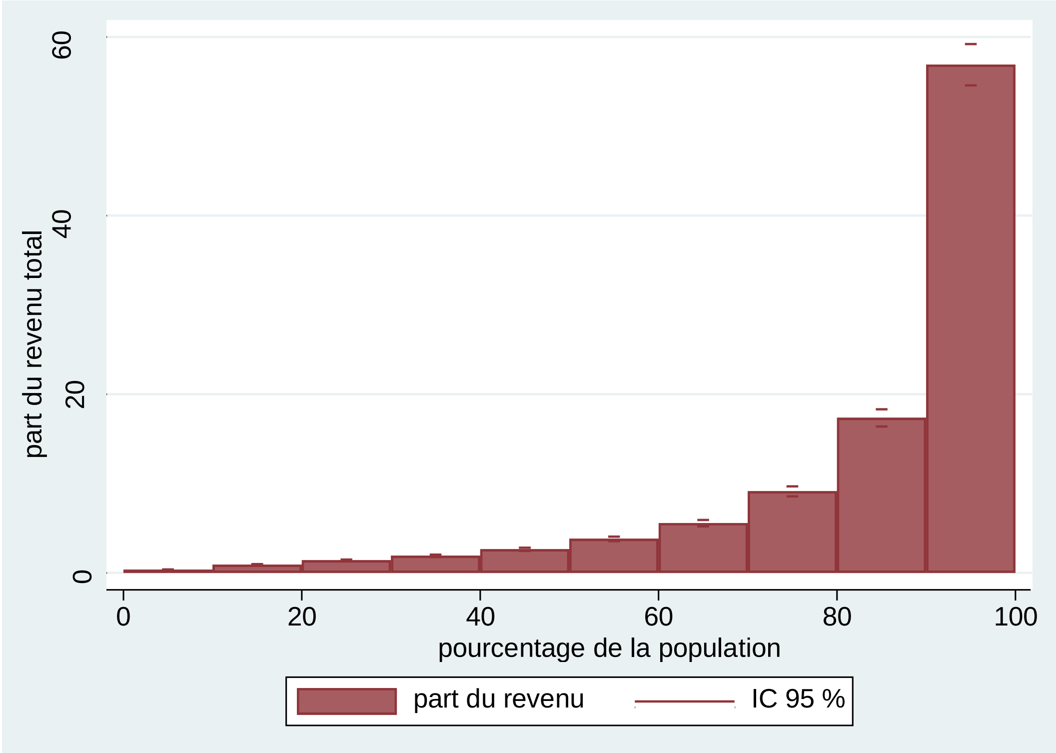
Percentile shares (percent)   Nombre d’obs =  17 710

| pcminc | Coef. | Erreur-type | [Interv. de | conf. à 95 %] |
|---|---|---|---|---|
| 0-10 | 0,3605923 | 0,016485 | 0,3282802 | 0,3929044 |
| 10-20 | 0,9175996 | 0,0355134 | 0,84799 | 0,9872093 |
| 20-30 | 1,398974 | 0,0508721 | 1,299259 | 1,498688 |
| 30-40 | 1,916194 | 0,0681962 | 1,782523 | 2,049866 |
| 40-50 | 2,643722 | 0,0921937 | 2,463013 | 2,824431 |
| 50-60 | 3,807283 | 0,1315998 | 3,549335 | 4,065232 |
| 60-70 | 5,567746 | 0,1830186 | 5,209011 | 5,92648 |
| 70-80 | 9,1327 | 0,2860643 | 8,571986 | 9,693415 |
| 80-90 | 17,35443 | 0,4894556 | 16,39505 | 18,31381 |
| 90-100 | 56,90076 | 1,182635 | 54,58268 | 59,21884 |

Les résultats montrent que le décile le plus riche a touché 57 % du revenu total en 2008, tandis que les 10 % les plus pauvres ont touché moins de 1 % du revenu total. On peut aussi présenter les estimations du tableau ci-dessus sous forme d’histogramme en utilisant la commande Stata suivante :

pshare histogram, name(p08, replace)

Figure 1




8 Pour plus de détails concernant l’ensemble des commandes Stata utilisées dans ce document, nous vous invitons à vous référer au menu d’aide du logiciel. Par exemple, en tapant « help pshare » dans la fenêtre de commande de Stata, on obtient des informations détaillées sur la commande « pshare ». La variable « pcminc » correspond au revenu par habitant, tandis que la variable « wgt » correspond aux poids (post-stratification) utilisés dans l’ensemble de données NIDS.

On peut également calculer un rapport interquantiles (ou interdéciles) pour comparer les revenus des différents quantiles. Par exemple, on peut comparer le revenu des 10 % les plus riches à celui des 10 % les plus pauvres en utilisant le ratio 90:10. D'après le tableau ci-dessus, ce ratio est d'environ 158. Cela signifie que les revenus perçus par les 10 % les plus riches sont 158 fois plus élevés que les revenus perçus par les 10 % les plus pauvres. La valeur d'un rapport interquantiles va de zéro à l'infini, et plus cette valeur est grande, plus les inégalités sont fortes. Il est cependant possible de normaliser le rapport interquantile de sorte qu'il soit compris entre zéro et un. Pour ce faire, on soustrait le revenu du quantile le plus pauvre de celui du quantile le plus riche, puis on divise le résultat ainsi obtenu par le revenu du quantile le plus pauvre. Par exemple, le rapport 90:10 normalisé correspond à la différence entre la valeur des revenus au 90e centile et la valeur des revenus au 10e centile, sous forme de proportion de la valeur des revenus au 10e centile. La valeur d'un rapport interquantile normalisé est de zéro lorsque les revenus des quantiles supérieur et inférieur sont égaux. La valeur d'un rapport interquantile normalisé atteint sa valeur maximale lorsque le revenu du quantile inférieur est égal à zéro. Cela signifie que personne dans le centile inférieur ne perçoit de revenu et que la valeur des revenus du quantile supérieur est positive. La valeur du rapport interquantile normalisé devient nulle si tous les membres de la population ont des revenus égaux. Cependant, un rapport interquantile nul ne signifie pas nécessairement que les revenus sont répartis de manière égale entre tous les individus de la population, car il peut toujours y avoir des variations de revenus à l'intérieur des quantiles.

Parmi les propriétés souhaitables évoquées ci-dessus, le rapport interquantile satisfait les propriétés d'anonymat, de normalisation et d'invariance à l'échelle, ainsi que le principe de population. En revanche, il ne respecte pas les principes de transfert (ni faible, ni fort). Les rapports interquantiles ne se décomposent pas : ils ne sont pas additivement décomposables et ne satisfont pas la propriété de décomposition cohérente en sous-groupes. Les rapports interquantiles présentent également une autre limite majeure : ils ne permettent de comparer que deux quantiles de revenus (uniquement les centiles sélectionnés), et ne rendent donc pas compte de l'ensemble de la distribution des revenus. On notera que différents types de rapports interquantiles sont utilisés dans la littérature pour mesurer les inégalités. Parmi les mesures les plus répandues, on trouve notamment la part du centile et du décile supérieurs dans le revenu total, ainsi que le ratio de Palma. Le ratio de Palma correspond à la part du décile supérieur dans les revenus divisée par celle des quatre déciles inférieurs. Le recours au ratio de Palma s'est développé ces dernières années (voir Doyle & Stiglitz, 2014). Son utilisation pour mesurer les inégalités repose sur l'observation empirique selon laquelle la part des revenus correspondant aux déciles « intermédiaires » (5 à 9) est relativement stable d'un pays à l'autre et dans le temps, et représente environ la moitié du revenu national brut. Ainsi, les variations des inégalités de revenus sont principalement dues à des changements au niveau des « extrémités » (Cobham, Schlögl & Sumner, 2016). Pour estimer le ratio de Palma, on peut utiliser la commande Stata suivante :

*pshare estimate pcminc [w=wgt], percentiles (40 90)*





```
. pshare estimate pcminc [w=wgt], percentiles(40 90)
(sampling weights assumed)

Percentile shares (proportion)   Nombre d'obs =  17 710
```

| pcminc | Coef. | Erreur-type | [Interv. de | conf. à 95 %] |
|---|---|---|---|---|
| 0-40 | 0,0459336 | 0,00166 | 0,0426798 | 0,0491874 |
| 40-90 | 0,3850588 | 0,0105345 | 0,3644101 | 0,4057076 |
| 90-100 | 0,5690076 | 0,0118264 | 0,5458268 | 0,5921884 |

*nlcom (Palma: _b[90-100] / _b[0-40])*

```
Palma: _b[90-100] / _b[0-40]
```

| pcminc | Coef. | Erreur-type | Z | P>\|z\| | [Interv. de | conf. à 95 %] |
|---|---|---|---|---|---|---|
| Palma | 12,38761 | 0,6726666 | 18,42 | 0,000 | 11,06921 | 13,70601 |

Selon l'estimation ci-dessus, la part du décile supérieur dans les revenus était 12,4 fois plus élevée que celle des quatre déciles inférieurs en Afrique du Sud en 2008.

### 4.2.2 Courbes de Lorenz

La courbe de Lorenz est une représentation graphique simple d'une distribution de revenus. C'est un graphique qui représente la part cumulée des revenus par rapport à la part cumulée des individus classés dans l'ordre. Pour obtenir une courbe de Lorenz, il faut d'abord classer la population du revenu le plus modeste au revenu le plus élevé. Ensuite, sur l'axe des ordonnées, on reporte la part cumulée des revenus perçus par chaque part cumulée de la population, l'axe des abscisses correspondant à la part cumulée de la population.

Pour ce faire, on utilise la commande Stata *clorenz* dans le package DASP. Il faut donc installer DASP. On utilise d'abord la commande *svyset* pour définir le plan d'échantillonnage. Dans les ensembles de données NIDS, les variables « psu » et « strat » correspondent respectivement aux unités primaires d'échantillonnage et aux strates.[9]

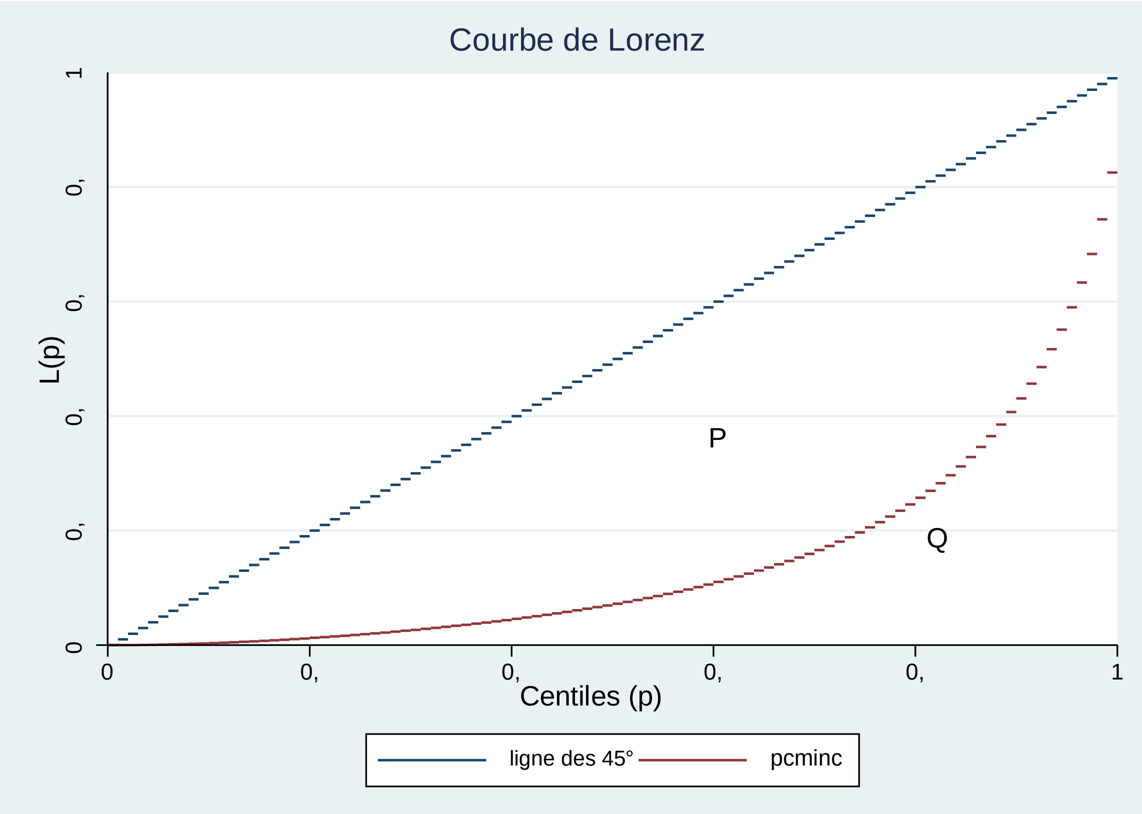
*svyset psu [pw=wgt], strata(strat)*

*Pour obtenir une courbe de Lorenz, on utilise la commande suivante :*

*clorenz pcminc, type(nor)*

9 L'enquête NIDS repose sur un échantillonnage en grappes stratifié à deux degrés (voir Leibbrandt, Woolard & de Villiers, 2009). À partir de l'échantillon principal de Stats SA pour 2003, composé de 3000 unités primaires d'échantillonnage (UPE), 400 UPE ont été sélectionnées lors de la première étape. Les UPE sont des zones géographiques bien définies qui comprennent un ou plusieurs districts du recensement de 2001 (Leibbrandt et al., 2009 : p.9). Les 53 conseils de district (CD) constituaient les strates explicites.

Figure 2a.



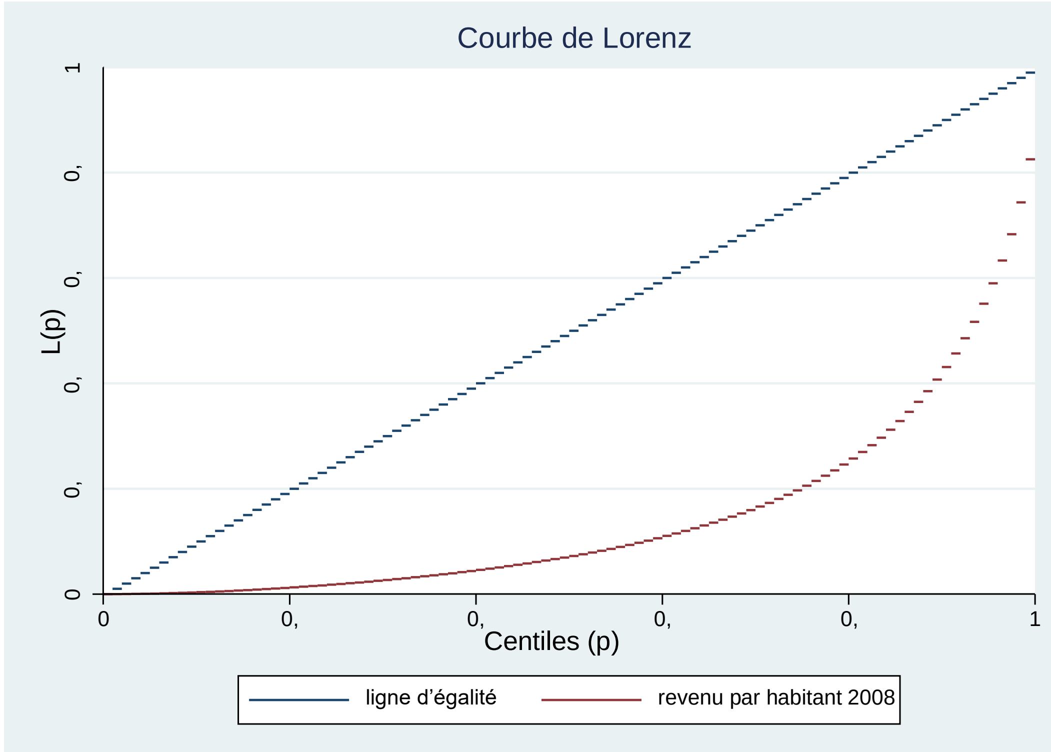

Après avoir généré le graphique en exécutant la commande *clorenz*, on peut cliquer sur l'onglet « start graph editor » et modifier les différentes parties du graphique selon nos besoins. Une autre possibilité consiste à utiliser des options supplémentaires dans votre commande. On peut par exemple utiliser l'option *legend* pour modifier le contenu des légendes du graphique ci-dessus :

clorenz pcminc, type(nor) legend( order(1 "line of equality" 2 "per capita income 2008"))





Figure 2b.



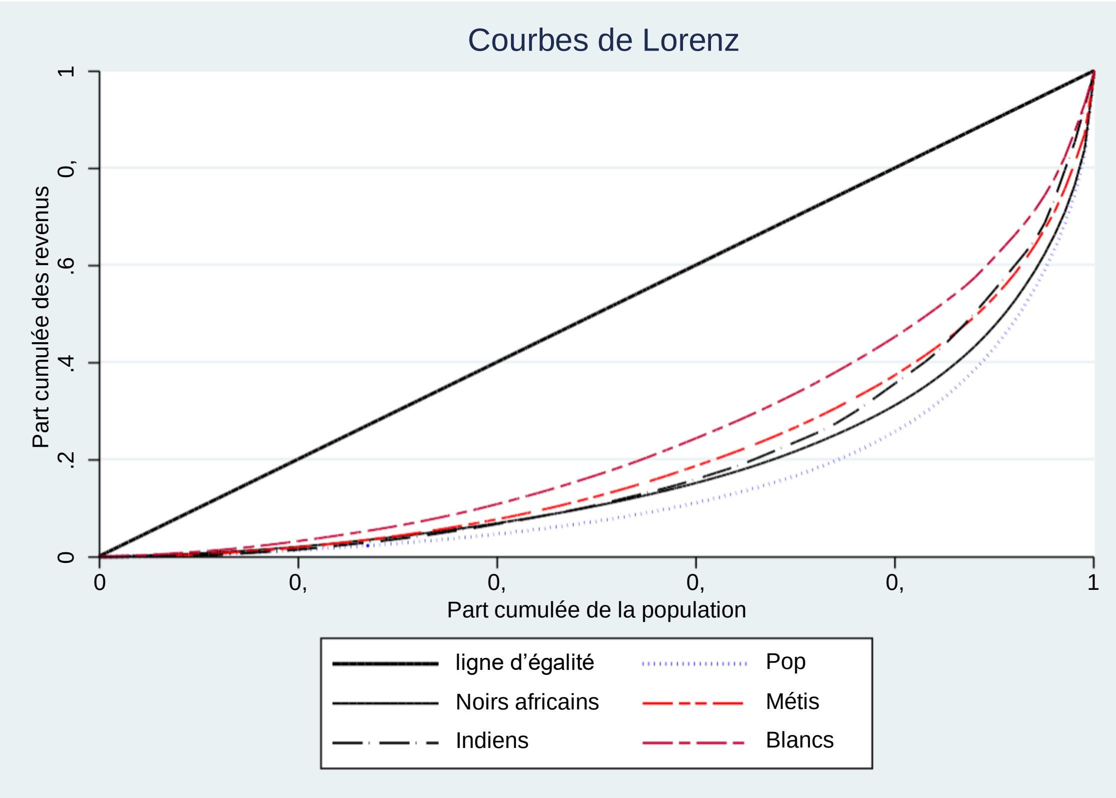

Si les revenus étaient distribués de manière égale au sein d'une population de taille n, alors chaque individu recevrait 1/n du revenu total. Dans ce cas, la courbe de Lorenz serait une droite à 45°. En réalité, les individus les plus modestes reçoivent moins de 1/n du revenu total et les individus aisés possèdent plus de 1/n du revenu total. Par conséquent, les courbes de Lorenz sont de forme convexe. Plus la courbe de Lorenz se rapproche de la ligne des 45°, plus le niveau d'inégalité est faible. On peut comparer différentes distributions de revenus en utilisant le concept de dominance de Lorenz. On dit qu'une distribution A domine une distribution B au sens de Lorenz si sa courbe de Lorenz se trouve en tout point au-dessus de la courbe de Lorenz de la distribution B (c'est-à-dire plus proche de la ligne des 45°). Dans ce cas, on peut affirmer de façon formelle que le niveau d'inégalité dans la population A est inférieur au niveau d'inégalité dans la population B. En revanche, si les courbes de Lorenz des deux distributions se croisent, il est impossible de comparer leur niveau d'inégalité en s'appuyant sur des courbes de Lorenz. On peut alors utiliser une courbe de Lorenz généralisée ou d'autres indices d'inégalité comme le coefficient de Gini pour comparer les niveaux d'inégalité des deux distributions.

On peut par exemple comparer le niveau des inégalités de revenus au sein des quatre groupes ethniques d'Afrique du Sud en utilisant la commande suivante :

*clorenz pcminc, type(nor) hgroup(race)*[10]

10 On peut modifier le graphique en utilisant l'éditeur de graphiques après avoir exécuté cette commande, ou en utilisant la commande suivante :
clorenz pcminc, hgroup(race) ///
lpattern("1" "." "1" "_--_#" "_.#" "_-" ) lc("black" "blue" "black" "red" "black") lwidth("medthick" "medthick") ///
xtitle("Cumulative population shares", size(small)) ytitle("Cumulative income shares", size(small)) ///
legend(order(1 "line of equality" 2 "pop" 3 "African" 4 "Coloured" 5 "Indian" 6 "White")) ///
saving(ineqrace.gph, replace) name(ineqrace, replace)

Figure 3.




La figure ci-dessus indique que parmi les groupes ethniques, c'est chez les Blancs que le niveau d'inégalité est le plus faible et chez les Noirs africains qu'il est le plus élevé. On peut dire que la distribution des revenus des Blancs domine celle du reste des groupes ethniques au sens de Lorenz. De même, la distribution des revenus des Métis domine celle des Noirs africains au sens de Lorenz. En revanche, on ne peut pas comparer le niveau d'inégalité des Indiens et des Métis puisque les deux courbes se croisent. Dans ce cas, on peut utiliser soit une courbe de Lorenz généralisée, soit d'autres indices d'inégalité tels que le coefficient de Gini pour comparer le degré d'inégalité entre groupes ethniques. Pour obtenir une courbe de Lorenz généralisée, on multiplie les coordonnées Y de la courbe de Lorenz par le revenu moyen de la population. En introduisant explicitement le revenu moyen, on compare les distributions sur la base d'un critère de bien-être. Le bien-être social est donc plus élevé dans le cas des distributions de revenus dont le revenu moyen est plus élevé, indépendamment du niveau d'inégalité. Pour obtenir des courbes de Lorenz généralisées par groupe ethnique, on peut utiliser la commande Stata suivante :

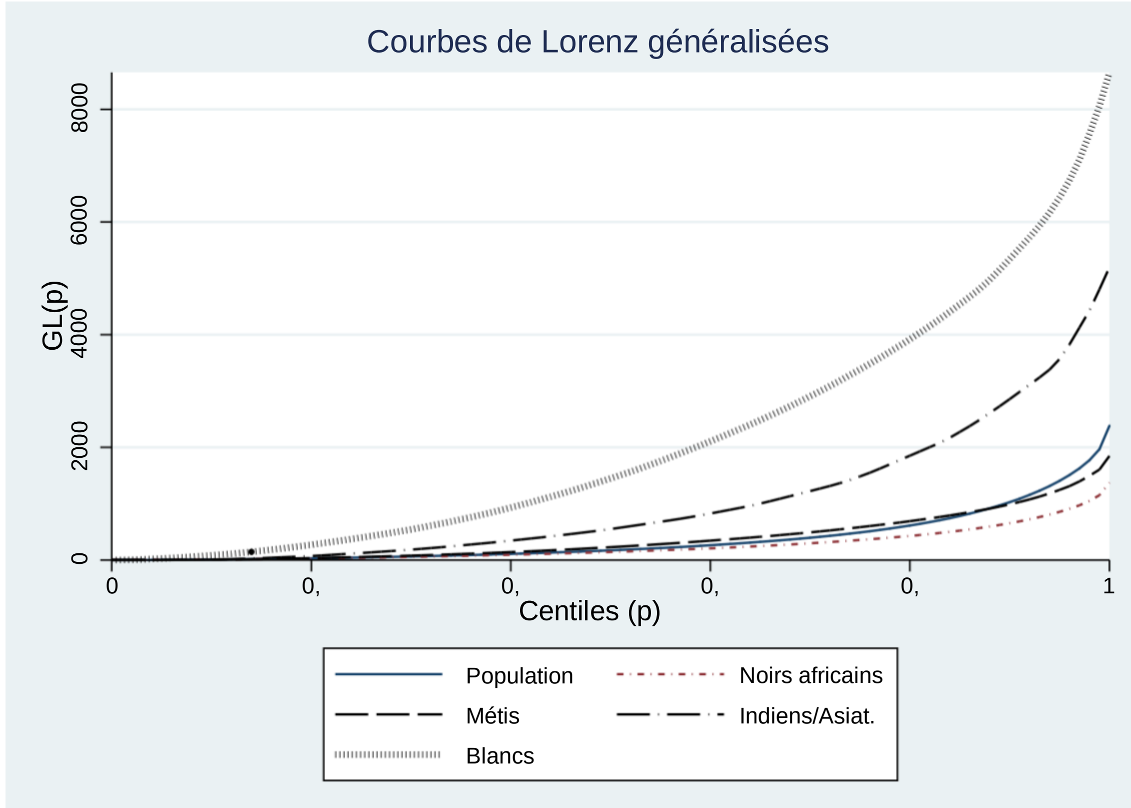
*clorenz pcminc, type(gen) hgroup(race)*

D'après les courbes de Lorenz généralisées (voir figure 4), le revenu moyen est le plus élevé chez les Blancs, suivi des Indiens, des Métis et des Noirs africains. On peut donc dire que c'est chez les Blancs que le bien-être est le plus élevé, puis chez les Indiens, les Métis, et qu'il est le plus faible chez les Noirs africains. Si les courbes de Lorenz généralisées se croisent, il est impossible de classer les distributions de revenus en se basant uniquement sur ces courbes.





Figure 4.




### 4.2.3 Le coefficient de Gini

Le coefficient de Gini est l'un des indicateurs d'inégalité les plus répandus et se calcule à partir d'une courbe de Lorenz. Il s'agit du rapport entre l'aire de la zone comprise entre la courbe de Lorenz et la ligne d'égalité, et l'aire de l'ensemble de la zone située sous la ligne d'égalité. Dans le cas de la Figure 2a., ce rapport se calculerait donc comme suit : Aire P/ (Aire P + Aire Q). La formule mathématique utilisée pour calculer le coefficient de Gini est la suivante :

$$G = \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{|y_i - y_j|}{2N^2\mu}$$

Où $y_i$ et $y_j$ indiquent le niveau de revenu de l'individu i et de l'individu j respectivement, $\mu$ est le revenu moyen, et N correspond à la taille de la population. Le coefficient de Gini est compris entre zéro, situation d'égalité parfaite où les revenus sont répartis de manière égale entre tous les membres de la population, et un, situation d'inégalité absolue où une seule personne concentre tous les revenus. Contrairement aux rapports interquantiles, le coefficient de Gini prend en compte les données de l'ensemble de la distribution de revenus. Pour estimer le coefficient de Gini correspondant à la courbe de Lorenz ci-dessus, on utilise la commande Stata suivante :

*igini pcminc*

```
. igini pcminc

Index: Gini index
Sampling weight : wgt
```

| Variable | Estimation | STE | LB | UB |
|---|---|---|---|---|
| 1: GINI_pcminc | 0,698286 | 0,014947 | 0,668888 | 0,727683 |

La valeur maximale du coefficient de Gini étant de 1 (inégalité absolue), la valeur de 0,69 est le signe d'un niveau d'inégalité élevé en Afrique du Sud. L'interprétation du coefficient de Gini est très intuitive. En multipliant le coefficient par deux et par le revenu moyen, on obtient l'écart de revenu attendu entre deux individus choisis au hasard au sein de la population.[11]

On peut également calculer le coefficient de Gini pour différents groupes de population. Par exemple, on peut estimer ce coefficient ventilé par groupe ethnique à l'aide de la commande Stata suivante :

*igini pcminc, hgroup(race)*

```
Index: Gini index
Sampling weight : wgt
Group variable   : race
```

| Groupe | Estimation | STE | LB | UB |
|---|---|---|---|---|
| 1: Noirs africains | 0,643054 | 0,016474 | 0,610653 | 0,675455 |
| 2: Métis | 0,592903 | 0,027139 | 0,539526 | 0,646280 |
| 3: Indiens/Asiatiques | 0,610705 | 0,064400 | 0,484043 | 0,737368 |
| 4: Blancs | 0,506923 | 0,028801 | 0,450277 | 0,563569 |
| Population | 0,698286 | 0,014947 | 0,668888 | 0,727683 |

Le coefficient de Gini est au plus bas chez les Blancs, suivis des Métis, des Indiens et des Noirs africains. Ces estimations suggèrent que les inégalités sont à leur plus haut niveau chez les Noirs africains, suivis des Indiens, des Métis puis des Blancs. Cependant, les larges intervalles de confiance qui entourent les estimations du groupe des Indiens/Asiatiques suggèrent que celles-ci ne sont pas très précises, du fait de la petite taille de l'échantillon de ce groupe de population.

On peut aussi calculer le coefficient de Gini ventilé par zone géographique. Le tableau suivant présente une estimation des inégalités de revenus par zone géographique (zone rurale/zone urbaine). D'après les coefficients de Gini obtenus, les inégalités de revenus sont plus importantes dans les zones urbaines que dans les zones rurales.

11 On peut récrire la formule du coefficient de Gini comme suit :

$$2\mu G = \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{|y_i - y_j|}{N^2}$$

Le côté droit de l'équation représente l'écart de revenus attendu entre deux individus sélectionnés au hasard au sein de la population.



*igini pcminc, hgroup(rural)*

```
Index: Gini index
Sampling weight: wgt
Group variable: rural
```

| Groupe | Estimation | STE | LB | UB |
|---|---|---|---|---|
| 1: 0. Zone urbaine | 0,669496 | 0,016767 | 0,636520 | 0,702473 |
| 2: 1. Zone rurale | 0,593243 | 0,028197 | 0,537785 | 0,648700 |
| Population | 0,698286 | 0,014947 | 0,668888 | 0,727683 |

En théorie, il est également possible de ventiler les inégalités par sexe. Toutefois, comme nous l'avons indiqué dans la section 3, estimer les inégalités de revenus en fonction du sexe nécessite d'avoir accès à des données individuelles détaillées sur la consommation et les revenus de chaque membre du foyer. Or, presque toutes les enquêtes menées auprès des ménages collectent les données de consommation au niveau du ménage. De même, à quelques rares exceptions près, les enquêtes sur les revenus et les dépenses de consommation collectent elles aussi les informations sur les revenus au niveau du ménage. De ce fait, on se contente souvent d'estimer les inégalités désagrégées en fonction du sexe du seul chef de ménage. Par exemple, dans notre cas, le tableau ci-dessous présente les inégalités de revenus ventilées par sexe du chef de famille. Les chiffres indiquent que les inégalités ne sont que légèrement plus élevées chez les individus qui vivent dans un ménage dont le chef est un homme, par rapport aux ménages dont le chef est une femme.

*igini pcminc, hgroup(hhhead)*

```
Index: Gini index
Sampling weight : wgt
Group variable    : hhhead
```

| Groupe | Estimation | STE | LB | UB |
|---|---|---|---|---|
| 1: chef_masculin | 0,699584 | 0,017137 | 0,665879 | 0,733290 |
| 2: chef_féminin | 0,681727 | 0,015050 | 0,652127 | 0,711327 |
| Population | 0,698286 | 0,014947 | 0,668888 | 0,727683 |

NIDS est l'une des rares enquêtes sur les revenus et les dépenses des ménages à recueillir des informations sur les revenus au niveau individuel (pour les adultes uniquement). On peut utiliser ces informations pour estimer les inégalités de revenus ventilées par sexe. Le tableau suivant montre les inégalités de revenus en fonction du sexe. Dans ce cas, la variable utilisée n'est pas le revenu par habitant mais le revenu individuel de chaque adulte vivant au sein du ménage. (Notons que les individus qui ne gagnent aucun revenu se voient attribuer une valeur de revenu nulle). On peut aussi effectuer une estimation similaire en utilisant les informations individuelles sur les revenus du travail ou les salaires issues des enquêtes sur les forces de travail.[12] Cela nous permet d'estimer les inégalités de rémunération ou de salaire chez les travailleurs salariés, ventilées par sexe. Cependant, dans la mesure où les ménages mettent en commun leurs revenus et autres ressources, l'utilisation directe de ces informations pour estimer les inégalités de revenus ou la pauvreté (c'est-à-dire l'utilisation des revenus au niveau individuel) peut s'avérer problématique (voir Section 3).

12 C'est l'une des façons d'estimer les résultats sur le marché du travail. Voir Wittenberg (2017) pour des estimations récentes des inégalités salariales en Afrique du Sud.

```
. igini totpinc, hgroup(gender)

    Index            : Gini index
    Sampling weight : w1_wgt
    Group variable   : gender
```

| Groupe | Estimation | STE | LB | UB |
|---|---|---|---|---|
| 1: 1. Homme | 0,642041 | 0,020633 | 0,601460 | 0,682621 |
| 2: 2. Femme | 0,650411 | 0,022168 | 0,606811 | 0,694010 |
| Population | 0,683918 | 0,016387 | 0,651688 | 0,716149 |

Une autre possibilité consiste à utiliser des indicateurs de bien-être mesurables au niveau individuel, comme le niveau d'études. On peut par exemple utiliser le nombre d'années de scolarité effectuées pour mesurer les inégalités en matière d'éducation selon le sexe (pour les personnes âgées de 15 ans et plus).[13]

```
. igini educ_yrs, hgroup(gender)

    Index            : Gini index
    Sampling weight : wgt
    Group variable   : gender
```

| Groupe | Estimation | STE | LB | UB |
|---|---|---|---|---|
| 1: Femme | 0,247800 | 0,005909 | 0,236179 | 0,259421 |
| 2: Homme | 0,229272 | 0,005900 | 0,217667 | 0,240877 |
| Population | 0,239044 | 0,005110 | 0,228993 | 0,249094 |

Les estimations ci-dessus montrent que le coefficient de Gini pour le nombre d'années de scolarité est légèrement plus élevé chez les femmes que chez les hommes, ce qui suggère que les inégalités en matière d'éducation sont plus importantes chez les femmes que chez les hommes. Cependant, contrairement aux données de revenu ou de consommation, qui sont des variables continues, les données sur le nombre d'années de scolarité sont des variables discrètes et incluent souvent un nombre important de valeurs nulles (dans la plupart des pays pauvres). Dans ce type de cas, il est conseillé d'utiliser la commande *ineqdec0*, écrite par des utilisateurs, qui permet de calculer le coefficient de Gini pour les données comprenant une quantité importante de valeurs nulles.

Le coefficient de Gini satisfait toutes les propriétés d'invariance (symétrie, principe de population, invariance à l'échelle et normalisation) ainsi que le principe de transfert. En revanche, il ne satisfait pas la propriété de sensibilité au transfert. Il peut être décomposable, mais avec un terme résiduel additionnel. Notons que le coefficient de Gini ne vérifie pas la propriété de cohérence des sous-groupes.

13 Toutefois, en Afrique du Sud, on limite souvent les données aux personnes âgées de 25 ans et plus, en partant du principe qu'à 25 ans la plupart des adultes ont terminé leur scolarité.



### *4.2.4 Les mesures d'entropie généralisée*

Si l'on a besoin d'un indice d'inégalité qui soit décomposable de manière additive et qui vérifie la propriété de cohérence des sous-groupes, on peut se tourner vers la classe des indices d'entropie. La formule mathématique correspondant à cette famille de mesures d'inégalité prend la forme suivante :

$$GE(\alpha)=\frac{1}{\alpha(\alpha-1)}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{y_i}{\mu}\right)^{-1}\right]$$

Où, $y_i$ indique le revenu individuel, $\mu$ est le revenu moyen et N représente la taille de la population. Dans la famille des indices d'entropie généralisée (GE), le paramètre $\alpha$ représente le poids donné aux différences entre les revenus à différents points de la distribution des revenus, et il peut prendre n'importe quelle valeur réelle. Si $\alpha$ a une valeur positive et élevée, l'indice d'entropie généralisée (GE) sera plus sensible aux changements dans la tranche supérieure de la distribution des revenus. À l'inverse, l'indice sera plus sensible aux changements dans la tranche inférieure de la distribution si la valeur de $\alpha$ est plus proche de zéro. Les mesures de l'indice d'entropie généralisée vont de zéro à l'infini, zéro représentant une distribution égalitaire (les revenus sont répartis de manière équitable entre tous les individus d'une population), tandis qu'une valeur plus élevée correspond à un haut niveau d'inégalité. La valeur de la limite supérieure dépend toutefois de la valeur spécifique attribuée à $\alpha$. Les valeurs de $\alpha$ les plus couramment utilisées sont 0,1 et 2.[14] L'indice GE (1) s'appelle l'indice T de Theil, et l'indice GE (0) s'appelle l'indice L de Theil (déviation logarithmique moyenne).[15]

La formule de l'indice T de Theil, ou GE (1), est la suivante :

$$T_T=\frac{1}{N}\sum_{i=1}^{N}\left(\frac{y_i}{\mu}\right)ln\left(\frac{y_i}{\mu}\right)$$

Tandis que la formule de l'indice L de Theil, ou GE (0), est la suivante :

$$T_L=-\frac{1}{N}\sum_{i=1}^{N}ln\left(\frac{y_i}{\mu}\right)$$

La limite supérieure pour GE (1) est ln(N), tandis que les valeurs correspondantes pour GE (0) sont non bornées. Dans les deux cas, les individus ayant des revenus nuls seront automatiquement exclus du calcul étant donné que le logarithme de zéro est indéfini.

14 Pour les valeurs négatives de $\alpha$, les mesures d'inégalité GE ($\alpha$) sont indéfinies en présence de revenus nuls. En pratique, on utilise donc des valeurs positives de $\alpha$.

15 On ne peut pas dériver les indices T et L de Theil de l'équation GE ($\alpha$) en substituant directement $\alpha$= 0 ou $\alpha$=1. Il faut pour cela avoir recours à la règle de L'Hôpital. Ce procédé consiste d'abord à dériver le dénominateur et le numérateur séparément, puis à calculer la limite du rapport de ces fonctions.

Cette exclusion est problématique si les revenus nuls correspondent à une valeur réelle. L'une des approches souvent utilisées consiste alors à remplacer les valeurs de revenus nulles par des valeurs peu élevées comme 1 (Jenkins & Jantti, 2005).

Pour calculer les indices d'inégalité T et L de Theil, on peut utiliser les commandes Stata suivantes :

*Pour GE(0)*

*ientropy pcminc, theta(0)*

*Pour GE(1)*

*ientropy pcminc, theta(1)*

Index : Entropy index
Parameter theta : 0
Sampling weight : wgt

| Variable | Estimation | STE | LB | UB |
|---|---|---|---|---|
| 1: entropy_pcminc | 1,054390 | 0,057380 | 0,941535 | 1,167244 |

Index : Entropy index
Parameter theta : 1
Sampling weight : wgt

| Variable | Estimation | STE | LB | UB |
|---|---|---|---|---|
| 1: entropy_pcminc | 1,018017 | 0,063061 | 0,893989 | 1,142045 |

Il est également possible de ventiler les indices de type GE($\alpha$) par groupe. Par exemple, l'indice GE(1) ventilé par groupe ethnique s'obtient de la manière suivante :

*ientropy pcminc, hgroup(race) theta(1)*

Index : Entropy index
Parameter theta : 1
Sampling weight : wgt
Group variable : race

| Groupe | Estimation | STE | LB | UB |
|---|---|---|---|---|
| 1: Noirs africains | 0,885423 | 0,073779 | 0,740315 | 1,030531 |
| 2: Métis | 0,700526 | 0,079994 | 0,543194 | 0,857858 |
| 3: Indiens/Asiatiques | 0,686213 | 0,175553 | 0,340935 | 1,031491 |
| 4: Blancs | 0,464310 | 0,056843 | 0,352510 | 0,576109 |
| Population | 1,018017 | 0,063061 | 0,893989 | 1,142045 |





Les estimations obtenues montrent que c'est parmi les Noirs africains que les inégalités sont les plus fortes, et parmi les Blancs qu'elles sont les plus faibles. Ce résultat est conforme aux estimations obtenues avec le coefficient de Gini. Cependant, on ne peut pas comparer les grandeurs obtenues à partir de l'indice d'entropie généralisée avec celles issues du coefficient de Gini. La valeur du coefficient de Gini est comprise entre 0 et 1, tandis que les valeurs des indices d'entropie généralisée vont de zéro à l'infini. Notons que ces différents indicateurs peuvent donner lieu à des classements différents du niveau d'inégalité pour une même distribution, car leur sensibilité aux différences entre les revenus à différents points de la distribution est variable.

Les indices d'entropie généralisée respectent tous les axiomes d'invariance : principe de population, invariance à l'échelle, normalisation et symétrie. En outre, pour $\alpha$ <2, les mesures d'entropie généralisée sont sensibles au transfert. Par ailleurs, utiliser les indices d'entropie généralisée pour mesurer les inégalités présente un avantage majeur : contrairement à l'indice de Gini, cette classe de mesures d'inégalité est décomposable de manière additive et satisfait l'axiome de cohérence des sous-groupes. On peut donc utiliser ces indices pour décomposer l'inégalité totale en deux composantes : les inégalités inter-groupes et les inégalités intra-groupes. Plusieurs facteurs peuvent être utilisés comme variable de classement, notamment l'origine ethnique, le sexe, la zone géographique et les sources de revenus. Par exemple, en prenant l'origine ethnique comme variable de classement, on peut décomposer les inégalités de revenus totales en Afrique du Sud en deux composantes : les inégalités « inter-groupes ethniques » et les inégalités « intra-groupes ethniques ». Pour décomposer l'indice GE(1) en composantes inter- et intra-groupes, on utilise la commande suivante* :

*dentropyg pcminc, hgroup(race) theta(1)*

Décomposition de l'indice d'entropie généralisée par groupes
Sampling weight : wgt
Group variable : race
Parameter theta : 1,00

| Groupe | Entropy index | Part de la population | (mu_k/mu)^theta | Contribution absolue | Contribution relative |
|---|---|---|---|---|---|
| 1: Noirs africains | 0,885423 | 0,756531 | 0,576466 | 0,386146 | 0,379312 |
| | 0,073779 | 0,025794 | 0,056654 | 0,058966 | 0,066516 |
| 2: Métis | 0,700526 | 0,097209 | 0,775567 | 0,052814 | 0,051880 |
| | 0,079994 | 0,015724 | 0,114612 | 0,012392 | 0,013490 |
| 3: Indiens/Asiat. | 0,686213 | 0,029545 | 2,180964 | 0,044217 | 0,043435 |
| | 0,175553 | 0,011521 | 0,782087 | 0,017998 | 0,017887 |
| 4: Blancs | 0,464310 | 0,116714 | 3,633287 | 0,196893 | 0,193409 |
| | 0,056843 | 0,018548 | 0,380633 | 0,037001 | 0,028353 |
| Intra | --- | --- | --- | 0,680071 | 0,668035 |
| | --- | --- | --- | 0,063617 | --- |
| Inter | --- | --- | --- | 0,350333 | 0,344133 |
| | --- | --- | --- | 0,017840 | --- |
| Population | 1,018017 | 1,000000 | --- | 1,018017 | 1,000000 |
| | 0,063061 | 0,000000 | --- | 0,063061 | 0,000000 |

Si l'on se base sur la contribution relative, environ 66 % des inégalités de revenus totales en Afrique du Sud en 2008 étaient dues à l'inégalité intra-groupes ethniques, contre 34 % pour l'inégalité inter-groupes ethniques.



*Toutes les observations comportant des données manquantes pour la variable revenu doivent être éliminées avant d'exécuter cette commande.*

On peut également décomposer les inégalités de revenus par source de revenus. Plusieurs méthodes (par régression ou non) permettent de le faire (voir par exemple Shorrocks, 1982, 2013 ; Fields, 2003). On peut notamment avoir recours au module DASP « dsineqs », qui utilise la méthode de décomposition basée sur la valeur de Shapley. Selon Shorrocks (2013 : p.101), cette méthode de décomposition consiste à calculer l'effet marginal sur les inégalités de « *l'élimination successive de chacun des facteurs contributifs, puis à attribuer à chaque facteur la moyenne de ses contributions marginales dans toutes les séquences d'élimination possibles*. » Cette méthode permet donc de décomposer les mesures d'inégalité sans résidu. Grâce à la valeur de Shapley, on peut décomposer les inégalités de revenus par source de revenus en utilisant les indices de Gini, d'Atkinson et d'entropie généralisée. Notons que si la procédure de décomposition par la valeur de Shapley peut être utilisée pour décomposer les inégalités à l'aide de l'indice de Gini, elle ne résout pas pour autant le problème d'incohérence des sous-groupes associé à cet indice.

Pour décomposer les inégalités de revenus par source de revenus, nous utilisons les données de l'enquête NIDS de 2008. Cinq sources de revenus sont prises en compte : les revenus salariaux (*wage*), les revenus provenant des aides sociales (*grants*), les revenus issus des transferts de fonds des travailleurs émigrés (*remittance*), les revenus du capital (*capital*) et les revenus provenant d'autres sources (*others*). La commande ci-dessous permet de décomposer les inégalités de revenus par source de revenus en utilisant l'indice GE(1).

*dsineqs wage other grants remittance capital, index(ge) theta(1)*

```
Décomposition de l'indice d'inégalité par sources de revenus (par la valeur de Shapley).
Temps d'exécution :        2,34 secondes


ineq index      :      1.484097
Sampling weight : w1_wgt
```

| Sources | Part du revenu | Contribution absolue | Contribution relative |
|---|---|---|---|
| 1: Pwageinc | 0,791420 | 0,998938 | 0,673094 |
| 2: Potherinc | 0,073475 | 0,185506 | 0,124996 |
| 3: Pgrantinc | 0,078498 | 0,066625 | 0,044893 |
| 4: Premitinc | 0,044207 | 0,142528 | 0,096037 |
| 5: Pcapitinc | 0,012400 | 0,090501 | 0,060980 |
| Total | 1,000000 | 1,484097 | 1,000000 |

On peut également décomposer les inégalités de revenus par source de revenus en utilisant l'indice de Gini.

*dsineqs wage other grants remittance capital, index(gini)*

```
Décomposition de l'indice d'inégalité par source de revenus (en utilisant la valeur de
Shapley).
Temps d'exécution :        3,44 secondes
ineq index      :      0.723094
Sampling weight : w1_wgt
```

| Sources | Part du revenu | Contribution absolue | Contribution relative |
|---|---|---|---|
| 1: Pwageinc | 0,791420 | 0,595506 | 0,823553 |
| 2: Potherinc | 0,073475 | 0,062187 | 0,086001 |
| 3: Pgrantinc | 0,078498 | 0,022013 | 0,030442 |
| 4: Premitinc | 0,044207 | 0,032681 | 0,045196 |
| 5: Pcapitinc | 0,012400 | 0,010707 | 0,014808 |
| Total | 1,000000 | 0,723094 | 1,000000 |





Le tableau ci-dessus montre de façon évidente que les revenus salariaux sont le principal vecteur d'inégalité de revenus en Afrique du Sud. D'après la décomposition de l'indice de Gini, environ 82 % des inégalités de revenus sont dues aux revenus salariaux. Ce constat n'a rien de surprenant dans la mesure où la part du revenu salarial dans le revenu total est proche de 80 %.

### 4.2.5 Le coefficient de variation

Le coefficient de variation (CV) est un autre outil souvent utilisé pour mesurer les inégalités. On l'utilise notamment pour analyser les inégalités spatiales et horizontales. La formule de base pour calculer ce coefficient est la suivante :

$$CV=\sqrt{\frac{\sum_{I}^{N}\left(Y-Y_{i}\right)^{2}/N}{\bar{Y}}}$$

Où $Yi$ indique le revenu individuel, $\bar{Y}$ est le revenu moyen, et N est le nombre d'individus. Les valeurs du CV vont de zéro à l'infini, une valeur élevée indiquant une distribution plus inégale des revenus. Parmi les propriétés souhaitables, le CV satisfait les principes d'anonymat, de normalisation, d'invariance à l'échelle et de transfert. Il est également décomposable de manière additive et vérifie la propriété de cohérence des sous-groupes. En revanche, il ne satisfait pas l'axiome de sensibilité au transfert. Le CV est également affecté par les valeurs extrêmes car il dépend du carré de la distance entre la valeur moyenne et les valeurs individuelles.[16] Nous reviendrons sur le CV dans la sous-section 4.5, lorsque nous aborderons les approches permettant de mesurer les inégalités spatiales.

### 4.2.6 La classe des indices d'Atkinson

Les indices d'inégalité que nous avons évoqués jusqu'à présent étaient des indices descriptifs, dérivés sans intégrer explicitement de fonction de bien-être social. Cependant, certains chercheurs estiment que ces mesures d'inégalité servent aussi à élaborer des politiques publiques et sont assorties de jugements de valeur implicites (Atkinson, 1970). Atkinson a donc proposé une mesure d'inégalité basée sur le bien-être, que l'on appelle la classe des indices d'inégalité d'Atkinson. La formule correspondant à ces mesures d'inégalité est la suivante :

$$I = 1 - \left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{y_i}{\mu}\right)^{(1-\varepsilon)}\right]^{\frac{1}{(1-\varepsilon)}}$$

16 D'autres mesures d'inégalité sont également affectées par la présence de valeurs extrêmes parmi les données. Par exemple, Cowell et Flachaire (2007) ont montré que les mesures d'entropie généralisée (GE) où $\alpha > 1$ étaient très sensibles à la présence de revenus élevés parmi les données. De même, les indices d'entropie généralisée où $\alpha <0$ et l'indice d'Atkinson où $\alpha > 1$ sont très sensibles à la présence de revenus très faibles dans les données. A l'inverse, le coefficient de Gini est moins sensible à la présence de valeurs extrêmes.

Où $y_i$ indique le revenu individuel, $\mu$ est le revenu moyen, et N représente la taille de la population. Le paramètre des indices d'inégalité d'Atkinson représente le degré d'aversion pour l'inégalité ($\varepsilon$) et peut prendre des valeurs allant de zéro à l'infini. Les valeurs les plus couramment utilisées sont 0,5, 1,5, 1 et 2. Le choix de ce paramètre est quelque peu arbitraire. Lorsque la valeur du paramètre d'aversion est élevée, cela signifie que le bien-être social est plus sensible à la variation du revenu d'un individu pauvre qu'à la même variation affectant un individu plus fortuné.

Tous les indices de la classe d'Atkinson ont une valeur comprise entre zéro (égalité parfaite) et un (inégalité maximale). Les indices d'Atkinson respectent tous les axiomes d'invariance (principe de population, invariance à l'échelle, symétrie et *normalisation*). Cette classe de mesures d'inégalité satisfait également le principe de transfert et l'axiome de sensibilité au transfert. Bien que ces indices d'inégalité vérifient la propriété de cohérence des sous-groupes, ils ne sont pas décomposables de manière additive. Cependant, comme indiqué dans la sous-section 4.2, on peut utiliser la procédure de décomposition par la valeur de Shapley pour décomposer les indices d'Atkinson sans effet résiduel.

Pour estimer l'indice d'Atkinson pour un paramètre $\varepsilon$ de valeur 1,5, on utilise la commande Stata suivante :

*iatkinson pcminc, epsilon(1.5)*

```
Index              : Atkinson index
Parameter epsilon  : 1,5
Sampling weight    : wgt
```

| Variable | Estimation | STE | LB | UB |
|---|---|---|---|---|
| 1: atk_pcminc | 0,809247 | 0,016304 | 0,777180 | 0,841315 |

Sachant que les valeurs de l'indice d'Atkinson sont comprises entre zéro et un, le chiffre de 0,809, proche de 1, indique un niveau d'inégalité élevé. Comme pour le coefficient de Gini, l'interprétation de l'indice d'Atkinson est très intuitive. Cet indice mesure la perte de bien-être occasionnée par les inégalités au sein d'une société. Par exemple, un indice de 0,809 signifie qu'environ 80 % du revenu actuel est perdu (gaspillé) du fait des inégalités. En d'autres termes, la société n'aurait besoin que de 19,1 % du revenu national actuel pour atteindre le même niveau de bien-être social si tous les revenus étaient distribués de manière égale.

Il est également possible d'estimer l'indice d'Atkinson en le ventilant par groupe ethnique comme suit :

*iatkinson pcminc, hgroup(race) epsilon(1.5)*





```
Index              : Atkinson index
Parameter epsilon  : 1,5
Sampling weight    : wgt
Group variable     : race
```

| Groupe | Estimation | STE | LB | UB |
|---|---|---|---|---|
| 1: Noirs africains | 0,713273 | 0,019869 | 0,674196 | 0,752351 |
| 2: Métis | 0,795259 | 0,038163 | 0,720201 | 0,870317 |
| 3: Indiens/Asiatiques | 0,926112 | 0,041680 | 0,844136 | 1,008087 |
| 4: Blancs | 0,697085 | 0,071384 | 0,556686 | 0,837484 |
| Population | 0,809247 | 0,016304 | 0,777180 | 0,841315 |

Ici, c'est chez les Indiens que les inégalités de revenus sont les plus fortes, suivis des Métis et des Noirs africains, tandis qu'elles sont à leur niveau le plus bas chez les Blancs. Les résultats de cette estimation sont donc différents de ceux que nous avons obtenus en utilisant le coefficient de Gini. Bien que les valeurs du coefficient de Gini et des indices d'Atkinson varient entre zéro et un, ces deux mesures peuvent malgré tout classer différemment un même ensemble de distributions dans la mesure où leur sensibilité aux différences entre les revenus à différents points de la distribution n'est pas la même. En outre, comme nous l'avons mentionné plus haut, les larges intervalles de confiance autour des estimations pour le groupe des Indiens suggèrent que ces dernières ne sont pas très précises, probablement du fait de la petite taille de l'échantillon en question.

## 4.3 Dynamiques des inégalités

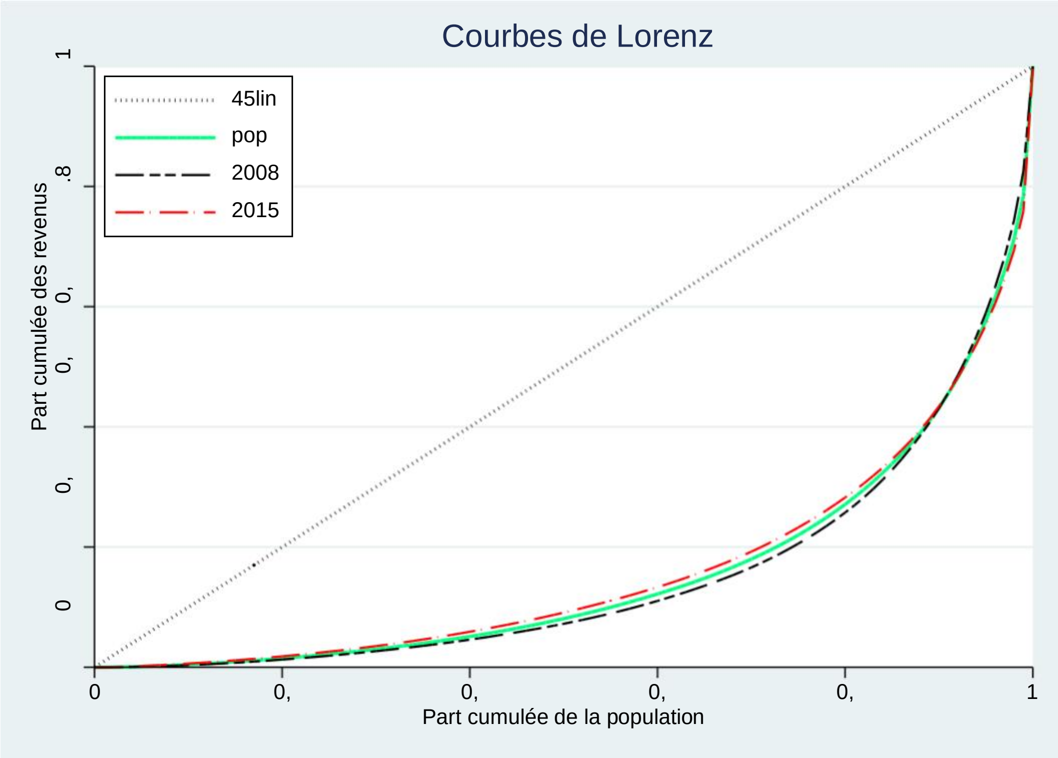
Pour évaluer l'impact d'une politique sur le bien-être social, il est nécessaire d'analyser les tendances en matière de pauvreté et d'inégalité. Pour pouvoir estimer ces tendances, il faut disposer de données sur au moins deux points dans le temps. Dans cette perspective, il est important que les données et les indicateurs de bien-être utilisés soient cohérents dans le temps. Comme nous l'avons vu dans la section 3, les variations du niveau d'inégalité mesuré au fil du temps peuvent être dus à des changements réels au sein de la distribution des revenus, mais aussi à d'autres facteurs tels que des modifications des modalités de collecte des données, un ajustement des prix ou tout autre changement méthodologique. Une fois que l'on maîtrise bien les données et les questions de mesure, on peut utiliser n'importe laquelle des mesures d'inégalité abordées plus haut pour comparer les niveaux d'inégalité dans le temps. On peut par exemple comparer le niveau des inégalités de revenus en 2008 et en 2015 grâce à des courbes de Lorenz en procédant de la manière suivante :

On commence par créer une variable de revenu en valeur réelle en prenant pour base les prix de 2008. En 2008, la valeur de l'IPC était de 63,6 (moyenne annuelle), contre 92,0 (moyenne annuelle) en 2015. Il faut donc multiplier notre variable *pcminc* 2015 par 63,6/92,0 pour obtenir une variable de revenu par habitant en valeur réelle, *real_pcminc*.[17] On utilise ensuite cette variable pour estimer l'évolution des inégalités dans le temps. La commande qui permet d'estimer les courbes de Lorenz pour plusieurs années, soit 2008 et 2015 dans notre cas, est la suivante :

*clorenz real_pcminc, hgroup(year)*

17 Les valeurs de *pcminc* pour 2008 n'auront pas besoin d'être ajustées, dans la mesure où nous avons choisi 2008 comme année de base pour l'ajustement en fonction de l'inflation. Pour 2008, *pcminc* et *real_pcminc* sont donc identiques.

Figure 5.




Comme les courbes de Lorenz des distributions de 2008 et de 2015 se croisent, il est impossible d'en tirer une quelconque conclusion quant aux tendances en matière d'inégalités de revenus sur la période. Pour comparer les inégalités dans le temps, on peut alors utiliser d'autres indices d'inégalité tels que le coefficient de Gini ou l'indice d'Atkinson (on peut choisir $\varepsilon$ =1,5 par exemple).

```
. igini real_pcminc,hgroup(year)

    Index             : Gini index
    Group variable    : year
```

| Groupe | Estimation | STE | LB | UB |
|---|---|---|---|---|
| 1: 2008 | 0,661186 | 0,017534 | 0,626792 | 0,695579 |
| 2: 2015 | 0,599787 | 0,032764 | 0,535519 | 0,664055 |
| Population | 0,630340 | 0,020799 | 0,589544 | 0,671137 |

```
. iatkinson real_pcminc , hgroup(year) epsilon(1.5)

    Index             : Atkinson index
    Parameter epsilon : 1,5
    Sampling weight   : wgt
    Group variable    : year
```

| Groupe | Estimation | STE | LB | UB |
|---|---|---|---|---|
| 1: 2008 | 0,809247 | 0,016206 | 0,777460 | 0,841035 |
| 2: 2015 | 0,718899 | 0,031568 | 0,656979 | 0,780820 |
| Population | 0,766667 | 0,018419 | 0,730538 | 0,802796 |





Les résultats obtenus à partir du coefficient de Gini et des indices d'Atkinson suggèrent que le niveau d'inégalité était plus bas en 2015 qu'en 2008, ce qui témoigne d'une baisse des inégalités de revenus sur la période concernée.

Il est également possible de décomposer le niveau d'inégalité en composantes intra-groupe et inter-groupes et d'en comparer l'évolution dans le temps. On peut ainsi décomposer les inégalités de revenus par groupe ethnique pour 2008 et 2015 et déterminer si la contribution de l'inégalité intra-groupe ou inter-groupes a évolué au fil du temps. Pour ce faire, il faut effectuer la décomposition séparément pour chaque année. Le code et les données de sortie correspondants sont présentés ci-dessous, ainsi que les résultats de l'estimation.

En 2008, les inégalités intra-groupe ethnique ont contribué pour environ 66 % à l'inégalité totale. Ce pourcentage a ensuite augmenté pour atteindre 75 % en 2015. On peut donc en conclure que la contribution des inégalités intra-groupe ethnique a augmenté au fil du temps, tandis que la contribution des inégalités inter-groupes ethniques a diminué.

Décomposition des inégalités par groupe ethnique pour 2008 :

```
. preserve

. keep if year == 2008 (27,105
observations deleted)

. dentropyg real_pcminc, hgroup(race) theta(1)

    Décomposition de l'indice d'entropie généralisée par groupes
    Sampling weight : wgt
    Group variable : race
    Parameter theta :  1,00
```

| Groupe | Indice d'entropie | Part de la population | (mu_k/mu)^theta | Contribution absolue | Contribution relative |
|---|---|---|---|---|---|
| 1: Noirs africains | 0,885423 | 0,756531 | 0,576466 | 0,386146 | 0,379312 |
| | 0,073779 | 0,025794 | 0,056654 | 0,058966 | 0,066516 |
| 2: Métis | 0,700526 | 0,097209 | 0,775567 | 0,052814 | 0,051880 |
| | 0,079994 | 0,015724 | 0,114612 | 0,012392 | 0,013490 |
| 3: Indiens/Asiat. | 0,686213 | 0,029545 | 2,180964 | 0,044217 | 0,043435 |
| | 0,175553 | 0,011521 | 0,782087 | 0,017998 | 0,017887 |
| 4: Blancs | 0,464310 | 0,116714 | 3,633287 | 0,196893 | 0,193409 |
| | 0,056843 | 0,018548 | 0,380633 | 0,037001 | 0,028353 |
| Intra | --- | --- | --- | 0,680071 | 0,668035 |
| | --- | --- | --- | 0,063617 | --- |
| Inter | --- | --- | --- | 0,350333 | 0,344133 |
| | --- | --- | --- | 0,017840 | --- |
| Population | 1,018017 | 1,000000 | --- | 1,018017 | 1,000000 |
| | 0,063061 | 0,000000 | --- | 0,063061 | 0,000000 |

```
. restore
```

Décomposition des inégalités par groupe ethnique pour 2015 :

```
. preserve

. keep if year == 2015 (18,480 observations deleted)

. dentropyg real_pcminc, hgroup(race) theta(1)

    Décomposition de l'indice d'entropie généralisée par
    groupes Sampling weight : wgt
    Group variable : race
    Parameter theta : 1,00
```

| Groupe | Indice d'entropie | Part de la population | (mu_k/mu)^theta | Contribution absolue | Contribution relative |
|---|---|---|---|---|---|
| 1: Noirs africains | 0,742994 | 0,782785 | 0,632436 | 0,367828 | 0,308164 |
| | 0,049069 | 0,024854 | 0,073726 | 0,052768 | 0,109988 |
| 2: Métis | 0,659462 | 0,093881 | 0,826108 | 0,051145 | 0,042849 |
| | 0,168030 | 0,018313 | 0,165617 | 0,020127 | 0,022090 |
| 3: Indiens/Asiat. | 0,643094 | 0,026580 | 2,106360 | 0,036005 | 0,030165 |
| | 0,071416 | 0,012041 | 0,648521 | 0,015403 | 0,016512 |
| 4: Blancs | 1,182039 | 0,096753 | 3,838578 | 0,439003 | 0,367793 |
| | 0,440893 | 0,015091 | 0,683283 | 0,235720 | 0,109109 |
| Intra | --- | --- | --- | 0,893981 | 0,748971 |
| | --- | --- | --- | 0,289221 | --- |
| Inter | --- | --- | --- | 0,294447 | 0,246685 |
| | --- | --- | --- | 0,019278 | --- |
| Population | 1,193613 | 1,000000 | --- | 1,193613 | 1,000000 |
| | 0,291455 | 0,000000 | --- | 0,291455 | 0,000000 |

## 4.4 Indices d'inégalité multidimensionnelle (Indices d'équipement des ménages)

Jusqu'à présent, nous nous sommes concentrés sur l'estimation des inégalités économiques au moyen d'une mesure unidimensionnelle du bien-être, à savoir le revenu par habitant ou la consommation.

Cependant, la section 2 a souligné le fait que les inégalités couvraient des dimensions multiples, comme l'éducation, l'équipement des ménages, la santé et bien d'autres encore, et qu'il était possible que les revenus ne soient pas un bon indicateur du bien-être individuel. Il est ainsi largement admis que les inégalités de revenus ne sont qu'une mesure approximative des inégalités de bien-être ou des inégalités économiques (Sen, 1992).

Selon Sen, les individus ont des façons très hétérogènes de convertir leurs revenus et autres ressources en bien-être. Leurs conditions de vie doivent donc être évaluées en tenant compte à la fois de leurs réalisations effectives en matière de bien-être (fonctionnements) et de leur capacité à faire (capabilités). Les réalisations effectives sont notamment le fait d'être bien nourri, éduqué et en bonne santé. En s'appuyant sur l'approche par les capabilités de Sen, des études récentes tentent de mesurer les inégalités en adoptant une perspective multidimensionnelle (voir Alkire & Foster, 2011).

En ce qui concerne la mesure des inégalités, des travaux récents utilisent des indicateurs de niveau de vie basés sur l'équipement des ménages pour estimer les inégalités dans une perspective multidimensionnelle (McKenzie, 2005 ; Wittenberg & Leibbrandt, 2017). Les biens que possèdent les ménages (télévision, réfrigérateur, bétail, etc.) et l'accès aux services de base (accès à l'eau, à un système d'assainissement, matériaux de construction pour le logement, etc.) sont ainsi utilisés pour mesurer les inégalités. Avant d'utiliser certains des indices d'inégalité évoqués plus haut, il faut d'abord combiner ces différents indicateurs pour construire un indice unique (souvent appelé « indice d'équipement »).

Dans la littérature, les méthodes généralement utilisées pour calculer l'indice d'équipement des ménages sont les approches statistiques comme l'analyse factorielle (AF), l'analyse en composantes principales (ACP) et l'analyse des correspondances multiples (ACM) (Filmer & Pritchett, 2001 ; Wittenberg & Leibbrandt, 2017). Si l'on a $k$ indicateurs de niveau de vie ($a_1$, $a_2$, …,$a_k$), on peut les combiner en un seul indice en utilisant la formule suivante :

$$Indice = w_1 a_1 + w_2 a_2 + \ldots + w_k a_k$$

où $w_1$ désigne les poids associés à chaque indicateur. Dans le cas de l'ACP, les poids sont établis à partir de la première « composante principale », une combinaison linéaire correspondant à la variance maximale de la distribution des biens d'équipement. On peut ainsi écrire chaque indicateur, $a_i$, sous forme de combinaison linéaire de $k$ facteurs ou composantes de la manière suivante :

$$a_1 = v_{11} A_1 + v_{12} A_2 + \ldots v_{1k} A_k$$

$$a_2 = v_{21} A_1 + v_{22} A_2 + \ldots v_{2k} A_k$$

$$a_k = v_{k1} A_1 + v_{k2} A_2 + \ldots v_{kk} A_k$$

où $A_1$, $A_2$,…,$A_k$ sont des composantes non observées qui ne sont pas corrélées entre elles. On peut alors démontrer que la solution va se présenter sous la forme suivante :

$$A_1 = v_{11} \tilde{a}_1 + v_{12} \tilde{a}_2 + \ldots v_{1k} \tilde{a}_k$$

Où $\tilde{a}_{1i}$ indique une variable d'équipement standardisée, $\tilde{a}_{1i} = \frac{a_{1i} - \bar{a}_1}{s_1}$

La première composante principale, $A_1$, est celle qui explique la majeure partie de la covariance commune des variables d'équipement. On peut considérer que la « richesse » est la variable sous-jacente non observée qui constitue le facteur commun ($A_1$).[18] Ainsi, un indice d'équipement plus élevé suppose une « richesse » plus importante.

L'utilisation de l'ACP et autres approches similaires pose toutefois le problème suivant : certains biens, comme le bétail (principalement détenu par les ménages ruraux), peuvent se voir attribuer un poids négatif. On risque donc de se retrouver à classer les ménages ruraux qui possèdent du bétail à un rang inférieur par rapport aux ménages qui ne possèdent aucun bien (Wittenberg & Leibbrandt, 2017). En outre, les indices d'équipement construits à l'aide de ces approches ont, par construction, des valeurs moyennes nulles (McKenzie, 2005 ; Wittenberg & Leibbrandt, 2017). Dans ce cas, les mesures d'inégalité classiques ne sont pas

18 Si l'on souhaite mesurer directement la richesse, il est nécessaire de recueillir des informations détaillées sur les actifs financiers et non financiers ainsi que sur les dettes. On peut alors calculer le patrimoine net en soustrayant la valeur totale des dettes de la valeur totale des actifs. Cette variable est utilisée pour estimer les inégalités de richesse. Malheureusement, ces informations figurent rarement dans les enquêtes sur les ménages.

adaptées. Pour résoudre ces problèmes, Wittenberg et Leibbrandt (2017) suggèrent de recourir à l'approche ACP non centrée (ACP NC) pour calculer l'indice d'équipement des ménages, adoptant ainsi une méthode initialement proposée par Banerjee (2010). En s'inspirant de Wittenberg et Leibbrandt (2017), on peut utiliser l'indice d'équipement obtenu avec l'approche ACP NC pour estimer les inégalités à l'aide d'indices d'inégalité classiques comme le coefficient de Gini. Afin d'illustrer notre propos, nous utilisons les données des enquêtes DHS de 1998 en Afrique du Sud ainsi que les variables listées ci-dessous pour créer un indice d'équipement des ménages en appliquant à la fois l'approche ACP et l'approche ACP NC. Dans les enquêtes DHS et d'autres enquêtes sur les ménages, les variables relatives aux biens d'équipement sont mesurées au niveau du ménage, et c'est donc à ce niveau que nous calculons notre indice d'équipement. Dans la mesure où il n'existe pas de méthode standard pour calculer les valeurs d'un indice d'équipement par habitant, tous les membres d'un même ménage se verront attribuer la même valeur d'indice d'équipement calculée au niveau du ménage.

### Variables utilisées pour le calcul de l'indice d'équipement (Enquête DHS, 1998)

| Variable | Obs | Moyenne | Ecart-type | Min | Max |
|---|---|---|---|---|---|
| | | | | | |
| eau courante | 12 247 | 0,353719 | 0,478143 | 0 | 1 |
| électricité | 12 247 | 0,616641 | 0,486225 | 0 | 1 |
| radio | 12 247 | 0,790071 | 0,407274 | 0 | 1 |
| télévision | 12 247 | 0,54495 | 0,497996 | 0 | 1 |
| réfrigérateur | 12 247 | 0,465665 | 0,49884 | 0 | 1 |
| voiture | 12 247 | 0,226831 | 0,4188 | 0 | 1 |
| pièces | 12 136 | 2,201714 | 1,103846 | 0 | 12 |
| téléphone | 12 247 | 0,255736 | 0,436292 | 0 | 1 |
| ordinateur | 12 247 | 0,051278 | 0,220573 | 0 | 1 |
| lave-linge | 12 247 | 0,186576 | 0,389587 | 0 | 1 |
| âne/cheval | 12 247 | 0,033559 | 0,180099 | 0 | 1 |
| moutons/bétail | 12 247 | 0,124765 | 0,330466 | 0 | 1 |

À l'exception de la variable « pièces », toutes les variables sont des variables indicatrices signalant que le ménage possède le bien en question. La variable « pièces », qui mesure le nombre de pièces du logement, est une variable de comptage. Nous utilisons la commande Stata *pca,* qui permet de calculer un indice d'équipement avec l'approche ACP, comme suit :

*pca water_inhouse* electricity *radio television refrigerator car rooms telephone* computer washing_machine *donkeyhorse sheepcattle [weight=pwt]*

*predict pcaindex*

La variable *pcaindex* est la variable de l'indice d'équipement créée sur la base de la première composante principale de l'ACP. Pour connaître les coefficients pour les variables d'équipement, on régresse la variable de l'indice d'équipement sur les variables indicatrices de chaque bien comme indiqué ci-dessous :

*reg pcaindex water_inhouse* electricity *radio television refrigerator car rooms telephone* computer washing_machine *donkeyhorse sheepcattle*



À la différence de l'ACP, il n'existe pas de commande Stata permettant de calculer un indice d'équipement en utilisant l'approche ACP NC. Par conséquent, il faut d'abord exécuter un fichier ado créé par Martin Wittenberg (ucpc.ado).[19] On exécute ensuite la commande suivante :

*ucpc water_inhouse* electricity *radio television refrigerator car rooms telephone* computer washing_machine *donkeyhorse sheepcattle [weight=pwt], gen(ucpcindex)*

*reg ucpcindex water_inhouse* electricity *radio television refrigerator car rooms telephone* computer washing_machine *donkeyhorse sheepcattle*

La variable d'indice d'équipement générée grâce à l'approche ACP NC est la variable *ucpcindex*. Les estimations des coefficients pour les variables de l'indice d'équipement sont présentées dans le tableau ci-dessous :

### Estimation des coefficients pour les variables d'équipement

| variables | ACP | ACP NC |
|---|---|---|
| eau courante | 0,729 | 0,569 |
| électricité | 0,690 | 0,219 |
| radio | 0,479 | 0,138 |
| Télévision | 0,699 | 0,271 |
| Réfrigérateur | 0,760 | 0,369 |
| Voiture | 0,770 | 1,211 |
| Pièces | 0,096 | 0,048 |
| Téléphone | 0,832 | 1,002 |
| Ordinateur | 0,955 | 15,048 |
| Lave-linge | 0,879 | 1,715 |
| Âne/cheval | -0,344 | 4,646 |
| Moutons/bétail | -0,408 | 0,494 |
| _cons | -2,750 | 0,000 |

Comme on peut le constater, les coefficients estimés pour les variables relatives au bétail sont négatifs dans le cas de l'approche ACP alors qu'ils sont positifs avec l'approche ACP NC. Une fois que nous avons notre variable d'indice d'équipement, nous pouvons calculer le coefficient de Gini ou toute autre mesure d'inégalité classique sur la base de la variable *ucpcindex*. Pour calculer le coefficient de Gini à partir de la variable *ucpcindex*, on utilise la commande suivante :

svyset cluster_num [pw=pwt]

igini ucpcindex, hsize(hhsize)

19 Le fichier ado se trouve en annexe.

```
. igini ucpcindex, hsize(hhsize)

    Index             : Gini index
    Household size    : hhsize
    Sampling weight : pwt
```

| Variable | Estimation | STE | LB | UB |
|---|---|---|---|---|
| 1: GINI_ucpcindex | 0,639234 | 0,006709 | 0,626067 | 0,652401 |

Dans la commande ci-dessus, nous utilisons l'option hsize(hhsize) car notre unité d'observation est le ménage. Pour obtenir une estimation des inégalités au niveau individuel, il nous faut donc pondérer les observations au niveau du ménage par la taille du ménage (hhsize). Le coefficient estimé à partir de la variable *ucpcindex* est de 0,64. Ainsi, sur la base de l'indice de Gini, le coefficient de Gini multidimensionnel était de 0,64 en 1998.

En se basant sur l'indice d'équipement calculé ci-dessus, on peut calculer des indices d'inégalité d'équipement ventilés par groupe comme l'origine ethnique ou la région (comme nous l'avons fait précédemment avec la variable de revenu). On peut également comparer les inégalités d'équipement dans le temps ou entre régions/pays. Cependant, lorsqu'on effectue ce type de comparaisons, il convient de faire preuve de prudence pour générer les indices d'équipement car les pondérations établies à l'aide des approches statistiques évoquées ci-dessus peuvent varier d'un pays à l'autre ou dans le temps (par exemple, la distribution des biens peut varier). Par exemple, pour comparer les inégalités d'équipement au fil du temps, il faut veiller à s'appuyer sur un ensemble commun de biens d'équipement. Deux approches sont alors possibles pour définir les poids : soit on génère un indice d'équipement après avoir regroupé les données dans le temps, soit on calcule les poids en utilisant les données d'une période spécifique, puis on applique le même ensemble de poids aux autres périodes.

## 4.5 Inégalités spatiales

Les inégalités spatiales sont les inégalités entre les unités géographiques d'un pays (ou d'une région). L'unité d'analyse est donc une unité géographique (c'est-à-dire une province, une municipalité, etc.) et tous les individus d'une même unité géographique se voient attribuer le même niveau de revenu (à savoir le niveau de revenu par habitant de cette unité géographique). Dans la lignée des travaux de Williamson (1965), l'approche classique pour mesurer les inégalités spatiales/régionales consiste à utiliser le coefficient de variation, abordé dans la section 5. Cependant, contrairement à la mesure des inégalités interpersonnelles, on utilise ici le coefficient de variation pondéré par la population, calculé à l'aide de la formule suivante :

$$CV_w=\frac{\sqrt{\sum_i^m (Y_i-\bar{Y}_w)^2 * p_i}}{\bar{Y}_w}$$

Où $CV_w$ est l'estimation du coefficient de variation pondéré par la population pour une région ou un pays donné, $Y_i$ est le revenu par habitant de la sous-région $i$ au sein de la région ou du pays, $\bar{Y}_w$ est le revenu moyen des sous-régions pondéré par la part de la population



($\bar{Y}_w = \sum_{i=1}^{m} p_i Y_i = Y/N$) ; où $Y$ est le revenu total de la région ou du pays et N est la taille de la population totale de la région ou du pays, $p_i$ est la part de la population totale correspondant à la sous-région *i* ($n_i/N$, *où* $n_i$ *est la taille de la population de la sous-région i)*, et $m$ est le nombre de sous-régions que compte la région ou le pays.

Une autre mesure souvent utilisée dans le cadre de l'analyse des inégalités spatiales est l'indice de Theil. Pour mesurer les inégalités spatiales, on utilise la formule suivante de l'indice T de Theil :

$$T_T = \sum_{i=1}^{N} p_i \left(\frac{y_i}{\bar{Y}_w}\right) ln \left(\frac{y_i}{\bar{Y}_w}\right)$$

Où $y_i$ et $p_i$ désignent, respectivement, le revenu par habitant de la sous-région *i* et la part de la population totale qu'elle représente, et $\bar{Y}_w$ est le revenu moyen des sous-régions du pays ou de la région, pondéré par la part de la population ($\bar{Y}_w = \sum_{i=1}^{m} p_i y_i$).

Bien que l'indice de Theil et le coefficient de variation soient tous deux largement utilisés dans la littérature sur la mesure des inégalités spatiales, certains auteurs suggèrent d'utiliser le coefficient de variation (voir notamment Portnov & Felsenstein, 2005 ; Lessmann, 2014). Par exemple, d'après Lessmann (2014), utiliser le coefficient de variation pondéré par la population est avantageux parce que cette mesure n'est pas sensible aux valeurs extrêmes uniques, qu'elle est indépendante du nombre et de la taille des unités spatiales, qu'elle est indépendante de la moyenne et qu'elle satisfait le principe de transfert. Cependant, les travaux récents de Gluschenko (2018) remettent en question le bien-fondé d'une pondération par la part de la population régionale, l'auteur suggérant plutôt d'utiliser le coefficient de variation non pondéré pour estimer les inégalités régionales. Les travaux de Gluschenko montrent que les indices d'inégalité pondérés par la population (c'est-à-dire le CV, l'indice de Theil et le coefficient de Gini) violent les trois axiomes fondamentaux de la mesure des inégalités : le principe de population, l'anonymat et le principe de transfert. En outre, Gluschenko (2018 : p.40) montre également que les indices d'inégalité pondérés par la population ne constituent qu'une mesure approximative des inégalités interpersonnelles dans l'ensemble de la population d'un pays plutôt qu'une mesure des inégalités régionales.

En nous appuyant sur les remarques de Gluschenko (2018), nous pouvons utiliser le coefficient de variation non pondéré pour calculer les inégalités spatiales au niveau de la province en Afrique du Sud, en prenant les municipalités comme unités spatiales. La formule permettant de calculer le CV non pondéré est la suivante :

$$CV_p=\frac{\sqrt{\sum_i^m (Y_i-\bar{Y}_w)^2/m}}{\bar{Y}}$$

Où $CV_p$ est le coefficient de variation estimé pour la province $p$, $Y_i$ est le revenu par habitant de la municipalité i dans la province $p$, $\bar{Y}$ est la moyenne des revenus par habitant des municipalités dans la province $p$ ($\bar{Y} = Y_1 + Y_2 +, \ldots, +Y_m/m$), et $m$ est le nombre de municipalités au sein de la province $p$. Nous utilisons les données du recensement de 2011 pour estimer les inégalités spatiales de revenu en Afrique du Sud.[20]

Nous comparons les inégalités spatiales entre les neuf provinces en prenant les municipalités comme unités d'observation. Nos données doivent donc être recueillies au niveau de la municipalité. Nous avons besoin des informations suivantes : le revenu par habitant pour chaque municipalité ($Y_i$), la moyenne des revenus par habitant des municipalités pour chaque province ($\bar{Y}$), et le nombre de municipalités dans chaque province ($m$).[21]

Une fois les données obtenues, nous calculons le numérateur du CV à l'aide des commandes suivantes :

*gen gap_square= ( y_bar)^2 /*

*bysort: P_PROVINCE: egen gapsq_sum=sum(gap_square) gen*

*numerator=sqrt(gapsq_sum)*

Ensuite, nous calculons le coefficient de variation pour chaque province à l'aide de la commande suivante :

*gen CV= numerator /y_bar*

Comme le nombre d'unités spatiales (municipalités) varie d'une province à l'autre, les valeurs maximales des estimations du CV varient en conséquence, ce qui rend difficile la comparaison des inégalités spatiales entre provinces. Pour résoudre ce problème, l'une des approches consiste à standardiser la valeur du CV en la divisant par sa limite supérieure, qui est donnée par $\sqrt{m-1}$ (Gluschenko, 2018).

20 Nous utilisons ici des données de recensement parce que la plupart des enquêtes sur les ménages ne sont pas représentatives au niveau des unités géographiques de rang inférieur.

21 Dans notre cas, comme les données sont de niveau individuel, le revenu par habitant de chaque municipalité peut être calculé à l'aide des commandes suivantes (la variable *perincome* indique le revenu par habitant au niveau individuel ; F00_NR est l'identifiant des individus, P_MUNIC est l'identifiant de la municipalité et P_PROVINCE est l'identifiant de la province) :

Taille totale de la population pour chaque municipalité :
bysort P_MUNIC: egen Mun_pop= count (F00_NR)

Revenu total au niveau de la municipalité :
bysort P_MUNIC:egen Ymt=sum(perincome)

Revenu par habitant de chaque municipalité :
gen yi=Ymt/Mun_pop
On peut également calculer le revenu par habitant de chaque province comme suit :
bysort P_PROVINCE: egen Pov_pop= count (F00_NR) bysort
P_PROVINCE:egen Ymp=sum(perincome)
gen Y_bar=Ymp/Pov_pop

Une fois que l'on a obtenu le revenu par habitant pour chaque municipalité et pour chaque province, on peut se contenter des données au niveau des municipalités. La commande suivante permet d'éliminer les observations répétées :
sort P_MUNIC
drop if P_MUNIC==P_MUNIC[_n-1]
On peut ensuite compter le nombre de municipalités au sein d'une même province en utilisant la commande suivante :
gen temp=1
bysort P_PROVINCE:egen m=sum(temp)





Après cette standardisation, les valeurs du CV sont comprises entre zéro et un. Pour effectuer la standardisation, on utilise la commande suivante :

*gen upprbond=sqrt( -1)*

*gen CV_stand= CV/upprbond*

Le tableau ci-dessous présente les estimations brutes et standardisées du coefficient de variation par province. En observant les valeurs standardisées du CV, on constate que c'est dans la province de Mpumalanga que les inégalités spatiales sont les plus importantes, suivie par les provinces du Nord-Ouest et de Gauteng, tandis que ces inégalités sont les plus faibles dans les provinces du Cap-Occidental et du Cap-Nord.

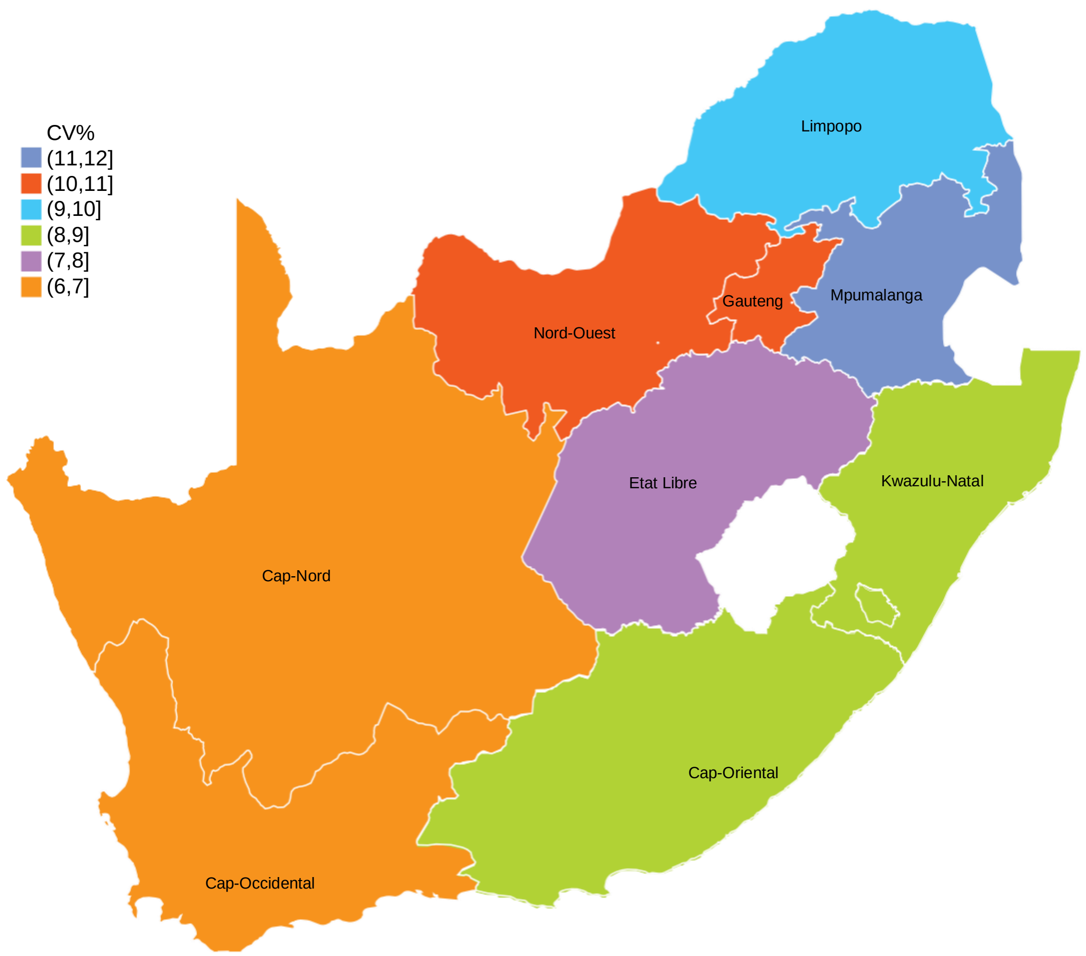
```
. tabstat CV CV_stand m,by(P_PROVINCE)

Summary statistics: mean
  by categories of: P_PROVINCE (Province)
```

| P_PROVINCE | CV | CV_stand | m |
|---|---|---|---|
| 1. Cap-Occidental | 0,2739636 | 0,0559226 | 25 |
| 2. Cap-Oriental | 0,4683907 | 0,075983 | 39 |
| 3. Cap-Nord | 0,3409061 | 0,0668572 | 27 |
| 4. Etat Libre | 0,3271644 | 9,0750567 | 20 |
| 5. Kwazulu-Natal | 0,6304823 | 0,0891637 | 51 |
| 6. Nord-Ouest | 0,4707559 | 0,1109582 | 19 |
| 7. Gauteng | 0,3317466 | 0,1105822 | 10 |
| 8. Mpumalanga | 0,4741826 | 0,1150062 | 18 |
| 9. Limpopo | 0,4826842 | 0,0985275 | 25 |
| Total | 0,452491 | 0,0853092 | 31,13675 |

Il est également possible d'établir une carte des inégalités spatiales estimées par province. L'annexe C contient les instructions de base pour réaliser une représentation cartographique à l'aide de Stata.

Figure 6.




Il convient de noter que tous les recensements ne collectent pas d'informations sur les revenus. Cependant, la plupart des recensements recueillent des informations sur l'accès aux services de base (comme l'électricité, l'eau, l'assainissement), le niveau d'éducation et les biens que possèdent les ménages. On peut alors utiliser ces variables pour construire un indice d'équipement et utiliser cette mesure plutôt que les revenus pour estimer les inégalités spatiales.

## 4.6 Inégalités horizontales

Le coefficient de variation est également très utilisé dans la littérature consacrée à l'analyse des inégalités horizontales. Les inégalités horizontales concernent les différences entre des groupes identitaires bien définis au sein d'une population (Stewart et al, 2010). Dans la mesure des inégalités horizontales, au lieu d'unités spatiales, des variables de classement comme l'origine ethnique sont utilisées comme unités d'analyse et tous les individus d'un groupe donné se voient attribuer le même niveau de revenu (à savoir le revenu moyen du groupe). Stewart et al. (2010) proposent d'utiliser le coefficient de variation du groupe





$$GCOV = \frac{1}{\bar{Y}} \left( \sum_{r=1} n_m(\bar{Y}_m - \bar{Y})^2 \right)^{½}$$

Où $m$ est le nombre de groupes ethniques ; $n_m$ est la part de la population que représente le groupe $m$ ; $\bar{Y}$ est la moyenne générale de la variable de revenu ; et $\bar{Y}_m$ est le revenu moyen du groupe $m$. Plus la valeur du GCOV est élevée, plus les inégalités entre les groupes sont importantes (c'est-à-dire plus les inégalités horizontales sont fortes).

La façon la plus simple de mesurer les inégalités horizontales consiste à examiner les revenus moyens ou médians par groupe de population. On peut par exemple comparer les revenus moyens et médians des différents groupes ethniques d'Afrique du Sud au cours de l'année 2015 en procédant de la manière suivante :

*tabstat pcminc [w=wgt], by(race) s(median mean)*

```
Summary for variables: pcminc
     by categories of: race (19 :Population Group : Section 0.0)
```

| groupe ethnique | p50 | moyenne |
|---|---|---|
| Noirs africains | 1000 | 2092,421 |
| Métis | 1410 | 2733,188 |
| Indiens/Asiatiques | 3223,333 | 6968,913 |
| Blancs | 5610 | 12699,97 |
| Total | 1221,982 | 3308,51 |

(GCOV) pondéré par la population pour mesurer les inégalités horizontales. Le coefficient de variation du groupe correspond au rapport de l'écart-type à la moyenne, l'unité d'analyse étant le groupe. De façon formelle, le GCOV peut s'écrire de la manière suivante :

Le tableau ci-dessus nous permet de constater que les revenus moyens et médians sont à leur niveau le plus bas chez les Noirs africains et à leur niveau le plus élevé chez les Blancs. Étant donné que le revenu moyen peut être influencé par la présence de valeurs extrêmes, il est préférable d'indiquer le revenu médian et d'effectuer les comparaisons sur cette base. En 2015, le revenu médian des Blancs était 5,6 fois plus élevé que celui des Noirs africains. De même, le revenu médian du groupe des Indiens/asiatiques était 3,2 fois plus élevé que celui des Noirs africains.

On peut également utiliser le sexe comme variable de classement et comparer l'écart entre les hommes et les femmes en matière de revenu moyen comme indiqué ci-dessous :

*tabstat pcminc [w=wgt], by(gender) s(median mean)*

```
(analytic weights assumed)

Summary for variables: pcminc
by categories of: gender
```

| Sexe | p50 | moyenne |
|---|---|---|
| femme | 1018,571 | 2774,455 |
| homme | 1416,667 | 3880,493 |
| Total | 1220 | 3306,908 |

Les revenus moyens et médians sont plus élevés chez les hommes que chez les femmes, le revenu médian des hommes étant 1,4 fois supérieur à celui des femmes.

Le ratio des revenus moyen et médian est la méthode la plus simple pour mesurer les inégalités horizontales. Cependant, ce type de mesure s’avère moins pratique quand le nombre de groupes est important (par exemple, dans le cas des groupes ethniques). Dans ce cas, il est préférable d’utiliser un indice permettant de synthétiser les inégalités horizontales entre les différents groupes de population. Parmi ces mesures, on peut citer le coefficient de variation du groupe (GCOV) pondéré par la population, suggéré par Stewart et al. (2010). En utilisant les données des enquêtes NIDS de 2008 et de 2015, on peut calculer le GCOV à l’aide des commandes Stata suivantes :[22]

*bys year:gen gap_square= Nm*(Ym_Y_bar)^2 bys*

*year: egen gapsq_sum=sum(gap_square) gen*

*numerator=sqrt(gapsq_sum)*

*gen GCOVr= numerator/Y_bar*

tabstat GCOVr,by(year)

Pour calculer les inégalités horizontales par sexe, on suit la même procédure que pour le groupe ethnique, mais en remplaçant la variable de l’ethnie par celle du sexe. Le tableau ci-dessous présente les estimations des inégalités horizontales par groupe ethnique et par sexe. Dans la mesure où la question de l’utilisation d’un coefficient de variation pondéré par la population pour mesurer les inégalités spatiales fait débat, nous présentons à la fois les estimations du coefficient de variation pondéré par la population et non pondéré par la population.[23]

22 Nous estimons un revenu moyen pondéré (Ym_bar) et la part ($n_m$) de la population correspondant à chaque groupe ethnique par année de la manière suivante :
*bys year race: asgen Ym_bar = pcminc, w(wgt)*
*bysort year race: egen pop_race=count(pid)*

On peut alors obtenir la part de la population correspondant à chaque groupe ethnique :
*gen Nm=pop_race/27105 if year==2015 replace*
*Nm=pop_race/18480 if year==2008*

La moyenne générale pour chaque année peut se calculer comme suit :
*bys year: asgen Y_bar= pcminc, w(wgt)*

Ne conservez que les observations au niveau de l’année et du groupe ethnique en créant un identifiant de groupe ethnique par année :
*egen raceyer=group (race year)*
*sort raceyer*
*drop if raceyer==raceyer[_n-1]*

23 Lors du calcul du CV non pondéré par la population, la variable gap_square en cas de classement par groupe ethnique se calcule à l’aide de la formule ci-dessous :
*bys year:gen gap_square = (Ym_Y_bar)^2/4*





| | GCOV groupe ethnique | | GCOV sexe | |
|---|---|---|---|---|
| année | non pondéré par la pop | pondéré par la pop | non pondéré par la pop | pondéré par la pop |
| 2008 | 1,46 | 0,781 | 0,136 | 0,135 |
| 2015 | 1,54 | 0,670 | 0,167 | 0,167 |

Les estimations du coefficient de variation des groupes ethniques suggèrent des tendances divergentes. En effet, les estimations du CV pondéré par la population indiquent que les inégalités horizontales ont diminué au fil du temps, tandis que le CV non pondéré par la population montre une tendance légèrement croissante. En ce qui concerne le sexe, les inégalités horizontales (écart entre les sexes) ont tendance à augmenter avec le temps, et les estimations sont plus ou moins les mêmes que l’on utilise le coefficient de variation pondéré ou non pondéré.

# 5. STRUCTURE D’UN RAPPORT PAYS

Le lecteur des Rapports Pays, tel que nous l’envisageons, est un décideur politique ou un économiste intéressé et qualifié mais qui ne connaît pas en détail la situation du pays qui l’intéresse en matière d'inégalité. Vous trouverez ci-dessous la liste des sections incluses dans chaque rapport, accompagnée d’une explication simple du contenu de chacune d’entre elles. Vous trouverez ensuite une liste des éléments fondamentaux que tout rapport doit idéalement contenir.

Sections du Rapport Pays

1. Introduction et contexte :

1.1. Fournir un bref aperçu de la situation du pays.

1.2. Fournir suffisamment d’éléments de contexte pour permettre au lecteur d’interpréter de façon pertinente les résultats qui seront présentés.

1.3. Cela inclut notamment une analyse sommaire de la situation économique globale du pays.

1.4. Parmi les informations pertinentes, citons les données démographiques telles que la taille de la population et la croissance démographique, l’espérance de vie à la naissance, les pyramides des âges, les niveaux d’éducation et d’alphabétisation, ainsi que la répartition géographique de la population entre zones rurales et urbaines.

1.5. Les statistiques relatives au contexte macro-économique doivent inclure le PIB et sa croissance, le PIB par habitant, les principaux secteurs industriels et leur contribution au PIB, ainsi que les secteurs qui emploient la majorité de la main-d’œuvre.

1.6. Mettre en évidence le contexte en matière d’inégalités.

1.7. Fournir les chiffres clés dans le domaine des inégalités qui ont été largement utilisés jusqu'au moment de la rédaction du rapport.

1.8. Signaler les principaux rapports ou études sur la question des inégalités afin de bien situer le rapport d'analyse et d'en souligner l'importance.

1.9. Mentionner le projet global dans lequel s'inscrit le rapport d'analyse (ACEIR).

2. Analyse de l'espace politique :

2.1. Passer en revue les principales politiques conçues pour avoir un impact sur les inégalités.

2.2. Passer en revue les principaux plans et dispositifs sociaux et/ou économiques du pays (plans nationaux de développement, etc.).

2.3. Passer en revue les politiques/initiatives pertinentes ou efficaces (sans entrer dans les détails de l'évaluation des politiques).

3. Données :

3.1. Passer en revue les données qui vont être utilisées dans le rapport, en insistant sur la diversité des sources de données, essentielle pour l'analyse d'une question transversale telle que les inégalités.

3.2. Inclure des informations sur le cadre d'échantillonnage, la représentativité des données, la taille de l'échantillon, les dates de collecte des données et l'organisation des enquêtes.

4. Définir le profil des inégalités, les analyser et les cartographier :

**Présenter chaque série de résultats sous forme de graphique ou de tableau.**

**Proposer un argumentaire cohérent qui explique la façon dont sont interprétés les résultats.**

4.1. Inégalités de consommation et/ou de revenus :

- Gini, Lorenz, Theil
- Tendances en matière d'inégalités
- Décomposition des inégalités par source de revenus
- Décomposition des inégalités par groupe de population

4.2. Marché du travail :

- Inégalités salariales
- Distribution des revenus du travail
- Accès au marché du travail
- Dynamiques, vulnérabilité & secteur informel

4.3. Inégalités de richesse :

- Indice d'équipement des ménages
- Terres
- Rendement des actifs financiers
- Indice de richesse
- 1 %, 0,1 %, 0,01 % les plus riches

4.4. Enjeux sociaux :

- Éducation (distance à parcourir pour aller à l'école, taux net de scolarisation, nombre d'années de scolarisation)
- Santé (distance par rapport aux établissements de santé, mesures anthropométriques, espérance de vie)
- Internet
- Transports
- Eau
- Electricité
- Assainissement
- Enlèvement des déchets
- Logement

4.5. Inégalités spatiales :

- Cartographie et dérivation des indices de privation multidimensionnelle (MDDI), des inégalités de revenus, des coefficients de variation et de toute autre mesure pertinente pour chaque pays

4.6. Perceptions/mesures subjectives des inégalités

4.7. Mobilité sociale :

- Analyse et vulnérabilité de la classe moyenne
- Dynamiques au sein de la distribution

5. Recommandations :

5.1. Points essentiels de la section 4.

5.2. Préciser les priorités en termes de groupes et de régions/unités géographiques.

6. Perspectives futures :

6.1. Difficultés liées aux données/techniques existantes

6.2. Hiérarchisation des besoins en matière de données

6.3. Harmonisation des mesures et des calculs de l'inégalité à l'échelle de la région et du continent

Nous partons du principe que chaque pays a accès à des données d'enquêtes sur les ménages représentatives au niveau national, et idéalement au niveau individuel. Il se peut toutefois que le type de données disponibles diffère d'un pays à l'autre et que certains types d'analyses ne soient pas réalisables en raison des contraintes liées aux données. Dans ce cas, il conviendra de préciser que les analyses nécessaires n'ont pas pu être réalisées du fait de l'insuffisance des données.

# 6. CONCLUSION

Dans ce Manuel, nous avons abordé quelques-unes des questions théoriques auxquelles les chercheurs sont susceptibles d'être confrontés lorsqu'ils se lancent dans une étude sur les inégalités. Après avoir défini la portée et la méthodologie de l'étude, on peut s'atteler à mettre en œuvre les analyses pertinentes et à en interpréter les résultats. En ce qui nous concerne, nous nous intéressons au niveau contemporain des inégalités de revenus ou de consommation et à leur évolution récente, au niveau individuel, dans chacun des pays participants.

La majeure partie du Manuel est consacrée aux exigences en matière de données, aux problèmes liés aux données et aux modalités de mise en œuvre des différents estimateurs.

Il s'agit notamment de l'estimation et de l'interprétation des mesures d'inégalité les plus courantes, comme le coefficient de Gini, le ratio de Palma, le rapport entre les différents centiles de la distribution des revenus, les coefficients de Theil et les coefficients d'Atkinson.

Enfin, la contribution finale de ce Manuel consiste à fournir la structure de base des différents éléments qui doivent figurer dans un rapport pays, y compris une sélection d'indicateurs essentiels que tout rapport de ce type doit idéalement inclure. De manière globale, une telle démarche est censée garantir la qualité scientifique de chaque rapport tout en assurant la comparabilité des rapports des différents pays.





# 7. REFERENCES


Alkire, S., & Foster, J. (2011). Counting and multidimensional poverty measurement. Journal of Public Economics, 95(7), 476-487.

Araar, A., & Duclos, J. Y. (2013). User manual DASP version 2.3. *Distributive Analysis Stata Package, PEP, CIRPÉE and World Bank, Université Laval, June.*

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of economic theory*, 2(3), 244-263.

Banerjee, A. K. (2010). A multidimensional Gini index. *Mathematical Social Sciences*, 60(2), 87-93.

Cobham, A., Schlögl, L., & Sumner, A. (2016). Inequality and the tails: the Palma Proposition and Ratio. *Global Policy*, 7(1), 25-36.

Cowell, F. A. (1985). Measures of distributional change: An axiomatic approach. *The Review of Economic Studies*, 52(1), 135-151.

Cowell, F. A., & Flachaire, E. (2007). Income distribution and inequality measurement: The problem of extreme values. *Journal of Econometrics,* 141(2), 1044-1072.

Doyle, M. W., & Stiglitz, J. E. (2014). Eliminating extreme inequality: A sustainable development goal, 2015–2030. *Ethics & International Affairs*, 28(1),
5-13.

Ezcurra, R., & Rodríguez-Pose, A. (2017). Does ethnic segregation matter for spatial inequality?. *Journal of Economic Geography*, 17(6), 1149-1178.

Fields, G. S. (2003). Accounting for income inequality and its change: A new method, with application to the distribution of earnings in the United States. In Worker well-being and public policy (pp. 1-38). Emerald Group Publishing Limited.

Filmer, D., & Pritchett, L. H. (2001). Estimating wealth effects without expenditure data—or tears: an application to educational enrollments in states of India. Demography, 38(1), 115-132.

Foster, J., Seth, S., Lokshin, M., & Sajaia, Z. (2013). *A unified approach to measuring poverty and inequality: Theory and practice*. The World Bank.

Gluschenko, K. (2018). Measuring regional inequality: to weight or not to weight?. *Spatial Economic Analysis*, 13(1), 36-59.

Jenkins, S. P., & Jantti, M. (2005). *Methods for summarizing and comparing wealth distributions* (No. 2005-05). ISER working paper series.

Leibbrandt, M., Woolard, I., & de Villiers, L. (2009). Methodology: Report on NIDS wave 1. *Technical paper*, 1.

Lessmann, C. (2014). Spatial inequality and development—is there an inverted-U relationship?. *Journal of Development Economics*, 106, 35-51.

McKenzie, D. J. (2005). Measuring inequality with asset indicators. J*ournal of Population Economics*, 18(2), 229-260.

Shorrocks, A. F. (1982). Inequality decomposition by factor components. *Econometrica: Journal of the Econometric Society*, 193-211.

Shorrocks, A. F. (2013). Decomposition procedures for distributional analysis: a unified framework based on the Shapley value. The Journal of Economic Inequality, 11(1), 99-126.

Sen, A. (1973). “On Economic Inequality”. New York: Norton.

Sen, A. (1992), Inequality Reexamined, Oxford University Press, Oxford.

Stewart, F., Brown, G., & Mancini, L. (2010). *Monitoring and measuring horizontal inequalities. Centre for Research on Inequality, Human Security and Ethnicity, CRISE Working Paper (4)*.

Williamson, J. G. (1965). Regional inequality and the process of national development: a description of the patterns. *Economic development and cultural change*, 13(4, Part 2), 1-84.

Wittenberg, M. (2017). Measurement of earnings: Comparing South African tax and survey data.

Wittenberg, M., & Leibbrandt, M. (2017). Measuring inequality by asset indices: A general approach with application to South Africa. *Review of Income and Wealth.*

Wittenberg, M. (2017). Wages and Wage Inequality in South Africa 1994–2011: Part 1–Wage Measurement and Trends. *South African Journal of Economics, 85(2)*, 279-297.





# 8. ANNEXES

## A. Instructions pour le téléchargement et l’installation du package DASP.

Pour télécharger le package DASP, consultez le site suivant :

*http://dasp.ecn.ulaval.ca/downloadhelp.htm*

**Suivez les étapes ci-dessous pour installer les modules DASP sur Stata**

1. Dézippez le fichier dasp.zip (ceci est un nom de fichier) dans le répertoire c:/

   Assurez-vous de bien avoir **c:/dasp/dasp.pkg** ou **c:/dasp/ stata.toc**

1. Dans la fenêtre de commande Stata, tapez la syntaxe suivante :

   net from c:/dasp

2. Puis tapez la syntaxe :

   net install dasp_p1.pkg, force replace
   net install dasp_p2.pkg, force replace
   net install dasp_p3.pkg, force replace
   net install dasp_p4.pkg, force replace

3. Créez un dossier intitulé « personal » dans le dossier ado (qui doit se trouver sur le disque c:/), s’il n’existe pas déjà. Copiez graph_header.idlg et profile.do dans ce dossier.

4. Fermez Stata, puis rouvrez le logiciel. Vérifiez que le menu DASP apparaît bien dans l’onglet « User » de la barre de menus supérieure.

## B. Fichier ado pour l'estimation des indices d'équipement des ménages avec l'approche ACP NC (de Martin Wittenberg)

Vous pouvez copier le fichier ado ci-dessous dans votre fichier do et l'exécuter. Utilisez ensuite la commande *ucpc* pour estimer les indices d'équipement avec l'approche ACP NC. Merci de bien vouloir indiquer que ce fichier ado a été écrit par Martin Wittenberg.

```
*! version 1.0.0 18dec2013
program ucpc, rclass
	syntax varlist(numeric min=2) [aw fw] [if] [in] [, GENerate(name)]

	if "`weight'" != "" {
		local wght `"[`weight'`exp']"'
	}
	if "`generate'" !="" {
		confirm new var `generate'
	}

// clean up varlist


	marksample touse
	quietly count if `touse'
	if (r(N) == 0) error 2000
	if (r(N) == 1) error 2001

	// local varlist : list uniq varlist
	foreach v of local varlist {
		quietly summ `v' if `touse', meanonly
			if r(mean) !=0 {
				local vlist `vlist' `v'
			}
			else {
				dis as txt "(`v' dropped because of zero mean)"

			}
	}

	if "`vlist'" == "" {
		dis as err "all variables dropped because of zero mean"
		exit 498
	}

	local varlist `vlist'
	local nvar : list sizeof varlist
	if `nvar' < 2 {
		error 102
	}
	foreach X of varlist `varlist'{
		tempvar temp`X'
		qui summ `X', meanonly
		gen `temp`X''=`X'/r(mean)
		local tempvlist `tempvlist' `temp`X''
		}

// create matrix to be analyzed
```

```
tempname C nobs Ev L Score
quietly matrix accum `C' = `tempvlist' if `touse' `wght' ,
nocons matrix colnames `C' = `varlist'
matrix rownames `C' = `varlist'

local nvar = colsof(`C')
quietly matrix symeigen `L' `Ev' = `C'
matrix `Score'=`L'[1...,1]
matrix colnames `Score' = scores
matlist `Score'
if "`generate'"!=""{

    tempvar index
    qui gen double `index' = 0
    forvalues i =1/`nvar' {
        gettoken v tempvlist: tempvlist
        quietly replace `index' = `index' + `L'[`i',1]*`v' if
        `touse'
    }
    qui gen `generate' =`index'
}
End


Exit
```



# C. Créer des cartes avec STATA

Pour générer des cartes, il faut d'abord obtenir les fichiers de forme (« shapefiles ») des unités géographiques concernées. Par exemple, dans notre cas (Figure 6 ci-dessus), nous nous sommes procurés les fichiers de forme des provinces en 2011. Ensuite, pour pouvoir créer des cartes à l'aide de STATA, il faut installer deux modules complémentaires, *spmap* et *shp2dta*, en utilisant les commandes STATA ssc install comme suit :

*ssc install spmap*

*ssc install shp2dta*

Où *spmap* est la commande graphique qui permet de transformer les données brutes en fichier de sortie STATA classique au format .gph.
Et *shp2dta* est la commande qui convertit les données spatiales (stockées dans un fichier .shp) en fichier STATA au format .dta, qui peut ensuite être utilisé par la commande *spmap*.

Pour utiliser les fichiers de forme des provinces (intitulés « PR_SA_2011.shp » dans notre cas), on utilise la commande *shp2dta* comme suit :

shp2dta using PR_SA_2011.shp, database(Province) coordinates(PRcoord) genid(PRid) genc(_c)

La commande ci-dessus génère deux ensembles de données : la base de données et les coordonnées.

Le fichier de base de données contient une variable avec les unités géographiques principales, dans notre cas les provinces. Nous avons nommé cette base de données « Province » dans la commande *shp2dta* ci-dessus.

Le fichier de coordonnées contient les coordonnées correspondant aux limites des unités spatiales concernées (c'est-à-dire les provinces). Nous avons nommé ce fichier « PRcoord ».

L'option genid(PRid) génère une variable intitulée PRid dans le fichier Province, qui attribue des numéros uniques à chaque unité géographique (ou Province). Enfin, genc(_c) crée des coordonnées x_y dans le fichier Province. Voyez ce que contient chaque ensemble de données. Assurez-vous que les noms de province dans l'ensemble de données Province correspondent bien aux données qui contiennent nos estimations du CV (dans notre cas, stockées dans le fichier CV_province) et que le même numéro d'identification, ou PRid, leur est attribué. On fusionne ensuite l'ensemble de données CV_province avec l'ensemble de données Province comme suit :

Use CV_province, clear

merge 1:1 PRid using Province

drop _merge

gen CV_standP=CV_stand*100

format CV_standP %4.0f

Puis on utilise la commande *spmap* de la manière suivante pour représenter les estimations du coefficient de variation sur une carte :

*spmap CV_standP using PRcoord , id(PRid) fcolor(BuRd) ocolor(balck ..) osize(thin ..) ///*

*legend(position(11)) legtitle("CV %") clmethod(eqint) name(CV,replace) ///*

*label(data(Province) xcoord(x_c) ycoord(y_c) label(PR_NAME) color(white black))*

Nous vous invitons à consulter le fichier d'aide de la commande *spmap* pour savoir quelles sont les différentes options spécifiées dans cette commande.