



Développement d'indices composites et politiques publiques : interactions, portée et limites méthodologiques

DIALGA Issaka et LE Thi-Hang-Giang

LEMNA, Laboratoire d'Economie et de Management de Nantes-Atlantique

Section Economie

Mots-clés : Indices Composites, Méthodologie, Développement Durable, Politiques publiques

Contacts: DIALGA Issaka, issaka.dialga@univ-nantes.fr

LE thi-hang-Giang, thi-hang-giang.le@univ-nantes.fr

Chemin de la censive du Tertre, 44322 Nantes cedex 3, France

Tel : +336 69 32 87 67 ; +337 58 20 84 36

Résumé

Les Indices Composites (IC) sont indispensables pour les décideurs publics. En amont, les IC peuvent servir à orienter les politiques publiques. En aval, les mêmes outils sont incontournables pour évaluer les performances de ces politiques qu'elles soient à l'échelle nationale, régionale ou locale. Cependant, ils sont sujets à débats tant il n'existe pas de règles standard pour leur construction. Cet article discute des aspects méthodologiques dans l'élaboration des IC. IL illustre les aspects les plus problématiques des IC en construisant un indicateur simple de Développement Durable (IDD) pour 15 pays choisis de façon aléatoire. En suivant les étapes essentielles (choix des variables, normalisation, pondération, agrégation et calcul des incertitudes), les résultats obtenus par la moyenne arithmétique montrent que dans l'ensemble, les pays à revenus élevés, exception faite des USA, ont un IDD supérieur 0,7 et se classent aux premières places. Les pays émergents sont les moins soutenables de l'échantillon, les PED étant dans une position intermédiaire. La méthode géométrique fournit des résultats plus nuancés. Avec cette méthode, ce sont les pays émergents qui deviennent leaders au classement et les pays riches, excepté l'Australie, ont vu leur indice se détériorer très fortement faisant d'eux les derniers. Cependant, une particularité apparaît. Quelque soit la méthode de pondération utilisée, la France a un classement stable ; deuxième dans la méthode arithmétique, elle est classée 9^{ème} par la méthode géométrique et ce, quelque soit la méthode de pondération utilisée. L'analyse de sensibilité indique que les scores et les classements des pays sont sensibles aux méthodes utilisées. Elle confirme la particularité de la méthode BOD. Les variations des scores entre celle-ci et les autres méthodes sont très énormes atteignant pour certains pays 150%. Le calcul de l'incertitude associée aux méthodes de pondération indique un niveau d'incertitude variant entre 1, 27% à 95,16% pour l'indice des PED et celui des pays émergents contre 23% en moyenne pour celui des pays développés.

1. Introduction

La nécessité de trouver des mesures alternatives au PIB a favorisé la florescence de toute une batterie d'indicateurs composites (IC) au début des années 90. Le PNUD a ouvert les vannes de ce nouveau champ exploratoire avec son Indice de Développement Humain (IDH) en 1990. Au regard des initiatives engagées dans le domaine de la construction des IC, il est fort à parier que cette tendance, déjà perceptible, va s'accroître plus encore, la demande sociale étant plutôt en faveur de ces mesures multidimensionnelles. Aujourd'hui, il est indéniable de constater que les actions en termes de politiques publiques sont largement tributaires de ces outils synthétiques et ce, à au moins deux niveaux. En amont, les IC interviennent comme phare pour éclairer le chemin que souhaite emprunter le décideur public au regard des aspirations sociales. En aval, les mêmes outils sont incontournables pour évaluer les performances de ces politiques qu'elles soient à l'échelle nationale, locale, régional et même pour comparer des ensembles territoriaux (voir l'IDH régional et l'Indice régional de Pauvreté Humaine – IPH – de la région Nord-pas de Calais) ou qu'elles soient dans une déclinaison sectorielle (politique de l'habitat, du transport, du logement, de l'énergie, etc.).

Cependant, ces indices synthétiques n'en demeurent pas moins sujets à de vives critiques (voir Saisana M. et Saltelli A., 2010 ; Klugman J. et al. 2011 ; Chiappini R., 2012). La plupart des griefs formulés portent sur des aspects méthodologiques qui ont conduit à leur construction. Ainsi, le choix de certaines variables et leur pondération comme composantes de l'IC relèveraient de la pure subjectivité sans aucune évidence empirique ni aucun fondement théorique défendable (Conseil de l'Europe, 2005 ; Chiappini R., 2012). La persistance de ces critiques tend à remettre en cause la légitimité locale¹ de ces d'indicateurs comme outils d'orientation des actions publiques et de leur suivi. Ce sentiment de rupture de la confiance à l'égard des chiffres est d'autant plus probable que le foisonnement de ces indicateurs complique parfois les choix des utilisateurs : quels indicateurs choisir, pour quelles actions, et à quelles fins ?

Or, comme ces IC sont au centre de l'action publique, en l'orientant d'abord, décide de ce qui convient de faire, puis l'évalue, l'appropriation de leur méthode de construction par un large public devient un enjeu démocratique et sans laquelle appropriation, ces outils synthétiques manqueraient de visibilité et les actions publiques, de légitimation de la part des mandataires

¹ Alors que l'IPH classe le Nord-pas de Calais comme l'une des régions les plus pauvres de la France, l'IDH régional le met en bonne position devant certaines régions « développées » de la France. Voir « Programme "indicateurs 21" région Nord-pas de Calais, sept.2010 ».

des élus qu'ils soient nationaux ou locaux. Pourtant, ces indicateurs composites restent sans doute l'un des meilleurs supports de communication et d'actions pragmatiques. Le rôle du scientifique dans ce cas de figure, en tant qu'éclaireur des pouvoirs publics, est d'accompagner prioritairement ce développement en exprimant explicitement les réserves, et en maintenant une « veille » de recherche sur les évolutions des débats (DREES², 2011). Fort de ce constat de besoins réels exprimés, effectuer une analyse sans complaisance des développements récents des IC en vue de déceler leurs forces et d'atténuer leurs faiblesses en est une source de motivation. C'est l'objet de cet article qui se propose de faire une revue non technique des récents développements sur les IC.

Dans la section 2, nous passons en revue les étapes de construction des IC. Nous nous attardons sur les aspects les plus problématiques à savoir le choix des variables, la normalisation et les pondérations et agrégations en faisant ressortir les implications qu'ils suscitent en termes d'incertitude et donc du crédit que l'on pourrait accorder à ces IC. Nous illustrons chaque étape (seules les approches les plus utilisées ont été retenues) par un indicateur de développement durable (IDD) très simple que nous nous proposons de construire en suivant les étapes qui seront décrites au fur et à mesure de l'élaboration de l'indice. Nous entendons montrer que, quelle que soit la méthode utilisée, l'indicateur reste soumis à des incertitudes faisant varier les scores et les classements sous-jacents des pays. Et de ce fait, le choix d'une méthode dans la construction d'un IC devrait faire l'objet d'une justification théorique cohérente sans laquelle l'indice souffrirait d'un manque de légitimation. Dans la section 3 nous discutons des résultats (scores et rangs des pays) obtenus suivant les combinaisons des approches retenues et leurs implications en termes de DD. Nous analysons la sensibilité de l'indicateur et des scores par rapport aux changements de méthodes. La section 4 conclut en offrant l'opportunité de mettre en lumière les points saillants de cette revue sur les indicateurs composites illustrée par un indicateur construit à cet effet.

2. Etapes de construction théorique d'un IC : portée, limites et alternatives

On entend par indicateur composite, une combinaison mathématique de plusieurs indicateurs représentant différentes dimensions d'un même concept (JRC, 2010). A partir de cette définition, les IC n'ont pas à *a priori* d'unités de mesure – et c'est ce qui est souhaitable; les différents éléments composant l'IC n'ayant pas forcément d'unités de mesure

² Direction de la Recherche, des Etudes, de l'Evaluation et des Statistiques

significativement commune. Suivant la nature des sous-indicateurs qui le composent, un IC peut résulter d'un et/ou de trois types d'indicateurs : les indicateurs quantitatifs-objectifs, les indicateurs qualitatifs-objectifs, et les indicateurs qualitatifs-subjectifs. Les premiers sont définis comme des valeurs directement mesurables, par exemple le revenu par habitant, le taux de chômage, la quantité de CO₂ émise. Les deuxièmes ne sont pas directement mesurables mais font appel à des réponses objectivement vérifiables comme la présence ou l'absence d'une norme de qualité. Les troisièmes relèvent de l'opinion, de l'appréciation telle que la satisfaction, la confiance...

2.1. Définition de l'IC et des variables composantes (Etape 1&2)

L'étape de définition est très cruciale car un indicateur peut laisser la marge à certaines ambiguïtés et provoquer des interprétations équivoques ou erronées. La définition d'un IC doit être en cohérence avec l'objectif et le phénomène que celui-ci désire représenter. Pour ce faire, les dimensions du phénomène doivent être définies par les variables les plus pertinentes ; elles-mêmes choisies suivant des critères qui peuvent être objectifs ou subjectifs (mais suivant tout de même une logique cohérente). Ces critères, pour être de qualité, doivent obéir à quatre exigences. Ils doivent être représentatifs par rapport à la question à laquelle ils se rapportent, riches en informations et univoques, permettre une interprétation normative claire et acceptée et enfin, ils doivent être non excessivement onéreux (Conseil de l'Europe, 2005). Ainsi, un indicateur de développement durable devrait par exemple inclure au moins les trois dimensions ; à savoir l'économie, le social et l'environnement (l'IDD obéit à cette définition), elles-mêmes déclinées en variables facilement cernables et interprétables. Dans notre cas illustratif, une seule variable est retenue dans chaque dimension : le Revenu National Brut par habitant en Parités de Pouvoir d'Achat (PPA US \$ constant 2005), l'indice de Gini (base 100) et la quantité de CO₂ émise par habitant (en tonne métrique/habitant). En outre, la déclinaison en sous-indicateurs dépend du degré de finesse de l'information que l'on souhaite fournir à travers l'IC tout en gardant à l'esprit que « trop d'informations tue l'information ».

Toutefois, comme souligné plus haut, la recherche d'un certain niveau de finesse dans l'information peut conduire à définir des concepts théoriquement incompatibles dans un même IC dont l'interprétation de celui-ci peut susciter des débats. En effet, le souci de représenter de façon exhaustive la richesse d'un pays – dans la dimension économique de l'IDD – peut conduire à définir dans un même indicateur des variables « stocks » – pour caractériser le patrimoine – et de variables « flux » comme la croissance de l'activité

économique du pays. Inversement, la complexité de certains phénomènes contraint parfois les constructeurs des IC à simplifier les domaines et les variables et à ne retenir ceux qu'ils jugent pertinents et représentatifs. Le capital humain représenté par les seuls taux de scolarisation et d'alphabétisation dans l'IDH illustre bien ces simplifications des réalités sociales. Une mauvaise définition de l'IC au départ aura de toute évidence des répercussions sur les autres étapes de la construction de l'IC notamment dans les étapes de l'analyse de la colinéarité, de « normalisation » et de « interprétation » de l'IC.

En somme, sans négliger pour autant les autres étapes, l'étape de définition de l'IC conditionne le succès de celui-ci car un cadre théorique mal construit conduit nécessairement à des résultats biaisés et difficilement interprétables et par voie de conséquence, à des politiques inadéquates. Cela ne sous-entend pas non plus qu'il faudrait partir des données disponibles et facilement accessibles pour élaborer un IC ; la définition des variables pertinentes devant orienter la mobilisation des données statistiques.

2.2. Sources et traitement des données manquantes (Étape 3)

La définition de l'IC par l'identification de ses sous-indicateurs et variables composantes doivent permettre de déterminer les types de données nécessaires à la construction de l'indicateur final. Conventionnellement, on distingue deux types de données : les données primaires et les données secondaires. Les données primaires sont récoltées directement via des enquêtes, des observations ou des expériences réalisées par les chercheurs pour une problématique précise. Les données secondaires sont disponibles avant l'étude et peuvent provenir des institutions de statistiques, des sources administratives ou des organismes de sondage.

Les données des institutions de statistiques : par le biais des enquêtes périodiques, ces institutions (l'INSEE pour la France, l'INSD pour le Burkina Faso, l'IWEPS pour la région Wallonne en Belgique par exemple) disposent parfois de certaines données spécifiques. Ces types de données ont l'avantage de refléter la réalité vécue. Cependant, le coût que cette méthode génère fait que les données d'enquêtes n'existent que de manière intermittente, suivant une périodicité de collecte définie par l'institution (tous les cinq ans, dix ans etc.). Ces données peuvent s'avérer non adaptées ou insuffisantes lorsqu'il s'agit de la construction d'un indicateur temporel (suivi de l'IC dans le temps) nécessitant des séries relativement longues. Aussi, parce que réalisées en général au niveau national, ces données d'enquêtes sont difficilement disponibles lorsque l'on étudie des phénomènes régionaux ou locaux.

Les sources administratives : les données issues de ces sources sont en général des compilations de diverses données quantitatives transmises périodiquement aux institutions de statistiques. Ces sources ont l'avantage de fournir des données fiables et parfois exhaustives car elles ne se limitent pas à un échantillon. Cependant, ces compilations sont pour la plupart des agrégats macro socioéconomiques. Elles ne permettent donc pas de répondre à des besoins spécifiques. Elles ne peuvent pas non plus servir lorsque l'échelle d'application de l'indicateur est beaucoup plus basse comme la région, la commune ou encore à l'échelle d'un village ou d'un quartier. En outre, dans nombre de pays, la transmission et l'accessibilité à ces données restent encore limitées pour plusieurs raisons : problèmes légaux, techniques, confidentialité, etc.

Les organismes de sondage : ils fournissent des données qualitatives et subjectives à partir des enquêtes d'opinion. Bien qu'elles permettent d'obtenir des jugements de valeur sur le phénomène sur lequel les individus sont interrogés, les données d'enquête d'opinion ont une étendue beaucoup plus limitée que celles fournies par les deux premières sources et varient fortement dans le temps.

Dans la pratique, les données souhaitées ne sont toujours pas disponibles en totalité. Pour contourner cette difficulté, les chercheurs utilisent plusieurs astuces statistiques. Les données manquantes peuvent être de trois types suivant leurs liens avec la variable d'intérêt d'une part et des autres variables observées d'autre part. En effet, les variables manquantes peuvent être de type « *Missing Completely At Random* » c'est-à-dire que les données manquantes ne dépendent ni des variables d'intérêt (Y), ni des autres variables observées (X_i); ($X_i \perp (X_j; Y) \forall i \neq j$). Dans ce cas, il est possible d'éliminer purement et simplement la variable dont les données manquent (*case deletion*) sans pour autant produire un indicateur biaisé. C'est le cas par exemple de la suppression de certains pays du classement PNUD lorsque que les données manquent. Cependant, cette suppression réduit la qualité de l'information révélée par l'IC surtout lorsque la variable représente une dimension forte. Il n'est pas non plus possible d'effectuer une étude comparative entre l'échantillon de départ et celui réduit. Dans ce cas, une substitution de la variable pour laquelle les données ne sont pas disponibles peut être envisagée.

Les données manquantes peuvent dépendre directement des variables d'intérêts (*Non Missing At Random*). Formellement le *Non Missing At Random* peut se caractériser par $X = f(Y)$ avec $X_i \perp X_j \forall i \neq j$. Il peut également s'agir des données qui dépendent d'autres variables mais ne dépendent pas des variables d'intérêt (*Missing At Random*). $X_i =$

$f(X_j) \forall i \neq j$ et $X_i \perp Y$. Dans ces deux cas de figure, l'élimination de la variable entraîne une perte partielle d'information. L'IC obtenu ne diffusera qu'une fraction de l'information réelle produite par le phénomène traité. En vue d'obtenir une information plus complète sur le phénomène, il faudra estimer les valeurs manquantes à l'aide des outils statistiques (utilisation des indicateurs tendanciels comme les moyennes, les médianes ou les modes) ou économétriques (régressions linéaires).

Ces approximations permettent certes de surmonter une difficulté mais soulève un autre problème : celle de la fiabilité de l'IC à cause des incertitudes que ces méthodes d'imputation ou d'extrapolation peuvent produire. En effet, les valeurs obtenues aléatoirement par l'imputation supposent que celles-ci sont équivalentes aux données observées. Or, une valeur unique imputée ne peut pas représenter toute l'incertitude. Considérer ces données équivalentes sous-estime cette incertitude ; ce qui a tendance à réduire la variance de l'échantillon et l'intervalle de confiance de l'indicateur (Donzé, 2001). Dans le même ordre d'idées, Saisana et Saltelli (2010) montrent que la plus ou moins grande considération de l'incertitude dans les données collectées peut faire varier très significativement la valeur de l'indicateur final. La qualité de l'indicateur dépend donc fortement de la qualité des données utilisées et celles-ci en cas d'imputation, dépendent à leur tour de la robustesse des outils mobilisés.

Dans notre exemple, les données brutes proviennent de la base de données de la banque mondiale, le World Development Indicators (WDI). Pour les données brutes nous retenons pour chaque pays l'année la plus récente où les données sont disponibles pour les trois variables, Le RNB/hbt et les émissions CO2 sont toutes de l'année 2008. Les pays de l'échantillon appartiennent à trois groupes de niveaux de revenus différents : 5 pays à revenu élevé (Australie, Allemagne, France, USA & Qatar) ; 5 à revenu intermédiaire ou les émergents (Brésil, Russie, Inde, Chine, Bulgarie) et 5 à revenu faible (Algérie, Burkina Faso, Burundi, Cambodge, Vietnam).

2.3. Analyse multi-variée (Etape 4)

L'analyse multi-variée vise à analyser la structure générale des données afin de déceler les éventuels liens de corrélation entre les sous-indicateurs (dans le cas de l'IDD, il s'agit d'analyser les liens entre les variables retenues). L'avantage de cette analyse est qu'elle permet déjà à cette étape de l'élaboration de l'IC de relever les incohérences dans la formulation de l'indicateur et d'en apporter des corrections au besoin – la pondération inverse

des sous-indices corrélés par exemple peut être une solution à ces problèmes de corrélation des variables d'entrée. En effet, si l'analyse révèle une corrélation négative entre deux sous-indicateurs par exemple, ces deux ne pourront pas logiquement tous composer l'indicateur final en ce sens que leurs effets se neutraliseront et constitueront de ce fait un biais dans certaines fonctions d'agrégation notamment la moyenne arithmétique. Au pire des cas, il faudrait opérer des pondérations différentes si l'on juge que lesdits indicateurs représentent des critères importants et distincts. Dans la pratique, des variables peuvent être corrélées entre elles (cf. tableau 1) et la non prise en compte de l'endogénéité des variables peut conduire à des estimateurs biaisés. Lorsque des variables sont corrélées entre elles, la méthode Analyse en Composantes Principales (ACP) fournit des poids permettant de prendre en compte ces interactions entre les variables. Les poids sont déterminés suivant trois étapes.

A l'étape 1, on vérifie qu'il existe des corrélations entre les variables; sinon l'ACP ne peut être appliquée pour fournir des poids aux sous-indicateurs ;

A l'étape 2, on sélectionne les variables qui expliquent le plus, la variance de l'échantillon. L'ACP procède à une combinaison linéaire de toutes les variables entretenant des relations entre elles. Elle en dégage donc les principales composantes qui peuvent se résumer à un, deux, trois facteurs ou plus suivant les différentes combinaisons linéaires. Pour arriver à déterminer les principaux facteurs, trois conditions généralement sont requises:

- 1) la valeur propre associée à la variable à retenir doit être ≥ 1 ;
- 2) la contribution individuelle de la variable à la variance totale doit être au moins $\geq 10\%$;
- 3) le cumul par ordre décroissant des variances des variables doit être $\geq 60\%$.

L'étape 3 consiste à obtenir les poids à partir d'une matrice de rotation permettant d'obtenir des coefficients liés aux interactions entre les variables que l'OCDE (2008) appelle facteurs de charge. A partir des variables retenues au regard de l'étape 2, on calcule les poids en faisant le carré des variances liées à l'interaction entre les différentes variables (facteurs de charge) divisé par la variance respective de chaque variable retenue à l'étape 2.

La littérature (voir annexe 1), propose plusieurs outils permettant d'effectuer une analyse multi-variée comme l'Analyse en Composante Principale (ACP), l'Analyse Factorielle des Correspondances (AFC) ou encore l'analyse de partitionnement.

Tableau 1 : Matrice de corrélation entre les variables de l'indicateur IDD

	RNB/habitant	Indice de Gini	Emission de CO ₂ /hbt
RNB/habitant	1,0000		
Indice de Gini	0,0023	1.0000	
Emission de CO ₂ /hbt	0,9086	0,1110	1,0000

Source : Calculs des auteurs

Concernant l'étape 1, nous pouvons constater que nos trois variables sont positivement corrélées entre elles. Cependant les corrélations ne sont pas toujours très fortes. Le coefficient de corrélation entre le revenu national brut (RNB) par habitant et l'indice de Gini est de moins de 1%, le coefficient de corrélation entre l'émission de CO₂ par habitant et l'indice de Gini est un peu plus grand que 10%, alors que la corrélation entre l'émission de CO₂ et le RNB par habitant est beaucoup plus importante, plus de 90% (les pays industrialisés et émergents émettent beaucoup plus de CO₂ de par l'importance de leur production totale).

2.4. Normalisation de l'indicateur (Etape 5)

Cette étape vise à unifier les unités de mesure lorsque les données de l'ensemble des variables peuvent être ramenées à une mesure commune ou équivalente ou de présenter un IC neutre c'est-à-dire exempt de toute unité de mesure. Ce dernier cas est le plus souhaitable en ce qu'il fournit des IC dont les classements sous-jacents sont stables. Selon qu'il s'agisse d'un indicateur d'alerte (existence d'un seuil critique pour un phénomène donné) ou selon que l'indicateur est à vocation comparative (indicateurs internationaux), différentes méthodes existent et permettent d'obtenir des échelles de référence. On peut citer le « *ranking* », la méthode centrée réduite ou la normalisation à des z-scores, la méthode par la distance à une référence ou le Denominator-Based Weight (DBW), ... Dans cette illustration, deux approches les plus utilisées sont présentées: La normalisation par la méthode Min-Max et la normalisation par rapport à un benchmark (une référence).

La normalisation par la méthode Min-Max : dans la pratique, elle est de loin la méthode la plus utilisée notamment dans la normalisation des indicateurs à vocation internationale comme l'IDH. Cette méthode permet de centrer l'indice entre les valeurs extrêmes de l'échantillon. Algébriquement la méthode min-max s'écrit : $SI_{ij}^t = \frac{I^t - \min(i')(I^t)}{\max(I)(I^t) - \min(i')(I^t)}$ où $\min(i')(I^t)$ est le plus faible score réalisé par une des entités. L'entité i' peut être différente

de i c'est-à-dire que le plus faible score peut être détenu par une autre entité que celle (i) pour laquelle l'on normalise le sous-indicateur j . $\max(I)(I^t)$ est le plus gros score réalisé par l'une des entités participant à l'étude. I peut être évidemment différent de i et est forcément différent de l'entité i sauf dans le cas où toutes les entités sont à la fois meilleures et mauvaises.

Par définition, le sous-indice ainsi normalisé se situe entre 0 et 1 et les classements de toutes les entités sont faits en référence à des positions relatives de l'indicateur dans cette fourchette. Contrairement à la méthode centrée réduite, la méthode min-max est très sensible aux valeurs extrêmes.

La normalisation par rapport à une référence (benchmark scale-ratio) :

Cette méthode associe des scores aux performances réalisées dans un domaine en référence à un seuil choisi de façon plus ou moins arbitraire. Ce seuil peut être la performance du pays de référence à l'année initiale: $SI_{ij}^t = \frac{I_{ij}^t}{I_{i,j=j}^{t_0}}$. Cependant, deux autres approches sont également utilisées : le seuil peut être $I_{i,j=j}^t$, c'est-à-dire la performance du pays de référence à l'année courante ; ou il peut être $I_{i,j}^{t_0}$, c'est-à-dire la performance du pays considéré à l'année initiale.

Les indices normalisés de l'IDD sont résumés dans le tableau 2. Les clarifications importantes sont à relever dans la normalisation des sous-indices de l'IDD :

- Les indices dimensionnels correspondant aux dimensions « social » et « environnement » sont dits « indices d'alerte » c'est-à-dire que l'on améliore le score final (IDD) en baissant les valeurs des variables (indice de Gini et émission de CO₂ par habitant). Pour cela, la normalisation se fait de la manière suivante : $I = \frac{Valeur_{max} - Valeur_{pays}}{Valeur_{max} - Valeur_{min}}$
- En revanche, l'indice se rapportant à la dimension « économie » est un indicateur d'abondance ou de prospérité : l'économie du pays sera d'autant plus forte que son revenu national par tête lui-même est soutenu : c'est à dire que l'on souhaite que le RNB/habitant croisse. Pour cela la méthode de normalisation respecte la formule habituelle ; d'où

$$I = \frac{Valeur_{pays} - Valeur_{min}}{Valeur_{max} - Valeur_{min}}$$

Tableau 2: Normalisation des variables par la méthode min-max

Pays	RNB/hbt	Indice de Gini	Emission de CO2/hbt
Australie	0,480	0,740	0,622
Algérie	0,100	0,734	0,935
Allemagne	0,503	0,996	0,805
Brésil	0,130	0,000	0,959
Bulgarie	0,172	1,000	0,865
Burkina Faso	0,008	0,576	0,998
Burundi	0,000	0,811	1,000
Cambodge	0,019	0,641	0,994
Chine	0,077	0,468	0,892
France	0,463	0,831	0,881
Inde	0,031	0,807	0,971
Qatar	1,000	0,520	0,000
Russie	0,262	0,476	0,755
USA	0,635	0,531	0,634
Vietnam	0,030	0,725	0,970

Source : Calcul des auteurs

Dans le domaine « économie », le Qatar a la meilleure performance (1,00) alors que la pauvreté qui persiste au Burundi se traduit par un score nul pour ce pays (0,00). Nous notons aussi que la plupart des pays de l'échantillon ont un score inférieur à 0,50, même pour les économies industrialisées telles que l'Australie et la France. A part le Qatar, seuls les Etats Unis et l'Allemagne parviennent à obtenir un score supérieur à 0,50.

Pour l'aspect « social », il n'y a pas de « meilleure » ni de « pire » performance grâce auxquelles on peut évaluer celles des autres pays quand l'on se base sur le coefficient Gini (les politiques sociales diffèrent selon les pays). Néanmoins, nous pouvons constater que certains pays obtiennent un meilleur score que d'autres, et que sur cet échantillon, c'est la Bulgarie qui tend vers une répartition plus égalitaire des revenus que le reste, alors que le Brésil reste assez inégalitaire. Les pays dits développés, comme l'Allemagne et la France, deux grands piliers de l'Union Européenne, avec l'Australie, reçoivent de bonnes notes dans ce domaine, sans doute grâce à l'efficacité de leurs politiques de protection sociale.

Quant au pilier « environnement », souvent mis en avant dans les discussions en rapport avec le développement durable, il paraît logique que les économies en effervescence, le Brésil, la Chine et la Russie, obtiennent les plus faibles scores avec la méthode min-max, et que les pays à capacités de production les moins importantes, donc qui émettent peu de CO₂, ont les scores plus élevés (Burkina Faso, Burundi).

2.5. Pondération des sous-indicateurs et Agrégation de l'IC (Etape 6&7)

Ces deux étapes sont intimement liées et difficilement dissociables dans la pratique d'autant que la méthode de pondération choisie impose de façon implicite le plus souvent la méthode d'agrégation à retenir. Cependant, certaines méthodes permettent de distinguer explicitement ces deux étapes.

La pondération par l'allocation budgétaire ou Budget Allocation Process : cette méthode consiste à demander à chaque expert ou acteur de répartir un budget de X montant entre différents domaines d'un phénomène. La moyenne des points ainsi attribués permet de calculer les poids des indicateurs, et le score composite est le résultat de leur somme pondérée. Evidemment, bien que l'allocation optimale de ce budget confère à l'IC une légitimité professionnelle, parce qu'émanant des jugements d'experts du domaine, il reste tout de même que les choix dépendent fortement de la perception du phénomène par l'« expert », et donc de son importance relative, faisant en sorte que la méthode devient en elle-même une méthode fondée sur la subjectivité implicite. C'est pourquoi il est parfois indiqué de vérifier la logique du jugement de valeur de l'acteur (expert ou toute autre partie prenante) en calculant un indice de cohérence³ de jugement de valeur (Saaty T.L., 1980 ; Mendoza G. A. et al. 2000). Lorsque la valeur de cet indice est supérieure au seuil tolérable de 10%, alors il existe une incohérence dans le jugement de valeur, et donc dans la répartition budgétaire de l'acteur qu'il faille déceler puis apporter des corrections en vue d'assurer une certaine harmonie et logique dans les scores finaux.

La méthode est discutable dans la mesure où l'avis de l'expert peut être différent de celui du public cible ou de la réalité vécue sur le terrain. Dans de tel cas de figure, la confrontation des allocations du même budget par les différents acteurs est nécessaire et devrait permettre de dégager un compromis sur les scores à attribuer à chaque sous indicateur.

En outre, lorsque le phénomène traité comporte plusieurs dimensions ou domaines entre lesquels l'expert doit allouer le budget X, la méthode fournit des pondérations insensées –la prise en compte d'un plus grand nombre de variables dans la construction d'un IC ne conduit pas nécessairement à un indicateur de qualité et représentatif du phénomène traité –le nombre

³ $I = x \frac{w_j}{w_{j'}}$ avec x le rapport entre des montants du budget alloués aux sous indicateur SI_j et $SI_{j'}$; w_j et $w_{j'}$ étant les poids relatifs des sous indicateurs SI_j et $SI_{j'}$ obtenus à partir de l'allocation du Budget X.

raisonnable de sous-indicateurs devant être compris entre une dizaine et une douzaine (Nardo M. et al.2005).

Pour illustrer cette méthode, nous avons eu recours à dix « experts⁴ ». Chacun d’entre eux a réparti 100 points entre les trois dimensions du développement durable (Economie, Social et Environnement). Le tableau 3 résume les résultats de la pondération par cette méthode:

Tableau 3: résumé de la pondération par le BAP

	Economie	Social	Environnement
Min	0,200	0,200	0,200
Moyenne	0,345	0,325	0,330
Médiane	0,350	0,300	0,300
Mode	0,200	0,300	0,300
Max	0,500	0,600	0,500

Source : Calculs des auteurs

Ainsi, le poids minimal donné à chacune de ces trois dimensions est de 1/5, le poids maximal est autour de 1/2. Le poids moyen ainsi que le poids médian sont autour de 1/3, ce qui montre que les « experts » ont tendance à appliquer une répartition égalitaire des poids entre les trois dimensions. Nous remarquons aussi que la valeur modale du poids donnée à « social » et « environnement » est plus importante que celle attribuée à la dimension « économie », ce qui reflète en quelque sorte la volonté de mettre en avant l’importance des deux premières dimensions aux dépens de la troisième lorsqu’on parle de développement durable.

La pondération par la maximisation des scores : Elle est directement issue de la méthode BOD utilisée pour la pondération des sous-indicateurs. La méthode BOD elle-même est une application de l’approche DEA, *Data Envelopment Analysis*, qui consiste à construire, à partir des performances meilleures, une frontière d’efficacité ou courbe des possibilités puis à déterminer les performances relatives des autres individus participant à l’étude par rapport à ce cadre de référence (*benchmark*) établi. Ainsi, la méthode BOD permet d’attribuer un poids relatif à un individu *i* au regard des scores obtenus par ceux parmi le groupe qui performant le mieux. En associant le score 1 à la meilleure performance, les individus les moins performants auront logiquement un score inférieur à l’unité mais strictement positif selon la formule suivante : $w_i = \frac{\text{Performance de } i}{\text{Benchmark ou frontière des possibilités}} \leq 1$. Ainsi défini, le poids relatif de chaque individu *i* dépend de sa performance réalisée comparativement à la situation « idéale » qui est le benchmark c’est-à-dire le niveau de performance souhaitable.

⁴ Nous avons sollicité l’avis de 10 étudiants du Master 2 Evaluation Economique du Développement Durable de l’Institut d’Economie et de Management de Nantes (IEMN-IAE) de l’université de Nantes.

Par définition, l'agrégation par cette méthode maximise les scores obtenus par le BOD tout en respectant la contrainte de probabilité c'est-à-dire la contrainte de non négativité des scores ou poids des sous-indicateurs (i) et le fait que la somme des sous-indicateurs pondérés de leur poids respectif doit être inférieure ou à la limite égale à l'unité ($\sum_{j=1}^J I_j w_j \leq 1$). Pour une série de sous-indicateurs représentant diverses dimensions d'un phénomène donné, la méthode accorde le bénéfice du doute à l'individu dont on souhaite évaluer sa performance globale en ne retenant parmi les scores engrangés que ceux des dimensions où l'individu performe le mieux. L'idée qui sous-entend cette méthode est que chaque individu (vu sous l'angle d'entité) a des priorités et cherche donc à maximiser ses actions dans la ou les dimensions qu'il juge primordiales. Formellement, l'IC est donné par le programme de maximisation suivante : $IC_i = \underset{w_{ij}=1,\dots,J}{Max} \{ \sum_{j=1}^J w_{ij} * SI_{ij} \}$; i l'individu (i= 1,...,I) ; j le sous indicateur (SI) représentant le domaine ou la dimension j (j= 1,..., J) et w_{ij} le poids relatif associé au sous indicateur j dans l'IC de l'individu i.

s.c

$$\begin{cases} 0 < \sum_{j=1}^J w_{ij} * SI_{ij} \leq 1 & (1) \\ w_{ij} > 0 & (2) \end{cases}$$

Comme les autres méthodes, l'agrégation par la maximisation de la meilleure performance peut présenter des faiblesses lorsque les dimensions du phénomène traité ne sont pas substituables. Autrement dit, lorsque l'étude porte sur une thématique présentant des dimensions d'importance égale (cas du DD) ou lorsqu'il s'agit des domaines complémentaires, la méthode de maximisation qui ne retient que des scores des domaines dans lesquels l'individu consacre plus ses efforts, ne permet pas de traiter le phénomène dans sa globalité d'autant que les autres dimensions jugées non prioritaires par l'individu sont occultées de l'IC final. Il peut y avoir à partir de ce moment un risque de déséquilibre dans l'appréhension du phénomène. La deuxième limite est inhérente à la méthode de pondération. En effet, en contraignant les scores dans une fourchette de [0,1], la méthode ne permet pas d'observer des pires performances et les individus qui excellent au-delà de la frontière de l'efficacité prédéfinie sont confinés à demeurer dans cette fourchette préétablie. Des individus pourraient présenter des scores médiocres en deçà de la limite inférieure imposée. Or, ces scores pourraient être une alerte dans l'analyse de la performance globale de l'entité. En bornant le poids des sous-indicateurs à 1, la méthode exclut la possibilité d'existence d'autres

situations meilleures. Ce qui peut conduire les entités se situant sur cette frontière à ne pas améliorer leur position de leader puisqu'elles sont considérées comme la référence vers laquelle les autres jugées moins performantes devraient tendre. De plus, la méthode devient complexe voire inopérante, lorsqu'il s'agit de représenter cette frontière des possibilités à partir d'un grand nombre de sous-indicateurs. La même difficulté se pose lorsque parmi les entités participant à l'étude, il y a « plusieurs meilleures » en même temps pour une même dimension.

Tout compte fait, la méthode BOD présente deux vertus essentielles : en accordant le doute à l'entité, la méthode lui offre une opportunité d'être excellent dans au moins une des dimensions du phénomène généralement là où l'entité concentre plus ses efforts. La seconde vertu de la méthode est que les poids des sous-indicateurs sont déterminés de façon endogène par le modèle lui-même, épargnant l'IC final de toute critique liée à la subjectivité dans la pondération de ses composantes.

Afin de mieux illustrer la vue d'ensemble de ces méthodes de pondération, nous présentons dans un même tableau (cf. tableau 4) les résultats de pondération obtenus avec les approches BAP, ACP et BOD. Avant de commenter les résultats, relevons quelques points liés à la méthodologie :

* Les poids fournis par le Budget Allocation Process (BAP) est la moyenne des points (100) répartis par nos dix « experts » entre les trois dimensions du DD (Economie, Social & Environnement).

* L'Analyse en Composantes Principales (ACP) a permis d'obtenir les poids relatifs aux trois dimensions de notre IC en raison de la contribution relative de chaque variable aux variations de la variance totale de l'indice. Les poids ici sont obtenus à partir des valeurs propres fournies par l'ACP ; la somme des poids ainsi répartis est égale à un.

* Les poids de la dernière colonne du tableau de pondération ont été obtenus par la méthode du Benefit Of the Doubt (BOD). Pour les dimensions « social » et « environnement », la logique du raisonnement est maintenue. Le ou les meilleurs pays ont été identifiés pour chacune des deux dimensions. Est meilleur le pays enregistrant le faible niveau d'inégalité (plus petite valeur de GINI parmi le groupe de pays à l'étude) et/ou le faible niveau d'émission de CO2 par habitant comparativement aux autres pays (plus petite valeur de CO2/habitant parmi le groupe de pays à l'étude). Le poids relatif est obtenu en rapportant

la valeur de référence (meilleure pratique) à la valeur du pays. Plus un pays s'écarte de la bonne pratique, plus son indice dimensionnel est pondéré d'une faible valeur. Inversement, les pays de référence ont un poids égal à 1 et ceux tendant vers ces derniers ont des pondérations relativement élevées pour la dimension considérée. Pour l'indice d'abondance (contribution positive à l'indice IDD) «économie», la formule est inversée. Les meilleurs sont le ou les pays enregistrant de meilleures performances économiques. Les poids relatifs des autres pays sont déterminés à l'aune de ces meilleures performances en rapportant le RNB par habitant du pays considéré à la meilleure valeur de cette variable parmi les groupes de pays à l'étude.

Tableau 4: Pondérations des dimensions suivant les méthodes BAP, ACP & BOD

Pays	BAP			ACP			BOD		
	RNB/hbt	IGINI	CO2/hbt	RNB/hbt	IGINI	CO2/hbt	RNB/hbt	IGINI	CO2/hbt
Australie	0,345	0,325	0,330	0,329	0,344	0,327	0,484	0,801	0,001
Algérie	0,345	0,325	0,330	0,329	0,344	0,327	0,107	0,798	0,007
Allemagne	0,345	0,325	0,330	0,329	0,344	0,327	0,506	0,996	0,002
Brésil	0,345	0,325	0,330	0,329	0,344	0,327	0,137	0,512	0,011
Bulgarie	0,345	0,325	0,330	0,329	0,344	0,327	0,178	1,000	0,003
Burkina Faso	0,345	0,325	0,330	0,329	0,344	0,327	0,016	0,712	0,189
Burundi	0,345	0,325	0,330	0,329	0,344	0,327	0,008	0,847	1,000
Cambodge	0,345	0,325	0,330	0,329	0,344	0,327	0,026	0,745	0,068
Chine	0,345	0,325	0,330	0,329	0,344	0,327	0,084	0,664	0,004
France	0,345	0,325	0,330	0,329	0,344	0,327	0,467	0,861	0,004
Inde	0,345	0,325	0,330	0,329	0,344	0,327	0,038	0,845	0,015
Qatar	0,345	0,325	0,330	0,329	0,344	0,327	1,000	0,686	0,000
Russie	0,345	0,325	0,330	0,329	0,344	0,327	0,267	0,667	0,002
USA	0,345	0,325	0,330	0,329	0,344	0,327	0,638	0,691	0,001
Vietnam	0,345	0,325	0,33	0,329	0,344	0,327	0,037	0,793	0,015

Les résultats reportés à la colonne 1 du tableau 4 (obtenus par la méthode BAP) proviennent du tableau 3. Les poids obtenus avec cette méthode sont répartis de façon plus égale tout comme ceux obtenus avec l'ACP (dont la somme fait également 1). Avec cette deuxième méthode, le poids relatif à la dimension « économie » vaut 0,329, celui de la dimension « environnement » est de 0,327 et le poids à attribuer à la dimension « sociale » est de 0,344 un peu plus important que les deux premiers. Les résultats fournis par la méthode BOD sont à analyser par dimension de l'IDD. En effet dans chaque dimension, les pays qui ont les meilleurs scores sont respectivement le Qatar (leader dans la dimension « économie »), la Bulgarie s'illustre bien dans l'égalité sociale et le Burundi est le plus respectueux de l'environnement parmi les pays de l'échantillon (faible émission de CO₂/hbt). Si nous calculons la moyenne de poids de chaque dimension pour l'ensemble de l'échantillon, nous obtenons les résultats suivants (dans le même ordre): 0,266 ; 0,774 ; 0,088. Donc la plupart des pays ont une bonne performance dans la dimension sociale et la pondération qui y est associée leur est avantageuse aussi. Par contre, la dimension environnementale a souvent une faible pondération, puisque les pays obtiennent des résultats médiocres dans ce domaine par rapport au meilleur pays. Même pour le Qatar et le Burundi, leaders respectifs dans les dimensions « économie » et « environnement », les pondérations de la dimension « social » sont toutes supérieures à 0,60.

La pondération - agrégation par la moyenne des sous-indicateurs ou Indicators Average : il peut s'agir d'une moyenne arithmétique, géométrique, harmonique, etc. Nous retenons ici la moyenne arithmétique et géométrique. En effet, la moyenne arithmétique est la méthode de pondération-agrégation la plus utilisée dans la pratique probablement expliquée par sa simplicité à être comprise par un plus grand public et par sa transparence. Cependant, la pondération égalitaire bien qu'en apparence « neutre » (même importance aux différentes dimensions du phénomène traité), peut être à l'origine de certaines discriminations. Autrement dit, c'est une méthode très sensible aux valeurs extrêmes qui peut fournir des résultats biaisés lorsque que les données comportent des valeurs aberrantes (*outliers*). Par ailleurs, cette méthode repose sur une hypothèse implicite qu'il existe une forte compensation (substituabilité parfaite) entre les différentes dimensions du phénomène d'une part et que d'autre part, les différentes dimensions sont d'importance égale. On risquerait ainsi de surestimer la situation ou les dimensions pour lesquelles les valeurs sont relativement faibles. La moyenne géométrique quant à elle tient compte de la non compensation parfaite des sous-indicateurs et récompense les entités (pays, régions, communes, etc.) qui performant dans tous

les domaines. Nous pouvons mieux voir comment la méthode de pondération-agrégation influence les résultats de calcul de l'indicateur composite et ainsi que le classement des pays en se reportant aux tableaux 5 et 6.

Tableau 5 : Scores et rangs pays selon différentes méthodes de pondération & l'agrégation par la moyenne Arithmétique

	EGAL		BAP		ACP		BOD	
	score	Rang	score	Rang	Score	rang	score	rang
Australie	0,614	4	0,611	4	0,616	4	0,826	6
Algérie	0,590	8	0,582	8	0,591	8	0,603	9
Allemagne	0,768	1	0,763	1	0,771	1	1,000	1
Brésil	0,363	15	0,361	15	0,357	15	0,028	15
Bulgarie	0,679	3	0,670	3	0,683	3	1,000	1
Burkina Faso	0,527	11	0,519	11	0,527	11	0,599	10
Burundi	0,604	5	0,594	6	0,606	5	1,000	1
Cambodge	0,551	10	0,543	10	0,551	10	0,545	12
Chine	0,479	14	0,473	14	0,478	14	0,321	14
France	0,725	2	0,720	2	0,726	2	0,935	5
Inde	0,603	6	0,593	7	0,605	6	0,698	8
Qatar	0,507	12	0,514	12	0,508	12	1,000	1
Russie	0,498	13	0,494	13	0,497	13	0,389	13
USA	0,600	7	0,601	5	0,599	7	0,772	7
Vietnam	0,575	9	0,566	9	0,576	9	0,591	11

Source : calcul des auteurs

Avec l'agrégation par la moyenne arithmétique, il n'est pas difficile de remarquer que les résultats ne varient pas beaucoup en fonction de la méthode de pondération. L'Allemagne occupe toujours la première place, que ce soit la méthode de pondération égale, BAP, ACP ou BOD. La France, la Bulgarie et l'Australie ont toujours de bons scores alors que la Russie, la Chine et le Brésil arrivent respectivement 13^{ème}, 14^{ème} et 15^{ème} quelle que soit la méthode de pondération utilisée. Cependant, pour le Burundi et le Qatar la situation est différente. Les leaders respectifs dans les domaines « environnement » et « économie » obtiennent les meilleurs scores dans l'échantillon avec le BOD, mais ont des résultats médiocres avec les autres méthodes de pondération.

Tableau 6: Scores et rangs pays selon différentes méthodes de pondération & l'agrégation par la moyenne Géométrique

	EGAL		BAP		ACP		BOD	
	Score	Rang	score	Rang	Score	rang	score	rang
Australie	0,604	3	0,602	3	0,606	3	0,550	9
Algérie	0,410	7	0,400	7	0,412	7	0,611	7
Allemagne	0,739	1	0,733	1	0,742	1	0,702	3
Brésil	0,000	13	0,000	13	0,000	13	0,000	13
Bulgarie	0,530	5	0,519	5	0,534	5	0,730	1
Burkina Faso	0,169	12	0,161	12	0,171	12	0,625	6
Burundi	0,000	13	0,000	13	0,000	13	0,000	13
Cambodge	0,230	11	0,220	11	0,232	11	0,646	5
Chine	0,318	8	0,311	8	0,319	8	0,487	10
France	0,697	2	0,692	2	0,698	2	0,595	8
Inde	0,289	9	0,278	9	0,293	9	0,730	2
Qatar	0,000	15	0,000	15	0,000	15	0,000	15
Russie	0,455	6	0,451	6	0,455	6	0,426	12
USA	0,598	4	0,599	4	0,597	4	0,483	11
Vietnam	0,275	10	0,265	10	0,278	10	0,681	4

Source : Calculs des auteurs

Avec l'agrégation par la moyenne géométrique, l'écart entre les résultats obtenus avec le BOD et les autres méthodes de pondération est plus visible. L'Allemagne n'arrive plus qu'à la troisième place avec le BOD alors qu'elle est première avec la pondération égale, le BAP et l'ACP. Nous observons le même décalage pour une grande partie des pays de l'échantillon, le BOD transforme pratiquement les rangs, certains pays qui sont mal classés avec les trois premières méthodes arrivent en haut du classement avec le dernier et inversement. Parmi les exceptions, une est particulièrement intéressante : le Qatar, qui maintient sa dernière place avec le BOD. Nous discuterons le cas du Qatar plus en détails dans la partie suivante.

3. Discussion des résultats et analyse de sensibilité :

Dans cette section nous discutons des résultats (rangs et scores des pays) obtenus suivant les combinaisons des approches retenues et leurs implications en termes de développement durable. Nous analysons également la sensibilité de l'indicateur et des scores par rapport aux changements de méthodes.

3.1. Discussion de l'Indice de développement durable

Pour rappel, les trois dimensions du développement durable - Economie, Social, Environnement – sont mesurées via les trois variables correspondantes respectives – Revenu National Brut par habitant en Parités de Pouvoir d'Achat (PPA US \$ constant 2005) ; Indice de GINI (base 100) ; Emissions CO2 par habitant (en tonne métrique).

L'agrégation par la moyenne arithmétique donne des résultats similaires entre les trois premières méthodes de pondération utilisées : pondération égale, BAP et ACP. Les scores obtenus par les pays ne varient pas beaucoup d'une méthode à une autre et le classement est pratiquement le même à deux exceptions près (le Burundi et les Etats Unis). L'Allemagne est toujours en tête du classement alors que les nouvelles puissances économiques – la Russie, la Chine et le Brésil occupent les dernières places, sans doute à cause du déséquilibre entre les efforts accordés à l'aspect économique et ceux déployés pour améliorer la situation sociale et environnementale. La dernière méthode de pondération, le BOD, induit quelques résultats différents sans véritablement bouleverser le classement, à l'exception du Qatar et du Burundi, qui sont récompensés par le BOD pour leurs performances respectives en économie et environnement. Ces chiffres illustrent clairement le risque de déséquilibre dans l'appréhension du phénomène quand la méthode de maximisation favorise un ou des aspects où le pays a plus d'avantages, alors que dans le développement durable les aspects sont d'importance égale. Nous pouvons également voir qu'avec le BOD, 4 pays se partagent le meilleur score de 1,000, dû à la contrainte des scores dans une fourchette de [0 ;1], alors qu'on ne peut certainement pas dire qu'ils ont la même performance au total.

Avec l'agrégation par la moyenne géométrique, alors les trois premières méthodes de pondération donnent lieu encore une fois aux résultats identiques entre elles en termes de classement, le BOD change quasiment et complètement les résultats en mettant en bonne position les pays mal placés avec les autres méthodes de pondération et inversement. Cependant un pays arrive dernier dans tous les classements, quelle que soit la méthode de pondération. C'est le Qatar, qui est tout de même leader dans la dimension économique, et qui a été récompensé par conséquent avec une première place avec la méthode d'agrégation arithmétique, la pondération BOD. En effet, le pays est pénalisé par ses émissions trop importantes de CO₂ par habitant, (note 0 pour l'aspect environnement donc 0 pour la note totale). Nous reconnaissons ici la prise en compte de la non compensation des sous-indicateurs, à la différence de l'agrégation par la méthode arithmétique.

A l'aune de ces résultats, on est tenté de conclure que l'atteinte des objectifs du DD sont plus facilement envisageables dans les vieilles nations industrialisées que dans les pays en développement, sans besoin d'intégrer dimension intergénérationnelle et de bonne gouvernance, autres piliers essentiels du DD.

3.2. Analyse de sensibilité et d'incertitude:

A partir des résultats obtenus ci-dessus, nous avons conduit des tests de sensibilité de l'indice et des scores par rapport aux méthodes de pondération, en retenant l'agrégation par la moyenne arithmétique. Dans le tableau 7, nous présentons les écarts cumulatifs imputables aux méthodes de pondérations (cf. annexe 3 pour les développements théoriques).

Tableau 7: Sensibilité de l'indice et des scores par rapport aux méthodes de pondération (agrégation: arithmétique)

	EGAL-BAP		EGAL-ACP		EGAL-BOD		BAP-ACP		BAP-BOD		ACP-BOD	
	variation score (%)	variation rang	variation score (%)	variation rang	variation score (%)	variation rang	variation score (%)	variation rang	variation score (%)	variation rang	variation score (%)	variation rang
Australie	0,43	0	-0,29	0	-34,50	-2	-0,72	0	-35,08	-2	-34,11	-2
Algérie	1,37	0	-0,23	0	-2,29	-1	-1,62	0	-3,71	-1	-2,06	-1
Allemagne	0,67	0	-0,41	0	-30,25	0	-1,09	0	-31,12	0	-29,71	0
Brésil	0,46	0	1,76	0	92,17	0	1,30	0	92,14	0	92,03	0
Bulgarie	1,36	0	-0,62	0	-47,26	2	-2,00	0	-49,28	2	-46,35	2
Burkina Faso	1,52	0	0,05	0	-13,52	1	-1,50	0	-15,27	1	-13,57	1
Burundi	1,67	-1	-0,36	0	-65,65	4	-2,06	1	-68,47	5	-65,06	4
Cambodge	1,53	0	-0,07	0	1,08	-2	-1,62	0	-0,45	-2	1,15	-2
Chine	1,25	0	0,21	0	32,99	0	-1,05	0	32,15	0	32,86	0
France	0,62	0	-0,17	0	-28,95	-3	-0,79	0	-29,75	-3	-28,74	-3
Inde	1,59	-1	-0,36	0	-15,73	-2	-1,98	1	-17,60	-1	-15,31	-2
Qatar	-1,45	0	-0,24	0	-97,41	11	1,19	0	-94,59	11	-96,94	11
Russie	0,69	0	0,17	0	21,83	0	-0,53	0	21,29	0	21,70	0
USA	-0,15	2	0,18	0	-28,70	0	0,32	-2	-28,51	-2	-28,93	0
Vietnam	1,55	0	-0,24	0	-2,73	-2	-1,82	0	-4,35	-2	-2,49	-2

On peut aisément constater que la pondération égale, le BAP et l'ACP donnent des résultats proches en termes de classement, sinon identiques. Il n'y a pratiquement pas de différence dans les rangs obtenus par les pays dans l'échantillon avec ces trois méthodes. Les différences sont légèrement plus perceptibles lorsqu'on regarde les scores. En moyenne, la variation de score d'une méthode à une autre n'est que de moins de 1%. La différence entre ces trois méthodes avec le BOD est par contre beaucoup plus notable. La variation de score entre le

BOD et chacune des trois est autour de 15%, les scores obtenus avec le BOD étant souvent plus élevés qu'avec les trois autres.

Après les tests de sensibilité, nous avons également réalisé l'évaluation de l'incertitude associée à l'indice de chaque pays en retenant l'agrégation par la moyenne arithmétique. Ne disposant pas de spécifications fonctionnelles reliant les différentes combinaisons utilisées (normalisations, pondérations et agrégations), seule l'incertitude liée aux changements de pondérations a été évaluée ici. Les simulations de Monte Carlo pourraient être la solution alternative mais elles requièrent un échantillonnage beaucoup plus large ; condition non satisfaite dans notre cas (voir annexe 2). Nous utilisons alors la formule de calcul de l'incertitude relative suivante: $\Delta X/\bar{X} = t \cdot \frac{s}{\sqrt{n}}/\bar{X}$ où \bar{X} est l'indice composite IDD, $t = 3,18$ la statistique de Student à 2 degrés de liberté ($ddl = n - 1$), s l'écart type de l'échantillon (ici l'écart type de l'IDD obtenu avec les 4 méthodes de pondération) et $n = 4$ le nombre de variables de l'étude (dans l'IDD, $n = 4$ représente le nombre de méthodes de pondération utilisées). L'incertitude absolue ΔX mesure l'erreur maximale que l'on est susceptible de commettre dans l'évaluation de l'indice \bar{X} . L'incertitude relative $\Delta X/\bar{X}$ mesure quant à elle l'importance de l'erreur maximale susceptible d'être commise par rapport à la valeur calculée de l'indice à un certain degré de confiance (dans notre application, le degré de confiance est de 95%).

Tableau 8 : calcul de l'incertitude relative associée à l'indice IDD

	EGAL	BAP	ACP	BOD	
	score	score	score	score	incertitude(%)
Australie	0,614	0,611	0,616	0,826	25,30
Algérie	0,590	0,582	0,591	0,603	2,40
Allemagne	0,768	0,763	0,771	1,000	22,44
Brésil	0,363	0,361	0,357	0,028	95,16
Bulgarie	0,679	0,670	0,683	1,000	33,85
Burkina Faso	0,527	0,519	0,527	0,599	10,89
Burundi	0,604	0,594	0,606	1,000	45,28
Cambodge	0,551	0,543	0,551	0,545	1,27
Chine	0,479	0,473	0,478	0,321	28,30
France	0,725	0,720	0,726	0,935	21,60
Inde	0,603	0,593	0,605	0,698	12,45
Qatar	0,507	0,514	0,508	1,000	61,71
Russie	0,498	0,494	0,497	0,389	18,16
USA	0,600	0,601	0,599	0,772	21,30
Vietnam	0,575	0,566	0,576	0,591	2,81

En changeant de méthode de pondération, l'incertitude (et inversement la crédibilité) associée à l'indice construit dans notre cas illustratif va de 1,26% pour l'IDD du Cambodge à plus de 95% pour celui du Brésil. Ce qui induit nécessairement des incertitudes au niveau des classements des pays. Sans vouloir prétendre contrôler toutes les sources d'incertitude, les premiers résultats de l'indice laisse penser que la méthode BOD est à l'origine de ces fortes variations des résultats de l'indice des pays en particulier pour le Brésil et le Qatar où l'incertitude dépasse les 50%. Ces résultats de calcul d'incertitude illustre bien la prudence qu'il est à observer dans les conclusions issues des analyses des indices composites construits en adoptant telle ou telle méthode de pondération. Il en va de même pour les autres approches dans la chaîne de construction de l'IC à savoir le choix de la méthode de normalisation, d'agrégation, etc.

4. Conclusion

Dans cet article, nous avons passé en revue les étapes de construction d'un indicateur composite, en les illustrant par l'élaboration d'un indicateur de développement durable simple qui comporte les trois dimensions habituelles : économique, sociale et environnementale. Chaque étape est illustrée par les approches les plus utilisées en pratique. Les résultats obtenus ont été discutés, ainsi que la sensibilité de l'indicateur et des scores et l'incertitude liée aux changements de méthodes. Quelle que soit la méthode utilisée, l'indicateur reste soumis à de fortes incertitudes et sont sensibles aux méthodes utilisées faisant en sorte que les classements sous-jacents des pays en soient également affectés. De ce fait, et puisque les IC définissent, orientent et évaluent l'action publique, le choix d'une méthode dans la construction d'un IC devrait faire l'objet d'une justification théorique cohérente sans laquelle l'indice souffrirait d'un manque de légitimation. Aussi, étant donné son rôle combien prépondérant dans la définition de ce qu'un pays, un territoire ou une ville souhaite être ou devenir, des recherches dans ce champ disciplinaire devraient être encouragées en vue de développer des méthodes permettant de rendre les IC plus robustes.

Références bibliographiques

1. Benzécri J.P., 1973, *L'Analyse des Données, Tome II : L'Analyse des Correspondances*, Bordas, Paris (1re édition 1973, 2e édition 1976, 3e édition 1980, 4e édition 1982)
2. Brand DA, Saisana M, Rynn LA, Pennoni F, et al. (2007), *Comparative Analysis of Alcohol Control Policies in 30 Countries*. PLoS Med 4(4): e151. doi:10.1371/journal.pmed.0040151
<http://www.plosmedicine.org/article/info:doi/10.1371/journal.pmed.0040151>
3. Cherchye, L., W. Moesen, N. Rogge, T. Van Puyenbroeck, M. Saisana, A. Saltelli, R. Liska, and S. Tarantola (2008), “*Creating composite indicators with DEA and robustness analysis: the case of the Technology Achievement Index*”, Journal of the Operational Research Society 59, 239-251.
4. Chiappini R. (2012), « *Les indices composites sont-ils de bonnes mesures de la compétitivité des pays ?* », LAREFI Working Paper N°2012-05, <http://lare-efi.u-bordeaux4.fr>
5. Conseil de l'Europe (2005), *Elaboration concertée des indicateurs de la cohésion sociale : Guide méthodologique*, Editions du Conseil de l'Europe F-67075 Strasbourg Cedex [http : //book.coe.int](http://book.coe.int)
6. Donzé L. (2001), « *L'imputation des données manquantes, la technique de l'imputation multiple, les conséquences sur l'analyse des données : l'enquête 1999 KOF/ETHZ sur l'innovation* », Centre de recherches conjoncturelles ETH, Ecole polytechnique fédérale de Zurich
7. Dumolard (2011), *Données géographiques, analyse statistique multivariée*, Ed Lavoisier, Paris, 2011
8. DREES (2011), « *Analyse critique du développement d'indicateurs composites : le cas de l'infarctus du myocarde après la phase aiguë* », document de travail, Séries Sources et Méthodes, COMPAQH – INSERM n° 19 – avril 2011
9. Gadrey J. (2002), « *De la croissance au développement: à la recherche d'indicateurs alternatifs* », working paper, Laboratoire CLERSE, université Lille 1
10. García, C., Kovacevic, M.: *Uncertainty and Sensitivity Analysis of the Human Development Index*. Human Development Research Paper 47. UNDP-HDRO, New York (2011)
11. Homma, T. and Saltelli, A. (1996), “Importance measures in global sensitivity analysis of model output”. Reliability Engineering and System Safety, 52(1), 1-17.

12. Jacques J. (2011), « Pratique de l'analyse de sensibilité : comment évaluer l'impact des entrées aléatoires sur la sortie d'un modèle mathématique », <http://labomath.univ-lille1.fr/~jacques/>
13. Klugman J., Rodriguez F. and Choi H. J (2011), « *The HDI 2010: new controversies, old critiques.* », *J Econ Inequal* (2011) 9: 249-288
14. Nardo M., Saisana M., Andrea Saltelli A. and Tarantola S. (2005), *Tools for Composite Indicators Building*, European Commission-JRC: I-21020 Ispra (VA) Italy.
15. OECD-EU-JRC (2008), *Handbook on Constructing Composite Indicators methodology and user guide*, OECD
16. Reginster I., Ruyters C. (cord.)(2011), « *Développement d'indicateurs complémentaires au PIB Partie 1 : Revue harmonisée d'indicateurs composites/synthétiques* », working paper N°4 Novembre 2011, Fiche 1.11.1, IWEPS.
17. Saaty, T. L. (1980), *The Analytic Hierarchy Process*, New York: McGraw-Hill
18. Saisana M. and Saltelli A. (2010), *Uncertainty and sensitivity Analysis of the 2010 Environmental Performance Index*, JRC Scientific and Technical Reports
19. Saltelli, A., Chan, K. and Scott, M. (2000a), *Sensitivity analysis, Probability and Statistics series*, New York: John Wiley & Sons.

ANNEXES

Annexe 1 : les outils de l'analyse multi-variée

L'ACP: l'ACP est une méthode statistique développée au 20^{ème} siècle par Hotelling et Thurstone (1934) en vue de présenter de manière résumée (en faisant ressortir les informations principales) des données quantitatives continues (Dumolard, 2011). L'ACP suppose que le tableau à résumer est « sans trou » c'est-à-dire sans valeur manquante, ce qui peut être vu comme un inconvénient dans nombre d'applications. C'est pourquoi l'étape de traitement des données manquantes doit précéder l'analyse multi-variée. L'ACP s'effectue en trois étapes (transformation du tableau de données et calcul des covariances, calcul des axes factoriels et de leurs variances et calcul d'aides à l'interprétation des résultats) dont la première est utile à l'analyse multi-variée. En effet, la première étape d'une ACP consiste à transformer un tableau initial D contenant des informations brutes (données primaires recueillies) en un tableau secondaire D' de dimensions réduites où les variables sont centrées sur leurs moyennes respectives (la valeur de chaque variable est remplacée par son écart à sa moyenne $D'_{ij} = D_{ij} - m_j$ où j est la variable) ou en un tableau D'' dans lequel les variables sont standardisées. Il n'y a ni différence de moyenne ni différence de variance ($D''_{ij} = (D_{ij} - m_j) / \sigma_j$ ou $D''_{ij} = D' / \sigma_j$) dans ce dernier cas. Cette transformation offre de façon inhérente, des données déjà normalisées résolvant du même coup les différences d'unités de mesure des variables. Cependant, il faudra tenir compte de cette standardisation « prématurée » dans l'étape de normalisation des sous indicateurs pour éviter une double normalisation des variables.

La matrice d'informations fournie par le tableau D' ou D'' est ensuite transformée en un tableau C où les variables sont mises en relation deux à deux. Lorsque les variables sont centrées (tableau D'), le tableau C représente un tableau de covariance. S'il s'agit d'une transformation standardisée (tableau D''), le tableau C correspond à une matrice de corrélation des variables de l'étude (coefficients de Bravais-Pearson). En particulier, lorsque les variables de départ sont de type qualitatif ordinal, le tableau D'' fournit des coefficients de corrélation de rangs de Spearman. La standardisation (tableau D'') a l'avantage de présenter des statistiques qui ne dépendent pas des unités de mesures initiales en ce sens que la standardisation (centrage et réduction) exprime l'écart de chaque valeur à sa moyenne en écart type de cette variable. A la limite, c'est l'écart type qui devient l'unité de mesure commune à toutes les variables.

La deuxième étape consiste à un calcul de pourcentage de variance de chaque variable et des axes factoriels pour des représentations visuelles. Les axes factoriels représentent les informations principales extraites et indiquent le degré d'explication ou du pouvoir explicatif des informations retenues sur le phénomène traité exprimé par le % des variances. Ainsi le premier axe factoriel (déterminé à partir des écarts des variances dans la matrice de corrélation C) va contenir l'information principale, les autres contenant les informations secondaires suivant un ordre décroissant de variance.

Enfin, l'étape 3 correspond à des représentations (le plus souvent graphiques) d'outils d'aides à l'interprétation des résultats. Ainsi, des axes factoriels ou encore des cercles de corrélation peuvent être construits pour visualiser la qualité de l'information fournie après les étapes 1 et 2 et d'en décider de la pertinence et du pouvoir des variables retenues pour expliquer le phénomène étudié.

L'une des limites de l'ACP est qu'à chacune de ses étapes, on perd une fraction plus ou moins importante de l'information fournie par les variables initiales (limites liées aux simplifications et à la hiérarchisation des priorités). En outre, la standardisation ne tient pas compte des différences de poids relatifs que peuvent avoir les unités observées. A titre d'exemple, dans une ACP de la variable démographique, la standardisation va supposer un poids égal en termes de population entre la France et le Luxembourg ! Pour lever cette limite il convient de raffiner l'analyse en calculant un coefficient de variation ($r_j = \sigma_j/m_j$). Ainsi, un r_j élevé indique une dissemblance entre les unités statistiques du point de vue de du caractère j considéré.

L'AFC : l'AFC elle, est développée par J.P. Benzecri (1957) pour traiter principalement (mais pas seulement) des statistiques qualitatives. Les étapes d'une AFC, à de différences près, sont identiques à celles d'une ACP d'autant que les deux analyses présentent des points communs. En effet, tout comme l'ACP, l'AFC est une analyse factorielle visant à (i) résumer l'information contenue dans de grands tableaux numériques, (ii) partir d'une représentation sous forme de nuage de points multidimensionnel, induire des résumés descriptifs hiérarchisés et enfin, (iii) leur donner une signification grâce à des aides à l'interprétation.

Cependant, l'AFC porte sur des tableaux dont les lignes et colonnes sont symétriques, croisant deux caractères jouant un rôle équivalent c'est-à-dire qu'on peut donc les permuter. Les tableaux peuvent contenir des valeurs nulles sans que l'analyse ne soit impossible, ce qui ne

l'est pas dans une analyse ACP. Enfin, la dernière particularité est que pour le croisement de caractères les sommes en lignes et en colonnes ont un sens.

En AFC, le passage d'une matrice D d'informations initiales à une matrice secondaire D' est donnée par les fréquences conditionnelles (en ligne i) ($D_{if} = f_{ij/i} = D_{ij}/D_{i.}$) et la moyenne

pondérée (de la colonne j)
$$m_j = \sum_{i=1}^n f_i \cdot f_{ij/i}$$

Où $f_{ij/i} = n_{ij} / n_i$. Fréquence en proportion du total de la ligne i (fréquence conditionnelle)

D_{ij} la valeur obtenue par croisement de la ligne i et de la colonne j

$D_{i.}$ effectif de la somme des valeurs de la ligne i

$f_i = n_i/n$, poids d'une ligne est la fréquence de la somme des valeurs de la ligne i sur l'effectif total. La somme de toutes les $f_i = 1$.

Le passage de la matrice D' à D'' (centrée et réduite) s'effectue en calculant la variance de la distribution de la manière suivante : $\sigma_j^2 = \sum_{i=1}^n f_i \cdot (f_{ij/i} - G_j)^2$

Où $G = m_1, m_2, \dots, m_p$, dans une matrice d'information D à n-lignes et p-colonnes, est le centre de gravité du nuage de n points-ligne.

Enfin, la matrice C de covariances est calculée par le produit matriciel $C = F_{ij/i} * F_{ij/j}$. Les majuscules désignent des matrices $F_{ij/i}$ celle des fréquences en ligne et $F_{ij/j}$ celle des fréquences en colonne. Comme les tableaux en AFC sont symétriques, dans le produit factoriel, il se passe comme si on multipliait une matrice par sa transposée. On obtient finalement une matrice symétrique dont on peut retenir soit la diagonale supérieure soit celle inférieure pour analyser les corrélations entre les différentes variables de l'étude.

Annexe 2 : L'Analyse d'incertitude des IC

L'Analyse de l'incertitude par la méthode de Monte Carlo

En considérant que toutes les étapes et les différentes méthodes de construction d'un IC génèrent des incertitudes qui se répercutent sur la variable résultat (ici le rang attribué à un pays par la valeur de son IC) l'analyse d'incertitude consiste à déterminer une fonction de distribution probabiliste permettant de relier les inputs (sous indicateurs) à l'output (rang) au moyen d'une combinaison aléatoire des différentes méthodes et étapes.

Diverses méthodes existent pour estimer l'incertitude de la variable résultat. Nardo et al. (2005) exposent l'estimation de cette incertitude par la méthode de Monte Carlo de la manière suivante :

La première étape est relative à la méthode utilisée pour le traitement des données manquantes dans la construction de l'IC. Les auteurs notent X_i ($i=1, \dots, k$) la variable aléatoire correspondant aux différentes étapes et méthodes. La variable aléatoire X_1 caractérise le traitement des données manquantes, et prend deux valeurs distinctes : 1 lorsque la méthode utilisée consiste à remplacer la variable dont les données manquent par une autre variable fortement corrélée à la première. A titre d'exemple, on pourra remplacer la variable « Investissement » par la variable « Epargne » et vice versa. X_1 prend le chiffre 2 lorsqu'on attribue un zéro à la variable dont les données manquent.

$$X_1 = \begin{cases} 1, & \text{si variable de remplacement} \\ 2, & \text{si attribution d'un zéro aux valeurs manquantes} \end{cases}$$

La deuxième variable aléatoire est relative à la méthode de normalisation des variables de départ.

$$X_2 = \begin{cases} 1 & \text{si } I = [I - \min(I)]/\text{benchmark} \\ 2 & \text{si } I = (I - \bar{I})/\sigma \\ 3 & \text{si données brutes} \end{cases}$$

Les auteurs supposent en outre que les deux variables aléatoires discrètes X_1 et X_2 sont distribuées de façon uniforme sur $[0; 1]$. En supposant le nombre aléatoire ζ , $X_1=1$ si $\zeta \in [0; 0,5)$ et $X_1=2$ si $\zeta \in [0,5; 1]$

De façon analogue, on définit X_3 comme étant la variable aléatoire représentant l'évènement « nombre de sous indicateurs isolés pour l'analyse » sachant que J sous indicateurs composent l'IC. Ainsi :

$$X_3 = \begin{cases} 0 & \text{si } \zeta \in \left[0; \frac{1}{J+1}\right) \text{ tous les sous indicateurs sont utilisés dans l'analyse} \\ 1 & \text{si } \zeta \in \left[\frac{1}{J+1}; \frac{2}{J+1}\right) \\ \dots & \dots \\ J & \text{si } \zeta \in \left[\frac{J}{J+1}; 1\right] \end{cases}$$

$\frac{1}{J+1}$ est la probabilité qu'aucun sous indicateur ne soit exclu de l'analyse tandis que $1 - \frac{1}{J+1}$ est la probabilité pour qu'au moins un sous indicateur soit exclu.

Notons que l'exclusion d'un sous indicateur fait référence à l'hypothèse que certains sous indicateurs ne peuvent être considérés dans certaines méthodes. A titre d'exemple, lorsque la méthode d'agrégation utilisée est la moyenne géométrique, tous les sous indicateurs dont la valeur est négative doivent être exclus de cette agrégation par définition d'une moyenne géométrique. En outre, en excluant le sous indicateur j de l'analyse de simulation, on isole la contribution de ce dernier à la formation de l'IC et permet de ce fait d'apprécier, toutes choses égales par ailleurs, l'importance relative de la dimension j dans l'explication du phénomène appréhendé par l'IC.

La variable aléatoire X_4 est utilisée pour capter l'incertitude liée à la méthode d'agrégation. Trois méthodes d'agrégation sont retenues : la méthode linéaire (LIN) c'est-à-dire la moyenne arithmétique, la méthode géométrique (GEM) et la méthode d'analyse multicritère (AMC) discutée en section 4.

$$\text{En} \qquad \qquad \qquad \text{effet,}$$

$$X_4 = \begin{cases} 1 & \text{si LIN (IC} = \sum w_j * SI_{ij}) \\ 2 & \text{si GEM (IC} = \prod_{j=1}^J (SI_{ij})^{w_j}) \\ 3 & \text{si AMC les scores de l'IC sont directement générés par la méthode} \end{cases}$$

La variable aléatoire X_5 est générée pour prendre en compte l'incertitude liée au système de pondération choisie. Trois systèmes de pondération ont été retenus par les auteurs. Il s'agit de la méthode du « Benefit of the Doubt (BOD) », de l'allocation budgétaire (BAP) et la méthode d'analyse hiérarchique (AHP) non développée dans cet article.

$$X_5 = \begin{cases} 1 & \text{si BAP} \\ 2 & \text{si AHP} \\ 3 & \text{si BOD} \end{cases}$$

La dernière variable aléatoire générée est X_6 . Elle permet de capter l'incertitude liée au jugement de l'expert notamment lorsqu'il y a de l'incohérence dans ses jugements de valeur se traduisant par des répartitions de points entre différents domaines de façon illogique. X_6 prend les valeurs $0, 1, \dots, N$ où N est le nombre d'experts participant à l'étude. Comme les experts sont choisis de façon aléatoire, chaque expert sélectionné est associé au poids que celui-ci attribue aux dimensions J .

Notons cependant que lorsque dans l'analyse, la variable $X_5=3$, la variable X_6 est d'office égale à zéro pour la simple raison que la méthode de pondération tirée au hasard ($X_5=3$ correspond à la méthode de BOD) ne fait pas intervenir le point de vue de l'expert. Les poids

dans ce cas sont déterminés de façon endogène c'est-à-dire déterminés par le modèle lui-même.

Etant données les six variables aléatoires ci-dessus générées, l'analyse de Monte Carlo consiste à définir une fonction probabiliste combinant les six variables aléatoires. Il est donc possible de générer N combinaisons à partir des X_i^l ($i=1, \dots, k$ avec $k=6$ dans notre cas et $l=1, 2, \dots, N$) variables aléatoires puis d'analyser l'impact de chaque combinaison sur la variable résultat c'est-à-dire sur la valeur de l'IC final ou sur le rang induit. Les échantillons X_i^l peuvent être obtenus à partir de plusieurs méthodes de randomisation telles que l'échantillonnage aléatoire simple, la méthode d'échantillonnage quasi-aléatoire, l'échantillonnage par stratification, etc. (Saltelli et al. 2000a). La variable de résultat (IC ou rang du pays) est liée aux variables aléatoires par la fonction de densité probabiliste mentionnée plus haut. A partir d'un seuil arbitraire fixé, il est possible de déterminer les caractéristiques de cette fonction de densité à partir du nombre de simulations N obtenues. En appliquant cette méthode de Monte Carlo à l'indice de développement technologique, Nardo et ses co-auteurs trouvent que le classement des pays varie lorsque l'on prend en compte toutes les incertitudes liées aux différentes étapes de la construction de l'IC. Nous reviendrons sur cet indicateur dans la section suivante en analysant les récents travaux qui ont porté sur celui-ci notamment ceux de Cherchye L. et al. (2008).

Annexe 3 : L'Analyse de sensibilité de l'IC.

L'analyse de sensibilité par construction d'intervalles de confiance

Il s'agit d'associer à la valeur de l'IC un intervalle de confiance dont la marge d'erreur est fixée au préalable. L'étape suivante consiste à faire varier les différentes méthodes puis d'apprécier la sensibilité de l'IC par rapport à ces changements. Pour un seuil de confiance donné, l'IC sera dit sensible aux méthodes, lorsque sa valeur calculée change d'intervalle suivant la méthode (normalisation, pondération, agrégation, etc.) utilisée.

Dans la pratique deux types d'intervalle de confiance peuvent être construites :

L'intervalle par la méthode des percentiles simples

Cette méthode fournit les bornes de l'intervalle à partir de la valeur estimée d'un paramètre (moyenne, médiane, etc.) d'une population mère à partir de laquelle un échantillonnage a été effectué par un tirage aléatoire avec remise.

$IC = \left[\hat{\theta}_{\left[\frac{\alpha}{2}\right]}; \hat{\theta}_{\left[1-\frac{\alpha}{2}\right]} \right]$ avec $\hat{\theta}$ la valeur estimée du paramètre considéré ; $1 - \frac{\alpha}{2}$ et $\frac{\alpha}{2}$ les percentiles de la distribution de l'échantillonnage et α la marge d'erreur. Dans le cas de la moyenne $\hat{\theta} = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b$ où B est le nombre d'échantillonnages obtenus par tirage aléatoire avec remise et $\hat{\theta}_b$ l'estimateur du paramètre θ .

La DREES (2011) fait remarquer que pour que cette méthode soit applicable, il est important que la distribution du paramètre que l'on souhaite construire ne suive pas une loi normale. Par ailleurs, un nombre élevé de ré-échantillonnages est requis pour obtenir une précision suffisante sur l'écart type de la distribution. Dans la pratique, il faut que B avoisine 1000. Cependant, l'intervalle de confiance se resserre au fur et à mesure que la taille de l'échantillon augmente. La restriction de l'intervalle donne une plus grande fiabilité à l'IC avec tout de même un risque que celui-ci n'appartienne pas à cet intervalle pour une certaine taille de l'échantillon d'une part et pour un certain degré de certitude donné (par exemple pour $\alpha=1$) d'autre part.

L'intervalle par la méthode de l'écart-type

Elle diffère de la méthode décrite ci-dessus par la distribution de l'échantillon. En effet, la méthode est applicable lorsque la distribution de l'échantillon obtenu suit une loi normale. Dans cette hypothèse, l'intervalle de confiance associé à l'IC est donné par :

$IC = \left[\hat{\theta} - u_{\left[1-\frac{\alpha}{2}\right]} * \hat{\sigma}_{\theta}; \hat{\theta} + u_{\left[1-\frac{\alpha}{2}\right]} * \hat{\sigma}_{\theta} \right]$ où $u_{\left[1-\frac{\alpha}{2}\right]}$ est la valeur de la distribution normale centrée réduite, $1-\alpha$ le degré de confiance retenu.

L'écart-type $\hat{\sigma}_{\theta}$ de la distribution est donné par : $\sqrt{\frac{\sum_{b=1}^B (\hat{\theta}_b - \hat{\theta})^2}{(B - 1)}}$

Contrairement à la méthode de percentiles simple, l'estimation de l'intervalle de confiance pour un IC par la méthode de l'erreur standard requiert un nombre de ré-échantillonnage relativement moins élevé. Il est très rare que plus de 200 répétitions soient nécessaires pour l'estimation de l'erreur standard ; 50 donnant déjà une précision assez suffisante sur le paramètre (DREES, 2011).

Lorsque l'IC est exprimé en pourcentage, il est possible, à partir d'un échantillon ayant servi à l'implémentation de l'IC, de lui associer un intervalle de confiance afin d'évaluer la

sensibilité de l'IC suite à un changement de méthode. Dans le cas d'un IC exprimé en %,

$$\text{l'intervalle de confiance est donné par : } IC = P_0 \pm t_\alpha * \sqrt{\frac{P_0 * Q_0}{n}}$$

Avec t_α la statistique de Student ; la valeur usuelle de t_α au seuil habituel de 5% vaut 1,96.

P_0 est la valeur de l'indicateur composite exprimée en pourcentage et Q_0 sa valeur complémentaire ($Q_0=100\% - P_0$).

Dans cette méthode, l'amplitude de l'intervalle est directement déterminée par la valeur de l'IC. En effet, pour un échantillon de taille n donnée, l'intervalle se resserre au fur et à mesure que la valeur de l'IC devient faible. Il est donc évident que si une méthode influence la valeur de l'IC, son intervalle de confiance se trouvera également affecté par voie de ricochet.

L'analyse de sensibilité par décomposition de variance

La méthode de la décomposition de la variance totale permet d'évaluer la solidité de l'output (IC), la variance étant une mesure de l'imprécision. L'analyse consiste à évaluer la contribution de chaque variable d'entrée (sous indicateurs) à la variance totale de l'indicateur composite d'une part et d'en déceler la part expliquée par les interactions entre les différents inputs (colinéarité, endogénéité, etc.) d'autre part. Cette décomposition permet de construire des indices de sensibilité de l'IC, une mesure quantitative de la sensibilité.

En considérant l'IC comme variable d'intérêt, et les sous indicateurs comme inputs, la première étape de la méthode consiste à spécifier une fonction liant l'output (ici l'indicateur composite) aux variables explicatives (les sous composantes de l'IC). Lorsque la spécification fonctionnelle liant la variable de sortie Y aux variables d'entrée X – supposées toutes indépendantes – est linéaire ($Y = \beta_0 + \sum_{i=1}^P \beta_i X_i$) un premier indice de sensibilité peut être construit (J. Jacques, 2011). L'indice SCR_i (*Standardized Regression Coefficient*) exprime la part de la variance de l'IC imputable à la variance de la variable X_i . $SCR_i = \frac{\beta_i^2 V(X_i)}{V(Y)}$. Où $\beta_i^2 V(X_i)$ est la variance de la variable X_i .

Nardo et al. (2005) notent qu'étant données les incertitudes liées aux différents niveaux de construction de l'IC, la forme fonctionnelle du modèle ne peut être linéaire ou forme additive. Les auteurs soutiennent que le modèle approprié pour cette analyse de sensibilité est un modèle non linéaire dont la spécification reste à déterminer. Cependant, ces modèles bien que non connus à l'avance, doivent vérifier les propriétés suivantes.

- Pour un nombre connu de n sous indicateurs, les modèles permettent une estimation de la variance totale expliquée par les n facteurs ;
- Les modèles permettent une analyse de sensibilité dans laquelle les inputs comportant des incertitudes sont considérés en groupes plutôt que traités individuellement et ce, dans l'optique d'estimer la part de la variance liée aux interactions entre les variables (effets cumulatifs des incertitudes, biais d'endogénéité des variables etc.).
- Les variances sont quantifiables et permettent une décomposition en variance principale et en variance secondaire (variance liée aux interactions entre les différentes variables explicatives) ;
- Elles (variances) sont facilement interprétables et explicables ;
- Enfin, elles devraient permettre de discuter du niveau de robustesse de l'IC.

Etant donnés X_i inputs (sous indicateurs), la contribution relative de la variable X_i à la variance totale de l'output Y (IC) est donnée par la variance de l'espérance conditionnelle de Y . $V_i = V_i(E_{X_i}(Y|X_i))$. Pour une valeur précise de $X_i=x_i^*$, il est possible de calculer la moyenne conditionnelle de Y . En particulier, lorsque X_i n'influence pas la variance principale de Y , $V_i=0$ et $V_i= V(Y)$ dans le cas où la variance de l'output est totalement expliquée par ce facteur X_i , tous les autres facteurs n'ayant aucun effet sur la variance totale.

La décomposition de la variance totale en variance principale et variance résiduelle est donnée par : $V(Y) = V_{X_i}(E_{X_i}(Y|X_i)) + E_{X_i}(V_{X_i}(Y|X_i))$. Ainsi, lorsqu'un facteur X est important dans la composition de la variance de Y , sa valeur résiduelle $E_{X_i}(V_{X_i}(Y|X_i))$ est faible. Inversement, lorsque la variance conditionnelle est faible, les niveaux d'interaction entre inputs et les effets cumulatifs des incertitudes sont énormes réduisant donc la variance conditionnelle.

En rapportant la variance conditionnelle à la variance totale, on obtient un premier indicateur de sensibilité de l'IC par rapport à la variable X_i .

$S_i = \frac{V_{X_i}(E_{X_i}(Y|X_i))}{V(Y)} = \frac{V_i}{V(Y)}$. S_i donne la contribution relative de la i -ième variable (seule) à la composition de la variable totale. Ainsi, plus le domaine ou la dimension traduit par la variable X_i est important, plus S_i sera grand. En particulier, lorsque la variable explique la quasi-totalité des variations de l'output, l'indicateur de sensibilité tend vers l'unité ($S_i \sim 1$). Dans ce cas précis, les incertitudes et les interactions sont négligeables (valeur résiduelle de la variance conditionnelle vaut zéro).

De façon analogue, il est possible de calculer des contributions relatives à la variance totale conditionnement à plusieurs variables explicatives interagissant entre elles.

Pour deux facteurs X_i et X_j donnés, la variance conditionnelle par rapport aux deux facteurs s'écrit : $V_{X_i X_j}(E_{X_{-ij}}(Y \setminus X_i, X_j))$. La variance résiduelle (résultant de l'interaction entre X_i et X_j) est donnée par : $V_{ij} = V_{X_i X_j}(E_{X_{-ij}}(Y \setminus X_i, X_j)) - V_{X_i}(E_{X_{-i}}(Y \setminus X_i)) - V_{X_j}(E_{X_{-j}}(Y \setminus X_j))$. Elle permet de déceler si dans le modèle les différentes variables explicatives entretiennent des relations entre elles. En absence de toute interaction entre X_i et X_j dans le modèle de construction de l'IC, V_{ij} vaut zéro. Autrement dit, toutes les variables explicatives – dans notre cas, tous les sous indicateurs – sont indépendantes et non colinéaires.

Pour k variables explicatives toutes indépendantes les unes des autres, la décomposition de la variance totale est donnée par la formule suivante : $V(Y) = \sum_i V_i + \sum_i \sum_{j>i} V_{ij} + \sum_i \sum_{j>i} \sum_{l>j} V_{ijl} + \dots + V_{12\dots k}$

Dans l'hypothèse d'indépendance totale entre les inputs, le modèle de décomposition de la variance totale est une somme des contributions marginales de chaque facteur.

$$\sum_{i=1}^k V_i = V(Y) \text{ et } \sum_{i=1}^k S_i = 1.$$

A partir de la décomposition en variances résiduelles, il est possible de calculer des indices de sensibilité relatifs aux interactions entre les X_i variables explicatives. Cependant, Nardo et al. (2005) montrent que le nombre n de ces indices est aussi élevé que le nombre de variables k est élevé ; $n = 2^k - 1$. Dans la pratique on calcule plutôt un indice condensé des interactions entre variables. Il donne l'effet marginal du facteur i dans l'explication de la variance totale de Y étant donnés les effets dus aux interactions des autres variables c'est-à-dire les variances résiduelles des $k - i$ facteurs.

Pour un IC comportant trois facteurs, l'indice marginal de sensibilité est :

$$S_{T1} = \frac{V(Y) - V_{X_2 X_3}(E_{X_1}(Y \setminus X_2, X_3))}{V(Y)} = S_1 + S_{12} + S_{13} + S_{123}. S_{T1} \text{ est le rapport de la somme des variances où intervient l'indicateur 1 de façon individuelle } (S_1) \text{ ou en interaction avec les autres indicateurs } (S_{12}, S_{13} \text{ et } S_{123}), \text{ et de la variance totale de l'indicateur composite.}$$

L'indice résiduel du facteur 2 s'écrit : $S_{T2} = S_2 + S_{12} + S_{23} + S_{123}$ et celui du troisième facteur vaut $S_{T3} = S_3 + S_{13} + S_{23} + S_{123}$

Homma et Saltelli (1996) montrent que $V_{X_2, X_3}(E_{X_1}(Y \setminus X_2, X_3))$ peut être généralisée comme suit : $V_{X_i}(E_{X_i}(Y \setminus X_i))$. Elle donne la contribution des $k - i$ variables à l'explication de la variance totale.

$$\text{Donc } S_{Ti} = \frac{V(Y) - V_{X_i}(E_{X_i}(Y \setminus X_i))}{V(Y)} = \frac{E_{X_i}(V_{X_i}(Y \setminus X_i))}{V(Y)} \text{ et } \sum_{i=1}^k S_{Ti} \geq 1.$$

Finalement, les deux indicateurs de sensibilité (S_i et S_{Ti}) permettent d'apprécier le degré d'adéquation globale du modèle ayant servi à la construction de l'IC et donc de la robustesse de celui-ci. En particulier, lorsqu'il existe une différence significative entre les indices S_i et S_{Ti} , ce résultat témoigne des effets d'endogénéité et de multi-colinéarité entre les facteurs X_i qu'il faille apporter des corrections (par déconstruction ou par changement de pondération ou encore par substitution de certaines variables par des variables non colinéaires) sans lesquelles l'IC final sera entaché de biais énormes.