Détermination des parités de pouvoir d'achat élémentaires dans le cadre du PCI-Afrique: La méthode CPD

Folly A. D. Adjogou et Mathieu B. Djayeola[1]

Résumé :

Les parités de pouvoir d'achat (PPA) sont destinées aux comparaisons transnationales d'agrégats économiques tels que le Produit intérieur brut ou ses emplois. Si l'interprétation économique des PPA est relativement simple, leur estimation requiert la maîtrise de procédures plutôt complexes, et le recours à la programmation économétrique dans un environnement informatique. Cet article présente de façon détaillée l'une des méthodes de calcul des PPA appelée méthode du Country Product Dummy (CPD). Trente ans après ses premières ébauches, dans un souci de vulgarisation à grande échelle, cet article expose sous forme matricielle les étapes méthodologiques du CPD afin de faciliter son opérationnalisation.

Mots clés : *Parité de pouvoir d'achat, CPD, Programme de comparaison internationale*

Summary:

Purchasing power parities (PPP) are intended for inter-country comparisons of economic aggregates such as Gross domestic product or its components. While the economic interpretation of the PPPs is relatively easy, their estimate requires mastering rather complex procedures, including econometric modelling in a data-processing environment. This article presents in a detailed way one of the PPP computation methods called Country Product Dummy (CPD) method. Thirty years after its first implementation, this paper shows its various methodological stages in a matrix form to facilitate its utilization.

Key words: *Parité de pouvoir d'achat, CPD, Programme de comparaison internationale*

[1]Ingénieurs Statisticiens Economistes assistants de recherche à la BAD: (f.adjogou@afdb.org) et (b.djayeola@afdb.org).

1. Introduction

La méthode Country Product Dummy (CPD), développée en 1973 par l'Economiste Robert Summers est une procédure de régression économétrique utilisée pour calculer des Parités de pouvoir d'achat (PPA). Dans leur forme la plus simple, les Parités de pouvoir d'achat sont des prix relatifs, ou autrement dit, des rapports de prix en monnaies nationales d'un même bien ou service dans différents pays. Les PPA sont utilisées essentiellement par les organisations internationales, les instances gouvernementales, les universités, les instituts de recherche et autres acteurs socio-économiques comme instrument pour la recherche économique et l'analyse politique dans les comparaisons internationales.
Dans le cadre du Programme de comparaison internationale pour l'Afrique (PCI-Afrique) qui est mis en œuvre en tant que partie intégrante du programme global du PCI 2003, la méthode CPD est utilisée pour estimer les PPA au niveau des positions élémentaires[2].
Ce papier explore une présentation matricielle du CPD tel qu'utilisé par le PCI-Afrique. Les notions exposées ont permis de développer une application Visual Basic qui génère les Parités de pouvoir d'achat.

2. Quelques méthodes de calcul de PPA

La méthode CPD a été utilisée dans les premières phases du Programme de comparaison internationale, principalement comme un outil pour l'agrégation des données de prix en dessous du niveau des positions élémentaires. Dans des phases plus récentes des travaux du PCI, la méthode Elteto-Koves-Szulc (EKS), a remplacé la CPD en tant que procédure de remplissage des données de prix manquants. On note cependant un regain d'intérêt dans la méthode CPD depuis les récents travaux de Prasada Rao (1995)[3]. Un certain nombre de propriétés de la méthode du CPD ont en effet été passées en revue par Rao (2001,2002)[4] en l'occurrence, la CPD pondérée a été utilisée pour obtenir un certain nombre de méthodes alternatives pour les comparaisons multilatérales. Diewert (2002)[5] montre

[2]Position élémentaire: Regroupement homogène de produits ou services ayants des caractéristiques communes
[3]Rao, D.S. Prasada (1995), "On the Equivalence of the Generalized Country-Product-Dummy (CPD) Method and the Rao-System for Multilateral Comparisons"
[4]D.S. Prasada (2001), "Weighted EKS and Generalised CPDMethods for Aggregation at Basic Heading Level and above Basic Heading Level"
[5]Hedonic Regressions : A Review of Some Unresolved Issues

comment de nouvelles spécifications et de nouvelles pondérations peuvent être utilisées dans le cadre de la CPD pour tirer des formules connues d'indices de prix. Diewert (2003, 2004)[6] a également examiné les propriétés variées de la méthode du CPD dans le cadre de la comparaison entre régions.

En effet, le choix de la méthode d'agrégation pour construire les PPA est débattu depuis deux décennies. Les premières études sur le thème des comparaisons internationales, telles que l'étude pionnière de Kravis, Heston et Summers (1982), ont fourni un large éventail de méthodes d'agrégation. Un travail de référence de l'OCDE et Eurostat a comparé les méthodes de Geary-Khamis (GK) et de Elteto-Koves-Szulc (EKS). L'agrégation s'effectue après qu'on a calculé la moyenne des indices des prix pour les biens et services de base pour obtenir des parités non pondérées pour des petits groupes de produits homogènes. Il faut ensuite pondérer et agréger les parités non pondérées s'appliquant aux groupes de produits pour obtenir des PPA et des valeurs réelles pour chaque catégorie de dépenses en remontant jusqu'au niveau du PIB total. Les deux méthodes diffèrent sensiblement. La méthode EKS considère que les pays sont un ensemble d'entités indépendantes : chaque pays reçoit un poids identique. Dans la méthode EKS, on obtient les prix en minimisant les différences entre d'une part, les PPA binaires au niveau multilatéral, et d'autre part les PPA binaires bilatérales. Les PPA de la méthode EKS ne sont guère différentes des PPA qui auraient pu être obtenues si chaque couple de pays avait été comparé individuellement. La méthode GK considère que les pays font partie d'un groupe. Chaque pays reçoit une pondération représentant sa part dans le PIB total et les prix qui sont calculés sont représentatifs plus généralement du groupe. Les deux méthodes présentent des avantages et des inconvénients:

- Pour des pays possédant des structures des prix très différentes de moyenne, l'approche GK produit des estimations des volumes (et du PIB par habitant) plus élevées que celles qui auraient été calculées si on avait eu recours à des prix plus spécifiques au pays. Cette surestimation est particulièrement importante lorsqu'on compare des pays présentant de fortes différences de niveau de revenu. Toutefois, les résultats de l'approche GK satisfont au critère d'additivité, ce qui implique que la valeur réelle des agrégats est égale à la somme de la valeur réelle de ses

[6] Diewert , E.W. (2004) "On the Stochastic Approach to Linking Regions in the ICP"

composantes. C'est un atout dans la perspective des comptes nationaux, car cela permet de comparer les structures de prix et de volume d'un pays à l'autre.
- La méthode EKS produit des résultats qui reflètent davantage les caractéristiques des prix de chaque pays. On obtient ainsi des estimations du PIB par habitant relativement semblables à celles résultant de l'utilisation de prix spécifiques. Cependant, ses résultats ne satisfont pas au critère d'additivité.

La méthode CPD constitue une alternative intéressante aux indices de prix standard utilisés dans les comparaisons internationales. Du fait qu'elle repose sur des techniques économétriques, elle confère aux utilisateurs une commodité qui leur permet d'exploiter plus efficacement les données de prix notamment dans le traitement des données manquantes. Au-delà de cette succincte présentation de la méthode CPD, ce travail revêt un caractère purement pédagogique.

L'accent qui est mis ici sur la présentation matricielle a pour but d'élucider les grandes étapes de la technique afin d'en permettre une mise en œuvre rapide et efficace. En effet, l'intérêt majeur du travail consiste à mettre à la disposition des statisticiens des prix et organismes internationaux ainsi qu'aux écoles de formation de statisticiens en Afrique, un outil didactique simple et concis pouvant conduire à des programmations informatiques de la méthode CPD selon le langage de préférence.

3. La matrice générale des prix

3.1 La matrice initiale des prix

La matrice générale des prix est constituée des données de prix moyens annuels disponibles pour l'ensemble des pays et pour tous les produits. Il s'agit d'une matrice qui dispose en colonnes les pays participant au Programme et en lignes, tous les produits identifiés par positions élémentaires. Soit F le nombre total de produits et G le nombre de pays participant au Programme. Après l'identification du pays de base, la matrice des données P se présente de la façon suivante:

$$P = \begin{array}{cc} & \begin{array}{cccc} 1 & 2 & \dots & G \end{array} \\ \begin{array}{c} 1 \\ 2 \\ \vdots \\ F \end{array} & \left[\begin{array}{cccc} \; & \; & \; & \; \\ \; & \; & \; & \; \\ \; & \; & \; & \; \\ \; & \; & \; & \; \end{array} \right] \end{array}$$

3.2 La matrice des prix renseignés

L'étape suivante consiste à éliminer les produits pour lesquels aucun des pays n'a fourni de prix. Soit N le nouveau nombre de produits renseignés, donc de lignes pour la matrice P. La nouvelle représentation matricielle est indiquée par la matrice Q suivante:

$$Q = \begin{array}{cc} & \begin{array}{cccc} 1 & 2 & \dots & G \end{array} \\ \begin{array}{c} 1 \\ 2 \\ \vdots \\ N \end{array} & \left[\begin{array}{cccc} \; & \; & \; & \; \\ \; & \; & \; & \; \\ \; & \; & \; & \; \\ \; & \; & \; & \; \end{array} \right] \end{array}$$

4. La matrice des prix par position élémentaire

4.1 La matrice de base des prix par position élémentaire

La matrice générale des prix Q présente les relevés de prix des N produits dans tous les G pays. Ces produits sont regroupés en différentes positions élémentaires. Soit x une position élémentaire. L'on désigne par n_x le nombre de produits appartenant à x. La matrice $P(x)$ ci-dessous présente les données de prix relatives à x. Elle est déterminante pour le calcul des PPA pour la position élémentaire x. En d'autre termes, chaque position élémentaire x a sa matrice de prix $P(x)$.

$$P(x) = \begin{array}{cc} & \begin{matrix} 1 & 2 & \ldots & G \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ \vdots \\ n_x \end{matrix} & \left[\begin{matrix} & & & \\ & & & \\ & & & \\ & & & \end{matrix} \right] \end{array}$$

4.2 La matrice finale des prix renseignés par position élémentaire

Il peut arriver qu'un pays, parmi les G pays de départ, ne dispose de relevés de prix pour aucun des produits de la position élémentaire x. Dans ce cas, ce pays est retiré de la suite des opérations matricielles afférentes à la position élémentaire x[7]. Soit px le nombre de pays disposant de données pour les produits relatifs à x. La matrice P(x) est transformée en une nouvelle matrice P*(x) qui se présente sous la forme suivante:

$$P^*(x) = \begin{array}{cc} & \begin{matrix} 1 & 2 & \ldots & p_x \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ \vdots \\ n_x \end{matrix} & \left[\begin{matrix} & & & \\ & & & \\ & & & \\ & & & \end{matrix} \right] \end{array}$$

C'est cette matrice $P^*(x)$ qui fera l'objet des prochaines opérations relatives à la position élémentaire x. On peut l'appeler "matrice récapitulative de x".

[7] En effect, ce pays est momentanément exclu et sa PPA par rapport au pays de base sera considérée comme "indisponible". Mais grâce à la propriété de la méthode CPD qui est basée sur un modèle et présentée en Section VI, ce pays sera réintégré plus tard et sa PPA pourra être estimée

5. Les variables indicatrices 'Country' et 'Product'

Les étapes suivantes conduisent à l'élaboration des variables indicatrices relatives aux pays et aux produits de la position élémentaire x. Ces variables indicatrices sont doublement importantes dans la procédure de régression. Elles traduisent le fait que la méthode CPD repose essentiellement sur le principe dual Pays et Produit et elles servent d'input à la technique du SWEEP OPERATOR[8]. Afin de mieux exploiter la méthode par l'analyse matricielle, trois types de vecteurs sont nécessaires: Le vecteur des prix, le vecteur des pays, le vecteur des produits.

5.1 Le vecteur des Prix

Considérons la matrice $P^*(x)$ indiquée ci-dessus. On redresse les relevés de prix en une matrice unicolonne appelée Prix_x.

Supposons que la matrice $P^*(x)$ s'écrive sous la forme $P^*(x) = (a_{i,j})$ avec i allant de 1 à nx (produit) et j allant de 1 à px (pays).(pays). Le redressement en colonne concerne uniquement les termes $a_{i,j}$ pour lesquels l'emplacement est non vide. Considérons l'exemple suivant:

$$P^*(x) = \begin{array}{cc} & \begin{array}{ccc} \quad 1 \quad & \quad 2 \quad & \quad 3 \quad \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{bmatrix} (1,1) & (1,2) & (1,3) \\ (2,1) & (2,2) & \underline{(2,3)} \\ (3,1) & \underline{(3,2)} & (3,3) \\ \underline{(4,1)} & (4,2) & \underline{(4,3)} \end{bmatrix} \end{array}$$

Les éléments de la matrice qui sont soulignés se réfèrent à des cellules vides. Par exemple le pays 3 ne dispose pas de relevé de prix pour les produits 2 et 4.On obtient le vecteur des prix suivant:

$$Vecteur\_prix = \begin{array}{cc} & Prix\_x \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{array} & \begin{bmatrix} (1,1) \\ (2,1) \\ (3,1) \\ (1,2) \\ (2,2) \\ (4,2) \\ (1,3) \\ (3,3) \end{bmatrix} \end{array}$$

[8] Voir page 14

Il faut signaler que le redressement en colonne a été effectué colonne par colonne en parcourant à la suite, toutes les colonnes de la matrice $P^*(x)$. Si le redressement est effectué ligne par ligne, la structure du vecteur des prix sera changée et l'on obtiendrait le vecteur suivant:

$$Vecteur\_prix = \begin{bmatrix} & \Pr ix\_x \\ 1 & (1,1) \\ 2 & (1,2) \\ 3 & (1,3) \\ 4 & (2,1) \\ 5 & (2,2) \\ 6 & (3,1) \\ 7 & (3,3) \\ 8 & (4,2) \end{bmatrix}$$

Ce changement dans la structure du vecteur des prix modifie les structures du vecteur des pays, du vecteur des produits et de la matrice des indicatrices qui sont obtenus à partir de la disposition des éléments dans le vecteur des prix. Mais ceci n'a aucun impact sur les résultats de la régression et donc sur l'estimation des Parités de pouvoir d'achat. En effet, ce changement de disposition des éléments du vecteur des prix aura simplement pour conséquence de modifier les positions des relevés de prix. Par exemple, le prix du produit 3 dans le pays 1 est considéré comme le 6ème relevé de prix dans le premier cas et le 3ème dans le second cas. Mais ceci n'occulte en rien l'information fondamentale qui est : ce prix est celui du bien 3 dans le pays 1 et son coût est connu.

De façon générale, soit f_x le nombre de relevés de prix pour la position élémentaire x, c'est-à-dire le nombre de cellules contenant des prix. Le vecteur colonne des prix se présente de la façon suivante:

$$Vecteur\_prix = \begin{bmatrix} & \Pr ix\_x \\ 1 & \\ 2 & \\ \vdots & \\ f_x & \end{bmatrix}$$

5.2 Le vecteur des pays

Pour chacun des prix dans la colonne des prix, on recherche le pays correspondant. On obtient la matrice unicolonne Pays_x. De façon générale, elle se présente comme suit:

$$Vecteur\_pays = \begin{array}{c} \begin{array}{cc} & Pays\_x \end{array} \\ \begin{bmatrix} 1 & \\ 2 & \\ \vdots & \\ f_x & \end{bmatrix} \end{array}$$

L'application est faite sur l'exemple précédent. On obtient comme vecteur des pays, le vecteur suivant:

$$Vecteur\_pays = \begin{array}{c} \begin{array}{cc} & Pays\_x \end{array} \\ \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 2 \\ 5 & 2 \\ 6 & 2 \\ 7 & 3 \\ 8 & 3 \end{bmatrix} \end{array}$$

5.3 Le vecteur des produits

Pour chacun des prix dans la colonne des prix, on identifie le produit correspondant. On obtient la matrice unicolonne Produit_x. La forme générale de cette matrice unicolonne est la suivante:

$$Vecteur\_produit = \begin{array}{c} \quad Produit\_x \\ \begin{bmatrix} 1 & \\ 2 & \\ \vdots & \\ f_x & \end{bmatrix} \end{array}$$

Lorsqu'on applique cette technique à l'exemple précédent, on obtient la représentation suivante:

$$Vecteur\_produit = \begin{array}{c} \quad Produit\_x \\ \begin{bmatrix} 1 & 1 \\ 2 & 2 \\ 3 & 3 \\ 4 & 1 \\ 5 & 2 \\ 6 & 4 \\ 7 & 1 \\ 8 & 3 \end{bmatrix} \end{array}$$

La suite de l'opération consiste à construire une matrice des variables indicatrices qui opérationnalise le processus.

5.4 La matrice des Indicatrices

A partir des matrices unicolonnes (vecteurs) Pays_x et Produit_x construites, on crée les variables indicatrices Country_i et Product_j avec i allant de 1 à n_x et j allant de 1 à p_x.

Country_i prend la valeur 1 si un prix donné a été relevé dans le pays i et prend la valeur 0 sinon. Product_j est égal à 1 si le prix en question correspond au produit j et est égal à 0 dans le cas contraire (s'il s'agit de tout autre produit). Considérons l'exemple cité plus haut. On obtient les "Dummy" variables suivantes: "c" correspond aux pays et "p" correspond aux produits. "c1" indique le pays 1 et "p1" indique le produit 1.

$$Dummies\_C\&P = \begin{bmatrix} & c1 & c2 & c3 & p1 & p2 & p3 & p4 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 3 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 4 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 5 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 6 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 7 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 8 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

5.5. La matrice de régression

Après l'étape de création des matrices unicolonnes et des variables indicatrices, on crée le vecteur colonne des logarithmes népériens des prix. Pour l'exemple traité précédemment, ce vecteur s'écrit de la façon suivante:

$$Vecteur\_Ln(\text{Pr}\,ix) = \begin{bmatrix} & LnP \\ 1 & Ln\left[(1,1)\right] \\ 2 & Ln\left[(2,1)\right] \\ 3 & Ln\left[(3,1)\right] \\ 4 & Ln\left[(1,2)\right] \\ 5 & Ln\left[(2,2)\right] \\ 6 & Ln\left[(4,2)\right] \\ 7 & Ln\left[(1,3)\right] \\ 8 & Ln\left[(3,3)\right] \end{bmatrix}$$

NB: Il est bien entendu que les éléments du vecteur des prix sont des réels positifs non nuls. Il est aberrant et absurde de trouver un prix nul ou négatif.

Cette variable constitue la variable d'intérêt pour le modèle d'estimation. Lorsqu'on combine les variables indicatrices des pays et produits et la colonne représentant les logarithmes népériens des prix, on obtient la matrice M suivante:

$$M=\begin{bmatrix} & c_1 & c_2 & \cdots & \cdots & c_{p_x} & p_1 & p_2 & \cdots & \cdots & p_{n_x} & \ln P \\ 1 & & & & & & & & & & & \\ 2 & & & & & & & & & & & \\ \vdots & & & & & & & & & & & \\ \vdots & & & & & & & & & & & \\ f_x & & & & & & & & & & & \end{bmatrix}$$

Toutes les colonnes de la matrice M ne participeront pas à la régression car la technique du SWEEP OPERATOR nécessite un format d'input spécifique. La matrice M sera privée de sa première colonne (C_1) qui correspond au pays de base. En effet, cette colonne sera exclue des variables explicatives du modèle afin d'éviter une multicolinéarité. La colonne de la variable d'intérêt est mise après toutes les autres variables afin de respecter la logique du SWEEP OPERATOR. Soit la matrice X* suivante:

$$X^* = \begin{bmatrix} c_2 & c_3 & \cdots & c_{p_x} \; p_1 & p_2 & \cdots & p_{n_x} & \ln P \end{bmatrix}$$

Posons

$$X = \begin{bmatrix} c_2 & c_3 & \cdots & c_{p_x} \; p_1 & {}_2 & \cdots & p_{n_x} \end{bmatrix}$$

$$y = [\ln P]$$

On a donc

$$X^* = [X \quad , \quad y]$$

6. La régression

La méthode d'estimation utilisée est le SWEEP OPERATOR. Le SWEEP OPERATOR est un algorithme développé par Beaton en 1964. Il est conçu pour des calculs de régression économétrique. Soit une régression d'une variable d'intérêt z sur une matrice A des variables explicatives. Le SWEEP OPERATOR consiste en des transformations sur des éléments de la matrice $S=(A^{*\prime}A^{*})$ où $A^{*}=[A \quad , \quad z]$

Dans le cas du CPD, on procède au calcul de la matrice $S=(X^{*\prime}X^{*})$. Il faut rappeler que $X^{*} = [\,X \quad , \quad y]$ où y est la variable d'intérêt ln_Prix et X représente les variables explicatives définies par les variables indicatrices "Country" et "Product". Il ne s'agit pas d'une régression de la variable y sur X étant donné qu'il existe une colinéarité entre les Pj. Mais l'on sous entend un modèle sans terme constant. C'est la matrice S qui constitue la matrice Input au SWEEP OPERATOR. La matrice S s'écrit:

$$S = \begin{bmatrix} X'X & X'y \\ y'X & y'y \end{bmatrix}$$

Après les transformations matricielles du SWEEP OPERATOR, on obtient la matrice D dont on récupère les coefficients de la régression. En effet, le modèle de régression peut s'écrire sous la forme:

$$\ln P = \sum_{i=2}^{p_x} a_i c_i + \sum_{j=1}^{n_x} b_j p_j$$

Les coefficients a_i et b_j constituent les paramètres à estimer. a_i mesure le logarithme espéré du ratio de prix P_{ji}/P_{j1} . Exp (a_i) est donc la Parité de pouvoir d'achat estimée du pays i par rapport au pays de base[9]. Ainsi, les coefficients de l'estimation constituent les logarithmes népériens des Parités de pouvoir d'achat. Le vecteur $\hat{\beta}$ de la matrice D est composé des $n_x + p_x - 1$ paramètres estimés et ses $(p_x - 1)$ termes constituent les coefficients de régression relatifs aux variables indicatrices 'Country'.

$$D = \begin{bmatrix} -(X'X)^{-1} & \hat{\beta} \\ \hat{\beta} & RSS \end{bmatrix} \quad ^{10}$$

[9] Voir Annexe

[10] RSS constitue le Residual Sum of Squares et $\hat{\beta}$ est la matrice des coefficients de l'estimation

Cette expression de la matrice D montre que le SWEEP OPERATOR procède, en réalité, à une estimation des paramètres du modèle par la méthode des moindres carrés ordinaires (MCO). L'on suppose que les hypothèses d'application d'une méthode MCO sont remplies et surtout que l'hypothèse de stationnarité pour la variable dépendante du modèle est valide.

La Parité de pouvoir d'achat pour un pays C est obtenue en calculant l'exponentielle du coefficient de régression correspondant à la variable indicatrice explicative relative au pays C.

6.1 Interprétation économique

La méthode CPD représente une simple approche par la régression économétrique pour expliquer les niveaux de prix des biens et services dans différents pays. De façon fondamentale, la méthode postule que le prix observé Pij d'un bien i dans un pays j est le produit de trois composantes:

- la Parité de pouvoir d'achat PPAj ou le niveau général des prix (exprimés en monnaies nationales) du pays j par rapport aux autres pays,
- le niveau de prix P_i du bien i par rapport aux autres biens et
- un terme d'erreur V_{ij}.

La formule de base peut être écrite sous la forme:

$$P_{ij} = PPA_j * P_i * V_{ij}$$

Sous forme logarithmique, la réécriture donne:

$$\begin{aligned} Ln(P_{ij}) &= Ln(PPA_j) + Ln(P_i) + Ln(V_{ij}) \\ &= \alpha_j + \beta_i + \mu_{ij} \end{aligned}$$

où $a_j = LnPPA_j$, $B_i = LnP_i$ *et* $\mu_{ij} = LnV_{ij}$. C'est pour estimer les paramètres a_i et B_j qu'intervient l'introduction des variables indicatrices 'Country' et 'Product' dans le modèle.

Les parités de pouvoir d'achat (PPA) sont des taux permettant de convertir les prix dans une monnaie commune tout en éliminant les différences de pouvoir d'achat entre monnaies. En d'autres termes, leur utilisation permet d'éliminer l'effet des différences de niveau des prix entre pays lors de la conversion.

6.2 Estimation de données manquantes

La méthode CPD a été conçue à la base pour traiter des données de prix manquantes dans le cadre de la comparaison internationale. Le traitement est basé sur les caractéristiques du pays par rapport aux autres pays où le prix du produit est disponible. En effet, en théorie il existe deux niveaux d'estimation de données manquantes:

1. Estimation des prix moyens manquants des produits à l'intérieur d'une position élémentaire. Elle est fondée sur le modèle et possède l'avantage de fournir des prix moyens estimés en phase avec la parité élémentaire déjà obtenue.
2. Estimation des parités élémentaires manquantes dues à l'absence de données pour certains pays dans les positions élémentaires.

La deuxième estimation "relance" dans la comparaison globale, les pays "exclus" dans la section II. 2). Sur une base purement économétrique, le CPD est utilisé pour estimer des Parités de pouvoir d'achat dans un pays au niveau de positions élémentaires pour lesquelles l'information disponible n'est pas assez suffisante pour calculer directement les PPA. Ainsi, les PPA disponibles (par position élémentaire et par pays) déterminent la variable d'intérêt du modèle. Le principe fondamental du CPD qui est la dualité Pays-Produit devient une dualité Pays-Position élémentaire.

La parité de pouvoir d'achat manquante d'un pays i pour une position élémentaire k est obtenue comme produit des exponentielles des coefficients estimés correspondants aux variables indicatrices relatives au pays i et à la position k. Le modèle s'écrit sous la forme:

$$\ln ppa_{ik} = \sum_{i=2}^{P} a_i c_i + \sum_{k=1}^{M} b_k pe_k$$

P désigne le nombre de pays et M, le nombre de positions élémentaires. c_i correspond au pays i et pe_k correspond à la position élémentaire k. Pour un pays i et relativement à la position élémentaire k, on a: $c_i = 1$; $pe_k = 1$ et donc: In $ppa_{ik} = a_i + b_k$ d'où $ppa_{ik} = \exp(a_i + b_k) = \exp(a_i)^* \exp(b_k)$.

7. Application avec Visual Basic

Dans le cadre du PCI-Afrique, le calcul des Parités de pouvoir d'achat au niveau des positions élémentaires a nécessité la mise au point d'une application utilisant la méthode Country Product Dummy. Il s'agit d'un programme écrit sous le langage Visual Basic d'Excel et qui parcourt les grandes étapes de la méthode CPD jusqu'à la génération des Parités de pouvoir d'achat des pays participants au PCI-Afrique. L'interface de l'application se présente comme suit:

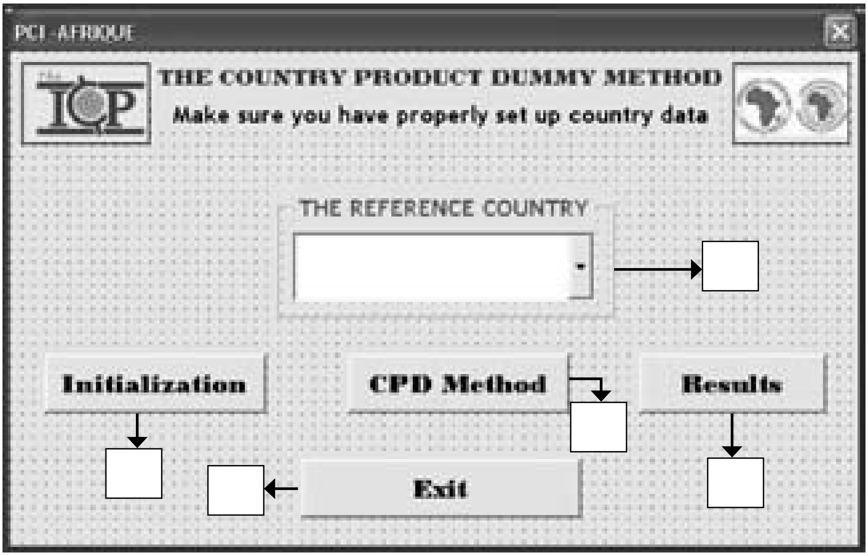

Figure 1: Maquette de l'application dans Visual Basic

On considère la matrice initiale des prix (P) sur la feuille de démarrage. du classeur Excel contenant l'application. Au départ, la colonne B est laissée vide sur la feuille. Le programme est exécuté à partir de la feuille de lancement.

- **L'étape 1** consiste à définir le pays de base pour le calcul des PPA. Lorsqu'un pays est choisi, 'ses données' sont automatiquement reportées dans la colonne B. En d'autres termes, la colonne B contient désormais les données relatives au pays de référence. On passe ensuite à l'étape d'initialisation.
- Au niveau du **bouton 2**, le programme lit les données sur les prix, par produit et par pays et supprime les lignes entièrement vides par produit. On obtient donc la matrice des prix renseignés Q sur une nouvelle feuille: 'delta'. On passe à l'étape CPD METHOD
- **Le bouton 7** constitue le cœur du programme. A cette étape, le programme invite l'opérateur à indiquer les positions élémentaires (1 à 155) dont il souhaite calculer les PPA. Pour chaque position élémentaire x, le programme retient à partir de la feuille 'delta' uniquement les

produits relatifs à cette position élémentaire. Il recopie ensuite cette matrice P(x) sur une feuille 'Posi_x'. Sur cette feuille, le programme génère la matrice P*(x), les vecteurs des prix, pays et produits. Une nouvelle feuille 'Pour_x' est créée sur laquelle le programme calcule la matrice des indicatrices, la matrice de régression et les valeurs estimées des Parités de pouvoir d'achat. On passe ensuite à l'étape RESULTS.
- **L'étape 4** correspond à la présentation des résultats. En effet les PPA sont affichées sur la feuille 'RESULTS' du classeur pour toutes les positions élémentaires spécifiées et pour l'ensemble des pays. Le bouton EXIT permet de mettre fin au programme.

A l'aide de cette application, les PPA peuvent ainsi être disponibles pour toutes les 155 positions élémentaires.

8. Conclusion

Parmi les méthodes usuelles d'estimation des PPA, la méthode CPD est appréciée pour le fait qu'elle repose sur des outils économétriques. Il est d'ailleurs recommandé par le Groupe consultatif technique du PCI pour le calcul des PPA au niveau des positions élémentaires.

Le modèle élémentaire du CPD peut être étendu pour inclure d'autres facteurs comme des caractéristiques qualitatives du produit : par exemple est-ce un produit représentatif ou non représentatif dans le pays de la collecte ? En ajoutant la variable de représentativité, le modèle obtenu est dénommé méthode Country Product Representativity Dummy (CPRD) et permet également de calculer les PPA dans une région donnée.

LE CPRD fournirait de meilleurs résultats par rapport à la méthode CPD lorsque l'information sur la représentativité des produits est disponible, fiable et complète. Et même dans ce cas, il vaut mieux s'en tenir au CPD lorsque, globalement les produits sont équi-caractéristiques.

Références bibliographiques

1. Adam, A. (2004). Estimation of the Country Product Dummy (CPD) equations, Paper presented at the Regional workshop of ICP-Africa in Yaoundé, Cameroon 2005;

2. Beaton, A.E., The Use of Special Matrix Operators in Statistical Calculus (Ed.D. thesis, Harvard University, 1964. Reprinted as Research Bulleting 64-51, Educational Testing Service, Princeton, New Jersey).

3 Diewert, E.W. (2002). Weighted Country Product Dummy Variable Regressions and Index Number Formulae mimeographed, Department of Economics, University of British Columbia, Vancouver.

4. Diewert, E.W. (2004) .On the Stochastic Approach to Linking Regions in the ICP. Paper prepared for the DECDG of the World Bank, Washington DC.

5. Noumon, C. N. Maxime (2006). Elaboration d'une méthode d'estimation des prix des enquêtes du Programme de comparaison internationale non relevés au cours du premier semestre 2005. Rapport de stage BAD/ ENSEA.

6. Rao, D.S. Prasada (2004). The Country-Product-Dummy Method: A Stochastic Approach to the Computation of Purchasing Power parities in the ICP. Paper presented at the SSHRC Conference on Index Numbers and Productivity Measurement, June 30-July 3, 2004, Vancouver, Canada.

7. Schreyer, Paul & Pilat, Dirk. (2001). Mesurer la productivité. Revue économique de l'OCDE n° 33, 2001/II.

8. The ICP Newsletter. (Juin 2006). Volume 3, Numero 2.

9. World Bank (2005), ICP Handbook. Chapter 10: The estimation of PPPs for basic headings, Rev 1.

Annexe: Quelques notions générales sur la méthode CPD)

Le CPD (Country Dummy Product) est une méthode d'estimation pour le calcul des indices spatiaux de prix, et des Parités de pouvoir d'achat (PPA), dans le cadre du Programme de comparaison internationale (PCI).

Le modèle de cette méthode peut être perçu comme la modélisation d'un certain nombre d'influences : l'influence des pays qui fournissent une estimation des PPA, et l'effet de chacun des produits qui fournissent une estimation des prix internationaux pour le produit concerné.

Soit p_{nc} le prix d'un produit n dans un pays *c* $(n = 1, 2, ...N ; c = 1, 2, ..., C)$. Le modèle statistique de base à l'origine de la CPD peut être décrit par l'équation:

$$p_{nc} = a_c b_n \mu_{nc}\ (n = 1, 2 \ldots, N\,;\ c = 1, 2, \ldots, C) \qquad \textbf{[1]}$$

b_n est l'effet du produit n; a_c est l'effet du pays c, et μ_{nc} est un effet résiduel. Le modèle postule que ces perturbations sont lognormalement distribuées avec $\ln(\mu_{nc}) \longrightarrow B(0, \sigma^2)$. Sous sa forme logarithmique, le modèle est linéaire et s'écrit:

$$\begin{aligned} E(\ln p_{nc}) &= \ln a_c + \ln b_n \\ &= \alpha_c + \gamma_n \end{aligned} \qquad \textbf{[2]}$$

Le paramètre a_c s'interprète comme le niveau général des prix dans le pays c relativement aux prix dans les autres pays concernés par la comparaison ; il est généralement exprimé par rapport à un pays de base; a_c représente dans ce cas, la PPA du pays c, exprimant le nombre d'unités monétaires du pays c, ayant le même pouvoir d'achat qu'une unité monétaire du pays de base. Dans ce cas la PPA du pays c est donnée par : $PPA_c = \exp(a_c)$

Le modèle de base du CPD peut être exprimé comme une équation de régression pour chaque prix observé correspondant à un produit *n* dans le pays *c*. Il est exprimé par l'équation [2] et peut être réécrit de la manière suivante :

$$\begin{aligned} y_{ij} = \ln p_{ij} &= \sum_{j=1}^{C} \ln a_j D_j + \sum_{i=1}^{N} \ln b_i D_i^* + \varepsilon_{ij} \\ &= \sum_{j=1}^{C} \alpha_j D_j + \sum_{i=1}^{N} \gamma_i D_i^* + \varepsilon_{ij} \end{aligned} \qquad [3]$$

p_{ij} est le prix du i^{eme} produit dans le j^{eme} pays ; ε_{ij} est un terme d'erreur aléatoire ; $(D_j)_{j=1,\ldots,c}$ et $(D_i)_{i=1,\ldots,N}$ sont deux ensembles de variables indicatrices. Chaque pays concerné par la comparaison, hormis le pays de base, est représenté par une variable indicatrice *D,* et chaque produit relatif à une position élémentaire est représenté par une variable indicatrice D* . L'équation **[2]** peut se mettre sous la forme:

$$y_{ij} = X_{ij}\beta + \varepsilon_{ij} \qquad [4]$$

avec $x_{nc} = \begin{bmatrix} D_1 & D_2 & \ldots & D_C & D_1^* & D_2^* & \ldots & D_N^* \end{bmatrix}$ et $\beta = (\alpha_1 \ \ \alpha_2 \ \ \ldots \ \ \alpha_C \ \ \gamma_1 \ \ \gamma_2 \ \ \ldots \ \ \gamma_N)$; les valeurs des indicatrices étant déterminées par les observations *ij*. En empilant les *nc* observations *(pour i = 1, ..., N et j = 1, ..., C)*, on obtient l'écriture matricielle du modèle:

$$Y = X\beta + \varepsilon \qquad [5]$$

Cette équation est un modèle de régression générale avec *NC* observations et *(N + C – 1)* variables explicatives. De plus la matrice X est de rang *N + C – 1*. Sous l'hypothèse que les perturbations sont identiquement et indépendamment distribuées, le meilleur estimateur sans biais de *ß*, et la matrice de covariance associée sont donnés par :

$$\hat{\beta} = (X'X)^{-1}X'Y \quad \text{et} \quad Var(\hat{\beta}) = \sigma^2(X'X)^{-1}$$

Etant donné ces expressions, on montre que :

$$\hat{a}_j = \frac{1}{N}\sum_{i=1}^{N}\left[\ln p_{ij} - \ln p_{i1}\right] \quad \text{ou} \quad a_j = \exp(\hat{a}_j) = \prod_{i=1}^{N}\left[\frac{p_{ij}}{p_{i1}}\right]^{1/N} \quad \text{et} \quad Var(\hat{a}_j) = \sigma^2\frac{2}{N}$$

Ainsi, en plus de l'estimation des paramètres inconnus, l'approche par la régression fournit également une estimation des écarts types pour tous les coefficients. En utilisant les résidus d'estimation pour chaque y_{ij} $\varepsilon_{ij} = y_{ij} - \hat{\alpha}_j - \hat{\gamma}_i$; *j = 1, ..., C et i = 1, ..., N*) et l'hypothèse selon laquelle les ε_{ij} sont indépendants et identiquement distribués de moyenne nulle et de variance σ^2 , un estimateur de σ^2 est donné par :

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^{C}\sum_{i=1}^{N}\varepsilon_{ij}}{(N-1)(C-1)}$$

L'estimateur $\hat{\alpha}_c$ est un estimateur sans biais de α_c et sa variance est donnée par :

$$EstVar(\hat{a}_j) = \hat{\sigma}^2\frac{2}{N}$$

Déduction des estimations des PPA: Etant donné que la PPA du pays c, est une fonction non linéaire de α_c donnée par $PPA_c = \exp(\alpha_c)$ il est d'usage, en pratique, de déduire un estimateur de PPA_c en utilisant la formule:

$$\dot{P}PA = \exp(\hat{\alpha}_c) \qquad [6]$$

Puis d'estimer sa variance par:

$$EstVar(\dot{P}PA_c) = EstVar(\hat{\alpha}_c).(\hat{\alpha}_c)^2 \qquad [7]$$

Il faut noter cependant que l'estimateur dans l'équation **[6]** n'est pas un estimateur sans biais (ESB), et la variance estimée dans l'équation **[7]** n'est qu'une approximation.

Une des limites de la méthode CPD est qu'elle n'est plus valide en présence d'autocorrélation spatiale dans les résidus, mais ce problème se résout facilement par l'utilisation d'une matrice d'autocorrélation.