



**Centre for Efficiency and Productivity Analysis**

**Working Paper Series**  
**No. 03/2004**

**Title**

The Country-Product-Dummy Method: A  
Stochastic Approach to the Computation of  
Purchasing Power Parities in the ICP

**Authors**

D.S. Prasada Rao

**Date: June, 2004**

**School of Economics**  
**University of Queensland**  
**St. Lucia, Qld. 4072**  
**Australia**

**THE COUNTRY-PRODUCT-DUMMY METHOD:  
A STOCHASTIC APPROACH TO THE COMPUTATION OF PURCHASING  
POWER PARITIES IN THE ICP**

*D.S. Prasada Rao*  
School of Economics  
University of Queensland  
Brisbane, Australia  
Email: [p.rao@economics.uq.edu.au](mailto:p.rao@economics.uq.edu.au)

June, 2004

A paper for presentation at the SSHRC Conference on Index Numbers and Productivity Measurement to be held during 30 June – 3 July, 2004 in Vancouver. The author wishes to thank Bettina Aten, Erwin Diewert, Alan Heston, Chris O'Donnell and Marcel Timmer for helpful discussions on the subject matter of this paper.

## **The Country-Product-Dummy Method: A Stochastic Approach to the Computation of Purchasing Power Parities in the ICP**

*D.S. Prasada Rao*

### Abstract

The Country-Product-Dummy (CPD) method is a regression-based method for computing spatial price index numbers and purchasing power parities (PPPs) within the International Comparison Program (ICP). It is essentially a hedonic regression model where the observed price is related to the country of origin and the characteristics of the product are represented by the commodity itself. The paper describes the unweighted and weighted CPD method and discusses a number of interesting properties of the method and argues that the CPD method provides a viable alternative to the standard index number formulae used in international comparisons. A major advantage of the CPD method is that it allows the use of a battery of powerful econometric techniques than can be used in handling data related problems including missing price data and also in computing standard errors associated with PPPs obtained.

JEL Classification: E 31 and C 19

Keywords: Purchasing Power Parities, International Prices, Country-Product-Dummy method, standard errors

### **1. Introduction**

There is a great range of methods available for purposes of aggregating price data in computing purchasing power parities (PPPs) of currencies. PPPs are essentially spatial price index numbers that provide measures of price level differences across countries or regions within a country. The Country Product Dummy (CPD) method represents a simple regression approach to measure price level differences in different countries. The CPD method was first proposed by Summers (1973) as a method for filling missing price data in the context of international comparisons. Subsequently the CPD method was used in the first few phases of the International Comparison Project (ICP) mainly as a tool for aggregating price data below the basic heading level. In the most recent phases of the ICP work, the Elteto-Koves-Szulc (EKS) method has replaced the CPD method as a procedure for aggregation and it has not been used even as a data filling procedure. There has been a revival of interest in the CPD method after the most recent work of Rao (1995) where a weighted-CPD was proposed and a link between the CPD method and the geometric variant of the Geary-Khamis method proposed in Rao (1990) was established. A number of interesting results on the CPD method have been reported in Rao (2001 and 2002) where the weighted CPD method was used in deriving a number of alternative methods for multilateral comparisons. Diewert (2002) shows how alternative specifications and weights can be used within the CPD framework to derive can be used in deriving a number of known index number formulae. Diewert (2003, 2004) has examined various properties of the CPD Method.

The main objective of this paper is to examine the unweighted and weighted versions of the CPD method and discuss a number of useful properties of the method. The paper argues that the CPD method provides a viable alternative to the standard index number methods such as

the Geary-Khamis method and the Elteto-Koves-Szulc (EKS) method. The CPD method has the additional advantage, as it is based on stochastic formulation, that it allows the use of a range of econometric tools and techniques that are not normally used in the computation of PPPs. The paper shows how the CPD method is useful in handling a number of data related problems including missing price data and also in computing standard errors associated with PPPs obtained.

## 2. Country-Product-Dummy (CPD) Method: The Basic Model

The following notation is used through the paper. Let  $p_{cn}$  represent the price of item  $n$  in country  $c$  ( $n=1,2,\dots,N$ ;  $c=1,2,\dots,C$ ).<sup>1</sup> The basic statistical model underlying the CPD method can be stated as:

$$p_{nc} = a_c b_n u_{cn} \text{ for } c = 1,2,\dots,C \text{ and } n=1,2,\dots,N \quad (1)$$

where  $a_c$  and  $b_n$  are unknown parameters to be estimated from price data and  $u_{cn}$  are independently and identically distributed random variables. In the present paper these disturbances are assumed to be lognormally distributed or that  $\ln u_{cn}$  's are normally distributed with mean 0 and a constant variance  $\sigma^2$ .

In logarithmic form the model is linear and

$$\begin{aligned} \ln p_{nc} &= \ln a_c + \ln b_n + \ln u_{cnc} \\ &= \alpha_c + \gamma_n + v_{nc} \end{aligned} \quad (2)$$

where  $v_{nc}$  are random disturbance terms which are independently and identically (normally) distributed with zero mean and variance  $\sigma^2$ . The CPD model can be seen as a simple fixed effects model where country-effects provide estimates of purchasing power parities and commodity-specific effects provide estimates of international prices.

The parameter  $a_c$  is interpreted as the general price level in country  $c$  relative to prices in other countries included in the comparison. It is possible to express  $a_c$  relative to a reference country (say country 1), then  $a_c$  represents the purchasing power parity of country  $c$  showing the number of country  $c$  currency units that have the same purchasing power as one unit of currency of country 1 or the reference country.<sup>2</sup> Therefore PPP for country  $c$  is given by:

$$PPP_c = \exp(\alpha_c) \quad (3)$$

Typically within the International Comparison Program (ICP) price data are aggregated at two levels of aggregation: (i) aggregation at the basic heading level and (ii) aggregation above the basic heading (BH) level. Basic heading is defined as the lowest level of commodity classification at which expenditure weights are available from household budget surveys. The main features of the data available for aggregation at these two levels are listed below.

---

<sup>1</sup> These prices may be considered as a single price observation. It is possible to consider a case where there are several price quotations available for each commodity in each country, a case considered in Diewert (2004). It is also possible to consider the single price observation as an annual average price for the commodity. These are considered in subsequent sections of the paper.

<sup>2</sup> If the model uses price data over time, then the parameter  $a_t$  represents a price index for period  $t$  with respect to a base period.

- Price data at the basic heading level may be incomplete. This is the case where not all countries provide price data for all the items listed in the price surveys. In fact price data at this level can be quite sparse.
- It is possible that data on individual price quotations, instead of just national average prices forming a single price quote, are available.
- No quantity data are available at the BH level. So no budget share weights can be attached to the price quotations.
- At the level above the basic heading, price data are usually complete. Prices refer to commodity subgroups (basic headings) and expenditures are available.

The main objective of the paper is to examine and describe the role of CPD at these two levels of aggregation and examine the main properties of the purchasing power parities resulting from the application of the unweighted and weighted CPD methods.

### 3. Unweighted CPD and Aggregation below the Basic Heading Level

The main distinguishing feature here is that there are no weights to be attached to price data. Price indices at the basic heading level are similar to the *elementary indices* used in the context of the consumer price index. Properties of the CPD method can be examined for the cases where all the items are priced in all the countries and the case where not all the items are priced in all the countries. The first case corresponds to the case where the price tableau – table of prices of all items in all countries – is complete.

#### 3.1 Complete Price Tableau

This is the case where  $p_{nc}$  are observed and reported for all  $n$  in all countries. Since there are no weights available under the BH level, the unknown parameters  $\alpha_c$  and  $\gamma_n$  can be estimated using simple unweighted or ordinary least squares by minimizing

The first order conditions for optimization with respect to  $\alpha_c$  and  $\gamma_n$  lead to the following system of  $C+N$  equations in as many unknowns.

$$\begin{aligned}\alpha_c &= \frac{1}{N} \sum_n \ln p_{nc} - \sum_n \gamma_n \quad \text{for } c = 1, 2, \dots, C \quad \text{and} \\ \gamma_n &= \frac{1}{C} \sum_c \ln p_{nc} - \sum_c \alpha_c \quad \text{for } n = 1, 2, \dots, N\end{aligned}\tag{4}$$

This system can be solved by imposing a linear restriction on the unknown parameters. For example, if  $\alpha_1 = 0$  is the restriction imposed, it can be easily shown that, for each  $c=2, \dots, C$

$$\alpha_c = \frac{1}{N} \sum_{n=1}^N [\ln p_{nc} - \ln p_{n1}] \quad \text{or} \quad a_c = \exp(\alpha_c) = \prod_{n=1}^N \left[ \frac{p_{nc}}{p_{n1}} \right]^{1/N}\tag{5}$$

Using the solution in (5), comparisons of price levels between two countries  $c$  and  $d$ , represented by  $PPP_{cd}$  can be derived as:

$$PPP_{cd} = \frac{a_d}{a_c} = \prod_{n=1}^N \left[ \frac{p_{nd}}{p_{nc}} \right]^{1/N}\tag{6}$$

The price level comparison,  $PPP_{cd}$ , in (6) satisfies the transitivity property and it is identical to the EKS (Elteto-Koves-Szulc) index used in the OECD and Eurostat comparisons for prices at the basic heading level.<sup>3</sup> The expression provided here is a special case of the expression given in Diewert (2004) where he allows for several price quotations for each commodity in each country.

### 3.1.2 The CPD Regression Model

The simple CPD model can be expressed as a regression equation for each price observation corresponding to commodity  $n$  in country  $c$ . The basic model,

$$\ln p_{nc} = \alpha_c + \gamma_n + v_{nc}$$

can be written as

$$y_{nc} = \ln p_{nc} = \alpha_1 D_1 + \alpha_2 D_2 + \dots + \alpha_C D_C + \eta_1 D_1^* + \eta_2 D_2^* + \dots + \eta_N D_N^* + u_{nc} \quad (7)$$

where  $D_c$  ( $c=1,2,\dots,C$ ) and  $D_n^*$  are respectively country and commodity dummy variables. Equation (7) can be written as

$$y_{nc} = x_{nc} \beta + v_{nc}$$

where  $x_{nc} = [D_1 \ D_2 \ \dots \ D_C \ D_1^* \ D_2^* \ \dots \ D_N^*]$  and  $\beta = (\alpha_1 \ \alpha_2 \ \dots \ \alpha_C \ \gamma_1 \ \gamma_2 \ \dots \ \gamma_N)'$  where the values of the dummy variables are determined at the observation  $nc$ .

Stacking all the NC observations (for  $c=1,2,\dots,C$  and  $n=1,2,\dots,N$ ), the model can be written as

$$y = X\beta + v \quad (8)$$

Model (8) is a general regression model with NC observations and  $(N+C)$  explanatory variables. It is easy to see that matrix  $X$  is of rank  $(N+C-1)$  as the first  $C$  dummy variables and the last  $N$  dummy variables sum to the same vector with elements equal to 1. Therefore, elements of parameter vector  $\beta$  are not readily identified. A single linear restriction will permit the estimation of parameters. In this paper, the restriction  $\alpha_1 = 0$  is imposed. This restriction implies that the currency of country 1 is taken as the numeraire and PPPs of all other countries are expressed relative to the currency of country 1.

The model under this restriction is essentially the same as (8) except that the first column of  $X$  and first element of  $\beta$  are dropped from the model. The restricted model is then given by

$$y = X^* \beta^* + v \quad (9)$$

where  $X^*$  is the same matrix as  $X$  but without the first column and  $\beta^*$  is the same vector as  $\beta$  without the first element  $\alpha_1$ .

Under the assumption of independently and identically distributed random disturbances, the best linear unbiased estimator of  $\beta^*$  and the associated covariance matrix are given by

$$\hat{\beta}^* = \left( X^{*'} X^* \right)^{-1} X^{*'} y \quad \text{and} \quad \text{Var}(\hat{\beta}^*) = \sigma^2 \left( X^{*'} X^* \right)^{-1} \quad (10)$$

<sup>3</sup> See Rao (2001) for a description of the EKS method for aggregation at the basic heading level.

Using the special structure of the matrix  $X^*$ , consisting of various dummy variables as columns, it can be shown that

$$\begin{aligned}
 X^{*'} X^* &= \left[ \begin{array}{cccc|cccc}
 N & 0 & \dots & 0 & 1 & 1 & 1 & 1 & \dots & 1 \\
 0 & N & \dots & 0 & 1 & 1 & 1 & 1 & \dots & \dots \\
 & & & & \cdot & & & & & \cdot \\
 & & & & \cdot & & & & & \cdot \\
 0 & \dots & \dots & N & 1 & 1 & \dots & \dots & \dots & 1 \\
 \hline
 1 & 1 & \dots & 1 & C & 0 & \dots & \dots & \dots & 0 \\
 1 & 1 & & 1 & 0 & C & \dots & \dots & \dots & 0 \\
 1 & 1 & & \cdot & & & & & & \\
 1 & 1 & & \cdot & & & & & & \\
 \cdot & \cdot & & \cdot & & & & & & \\
 \cdot & \cdot & & \cdot & & & & & & \\
 1 & \cdot & \dots & 1 & 0 & \cdot & \dots & \dots & \dots & C
 \end{array} \right] \\
 &= \begin{bmatrix} NI_{C-1} & J_{C-1 \times N} \\ J_{N \times C-1} & CI_N \end{bmatrix}
 \end{aligned} \tag{11}$$

where  $I$  is an identity matrix of a given order and  $J$  is a rectangular matrix (of appropriate order) whose elements are all equal to 1.

Using the special structure of  $X^{*'} X^*$  and Theorem 8.3.5 of Graybill (1983), it can be shown that

$$\left( X^{*'} X^* \right)^{-1} = \begin{bmatrix} \frac{1}{N} I_{C-1} + \frac{1}{N} J_{C-1 \times C-1} & \frac{1}{NC - (N+C)} J_{C-1 \times N} \\ \frac{1}{NC - (N+C)} J_{N \times C-1} & \frac{1}{C} I_N + \frac{C-1}{CN} J_{N \times N} \end{bmatrix}$$

Given this expression, it can be shown that

$$\hat{\alpha}_c = \frac{1}{N} \sum_{n=1}^N [\ln p_{nc} - \ln p_{n1}] \quad \text{or} \quad a_c = \exp(\hat{\alpha}_c) = \prod_{n=1}^N \left[ \frac{p_{nc}}{p_{n1}} \right]^{1/N}$$

and

$$\text{Var}(\hat{\alpha}_c) = \sigma^2 \frac{2}{N}$$

In addition to the estimates of the unknown parameters, the regression approach also provides estimated standard errors for all the coefficients. Using the least squares residuals defined for each observation  $y_{nc}$  as

$$e_{nc} = y_{nc} - \hat{\alpha}_c - \hat{\gamma}_n \text{ for } c = 1, 2, \dots, C \text{ and } n = 1, 2, \dots, N$$

Since all the  $v_{nc}$ 's are assumed to be independently and identically distributed with mean zero and variance  $\sigma^2$ , an unbiased estimator of  $\sigma^2$  is given by

$$\hat{\sigma}^2 = \frac{\sum_{c=1}^C \sum_{n=1}^N e_{nc}^2}{CN - (C + N - 1)}. \quad (12)$$

Estimator  $\hat{\alpha}_c$  is an unbiased estimator of  $\alpha_c$  and its variance<sup>4</sup> is given by

$$\text{Est.V}(\hat{\alpha}_c) = \frac{2}{N} \hat{\sigma}^2 \quad (13)$$

It is possible to derive the estimated variance of each element of  $\hat{\beta}$ .

An interesting feature of the expression in (13) is that the estimated standard errors of the  $\hat{\alpha}_c$ 's are the same for all  $c=1, 2, \dots, C$ . Therefore, there is little that distinguishes the estimators of the logarithmic price level differences in terms of standard errors. This is mainly due to the fact that all commodities are priced in all the countries and disturbances in equation,  $v_{nc}$ , are independently and identically distributed.

It should be noted that the estimated variance of  $\hat{\alpha}_c$  in (13) differs from the variance expression reported in equation (17) of Diewert (2004). This difference is essentially due to the different normalizations used by Diewert (2004) and the one used in this paper. A more appropriate comparison is between the standard errors of the estimates of the PPP between currencies of countries  $c$  and  $d$ . This is given by

$$\ln \text{PPP}_{cd} = \hat{\alpha}_d - \hat{\alpha}_c$$

and its estimated variance can be shown to be

$$\text{Est. Var}(\hat{\alpha}_d - \hat{\alpha}_c) = \frac{2}{N} \hat{\sigma}^2 \quad (14)$$

Note that covariance between  $\hat{\alpha}_d$  and  $\hat{\alpha}_c$  is equal to  $(1/N) \hat{\sigma}^2$ . This expression is essentially the same as the expression resulting from Diewert (2004).<sup>5</sup>

### 3.1.3 Estimation of Purchasing Power Parities of Currencies

Since  $\text{PPP}_c$ , purchasing power parity of currency of country  $c$ , is a non-linear function of  $\alpha_c$  given by

$$\text{PPP}_c = \exp(\alpha_c).$$

It is a standard practice to derive an estimator of  $\text{PPP}_c$  using

$$\text{PPP}_c^\hat{} = \exp(\hat{\alpha}_c), \quad (15)$$

and to estimate its variance by

$$\text{Est Var}(\text{PPP}_c^\hat{}) \approx \text{Est Var}(\hat{\alpha}_c) \bullet (\hat{\alpha}_c)^2 \quad (16)$$

<sup>4</sup> Note that this expression is a special case of the expression (17) in Diewert (2004) which is a

<sup>5</sup> Diewert's expression and the expression in (13) differ by a factor  $(C-1)/C$ .



However, estimator in (15) is not an unbiased estimator and the estimated variance in (16) is only an approximation.

Under the assumption that  $v_{nc}$ 's are normally distributed, or the original disturbances  $u_{nc}$ 's in (1) are lognormally distributed, it is possible to derive *uniformly minimum variance unbiased estimators (UMVUE)* of  $PPP_c$ 's using the approach followed in Prasada Rao and Selvanathan (1992). These details with some numerical illustrations are also available in Selvanathan and Rao (1994).

### 3.1.4 The CPD Model with spatially auto-correlated error structure

The CPD model described in equations (1) and (2) and the subsequent econometric estimation of parameters is based on the assumption that the disturbances,  $v_{nc}$ 's, are assumed to have a mean equal to zero and have the same variance and are not autocorrelated. Under these assumptions least squares estimates of the parameters,  $\alpha_c$ , are the best linear unbiased estimators. Generally it is possible to test and adjust for the presence of heteroscedastic disturbances, a topic that is dealt with in the next section where weighted CPD approach is described in some detail.

In the context of the CPD model, the possible presence of autocorrelation among disturbances across countries for a specific commodity, the presence of spatial autocorrelation, has not been investigated. Aten (1996) has demonstrated the existence of spatial autocorrelation and analysed the presence of patterns in relative price structures across geographical regions. Presence of spatially autocorrelated disturbances implies that the use of ordinary least squares no longer provides the most efficient estimates of the parameters involved.

In this paper we briefly touch upon some of the main features of the generalised CPD model with spatially autocorrelated models. Before embarking on the testing and specification issues, it is useful to find an interpretation of the disturbances in the CPD model.

From (2), the model is

$$y_{nc} = \ln p_{nc} = \alpha_c + \gamma_n + v_{cn}$$

which can be rewritten as

$$\begin{aligned} v_{nc} &= y_{nc} - \alpha_c - \beta_n \\ &= \ln p_{nc} - \alpha_c - \beta_n \\ &= \ln \left[ \frac{p_{nc} / PPP_c}{b_n} \right] \end{aligned} \quad (17)$$

From equation (16) it is evident that the disturbance term for n-th commodity in country c is the logarithm of domestic price of commodity n,  $p_{nc}$ , converted to a common currency unit using PPPc, expressed relative to the international average price of commodity n. Thus the disturbance term provides a measure of price levels relative to international average prices, for each commodity in each country.

The assumption that  $v_{nc}$ 's are identically and independently distributed over all c and n is probably too restrictive. It is possible that domestic prices relative to some international average price may exhibit correlation with the relative price in a neighbouring country thus exhibiting spatial autocorrelation. This phenomenon was investigated by Aten (1996) who found significant spatial autocorrelation for a number of commodity groups.

The CPD regression model can be modified to accommodate the presence of spatially autocorrelated errors. Suppose the model after imposing the identifying restriction  $\alpha_1 = 0$  is written as:

$$y_{nc} = x_{nc} \beta + v_{nc}$$

Suppose all the C observations for each commodity are stacked into a vector, then the model is given by

$$y_n = x_n \beta + v_n \quad \text{for } n = 1, 2, \dots, N \quad (18)$$

where

$$y_n = [y_{n1} \ y_{n2} \ \dots \ y_{nC}]';$$

$$x_n = \begin{bmatrix} x_{n1} \\ x_{n2} \\ \vdots \\ x_{nC} \end{bmatrix} \quad \text{and} \quad v_n = \begin{bmatrix} v_{n1} \\ v_{n2} \\ \vdots \\ v_{nC} \end{bmatrix}$$

Suppose the vector disturbances  $v_n$  is spatially autocorrelated with the following structure:

$$v_n = \rho_n W v_n + \varepsilon_n \quad \text{with } \varepsilon_n \sim (0, \sigma^2 I_C) \quad (19)$$

Here W is a square matrix of order C x C representing spatial weights.<sup>6</sup> Under the assumption that  $\rho_n$  is such that  $[I - \rho_n W]^{-1}$  exists, the model in (18) can be transformed using  $[I - \rho_n W]$  provided  $\rho_n$  is known. Under this assumption, the transformed model is given by

$$[I - \rho_n W] y_n = [I - \rho_n W] x_n \beta + \varepsilon_n \quad (20)$$

where  $\varepsilon_n \sim (0, \sigma^2 I_C)$ .

Stacking all the observations over all the commodities,  $n=1, 2, \dots, N$ , the model can be written as

---

<sup>6</sup> For more details on the choice of weights in modelling spatial autocorrelation, see Aten (1996).

$$y^* = X^* \beta + \varepsilon$$

where elements of  $\varepsilon$  are independently and identically distributed. Then the best linear unbiased estimator of  $\beta$  is given by

$$\begin{aligned} \hat{\beta} &= \left( X^{*'} X^* \right)^{-1} X^{*'} y^* \\ &= \left[ \sum_{n=1}^N x_n' (I - \rho_n W)' (I - \rho_n W) x_n \right]^{-1} \left[ \sum_{n=1}^N x_n' (I - \rho_n W)' (I - \rho_n W) y_n \right] \end{aligned} \quad (21)$$

and the estimated variance-covariance matrix of the estimated parameters is given by

$$\text{Est. Cov.}(\hat{\beta}) = \hat{\sigma}^2 \left[ \sum_{n=1}^N x_n' (I - \rho_n W)' (I - \rho_n W) x_n \right]^{-1} \quad (22)$$

The derivation of estimators of  $\alpha_c$  and  $\gamma_n$  is fairly straightforward if  $\rho_n$  ( $n=1,2,\dots,N$ ) are all known.

There are two general issues to be considered here. First, it is necessary to test for the presence of spatially autocorrelated errors and secondly, the estimation procedure above requires estimates of  $\rho_n$ 's. These two issues are considered briefly below.

### ***Testing for the presence of Spatial Autocorrelation***

There are a number of standard test procedures available. In this discussion here, a simple test based on Moran's I-statistic, following Aten (1996), is described here. First run the simple least squares on the CPD model and derive the least squares residuals

$$\hat{v}_{nc} = y_{nc} - \hat{\alpha}_c - \hat{\gamma}_n.$$

Then using the residuals along with a pre-specified spatial autocorrelation matrix  $W$ , the Moran's I-statistic is defined as

$$I = \frac{C \sum_{c=1}^C \sum_{d=1}^D w_{cd} (\hat{v}_{nc} - \bar{\hat{v}}_n) (\hat{v}_{nd} - \bar{\hat{v}})}{S_0 \sum_{c=1}^C (\hat{v}_{nc} - \bar{\hat{v}})^2} \quad (23)$$

where  $S_0 = \sum_c \sum_d w_{cd}$

Under the null hypothesis that the disturbances are not spatially autocorrelated,  $I$  is asymptotically normally distributed with a known mean and variance. Based on this a standard test based on normal distribution can be used.

### *Estimation of $\rho_n$*

It is possible to derive maximum likelihood estimators of  $\beta$  and  $\rho_n$  ( $n=1,2,\dots,N$ ) under some assumptions regarding the distribution of  $\varepsilon_n$ . Otherwise, an iterated least squares method can be employed. For this, start with some initial values of  $\rho_n$ , denoted by  $\rho_n(0)$ , for  $n=1,2,\dots,N$ . Then let

$$\tilde{\Omega}_n(0) = [I - \rho_n(0)W]$$

and derive estimates of the unknown parameters

$$\tilde{\beta}(0) = \left[ \sum_{n=1}^N x_n' \tilde{\Omega}_n(0)' \tilde{\Omega}_n(0) x_n \right]^{-1} \sum_{n=1}^N x_n' \tilde{\Omega}_n(0)' \tilde{\Omega}_n(0) y_n$$

Once the parameters are estimated, define least squares residuals as

$$\hat{v}_n(0) = y_n - x_n \tilde{\beta}(0).$$

Using the estimated residuals, the next step values of  $\rho_n$  are derived using:

$$\tilde{\rho}_n(1) = \frac{\hat{v}_n(0)' \tilde{\Omega}_n(0) \hat{v}_n(0)}{\hat{v}_n(0)' \tilde{\Omega}_n(0)' \tilde{\Omega}_n(0) \hat{v}_n(0)} \quad (24)$$

The iterative procedure is continued until it converges. Various results on the convergence are available in the literature. Rao (2001) has numerical illustration showing estimated purchasing power parities under the assumption of spatially autocorrelated structures for aggregation above the basic heading parities.

The role of spatially autocorrelated errors in the context of deriving parities at the basic heading level using complete or incomplete price data needs to be further explored. Given the strong spatial autocorrelation present for most commodity groups reported in Aten (1996), using a CPD model with spatially autocorrelated errors appears to be a natural next step.<sup>7</sup>

### **3.2 CPD with Incomplete Price Tableau**

The case of complete price data at the item level is rarely encountered in practice. In fact the general experience in international comparison exercises is that only a few items are priced in each of the participating countries resulting in a rather sparse price tableau. This section examines the nature and role of the CPD method in this context and it is contrasted with the alternative aggregation method based on variants of the EKS method used in international comparisons at the OECD and Eurostat.

#### **3.2.1 CPD framework for incomplete price data**

---

<sup>7</sup> However, any empirical application of this method requires raw price data collected as a part of the ICP. The author of this paper has not been privileged enough to have access to such data.

Let  $N_c$  represent the set of items, as well as the number of items, priced in country  $c$ .<sup>8</sup> Obviously  $N_c \leq N$ . Similarly let  $C_n$  represent the set of countries, and number of countries, in which item  $n$  is priced. Again  $C_n \leq C$ . Further both  $N_c$  and  $C_n$  are both assumed to be strictly positive. This implies that there is at least one item priced in each country and each commodity is priced in at least one country.<sup>9</sup>

In this case, the basic CPD model discussed in Section 3.1 can be applied only for those price observations collected in each country. Thus

$$y_{nc} = \ln p_{nc} = \alpha_c + \gamma_n + v_{nc} \quad (25)$$

is applied for all observations such that  $n \in N_c$  for  $c = 1, 2, \dots, C$ .

The CPD model makes use of all the price observations in the data. Let  $K$  denote the total number of price observations, then

$$\sum_{c=1}^C N_c = \sum_{n=1}^N C_n = K$$

As before, model (22) can be expressed in a standard regression format as:

$$y_{nc} = \sum_{c=1}^C \alpha_c D_c + \sum_{n=1}^N \gamma_n D_n^* + v_{nc} \quad \text{for all } n \in N_c \text{ for } c=1, 2, \dots, C$$

To facilitate identification of the parameters in this regression model, the simple linear restriction  $\alpha_1 = 0$  is imposed. Under this assumption of independently and identically distributed error terms the best linear unbiased estimator of the parameters is given by

$$\hat{\beta} = \begin{bmatrix} \hat{\alpha} \\ \hat{\gamma} \end{bmatrix} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \quad (26)$$

where  $\hat{\alpha} = (\alpha_2 \ \alpha_3 \ \dots \ \alpha_C)'$  and the matrix  $(\mathbf{X}'\mathbf{X})$  is given by

---

<sup>8</sup> This is used just to simplify the notation and should not cause any confusion. With the aim of keeping the exposition in the paper brief and simple, in this only one price quotation is assumed to be available for each commodity priced. A more general case with possible multiple price observations is considered by Diewert (2004),

<sup>9</sup> In fact, Diewert (2004) notes that if an item is priced in only one country, that price will have no effect on the estimates of  $\alpha_c$  and hence the purchasing power parities.

$$\left[ \begin{array}{cccc|cccc}
N_2 & 0 & \cdot & \cdot & 0 & d_{21} & d_{22} & \cdot & \cdot & d_{2N} \\
0 & N_3 & \cdot & \cdot & 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
0 & \cdot & \cdot & \cdot & N_c & d_{c1} & d_{c2} & \cdot & \cdot & \cdot & d_{cN} \\
\hline
d_{21} & \cdot & \cdot & \cdot & d_{c1} & C_1 & 0 & \cdot & \cdot & \cdot & 0 \\
d_{22} & \cdot & \cdot & \cdot & \cdot & 0 & C_2 & \cdot & \cdot & \cdot & 0 \\
0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
d_{2N} & \cdot & \cdot & \cdot & d_{cN} & 0 & \cdot & \cdot & \cdot & \cdot & C_N
\end{array} \right]; \text{ and } X'y = \left[ \begin{array}{c}
\sum_{n \in N_2} y_{2n} \\
\cdot \\
\cdot \\
\sum_{n \in N_c} y_{cn} \\
\sum_{c \in C_1} y_{c1} \\
\cdot \\
\cdot \\
\sum_{c \in C_N} y_{cn}
\end{array} \right]$$

where  $d_{cn} = 1$  if commodity  $c$  is priced in country  $n$ ; and  $= 0$  otherwise.

The top right-hand corner matrix is like an indicator/incidence matrix showing whether or not a commodity is priced in a particular country. By definition

$$\sum_{n=1}^N d_{cn} = N_c$$

$$\sum_{c=1}^C d_{cn} = C_N$$

### 3.2.2 Feasibility of the CPD Method

In the case where the price tableau is incomplete, it is not obvious that the CPD method is feasible in the sense that the matrix  $(X'X)^{-1}$  in (23) may not exist. A few comments in the form of propositions are made without formal proofs.<sup>10</sup>

**Proposition 1:** The CPD Method is feasible and  $(X'X)$  is non-singular when the price tableau is complete.

Given incomplete price data, non-singularity of  $(X'X)$  and the feasibility of the CPD are not always guaranteed. Then it is useful to identify some conditions under which the CPD approach provides unique solutions to all the unknown  $\alpha$ 's and  $\gamma$ 's.

**Proposition 2:** A sufficient condition for  $(X'X)$  in (23) to be non-singular is that there is at least one country where all the commodities are priced, i.e., there is a country  $c$  with  $N_c = N$ .

<sup>10</sup> Some of the proofs are straightforward but others require easy but lengthy proofs.

In order to provide an intuitive background to the next proposition, which provides a necessary and sufficient condition, it is useful to describe the price data availability using graph theoretic concepts.

Let  $G$  denote a *graph* with countries as *vertices*. In this case the graph will have  $C$  vertices representing all the countries. Connect country  $c$  with country  $d$  using an “edge” or line if there is an item that is commonly priced in both countries. Connect two countries  $c$  and  $d$  if there exists an  $n$  such that  $d_{cn} = 1 = d_{dn}$ . If two countries are connected by an edge, then the price data can form the basis for a direct price comparison across the two countries.

Construct graph  $G$  by connecting all eligible pairs of countries based on the observed price data. Let this graph be known as the “*price-data*” graph.

A graph  $G$  is said to be *connected* if any pair of countries  $a$  and  $b$  can be linked through a sequence of countries that are all connected by edges. If price-data graph is connected, it is possible to make a price level comparison between any pair of countries ( $a,b$ ) either directly or indirectly through chained comparisons.<sup>11</sup>

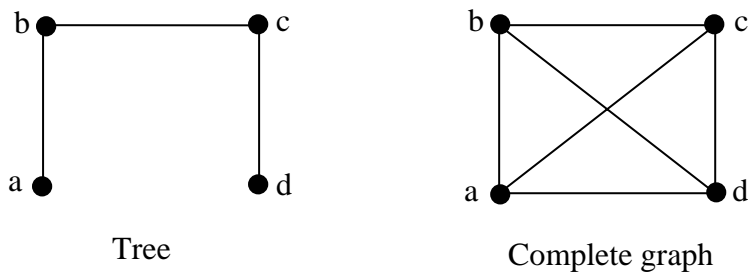
**Proposition 3:** A necessary and sufficient condition for  $(X'X)$  in (23) to be non-singular is that the price-data graph, associated with the incomplete price tableau, is connected.

It is quite easy to see the necessary condition part of the proposition. Suppose the price-data graph is not connected, then the  $C$  countries in the comparison can be divided into two groups  $C_1$  and  $C_2$  such that no items priced in group 1 countries are priced any of the group 2 countries. If these two country groups have no common items priced then there is no basis for meaningful price comparisons across these country-groups and in this case the CPD method breaks down. Proof of sufficiency is more complicated and is not included here.

**Comment on Proposition 3:** Proposition 3 shows that it is feasible to construct price comparisons when the price-data graph is connected. But comparisons can be very sensitive to the degree of connectivity or the degree of overlap of price data underlying a price comparison between two countries that are connected by an edge. Similarly if the price-data graph is based on a spanning tree then such comparisons may be weaker. The following graphs show two graphs with 4 countries.

---

<sup>11</sup> Note that “spanning trees” used by Hill (1999) are all examples of connected graphs.

**Figure: Examples of graphs**

The complete graph implies that all direct binary comparisons are feasible on the basis of observed price data where as in the case of tree-graph, price data has no over lap for countries 1 and 3, 3 and 4 and for 2 and 4.

It should be possible to relate the “degree of connectivity” of price data to the estimated standard errors resulting from the application of the CPD method whenever it is applicable.

### ***3.2.3 CPD Fillers for Unobserved Prices and the Application of the EKS Method***

The main motive behind the formulation of CPD method by Summers (1973) was to provide a statistical tool that can be used in filling gaps in price data and prepare a complete price tableau. Once all the holes are filled, the resulting data can be used in computing purchasing power parities as simple unweighted geometric averages of price relatives, which is same as the EKS procedure used for aggregation at the basic heading level.

In situations with incomplete price data, there are two possible ways of adapting the EKS method. The aim of this section is to demonstrate how the use of CPD in the presence of price gaps is superior to the EKS and its variants that are commonly used. Two such variants are considered here.

#### **Variant of EKS Method – I**

Under this variant of the EKS method, a bilateral comparison between two countries c and d is derived using price data for all those commodities that are priced in both countries. Thus

$$PPP_{cd} = \prod_{n \in N_c \cap N_d} \left[ \frac{P_{nd}}{P_{nc}} \right]^{1/n_{cd}} \quad (27)$$

where  $n_{cd}$  is the number of items that are commonly priced in countries c and d.

In cases where there are no commonly priced items, PPP in (27) not defined and, therefore, derived using an arbitrarily selected link country to link c and d.<sup>12</sup>

<sup>12</sup> Existence of such a link can be guaranteed if the price data graph is connected.



Since PPPs in (27) are not transitive, the EKS method is used in deriving

$$\text{EKS - PPP}_{cd} = \prod_{a=1}^C [\text{PPP}_{c,a} \bullet \text{PPP}_{a,d}]^{1/C} \quad (28)$$

Indices in (28) are transitive and are commonly used as PPP's at the basic heading level.

The main drawback of this procedure is that it does not utilize price data available efficiently and where indirect comparisons are derived choice of the link country is arbitrary. In this case, the CPD method provides a superior alternative in that it uses all the price data available simultaneously and no arbitrary choice regarding a link need to be made.

### **Variant of EKS Method - II**

Another approach to the problem of missing data is to use prices predicted from the CPD method to fill all the gaps and then employ the standard EKS method to aggregate price data leading to PPPs of currencies. While this approach appears logical and is consistent with the general thrust of the original formulation of the CPD method by Summers (1973), it will be shown below that such an approach does not add any value and that PPPs from the CPD gap-filled data are in fact the same as the PPPs resulting from the CPD method applied to the incomplete price tableau. This result follows from the following two propositions.

**Proposition 4:** If the price tableau is complete then the CPD and EKS parities are the same.

This is a well-established result, the basic proof of this was presented in the earlier sections.<sup>13</sup>

**Proposition 5:** Estimators of logarithms of PPP's ( $\alpha$ 's) from the CPD applied to the incomplete price tableau are identical to the estimates from the CPD applied to the complete tableau with gaps in price filled using CPD on incomplete price tableau.

Proof of this proposition follows from the properties of the basic regression model. Following equation (22), let the regression model for the incomplete tableau, after setting  $\alpha_1 = 0$ , be given by

$$y = X\beta + v \quad (29)$$

based on  $K = \sum_{c=1}^C N_c$  price observations. Estimator of  $\beta$  is given by

$$\hat{\beta} = (X'X)^{-1} X'y.$$

Let  $X^*$  be a matrix of values of dummy variables of countries and items for which price data are not available. Thus  $X^*$  has  $(NC-K)$  rows and  $(C-1+N)$  columns. The missing price data can be predicted as:

$$\hat{y}^* = X^* \hat{\beta}.$$

---

<sup>13</sup> Ferrari and Riani (1996) and Diewert (2004) also provide proofs of this statement.

Now run a CPD regression by augmenting the predicted prices for unobserved prices in model (29). The new regression model with augmented observations is:

$$\begin{pmatrix} y \\ y^* \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \mathbf{X}^* \end{pmatrix} + \begin{pmatrix} \mathbf{v} \\ 0 \end{pmatrix} \quad (30)$$

leading to the model

$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\beta + \tilde{\mathbf{v}}$$

and the least squares estimator for the pooled model is

$$\tilde{\beta} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}'\tilde{\mathbf{y}}$$

where  $(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}) = (\mathbf{X}'\mathbf{X} + \mathbf{X}^*\mathbf{X}^*)$  and  $\tilde{\mathbf{X}}'\tilde{\mathbf{y}} = \mathbf{X}'\mathbf{y} + \mathbf{X}^*\mathbf{y}^*$ .

Therefore, the least squares estimator from the augmented data is

$$\tilde{\beta} = \left[ \mathbf{X}'\mathbf{X} + \mathbf{X}^*\mathbf{X}^* \right]^{-1} \left( \mathbf{X}'\mathbf{y} + \mathbf{X}^*\mathbf{y}^* \right) = \left[ \mathbf{I} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}^*\mathbf{X}^* \right]^{-1} (\mathbf{X}'\mathbf{X})^{-1} \left( \mathbf{X}'\mathbf{y} + \mathbf{X}^*\mathbf{y}^* \right)$$

Using the fact that  $\hat{\mathbf{y}}^* = \mathbf{X}^*\hat{\beta} = \mathbf{X}^*(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ , it follows that  $\tilde{\beta} = \hat{\beta}$ .

This implies that the estimator of  $\beta$  is the same whether the incomplete price tableau or the filler-augmented price tableau are used.

Propositions (4) and (5) combined imply that the use of the EKS method on filler-augmented price data results in the same parities as the parities from the CPD method applied to the incomplete price tableau.

In summary, the discussion thus far suggests that the best course of action in the presence of gaps in price data is to simply use the CPD regression on the available price data. Measures of reliability in the form of standard errors, reflecting the sparseness of data, are also available from the CPD approach.

#### 4. Weighted CPD and aggregation above the Basic Heading Level

All the models described in Section 3 share the common feature that all the observations are accorded the same importance in the estimation of the parameters of the CPD model. There are many practical situations where use of weights can lead to improved and more efficient estimators. This section is devoted to the treatment of weighted CPD model with some applications.

It is useful to start with the CPD model where the parameter for country 1 is restricted to zero as a way of identifying the parameters. Then the basic CPD model

$$\ln p_{nc} = \alpha_c + \gamma_n + v_{nc}$$

can be written as

$$y_{nc} = \ln p_{nc} = \alpha_2 D_2 + \dots + \alpha_C D_C + \eta_1 D_1^* + \eta_2 D_2^* + \dots + \eta_N D_N^* + u_{nc} \quad (31)$$

where  $D_c$  and  $D_n^*$  are respectively country and commodity dummy variables. Equation (7) can be written as

$$y_{nc} = x_{nc} \beta + v_{nc}$$

where  $x_{nc} = [D_2 \quad D_C \quad D_1^* \quad D_2^* \quad D_N^*]$  and  $\beta = (\alpha_2 \quad \alpha_C \quad \gamma_1 \quad \gamma_2 \quad \gamma_N)'$  where the values of the dummy variables are determined at the observation  $nc$ .

Stacking all the NC observations (for  $c=1,2,\dots,C$  and  $n=1,2,\dots,N$ ), the model can be written as

$$y = X\beta + v$$

Here  $y$  is a vector of order  $NC \times 1$  and  $X$  is a matrix of observations (in the form of dummy variables) of order  $NC \times (C-1+N)$  and  $\beta$  is a column vector of order  $(C-1+N) \times 1$ .

The approach used in Section 3 essentially derives estimators of  $\beta$  through the minimization of  $\sum_{c=1}^C \sum_{n=1}^N [y_{nc} - \alpha_c - \gamma_n]^2$  with  $\alpha_1 = 0$  or  $[y - X\beta]' [y - X\beta]$  with respect to  $\beta$ .

The weighted least squares approach suggests that each observation corresponding to a given country and commodity,  $n$  and  $c$  be given a prespecified weight in the least squares estimation. Suppose  $w_{nc}$  is a weight to be given to the observation  $nc$ , then the weighted least squares (WLS) requires the minimization of:

$$\sum_{c=1}^C \sum_{n=1}^N w_{nc} [y_{nc} - \alpha_c - \gamma_n]^2 \quad \text{with } \alpha_1 = 0 \quad \text{or} \quad [y - X\beta]' W [y - X\beta] \quad (32)$$

where  $W$  is a diagonal matrix with elements  $w_{nc}$  which are all assumed to be positive. The weighted least squares estimator of  $\beta$  is given by

$$\hat{\beta} = [X'WX]^{-1} X'Wy \quad (33)$$

and the variance-covariance matrix of  $\hat{\beta}$ , under the assumption of independently and identically distributed random disturbances  $v_{nc}$  with mean 0 and variance  $\sigma^2$  can be shown to be

$$\text{Var}(\hat{\beta}) = \hat{\sigma}^2 (X'WX)^{-1} X'WWX(X'WX)^{-1} \quad (34)$$

The weighted least squares estimator is similar to the M-estimators used in the estimation of parameters using weights matrices within the least squares approach (Davidson and Mackinnon, 1993, pp. 587-596). The use of weights is also consistent with the approach used in Theil (1967).

The estimator  $\hat{\beta}$  is an unbiased estimator but whether it is efficient or not depends upon the nature of the  $W$  matrix. Once  $W$  matrix is given, it is a simple exercise to compute the weighed least squares estimator. The main practical issue that arises here is the choice of the  $W$  matrix – which weights to use and why?. In the following discussion a few instances where the use of weights arises naturally are discussed.

#### 4.1 National average prices and the use of weights

In the exposition thus far, a single price observation is assumed for each commodity in each country whenever an observation is made. In practice, a single price quotation is somewhat unrealistic. The general practice is that a number of price quotes are obtained for each commodity and then an average price is reported for use in the international comparison work. If individual quotes are available, then these can be incorporated into the unweighted CPD model as observations along the lines suggested by Diewert (2004).

Suppose the number of price quotes used in reporting a price for commodity  $n$  in country  $c$  is equal to  $K_{nc}$  for  $c=1,2,\dots,C$  and  $n=1,2,\dots,N$ . Further let  $\bar{y}_{nc}$  represent the average of logarithm of the price quotes

$$\bar{y}_{nc} = \frac{1}{K_{nc}} \sum_{k=1}^{K_{nc}} \ln p_{nck}$$

where  $p_{nck}$  represents the  $k$ -th price quote for commodity  $n$  in country  $c$ .<sup>14</sup> In this situation, it is intuitively obvious that the difference price observations  $\bar{y}_{nc}$  have different levels of reliability depending upon the number of price quotes that are used. Under the assumption that the individual price quotes are independently and identically distributed with variance  $\sigma^2$ ,

$$\text{Var}(\bar{y}_{nc}) = \frac{1}{K_{nc}} \sigma^2$$

and, therefore the CPD model based on  $\bar{y}_{nc}$  will have heteroskedastic disturbances. In this case a natural choice of weights is simply  $w_{nc} = K_{nc}$ . The resulting estimator is more efficient than the unweighted least squares estimator used in Section 3.

#### 4.2 Weights based on “representativeness” of items priced

In the context of international comparisons of prices, price data are collected for items in a product list which are selected on the basis of their “comparability” across countries included. While the items included in the product lists are available in the countries concerned, they are not always “representative” of the consumption of the general population in terms of their share in the total volume of sales for similar products. The unweighted CPD method in Section 3 treats all the price data as equally important whether or not the items priced are considered representative in a given country. The OECD and Eurostat comparisons in the recent years have modified their EKS procedure to take this phenomenon into account.

---

<sup>14</sup> Note that this is not the same as using logarithm of average price,

$$\ln \bar{p}_{nc} \text{ where } \bar{p}_{nc} = \frac{1}{K_{nc}} \sum_{k=1}^{K_{nc}} p_{nck}.$$

In this case, the weights need to be worked out carefully. This has some implications as to how the national statistical offices should report the prices.

For purposes of making a distinction between comparable and representative items, it may be possible to accord an arbitrarily selected weight. For example, it is possible to consider the following weighting scheme with

$$\begin{aligned} w_{nc} &= 2 \text{ if commodity } n \text{ is considered to be representative in country } c \\ &= 1 \text{ if commodity } n \text{ is available but not representative in country } c \end{aligned}$$

Essentially, this weighting scheme implies twice the weight to items that are representative than those that are not. The implied weighting matrix is a diagonal matrix with 2's and 1's and it can be used with observed  $X$  and  $y$  to yield weighted CPD indices. In this case it is possible that the weighted CPD indices may not be efficient when compared to unweighted indices.

### ***4.3 Expenditure Weights and Aggregation above the Basic Heading Level***

For purposes of aggregation above the basic heading level, we have expenditure share weights available for each basic heading in each country. If  $w_{ij}$  represents the share of  $i$ -th basic heading in  $j$ -th country then it is possible to obtain PPPs by minimizing

$$\sum_{c=1}^C \sum_{n=1}^N w_{nc} [y_{nc} - \alpha_c - \gamma_n]^2 = \sum_{c=1}^C \sum_{n=1}^N w_{nc} [\ln p_{nc} - \alpha_c - \gamma_n]^2 \quad (35)$$

with respect to  $\alpha_c$  and  $\gamma_n$ .

Once the expenditure share weights are given, estimation of parameters in (35) is fairly straightforward. According weights in the form of (35) is more in line with Theil (1967)'s approach and provides a point of departure from the general stochastic approach. Here there is no assumption of any heteroskedasticity in the error terms as it was assumed in the work on stochastic approach due to Clements and Izan (1981) and Selvanathan and Rao (1994). The use of weights in (35) is consistent with the standard index number approach where the index is expected to track the trends and levels in prices of items that are considered important in the budget.

#### ***4.3.1 Equivalence of weighted CPD Parities and the Rao Index for Multilateral Comparisons***

The normal equations resulting from the minimization of weighted residuals in (34) has a definite advantage. These normal equations can be shown to be identical to the equations used in defining the Rao (1990) system of PPPs. Rao (1995) has shown that PPPs resulting from the use of weighted CPD method are equivalent to a system of expenditure-share weighted log-change system similar to the Geary-Khamis system. The PPPs from the weighted CPD method can be derived from the following system of normal equations:

$$\hat{\alpha}_c = \sum_{n=1}^N w_{nc} (\ln p_{nc} - \hat{\gamma}_n) \quad \text{for } c=1,2,\dots,C; \text{ and}$$

$$\hat{\gamma}_n = \sum_{c=1}^C \frac{w_{nc}}{\sum_{c=1}^C w_{nc}} (\ln p_{nc} - \alpha_c) \text{ for } n=1,2,\dots,N \quad (36)$$

Equations in (36) are identical to the system of equations that define PPPs and international prices in the system provided in Rao (1990).

This result provides a very useful link between the type of index number formulae used in international comparisons and the CPD method which is essentially a stochastic approach to the construction of index numbers. Results in Rao (1990) can be used in interpreting the PPPs resulting from the expenditure share weighted CPD method described here. A very useful corollary is that when this method is applied for binary comparisons, the resulting index is in fact a superlative index which is very similar to the Tornqvist index.

### 4.3.2 CPD and the Tornqvist Index Numbers

The weighted CPD technique can be used in generating Tornqvist index numbers for binary comparisons, thus providing multilateral generalisations that are identical to those proposed in Caves, Christensen and Diewert (1982). Suppose we start with the CPD specification in (2). Suppose we are interested in a binary comparison between two countries  $c$  and  $d$ . The CPD model implies:

$$\begin{aligned} \ln p_{nc} &= \alpha_c + \gamma_n + v_{nc} \\ \ln p_{nd} &= \alpha_d + \gamma_n + v_{nd} \end{aligned}$$

If the principal objective is to compare price levels in countries  $c$  and  $d$  then  $(\alpha_d - \alpha_c)$  is of interest. This difference can be estimated directly by taking the difference of the above equations for  $c$  and  $d$  resulting in

$$\ln p_{nd} - \ln p_{nc} = \alpha_d - \alpha_c + (v_{nd} - v_{nc}) \quad (37)$$

Given price and expenditure share data, a direct price comparison can be obtained by minimizing the weighted sum of squares of residuals from (36). If average of expenditure shares in countries  $c$  and  $d$  are used to weight the residuals, the problem is one of minimizing

$$\sum_{n=1}^N \left( \frac{w_{nc} + w_{nd}}{2} \right) [\ln p_{nd} - \ln p_{nc} - (\alpha_d - \alpha_c)]^2$$

with respect to  $(\alpha_d - \alpha_c)$ . The resulting estimator of this difference leads to the standard Tornqvist index. If equation (36) is applied to data involving all pairs of countries  $c$  and  $d$  ( $c, d = 1, 2, \dots, C$ ) then the weighted least squares estimates coincide with the CCD indices proposed in Caves et al. (1982). Proof of this result can be found in Selvanathan and Rao (1992).

This result provides another application of the weighted CPD and the result stated above provides a link between the method and the Tornqvist index which is a superlative index.

## 5. Concluding Remarks

The main objective of the paper is to describe the country-product-dummy method for international comparisons and derive a number of important properties of the price comparisons derived using this method. The weighted and unweighted CPD variants are discussed in detail with particular attention to the cases where there are gaps in the price data. Where applicable, the paper provided a comparison of the CPD and EKS methods for international comparisons and demonstrated the clear advantages in using the CPD method. The paper also shows how the weighted CPD method can be useful in dealing with various data related issues. A link between the weighted CPD and more traditional indices such as those due to Caves *et al.*(1982) and Rao (1990) shows that indices derived using weighted CPD may have attractive economic theoretic properties. A clear advantage of the CPD method is that a range of econometric techniques and model specifications can be used in conjunction with the basic idea underlying the CPD method, these in turn lead to statistically improved estimates of the underlying parities. In particular, the paper has shown how the presence of geographical patterns in relative prices can be incorporated in to the CPD model using spatially autocorrelated structures.

The CPD method, the weighted version in particular, is being increasingly used in deriving spatial comparisons due to its ability to handle price quotations. Recently, Aten and Menezes (2002), Rao (2003) and Deaton *et al.* (2004) use the weighted CPD regression method to compare price levels using household data on unit values. Research on other analytical properties including variations of the CPD method for international comparisons is being actively pursued. An important aspect that deserves further research is on the possible applications of the CPD method for simultaneous temporal-spatial comparisons. In conclusion, the weighted and unweighted CPD methods provide an excellent tool for compiling purchasing power parities in the International Comparison Program.

### References:

- Aten, B (1996), "A Mapping of International Prices", *Review of Income and Wealth*.
- Aten, B. and T.A. Menezes (2002), "Poverty Price Levels: An Application to Brazilian Metropolitan Areas", Conference on the International Comparison Program, 11-15 March, 2002, Washington DC, World Bank.
- Caves, D.W., Christensen, L.R. and Diewert, W.E. (1982) "Multilateral Comparisons of Output, Input and Productivity Using Superlative Index Numbers", *Economic Journal*, 92, 73-86.
- Clements, K.W. and H.Y. Izan (1981) "A Note on Estimating Divisia Index Numbers", *International Economic Review*, 22, 745-47.
- Davidson, r. and J.G. Mackinnon (1993), *Estimation and Inference in Econometrics*, Oxford University Press.
- Deaton, A, J. Friedman and V. Alatas (2004), "Purchasing power parity exchange rates from household survey data: India and Indonesia", Paper presented at the World Bank Seminar on Global Poverty Measurement, January, 2004, Washington DC.

- Diewert, E.W. (2002) Weighted Country Product Dummy Variable Regressions and Index Number Formulae, mimeographed, Department of Economics, University of British Columbia, Vancouver.
- 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.
- Ferrari, G., G. Gozzi and M. Riani (1996), "Comparing GEKS and EPD Approaches for Calculating PPPs at the Basic Heading level", in *Improving the Quality of Price Indices: CPI and PPP*, Eurostat.
- Graybill, F.A. *Matrices with Applications in Statistics (2<sup>nd</sup> Edition)*, Belmont, California: Wadsworth International Group.
- Hill, R.J. (1999), "Chained PPPs and Minimum Spanning Trees" in Lipsey and Heston (eds.) *International and Interarea Comparisons of Prices, Income and Output*, NBER, Chicago University Press, pp. 327-364.
- Rao, D.S. Prasada (1990), 'A System of Log-Change Index Numbers for Multilateral Comparisons', in *Comparisons of Prices and real Products in Latin America* (eds. Salazar-Carrillo and Rao), Contributions to Economic Analysis Series, Amsterdam: North-Holland Publishing Company.
- Rao, D.S. Prasada (1995), "On the Equivalence of the Generalized Country-Product-Dummy (CPD) Method and the Rao-System for Multilateral Comparisons", Working Paper No. 5, Centre for International Comparisons, University of Pennsylvania, Philadelphia.
- Rao, D.S. Prasada (2001), 'Weighted EKS and Generalised CPD Methods for Aggregation at Basic Heading Level and above Basic Heading Level', Paper presented at the Joint World Bank-OECD Seminar on Purchasing Power Parities: real Advances in Methods and Applications, 30 January-1 February, 2001, Washington DC.
- Rao, D.S. Prasada (2003), *Purchasing Power Parities for the Measurement of Global and Regional Poverty*, A Report prepared for the DECDG of the World Bank, Washington DC, May.
- Rao, D.S. Prasada and E.A. Selvanathan (1992) "Uniformly Minimum Variance Unbiased Estimators of Theil-Tornqvist Index Numbers" *Economics Letters*, 39: 123-127.
- Selvanathan, E.A. and D.S. Prasada Rao (1992) "An Econometric Approach to the Construction of Generalised Theil-Tornqvist Indices for Multilateral Comparisons", *Journal of Econometrics*, 54, 335-346.
- Selvanathan, E.A. and D.S. Prasada Rao (1994), *Index Numbers: A Stochastic Approach*, Ann Arbor: The University of Michigan Press.
- Summers, R. (1967), *Economics and Information Theory*, Amsterdam, North-Holland.
- Theil, H. (1967), *Economics and Information Theory*, Amsterdam, North-Holland.