Review of Income and Wealth
Series 50, Number 1, March 2004

A METHOD OF CALCULATING REGIONAL CONSUMER PRICE DIFFERENTIALS WITH ILLUSTRATIVE EVIDENCE FROM INDIA

BY D. COONDOO, A. MAJUMDER

Indian Statistical Institute

AND

R. RAY*

University of Tasmania

In this paper we propose a method of estimating multilateral regional price index numbers from a given household level data set on item-wise unit values/prices. The method is closely related to the Country-Product Dummy variable model of Summers (1973). This method is likely to be particularly useful in studies of regional comparisons of poverty and inequality, optimal commodity taxes and tax reforms. To illustrate the method, we use it to calculate the regional consumer price index numbers for Eastern, Western and Southern India (taking Northern India as the reference region) separately for three categories of rural and urban households, viz., all households and those below and above the poverty line, using household level unit records of the NSS 50th round (1993–94) Consumer Expenditure Survey.

1. INTRODUCTION

The measurement of differences in consumer price levels over time, across regions or population groups is important for a variety of reasons including policy making in business and government. Suitably defined consumer price index numbers measuring differentials in consumer price levels are essential for comparison of real income or level of living/consumer expenditure pattern over time, across regions or across well-defined population groups. For example, in large countries like India or the U.S. or within a group of countries (for example, the OECD, the E.U., the ASEAN or the SAARC) there may be considerable regional heterogeneity in the level and pattern of consumer expenditure together with a regional differential in the consumer price levels and hence a meaningful comparison of the regional levels/patterns of consumption in real terms would call for the computation of a set of consistent consumer price indices measuring the extent of price differentials. There is a significant literature on the measurement

Note: The authors are grateful to Professors Nikhilesh Bhattacharya and Probal Chaudhuri for their helpful advice during the work on this paper. Helpful comments from Professors D. S. P. Rao and J. K. Ghosh and two anonymous referees on an earlier version are gratefully acknowledged. The authors also thank Geoffrey Lancaster for his painstaking and skilful research assistance. Ranjan Ray acknowledges the financial assistance from an Australian Research Council (Large) Grant. The usual disclaimer applies.

*Correspondence to: R. Ray, School of Economics, University of Tasmania, GPO Box 252-85, Hobart, Tasmania 7001, Australia (ranjan.ray@utas.edu.au).

of regional cost of living based mostly on the U.S. data (e.g. Moulton, 1995; Kokoski, Moulton, and Zeischang, 1999; Koo, Phillips, and Sigalla, 2000).

As is well-known, when more than two regions/countries/population groups are involved in a comparison of real income levels, the price index number problem is resolved in one of two major ways. The most straightforward approach is to use binary price index numbers for pair-wise comparison of real income levels and then attempt to get a consistent ordinal ranking of regions/countries/population groups such that transitivity is obeyed. Examples include Sen (1976), Bhattacharya, Joshi, and Roychowdhury (1980), Bhattacharya, Chatterjee, and Pal (1988), Coondoo and Saha (1990) and Deaton and Tarozzi (2000). Use of this binary methodology, however, does not guarantee transitivity of price level comparisons except under trivial and simplifying assumptions. The other approach, viz. the approach of the International Comparison Project (ICP) of the UN Statistical Office and the World Bank popularized the methodology of multilateral price level comparisons whereby a set of internally (transitivity) consistent price index numbers, known as Purchasing Power Parities (PPP) are estimated from a set of region/country/population group-specific price and quantity data for a common set of items/item groups (Geary, 1958; Khamis, 1972; Kravis, Heston, and Summers, 1978).

The methodology of multilateral price comparison has thrived over time in terms of theoretical development as well as a wide variety of applications (e.g. Balk, 1996; Hill, 1997; Prasada Rao, 1997; Diewert, 1999). However, like binary price index numbers, computation of a set of multilateral price index numbers also requires a data set of region/country/population group-specific prices and quantities of a set of items/commodities of uniform quality specifications, which is difficult to obtain. To resolve the data problems arising from quality variation of items across regions/countries/population groups and from gaps in the available price data, the Country Product Dummy (CPD) methodology was proposed (Summers, 1973). The CPD procedure, which is essentially an implementation of the *hedonic* approach (used to explain observed variations in the price of an item in terms of the quality attributes of the item) offers a regression analysis-based econometric methodology of construction of a set of multilateral price index numbers (Kokoski, Moulton, and Zeischang, 1999). Over the years the CPD methodology has undergone immense theoretical improvements (see e.g. Prasada Rao (2001) where the equivalence between a generalized CPD procedure and some standard multilateral price index number formulations has been discussed).

The literature on multilateral price index numbers is mostly concerned with the construction of PPPs/exchange rates from item/item group-specific price and quantity/expenditure share data available at the *region/country/population group* level using the CPD methodology. One may alternatively use micro-level data (e.g. household level data on commodity prices/unit values available from country-wide consumer expenditure surveys) in the estimation of multilateral consumer price index numbers based on the CPD methodology (e.g. Aten and Menezes, 2002).[1] Given the fact that such a body of micro-level data contains a huge and valuable price information, it is worthwhile to explore if household level price/unit value

[1]A referee has brought this study to our notice.

data set covering households belonging to more than one region/population groups (e.g. a set of countries, states/provinces within a country, districts within a state/province or social/occupation groups, income groups etc.) can be utilized to measure the extent of differential in the levels of consumer prices across regions/population groups by estimating an appropriately defined set of multilateral consumer price index numbers.

In this paper we present a regression analysis-based econometric procedure for estimation of a set of multilateral consumer price index numbers based on cross-sectional household level data set on item/item group-specific prices/unit values or expenditures (which is typically obtained from a nationwide consumer expenditure survey). As explained later, our approach is primarily based on the concept of *quality equation* (henceforth called the price equation interchangeably) of Prais and Houthakker (1955), which relates the observed price/unit value paid by a household for an item/item group to the household's level of living (as measured by the per capita real income/total consumer expenditure (PCE) of the household). In this particular study we concentrate on the inter-regional comparison. The same framework is applicable to inter country/population group comparisons and also to inter temporal and seasonal comparisons. Briefly, given the set of items/item groups (henceforth to be called *items*) for which household level consumer expenditure data are available for all regions under consideration, we specify and estimate the set of item-specific price equations using the pooled data for all regions in the first stage of estimation. A set of regional dummy variables is introduced in these item-specific price equations such that the estimated region-specific price equations for each item are obtained. In the second stage, the set of multilateral consumer price index numbers are obtained, again by using regression analysis, by relating the slope and intercept terms of the estimated item-specific price equations separately for individual region. As explained below, our suggested procedure belongs to the CPD methodology in a generic sense as the *price equation* essentially shares the *hedonic* feature, which is central to the idea of the CPD methodology. There are, however, a few basic differences: (1) we use the household PCE and attributes as surrogates for quality of items consumed by a sample household, whereas the CPD procedure basically tries to decompose observed variations in prices across countries/regions/population groups as well as items/item groups into two parts—one due to the country/region/population group effect and the other due to the product/item effect; (2) whereas the CPD procedure utilizes a data set consisting of a total of ($M \times R$) sample observations (where M = number of items and R = number of countries/regions/population groups), our procedure is designed to utilize the entire set of household level data on item-specific price/unit value that may be available from a large scale consumer expenditure survey; and (3) whereas the CPD procedure can be used for multilateral comparison of levels of different types of prices (e.g. consumer prices, prices of capital goods, industrial raw materials etc.) alike, the scope of our procedure (being based on the *quality equation*) is essentially limited to comparison of consumer price levels.

The paper is organized as follows: Section 2 specifies the price equation and explains the proposed model with reference to the CPD model; Section 3 sets out the estimation procedure; Section 4 enumerates the distinctive features of the proposed procedure; Section 5 describes the data used for the illustrative exercises done and presents the results; and Section 6 concludes.

2. The Proposed Model

The basic CPD model consists of a single linear regression equation of which the dependent variable is p_{jr}, the natural logarithm of the observed price of the j-th commodity/item ($j = 1, 2, \ldots, M$) for the r-th country ($r = 0, 1, \ldots, R$), and there are two sets of explanatory variables—one being the set of M commodity/item dummy variables $D_1, D_2, \ldots, D_M$ and the other being the set of R country-specific dummy variables $S_1, S_2, \ldots, S_{\hat{R}}$ ($r = 0$ denoting the *numeraire* country). The linear regression equation specification of the basic CPD model is thus as follows:

$$p_{jr} = \sum_{r=1}^{R} b_r S_r + \sum_{j=1}^{M} z_j D_j + \varepsilon_{jr}, \tag{1}$$

where b_r's and z_j's are the ($R + M$) regression parameters of the model and ε_{jr} is the equation random disturbance term. The country coefficients, i.e., the b_r's, measure the natural logarithms of the country parity and the commodity coefficients, i.e. the z_j's measure the natural logarithms of the price of individual commodities/items in the numeraire country's currency. It may be noted that this model was originally used for *filling gaps in available price information rather than for estimating purchasing power parities (PPP)* and it does not make use of any quantity or value data.

Prasada Rao (1996) generalized the estimation procedure of this model by making use of quantity and value data and Prasada Rao (2001) proposed a generalized estimation procedure in which a weighted residual sum of squares is minimized, the weights being the expenditure share of a commodity/item for the given country. Hill, Knight, and Sirmans (1997), Kokoski, Moulton, and Zeischang (1999) and Triplett (2000) proposed the use of CPD model for incorporating quality adjustment in the estimation of PPP for regional price level comparison. The model used for making quality adjustments is given by the following single linear regression equation:

$$p_{jr} = \sum_{r=1}^{R} b_r S_r + \sum_{j=1}^{M} z_j D_j + \sum_{q=1}^{Q} \theta_q C_{qjr} + \varepsilon_{jr}, \tag{2}$$

where C_{qjr}'s, $q = 1, 2, \ldots, Q$, are the set of quality characteristics that are deemed to be relevant for a given price level comparison problem.

In contrast to the CPD methodology mentioned above, our proposed procedure is based on a set of M price equations (i.e. quality equations) relating to the individual commodities/items. An individual price equation, as specified below, can be regarded as a variant of equation (2), broadly speaking. The price equation for the j-th item is specified as follows:

$$(p_{jrh} - \Pi_r) = \alpha_j + \sum_i \delta_{ji} n_{irh} + (\lambda_j + \eta_{jr})(y_{rh} - \Pi_r) + \varepsilon_{jrh}, \quad j = 1, 2, \ldots, M \tag{3}$$

where p_{jrh} denotes the natural logarithm of the nominal price/unit value for the j-th item paid by the h-th sample household of region r, y_{rh} denotes the natural logarithm of the nominal per capita income/PCE of the h-th sample household of region r, n_{irh} denotes the number of household members of the i-th age-sex category present in the h-th sample household of region r, $i = 1, 2, 3, 4$ denoting adult male, adult female, male child and female child, respectively, and ε_{jrh} denotes the random equation disturbance term. α_j, δ_{ji}, λ_j, η_{jr}, Π_r are the parameters of the model. We assume that the random disturbance terms associated with individual observations follow the standard assumptions of the classical linear regression model.

Before providing interpretation of the parameters, let us explain the rationale of the set up. As already mentioned, the basic premise of the present approach is the concept of quality equation due to Prais and Houthakker (1955) in which the price/unit value for a commodity/item paid by a household is taken to measure the quality of the commodity/item group consumed and hence the price/unit value is postulated to be an increasing function of the level of living of the household. In so far as a broad measure of a household's level of living, ceteris paribus, is the effective per capita income/PCE, PCE and household demographics should be the basic explanatory variables of the price equation to be estimated on the basis of household level data.[2] Further, when the sample consists of households belonging to more than one region, both the price/unit value of individual items and PCE should be measured in real terms so as to adequately capture the possible effect of differentials in levels of consumer prices across regions on the quality sensitivity of the households. The parameters $(\Pi_r - \Pi_0)$, $r = 1, 2, \ldots, R$ denote a set of logarithmic price index numbers for individual regions measuring the regional price level relative to that of the reference *numeraire* region ($r = 0$). In principle, thus, Π_r's may be interpreted as *the natural logarithm of the value of a reference basket of items/commodities purchased at the prices of region r.* The l.h.s. of (3) thus measures the logarithm of the price/unit value paid in real terms (i.e. at region 0 prices) and $(y_{rh} - \Pi_r)$ on the r.h.s. of (3) measures the logarithm of real PCE (i.e. measured at region 0 prices). The parameter $(\lambda_j + \eta_{jr})$ is known as the *quality elasticity* for item j for households of region r as it measures the percentage by which the price/unit value paid for commodity/item group j increases in response to 1 percent change in real PCE, ceteris paribus. Whereas λ_j denotes the common part of quality elasticity, η_{jr} measures the regional differential in this elasticity. Normally, one would expect $(\lambda_j + \eta_{jr}) \geq 0$ for all j and r. Next, parameters δ_{ji}'s measure the partial effect on the price/unit value of commodity/item group j of a change in the household size and composition due to a change in the number of i-th type household member, given the real PCE. For example, addition of a male child, ceteris paribus, may induce a household to opt for cheaper/inferior variety of some items in order to increase the level of consumption of male child-specific items.

[2]The rationale for inclusion of household PCE as an explanatory variable in the quality equation is as follows: For an item like "rice," a variety of which may be available in the market, a richer household is likely to prefer and purchase rice of a superior quality, thereby paying a higher unit value for "rice." For a composite item group like, say, "milk and milk products," it is expected that a richer household will consume relatively expensive items of the group and hence pay a higher unit value for the composite item group.

Finally, α_j denotes the average price/unit value of item j measured at *numeraire* region (i.e. region 0) prices. To sum up, our proposed procedure seeks to estimate a set of logarithmic multilateral price index numbers in the form of $(\Pi_r - \Pi_0)$'s by estimating the M price equations in (3) by utilizing observed inter-household variations in the nominal price/unit value of items present in household level data thrown up by one or more household expenditure survey(s). Here two deterministic sources of such inter-household price variations are recognized—viz. variation in price level across regions and inter-household variations in level of living (measured by the (logarithm of) PCE and household size and composition).

3. Estimation Procedure

Under the CPD methodology a single linear regression equation of the form (1) or (2) is estimated on the basis of $(R + 1)M$ data points. Our proposed procedure, on the other hand, involves estimation of the set of M price equations (3) based on $\sum_{j=1}^{M}\sum_{r=0}^{R} N_{jr}$ observations, where N_{jr} is the number of sample households of region r reporting a price/unit value for item j.[3] It may be noted that (i) the set of equations (3) is nonlinear in parameters and (ii) the set of parameters Π_r's appears in all the individual price equations. While it may be possible to devise an appropriate non-linear systems approach, such an estimation procedure may turn out to be computationally heavy. We suggest below a two-stage estimation procedure such that at both stages Ordinary Least Squares (OLS) method may be used.

The first stage involves estimation of the individual price equations separately on the basis of pooled data set of all the regions. For this, the price equation (3) for an individual item may be expressed as the following linear regression equation:

$$p_{jrh} = \alpha_j^* + \sum_{i=1}^{4} \delta_{ji} n_{irh} + \sum_{p=1}^{R} \phi_{jp} S_p + \lambda_j y_{rh} + \sum_{p=1}^{R} \eta_{jp} y_{ph} S_p + \varepsilon_{jrh}, \tag{4}$$

$$j = 1, 2, \ldots, M; \quad r = 0, 1, \ldots, R; \quad h = 1, 2, \ldots, N_r$$

Estimation of (4) will yield estimates of the parameters α_j^*'s, δ_{ji}'s, ϕ_{jp}'s, λ_j's and η_{jp}'s appearing in (4). Of these, α_j^*'s and ϕ_{jp}'s are functions of the parameters appearing in (3) and their explicit forms can be obtained by examining the equivalence between equations (3) and (4). To see this equivalence, let us rewrite equation (3) as

$$\begin{aligned}
p_{jrh} &= \alpha_j + \Pi_r + \sum_i \delta_{ji} n_{jrh} + (\lambda_j + \eta_{jr})(y_{rh} - \Pi_r) + \varepsilon_{jrh} \\
&= \alpha_j + \Pi_r + \sum_i \delta_{ji} n_{jrh} - (\lambda_j + \eta_{jr})\Pi_r + \lambda_j y_{rh} + \eta_{jr} y_{rh} + \varepsilon_{jrh} \\
&= \alpha_j + (1 - \lambda_j)\Pi_0 + \sum_i \delta_{ji} n_{jrh} + \{1 - (\lambda_j + \eta_{jr})\}\Pi_r - (1 - \lambda_j)\Pi_0 \\
&\quad + \lambda_j y_{rh} + \eta_{jr} y_{rh} + \varepsilon_{jrh}.
\end{aligned}$$

[3]Note that it is not necessary that all sample households of all the regions have price/unit value for all the M items.

Hence,

$$p_{jrh} = \{\alpha_j + (1-\lambda_j)\Pi_0\} + \sum_i \delta_{ji} n_{jrh} + [\{1-(\lambda_j + \eta_{jr})\}\Pi_r - (1-\lambda_j)\Pi_0] + \lambda_j y_{rh} + \eta_{jr} y_{rh} + \varepsilon_{jrh} \tag{5}$$

Comparing equations (4) and (5) which are identical equations, we get

$$\alpha_j^* = \alpha_j + (1-\lambda_j)\Pi_0 \tag{6a}$$

$$\phi_{jp} = \{1-(\lambda_j + \eta_{jp})\}\Pi_p + \alpha_j - \alpha_j^* = \{1-(\lambda_j + \eta_{jp})\}\Pi_p - (1-\lambda_j)\Pi_0. \tag{6b}$$

Note from equation (3) that $(\lambda_j + \eta_{jr})$ is the slope coefficient for region r ($\neq 0$), λ_j is that for the *numeraire* (i.e. the reference) region, α_j^* is the intercept term for the *numeraire* region and ϕ_{jr} is the differential intercept term for region r ($\neq 0$) of the price equation for item j. Thus, $\exp(\phi_{jr})$ is the price relative of item j for region r ($\neq 0$) with the *numeraire* region taken as the base. Note that this model (i.e. equation (4)) reduces to the basic CPD model when $\phi_{jr} = \phi_j$ for all j, $\eta_{jr} = 0$ for all j and r, and $\lambda_j = 0$ for all j. Thus, equation (4) is a generalized version of the basic CPD model.[4]

Given the estimated values $\hat{\phi}_{jp}$, $\hat{\lambda}_j$ and $\hat{\eta}_{jp}$ of the corresponding parameters obtained from the first stage estimation, equation (6b) is used to define the linear regression equation to be estimated in the second stage for obtaining the estimates of Π_p's. Thus, rewriting equation (6b), we get the following dummy variables regression equation:

$$\hat{\phi}_{jr} = \sum_{p=1}^{R} \Pi_p \{1-(\hat{\lambda}_j + \hat{\eta}_{jp})\} S_p - \Pi_0(1-\hat{\lambda}_j) + \varepsilon_{jr}. \tag{7}$$

and the pooled set of estimated item-specific $\hat{\phi}_{jr}$'s of all the regions are regressed on region dummies $S_1, S_2, \ldots, S_R$ and $(1-\hat{\lambda}_j)$'s.[5]

Finally, it may be noted that equation (7) is derived from the equation system (6a)–(6b) which is a system of $(R+1)$ equations in $(R+2)$ unknowns, viz., $\Pi_0, \Pi_1, \ldots, \Pi_R$ and α_j, for every j. Thus, each Π_r is a linear function of (every) α_j (which is unidentifiable and hence non-estimable, given the model). That is, the estimated Π_r's will have α_j's confounded in them thus affecting the magnitude of these estimates. Actually, the Π_r's estimated for a given data set will contain an additive component which is some kind of an average of the non-estimable α_j's, say $\bar{\alpha}$.

[4]Note that the term $\sum_{i=1}^{4} \delta_{ji} n_{irh}$ in (4) corresponds to the household suffix h in the l.h.s. and does not affect the basic structure of the CPD model.

[5]It may be noted that (7) is actually an alternative dummy variables representation of

$$\hat{\phi}_{jr} = \Pi_r\{1-(\hat{\lambda}_j + \hat{\eta}_{jr})\}S_r - \Pi_0(1-\hat{\lambda}_j) + \varepsilon_{jr}, \quad r = 1, 2, \ldots, R$$

which constitutes a system of R linear regression equations in each of which the term $-\Pi_0(1-\hat{\lambda}_j)$ appears. In other words, Π_0 in the present model is over-identified as R different estimates of this parameter may be obtained by estimating the above equation separately for $r = 1, 2, \ldots, R$. To resolve this over-determinacy of Π_0, we propose estimation of the dummy variables regression equation (7) instead, which ensures that a single estimate of Π_0 is obtained. The number of observations used in this second stage estimation thus equals the number of items times the number of regions.

Thus, while the estimates of Π_r's *will not have any obvious interpretation*, their differences will unambiguously measure the logarithm of the price index number of one region with respect to another (as the $\overline{\alpha}$ will cancel out when the difference is taken).

4. Some Features of the Proposed Procedure

Let us next briefly enumerate some of the distinctive features of the proposed procedure. First, the estimates of the set of multilateral logarithmic consumer price index numbers $(\Pi_r - \Pi_0)$'s, being based on a dummy variables linear regression model, will be invariant to the choice of the *numeraire* (i.e. the reference) region due to the properties of the dummy variables regression model. This implies that the resulting multilateral price index numbers will automatically satisfy the required *circular* (i.e. the transitivity) consistency.

Second, an advantage of the proposed procedure is that this procedure can include items of expenditure like service items (e.g. consumption of education services, health services, personal care etc.) for which price/unit value may not be well-defined. How this can be accomplished is explained in the Appendix. It may be noted that no other procedure of multilateral consumer price index number construction can incorporate items of consumption for which price/unit value and quantity of consumption are not well-defined.

Third, unlike other procedures of multilateral consumer price index number construction, the procedure proposed here does not require that price data for all items must be available for all regions for the procedure to work. The proposed procedure will work even if price data for some items are not available for some regions. As already described, the first stage of this method involves estimation of the item-specific (logarithmic) price equations based on item-specific price data for all regions pooled together. In the second stage the region-specific (logarithmic) price index numbers are estimated (based on a linear regression equation with region dummy variables) using region-specific item-wise intercept and slope differentials of the price equations estimated in the earlier stage. Therefore, if, say, for item j ($j = 1, 2, \ldots, M$) price data for some region p ($p = 0, 1, 2, \ldots, R$) are not available, the first stage estimation will not yield the estimate of the corresponding ϕ_{jp}. This will, however, not hamper the second stage estimation (so long as the item is available in other regions) as the second stage estimation will be based on $\sum_p M_p$ observations, M_p being the number of items available for the p-th region.

Finally, although we have suggested here a simple two-stage estimation strategy for estimation of the set of multilateral consumer price index numbers, under our proposed procedure it may not be difficult, in principle, to devise an appropriate iterative nonlinear estimation technique for simultaneous estimation of the set of M item-specific price equations in (4) under reasonable and realistic assumptions about the random disturbance terms. If this is done, our proposed procedure will yield standard errors of the set of estimated multilateral consumer price index numbers. In this context it may be mentioned that estimation of standard error of computed price index numbers has been an issue of major concern. For example, Clements and Izan (1981) and Selvanathan and Rao (1994) tried to devise econometric approaches to price index number estimation which yield standard error of estimated price index number in a natural way. Diewert (1995), however, criticized these formulations as being based on rather unrealistic statistical assumptions and Diewert (2002) proposed a regression based CPD approach which yields standard error of estimated price index numbers under plausible assumptions. Note, however, that it may not be conceptually difficult to evolve a satisfactory estimation procedure for our proposed model, although the implementation of such a procedure may be computationally very heavy.

5. An Illustrative Application of the Proposed Procedure

For the purpose of illustration, we applied the proposed procedure to the Indian household level consumer expenditure data thrown up by the 50th round of the all-India consumer expenditure survey conducted by the National Sample Survey Organization during the period July 1993 to June 1994. This survey covered approximately 115,000 households all over the country, of which 70,000 were rural households and 45,000 were urban households.

Using this data set, multilateral regional consumer price index numbers were estimated for all households and also separately for households living below and above the poverty line. In each of these cases, separate sets of regional consumer price index numbers were estimated for the rural and urban sectors. With regard to the formation of the regions, 25 States of the country were covered in this illustrative exercise and these were classified into four geographical regions: North ($r = 0$, i.e. taken as the *numeraire* or reference region), South ($r = 1$), East ($r = 3$), and West ($r = 4$). Table 1 presents State composition of these regions.

With regard to the commodity/item group coverage of the estimated price index numbers, items for which information on both *expenditure* and *quantity* was available were considered because for these items *price/unit value* could be readily calculated by dividing expenditure by the corresponding quantity consumed.[6] By this criterion a subset of 45 food items was selected for the exercise. Table 1 also gives the list of items covered.

As already explained, estimation of commodity/item group-specific price equations required household level data on *effective* PCE which was obtained by dividing observed household total consumer expenditure (taken as sum total of expenditures on all food and nonfood items) by effective household size computed by using appropriate household equivalence scale. State-specific household equivalence scales estimated by Meenakshi and Ray (2002, Table 3) were used for this purpose. It may be mentioned that use of such State-specific household equivalence scales was necessary as considerable heterogeneity in the consumption effects of household demographics across States is known to exist in India that may have significant effect on regional consumer price disparities via the strong link between household demographics and price effects often cited in the literature (see, e.g.

[6]Since the exercise is for illustrative purpose only, we confined ourselves to such items. As already mentioned, the proposed procedure can also accommodate items for which only expenditure data are available.

TABLE 1

LIST OF STATES AND ITEMS COVERED

Regions/States			
North (*r* = 0: Reference Region)	South (*r* = 1)	East (*r* = 2)	West (*r* = 3)
Haryana	Andhra Pradesh	Arunachal Pradesh Nagaland	Goa
Himachal Pradesh	Karnataka	Assam Orissa	Gujarat
Jammu & Kashmir	Kerala	Bihar Sikkim	Maharashtra
Madhya Pradesh	Tamil Nadu	Manipur Tripura	Rajasthan
Punjab		Meghalaya West Bengal	
Uttar Pradesh		Mizoram	

Items covered

Rice	Vanaspati + margarine	Cauliflower
Wheat	Mustard oil	Cabbage
Other cereals	Groundnut oil	Brinjal
Cereal substitutes	Coconut oil	Tomato
Arahar	Other edible oils	Green chilli
Gram	Goat meat + mutton	Other vegetables
Moong	Beef + buffalo meat	Fruits-fresh
Masoor	Pork	Fruits-dry
Urd	Other meat	Sugar (crystal)
Khesari	Chicken	Other sugar items
Peas	Fish (fresh + dry)	Salt
Soybean	Potato	Spices
Other pulses and pulse products	Onion	Tea + coffee
Liquid milk	Other root vegetables	Processed food 1 (biscuits etc.)
Milk products	Gourd all varieties	Processed food 2 (pickle etc.)

Barten, 1964). Finally, estimation of regional consumer price index numbers separately for households living below and above poverty line required State-specific poverty lines. The required information was taken from Dubey and Gangopadhyay (1998) who constructed the poverty lines separately for rural and urban areas in each State.

Let us next present the results of our illustrative exercise. As already mentioned, the first stage of the exercise involved estimation of item-specific price equations involving coefficients of demographic variables (δ's), coefficients of regional intercept dummies (ϕ's), coefficient of the PCE variable or the *common* quality elasticity (λ's) and region-specific differential slope coefficients or the *differential* quality elasticity (η's).[7] The salient features observed in these results are briefly summarized below.

For each of the four population sub-groups, viz. rural poor households, rural non-poor households, urban poor households and urban non-poor households, the estimate of *common* quality elasticity (λ's) turned out to be positive and statistically significant in all cases barring a very few exceptions. Most of the esti-

[7]The estimated parameters of the price equations (i.e. equation (4)) of different categories of households have not been presented here for reasons of space. These may be made available to interested readers, if requested.

mated coefficients of regional intercept dummies (ϕ's) and those of the region-specific differential quality elasticity (η's) were also found to be statistically significant, indicating the presence of regional differentials in the level of prices in the data. With regard to the effect of demographic variables on the prices/unit values (i.e. estimated δ's), both positive and negative (statistically significant) values of these were obtained, which perhaps provided justification for inclusion of these variables in the price equations.

Coming next to the results of the second stage estimation (i.e. the estimation of equation (7), which yielded estimates of region-specific Π's), estimated region-specific Π's are presented in Table 2. In fact, in this table for each of the six population groups (viz. households below and above poverty line and all households of rural and urban India), two different sets of estimated Π's have been presented. One set was obtained from the OLS estimation of equation (7). The other set arose out of the jackknife exercise done to examine the stability of the estimated Π's[8] The results of the jackknife procedure presented confirm that all the estimated Π's are indeed statistically significant and stable.

It may be mentioned here that according to our formulation $\phi_{jr} + (1 - \lambda_j)\Pi_0$ should bear a proportional relationship with $\{1 - (\lambda_j + \eta_{jr})\}$ across j for each r, the factor of proportionality being Π_r (vide equation (6b)). This means that if a given data set supports our model specification, then given the value of Π_0, for every region r, the scatter diagram of estimated values of these composite parameters for individual items (i.e., the former plotted against the latter) will show a linear relationship passing through the origin and the slope of the line will be the value of Π_r for that region. We verified this implied proportionality relationship for each of the cases of price level comparison. In all cases this proportionality was strongly confirmed, thus lending a rather strong empirical support for our proposed procedure.

Table 3 presents the estimates of region-specific consumer price index numbers for each of the three different types of households separately for the rural and the urban sector (with *North* taken as the reference region in each case). The following features of these results are worth noting:

(1) For the *all households* group in both rural and urban sector, the price index number for the East is the highest among all regions. East is followed by North, West and South, in descending order, South being the cheapest.

(2) For the *households above poverty line* group in both the sectors, the same pattern as mentioned in (1) is observed.

[8]The second stage estimation was done using SHAZAM. The jackknife procedure involved repeated estimation of equation (7) each time omitting a different observation of the given set. For the linear regression model $Y = X\beta + \varepsilon$, the jackknife estimate (dropping the t-th observation) is $\hat{\beta}_{(t)} = \hat{\beta} - (X'X)^{-1} X_t' e_t^+$, where $\hat{\beta}$ is OLS(β), X_t: row vector for the t-th observation and $e_t^+ = e_t/(1 - K_t)$ e_t and K_t being the OLS residual for the t-th observation and the t-th diagonal element of the matrix $X(X'X)^{-1}X'$, respectively. A total of $N(k + 1)$ coefficient vectors are generated, each corresponding to a separate regression with the t-th observation dropped. The average of these $N(k + 1)$ coefficient vectors is reported in Table 2 (see Judge *et al*. (1988) for more details).

TABLE 2

ESTIMATED Π COEFFICIENTS FOR DIFFERENT REGIONS AND HOUSEHOLD GROUPS: RURAL AND URBAN INDIA

		OLS Coefficients					Jackknife Coefficients			
Household Group	No. of observations	North ($\hat{\pi}_0$)	South ($\hat{\pi}_1$)	East ($\hat{\pi}_2$)	West ($\hat{\pi}_3$)	R^2	North ($\hat{\pi}_0$)	South ($\hat{\pi}_1$)	East ($\hat{\pi}_2$)	West ($\hat{\pi}_3$)
Rural India										
All households	123	11.125	11.115	11.216	11.121	0.9489	11.208	11.199	11.302	11.208
		(0.229)	(0.238)	(0.251)	(0.242)		(0.522)	(0.520)	(0.545)	(0.526)
Households above poverty line	123	11.277	11.251	11.358	11.275	0.9683	11.316	11.287	11.397	11.315
		(0.181)	(0.190)	(0.196)	(0.195)		(0.352)	(0.352)	(0.364)	(0.359)
Households below poverty line	120	9.598	9.677	9.694	9.687	0.9998	9.502	9.512	9.593	9.623
		(0.051)	(0.012)	(0.083)	(0.070)		(0.128)	(0.167)	(0.152)	(0.137)
Urban India										
All households	123	11.139	11.068	11.179	11.099	0.9508	11.194	11.123	11.237	11.155
		(0.227)	(0.235)	(0.248)	(0.238)		(0.350)	(0.351)	(0.369)	(0.364)
Households above poverty line	123	11.163	11.085	11.205	11.114	0.9646	11.152	11.073	11.193	11.102
		(0.188)	(0.199)	(0.206)	(0.200)		(0.219)	(0.224)	(0.233)	(0.231)
Households below poverty line	117	9.982	10.083	9.976	9.990	0.9979	9.936	9.998	9.925	9.924
		(0.052)	(0.073)	(0.077)	(0.047)		(0.091)	(0.119)	(0.108)	(0.098)

Note: Figures in parenthesis are the standard errors.

TABLE 3

ESTIMATED PRICE INDEX NUMBERS FOR DIFFERENT REGIONS AND HOUSEHOLD GROUPS: RURAL AND URBAN INDIA (BASE: NORTH = 1.0)

	Rural India			Urban India		
Household Group	South ($e^{(\hat{\pi}_1-\hat{\pi}_0)}$)	East ($e^{(\hat{\pi}_2-\hat{\pi}_0)}$)	West ($e^{(\hat{\pi}_3-\hat{\pi}_0)}$)	South ($e^{(\hat{\pi}_1-\hat{\pi}_0)}$)	East ($e^{(\hat{\pi}_2-\hat{\pi}_0)}$)	West ($e^{(\hat{\pi}_3-\hat{\pi}_0)}$)
All households	0.990	1.095	0.996	0.931	1.041	0.961
Households above poverty line	0.974	1.084	0.998	0.925	1.043	0.952
Households below poverty line	1.082	1.101	1.092	1.058	0.994	1.009

(3) For the *households below poverty line* group of the rural sector, all price index numbers are greater than 1, indicating the price level for North to be the lowest. North is followed by, South, West and East, in ascending order. For the urban sector the ordering of region is somewhat different. In this case, East is followed by North, West and South, in ascending order. Thus, South turns out to be the most expensive region, so far as urban poor are concerned.

(4) On the whole, the pattern of regional consumer price differentials is sensitive not only to the rural-urban divide, but perhaps more crucially to whether a household lives below or above the poverty line.

6. CONCLUSION

In this paper we have proposed a simple and straightforward regression-based econometric methodology of estimation of a set of multilateral consumer price index numbers from a given set of household level data on item-specific prices/unit values. The working of the proposed procedure has been illustrated by applying it on the Indian household level item-specific price/unit value data thrown up by the NSS 50th round household consumer expenditure survey (July 1993 to June 1994). The results relating to regional consumer price differentials for groups of households below and above the poverty line and all households in the rural and urban sectors turn out to be sensible and robust.

Some of the distinctive technical features of the proposed procedure are as follows: (i) the method, being based on household level data, is capable of bringing out differentials in consumer price levels across various well-defined groups of households implicit in the given data set in a robust manner; (ii) in view of an adequate stochastic specification of the price equations, it is possible to evolve a satisfactory estimation strategy such that standard errors of the estimated price index numbers can be obtained; (iii) as explained in the Appendix, our proposed procedure can be easily extended such that items of consumption for which quantity of consumption is not well-defined and therefore only expenditure data are available (such as consumption of services like recreation, education, health and medical care etc.) can also be included in the multilateral consumer price index number being compiled/estimated—a feature shared by none of the existing procedures; and (iv) being based essentially upon the CPD approach (which originally was devised to fill up gaps in the available price data required for construction of multilateral price index numbers), the proposed procedure will work even when all goods (and hence data on all prices) are not available for all the regions/population groups.

The empirical potentialities of the proposed procedure seem to be immense. Being essentially econometric in nature and meant to be used on a set of household level data on price/unit value of items of consumer expenditure, our proposed method can be conveniently used for a wide variety of consumer price level comparisons by defining appropriate population (household) groups (across which price level comparison is to be done). Thus, for example, one may cross-classify a given population of households (say, the Indian households) by rural-urban status, region of location (i.e. State) etc. and perform a price level comparison exercise. Such an exercise will, no doubt, involve heavy computations, but is feasible and straightforward. One may even get inter-temporal price level comparison done by pooling household level data on prices/unit values obtained from household consumer expenditure surveys done at different times and treating the set of sample households of individual surveys as different population groups.

Appendix: Extension of the Proposed Procedure

The multilateral consumer price index number estimation procedure that we have proposed in Section 2 is well-defined when the items covered are all measured in the same unit (e.g. in kilograms or grams as we have chosen in our illustrative exercise). The reason for this has been explained in the ultimate paragraph of Section 3—viz. given our procedure, *the estimated Π_r's will have α_j's confounded in them (thus affecting the magnitude of these estimates). Actually, the Π_r's estimated for a given data set will contain an additive component which is some kind of an average of the non-estimable α_j's, say $\bar{\alpha}$.* Thus, in case data on prices/unit values of items measured in a variety of quantity units are used, $\bar{\alpha}$ will be a meaningless quantity and hence so will be the estimated Π_r's. However, this need not be a shortcoming of our proposed procedure and, as we show here, the procedure can be extended such that it may cover not only prices of items, quantities of which are measured in a variety of units (e.g. kg, liter, meter, number of pieces, pair, dozen etc.), but also items such as service items for which only expenditure data are available. In what follows, we explain these extensions.

Inclusion of Item Groups Having Different Measurement Units for Quantity

Let us assume that there are G item groups and the g-th group ($g = 1, 2, \ldots, G$) has M_g items (all of which are measured in the same quantity unit). Equation (4) may now be written as

$$p_{jgrh} = \alpha^*_{jg} + \sum_{i=1}^{4} \delta_{jgi} n_{irh} + \sum_{p=1}^{R} \phi_{jgp} S_p + \lambda_{jg} y_{rh} + \sum_{p=1}^{R} \eta_{jgp} y_{ph} S_p + \varepsilon_{jgrh} \tag{4'}$$

$$j = 1, 2, \ldots, M_g; \quad g = 1, 2, \ldots, G; \quad r = 1, 2, \ldots, R; \quad h = 1, 2, \ldots, N_r$$

Correspondingly, equation (3) is now written as:

$$p_{jgrh} = \alpha_{jg} + \Pi_r + \sum_i \delta_{jgi} n_{irh} + (\lambda_{jg} + \eta_{jgr})(y_{rh} - \Pi_r) + \varepsilon_{jgrh}$$

$$= \alpha_{jg} + \Pi_r + \sum_i \delta_{jgi} n_{irh} - (\lambda_{jg} + \eta_{jgr})\Pi_r + \lambda_{jg} y_{rh} + \eta_{jgr} y_{rh} + \varepsilon_{jgrh}$$

$$= \alpha_{jg} + (1 - \lambda_{jg})\Pi_0 + \sum_i \delta_{jgi} n_{irh} + \{1 - (\lambda_{jg} + \eta_{jgr})\}\Pi_r - (1 - \lambda_{jg})\Pi_0$$

$$+ \lambda_{jg} y_{rh} + \eta_{jgr} y_{rh} + \varepsilon_{jgrh}.$$

Hence,

(5′)
$$p_{jgrh} = \{\alpha_{jg} + (1 - \lambda_{jg})\Pi_0\} + \sum_i \delta_{jgi} n_{irh} + [\{1 - (\lambda_{jg} + \eta_{jgr})\}\Pi_r - (1 - \lambda_{jg})\Pi_0]$$
$$+ (\lambda_{jg} y_{rh} + \eta_{jgr} y_{rh} + \varepsilon_{jgrh})$$

Comparing (4′) and (5′) as before, we see that the two equations are identical with

(6a′)
$$\alpha^*_{jg} = \alpha_{jg} + (1 - \lambda_{jg})\Pi_0$$

(6b′)
$$\phi_{jgp} = [1 - (\lambda_{jg} + \eta_{jgp})]\Pi_p + \alpha_{jg} - \alpha^*_{jg}.$$

Now, we have from the first stage estimation of equation (4′) the estimated parameters $\hat{\alpha}^*_{jg}$, $\hat{\delta}_{jgi}$, $\hat{\phi}_{jgp}$, $\hat{\lambda}_{jg}$, $\hat{\eta}_{jgp}$. Given these, we may specify the following second stage dummy variables regression equation for estimating Π_p, $p = 0, 1, 2, \ldots, R$:

(7′)
$$\hat{\phi}_{jgr} = \sum_{p=1}^{R} \Pi_p \{1 - (\hat{\lambda}_{jg} + \hat{\eta}_{jgp})\} S_p - \Pi_0 (1 - \hat{\lambda}_{jg}) + \varepsilon_{jgr}, \quad g = 1, 2, \ldots, G$$

where $S_p = 1$ for $p = r$ and 0 otherwise. Note that (7′) may be estimated separately for each of the G item groups yielding G different sets of estimates of the Π's, viz., $\hat{\Pi}_{0g}, \hat{\Pi}_{1g}, \ldots, \hat{\Pi}_{Rg}$, $g = 1, 2, \ldots, G$. We have to ensure $\Pi_{pg} - \Pi_{0g} = \varphi_p$, $p = 1, 2, \ldots, R$ for every $g = 1, 2, \ldots, G$. To do so, we rewrite (7′) as follows:

(8′)
$$\hat{\phi}_{jgr} = \sum_{p=1}^{R} (\Pi_{0g} + \varphi_p)\{1 - (\hat{\lambda}_{jg} + \hat{\eta}_{jgp})\} S_p - \Pi_{0g}(1 - \hat{\lambda}_{jg}) + \varepsilon_{jgr}$$

$$= \sum_{p=1}^{R} \varphi_p \{1 - (\hat{\lambda}_{jg} + \hat{\eta}_{jgp})\} S_p + \Pi_{0g}\left[\sum_{p=1}^{R} \{1 - (\hat{\lambda}_{jg} + \hat{\eta}_{jgp})\} S_p - (1 - \hat{\lambda}_{jg})\right] + \varepsilon_{jgr}$$

$$= \sum_{p=1}^{R} \varphi_p \{1 - (\hat{\lambda}_{jg} + \hat{\eta}_{jgp})\} S_p + \sum_{k=1}^{G} \Pi_{0k}[-\hat{\eta}_{jkr}] S'_k + \varepsilon_{jgr}$$

where S'_k's are item group dummy variables. Let us call $X_{jgp} = \{1 - (\hat{\lambda}_{jg} + \hat{\eta}_{jgp}) S_p$ and $Z_{jkr} = [-\hat{\eta}_{jkr}] S'_k$ and write (8′) as

(9′)
$$\hat{\phi}_{jgr} = \sum_{p=1}^{R} \varphi_p X_{jgp} + \sum_{k=1}^{G} \Pi_{0k} Z_{jkr} + \varepsilon_{jgr}$$

for $j = 1, 2, \ldots, M_g$; $g = 1, 2, \ldots, G$; $r = 1, 2, \ldots, R$; $h = 1, 2, \ldots, N_r$.

Inclusion of Item Groups Having Data on Expenditure Only

Let there be M_0 item groups for which only expenditure data are available (e.g. item group sub-totals like expenditure on transport services, education. recreation, health etc.). Let us denote these together as item group $g = 0$. Consider the value share function for the j-th item of this group and specify the empirical value share function to be of the form

(a) $$w_{j0rh} = \alpha^*_{j0} + \sum_{i=1}^{4} \delta_{j0i} n_{irh} + \sum_{p=1}^{R} \phi_{j0p} S_p + \lambda_{j0} y_{rh} + \sum_{p=1}^{R} \eta_{j0p} y_{ph} S_p + \varepsilon_{j0rh}$$

where w_{j0rh} denotes the observed value share for the j-th item group for the h-th household of region r. Given the presence of regional price variations, we may alternatively express the value share function as

(a′) $$w_{j0rh} = \alpha_j + \sum_{i=1}^{4} \delta_{j0i} n_{irh} + (\lambda_{j0} + \eta_{j0r})(y_{rh} - \Pi_r) + \varepsilon_{j0rh}$$

$$= (\alpha_j + (1 - \lambda_{j0})\Pi_0) + \sum_{i=1}^{4} \delta_{j0i} n_{irh} + [\{-\lambda_{j0} + \eta_{j0r}\}\Pi_r - (1 - \lambda_{j0})\Pi_0]$$
$$+ \lambda_{j0} y_{rh} + \eta_{j0r} y_{rh} + \varepsilon_{jorh}.$$

Since (a) and (a′) are equivalent expressions, we have

(b) $$\alpha^*_{j0} = (\alpha_j + (1 - \lambda_{j0})\Pi_0)$$

and

(c) $$\phi_{j0r} = [\{-\lambda_{j0} + \eta_{j0r}\}\Pi_r - (1 - \lambda_{j0})\Pi_0]$$

Now, given the first stage estimation of equation (a), the parameter estimates $\hat{\phi}_{j0r}$, $\hat{\lambda}_{j0}$, η_{j0r} are available. Using these, Π_r, $r = 0, 1, 2, \ldots, R$ may, in principle, be estimated from the following dummy variables regression equation

(d) $$\hat{\phi}_{j0r} = \sum_{p=1}^{R} \Pi_p (-\hat{\lambda}_{j0} + \hat{\eta}_{j0p}) S_p - \Pi_0 (1 - \hat{\lambda}_{j0}) + \varepsilon_{j0r}$$

Let us call this estimate of Π's as $\hat{\Pi}_{p0}$, $p = 0, 1, 2, \ldots, R$. To ensure that $\hat{\Pi}_{p0} - \hat{\Pi}_{00} = \hat{\varphi}_p$, $p = 1, 2, \ldots, R$ holds (where $\hat{\varphi}_p$'s are estimated from (9)), we may rewrite (d) as

(e) $$\hat{\phi}_{j0r} = \sum_{p=1}^{R} \varphi_p (-\hat{\lambda}_{j0} + \hat{\eta}_{j0p}) S_p + \Pi_{00} \left[\sum_{p=1}^{R} (-\hat{\lambda}_{j0} + \hat{\eta}_{j0p}) S_p - (1 - \hat{\lambda}_{j0}) \right] + \varepsilon_{j0r}$$

$$= \sum_{p=1}^{R} \varphi_p (-\hat{\lambda}_{j0} + \hat{\eta}_{j0p}) S_p + \Pi_{00} (\hat{\eta}_{j0r} - 1) + \varepsilon_{j0r}$$

$$= \sum_{p=1}^{R} \varphi_p X_{j0p} + \Pi_{00} Z_{j0r} + \varepsilon_{j0r}$$

Now, equation (e) may be included in equation (9′), and (9′) may be modified as

$$(9'') \qquad \hat{\phi}_{jgr} = \sum_{p=1}^{R} \varphi_p X_{jgp} + \sum_{k=0}^{G} \Pi_{0k} Z_{jkr} + \varepsilon_{jgr}$$

for $j = 1, 2, \ldots, M_g$; $g = 0, 1, 2, \ldots, G$; $r = 1, 2, \ldots, R$; $h = 1, 2, \ldots, N_r$, which is the required extended form of the second stage equation.

It may, however, be mentioned that although the procedure is conceptually reasonable, its applicability should crucially depend upon sensibility of the estimates obtained. This is essentially because information on expenditure share for these items thrown up by an expenditure survey may be unreliable.

References

Aten, Betina and T. Menezes, "Poverty Price Levels: An Application to Brazilian Metropolitan Areas," World Bank ICP Conference, Washington, D.C., March 11–15, 2002.

Balk, B. M., "A Comparison of Ten Methods for Multilateral International Price and Volume Comparison," *Journal of Official Statistics*, 12, 199–222, 1996.

Barten, A. P., "Family Composition, Prices and Expenditure Patterns," in P. E. Hart, G. Mills and J. K. Whitaker (eds), *Econometric Analysis for National Economic Planning*, Butterworths, London, 277–92, 1964.

Bhattacharya, N., G. S. Chattejee, and P. Pal, "Variations in Level of Living Across Regions and Social Groups in India, 1963–64 and 1973–74," in T. N. Srinivasan and P. K. Bardhan (eds), *Rural Poverty in South Asia*, Oxford University Press, 1988.

Bhattacharyya, S. S., P. D. Joshi, and A. B. Roychowdhury, "Regional Price Indices Based on NSS 25th Round Consumer Expenditure Data," *Sarvekshana, Journal of the NSS Organisation*, 3(4), 107–21, 1980.

Clements, K. W. and H. Y. Izan, "A Note on Estimating Divisia Index Numbers," *International Economic Review*, 22, 745–7, 1981.

Coondoo, D. and S. Saha, "Between-State Differentials in Rural Consumer Prices in India: An Analysis of Intertemporal Variations," *Sankhya*, Series B, 52(3), 347–60, 1990.

Deaton, A. S. and A. Tarozzi, "Prices and Poverty in India," Research Program in Development Studies, Princeton University, 2000.

Diewert, W. E., "On the Stochastic Approach to Index Numbers," Discussion Paper No. 95-31, Department of Economics, University of British Columbia, Vancouver, BC, V6T1Z1, Canada, 1995.

———, "Axiomatic and Economic Approaches to International Comparisons," in A. Heston and R. E. Lipsey (eds), *International and Interarea Comparisons of Income, Output and Prices*, University of Chicago Press, Chicago, 13–87, 1999.

———, "Weighted Country Product Dummy Variable Regressions and Index Number Formulae," Discussion Paper No. 02-15, Department of Economics, University of British Columbia, Vancouver, BC, V6T1Z1, Canada, 2002.

Dubey, A. and S. Gangopadhyay, "Counting the Poor: Where are the Poor in India?" *Sarvekshana*, Analytical Report No. 1, Department of Statistics, Government of India, 1998.

Geary, R. C., "A Note on Comparison of Exchange Rate and Purchasing Power Parities Between Countries," *Journal of the Royal Statistical Society*, 121(Part 1), 97–9, 1958.

Hill, R. J., "A Taxonomy of Multilateral Methods for Making International Comparisons of Prices and Quantities," *Review of Income and Wealth*, 43(1), 49–69, 1997.

Hill, R. C., J. R. Knight, and C. F. Sirmans, "Estimating Capital Asset Pricing Indexes," *Review of Economics and Statistics*, 79, 226–33, 1997.

Judge, G., R. Hill, W. Griffiths, H. Lutkepohl, and T. Lee, *Introduction to the Theory and Practice of Econometrics*, 2nd edition, Wiley, 1988.

Khamis, S. H., "Properties and Conditions for the Existence of a New Type of Index Numbers," *Sankhya*, Series B, 32 (Parts 1 & 2), 81–98, 1972.

Kokoski, M. F., B. R. Moulton, and K. D. Zeischang, "Interarea Price Comparisons for Heterogeneous Goods and Several Levels of Commodity Aggregation," in R. E. Lipsey and A. Heston (eds), *International and Interarea Comparisons of Prices, Income and Output*, National Bureau of Economic Research, Chicago University Press, 327–64, 1999.

Koo, J., K. R. Phillips, and F. D. Sigalla, "Measuring Regional Cost of Living," *Journal of Business and Economic Statistics*, 18(1), 127–36, 2000.

Kravis, I. B., A. Heston, and R. Summers, *International Comparison of Real Product and Purchasing Power*, John Hopkins University Press, Baltimore, 1978.
Meenakshi, J. V. and R. Ray, "Impact of Household Size, Family Composition and Socio Economic Characteristics on Poverty in Rural India," *Journal of Policy Modeling*, 24, 539–59, 2002.
Moulton, B. R, "Interarea Indexes of the Cost of Shelter using Hedonic Quality Adjustment Techniques," *Journal of Econometrics*, 68(1), 181–204, 1995.
Prais, S. J. and H. S. Houthakker, *The Analysis of Family Budgets*, Cambridge University Press, Cambridge (2nd edition, 1971), 1955.
Prasada Rao, D. S., "On the Equivalence of the Generalised Country-Product-Dummy (CPD) Method and the Rao-System for Multilateral Comparisons," Working Paper No. 5, Center for International Comparisons, University of Pennsylvania, Philadelphia, 1996.
———, "Aggregation Methods for International Comparison of Purchasing Power Parities and Real Income: Analytical Issues and Some Recent Developments," *Proceedings of the International Statistical Institute*, 51st Session, 197–200, 1997.
Prasada Rao, D. S., "Weighted EKS and Generalised CPD Methods for Aggregation at Basic Heading Level and Above Basic Heading Level," Joint World Bank–OECD Seminar on Purchasing Power Parities—Recent Advances in Methods and Applications, Washington D.C., 2001.
Selvanathan, E. A. and D. S. P. Rao, *Index Numbers: A Stochastic Approach*, The University of Michigan Press, Ann Arbor, 1994.
Sen, A., "Real National Income," *The Review of Economic Studies*, 43(1), 19–39, 1976.
Summers, R., "International Price Comparisons Based Upon Incomplete Data," *Review of Income and Wealth*, 19(1), 1–16, 1973.
Triplett, J. E., "Hedonic Valuation of 'Unpriced' Banking Services: Application to National Accounts and Consumer Price Indexes," Draft paper presented at NBER, Summer Institute, 2000.