INDEX NUMBER THEORY AND MEASUREMENT ECONOMICS

By W.E. Diewert, January, 2015.

CHAPTER 5: The Theory of the Cost of Living Index: The Many Consumer Case

1. Introduction

In previous chapters, we have considered the theory of the cost of living index for only a single consumer or household. In this chapter[1], we consider some of the problems involved in the construction of a superlative index when there are many households or regions in the economy and our goal is the production of a national index. In our algebra below, we allow for an arbitrary number of households, H say, so in principle, each household in the economy under consideration could have its own consumer price index. However, in practice, it will be necessary to group households into various classes and within each class, it will be necessary to assume that the group of households in the class behaves as if it were a single household in order to apply the economic approach to index number theory. Our partition of the economy into H household classes can also be given a regional interpretation: each household class can be interpreted as a group of households within a region of the country under consideration.

2. Plutocratic Cost of Living Indexes and Observable Bounds

In this section, we will consider an economic approach to the CPI that is based on the *plutocratic cost of living index* that was originally defined by Prais (1959). This concept was further refined by Pollak (1980; 276) (1981; 328) who defined his *Scitovsky-Laspeyres cost of living index* as the ratio of total expenditure required to enable each household in the economy under consideration to attain its base period indifference surface at period 1 prices to that required at period 0 prices. In the following paragraph, we will make various assumptions and explain this concept more fully.

Suppose that there are H households (or regions) in the economy and suppose further that there are N commodities in the economy in periods 0 and 1 that households consume *and* that we wish to include in our definition of the cost of living. Denote an N dimensional vector of commodity consumption in a given period by $q \equiv (q_1,q_2,\ldots, q_N)$ as usual. Denote the vector of period t market prices faced by household h by $p_h^t \equiv (p_{h1}^t,p_{h2}^t,\ldots,p_{hN}^t)$ for t = 0,1. Note that we are *not* assuming that each household faces the same vector of commodity prices. In addition to the market commodities that are in the vector q, we assume that each household is affected by an M dimensional vector of *environmental*[2] *or demographic*[3] *variables or public goods*, $e \equiv (e_1,e_2,\ldots,e_M)$. We suppose that there are H households (or regions) in the economy during periods 0 and 1 and the preferences of household h over different combinations of market commodities q and environmental variables e can be represented by the continuous utility function $f^h(q,e)$ for h =

[1] Much of the material in this section is based on Diewert (1983) (2000) (2001).
[2] This is the terminology used by Pollak (1989; 181) in his model of the conditional cost of living concept.
[3] Caves, Christensen and Diewert (1982; 1409) used the terms *demographic variables* or *public goods* to describe the vector of conditioning variables e in their generalized model of the Konüs price index or cost of living index.

1,2,…,H.[4] For periods t = 0,1 and for households h = 1,2,…,H, it is assumed that the observed household h consumption vector $q_h^t \equiv (q_{h1}^t,\ldots,q_{hN}^t)$ is a solution to the following household h expenditure minimization problem:

$$(1)\ \ \min_q \{ p_h^t \bullet q : f^h(q,e_h^t) \geq u_h^t \} \equiv C^h(u_h^t,e_h^t, p_h^t)\ ;\ t = 0,1;\ h = 1,2,\ldots H$$

where e_h^t is the environmental vector facing household h in period t, $u_h^t \equiv f^h(q_h^t,e_h^t)$ is the utility level achieved by household h during period t and C^h is the cost or expenditure function that is dual to the utility function f^h.[5] Basically, these assumptions mean that each household has *stable preferences* over the same list of commodities during the two periods under consideration, the same households appear in each period and each household chooses its consumption bundle in the most cost efficient way during each period, conditional on the environmental vector that it faces during each period. Note again that the household (or regional) prices are in general different across households (or regions).

With the above assumptions in mind, we generalize Pollak (1980) (1981) and Diewert (1983; 190)[6] and define the class of *conditional plutocratic cost of living indexes*, $P^*(p^0,p^1,u,e_1,e_2,\ldots,e_H)$, pertaining to periods 0 and 1 for the arbitrary utility vector of household utilities $u \equiv (u_1,u_2,\ldots,u_H)$ and for the arbitrary vectors of household environmental variables e_h for h = 1,2,…,H as follows:

$$(2)\ \ P^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u,e_1,e_2,\ldots,e_H) \equiv \sum_{h=1}^H C^h(u_h,e_h,p_h^1) / \sum_{h=1}^H C^h(u_h,e_h,p_h^0)\ .$$

The numerator on the right hand side of (2) is the sum over households of the minimum cost, $C^h(u_h,e_h,p_h^1)$, for household h to achieve the arbitrary utility level u_h, given that the household h faces the arbitrary vector of household h environmental variables e_h and also faces the period 1 vector of prices p_h^1. The denominator on the right hand side of (2) is the sum over households of the minimum cost, $C^h(u_h,e_h,p_h^0)$, for household h to achieve the *same* arbitrary utility level u_h, given that the household faces the *same* arbitrary vector of household h environmental variables e_h and also faces the period 0 vector of prices p_h^0. Thus in the numerator and denominator of (2), only the price variables are different, which is precisely what we want in a theoretical definition of a consumer price index.

We now specialize the general definition (2) by replacing the general utility vector u by either the period 0 vector of household utilities $u^0 \equiv (u_1^0,u_2^0,\ldots u_H^0)$ or the period 1 vector of household utilities $u^1 \equiv (u_1^1,u_2^1,\ldots u_H^1)$. We also specialize the general definition (2) by replacing the general household environmental vectors $(e_1,e_2,\ldots e_H) \equiv e$ by either the period 0 vector of household environmental variables $e^0 \equiv (e_1^0,e_2^0,\ldots e_H^0)$ or the period 1 vector of household environmental variables $e^1 \equiv (e_1^1,e_2^1,\ldots,e_H^1)$. The choice of the base period vector of utility levels and base period environmental variables leads to the *Laspeyres conditional plutocratic cost of living index*,

[4] We assume that each $f^h(q,e)$ is continuous and increasing in the components of q and e and is concave in the components of q.

[5] In order to minimize notational clutter, in this section we use the notation $p \bullet q \equiv \sum_{n=1}^N p_n q_n$ as the inner product between the vectors p and q, rather than write out the summations.

[6] These authors provided generalizations of the plutocratic cost of living index due to Prais (1959). Pollak and Diewert did not include the environmental variables in their definitions of a group cost of living index.

$P^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u^0,e^0)$[7], while the choice of the period 1 vector of utility levels and period 1 environmental variables leads to the *Paasche conditional plutocratic cost of living index*, $P^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u^1,e^1)$. It turns out that these last two indexes satisfy some interesting inequalities, which we derive below.

Using definition (2), the *Laspeyres plutocratic conditional cost of living index*, $P^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u^0,e^0)$, may be written as follows:

$$
\begin{aligned}
(3)\ \ & P^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u^0,e_1^0,e_2^0,\ldots,e_H^0) \\
& \equiv \textstyle\sum_{h=1}^H C^h(u_h^0,e_h^0,p_h^1) / \sum_{h=1}^H C^h(u_h^0,e_h^0,p_h^0) \\
& = \textstyle\sum_{h=1}^H C^h(u_h^0,e_h^0,p_h^1) / \sum_{h=1}^H p_h^0\bullet q_h^0 \quad \text{using (1) for } t = 0 \\
& \leq \textstyle\sum_{h=1}^H p_h^1\bullet q_h^0 / \sum_{h=1}^H p_h^0\bullet q_h^0 \\
& \qquad \text{since } C^h(u_h^0,e_h^0,p_h^1) \equiv \min_q \{ p_h^1\bullet q : f^h(q,e_h^0) \geq u_h^0 \} \leq p^1\bullet q_h^0 \text{ and } q_h^0 \\
& \qquad \text{is feasible for the cost minimization problem for } h = 1,2,\ldots,H \\
& \equiv P_{PL}
\end{aligned}
$$

where P_{PL} is defined to be the observable (in principle) *plutocratic Laspeyres price index*, $\sum_{h=1}^H p_h^1\bullet q_h^0 / \sum_{h=1}^H p_h^0\bullet q_h^0$, which uses the individual vectors of household or regional quantities for period 0, $(q_1^0,\ldots,q_H^0)$, as quantity weights.[8] If prices are equal across households (or regions), so that

$$(4)\ \ p_h^t = p^t \quad \text{for } t = 0,1 \text{ and } h = 1,2,\ldots,H,$$

then the plutocratic (or disaggregated) Laspeyres price index P_{PL} collapses down to the usual *aggregate Laspeyres index*, P_L; i.e., if (4) holds, then P_{PL} in (3) becomes

$$
\begin{aligned}
(5)\ \ P_{PL} & \equiv \textstyle\sum_{h=1}^H p_h^1\bullet q_h^0 / \sum_{h=1}^H p_h^0\bullet q_h^0 \\
& = p^1\bullet \textstyle\sum_{h=1}^H q_h^0 / p^0\bullet \sum_{h=1}^H q_h^0 \\
& = p^1\bullet q^0 / p^0\bullet q^0 \\
& \equiv P_L
\end{aligned}
$$

where the total quantity vector in period t is defined as

$$(6)\ \ q^t \equiv \textstyle\sum_{h=1}^H q_h^t \quad \text{for } t = 0,1.$$

The inequality (3) says that the theoretical Laspeyres plutocratic conditional cost of living index, $P^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u^0,e^0)$, is bounded from above by the observable (in principle) plutocratic or disaggregated Laspeyres price index P_{PL}. The special case of inequality (3) when the equal prices assumption (4) holds was first obtained by Pollak (1989; 182) for the case of one household

[7] This is the concept of a cost of living index that Triplett (2000; 27) found most useful for measuring inflation: "One might want to produce a COL *conditional* on the base period's weather experience…. In this case, the unusually cold winter does not affect the *conditional* COL subindex that holds the environment constant. … the COL subindex that holds the environment constant is probably the COL concept that is most useful for an anti-inflation policy." Hill (1999; 4) endorsed this point of view.

[8] Thus the plutocratic Laspeyres index can be regarded as an ordinary Laspeyres index except that each commodity in each region is regarded as a separate commodity.

with environmental variables and by Pollak (1980; 276) for the many household case but where the environmental variables are absent from the household utility and cost functions.

In a similar manner, specializing definition (2), *the Paasche conditional plutocratic cost of living index*, $P^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u^1,e^1)$, may be written as follows:

$$
\begin{aligned}
(7)\quad & P^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u^1,e_1^1,e_2^1,\ldots,e_H^1) \\
& \equiv \textstyle\sum_{h=1}^H C^h(u_h^1,e_h^1,p_h^1) / \sum_{h=1}^H C^h(u_h^1,e_h^1,p_h^0) \\
& = \textstyle\sum_{h=1}^H p_h^1\bullet q_h^1 / \sum_{h=1}^H C^h(u_h^1,e_h^1,p_h^0) && \text{using (1) for } t = 1 \\
& \geq \textstyle\sum_{h=1}^H p_h^1\bullet q_h^1 / \sum_{h=1}^H p_h^0\bullet q_h^1 && \text{using a feasibility argument} \\
& \equiv P_{PP}
\end{aligned}
$$

where P_{PP} is defined to be the *plutocratic or disaggregated (over households) Paasche price index*, $\sum_{h=1}^H p_h^1\bullet q_h^1 / \sum_{h=1}^H p_h^0\bullet q_h^1$, which uses the individual vectors of household quantities for period 1, $(q_1^1,\ldots,q_H^1)$, as quantity weights.

If prices are equal across households (or regions), so that assumptions (6) hold, then the disaggregated Paasche price index P_{PP} collapses down to the usual aggregate Paasche index, P_P; i.e., if (6) holds, then P_{PP} in (7) becomes

$$
\begin{aligned}
(8)\quad P_{PP} & \equiv \textstyle\sum_{h=1}^H p_h^1\bullet q_h^1 / \sum_{h=1}^H p_h^0\bullet q_h^1 \\
& = p^1\bullet \textstyle\sum_{h=1}^H q_h^1 / p^0\bullet \sum_{h=1}^H q_h^1 \\
& = p^1\bullet q^1 / p^0\bullet q^1 \\
& \equiv P_P.
\end{aligned}
$$

Returning to the inequality (7), we see that the theoretical Paasche conditional plutocratic cost of living index, $P^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u^1,e^1)$, is bounded from below by the observable plutocratic or disaggregated Paasche price index P_{PP}. Diewert (1983; 191) first obtained the inequality (7) for the case where the environmental variables were absent from the household utility and cost functions and prices were equal across households.

In the following section, we shall show how to obtain a theoretical plutocratic cost of living index that is bounded from above and below rather than the theoretical indices that just have the one sided bounds in (3) and (7).

3. The Fisher Plutocratic Price Index

Using the inequalities (3) and (7) and the continuity properties of the conditional plutocratic cost of living $P^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u,e)$ defined by (2), it is possible to modify the method of proof used by Konüs (1924) and Diewert (1983; 191) and establish the following result:[9]

[9] Note that the household cost functions must be continuous in the environmental variables which is a real restriction on the types of environmental variables which can be accommodated by the result. Thus if household preferences change discontinuously as the season of the year changes, then Proposition 1 would not be valid.

Proposition 1: Under our assumptions, there exists a reference utility vector $u^* \equiv (u_1^*,u_2^*,\ldots,u_H^*)$ such that the household h reference utility level u_h^* lies between the household h period 0 and 1 utility levels, u_h^0 and u_h^1 respectively for h = 1,…,H, and there exist household environmental vectors $e_h^* \equiv (e_{h1}^*,e_{h2}^*,\ldots,e_{hM}^*)$ such that the household h reference mth environmental variable e_{hm}^* lies between the household h period 0 and 1 levels for the mth environmental variable, e_{hm}^0 and e_{hm}^1 respectively for m = 1,2,…,M and h = 1,…,H, and the conditional plutocratic cost of living index $P^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u^*,e^*)$ evaluated at this intermediate reference utility vector u^* and the intermediate reference vector of household environmental variables $e^* \equiv (e_1^*,e_2^*,\ldots,e_H^*)$ lies between the observable (in principle) plutocratic Laspeyres and Paasche price indexes, P_{PL} and P_{PP}, defined above by the last equalities in (3) and (7).

Proof: Utilize the method of proof used in Proposition 1 in Chapter 4 by defining g((λ) for $0 \le \lambda \le 1$ by $g(\lambda) \equiv P^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1, (1-\lambda)u^0+\lambda u^1, (1-\lambda)e_1^0+\lambda e_1^1,\ldots, (1-\lambda)e_H^0+\lambda e_H^1)$. Note that $g(0) = P^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u^0,e_1^0,e_2^0,\ldots,e_H^0)$ and $g(1) = P^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u^1,e_1^1,e_2^1,\ldots,e_H^1)$. There are 24 = (4)(3)(2)(1) possible a priori inequality relations that are possible between the four numbers g(0), g(1), P_{PL} and P_{PP}. However, the inequalities (3) and (7) above imply that $g(0) \le P_L$ and $P_P \le g(1)$. This means that there are only six possible inequalities between the four numbers:

(9) $g(0) \le P_{PL} \le P_{PP} \le g(1)$;
$g(0) \le P_{PP} \le P_{PL} \le g(1)$;
$g(0) \le P_{PP} \le g(1) \le P_{PL}$;
$P_{PP} \le g(0) \le P_{PL} \le g(1)$;
$P_{PP} \le g(1) \le g(0) \le P_{PL}$;
$P_{PP} \le g(0) \le g(1) \le P_{PL}$.

Using the assumptions that: (a) the consumer's utility function f is continuous over its domain of definition; (b) the utility function is increasing in the components of q and hence is subject to local nonsatiation and (c) the price vectors p^t have strictly positive components, it is possible to use Debreu's (1959; 19) Maximum Theorem (see also Diewert (1993; 112-113) for a statement of the Theorem) to show that the household cost functions $C^h(u_h,e_h,p_h)$ will be continuous in the components of u_h and e_h. Thus using definition (2), it can be seen that $P^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u,e_1,e_2,\ldots,e_H)$ will also be continuous in the components of the vectors $u,e_1,e_2,\ldots,e_H$. Hence g(λ) is a continuous function of λ and assumes all intermediate values between g(0) and g(1). By inspecting the inequalities (9) above, it can be seen that we can choose λ between 0 and 1, λ^* say, such that $P_{PL} \le g(\lambda^*) \le P_{PP}$ or such that $P_{PP} \le g(\lambda^*) \le P_{PL}$. Q.E.D.

The above result tells us that *the theoretical national plutocratic conditional consumer price index* $P^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u^*,e^*)$ lies between the plutocratic or disaggregated Laspeyres index P_{PL} and the plutocratic or disaggregated Paasche index P_{PP}. Hence if P_{PL} and P_{PP} are not too different, a good point approximation[10] to the theoretical national plutocratic consumer price index will be the *plutocratic or disaggregated Fisher index* P_{PF} defined as $P_{PF} \equiv [P_{PL}\, P_{PP}]^{1/2}$.

[10] Recall Proposition 2 in Chapter 1 for a more formal justification for the use of the Fisher index as an appropriate average of the Paasche and Laspeyres indexes.

The plutocratic Fisher price index P_{PF} is computed just like the usual Fisher price index, except that each commodity in each region (or for each household) is regarded as a separate commodity. Of course, this index will satisfy the time reversal test.

Since statistical agencies do not calculate Laspeyres, Paasche and Fisher price indexes by taking inner products of price and quantity vectors, it will be useful to obtain formulae for the Laspeyres and Paasche indices that depend only on price relatives and expenditure shares. In order to do this, we need to introduce some notation. Define the *expenditure share of household h on commodity i in period t* as

$$(10)\quad s_{hi}^t \equiv p_{hi}^t q_{hi}^t / \sum_{n=1}^N p_{hn}^t q_{hn}^t ; \qquad t = 0,1 ;\ h = 1,2,\ldots,H ;\ i = 1,2,\ldots,N.$$

Define the *expenditure share of household h in total period t consumption* as:

$$\begin{aligned}(11)\quad S_h^t &\equiv \sum_{n=1}^N p_{hn}^t q_{hn}^t / \sum_{k=1}^H \sum_{n=1}^N p_{kn}^t q_{kn}^t \\ &= p_h^t \bullet q_h^t / \sum_{k=1}^H p_k^t \bullet q_k^t \qquad t = 0,1 ;\ h = 1,2,\ldots,H.\end{aligned}$$

Finally, define the *national expenditure share of commodity i in period t* as:

$$\begin{aligned}(12)\quad \sigma_i^t &\equiv \sum_{h=1}^H p_{hi}^t q_{hi}^t / \sum_{k=1}^H p_k^t \bullet q_k^t && t = 0,1 ;\ i = 1,2,\ldots,N \\ &= \sum_{h=1}^H [p_{hi}^t q_{hi}^t / p_h^t \bullet q_h^t][p_h^t \bullet q_h^t / \sum_{k=1}^H p_k^t \bullet q_k^t] \\ &= \sum_{h=1}^H s_{hi}^t\ p_h^t \bullet q_h^t / \sum_{k=1}^H p_k^t \bullet q_k^t && \text{using definitions (10)} \\ &= \sum_{h=1}^H s_{hi}^t S_h^t && \text{using definitions (11).}\end{aligned}$$

The *Laspeyres price index for household h* (or region h) is defined as:

$$\begin{aligned}(13)\quad P_{Lh} &\equiv p_h^1 \bullet q_h^0 / p_h^0 \bullet q_h^0 && h = 1,2,\ldots,H \\ &= \sum_{i=1}^N (p_{hi}^1 / p_{hi}^0)\, p_{hi}^0 q_{hi}^0 / p_h^0 \bullet q_h^0 \\ &= \sum_{i=1}^N s_{hi}^0 (p_{hi}^1 / p_{hi}^0) && \text{using definitions (10).}\end{aligned}$$

Referring back to (3), the *plutocratic national Laspeyres price index* P_{PL} can be rewritten as follows:

$$\begin{aligned}(14)\quad P_{PL} &\equiv (\sum_{h=1}^H p_h^1 \bullet q_h^0) / (\sum_{h=1}^H p_h^0 \bullet q_h^0) \\ &= \sum_{h=1}^H [p_h^1 \bullet q_h^0 / p_h^0 \bullet q_h^0][p_h^0 \bullet q_h^0 / \sum_{h=1}^H p_h^0 \bullet q_h^0] \\ &= \sum_{h=1}^H [p_h^1 \bullet q_h^0 / p_h^0 \bullet q_h^0]\, S_h^0 && \text{using definitions (11) with } t = 0 \\ (15)\quad &= \sum_{h=1}^H S_h^0 P_{Lh} && \text{using definitions (13)} \\ &= \sum_{h=1}^H S_h^0 \sum_{i=1}^N s_{hi}^0 (p_{hi}^1 / p_{hi}^0) && \text{using the last line of (13)} \\ (16)\quad &= \sum_{h=1}^H \sum_{i=1}^N S_h^0 s_{hi}^0 (p_{hi}^1 / p_{hi}^0) && \text{rearranging terms.}\end{aligned}$$

Equation (15) shows that the plutocratic national Laspeyres price index is equal to a (period 0) regional expenditure share weighted average of the regional Laspeyres price indices. Equation (16) shows that the national Laspeyres price index is equal to a period 0 expenditure share weighted average of the regional price relatives, (p_{hi}^1/ p_{hi}^0), where the corresponding weight, $S_h^0 s_{hi}^0$, is the period 0 national expenditure share of commodity i for household h.

The *Paasche price index for region h* (or household h) is defined as:

$$
\begin{aligned}
(17)\quad P_{Ph} &\equiv p_h^1 \bullet q_h^1 / p_h^0 \bullet q_h^1 && h = 1,2,\ldots,H \\
&= 1 / \textstyle\sum_{i=1}^N (p_{hi}^0/ p_{hi}^1)\, p_{hi}^1 q_{hi}^1 / p_h^1 \bullet q_h^1 \\
&= 1 / \textstyle\sum_{i=1}^N s_{hi}^1 (p_{hi}^1/ p_{hi}^0)^{-1} && \text{using definitions (10)} \\
&= [\textstyle\sum_{i=1}^N s_{hi}^1 (p_{hi}^1/ p_{hi}^0)^{-1}]^{-1} .
\end{aligned}
$$

Referring back to (7), the *plutocratic national Paasche price index* P_{PP} can be rewritten as follows:

$$
\begin{aligned}
(18)\quad P_{PP} &\equiv (\textstyle\sum_{k=1}^H p_k^1 \bullet q_k^1)/(\sum_{h=1}^H p_h^0 \bullet q_h^1) \\
&= 1 / \{\textstyle\sum_{h=1}^H [p_h^0 \bullet q_h^1 / p_h^1 \bullet q_h^1][p_h^1 \bullet q_h^1 / \sum_{k=1}^H p_k^1 \bullet q_k^1]\} \\
&= 1 / \textstyle\sum_{h=1}^H [p_h^1 \bullet q_h^0 / p_h^0 \bullet q_h^0]^{-1} S_h^1 && \text{using definitions (11) with } t = 1 \\
(19)\quad &= [\textstyle\sum_{h=1}^H S_h^1 P_{Ph}^{-1}]^{-1} && \text{using definitions (17)} \\
&= [\textstyle\sum_{h=1}^H S_h^1 \sum_{i=1}^N s_{hi}^1 (p_{hi}^1/ p_{hi}^0)^{-1}]^{-1} && \text{using the last line of (17)} \\
(20)\quad &= [\textstyle\sum_{h=1}^H \sum_{i=1}^N S_h^1 s_{hi}^1 (p_{hi}^1/ p_{hi}^0)^{-1}]^{-1} && \text{rearranging terms.}
\end{aligned}
$$

Equation (19) shows that the national plutocratic Paasche price index is equal to a (period 1) regional expenditure share *weighted harmonic mean* of the regional Paasche price indices. Equation (20) shows that the national Paasche price index is equal to a period 1 expenditure share *weighted harmonic average* of the regional price relatives, (p_{hi}^1/ p_{hi}^0), where the weight for this price relative, $S_h^1 s_{hi}^1$, is the period 1 national expenditure share of commodity i in region h.

Of course, the share formulae for the plutocratic Paasche and Laspeyres indices, P_{PP} and P_{PL}, given by (16) and (20) can now be used to calculate the plutocratic Fisher index, $P_{PF} \equiv [P_{PP}\, P_{PL}]^{1/2}$.

If prices are equal across regions, the formulae (16) and (20) simplify. The formula for the plutocratic Laspeyres index (16) becomes:

$$
\begin{aligned}
(21)\quad P_{PL} &= \textstyle\sum_{h=1}^H \sum_{i=1}^N S_h^0 s_{hi}^0 (p_{hi}^1/ p_{hi}^0) \\
&= \textstyle\sum_{h=1}^H \sum_{i=1}^N S_h^0 s_{hi}^0 (p_i^1/ p_i^0) && \text{using assumptions (4)} \\
&= \textstyle\sum_{i=1}^N \sigma_i^0 (p_i^1/ p_i^0) && \text{using (12) for } t = 0 \\
&= P_L
\end{aligned}
$$

where P_L is the usual aggregate Laspeyres price index based on the assumption that each household faces the same vector of commodity prices; see (5) for the definition of P_L. Under the equal prices across households assumption (4), the formula for the plutocratic Paasche index (20) becomes:

$$
\begin{aligned}
(22)\quad P_{PP} &= [\textstyle\sum_{h=1}^H \sum_{i=1}^N S_h^1 s_{hi}^1 (p_{hi}^1/ p_{hi}^0)^{-1}]^{-1} \\
&= [\textstyle\sum_{h=1}^H \sum_{i=1}^N S_h^1 s_{hi}^1 (p_i^1/ p_i^0)^{-1}]^{-1} && \text{using assumptions (4)} \\
&= [\textstyle\sum_{i=1}^N \sigma_i^1 (p_i^1/ p_i^0)^{-1}]^{-1} && \text{using (12) for } t = 1 \\
&= P_P
\end{aligned}
$$

where P_P is the usual aggregate Paasche price index based on the assumption that each household faces the same vector of commodity prices; see (8) for the definition of P_P.

Thus with the assumption that commodity prices are the same across regions, in order to calculate national Laspeyres and Paasche indexes, we require only "national" price relatives and national commodity expenditure shares for the two periods under consideration. However, if there is regional variation in prices, then the simplified formulae (21) and (22) are not valid and we must use our earlier formulae, (16) and (20).

4. Democratic versus Plutocratic Cost of Living Indexes

The plutocratic indexes considered above weight each household in the economy according to the size of their expenditures in the two periods under consideration. Instead of weighting in this way, it is possible to define theoretical indices (and "practical" approximations to them) that give each household or household group in the economy an *equal weight*. Following Prais (1959), we will call such an index a *democratic index*. In this section, we will rework the plutocratic index number theory developed in sections 2 and 3 above into a democratic framework.

Making the same assumptions as in section 2 above, we define the class of *conditional democratic cost of living indices*, $P_D^*(p^0,p^1,u,e_1,e_2,\ldots,e_H)$, pertaining to periods 0 and 1 for the arbitrary utility vector of household utilities $u \equiv (u_1,u_2,\ldots,u_H)$ and for the arbitrary vectors of household environmental variables e_h for $h = 1,2,\ldots,H$ as follows:

$$(23)\quad P_D^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u,e_1,e_2,\ldots,e_H) \equiv \sum_{h=1}^{H} [1/H]C^h(u_h,e_h,p_h^1)/C^h(u_h,e_h,p_h^0).$$

Thus P_D^* is a simple unweighted arithmetic average of the individual household *conditional cost of living indexes*, $C^h(u_h,e_h,p_h^1)/C^h(u_h,e_h,p_h^0)$. In the numerator and denominator of these conditional indexes, only the price variables are different, which is precisely what we want in a theoretical definition of a consumer price index. If the vector of environmental variables, e_h, is not present in the cost function of household h, then the conditional index $C^h(u_h,e_h,p_h^1)/C^h(u_h,e_h,p_h^0)$ becomes an ordinary Konüs true cost of living index of the type defined earlier in Chapter 3.

We now specialize the general definition (23) by replacing the general utility vector u by either the period 0 vector of household utilities $u^0 \equiv (u_1^0,u_2^0,\ldots u_H^0)$ or the period 1 vector of household utilities $u^1 \equiv (u_1^1,u_2^1,\ldots u_H^1)$. We also specialize the general definition (23) by replacing the general household environmental vectors $(e_1,e_2,\ldots e_H) \equiv e$ by either the period 0 vector of household environmental variables $e^0 \equiv (e_1^0,e_2^0,\ldots e_H^0)$ or the period 1 vector of household environmental variables $e^1 \equiv (e_1^1,e_2^1,\ldots,e_H^1)$. The choice of the base period vector of utility levels and base period environmental variables leads to the *Laspeyres conditional democratic cost of living index*, $P_D^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u^0,e^0)$, while the choice of the period 1 vector of utility levels and period 1 environmental variables leads to the *Paasche conditional democratic cost of living index*, $P_D^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u^1,e^1)$. It turns out that these two democratic indexes satisfy some interesting inequalities, which we derive below.

Specializing definition (23), the *Laspeyres democratic conditional cost of living index*, $P_D^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u^0,e^0)$, may be written as follows:

(24) $P_D{}^*(p_1{}^0,\ldots,p_H{}^0,p_1{}^1,\ldots,p_H{}^1,u^0,e_1{}^0,e_2{}^0,\ldots,e_H{}^0)$

$\equiv \sum_{h=1}^H [1/H]\, C^h(u_h{}^0,e_h{}^0,p_h{}^1) / C^h(u_h{}^0,e_h{}^0,p_h{}^0)$

$= \sum_{h=1}^H [1/H]\, C^h(u_h{}^0,e_h{}^0,p_h{}^1) / p_h{}^0\bullet q_h{}^0$ using (1) for t = 0

$\leq \sum_{h=1}^H [1/H]\, p_h{}^1\bullet q_h{}^0 / p_h{}^0\bullet q_h{}^0$

since $C^h(u_h{}^0,e_h{}^0,p_h{}^1) \equiv \min_q \{ p_h{}^1\bullet q : f^h(q,e_h{}^0) \geq u_h{}^0 \} \leq p^1\bullet q_h{}^0$ and $q_h{}^0$ is feasible for the cost minimization problem for h = 1,2,…,H

$\equiv P_{DL}$

where P_{DL} is defined to be the observable (in principle) *democratic Laspeyres price index*, $\sum_{h=1}^H [1/H]\, p_h{}^1\bullet q_h{}^0 / p_h{}^0\bullet q_h{}^0$, which uses the individual vectors of household or regional quantities for period 0, $(q_1{}^0,\ldots,q_H{}^0)$, as quantity weights.

In a similar manner, specializing definition (23), *the Paasche conditional democratic cost of living index*, $P_D{}^*(p_1{}^0,\ldots,p_H{}^0,p_1{}^1,\ldots,p_H{}^1,u^1,e^1)$, may be written as follows:

(25) $P_D{}^*(p_1{}^0,\ldots,p_H{}^0,p_1{}^1,\ldots,p_H{}^1,u^1,e_1{}^1,e_2{}^1,\ldots,e_H{}^1)$

$\equiv \sum_{h=1}^H [1/H]\, C^h(u_h{}^1,e_h{}^1,p_h{}^1) / C^h(u_h{}^1,e_h{}^1,p_h{}^0)$

$= \sum_{h=1}^H [1/H]\, p_h{}^1\bullet q_h{}^1 / C^h(u_h{}^1,e_h{}^1,p_h{}^0)$ using (1) for t = 1

$\geq \sum_{h=1}^H [1/H]\, p_h{}^1\bullet q_h{}^1 / p_h{}^0\bullet q_h{}^1$ using a feasibility argument

$\equiv P_{DP}$

where P_{DP} is defined to be the *democratic Paasche price index*, $\sum_{h=1}^H [1/H]\, p_h{}^1\bullet q_h{}^1 / p_h{}^0\bullet q_h{}^1$, which uses the individual vector of household h quantities for period 1, $q_h{}^1$, as quantity weights for term h in the summation of individual household Paasche indices. Thus, we see that the theoretical Paasche conditional democratic cost of living index, $P_D{}^*(p_1{}^0,\ldots,p_H{}^0,p_1{}^1,\ldots,p_H{}^1,u^1,e^1)$, is bounded from below by the observable (in principle) democratic Paasche price index P_{DP}. Diewert (1983a; 191) first obtained the inequality (25) for the case where the environmental variables are absent from the household utility and cost functions and prices are equal across households.

We now show how to obtain a theoretical democratic cost of living index that is bounded from above and below by observable indexes. Using the inequalities (24) and (25) and the continuity properties of the conditional democratic cost of living $P_D{}^*(p_1{}^0,\ldots,p_H{}^0,p_1{}^1,\ldots,p_H{}^1,u,e)$ defined by (23), it is possible to modify the method of proof used by Konüs (1924) and Diewert (1983; 191) and establish the following result:

Proposition 2: Under our assumptions, there exists a reference utility vector $u^* \equiv (u_1{}^*,u_2{}^*,\ldots,u_H{}^*)$ such that the household h reference utility level $u_h{}^*$ lies between the household h period 0 and 1 utility levels, $u_h{}^0$ and $u_h{}^1$ respectively for h = 1,…,H, and there exist household environmental vectors $e_h{}^* \equiv (e_{h1}{}^*,e_{h2}{}^*,\ldots,e_{hM}{}^*)$ such that the household h reference mth environmental variable $e_{hm}{}^*$ lies between the household h period 0 and 1 levels for the mth environmental variable, $e_{hm}{}^0$ and $e_{hm}{}^1$ respectively for m = 1,2,…,M and h = 1,…,H, and the conditional democratic cost of living index $P_D{}^*(p_1{}^0,\ldots,p_H{}^0,p_1{}^1,\ldots,p_H{}^1,u^*,e^*)$ evaluated at this intermediate reference utility vector u^* and the intermediate reference vector of household environmental variables $e^* \equiv (e_1{}^*,e_2{}^*,\ldots,e_H{}^*)$ lies between the observable (in principle) democratic Laspeyres and Paasche price indexes, P_{DL} and P_{DP}, defined above by the last equalities in (24) and (25).

The above result tells us that *the theoretical national democratic conditional consumer price index* $P_D^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u^*,e^*)$ lies between the democratic Laspeyres index P_{DL} and the democratic Paasche index P_{DP}. Hence if P_{DL} and P_{DP} are not too different, a good point approximation to the theoretical national democratic consumer price index will be the *democratic Fisher index* P_{DF} defined as:

(26) $P_{DF} \equiv [P_{DL}\, P_{DP}]^{1/2}$.

The democratic Fisher price index P_{DF} will satisfy the time reversal test.

Again, it will be useful to obtain formulae for the democratic Laspeyres and Paasche indexes that depend only on price relatives and expenditure shares. Recall (13) and (17), which defined the Laspeyres and Paasche price indexes for household h, P_{Lh} and P_{Ph}. Substituting (13) into the definition of the democratic Laspeyres index, P_{DL}, leads to the following share type formula:[11]

(27) $P_{DL} = \sum_{h=1}^{H} [1/H] \sum_{i=1}^{N} s_{hi}^0\, (p_{hi}^1/p_{hi}^0)$.

Similarly, substituting (17) into the definition of the democratic Paasche index, P_{DP}, leads to the following share type formula:

(28) $P_{DP} = \sum_{h=1}^{H} [1/H] [\sum_{i=1}^{N} s_{hi}^1\, (p_{hi}^1/p_{hi}^0)^{-1}]^{-1}$.

The formula for the democratic Laspeyres index in the previous paragraph simplifies if we can assume that each household faces the same vector of prices in each of the two periods under consideration. Under this condition, we may rewrite (27) as

(29) $P_{DL} = \sum_{i=1}^{N} s_{di}^0\, (p_i^1/p_i^0)$

where the *period 0 democratic expenditure share for commodity i*, s_{di}^0, is defined as follows:

(30) $s_{di}^0 \equiv \sum_{h=1}^{H} [1/H]\, s_{hi}^0$; $\quad i = 1,\ldots,N.$

Thus s_{di}^0 is simply the arithmetic average (over all households) of the individual household expenditure shares on commodity i during period 0. The formula for the democratic Paasche index does not simplify in the same way, under the assumption that households face the same prices in each period, due to the harmonic form of averaging in (28).

Our conclusion at this point is that democratic and plutocratic Laspeyres, Paasche and Fisher indexes can be constructed by a statistical agency provided that information on household specific

---

[11] Comparing the formula for the democratic Laspeyres index, P_{DL}, with the previous formula (16) for the plutocratic Laspeyres index, P_{PL}, we see that the plutocratic weight for the ith price relative for household h is $S_h^0 s_{hi}^0$ whereas the corresponding democratic weight is $(1/H)s_{hi}^0$. Thus households that have larger base period expenditures and hence bigger expenditure shares S_h^0 get a larger weight in the plutocratic index as compared to the democratic index.

price relatives p_{hi}^1/p_{hi}^0 and expenditures is available for both periods under consideration. If expenditure information is available only for the base period, then only the Laspeyres democratic and plutocratic indices can be constructed.

It is now necessary to discuss a practical problem that faces statistical agencies: namely, that existing household consumer expenditure surveys, which are used in order to form estimates of household expenditure shares, *are not very accurate*. Thus the detailed commodity by region expenditure shares, $S_h^0 s_{hn}^0$ and $S_h^1 s_{hn}^1$, which appear in the formulae for the plutocratic Laspeyres and Paasche indexes are generally measured with very large errors. Similarly, the individual household expenditure shares for the two periods under consideration, s_{hn}^0 and s_{hn}^1, which are required in order to calculate the democratic Laspeyres and Paasche indices defined by (27) and (28) respectively, are also generally measured with substantial errors. Hence, it may lead to less overall error if the regional commodity expenditure shares s_{hn}^t are replaced by the national commodity expenditure shares σ_n^t defined by (12). Whether this approximation is justified would depend on a detailed analysis of the situation facing the statistical agency. In general, complete and accurate information on household expenditure shares will not be available to the statistical agency and hence statistical estimation and smoothing techniques may have to be used in order to obtain expenditure weights that will be used to weight the price relatives collected by the agency.

Problems

1. Define the *conditional cost of living for household h* between periods 0 and 1, conditioning on the household h utility level u_h and the household h environmental vector e_h, as follows:

(a) $P^h(p_h^0,p_h^1,u_h,e_h) \equiv C^h(u_h,e_h,p_h^1) / C^h(u_h,e_h,p_h^0)$ for $h = 1,2,...,H$.

Show that the national *conditional plutocratic cost of living index*, $P^*(p^0,p^1,u,e_1,e_2,\ldots,e_H)$ defined by (2) above can be written as a share weighted average of the individual household (or regional) conditional cost of living indexes $P^h(p_h^0,p_h^1,u_h,e_h)$ defined by (a) above.

2. Definition (23) defines the class of *arithmetic conditional democratic cost of living indexes*:

$$\text{(a) } P_D^*(p^0,p^1,u,e_1,e_2,\ldots,e_H) \equiv \sum_{h=1}^H [1/H]C^h(u_h,e_h,p_h^1)/C^h(u_h,e_h,p_h^0) = \sum_{h=1}^H [1/H]\, P^h(p_h^0,p_h^1,u_h,e_h)$$

where $P^h(p_h^0,p_h^1,u_h,e_h)$ is the *conditional cost of living for household h* defined in problem 1 above. However, instead of taking an arithmetic average of the individual household cost of living indexes, we could take an *equally weighted harmonic mean*. Thus we define the class of *harmonic conditional democratic cost of living indexes* as follows:

$$\text{(b) } P_{DH}^*(p^0,p^1,u,e_1,e_2,\ldots,e_H) \equiv \{\sum_{h=1}^H [1/H]\, [P^h(p_h^0,p_h^1,u_h,e_h)]^{-1}\}^{-1}.$$

Work out an observable bound for the Paasche type harmonic conditional democratic cost of living index $P_{DH}^*(p^0,p^1,u^1,e_1^1,e_2^1,\ldots,e_H^1)$, which uses period 1 data for $u,e_1,e_2,\ldots,e_H$. Express this bound in expenditure share form and compare your answer to formula (28).

References

Caves, D.W., L.R. Christensen and W.E. Diewert (1982), "The Economic Theory of Index Numbers and the Measurement of Input, Output and Productivity", *Econometrica* 50, 1393-14-14.

Debreu, G. (1959), *Theory of Value*, New York: John Wiley and Sons.

Diewert, W.E. (1983), "The Theory of the Cost of Living Index and the Measurement of Welfare Change", pp. 163-233 in *Price Level Measurement*, W.E. Diewert and C. Montmarquette (eds.), Ottawa: Statistics Canada, reprinted as pp. 79-147 in *Price Level Measurement*, W.E. Diewert (ed.), Amsterdam: North-Holland, 1990.

Diewert, W.E. (1993), "Duality Approaches to Microeconomic Theory", pp. 105-175 in *Essays in Index Number Theory*, Volume 1, W.E. Diewert and A.O. Nakamura (eds.), Amsterdam: North-Holland.

Diewert, W.E. (2000), "Notes on Producing an Annual Superlative Index Using Monthly Price Data", Discussion Paper 00-08, Department of Economics, University of British Columbia, Vancouver, Canada, V6T 1Z1, 30 pp.

Diewert, W.E. (2001), "The Consumer Price Index and Index Number Purpose", *The Journal of Economic and Social Measurement* 27, 167-248.

Hill, T.P. (1999), "COL Indexes and Inflation Indexes", paper tabled at the 5th Meeting of the Ottawa Group on Price Indices, Reykjavik, Iceland, August 25-27, 1999.

Konüs, A.A. (1924), "The Problem of the True Index of the Cost of Living", translated in *Econometrica* 7, (1939), 10-29.

Pollak, R.A. (1980), "Group Cost-of-Living Indexes", *American Economic Review*70, 273-278.

Pollak, R.A. (1981), "The Social Cost-of-Living Index", *Journal of Public Economics* 15, 311-336.

Pollak, R.A. (1989), "The Treatment of the Environment in the Cost-of –Living Index", pp. 181-185 in R.A. Pollak, *The Theory of the Cost-of-Living Index*, Oxford: Oxford University Press.

Prais, S.J. (1959), "Whose Cost of Living?", *The Review of Economic Studies* 26, 126-134.

Triplett, J.E. (2000), "Should the Cost-of-Living Index Provide the Conceptual Framework for a Consumer Price Index?", in *Proceedings of the Ottawa Group Fifth Meeting,* Reykjavik, Iceland, August 25-27, 1999, Rósmundur Gudnason and Thóra Gylfadóttir (eds.), Reykjavik: Statistics Iceland. Available on the web at: http://www.statcan.ca/secure/english/ottawagroup/weblist.htm