A NEW SYSTEM OF LOG-CHANGE INDEX NUMBERS FOR MULTILATERAL COMPARISONS

D.S. Prasada Rao

No. 9 - October 1980

WORKING PAPERS IN ECONOMETRICS AND APPLIED STATISTICS

Department of Economic Statistics
A NEW SYSTEM OF LOG-CHANGE INDEX NUMBERS FOR MULTILATERAL COMPARISONS

D. S. Prasada Rao
University of New England, Armidale, Australia

Several log-change index numbers are available in the literature for price and quantity comparisons, some of the important contributions are due to Kloek-Theil (1965), Theil (1973), Sato (1974) and Theil (1974). All the log-change index numbers suggested so far are basically for binary comparisons though multilateral comparisons can be derived indirectly using the chain-base method. Much of the work by Theil centers around construction of an index number which satisfies the factor reversal test to a desired degree of accuracy. Sato (1974) shows that there are infinitely many log-change index numbers which are equally efficient from this point of view.

In this paper a new system of log-change index numbers for multilateral comparisons is proposed, using an altogether different conceptual framework. This makes use of the concepts 'exchange rate' and 'average price', which were introduced by Geary (1958). Geary (1958) proposes one possible algebraic formalization of these concepts in terms of price and quantity data, and various properties of that system were derived by Khamis (1970) and Prasada Rao (1971). This system has been used for international comparisons by Kravis et al [see Kravis, Kenessey, Heston and Summers (1975) and Kravis, Heston and Summers (1978)] as part of a joint project of the United Nations and the World Bank.

The present paper outlines a system of log-change index numbers derived through another formalization of these concepts, exchange rate and average price, different from that of Geary (1958). In Section 1 the
notation and definitions are developed and the alternative system of log-change index numbers is defined. In Section 2 the viability of this system is established by proving the existence of meaningful comparisons under the new system. The last section contains a numerical illustration derived using data from India.

1. Definitions and Notations

We consider the problem of construction of price and quantity index numbers for M population groups involving N commodities. These population groups may represent different countries or sections of populations within a country defined according to certain social and economic characteristics, or alternatively, the same group of people at different points in time. Let $p_{ij}$, $q_{ij}$, $i = 1, 2, ..., N$ and $j = 1, 2, ..., M$, respectively denote price and quantity data for the $i$-th commodity in the $j$-th population group. Let $R_j, j = 1, 2, ..., M$ represent the exchange rate of the currency of the $j$-th population group, in terms of a common currency unit and $P_i, i = 1, 2, ..., N$ represent the average price of the $i$-th commodity expressed in the common currency unit, the average being taken over the M groups.

Since the exchange rate of the $j$-th currency is inversely related to the general price level in the $j$-th group, the price index number for group $k$ with the $j$-th group as a base, denoted by $I_{jk}$, is defined as

$$I_{jk} = \frac{R_j}{R_k}$$

for all $j$ and $k$, and a system of log-change index numbers can be defined by  

$$I^*_{jk} = \log R_j - \log R_k.$$

† All logarithms used in this paper are natural logarithms.
3.

Obviously the $R_j$'s and $P_i$'s are determined by the price and quantity data. A particular index number system can be derived by specifying functional relationships between the $R_j$'s, the $P_i$'s and data on prices and quantities. The new system we propose in this paper is defined by the following $(M+N)$ equations. These are, for $j = 1, 2, \ldots, M$ and $i = 1, 2, \ldots, N$.

$$R_j = \left\{ \prod_{i=1}^{N} \left( \frac{P_i}{P_{ij}} \right) v_{ij} \right\}$$

...(1)

$$P_i = \left\{ \prod_{j=1}^{M} \left( \frac{R_j}{P_{ij}} \right)^* v_{ij} \right\}$$

where $v_{ij}^* = \frac{v_{ij}}{\sum_{k=1}^{M} v_{ik}}$ and $v_{ij} = \frac{P_{ij}q_{ij}}{\sum_{i=1}^{N} P_{ij}q_{ij}}$.

Here $v_{ij}$ represents the share of the $i$-th commodity in the $j$-th group's total expenditure and $v_{ij}^*$ is a measure of importance attached to the $i$-th commodity over different population groups.

We briefly present below the system proposed by Geary (1958), henceforth referred to as the Geary-Khamis system due to the important contributions of Khamis which made this system popular, and compare their system with the one proposed above. The Geary-Khamis system is defined for $j = 1, 2, \ldots, M$ and $i = 1, 2, \ldots, N$, by

$$R_j = \left\{ \prod_{i=1}^{N} \frac{P_i}{P_{ij}} v_{ij} \right\}$$

...(2)

$$P_i = \left\{ \prod_{j=1}^{M} \left( \frac{R_j}{P_{ij}} \right)^* q_{ij} \right\}$$
where \( q_{ij}^* = \frac{q_{ij}}{\sum_{k=1}^{M} q_{ik}} \). Here \( q_{ij}^* \) reflects the importance of the \( i \)-th commodity across the \( M \) groups.

These two systems are well defined only if

(i) for all \( i \) and \( j \), \( p_{ij} > 0 \) and

(ii) for all \( i \) and \( j \), \( q_{ij} > 0 \) and for each \( i \), \( q_{ij} > 0 \)

for at least one \( j \).

Comparison of these two systems is straightforward. The new system as defined by equation system (1) is designed to be less sensitive to the presence of population groups whose consumption patterns are extreme. This is achieved by making use of weighted geometric means instead of weighted arithmetic means as in (2). Further, we make use of expenditure shares to define all of the weights used in (1) whereas the Geary-Khamis system makes use of expenditure shares only as weights to define the \( R_j \)'s and uses quantity shares to define the \( P_i \)'s. Generally expenditure share weights \( v_{ij}^* \) are more stable than quantity share weights \( q_{ij}^* \) and hence system (1) can be expected to perform better than the Geary-Khamis system (2). It has been pointed out before in Prasada Rao (1972) that some extreme quantity vectors can distort price comparisons to a large extent. This kind of distortion can be expected to be reduced through the use of log-linear equations and expenditure share weights.

2. Viability of the New System

The new system defined by (1) is a system of \((M+N)\) log-linear equations in as many unknowns. This system would be viable only if the existence of positive solutions for the \( R_j \)'s and \( P_i \)'s are guaranteed. Further this set of solutions must be unique if the system is to be
meaningful. Since price indices are defined as ratios of $R_j$'s, solutions for the $R_j$'s should be unique up to a factor of proportionality. In what follows, we prove that the new system is viable in the sense described above provided quantity vectors for different population groups satisfy some very mild conditions.

Taking logarithms on both sides of the $(M+N)$ equations in (1), we have for $k = 1, 2, \ldots, M$ and $i = 1, 2, \ldots, N$

$$R_k^* = \sum_{i=1}^{N} \left[ P_i^* - \log p_{ik} \right] v_{ik}$$

and

$$P_i^* = \sum_{j=1}^{M} \left[ R_j^* + \log p_{ij} \right] v_{ij}$$

where $R_k^* = \log R_k$ and $P_i^* = \log P_i$.

This is a system of $M+N$ linear non-homogeneous equations in $(M+N)$ unknowns. Substituting for $P_i^*$ in $R_k^*$, we get, for $k = 1, 2, \ldots, M$

$$R_k^* = \sum_{j=1}^{M} \sum_{i=1}^{N} R_j^* v_{ij} v_{ik} + \sum_{j=1}^{M} \sum_{i=1}^{N} \log p_{ij} v_{ij} v_{ik} - \sum_{i=1}^{N} \log p_{ik} v_{ik}$$

or

$$R_k^* = \sum_{j=1}^{M} \sum_{i=1}^{N} R_j^* v_{ij} v_{ik} + \sum_{j=1}^{M} \left[ \log p_{ij} - \log p_{ik} \right] v_{ij} v_{ik}^*$$

since $\sum_{j=1}^{M} \sum_{i=1}^{N} \log p_{ik} v_{ik} v_{ij}^* = \sum_{i=1}^{N} \log p_{ik} v_{ik} \sum_{j=1}^{M} v_{ij}^* = \sum_{i=1}^{N} \log p_{ik} v_{ik}^*$

This is a system of $M$ linear equations in $M$ unknowns $R_k^*$. We can rewrite this as a system of linear equations of the form

$$AR^* = b \tag{3}$$
where $A$ is a matrix of order $M \times M$, with typical element $a_{jk}$ of the form $a_{jk} = \delta_{jk} - \Sigma v_{ik} v_{ij}$.

\[ \delta_{jk} = 1 \text{ if } j = k \text{ and } 0 \text{ if } j \neq k; \quad R^* = [R_1^* \ R_2^* \ldots \ R_M^*]' \quad \text{and} \]

\[ b = [b_1 \ b_2 \ldots b_M]' \quad \text{where} \quad b_k = \sum_{j=1}^{M} \sum_{i=1}^{N} (\log p_{ij} - \log p_{ik}) v_{ik} v_{ij}^* . \]

A few properties of equation system (3), the proofs of which are easy to see, are stated below.

**Property 1:** The elements of each row in matrix $A$ add up to zero. Since $A$ is symmetric, this implies that the elements of each column of $A$ add up to zero.

**Property 2:** In view of property 1, $A$ is singular and hence $R(A) \leq M-1$, where $R(A)$ denotes the rank of the matrix $A$.

**Property 3:** The elements of the column vector $b$ add up to zero.

**Result 1:** If $R(A) = M-1$, then equation system (3) is consistent in that there exists a solution for $R^*$ such that $AR^* = b$.

**Proof:** Let $A^*$ represent the augmented matrix by adding column vector $b$ to the columns of matrix $A$. We have

\[ A^* = [A:b] \]

To establish consistency of (3), it is enough if we show that $R[A] = R[A^*]$. Since $A^*$ is a matrix of order $M \times (M+1)$ and from properties 2 and 3, we have that all columns add up to zero in $A^*$, we have $R(A^*) \leq M-1$. If $R(A) = M-1$, we have $R(A^*) = M-1 = R(A)$, since $R(A) \leq R(A^*)$, and this proves the result.
We now derive some necessary and sufficient conditions under which \( R(A) = M-I \). If \( R(A) = M-I \), then there is no unique solution for \( R^* \), but there exists a solution which is unique up to a factor of proportionality. A few definitions are required before we state the main result.

**Definition:** Let \( G \) be a graph with \( M \) groups as vertices and join vertices \( j \) and \( k \) by an 'edge' if there exists at least one commodity \( i \) such that \( q_{ij} \) and \( q_{ik} \) are both positive.

This graph is known as the adjacent graph of the data.

**Definition:** A graph \( G \), as defined above, is said to be connected if we can pass from any vertex to any other vertex through a series of edges.

The following result states the necessary and sufficient condition under which \( R(A) = M-I \) and a meaningful solution exists for the system defined by (3) and hence (1).

**Main Result:** The log-linear equation system (1) has solution for \( R \) which is positive and unique up to a factor of scalar multiplication through system (3) if and only if the quantity data is such that the 'adjacent graph' \( G \) of the data is connected.

**Proof:** Proof of this result can be constructed by retracing the steps in the proof of a similar result in Prasada Rao (1971) and hence it is omitted.

The conditions for existence stated in the main result above are very mild and are generally satisfied by most price-quantity data sets. If the condition of connectedness of the 'adjacent graph' of the data is not satisfied then we can divide \( M \) groups into two non-empty subsets such that no commodity consumed in one subset of population groups is consumed by any population group in the other subset. Then we have a situation where the two subsets have completely different consumption patterns.
Obviously in such a situation the usual problem of price and quantity index number construction is not meaningful. Such situations call for a different methodology.

For purposes of comparison with log-change index numbers due to Kloek-Theil (1965) and Theil (1973), we derive price index numbers from our system in an explicit form for the special case \( M=2 \). Solving equation system (3) for \( R_1^* \) and \( R_2^* \) and hence deriving \( R_1 \) and \( R_2 \), we have

\[
I_{12} = \frac{R_1}{R_2}
\]

or

\[
= N \left[ \frac{P_{i2}}{P_{i1}} \right] \frac{v_{i2}^* v_{i1}}{v_{i1}^* v_{i2}}
\]

where \( v_{i2}^* = \frac{v_{i2}}{v_{i1} + v_{i2}} \). This means that the new system leads to an index number which is a weighted geometric mean of the price relatives \( \frac{P_{i2}}{P_{i1}} \), \( i = 1, 2, \ldots, N \) where the weights are based on harmonic means of \( v_{i1} \) and \( v_{i2} \) for each \( i \). The Kloek-Theil (1965) index is

\[
I_{12} = \frac{N}{i=1} \left[ \frac{P_{i2}}{P_{i1}} \right] \frac{v_{i1}^* v_{i2}^*}{v_{i1}^* v_{i2}^*} \frac{v_{i2}^* + v_{i1}^*}{v_{i2}^* + v_{i1}^*}
\]

This is a weighted geometric mean of price relatives which is based on weights derived from the arithmetic means of \( v_{i1} \) and \( v_{i2} \) for all \( i \). Theil (1973) suggests an index number of the form

\[
I_{12} = \frac{N}{i=1} \left[ \frac{P_{i2}}{P_{i1}} \right] w_i
\]
where the weights \(w_i\) are defined as

\[
\frac{\left(\frac{v_{i1} + v_{i2}}{2}\right)^{1/3}}{\sum_i \left[\frac{v_{i1} + v_{i2}}{2}\right]^{1/3}}
\]

Theil proposed this index since it comes very close to satisfying the factor reversal test and in fact it satisfies the factor reversal test to the fifth order of smallness. From this point of view neither our index nor the Kloek-Theil index are accurate to that degree. In fact, using Sato (1974)'s main result, our index can be shown to satisfy the factor reversal test to the third order of smallness. But our new index is meant to yield consistent multilateral comparisons and is not designed specifically from the point of view of the factor reversal test. It just shows that, when binary comparisons are involved the new system presented here simplifies to a simple weighted geometric mean, where the weights are slightly different from those used by either the Kloek-Theil index or the Theil index.

3. A Numerical Illustration

We illustrate here that the new system defined by equation system (1) is indeed viable, using some Indian Data. We consider the decile groups of the rural sector of India as our population groups. We take the National Sample Survey 18th round data on consumption expenditure in these groups for illustration. We take 56 items of consumption for which price and quantity information is available. The following table presents price index numbers based on the new system proposed here, and the index numbers from the Geary-Khamis system along with the more conventional indices defined by Laspeyres and Paasche.
### Price Index Numbers based on different Methods
(Base Group 0-10)

<table>
<thead>
<tr>
<th>Decile Group</th>
<th>Laspeyres' Index</th>
<th>Paasche Index</th>
<th>Geary-Khamis Index</th>
<th>New Index Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-10</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>10-20</td>
<td>103.61</td>
<td>103.44</td>
<td>103.66</td>
<td>103.62</td>
</tr>
<tr>
<td>20-30</td>
<td>106.27</td>
<td>105.88</td>
<td>106.31</td>
<td>106.06</td>
</tr>
<tr>
<td>30-40</td>
<td>106.59</td>
<td>106.15</td>
<td>106.38</td>
<td>106.25</td>
</tr>
<tr>
<td>40-50</td>
<td>108.33</td>
<td>107.77</td>
<td>107.85</td>
<td>107.82</td>
</tr>
<tr>
<td>50-60</td>
<td>110.25</td>
<td>107.31</td>
<td>109.62</td>
<td>108.59</td>
</tr>
<tr>
<td>60-70</td>
<td>109.69</td>
<td>109.62</td>
<td>109.57</td>
<td>109.58</td>
</tr>
<tr>
<td>70-80</td>
<td>111.41</td>
<td>110.80</td>
<td>110.78</td>
<td>110.78</td>
</tr>
<tr>
<td>80-90</td>
<td>110.90</td>
<td>111.55</td>
<td>111.37</td>
<td>111.47</td>
</tr>
<tr>
<td>90-100</td>
<td>113.43</td>
<td>115.62</td>
<td>115.46</td>
<td>115.73</td>
</tr>
</tbody>
</table>

Though the Geary-Khamis indices and index numbers from the new system look very similar, the main effect of using geometric means and expenditure share weights can be seen after close scrutiny. Our system generally produces a lower magnitude for the price index for lower decile groups (0-10 group to 50-60 group) than does the Geary-Khamis index and higher values for the price index for upper decile groups (60-70 to 90-100 groups). This is the type of effect which this system is designed to bring about. One could expect more noticeable differences if price and quantity data were more heterogeneous than the data used here. Such situations are typically confronted when international comparisons are involved.
References


