HARMONIZATION OF DATA: LESSONS FROM MAJOR INTERNATIONAL ORGANIZATIONS

Chandra Sitahal Aleong
and
John Aleong

Introduction

This paper discusses the importance of research into (a) the design, maintenance and implementation of a metadata system that would integrate national and regional data and (b) the development of statistical measures like the Consumer Price Index (CPI), that are meaningful and relevant to current CARICOM economies. The end product is a statistical metadata repository that is robust enough to accommodate changes over time.

The paper is structured in the following way. First, there will be an overview of the strategic plans of two organizations that have had to design, maintain and enhance very sophisticated data systems. Developments in managing large censuses; issues of privacy and confidentiality; standardization of statistical terminology and methodology; and the need for common legislation that would enable all of this to be accomplished are some of the issues discussed. Finally there will be an examination of the existing databases of CARICOM countries and a discussion of the computation, the problems, and the need for further research on the consumer price index and its adaptation to CARICOM.

Strategic Goals

Having a metadata system requires a multi-pronged approach, but one of the first steps is the need to delineate the strategic advantages. While this might seem to be obvious, unless there is complete buy in by all the parties there can be no solution to the problems of integration and harmonization. Challenging questions like, where does the region want to be five, ten, fifteen years from now and what type of database would meet these needs, should be at the centerpiece in the planning taking place at the regional as well as at the national level. The commitment must originate from the head of state of each of the countries and should provide the underpinnings for all the decisions that would make the data infrastructure materialize. In this environment, putting together a meta-database would be time consuming but very feasible.

Political and national associations like Census Bureau and Statistics Canada have been successful in collecting and disseminating massive amounts of data and are the central repositories for state and provincial data. The lack of a central force in the Caribbean equivalent to a federation of states or provinces might provide obstacles in the writing and enforcement of legislation to make a data repository materialize. However a good example for CARICOM is the North-
American Free Trade Association (NAFTA). It demonstrates that despite very vocal nationalist roadblocks, the motivation to find commonalities for growth outweighed the differences. Although not touted as the raison d'être for NAFTA, reliable data is surely one of the underlying ingredients in determining the feasibility of trade and alliances.

**Strategic Goals: U.S. Census Bureau and Statistics Canada**

The strategic goals and objectives of the U.S. Census Bureau and Statistics Canada (SC) will be discussed together because they are similar. Some of the more important ones are as follows: to meet the needs of policymakers, businesses, nonprofit organizations and, most importantly, the public. Reliable data is necessary to provide the U.S. and Canadian populations with current measures of its status as well as with credible measures of the economy. Data from these two organizations support the economic and political foundations of the country by producing benchmark measures of how the economy and population are faring. In the U.S. the Chairman of the Federal Reserve relies on a multiplicity of economic indicators to determine whether inflation is increasing. This critical data is used as a determinant for changing interest rates, which has a spiral effect not only on the U.S. economy, but also on the global economy. The ramifications as a result of this change are immense so the data has to be credible. Political battles are fought on the basis of this data, hence the scrutiny and involvement of the average citizen, lobby groups, investors, and others both nationally and internationally. The same involvement does not seem to exist among the citizens of the CARICOM community. There is a gap between policy makers and the average citizen in the knowledge and concern with data.

In this environment currencies and interest rates are changed without the involvement that an educated population deserves. This has to change if all citizens are to participate and benefit from development.

Another goal of the databases of Statistics Canada and Census Bureau is to meet constitutional and legislative mandates such as funding for social security and other retirement benefits, education and healthcare. This requires monitoring of population and economic trends for planning and extrapolation purposes. Budgeting for the short term requires payment of significant amounts of money and these estimates have to be integrated with policies of the Treasury Department and other bank and financial institutions. Projections for the long-term might mean an overhaul in the nature of financing Social Security, Medicare and other retirement benefits.

Education funding is based on population demographics such as the size of the K-12 group for financing primary and secondary schools and the 18 to 24 year old cohort for higher education support. Healthcare is another very controversial topic for most U.S. and Canadian citizens because of cutbacks by companies and governments in the financing of medical insurance. This has triggered heated discussions between unions and management as well as federal and state legislatures as to who picks up the tab. Many of these negotiations are pegged to indices like the unemployment and inflation rates and the consumer price index. Data that would result in decisions that are fair and consistent is the fuel on which developed economies run. Without it there will be a perpetuation of a decision making process from which only an elite few benefit.
This highlights the need for objective statistical information for the efficient operations of an open and democratic society. Reliable data on economic performance, development of fiscal, monetary, foreign exchange, international trade and tariff policies are major ingredients in the functioning of progressive societies. Other more immediate objectives are the development of policies and programs to assist small businesses and the support of the development and the evaluation of government programs on economic and social welfare programs. Other reasons are the promotion of domestic and international competitiveness, the support of tourism strategies and programs, the monitoring of the justice system’s effectiveness and efficiency, the selection of sites for schools and public transportation, and the development of programs such as day care and subsidized housing.

Data generation and analysis are of importance not only for national planning reasons but also for regional and international purposes. In the case of the North American Free Trade Area (NAFTA) or the European Common Market, sound statistics are the basis for improving the allocation of government program funding by determining their social and economic effects region by region. One example pertinent to common market discussions is the fairness of payments and other fiscal transfers. Another is the issuance of inflation-indexed contracts and entitlements. The planning and consolidation of data by the departments of transportation and environment across Canada and the U.S. in association with NAFTA, are excellent examples of the synchronization that is possible when data is integrated and harmonized.

Another type of projection that is done with harmonized data relates to population projections. The U.S. Census Bureau estimates total resident population for each state and county with a ‘component of population change’ method. The assumption here is that the components of population change can be closely approximated in a demographic change model. Each component of the population is represented with data that is symptomatic of some aspect of population change. For example, birth certificates represent additions, death certificates and other medical records reductions. Population data forms the basis for planning at every level, and the modeling and analysis of data is done on a real-time basis with the appropriate technology. Participation by all users of the process, a common protocol, and clear understanding of terms and concepts become vital to the generation and final interpretation of the data.

Clients and stakeholders play a major role in ensuring that programs meet emerging information needs. Statistics on major sectors of the economy such as agriculture, the environment, manufacturing, and the mineral producing segments as well as data on social conditions, travel and tourism, provide input for short and long term budgeting. Every five or ten years a census provides an updated national picture of the region’s work life, family structure, ethnicity, housing and education system. From the location of schools, to the invitation of new business ventures, to the investment patterns of the financial sector there is need for continuous updates based on emerging needs. The economies of the U.S., Canada, and the nations that comprise the European Union, as well as Japan and Singapore are examples of the integration of all sectors of the economy and the successful use of data to encourage growth.

Continuous Improvement and Technology

Timeliness and the continuous improvement in the quality of data are concurrent with meeting the goals defined above. Thus, as soon as one
census is completed by the United States Census Bureau and Statistics Canada changes and improvements are made for the next one. The deliberations are exhaustive and the resources deployed are budgeted for by the federal and state legislatures. Innovation, promoting data usefulness, minimizing the burden on respondents, respecting privacy, and ensuring confidentiality are constant concerns of every citizen from Congress on down. The last decennial census in the U.S. involved major debates by the U.S. Congress about the rationale for using samples to estimate inner city populations and door-to-door headcount. The main point here is to involve all parties and to continuously improve the process.

Intrinsic to all of this is state-of-the-art computing technology and expertise in survey methodology and statistical methods. While some of the information comes from existing administrative data, most is collected through businesses and household surveys. Survey respondents are coached on their role in the overall process and are guaranteed confidentiality. Finding innovative ways to reduce the time spent conducting and completing surveys is continuous and is facilitated by the declining cost of computing technology. In addition, hardware and software costs enable data collection and integration to be decentralized.

Countries that are currently in the process of developing databases have an enormous advantage due to the Internet. The collection of data and analysis and dissemination of statistical reports to large numbers of users can be accomplished in a short period relative to the process even five or ten years ago. The databases are user-friendly and users are educated on how to access and interpret the data. Statistics Canada has a long-standing program of education outreach so that statistical concepts are introduced to school age children and social and economic data has become part of the Canadian school curriculum.

Education

One of the ‘best practices’ used in fostering harmonization and integration is education. Instructing citizens of the cost implications of redundant and inaccurate information, and training survey administrators in the concepts, definitions and classification system of the information system are standard practices for Statistics Canada and the Census Bureau. Differences in terminology, such as years of post-secondary education, could lead to the collection of completely different data in a CARICOM country versus a North American country. Conceptual differences also lead to misunderstanding. Errors in definitions such as size of family, as in the case of the extended family, pose differences in interpretation and require documentation at the input, output, as well as the interpretation level. In cases where there is great scope for ambiguities, best practices are identified and documented.

Legislation

One of the goals of CARICOM is to produce statistics that would help the Caribbean nations better understand the region. A system that is robust enough to handle data on every economy in the region entails collecting information on the population, resources, society and culture of all peoples. This responsibility belongs to all the countries of the region with CARICOM being the central agency for collection and distribution of all statistical data. This cannot happen without legislation. Statistics Canada and the Census Bureau have specific legal acts that were passed by federal and state
legislatures to collect and consolidate data. They perform this function for the whole of Canada and the United States as well as for each of the provinces and states. They collaborate to avoid the duplication of information collected by departments of government. Most importantly they promote and develop integrated social and economic statistics pertaining to the whole country as well as to each region.

The US Census Bureau is authorized by Title 13 of the United States Codes to collect data using censuses and surveys. Section 9 of the same code authorizes the Census Bureau to keep the data collected under Title 13 confidential. Section 214 of Title 13 and Sections 3559 and 3571 of Title 18 of the United States Code provide for the imposition of penalties of up to five years in prison and up to $250,000 in fines for wrongful disclosure of confidential census information. Data collected by Statistics Canada through their surveys, census, or any other methods are confidential and cannot be released with any identification of the individuals or groups unless there is prior consent. The statistics and the privacy acts are well documented and are available at http://www.statcan.ca/english/about/statact.htm.

The CARICOM website lists only a few countries where there is documentation of the legal framework to collect data by censuses and surveys pursuant to the statistics acts of the individual member countries. The countries which are documented by the Caribbean Community Secretariat through the CARICOM Statistics website, http://www.caricomstats.org/projects.html, are Anguilla, Antigua and Barbuda, The Bahamas, Barbados, Belize, Guyana, Trinidad and Tobago, and St. Kitts and Nevis. There is need for legislation, which would authorize CARICOM to collect data by census, surveys and other methods for all CARICOM countries. This legislation should ensure the confidentiality of the data collected and give guidelines for the dissemination of the data and the imposition of penalties for the ‘wrongful disclosure of confidential census information’.

Summary of Role of Strategic Planning, Education and Legislation

CARICOM countries have measures of their populations or economies but all the supporting research suggests that there is no common protocol or repository to coordinate and plan for the future shape of a regional database. Having the necessary legislative acts and procedures, as well as the educational framework as part of an overall strategic plan would remove some of the major roadblocks to harmonization and integration.

Henry Mintzberg, one of the great thinkers in the field of strategy states, if the grass roots are not involved in strategic planning, the top echelons can develop the most beautiful plans but they will not be implemented because of lack of commitment. There must be trust and clear understanding of what is needed for the generation of the end product, in this case data. In the case of CARICOM a strategic plan to develop a metadata repository must be incorporated into an umbrella strategic plan for the region. The political, legislative, economic, social, environmental and other entities must be integrated because they all need reliable data for planning and implementation. Once this is accomplished the methodology for computing and analyzing data can be resolved.
Statistical Methodology

Data Access

The best data infrastructure should provide data in a format that the user could access and manipulate to meet a multiplicity of needs. A review of the CARICOM web sites showed data only for a few countries.

Table 1 shows documentation that was available for major categories of statistics. Balance of payments documentation was found only for St. Kitts and Trinidad and Tobago; business surveys only for the Bahamas; and household surveys only for Antigua and Barbuda etc. The issue here is the paucity of readily accessible documentation for individual countries.

<table>
<thead>
<tr>
<th>Type of Sample Survey Statistics</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balance of Payments</td>
<td>St. Kitts and Nevis, Trinidad and Tobago</td>
</tr>
<tr>
<td>Business Surveys</td>
<td>The Bahamas</td>
</tr>
<tr>
<td>Household Surveys</td>
<td>Antigua and Barbuda</td>
</tr>
<tr>
<td>Indices</td>
<td>Antigua and Barbuda, Barbados, St. Kitts and Nevis, Trinidad and Tobago</td>
</tr>
<tr>
<td>Merchandise Trade</td>
<td>Barbados, Dominica, Trinidad and Tobago</td>
</tr>
<tr>
<td>National Accounts</td>
<td>The Bahamas, Barbados, St. Kitts and Nevis, Saint Lucia, Trinidad and Tobago</td>
</tr>
</tbody>
</table>

Source: http://www.caricomstats.org/index.htm

Contextual knowledge of the CARICOM countries suggests that the statistical offices or the central bank in each country have data that has been collected and formatted for internal use as well as for use by lending agencies and international organizations. One of the problems with access to data seems to be the deep-seated belief that data and information should not be in the public domain but belongs to those in positions of power. Open access to the enormous database of the Census Bureau and Statistics Canada requires a completely different mindset; one that acknowledges that the data infrastructure is paid for by the taxpayer and should be accessible to as many users as possible. Potential users of data for CARICOM countries must find the middleman to guide him/her through the maze before finding the relevant information for analysis.
Methodology

Statistical methodology should reflect the latest research in the field with use of a recognizable protocol, consistent criteria, and relevance for the circumstance. One of the first steps in setting up a harmonized database is the standardization of methodologies. A review of the databases on the CARICOM websites showed considerable variability in the statistical methodology among CARICOM countries with respect to collecting the same data. For example consider the computation of the consumer price index (CPI).

Table 2 shows that the CPI was computed monthly for Trinidad and Tobago but quarterly for Tobago. It also shows for Trinidad and Tobago and Antigua and Barbuda that the sample design and the calculation of the CPI are different. The frequency of data collection and the type of data collected are also different; so are the base years and the timeliness. Even the baskets of goods and services and the weights used are different. A close reading of the methodology shows details on how to calculate the aggregation formulas for Trinidad and Tobago while there are few details for Antigua and Barbuda.

While Table 2 shows the obvious lack of consistency in all the criteria, there are inherent problems in the way the CPI is being computed. This compounds the interpretation of this widely used measure of the economy. A discussion of issues related to the CPI is relevant in a discussion of harmonization and integration of a meta-database because once the strategy, education and legislation are in place continuous improvement requires research and development of current methods of computing indices and measures.

The CPI: Issues and Methods of Computation

The CPI measures changes in prices and is an index of the rate of change in the economy over time. The problem with its computation is the speed with which prices change in a dynamic economy. The rapid proliferation of products as well as the constant improvement in the quality of products also makes computation more difficult. Its widespread use as a basis of comparison of the effect of monetary and fiscal policies and economic progress over time makes its accuracy very important. As stated earlier, spending levels for union contracts and Social Security payments are indexed to the CPI. The political ramifications will be immense in the U.S. if there is a move to reduce the index due to substitution and quality improvements.

A description of several methods of computing price indices such as the Laspeyres, Paasche, Fischer and others are an attempt at demonstrating that there is no ‘one size fits all’ and that there is need for CARICOM countries to discuss the design of a meta-database as well the mechanics of how statistical measures are computed.

A true cost of living index (COLI) should be calculated at the household level and then at the social level. Konus (1939) states that the COLI for a single household is the ratio of the minimum cost of achieving a certain reference utility in a base period given prices at that time with a later current period, given changes in price during the interval. In effect it compares a base period with the current period. Laspeyres index $(P_L)$, which computes the cost of purchasing a fixed basket of goods in the base period and then the cost of the same basket in the present overestimates the COLI, since it does not
<table>
<thead>
<tr>
<th>Item</th>
<th>Trinidad &amp; Tobago</th>
<th>Tobago</th>
<th>Antigua and Barbuda</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>Monthly</td>
<td>Quarterly</td>
<td>Monthly, quarterly and annually</td>
</tr>
<tr>
<td>Timeliness of Data</td>
<td>3 weeks after reference month</td>
<td>Last week after the reference month</td>
<td>Different</td>
</tr>
<tr>
<td>Basket of goods and services</td>
<td>National index for income TT$400-TT$4600</td>
<td>Separate index</td>
<td>Barbuda excluded</td>
</tr>
<tr>
<td>Source</td>
<td>Household budgetary survey</td>
<td>Household income survey 1998-1999</td>
<td>N=660 households</td>
</tr>
<tr>
<td>Sample size</td>
<td>N=2433 households/1712 households used</td>
<td>Not clear. Six months data collected over 2 weeks</td>
<td>State 1: primary sampling units (pps) proportional to size. Stage 2: clusters of households from pps in Stage 1</td>
</tr>
<tr>
<td>Base year</td>
<td>September 1993</td>
<td>October 2001</td>
<td></td>
</tr>
<tr>
<td>Basket of goods and services</td>
<td>N.A.</td>
<td>Data collected monthly for two weeks beginning on Tuesday</td>
<td>Different</td>
</tr>
<tr>
<td>Data collection System</td>
<td>N.A.</td>
<td>Food -214.2</td>
<td></td>
</tr>
<tr>
<td>Weights</td>
<td>Food -217</td>
<td>Area weights Item weights Section weights</td>
<td>Laspeyres formula</td>
</tr>
<tr>
<td>Aggregation</td>
<td>N.A.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: http://www.caricomstats.org/index.htm

N.A. - Not Available.
account for the substitution of goods. Substitution occurs at three levels: the elementary, commodity and the outlet level. When the price of a commodity can be purchased at different outlets for heterogeneous prices, what price should be used in the computation of the index? The Paasche index, which estimates the cost of a fixed basket of goods now and also in the past, underestimates the COLI. Diewert (1998) justified and suggested the use of the Fisher ideal price index \( P_F \), which is the geometric average of the Paasche and the Laspeyres indexes. The Tornqvist index, \( TI \) suggested by the Boskin (1997) Commission, as another good index, is a weighted geometric mean of price relatives at the ‘stratum’ and ‘entry level item level.’ It is concurrently tied to geometric means of price relatives at the elementary aggregation level.

In summary, the Boskin (1997) CPI Commission states that the CPI overestimates the US Cost of Living Index (COLI) by about 1.1% points per year (i.e. a range of 0.8 to 1.6 points per year). The 1.1% bias can be broken down as follows: upper level substitution .15%; lower level substitution .25%; quality change and new product bias .6%; new outlets .1% giving a total of 1.1% upward bias in the CPI. This is a very significant amount when compounded over time. For example when the CPI is at 3%, the COLI is really about 2%.

Mathematically, elementary substitution bias \( B_L \) is defined as the difference between the Laspeyres price index \( P_L \) and the Fisher price index \( P_F \). Let \( e \) equal the inflation adjusted percentage change in price. Appendix 1 shows that

\[
B_L = P_L - P_F = .5P_L * e
\]

Variance \( e \) where \( P_L = (1+i) \).

Suppose the variance of the inflation adjusted percentage change in price is .005, the rate of inflation, \( i \) is equal to 2%, then the Laspeyres index is 1.02, and the bias is .00255 or .255 % points. For commodity substitution bias, the bias is among different commodity prices. If variance of \( e = .005 \), \( i = 2\% \), then the commodity substitution bias is an additional .255% points. Hence the commodity substitution bias is .5% points. For outlet substitution bias \( B_o \) suppose a low cost retailer moves into an area and let \( s \) equal the market share taken by this retailer in the current period, while \( d \) equals the percent discount of the low cost retailer over the traditional retailer, then the outlet substitution bias

\[
B_o = .5*(1+i)sd.
\]

If inflation \( i = 2\% \) then \( P.L = 1.2 \). If \( s = .05 \) is the proportion of new goods which are not in the basket, and \( d = 20\% \) is the decline in prices of new goods from the initial imputed prices, then the new good bias is \( .5X1.02X.05X.2 = .051 \) or .51 percentage points.

**Implications**

Bias in computing methodologies has serious implications for the budget, debt, poverty rates and other economic indicators. The Boskin report recommended improving the index by using hedonic regression models for adjustments of quality. Appendix 2 gives a brief discussion of hedonic regression models. Appendix 3 gives a summary of some additional methodological problems with the consumer price index.

The Bureau of Labor Statistics did the following to address biases in the CPI; use of the geometric mean for low-level estimation; updating weights frequently; hedonic adjust-
ments for the CPI; and use of superlative and consistent indices. The European Union has a Harmonized Index of Consumer Prices (HICP), which is mandatory for member countries and allows for international comparisons. Meanwhile in Trinidad, the Central Bank is addressing some of the issues in the computation of the CPI. The discussion indicates progress because of the importance of finding indices that are relevant to each country and to the region as a whole.

**Future Research**

There is need for more collaboration and interaction among organizations, such as the statistical offices of each country, the central banks, the international agencies and others to consolidate and harmonize data. Some of this work has begun at the Central Bank in Trinidad and Tobago and is most likely being done in the other countries. Agreement of a protocol; standardization of methodology; publication of meaningful price indices; research into alliances with other data collection bodies in the U.S. Canada and Europe are some of the areas for further work. In addition, a region-wide strategic plan and common legislation will set the overall framework of what needs to be accomplished and will give the design and implementation of a meta-database the rigor and exposure needed for its success.

Electronic data collection is another area for further work. Price quotes at retail outlets should be done electronically and be fed into a central database. Diewvert in Triplett (1990) suggested scanner data or detailed computerized data be used to make estimates of substitution biases for new goods as well as for aggregation at various levels. Data collected by survey mail questionnaires should also be done electronically. Statistics Canada is implementing this for all surveys. Most important is the need for research into the multiplicity of uses for the Internet to lower cost; increase avenues for data collection; educate users and participants; and other areas that lend themselves to the use of this medium.

As indicated earlier, there is need for research into a family of indices rather than a single index as measures of the status of the economy. The current CPI and a lagged CPI can be used to reduce the substitution bias and incorporate quality adjustments as recommended by the Boskin (1996) Commission. For seasonal products with large price variability, there is need for more than one index; one for non-seasonal commodities; one annual index that compares March of one year with March of the previous year; and another index which compares a moving total of 12 months with 12 base months. In the Stigler (1961) Commission report, Griliches (1961) suggested using the hedonic regression method for updating prices for quality changes in the automobile industry. The U.S Bureau of Labor Statistics (BLS) adopted hedonic regression in 1980 and it is used in the electronic, housing, television, apparel and computer industries. Hedonic quality adjustment in the CPI needs to be investigated for other products and services.

Another area for further research is the service sector. How do we define and measure an output? There are difficulties in measuring services in the banking, construction, telecommunications and medical industries because it is difficult to measure the rapid changes in price and quality improvement. Another service area that poses measurement problems is teaching, which depends heavily on consumer characteristics, participation, and the delivery system. For example is output for teaching online for mature students the same as in-class primary school teaching?
Price and quantity measures have to be done on an industry-by-industry basis and a product-by-product basis. The BLS is engaged in an international product classification system, which should result in a service sector product classification system that may define a service output more clearly. Since most economies are moving from production of goods to production of services the demand for understanding how to measure output is critical.

These are just a few of the areas that need further research and are under constant review. The collection, coordination, analysis and distribution of data and its appropriate measures and indices, is a never-ending process.

Conclusion

The paper shows that there are several countries and institutions that have built meta data systems that integrate national and regional data. As indicated above it is an ongoing process to collect data that is reliable and to identify methodology that gets to the core of explaining what is really happening in a dynamic economy. The challenges are enormous even when the population is fairly homogeneous, and the economy is centralized. In more complex economies like the U.S., and Canada, it helps that a federal government has the clout to pass legislation demanding that states and provinces follow. The ultimate goal is to get data that is unbiased and consistent that allows decision makers to make fair and informed decisions.

CARICOM countries have to surmount the problems inherent with building a database as well as overcome the colonial and nationalistic tendencies given their history. They respond more readily to the demands of the current providers of financing rather than to the demands of what could foster a successful long-term coalition of states. This vision is myopic because it is denying generations the benefits of building a data infrastructure that would lead to a stronger coalition and increased standards of living.

In conclusion it is an understatement to say that there is much work to be done to integrate and harmonize data for the CARICOM countries. A strategic plan with clearly defined goals, a time line for implementation, adequate resources, assigned responsibilities, and evaluation of performance, are essential. However, there are lessons and ‘best practices’ in the work accomplished by Census Bureau and Statistics Canada. Without a data infrastructure CARICOM might not be able to accomplish the goals of its mission statement ‘to improve the quality of life for people in the region’.
APPENDIX I

Biases of the Estimated Laspeyres Price Index $P_L$

Let $(p_n^t, q_n^t)$ be the price and quantity for outlet $n$ in period $t$ where $t = 0$, or $1$, and $n = 1, \ldots, N$. The Laspeyres Price Index $(P_L)$, the Paasche Price Index $(P_p)$ and the Fisher Price Index $(P_F)$ are defined as

$$P_L = \Sigma s_n^0 r_n$$ (1)

$$P_p = (\Sigma s_n^1 r_n^{-1})^{-1}$$ (2)

$$P_F = (P_L P_p)^{-1}$$ (3)

where the outlet expenditures shares for period $t$ is given by

$$s_n^t = (p_n^t \cdot q_n^t) / \Sigma p_n^t \cdot q_n^t$$ (4)

Let outlet expenditures shares be equal in both periods, then $s_n^t = 1/N$ for $t = 0, 1$ and $n = 1, \ldots, N$.

$$P_L = \Sigma (1/N) r_n = (1 + i)$$ (5)

where $(1 + i)$ is the mean of the price relatives, and $i$ is the estimated inflation rate.

Let the price ratio $r_n$ deflated by the mean $l + i$ is equal to $\varepsilon_n$.

Therefore $l + \varepsilon_n = r_n / (1 + i)$ and with equation (5),

$$\Sigma \varepsilon_n = 0$$

$$P_L = \Sigma s_n^0 r_n = (1/N) \Sigma r_n = (1/N) \Sigma \varepsilon_n$$

$$P_p = (\Sigma s_n^1 r_n^{-1})^{-1} = (\Sigma 1/N r_n^{-1})^{-1} = ((1/N) \Sigma r_n^{-1})^{-1}$$

$$P_F = (P_L P_p)^{-1}$$

Since the arithmetic mean is greater than the harmonic mean then $P_L > P_p$ unless all the price ratios are equal in which case $P_L = P_p = 1 + i$.

Since the geometric mean lies between the arithmetic and the harmonic mean then $P_L > P_F > P_p$, unless the price ratios are equal in which case $P_L = P_F = P_p = 1 + i$.

The bias is equal to $P_L - P_F$. Dievert (1998) shows that the bias, $P_L - P_F = 5(1 + i)s\varepsilon$. 
Hedonic Regression (HR) is used to separately identify and value the changes in the quality of goods and services. This helps to separate quality changes in the index. In HR, price is a function of measurable variables.

Consider the following multiple regression equations

\[ P_i = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \delta_1 t_1 + \delta_2 t_2 + \epsilon_i \]

\[ \epsilon_i \sim NID(0, \sigma^2) \]

In this multiple regression models, the partial regression coefficients are \( \beta \), which may be interpreted as shadow or implicit prices, \( x_i \) are the cost or features of the goods, which may be either a continuous variable or a dummy variable.

The partial regression coefficients \( \delta_1 \) and \( \delta_2 \) are the price index for time 1 and time 2 respectively. The partial regression coefficients are usually estimated by ordinary least squares. But after tests of the assumptions underlying these models, weighted least squares or some nonparametric method may be used.

Other HR models are the multiplicative models:

\[ \ln P_i = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \epsilon_i \]

The independent variables may be month, outlet types, and area.
APPENDIX 3

Problems with the CPI

1. The CPI at the final stage of aggregation is not a true Laspeyres index, since the expenditure weights, which refer to a base year are selected annually while the prices are collected monthly or quarterly. The base period weights and prices should coincide. The CPI turns out to be a weighted arithmetic average of the price relatives, which is the Young index.

2. At the beginning stages of aggregation, unweighted averages of price relatives (arithmetic, Carli index; geometric, Jevons index; harmonic, Jevons-Coggeshall) are used. However as stated earlier these unweighted indices suffer from a serious upward bias problem. For more details see the “Consumer Price Index Manual: Theory and Practice”, ILO, Geneva, April 2004.

3. The standard fixed based CPI has difficulty incorporating quality changes and new goods in the base basket of goods. There is also some difficulty incorporating hedonic regression in the CPI. Chained indices rather than fixed based indices are used. However these chained Cali indices, which are un-weighted arithmetic averages of price relatives, magnify the positive bias.

4. Seasonal commodities pose a problem for statistical agencies, since as stated in problem one, the expenditure weights refer to a base year but prices, which are selected monthly or quarterly for seasonal goods vary considerably. Using a short-term index and also a seasonally adjusted long-term index can solve the problem of seasonal weights.

5. There are problems measuring complex services. More attention is paid to measuring goods rather than services. In the global economy, these are many services, for example, financial, computing, insurance that are either ignored or not measured accurately.

6. As discussed in problem four there is need for multiple indices to measure the effect of seasonal goods. Another usage of multiple indices is the measurement of owner occupied housing, should it be based on the acquisition approach, the rental equivalence approach or the user cost approach? Maybe there is need to provide indices for each approach. For further discussion on these problems see Diewert (2003).
References


