Session 3 - Elementary aggregation, superlative indexes

Chair: Keith Woolford, Australian Bureau of Statistics

Summary of session

Aggregate index outcomes can be significantly influenced by the choice of index number formula at both the elementary aggregate level and at higher levels of aggregation. In general, the magnitude of any differences delivered by the alternative formulae is a function of the dispersion in the rates of price change or of the change in price dispersion.

The first two papers in this session (by Balk and Silver respectively) are concerned primarily with choosing the formula to be used to compile indexes for elementary aggregates. The third paper, presented by Greenlees, focuses on higher level aggregation and describes the calculation of the new superlative consumer price index compiled by the U.S. Bureau of Labor Statistics (the C-CPI-U).

Balk investigates alternative elementary aggregate formula from a sampling perspective. The population for which the elementary aggregate index is to be constructed is classified as either homogeneous or heterogeneous and some practical guidelines for making this distinction are provided. Alternative sample index formulae are then assessed against relevant alternative target population price indexes using feasible alternative sample designs (simple random sampling and sampling with probability proportional to size). For each sample/population index formula pair, Balk categorises any relative bias in the sample formula into a technical component (which trends to zero as the sample size increases) and a structural component (which is independent of sample size).

In developing his sample approach, Balk outlines a logical progression of steps that should be followed by the prices statistician. First, classify the elementary aggregate as being homogeneous or not. Second, select the appropriate population target price index and finally, according to the sample design, choose that sample price index formula which provides the least biased (or approximately unbiased) estimator of the population target price index.

Silver outlines the nature of the relationships between the various formulae in terms of price dispersion and tries to explain the existence of price dispersion and its persistence over time by reference to search cost and menu cost theories and signal extraction models. He then proceeds to investigate the phenomena using a scanner data set containing some 70,000 observations on television sets over 51 months. Hedonic techniques are used to construct 'heterogeneity-controlled' indexes to compare with the more traditional formulae.

The paper provides some interesting insights into price dispersion both within months and over time and concludes by drawing out some practical advice to assist in choosing the elementary aggregate formula and in the selection and replacement of representative items.

Greenlees provides a comprehensive description of the methodology used to calculate the Chained Consumer Price Index for All Urban Consumers (C-CPI-U) that has been published by the U.S. Bureau of Labor Statistics since August 2002. This superlative price index is

constructed using the Tornqvist formula and is published in parallel with the more longstanding CPIs compiled for Urban Wage Earners and Clerical Workers (CPI-W) and All Urban Consumers (CPI-U). The CPI-W and CPI-U are compiled using the more conventional Laspeyres type fixed-base formula and are not generally revisable (at least in response to new weighting information becoming available). By comparison, the C-CPI-U is first published in preliminary form and is subject to two subsequent revisions as more recent expenditure data becomes available.

In addition to describing how the index is calculated, the paper discusses the various issues that were addressed in developing this new measure and comments on differences in outcomes compared to the CPI-U. The latter appear to be rather high (0.8 percent in 2000 and 0.3 percent in 2001).

Recommendations for statistical agencies

These three papers provide a number of valuable insights into the issues involved in selecting the most appropriate formula for elementary and higher level index aggregation Three recommendations could be drawn from this material and the discussions:

1. The choice of elementary aggregate formula is important, especially when the price dispersion within the elementary aggregate is high. It should be made carefully according to the way in which representative products are sampled.

2. One should exercise care in using the relative of arithmetic mean prices formula (Dutot) for other than homogeneous commodities.

3. The method used by the U.S. Bureau of Labor Statistics to calculate the C-CPI-U and the issues addressed in its design are a sound starting point for any national statistical agency contemplating the construction of a superlative index.