The Consumer Price Index: Recent Developments

Presentation at the Economic Measurement Group Workshop Asia 2013: Data Gaps and Economic Measurement, October 16, 2013

Erwin Diewert

University of British Columbia and

University of New South Wales

Introduction

- **The 2004 International Labour Office *Consumer Price Index Manual: Theory and Practice* summarized the state of the art for constructing Consumer Price Indexes (CPIs).**
- **In the intervening decade, there have been some significant new developments which are reviewed in this paper.**
- **The *CPI Manual* recommended the use of chained superlative indexes for a month to month CPI. However, subsequent experience with the use of monthly scanner data has shown that a significant chain drift problem can occur.**
- **We will discuss two recently developed methods that can deal with the chain drift problem:**

  (1) Rolling Year GEKS and

  (2) The Weighted Time Product Dummy Method

Alternative Approaches to the Target Index

The following 4 approaches were explained in the *ILO Manual*:

- **Fixed basket and averages of fixed basket approaches;**
- **The test or axiomatic approach;**
- **The stochastic approach and**
- **The economic approach.**

These four approaches are explained in the paper. Basically, these four approaches lead to the Fisher Ideal Index or the Törnqvist-Theil indexes as suitable target indexes. Furthermore, the *Manual* endorsed chained versions of these indexes as target indexes.

The Chain Drift Problem and Possible Solutions

Fixed base **(or direct)** ***indexes*** **for 3 periods can be constructed as follows ($P(p^0,p^1,q^0,q^1)$ is your favorite index formula):**

- $P^0 \equiv 1$; $P^1 \equiv P(p^0,p^1,q^0,q^1)$; $P^2 \equiv P(p^0,p^2,q^0,q^2)$.

Thus the prices in period 2, p^2, are compared ***directly*** **with the prices in period 0, p^0.**

The price levels, P^0, P^1 and P^2, using ***chained indexes*** **can be constructed as follows:**

- $P^0 \equiv 1$; $P^1 \equiv P(p^0,p^1,q^0,q^1)$; $P^2 \equiv P(p^0,p^1,q^0,q^1)P(p^1,p^2,q^1,q^2)$.

Fixed base and chained indexes will coincide if the index formula satisfies the circularity test. But P_F and P_T do not!

The Chain Drift Problem and Possible Solutions (cont)

- ***Chain drift*** **occurs when an index does not return to unity when prices in the current period return to their levels in a distant base period; i.e., we want the index to satisfy Walsh's Multiperiod Identity Test:**

(27) $P(p^0,p^1,q^0,q^1)P(p^1,p^2,q^1,q^2)P(p^2,p^0,q^2,q^0) = 1.$

- **The *Manual* did not take into account how severe the chain drift problem could be in practice.**
- **The problem is mostly caused by periodic *sales* of products. An example will illustrate the problem.**
- **Suppose that we are given the following price and quantity data for 2 commodities for 4 periods (see next slide):**

The Chain Drift Problem: An Example

Period t	p_1^t	p_2^t	q_1^t	q_2^t
1	1.0	1.0	10	100
2	0.5	1.0	5000	100
3	1.0	1.0	1	100
4	1.0	1.0	10	100

- **Item 2 sells 100 units at a constant price of 1 every period.**
- **Item 1 sells 10 units in period 1 at the "normal" price of 1**
- **But in period 2, item 1 goes on sale at half price and the sales volume shoots up to 5000 units sold.**
- **In period 3, item 1 is back to its normal price of 1 but customers have stocked up on the item in period 2 so sales are down to 1 unit in period 3.**
- **In period 4, the effects of the sale have worn off and item 1 is back to the price and quantity sold of period 1.**

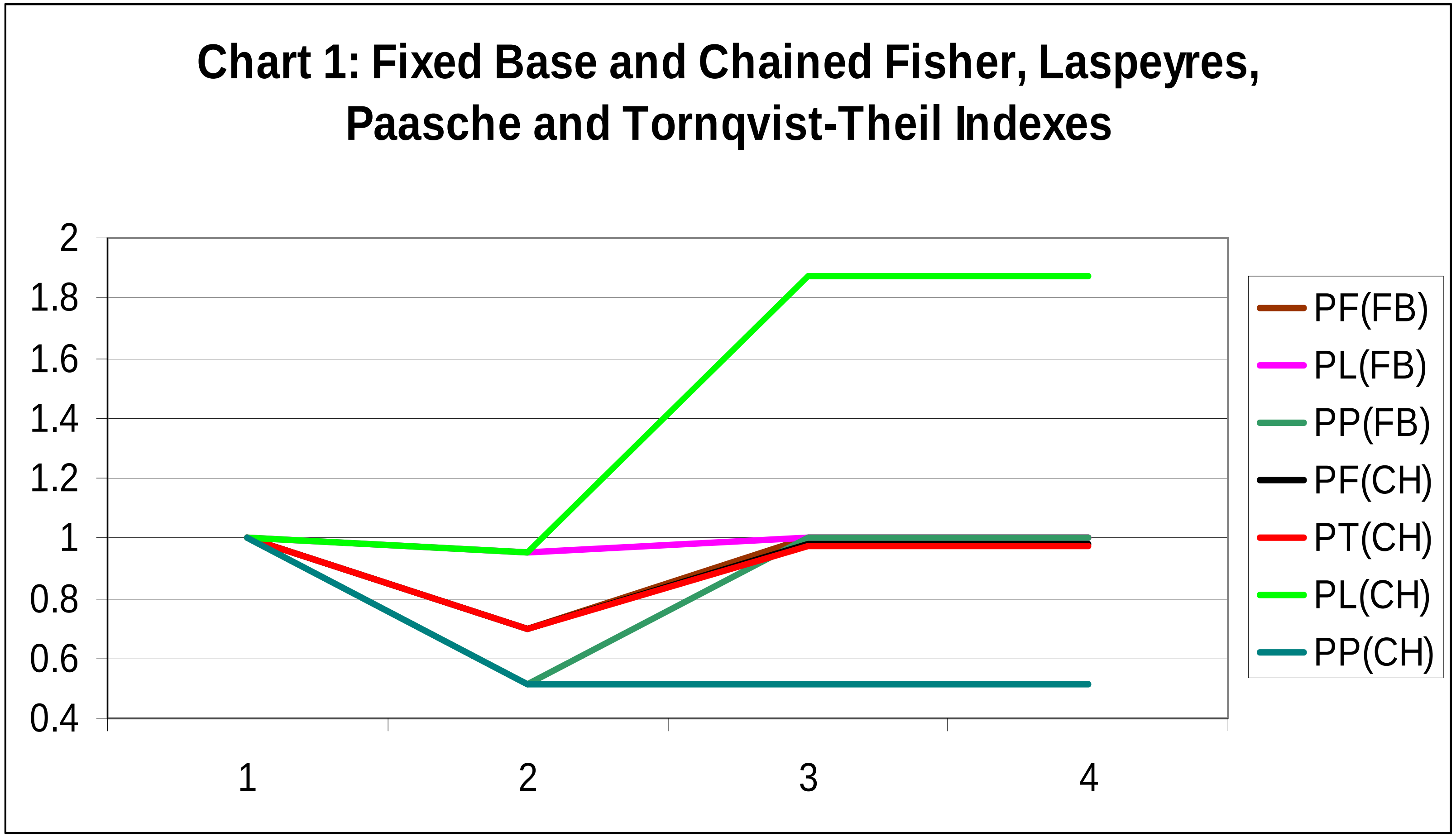
The Chain Drift Problem: An Example (cont)

Table 2: Fixed Base and Chained Fisher, Törnqvist-Theil, Laspeyres and Paasche Indexes

Period	$P_{F(FB)}$	$P_{L(FB)}$	$P_{P(FB)}$	$P_{F(CH)}$	$P_{T(CH)}$	$P_{L(CH)}$	$P_{P(CH)}$
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	0.698	0.955	0.510	0.698	0.694	0.955	0.510
3	1.000	1.000	1.000	0.979	0.972	1.872	0.512
4	1.000	1.000	1.000	0.979	0.972	1.872	0.512

- **The fixed base indexes all end up where they started in period 4 so there is no chain drift problem with them.**
- **But the chained Laspeyres ends up at 1.872 (way too high) and the chained Paasche ends up at 0.512 (way too low).**
- **But the chained Fisher and Törnqvist-Theil indexes end up 2-3% too low in period 4. This is a significant bias.**

The Chain Drift Problem: An Example (cont)



The Chain Drift Problem: Possible Solutions

There are at least three possible solutions to the chain drift problem that is associated with the use of a superlative index in a situation where monthly scanner data is available to the statistical agency for components of the CPI:

- **Stick to the usual annual basket Lowe index that uses annual expenditure weights from a past year;**
- **Use Rolling Year GEKS to control for chain drift or**
- **Use the Weighted Time Dummy Product method to control for chain drift.**

The problem with the use of the Lowe index is that it will usually generate upward substitution bias equal to 0.15 to 0.4 percentage points per year, which is significant.

[There is also the solution of Handbury, Watanabe and Weinstein which is to use year over year monthly superlative indexes. But this index is not suitable for all purposes.]

What is GEKS?

- **Suppose we have price and quantity information for a component of the CPI on a monthly basis for a sequence of 13 consecutive months.**
- **Now pick one month (say month k) in this augmented year as the base month and construct Fisher price indexes for all 13 months relative to this base month.**
- **Denote the resulting sequence of Fisher indexes as $P_F(1/k)$, $P_F(2/k)$, ..., $P_F(13/k)$.**
- **The final set of GEKS indexes for the 13 months is simply geometric mean of all 13 of the specific month indexes; i.e., the final set of *GEKS indexes for the months in the augmented year* is any normalization of the following sequence of indexes:**

(28) $[\prod_{k=1}^{13} P_F(1/k)]^{1/13}, [\prod_{k=1}^{13} P_F(2/k)]^{1/13}, \ldots , [\prod_{k=1}^{13} P_F(13/k)]^{1/13}$.

Properties of GEKS Indexes

- **They satisfy Walsh's multiperiod identity test so that if any two months in the augmented year have exactly the same price and quantity vectors, then the above index values will coincide for those two months; i.e., *the above indexes are free from chain drift*.**
- **The above indexes do not asymmetrically single out any single month to play the role of a base period; all possible base months contribute to the overall index values.**
- **The above indexes make use of all possible bilateral matches of the price data between any two months in the augmented year.**
- **Strongly seasonal commodities make a contribution to the overall index values.**

Rolling Year GEKS

- **The major problem with the GEKS indexes defined by (28) is that the indexes change as the data for a new month becomes available. Typically, CPIs cannot be revised.**
- **Add the price and quantity data for the most recent month to the current year data and drop the oldest month from the data set.**
- **The GEKS indexes for the new augmented year are calculated in the usual way and the ratio of the index value for the last month in the new augmented year to the index value for the previous month in the new augmented year is used as an *update factor* for the value of the index for the last month in the previous augmented year.**
- **This is the Rolling Year GEKS method. IDF (2009)**
- **It is free from chain drift to a high degree of approximation.**

The Weighted Time Product Dummy Model

- **Robert Summers (1973) invented the (unweighted) Country Product Dummy Method for constructing indexes across countries.**
- **Several people saw that the method could be adapted to time series index construction (Aizcorbe, Corrado and Doms (2003), de Haan and Krsinich (2012) and Diewert (2012)).**
- **And weighted versions of the method were also suggested by Rao (2004) and Diewert (2004) (2005).**
- **We will try to explain the method in the next couple of slides.**

The Time Product Dummy Model (cont)

Let p_{tnk} denote the price of item n in outlet k in time period t for t = 1,…,T; n = 1,…,N; k = 1,…,K. The basic statistical model that is assumed is the following one:

$$(29)\quad p_{tnk} = a_t b_n u_{tnk}\ ; \qquad t = 1,\ldots,T;\ n = 1,\ldots,N;\ k = 1,\ldots,K$$

where the a_t and b_n are unknown parameters to be estimated and the u_{tnk} are independently distributed error terms with means 1 and constant variances.

The parameter a_t is to be interpreted as the *average level of prices* (over all items in this group of items) in time period t and the parameter b_n is to be interpreted as *multiplicative units of measurement factor* that is specific to product n.

The Time Product Dummy Model (cont)

- **Taking logarithms of both sides of (29) leads to the following model:**

(30) $y_{tnk} = \alpha_t + \beta_n + \varepsilon_{tnk}$; $t = 1,\ldots,T; n = 1,\ldots,N; k = 1,\ldots,K$

where $y_{tnk} \equiv \ln p_{tnk}$, $\alpha_t \equiv \ln a_t$, $\beta_n \equiv \ln b_n$ **and** $\varepsilon_{tnk} \equiv \ln u_{tnk}$.

- **The model defined by (30) is a linear regression model where the independent variables are dummy variables. The least squares estimators for the** α_c **and** β_n **can be obtained by solving the following least squares minimization problem:**

(31) $\{\sum_{t=1}^{T} \sum_{n=1}^{N} \sum_{k=1}^{K} [y_{tnk} - \alpha_t - \beta_n]^2\}$.

- **We also require a normalization on the** α_t **and** β_n **such as** $\alpha_1 = 0$.
- **The above model is a completely balanced panel model. In real life, there is sample attrition and so we need to deal with this as well as to introduce economic weighting into the model.**

The Weighted Time Product Dummy Model

- **For product n in time period t, we assume that there are K(t,n) outlets that have transactions in product n and that the *unit value price* for the kth such transaction is p_{tnk} and the associated *quantity transacted* is q_{tnk} for k = 1,2,…,K(t,n).**

- **Again, $y_{tnk} \equiv \ln p_{tnk}$ is the logarithm of the price p_{tnk}. For each time period t, we use the prices and quantities p_{tnk} and q_{tnk} in order to form the following *period t expenditure shares* across all products n and all outlets k:**

$$(33)\quad s_{tnk} \equiv p_{tnk}q_{tnk} / \sum_{i=1}^{N} \sum_{j=1}^{K(t,n)} p_{tij}q_{tij} ; \qquad t = 1,...,T ; \quad n = 1,\ldots,N ; k = 1,\ldots,K(t,n).$$

- **For each time period t, these expenditure shares sum up to 1:**

$$(34)\quad \sum_{n=1}^{N} \sum_{k=1}^{K(t,n)} s_{tnk} = 1 ; \qquad t = 1,\ldots,T.$$

The Weighted Time Product Dummy Model (cont)

- **The *Weighted Time Product Dummy* (WTPD) counterpart to the unweighted least squares minimization problem (31) above is:**

(35) $\{\sum_{t=1}^{T} \sum_{n=1}^{N} \sum_{k=1}^{K(t,n)} s_{tnk} [y_{tnk} - \alpha_t - \beta_n]^2\}$.

- **Again, the parameters α_t and β_n cannot be uniquely identified so we will choose to set the price level in period 1, $a_1 \equiv \exp[\alpha_1]$, equal to 1, which implies the following normalization $\alpha_1 = 0$ on the parameters appearing in (35).**
- **The WTPD generates price levels that are free of chain drift.**

The Rolling Year Weighted Time Product Dummy Model

- **The WTPD price level estimates suffer from the same problem that the GEKS estimates suffer from: the addition of one more period to the sample will change all of the estimates.**
- **Thus Ivancic, Diewert and Fox (2009) proposed a Rolling Year approach to the Weighted Time Product Dummy (RYTPD) estimation procedure; i.e., set T = 13 and as a new month's data is added, delete the data for the oldest month in the sample, obtain new WTPD estimates and use the month over month movement in the estimated price levels for the last two months to update the previous estimates.**
- **Both the RYGEKS and RYWTPD methods seem to be largely free of chain drift.**
- **There is tendency for the RYWTPD estimates to be slightly less than their RYGEKS counterparts. The reasons for this are not yet fully understood. But …. See the paper!**

Elementary Indexes: New Developments

- **The new development here is the suggestion that the (unweighted) Rolling Year TPD method be used in place of the Jevons index (which has the best axiomatic properties when there are no strongly seasonal commodities or missing observations).**
- **Diewert (2012) and de Haan and Krsinich (2012) both suggested this method.**
- **For countries that use a fixed base month in their index (like the Eurostat HICP and the RPI in the UK), this suggestion will eliminate the massive asymmetry in their present indexes; i.e., the base month in these indexes plays an asymmetric role.**

New Approaches to Quality Adjustment

- **A problem with the RYGEKS and RYWTPD methods described above is that these methods do not deal adequately with the introduction of new products.**
- **De Haan and Krsinich (2012) (2013) invented a method that deals with this problem. The basic building block in their method is a time dummy hedonic regression model that uses the data for two periods.**
- **The dependent variable in the model is the logarithm of the item price and a time dummy and various characteristics of the product enter the regression as independent variables.**
- **The time dummy coefficient and the characteristic "prices" are the result of a weighted least squares minimization problem. If an item appears in both periods under consideration, the weights in the weighted regression are the (arithmetic) average of the expenditure shares for the item in the two periods; if the item appears in only one of the two periods, one half of the expenditure share on the item for that period is used as the weight.**

New Approaches to Quality Adjustment (cont)

- **The resulting bilateral price index turns out to equal the usual Törnqvist index if all items are present in both periods but for unmatched items, an imputed price for the missing price enters the index number formula and this imputed price is obtained as a predicted price using hedonic regression.**
- **Thus when there are unmatched items in the two periods under consideration, we obtain a generalization of the usual Törnqvist index that makes use of imputed prices from the hedonic regression and hence de Haan and Krsinich (2012) (2013) call the resulting bilateral index number formula the *Imputation Törnqvist index*.**
- **They proposed the following variation of the Rolling Year GEKS method: instead of using bilateral Fisher indexes as the basic building blocks, the Fisher indexes are replaced by bilateral Imputation Törnqvist indexes.**
- **They call the resulting indexes ITRYGEKS indexes.**

Other Problems Needing Attention!

- Should the CPI be compiled on a domestic, national or household inflation basis?
- What is the appropriate treatment of Owner Occupied Housing (OOH) in the CPI?
- How exactly should financial services be treated in the CPI?
- How can strongly seasonal commodities make a contribution to the month to month CPI?
- Is it possible to construct real time CPIs using current month information on prices and older information on household expenditure shares that will approximate a superlative CPI that is constructed later when additional data on expenditures shares become available?