Chain Drift in Some Price Index Estimators

Janice Lent
Bureau of Transportation Statistics,* 400 7th St. S.W., Rm. 3430, Washington, DC 20590

1. Introduction

Price index *chaining* is estimating long-term price changes as products of shorter-term changes ("links"). For example, suppose the price of a widget moves as shown below for time periods 1 through 4:

Time:	1	2	3	4
Price:	1.00	1.10	0.55	1.10

A direct measure the change between periods 1 and 4 is

$$I_{1,4} = \frac{1.10}{1.00} = 1.10.$$

The corresponding chained measure is

$$\begin{aligned} I'_{1,4} &= I_{1,2}I_{2,3}I_{3,4} \\ &= \left(\frac{1.10}{1.00}\right)\left(\frac{0.55}{1.10}\right)\left(\frac{1.10}{0.55}\right) = 1.10. \end{aligned}$$

In practice, the intermediate links $I_{t-1,t}$ may be *estimated* changes for a *category* of items; in this case, we generally have $I_{1,4} \neq I'_{1,4}$.

For practical reasons, index chaining is widely used by government statistical agencies in computing price indexes. Both the universe of available goods and the sample of items used for index computation are in constant flux; new products are introduced daily, and the sample is routinely rotated to keep pace. Short-term changes may therefore be measured more accurately than long-term changes: the samples for two consecutive months, for instance, contain many more comparable items than the samples for two months one year apart. Chained index estimators, however, are subject to systematic biases relative to their direct counterparts. The magnitude and direction of the "chain drift" depend on the index aggregation formula, the economic behavior of the purchasing population, and the properties of the sample survey data used in estimation.

The theory of index chaining originated with Divisia (1925), who used integral calculus to formulate a chained index with arbitrarily short links. Richter (1966) presented invariance axioms for a variety of index numbers, including price indexes. While the Divisia index satisfied his axioms, it lacked independence of the "path"—the series of intermediate links. More recently Forsyth (1978) and Forsyth & Fowler (1981) expanded this idea, concluding that the choice of chaining interval (or link length) came down to a choice between "transitivity and representativity," direct indexes providing the former and chained indexes more of the latter.

In this paper, we examine the discrepancies ("drift factors") between the chained and direct versions of the geometric mean and Törnqvist index formulas (defined below). After briefly introducing the index formulas, we analyze the drift factors in terms of (1) assumptions about the population's economic behavior and (2) statistical correlations between components of the index estimators. We then use our analytical results, together with an empirical investigation, to examine the magnitude of and reasons for the chain drift in some price index estimators. Finally, we relate our research to the "transitivity vs. representativity" trade-off identified by Forsyth & Fowler, noting in particular the impact of the survey data to be used in estimating the chosen index formula.

2. Introduction to Price Indexes

A *consumer price index* (CPI) is a measure of change in the value of a monetary unit. In estimating a CPI, government statistical agencies generally adopt one of two approaches: (a) the fixed market basket (or Laspeyres-type) approach, or (b) the cost of living index (COLI) approach.

For approach (a), we simply select a collection of consumer goods and services (in fixed quantities) and track the total price of this "market basket" across time. The fixed market basket idea underlies the Laspeyres index

$$L_{t_0,t} = \frac{\sum_{i=1}^{N} q_{i,t_0}p_{i,t}}{\sum_{i=1}^{N} q_{i,t_0}p_{i,t_0}} = \sum_{i=1}^{N} w_{i,t_0}\left(\frac{p_{i,t}}{p_{i,t_0}}\right), \quad (2.1)$$

where, for each item i in the population of N goods and services, $p_{i,j}$ and $q_{i,j}$ represent the price and quantity purchased, respectively, in time period j and $w_{i,j} = q_{i,j}p_{i,j}/\sum_k q_{k,j}p_{k,j}$. Replacing q_{i,t_0} in formula 2.1 with $q_{i,B}$, where B is a "base period" prior to t_0, yields the Modified Laspeyres index, which, until recent years, served as the target population index for the U.S. CPI. In practice, data on quantities purchased are not available; estimated expenditure shares $w_{i,j}$ may be computed, but these generally become available only on a lagged basis. Hence the use of a Modified Laspeyres index, rather than a pure Laspeyres, was due primarily to data availability issues.

*Research for this paper was performed at the Bureau of Labor Statistics (BLS). Opinions expressed are those of the author and do not constitute policy of the BLS or the BTS.

In the COLI approach to index calculation, we seek to estimate the change in the cost of a fixed level of consumer satisfaction or "utility," rather than the change in the cost of a fixed collection of goods and services. Motivated by economic utility theory, this approach allows for the possibility that, as relative prices change, consumers may revise their market baskets to obtain a constant level of utility across time. That is, they need not maintain a fixed market basket in order to maintain a fixed level of satisfaction, since alternative bundles may provide equal satisfaction. In economic terms, this phenomenon is called *substitution;* the *elasticity of substitution* (which we denote by η) is a measure of the extent to which consumers adjust their purchases in response to price change while maintaining a fixed level of utility.

The fixed market basket assumption is equivalent to setting $\eta = 0$, while setting $\eta = 1$ implies the assumption that consumers substitute towards lower priced goods in such a way that the proportion of their total expenditures for each item (or item category) remains constant across time. Under the latter assumption, the geometric mean (or Jevons) index, defined as

$$G_{t_0,t} = \prod_{i=1}^{N} \left(\frac{p_{i,t}}{p_{i,t_o}} \right)^{w_{i,t_0}}, \qquad (2.2)$$

where w_{i,t_0} is as for equation 2.1, is thought to accurately reflect changes in the true cost of living.

The U.S. CPI is often used as an inflator to convert monetary figures to constant dollars, and the fixed-basket approach is deemed inconsistent with this usage. In the past several years, the Bureau of Labor Statistics (BLS) has therefore made a series of changes to the CPI in an effort to convert it from a fixed market basket index to a COLI. Developments in price index theory, along with technological advances (e.g., the use of electronic scanners) that have enlarged the pool of available price data, have made a COLI seem a more feasible target than it formerly was. Economic theory suggests that some population index formulas provide estimates of the true change in the cost of living, accounting for substitution, regardless of the value of η. One such "superlative" index (see, for example, Diewert 1987) is the Törnqvist index, given by

$$T_{t_0,t} = \prod_{i=1}^{N} \left(\frac{p_{i,t}}{p_{i,t_o}} \right)^{\left(w_{i,t_0}+w_{i,t}\right)/2}. \qquad (2.3)$$

Unlike the Laspeyres and geometric mean indexes, formula 2.3 involves expenditure shares ($w_{i,j}$) from both periods t_0 and t. By using this additional data (which, in practice, is often unavailable), we may correctly account for consumer substitution behavior, avoiding the uncertainty associated with an assumed elasticity. (For more on COLI estimation, see Shapiro and Wilcox 1997 or Dorfman, Leaver, and Lent 1999.)

3.Estimating Indexes from Sample Data

Statistical agencies often compute price index estimates through a series of aggregation stages. In the first stage of aggregation, similar goods and services bought by consumers living in particular geographic areas are grouped to together to form *item strata.* Examples of item strata include uncooked ground beef in Philadelphia and breakfast cereal in Atlanta. For each item stratum, a sample of items is selected from representative outlets; an expendure weight w_{iB} may be estimated for each sample item i in some base period B. Field economists then track the prices of the sample items across time. Monthly sample price data and base period expenditure data may then be aggregated within each item strata using an estimator of a particular index formula. For item stratum i, formula 2.2 may be estimated as

$$P_{i,t_0,t} = \prod_{k=1}^{n} \left(\frac{p_{k,i,t}}{p_{k,i,t_o}} \right)^{\widehat{w}_{k,i,B}}$$

where n is the sample size, and $p_{k,i,j}$ is the price of the k^{th} sample item in stratum i at time j. Note that, under the assumption of unitary elasticity, $\widehat{w}_{k,i,B}$ (based on expenditure data for period B) may be considered an estimator of w_{k,i,t_0}.

The first stage of aggregation results in a collection of sub-indexes $P_{i,t_0,t}$. These sub-indexes are further aggregated to form higher level indexes. The higher-level aggregation may involve expenditure information collected from a separate survey, such as the U.S. Consumer Expenditure Survey (CEX, a household survey). For example, we may apply a Törnqvist formula at the upper level:

$$\widehat{T}_{t_0,t} = \prod_{i} \left(\frac{P_{i,t}}{P_{i,t_o}} \right)^{\left(\widehat{w}_{i,t_0}+\widehat{w}_{i,t}\right)/2},$$

where the weights $\widehat{w}_{i,t_0}$ and $\widehat{w}_{i,t}$ are estimated from household survey data.

4. Chain Drift Factors for Geometric Mean Indexes

We define the drift factor of a chained index as the ratio of the chained index to the direct index. Here we examine the drift factors for the geometric mean and Törnqvist index formulas and offer possible economic interpretations of the factors. (For an investigation of the Laspeyres drift factor, see Szulc 1983.) We consider only index formulas used to perform upper-level aggregation.

Notation: Let i index the set of item strata, and let $w_{i,j}$ denote the expenditure share weight of item stratum i in month j. Let $P_{i,j}$ denote the sub-index for the i^{th} item stratum, measuring change from the base period B (implied) to month j. In general, the index estimator I_{j_1,j_2} indicates price change between month j_1 and j_2.

Computation of Drift Factors
We define the binary geometric mean index measuring change from month t_0 to month t as

$$G_{t_0,t} = \prod_i \left(\frac{P_{i,t}}{P_{i,t_0}}\right)^{w_{i,t_0}};$$

we define the corresponding monthly chained geometric mean as

$$G'_{t_0,t} = \prod_{j=t_0+1}^{t} \prod_i \left(\frac{P_{i,j}}{P_{i,j-1}}\right)^{w_{i,j-1}}.$$

Thus

$$\begin{aligned} G'_{t_0,t} &= \prod_i \prod_{j=t_0+1}^{t} \left(\frac{P_{i,j}}{P_{i,j-1}}\right)^{w_{i,j-1}} \\ &= G_{t_0,t}\left\{\prod_i f_{i,G}\right\}, \end{aligned}$$

where

$$f_{i,G} = \prod_{j=t_0+1}^{t-1} \left(\frac{P_{i,t}}{P_{i,j}}\right)^{w_{i,j}-w_{i,j-1}}$$

is the geometric mean's "drift factor" for the i^{th} item/area. To see the direction of the bias, we may write

$$G'_{t_0,t} = G_{t_0,t}\left\{\prod_{j=t_0+1}^{t-1}\left[\frac{\prod_i \left(\frac{P_{i,t}}{P_{i,j}}\right)^{w_{i,j}}}{\prod_i \left(\frac{P_{i,t}}{P_{i,j}}\right)^{w_{i,j-1}}}\right]\right\}.$$

Empirical evidence suggests that, for the "average" U.S. consumer, $0 < \eta < 1$; that is, some substitution occurs, but not enough to render item expenditure shares constant. When we have $\eta < 1$, and the prices and weights follow consistent trends, we expect

$$\prod_i \left(\frac{P_{i,t}}{P_{i,j}}\right)^{w_{i,j}} > \prod_i \left(\frac{P_{i,t}}{P_{i,j}}\right)^{w_{i,j-1}}, \qquad (4.1)$$

for $j = t_0+1, ..., t-1$, because the index on the right-hand side is based on outdated weights. Thus we expect an upward chain drift for the geometric mean.

Similarly, we may calculate the drift factors $f_{i,T}$ for the chained Törnqvist index. We define the binary and monthly chained Törnqvist indexes, respectively, as

$$T_{t_0,t} = \prod_i \left(\frac{P_{i,t}}{P_{i,t_0}}\right)^{(w_{i,t_0}+w_{i,t})/2}$$

and

$$T'_{t_0,t} = \prod_{j=t_0+1}^{t} \prod_i \left(\frac{P_{i,j}}{P_{i,j-1}}\right)^{(w_{i,j-1}+w_{i,j})/2}.$$

Then, by algebraic manipulation similar to that which we used to obtain $f_{i,G}$, we have

$$\begin{aligned} T'_{t_0,t} &= \prod_i \prod_{j=t_0+1}^{t} \left(\frac{P_{i,j}}{P_{i,j-1}}\right)^{(w_{i,j-1}+w_{i,j})/2} \\ &= T_{t_0,t}\left\{\prod_i f_{i,T}\right\}, \end{aligned}$$

where the factor $f_{i,T}$ is defined as

$$\left\{\prod_{j=t_0+1}^{t-1}\left[\left(\frac{P_{i,t}}{P_{i,j}}\right)^{w_{i,j}-w_{i,j-1}} \left(\frac{P_{i,t_0}}{P_{i,j}}\right)^{w_{i,j+1}-w_{i,j}}\right]\right\}^{1/2}$$

Note that

$$f_{i,T} = \left\{f_{i,G} \cdot \prod_{j=t_0+1}^{t-1} \left(\frac{P_{i,t_0}}{P_{i,j}}\right)^{w_{i,j+1}-w_{i,j}}\right\}^{1/2}.$$

Again, we write the bias factor in terms of indexes to identify the direction of the bias:

$$\begin{aligned} T'_{t_0,t} &= T_{t_0,t}\left\{\prod_{j=t_0+1}^{t-1} \frac{\prod_i \left(\frac{P_{i,t}}{P_{i,j}}\right)^{w_{i,j}}}{\prod_i \left(\frac{P_{i,t}}{P_{i,j}}\right)^{w_{i,j-1}}}\right\}^{1/2} \\ &\cdot \left\{\prod_{j=t_0+1}^{t-1} \frac{\prod_i \left(\frac{P_{i,j}}{P_{i,t_0}}\right)^{w_{i,j}}}{\prod_i \left(\frac{P_{i,j}}{P_{i,t_0}}\right)^{w_{i,j+1}}}\right\}^{1/2}. \end{aligned}$$

Given that elasticity is less than one and the prices and weights follow consistent trends, we expect both inequality 4.1 and

$$\prod_i \left(\frac{P_{i,j}}{P_{i,t_0}}\right)^{w_{i,j}} < \prod_i \left(\frac{P_{i,j}}{P_{i,t_0}}\right)^{w_{i,j+1}}$$

for $j = t_0+1, ..., t-1$. So the direction of the chain drift for the Törnqvist index is indeterminate.

We note the following about the drift factors $f_{i,T}$, as compared to the $f_{i,G}$ of the geometric mean:

1. The drift factor in each (multiplicative) term of $f_{i,T}$ contains two price relatives which, if prices and shares are steadily increasing or decreasing, should neutralize each other.

2. The two products of price relatives in $f_{i,T}$ are raised to the 1/2 power; thus if their product is close to 1, $f_{i,T}$ will be even closer to 1.

5. Effect of Correlations on Chain Drift

Following Forsyth and Fowler (1981), we may also examine the drift in the geometric mean and Törnqvist indexes in terms of the covariances between the ratio of sub-indexes $r_{i;j_1,j_2} = P_{i,j_2}/P_{i,j_1}$ and the expenditure weights $w_{i,j}$. Here the $w_{i,j}$ represent expenditure shares for the m item strata. We denote the usual covariance between two vectors $\mathbf{x}$ and $\mathbf{y}$ of order m by

$$c(\mathbf{x}, \mathbf{y}) = \frac{1}{m} \sum_{k=1}^{m} x_k y_k - \frac{1}{m^2} \left(\sum_{k=1}^{m} x_k \right) \left(\sum_{k=1}^{m} y_k \right).$$

Writing $\mathbf{r}_{j_1,j_2} = \{r_{1;j_1,j_2}, ..., r_{m;j_1,j_2}\}$ and $\mathbf{w}_j = \{w_{1,j}, ..., w_{m,j}\}$, we have

$$\begin{aligned}
\ln\left(\frac{G'_{t_0,t}}{G_{t_0,t}}\right) &= m \sum_{j=t_0+1}^{t} c(\ln \mathbf{r}_{j-1,j}, \mathbf{w}_{j-1}) \\
&\quad - mc(\ln \mathbf{r}_{t_0,t}, \mathbf{w}_{t_0}) \\
&= m \sum_{j=t_0+1}^{t} c(\ln \mathbf{r}_{j-1,j}, \mathbf{w}_{j-1}) \\
&\quad - mc\left(\sum_{j=t_0+1}^{t} \ln \mathbf{r}_{j-1,j}, \mathbf{w}_{t_0} \right) \\
&= m \sum_{j=t_0+1}^{t} c(\ln \mathbf{r}_{j-1,j}, \mathbf{w}_{j-1} - \mathbf{w}_{t_0}).
\end{aligned}$$

When both prices and weights follow smooth and consistent trends with elasticity is less than 1, $\ln \mathbf{r}_{j-1,j}$ and $\mathbf{w}_{j-1} - \mathbf{w}_{t_0}$ will be positively correlated (i.e., expenditure shares will rise for those item strata whose prices are rising more quickly). In this case, we will have

$$c(\ln \mathbf{r}_{j-1,j}, \mathbf{w}_{j-1} - \mathbf{w}_{t_0}) > 0,$$

for all j, giving a positive drift factor. When prices "bounce" from period to period, however, we may have

$$c(\ln \mathbf{r}_{j-1,j}, \mathbf{w}_{j-1}) < c(\ln \mathbf{r}_{j-1,j}, \mathbf{w}_{t_0}) < 0.$$

This occurs because, at the population level,

$$r_{i;j-1,j} = P_{i,j}/P_{i,j-1},$$

while

$$w_{i,j-1} = \frac{P_{i,j-1} Q_{i;j-1}}{\sum_i P_{i,j-1} Q_{i;j-1}},$$

where $Q_{i,j}$ represents quantities of items purchased in item stratum i in time period j. Note that $P_{i;j-1}$ appears in the numerator of $w_{i;j-1}$ and in the denominator of $r_{i;j-1,j}$. Given $\eta < 1$ (i.e., insufficient substitution to render expenditure shares constant), this may result in a relatively strong negative correlation between $\mathbf{r}_{j-1,j}$ and $\mathbf{w}_{j-1}$ (stronger than any negative correlation between $\mathbf{r}_{j-1,j}$ and $\mathbf{w}_{t_0}$). In this case, the geometric mean index will display a downward chain drift. In practice, however, $\mathbf{r}_{j-1,j}$ and $\mathbf{w}_{t_0}$ may be estimated using data from different surveys—perhaps even from different time periods—and the negative correlation between the estimators may be quite weak.

A similar development for the Törnqvist index yields

$$\begin{aligned}
\ln\left(\frac{T'_{t_0,t}}{T_{t_0,t}}\right) &= m \sum_{j=t_0+1}^{t} c\left(\ln \mathbf{r}_{j-1,j}, \frac{\mathbf{w}_{j-1} + \mathbf{w}_j}{2}\right) \\
&\quad - mc\left(\ln \mathbf{r}_{t_0,t}, \frac{\mathbf{w}_{t_0} + \mathbf{w}_t}{2}\right) \\
&= \frac{m}{2} \sum_{j=t_0+2}^{t} c(\ln \mathbf{r}_{j-1,j}, \mathbf{w}_{j-1} - \mathbf{w}_{t_0}) \\
&\quad - \frac{m}{2} \sum_{j=t_0+1}^{t-1} c(\ln \mathbf{r}_{j-1,j}, \mathbf{w}_t - \mathbf{w}_j).
\end{aligned}$$

6. Empirical Results from CPI Data

The analyses presented in Sections 4 and 5 hold both at the population level and in the practical case in which the index components ($w_{i,j}$ and $P_{i,j}$) are sample-based estimators. In this section, we present results of an empirical analysis, based on upper-level aggregation data (subindexes and item stratum weights) for the U.S. CPI. The expenditure share weights are based on CEX data, while the prices used to compute the subindexes are from the CPI monthly survey of retail and service outlets. The subindexes themselves were computed by a chained modified Laspeyres formula.* The data suggest that, at the item stratum level, we have $\eta < 1$, i.e., substitution across item strata is insufficient to render stratum expenditure shares constant across time. While both surveys are conducted monthly, the CEX employs a quarterly sample design, and monthly estimates computed from CEX data are subject to high sampling variability as well as potential deficiencies in coverage. In our analysis, we test several methods of smoothing the monthly weights, and we note the effect of the smoothing for various index formulas. Intuitively, we expect weight smoothing to (1) weaken the negative correlations between the weights and

*Some of the subindexes are believed to be subject to an upward "formula bias," as discussed by Reinsdorf 1998.

subindexes, and (2) reduce the effect of "price bouncing" on the estimates, rendering them more reflective of long-term trends. A stronger smoothing algorithm (e.g., a simple 13-month moving average) should have a more pronouced effect than a weaker algorithm (e.g., a weighted moving average which assigns greater weight to the central month).

Tables 1a, 2a, and 3a below give values of monthly-chained long-term Laspeyres, geometric mean, and Törnqvist estimators, where the base month is June 1987 (i.e., the index for June 1987 is set to 100). The annual estimates are measures of June-to-June price change. Tables 1b, 2b, and 3b give differences between these monthly-chained indexes and comparable annually-chained indexes. In each case, an upward chain drift in the Laspeyres index is expected based on previous research (see, for example, Szulc 1983); we only compare the magnitude of the drift for the alternative weight estimation schemes.

For Tables 1a and 1b, the CEX weights are "pure" monthly weights, i.e., computed using expenditure data for each specific reference month. As Table 1b indicates, chaining monthly with these weights creates an upward drift for the Laspeyres index and a downward drift for the geometric mean. The Törnqvist shows no discernible drift in either direction. Further investigation of the correlations between weights and price relatives reveals that, in this case,

$$c(\ln \mathbf{r}_{j-1,j}, \mathbf{w}_{j-1}) < c(\ln \mathbf{r}_{j-1,j}, \mathbf{w}_{t_0}) < 0,$$

accounting for the downward drift in the geometric mean index. Since this drift increases the differences between the geometric mean index values and those of the "superlative" Törnqvist index, the downward chain drift is undesirable. We note that it is a consequence of a relatively "tight" connection between the expenditure weights and the subindexes.

One way to weaken the negative correlation between the $\mathbf{r}_{j-1,j}$ and the $\mathbf{w}_{j-1}$ is to replace the pure monthly weights with moving averages of monthly expenditure shares. Tables 2a and 2b give values computed using 13-month moving averages of estimated expenditure shares rather than actual monthly weights. With these weights, the downward chain drift for the geometric mean index disappears, while the upward drift for the Laspeyres index is exacerbated. Interestingly, the Törnqvist index again shows no consistent drift. The simple moving average effectively removes the correlation between the weights and subindexes.

A positive correlation between weights and subindexes, however, is favorable for the chained geometric mean, since it brings it closer to the superlative Törnqvist. For Tables 3a and 3b, we smoothed the monthly expenditure shares using a *weighted* moving average rather than a simple moving average. The weights for the various months in the moving average were determined by a discretized Epanichnekov density (described in the Appendix) with $a = b = 6$ and $d = 1$. This weighting method assigns more weight to the "target" month, i.e., the month in the center of the moving average; thus it is less drastic in its smoothing than the simple moving average. The empirical results show that the positive correlation between the weights and the subindexes creates a slight upward drift in the monthly chained geometric mean; indeed the values in the second and third columns of Table 3a are nearly identical. For the geometric mean indexes, the Epanichnekov weights appear to assign sufficient weight to the data from the target month, while effectively neutralizing the negative correlation between weights and subindexes—a good "representativity vs. transitivity" balance.

Table 1a
Monthly-chained Indexes, June 1987 = 100
Pure Monthly Weights

Year	$\widehat{L}_1^m$	$\widehat{G}_1^m$	$\widehat{T}_1^m$
88	104.45886	103.50361	103.72109
89	110.44233	108.38743	109.05211
90	116.98616	112.40391	113.39736
91	123.01300	116.79749	118.45881
92	127.81897	119.47612	121.46332
93	132.43455	122.60377	124.74208
94	136.62502	125.40653	127.70174
95	141.56401	128.79049	131.10151

Table 1b
Differences Between Monthly-chained and Annually-chained Indexes
Pure Monthly Weights

Year	$\widehat{L}_1^m - \widehat{L}_1^a$	$\widehat{G}_1^m - \widehat{G}_1^a$	$\widehat{T}_1^m - \widehat{T}_1^a$
88	0.74534	0.14887	0.19399
89	1.45315	0.10760	0.63569
90	3.21092	-0.35100	0.34308
91	4.12270	-0.64732	0.64133
92	5.36603	-1.14196	0.43952
93	6.32071	-1.31495	0.36622
94	7.05079	-1.62927	0.21879
95	7.90717	-1.93506	-0.01858

Table 2a
Monthly-chained Indexes, June 1987 = 100
13-Month Moving Average Weights

Year	$\widehat{L}_{13}^m$	$\widehat{G}_{13}^m$	$\widehat{T}_{13}^m$
88	104.65003	103.63963	103.65491
89	110.63106	108.47516	108.53214
90	117.85420	113.05507	113.12406
91	124.13384	117.74416	117.87067
92	129.57824	120.98451	121.11773
93	134.30384	124.16689	124.32016
94	138.70588	127.07846	127.22922
95	143.88275	130.64277	130.78537

Table 2b
Differences Between Monthly-chained and Annually-chained Indexes
13-Month Moving Average Weights

Year	$\widehat{L}^m_{13} - \widehat{L}^a_{13}$	$\widehat{G}^m_{13} - \widehat{G}^a_{13}$	$\widehat{T}^m_{13} - \widehat{T}^a_{13}$
88	0.67162	-0.04718	-0.02453
89	1.36831	-0.14705	-0.10010
90	3.79342	-0.03352	-0.08025
91	5.00448	-0.02584	-0.06499
92	6.79891	-0.03892	-0.08445
93	7.91178	-0.09041	-0.12274
94	9.03963	-0.07608	-0.17206
95	10.25818	-0.07802	-0.21033

Table 3a
Monthly-chained Indexes, June 1987 = 100
13-Month Weighted Moving Average Weights

Year	$\widehat{L}^m_w$	$\widehat{G}^m_w$	$\widehat{T}^m_w$
88	104.64159	103.63711	103.63201
89	110.62377	108.47530	108.47751
90	117.89582	113.15578	113.14536
91	124.20280	117.87946	117.87453
92	129.59878	121.15491	121.14383
93	134.31869	124.34287	124.33107
94	138.75436	127.27609	127.25665
95	143.96570	130.87415	130.83972

Table 3b
Differences Between Monthly-chained and Annually-chained Indexes
13-Month Weighted Moving Average Weights

Year	$\widehat{L}^m_w - \widehat{L}^a_w$	$\widehat{G}^m_w - \widehat{G}^a_w$	$\widehat{T}^m_w - \widehat{T}^a_w$
88	0.64036	-0.07478	-0.06374
89	1.34350	-0.16691	-0.17680
90	3.82313	0.05559	-0.08813
91	5.05632	0.09203	-0.10766
92	6.80863	0.12654	-0.10542
93	7.90868	0.07726	-0.14103
94	9.11015	0.16219	-0.13911
95	10.38366	0.21524	-0.13007

7. Conclusions

Under the assumption that $\eta < 1$ (as in the data we analyzed), we may draw several conclusions from the results above. First, the Törvquist index appears remarkably robust to chain drift, while the Laspeyres index is prone to severe upward drift. The case of the geometric mean index is more complex. It suffers a downward chain drift when the weights and subndexes are negatively correlated (as in Tables 1a and 1b). When the correlation is effectively removed, there is no noticeable drift, yet the geometric mean is downwardly biased relative to the Törnqvist (as in Tables 2a and 2b). In this case, we have near transitivity with a lack of representativity. The ideal weights for the geometric mean index are positively correlated with the subindexes (as in Tables 3a and 3b). The resulting upward drift neutralizes the downward bias relative to the Törnqvist; that is, it increases the representativity of the geometric mean indexes.

Acknowledgment: The author gratefully acknowledges the helpful comments and suggestions of Alan Dorfman on the research presented in this paper.

References

Diewert, W. E. 1987. "Index numbers." In *The New Palgrave: A Dictionary of Economics,* (edited by J. Eatwell, M. Milate, and P. Newman), MacMillan, London.

Divisia, F. 1925. L'Indice Monétaire et la Théorie de la Monnaie. *Rev. Econ. Politique,* Paris, July-August, Sept.-Oct., Nov.-Dec., Jan. -Feb. 1926, Paris. (Reference obtained from Forsyth and Fowler 1981.)

Dorfman, A., Leaver, S., and Lent, J. (1999). "Some Observations on Price Index Estimators," *Proceedings of the Federal Committee on Statistical Methodology Research Conference, Monday B Sessions,* pp. 56-65.

Forsyth, F. 1978. "The Practical Construction of a Chain Price Index Number," *The Journal of the Royal Statistical Society, Series A,* Vol. 141, Part 3, pp. 348-358.

Forsyth, F. and Fowler, R. 1981. "The Theory and Practice of Chain Price Index Numbers," *The Journal of the Royal Statistical Society, Series A,* Vol. 144, Part 2, pp. 224-246.

Reinsdorf, M. B. 1998. "Formula Bias and Within-Stratum Substitution Bias in the U.S. CPI. *The Review of Economics and Statistics, Vol. LXXX,* Number 2 (May 1998), pp. 175-187.

Richter, M. K. 1966. "Invariance Axioms and Economic Indexes," *Econometrica,* Vol. 34, No. 4 (October, 1966), pp. 739-755.

Shapiro, M. and Wilcox, D. 1997 "Alternative Strategies for Aggregating Prices in the CPI," *Federal Reserve Bank of St. Louis Review,* May/Jun 1997, pp. 113- 125.

Szulc, Bohdan, J. 1983. "Linking Price Index Numbers," *Price Level Measurement: Proceedings from a conference sponsored by Statistics Canada,* (W. Erwin Diewert and Claude Montmarquette, eds.). Ottawa, pp. 537-566.

Appendix: Computing Smoothed Weights

The moving average weights used in estimating expenditure shares for the indexes in Tables 3a and 3b are based on the "discretized Epanichnekov density," which is given by

$$f_1(a,b;x) = \frac{3\left[(a+1)^2 - x^2\right]}{(2b+1)\left[3(a+1)^2 - b(b+1)\right]},$$

$x = 0, \pm 1, ..., \pm b$, where a and b are integers, and $b \leq a$. When $a = b$, we have

$$f_1(b;x) = \frac{3\left[(b+1)^2 - x^2\right]}{(b+1)(2b+1)(2b+3)},$$

$x = 0, \pm 1, ..., \pm b$. (In this case, we suppress the subscript a in our notation.) Let $X_{a,b,1}$ be a random variable following a discretized Epanichnekov distribution. Since the densities above are symmetric about 0, $E\left[X_{a,b,1}\right] = 0$, and thus

$$\begin{aligned} \sigma^2\left[X_{a,b}\right] &= E\left[X^2_{a,b}\right] \\ &= \frac{b(b+1)\left[(a+1)^2 - \left(3b^2+3b-1\right)/5\right]}{3(a+1)^2 - b(b+1)}. \end{aligned}$$

When $a = b$,

$$\sigma^2\left[X_{b,1}\right] = \frac{b(b+2)}{5}.$$

Setting a equal to a value strictly greater than b results in a higher variance (i.e., a "flatter" curve).

A more general form of the density is given by

$$f_d(b;x) = c(a,b,d)\left[1 - \left(\frac{x}{a+1}\right)^2\right]^d,$$

$x = 0, \pm 1, ..., \pm b$, where $d \geq 0$, and $c(a,b,d)$ is a normalizing constant that must be computed for the chosen parameters.