Estimating Dynamic Price Indexes

Alan H. Dorfman and Mary Kokoski
U.S. Bureau of Labor Statistics
2 Massachusetts Avenue, NE, Washington, DC 20212
(Dorfman.Alan@bls.gov)

Opinions are the authors' and do not reflect or imply policy of the US Bureau of Labor Statistics

Abstract
Price indexes are summary statistics meant to convey a comparison of prices at one time (or place) to another. The raw ingredients from which a standard price index is constructed are the ratios of the price of an item at one time period to the price of the same item at another. Goods which disappear after the initial time period, or first appear in the second, thwart constructing these ratios. We investigate empirically approaches to handling this problem, in particular, the use of time dummy hedonic indexes.

Key Words: hedonic price indexes, multiple regression, survey sampling, Törnqvist Index

1. Introduction

Comparisons of prices at different times or places—the construction of price indexes—is complicated by several factors, chiefly the heterogeneity of the goods sold and the shifting amounts expended on them. These problems are handled, at least in principle, by the construction of what are called *superlative* price indexes, which have many ideal properties from the standpoint of economic theory. An important example of a superlative index is the *Törnqvist Index*.

Another level of difficulty ensues if the physical nature of the goods changes from one time or place to another, or if they disappear entirely, or if new goods appear for the first time. One approach to handling these complications is *hedonic regression.* It turns out that a particular form of a hedonic regression index—that based on a time dummy regression model with specific weights—is a generalization of the Törnqvist Index. The question naturally arises whether this generalization, termed the *Dynamic Törnqvist*, shares the superlative properties of the original Törnqvist. We investigate this and related questions empirically using a large population of detailed scanner data.

2. Background on Price Indexes

Suppose that consumer/merchant transactions yield a population of heterogeneous items i sold with prices p_i^y at time y, with q_i^y the corresponding quantity sold. If we want to know how the prices of these items are changing over time, it is natural to consider the price ratios $\frac{p_i^{y+1}}{p_i^y}$ in order to compare the two periods y and $y+1$. We would like to combine these price ratios into an overall *price index.*

There exist many formulas for doing this, with competing considerations of meaningfulness and practicality playing a role in formula selection. Here, for brevity, we will focus mainly on (weighted) geometric means, that is, indexes of the form

$$I^{y,y+1} = \prod_i \left(\frac{p_i^{y+1}}{p_i^y}\right)^{w_i} .$$

What should the weights w_i be? One possibility is to simply let them be constant: $w_i = 1/N$, where N is the number of items i in the population. More sophisticated indexes are based on expenditure shares: Let $e_i^y = p_i^y q_i^y$ be the *expenditures* (for period y) and $s_i^y = \frac{e_i^y}{\sum_i e_i^y}$ the corresponding *expenditure shares.*

Then we might consider Single Period Expenditure Weighting : $G^{y,y+1} = \prod_i \left(\frac{p_i^{y+1}}{p_i^y}\right)^{s_i^y}$ or Dual (period) Expenditure Weighting :

$$T^{y,y+1} = \prod_i \left(\frac{p_i^{y+1}}{p_i^y}\right)^{\left(s_i^y + s_i^{y+1}\right)/2} .$$

As a rule the first, $G^{y,y+1}$, is more practicable from the standpoint of timely data processing. The second, $T^{y,y+1}$, is referred to as the Törnqvist Index, after the economist who originally suggested it, and has desirable properties entitling it to enter the category of

Superlative Price Indexes (Diewert 1976). Thus the Törnqvist Index can be taken as a gold standard of price indexes. (There are others that also qualify. The superlative indexes typically differ from each other only minutely in the numerical results they yield.)

It is generally regarded as impossible to collect data on *all* the items in the population of transactions. With some exceptions, what governments call *indexes* are really *index estimators* based on a *sample*. For example, if we select sample *s* by probability proportional to size sampling, with size measure w_i,

then the *index estimator* $\hat{I}^{y,y+1} = \prod_{i \in s} \left(\frac{p_i^{y+1}}{p_i^y} \right)^{1/n}$ (n the sample size)
is approximately design unbiased for

$$I^{y,y+1} = \prod_i \left(\frac{p_i^{y+1}}{p_i^y} \right)^{w_i} .$$

In what follows we shall be interested in both indexes and index estimators.

3. **The Problem of non-matched items**

The above definitions of price index rely on being able to match up transactions from one period to the other. This assumes each item is sold in both periods. Suppose there are items i which are available at time y (for which we get prices p_i^y) but which then *disappear*, that is, the items are not available and therefore have no prices at the succeeding time $y + 1$. Or suppose there are *new* items for which there are prices p_i^{y+1} at time $y + 1$, but no prices at time y. In these cases, it is impossible to form the price ratios $\frac{p_i^{y+1}}{p_i^y}$ on which the indexes G, T are based. (See figure 1)

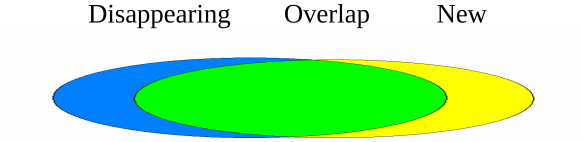

Figure 1

How then to define a price index?

One approach, not discussed in this paper, but common where there are but few missing items, is the imputation of prices, using, for example, hedonic regression in a manner different from, although connected to, that described below. For example, Liegey (1994), describes the use of hedonic regression for quality adjustment in the U.S. CPI.

There are at least three other possibilities:
1. *Matched goods approach.* We construct the index based on the Overlap, the set of *continuing* items, for which the price ratios can be calculated. There are two problems with this:
(a) these goods may be intrinsically different from the disappearing and new goods, so that the resulting index is somehow a distortion;
(b) there may be *no* overlap. This is unlikely in the ordinary *time-based* indexes which we have emphasized so far. But there is often considerable interest in constructing *across area* indexes, indexes that compare prices in one place to those in another. Price economists often regard as distinct items sold in different outlets. (Cheerios sold in a convenience store may not be in all respects the same as "the same box" of Cheerios sold in a superstore. If so, then items in the two areas are by definition distinct.) We are then stuck with a situation such as in Figure 2.

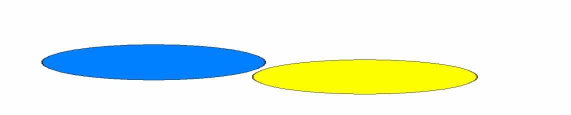

Figure 2.

2. *Unit value approach.* Group similar items, get average price ("unit value") for each group, use ratio of average prices as raw ingredients for indexes. Thus instead of price relatives of items, we build the index from price relatives of an item-group:

$$\frac{\bar{p}_g^{y+1}}{\bar{p}_g^y} = \frac{\sum_{i' \in g} p_{i'}^{y+1} q_{i'}^{y+1} \Big/ \sum_{i' \in g} q_{i'}^{y+1}}{\sum_{i'' \in g} p_{i''}^{y+1} q_{i''}^{y+1} \Big/ \sum_{i'' \in g} q_{i''}^{y+1}} .$$

The items i' in the numerator need not be the same as the items i'' in the denominator. The $\bar{p}_g^y$ are the "unit values of item-group g". Then the index is given by

$$G_{uvb}^{y,y+1} = \prod_g \left(\frac{\bar{p}_g^{y+1}}{\bar{p}_g^y} \right)^{w_g} .$$

The problems are: (a) we may be averaging prices of dissimilar items for which could expect price trends to differ, and (b) the formula for the unit value requires explicit measures of quantity. From a practical standpoint, getting quantity measurements is often difficult and it turns out to be easier in practice to get, or at least estimate, expenditures.)

3. *Hedonic regression indexes.* This is a major departure from the above approaches. Items are regarded as composed of priceable attributes and item prices can, therefore, be disaggregated into component prices of item attributes. These implicit prices cannot be directly observed, and thus are estimated by statistical regression techniques. It is assumed that the attributes have a kind of permanence across time (or space) that the items themselves may lack. Economic theory does not prescribe the form of the statistical regression used to derive the implicit prices of characteristics, and a variety of approaches exist (Triplett 1987). One promising approach is to use the *time dummy regression model*

$$\log(p_i^{y^*}) = \alpha + \gamma\, I\left(y^*\right) + \sum_j \beta^{(j)} x_{ij} + e_i\ , \quad (1)$$

where the x_{ij} are quantitative or categorical variables reflecting properties of the item sold

The coefficients $\beta^{(j)}$ are assumed constant across time

$I(y^*)$ is a "time-indicator", $I(y^*) = \begin{cases} 1, & y^* = y+1 \\ 0, & y^* = y \end{cases}$.

Of major interest is the parameter γ , which reflects changes across time.

A *hedonic price index* is then given by

$$H^{y,y+1} = \exp(\hat{\gamma}),$$

where $\hat{\gamma}$ is obtained from fitting the regression model (1).. This appears very different from the formulas for *G, T.* The following result, described by De Haan, and which we put in the form of a theorem, is therefore rather surprising.

De Haan's Theorem (2003) In the case of *full overlap* (no new, no disappearing items), if the hedonic regression model (1) is fitted using weights w_i, then

$$H^{y,y+1} = I^{y,y+1} = \prod_i \left(\frac{p_i^{y+1}}{p_i^{y}} \right)^{w_i} .$$

In particular, with the "superlative weights"

$$w_i = \left(s_i^{y} + s_i^{y+1}\right)/\,2\,, \qquad (1)$$

we have $H^{y,y+1} = T^{y,y+1}$, and the hedonic index is a superlative index *ipso facto*. An earlier constant weighted version of this theorem is found in Triplett (2001).

What is especially astonishing is that the equality holds regardless of the details of the (time dummy) regression model employed! The model could have many or few parameters, so that quantities like R^2, the variance of $\hat{\gamma}$, could differ across models fitted, but the index value would be unchanged.

Now we can (in principle) use the superlative weights where there is partial overlap (or even *no* overlap), letting $s_i^{y+1} = 0$ for disappearing goods, and $s_i^{y} = 0$ for new goods. The resulting hedonic index has been termed a *Dynamic Törnqvist Index* (Dalen, 2001). There is this difference from the full overlap case: For such "incomplete" populations, the model matters. The index can vary depending on the attributes represented in the model.

Should the resulting hedonic index still be regarded as superlative? That is, is a *dynamic* Törnqvist superlative the way the standard Törnqvist is? Superlative indexes are defined in the context of matching items. In the present context of non-matching items, we take an indirect pragmatic approach, classifying dynamic indexes as "quasi-superlative" when they meet a minimal condition, described in Section 4.

If we take a probability approach to sampling and sample $pps(w_i)$, and construct a hedonic index through unweighted regression (OLS), the result will be an index estimating the corresponding population dynamic Törnqvist.

4. The Dynamic Törnqvist

Since in the case of a full overlap population, the Törnqvist is known to be a superlative index, we begin to get a handle on the behavior of the dynamic Törnqvist, i.e. the hedonic index with weights (1), by considering the following scenario:

Suppose a population *P* of *N* items where prices and expenditures, both time periods, are known for all items. Thus the traditional Törnqvist (or equivalently a dummy model hedonic index with weights (1)) – the gold standard – is calculable. Note that for this population there would be $Q = 2N$ prices, one each period, for each of the *N* items. Now, for some fixed $M < Q$, take at random different versions of sub-populations (from the point of view of the large population these will actually be very large random samples), some with complete overlap, some with partial or no overlap, all having *M* prices. Thus a complete overlap sub-population will contain *M/*2 items, the complete non-overlap *M* items, and the partial overlap something in between. From the point of view of prices contributing to index calculation, all will contain the same amount of data.

If a given index (in particular, a Dynamic Törnqvist, that is, a hedonic index with specified regression model) using a given amount of data (*M* prices) in the incomplete overlap case, approximates the full *N* population Törnqvist as well as the full overlap Törnqvist which also relies on *M* prices, then we call that index *quasi-superlative*. We then ask: is the Dynamic Törnqvist quasi-superlative?

This leads to the following questions, which we have investigated empirically:

How do the dynamic Törnqvist indexes for the sub-populations compare to the full population Törnqvist, under various conditions of overlap?

How much effect on hedonic indexes do different models have in the cases of incomplete overlap?

How close to full population Törnqvist are the hedonic indexes compared to matched model or unit value approaches?

How do various versions of sample-based index estimators do, when we sample according to some probability scheme from the sub-populations.?

5. A Simulation Study

We performed a simulation study using Cereal Scanner Data, for which 6 time periods of prices and quantities were available, that is five pairs of periods. For a full description, see Dorfman, et al.(2006). For each item, there are a number of attributes, which can be translated into variables in a regression model:

A. *base information*:

(a) area where it was purchased,
(b) the supermarket chain,
(c) the manufacturer

B. *strata information* : there were ten *strata* of cereals. These were reflective of long term price trends, and grouped into broad natural divisions corresponding to type of cereal–hot, cold, fruity, or sweet.

This leads to the three regression models being considered in our study: (i) just the basic information in A – notated in tabulations as "basic", (ii) just the stratification on cereal types in B – "strat", and (iii) a full model using the regressors in both A and B – "basic+strat".

There were different amounts of price data actually available for each of the five (matched) pair of time periods. To keep things uniform, for each pair of time periods, we randomly selected 200 "full populations" having *N* =12000 items, hence *2N =24000* price measurements. From each of these, three sub-populations were randomly selected, each having the same number (12000) of price (and quantity) measurements, namely: (a) the "half population" having *M/*2 = 6000 items with complete matching, (b) the "overlap population" with 8000 items, of which 4000 overlap, 2000 are new, and 2000 disappear, and (c) the "separate population", with *M* = 12000 items, 6000 old, 6000 new (this situation would parallel the usual spatial index situation; see above.)

Additionally, from these we took samples of size n = 200, sampling *pps*(w_i), with $w_i = \left(s_i^y + s_i^{y+1}\right)/2$ on any overlap, $w_i = s_i^y / 2$ on any disappearing goods, and $w_i = s_i^{y+1} / 2$ on new goods. Of primary interest is measuring the extent to which indexes and index estimators differ from the gold standard Törnqvist. Our main measure of this difference was the empirical root mean square error (*rmse*)

$$r = \sqrt{\frac{\sum_k \left(I_{k,subpop}^{y,y+1} - T_{k,full}^{y,y+1}\right)^2}{200}}$$

. We also considered the empirical bias

$$b = \frac{\sum_k \left(I_{k,subpop}^{y,y+1} - T_{k,full}^{y,y+1}\right)}{200}.$$

6. Results

Results (multiplied by a thousand) at the population level are given in Tables 1 – 2. For each pair of years, within each sub-population, the minimum *rmse* is noted in red and likewise for the (absolute value of) the bias. Instances where an index for the overlap or separate sub-populations does better than the half-population index are italicized. A similar notation is used in Tables 3 and 4 for sample estimates; in this case the comparison is to the half-population geomean.

It is clear that the largest differences in *rmse* depend on the *amount of overlap* in populations: the greater the overlap, the smaller the difference between the sub-population indexes and the full population Törnqvist. In particular, none of the overlap indexes has smaller *rmse* than the half-population Törnqvist, and the separate population indexes do even worse.

Differences among the within sub-population indexes are not nearly so pronounced. There are no clear winners among competing models/indexes. Usually, but not in every instance, a hedonic index does better than the index using just matched items. The full basic + strat model seems to have a slight edge over less full regression models The unit value based (uvb) is not usually as good as the other indexes.

Results are a bit less sharp for biases (Table 2). We regard the *rmse* as the more important indicator.

The *rmse* situation under sampling (Table 3) is not so sharp as at the population level. In the year-pairs 95-96 and 98-99, the hedonic indexes for samples from the overlap population actually do better than the geomean estimator in samples from the half population. Sometimes, in the overlap population, the matched model index is appreciably better than the hedonic indexes, sometimes the reverse.

In all instances the samples from the separate populations yield worse *rmse* than those from the half-population. In the separate population, there are instances where fairly large differences arise between hedonic indexes based on different models, but these differences fade under sampling. Also, things are less clear cut with respect to bias.

7. Discussion

The differences in *rmse* between, on the one hand, the Dynamic Törnqvist Indexes, that is, the Hedonic Indexes, in the Overlap and Separate Populations, and, on the other, the standard Törnqvist in the Half Population, inclines us to think of these as quite distinct. If, in this rather simple situation, they fall short, then the hedonic indexes under sizable non-overlap do not qualify as quasi-superlative. Since the name “Törnqvist” is is associated with “superlative”, we are inclined to think that the name “Dynamic Törnqvist” is a misnomer. They might be the best we can do, given the data, but the main point might be to seek to improve the data. For example, some careful matching of items in across-area indexes (the separate case) might be the best support of a sound index.

The sampling that was done in this study was rather simple-minded: the segments “matching”, “disappearing”, and “new” were taken as given and within each sampling was done according to expenditure share. Expanding this work to determine the consequences of a more realistic sampling scheme would be desirable.

References

Dalen, J. (2001) “Statistical targets for Price Indexes in Dynamic Universes”, International Working Group on Price Indexes, Sixth Meeting

de Haan, J. (2003) “Direct and Indirect Time Dummy Approaches to Hedonic Price Measurement”, International Working Group on Price Indexes, Seventh Meeting.

Diewert, W.E. (1976) “Exact and Superlative Index Numbers”, *Journal of Econometrics* (May 1976), 115-145

Dorfman, A.H., Lent, J., Leaver, S.G. and Wegman, E. (2006), “On Sample Survey Designs for Consumer Price Indexes”, *Survey Methodology*, to appear.

Liegey, P. (1994) “Apparel Price indexes: Effects of Hedonic Adjustment”, *Monthly Labor Review*, (May, 1994), 38-45.

Triplett, J. E. (1987), “Hedonic functions and hedonic indexes,” in Eatwell,J., M. Milgate, and P. Newman (eds.) *New Palgrave Dictionary of Economics*, vol. 2, 630-634.

Triplett, J.E. (2001), Handbook on Quality Adjustment of Price Indexes for Information and Communication Technology Products (draft), OECD, Paris.

.

Table 1. *rmse*, three sub-populations

	95-96	96-97	97-98	98-99	99-00
			Half		
	4.26	4.22	4.55	6.52	7.60
			Overlap		
matched model	6.20	5.76	6.77	9.09	10.53
hedonic basic	5.83	8.10	5.65	8.04	9.05
" basic+strat	5.60	7.93	5.42	7.89	9.52
hedonic strat	6.36	7.92	5.80	8.00	10.09
			Separate		
hedonic basic	16.95	29.73	11.35	12.02	14.89
" basic+strat	15.05	29.03	10.15	11.39	15.93
hedonic strat	23.31	28.33	11.14	13.13	22.80

Table 2. *bias*, three sub-populations

	95-96	96-97	97-98	98-99	99-00
			Half		
	0.69	0.15	-0.14	-0.40	1.13
			Overlap		
matched model	1.14	*-0.12*	*0.03*	*0.27*	*0.92*
hedonic basic	-2.64	-5.90	0.92	-1.19	1.20
" basic+strat	-2.12	-5.73	*-0.12*	*-0.05*	3.02
hedonic strat	-3.64	-5.62	-0.44	0.59	4.39
			Separate		
hedonic basic	-14.81	-27.90	3.78	-3.17	-3.65
" basic+strat	-12.66	-27.23	-0.86	3.87	7.87
hedonic strat	-21.71	-26.51	-2.81	4.86	17.27

Table 3. *rmse*, samples from 3 sub-populations

	95-96	96-97	97-98	98-99	99-00
			half		
geomean	24.43	14.86	12.09	47.95	17.80
uvb	31.00	17.28	12.19	46.63	20.14
			overlap		
matched model	27.77	16.27	14.50	50.63	21.40
uvb	32.39	22.44	26.53	50.52	40.39
hedonic basic	*21.80*	23.06	25.50	*42.18*	38.90
"basic+strat	*21.63*	22.28	24.50	*41.37*	36.92
hedonic strat	*22.35*	22.36	24.53	*43.21*	36.93
			separate		
uvb	40.19	41.85	51.89	69.50	46.53
hedonic basic	42.74	51.99	54.24	62.08	55.84
"basic+strat	39.48	48.25	50.40	61.19	46.78
hedonic strat	37.27	43.97	47.19	62.94	49.10

Table 4. *bias*, samples from 3 sub-populations

	95-96	96-97	97-98	98-99	99-00
			half		
geomean	21.45	7.63	-1.28	46.38	12.08
uvb	28.77	12.57	4.91	45.15	16.31
			overlap		
matched model	23.60	8.24	*0.33*	47.86	*10.18*
uvb	*19.70*	*0.26*	11.25	43.80	28.80
hedonic basic	*-0.96*	*-6.17*	9.46	29.91	24.40
“basic+strat	*-0.69*	*-5.38*	8.04	29.29	23.22
hedonic strat	*-2.57*	*-5.54*	5.71	29.59	22.78
			separate		
uvb	*13.81*	-10.36	22.85	51.04	*8.34*
hedonic basic	*7.15*	-21.93	24.15	40.23	*-10.89*
“basic+strats	*6.86*	-20.54	22.14	41.85	*-4.93*
Hedonic strat	*-3.09*	-17.13	15.11	41.76	*0.57*