Tracking Consumer Energy Price Change: An Overview of Federal Data Sources and Methodologies

Janice Lent[1], Joseph Ayoub[2]
[1]Energy Information Administration, EI-70, 1000 Independence Ave., S.W., Washington, DC 20585
[2]Energy Information Administration, EI-64, 1000 Independence Ave., S.W., Washington, DC 20585

Abstract
The Energy Information Administration (EIA) is researching estimation methods with the goal of developing an experimental Energy Consumer Price Index (ECPI), based almost entirely on EIA data. For some major energy sources, EIA collects universe or large-sample price and sales data, which can be used to compute price indexes with very low sampling error. Also, EIA's model-based projections of future energy prices and consumption levels can be used to develop CPI forecasts for some energy components. Because the experimental indexes are being computed in a research environment rather than in a large-scale production environment, the process of incorporating data from new energy surveys will be streamlined. This paper provides background information and preliminary results of EIA's price index estimation research.

Key Words: price index, Fisher, Törnqvist

1. Data Sources for Consumer Energy Prices and Consumption

1.1 Introduction
Price index estimation requires data on both prices paid and quantities of goods and services purchased by consumers. With regard to energy prices, a vast array of data sources is publicly available. Price data appropriate for use in a consumer price index, i.e., data on prices paid by ultimate consumers, including all applicable taxes and distribution costs, must be distinguished from other types of price data. Moreover, household energy consumption (e.g., number of gallons of gasoline purchased by households) is often difficult to estimate from the available data sources.

Numerous government and non-government entities collect and publish information on consumer energy prices and consumption. The American Automobile Association, for example, makes gasoline and on-highway diesel fuel prices available on its daily Fuel Gauge Report website. Here we focus primarily on estimates published by the Energy Information Administration (EIA) and the Bureau of Labor Statistics (BLS), the principal federal publishers of energy and price index data, respectively.

1.2 Energy Price and Expenditure Data from Non-EIA Sources
The BLS uses data from three sources for its Consumer Price Index (CPI) programs. The Consumer Expenditure Survey (CEX) and the Telephone Point of Purchase Survey (TPOPS) are household surveys conducted by the Census Bureau and sponsored by BLS. In addition, BLS field economists collect price data from a sample of retail outlets each month for use in the CPI. Data from all three surveys are used in estimating consumer expenditures and expenditure shares for individual items or item categories; the estimated shares are used as weights in the CPI.

The BLS all-items CPI (or CPI-U) is designed to reflect inflation experienced by all urban consumers in the U.S. The sample of retail outlets from which prices are collected for the CPI is constructed using data from the TPOPS. Although the CPI sampling frame covers the entire country, the outlets identified through the TPOPS are those from which metropolitan and non-metropolitan urban households purchase items. For certain categories of goods and services, however, the TPOPS sampling frame is not used. Electricity and piped utility natural gas are two of the "non-TPOPS" categories in the CPI. For these categories, the BLS regional offices collect price data from utility companies in sampled geographic areas. For further details on the sampling and estimation methods used for the BLS CPI, see the *BLS Handbook of Methods* (2005). The target index formulas used are described in the next subsection.

1.3 Background on Alternative Price Index Formulas

In general, a price index measures inflation/deflation or change in the purchasing power of a currency. In addition to broad inflation measures, price indexes for specific collections of goods and services are frequently computed, either as inputs to an aggregate index or as economic analysis tools in their own right. The economics literature contains a wide variety of price index formulas that may be accepted as estimation targets. The "textbook" Laspeyres formula, for example, is given by

$$L = \frac{\sum_{i=1}^{N} q_{i,1} p_{i,2}}{\sum_{i=1}^{N} q_{i,1} p_{i,1}}$$

where N is the number of items in the target population, and, for $t \in \{1, 2\}$, $p_{i,t}$ and $q_{i,t}$ denote the price and quantity purchased, respectively, of item i in time period t, $i = 1, 2, ..., N$. Note that the index represents a comparison between prices in two arbitrary but discrete time *periods* 1 and 2 (e.g., months, years). The classical index formulas also rely on the implicit assumption that the collection of N items remains the same for the two reference periods. We may also write

$$L = \sum_{i=1}^{N} w_{i,1} \left(\frac{p_{i,2}}{p_{i,1}} \right),$$

where $w_{i,t} = \frac{q_{i,t}\, p_{i,t}}{\sum_{i=1}^{N} q_{i,t}\, p_{i,t}}$, the expenditure share associated with item i in time $t \in \{1,2\}$.

The ratio $p_{i,2} / p_{i,1}$ is often called the *price relative* for item i. The Laspeyres index is an arithmetic mean of the price relatives, weighted by the first period expenditure shares. The Paasche index is similar to the Laspeyres, but it is based on quantity measures from the second period. We define it as

$$P = \frac{\sum_{i=1}^{N} q_{i,2}\, p_{i,2}}{\sum_{i=1}^{N} q_{i,2}\, p_{i,1}} = \frac{1}{\sum_{i=1}^{N} w_{i,2} \left(\frac{p_{i,2}}{p_{i,1}} \right)^{-1}},$$

the harmonic mean of the price relatives with expenditure share weights from the second period. In the economics literature, certain price index formulas have come to be known as "superlative indexes," because they approximate a true cost of living index under relatively weak assumptions regarding consumer buying behavior. (See Fisher 1922 and Diewert 1976.) The superlative index formulas include the Fisher index, defined as

$$F = \sqrt{LP},$$

and the Törnqvist index, defined as

$$T = \prod_{i=1}^{N} \left(\frac{p_{i,2}}{p_{i,1}} \right)^{w_i},$$

where

$$w_i = \frac{w_{i,1} + w_{i,2}}{2} .$$

Although the Fisher and Törnqvist formulas share some properties in common, they differ in their algebraic forms and in regard to sensitivity to extreme price values (see Lent 2004). In many practical applications, estimating either of the superlative indexes is impossible, because data on the second-period quantities $q_{i,2}$ and expenditure shares $w_{i,2}$ are unavailable at the time of index publication. In such cases, the Laspeyres index or the weighted Geometric Mean (or "geomean") index,

$$G = \prod_{i=1}^{N} \left(\frac{p_{i,2}}{p_{i,1}} \right)^{w_{i,1}} ,$$

may be used. The Laspeyres index approximates the Fisher index when quantities purchased remain fairly consistent over time, i.e., when consumers buy a "fixed market basket" of items. In the case $q_{i,1} = q_{i,2}$ for all *i*, we have $L = P = F$. Similarly, the geomean index approximates the Törnqvist when the expenditure shares remain similar over time. When $w_{i,1} = w_{i,2}$ for all *i*, we have $G = T$.

A simpler index formula is often used for aggregating prices within narrowly-defined categories of very similar items measured in the same units. For such a category *c* and for $t \in \{1,2\}$ let

$$q_{c,t} = \sum_{i \in c} q_{i,t} .$$

The unit value index for category *c* is defined as

$$u_c = \frac{\sum_{i \in c} q_{i,2}\, p_{i,2} / q_{c,2}}{\sum_{i \in c} p_{i,1} q_{i,1} / q_{c,1}} .$$

In words, the unit value index is the average price paid for an item in category *c* during time period 2 divided by the average price paid for an item in category *c* during time period 1. For more on the use of the unit value index in price index estimation, see Balk (1999).

Due to data availability and other considerations, government statistical agencies often compute national-level price indexes by implementing multiple stages of aggregation (see, for example, Dorfman et al. 1999 or Balk 2003). The formulas and estimation methods used for the different aggregation levels often differ. In the BLS CPI program, for example, the geomean index is used for aggregating price ratios within narrowly defined categories; for higher level aggregation, the Laspeyres and Törqvist formulas are used. In the research described below, we used the unit value index formula to compute regional-level price indexes for the various energy components. We then used *L*, *P*, *F*, *T*, and *G* to aggregate the unit value subindexes to the national level. (The unit value subindexes take the place of the price relatives in the formulas *L*, *P*, *F*, *T*, and *G*.) We reasoned that, if *L* or *G* provided a close approximation to *F* and *T*, they would be preferable, due to ease of computation.

2. Empirical Research Methods and Preliminary Results

2.1 Energy Price and Sales Data Available from EIA

The EIA Regional Short-term Energy Modeling (RSTEM) system produces estimates of residential-sector average prices by region, including applicable taxes and distribution costs, for the following energy products:

- Electricity
- Piped (utility) natural gas
- Heating oil
- Propane
- Gasoline, all grades
- Gasoline, regular grade
- On-highway diesel fuel

The regions for which average prices are computed differ by commodity: electricity, natural gas, and heating oil prices are published by Census region and division, while gasoline and on-highway diesel prices are published by Petroleum Allocation for Defence Districts (PADD's).[1]

Monthly residential *consumption* estimates are also available, from RSTEM or other EIA sources, for the following energy products:

- Electricity
- Piped (utility) natural gas
- Heating oil
- Renewable sources (in BTU's), including geothermal, biofuels and biomass, and solar

Household utility consumption estimates are obtained through EIA's surveys of residential energy providers. Household vehicle fuel consumption, however, is more difficult to estimate, because gas stations normally do not know the percentage of their gasoline that is purchased by household vs. commercial vehicles (e.g., taxis and family cars may be refueled at the same stations, their drivers often paying the same prices at the pump). Our method of estimating household vehicle fuel consumption from the available data is described below.

Conversely, renewable energy consumption estimates are available in BTU's, but the price per renewable-source BTU cannot be directly estimated from current data. EIA does collect data on prices of renewable energy equipment sold to customers in the residential sector, e.g., solar thermal collector systems (EIA-63A), photovoltaic modules and systems (EIA-63B), geothermal heat pumps (EIA-902), and alternative fueled vehicles (EIA-886). The problem of translating equipment prices into actual energy prices, however, is beyond the scope of the research described here.

2.2 Methods and Assumptions Used in Estimating Price Index Weights

As discussed above, estimated expenditures (prices multiplied by quantities purchased) for commodities purchased are used as weights in the commonly used price index formulas. Because of the data gaps discussed above, some assumptions were used in estimating levels of residential purchases of gasoline and of #2 distillate fuel oil. For each fuel type, we estimated adjustment factors that could be applied to the EIA estimates of fuel sales by prime suppliers[2] (including sales to non-residential customers) to reduce them to levels representative of household sales levels. In the case of gasoline, the most recent available household consumption estimates were from the 2001 National Household Transportation Survey (NHTS).[3] For each PADD, we calculated an adjustment factor based entirely on annual data from 2001. Because the NHTS consumption estimates were published by Census Division rather than by PADD, we computed PADD-level factors as weighted averages of Census Division-level factors. We used state-level annual (2001) gasoline sales estimates to weight the Census Division-level factors in the calculations. Specifically, for each PADD *P*, we computed a household sales adjustment factor for gasoline, $\hat{f}_{g,P}$, as follows:

---

[1] Definitions of these geographic areas are available at http://www.eia.doe.gov/glossary/glossary_c.htm and http://www.eia.doe.gov/glossary/glossary_p.htm.

[2] Prime suppliers are dealers who sell to other dealers, end-use customers, or both. They include producers, importers, and wholesalers. The prime supplier sales volumes are used here as a proxy for total sales. EIA also publishes estimates of "product supplied," which is often used as a proxy for total consumption.

[3] The NHTS motor fuel volume estimates were given in "gasoline equivalents," i.e., volumes of other fuels used had been converted into gasoline equivalents. Because no data were available on the volume proportions for gasoline, diesel, and other fuels, we treated the gasoline equivalents as gallons of gasoline in this research. The values of the adjustment factors computed are clearly higher than they would have been if estimates of gasoline consumption alone had been used. For more information on the NHTS, see http://nhts.ornl.gov/.

$$\hat{f}_{g,P} = \frac{\sum_{s \in P} \hat{g}_s \left(\frac{\hat{g}_{h,D,s \in D}}{\hat{g}_{D,s \in D}} \right)}{\sum_{s \in P} \hat{g}_s},$$

where $\hat{g}_s$ = estimated gallons of gasoline sold by prime suppliers in state *s*;

$\hat{g}_{h,D,s \in D}$ = estimated gallons of gasoline equivalents purchased by households in Census Division *D* containing state *s*;

$\hat{g}_{D,s \in D}$ = estimated gallons of gasoline equivalents sold by prime suppliers in Census Division *D* containing state *s*.

The estimates $\hat{g}_s$ and $\hat{g}_{D,s \in D}$ were obtained from EIA data, while $\hat{g}_{D,s \in D}$ was based on NHTS data. Total residential purchases for PADD *P* was then estimated as

$$\hat{g}_{h,P} = \hat{f}_{g,P} \hat{g}_P ,$$

where $\hat{g}_P =$ estimated gallons of gasoline sold by prime suppliers in PADD *P*, based on EIA data. Thus we assumed that the adjustment factor $\hat{f}_{g,P}$, calculated from 2001 annual data, could be applied to subsequent months and years—a strong assumption. The adjustment factors varied by PADD and generally ran between 0.75 and 0.95.

For #2 distillate fuel oil, EIA publishes PADD-level annual estimates of volumes purchased by households and monthly estimates of total volumes sold by prime suppliers. We used the annual estimates to compute annual household adjustment factors and then applied these to the monthly total sales estimates to compute monthly estimates of household purchases. Specifically, for each month *m* in year *y*, we estimated the household sales of #2 distillate fuel oil within each PADD *P* as follows:

$$\hat{d}_{P,m,h} = \hat{d}_{P,m} \left(\frac{\hat{d}_{P,y,h}}{\hat{d}_{P,y}} \right),$$

where $\hat{d}_{P,m,h}$ = gallons of #2 distillate fuel oil sold to households in PADD *P* during month *m* in year *y*;

$\hat{d}_{P,m}$ = total gallons of #2 distillate fuel oil sold by prime suppliers in PADD *P* during month *m* in year *y*;

$\hat{d}_{P,y,h}$ = gallons of #2 distillate fuel oil sold to households in PADD *P* during year *y*; and

$\hat{d}_{P,y}$ = total gallons of #2 distillate fuel oil sold by prime suppliers in PADD *P* during year *y*;

The implicit assumption is that the proportion of fuel oil sales accounted for my household consumption remains constant over the year. In fact, household fuel use tends to be more seasonal than commercial or industrial use. Estimating $\hat{d}_{P,m,h}$ by this method may therefore dampen the seasonality in the resulting index series.

2.3 Preliminary Results of Empirical Research

Figure 1 shows the electricity component computed with the five alternative index formulas based to January 1998. The index projections shown run through 2009 and are based on price and sales data from EIA's Regional Short-term Energy Models (RSTEM). The experimental Fisher and Törnqvist series run very close together; the Laspeyres index runs above them, while the Paasche and geomean run low. Figure 1 also shows the BLS electricity CPI, which runs fairly close to our experimental Fisher and Törnqvist series. Initially, the BLS CPI runs below our Fisher and Törnqvist, but it increases more quickly and eventually starts to run above our series. (Projections for the BLS CPI are not available.) For natural gas, we did find some fairly pronounced differences between the series computed by the different formulas and between the BLS CPI and our experimental indexes. The alternative series for natural gas are shown in Figure 2. The BLS natural gas CPI (not seasonally adjusted) clearly doesn't show the same seasonal pattern that is reflected in the EIA data. The differences between these series indicates that natural gas prices follow different patterns of change in different parts of the country. The weights we apply to the price ratios can therefore make a substantial difference in the pattern of price change indicated. Here again, however, we see the experimental Fisher and Törnqvist series running close together.

Motor fuel is the largest component for which we faced challenges due to data gaps. As described above, we relied on some assumptions in deriving the expenditure share weights. For this component, however, the weights had very little impact on the pattern of price change indicated by the series. Figure 3 shows all of the experimental EIA series running very closely together, regardless of the price index formula or the weights used. Moreover, the BLS CPI runs practically on top of our experimental series. This phenomenon indicates that the price movements during this ten-year period were very similar in different parts of the country. The price levels, of course, were different. We know, for example, that gas prices in the Midwest were considerably lower than prices in California. But, in terms of percentage change from month to month, the prices were moving in tandem, essentially responding to changes in global oil prices. In this type of economic situation, the use of different formulas and weights has little impact on the general movements of the series.

Figure 1. EIA Preliminary and BLS CPI Component Series for Electricity

January 1998 = 100

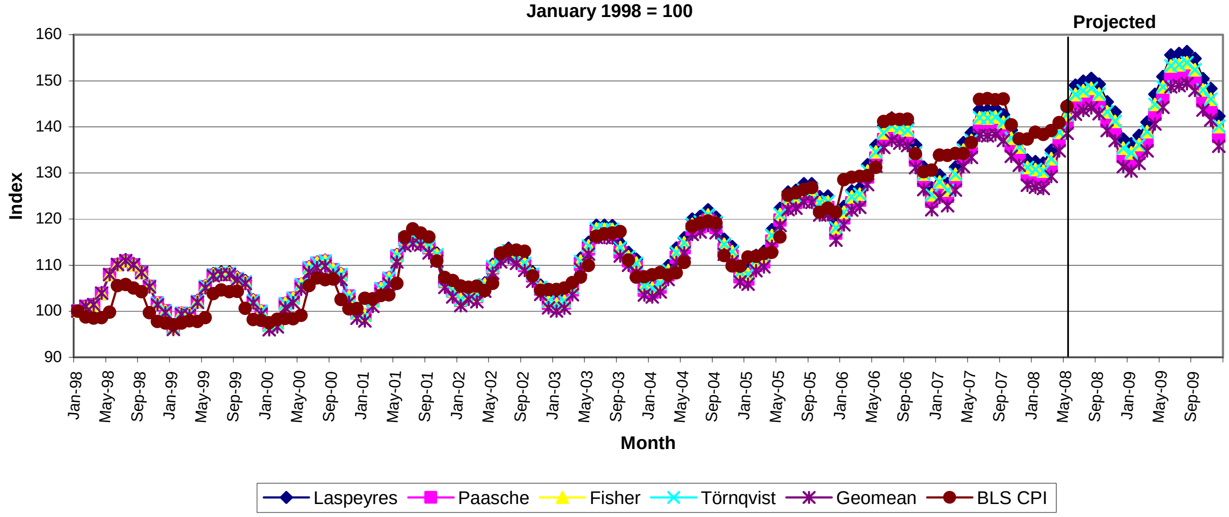


Figure 2. EIA Preliminary and BLS CPI Component Series for Natural Gas

January 1998 = 100

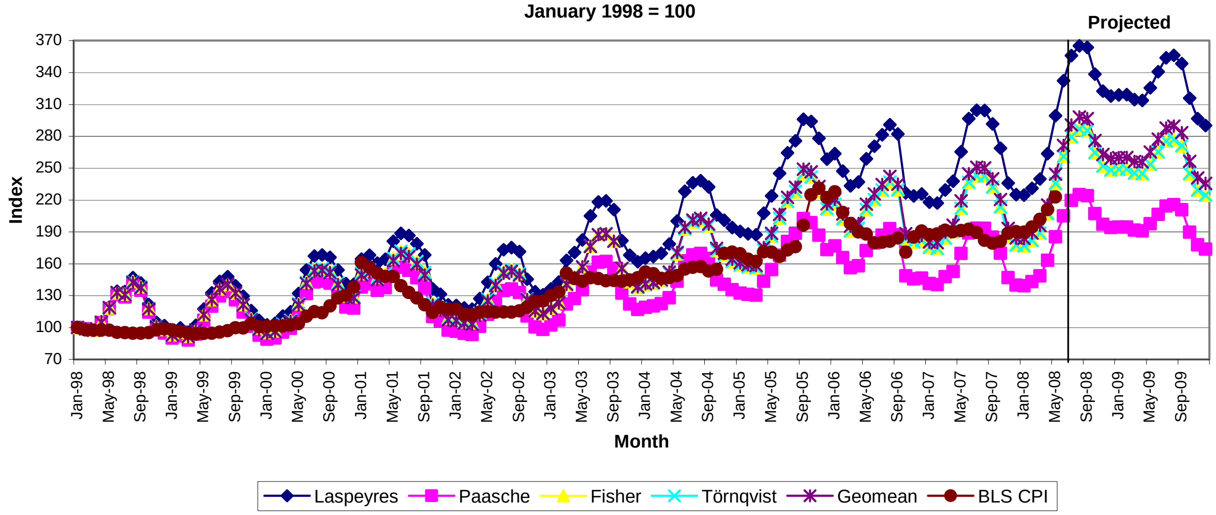


Figure 3. EIA Preliminary and BLS CPI Component Series for Motor Gasoline
January 1998 = 100

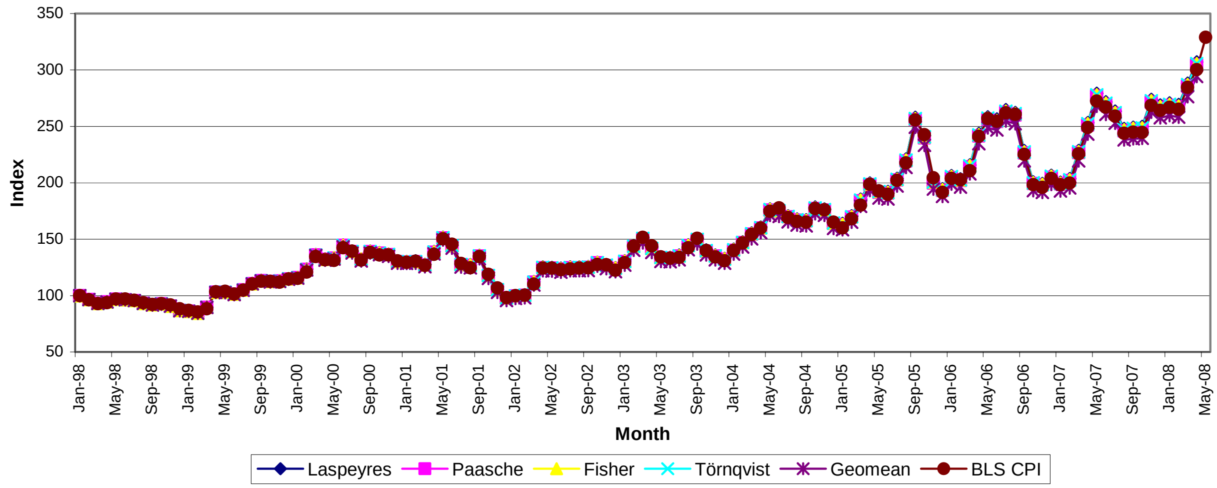
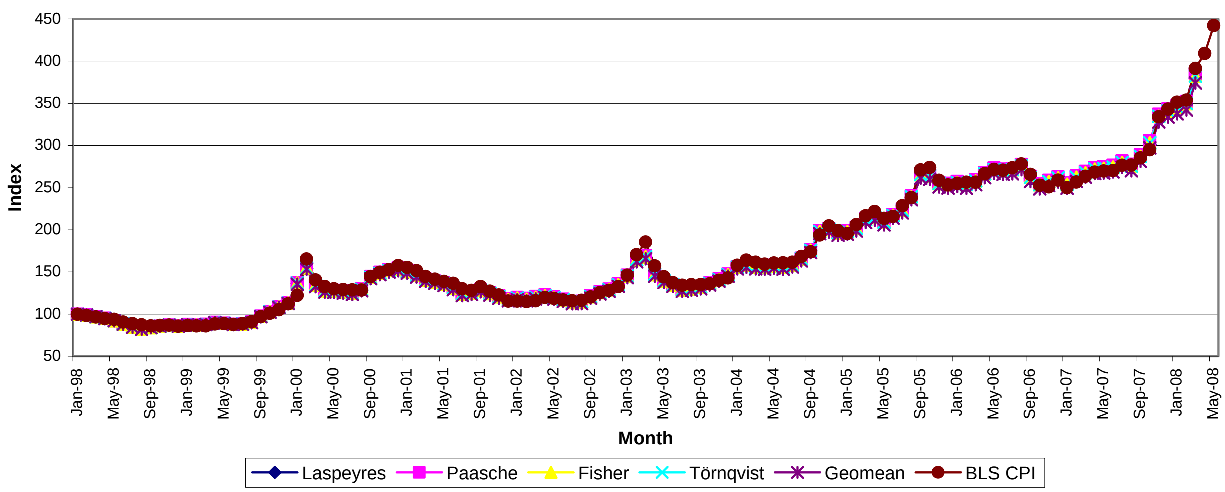


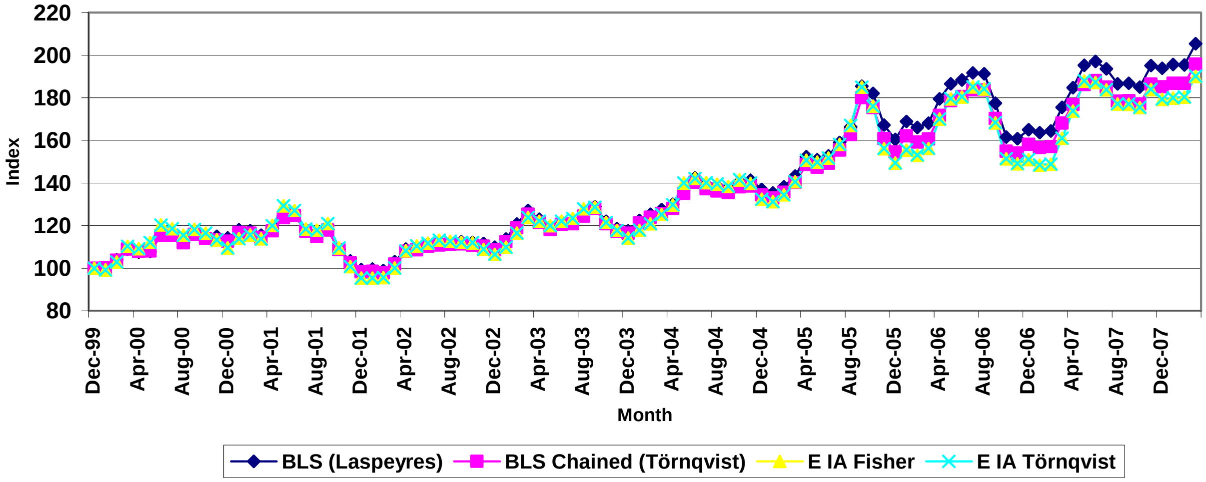
Figure 4 shows the experimental series for heating oil (#2 distillate). In some parts of the Northeast, this component accounts for a fairly large share of household energy expenditures, although this is not the case at the national level. Although the pattern of price change for distillate differed somewhat from the pattern for gasoline, the results for this component were similar in that all of the index series followed very similar paths. Prices across the country were essentially moving together. Also, for this category, the Northeast regions account for most of the expenditures, so price movements in the Northeast regions dominate the national index movements. For distillate, as for gasoline, the BLS CPI component was very similar to our experimental CPI's.

Figure 4. EIA Preliminary and BLS CPI Component Series for #2 Distillate Fuel Oil
January 1998 = 100



We aggregated the Fisher and Törnqvist indexes for the four dominant energy components to get the experimental CPI's shown in Figure 5. This figure shows the experimental EIA Fisher and Törnqvist series as well as the two energy CPI's that BLS publishes. For the all energy component, BLS publishes one index series (the CPI-U) based on the Laspeyres formula and another (the Chained CPI-U) based on the Törnqvist. Here we see the BLS Törnqvist series running slightly below the BLS Laspeyres for the later years of the series. The EIA Fisher and Törnqvist series run together, just slightly below the BLS Tornqvist. Although the reasons for the differences are not entirely clear, the BLS and EIA series are based on different data sources and different estimation methods. The fact that they track each other so closely was actually unexpected.

Figure 5. EIA Preliminary and BLS CPI Series for All Energy
December 1999=100



3. Conclusions and Plans for Future Research

In conclusion, we have a proof of concept for the idea of estimating an ECPI based on EIA data. We can also see that the use of the Fisher or the Tornqvist formula is necessary in this application. The other indexes (Laspeyres, Paasche, geomean.) ran too high or too low for some of the energy components we examined. In our future research, we'd like to enhance our experimental index series by adding some new components such as propane and possibly wood, kerosene, or coal, though these three energy sources account for very small expenditure shares in the residential sector. We'd also like to research additional forecasting possibilities, especially for the petroleum-based products. Because the weights had very little impact on the indexes, we might be able to project fixed the indexes forward using fixed weights.

In the future, we would like to include data on renewables in our index. Because most renewable energy is currently used for electric power generation, most household consumption of renewables in the U.S. is indirect, making prices and quantities for renewables difficult to estimate. Residential use of alternative fuels is still relatively limited in the U.S., and EIA does not yet have reliable data on the quantities of such fuels purchased by households. In the future, however, as the use of alternative motor fuels increases, better estimates of quantities and prices may become available. Similarly, direct household use of renewable energy sources may become more common in future decades. EIA's small-scale index estimation system should allow timely introduction of data from new sources in the index estimates.

References

Balk, B. (1999), "On the Use of Unit Values as Consumer Price Subindices," *Proceedings of the Fourth International Working Group on Price Indices*, BLS, Washington, D.C.

Balk, B. (2003), "Price Indexes for Elementary Aggregates: The Sampling Approach, *Proceedings of the Seventh Meeting of the International Working Group on Price Indices* (*Ottawa Group*), Paris.

BLS Handbook of Methods (2005), http://stats.bls.gov/bls/descriptions.htm

Diewert, W. E. (1976). "Exact and Superlative Index Numbers," *Journal of Econometrics,* 4, pp. 115-145.

Dorfman, A.H., Leaver, S.G., 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

Fisher, I. (1922). *The Making of Index Numbers.*

Lent, J. (2004). "Effects of Extreme Values on Price Indexes: The Case of the Air Travel Price Index," *Journal of Transportation and Statistics, Volume 7, Numbers 2/3,* pp. 41-52. U.S. Department of Transportation.

Disclaimer: Opinions expressed in this paper are those of the authors and do not constitute policy of the Energy Information Administration.