A transaction price index for air travel

Research on a price index estimator based on data from a U.S. Department of Transportation survey involves testing unique imputation and across-time matching procedures; the resulting experimental index is compared with the official CPI series and the consumer expenditure deflator series used in National Accounts computations

Janice Lent
and
Alan H. Dorfman

Janice Lent is a mathematical statistician with the U.S. Research and Innovative Technology Administration, Washington, DC; Alan H. Dorfman is a mathematical statistician in the Office of Survey Methods Research, Bureau of Labor Statistics. The opinions expressed in this article are those of the authors and do not constitute policy of either the Research and Innovative Technology Administration or the Bureau of Labor Statistics.

Special discount airfares, facilitated by the Internet and "frequent-flyer" programs, complicate efforts to measure changes in the price of commercial air travel. Endeavoring to fill their flights, airlines offer a variety of discount fares through several media (credit card points, supermarket coupons, and the like). The official Consumer Price Index (CPI) for commercial air travel, however, is based on prices listed by the airlines in SABRE, a reservation system used by many travel agencies. Thus, the CPI fails to reflect price changes that may be effected through special discounted prices and frequent-flyer awards. This article reports on a study whose aim was to produce an index series based on actual prices paid by consumers. The most promising data set currently available for that purpose is the Transportation Department's Data Bank 1B, which contains data from the quarterly Passenger Origin and Destination (O&D) Survey, collected by the U.S. Government's Bureau of Transportation Statistics. These data are itinerary based: each observation consists of a fare (the actual fare paid, including tax), a sequence of airports and carriers, and other details of an itinerary traveled by a passenger or a group of passengers.

The Department of Transportation is developing plans to improve and expand the O&D Survey. The additional data that the Department plans to collect will greatly enhance analysts' ability to compute detailed price indexes; among the new data is detailed information regarding the sale of the airline ticket, as well as transaction fares for flights in the recorded itineraries. The Department also plans to improve the timeliness of the survey data. Currently, the data become available with a lag of 3 to 6 months—too late to be used in computing the airfare component of the CPI. This article examines research aimed at computing price indexes from the current O&D Survey data. The Bureau of Transportation Statistics will soon be publishing the new quarterly experimental Air Travel Price Index (ATPI) series, computed at a variety of aggregation levels.

A secondary goal of the research is to test the feasibility of computing price indexes from non-matched samples of customized items. The sample for the O&D Survey is selected independently each quarter and is a 10-percent sample of airline tickets from reporting carriers, both foreign and domestic. Each ticket having a serial number that ends in "0" is selected for the sample. For the purpose of this research, the O&D sample is treated as a simple random sample. Because the quarterly samples are independently selected and airline itineraries are customized, matching the data across time is the primary challenge. Large data sets (containing, for example, scanned-in data) with the prices of other types of customized items may well become available in the future. The current research will provide insight into the potential usefulness and limitations of such data sets for price index computation.

The next section compares the ATPI with two important airfare price indexes currently in use.

Following that comparison, the methodological research undertaken in the development of the ATPI is discussed. Then, time plots of ATPI series, computed for research purposes, are presented. A discussion of possible directions for further investigation rounds out the text of the article. Most formulas and technical details are relegated to the appendix.

Comparison of airfare indexes

This section compares and contrasts the ATPI with two important airfare index series:

1. the BLS Consumer Price Index (CPI) for airline fares
2. the consumer expenditure deflator for airline fares, computed by the Bureau of Economic Analysis (BEA) and used in National Accounts estimation.

Comparing the experimental ATPI with the CPI. The Bureau of Labor Statistics currently publishes several price indexes for airfares: (1) a Consumer Price Index (CPI), (2) a Producer Price Index, and (3) international import and export price indexes. Because the CPI is perhaps the best-known and most widely used of the BLS price indexes, this section focuses on a comparison between the ATPI and the airfare component of the CPI. The CPI measures changes in the prices paid by consumers for airline trips, including taxes and all distribution costs paid by the consumers. The experimental ATPI series are similar to the BLS CPI in that the prices they measure include taxes paid, as well as fares received by the airline. The ATPI prices, however, exclude any distribution costs that were not received by the carriers (for example, travel agents' fees). The CPI includes trips purchased from foreign carriers, while the current ATPI series do not include data from foreign carriers.[1] CPI air-travel prices are gathered monthly from the SABRE system, while information on ATPI prices and quantities come from the O&D Survey.

The sample for the CPI airfare component is drawn from a subset of the O&D Survey data. Conceptually, the CPI excludes business trips, but because such trips cannot be identified on the sampling frame (information on the purpose of a trip is not collected in the O&D Survey), they cannot be screened out of the sample. Thus, both the CPI and ATPI samples include personal trips as well as business trips.[2]

Another important difference between the ATPI and the airfare CPI lies in the target index formulas used. 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 period t, for $i = 1, 2, ..., N$. Note that the index represents a comparison between prices in two arbitrary, but discrete, *periods* 1 and 2 (for example, months or 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. Index estimators, in contrast, must allow for the continual flow of goods and services on and off the market, as well as for the fact that information on prices and quantities normally are available only for a sample of items in the population.[3]

The Laspeyres formula, which measures changes in the price of a "fixed market basket" of items, is commonly used by government statistical agencies. Economic theory suggests, however, that other formulas may provide better approximations of changes in the cost of living, because consumers do not purchase the same set of items (a fixed market basket) in each survey period. Rather, they tend to alter their buying habits in response to changes in relative prices—for example, buying a particular brand of a product when that brand is on sale. Formulas such as the Jevons (or geometric mean), Fisher, and Törnqvist indexes are often considered more appropriate, given a "dynamic" market basket. (See the appendix for definitions of these formulas.)

The Fisher and Törnqvist indexes in particular are known as "superlative" indexes, because they approximate the change in the cost of living (that is, the cost of obtaining a fixed level of "utility") under relatively weak assumptions concerning consumer buying behavior.[4] The Jevons and Laspeyres formulas are often more practical, however, because they require less information on consumer expenditures than do the superlative formulas.

Since January 1999, the airfare CPI has been based on a weighted Jevons index formula within each sample geographic area, with sampling weights obtained from O&D Survey data. At the upper level of aggregation (aggregating across geographic areas), the CPI employs a modified version of the Laspeyres index, with weights estimated from Consumer Expenditure Survey data. The implementation of the Jevons index (replacing the Laspeyres index) at the lower level of aggregation in the CPI was motivated in part by empirical research.[5]

In the course of the ATPI research, indexes based on the Jevons, Laspeyres, Fisher, and Törnqvist formulas were computed. The Jevons index estimates were severely biased downward relative to the Fisher and Törnqvist estimates. Moreover, the Fisher index series proved more robust to extreme fare values[6] than did the Törnqvist series. Accordingly, the Fisher formula is the most desirable for the air-travel application[7] and is thus the one presented in this article.

The ATPI series also differ from the BLS CPI series in the definitions of their reference periods. From the current O&D

Survey data, only quarterly indexes can be computed, and the reference quarter is the quarter in which the airline ticket was used for travel.[8] The BLS CPI is a monthly survey, and the Bureau collects prices at which tickets are being sold (not necessarily used) during the reference month. Moreover, the scope of the ATPI is slightly wider than that of the airfare CPI. The CPI covers only trips that originate in the United States, whereas the O&D Survey encompasses trips originating in foreign countries, provided that they include at least one stop within the United States. Indexes with more limited scope may, of course, be computed by aggregating selected subsets of the data. For 1998–2003, the ATPI series for itineraries of flights originating in the United States (see later) shows a trend similar to that of the airfare CPI, although the differing formulas and reference periods result in different seasonal patterns for the two series.

Comparing the experimental ATPI with the BEA consumer expenditure deflator for airfares. The BEA computes chain-type price indexes for commodity categories for use in producing the National Income and Product Accounts estimates. For deflating consumer air-travel expenditure estimates, the BEA computes an index series based on both Department of Transportation data on total airline revenue per passenger mile flown and the BLS airfare CPI.

Results presented later indicate that the BEA deflator, which relies on measures of average revenue per passenger mile, does not provide a good approximation to a price index when the airline industry is undergoing a period of structural change. Airline financial data collected by the Bureau of Transportation Statistics show that the length of the average airline trip has been increasing in recent years,[9] and longer trips generally cost less per mile than shorter ones. Moreover, the overall quality of air-travel service has decreased with the emergence of low-cost carriers and the use of smaller, regional jets for cross-country flights. Both of these factors exert a downward pressure on the revenue that airlines collect per passenger mile, although they are not by themselves evidence of actual deflation.

Estimation method and research results

For the purpose of computing a price index, the peculiarity of the quarterly O&D Survey data is the absence of across-time matching of individual itineraries. In general, price index formulas are based on the direct comparison of prices of identical items in different periods. In the O&D Survey, the sample of tickets priced in time t is selected independently of the sample priced in time $t-1$. Moreover, some information that may affect the fares (for example, the time of day of the flight and the date the ticket was sold) is not collected through the survey. Thus, the survey cannot directly compare fares for identical air-service itineraries in different quarters. This section describes research on methods of addressing this primary obstacle to the use of O&D Survey data for index estimation. First, two stages of record matching are outlined: itinerary- and segment-level matching. Because the O&D data provide only itinerary-level airfares, fares for segment-level matching must be estimated. Alternative imputation methods are therefore discussed and compared. Finally, the results of a test designed to compare unit-value indexes computed from imputed segment-level fares against those computed from itinerary-level fares collected in the O&D Survey are presented.

Matching prices across time for index calculation. To circumvent the across-time matching problem, each quarterly sample can be divided into detailed categories, and a *unit-value index* (average price in time t, divided by average price in time $t-1$) computed for each category. The unit-value indexes are treated as elementary aggregates, which may be further aggregated with the use of standard index formulas (for example, Laspeyres, Paasche, Fisher, and Törnqvist formulas). Unit-value indexes are appropriate only for aggregating prices of items that are similar (for instance, round-trip United Airlines coach service from Boston to San Francisco with one stop in Chicago).

The first stage of matching is *itinerary-level* matching, in which the itineraries are cross-classified by the following variables:

(1a) sequence of origination and destination airports (that is, origination airport, first destination airport, second destination airport, and so on)
(1b) sequence of classes of service (that is, class of service for first segment, second segment, and so on)
(1c) sequence of operating carriers

Itineraries that are identical in characteristics 1a through 1c form a *first-stage unit-value category*. Note that trips within the first-stage category must have exactly the same number of trip *segments*, or flights.[10] As the number of segments increases, the percentage of categories appearing in both of two consecutive quarterly databanks decreases. For trips with eight segments, less than 2 percent of the unit-value categories could be matched across consecutive quarters. As a result, the first-stage matching procedure was performed only for trips with eight or fewer segments. (Just 0.15 percent of the itineraries in the O&D Survey databanks comprise nine or more segments.)

The second-stage matching procedure is *segment-level* matching. Itineraries not matched in the first stage are broken into individual segments. Because only itinerary-level fares are available in the databanks, the second-stage matching procedure involves imputing (that is, estimating) a fare for each segment. Two alternative methods of imputation are discussed in the next subsection. After the fares for second-stage matching are imputed, the trip segments are cross-classified by the following variables to form *second-stage unit-value categories:*

(2a) segment-level origination and destination airports
(2b) class of service
(2c) round-trip itinerary or non-round-trip itinerary
(2d) itinerary of U.S. origin or of foreign origin
(2e) operating carrier

Unit-value indexes are computed for these segment-level categories and are then matched from quarter to quarter.

The entire matching process, involving both first- and second-stage matching, is performed separately for each pair of consecutive quarters, to create a "rolling" sample. The extent to which the segment-level matching increases the percentage of trip segments that can be matched across quarters depends on the second-stage fare imputation method used. It is expected, however, that a small percentage of segments will always be omitted from the index computations due to incomplete matching.

Second-stage fare imputation methods. Two methods of second-stage fare imputation were compared and designated the "single-segment matching method" and "proportionate distance method," respectively. Of the two methods, the single-segment matching method clearly has the lower potential for introducing bias, but it results in a lower matching percentage.

For the single-segment matching method, the proportion of the fare contributed by each segment is estimated on the basis of the relative values of fares for single-segment itineraries similar to those of the individual segments. Let M_i be the number of segments in an unmatchable itinerary i. For each $m = 1,...,M_i$, segment m is matched to a set of single-segment itineraries having the same origination airport, destination airport, and class of service. Let $\overline{p}'_{im}$ denote the average fare, excluding fares with a value of zero, for single-segment itineraries thatmatch segment m of itinerary i and p_i denote the fare for itinerary i. Then, for this method, the imputed fare for segment m is

$$\hat{p}_{1im} = p_i \left(\frac{\overline{p}'_{im}}{\sum_{l=1}^{M_i} \overline{p}'_{il}} \right).$$

Clearly, in order to impute a fare by the single-segment matching method, *each* of the segments in itinerary i must be able to be matched to at least one nonzero fare for a similar one-segment itinerary.

The alternative second-stage imputation method examined assigns fares on the basis of the proportion of total mileage represented by the individual segments within the itinerary. That is, the imputed fare for segment m in itinerary i is

$$\hat{p}_{2im} = p_i \left(\frac{d_{im}}{\sum_{l=1}^{M_i} d_{il}} \right),$$

where p_i is as before and d_{im} is the distance[11] traveled in segment m of itinerary i (available in the databank).

Each of the methods described has its limitations. Because the proportionate distance method uses only relative distances to divide the fare among the segments, it can reasonably be applied only to itineraries in which all segments were flown in the same class of service. The restriction imposed by the single-segment matching method, though, is even more severe: if just one segment in itinerary i has no comparable one-segment itineraries in the quarterly databank, the method cannot be used to impute fares for *any* of the segments in the itinerary.

Both methods, however, allow for an implicit form of imputation within second-stage unit-value categories. Suppose, for example, that a particular segment does not qualify for single-segment matching imputation. When this situation arises because *another* segment in the itinerary could not be matched to a similar single-segment itinerary, there may be fare values in the unit-value category into which the segment falls. The segment then implicitly receives an imputed value equal to the average imputed fare for that category.[12] That is, the segment's missing fare does not affect the average for the category, but the segment still contributes to the category's *weight* in the aggregate indexes. Similarly, a segment that fails to qualify for proportionate distance imputation because of disparate class-of-service codes within the itinerary may fall into a unit-value category that contains fare values and be implicitly imputed.

The clear disadvantage of the proportionate distance method relative to the single-segment matching method is that it fails to account for price pressures *other than* the distance of the flight (for example, airline "overhead" costs, and supply and demand). Note, however, that although this deficiency undoubtedly leads to biased fare estimates (generally speaking, assigning too large a proportion of the fare to longer flights), it does not imply that the proportionate distance method yields unit-value indexes that are significantly biased relative to those computed by single-segment matching. The initial thinking was that if the bias pattern were relatively constant across time, then the unit-value indexes computed by the proportionate distance method—and thus the aggregate indexes—would closely approximate those computed by single-segment matching. This hypothesis was tested with data from a four-quarter test period stretching from the third quarter of 1999 to the second quarter of 2000.

Testing revealed that, within the itineraries not matched in the first stage, roughly 53 percent to 54 percent of the segments qualified for single-segment matching imputation, whereas the percentage qualifying for proportionate distance imputation hovered around 85 percent. As expected, proportionate distance imputation consistently allowed the imputation of a higher overall percentage of segment-level fares and the matching of more passenger flight segments across quarters, thus reducing the potential for index bias resulting from the omission of certain itineraries or segments.

In general, roughly 84 percent of itineraries, representing about 75 percent of passenger flight segments, are matched in the first stage. (Because itineraries comprising large numbers of segments are less likely to be matched in this stage, the percentage of itineraries matched is expected to exceed the percentage of segments matched.) About 75 percent of the passenger segments not matched in the first stage are matched in the second stage under single-segment imputation. The newly matched segments include segments whose fares have been implicitly imputed, as described earlier. The matched segments represent approximately 18 percent of passenger flight segments in the databanks. For single-segment matching imputation, the resulting total percentage of segments matched is about 93 percent to 94 percent. Proportionate distance imputation provides a total matching percentage of roughly 98 percent. It is important to note, however, that under single-segment matching imputation, a larger percentage of segments receives implicit imputation: about 21 percent of second-stage segments (roughly 5.25 percent of all segments) are implicitly imputed under this method, compared with about 9 percent (2.25 percent of all segments) for proportionate distance imputation.

Nonetheless, fares imputed by the proportionate distance method do indeed appear to be somewhat biased relative to those imputed by the single-segment matching method. On the one hand, results of t-tests indicated that, at low levels of aggregation (at which the indexes were subject to high variances), the differences between the unit-value indexes computed by the two methods were not significant. For t-statistics based on unit-value indexes within "city of origin" categories, for example, p values generally ran between 0.02 and 0.8. On the other hand, at higher levels of aggregation, significant differences were sufficiently common to raise concern, even given the magnitude of the sample sizes. Within "class of service" categories, p values below 0.05 appeared for three of the six major categories in one of the quarter-to-quarter test periods (the first to second quarter of 2000). Moreover, even at low levels of aggregation, the t-statistics revealed that distance-based imputation yielded consistently higher unit-value indexes than did single-segment matching imputation.

An examination of the differences between fares imputed by the two methods indicated that proportionate distance imputation generally overestimated fares for longer flights and underestimated fares for shorter flights. This was expected, because the method fails to account for airline overhead costs associated with individual flights. The majority of flights recorded in the databanks have distances less than the average distance (that is, the mean flight distance exceeds the median distance), so we expect distance-based imputation to *underestimate* fares for the majority of flights. This tendency, along with the general one of the unit-value indexes to exceed unity, may account for the upward bias of the unit-value indexes computed through proportionate distance imputation. To see this relationship, let $\hat{f}_{1,t_1}$ and $\hat{f}_{1,t_2}$ represent the single-segment imputed fares for periods 1 and 2, respectively, and suppose that $\hat{f}_{1,t_1} < \hat{f}_{1,t_2}$ (in other words, fares increased between periods t_1 and t_2). Let d represent the absolute value of the bias (assumed constant and additive) of the distance-based imputed fares relative to those computed by single-segment matching. (That is, for $i \in \{1,2\}$, let $\hat{f}_{2,t_i} = \hat{f}_{1,t_i} - d$.) Then, with $d < \hat{f}_{1,t_1} < \hat{f}_{1,t_2}$, it follows that

$$\frac{\hat{f}_{2,t_2}}{\hat{f}_{2,t_1}} = \frac{\hat{f}_{1,t_2} - d}{\hat{f}_{1,t_1} - d} > \frac{\hat{f}_{1,t_2}}{\hat{f}_{1,t_1}},$$

giving an upward bias for the unit-value indexes computed by proportionate distance imputation.

The assumption of a constant additive bias is, of course, a strong one. It is also possible that the upward direction of the bias of the unit-value *indexes* computed by proportionate distance imputation indicates that the bias pattern of the *fares* imputed by this method is changing gradually over time. Specifically, the upward bias of the imputed fares for long-distance flights may be increasing, perhaps indicating that factors other than distance were exerting an increasing influence on the prices of airline flights over the period examined. It is therefore possible that, during other periods—especially those marked by rapidly increasing fuel costs—the direction of the bias of the unit-value indexes changes.

In sum, the test results indicated cause for concern about the potential bias of unit-value indexes computed by the proportionate distance method, relative to those computed by single-segment matching imputation. Although the proportionate distance method yielded a higher overall matching percentage, the difference in matching percentages was not sufficient to warrant the use of that method in view of its evident deficiencies.

Comparing first- and second-stage unit-value indexes. Under single-segment matching imputation, second-stage unit-value indexes were compared with unit-value indexes obtained from first-stage (itinerary-level) matching. Using only observations that matched in the first stage, the following indexes were computed for each first-stage category c:

i. a first-stage unit-value index $u_{c,1,2}$ (as discussed earlier in the section; see the appendix for the formula) and

ii. an index , $u^{(s)}_{c,1,2}$ based on unit values computed through second-stage matching. (Again, see the appendix for the formula.)

Note that $u^{(s)}_{c,1,2}$ reflects a price change for an *itinerary-level* (first-stage) category, but is computed by aggregating *segment-level* (second-stage) unit-value indexes for the various segment-level categories that correspond to the itinerary-level category. For example, the first-stage category comprising restricted coach itineraries for United Airlines round-trip service from Washington's Reagan National Airport to Chicago's O'Hare Airport has two corresponding segment-level categories, one for each segment of the itinerary: (1) United restricted coach service from Washington Reagan to Chicago O'Hare within a round-trip itinerary and (2) United restricted coach service from Chicago O'Hare to Washington Reagan within a round-trip itinerary.

To examine the effects of segment-level imputation and matching relative to those of itinerary-level matching, the distributions of $u^{(s)}_{c,1,2}$ and $u_{c,1,2}$ for the second through the fourth quarters of 2000 were compared. Histograms [13] showed that the distributions were similarly shaped (slightly positively skewed) and that the distribution of the differences $u_{c,1,2} - u^{(s)}_{c,1,2}$ was roughly symmetric about zero. For the three quarter-to-quarter changes tested, the numbers of first-stage categories, shown in table 1, hover around 300,000. In each case, the mean difference $u_{c,1,2} - u^{(s)}_{c,1,2}$ is statistically significant, due to the large number of categories. The Fisher indexes computed from the two sets of subindexes, however, differ only in the third decimal place, as indicated in the table.

Chart 1 summarizes the current two-stage procedure in flowchart form. Note that the current experimental process does not include a "quality adjustment" step to account for changes in the real values of itineraries flown in different periods (due, for example, to changes in food served or seating space). Quality adjustment is not practical here, because the data needed for such adjustment (for instance, by hedonic regression) are not collected in the current O&D Survey. More importantly, we have no reason to believe that the collection of itineraries matched in later quarters is qualitatively any better than the collections matched in earlier quarters. Rather, the unmatched flights and itineraries simply represent unusual travel routes flown in particular quarters. Thus, the systematic downward bias that the absence of quality adjustment may induce for items whose quality is generally improving with the introduction of new models [14] is unlikely to occur in the application presented here.

Table 1. Fisher indexes computed by aggregating first-stage ($u_{c,1,2}$) and second-stage ($u^{(s)}_{c,1,2}$) unit-value indexes

Index period, 2000	First-stage Fisher index	Second-stage Fisher index	Number of categories
First quarter to second quarter	1.02679	1.02866	287,727
Second quarter to third quarter	1.02468	1.02202	325,445
Third quarter to fourth quarter	1.00613	1.01036	312,343

Experimental index series

This section examines some ATPI series, based on the Fisher index formula, for several class-of-service and point-of-origin categories. Indexes based only on first-stage matching are labeled "preliminary," while those based on both first- and second-stage matching are labeled "final." The index series to be presented were computed solely with data from U.S. carriers; that is, only itineraries flown entirely on U.S. carriers are in scope for these series. Except where otherwise stated, the index series shown are referenced to the first quarter of 1995.

The discussion accompanying the charts that follow is intended to highlight interesting features of the index series. In interpreting the series, readers should bear in mind the scope of the O&D Survey, as well as the exclusion of foreign-carrier flights from the data. The survey covers all air itineraries having some U.S. component and being flown on all carriers reporting. Thus, the index series computed for foreign points of origin cover, not *all* itineraries originating from those points, but only the itineraries that include some U.S. destination or "stopover" points.

The "class of service" variable for the O&D Survey underwent a standardization process in 1997–98, and the change in reporting codes may be responsible for some of the movements observed in the index series. Accordingly, in the discussion that follows, special attention is given to the portion of the series between the fourth quarter of 1998 and the second quarter of 2003. Tables 2 and 3 summarize the percent changes over this period.

Primary ATPI series compared with BLS and BEA airfare index series. Chart 2 compares the ATPI series with the BLS CPI series and the BEA Personal Consumption Expenditure Deflator for airfares.[15] The top panel shows all series referenced to the first quarter of 1995, the bottom panel to the fourth quarter of 1998. The BLS index differs in its seasonal pattern from both the BEA index and the ATPI, due to its different definition of the reference period (the date of sale rather than the date of the flight). Consequently, just the long-term trends, and not the quarterly movements, of the different index series are comparable. The BLS CPI covers

Chart 1. Two-state matching procedure for Passenger Origin and Destination (O&D) Survey airfare index computation

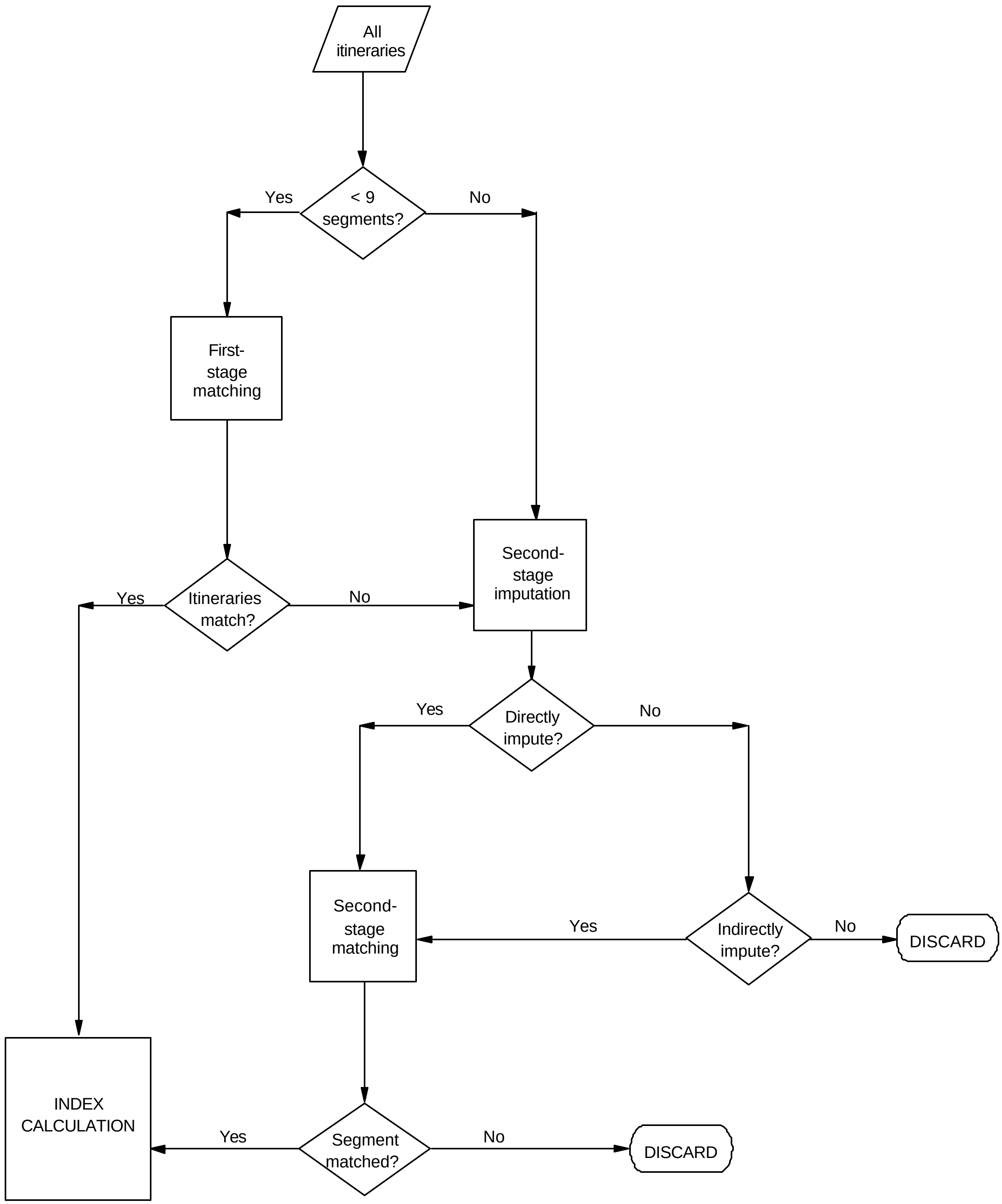


Table 2. Percent change for major index series, fourth quarter 1998 to second quarter 2003

Series	Percent change
BLS CPI for airline fares	15.4
BEA personal consumption expenditure deflator for airfares	–11.7
Full-scope ATPI	6.6
U.S.-origin ATPI	6.8
Foreign-origin ATPI	4.4
Restricted coach class ATPI	9.8
Unrestricted coach class ATPI	–9.2
Restricted first-class ATPI	7.1
Unrestricted first-class ATPI	1.4
Restricted business-class ATPI	42.1
Unrestricted business-class ATPI	11.4

only itineraries originating in the United States and is comparable, therefore, to the "U.S.-origin-only" ATPI.

Before the third quarter of 1996, the BLS modified Laspeyres index suffered from an upward "formula bias."[16] Thus, we expect the BLS index to run above the U.S.-origin ATPI for the period from the first quarter of 1995 to the third quarter of 1996. For the period from the fourth quarter of 1998 to the second quarter of 2003, the BLS index is based on the hybrid Jevons/Modified Laspeyres formula.[17] The BLS index increased 15.4 percent during this period, while the U.S.-origin ATPI increased 6.8 percent and the full-scope ATPI increased 6.6 percent. This difference is probably due mainly to (1) the different target formulas used (Fisher or Jevons/Modified Laspeyres) and (2) the ATPI's inclusion of special discount fares that involve differential pricing (for example, frequent-flier awards and Internet specials), combined with consumers' increasing use of special discount tickets during the period. The U.S.-origin ATPI also shows a sharper drop in the last two quarters of 2001—a more pronounced "9/11 effect"—than is seen in the airfare CPI.

Chart 2 also compares the ATPI series with the quarterly BEA Personal Consumption Expenditure Deflator for airfares. In the top panel, which shows all series referenced to the first quarter of 1995, the BEA series runs above the U.S.-origin ATPI series for most of the period shown, but the two series cross in the fourth quarter of 2000, when the BEA series begins a steep decline. The bottom panel of chart 2 shows the BEA series running consistently below the ATPI. Research has revealed that the average distance flown per airline itinerary has been steadily increasing in recent years, which has naturally led to a decline in air carrier revenues per passenger mile.[18] Because the BEA index is driven largely by a measure of revenue per passenger mile, we expect the increase in distance, along with a corresponding increase in the percentage of passengers choosing "no-frills" air-travel service, to push the BEA series below the ATPI series during the 1999–2003 period.

Comparing final and primary ATPI series. The top panel of chart 3 shows the preliminary and final ATPI series for U.S. and foreign points of origin. As expected at this level of aggregation, the two series are virtually indistinguishable. The same holds for the series (not shown) for foreign and domestic points of origin combined.

The remaining three panels of chart 3 show preliminary and final series by class of service for domestic points of origin. Index values for 1997–98 must be interpreted with caution, because the reporting codes were changed during this period. A variety of reporting codes previously used were standardized to produce the basic categories of first class, business class, and coach. Each of these categories is further divided into restricted and unrestricted tickets; the price for restricted tickets carried some restrictions for the purchasers. (For example, advance booking was required, and there was an added fee for a change in schedule.) Again, in general, little difference is found between the preliminary and final versions of the experimental series. Whatever differences there are are especially small for the largest category: restricted coach (second panel of chart 3). For unrestricted coach service, the preliminary and final series are similar, except that (1) the final series shows a less severe "break" (in this case, an upward jump) between the fourth quarter of 1997 and the first quarter of 1998, and (2) the final series shows a more pronounced drop from the terrorist attacks of 9/11 in 2001.

The restricted coach index is conceptually the closest substitute for a consumer price index that has been produced from the O&D Survey data: it reflects movements in fares paid by the most price-conscious buyers. The final restricted coach series increased by 2.6 percent from the first quarter of 1995 to the second quarter of 2003. From the fourth quarter of 1998 to the second quarter of 2003, however, it increased by 9.8 percent, closer to the increase indicated by the official airfare CPI (See chart 2.) The unrestricted coach series displays an unusual downward spike from the third quarter of 1995 to the second quarter of 1996; because a number of class-of-service code systems were in use during that period, the odd movement may be associated with variability in coding. Over the entire period from the first quarter of 1995 to the second quarter of 2003, the final unrestricted coach series increased 16.4 percent, while the restricted coach series increased by 2.6 percent, as just noted. Over the period from the fourth quarter of 1998 to the second quarter of 2003, however, the trend was reversed: the unrestricted coach series decreased 9.21 percent, while the restricted coach series increased the aforementioned 9.8 percent.

The series for business-class service appear in the third panel of chart 3. For these categories, the differences between the preliminary and final versions of the series are noticeable, but not extreme. Moreover, the final series runs slightly above the preliminary series for restricted business-class service and slightly below the preliminary series for unrestricted business-

Table 3. Percent changes for point-of-origin ATPI series, fourth quarter 1998 to second quarter 2003

City or area	Percent change
Chicago	–0.78
Los Angeles	3.1
New York	4.4
Montreal, Canada	18.1
Toronto, Canada	19.2
Vancouver, Canada	24.0
Canada	18.7
Frankfurt, Germany	10.1
London, England	13.8
Tokyo, Japan	3.2
Houston	8.8
Minneapolis	17.1
Washington, DC	14.1
Detroit	18.6
Charleston, SC	23.5
Colorado Springs	7.0
Des Moines	–1.3
Albany	10.8
Dayton	7.2
Tucson	4.0

class service, indicating that there is no systematic bias associated with the unit-value indexes produced through second-stage matching. The "big dipper" movement of the restricted business-class series during 1997–98 may be due in part to the earlier mentioned changes in reporting codes. Changes in frequent-flier upgrade behavior also may be partly responsible.

The bottom panel of chart 3, showing the series for first-class service, reveals almost no difference between the preliminary and final versions of the series for restricted first-class service, except for a slight divergence during the 1997–98 break. The series for unrestricted first-class service are similar to those for unrestricted coach (second panel of the chart): the final series differs from the preliminary one only in that it suffers a milder 1997–98 break. Moreover, during the period from the fourth quarter of 1998 to the fourth quarter of 2000, the restricted first-class series displays movements similar to those of the series for restricted coach service. This similarity may reflect a growing number of frequent-flier passengers who upgraded and flew first class during the period, together with an increase in the number of coach service seats classified as first class by some carriers when reporting data for the O&D Survey.[19]

The indexes shown in the last three panels of chart 3 generally indicate steeper price increases for unrestricted air service than for restricted service. Because special discount fares apply almost exclusively to restricted service, these indexes provide evidence that the divergence of the BLS and ATPI series (see chart 2) is due in part to the O&D Survey's inclusion of such discount fares.

Index series by place of origin. This section examines O&D Survey index series computed for various cities of origin in a passenger's itinerary. These series are local-area economic indicators reflecting changes in the airfare component of the cost of living for residents of the cities in question. Particular cities, representing a wide range of geographic areas and sizes, were selected to serve as examples. Note that, for these detailed itinerary-level points of origin, second-stage matching is not practical due to the small number of segments in most of the resulting second-stage categories. For these characteristics, the preliminary series are therefore final.

The series in the top panel of chart 4 for the three largest U.S. cities indicate similar price movements for itineraries originating in these cities. The series run roughly parallel to, though slightly above, the U.S. Origin ATPI series shown in chart 2. Much more disparity appears in the movements of the series for Canadian[20] cities of origin (middle panel of chart 4), with Toronto exhibiting the largest increase by far over the period shown. Except for the "9/11 effect," the Canadian city index series tend to gradually level off during the later years of the period. Interestingly, the Toronto series displays a much more pronounced 9/11 effect than the series for the other Canadian cities.

The most striking feature of the index series for large overseas cities of origin (bottom panel of chart 4) is the seasonal pattern. The third-quarter spikes indicate a predominance of vacation travelers paying peak-season fares. Price movements for overseas cities of origin are confounded with changes in currency exchange rates, which may account for some of the overall decrease in the series shown in the chart. Except for seasonality, these series, like those for U.S. and Canadian cities, tend to level off from the fourth quarter of 1998 to the second quarter of 2003. One possible exception, however, is the Frankfurt series, which shows an unusual increase in the final 2 years of the period.

The Houston series (see top panel of chart 5) is similar to the series for Los Angeles (chart 4), except that it shows larger increases in the first quarters of 2001 and 2003. Similarly, the series for Detroit and Minneapolis (top panel of chart 5) track each other quite closely, perhaps due to geographic proximity and the dominance of the same air carriers in the two cities. Although these two series run well below those for the larger cities, they display the same "leveling" trend during the final years shown and a much less pronounced 9/11 effect. For the period from the fourth quarter of 1998 to the second quarter of 2003, the series for Detroit, Minneapolis, and Washington, DC (again, top panel of chart 5), show some of the larger increases among the "city of origin" series examined. The Washington index increased 14.1 percent over this period, while the Detroit and Minneapolis series increased 18.6 percent and 17.1 percent, respectively. In the latter two series, however, the increases followed steady declines seen in the previous couple of years. The city index with the largest decrease (amongthose shown) for the period from the fourth quarter of 1998 to the second

Chart 2. BLS hybrid airfare CPI and primary ATPI series, not seasonally adjusted, 1995–2003

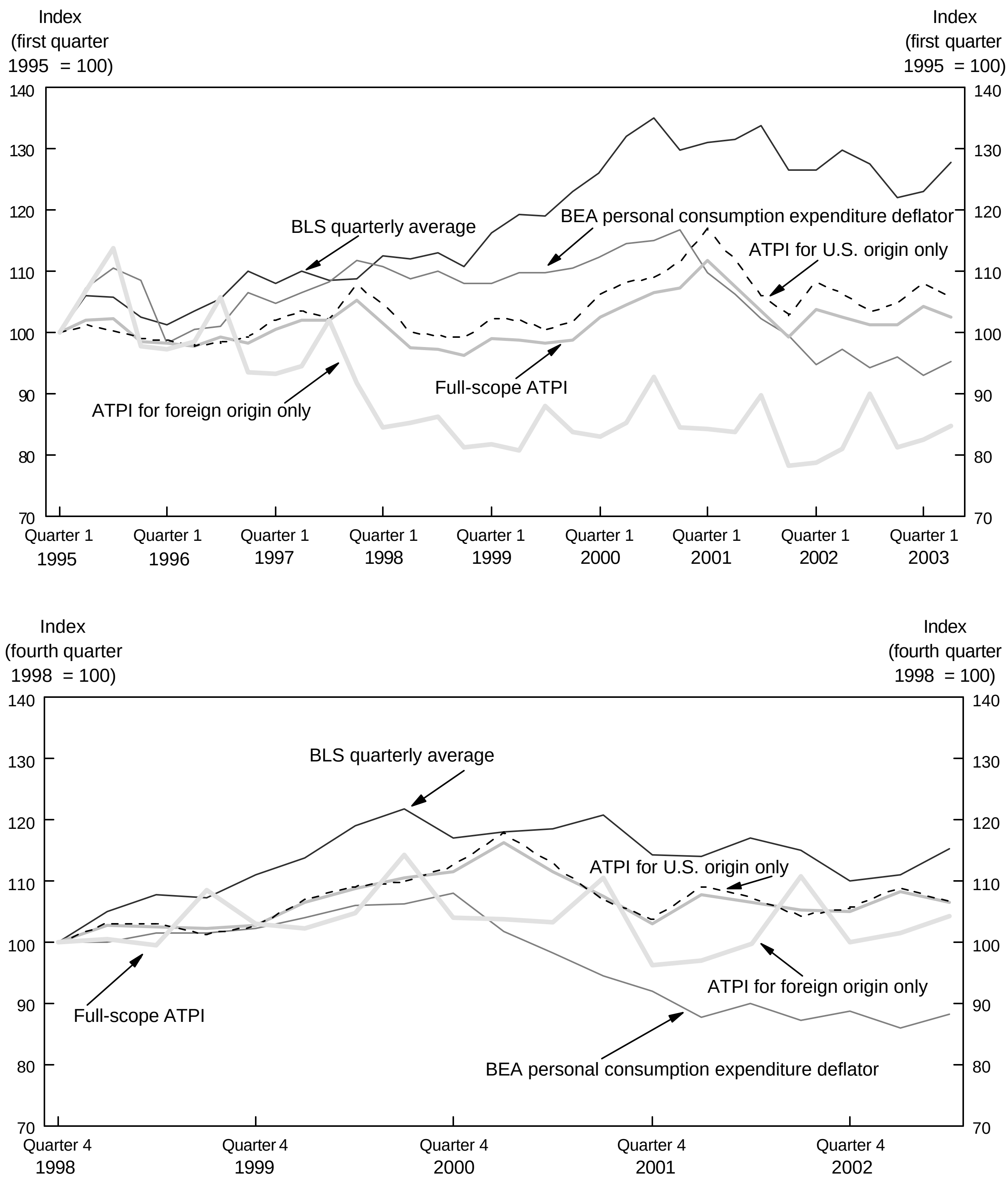


Chart 3. Preliminary and final ATPI series for U.S. and foreign points of origin or for U.S. points of origin alone, 1995–2003

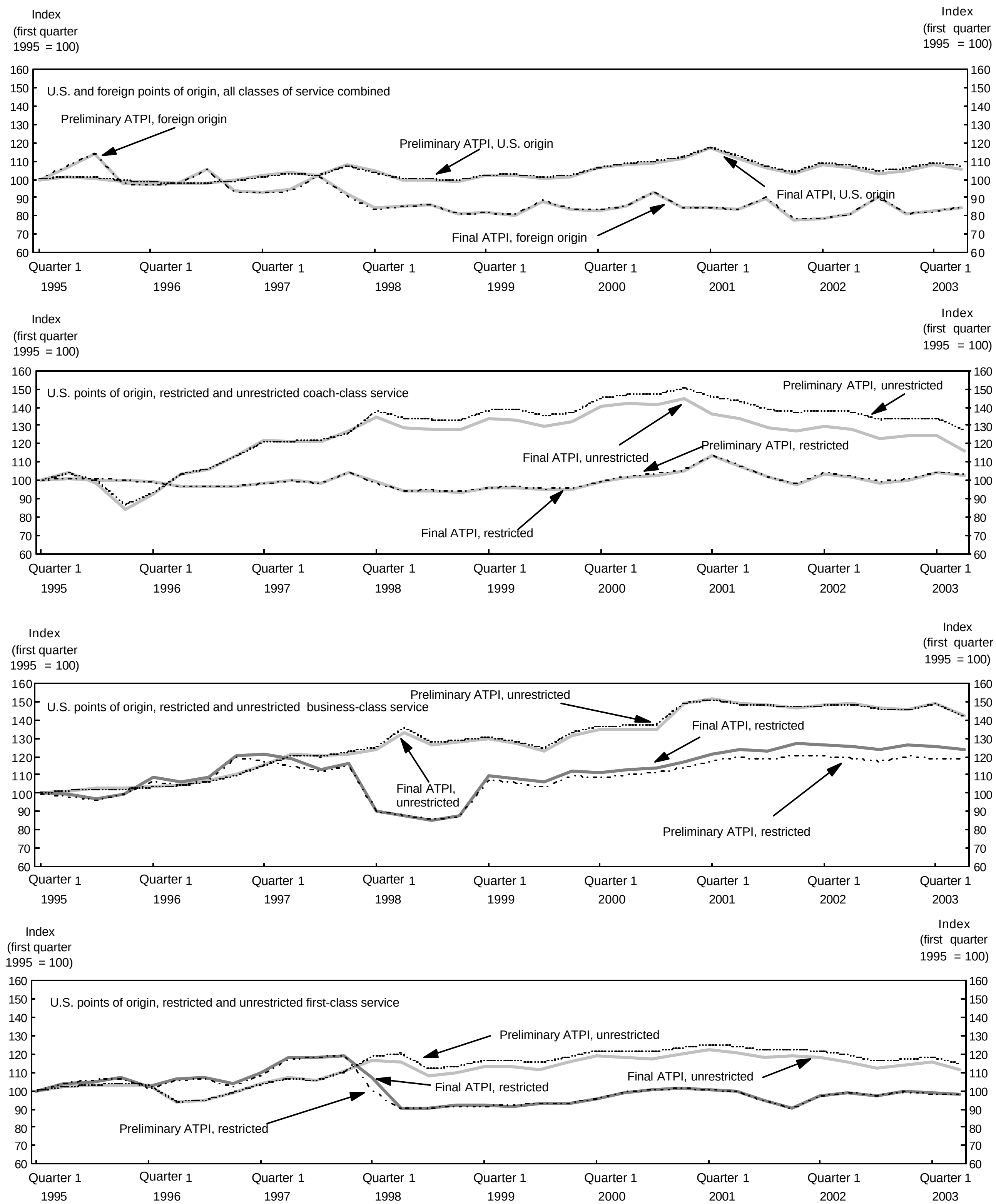


Chart 4. ATPI series for large cities of origin, all classes of service combined, 1995–2003

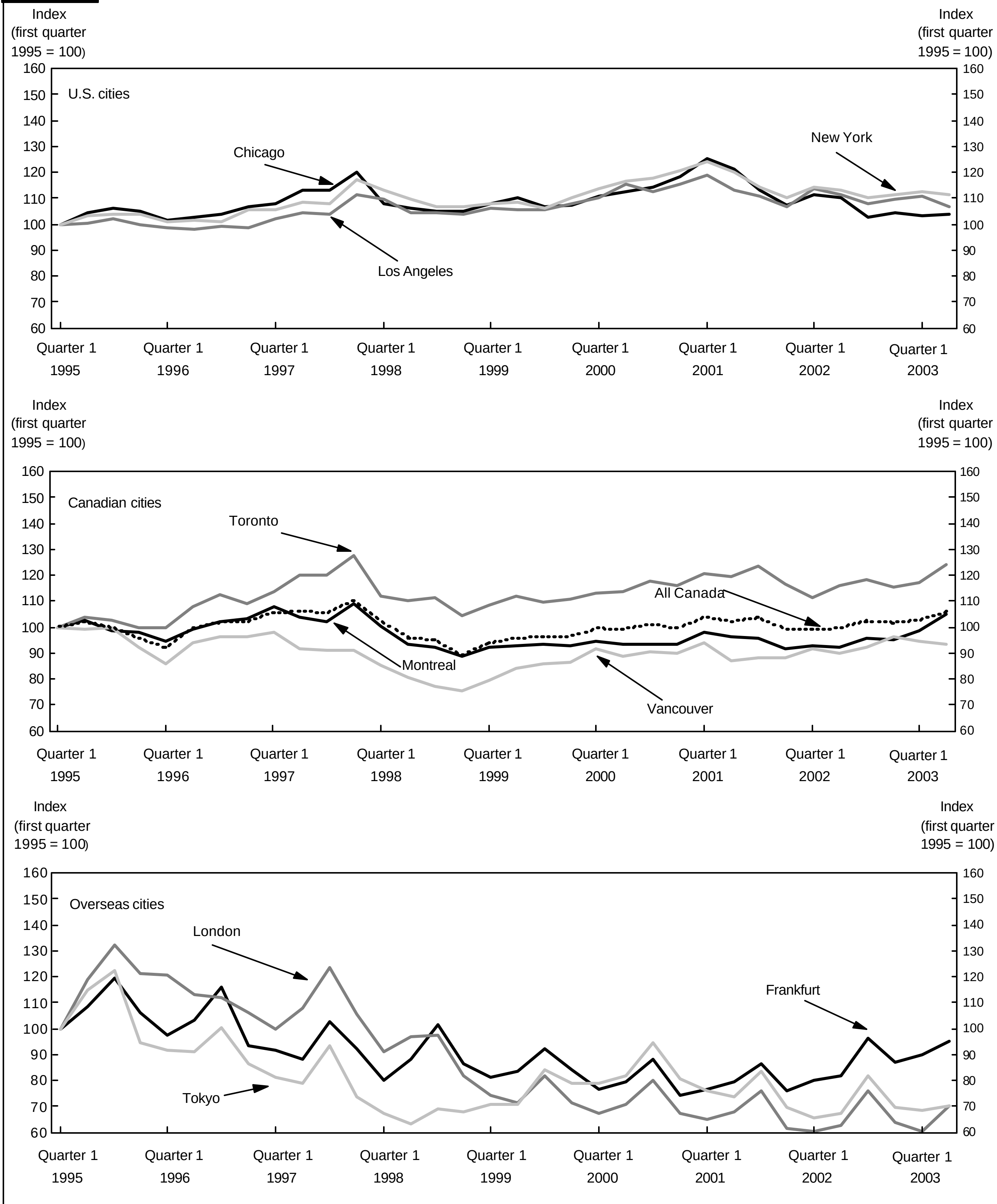
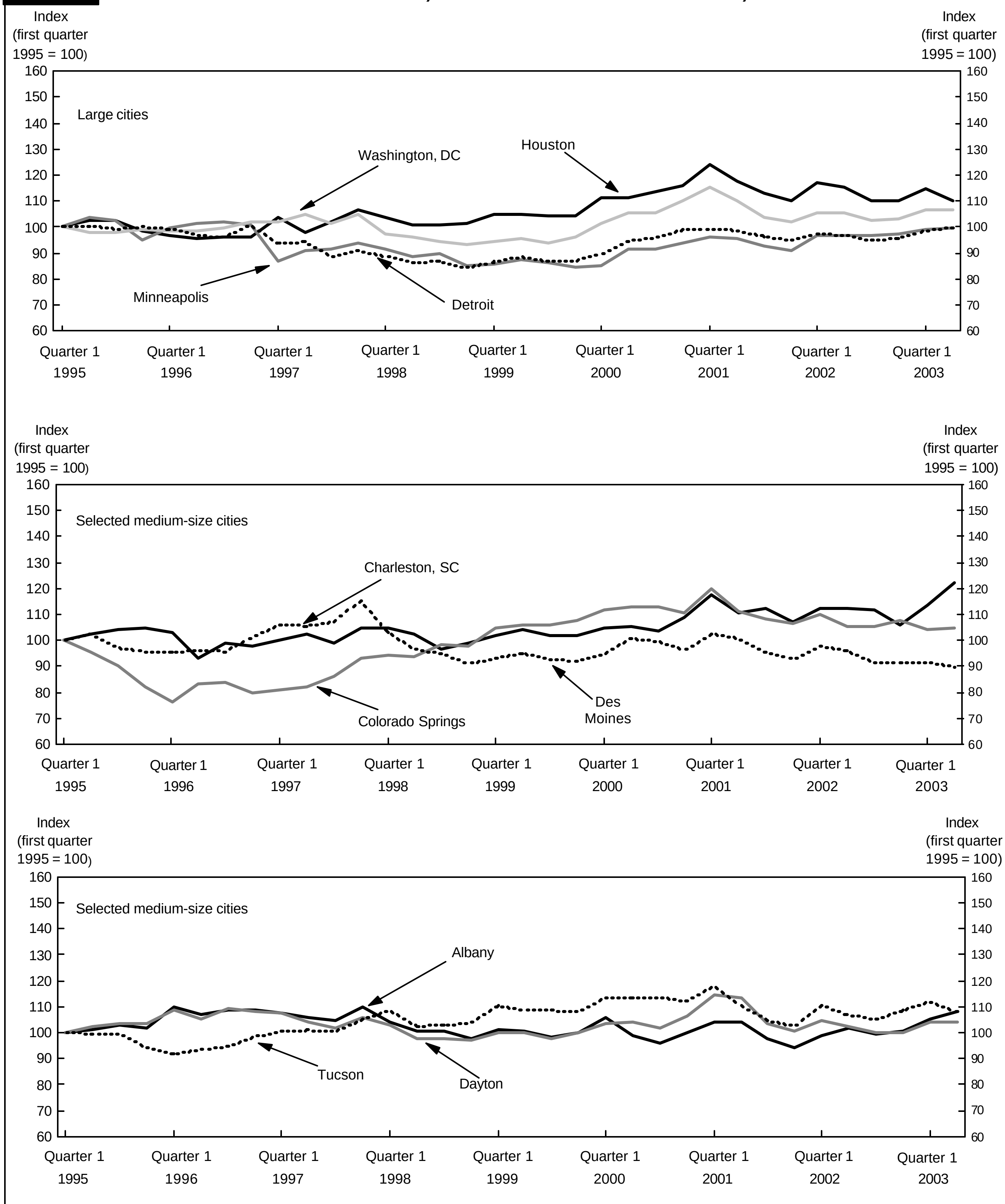


Chart 5. ATPI series for selected U.S. cities, all classes of service combined, 1995–2003



quarter of 2003 is that of Des Moines, Iowa, with a drop of 1.3 percent. (See middle panel of chart 5.) The series for Charleston, South Carolina, however, gradually climbed 23.5 percent during the same period. The index series for Colorado Springs, Colorado (again, middle panel of chart 5), reflects the impact of Western Pacific, a low-cost airline that began offering discount service from Colorado Springs to Dallas-Fort Worth in the second quarter of 1995. In 1995–96, Western Pacific expanded its operations to other markets, including Seattle and Washington, DC. Larger airlines responded by lowering fares and expanding service in markets served by Western Pacific, which then was forced to curtail its operations, ultimately ceasing all operations in the early part of 1998.

The series for Albany, New York, and Dayton, Ohio (bottom panel of chart 5), track each other fairly closely, except for the dip in the Albany series in the second and third quarters of 2000. Their similarity may reflect regional economic impacts and similar servicing carriers, as do the Detroit and Minneapolis series shown in the top panel of the chart. The Tucson, Arizona, series (bottom panel of chart 5) is atypical, displaying movements somewhat similar to those seen in the Colorado Springs series, though less dramatic. In the case of Tucson, however, there is no firm evidence of a "discount carrier" effect on the index series. (The presence of Reno Air in the Tucson market may have exerted a downward pressure on airfares from Tucson, but Reno did not exit the market until the second quarter of 1999, well after fares had begun to increase.) BLS employment and unemployment data[21] indicate a general economic downturn in Tucson in 1996–97, characterized by increased unemployment levels and rates; this decline seems the most likely explanation for the contemporaneous dip in airfares.

Additional developments

The Bureau of Transportation Statistics' ATPI research project has involved numerous specific methodological studies. In one such study, an empirical investigation into alternative chaining intervals revealed no evidence of chain drift in the quarterly chained Fisher series presented in this article.[22] A study of sensitivity to extreme values showed the Fisher index estimator to be more robust than the Törnqvist for the airfare application. In the future, the expanded O&D Survey data will offer the possibility of using shorter chaining intervals—for example, months or even weeks—and of producing timely monthly indexes. Other areas for future research include standard error estimation for the index series and the development of seasonal adjustment methods. □

Notes

ACKNOWLEDGMENTS: Suggestions from Ed Wegman of George Mason University, Sylvia Leaver of the Bureau of Labor Statistics, and Othmar Winkler of Georgetown University have greatly enhanced the work presented in this article. We also thank the *Monthly Labor Review* referees for their helpful comments.

[1] Although the experimental ATPI series were originally computed with data from both foreign and domestic carriers, legal concerns from within the Department of Transportation required the suppression of the data from foreign carriers. Overall, the removal of those data resulted in minimal changes to the series presented.

[2] For a detailed discussion of CPI methods, see *BLS Handbook of Methods* (Bureau of Labor Statistics, 1997), Chapter 17, "The Consumer Price Index," available on the Internet at **www.bls.gov/opub/hom/home.htm**.

[3] For a discussion of alternative index formulas, see, for example, Irving Fisher, *The Making of Index Numbers: A Study of Their Varieties, Tests, and Reliability* (New York, Sentry Press, 1922); W. Erwin Diewert, "Index Numbers," in John Eatwell, Murray Milgate, and Peter Newman, eds., *The New Palgrave: A Dictionary of Economics, Vol. II* (London, MacMillan, 1987), pp. 767–80; or Brent R. Moulton, "Basic components of the CPI: estimation of price changes," *Monthly Labor Review*, December 1993, pp. 13–24 (on the Internet at **http://www.bls.gov/opub/mlr/1993/12/art2full.pdf**). For more information on price index concepts and design, see Charles L. Schultze and Christopher Mackie, eds., *At What Price? Conceptualizing and Measuring Cost-of-Living and Price Indexes* (Washington, DC, National Academy Press, 2002).

[4] See W. Erwin Diewert, "Exact and Superlative Index Numbers," *Journal of Econometrics,* May 1976, pp. 115–45.

[5] See, for example, Ana M. Aizcorbe and Patrick C. Jackman, "The commodity substitution effect in CPI data, 1982–91," *Monthly Labor Review,* December 1993, pp. 25–33; and Matthew D. Shapiro and David Wilcox, "Alternative Strategies for Aggregating Prices in the CPI," *Federal Reserve Bank of St. Louis Review,* May/June 1997, pp. 113–25. The methods and results described in these articles have been called into question by Alan H. Dorfman, Sylvia G. Leaver, and Janice Lent, "Some Observations on Price Index Estimators," *Proceedings of the Federal Committee on Statistical Methodology Research Conference, Monday B Sessions* (Federal Committee on Statistical Methodology, 1999), pp. 56–65.

[6] The O&D Survey fares include zero-value fares (for example, for frequent-flier awards), which are imputed as $0.01. These imputations often result in extreme values for the unit-value indexes that serve as the "atoms" of the indexes presented in this article. (See later.)

[7] For a discussion of the performance of the index formulas with respect to outliers in air-travel application, see Janice Lent, "Effects of Extreme Values on Price Indexes: The Case of the Air Travel Price Index," *Journal of Transportation and Statistics*, vol. 7, nos. 2–3, 2004, pp. 41–52.

[8] If an itinerary straddles multiple quarters, it is counted in the

quarter in which the first ticket in the itinerary is used.

[9] See, for example, the findings of Steven Anderson and Richard Leonard, "Domestic Airline Industry Passenger Price Trends," internal document, Bureau of Transportation Statistics, April 26, 2004.

[10] In the terminology used in this article, one segment involves exactly one aircraft takeoff and landing. Due to reporting deficiencies in the O&D Survey, some multiple-stop flights are currently being reported as nonstop flights, and the actual number of stops cannot always be determined. The Bureau of Transportation Statistics is working to correct this data-reporting problem.

[11] Tests also were run that used the square root of the distance in place of the distance. The "square root of proportionate distance" method produced the same type of bias as the proportionate distance method, although the severity of the bias was somewhat reduced.

[12] For formulas detailing the method of implicit imputation, see the appendix.

[13] Copies are available from the authors upon request.

[14] At least one researcher has identified such a bias. (See Jan De Haan, "Generalised Fisher Price Indexes and the Use of Scanner Data in the Consumer Price Index (CPI)," *Journal of Official Statistics*, March 2002, pp. 61–85.)

[15] A description of the BLS estimation method is available on the Agency's Web site at **http://www.bls.gov/cpi/cpifacaf.htm**.

[16] See, for example, Robert B. McClelland, "Evaluating Formula Bias in Various Indexes Using Simulations," 1996; on the Internet at **http://www.bls.gov/ore/pdf/ec960140.pdf**; or Brent R. Moulton, "Bias in the Consumer Price Index: What Is the Evidence?" BLS working paper no. 294, 1996; on the Internet at**http://www.bls.gov/ore/pdf/ec960170.pdf**.

[17] For a discussion of the change in index formulas, visit **www.bls.gov/cpi/cpiadd.htm#4_1** on the Internet.

[18] See Anderson and Leonard, "Passenger Price Trends."

[19] For production purposes, the Bureau of Transportation Safety will define a new class of service comprising services provided by carriers that offer only one class of service. Thus, itineraries flown on carriers, such as Southwest, that report all their seats as first class will not be categorized as first class.

[20] Note that, for all points of origin outside the United States, the O&D Survey indexes cover only itineraries incorporating some U.S. component.

[21] Visit **http://146.142.4.24/servlet/urveyOutputServlet?series_id=lausm85200003** for the Tucson employment and unemployment figures.

[22] For details, see Janice Lent, "Chain Drift in Experimental Index Series," *Proceedings of the Section on Survey Research Methods, American Statistical Association*, Alexandria VA, Joint Statistical Meetings, San Francisco, CA, Aug. 8–12, 2003 (published on CD only).

APPENDIX: Formulas for price index estimation

Price index estimators. A measure of relative change in the price of a particular item j between periods 1 and 2 is the price ratio, $p_{j,2}/p_{j,1}$, where $p_{j,t}$ represents the price of item j at time $t \in \{1,2\}$. Because each quarterly O&D Survey sample is independently drawn, it is impossible to match each individual itinerary with an identical one in the following (or previous) quarter and compute individual price ratios. This article therefore presents a method for computing unit-value indexes for itineraries (or, in the second stage, segments) within each unit value category $c \in C_{1,2}$, where $C_{1,2}$ is the collection of categories populated by sample units in quarters 1 and 2. (See text for definitions of categories.) For simplicity, it is assumed that prices are available for all observations in the data set.

Let $q_{j,t}$ be the quantity of item j purchased in period t. For the O&D data, the item is an itinerary and $q_{j,t}$ is the number of passengers flying the same itinerary at the same fare. (The variable denoting the number of passengers is included in each O&D Survey itinerary record.) Because the O&D sample is self-weighting, we may directly apply the standard population price index formulas. Let

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

The unit-value index estimator for category c is defined as

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

In words, the unit-value index is the average price paid for an item in category c during period 2, divided by the average price paid for an item in category c during period 1.

Once the unit-value index estimates are computed for all $c \in C_{1,2}$, they are treated as price ratios in the standard index formulas. For $t \in \{1,2\}$, let

$$w_{c,t} = \frac{\sum_{j \in c} p_{j,t} q_{j,t}}{\sum_{c' \in C_{1,2}} \sum_{j \in c'} p_{j,t} q_{j,t}}$$

be the expenditure share for category $c \in C_{1,2}$ during period t. (Note that $w_{c,t}$ is dependent on $C_{1,2}$ and would be more clearly denoted by $w_{C(1,2),t}$. For ease of notation, however, this dependence is left implicit; note also that all indexes described in this appendix indicate price changes between periods 1 and 2.) Then the following indexes may be estimated for all desired categories of aggregation $C_{1,2}$:

Laspeyres index:

$$\hat{L} = \sum_{c \in C_{1,2}} w_{c,1} u_{c,1,2}.$$

Paasche index:

$$\hat{P} = \frac{1}{\sum_{c \in C_{1,2}} \left(w_{c,2} / u_{c,1,2} \right)}.$$

Fisher index:

$$\hat{F} = \sqrt{\hat{L}\hat{P}}.$$

Jevons (or geometric mean) index with weights from period 1:

$$\hat{G} = \prod_{c \in C_{1,2}} u_{c,1,2}^{w_{c1}}.$$

Törnqvist index:

$$\hat{T} = \prod_{c \in C_{1,2}} u_{c,1,2}^{(w_{c1} + w_{c2})/2}.$$

Implicit imputation through unit-value indexes. When some prices are missing from the data set, they may be implicitly imputed through the computation of unit-value indexes. As noted in the text of this article, such imputation occurs in the computation of second-stage unit-value indexes. Let c' be the set of observations in category c with nonmissing price values, and let

$$\overline{p}_{c,t} = \frac{\sum_{j \in c'} q_{j,t} p_{j,t}}{\sum_{j \in c'} q_{j,t}}$$

be the average of the nonmissing prices in category c. Then the unit-value index for category c is defined as

$$u'_{c,1,2} = \frac{\overline{p}_{c,2}}{\overline{p}_{c,1}}.$$

The weight for category c in time t is

$$w'_{c,t} = \frac{\overline{p}_{c,t} q_{c,t}}{\sum_{c \in C_{1,2}} \overline{p}_{c,t} q_{c,t}},$$

where $q_{c,t}$ is the total quantity of items in category c at time t (including those items with missing prices). The Laspeyres, Paasche, Fisher, Jevons, and Törnqvist indexes are then calculated from their given formulas, but with $q_{c,t}$ and $w'_{c,t}$, for $t \in \{1,2\}$, replacing $u_{c,1,2}$ and $w_{c,t}$, respectively.

Using second-stage unit values to compute indexes for first-stage categories. The second-stage unit-value index $u_{c,1,2}^{(s)}$ for a first-stage category[1] c is calculated as follows:

Let K_c denote the collection of second-stage (segment-level) categories k corresponding to category c. For a given quarter t, let

$$\overline{p}_{k,t} = \sum_{l=1}^{q_{k,t}} \frac{\hat{p}_{l,k,t}}{q_{k,t}},$$

where $q_{k,t}$ is the number of passenger itinerary segments (possibly from itineraries in different first-stage categories) in second-stage category k for quarter t and, for $l = 1,...,q_{k,t}$, $\hat{p}_{l,k,t}$ is the imputed price of segment l in category k. Then

$$u_{c,1,2}^{(s)} = \frac{\sum_{k \in K_c} \overline{p}_{k,2}}{\sum_{k \in K_c} \overline{p}_{k,1}}.$$

As noted in the text, a second-stage category k may correspond to many first-stage categories c; that is, it may be that $k \in K_{c_1}$ and $k \in K_{c_2}$, where $c_1 \neq c_2$. Note also that $u_{c,1,2}^{(s)}$ is a Fisher index indicating price change from period 1 to period 2 for itineraries in category c, with the quantity associated with each $\overline{p}_{k,t}$ set equal to unity and the segment-level unit-value indexes serving as price relatives. That is,

$$u_{c,1,2}^{(s)} = \sum_{k \in K_c} \left(\frac{\overline{p}_{k,1}}{\sum_{k \in K_c} \overline{p}_{k,1}} \right) u_{k,1,2},$$

where

$$u_{k,1,2} = \frac{\overline{p}_{k,2}}{\overline{p}_{k,1}}.$$

To compute the Fisher indexes shown in table 1, the $u_{c,1,2}^{(s)}$ were aggregated with the use of the Fisher formula, with expenditure share weights $w_{c,t}$ computed from itinerary-level data, as described in the text.

Note to the appendix

[1] See text for a list of the variables that define a first-stage category.