THE USE OF SCANNER DATA IN RECONCILING
TIME-SERIES (CONSUMER PRICE INDEX)
AND
GEOGRAPHIC (PLACE-TO-PLACE) PRICE COMPARISONS

William J. Hawkes

Presented at the Fourth Conference of the
International Working Group on Price Indexes
Washington D. C., U. S. A.
April 22-24, 1998

ABSTRACT

Prices for an array of goods and services in a particular geographic area at a particular point in time can be compared against a corresponding array of prices, either in the same geographic area at some earlier point in time, or for a different geographic area at the same point in time. Current practice in most statistical agencies is to produce time-series price Indexes, on the one hand, and interarea price comparisons, on the other, using procedures that are not designed to maximize consistency between these intertemporal and interspatial measures. The recent availability of scanner and other point-of-sale data for certain expenditure categories in certain outlet types in many industrialized countries makes it possible to develop procedures for both geographic and time-series price comparisons that will enhance their quality and consistency. This paper will describe such procedures and discuss how they can be utilized at both the individual country (interarea cost-of-living measures) and international (purchasing power parity) levels.

I thank Bill Cook and Mary Kokoski for their comments on an earlier draft of this paper, and Marshall Reinsdorf and Kimberly Zieschang for discussions on this topic.

1. Introduction

A useful definition of a price index is given by Peter Hill in chapter 16 of Systems of National Accounts (Hill, 1993, p.381):

> "A price index is an average of the proportional changes in the prices of a specified set of goods and services between two periods of time."

Hill goes on to state that:

> "It is possible to compare prices and volumes between countries using the same general methodology as for intertemporal comparisons within a single country."
>
> Peter Hill [1993; 393]

Although Hill is writing of price comparisons between countries, individual cities or regions within a country can be compared as well. Between-country price indexes are usually interpreted as "purchasing power parities," and within-country indexes as "interarea cost-of-living comparisons"; but there is in fact little methodological or formulaic difference between the two. Thus Hill's "price index" definition might well be re-stated as "a price index is an average of the proportional differences in the prices of a specified set of goods and services between two price regimes", where "regime" is used to denote a particular place and time.

Purchasing power parities, however, differ from intra-country measures in two important respects: 1) they measure the relationship between different currency denominations rather than cost-of-living differences and 2) they are often used as deflators for nominal product, income, or expenditures in order to compare "real" quantities or volumes across countries. Much of the literature on purchasing power parities, accordingly, focuses on these "real volume" output measures (or individual country shares of "real volumes"), rather than on the purchasing power parities themselves. Balk (1996, p. 210) for example, notes that Diewert's (1987) and his tests for appropriate multilateral geographic comparisons:

> "…emphasize the fact that the primary purpose of any international comparison is to make volume comparisons. Purchasing power parities play only an intermediate role, namely as currency convertors. Consequently, the tests are phrased in terms of volume shares,…"
>
> Bert Balk [1996;210]

In general, neither time-series nor geographic price index comparisons are transitive; i. e., index change in prices between periods 1 and 2, multiplied by the index change between periods 2 and 3, will in general not equal the index change calculated directly between periods 1 and 3; nor will the price difference between areas A and B, multiplied by the difference between areas B and C, equal the difference between areas A and C. As Balk (1995) points out, Frisch (1930), Eichhorn and Vollner (1978) and others have shown that price indexes involving the use of period-specific (or country-specific) quantities for more than two periods (or countries) cannot be transitive[1] while still meeting other desirable objectives or tests.

Temporal transitivity is usually assured for official Consumer Price Indexes by construction; i.e., month-to-month indexes are always chained across successive months, while "direct" measures of month-to-year ago change, for example, are simply not computed (or not published). Geographic transitivity can be satisfied by various methods[2] of "averaging" or "pooling" prices and quantities across counties in such a way that the bilateral country-to-country price ratios are all distorted slightly in order to make them mutually consistent – much as the Pythagorean musical ratios of perfect thirds, fourths, and fifths are all distorted slightly by the device of "equal temperament" in order to make all key signatures isomorphic and harmonious throughout the scale. The resulting "loss of characteristicity" [in Drechsler's (1973) formulation] from the bilateral (i. e., country-to-country) comparisons is similar to the dissonance induced into the pure tonal ratios by the "equal temperament" solution.

Since the current period prices in a particular geographic area can be compared with prior period prices in the same area or with current period prices in different area, the question immediately arises as to whether temporal and spatial measures are mutually transitive. This involves simultaneous comparison of prices across time and across geographic areas. For example, consider the following question:

Did prices in New York between 1996 and 1997 increase faster than they did in Los Angeles? This question can be answered in two different ways.

1) Compare the CPI trends between 1996 and 1997 for Los Angeles and New York.

2) Compare the interarea price differences between Los Angeles and New York for 1996 and 1997.

Conceptually, (1) and (2) purport to measure the same thing, namely the differential price change between New York and Los Angeles:

---

[1]In his recently published doctoral thesis, Bernhardt Olt (1996;46-50) considers several "weak circularity" tests, and goes on to discuss whether "approximate" or "near" circularity can be achieved, for a Fisher, index to a first-order approximation. His conditions, however, seem highly restrictive. Olt also cites Diewert (1993) on Fisher's claim of approximate circularity. Fisher (1922), having discovered intransitivity, tries to turn it into a virtue; Samuelson and Swami (1974), somewhat cavalierly, note that "one must not expect to be able to make the naïve measurements that untutored common sense always longs for."

[2]Balk (1996) discusses 10 different methods for interarea price comparisons. Each is unsatisfactory with regard to at least one criterion. Diewert (1996) has also discussed 10 alternative measures in a recent comprehensive survey paper, identifying 4 of these measures as "best".

$$\frac{I_{97/96,NY}}{I_{97/96,LA}} = \frac{I_{NY/LA,97}}{I_{NY/LA,96}}$$

However, in general this conceptual identity does not hold in actual practice, with the geographic and time-series view neither consistent or transitive. Curiously, there has been relatively little attention paid to this problem[3], which will be the central topic of this paper.

The paper proceeds as follows: Section 2 describes differences between traditional geographic and time-series views of price comparisons that restrict their comparability and, hence, their rectangular transitivity, beyond the mathematical or logical constraints described by Eichhorn and Vollner. Some of these factors have been described previously, but others are newly identified by this paper. "Real life" examples of such intransitivity are provided. Section 3 demonstrates how scanner data can be used to quantify the effect of several of these factors. Section 4 illustrates how scanner data can be reconfigured so as to help reduce these effects and to enhance the rectangular transitivity of geographic and time-series results – and in the process, to improve both the geographic and the time-series measurements themselves. Section 5 concludes.

2) Differences between Geographic and Intertemporal Price Indexes

As has been noted by Balk (1996), Diewert (1996), Hill (1993), and others, intertemporal and interspatial price comparisons, especially those involving several time periods or several geographic areas (i.e. those characterized as "multilateral" in scope) differ in several important respects. Balk, for example, notes that:

> "-- Time proceeds continuously whereas the number of countries involved in a comparison is fixed.
> -- Unlike time periods, countries do not have a natural order.
> -- In an intertemporal comparison, the time periods considered are usually of the same size (one compares months with months, years with years, etc.). Countries, however are by nature not equally "important" (with respect to area, population, economic potential, etc.).

---

[3]Caves, Christensen, and Diewert (1982) point out the applicability of their formulation to combinations of time-series and cross-sectional data, citing one example of a multilateral productivity index for a panel of airline firms. Sultan Ahmad (1997) has written a Working Paper for the World Bank on "Intertemporal and Interspatial Comparisons of Income". Kokoski, Moulton, and Zieschang (1996) mention discrepancies between GDP measures obtained via purchasing power parity rates and those obtained by using implicit price deflators, but provide no specific case studies. With these exceptions, substantive references to this problem (especially with examples) are hard to find.

> -- More than in intertemporal comparisons, there is in international comparisons a strong desire to aggregate the entities and use such aggregates also in comparisons (e.g., the European countries compared to the European Community as a whole, the European Community compared to the United States)." Bert Balk[1996; 200]

In light of the foregoing list, it is not surprising that geographic (Place-to-Place) and intertemporal (Time-Series) price comparisons have evolved to meet the needs of somewhat different constituencies. Partly for this reason, achieving consistency between the two views has rarely been viewed as a primary objective by producers or users of either set of data. Further, most official price indexes have been designed to measure price changes across time rather than price differences within or between countries. Accordingly, official intertemporal Price Index programs have generally been well-funded, while interarea price comparisons have either been left to the private sector, forced to "make-do" with leftovers from intertemporal Indexes; or severely under-funded using a makeshift separate survey framework.

Nonetheless, a number of major interarea price comparison programs do exist, including the international Purchasing Power Program administered by Eurostat within the European Community, and by the Organization for Economic Cooperative Development (OECD) across the entire set of OECD countries. In the U. S., the Bureau of Labor Statistics formerly conducted a "family budget" program which reported costs for specified family budgets at various utility levels across major geographic regions in the U. S. This program was discontinued around 10 years ago for a variety of budgetary, conceptual and technical reasons.

The U. S. Bureau of Labor Statistics has also published, for many years, a series on "Average Prices" for a number of food items, as well as gasoline and utilities, both nationally and for major regions of the country. Comparisons of the resulting estimates with CPI historical series have been presented and discussed in a series of papers by Reinsdorf [1992,1993,1994].

More recently, the BLS has published a number of papers on "experimental" place-to-place indexes, [Kokoski (1991); Primont and Kokoski (1991), Kokoski, Cardiff, and Moulton (1994); Kokoski, Moulton, and Zieschang (1996)] drawing upon (as is also true for the "average Food Prices" program cited above), prices that happen to be collected for the Consumer Price Index program. Because each item in each outlet for which data are obtained in the CPI is explicitly matched across time, while no effort is made within the CPI framework to match items across outlets or across geography, such interarea price comparisons are difficult to make. Nonetheless, considerable interest has been expressed, by users, in place-to-place price comparisons; and these new experimental indices have been developed in recognition of this user interest.

Discrepancies between time-series and place-to-place price indexes can be found in any comparison of the two data sets. An example drawn from the international Purchasing Power Parity program is shown in Table 1. Purchasing Power Parities are currently obtained once every three years for all OECD countries. These figures are then trended forward between survey years by using time-series Price Indexes in each country. The resulting "implied" PPP deflators for Gross Domestic Product 1993, by country, differ from the "actual" PPP deflators by an average of 2-3% percentage points. Given the procedural differences between the two series (for example, PPP consumer prices are generally collected only for the principal city in each country and only for a limited period of time), such differences are not surprising.

Following George Steiner (1978), I want to draw a distinction between "contingent" and "ontological" difficulties. "Contingent" difficulties, in a price index context, are those that can be overcome by hard work and more money; "ontological" difficulties are those issues, deeply rooted in the nature of price index theory, that are mathematically or conceptually intractable.

The conflict between transitivity and other desirable properties of a price index, described above, is an example of such an "ontological" difficulty. But simply to prove the impossibility of "perfect" transitivity is not enough. A more useful question to address is which bilateral price index formula comes closest to achieving transitivity when introduced into a multilateral context, under what circumstances is this minimization likely to happen, and how large are these discrepancies likely to be? More empirical, as well as theoretical, work is needed on these important issues.[4]

One factor that has not been discussed in the literature is the difference in the procedures generally used in constructing elementary aggregates for time-series and geographic comparisons. Typically, time-series aggregation of individual observations to the "item stratum" level starts with price and quantity measures at the item-outlet level to develop "elementary aggregates" by area and commodity i.e. price ratios are computed between successive months for a specific item of a commodity in a specific outlet and averaged in some way to construct an "elementary aggregate". In comparing prices between geographic areas, however, there is no natural "match" between individual outlets in one area and individual outlets in another area. Thus one is forced to aggregate prices and quantities across outlets, in effect constructing a "unit value" price for each item at the area level, in contrast to the binary "weaving together" of individual outlet prices across time using Laspeyres, weighted geometric means, or other formulas. [See Dalen (1992; 135) for a characterization of these two different approaches to construct "elementary aggregates".]

A second example is in aggregating the individual item[5] prices. Again, in time-series practice, individual items are also matched across time. But in an interarea comparison,

---

[4] Fisher (1992) devotes several pages to these questions. His term for such "ontological" difficulties is "proper" or "legitimate deviations".

[5] There is some confusion in the literature as to what is meant by an "item", a "unique item", a "specification item", a "variety", an "SKU" (stock-keeping unit), and a commodity. Without intending to add to the confusion, I use "item" to denote a particular entity, defined by brand, size, and other characteristics, that is a member of the set that will be called a "commodity". In this taxonomy, Taster's Choice Decaffeinated Freeze Dry Coffee in the 10 ounce size is an "item", while coffee is a "commodity". I leave until later in this paper the very important question of how many intermediate levels one might wish to interpose between "item" and "commodity", and how homogeneous in characteristics items constituting a "commodity"

individual items often cannot be matched. Thus in interarea index construction the usual procedure is simply to compute an "average price" at the commodity level (perhaps using hedonics or other factors to adjust for quality and size differences). Thus a unit value is also constructed across the item dimension for purposes of interarea, but not intertemporal, comparison. As a final example, the hedonic adjustments for various item characteristics that, in the U. S., have been developed for interspatial comparisons are not used in the Consumer Price Index. Such differences all serve to reduce comparability between the two measures.

The following section illustrates these issues with an extended example using scanning data.

3) Quantifying sources of difference between intertemporal and interspatial indexes using scanning data

As described in several papers [Bradley, Cook, Leaver, and Moulton (1997), Dalen (1997), DeHaan and Opperdoes (1997), Hawkes (1997), Ilver, Ionnides, and Haworth (1997) and Scobie (1997)] presented or distributed at last year's conference of this group at the Hague, the recent advent of scanner-based and other point-of-sale data in many countries offers the promise of a substantially enlarged data set for which weekly prices and quantities are available for each item in certain expenditure categories in certain outlet types.[6]

This analysis makes use of the same U. S. Coffee data set[7] described earlier by Reinsdorf (1996), and Hawkes (1997). Two geographic markets (Metropolitan Chicago and Metropolitan Washington) and two months (December 1993 and December 1994) were selected for study. The data set summarized in this paper involves a total of 150 supermarkets (individually tabulated but with their identities undisclosed) for some 940 individual coffee Universal Product Codes found in one or both geographic markets in one or both years, for a total of around 150,000 observations (items x outlets x weeks). Prices and quantities were tabulated for each item in each week. For reasons previously described by Reinsdorf and Hawkes, weekly data were aggregated to 4 week summaries designated "December 1993" and "December 1994."

The number of separate UPC's in the data set—some 940 individual coffee items, is substantial. Each UPC is distinguished by brand, sub-brand, size, grind, caffeine content, container, form (i.e. whole bean, ground, spray dried, freeze-dried, flavor, yield (generally based on number of cups per ounce) and specialty type (e.g. Kenyan, Jamaican, etc.). These

---

should be. In general, a Universal Product Code is used to designate one particular item, as defined above.

[6] See also papers by Bradley (1996), Bradley and Verdon (1996), Silver (1995), and Magnien and Pougnard (1998) on uses of scanner data in Consumer Price Indexes.

[7] The data were provided by the A. C. Nielsen Company and, in aggregated form, are part of its ScanTrack ® service.

classification criteria are consistent with Uniform Code Council coding standards for product handled by the U. S. grocery industry.

It is interesting to note that only 200 of the 940 items were found in both markets in both years, although these 200 items account for around 65% of total dollar expenditures (Table 2 and 3). A much higher proportion of both items and expenditures are matched within a particular geographic market between both years than between markets in a particular year. Only 3% of expenditures were for items that were “new” or “discontinued” between Dec. 1993 and Dec. 1994 at the geographic market level, while more that 30% of expenditures were for the items that were found in one geographic area but not the other. This is one obvious reason for differences between geographic and time-series “Fisher Index” measurements, which are based on identical items across time periods or between markets.. Another obvious reason, as noted earlier, is the need to compute a “unit value” at the item level across outlets for purposes of price comparisons across geographic areas, but not for purposes of comparison across time.

The effect of these differences is shown on Table 4. If the time-series (year ago) price changes are computed, at the individual outlet level, for each item matched in the same outlet across time (which is the usual CPI procedure), the resulting “Fisher” index estimate of relative price change between Chicago and Washington from 1996 to 1997 differs by 5.0% from the corresponding relative price change computed from comparing Chicago and Washington prices (also using a Fisher formula) across the items matched between the two areas. If we also compute the time-series index for each area by aggregating across outlets using unit values for each individual item, the 5.0% difference narrows to 3.0%. However, restricting both time-series and geographic data sets to the 200 common items found in both areas in both years causes the discrepancy to increase again, this time to 5.5%. Clearly, limiting the data set to the artificial group of items that happen to be in common to both time-periods and geographic areas not only does not improve transitivity; it worsens it. In fact, making the data set artificial in this manner serves to distort the effect of intertemporal cross-elasticities by deleting items that are part of each city’s analysis for no better reason than that they happen not to be sold in the other city.

Table 5 provides some insight into possible alternative approaches. The first approach is to see whether another Index form might help. It turns out that although a Tornqvist formulation reduces the discrepancy, and the Sato-Vartia II[8] formulation reduces it still further, the improvement is marginal and suggests, again, that "superlative" and "ideal" indexes are, in actual practice, all good approximations of each other. Help will have to come from some other source than a better formula.

As Table 5 above demonstrates, intransitivity can be set to zero by giving up some "characteristicity", as termed by Drechsler (1973). While there are several choices for forcing "rectangular transitivity"[9] (especially in this simple, double-bilateral example), I have elected to show the "CCD" outcomes [due to Caves, Christensen, and Diewert (1982)], and also the new, weighted "KMZ" outcomes using a method proposed by Kokoski, Moulton, and Zieschang (1996). In this particular instance, the two procedures lead to very similar results. Where individual geographic weights differ, the KMZ method has the attractive property of reflecting the "irrelevance of tiny countries" (Balk, 1996) and of providing a closer approximation to additivity for individual commodity or product components summing to a market basket total.

Despite the appeal of such multilateral adjustment methods, one might wish to minimize disturbance to the original bilateral price ratios by making these ratios themselves more consistent across time and space. One way to do this is to improve the match of items across geography; and this is the subject of the next section.

4) Improving Transitivity with Scanner Data

The UPC item code structure described earlier is somewhat more detailed that that used by the U. S. Bureau of Labor Statistics in determining which particular item is to be priced in a particular store. The BLS standards do not distinguish among grinds (e.g. drip, percolator, etc.) or flavors (e.g. vanilla, raspberry, Irish cream, etc.) unless grinds or flavors are price-determining characteristics at the individual outlet.

---

[8] In last year's paper, I expressed a belief that transitivity ought to be enhanced by the use of a Sato-Vartia II price index formula. This formula resembles a Tornqvist index except that it uses a particular form of logarithmic, rather than arithmetic, mean of item shares across periods or areas being compared. The Sato-Vartia II formula is attractive in two respects:

A) it downweights items whose shares differ widely between the two periods or markets,
B) for CES preferences, it is price and quantity permutation-independent at the outlet level. This implies that the Dalen-drawn distinction between elementary aggregation approaches ought to matter less, for the Sato-Vartia, than for other index forms.

It is true, in this data set, that the Sato-Vartia II, in fact, outperforms both the Fisher and the Tornqvist. But the difference is so slight that it is not worth the extra computational effort.

[9] For example the procedures described by Geary (1958), Khamis (1972), and Kravitz et al (1975), (1978), (1982).

Accordingly, an attractive way to reduce the number of separate items (and to increase the proportion of matched items) is to apply the same rules to the set of UPC items as that used by BLS to select the "unique items" to be priced.

The results of this exercise are shown on Table 6. The original 940 item data set is reduced by around 30 by combining items distinguished only by grind but not by price. Collapsing items across flavors (where flavor is not a price-determining factor) compresses the data set by another 230 items. These two actions result in a reduced set of around 680 items.

A third step opportunity for data reduction reflects the fact that many retailers have their own "house brand" or "controlled" label. The name of the specific house brand will vary from one retailer to another, but many commonalties in product attributes exist across retailers. Of the 680 items identified through the two procedures just described, 181 are "controlled brand" merchandise. If these are matched according to exact characteristics, this number drops to 107; and much of this reduction takes place across cities, since most of the retailers in the data set do not operate in both geographic areas.

Linking "bonus pack" and "pre-priced" items with their parent codes reduces the data set further, down to a new total of 599 items. However, even with 599 items there are still numerous non-matches, especially across the two cities. Two opportunities for further reduction, both in the spirit of existing BLS procedures, come to mind.

The first is the "time-series" CPI matching procedure used by BLS to obtain a "comparable substitute" for an items being priced each month in an outlet when the item is no longer handled by the outlet. This procedure is described by Moulton and Moses (1997). Under this procedure, a "comparable substitute" is an item which differs from the previously designated item only with regard to one or more particular attributes that have a low ranking in a hierarchical "disaggregation" process that distinguishes individual items by successively less important price-distinguishing characteristics. For example, consider a case where three items of different brands (X, Y, and Z) of coffee, all caffeinated, all in cans, all 13 ounce, all "regular yield", all from the same country of origin, and all selling at $3.19, are all handled by the same outlet. If this particular item of Brand X has been the item designated for pricing each month, and if Brand X is then discontinued by the store, then the corresponding item in Brand Y or Z may be used as a "comparable substitute." In this instance, the difference between the new item's current price and the old item's previous price is regarded as a "real" price change. Obviously the same reasoning could be applied in comparing items across geographic areas, although "price matching" across outlets is more complicated than it is within outlets. Nonetheless, this "specification matching" algorithm is adaptable to interarea and, for that matter, to international item matches. In fact, since with scanning data historical prices are available for every coffee item in every sampled outlet, the "comparable substitution" procedure should be applicable to a higher proportion of food items than it is now in the CPI [slightly over 50% of all item substitutions for 1995 as reported by Moulton and Moses (1997)].

In some instances CPI time-series replacement items are grafted into the old item via "direct" price adjustment, as, for example, when a new size replaces an old one; and this procedure, too, can be replicated using scanning data.

A second approach is the "hedonic" adjustment procedure used by the U. S. Bureau of Labor Statistics in its experimental interarea price comparison program. At the Conference on Research on Income and Wealth in 1996, Kokoski, Moulton, and Zieschang presented an important paper that, unfortunately, seems not to have been widely distributed outside the Bureau of Labor Statistics. KMZ (1996) characterize their work in the following abstract:

> "We derive a general form of Tornqvist multilateral (transitive place to place index numbers and a new variant of regression methodology for imposing transitivity while minimally adjusting the initial system of bilateral index comparisons. We show that when several levels of item aggregation are to be published in a system of Tornqvist interarea parities, the adjusted, transitive Tornqvist parities at each level of aggregation preserve the aggregation rule in the unadjusted data. Finally, the method incorporates characteristics-based, hedonic quality adjustment as an integral feature. We apply the method to a subset of commodity price and expenditure data for the 44 areas of the United States covered by the Consumer Price Index. In closing, we also discuss an application of the method that makes time series and geographical comparisons consistent with one another, and note that it permits decentralization of calculation in a way that may have distinct advantages for compiling international price comparisons."
>
> Kokoski, Moulton, and Zieschang [1996;ii]

In this procedure, characteristics and price data for, say, a twelve months period are pooled to develop regression-based "quality equalization" factors for various product attributes. This procedure can be thought of as an extension in continuous form of the dichotomous "comparable substitution" procedure, which, in effect, assigns a pseudo-hedonic coefficient of either zero or one on the replacement item depending on whether it is judged a "comparable" or "non-comparable" substitute. But the hedonic procedure is more flexible, and it also provides an automatic way for "new" or "unmatched" items to enter into both interarea and intertemporal comparisons.

However, the hedonic approach is not without difficulties of its own, notably to the extent that the hedonic coefficients are based on subjective consumer preferences as well as objective attributes and, as such, are likely to shift over time. It is also possible for the hedonic scores to be correlated across characteristics, resulting in biased corrections for individual items. Accordingly, a more robust procedure may be to use the hedonic coefficients to identify price-determining characteristics and then to combine items into clusters or groups according to these characteristics. In the "grouped data" configuration presented on Table 7, I have extended the "comparable substitution" procedure to classify roasted coffee items into around 50 mutually exclusive categories, where each category classifies coffee items based two or more levels of each of the following characteristics:

1) Caffeine vs. Decaffeinated
2) "Specialty" coffees, roasts, and countries of origin
3) Whole bean versus ground coffee
4) "Extended yield" and regular yield categories
5) Package size

This grouping procedure ensures that most coffee items can be used in both geographic and time-series comparisons. The resulting groups are, in effect, then considered market segments or sub-commodities, summarized across time and geography using Tornqvist formulations. The "intransitivity" between intertemporal and interspatial price comparisons is, under this procedure, reduced from nearly 6.0% to 0.6%.

While I have not as yet applied the KMZ hedonic adjustment procedure to the same data set, I expect similar results to the "grouped category" procedure described above.[10]

---

[10] One difference between interarea comparisons and those involving both areas and time periods is the following: in interarea comparisons, all pairs of area comparisons can be thought of as equally important, or at least of an importance commensurate with the economic importance of these areas. But in evaluating geographic and time-series results, the "diagonals"; i.e. the relationship between prices for an area "A" in period 1 and area "B" in period 2, are never examined. Thus "closed transitivity" is important in a triangular sense for interarea comparisons, but only in a retangular sense for combined interarea and intertemporal comparisons. By "closed transitivity" I mean an eventual return to a starting point. For example, in interarea comparisons, if we require that $(I_{A,B})(I_{B,C})(I_{C,A})=1$, this is an example of "closed (triangular) transitivity". In combined interarea and interspatial comparisons, we require $(I_{1,A})(I_{A,B})(I_{B,2})(I_{2,1})=1$ for rectangular closure. But this has a different solution if we also impose "triangular" transitivity on the data than if we do not. For this reason, making the "triangular" comparisons "closed transitive" is necessary even if the diagonals are never used. The CCD and KMZ solutions do satisfy triangular as well as rectangular closure for combined time-series and geographic comparisons.

5) Summary

The dramatic reduction in rectangular intransitivity resulting from the construction of similar "item groupings" may be an artifact of this particular data set, and needs to be replicated across other categories, geographic areas, and time periods. In any event, three important questions remain:

1) Is the "grouped item" procedure to be preferred over the "hedonic adjustment" procedure? Obviously there is no simple answer to this question. Computing and re-evaluating hedonic adjustments would seem a formidable task to complete each month in a production environment. If the hedonics are carried over from prior periods, one might argue that the hedonics themselves have lost some "characteristicity" to the extent that they make use of data pooled across time and geography.

   An attractive alternative might be to use the implied hedonic adjustments on each level of each characteristic as a way of defining the item groups; i.e. of determining which "comparable substitutes" are in fact most comparable, and then extending the solution of "comparable substitutes" to define the item groups. These item groups, defined in advance, could then be used each month in a production environment.

2) Even if the "grouped item" method produces "approximate transitivity", should the "grouped item" discrepancy then be further reduced to zero by the computation of pooled shares and prices as described by CCD and KMZ? Moreover, the foregoing discussion considers only intra-commodity transitivity, even through between-commodity intransitivity is also a problem.[11] My suggestion is that the KMZ (or CCD) transitivity procedure should be used across areas but not across time periods. The reason is that, for policy reasons, time-series (CPI) estimates are produced in "real time" and rarely revised. When this constraint does not always hold, as, for example, with GDP deflators, the fact that time unfolds continuously means that intertemporal adjustments, in any case, must still be made across what KMZ term a "moving window".

3) How many "sub-commodities" does "coffee" (or any other commodity) comprise? The BLS coffee item-stratum is divided into 2 "entry-level items" – roasted versus instant. This is a useful first cut, but clearly does not distinguish among important market segments, such as decaffeinated coffee, freeze dried coffee, and "gourmet" or specialty brands generally sold or offered in whole-bean form. Yet 940, or even 599 individual items, is too many to match across only 2 areas and only 2 points in time. The 50 groups that I have designated for Roasted Coffee represent one attempt at a

[11] Fisher (1922) claims that, based on his data set of 36 commodities across 6 years, between-commodity intransitivity across time is of minor consequence – typically around 0.3%.

workable compromise. But clustering criteria need to be made explicit for all categories, and this is not an easy task.[12]

The Marketing Research industry has also sought to address this issue, with modest success. Possible criteria for marketing segmentation include price levels (although “special” or “sale” prices may need to be excluded for this purpose), consumer item-or-brand-switching behavior, characteristics of purchasing households, and physical characteristics of the product itself. Some references in the Marketing Research literature on this topic include Foekens, Lieflang, and Wittink (1997), Blattburg and Wisnewski (1989), Cooper (1988), Novak (1993), and Siddarth (1995). Novak applies several different market structure models to an instant coffee data set originally analyzed by Grover and Srinivasan (1987). Understandably, the Marketing Research literature focuses on brand but not size as a characteristic, while the “hedonic” approach largely disregards brand identity as a price-determining characteristic, except for distinguishing “store brands” from “national brands”.

Scanner data will help to shed light on these important issues, and will facilitate the continued development of procedures designed to improve both the quality and the transitivity of geographic and time series price comparisons.

[12] Ideally, one would hope to choose sub-categories, with stable boundaries across time, such that homogeneity (and substitution) within each sub-category is maximized.

REFERENCES

Ahmed, S. (1993), “Intertemporal and Interspatial Comparisons of Income,” Policy Research Working Paper WPS 1157, World Bank, Washington D. C.

Balk, B. M. (1995), “Axiomatic Price Index Theory: A Survey,” International Statistical Review, 63, 1, 69-93.

Balk, B. M. (1996), “A Comparison of Ten Methods for Multilateral International Price and Volume Comparison,” Journal of Official Statistics, Vol. 12, No 2, 199-222.

Balk, B. M. (1994), “On the First Step in the Calculation of a Consumer Price Index,” paper presented at the International Conference on Price Indices, October 31-November 2, Ottawa: Statistics Canada.

Blattberg, R. C., Wisniewski, K. J. (1989), “Price Induced Patterns of Competition,” Marketing Science 8, 291-309.

Bradley, R. (1996), “The Use of Scanner Data as a Means to Reduce the Mean Squared Error in the CPI,” paper presented at the American Statistical Association Annual Meeting.

Bradley, R. and Verdon S. (1996) “Time Substitution and Uncertainty Bias in the COL,”U. S. Bureau of Labor Statistics memorandum.

Bradley, R., Cook B., Leaver S. and Moulton B. (1997), “An Overview of Research on Potential Uses of Scanner Data in the U. S. CPI.”*

Caves, D. W., Christensen L. R. and Diewert W. E.(1982), “Multilateral Comparisons of Output, Input, and Productivity Using Superlative Index Numbers,” The Economic Journal 92, 77-86.

Cooper, L. G. (1988), “Competitive Maps: The Structure Underlying Asymmetric Cross Elasticities,” Management Science 34, 707-723.

Dalen, J. (1992), "Computing Elementary Aggregates in the Swedish Consumer Price Index,” Journal of Official Statistics 8, pp. 129-147.

Dalen, J. (1997) “Experiments with Swedish Scanner Data,” *.

Diewert, W. E. (1987), "Index Numbers,” in The New Palgrave: A Dictionary of Economics, Vol. 2, J. Eatwell, M. Malgate and P. Newman (eds.), London: The MacMillan Press, pp. 767-780.

Diewert, W. E. (1993), "The Early History of Price Index Research," in Essays in Index Number Theory, Vol. 1, W. E. Diewert and A.O. Nakamura (eds.), Amsterdam: North-Holland.

Diewert, W. E. (1995), "Price and Volume Measures in the System of National Accounts," Discussion paper No. 95-02, Department of Economics, University of British Columbia, Vancouver, Canada.

Diewert, W. E. (1996), "Axiomatic and Economic Approaches to International Comparisons," Discussion Paper no. 96-11, Department of Economics, University of British Columbia, Vancouver, Canada.

Drechsler (1973), "Weighting of Index Numbers in Multilateral International Comparisons," The Review of Income and Wealth, Series 19 (March), 17-47.

Eichhorn, W. and Voeller J. (1976), Theory of the Price Index, Lecture Notes in Economics and Mathematical Systems, Vol. 140, Berlin: Springer-Verlag.

Fisher, I. (1922), The Making of Index Numbers, Boston: Houghton Mifflin.

Foekens, E. W., Leeflang, P., Wittnik, D.(1997), "Hierarchical Versus Other Market Share Models for Markets with Many Items," International Journal of Research in Marketing 14, 359-378.

Frisch, R. (1930), "The Problem of Index Numbers," Journal of the American Statistical Association, December.

Geary, R. B. (1958), "A Note on Comparisons of Exchange Rates and Purchasing Power Between Countries," Journal of the Royal Statistical Society A, 121, 97-99.

Grover, P. E. and Srinivasan (1987), "A Simultaneous Approach to Market Segmentation and Market Structuring," Journal of Marketing Research, 24 (May), 139-153.

Haan, J. de and Opperdoes E. (1997), "Estimation of the Coffee Price Index Using Scanner Data: Simulation of Official Practices."*

Haan, J. de and Opperdoes E. (1997), "Estimation of the Coffee Price Index Using Scanner Data: the Choice of the Micro Index." *.

Hawkes, W. J. (1997), "Reconciliation of Consumer Price Index Trends with Corresponding Trends in Average Prices For Quasi-Homogeneous Goods Using Scanning Data."*.

Hill, P. (1993), "Price and Volume Measures." in Systems of National Accounts 1993, Eurostat, IMF, OECD, UN and World Bank, Brussels, Luxembourg, New York, and Washington D. C., pp. 379-406.

Khamis, S. H. (1972), “A New System of Index Numbers for National and International Purposes,” Journal of the Royal Statistical Society A, 135, 96-121.

Kokoski, M. F., Cardiff P. and Moulton B. R. (1994), "Interarea Price Differences for Consumer Goods and Services: An Hedonic Approach Using CPI Data,” working paper 256, Bureau of Labor Statistics.

Kokoski, M. F. (1991), “New Research on Interarea Consumer Price Differences,”Monthly Labor Review, July 1991, 31-34.

Kokoski, M. F.(1993), "Quality Adjustment of Price Indexes,” Monthly Labor Review, December 1993.

Kokoski, M.F., Moulton, B., Zieschang, K. (1996), “Interarea Price Comparisons for Heterogeneous Goods and Several Levels of Commodity Aggregation,” working paper 291, Bureau of Labor Statistics.

Kravis, I. B., Kenessey, Z., Heston, A. and Summers R. (1975), A system of International Comparisons of Gross Product and Purchasing Power, Baltimore: The John Hopkins University Press.

Kravis, I. B., Sumers, R., and Heston, A. (1982), “Comments on ‘Selected Problems of Intercountry Comparisons on the Basis of the Experience of the EEC,’” Review of Income and Wealth, Series 28, 407-410.

Magnien, François and Pougnard, Jacques (1998), “Etude de Chaînage d’Indices de Prix à l’Aide de Micro-Données,” Journées de Méthodologie Statistique, March 1998.

Moulton, B. R. and Moses, K. E.(1997), “Addressing the Quality Change Issue in the Consumer Price Index,” U. S. Bureau of Labor Statistics.

Novak, T. P. (1993), “Log – Linear Trees: Models of Market Structure in Brand Switching Data,” Journal of Marketing Research 30, 267-287.

Olt, B. (1996), Axiom und Struktur in der Statistischen Preisindextheorie, Peter Lang, Frankfurt, Germany.

Primont, D. F. and Kokoski M. F. (1991), "Differences in Food Prices Across U. S. Cities: Evidence from CPI Data,” working paper 209, Bureau of Labor Statistics.

Reinsdorf, M. (1992), “Changes in Comparative Price and Changes in Market Share,” Managerial and Decision Economics, Vol. 13, 233-245.

Reinsdorf, M. (1996), "Constructing Basic Components for the U. S. CPI from Scanner Data: A Test Using Data on Coffee,” Working Paper 277, U. S. Bureau of Labor Statistics.

Reinsdorf, M. (1993), "The Effect of Outlet Price Differentials on the U. S. Consumer Price Index," in M. F. Foss, M. E. Manser and A. H. Young (eds.), Price Measurements and their Uses, NBER Studies in Income and Wealth, vol. 57, Chicago, University of Chicago Press.

Reinsdorf, M., (1994), "Price Dispersion, Seller Substitution and the U. S. CPI," Working Paper 52, Bureau of Labor Statistics.

Samuelson, D.P. and Swany (1974), "Invariant Economic Index Numbers and Canonical Duality: Survey and Synthesis," American Economic Review 64, 566-593.

Scobie, H. (1997), "Potential Uses of Scanner Data in the Production of Price Indexes – a Case Study using Coffee Data," Statistics Canada.

Siddarth, S., Bucklin, R. E., Morrison, D. G. (1995), "Making the Cut: Modeling and Analyzing Choice set restriction in Scanner Panel Data," Journal of Marketing Research 32, 255-266.

Silver, M. (1995), "Elementary Aggregates, Micro-Indices and Scanner Data: Some Issues in the Compilation of Consumer Price Indices," Review of Income and Wealth, Series 41, Number 4, December 1995, Cardiff Business School, University of Wales College of Cardiff, England.

Steiner G. (1978), On Difficulty and Other Essays, New York and Oxford: Oxford University Press, 18-47.

* Proceedings of the Third Meeting of the International Working Group on Price Indexes, Balk (ed.), Research Paper no. 9806, Statistics Netherlands, Division Research and Development, Department of Statistical Methods.

TABLE 1

COMPARISON OF PURCHASING POWER PARITIES FOR GROSS DOMESTIC PRODUCT ESTIMATED BY GEOGRAPHIC AND BY TIME-SERIES METHODS FOR 1993

		GDP PURCHASING POWER PARITIES (LOCAL CURRENCY RELATIVE TO U. S. $)			
		1990	1993 ESTIMATES		
COUNTRY	1990-1993 TREND IN GDP DEFLATOR	GEOGRAPHIC BASED ESTIMATES	GEOGRAPHIC BASED (DIRECT)	TIME-SERIES BASED* (IMPLIED)	% DISCREPANCY TIME-SERIES ESTIMATE
U.S.	109.7	--	--	--	--
Canada	105.3	1.30	1.26	1.25	-0.8%
Japan	105.2	195	184	187	+1.6%
Australia	105.0	1.39	1.35	1.33	-1.5%
New Zealand	106.5	1.61	1.51	1.56	+3.3%
Austria	112.5	14.0	13.9	14.3	+2.9%
Denmark	105.4	9.39	8.79	9.02	+2.6%
Finland	105.7	6.38	6.09	6.15	+1.0%
France	108.1	6.61	6.57	6.51	-0.9%
Germany	109.4	2.09	2.10	2.08	-1.0%
Italy	117.3	1421	1534	1519	-1.0%
Netherlands	107.2	2.17	2.16	2.12	-1.9%
Norway	104.8	9.73	8.93	9.30	+4.1%
Portugal	133.0	104	117	126	+7.7%
Spain	119.4	110	117	120	+2.6%
Sweden	111.6	9.34	9.83	9.50	-3.4%
Switzerland	110.4	2.20	2.13	2.21	+3.8%
Turkey	436.0	1491	5990	5926	-1.1%
U. K.	114.8	.602	.637	.630	-1.1%
				Root Mean Square Discrepancy:	2.9%

* 1990 PPP trended forward by 1990-93 GDP deflator relative to U. S.

TABLE 2

CLASSIFICATION OF INDIVIDUAL COFFEE ITEMS (UNIVERSAL PRODUCT CODES) IN SCANTRACK METRO CHICAGO AND METRO WASHINGTON GEOGRAPHIC AREAS FOR DECEMBER 1994

748 Coffee Items (UPC'S) Sold in Chicago or Washington during December 1994

	NUMBER OF ITEMS	% 0F ITEMS	% of COFFEE DEC. 1994 $ EXPENDITURES
Total Items Sold in Chicago	530	100.0%	100.0%
Items Also Sold in Washington	277	52.2%	61.7%
Items Unique to Chicago	253	47.8%	38.3%
Items Also Sold in December 1993	450	84.9%	96.7%
Items Unique to December 1994	80	15.1%	3.3%
Items Sold in Both Areas Both Years	200	37.7%	58.5%
Total Items Sold in Washington	495	100.0%	100.0%
Items Also Sold in Chicago	277	56.0%	74.4%
Items Unique to Washington	218	44.0%	25.6%
Items Also Sold in December 1993	411	83.0%	96.1%
Items Unique to December 1994	84	17.0%	3.9%
Items Sold in Both Areas Both Years	200	40.4%	69.4%

TABLE 3

COMPOSITION OF SCANTRACK COFFEE ITEMS ACCORDING TO USABILITY FOR PRICE INDEX COMPARISONS* BETWEEN METRO AREAS (METRO CHICAGO AND WASHINGTON) AND BETWEEN YEARS (DECEMBER 1993 AND 1994)

USABILITY CATEGORY	NUMBER OF ITEMS	%OF ITEMS	%OF EXPENDITURES
All Items	940	100.0	100.0
Items in Chicago Both Years	450	47.8	97.2
Items in Washington Both Years	411	43.7	96.7
Items in Both Metro Areas in 1993	261	27.8	68.1
Items in Both Metro Areas in 1994	277	29.5	68.8
Items in Both Metro areas Both Years	200	21.3	65.1

*All comparisons on this table involve the use of unit values to aggregate across weeks within each month, and to aggregate across outlets within each metro area. When index comparisons are made without aggregating across outlets, i. e. when items are matched on an individual outlet basis between December 1993 and December 1994, 83.7% of 1993 total expenditures, and 83.4% of December 1994 total expenditures, are usable.

TABLE 4

RECTANGULAR INTRANSITIVITY

BETWEEN TIME-SERIES AND GEOGRAPHIC COFFEE PRICE INDEX COMPARISONS

FOR SCANTRACK METRO CHICAGO AND WASHINGTON GEOGRAPHIC AREAS IN DECEMBER 1993 AND 1994

COMPARISONS		TIME-SERIES INDEX (DEC. 1994 PRICES AS % OF 1993 PRICES)			GEOGRAPHIC INDEX (CHICAGO PRICES AS % OF WASHINGTON PRICES)			DISCREPANCY BETWEEN TIME SERIES AND GEOGRAPHIC COMPARISONS
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
TIME-SERIES	GEOGRAPHY	WASHINGTON	CHICAGO	(2)÷(1)	DEC. 1993	DEC1994	(5)÷(4)	[(3)÷(6)]-1.000
A	B	146.2	164.3	1.125	94.7	101.5	1.071	+5.0%
C	B	147.2	162.4	1.102	94.7	101.5	1.071	+3.0%
D	D	142.3	160.7	1.129	94.5	101.1	1.070	+5.5%

A = Fisher Indexes, based on outlet-level prices and quantities for items sold in the same outlets in both months.
B = Fisher Indexes, based on area-level prices and quantities for items sold in both geographic areas in the same month.
C = Fisher Indexes, based on area-level prices and quantities for items sold in the same geographic area in both months.
D = Fisher indexes, based on area-level prices and quantities for items sold in both geographic areas in both months.

Note: All time-series comparisons are “direct”; i. e. indexes have not been chained through intervening months. All prices were computed by using unit values to aggregate across weeks.

TABLE 5

RECTANGULAR INTRANSITIVITY
BETWEEN TIME-SERIES AND GEOGRAPHIC COFFEE PRICE INDEX COMPARISONS
FOR SCANTRACK METRO CHICAGO AND WASHINGTON GEOGRAPHIC AREAS IN DECEMBER 1993 AND 1994
BASED ON INDIVIDUAL ITEMS SOLD IN BOTH GEOGRAHIC AREAS IN BOTH MONTHS

INDEX BASIS	TIME-SERIES INDEX (DEC. 1994 PRICES AS % OF1993 PRICES)			GEOGRAPHIC INDEX (CHICAGO PRICES AS % OF WASHINGTON PRICES)			DISCREPANCY BETWEEN TIME SERIES AND GEOGRAPHIC COMPARISONS
	(1) WASHINGTON	(2) CHICAGO	(3) (2)÷(1)	(4) DEC. 1993	(5) DEC. 1994	(6) (5)÷(4)	(7) [(3)÷(6)]-1.000
FISHER	142.34	160.74	1.1293	94.51	101.13	1.0701	+5.53%
SATO-VARTIA II	142.44	160.55	1.1271	94.73	101.40	1.0704	+5.30%
TORNQVIST	142.29	160.50	1.1280	94.65	101.19	1.0691	+5.50%
CCD*	143.66	157.76	1.0981	93.28	102.43	1.0981	.00
KMZ**	143.18	156.95	1.0962	93.45	102.44	1.0962	.00

* Caves, Christensen, and Diewert (1982) method for forcing transitivity.

** Kokoski, Moulton, and Zieschang (1996), method for forcing transitivity.

TABLE 6

CONSOLIDATION OF INDIVIDUAL UPC COFFEE ITEMS USING CPI "SPECIFICATION ITEM" CRITERIA AND STORE BRAND (CONTROLLED BRAND) ITEM UNIFICATION

	TOTAL COFFEE	ROASTED COFFEE	INSTANT COFFEE
Total UPC Items	940	735	205
Minus Item Consolidations by Grind	-30	-30	0
Minus Item consolidation by Flavor	-232	-207	-25
Equals Total "CPI"-Type Items"	678	498	180
Minus Item Consolidations for store brands	-74	-45	-29
Equals Total CPI-Type Items with store brand consolidations	604	453	151

TABLE 7

EFFECT OF ALTERNATIVE METHODS FOR REDUCING RECTANGULAR INTRANSITIVITY FOR ROASTED COFFEE ITEMS

INDEX BASIS	TIME-SERIES INDEX DEC. 1994 PRICES AS % OF1993 PRICES)			GEOGRAPHIC INDEX (CHICAGO PRICES AS % OF WASHINGTON PRICES)			DISCREPANCY BETWEEN TIME SERIES AND GEOGRAPHIC COMPARISONS
	(1) WASHINGTON	(2) CHICAGO	(3) (2)÷(1)	(4) DEC. 1993	(5) DEC. 1994	(6) (5)÷(4)	(7) [(3)÷(6)]-1.000
TORNQVIST FOR MATCHED ITEMS	153.42	178.40	1.1628	91.65	100.58	1.0975	+5.94%
CCD* FOR MATCHED ITEMS	154.71	174.92	1.1307	89.77	101.05	1.1307	0.00%
KMZ** FOR MATCHED ITEMS	154.41	173.86	1.1260	90.70	102.12	1.1260	0.00%
TORNQVIST FOR GROUPED ITEMS	154.86	176.42	1.1392	89.18	102.44	1.1325	+0.60%

* Caves, Christensen, and Diewert (1982) method for forcing transitivity.

** Kokoski, Moulton, and Zieschang (1996) method for forcing transitivity.