IMF Working Paper

WP/99/78

INTERNATIONAL MONETARY FUND

Statistics Department

Price Imputation and Other Techniques for Dealing with Missing Observations, Seasonality and Quality Change in Price Indices

Prepared by Paul A. Armknecht and Fenella Maitland-Smith[1]

Authorized for distribution by Adriaan M. Bloem

June 1999

Abstract

Price index compilers frequently face situations where price observations are missing due to seasonal unavailability, supply shortages, or the discontinuation of products. Incorrect treatment of such situations can result in biased price indices. This paper presents statistical imputation techniques that index compilers can use to prevent bias and suggests the extension of these same techniques to assist with adjustments for quality differences. The use of additional procedures for dealing with some of the problems caused by seasonal commodities is also discussed.

JEL Classification Numbers: C43, C63, C81

Keywords: Price indices, imputation techniques, seasonal goods, quality adjustment

Authors' E-Mail Addresses: parmknecht@imf.org, fmaitlandsmith@imf.org

---

[1]Paul A. Armknecht is a technical assistance advisor and Fenella Maitland-Smith is an economist, both in the Real Sector Division.

Contents Page

Tables

I. INTRODUCTION

As part of the Statistics Department's technical assistance in price index compilation, staff are frequently confronted with questions about how to deal with missing price observations—a common problem in almost all countries. Missing prices present a problem for index compilers because if not treated correctly they can affect index levels and changes and can cause bias. The purpose of this paper is to provide methodological approaches that can be used for resolving the problems of missing or disappearing items in the measurement of price change.

The issue of missing prices arises because price index compilation is based on statistical sampling techniques, i.e., since it is not possible to perform a complete enumeration of all prices paid in relevant transactions, it is necessary to take representative samples. Sampling is usually necessary in several dimensions of price collection. For example, in the case of a consumer price index, a sample of geographical areas (shopping centers) is selected, samples of retail outlets are drawn from the universe of outlets, samples of varieties (brands) of an item are selected, etc. The timing of price collection may be subject to sampling in that prices are collected on certain days each month. Thus, if the chosen variety of an item is not available for pricing in one of the chosen outlets on the day of price collection, this may be an entirely localized phenomenon and not representative of the universal availability of the variety. In this case, the fact that a price could not be collected should not have undue influence on the overall price index. If, on the other hand, the local unavailability is representative of a wider phenomenon and consumers are forced to permanently substitute a replacement variety, then the index should reflect any true price change associated with this substitution, but not price change due to quality differences.

From the economic theory viewpoint, a price index should reflect price changes experienced in maintaining a constant level of consumer utility or satisfaction. Thus, any price change associated with increasing, or decreasing, utility (quality) should not affect a price index. When products disappear from the market consumers will generally seek substitutes, and when this occurs it is important for the index compiler to reflect this shift, but distinguish between pure price change and price change due to quality differences. Quality changes are, in fact, volume changes and their price effects must be excluded from price indices.

A. Causes of Missing Prices

There are many reasons why a price collector may find that a price observation is not available in any period and it is important to establish whether the unavailability is likely to be temporary or permanent. A price may be considered as temporarily unavailable if the same commodity is likely to return to the market—some seasonal commodities show fairly predictable temporary unavailability. Permanent unavailability, on the other hand, occurs when a commodity is withdrawn from the market with no prospect of returning in the same form.

Items may be temporarily unavailable due to supply shortages caused by factors such as the seller underestimating demand, strikes by factory or transportation employees, or supply problems with imported commodities. In these cases a price observation may be unavailable in the current period but the collector has information to suggest that the same commodity will become available again, although it may not be clear when this will happen.

Seasonal commodities often show seasonal patterns of availability, i.e., they are temporarily unavailable over roughly the same months each year so that their unavailability is to a large extent predictable. Primary examples of this phenomenon are fresh fruit, fresh vegetables, clothing and some types of sporting equipment which are available in the market for short periods only each year[2].

Particular care must be taken when dealing with temporary unavailability as bias can arise when prices become available again, especially if a modified Laspeyres formula is being used (this is explained in section IB).

In practically every period for which a price index is compiled, some varieties of a commodity disappear and will not be sold again, i.e., they become permanently unavailable. In such situations replacement varieties should be selected and introduced into the sample with great care being taken to identify any differences in quality between the original and replacement varieties. If a quality difference does exist, then it is crucial to ascertain whether the difference has a value—if not, then the price of the replacement variety can be compared directly with the price of the original. If the quality difference does have a value, however, then this price effect must be removed from the index calculations using an adjustment technique. If price effects due to quality change are not adjusted for, then the price index will be biased.

B. Current Practices and Their Implications

The methods most frequently used to deal with the missing prices problem in member countries generally fall under one of the following headings: i) taking no action; ii) ensuring that samples are comparable—with particular reference to use of the modified Laspeyres formula; iii) carrying forward the last available price; and iv) imputing prices.

If no action is taken to accommodate missing prices, the outcome will depend on the method used to compile the elementary aggregates of the price index, where the method may be either the ratio of averages, the average of relatives, or the geometric mean. Taking the ratio of averages first—the average price of a sample of observations (e.g., varieties of washing soap) in the current period is compared to the average price in the previous (or base) period, and in order to obtain a correct result, it is vital that the samples of varieties are the same in both periods. If a price is missing in the current period and no action is taken the index will be biased. For example, if prices for three varieties (A=$3, B=$3, C=$2) were

---

[2]Seasonal commodities may present additional problems for index compilers, such as extreme price movements and frequent specification changes, and these are also discussed in this paper as they often interact with the problem of missing prices (see section III).

available in the previous period giving an average price of $2.67, whereas in the current period prices are available for two varieties only (B=$3, C=$2) the average price will fall to $2.50 even though none of the variety prices has actually fallen. The average price, and thus the index level, has fallen purely due to a change in the sample composition, and the index is now a biased indicator of price change.[3]

Rather than comparing the average prices of two periods, a price index may be constructed by comparing each individual price with its corresponding price in the previous (or base) period to give a price relative for each observation. These price relatives are then weighted together, using the standard Laspeyres (long-term relative) or the modified Laspeyres (short-term relative) formulae,[4] to give an index value—this is the average of relatives formulation. If a price is missing in the current period, then its price relative cannot be calculated (possibly causing computer error). If this price relative has a specific weight, then omitting the relative from the next level of aggregation may cause errors if the weights of the other price relatives are not adjusted accordingly.

The third method of averaging, which is being introduced by more and more countries for estimating the elementary aggregate indices, is to calculate a geometric mean of prices in both periods and then derive the price relative or, alternatively, calculate a geometric average of the price relatives. Both calculations will yield the same results. However, the problem of

---

[3]The situation described here assumes that we are dealing with a sample, and that the sampled variety is temporarily missing but will be available again at some point in the future. This should be dealt with by using one of the price imputation methods discussed in section II. If the variety is permanently unavailable, then a replacement should be selected. Replacement techniques are discussed in section IV.

[4]The standard Laspeyres formula is:

$$I_{i,t} = \sum_{i=1}^{n} \left(\frac{p_{i,t}}{p_{i,o}}\right) \frac{p_{i,o} q_{i,o}}{\sum^{n} p_{i,o} q_{i,o}}$$

The modified Laspeyres formula is:

$$I_{i,t} = \sum_{i=1}^{n} \left(\frac{p_{i,t}}{p_{i,t-1}}\right) \left(\frac{p_{i,t-1}}{p_{i,0}}\right) \frac{p_{i,0} q_{i,0}}{\sum_{i=1}^{n} p_{i,0} q_{i,0}} = \sum_{i=1}^{n} \left(\frac{p_{i,t}}{p_{i,t-1}}\right) I_{i,t-1}\, w_{i,0}$$

In using the short-term relative in the modified Laspeyres formula, it is important that the base period weight ($w_{i,0}$) is updated for price change by the previous period index ($I_{i,t-1}$). This is necessary to ensure that the implicit quantity weights remain fixed, to avoid the problems associated with a Sauerbeck index. A weighted version of the Sauerbeck index would use the formula:

$$S_{i,\tau} = \prod_{\tau=1}^{t} \sum_{i=1}^{n} \left(\frac{p_{i,t}}{p_{i,t-1}}\right) w_{i,0}$$

In this index the implicit quantity weights would change as the previous period price changes, which usually results in a bias. For a more complete discussion of this issue see Lequiller and Zieschang (1994).

missing observations is still the same as in the previous two cases. The sample of observations used from each period is required to have the same number of observations for computing the geometric mean of prices in order to avoid a biased measure of price change. In the case where a weighted geometric mean is used, the weights for missing observations would also need to be distributed to the remaining observations to avoid any bias.

Returning to the ratio of averages method, special care is needed when comparisons are made between the average price of the current period, t, with that of the previous period, t-1 (the modified Laspeyres formula), as opposed to that of the base period (the standard Laspeyres formula). Some countries are careful to ensure that the problem of non-comparable samples described above does not occur, i.e., when a price is missing, the corresponding price is removed from the sample in the previous period, t-1, so that the samples are of the same composition in both periods. If the price returns a few months later, t+3, again, care is taken to ensure that samples of the same composition are compared. Thus, during the first period that the price is again available, t+3, it is not included in the sample, but during the second period, t+4, it is included in both the sample of the first and second periods. Although the index movement between t+3 and t+4 will be correct, the movement between t+2 and t+3 (or more correctly, between t and t+3) will be incorrect as it will not reflect any difference between the price at t and the price at t+3, and the index will be biased.

This situation can easily occur when a seller exhausts old stock of a product, experiences a gap of several months before receiving new stock and then sells the new stock at a higher price. This is similar to the situation with many seasonal goods which are unavailable during the winter, for example, before appearing on the market at a high price in the spring, experiencing a steady fall in price over the summer before disappearing from the market in the fall. The method described above will fail to reflect the high price in the spring and the index will suffer a systematic downward bias. The correct method to use in this situation is to impute prices in all months when observations cannot be made (see the techniques described in section II).

A common treatment of missing prices is to carry forward the last available price to the months when prices are not available. Although this does provide a price in the months when observations cannot be made, it is likely to mean that monthly movements in the index are biased, since the sub-indices in question will remain flat when prices are not available (if prices in general are rising, the bias will be downwards, whereas if prices are falling the bias will be upwards). There is also likely to be a large step-change in the index when the price becomes available again or when the new season starts. This method is not recommended, particularly with high inflation or where monthly movements in the price index (as opposed to annual movements) are used as a major indicator.

Although carrying forward a price is better than simply adjusting the sample composition to ensure comparability and is undoubtedly better than taking no action at all, by far the best solution is to estimate and impute a price. Imputation makes use of the best available information to provide an unbiased estimate of price movement.The rest of this paper is concerned with imputation methods and their application.The principles of imputation are described in Section II. Section III describes how these techniques can be applied to the special case of seasonal unavailability and discusses some of the additional problems caused by seasonal commodities and their possible solutions. Section IV explains how the estimation techniques described in Section II can be used to make the adjustments necessary when the quality of a good or service changes. The final section provides concluding remarks on those methods that are preferred by the Statistics Department of the Fund and those that should be avoided.

II. IMPUTATION TECHNIQUES

As the preceding discussion indicates, it is good practice to estimate a value whenever a price or an index is missing and such imputations should always be made using the best information available.

Imputation can be done in such a way as to be implicit or explicit. In the case of implicit imputation, the missing index and its weight are simply omitted from all calculations, resulting in the weight being automatically redistributed, proportionally, over the other indices in the group. Thus the group index is calculated using only those indices that are available and the missing index implicitly assumes the same value as the group index. In the case of explicit imputation the index (or price) for the missing item is explicitly recorded, but flagged as an imputation, and the group index (or average price) is calculated using this imputed value.The imputation of prices is necessarily explicit, whereas the imputation of indices can be implicit or explicit.

It is recommended that imputations should always be made explicitly. In cases where the modified Laspeyres formula is used, an imputed price or index will be needed for index calculation in the following month. In all cases, explicit imputation is preferred so that index movements are fully understood and that imputed indices/prices are always flagged or marked so as to be readily identifiable. Imputed prices could be entered directly onto price collection forms, and flagged, to aid transparency and to avoid confusion[5].

A. Imputation of Indices or Price Relatives

Imputations can be made using the index (movement) of a group, sub-group, or single product (item). The choice depends on which is felt to most closely represent the movement of the index for which prices are missing.

If a **group** index (movement) is used, then the imputation can be viewed as both implicit **and** explicit. In the case of implicit imputation, the group index is calculated using

---

[5]This is not meant to imply that price collectors perform the imputations, but rather that imputations are done at headquarters and then entered onto forms at a later date. This is easy to implement where new price collection forms, including previous data, are computer-generated each month.

only those indices that are available and the missing index implicitly assumes the same value as the group index (see calculation A). In the case of explicit imputation, the group index is calculated as described in the implicit case, but the index for the missing item is explicitly recorded, and flagged as an imputation. The group index is then "recalculated" using this imputed value (calculation A followed by B). Of course, the level of the group index is unaffected by this explicit imputation, and will be the same in both the implicit and the explicit case. Again, it is recommend that imputations be made explicitly in all cases, so that index movements are fully understood. If a **product** (or item) index is used to supply a value for imputation, then the process is necessarily explicit, as the group index will be affected by the imputation (calculation C).

Using group indices, the implicit (and first stage of the explicit) process is to calculate group indices using the available sub-indices (products A and B below) :

Calculation A

	Dec 97 weight (w)	Dec 97	**Long-term indices (I)** Jan 98	Feb 98	Mar 98	Apr 98	May 98
Product A	400	100	110	115	130	135	145
Product B	600	100	115	120	125	130	137
Product C	500						
$I_A*w_A + I_B*w_B$		100,000	113,000	118,000	127,000	132,000	140,200
Long-term group index		100	113	118	127	132	140
Monthly change (short-term)			1.13	1.04	1.08	1.04	1.06

The second stage of the explicit process uses the group indices calculated above to recalculate group indices. Firstly impute the group indices from calculation A into the line for product C, and then relaculate the group indices (which will remain the same):

Calculation B

	Dec 97 weight (w)	Dec 97	**Long-term indices (I)** Jan 98	Feb 98	Mar 98	Apr 98	May 98
product A	400	100	110	115	130	135	145
product B	600	100	115	120	125	130	137
product C	500	**100**	**113**	**118**	**127**	**132**	**140**
$I_A*w_A + I_B*w_B + I_C*w_C$		150,000	169,500	177,000	190,500	198,000	210,300
Long-term group index		100	113	118	127	132	140.2
Monthly change (short-term)		1.13	1.04	1.08	1.04	1.06	

If another similar product (A) is used to provide an index, rather than using the group index, then the process is explicit. Firstly, the indices for product A are imputed into the line for product C, and then group indices are calculated. Note that these group indices are different to those obtained in calculations A and B:

Calculation C

	Dec 97 weight (w)	Dec 97	**Long-term indices (I)** Jan 98	Feb 98	Mar 98	Apr 98	May 98
product A	400	100	110	115	130	135	145
product B	600	100	115	120	125	130	137
product C	500	**100**	**110**	**115**	**130**	**135**	**145**
$I_A*w_A + I_B*w_B + I_C*w_C$		100,000	168,000	175,500	192,000	199,500	212,700
Long-term group index		100	112	117	128	133	142
Monthly change (short-term)			1.12	1.04	1.09	1.04	1.07

B. Imputation of Prices

Imputation is generally done at the first level of aggregation, i.e., indices or price relatives are imputed rather than prices, but imputation of prices may be necessary where the lowest level aggregation is done using a ratio of unweighted average prices formula. For example, if CPI prices are collected for a particular item from several outlets in a city, and unweighted averages are calculated for the city for the current and previous months, then missing prices should be imputed.

The steps involved in the calculation are as follows:

1. calculate **monthly** movement of comparable averages (using the same observations in each month), where the previous month's average may contain imputed prices but the current month's average will consist of collected prices only. This could be done at the level of the item at the city, region or national level depending on how many prices are actually available:

	Dec 97	**Prices** Jan 98		Feb 98	
outlet A	10	10	--✓--	11	
" B	10	10	--✓--	10	
" C	9	10	--✓--	11	
" D	12	12		..	
Comparable averages:					
- current month				10.67[a]	a = (11+10+11)/3
- previous month		10[b]			b = (10+10+10)/3
Monthly movement				1.067[c]	c = a / b

2. multiply previous month's price (this may itself have been imputed) by movement in comparable average, to give imputed price:

   Imputed price = previous month's price x movement in comparable averages
   = 12 x 1.067
   = 12.8

3. recalculate averages using all prices (collected and imputed) in both the current and previous month (the movement will be the same as the movement obtained in step 1, at the level used in step 1):

		Prices			
		Dec 97	Jan 98		Feb 98
outlet	A	10	10	--✓--	11
"	B	10	10	--✓--	10
"	C	9	10	--✓--	11
"	D	12	12	--✓--	**12.8**
Averages			10.5		11.2
Monthly movement					1.067

If a price is missing for several months, an imputation must be made every month, i.e., imputed prices must be updated. This is the mechanism that needs to be employed in the case of seasonal commodities which are only available for a few months (strawberries). The last available, end of season, prices are continuously updated so that, when strawberries return to the market the following year, each new price collected has an imputed price in the previous month. Using this method, the first available, new season, prices can be used in the index. The indices may show large peaks in certain months as a result of this method, but this is correct and reflects market conditions. The peaks should fall in the same months each year with the result that they do not affect the 12 month rate of change.

		Prices					
		Dec 97	Jan 98	Feb 98		Mar 98	
outlet	A	10	10	11	--✓--	11	
"	B	10	10	10	--✓--	10	
"	C	9	10	11	--✓--	12	
"	D	12	12	**12.8**		..	
Comparable averages:							
- current month						11[a]	a = (11+10+12)/3
- previous month				10.67[b]			b = (11+10+11)/3
Monthly movement						1.031[c]	c = a / b

Imputed price = previous month's price x movement in comparable averages
= 12.8 x 1.031
= 13.2

		Dec 97	Prices Jan 98	Feb 98		Mar 98
outlet	A	10	10	11	--✓--	11
"	B	10	10	10	--✓--	10
"	C	9	10	11	--✓--	12
"	D	12	12	**12.8**	--✓--	**13.2**
Averages				11.2		11.6
Monthly movement						1.031

III. Treatment of Seasonal Commodities in Price Indices

In almost every economy there are a number of products that are consistently unavailable during certain periods of the year, and this occurs on a regular basis year after year. The problem of seasonal unavailability can be handled using the imputation techniques described in Section II, although there are alternative solutions such as the use of variable weights (described below). Seasonal commodities may present additional problems for index compilers, such as extreme price movements and frequent specification changes, and these are also discussed as they often interact with the problem of missing prices. The following discussion uses two commodity groups (fruit and vegetables, and clothing) as examples.

A. Possible Problems

There are several features that might appear in the price patterns of seasonal commodities that can cause problems during index compilation:

A. seasonal unavailability (commodities not available for some months of each year):
- what to do in missing months - impute prices, or use zero weights?
- how to re-introduce prices when season starts?
- how to deal with any changes in quality at start of new season?

B. variation in the seasonal pattern from year to year (the season sometimes starts early or late):
- in which month should price collection start - in exactly the same month each year or, should there be flexibility?

C. a seasonal pattern that shows extreme price movements:
- should the full extent of the movements be allowed to influence the overall price index?
- annual percentage changes will be unusual if problem B exists too.

D. seasonal changes in product ranges:
- low prices of seasonal sales (to clear stocks of old products) are likely to be followed by introduction of new products
- how to deal with any changes in quality at start of new season?

Fruit and vegetables

In the case of seasonal fruit and vegetables, prices are strongly dependent on supply conditions, such as weather, and are thus liable to show extreme movements (C) and shifts in seasons (B). Their availability is also seasonal (A). Changes in specification or quality are not common and do not therefore cause the same problems as are experienced with seasonal clothing.

Clothing

In many countries the price patterns of clothing are the result of seasonal sales (for example, January and July), and seasonal availability (for example, winter coats, bathing suits). So, the problem of extreme price movements (C) is combined with seasonal unavailability (A).

The situation is further complicated by seasonal changes in product ranges (D) or, in other words, the effect of fashion. New stocks of clothes arrive in the shops immediately after the sales of existing inventory at discounted prices, and are often of different styles to those that have disappeared during the period of price reductions. So, there is a question of whether the new styles are of a different quality to the old styles. Of course, for seasonal items, comparisons must be made between the new products and the old products that disappeared at the end of the previous, comparable, season, maybe six months earlier.

B. Using Imputation Techniques

If imputation is used, then all items have fixed weights, and prices are imputed for the months when prices are not available using the techniques described in Section II. This should avoid the problem of bias in monthly movements and should result in a smaller step-change in the index when the new season starts. It is important that the index reflects the full extent of the price difference between the last month in which a price was imputed and the first month of the new season. If this difference is not reflected, annual movements in the price index will not correctly reflect price changes over the year. Allowing this price difference to be reflected may cause strange movements in the index between the last month in which a price was imputed and the first month of the new season, but this should be allowed to happen.

If, however, there is a tendency for seasons to shift (B), and prices are collected as soon as the new product appears, these step changes will occur in different months each year and will therefore cause the index to show strange annual changes. In this case it would be wise to collect prices in exactly the same months each year, regardless of temporary shifts in seasonal availability.

C. Using Variable Weights

Instead of using imputation to solve the problem of seasonal unavailability, a system of variable weights can be used whereby items have different weights in different months according to consumption[6], but within a fixed group weight. Thus, zero weights are used in the months when prices are not available (A). This approach, to some extent, answers the question of how to treat extreme price movements (C) as the very high prices common at the start of a season are likely to receive a low weight and thus have a reduced effect on the overall index. A system of variable weights will, by definition, result in prices being collected in fixed months each year (B), but this rigidity can prove a problem if seasons shift so much that prices cannot be found in a month when they have a weight. In this case, imputation may be necessary.

This method needs careful presentation as monthly movements in the group index will not only reflect price movements, but also weighting shifts. In addition, it is important that a fixed price reference base is maintained, to avoid the drift bias[7] that could arise when the weights change from month to month. The method will also require the construction of special software, different from that used for ordinary commodities.

The following example, using fruit and vegetables, will demonstrate how a variable weighting system might be applied. First, consumption data are analyzed for recent years. It is sensible to construct the weights using the average of several years' consumption to prevent any exceptional seasons from having too large an effect. So, the total quantity consumed in each month over the last three years might be as follows:

---

[6]It is interesting to note that, even with a system of fixed weights, when missing prices are imputed the weight of a missing item is implicitly redistributed amongst those items still available, effectively creating a system of variable weights. This analogy only holds, however, during months when some items are **completely** unavailable, but not when relative consumption changes between available items.

[7]Drift bias occurs when new weights and a new price reference base are introduced simultaneously and then linked to the index of the previous period. If price and quantity movements are negatively correlated, the index will be biased upwards (see Schultz, 1983). This effect is accentuated with frequent chaining, and for this reason, it is important that monthly chained indices are not used in the context of variable weights.

Table 1. Consumption of fruit 1995–97 ('000 tons)

	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec	**Total**
Apples	60	60	60	60	50	40	20	20	80	80	60	60	**650**
Cherries	-	-	-	-	-	10	20	25	20	10	-	-	**85**
Plums	-	-	-	-	-	10	30	25	20	10	-	-	**95**
Melons	5	5	5	10	20	20	30	30	20	20	15	10	**190**
Grapes	10	10	10	20	30	40	40	50	20	20	20	20	**290**

The following formula[8] will be used (summation is over items, not months):

$$(1) \qquad I_{m,y} = \frac{\sum p_{m,y} q_{m,0}}{\sum \bar{p}_0 q_{m,0}} = \sum \frac{p_{m,y}}{\bar{p}_0} \frac{\bar{p}_0 q_{m,0}}{\sum \bar{p}_0 q_{m,0}}$$

in which **p** stands for price and **q** for monthly quantities (as in Table 1), **m** stands for the month, **y** for the current year, **0** for the base period (here 1995-97), and **p** with a bar for the average price over all months in the base period[9]. Thus, although the monthly quantities (basket) are the same in the numerator and the denominator, the basket changes from month

---

[8]This formula is related to a Rothwell formula which uses variable weights and makes direct price comparisons with prices twelve months earlier:

$$I_{m,y}^{Rothwell} = \frac{\sum p_{m,y} q_{m,0}}{\sum p_{m,0} q_{m,0}} = \sum \frac{p_{m,y}}{p_{m,0}} \frac{p_{m,0} q_{m,0}}{\sum p_{m,0} q_{m,0}}$$

But formula (1) has a fixed price reference base and thus, in a sense, it can be viewed as a variant of a Rothwell formula which is re-referenced to a fixed base each month.

[9]Average prices should be calculated as weighted averages (summation over all months in the base period):

$$\bar{p}_0 = \frac{\sum p_{m,0} q_{m,0}}{\sum q_{m,0}} = \sum p_{m,0} \frac{q_{m,0}}{\sum q_{m,0}}$$

to month. The change in the index from month to month is therefore not only a pure price change but also reflects changes in the monthly basket of commodities.

Relative expenditure weights are calculated using average prices for the base period:

(2) $$w_{m,0} = \frac{\bar{p}_0 q_{m,0}}{\sum \bar{p}_0 q_{m,0}}$$

Price relatives can be calculated using a convenient price reference base period, for example, to correspond with the price reference base of the overall index (commonly either a whole year or a single month). In any case, for practical reasons it might be preferable to use an annual average price as the price reference base rather than a single month, since this avoids the problems associated with commodities which are never available in the single reference month. If the overall base period is, for example, annual average of 1997, then the price relatives will be of the form:

(3) $$\frac{p_{m,y}}{\bar{p}_{1997}}$$

In order to use price relatives based on a different period to their corresponding weights (which relate to the period 1995-97), a link factor is needed:

(4) $$\frac{\bar{p}_{1997}}{\bar{p}_0}$$

Thus, the full calculation becomes (which is equivalent to (1)):

(5) $$I_{m,y} = \sum \frac{p_{m,y}}{\bar{p}_{1997}} \frac{\bar{p}_{1997}}{\bar{p}_0} \frac{\bar{p}_0 q_{m,0}}{\sum \bar{p}_0 q_{m,0}}$$

To obtain the relative weights as given in (2), monthly quantity data in Table 1 are first multiplied by average prices of 1997, as shown in Table 2:

Table 2: Quantities multiplied by average prices of 1997

	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec	Avg. Prices 1997
Apples	731	731	731	731	609	487	244	244	975	975	731	731	12.18
Cherries	0	0	0	0	0	388	776	971	776	388	0	0	38.82
Plums	0	0	0	0	0	34	101	84	67	34	0	0	3.37
Melons	474	474	474	947	1895	1895	2842	2842	1895	1895	1421	947	94.74
Grapes	476	476	476	952	1428	1903	1903	2379	952	952	952	952	47.59
Total	**1681**	**1681**	**1681**	**2630**	**3932**	**4707**	**5867**	**6520**	**4665**	**4243**	**3104**	**2630**	

These weights can now be normalized by dividing each column by its total (and multiply them by 100), thus leading to Table 3:

Table 3: Normalized monthly weights

	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
Apples	43.5	43.5	43.5	27.8	15.5	10.4	4.2	3.7	20.9	23.0	23.6	27.8
Cherries	0	0	0	0	0	8.2	13.2	14.9	16.6	9.1	0	0
Plums	0	0	0	0	0	0.7	1.7	1.3	1.4	0.8	0	0
Melons	28.2	28.2	28.2	36.0	48.2	40.2	48.4	43.6	40.6	44.7	45.8	36.0
Grapes	28.3	28.3	28.3	36.1	36.3	40.4	32.4	36.5	20.4	22.4	30.7	36.2
Total	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**

Formula (5) is now applied to derive index numbers using 1997 as a base.

The next step is to aggregate this result into the overall index. As explained earlier, aggregation in the overall index uses the normal system of fixed weights. In other words, if weights are variable inside the category of seasonal commodities (in our example, fruit), the category itself has a fixed weight in the overall index.

D. Treatment of Clothing

Seasonal clothing is a special case as it combines the problem of extreme price movements (C), with seasonal unavailability (A), and seasonal changes in product ranges (D).

In the worst case, extreme discounting (to clear old stocks) is followed by several months of unavailability before new stocks of new styles are available.

The first choice to be made is whether or not to include discounted prices in the price index. In the case of a product showing temporary discounts or promotions, if it seems likely that the price will return to its normal level after a short period, then the discounted price **should certainly be included** in the index and no special method is needed to handle it.[10] If however the discounting is seasonal in nature and intended to help the retailer to clear stocks of old styles, then the issue is more problematical because, unless specific procedures are applied to ensure a return to normal prices, the index will be subject to a systematic downward bias.

Seasonal discounting is well established in many countries but is just starting to appear in others, for example, in some transition countries. Statistical offices could start collecting discounted prices from outlets and use them as part of a research project based on the method described below with a view to introducing discounted prices into their price indices.

First, those clothing items which are subject to seasonal discounting should be distinguished from other clothing items and the dates during which discounting occurs should be identified. Price collectors should record the prices of these products as for any other products, but during the discounting period, prices will be recorded with a clear marking. When the discounting period is finished and item has disappeared from the shop, the price collector should not replace it, but instead wait for the start of the next season.[11] During this period when a seasonal clothing item is not available, the following procedure should be employed:

i. the last available **normal** price should be imputed in the first month when the item is unavailable. In other words, when the discounting period is over, the index will return to "normal";

ii. this last normal price will be moved forward during the period of unavailability by applying the monthly movements for clothing items for which prices are available;

iii. when the normal season resumes, a new product as similar as possible to the product that was priced during the previous season should be selected and its price directly compared to the old product. It is important that the index reflects the full extent of the price difference between the last month in which a price was imputed and the first month of the new season;

---

[10]However, if the products which are discounted are faulty or shop-soiled then their prices should certainly **not** be included.

[11]This assumes that the product shows seasonal unavailability (A). If this is not the case, then the product should be replaced immediately.

iv. pricing should continue using this new product.

As a result of these procedures, the index movement for clothing items affected by seasonal discounting will move similarly to other items during the "normal" season, will decrease sharply during the discounting season, return to the last normal season level just after the discounting season, move in parallel with other non-seasonal commodities during the non availability period, and be moved by annual comparisons when the normal season resumes.

Although a simpler approach might be to impute the discounted price during the period of unavailability, this should be avoided as it will result in an extremely sharp upward movement when the new normal price is compared to the imputed sale price, which will lead to particularly erratic index movements when the seasonal patterns vary (shift in time) from year to year.

IV. Methods for Making Quality Adjustments in Price Indices [12]

The discussion so far has concentrated on the use of imputation for estimating prices for missing observations, but imputation of prices can also play a vital role when products disappear and replacements are introduced into a price index. When such replacements are made, care must be taken to distinguish between price change and quality change, otherwise, quality improvements (or degradations) could be treated as price changes leading to a potential bias that results in an overstatement (or understatement) of true price change. The aim of a price index is to measure pure price change and exclude changes due to improvements or deterioration in the quality of products.

The need for quality adjustments in price index computation generally occurs when a priced variety is permanently missing, i.e., the shop has discontinued carrying the commodity or it is no longer being produced. In this case, a product or service as similar as possible to the previous variety should be selected as a replacement.

When the replacement variety has been selected, a decision must be made as to whether the new and old variety are of comparable quality. This decision can be made based on product or service characteristics between the varieties. Do they appear to be the same or does one have some feature that makes is better or worse than the other? If there is no distinguishable difference between the two, then they are comparable and the new variety's price can be used directly in the index with no further adjustment. To the extent that varieties can be chosen which are strictly comparable to the old variety, the substitution will not result in any special procedures for calculating the index.

---

[12] This section is based on an earlier, unpublished paper by Paul A. Armknecht and Brent R. Moulton entitled "Quality Adjustment in Price Indices: Methods for Imputing Price and Quality Changes," prepared in October 1995.

If the replacement variety is not comparable, then some adjustment for quality will have to be made. An estimate is needed for the value of the quality difference.

A. Quality Adjustment Techniques

Techniques for quality adjustment in price indices have been discussed recently in the economic literature, particularly following the release of the Boskin report[13] in the U.S. This paper classifies the techniques available for quality adjustments into two classes—direct adjustment and imputation.

Direct adjustment

Direct adjustment involves assigning a monetary value to the quality difference and then adjusting the price of the observation for the quality differential. Such adjustments, in practice, are handled in three ways. First the data collector determines the value either through direct knowledge or in consultation with personnel in the outlet where the product or service is sold. Alternatively, analysts who are knowledgeable about the product could assign a value or, in the case of an improvement to an existing product, could contact the producer directly to obtain an estimate of the difference in production cost and normal mark up so that an estimate of the quality change can be made (this can be relatively simple in cases where an optional feature has become standard) . The third way in which the value of the quality difference can be obtained is to estimate the value of the change in a product characteristic using hedonic regression models.

Each of these methods has its limitations. Reliance on an individual's knowledge concerning the products is open to subjectivity. Estimates by producers of their marginal costs for changes provides better data but producers are often unwilling to provide such information because it takes considerable time to evaluate and often it is considered to be proprietary and confidential. Hedonic methods require large databases with a wide range of product characteristics. Such databases are just not readily available in most national statistical offices but may be obtained for some products through the increasing availability of scanner data. It should be noted, however, that such databases involve substantial design and maintenance costs. In addition, hedonic models need to be re-estimated periodically.

---

[13] Advisory Commission to Study the Consumer Price Index, *Toward a More Accurate Measure of the Cost of Living*, Final Report to the Senate Finance Committee, U.S. Government Printing Office, December 4, 1996 (referred to as the Boskin report). Also see Brent R. Moulton, "Bias in the Consumer Price Index: What is the Evidence?" *Journal of Economic Perspectives* (Fall 1996), Vol. 10, No. 4, pp. 159-77.

Imputations

The other class of techniques for making quality adjustments is by imputation using price changes for comparable varieties of products. The actual adjustments made depend on the index formula that is used.[14]

One of the preferred imputation methods is referred to as overlap pricing. In this instance, prices for the old and the new varieties are observed in the market place simultaneously, and this period is known as the overlap period. The price for the old variety is used in calculating the price index in the overlap period and the price for the new variety is used in the next period[15]. The price differential between the two varieties never affects the index, and the adjustment for quality change is equal to this differential. In this case, the market has determined the value of the quality adjustment.[16] This method can only be applied if the old and new varieties are available in the same time period, and this is frequently not the case.[17]

A more frequently used method is called "linking" or "splicing". If there is no overlap period for price observations for the old and new varieties, then the price of the old variety should be estimated for the current period to create an overlap price. This estimated price is obtained by using the price change of similar items for the current period to impute a price change for the old variety. The price difference between the imputed price for the old variety and the price of the new variety in the current period is the estimated quality adjustment.

This procedure is used frequently because of its simplicity, but there may be an inherent bias built into the methodology, particularly where major model changes are occurring. Major price changes frequently occur at the time new varieties or models of a

---

[14] Most countries use some form of a Laspeyres price index. If price movements are estimated using long-term relatives (the standard Laspeyres formula), then the base price may be adjusted proportionately for the estimated quality difference. If price movements are estimated using short-term relatives (modified Laspeyres), then the adjustment would be implicitly made by using the price change of the old variety this month and the new variety next month. Some examples of adjustments that are used with short-term relatives are presented in more detail in the next section.

[15]If the standard Laspeyres formula is used to compute a long-term relative, the base price should be adjusted by the ratio of the new variety's price to the old variety's price in the overlap period.

[16]The market price of new features can also be used for making direct quality adjustments. For example, when radial tires became a standard feature on new automobiles in the US, the price of adding the optional radial tires was used as the value of the quality adjustment.

[17] Another important factor is whether the old variety is only available at clearance prices. This occurs when only a few are available and the shops in which they are sold have deep discount prices in an effort to sell quickly the remaining supply of the old variety. A careful evaluation must be made as to whether such discount prices adequately reflect the quality difference.

product are introduced, and this is quite common with new vehicles, electronic equipment and apparel items. As the new varieties are introduced, a substantial supply of similar products are still available which show little price change or, in fact, may actually be declining in price. The use of the price change for these products to impute price change for new models would cause a downward bias in the price index due to an underestimation of the true price change for the new varieties.

If possible, it is better to use an imputation procedure based on the price change between old and new models, of the same item, where the old and new models are of comparable quality, [18] i.e., any price difference between the old and new models can be reflected in the index as genuine price change. Thus, for the variety where a quality change **has** occurred, a price can be imputed using the price change between the old and new models of the closest variety for which old and new models were comparable. This is referred to as "class mean" imputation in which a substrata with distinguishable price change attributes has been identified within an existing strata. The standard procedure—"overall mean" imputation—in which all observations in the strata are used, is not applicable for all cases. This additional imputation procedure adds more complexity, but reduces two types of bias: (1) bias from ignoring quality change altogether and including all price movements as price change and (2) bias from over adjusting for quality change by including some pure price change as quality change.[19]

A cautionary note is presented here as a reminder. No matter how trivial the effects of quality change may be or how sophisticated the models and procedures used to deal with it, adjustment for quality change is still an art for price index practitioners. Ignoring quality changes can result in substantial overstatement of price change as price increases due to quality improvements are included in our price indices. If one tries to correct for this problem through extensive use of linking observations when replacement varieties are introduced, an over adjustment can occur with some pure price change being counted as quality change and excluded from the price indices[20].

---

[18]Varieties of comparable quality are those items whose replacements have been declared comparable after review by knowledgeable analysts or have already been quality adjusted through one of the "direct" methods. They are new varieties for

[19] The case of apparel items, as discussed earlier, usually requires additional procedures to control substitutions depending on the season of the year, e.g., fall/winter versus spring/summer clothing.

[20]A pertinent example is products with annual model changes. Towards the end of the model year, prices are reduced to clear stocks. When the new models appear, if they are comparable to the old, no problems occur because the price of the replacement model can be directly compared with the price of the old product. If, however, the new model is not comparable to the old and its price is linked to the discounted price of the old, a downward bias will occur in the index since all the price difference between the two models is treated as being due to quality changes whereas some is due to genuine price effects, i.e., the end of year discounting. A more appropriate approach would be to compare the price of the replacement

(continued...)

The approach articulated and presented here provides a middle ground that has practicality for many countries. The following section provides some examples of how this procedure can be applied.

B. Using Imputation Techniques for Quality Adjustment

Consider the case where a modified Laspeyres formula is being used, i.e., if the following estimator is used at the elementary aggregate level:

$$I_t = \Sigma\, w_{i,o}\,(P_{i,t} / P_{i,t-1}) * I_{t-1}$$

where I_t is the current elementary index, $w_{i\,o}$ is the expenditure share of variety i within the item stratum during the base period, $P_{i\,t}$ is the current average price of variety i, and $P_{i\,t\text{-}1}$ is the average price of variety i in the previous period. If there are no expenditure weights within the stratum, then $w_{i\,o}$ would be replaced with 1/n, where n is the number of varieties for which prices are collected within the stratum. If there is only one variety selected for the stratum, then $w_{i,o} = 1$.

Three examples are presented below. The first uses the overlap price imputation procedure, the second uses a linking procedure with overall mean imputation, while the third demonstrates the linking procedure with class mean imputation. In these examples, there are prices collected for several varieties of an item and one item experiences a quality change.

Overlap Price Imputation

As previously discussed, information for making direct quality adjustments is often not available. More realistically, the price information for the two varieties will have to be used to estimate the quality difference. If the old and new variety were available at the same time (overlap pricing), then the price difference between the two can be used as an estimate of the quality difference.

The following example illustrates the use of the overlap procedure.

---

[20](...continued)
to the last available normal price of the old model (as discussed in section III D) or to use the class mean imputation technique for which an example appears in the next section.

Variety	Price Index Month 1	Average Price in Month 1	Average Price in Month 2	Month 2 Price Relative	Price Index Month 2
Prod 1		150	160	1.067	
Prod 2		225	250	1.111	
Prod 3		(140)	-	-	
Sub 1		160	180	1.125	
All items	133.3			1.101	146.8

Within this item index the sample consists of three varieties of products. When the data collector visits the shop in Month 2 to collect prices, Prod 3 is permanently unavailable. The shop manager and the data collector determine that Sub 1 with a price of 180 is the most similar product to Prod 3. Review of the differences between the products indicates that Sub 1 has a couple of added features and is not comparable in quality to Prod 3. The shop manager was able to determine that Sub 1 was available during the last price collection period and its price was 160. The price change for Sub 1 (180/160) can now be used to calculate the index this month. In this example it is assumed that the observations are equally weighted in calculating the index. The quality difference that has been excluded from the index is the price differential between Sub 1 and Prod 3 (160 - 140 = 20). If Sub 1 had been compared directly to Prod 3, then the item index would have risen by a larger amount.

Overall Mean Imputation

Frequently, the price of the replacement variety will not be known for the previous pricing period. An estimate of the price of the missing item can still must be made. The following example illustrates how to make such an estimate and, therefore, implicitly adjust for quality differences.

Variety	Price Index Month 1	Average Price in Month 1	Average Price in Month 2	Month 2 Price Relative	Price Index Month 2
Prod 1		150	160	1.067	
Prod 2		225	250	1.111	
Prod 3		140	*(152)*	1.089	
Sub 1		---	180		
All items	133.3			1.089	145.1

To get the required price ratio one can estimate (impute) the price of the discontinued variety by multiplying the price of Prod 3 in the previous period by the average price change of the other varieties available this month. For Prod 1 and Prod 2, the price change is calculated between Month 1 and Month 2 as (1.067 + 1.111)/2 = 108.9. This is the price relative for the item[21] and can be multiplied by the price of Prod 3 in the previous period (140) to get an estimated (imputed) price in Month 2 of 152. The amount of the quality adjustment is the difference between the price for Sub 1 and this estimated price (180 - 152 = 28).

Class Mean Imputation

The previous example assumed that there was no difference in pricing strategy among the product varieties within the stratum. Assume that Prod 1 represents a variety that was

[21] Note that in this and subsequent examples the price relative for the item is based on the average price change of the two varieties for which actual prices were collected. An alternative approach would be to calculate the average prices of Prod 1 and Prod 2 and compute the price relative using these average prices.

If the price levels within a strata vary substantially, the average of price relatives may be used so that the high priced items do not dominate the relative price change. For example, if we had a stratum with one variety that increased from 10 to 11 and another variety that increased from 100 to 102, using the average of price relatives would give a result of 6 % while using the change in average prices would yield 2.7%. Ideally, we would like to have weights for each variety to derive a weighted average, but this does not usually occur. If each variety is equally representative of the universe of expenditures, then the average of relatives (6%) should be used. If the higher priced item is representative of about 90 percent of the universe, then the change in average prices should be used.

available in both periods and that Prod 2 represents a new model in month 2 that was deemed comparable in quality to the previous model priced in month 1. Since Sub 1 is not comparable in quality to Prod 3, the price change could be imputed by using only the observed experience from Prod 2. In such a case the price relative would be 1.111, the imputed price would be 156. Then calculate the price relative for the item index would be calculated as the average price relative of the three products (i.e., including the price relative for the imputed variety) as shown in the following table.

Variety	Price Index Month 1	Average Price in Month 1	Average Price in Month 2	Month 2 Price Relative	Price Index Month 2
Prod 1		150	160	1.067	
Prod 2		225	250	1.111	
Prod 3		140	*(156)*	1.111	
Sub 1		---	180		
All items	133.3			1.096	145.1

In this example, the amount of the implicit quality adjustment is the difference between the prices for Sub 1 and Prod 3 (180 - 156 = 24).

Summary

The results in these three examples are different because the information that is available is different. In the first example an actual observation of the prices for each variety during the same time period was available. In the second example there were no overlapping observations so one was estimated using all the information that was available. In the third example the sample is partitioned into substrata and only the information that was most relevant to the case at hand was used to estimate the overlap price.

The following table summarizes what effect each of these procedures would have on the strata price index. In month 1 the price index stays the same. In month 2 the strata index changes due to the type of imputation used.

Table 4. Effect of quality adjustment imputation on price change

	Month 1	Month 2	%change
No quality Adj.	133.3	153.9	15.5
Overlap Price	133.3	146.8	10.1
Overall Mean	133.3	145.1	8.8
Class Mean	133.3	146.1	9.6

The first line of table 4 shows that if no quality adjustment were made in the above example the price change would be 15.5%. Using different sources of information, adjustments have been made that result in reductions of about 6 percentage points for quality change. In this example the overall mean imputation procedure and the class mean procedure yield different results because of the different assumptions about the best estimator for underlying price trends within the strata.

If the change in average prices for the varieties in each strata were used to calculate the imputed prices instead of the average of the price relatives, the general observation about the effects would be the same but the specific results would differ. The overall percent change with no quality adjustment would be 14.6%, overlap pricing would result in a change of 10.3%, overall mean imputation would show a change of 9.1%, and class mean imputation would show a change of 9.9%. Usually, the change in average prices would be used in strata which contain relatively homogeneous varieties and for which the price levels of the observations are not that different.

V. Conclusions

The measurement of price change is made more difficult when sample observations are unavailable in the market at different periods during the year. While there are numerous reasons for the absence of commodities and services from the market place, the primary reasons can be categorized into three types—seasonal effects, discontinued products and temporary supply shortages. Whenever a product in a sample of prices is missing, it is important to maintain the continuity of the sample by either estimating the missing price if it is thought that the product will return to the market, or by finding a replacement product that is as similar as possible if it is thought that the product will not return.

If the prices are unavailable because of seasonal patterns or short-term disruptions in supply, it is necessary to estimate the current price of the missing product using the price trends of similar products because it is certain that the missing product will become available in the future. However, if the sampled products are discontinued with little likelihood of again appearing, they should be replaced, if possible, with products that are of the same quality. If the only possible replacements are of a different quality, then the compiler needs to adjust for the value of the quality difference so that it is not reflected as price change.

There are a number of techniques that can be used to estimate missing prices, either for individual observations or for aggregate indices. The methods which are recommended in this paper are those that use the price trends from available sample observations. In each case, the imputation technique involves the use of price trends which would be the most representative of the trend for the missing observation. These will differ depending on the reason that the price is unavailable. One frequently used method, which holds prices or indices constant at previous levels, is discouraged unless there are clear indications that prices have not changed.

In the case of the permanent disappearance of a product in which the selected replacement is of a different quality, imputation techniques can be used to provide an implicit solution to the quality adjustment process. The price trend of similar products can be used to represent the price trend of the discontinued product in the sample. No comparison is made between the price of the old product and the new product. The difference in price between the imputed price of the old product and the price of the new product is the implicit quality adjustment. This technique is preferred rather than making a direct comparison between the prices of the old and new product or allowing for no price movement whatsoever by using the last available price for the old product.

References

Advisory Commission to Study the Consumer Price Index. 1996. *Toward a More Accurate Measure of the Cost of Living*, Final Report to the Senate Finance Committee, December 4.

Armknecht, Paul and Brent Moulton. 1995. "Quality Adjustment in Price Indices: Methods for Imputing Price and Quality Change," unpublished manuscript, October.

Armknecht, Paul, Brent Moulton and Kenneth Stewart. 1994. "Improvements to the Food at Home, Shelter and Prescription Drug Indexes in the U.S. Consumer Price Index," *International Conference on Price Indices*, October 31-November 2, Ottawa: Statistics Canada, pp. 50-72.

Armknecht, Paul and Donald Weyback. 1989. "Adjustments for Quality Change in the U.S. Consumer Price Index," *Journal of Official Statistics*, Vol. 5, No.2, pp. 107-123.

Kokoski, Mary. 1993. "Quality Adjustment of Price Indexes," *Monthly Labor Review*, December, pp. 34-46.

Lequiller, Francois I. and Kimberly D. Zieschang. 1994. "Drift in Producer Price Indices for the Former Soviet Union Countries," *International Monetary Fund Staff Papers*, September, pp. 526-32.

Liegey, Paul. 1993. "Adjusting Apparel Indexes in the CPI for Quality Differences," in *Price Measurements and Their Uses*, M. Foss, M. Manser and A. Young, eds. NBER Studies in Income and Wealth Vol. 57. Univ. of Chicago Press, pp.209-226.

Liegey, Paul. 1994. "Apparel Price Indexes: Effects of Hedonic Adjustment," *Monthly Labor Review*, May, pp. 38-45.

Moulton, Brent. 1996. "Bias in the Consumer Price Index: What is the Evidence," *Journal of Economic Perspectives*, Fall, Vol. 10 No. 4, pp. 159-77.

Reinsdorf, Marshall., Paul Liegey and Kenneth Stewart. 1996. "Avoiding Downward Bias from Linking in the U.S. Consumer Price Index," BLS Working Paper 276, April.

Schultz, Bohdan J. 1983. "Linking Price Index Numbers," in *Price Level Measurement*, W.E. Diewert and C. Montmarguette, eds., Ottawa: Statistics Canada, 537-66.

Turvey, Ralph, et al. 1989. "Chapter 6. Computation," *Consumer Price Indices: An ILO Manual*, International Labor Office, Geneva, pp. 85-111.