ESTIMATING VARIANCES FOR THE U.S. CONSUMER PRICE INDEX FOR 1987-1991

Sylvia G. Leaver and David C. Swanson
U.S. Bureau of Labor Statistics
Sylvia G. Leaver, U.S. Bureau of Labor Statistics, 2 Massachusetts Avenue, N.E., Washington, D.C. 20212

KEY WORDS: random groups, expenditure weights, linearization

In this paper we present estimates of the sampling variance of price change from the Consumer Price Index, for the years 1987 through 1991. These estimates are the first published estimates of the Consumer Price Index's variance since its revision in 1987. This paper presents estimates of 1-, 2-, 6-, and 12-month price change variance, computed at the national level for All Items, seven major product groups, and three housing subgroups, and at the regional level for All Items.

1. Introduction and Findings

This paper marks the first release of variance estimates of price change from the Consumer Price Index since its 1987 revision. These estimates are for January 1987 through December 1991. They follow estimates for the variance at the U.S. level of the index and for price change for 1978-86 which were published in these proceedings in August 1991.

Estimates of price change standard error are quite stable for the All Items index and most major groups at the U.S. level. However, they are in most cases larger and more variable than those estimated for the 1978-86 period. Standard errors exhibit an upward trend and considerable variability for some item groups and for Apparel and Upkeep in particular.

The methodology for estimating variances is given and sources of increased variability in variance estimates are discussed.

2. Background: The Consumer Price Index

The CPI is estimated on a monthly basis for all consumer items for the total urban U.S., as well as at numerous other levels defined by the geographic area, group of items, and population (type of consumer). There are two populations for which the CPI is currently computed: urban wage earners and clerical workers (CPI-W) and all urban consumers (CPI-U). This report focuses on estimates of variance of price change computed for the all urban consumers population.

For a full discussion of the CPI since 1987, we refer the reader to the *BLS Handbook of Methods* (1988). A description of the features of the CPI which pertain to estimating variances is also given in Leaver (1990) and Leaver, Johnstone, and Archer (1991).

We shall refer to the following item classifications:

(a) The *item stratum* is the most refined classification of commodities or services for which estimates of expenditures are computed and indexes are published. Examples are bread and college tuition.

(b) Item strata are grouped into *expenditure classes* (EC's). Examples include bakery products and educational expenses.

(c) Expenditure classes are combined into seven published *major groups*: Food, Housing, Apparel and Upkeep, Transportation, Medical Care, Entertainment, and Other Commodities and Services.

The geographic areas for which the CPI is published and for which variances are computed are the following:

(a) The *index area* is the basic geographic area for which a fixed sample of commodities and services is priced monthly or bimonthly. Index areas are metropolitan statistical areas (MSAs), or groupings of MSAs and areas which are the urban parts of non-metropolitan counties. Examples are the Chicago MSA and medium sized metropolitan areas in the Northeast.

b) The four Census *regions* are Northeast, North Central, South, and West.

c) The urban *U.S.*

The CPI is a Laspeyres index which is a ratio of the costs of purchasing a set of items of constant quality and quantity in two different time periods. Let $IX(i,m,t,0)$ denote the index for item stratum i in index area m, which compares expenditures between time t and time 0, the base or reference period. The base period is June 1983 for every item group reported here, but is December 1982 for Homeowner's Equivalent Rent. Then

$$IX(i,m,t,0) = 100 * CW(i,m,t)/CW(i,m,0),$$

where $CW(i,m,t)$ and $CW(i,m,0)$ denote estimates of expenditures, termed *cost weights*, for time t and for the base or reference period 0, respectively, for the stratum-area.

$IX(i,m,t,0)$ may also be expressed in terms of the index for a previous period:

$$IX(i,m,t,0) = 100 * [R(i,m,t,t-1) * CW(i,m,t-1)]/CW(i,m,0),$$

where $CW(i,m,t-1)$ denotes the cost weight for the stratum-area at time t-1 and $R(i,m,t,t-1)$ denotes its estimate of price change, termed a one-period relative, from time t-1 to time t, computed by:

$$R(i,m,t,t-1)=\sum_{qj} W_{iqj}\,(P_{iqjt} / P_{iqja})\,/\,\sum_{qj} W_{iqj}\,(P_{iqjt-1} / P_{iqja}),$$

where P_{iqjt} is the price of the q^{th} quote, i.e., a sample item in a sample outlet, in the j^{th} outlet in time period t for item stratum i; P_{iqja} is the price of the q^{th} quote in the j^{th} outlet in the outlet expenditure frame development reference period a; and W_{iqj} is a composite sampling weight for the q^{th} quote in

the j^{th} outlet for item stratum i. The sum given here is over all quotes and outlets for the item stratum in the index area.

Cost weights for item strata are updated on a monthly or bimonthly basis using one period price relatives defined above. They are summed to estimate cost weights for higher level item aggregates (HLIAs) such as expenditure classes, major groups, and all items, denoted I, and higher level geographic aggregates (HLGAs), such as regions and all U.S., denoted M, by:

$$CW(I,M,t) = \sum_{m \in M} \sum_{i \in I} CW(i,m,t).$$

Thus the formula for the index for any higher level item-area aggregate is:

$$IX(I,M,t,0) = 100 * CW(I,M,t) / CW(I,M,0).$$

The CPI sample uses a multi-stage probability design. The sample for any index area consists of one or more primary sampling units (PSUs), which are either MSAs or the urban parts of non-MSA counties. [See Dippo and Jacobs (1983).] For purposes of variance computation and operational manageability, samples for all index areas are split into two or more disjoint subsets or replicates.

The item-outlet sample for the commodities and services component of the CPI was redesigned with the 1987 revision to allocate sample resources to minimize the variance of 6-month price change for the All Items index at the U.S. level, subject to operational cost limits. [See Leaver et al., (1987).] This allocation was based on models for components of sampling variance and operational costs. Because the dominant component of price change sampling variability was the between-outlet component, the general tendency of the revised design was to shift sample resources away from sampling multiple items within each outlet to sampling more outlets. The revised design also tended to shift sample resources away from smaller sample cities towards larger sample cities, and away from less variable item groups towards more variable item groups. Approximately one fifth of the sample cities were introduced into the CPI sample with the revised design each year, beginning in 1986. Two additional replicate panels for the Chicago MSA and the New York City portion of the New York MSA were introduced in March of 1990 and 1991, respectively. Thus, the revision sample design was not fully introduced until March 1991.

3. Estimating Variances of the Index and Price Change

Variance estimates for the CPI series starting with the 1987 revision differ from those computed for 1978-1986 in two important ways. The first difference is that the revision expenditure estimates for December 1986 (which were updated from the 1982-84 Consumer Expenditure Survey) were independently estimated for each replicate. Thus the variances computed for the 1987 revision index series directly incorporate the sampling variance attributable to the estimation of expenditure weights from the 1982-84 CE Survey. Variances for 1978-1986 were estimated by first estimating the conditional variance of the index or price change given the December 1977 expenditure weights, then estimating the unconditional variances of the December 1977 expenditure weights, and then combining the conditional index or price change variances with the variances of the expenditure weights to produce unconditional index or price change variances.

The second difference between the estimates is that the 1987 revised CPI variances incorporate between index area covariances for HLGAs such as regions, city-size classes, and All Cities.

Variances of price change for the CPI were computed using a hybrid methodology, combining random group variance estimation for cost weight variances with linearization for price change variances.

As noted above the full sample for any item and index area comprises two or more replicate panels, half of which were designated "odd" and the other half "even."

Each index area is in one of four Census regions. Each region can further be divided into two major areas, one composed of the self-representing (A) index areas and one composed of the non-self-representing (non-A) index areas; thus there are eight major areas in the Nation. For each HLGA larger than one index area, estimates of between-index-area covariances for each pair of different index areas in the same major area in the HLGA were included in variance estimation.

To estimate the variance of price change, consider the 2n x 1 vector $CW_f(I,MA,t,t')$ of full sample cost weights for an item or item aggregate I, whose elements are the cost weights for each of n index areas in major area MA in months t and t':

$$CW_f(I,MA,t,t') =$$

$$[CW_f(I,m_1,t),...,CW_f(I,m_n,t),CW_f(I,m_1,t'),...,CW_f(I,m_n,t')]^T$$

Similarly denote $CWA_1(I,MA,t,t')$ and $CWA_2(I,MA,t,t')$ to be the corresponding vectors of average replicate cost weights for the item and major area, with 1 denoting the average of r/2 odd replicates and 2 denoting the average of r/2 even replicates:

$$CWA_1(I,MA,t,t') = 2/r \sum_{j\ odd} CW_j(I,MA,t,t')$$

$$CWA_2(I,MA,t,t') = 2/r \sum_{j\ even} CW_j(I,MA,t,t').$$

Let $A_{M,MA}$ be the 2 x 2n area aggregation matrix for any area aggregate M with component index areas belonging to major area MA where

$A_{M,MA}(1,j) =$ 1, if index area j∈M,
0, otherwise; j= 1,...,n
$A_{M,MA}(1,j) =$ 0, j= n+1,...,2n

$A_{M,MA}(2,j) =$ 0, j= 1,...,n, and
$A_{M,MA}(2,j) =$ 1, if index area j∈M,
0, otherwise; j=n+1,...,2n

The 2 x 2 covariance matrix of cost weights for any area aggregate M within a major area MA, $W(I,M,MA,t,t')$, is estimated by:

$$W(I,M,MA,t,t') = A_{M,MA}\ 1/2\{\sum_{j=1}^{2} DCWj\ DCWj^T\}\ A_{M,MA}^T,$$

where $DCWj$ is the difference vector:
$DCW_j = [CWA_j(I,MA,t,t') - CW_f(I,MA,t,t')], j= 1,2.$

Under the assumption that the cost weights for index areas are independent between major areas, the cost weight covariance matrix $W(I,M,t,t')$ for any HLGA M comprising index areas in more than one major area, such as All Cities, regions, and size classes, is computed by summing $W(I,M,MA,t,t')$ over all major areas:

$$W(I,M,t,t') = \sum_{MA=1}^{8} W(I,M,MA,t,t')$$

The variance of a k-month price change from month t-k to month t for item aggregate I and area aggregate M,

$$PC(I,M,t,t\text{-}k) = 100 * [(CW(I,M,t)/CW(I,M,t\text{-}k))\text{-}1]$$

can be estimated by a first order Taylor series approximation of the ratio of two cost weights at times t and t'=t-k :

$$Var[PC(I,M,t,t\text{-}k)] = L(I,M,t,t\text{-}k)\ W(I,M,t,t\text{-}k) L(I,M,t,t\text{-}k)^T$$

where $W(I,M,t,t\text{-}k)$ is as defined above and the 2 x 1 linear transformation vector $L(I,M,t,t\text{-}k)$ is given by:

$$L(I,M,t,t\text{-}k) = 100 * [1/CW(I,M,t\text{-}k), - CW(I,M,t)/(CW(I,M,t\text{-}k))^2].$$

4. Findings

Figure 1 displays the standard error of 1-month price change at the U.S. level for All Items, and Housing, which represents over 40% of consumer expenditures nationally, plotted as a function of time for the months January 1978 - December 1991. Figures 2-8 display the standard error of 12-month price change for All Items, 7 major groups, and 3 housing subgroups as a function of time for the same period. Units of measure are percentages.

In nearly every case, the standard error of price change exhibits a gradual climb and then either a leveling off or decrease over the 5 year period of this study, denoted in the graphs as months 109-168. Some major groups, most notably Apparel, Housing, and all three subgroups within Housing, exhibit seasonal behavior in their standard errors also.

An abrupt jump in 1- and 2-month price change standard error, due to data capture and sample initiation difficulties in the first three months of 1987, occurs in Food at the outset of the revision. Figure 11 graphs the number of quotes used in the commodities and services component of the CPI over the period of the study. The continuing decline in the number of quotes after 1987 reflects the revised sample design described earlier.

Additional abrupt jumps occurred in Entertainment, Medical Care, Other Commodities and Services, and Shelter, which were due to real and substantial price change in item strata. In particular, Lodging While Out of Town is the single most important contributor to Shelter variance and accounts for the climb and higher degree of fluctuation seen in the Shelter, Housing, and All Items standard error graphs. This stratum, whose weight was substantially increased after the revision of the 1987 item sample design, is currently undergoing pricing sample augmentation.

While more variable than corresponding estimates for the 1978-86 period, price change variance for All Items, Food, and Housing excluding Lodging While Out of Town remain generally stable and uniform over the 1987-91 period, especially when compared to the actual values of price change estimated for that period.

Figures 9 and 10 display the All Items price change ± two standard errors over time for 1- and 12-month intervals, respectively. This stability in price change standard error is due to the high correlation of the index between months, that is, between the numerator and denominator indexes of price change estimates.

Among the major groups, price change standard errors for Apparel and Upkeep are most extreme. They also exhibit greater seasonal variation than they did in 1978-86, but do not climb as steeply. The reasons for this behavior are partly due to the seasonal, ephemeral character of items in this major group, and the attendant difficulty in maintaining a price series for them.

Other sources of nonsampling error which are confounded in price change sampling error estimates include variability introduced in item substitution, stratum-level price relative imputation, and within stratum imputation of current and reference period prices. Stratum-level price relative imputations, shown for All Items in both weighted and unweighted form in Figure 12, increased for both full sample and replicate relatives over the period of the study. Of particular note is the frequency with which replicate price

relatives alone were imputed. These increases are being investigated but are in part due to changes made with the revision sample redesign in which the minimum number of sample outlets per index area replicate per CPOPS category dropped from 2 to 1. Larger than expected outlet out-of-scope rates have resulted in the complete loss of sample in some replicates and attendant thinning of full sample data. Variance estimates using imputed replicates may overstate true price change sampling variability; however, within-stratum imputation, which is also currently being researched, may counteract this effect.

Table 1 gives median price change and standard errors for 1-, 2-, 6-, and 12-month intervals for 1987-1991. It also gives ratios comparing 2-, 6-, and 12-month price change standard errors to 1-month price change standard errors, and ratios comparing median standard errors before and after the 1987 revision.

Revision price change standard errors for All Items and many item groups are larger than for the prerevision period, with increases ranging from 5% to 60% for 6-month standard error. Four major groups (Food, Fuels and Utilities, Medical Care and Entertainment) show improvements in 12-month price change standard errors, with decreases ranging between 17% and 24%.

As expected, price change standard errors increase with lag, though not linearly with respect to the length of the interval. The median standard error for 12-month price change is approximately 1.9 times the median standard error for 1-month price change and approximately 1.4 times the median standard error for 2-month price change for the All Items index. Similar behavior holds for most other major groups at the All U. S. and regional levels.

Table 1 - Median Price Change, Median Price Change Standard Errors by Major Group for 1-, 2-, 6- and 12 Month Intervals

Major Group	1-Mo Med PC	1-Mo Med SE	Ratio Rev/ Pre87	2-Mo Med PC	2-Mo Med SE	Ratio 2/1	Ratio Rev/ Pre87	6-Mo Med PC	6-Mo Med SE	Ratio 6/1	Ratio Rev/ Pre87	12-Mo Med PC	12-Mo Med SE	Ratio 12/1	Ratio Rev/ Pre87
U. S. All Cities															
All Items	0.370	0.074	1.43	0.748	0.103	1.39	1.49	2.316	0.130	1.75	1.43	4.635	0.143	1.93	1.16
Food	0.302	0.081	1.03	0.599	0.100	1.22	1.01	2.457	0.131	1.60	0.93	5.073	0.157	1.92	0.83
Housing	0.294	0.146	1.31	0.680	0.211	1.45	1.43	1.938	0.242	1.66	1.40	3.872	0.282	1.93	1.25
Shelter	0.370	0.202	1.17	0.790	0.298	1.48	1.30	2.314	0.327	1.62	1.29	4.834	0.398	1.97	1.13
Fuels & Util	0.110	0.159	0.98	0.346	0.242	1.52	1.06	1.434	0.367	2.31	1.05	3.052	0.357	2.25	0.80
HH Furn & Opn	0.145	0.244	1.62	0.244	0.293	1.20	1.54	0.845	0.410	1.68	1.60	1.948	0.505	2.07	1.54
Apparel	-.042	0.480	1.87	0.581	0.699	1.46	1.99	1.667	1.026	2.14	1.94	4.286	0.892	1.86	1.28
Transportation	0.324	0.077	1.34	0.723	0.100	1.31	1.24	2.108	0.174	2.27	1.41	4.037	0.206	2.69	1.18
Medical Care	0.604	0.100	0.96	1.246	0.137	1.38	0.88	3.802	0.223	2.24	0.81	8.093	0.308	3.09	0.76
Entertainment	0.369	0.191	1.13	0.748	0.258	1.35	1.10	2.290	0.364	1.90	0.96	4.568	0.436	2.28	0.81
Other C&S	0.441	0.116	1.38	1.000	0.155	1.34	1.21	3.781	0.281	2.42	1.28	7.623	0.392	3.39	1.17
Region 1 - Northeast															
All Items	0.368	0.161	1.59	0.824	0.236	1.46	1.77	2.646	0.317	1.97	1.82	5.322	0.321	1.99	1.32
Region 2 - North Central															
All Items	0.364	0.101	1.17	0.763	0.156	1.54	1.36	2.213	0.184	1.82	1.25	4.214	0.232	2.30	1.18
Region 3 - South															
All Items	0.304	0.105	1.11	0.646	0.138	1.32	1.07	2.124	0.180	1.72	1.10	4.215	0.200	1.91	0.91
Region 4 - West															
All Items	0.400	0.130	1.07	0.720	0.184	1.42	1.04	2.177	0.251	1.93	1.17	4.599	0.306	2.36	1.07

It is also useful to note that standard errors of 1-month price change for All Items are, in times of low inflation, on the same order as the price change estimates themselves. Figures 9 and 10 depict interval estimates of price change for 1 and 12 month lags for the revision years. In most cases for 6- and 12-month lags, the standard errors of small price changes are much smaller than the price change estimates themselves.

5. Conclusions

The estimates of price change standard error for 1987-91 raise important issues regarding the sample design for the survey. In particular, sample attrition and insufficiency are principal causes of the observed increases in price change standard error. Extreme price changes in a few quotes can have a dramatic effect on the variance of a major group. This

is particularly the case when an item is heavily weighted or when significant imputation occurs. This points to the need to scrutinize imputation assumptions and account for effective stratum relative importance in sample allocation.

Standard error of price change for All Items and the most important major groups appears to be a fairly stable measure over the period of this study, which at the All Items level has been a period of relatively stable inflation. This finding is largely attributable to the high correlation of lagged indexes over time. However, the similarities in magnitude of standard errors and price change for small values of price change, particularly for 1- and 2- month lags, indicate the need for caution in inferring level and direction of price change trends.

6. Acknowledgments

The authors would like to acknowledge Michael P. Cohen for his development of the computing formulae for price change variances, Anne Asher, Andrew Chen, Richard P. Duquette, Larry Goldsmith, Helen Hsu, James E. Johnstone, Peter Lorenzetti, Leanna McEnearney, Virasack Savanh, and Connie Phillips for replicate index series construction and variance estimation systems development, and William Smyth for production processing of variance estimates. Richard Kerr, Claire McAnaw, Mary McMichael, and Frank Ptacek provided valuable information in researching particular instances of price change.

The authors also wish to acknowledge Paul A. Armknecht and the late Curtis A. Jacobs for their support and encouragement in this effort, and to thank Richard Valliant, Stuart Scott, Marshall Reinsdorf, Janet Williams, and Walter Lane for their careful reading of this and earlier drafts of this paper and for their helpful comments.

7. References

Bureau of Labor Statistics, *BLS Handbook of Methods*, U.S. Government Printing Office, 1988, 154-208.

Dippo, Cathryn S., and Jacobs, Curtis A., "Area Sample Redesign for the Consumer Price Index," *Proceedings of the Survey Research Methods Section*, American Statistical Association, 1983, 118-23.

Leaver, Sylvia G., Cohen, Michael P., Weber, William L., and Archer, Kenneth P. "Item-Outlet Sample Redesign for the 1987 U.S. Consumer Price Index," *Proceedings of the 46th Session, International Statistical Institute*, Tokyo, 1987, Vol. 52, Book 3, 173-185 .

Leaver, Sylvia G., "Estimating Variances for the U.S. Consumer Price Index for 1978-1986," *Proceedings of the Survey Research Methods Section*, American Statistical Association, 1990, 290-295.

Leaver, Sylvia G., Johnstone, James E., and Archer, Kenneth P., "Estimating Unconditional Variances for the U.S. Consumer Price Index for 1978-1986," *Proceedings of the Survey Research Methods Section*, American Statistical Association, 1991, 614-619.

Figure 1. 1-Month Price Change Standard Error vs Time for All Items and Housing

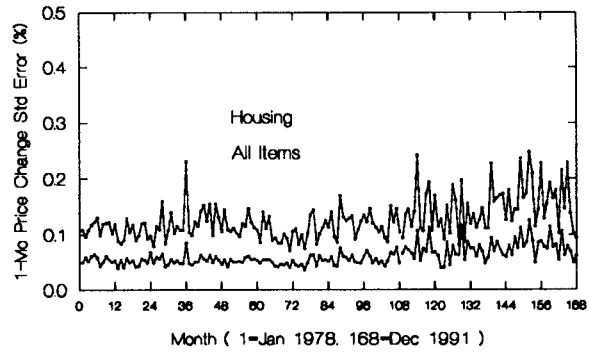

Figure 2. 12-Month Price Change Standard Error vs Time for All Items and Housing

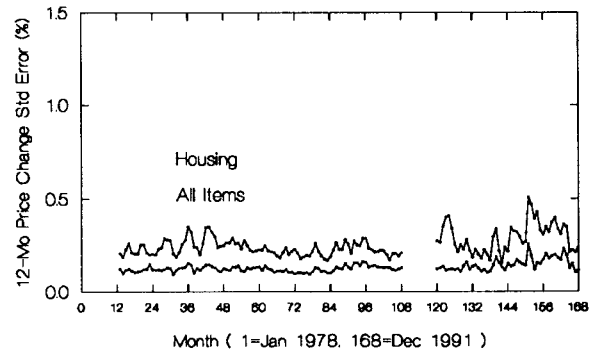

Figure 3. 12-Month Price Change Standard Error vs Time for Apparel, Other C&S

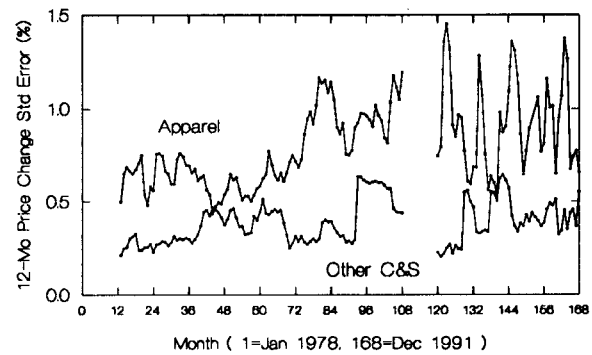

Figure 4. 12-Month Price Change Standard Error vs Time for Medical Care, Transportation

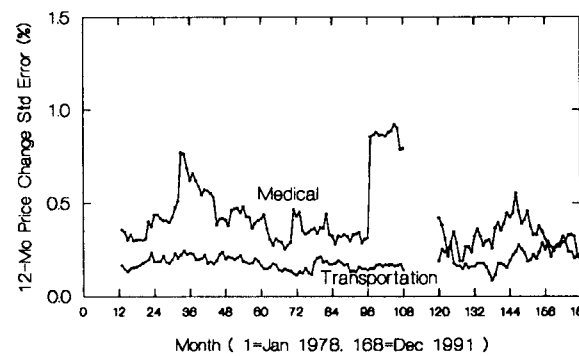

Figure 5. 12-Month Price Change Standard Error vs Time for Food, Entertainment

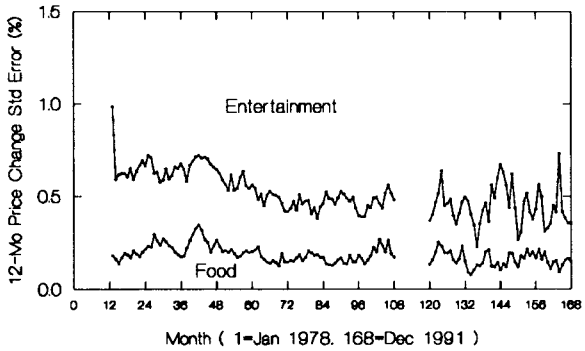

Figure 6. 12-Month Price Change Standard Error vs Time for Shelter

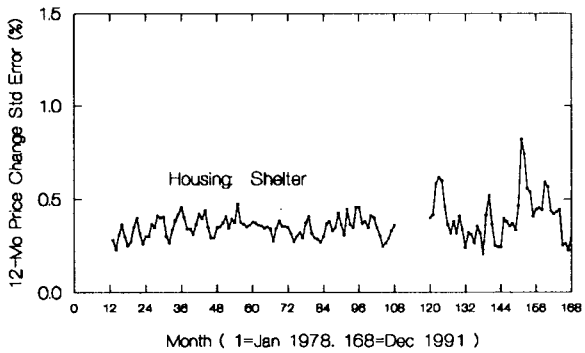

Figure 7. 12-Month Price Change Standard Error vs Time for Fuels and Utilities

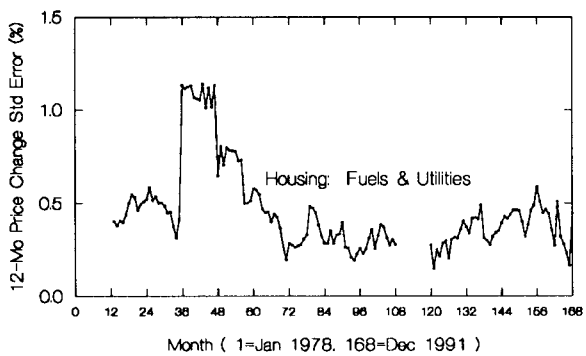

Figure 8. 12-Month Price Change Standard Error vs Time for HH Furnishings & Operation

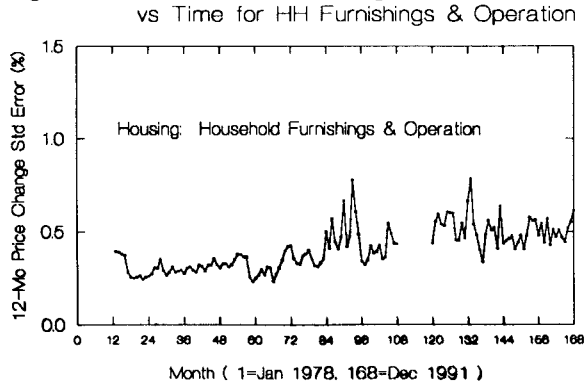

Figure 9. 1-Month Price Change + 2 Standard Errors vs Time for All Items

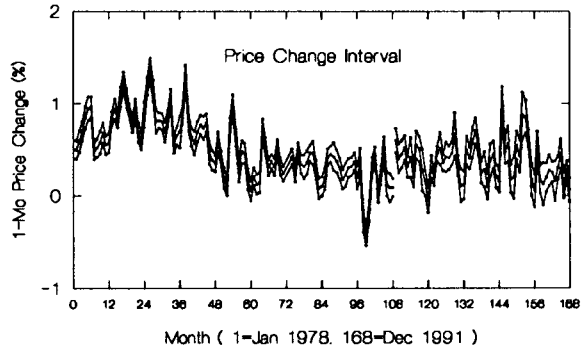

Figure 10. 12-Month Price Change + 2 Standard Errors vs Time for All Items

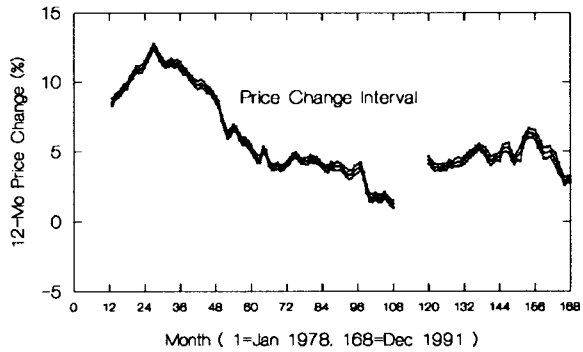

Figure 11. Usable Quotes vs Time for All Items

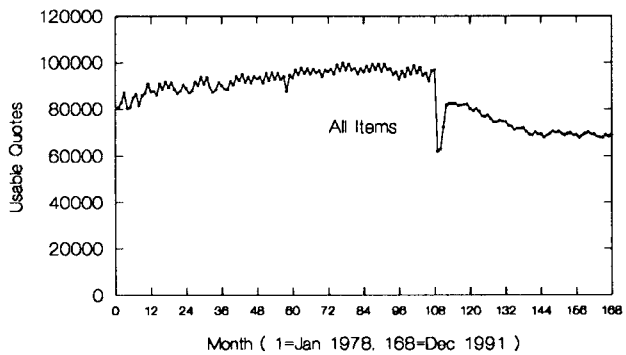

Figure 12. Price Relative Imputation Rates vs Time for All Items