ESTIMATING VARIANCES FOR THE U.S. CONSUMER PRICE INDEX FOR 1978-1986

Sylvia G. Leaver, U.S. Bureau of Labor Statistics
600 E Street, N.W., Room 5217, Washington, D.C. 20212

KEY WORDS: price index, variance, replication

In this paper we present initial results of a multi-phase project to compute estimates of the variance of the Consumer Price Index (CPI), and variance estimates of price change reflected by the index for the years 1978 through 1986, the period between the last two major revisions of the CPI. The project includes estimation of index and price change variances which are conditional on December 1977 expenditure weights derived from the 1972-1974 Consumer Expenditure Surveys, estimation of variances of the December 1977 expenditure weights, and combination of these estimates to produce unconditional variances for the index and price change for the 9-year period. Conditional variances for indexes were estimated via replication using random groups, while variance estimates for price change were estimated via a linear approximation using the index variance component estimates computed by replication.

1. Introduction and Findings

This paper marks the first release of estimates of variance for the Consumer Price Index. These estimates are for the time period January 1978 through December 1986. Estimates are given for the variance at the U.S. level of the index and for price change, that is, the percentage change in the index, for 1-, 2-, 6-, and 12 month intervals for the same months, at the all items level and for seven published major groups.

The variance of the CPI is seen to be an increasing function of time and value of the index, for the U.S. All Items index as well as for the indexes published at the U.S. level for major groups of commodities and services and housing. Variance of price change, however, exhibits varying behavior over time. It is quite stable for the All Items index and most major groups but exhibits an upward trend and considerable variability for Apparel in particular. Sources of the variability of variance estimates are discussed, and further estimation of cost weight variances and unconditional index and price change variances is described.

2. Background

2.1 The Consumer Price Index

For a full discussion of the CPI for 1978-1986, we refer the reader to the BLS Handbook of Methods, Vol. II (1984). The following is a brief description of the features of the CPI which pertain to estimating variances.

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 presently computed: urban wage earners and clerical workers (CPI-W) and all urban consumers (CPI-U). This report focuses on estimates of variance of the index and price change computed for the all urban consumers population.

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.* 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.

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

c) The urban *U.S.*

The CPI is a modified 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 time t for item stratum i in index area m, where 0 represents the base or reference period. Then

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

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, on stratum i.

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

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

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

$$CW(i,m,t) = CW(i,m,0) * \prod_{s=1}^{t} R(i,m,s,s-1)$$

Here $R(i,m,t,t-1)$ is computed by:

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

where P_{iqjt} is the price of the qth quote, i.e., a sample item in a sample outlet, in the jth outlet in time period t for item stratum i; P_{iqja} is the price of the qth quote in the jth outlet in the outlet expenditure frame development reference period a; and W_{iqj} is a composite weight for the qth quote in the jth outlet for item stratum i, the inverse of its probability of inclusion in the CPI sample. 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, and higher level geographic aggregates (HLGAs), such as regions and all U.S. by:

$$CW(I,M,t) = \Sigma_{m\in M} \Sigma_{i\in I} CW(i,m,t), \text{ where}$$

m denotes the index area in HLGA M and i denotes the item stratum in HLIA I.

This gives the computing formula for the index for any higher level item-area aggregate as:

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

The CPI is computed from measurements of price change on sampled commodities and services and housing, observed in sampled outlets in sampled geographic areas across the United States. Its sample is a multi-stage probability design. The sample for any index area consists of one or more primary sampling units (PSUs) which are either metropolitan statistical areas (MSAs), or the urban parts of non-MSA counties. [See Dippo and Jacobs, 1983.] Index areas can be classified into two types: self-representing areas consisting of one or more PSUs which were selected with certainty, such as New York, Los Angeles, and Chicago; and non-self-representing areas, whose sample consists of sets of two or more PSUs selected according to a probability sample. For purposes of variance computation and operational manageability, samples for all index areas are split into two or more disjoint subsets or replicates, historically called half-samples.

CPI item and outlet selection is performed independently for each PSU-replicate. Specific items are selected from each item stratum by a systematic probability proportional to size (PPS) procedure. Here the item weight is its relative importance with respect to expenditures within the stratum as measured by the Consumer Expenditure Survey.

Most of the sample frames and weights used in outlet selection are derived from the Current Point of Purchase Survey (CPOPS), a household survey conducted by the U.S. Bureau of the Census for the BLS. In this survey, each household is asked the names and addresses of the outlets from which purchases were made, and the dollar amount of those purchases by outlet, for each of a set of item classes known as CPOPS categories. A CPOPS category is a broad category of items which are normally sold in the same kinds of retail outlets. Each item within a stratum belongs to one and only one CPOPS category. Outlet frames and selection weights are derived from CPOPS survey data for each PSU, replicate, and CPOPS category. Here, the selection weight for a given outlet in a given CPOPS category is proportional to the ratio of the outlet's reported expenditures to total reported expenditures for the category.

In outlet selection for each PSU-replicate, outlets are sampled from frames corresponding to the CPOPS categories of selected items by a systematic PPS procedure, where the selection weights are those described above. Sampled items then are priced for the CPI on a monthly, bimonthly, or seasonal basis in the outlets accordingly selected.

3. Estimating Variances of the Index and Price Change

Conditional variances of the CPI were computed using a random group estimation method. As noted above, the sample for the CPI in any index area is partitioned into two or more disjoint replicates, also termed random groups. In self-representing index areas, comprising only one PSU, these replicate panels are disjoint subsets of the sample for the PSU. The sample for most self-representing index areas consists of two replicates. In non-self-representing index areas, each replicate consists of the sample for one or more of the sample PSUs in the index area. The number of replicates for non-self-representing index areas ranges from two to four.

Beginning in January 1978, price relatives for each item stratum were computed separately for the full sample and for each replicate in every CPI index area to produce separate cost weight series for all item strata for each replicate as well as the full sample for each area.

Cost weights for HLIAs were constructed for each index area at full sample and replicate levels using the index aggregation methodology described above. Item stratum

and HLIA index variance estimates were then computed for each index area using the following formula:

$$Var\,[IX(I,m,t,0)] = \frac{\sum_{j=1}^{r} [CW_j(I,m,t)-CW_f(I,m,t)]^2}{r(r-1)\,CW(I,m,0)^2}$$

where I denotes the item stratum or HLIA, r denotes the total number of replicates in the index area, $CW_f(I,m,t)$ and $CW_j(I,m,t)$ denote the cost weights for the full sample and replicate j, respectively in the index area.

Index variance estimates for HLGAs such as regions and the national level were then derived, assuming independence of indexes between areas, by summing over index areas within the HLGA:

$$Var\,[IX(I,M,t,0)] = \sum_{m \varepsilon M} Var\,[IX(I,m,t,0)]\,.$$

Variances of k-month percentage price change, defined as

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

were estimated based on a Taylor-series linear approximation of the ratio IX(I,M,t)/IX(I,M,t-k):

$$Var\,[100 * IX(I,M,t,0)/\,IX(I,M,t\text{-}k,0)] =$$
$$100^2 * [IX(I,M,t\text{-}k,0)]^{-2}\,\{Var[IX(I,M,t,0)]$$
$$+ [IX(I,M,t,t\text{-}k)]^2\;\; Var[IX(I,M,t\text{-}k,0)]$$
$$- 2\,IX(I,M,t,t\text{-}k)\,Cov\,[IX(I,M,t,0),\,IX(I,M,t\text{-}k,0)]\}$$

where $Cov\,[IX(I,M,t,0),\,IX(I,M,t\text{-}k,0)]$ is the covariance of the indexes in months t-k and t estimated by the analogous formulae:

$$Cov\,[IX\,(I,m,t,0),\,IX\,(I,m,t\text{-}k,0)] =$$

$$\frac{\{\sum_{j=1}^{r} (CW_j(I,m,t)-CW_f(I,m,t))(CW_j(I,m,t\text{-}k)-CW_f(I,m,t\text{-}k))\}}{r\,(r\text{-}1)}$$

$$Cov\,[IX(I,M,t,0),\,IX(I,M,t\text{-}k,0)] =$$
$$\sum_{m \varepsilon M} Cov\,[IX(I,m,t,0),\,IX(I,m,t\text{-}k,0)]\,.$$

We note here that the December 1977 cost weights for the full sample and all replicates for any series in any index area were equal in value. Variances computed using these cost weight series alone are termed conditional because they do not reflect the variance of the index or price change due to sampling variation of the initial expenditure levels. That is, they are conditional on the values of December 1977 expenditure estimates. Thus, the results presented in this paper are expected to be underestimates of unconditional variances. Additionally, for the time period covered by this report, base period expenditure estimates were not separately estimated for full and replicate samples, and thus are not distinguished.

It is also important to note that for the database from which these variances were estimated, cost weights for the Homeowner's Equivalent Rent (REQ) item stratum were computed for every month in the 9-year period, even though the REQ stratum was not officially incorporated in the index before January 1983. These REQ cost weights were used in computation of the index and price change and their variances for All Items and Housing for each month. Thus the estimates of the index and price change for All Items and Housing given here will not correspond exactly with their published values.

Later phases of the CPI variance estimation activity will include estimation of the variances of the expenditure weights for December 1977, estimation of between item stratum and between time covariances, and the combination of these with conditional variances given here to produce unconditional estimates of index and price change variances for 1978-1986. Results of these activities will follow in later reports.

4. Findings

Figures 1-2 display the standard error of the CPI at the U.S. level for the all items and the seven major groups, plotted as a function of time for the 108 months, January 1978 - December 1986. Figures 4-7 display the standard error of 12-month price change for All Items and major groups as a function of time for the same period.

It is interesting to note that index standard errors for the All Items index and most major groups exhibit a gradual climb over the time period of this study. This is also reflected in the plots of index standard error versus value of the index shown for All Items, Food, and Apparel in Figure 3. Some major groups, most notably Housing and Entertainment, exhibited apparent seasonal behavior in their standard errors also.

The steady increase in index variance observed in all major groups is consistent with the findings of Valliant and Miller (1989) who considered a Laspeyres-type index computed by chaining month-to-month estimates of price change. The relative variance of this estimator was shown to grow over time.

Abrupt jumps in index variance occurred in Medical Care in September 1980, September 1985 and January 1986 and in Other Commodities and Services in May-June 1981 and September 1985. The nature of these jumps was investigated and attributed to real, substantial price change in sample observations in specific strata in one or two index areas in hospital and other medical care services, health insurance, personal services, and tuition and other school fees expenditure classes. The effect of the price change in the hospital and other medical care services EC was magnified in variance calculations by the fact that the price movement for the other hospital and medical care services

stratum is used, in conjunction with another medical care stratum, as a regular source of imputation for four health insurance strata. The effects of these apparent discontinuities in what otherwise is a generally smooth function can be seen in the corresponding locations in graphs of price change standard error for these major groups (see Figures 6 and 7 for 12-month price change.)

While index variance was seen to climb steadily with respect to time and value of the index, behavior of price change variance was for All Items and several major groups generally stable and uniform over the 108-month period, especially when compared to the actual variability of price change estimated for that period. Figures 8-9 display price change and price change standard error versus time for All Items for 1- and 12-month lags, respectively. This stability is attributable to the high correlation of the index between months. Figure 10 displays the 1- and 12-month lagged correlations of the All Items index. Similar results obtained for most major groups.

Price change variances for Apparel and Upkeep are a notable exception to the above in that they appear to exhibit a more seasonal variation and climb with greater speed and variability over the interval of the study. The reasons for this behavior are not entirely well understood, though it is well recognized that price collection for items in this major group poses greater difficulties than in other major groups, especially with respect to availability of items. Comparisons of numbers of usable quotes, the actual number of price change observations used in computing the price relatives for strata, as shown in Figures 11-12, however, indicate the effect of sample size in major group price change standard error. Usable quotes for Apparel and Upkeep and Entertainment are approximately one tenth the number for Food in any month.

As expected, average price change variance increased with the length of the interval, though in most cases, not linearly so. The average standard error for 12-month price change was approximately 2.2 times the standard error for 1-month price change and approximately 1.6 times the standard error for 2-month price change for the All Items index. Similar behavior obtained for most other major groups. Medical and Other Commodities and Services, for reasons discussed above, were notable exceptions. Average price change and average price change standard errors for 1-, 2-, 6-, and 12-month lags and the ratios of average price change standard errors to that for 1-month price change for 2-, 6- and 12-month lags are given in Table 1.

Table 1

Major Group	1-Mo Avg PC	1-Mo Avg SE	2-Mo Avg PC	2-Mo Avg SE	Ratio 2/1	6-Mo Avg PC	6-Mo Avg SE	Ratio 6/1	12-Mo Avg PC	12-Mo Avg SE	Ratio 12/1
All Items	0.516	0.053	1.040	0.070	1.33	3.183	0.089	1.69	6.560	0.115	2.18
Food	0.469	0.078	0.936	0.098	1.26	2.762	0.140	1.80	5.445	0.180	2.32
Housing	0.547	0.114	1.102	0.151	1.32	3.407	0.172	1.50	7.049	0.216	1.89
Apparel	0.278	0.269	0.592	0.368	1.37	1.764	0.558	2.07	3.474	0.727	2.70
Transportation	0.500	0.061	1.014	0.084	1.37	3.158	0.126	2.08	6.780	0.165	2.72
Medical	0.713	0.116	1.433	0.169	1.45	4.384	0.298	2.57	9.042	0.426	3.67
Entertainment	0.449	0.182	0.905	0.249	1.37	2.748	0.382	2.10	5.674	0.517	2.84
Other C&S	0.643	0.102	1.296	0.144	1.42	3.998	0.246	2.42	8.244	0.352	3.46

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. It is only for 6- and 12-month lags that in most cases the standard errors of small price changes do not overwhelm the price change estimates themselves.

5. Conclusions

The estimates of standard error for the CPI and price change presented in this paper are preliminary and should be regarded as underestimates of their final values. However, even in their preliminary form, these estimates raise important questions with respect to the sample design and index estimation for the survey. Extreme changes in price in a few quotes can have a dramatic effect on the variance of a major group and lower level index or price change estimators. This is particularly the case 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 appears to be a fairly stable measure over the period of this study, which included periods of extremely high and extremely low inflation. This is a comforting finding, which 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

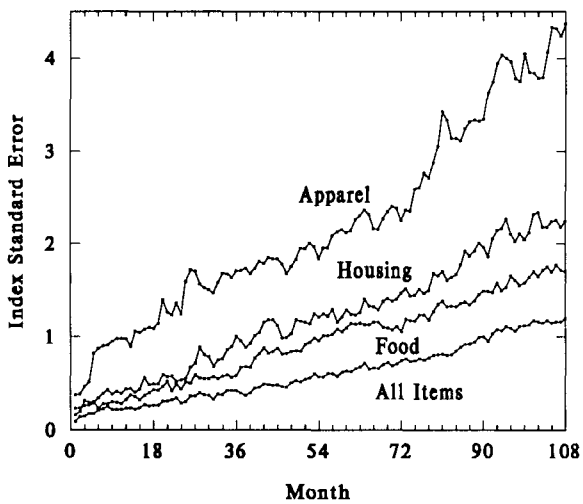
The author would like to acknowledge Michael P. Cohen for his development of the computing formulae for index and price change variances, James E. Johnstone for his work in computer systems analysis and programming of the CPI variance estimates, and Kenneth P. Archer, David L. Swanson, Rita Smith, and Donald Jones for their assistance in production processing of variance estimates. Daniel H. Ginsburg provided valuable information in researching particular instances of price change. Special thanks go to Gary R. Zion for his help in the graphical display of these results. The author also wishes to acknowledge and thank the late Curtis A. Jacobs, Richard Valliant, Paul Armknecht, Janet Williams, and Stuart Scott for their careful reading of this and earlier drafts of this paper and for their helpful comments.

7. References

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

Valliant, R. and Miller, S. M. (1989), "A Class of Multiplicative Estimators of Laspeyres Price Indexes", *Journal of Business and Economic Statistics*, 77, 387-394.

Figures 1-2. Index Standard Error vs Time for All Items, 7 Major Groups



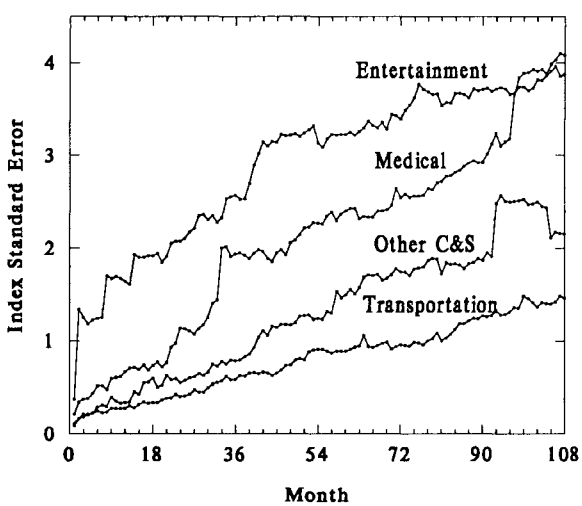


Figure 3. Index Standard Error vs Index for All Items, Food, Apparel

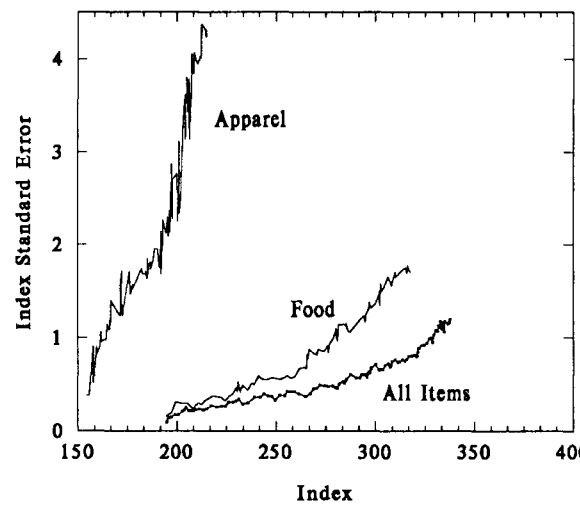


Figure 4. 12-Month Price Change Standard Error vs Time for All Items and Housing

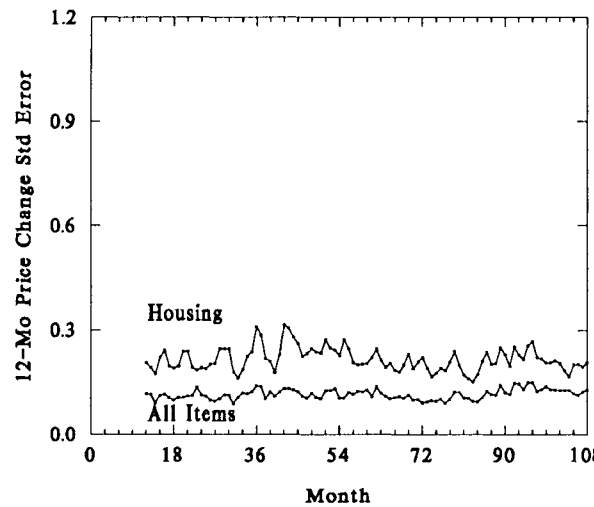


Figure 5. 12-Month Price Change Standard Error vs Time for Food and Apparel

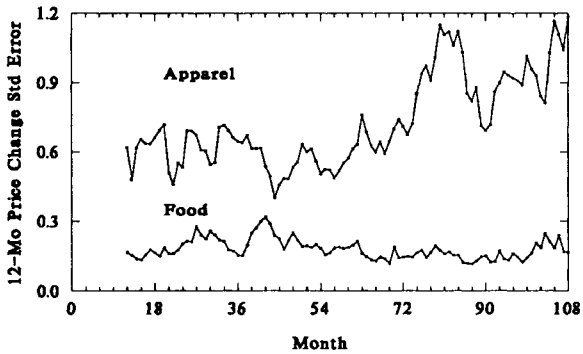


Figure 6. 12-Month Price Change Standard Error vs Time for Transportation, Medical Care

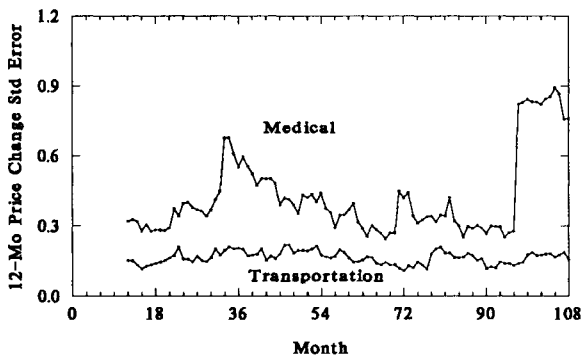


Figure 7. 12-Month Price Change Standard Error vs Time for Entertainment, Other C&S

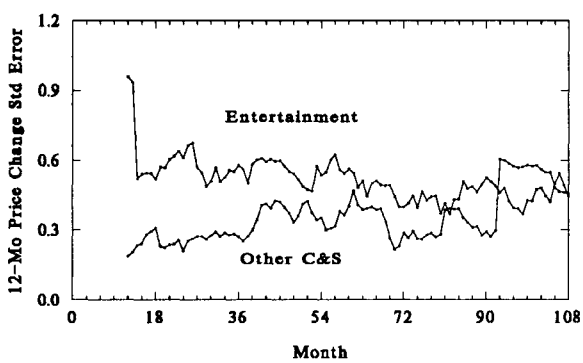


Figure 8. 1-Month Price Change, Price Change Standard Error vs Time for All Items

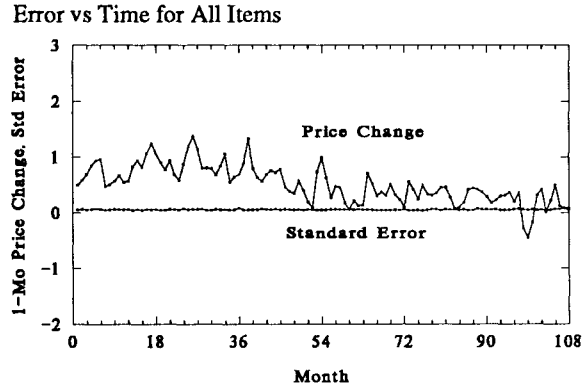


Figure 9. 12-Month Price Change, Price Change Standard Error vs Time for All Items

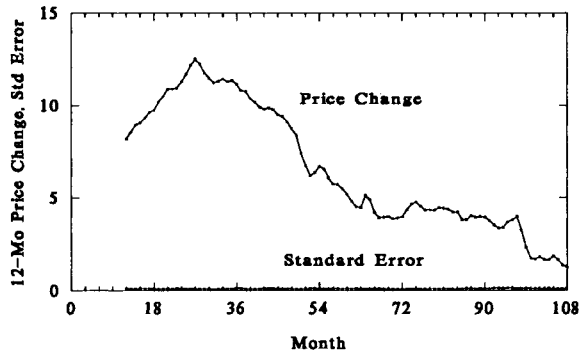
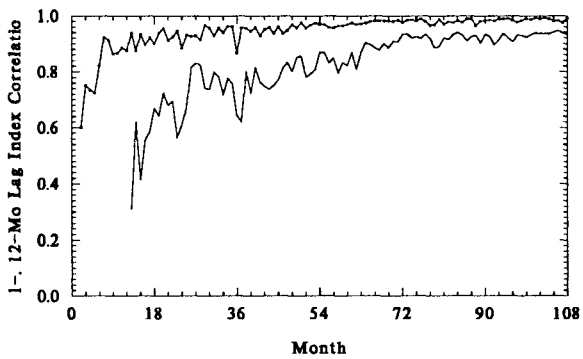


Figure 10. Index Correlation vs Time for All Items, for 1- and 12-Month Lags



Figures 11-12. Usable Quotes vs Time for Selected Major Groups