BLS WORKING PAPERS

U.S. DEPARTMENT OF LABOR
Bureau of Labor Statistics

OFFICE OF PRICES AND LIVING CONDITIONS

Small Sample Bias in Geometric Mean and Seasoned CPI Component Indexes

Robert McClelland, U.S. Bureau of Labor Statistics
Marshall Reinsdorf, Federal Deposit Insurance Corporation

Working Paper 324
August 1999

Small Sample Bias in Geometric Mean and Seasoned CPI Component Indexes

Robert McClelland
Marshall Reinsdorf

Initial version: July 1997

Revised version: November 1998

The views expressed here do not reflect the policy of the Bureau of Labor Statistics or the views of other BLS staff members.

I. Introduction

In 1978, in response to a recommendation of the Stigler Commission, the Bureau of Labor Statistics (BLS) began probability sampling outlets and product versions to arrive at the set of items to be priced for the Consumer Price Index (CPI). This important advance made the CPI samples more representative of consumers' outlet and product version purchasing patterns. It also necessitated a change in the formula used to calculate CPI component indexes.

The concept underlying the CPI component index formula adopted in 1978 is the "modified" Laspeyres index. The modified Laspeyres index uses the market basket quantities from a base period *B* to measure price change beginning in a later "link month" *l*. A lack of information prevents the calculation of such an index at the lowest level of aggregation, however. BLS collects data on expenditures during period *B*, but it does not collect the period *B* price data needed to convert those expenditures into quantities. Consequently, in the index areas where the entire sample has the same link month, the working CPI component index formula adopted in 1978 can be simplified to:

$$L_{l,t} = \sum_{i \in U} w_{iB} (P_{it} / P_{il}) \tag{1}$$

where U is the universe or population of "priceable" items covered by the component index, w_{iB} is a weight based on consumer expenditures during time period B, P_{it} is the current price of unique item i, and P_{il} is its price in the link month l. The elements of U are typically product-outlet combinations, such as a particular size box of regular flavor Cheerios in a particular store. Letting E_{iB} represent the expenditures on unique item i during some base period B, a precise definition for w_{iB} is: $w_{iB} = E_{iB} / \sum_j E_{jB}$.

Equation (1) tends to have a higher value than the modified Laspeyres index that it seeks to mimic.[1] This problem, which the Advisory Commission to Study the CPI (the "Boskin Commission") termed "formula bias," led BLS to abandon equation (1) in January 1995 for the food at home portion of the CPI, and in June 1996 for all other commodities and services.

The index formula adopted in lieu of equation (1) is known as the "seasoned" index. It delays use of the new sample in the CPI until month $m = l + 2$. From the original link month l to the delayed link month m only the old sample of outlets and products is used in the CPI. The seasoned index formula is then,

$$S_{m,t} = \frac{L_{l,t}}{L_{l,m}} \tag{2}$$

where $L_{l,m}$ and $L_{l,t}$ are calculated as defined by equation (1). Simulations by McClelland (1997) indicate that the seasoned index approximates the modified Laspeyres index concept quite well.

The recent report by the Boskin Commission favors yet another formula for calculating CPI basic component indexes, the geometric mean index.[2] Like the seasoned index, this index avoids "formula bias." This index also allows for a certain amount of substitution behavior. Although such substitution behavior is often plausible, many examples exist where it is not. The geometric mean index may be written as:

$$G_{m,t} = \prod_{i \in U} (P_{it}/P_{im})^{w_{iB}} \tag{3}$$

II. Sample Estimation of CPI Component Indexes

Since collecting a census of the prices in U is generally impractical, BLS has employed sample estimators of the indexes described by equations (1), (2) and (3). Let a sample M of n

---

[1] See Reinsdorf, 1994.

[2] The Bureau of Labor Statistics recently announced that this formula would be used for some of the basic component indexes.

items be drawn with replacement (so that the same item can be selected twice.) Also, let the probability of choosing the arbitrary item *i* on any particular draw equal w_{iB}. Then,

$$\hat{L}_{l,t} = (1/n) \sum_{i \in M} (P_{it}/P_{il}) \quad (4)$$

is an unbiased sample estimator of $L_{l,t}$.[3] That is, if one were to draw repeated samples of size *n* with any item *i* have a selection probability for any particular observation of w_{iB}, the average of the sample outcomes would converge to $L_{l,t}$ as the number of sample replicates grows large.

In contrast to equation (1), equations (2) and (3) are non-linear. This makes the expected value of their sample estimators depend on the sample size. We define *small sample bias* as the difference between the expected value of the sample estimator given some small *n* and the expected value of the estimator given an infinite sample. Since all the estimators that we consider are consistent, as the sample grows very large their expected value always converges to the population value of the concept that they seek to estimate.

To illustrate the effect of sample size on an index estimator's expected value, suppose that in *U* the triples $\{P_{il}, P_{im}, P_{it}\}$ have four possible values, {1,1,2}, {1,2,1}, {2,1,2} and {2,2,1}. Assuming that these values are equally likely and that all items have the same expenditures in period *B*, $L_{m,t} = 1.25$. Also, with an infinite sample, $\hat{L}_{l,m}$ and $\hat{L}_{l,t}$ equal 1.125 with probability 1. Consequently, with an infinite sample $E[\hat{S}_{m,t}|n{=}\infty] = 1$, where $E[\cdot]$ denotes an expected value and $\hat{S}_{m,t}$ equals $\hat{L}_{l,t}/\hat{L}_{lm}$. On the other hand, with $n = 2$, there are 16 possible samples with index outcomes: {2, 1, 2, 1.25; 1, 0.5, 0.8, 0.5; 2, 0.8, 2, 1; 1.25, 0.5, 1, 0.5}. Averaging these values

[3] In the interest of readability, equation (4) omits the weights that BLS sometimes applies to sampled items. The presence of these weights would make $\hat{L}_{m,t}$ biased in small samples; see Adelman (1958, pp. 241-2).

implies that $E[\hat{S}_{m,t}|n{=}2] = 1.13125$.[4] Of course, $E[\hat{S}_{m,t}|n{=}1] = 1.25$, because the formula for combining price quotes does not matter when the sample contains a single observation.

The sample estimator of the geometric mean index is defined as,

$$\hat{G}_{m,t} \ = \ \prod_{i\in M} (P_{it}/P_{im})^{1/n} \qquad (5)$$

Under the same assumptions used to illustrate the effect of sample size on $E[\hat{S}_{m,t}|n]$, $E[\hat{G}_{m,t}|n{=}\infty] = 1$ and $E[\hat{G}_{m,t}|n{=}1] = 1.25$. Also, $E[\hat{G}_{m,t}|n{=}2] = 1.125$ since the sample index equals 1 half the time, 2 a quarter of the time and 0.5 a quarter of the time.

The exact expected value of the sample geometric mean index is easy to calculate. On each of the n independent draws needed to select a sample of size n, exactly one item in U is chosen and its price relative is raised to the $1/n$ power. Since any item $i \in U$ has a probability of selection on a particular draw of w_{iB}, the expected value of the contribution of any particular draw to $\hat{G}_{m,t}$ is $\sum_{i\in U} w_{iB}(P_{it}/P_{im})^{1/n}$. The expected product of all n draws is simply the product of their expected values because they are independent. Hence, the expected value of the product in equation (5) is:

$$E[\hat{G}_{m,t}|n] \ = \ \left[\sum_{i\in U} w_{iB}(P_{it}/P_{im})^{1/n}\right]^{n}. \qquad (6)$$

If $w_{iB} = w_{im}$ for all i, the right side of equation (6) equals the exact cost of living index for a CES utility function with an elasticity of substitution of $1 - 1/n$. A substitution elasticity of 1 is needed for $w_{iB} = w_{im}$, so equation (5) is not an unbiased estimator of a cost of living index. It tends to overestimate a cost of living index when the substitution elasticity is at least $1 - 1/n$ and prices have transitory disturbances around a common trend. Conversely, if this elasticity is less than $1 - 1/n$ and prices have different trends, it will tend to underestimate a cost of living index.

---

[4] With $n = 2$ and no replacement, there are six possible samples with index outcomes: {1, 2, 1.25, 0.8, 0.5, 1}. Hence $E[\hat{S}_{m,t}|n{=}2] = 1.09167$.

III. Performance of the Sample Estimators of Seasoned and Geometric Mean Indexes

BLS calculates the CPI from a cross-section of about 80,000 quotes covering 206 different item strata in 44 areas. Since each item stratum in each area furnishes a separate component index, the average sample size for a CPI component index is around nine. Use of such small samples could plausibly lead to a non-negligible small sample bias in either the seasoned index now used to correct formula bias or the geometric mean index alternative.

To obtain empirical evidence on whether this bias is large enough to be of practical concern, we simulated CPI sampling procedures for 159 indexes accounting for about 65 percent of the CPI market basket. These indexes comprise nearly all of the "Commodities and Services" portion of the CPI except for women's apparel, apparel services, and used cars. (These indexes were excluded because extreme seasonal changes in sample composition and other data problems made them very difficult to simulate satisfactorily.) The "population" from which the simulated samples were drawn was a data set of prices and expenditures weights used to calculate the official CPI from January 1993 to November 1994. This data set, which McClelland (1997) also uses to investigate formula bias and small sample bias, pools prices and expenditure weight data from all CPI index areas. In the simulations, we treat this geographically pooled data set as if it were the population of prices and expenditure weights in a single area.

We assume that the original link month l is January 1993 and that the delayed link month m is March 1993 for all quotes. The simulated indexes therefore compare March 1993 prices with prices in the 18 month period starting April 1993 and ending September 1994. We set the sample size for each item stratum equal to the number of quotes in the national CPI sample for that stratum divided by 44 and rounded down to the nearest integer. See the appendix for additional information on the data.

The simulations departed from actual BLS procedures in one respect. In the simulations we set the sampling probabilities equal to the desired item weights so that the simulated samples would allow unbiased estimation of the 1978 CPI formula. (See footnote 2.) In actual practice, BLS often weights the items in the sample, primarily because the expenditures used to set the outlet sampling probabilities cover broader aggregates than the required outlet weights.

We calculated the expected value of the sample geometric index from equation (6). No analogous simple expression exists for the expected value of the sample seasoned index. For sample size $n \leq 8$, we therefore considered an exhaustive list of all possible combinations of n quotes drawn from U. We averaged the seasoned indexes implied by these combinations, with each index weighted by the probability of selection of the combination used to calculate it. Because of the computational burden for n larger than 8, these cases were simulated with 20,000 draws of samples of size n, with each item's probability of selection for any particular draw of an observation equal to w_{iB}. We then averaged the 20,000 simulated index outcomes to get the expected value of the item stratum index.

IV. Results

Both the geometric mean index and the seasoned index have positive small sample bias in our simulations. Figure 1 depicts the small sample bias of the aggregate geometric mean index. This aggregate index weights each of the 159 simulated item strata in proportion to its weight in the CPI-U as of December 1994. The months on the horizontal axis start two months after month m, the starting point for price change measurement. Only even months appear because many item strata have their prices collected every other month. In months when BLS collects no data for these strata, their prices are assumed to retain their value from the previous month. Rather than graph the indexes implied by this assumption, we chose to graph only months when all prices were actually observed.

The top line in Figure 1 shows the bias that would result from using a sample size of one for all item strata. The next lines show that doubling the sample sizes to two and again to four almost cuts small sample bias in half. Missing values are one reason why doubling the sample size fails to reduce the bias by exactly half. When a sample of size two or more is selected, one of the sampled items may have a missing price. This has the effect of reducing the sample size by one.[5] Thus, a small proportion of the indexes averaged to obtain the lines labeled “two quotes” and “four quotes” are based on samples of fewer than two or four quotes.

The lowest line in Figure 1 shows the bias that results from using average CPI sample size for each item stratum.[6] It is considerably below the “four quotes” line because CPI samples average eight or nine quotes, with many more quotes for item strata that are especially noisy or that have a high weight. On the other hand, the small sample bias is still large enough to be a concern. After a year it is over 0.1 percent and after 18 months it is nearly 0.2 percent. This figures should be multiplied by 0.7 to obtain a projected effect on the all-items CPI, because after the 1998 revision the Housing portion of the CPI will be unaffected by small sample bias. Compared with the estimated effect of about 0.25 percent per year from the adoption of the geometric mean formula in all 206 strata,[7] an additional reduction of perhaps 0.08 percent per year from simply increasing the sample sizes is quite substantial.

The small sample bias of the seasoned index is displayed in Figure 2. Its small sample bias is significantly less than that of the geometric mean index. What is more, its bias tends to fluctuate seasonally rather than grow indefinitely over time.

Figure 3 brings these contrasts into sharper focus. It compares the performance of the geometric mean and seasoned indexes based on the average sample sizes in use in the CPI. After

---

[5] We simply disregarded samples where every price was missing. This amounts to imputing their index by averaging all the indexes from samples with at least one non-missing value.

[6] Brent Moulton has pointed out that the harmonic mean might be preferable to the arithematic mean. However, in practice there is little difference between the two.

[7] See Moulton and Smedley, 1996.

18 months the seasoned index has less than a quarter of the small sample bias of the geometric mean index.

One reason why the sample geometric mean index has a larger bias is that the difference between the sample size one index (which is the same for both formulas) and the sample size infinity index is larger for the geometric mean index formula. For example, after 12 months the sample size 1 index equals 102.8, the seasoned population index equals 102.3 and the geometric mean population index equals 101.8. After 18 months their respective values are 104.3, 103.8, and 103.0.

Figure 4 graphs the small sample bias using the average CPI sample sizes as a percentage of the difference between the sample size 1 and sample size infinity indexes. The average CPI sample sizes leave around 12 percent of the maximum possible small sample bias uncorrected with a geometric mean formula and about 8 percent uncorrected with a seasoned index formula. Thus about half the difference between the geometric mean and seasoned index small sample bias results from the fact that the geometric mean index has farther to fall and about half reflects the lower proportion of the maximum possible distance that it does fall.

Figure 4 also reveals an anomaly in the behavior of the seasoned index. The percentage of the maximum possible small bias that is uncorrected should rise over time due to sample attrition. Such a pattern is evident for the geometric mean index, but not for the seasoned index.

V. Explaining the Behavior of the Seasoned Index Small Sample Bias

The contrast between the geometric mean index small sample bias and seasoned index small sample bias is quite striking. The former continues to grow even after a year has passed, while the latter stops growing after six months.

The lack of growth of the seasoned index small sample bias can be explained by considering a second order Taylor series expansion. Let $\mathrm{Corr}(\cdot)$ denote a correlation, and $\mathrm{cv}(\cdot)$ denote a coefficient of variation (i.e. a standard deviation divided by the mean). Then a second order Taylor series expansion for $E(\hat{L}_{l,t}/\hat{L}_{l,m})$ is:

$$E(\hat{L}_{,t}/\hat{L}_{l,m}) \approx E(\hat{L}_{l,t})/E(\hat{L}_{l,m})$$

$$+ [E(\hat{L}_{l,t})/E(\hat{L}_{l,m})][[\text{cv}(\hat{L}_{l,m})]^2 - \text{Corr}(\hat{L}_{l,t},\hat{L}_{l,m})\text{cv}(\hat{L}_{l,t})\text{cv}(\hat{L}_{l,m})]. \qquad (7)$$

Since the population seasoned index equals $E(\hat{L}_{l,t})/E(\hat{L}_{l,m})$, the seasoned index small sample bias approximately equals:

$$SSB_{seasoned} \approx S_{m,t}[[\text{cv}(\hat{L}_{l,m})]^2 - \text{Corr}(\hat{L}_{l,t},\hat{L}_{l,m})\text{cv}(\hat{L}_{l,t})\text{cv}(\hat{L}_{l,m})]. \qquad (8)$$

Figure 5 depicts the actual simulated small sample bias of the seasoned index and the predicted small sample bias using equation (8). The approximation formula is able to predict the actual small sample bias quite well except when the index contains extremely divergent price changes. This is why a handful of very noisy strata (many of which involve fresh agricultural products) cause a slight underprediction of the average bias in a few of the months. One indication of the influence of such outliers is that in July 1994 the weighted average of the 159 simulated small sample biases equals their weighted third quartile. The (weighted) median small sample bias is always very close to its predicted value and also close to zero.

Equation (8) implies that the major determinants of small sample bias in $\hat{S}_{m,t}$ are $\text{cv}(\hat{L}_{l,m})$, which has a positive effect, and $\text{Corr}(\hat{L}_{l,t},\hat{L}_{l,m})$, which has a negative effect. Also, since the correlation is usually positive, $\text{cv}(\hat{L}_{l,t})$ has a negative effect on the seasoned index's small sample bias. If prices that rise unusually rapidly between times l and m continue to rise quickly between times m and t, $\text{cv}(\hat{L}_{l,t})$ will tend to be large and $\hat{S}_{m,t}$ will tend to be small.

The coefficient of variation of the sample outcomes for the $\hat{L}_{l,m}$ is large when the sample size is small and the price relatives (P_{im}/P_{il}) differ greatly from each other. Paradoxically, its value can change with t because items with missing prices at time t must be dropped from any sample in which they appear before calculating the $\mathrm{cv}(\hat{L}_{l,m})$ that predicts the small sample bias of $\hat{S}_{m,t}$. Figure 6 shows, however, that $\mathrm{cv}(\hat{L}_{l,m})$ remains approximately constant over time.

This constant coefficient of variation does not imply a constant small sample bias because the correlation coefficient in equation (8) drops substantially in the first few months. In later months the downward trend in this correlation coefficient becomes quite gradual, and $\mathrm{cv}(\hat{L}_{l,t})$ rises; see Figures 7 and 8. The upward trend of $SSB_{seasoned}$ therefore vanishes, and seasonal fluctuations become important.

VI. Explaining the Behavior of Small Sample Bias for the Geometric Mean Index

Since $\log \hat{G}_{m,t} = (1/n) \sum_{i \in M} \log(P_{it}/P_{im})$, the central limit theorem suggests that the distribution of $\log \hat{G}_{m,t}$ in repeated independent samples will be approximately normal. Letting $\mathrm{Var}(\cdot)$ represent a variance, the textbook formula for the expected value of a lognormal distribution implies that:

$$E[\hat{G}_{m,t}] \approx \exp\{E[\log \hat{G}_{m,t}]\} \exp\{0.5 \,\mathrm{Var}[\log \hat{G}_{m,t}]\}$$

$$= G_{m,t} \exp\{0.5 \,\mathrm{Var}[\log \hat{G}_{m,t}]\}. \quad (9)$$

An approximate expression for the geometric mean index small sample bias is, then,

$$SSB_{geo} \approx G_{m,t} \left[\exp\{0.5 \,\mathrm{Var}(\log \hat{G}_{m,t})\} - 1\right]. \quad (10)$$

Empirical tests of this approximation confirm that it works well. Its ability to predict the small sample bias of the average of the sample geometric mean indexes for the 159 item strata is evident from the near perfect agreement of the top two lines in Figure 9.

The “predicted small sample bias” line in Figure 9 shows that the variance of the sample outcomes for $\log \hat{G}_{m,t}$ rises as t moves farther from m. One reason why samples with different compositions disagree by larger amounts in later months is that the discrepancies between the price relatives P_{it}/P_{im} tend to be larger when t is higher. The rising variance of $\log \hat{G}_{m,t}$ across samples could also be caused by missing values becoming more frequent, which would reduce the sample sizes.

The bottom two lines of Figure 9 show that the former of these effects is far more important than the latter one. The third line from the top predicts how SSB_{geo} would behave if sample sizes never changed, and the bottom line predicts how SSB_{geo} it would behave if the discrepancies between the P_{it}/P_{im} did not grow. By September 1994 (which is 18 months after month m) sample attrition seems to raise SSB_{geo} by just over 0.01 percent.

The bottom two lines in Figure 9 are calculated as follows. Letting M_t represent an average of the log **prices** in the sample at time t,

$$\log \hat{G}_{m,t} = \{(1/n) \sum_{i \in M} \log P_{it}\} - \{(1/n) \sum_{i \in M} \log P_{im}\}$$

$$= M_t - M_m. \quad (11)$$

Using the textbook result for the variance of a difference:

$$\mathrm{Var}[\log \hat{G}_{m,t}] = \mathrm{Var}\, M_t + \mathrm{Var}\, M_m - 2\, \mathrm{Corr}(M_t, M_m) \sqrt{\mathrm{Var}\, M_t} \sqrt{\mathrm{Var}\, M_m} \quad (12)$$

A tendency of the highest month m prices to rise the fastest would elevate Var M_t. Yet no such pattern is apparent in the data. Instead, the similarity of Var M_t and Var M_m in Figure 10 implies that sample attrition causes almost all the change in both variances because changes in Var M_m can come only from changes in sample composition. Figure 10 also plots a least squares prediction of Var M_m that uses the average percentage of observations that are missing as the only explanatory variable. Even though a weighted average across item strata of missing value frequencies is a rather crude measure of their aggregate importance, in most months changes in this variable can be used to predict changes in Var M_m.

In contrast to the variances in equation (12), there is no reason to expect sample attrition to have much effect on the correlation coefficient. This variable depends on how much the P_{it}/P_{im}

differ from each other. For example, if every price in U rises by the same proportion between times m and t, the correlation in equation (12) will equal 1.

Holding Corr(M_t, M_m) constant at the value it has when t = May 1993 while calculating the two variances in equation (12) based on the samples in use in later months prevents almost all of the rise in the predicted small sample bias except in the final months. On the other hand, holding the variances constant while letting Corr(M_t, M_m) change reduces the predicted growth of small sample bias by very little except in July and September of 1994. These results are consistent with the findings in Section V that cv($\hat{L}_{l,m}$) is approximately constant while cv($\hat{L}_{l,t}$) grows.

The tendency of the price relatives P_{it}/P_{im} within a stratum to differ by increasing amounts as time goes on is at odds with the assumption of a common price trend in the model of Reinsdorf and Moulton (1997, p. 407). Many CPI item strata contain both similar and dissimilar items. Prices of items that are poor substitutes and that are produced with different technologies may have little tendency to move together.

The importance of shifts in relative prices in explaining the behavior of the geometric mean index implies that its accuracy as a cost of living index depends on consumers' substitution behavior. Shifting relative prices create opportunities for substitution. In effect, the geometric mean index assumes that consumers take advantage of those opportunities, while the seasoned index assumes that they do not.

A substitution elasticity of 1 makes the geometric mean index an exact cost of living index. A typical CPI item stratum probably contains some items that are more substitutable than a unitary elasticity would imply and other items that are less substitutable. Whether the unitary elasticity assumptions usually yields accurate cost of living indexes because it is, on average, correct, is unclear. The greatest changes in relative prices are likely to occur between the least substitutable items. If the only changes in relative prices involve items with low substitutability, a geometric mean index could underestimate the change in the cost of living.

VII. Correction for Area-Specific Price Effects

The estimates of small sample bias above are potentially too high because the variance of the population of price changes used to simulate the CPI sample estimators may be too large. Specifically, CPI price data from many areas were pooled to form the population of price changes used to simulate an index for an item stratum in an area. If areas differ in their price behavior, the variance of the price changes in this kind of simulated population would tend to exceed the true within-area variances.

The approximate expression for the geometric mean index small sample bias can be used to investigate this effect. Let the available CPI data for an item stratum come from A areas, let a denote a particular area, n_a denote the number of quotes in the data set U_a from area a, let N_A denote the total number of areas, and let n denote the total number of quotes. Letting the variance estimate for each area be weighted by its degrees of freedom $n_a - 1$, a weighted average of the within-area price change variances is:

$$V^* = [1/(n - N_A)]\{\sum_{a \in A} \sum_{i \in U_a} \{ \log(P_{it}/P_{im}) - (1/n_a)[\sum_{i \in U_a} \log(P_{it}/P_{im})]\}^2 \}. \qquad (13)$$

Now let k denote a "shrinkage factor" equal to the ratio of the mean within-area variance to the variance of the log price relatives around their overall mean, or $V^*/\text{Var}[\log(P_{it}/P_{im})]$. Then the mean within-area variance implies a small sample bias for the geometric mean index of approximately:

$$SSB^*_{geo} \approx G_{m,t}\ [\exp\{0.5\ (k)\text{Var}(\log \hat{G}_{m,t})\} \qquad (14)$$

Figure 11 reports the small bias estimates implied by equation (z), with k set equal to 1 for any item stratum where V^* exceeds $\text{Var}[\log(P_{it}/P_{im})]$. Preventing k from exceeding 1 probably leads to an overadjustment for area specific variance components. Hence area-specific price effects probably distort the estimates of small sample bias even less than figure 11 implies.

VIII. Conclusion

BLS calculates the CPI by averaging approximately 9000 component indexes representing 206 individual item strata in 44 local areas. The use of so many cells results in small sample sizes in many individual cells. This has the effect of raising the expected values of the seasoned and geometric mean indexes because their formulas are non-linear.

The geometric mean index seems to be particularly susceptible to small sample bias. The effect on the CPI of changing to the geometric mean formula would be significantly greater if that change were accompanied by an increase in sample sizes. Furthermore, even if the seasoned index formula were retained for all the item strata were it is presently used, an increase in sample sizes would probably reduce the growth rate of the CPI by several hundredths of a percent per year. In some cases, such an increase in sample size may be attainable by pooling samples from more than one area and calculating a combined index. This would be a particularly attractive solution if locality has little effect on prices in the small number of strata that cause most of the aggregate small sample bias. It may also be useful to increase the sample size of just the indexes most heavily affected. Figure 5 shows that the mean bias is about equal to the third quartile, implying that a relatively small number of indexes may be responsible for most of the bias. In this case the total amount of bias could be reduced without increasing the total number of quotes simply by reallocating quotes from less sensitive to more sensitive indexes.

References

Adelman, Irma. 1958. “A New Approach to the Construction of Index Numbers.” *Review of Economics and Statistics*, vol. XL, 240-9.

Armknecht, Paul A., Moulton, Brent R. and Stewart, Kenneth J. 1995. "Improvements to the Food at Home, Shelter and Prescription Drug Indexes in the U.S. Consumer Price Index." BLS Working Paper 263.

Boskin, Michael J., Dulberger, Ellen R., Gordon, Robert J., Griliches, Zvi and Jorgenson, Dale. 1996. *Final Report of the Advisory Commission to Study the Consumer Price Index.* US Government Printing Office for the Committee on Finance, U.S. Senate, 104th Congress, Second Session.

Reinsdorf, Marshall. 1993. "The Effect of Outlet Price Differentials on the U.S. Consumer Price Index." In *Price Measurements and Their Uses*., Murray F. Foss, Marilyn E. Manser and Allan H. Young, eds., National Bureau of Economic Research Studies in Income and Wealth Volume 57, University of Chicago Press (Chicago): 227-260.

Reinsdorf, Marshall. 1994. “Price Dispersion, Seller Substitution and the U.S. CPI.” BLS Working Paper 252.

Reinsdorf, Marshall B. and Moulton, Brent R. 1997. “The Construction of Basic Components of Cost of Living Indexes.” In *The Economics of New Goods*., Timothy F. Breshahan and Robert J. Gordon, eds. National Bureau of Economic Research Studies in Income and Wealth Volume 58, University of Chicago Press (Chicago): 397-436.

APPENDIX
DATA SET CONSTRUCTION

The prices used in this analysis are the prices of items in the commodities and services section that are used in calculating the CPI. Item classes in which the average number of quotes was less than two were deleted --calculated by dividing the total number of quotes by 44, the total number of geographic areas for which indexes are created-- as were strata with an unusually large number of missing prices or with anomalous price behavior. The total weight of these three classes in the CPI is less than 0.5 percent. The number of remaining item strata is 159.

Any quote that was missing in January 1993 or for which the number of quotes in the area was unknown was deleted. If the price for a given quote was missing for every month from January to November of 1994, then the quote was deleted. Because approximately 20 percent of the sample rotates every year this left about 60 percent of the original sample. In addition, any quote in which a noncomparable substitution was made was deleted. If an index was missing because all of the values were missing, the index was imputed using the average of all other indexes.

For the 82 item groups where prices are collected monthly, the index is calculated from March 1993 to later months ending in September 1994. For the 77 groups where prices were collected bimonthly, even-month indexes run from February through August and odd-month indexes run from January through September. The even-numbered months were treated as odd-month quotes (i.e., quotes from February were combined with the January quotes.) There were also several large metropolitan areas that collect the bimonthly items on a monthly basis. Odd-month quotes of this type were used and even-month quotes were ignored. To avoid any potential problem from chaining, indexes between March 1993 and later months were formed by comparing the prices in March 1993 with that month. As with BLS policy, if at least one quote existed in both periods of the index, then missing values were imputed using the index formed by the existing quotes.

The indexes were aggregated using the relative importance weights for December 1994. These weights are the product of the index in a given period and the item's expenditure share estimated from the 1982-4 Consumer Expenditure Survey. While this method is slightly inaccurate, the inaccuracy is only a function of how the relative inflation rates across goods has changed from March 1993 to December 1994. The outlet expenditures used for the expenditure weights were taken to be the average expenditures in the area from which a price quote was taken. All the adjustment factors in the weights, such as the percent of POPS correction, were ignored. For the Monte Carlo simulations, quotes were drawn from the population 20,000 times, with the probability of an outlet being selected was proportional to its expenditures, as defined above.

Official Bureau procedure is to select randomly one entry level item (ELI) to represent the item strata of goods and then randomly select outlets from that ELI proportional to their expenditures. Here, we do not separate the quotes by ELI, but simply sample outlet/quote combinations from the entire item stratum.

Figure 1
Small Sample Bias of the Geometric Mean Index

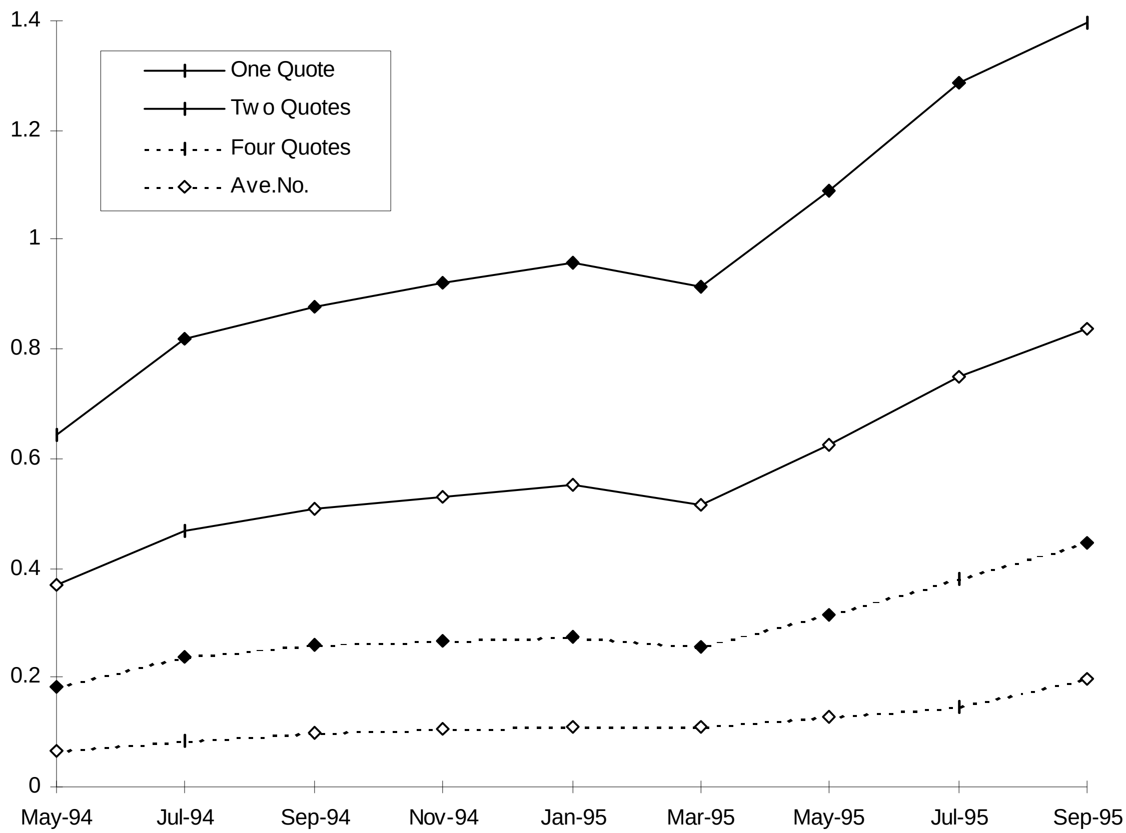


Figure 2

Small Sample Bias for Seasoned Method

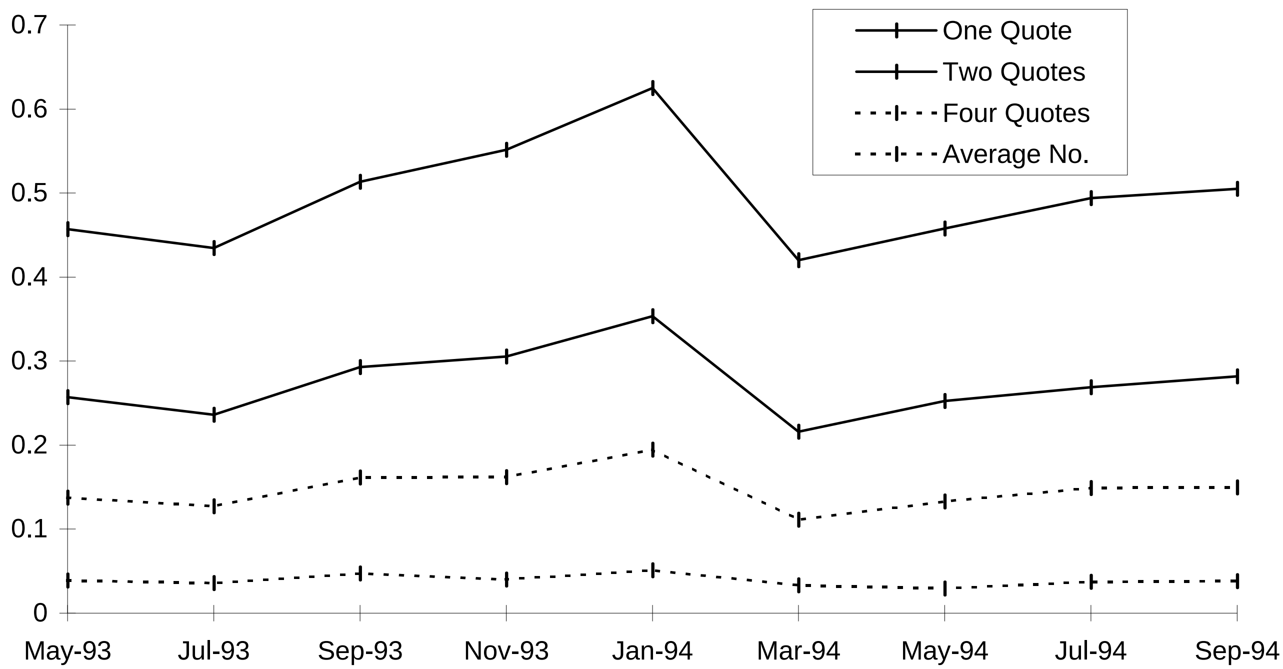


Figure 3: Small Sample Bias using Average CPI Sample Sizes

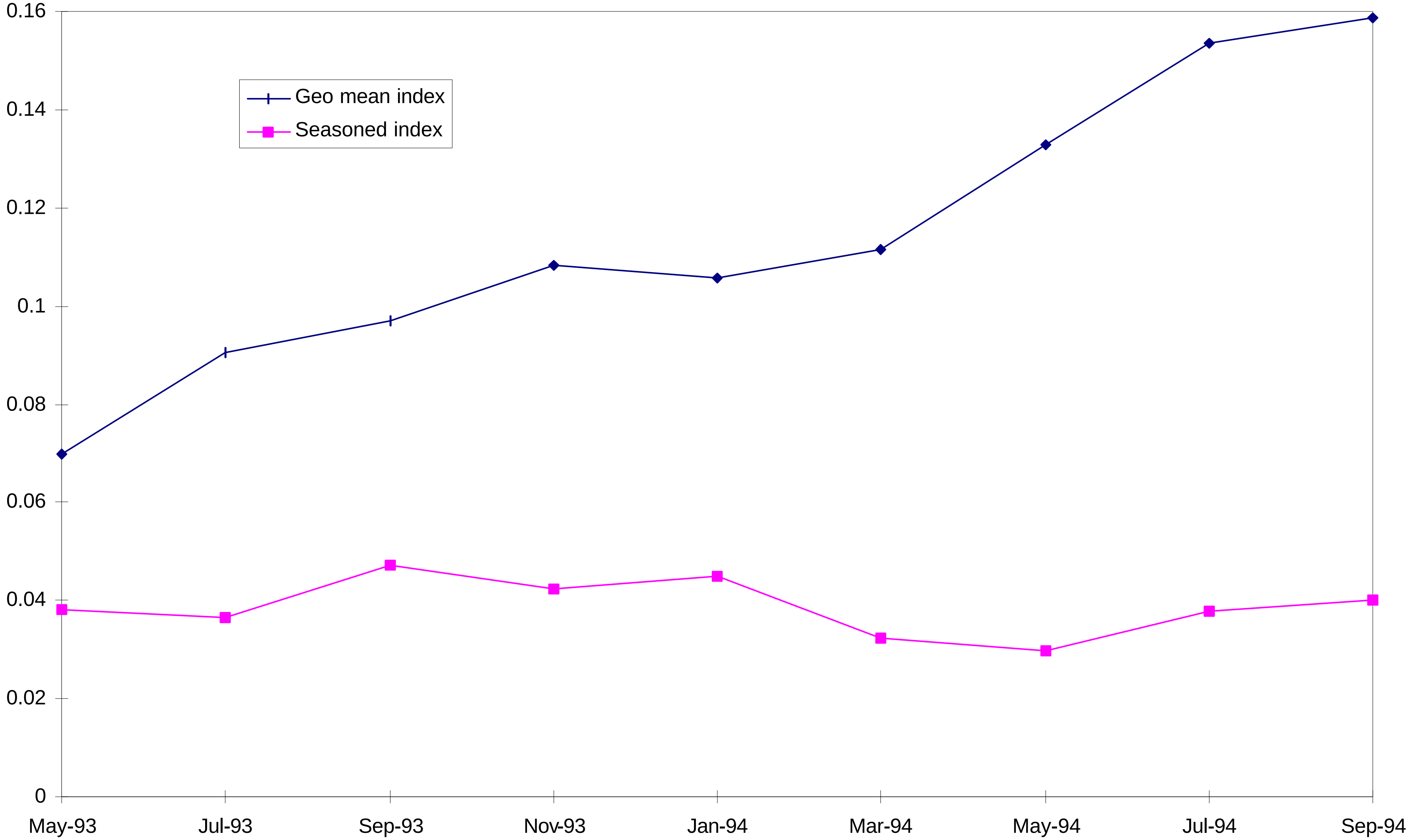


Figure 4: Percent of Maximum Possible Small Sample Bias Remaining in the CPI

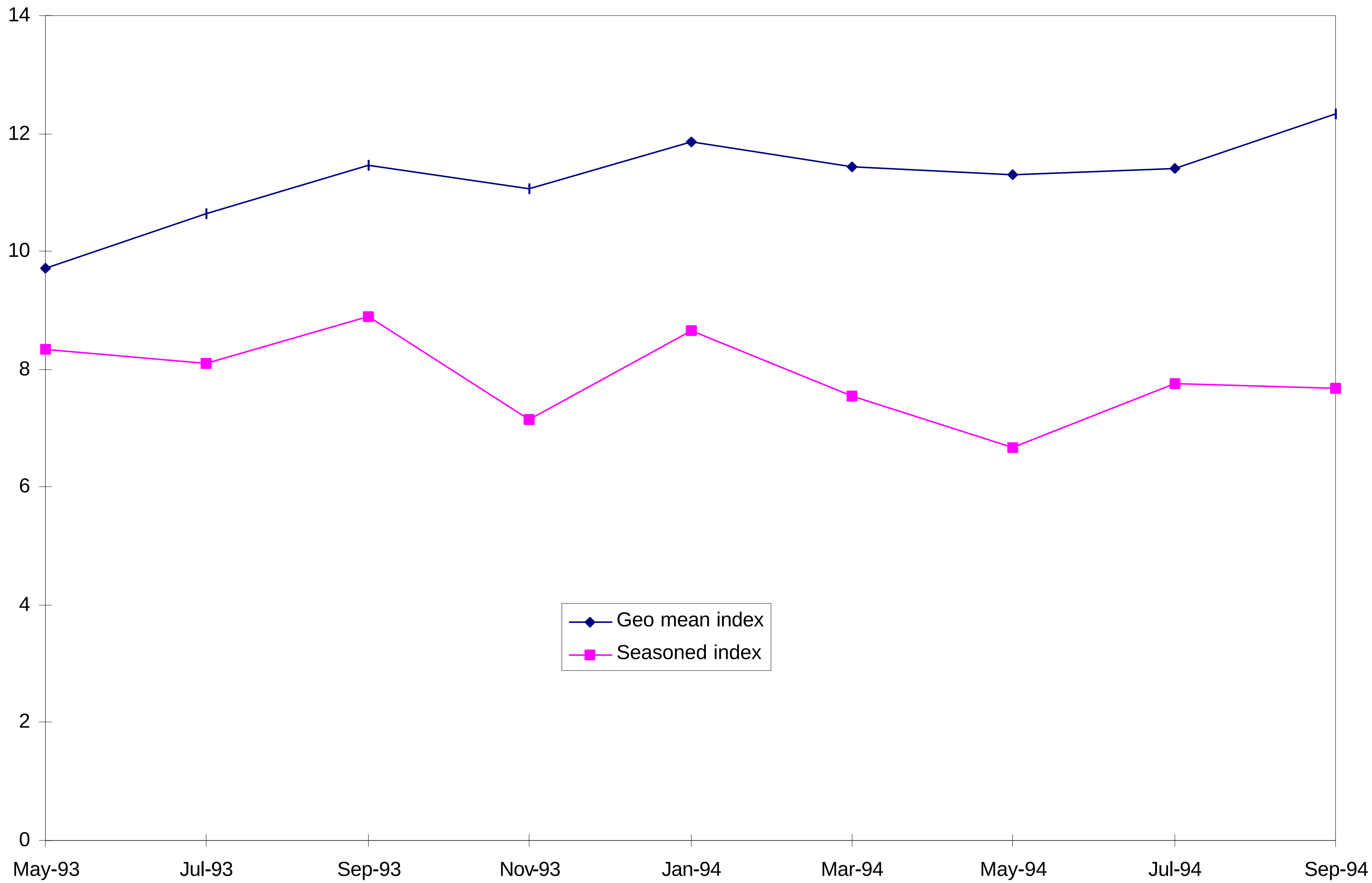


Figure 5: Actual and Predicted Small Sample Bias, Seasoned Index

(Note: All Measures use CPI Item Stratum Weights)

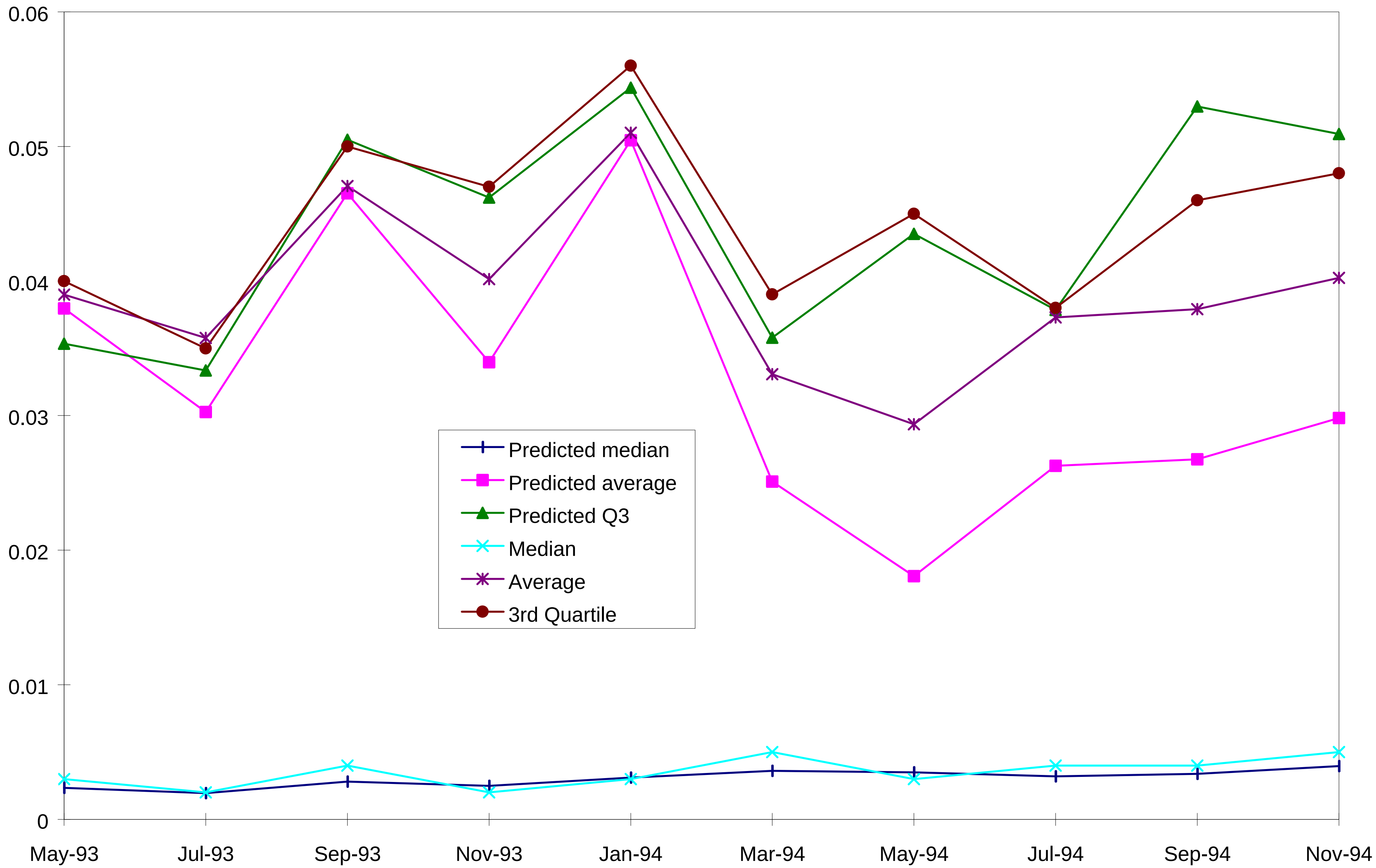


Figure 6: CV of Sample Averages of Link Month Relatives

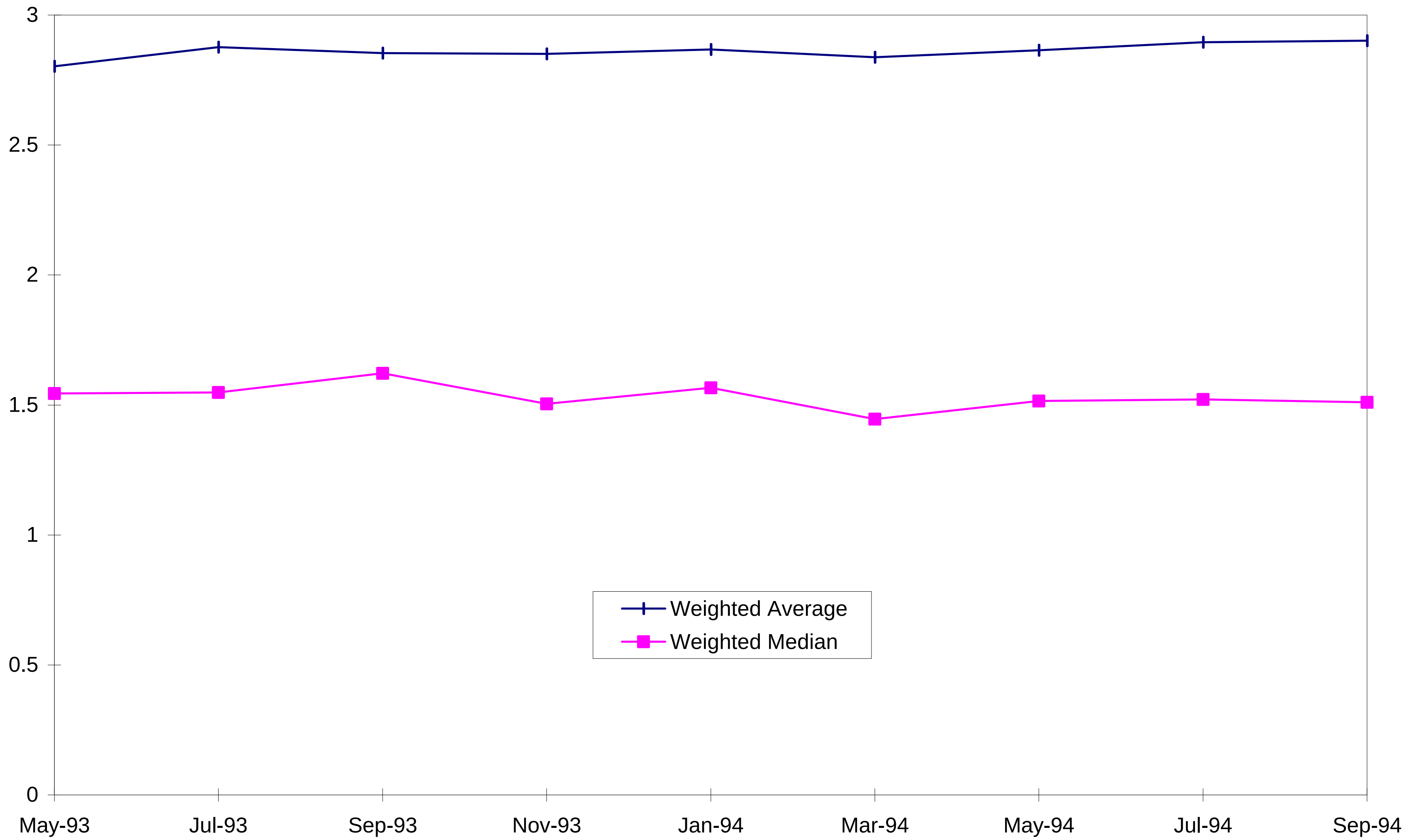


Figure 7: Correlation of Sample Averages of Link and Comparison Month Relatives

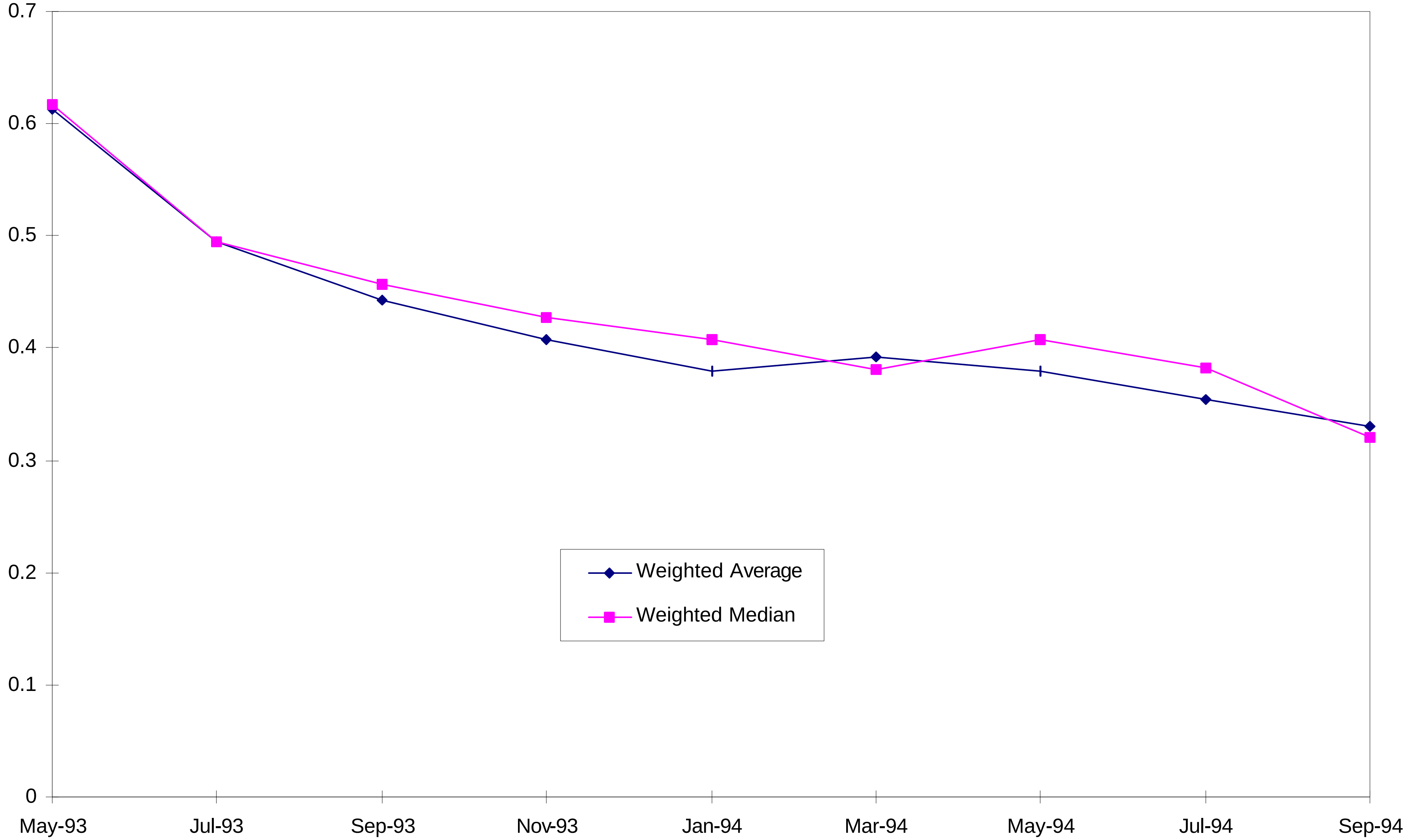


Figure 8: CV of Sample Averages of Comparison Month Relatives

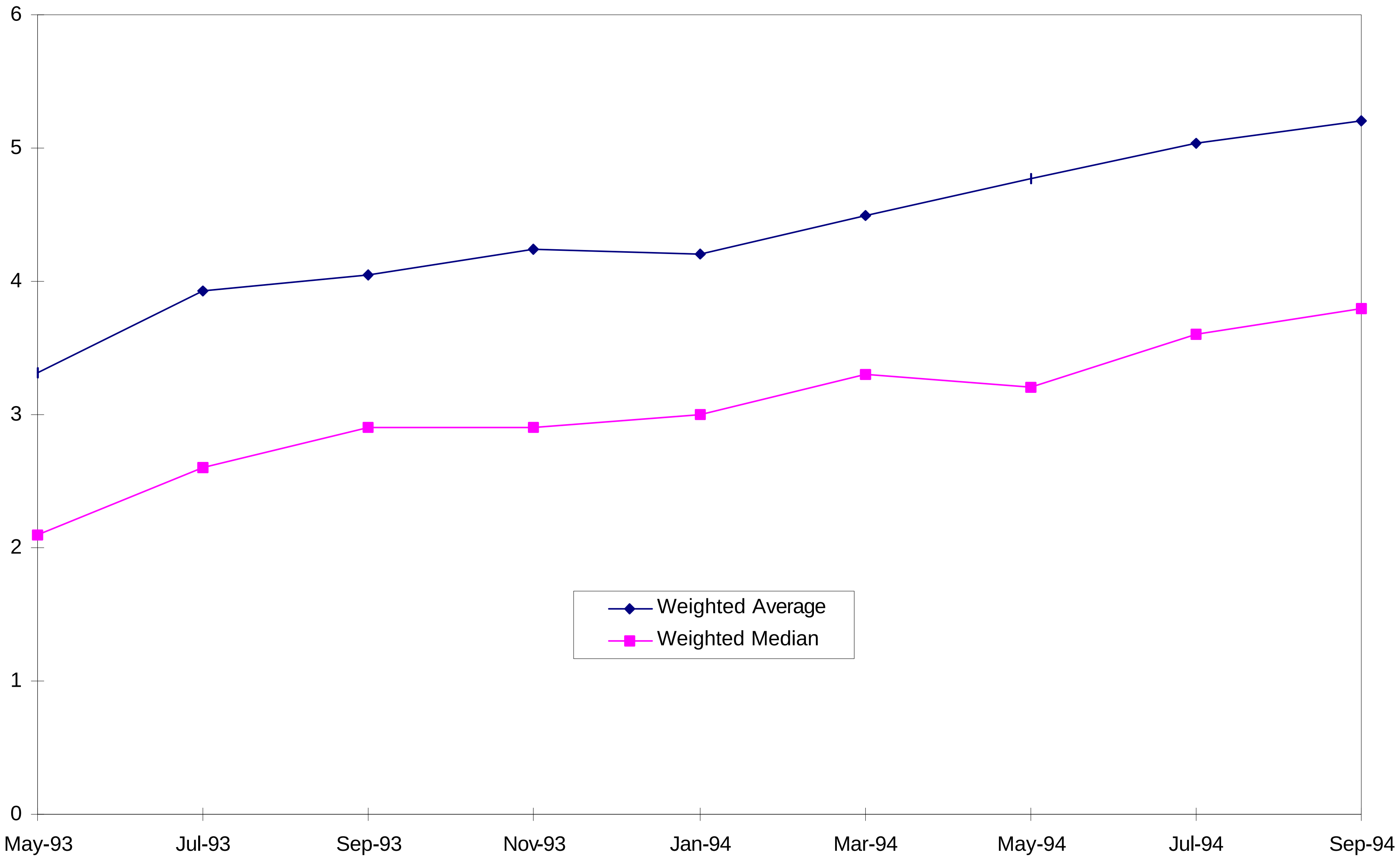


Figure 9: Actual and Predicted Small Sample Bias, Geometric Mean Index

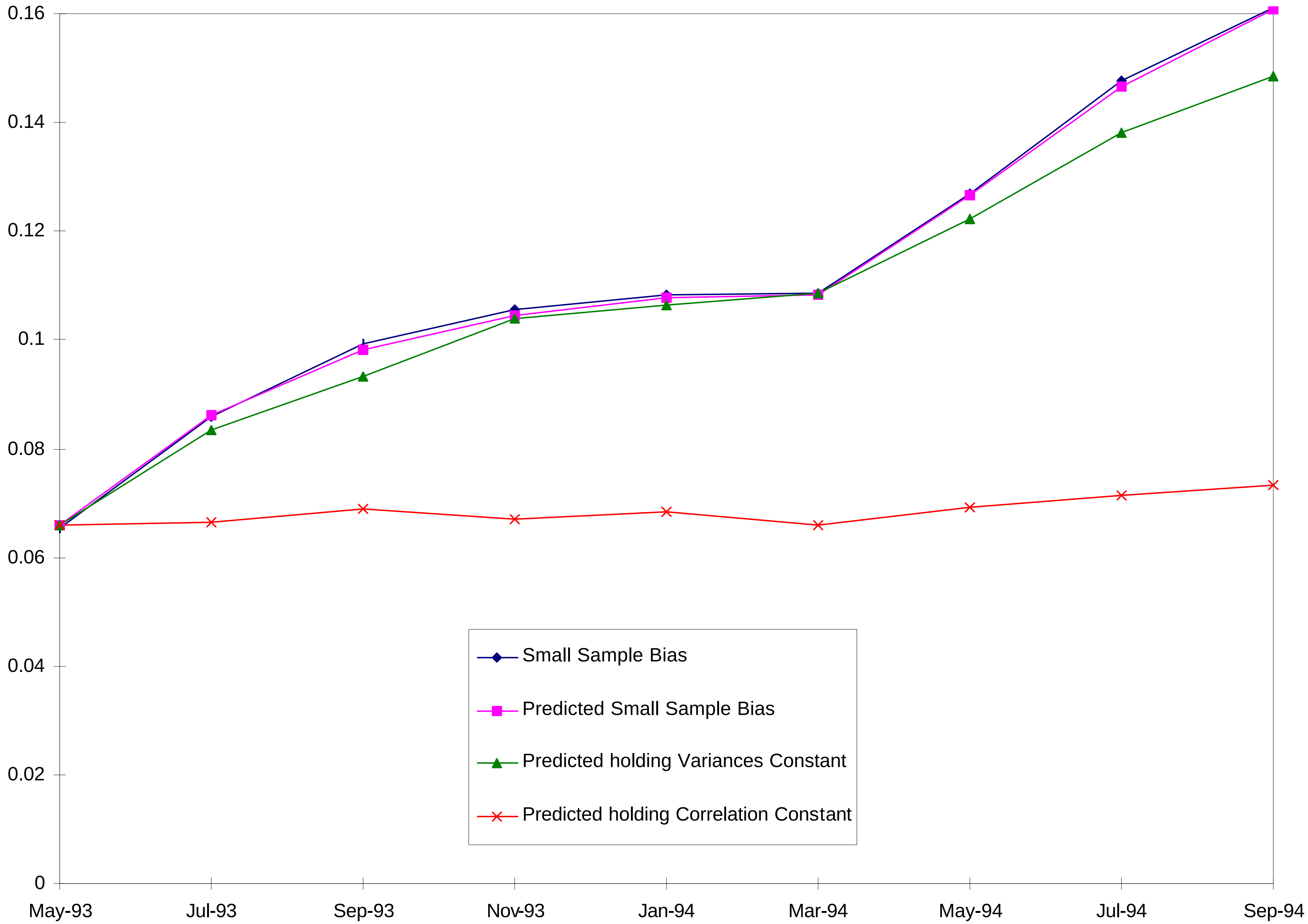
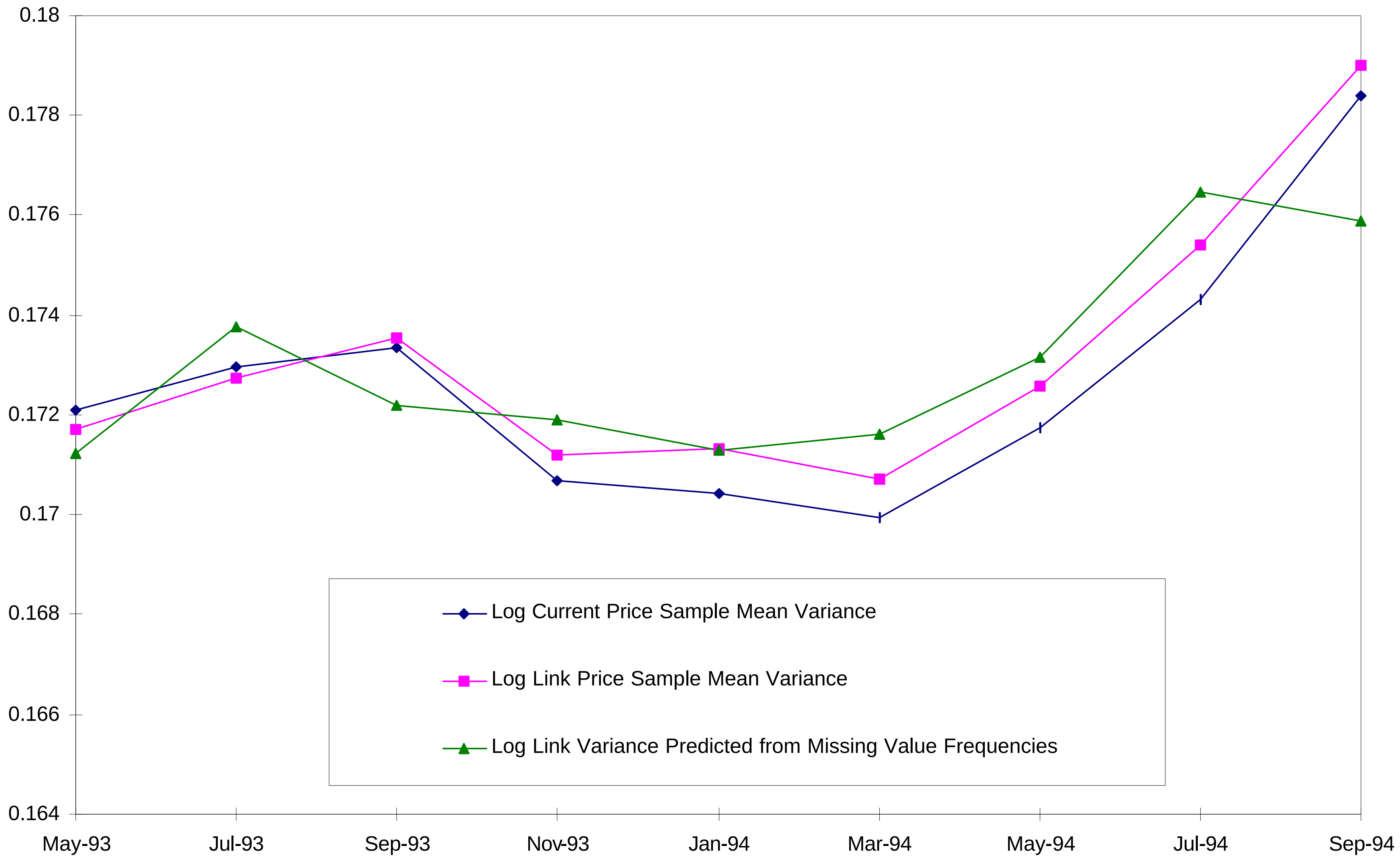


Figure 10: Variances of Sample Means of Log Prices



1	2	3	4	5	6	7	8	9