

Revising Replicate Selection in the CPI Variance System

Owen J. Shoemaker and Fred Marsh III

U.S. Bureau of Labor Statistics, 2 Mass Ave., NE, Room 3655, Washington, DC 20212

shoemaker_o@bls.gov marsh.fred@bls.gov

Abstract

The Consumer Price Index (CPI) program at the Bureau of Labor Statistics (BLS) is actively considering streamlining its CPI Estimation System to produce a more efficient and more flexible overall operation. This New Estimation System will entail the elimination of the replicate structure, which currently provides the necessary replicate values for the CPI's Stratified Random Group (SRG) Variance System. In order for BLS to continue using its SRG methodology, it will become necessary to create these needed replicates "dynamically" (by random assignment) each month from full sample values. In this paper, we will investigate and compare as well as produce the results from the use of dynamically constructed replicate price changes and compare these variance results with the currently computed CPI variances. At least two random methodologies for selecting the replicate values will be analyzed and evaluated. A more robust variance system is the hoped for objective.

Key Words: Stratified Random Groups, Replicates

Any opinions expressed in this paper are those of the author and do not constitute policy of the Bureau of Labor Statistics.

Background

The BLS is in process of implementing a new estimation system for the CPI, and one of its component features is the elimination of the replicate structure upon which the CPI's official SRG Variance System is built. The current method of assigning and producing replicates is sample-based, with the replicates a part of the original sampling process. All too often this sample-based structure becomes too inflexible and overly restrictive in the sample size requirements for each replicate. In the larger cities (A's), the replicates are sub-samples of the full samples in each Index Area. In the medium (X's) and smaller (D's) cities, the replicates are formed from one or more usually two smaller PSU's within the given X or D Index Area. By giving more flexibility and more balance to the replicate structure process, BLS hopes to improve the variance system as well as streamline the overall estimation system.

When the New Estimation System is installed and implemented, this embedded, sample-based replicate structure will be replaced by a dynamically loaded process within the new variance system itself. BLS can either do this new assigning of replicates to quotes (1) randomly, or (2) systematically, or (3) a combination of either (1) or (2) with the requirement that once the replicate is assigned a replicate number it keeps that replicate number assignment until that quote is lost or rotated out of the system. The replicate assigning methodology chosen by BLS is (3), with the original randomization process to

be done systematically. In this paper we will compare these three replicate-assigning methods to our current system, but we will concentrate our final analysis on Method 3. Currently, in both full samples and replicate sub-samples, the quotes (and housing units) retain their replicate assignment number throughout the life of the quote (or housing unit). For the sake of continuity with the current variance system and to guarantee as much as possible that the replicates will be valid estimates of the full-sample values, BLS has adopted Method 3 as the required methodology for the new replicate structure.

The Stratified Random Group (SRG) Methodology

The CPI Variance System currently uses the following Stratified Random Group methodology to calculate variances (and so standard errors) for all of its aggregate Area(A)—Item(I) combinations when a replicate structure is present:

$$(1) \quad \text{Var}(A, I, t, t-k) = \sum_{a \in A} \frac{1}{N_a(N_a-1)} \sum_{r \in R_a} [PC(A, I, r(a), t, t-k) - PC(A, I, f, t, t-k)]^2$$

Here R_a is the set of replicates in Area= a , and N_a denotes the number of replicates in Area= a . PC is a price change between time t and time $t-k$. $PC = (\text{PREL}-1)*100$, with each PREL, or price relative, being a ratio of sums of cost weights.

For Special Relative Calculations (SRC's), below the Basic Cell level of Index_Area—Item_Stratum, BLS uses a jackknife formulation to calculate the variances.

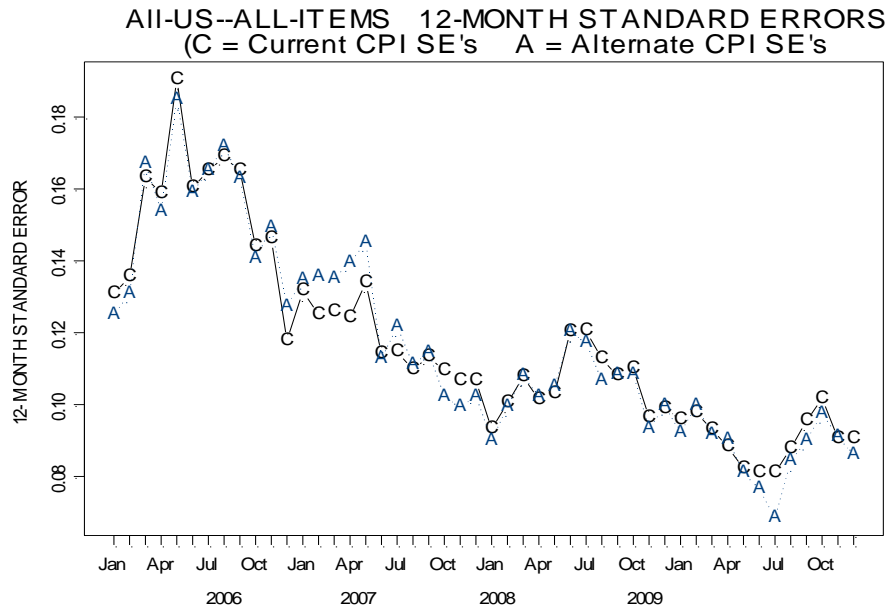
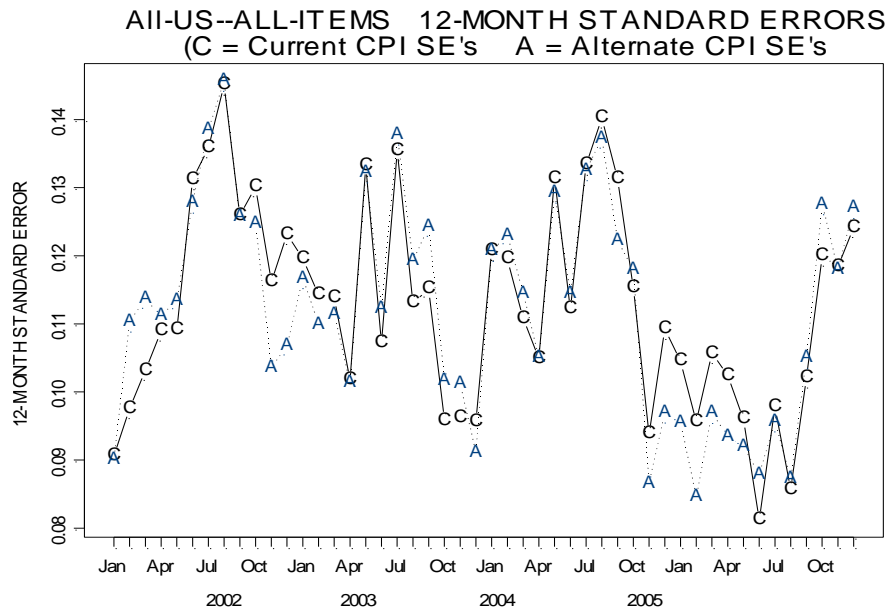
In current usage, whenever an aggregate Area—Item combination includes Items which span more than one of the eight Major_Groups (of Items), we break out the Self-Representing Areas (S_A) into Major_Group-level random groups, while leaving the Non-Self-Representing Areas (N_A) computed as in (1). The formula for these higher-level aggregates, which includes All-US—All-Items, expands out as follows:

$$(2) \quad \text{Var}(A, I, t, t-k) = \sum_{a \in S_A} \frac{1}{N_a(N_a-1)} \sum_{i \in I} \sum_{r \in R_a} [PC(A, i, r(a), t, t-k) - PC(A, I, f, t, t-k)]^2 \\ + \sum_{a \in N_A} \frac{1}{N_a(N_a-1)} \sum_{r \in R_a} [PC(A, I, r(a), t, t-k) - PC(A, I, f, t, t-k)]^2$$

Since the current replicate structure is sample-based, formula (2) has seemed appropriate for our higher-level aggregates whenever Area selection was not the first stage of sampling. However, in the new estimation system, where the replicate structure is to be dynamically determined, the sample-based rationale for splitting out the two sets of Areas is no longer applicable and we can use the simpler formulation in (1) for all of our SRG calculations. The 38 Areas are our random groups, with the quotes within each group being randomly (and systematically and permanently) allotted to its selected

replicate group. (The following graphs show how little difference there has been historically between the two estimates.)

FIG 1 CPI SE's (C=Breakout by MG) vs. CPI SE's (A=No Breakout)



MEAN (Current) = 0.1149 MEAN (Alternate) = 0.1144
C Larger = 54 # C Smaller = 42 CORR=0.967
Mean DIFF (C-A) = 0.0006 Mean Abs DIFF = 0.0045

The graphs in FIG 1 clearly show how close the two formulations have tracked with each other over the eight years represented in these two graphs. During these 96 months, from Jan '02 through Dec '09, the mean difference between the two sets of estimates of 12-month standard errors for the All-US—All-Items indexes was a mere 0.0006, with the current standard errors (where the random groups are broken out into Major Groups) actually averaging higher across these 96 months. Note further the high correlation (0.967) and, just to accentuate the point, a paired-comparison t-test was run between these two sets of standard errors, with a non-significant p -value result of 0.3297. With the sample-based rationale gone and the difference between the two sets of standard errors nearly indistinguishable, there is no reason to utilize the break-out by Major Groups any longer, and the new variance system will reflect that.

Comparison of the Replicate Selection Methods

The replicate scheme that will be deployed in the new variance system will try to look and act very much like current practice. The number of replicates in each Index_Area will be parameterized in the new system and so have the flexibility to be reset either higher or lower depending on extant criteria at the time of implementation. But the total number of replicates will probably not deviate too much from the current system's replicate structure. For our purposes here, in comparing the various new replicate selection methods, we will keep with the same number of replicates in each Index_Area as is currently the case. However, instead of having the replicates assignments being sample-based, we will first take the entire set of usable (eligible) quotes in all $38 \times 211 = 8018$ Index_Area—Item_Stratum cells and assign each quote a random number r (between zero and one). In Method 1, we simply partition the probability space equally according to the total number of replicates (N_R) in the cell. If $N_R = 2$ then if $r < 0.5$ the quote is assigned Rep 1, if $r \geq 0.5$ the quote is assigned to Rep 2; for $N_R = 4$ the partition is by increments of 0.25, etc. For Method 2, we sort by r within each Index_Area—Item_Stratum and assign the replicates to each quote systematically. If $N_R = 2$, we assign in systematic order: 1, 2, 1, 2, 1, etc, until all quotes in that cell are assigned; if $N_R = 4$, the assignments run 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, etc, and so forth for any given N_R . (When new quotes come in in subsequent months the sequence in the basic cell is picked up from where it left off.) Method 1 will be referred to as Randomized, Method 2 as Systematic. For Method 3, which we are calling Combo Plus, and which is the method BLS has chosen to use in the new variance system, Method 2 is first implemented but instead of re-doing the randomization (or systematic) selection process in the subsequent months, we let each new quote hold onto its initial replicate assignment and keep that assignment for the life of the quote. The CPI is essentially a chained index, with each new price relative a function of a ratio of a set of weighted quotes at time t over that same set of weighted quotes at time $t-1$. Method 3 is clearly more in line with current practice and maintains the connection across time of any given quote with itself. We are interested in the differences that can occur if the replicate selection process is allowed to be produced randomly (or systematically) each new month, but we know we want to have the individual quote remain in the same replicate assignment throughout its time in the index. The graphs on the following two pages give a clear picture of the differences that would occur if either Method 1 or 2 were used instead of Method 3.

FIG 2-4 ALL-ITEMS – HOUSING -- FOOD

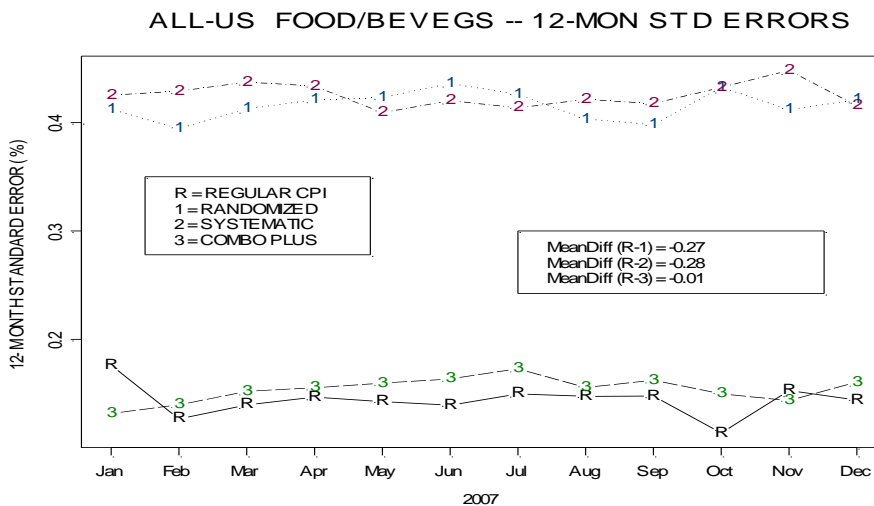
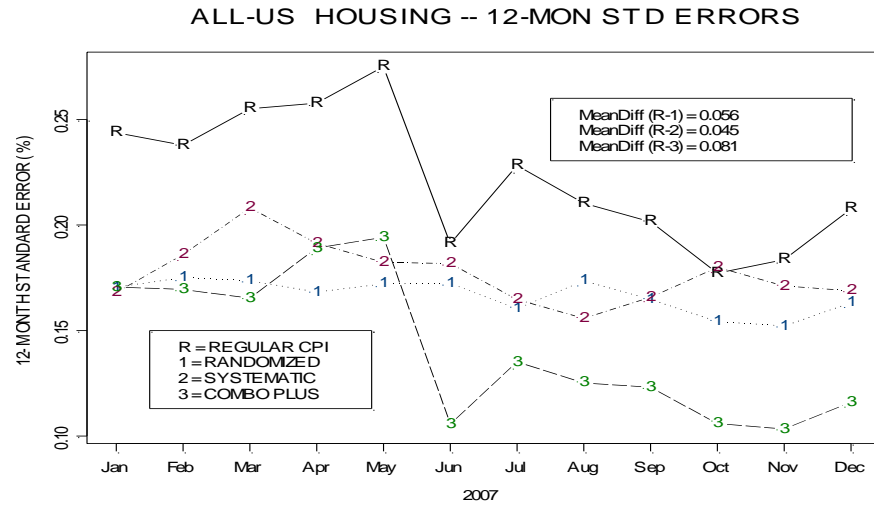
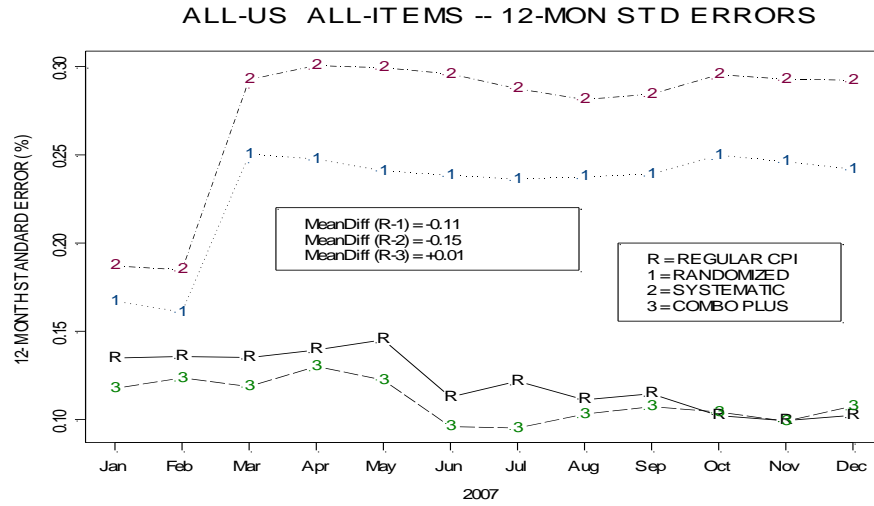
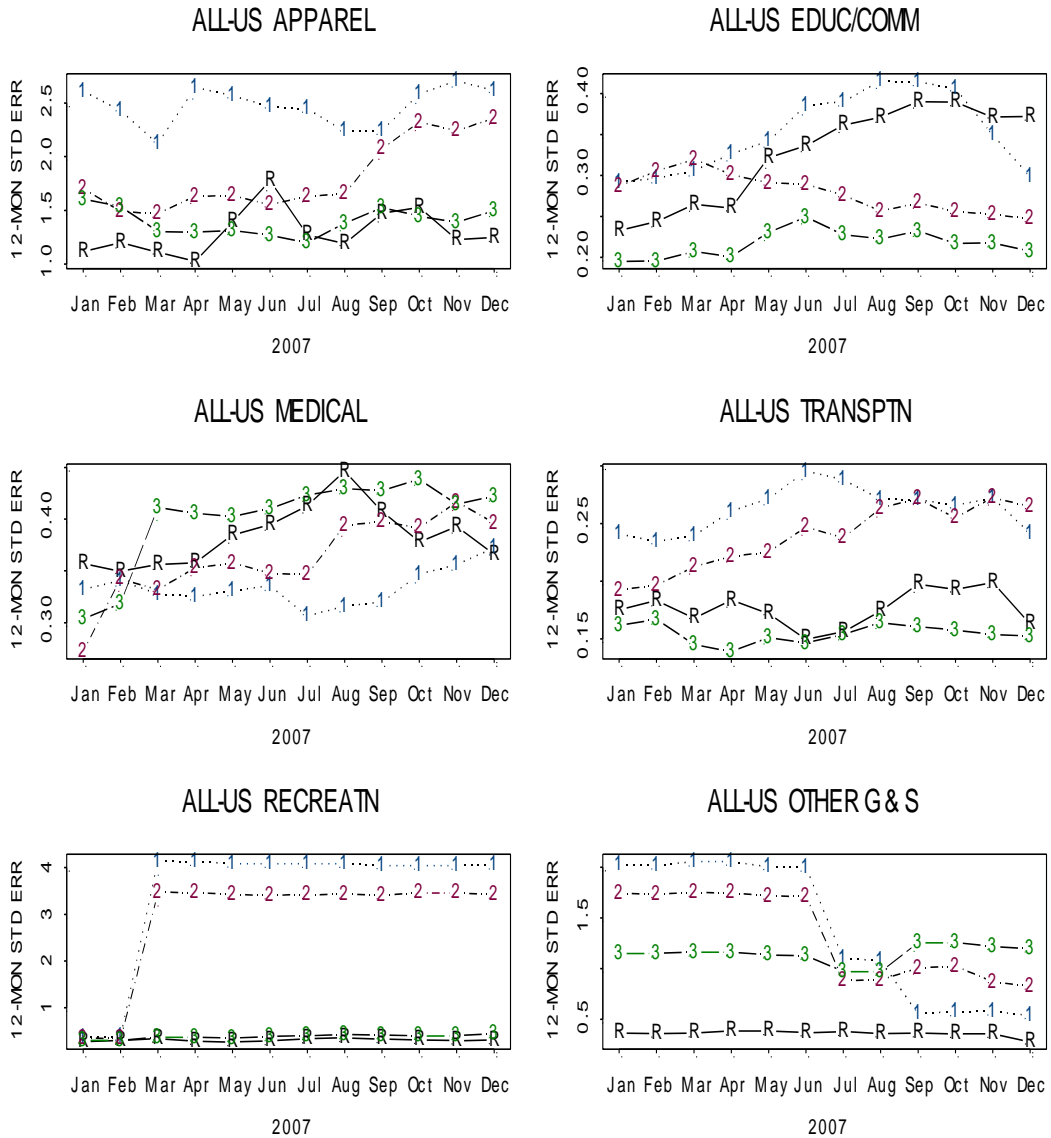


FIG 5-10 12-MON STD ERRORS -- 6 MAJOR GROUPS
(R = Reg CPI 1=Randomized 2=Systematic 3=Combo Plus)



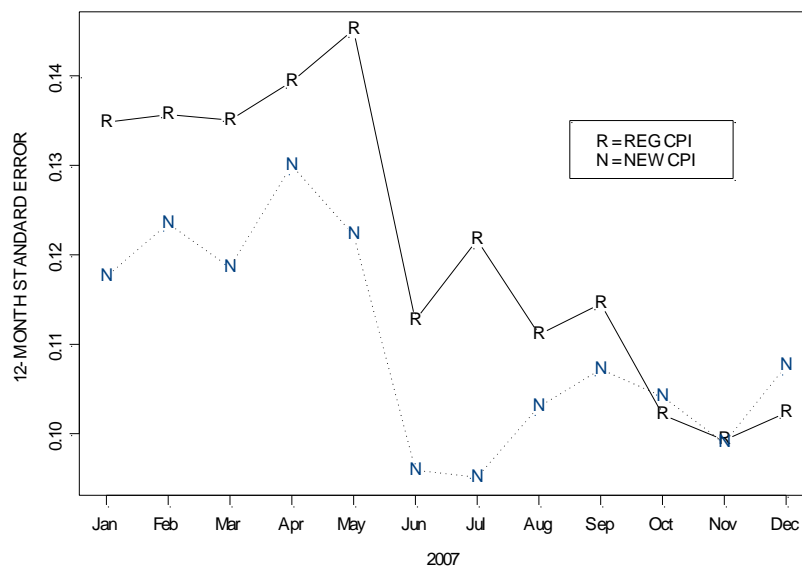
The pitfalls and inadequacies inherent in the use of Methods 1 or 2 are apparent in nearly all the charts, from the ALL-ITEMS one through all eight of the Major Group categories, save maybe MEDICAL and EDUC/COMM. Without the ability for a given set of quotes in any one replicate to hold its replicate assignment over time, at least two aberrant and variance-increasing effects result. First of all, if a quote can be bounced from one replicate to another, it may be unable to compensate naturally for a large dip (or rise) in price by returning to its original price in the next collection period. This can be particularly true when sales prices occur. The second main aberration, seen dramatically in RECREATION and somewhat in OTHER G & S, is the occurrence of an outlier

which can get shifted to another replicate assignment and throw off the chain dramatically thereon out. In fact, it seems clear that the outlier affect in RECREATION continued to ramify its deleterious effects all the way up to the ALL-ITEMS level. Method 3, in all of these instances, was able to smooth these outliers and sales prices to produce a chained set of standard errors that not only nicely mimics the current CPI's standard errors but comes in fractionally less overall (see the R-3 difference in FIG 2).

In HOUSING and MEDICAL we see Methods 1 and 2 producing consistently lower standard errors, but again Method 3 for these two categories tracks more closely (and certainly correlatively very closely in HOUSING) to the Regular CPI standard errors. Housing is a special case, since it contains the two biggest components in the CPI, RENT and Owners' Equivalent Rent (OER). These two item-strata constitute a full 30% of the CPI's expenditures, with HOUSING itself having a relative importance of 40%. Why Methods 1 and 2 produce lower standard errors in MEDICAL is not immediately clear, but it's important to note that it is Method 3 that most closely tracks with Regular CPI in the MEDICAL category. For RENT and OER, BLS runs an entirely separate Price Relative Calculation (PRC), apart from the Commodities & Services (C&S) PRC. C&S constitutes the remaining 70% of the Index. Housing quotes (or housing units) are collected in six six-month panels, and so any month-to-month fluctuation of quotes from one replicate to another (as happens in Methods 1 and 2) doesn't affect the Housing variances much if at all.

Method 3 performs clearly, even dramatically, better than either Method 1 or 2 in nearly all categories of comparison. Method 3 was fore-ordained to be chosen as the new replicate selection scheme, due to its closeness in methodology to current practice, but it is important to show that it cannot be simply replaced by a straight-forward randomization process each and every month. We now want to look exclusively and clearly at the comparisons between Regular CPI and New (Method 3) CPI standard errors.

FIG 11 ALL-US--ALL-ITEMS 12-MON STD ERRORS

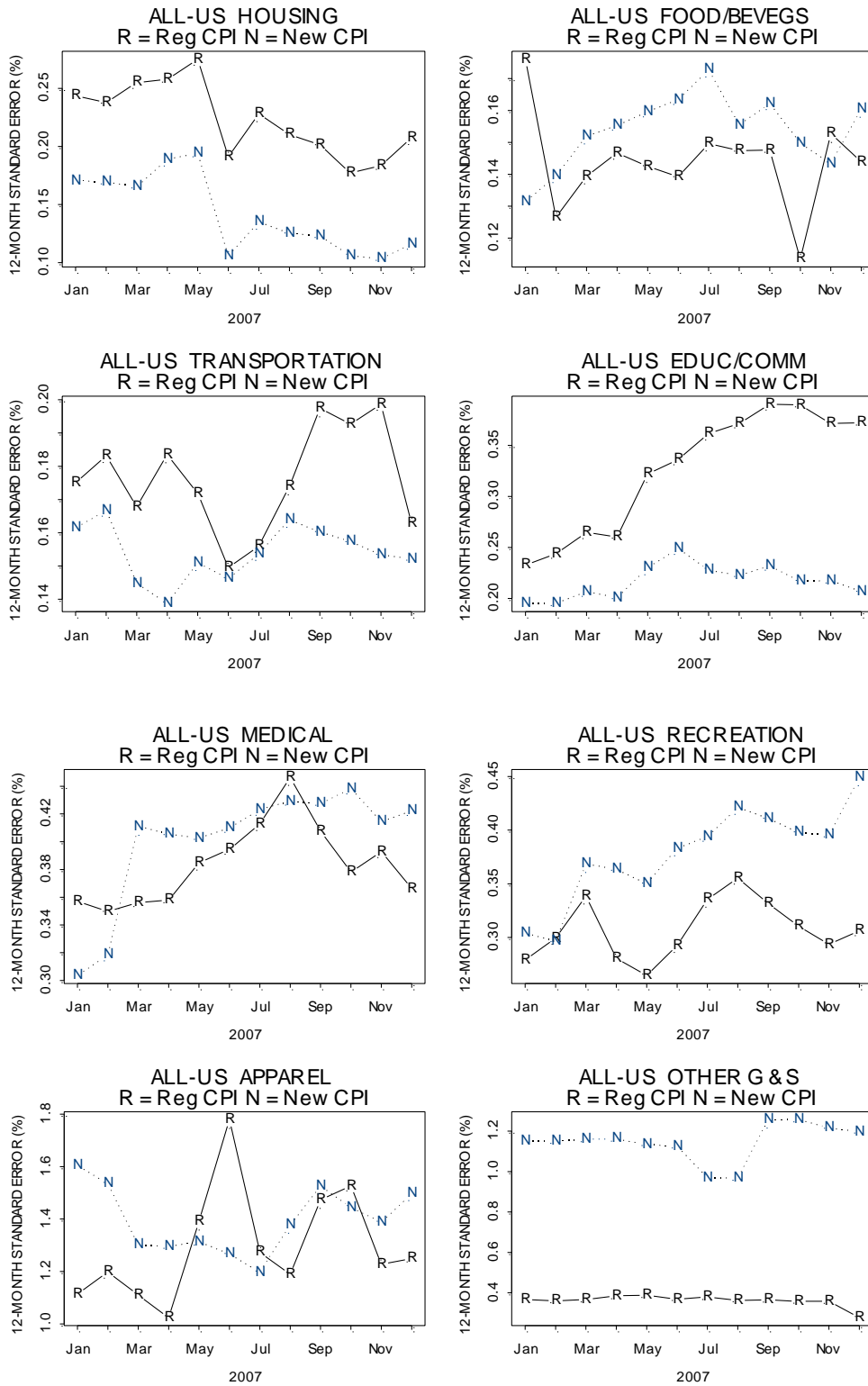


Regular CPI versus New (Method 3) Standard Errors

In FIG 11, we distill the comparison down to the twelve 12-month differences between Regular CPI standard errors for 2007 and New (Method 3) CPI standard errors for 2007. The new variance system, at least in this one yearly set of standard errors, for this one random seed, is producing results significantly lower than their Regular CPI counterparts. A paired-comparison t -test yields a p -value = 0.0028. However, the two sets of results are highly correlated (CORR = 0.802). In the two sets of Major Group comparisons below (FIG 12-13), we can see where the new standard errors are sometimes higher, sometimes lower than the official standard errors, with the balance clearly coming out on the side of lower standard errors under the new system. Several factors may be accounting for this performance difference. First of all, the cost weights (Index Value x Aggregate Weight = Cost Weight) for all the replicate values in each AREA-ITEM cell begin equal to their respective full-sample cost weight. In the current system, individual and therefore different weights are developed for the replicates as well as for the full-samples. This small portion of the variance that would be due to these initial differences is eliminated in the new system. Since the replicates are no longer a part of the sampling scheme, their weights can only be defined as being equal to their full-sample counterparts. And so an added measure of variability is perforce eliminated. Secondly, the systematic nature of the replicate assignment structure better guarantees balanced sample sizes in the corresponding replicates within each AREA-ITEM combination. Currently, due to rotation or initiation issues, or the early loss of outlets in the initiation process, we often discover imbalances in the replicates when it comes to sample size, which more often than not results in higher variances from that fact alone. Thirdly, we will no longer be tying the replicate assignments to the smaller PSU's in the X's and D's, and in the X's (Regional Medium-Sized Cities) in particular we may achieve a significant lowering of variances from those sectors. We will look at the comparisons of our two sets of standard errors in these three city-sized sectors (A, X and D) in the next section. So, while the new variance system seems to be producing similar estimates as the current variance system, these several changes all contribute to that significant lowering of the overall standard error we see in FIG 11.

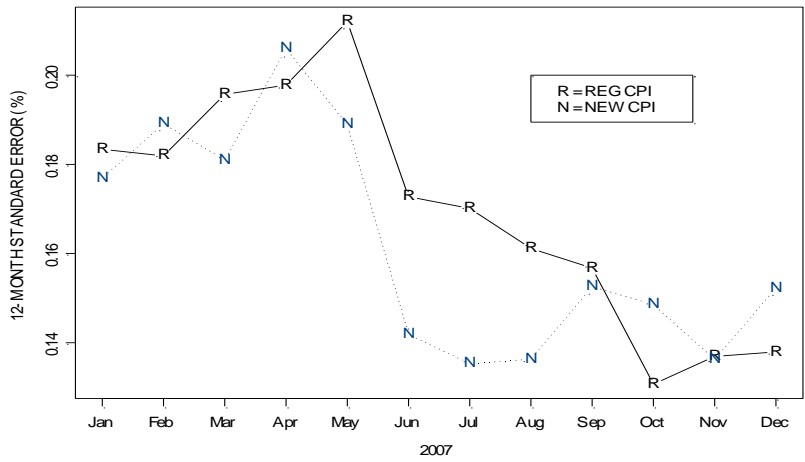
On the following page, FIG 12-13 displays the similarities and differences between the two sets of 12-month standard errors for all eight of the Major Groups in 2007. The Major Groups have been arranged in order of relative importance. As we have mentioned before, HOUSING, which includes Rent and OER, has a relative importance of 40%. TRANSPORTATION is 20% and FOOD & BEVEGS is 16.5%. Then EDUCATION & COMMUNICATIONS is 5.5%, MEDICAL is 5%, RECREATION is also around 5%, APPAREL 4% and OTHER GOODS & SERVICES is 3.5%. These relative importances need to be taken into account when assessing which major group are contributing what to the overall ALL-US—ALL-ITEMS variance. HOUSING, TRANSPORTATION, EDUC/COMM all have clearly lower variances using the new system, with RECREATION and OTHER G & S clearly higher in the new system. FOOD, APPAREL and MEDICAL are a wash. And only in OTHER G & S, where the relative importance is the smallest, do we see a possible problem with worrying whether the two sets of standard errors are producing essentially similar results. Some outlier or outliers has most probably confounded the results in OTHER G & S, but exactly where and how is not immediately clear.

FIG 12-13 12-MONTH STANDARD ERRORS BY MAJOR GROUP

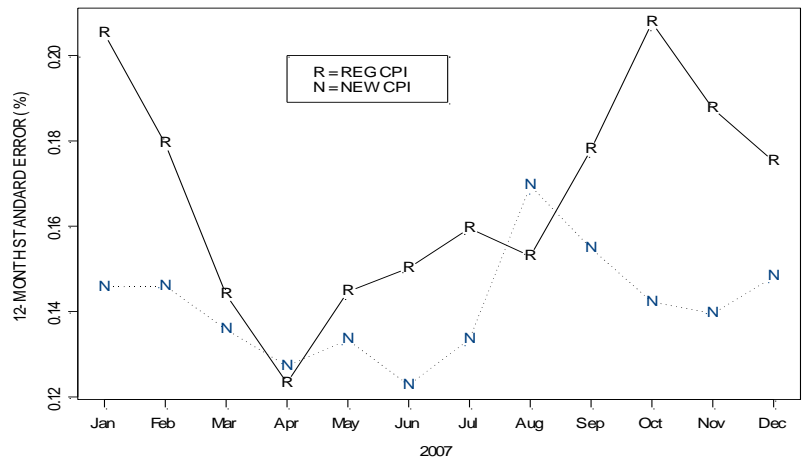


Comparisons by City Class Size (FIG 14)

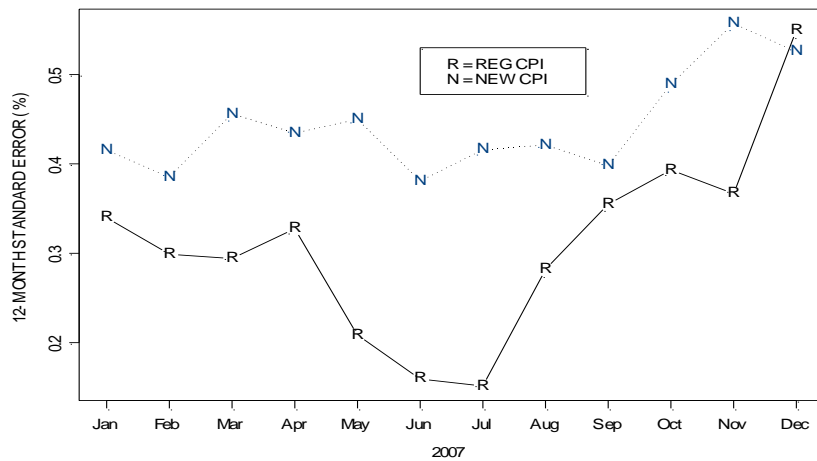
A-SIZED CITIES ALL-ITEMS -- 12-MON STD ERRORS



X-SIZED CITIES ALL-ITEMS -- 12-MON STD ERRORS



D-SIZED CITIES ALL-ITEMS -- 12-MON STD ERRORS



The application of the new variance system, using Method 3, appears to be achieving a significant performance improvement in the X-sized Index_Areas. By de-coupling the replicate selection process from the individual X PSU's and assigning the replicates systematically, and thus more balanced, across replicates, the new standard errors are significantly lower. In the A-Sized Index_Areas, on the other hand, the new standard errors are *not* significantly different from the regular CPI standard errors, albeit somewhat lower on average across these particular twelve months. The D-Sized Index_Areas round out and confound the analysis a bit by coming in significantly higher than regular CPI. However, the D-Sized areas account for only 5.7% of the entire CPI, and their lack of robust sample sizes may be contributing to these differences. In the table below are compiled the comparative City Class Size summary statistics.

FIG 15 City Class Size Summary Statistics

	A000	X000	D000
RELATIVE IMPORTANCE	57.7%	36.6%	5.7%
MEAN SE (REG CPI)	0.170	0.167	0.311
MEAN SE (NEW CPI)	0.162	0.142	0.444
Mean Diff (REG-NEW)	0.008	0.026	-0.133
CORRELATION	0.753	0.368	0.623
P-VALUE (Difference)	0.164	0.004	0.001

Summary of Methodology

The implementation of the new variance system using Method 3 (Combo Plus) will result in the following changes and differences in methodology:

- Using one random seed, the complete set of usable and eligible quotes (or housing units in Rent and OER) will be given a random number (r) between zero and one. This same random seed will be used in all successive months, in order to simplify reproducibility of the variances.
- In each of the 8,018 (38 Index Areas x 211 Item Strata) basic cells, the random numbers assigned to the quotes will be sorted in descending order, with each quote being *systematically* assigned a permanent replicate number assignment. If $N_R = 4$, the sequence of assignments would run 1, 2, 3, 4, 1, 2, 3, etc. When new quotes are added to the basic cell in subsequent months, the new quotes are sorted by their new random number, and then assigned systemically a replicate number, picking up where the assignment sequence in that basic cell had left off in the previous month.
- Each new replicate, when the new variance system is first implemented, will be assigned an initial cost weight that is identical to its full-sample cost weight in its particular basic cell. When a new set of biennial weights is added to the system, this process is repeated. I.e., each replicate cost weight will be set equal to its full-sample counterpart in each basic cell.
- With the weights in place and the replicate assignments made for all usable (eligible) quotes, the new variance process moves to obtaining price relatives (using either a Laspeyres or Geomeans formulation) for each replicate in each of

the basic cells. The subset of usable quotes in each replicate in each basic cell is then fed into either the C&S PRC (Price Relative Calculation) or the Housing PRC (for the Rent and OER quotes only) and replicate price relatives are produced. Where quotes have been imputed in the full-sample those same quotes will be re-imputed in the replicates.

- The new replicate price relatives then update the cost weights ($CW_t = CW_{t-1} * PREL_t$), which in turn are then aggregated up within the Index Areas to all the requested upper-level aggregates, all the way up to All-US—All-Items.
- Finally, the Stratified Random Group (SRG) methodology, without any break-out by Major Groups, is applied to this completed replicate structure and new variances (and standard errors) are calculated for all the lower- and higher-level aggregates, producing a continuing set of 1-, 2-, 6- and 12-month standard errors (albeit with the -2, 6- and 12-month standard errors coming in the first year in staggered fashion).

Summary of Results

Method 3 (Combo Plus) has been shown to be a doable, robust, and appropriate methodology for producing variances for the CPI. It preserves continuity with the current system and brings improved performance. We have worked with only 24 months of data (2006-2007) in order to produce a full set of 12-month standard errors for the twelve months of 2007, but these initial, and admittedly incomplete, results clearly demonstrate that the new variance system, using Method 3, conforms in scale to the current CPI variances and in general is producing lower results. Methods 1 and 2 have been shown to be both deficient and poor performers in almost all settings. Using Method 3, BLS can be confident that the new variance system will provide both historical continuity and improved variance performance.