

TECHNICAL WORKING PAPER SERIES

GENERATING NON-STANDARD
MULTIVARIATE DISTRIBUTIONS WITH
AN APPLICATION TO Mismeasurement
IN THE CPI

Matthew D. Shapiro
David W. Wilcox

Technical Working Paper 196

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 1996

We are grateful to Francis X. Diebold for suggesting Method II to us. Shapiro gratefully acknowledges the support of the Alfred P. Sloan Foundation. This paper is part of NBER's research programs in Monetary Economics and Productivity. Any opinions expressed are those of the authors and not those of the Board of Governors of the Federal Reserve System, other members of its staff, or the National Bureau of Economic Research.

© 1996 by Matthew D. Shapiro and David W. Wilcox. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

NBER Technical Working Paper 196
May 1996

GENERATING NON-STANDARD
MULTIVARIATE DISTRIBUTIONS WITH
AN APPLICATION TO MISMEASUREMENT
IN THE CPI

ABSTRACT

This paper shows how to generate the joint distribution of correlated random variables with specified marginal distributions. For cases where the marginal distributions are either normal or lognormal, it shows how to calculate analytically the correlation of the underlying normal distributions to induce the desired correlation between the variables. It also provides a method for calculating the joint distribution in the case of arbitrary marginal distributions.

The paper applies the technique to calculating the distribution of the overall bias in the consumer price index. The technique should also be applicable to estimation by simulated moments or simulated likelihoods and to Monte Carlo analysis.

Matthew D. Shapiro
Department of Economics
University of Michigan
Ann Arbor, MI 48109
and NBER

David W. Wilcox
Board of Governors of the
Federal Reserve System
20th Street and Constitution Avenue
Washington, DC 20551

1. Introduction

Statisticians and econometricians often wish to study multivariate random variables that are non-standard in the sense that they do not have closed-form distribution functions. The multivariate normal distribution—the most common multivariate distribution—affords a closed-form solution. Yet, there are many cases where normality is clearly a poor approximation, for example, when the range of the variable is restricted in at least one dimension or when at least one of the marginal distributions is clearly asymmetric. In such cases, the analyst will need more flexibility in specifying the joint distribution than is allowed by well-known functional forms.

We imagine that the analyst builds the multivariate distribution from information about its components. In particular, suppose the analyst specifies the following:

1. The marginal distribution of each of the components, and
2. The correlation among the components.

Such information about individual variables and their correlations might be available from previous studies, from economic theory, or from subjective priors. Analysts will likely find it easier to specify marginal distributions and correlations than joint distributions. Moreover, plausible marginal distributions and correlations frequently imply a joint distribution of unknown functional form.

In this paper, we propose two methods for calculating the distribution of a multivariate random variable with specified marginal distributions and correlations. Method I is applicable to the circumstance in which some of the components have

normal marginal distributions and the others have lognormal marginal distributions. For this case, we are able to show analytically how to impose the desired correlation among the variables. Method II is applicable when the components have any marginal distribution.

These methods for calculating the distribution functions of correlated random variables with arbitrary marginal distributions should be useful in a wide range of applications. We developed this method in order to study the distribution of measurement error in the consumer price index. We had information about the distribution of several components of measurement error in the CPI and used this information to construct the distribution of the overall measurement error.

Our method could also be used for estimation by simulated moments or simulated likelihoods where the variates have different marginal distributions. Moreover, it would be useful for Monte Carlo analysis where the analyst desires to impose different marginal distributions for the component shocks.

The remainder of the paper is organized as follows. The next section discusses the two methods for calculating the joint distribution. The following section applies the method to the case measurement errors in the consumer price index. In addition to presenting the results of using the joint distribution to aggregate the components of measurement error, we discuss pitfalls in analyses that are not based on well-specified probability statements. The final section offers some conclusions.

2. Methods for Calculating the Joint Distribution of Correlated Variables.

In this section, we discuss how to calculate the joint distribution of variables with specified marginal distributions. If the variables were independent, it would be

simple to calculate their joint distribution by Monte Carlo techniques. One would simply take independent draws from the specified distributions and tabulate the results. With correlated variables, it is necessary, however, to draw from the joint distribution. We discuss two methods for calculating the desired joint distribution while preserving the specified marginal distributions. The first method is for the case where the marginal distributions are either normal or lognormal. The second method can be used with any marginal distributions.

2.1. Method I: Normal and Lognormal Marginal Distributions.

The first method for calculating the joint distribution uses the moment-generating function of the multivariate normal. We calculate the correlation between the underlying normal distributions that induces the desired correlation between either a normal variable and a lognormal variable or two lognormal variables. Let (x,y,z) be distributed trivariate normal. The moment generating function of a trivariate normal is

$$\begin{aligned} m(t_1,t_2,t_3) &\equiv E\left[\text{Exp}[t_1 x + t_2 y + t_3 z]\right] \\ &= \text{Exp}\left[t_1 \mu_x + t_2 \mu_y + t_3 \mu_z + \frac{1}{2}(t_1^2 \sigma_x^2 + t_2^2 \sigma_y^2 + t_3^2 \sigma_z^2) \right. \\ &\quad \left. + \rho_{xy} \sigma_x \sigma_y t_1 t_2 + \rho_{yz} \sigma_y \sigma_z t_2 t_3 + \rho_{xz} \sigma_x \sigma_z t_1 t_3\right] \end{aligned} \quad (1)$$

where the parameters μ , σ , and ρ are the means, standard deviations, and correlations of x , y , and z . We need to determine how these parameters relate to the correlation of x and $\text{Exp}[y]$, that is, of a normal variable and a lognormal variable. To do so, we use the moment generating function. The derivative evaluated at zero of the moment generating function gives the expectation of a normal variable, while its level evaluated at one gives the expectation of a lognormal variable. For example, the expectation of $x \cdot \text{Exp}[y]$ is

$$\begin{aligned}
 E[x \cdot \text{Exp}[y]] &= \left. \frac{\partial m(t_1, t_2, t_3)}{\partial t_1} \right|_{t_1=0, t_2=1, t_3=0} \\
 &= (\mu_x + \rho_{xy} \sigma_x \sigma_y) \text{Exp}[\mu_y + \frac{1}{2} \sigma_y^2].
 \end{aligned} \tag{2}$$

Note that the mean of $\text{Exp}[y]$ is $\text{Exp}[\mu_y + \frac{1}{2} \sigma_y^2]$ and that the standard deviation of

$\text{Exp}[y]$ is $\text{Exp}[\mu_y] \text{Sqrt}[(\text{Exp}[\sigma_y^2])^2 - \text{Exp}[\sigma_y^2]]$.¹ Equation (2) and some

manipulation yields that the correlation of x and $\text{Exp}[y]$ is

$$\text{corr}[x, \text{Exp}[y]] = \frac{\text{Exp}[\frac{1}{2} \sigma_y^2] \rho_{xy} \sigma_y}{\text{Sqrt}[\text{Exp}[\sigma_y^2] (\text{Exp}[\sigma_y^2] - 1)]}. \tag{3}$$

Given values of σ_y and $\text{corr}[x, \text{Exp}[y]]$, one can solve equation (3) for the correlation ρ_{xy} of the underlying normal variables that induces the desired correlation between the normal and lognormal variables.

One can use a similar calculation to induce the correlation between two lognormal variables. From the moment generating function, the expectation of $\text{Exp}[y] \cdot \text{Exp}[z]$ is

$$\begin{aligned}
 E[\text{Exp}[y] \cdot \text{Exp}[z]] &= m(0, 1, 1) \\
 &= \text{Exp}[\mu_y + \mu_z + \frac{1}{2} (\sigma_y^2 + 2 \rho_{yz} \sigma_y \sigma_z + \sigma_z^2)].
 \end{aligned} \tag{4}$$

Using equation (4), the expression for the correlation of $\text{Exp}[y]$ and $\text{Exp}[z]$ is

¹The first and second moments of the lognormal variable can be calculated from m evaluated at t_2 equal to one and two.

$$\text{corr}[\text{Exp}[y], \text{Exp}[z]] = \frac{\text{Exp}[\frac{1}{2}(\sigma_y^2 + \sigma_z^2)](\text{Exp}[\rho_{yz} \sigma_y \sigma_z] - 1)}{\text{Sqrt}[\text{Exp}[\sigma_y^2](\text{Exp}[\sigma_y^2] - 1) \text{Exp}[\sigma_z^2](\text{Exp}[\sigma_z^2] - 1)]} \quad (5)$$

One can solve equation (5) for the correlation ρ_{yz} between the underlying normal variables that induces the desired correlation between the lognormal variables.

2.2. Method II: Arbitrary Marginal Distributions.

The second method for calculating the joint distribution is more general, but more cumbersome. Unlike the first method, which exploits the functional form of the normal and lognormal marginal distributions, this method can be used to analyze arbitrary, correlated random variables. To induce a correlation between two variables with different marginal distributions, do the following. First, draw independent samples from each of the distributions. Second, sort the two samples. Third, calculate the correlation between the sorted samples. Call this correlation ρ_{\max} . Fourth, unsort both samples and resort a fraction $\lambda = \frac{\rho}{\rho_{\max}}$, where ρ is the desired correlation

between the variables. In general, ρ_{\max} will be less than one. Thus, ρ_{\max} puts a ceiling on the range of the admissible values for ρ . (In the case of a normal variable and a lognormal variable, the ceiling is also relevant. In that case, ρ_{\max} is given by equation (3) evaluated at $\rho_{xy}=1$.)

In the special case of the normal and lognormal variables studied under method I, this algorithm will give the exact value of ρ owing to the linearity of equation (3) in ρ_{xy} . For other distributions, it may be necessary to iterate to find the fraction λ that generates the desired ρ .

This procedure is straightforward to generalize for inducing correlations among more than two variables.

2.3. Discussion of Method I versus Method II.

The joint distribution of two or more random variables is not uniquely determined by the marginal distributions of the components and the correlations among them. Indeed, Method I and Method II yield different joint distributions. For the special case of normal and lognormal distributions, Method I delivers a homogeneous sample: Every realization is drawn from the same population distribution and the population distribution itself exhibits the desired correlation pattern across components. By contrast, Method II delivers a heterogeneous sample: The sample realizations are drawn from two different population distributions. Realizations from the first distribution exhibit zero correlation across components while realizations from the second distribution exhibit correlation equal to ρ_{\max} . The joint distribution is formed by blending these two samples to achieve the desired overall correlation between the components.

The joint distribution of generated by Method I will be a smooth function, while the distribution generated by Method II will have a ridge. In many contexts, the smoothness of the distribution function generated by Method I will be desirable. But in other contexts, Method II will suffice notwithstanding the ridge. For example, in the next section we use the multivariate distribution to study the sum of the random variables. The distribution of the sum of random variables will be smooth even if their joint distribution has the ridge induced by Method II.

3. Application to Mismeasurement in the CPI.

There are several measurement problems in the consumer price index. These include the effects of substitution across items, within items, across outlets, and the effects of new items and unmeasured changes in quality. In Shapiro and Wilcox (1996), we follow the tradition in the small, but growing, “CPI bias” literature of estimating the extent of these effects component-by-component and then aggregating them to get an estimate of the extent of the overall bias in the CPI as a measure of the cost of living. In this section, we describe how we use the methods discussed in Section 2 to calculate the distribution of the overall bias in the CPI.

We state the bias in the CPI in terms of a probability distribution for several reasons. First, doing so highlights that there is uncertainty about the magnitude of the effects. Second, this approach allows us to be explicit about our beliefs concerning the magnitude of the effects and the uncertainty surrounding them. Third, specifying the probability distribution of the component effects allows us to aggregate the effects rigorously and to calculate the probability distribution of the total effect. Fourth, we have great flexibility in specifying the probability distributions. For example, we can allow the uncertainty about the size of the effect to be asymmetric, to be such that the effect is always positive, or allow for correlations between the effects.

3.1. Specifying the Component Distributions

To specify the probability distributions of the effects, we examine a wide range of previous studies. While the parameter values are based on a reading of the evidence and considerations of economic theory, they are admittedly subjective. In specifying the distributions, we considered four features: the central tendency, the

shape, the dispersion, and the correlation of the effects. Our judgements about the central tendency of the distributions are based on our reading of the best available estimates for the various effects. The evidence on the other features of the probability distributions is almost non-existent, so the specifications are almost entirely judgmental. Unless we had a good reason for doing otherwise, we assume that the uncertainty about the effects was symmetric. In two cases, there is good reason to believe that the effect is bounded below by zero and the uncertainty is asymmetric. We parameterize the symmetric distributions as normal and the bounded, asymmetric distribution as lognormal. These are chosen for ease of computation and because they capture the basic shapes, not because there is particular evidence about the functional form of the uncertainty. We now briefly summarize the component effects and the source of evidence about them.²

Across-strata effect. This bias arises because the CPI is a modified Laspeyres index and therefore does not take into account that consumers substitute across goods and services when relative prices change. Aizcorbe and Jackman (1993) have reaggregated the strata of the CPI using a Tornqvist index.³ Diewert (1976) shows a Tornqvist index is a second-order approximation to the true cost of living index. The estimate of the across-strata effect is based on the difference between the growth in the Tornqvist index and a modified Laspeyres index.

Within-strata effect. The BLS again assumes a modified Laspeyres index when it aggregates individual price observations to form the price indexes for the

²See Shapiro and Wilcox (1996) for a detailed discussion of the theory underlying these effects and of the evidence.

³In the terminology of the CPI, a stratum is an item within a geographical area.

strata. The within-strata effect arises for two, intertwined reasons. First, consumers alter their shopping patterns in response to changes in the relative prices offered at different outlets for the same good. The modified Laspeyres formula ignores this within-stratum substitution. Second, the method of aggregating individual prices can impart biases. (This second issue has been termed “formula bias.”) These two issues cannot be analyzed separately. In particular, the magnitude of the bias induced by the method of aggregation is a function of the degree to which consumers substitute across outlets. Hence, we treat within-strata substitution and the issue of the aggregation formula as a single effect. To quantify this bias, we rely on the difference at the substratum level between a modified Laspeyres index and an alternative index that provides an approximation to the true cost of living.⁴

New item and new outlet effects. The new item effect arises because BLS procedures implicitly assume that a new item delivers no consumer surplus at the price that it is introduced into the index. Similarly, the new outlet effect arises because the BLS assumes that any price differential at a new outlet reflects a difference in quality, not an increase or decrease in the cost of living. There is no systematic evidence on the new item effect, although there are several dramatic examples of its importance.⁵ Our estimate is thus conjectural, although it derives from an assessment of what fraction of strata in the CPI have frequent introduction of new items. For the new

⁴See Moulton (1993), Reinsdorf (1996), Moulton and Smedley (1995), and Reinsdorf and Moulton (1996) for both theory and empirical results, and, for a related calculation, Armknecht, Moulton and Stewart (1995). Shapiro and Wilcox (1996) discuss the details of this computation.

⁵Examples of new goods that generate significant consumer surplus include new models of automobiles, generic drugs, and new brands of breakfast cereals.

outlet effect, we rely on Reinsdorf (1993) and Lebow, Roberts, and Stockton's (1994) analysis of Reinsdorf's evidence.

Quality change. The BLS attempts to make adjustments for changes in quality of items in the CPI, but the adjustments are incomplete and imperfect. Evidence on quality change must be collected item by item, so it is necessarily fragmentary. We base our estimate on systematic evidence for durable goods (Gordon, 1990) and other fragments of evidence which largely suggest that unmeasured quality change is positive.

Distribution of the components: Uncertainty, Shape, and Correlation. The top panel of Table 1 reports the means, tail probabilities, and the functional forms that we specified for the distribution of each of the effects discussed above. It also shows the standard deviation implied by the specifications. The means are based on the evidence discussed in the previous paragraphs. The specification of the dispersion, shape, and correlation of the distributions is largely judgmental. In parameterizing the dispersion, we attempt to convey our degree of confidence about the magnitude of the effect. We specify the dispersion in terms of the 10 percent tail probabilities because we found them easier to assess subjectively than standard deviations. For the normally distributed effects, we specify both tails; for the lognormally distributed effects, we specify the upper tail.⁶ For effects with better information (especially the across- and within-strata effects), our distributions are fairly tight. For the new item and quality

⁶Unlike the normal, the lognormal is not uniquely determined by specifying the mean and tail probabilities. There are two distributions corresponding to a given mean and tail probability. We chose the one that had the typical lognormal shape. (The other has a spike close to zero and a long, flat tail.) See notes to Table 1 for the parameters of the distribution of the lognormal variables.

change effects, which are largely driven by innovation in the economy at large, we specify wider distributions.

For the new items and new outlets effect, we have a strong belief that the effects are positive, so we assume lognormality. For the other effects, following the principal of insufficient reason, we assume normality.

Finally, we specify that a subset of the effects are correlated. This specification is based solely on theoretical considerations. Three of the effects—within-strata, new goods, and new outlets—are functions of how readily consumers switch their spending patterns. Hence, if we underestimate the extent to which consumers economize by finding the cheapest good within a stratum, we are also likely to underestimate the extent to which consumers search for the lowest-price outlet. Without much specific justification, we assign a value of 0.25 to the correlation among the estimates of these effects. We consider below how sensitive the results are to alternative assumptions about this correlation.

3.2. Results

To calculate the distribution of the overall bias in the CPI, we use Method I as follows: We calculate the means and standard deviations of the multivariate normal corresponding to the normal and lognormal marginal distributions given in Table 1. We calculate the correlation of the multivariate normals required to induce the desired correlations given in Table 1 using equations (3) and (5). We draw a replication from the multivariate normal,⁷ exponentiate the lognormal components, and sum the

⁷In practice, we draw from five independent, univariate normals and use the Cholesky factorization of the covariance matrix to calculate the linear combination yielding the multivariate normal.

components. We replicate this procedure many times to give a good estimate of the distribution of the sum of the components. Method II gives identical results (up to simulation error). The results are based on 100,000 replications (50,000 draws from the underlying multivariate normal and their conjugate values).

Table 2 reports the mean, standard deviation, and percentiles of the distribution of the total bias for different assumptions about the correlation of the correlated effects. The first column of numbers gives the results for the 0.25 correlation of the second, third, and fourth components. The second and third columns report results for correlations of 0.0 and 0.5. Figure 1 plots the distribution under the three assumptions about the correlation.⁸

The mean estimate of the total bias is 1.0 percentage point per year. Consider first the distribution for the baseline specification with the correlations equal to 0.25. The median is 0.97, indicating only slight right skewness.⁹ Figure 1 shows that, despite the decidedly nonnormal shape of the lognormal components, the distribution of the total bias resembles a normal distribution. There is substantial uncertainty concerning the point estimate of the total bias. The standard deviation is 0.38. The 80-percent confidence interval runs from 0.56 to 1.47 percentage points.

Varying the correlation among the component effects has the expected impact on the distribution of the total bias. Lowering the correlation reduces the dispersion of the estimate and makes the distribution less skewed. Raising the correlation has the opposite effect. (See Figure 1.) Nonetheless, over the range of correlations

⁸The plotted distributions are smoothed using a normal kernel.

⁹The coefficient of skewness is 0.7.

considered, the quantitative impact on the distribution is fairly small, which is perhaps fortunate given that we have to guess the size of the correlation.

3.3. Probability Distributions versus “Ranges”.

Other analysts (the CPI Commission appointed by the Senate Finance Committee and Lebow, Roberts, and Stockton) express their estimates of the components of the biases in terms of ranges.¹⁰ Several questions arise in connection with the use of ranges, all of which are addressed by our use of the distributions.

- Where is the probability concentrated? Is it uniform across the interval, concentrated in the center, skewed?
- Is there zero probability outside of the range?
- How are ranges to be aggregated?

This last point is particularly troublesome. Analysts using ranges for the components have also reported a range for the total that is the outer envelope of the ranges of the components. Reporting ranges for both the components and the aggregate perhaps leaves the impression that the distribution within each is similar. Yet, as Figure 1 illustrates, the aggregate will necessarily resemble a bell curve. The central limit theorem has some bite even for sums of a only few elements.

In addition to the ranges, the Commission also reports point estimates in its interim report. In some cases, the point estimates are not close to the midpoint of the range. For example, the Commission’s point estimate for quality change is at the lower bound of the range. Hence, the Commission’s point estimates apparently do not

¹⁰At this point, only the preliminary report of the Commission is available.

correspond to the mean of the estimates. Instead, they might be interpreted as a cautious adjustment to indexation—that is, one that could be made with a high level of confidence of not overadjusting.

Our method allows such policies to be derived based on explicit probability calculations. Indeed, some policymakers have suggested that indexation of items in the Federal budget be modified in light of the overstatement of the increase in the cost of living by the consumer price index.¹¹ Assessing such proposals is far beyond the scope of our research. Nonetheless, the probability distribution we provide for the overall bias in the CPI will be relevant for policymakers confronted with such issues.

4. Conclusion

In this paper, we offer two methods for calculating the joint distribution of random variables with diverse marginal distributions. The methods allow for correlation among the variables. In one method, we show how to use the moment generating function of the multivariate normal distribution to calculate the joint distribution of correlated variables with normal and lognormal marginal distributions. In the other method, we show how to calculate the joint distribution for variables with arbitrary marginal distributions.

¹¹See Daniel Patrick Moynihan, "The CPI: An Easy Fix..." Washington Post September 26, 1995, opinion page. It is not clear from this op-ed piece whether Senator Moynihan is suggesting technical adjustments in the CPI that would reduce the increase in the CPI or legislative adjustments in indexation formulas relative to the CPI. Several governors also endorsed a change in CPI indexation to address as part a budget deal (see Judith Havemann, "Governors Recommend CPI Changes," Washington Post, December 5, 1994, page A9.

Our methods should have wide application in econometrics—both for estimation by simulation and for Monte Carlo analysis. In this paper, we show how we apply our methods to the problem of aggregating estimates of component biases in the CPI. We recommend that this method be used as a way for reporting assessments of biases in the CPI and for other, similar applications. Our method requires the analyst to be explicit concerning the uncertainty of the estimates. Highlighting this uncertainty is important both for the evaluation of official statistics and for related policy recommendations. The method allows for consistent aggregation of the effects. Uncertainty about the total is derived explicitly from uncertainty about the components. This quantification of the uncertainty about the magnitude of biases in the CPI should be helpful to policymakers debating adjustments to the indexation of tax brackets and benefits, and to private agents in designing indexed contracts.

REFERENCES

Advisory Commission to Study the Consumer Price Index. "Toward a More Accurate Measure of the Cost of Living: Interim Report to the Senate Finance Committee," mimeo. Senate Finance Committee, Washington, D.C., 1995.

Aizcorbe, Ana M., and Patrick C. Jackman. "The Commodity Substitution Effect in CPI Data, 1982-1991," Monthly Labor Review 116 (December 1993) pp. 25-33.

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

Diewert, W. Erwin. "Exact and Superlative Index Numbers," Journal of Econometrics 4 (1976) 115-125.

Gordon, Robert J. The Measurement of Durable Goods Prices. Chicago: University of Chicago Press, 1990.

Lebow, David E., John M. Roberts, and David J. Stockton. "Monetary Policy and 'The Price Level,'" mimeo. Board of Governors of the Federal Reserve System, Washington, D.C., 1994.

Moulton, Brent R. "Basic Components of the CPI: Estimation of Price Changes," Monthly Labor Review 116 (December 1993) 13-24.

_____ and Karin E. Smedley. "A Comparison of Estimators for Elementary Aggregates of the CPI" mimeo, Bureau of Labor Statistics, 1995.

Reinsdorf, Marshall. "The Effect of Outlet Price Differentials on the U.S. Consumer Price Index" in M. Foss, M. Manser, and A. Young, eds., Price Measurements and Their Uses, NBER Studies in Income and Wealth, vol. 57. Chicago: University of Chicago Press, 1993.

_____. "The 'Formula Bias' and Seller Substitution Bias in the U.S. CPI," mimeo, Bureau of Labor Statistics, 1996.

_____ and Brent R. Moulton, "The Construction of Basic Components of Cost of Living Indexes," in The Economics of New Goods, T. Bresnahan and R. Gordon, eds., (1996) forthcoming.

Shapiro, Matthew D. and David W. Wilcox, "Mismeasurement in the Consumer Price Index: An Evaluation," NBER Macroeconomics Annual 11 (1996) forthcoming.

Table 1

Parameterization of Distributions: Components of CPI Bias

| | Component | Distribution | Mean | Standard Deviation | 10-Percent Tail ^a |
|---|----------------|------------------------|------|--------------------|------------------------------|
| 1 | Across-strata | Normal | 0.20 | 0.122 | 0.0, 0.4 |
| 2 | Within-strata | Normal | 0.25 | 0.152 | 0.0, 0.5 |
| 3 | New Item | Lognormal ^b | 0.20 | 0.182 | 0.4 |
| 4 | New Outlet | Lognormal ^c | 0.10 | 0.091 | 0.2 |
| 5 | Quality Change | Normal | 0.25 | 0.182 | -0.05, 0.55 |

| Correlation Matrix | | | | |
|--------------------|---|------|------|---|
| | 1 | 2 | 3 | 4 |
| 1 | 1 | | | |
| 2 | 0 | 1 | | |
| 3 | 0 | 0.25 | 1 | |
| 4 | 0 | 0.25 | 0.25 | 1 |
| 5 | 0 | 0 | 0 | 0 |

Source: Shapiro and Wilcox (1996).

^aTwo-tailed for normal distribution and total; one-tailed for lognormal.

^bTable gives mean, standard deviation, and correlation (below) of level of component. Lognormal parameters (mean and standard deviation of underlying normal) are $\mu=-1.9110$ and $\sigma=.7755$.

^cLognormal parameters are $\mu=-2.603$ and $\sigma=.7755$.

Table 2

Total Bias in the CPI

| Distribution of Total Bias | | | | |
|-----------------------------------|----|------|------|------|
| Mean | | 1.0 | 1.0 | 1.0 |
| Standard Deviation | | 0.38 | 0.34 | 0.42 |
| Percentiles: | 1 | 0.25 | 0.30 | 0.21 |
| | 5 | 0.45 | 0.49 | 0.41 |
| | 10 | 0.56 | 0.60 | 0.53 |
| | 25 | 0.75 | 0.78 | 0.72 |
| | 50 | 0.97 | 0.98 | 0.96 |
| | 75 | 1.21 | 1.20 | 1.23 |
| | 90 | 1.47 | 1.42 | 1.51 |
| | 95 | 1.64 | 1.56 | 1.71 |
| | 99 | 2.08 | 1.91 | 2.23 |
| Assumed correlation of components | | 0.25 | 0.0 | 0.5 |

Note: Table shows mean, standard deviation, and percentiles of the distribution of total bias in the CPI aggregated from the distributions of the component effects under alternative assumptions about the correlation of components 2, 3 and 4.

Figure 1

**Distribution of Total Bias in CPI:
Alternative Correlations of Components**

