 Journal of Business & Economic Statistics, January 1986, Vol. 4, No. 1

Effects of Rotation Group Bias on Estimation of Unemployment

Gary Solon
Department of Economics, University of Michigan, Ann Arbor, MI 48109

Previous analysis of rotation group bias in the Current Population Survey has concluded that if the biases are additive, the ratio and composite estimators of month-to-month change in unemployment are unbiased. This article shows that if the biases contain a multiplicative aspect, both estimators of change are then biased. The article also presents some empirical results that cast doubt on the validity of a purely additive model.

KEY WORDS: Response error; Unemployment statistics; Current Population Survey.

1. INTRODUCTION

The Current Population Survey (CPS), the U.S. Census Bureau's monthly survey of households, has a rolling panel structure. A household whose address is selected for the sample is interviewed for four consecutive months, rotated out of the sample for eight months, and then interviewed for another four months before being retired from the sample. (If the household moves from the selected address, however, it is not followed. Instead, the CPS interviews the household that moves into the selected address.) As a result, in any particular month, the CPS sample consists of eight "rotation groups": the households with addresses in the survey for the first time, for the second time, . . . , for the eighth time.

Since all eight rotation groups are random samples of the population, they can be used to generate eight separate estimates of a population characteristic such as the number unemployed. In principle, the estimates should differ only by random error. In practice, though, the estimates from different rotation groups show sizable systematic differences. For example, several studies have found that the unemployment reported by the first rotation group tends to be about 10% higher than that reported by the full sample. Why responses vary systematically with time in the sample is unknown, but possible factors include conditioning of respondents or interviewers by repeated contacts, differences among rotation groups in the length and content of the questionnaire and differences in nonresponse, which household member is interviewed, and whether the interview is conducted by telephone or in person. (Further discussions of rotation group bias appear in Bailar 1975, U.S. Bureau of the Census 1978, and McCarthy 1979.)

A previous paper by Bailar (1975) assumed an additive model of rotation group bias and found that, although the estimation of *levels* of population characteristics is then biased, estimation of month-to-month *changes* is not. The current article allows for a multiplicative aspect to rotation group bias. Section 2 shows that with a multiplicative (or mixed additive and multiplicative) model, change estimation also is biased, and the bias takes a particularly complicated form in the case of the "composite estimator" used by the Census Bureau. Section 3 presents empirical evidence on the validity of the additive bias model for unemployment statistics. Section 4 summarizes and discusses the article's findings.

2. MODELS OF ROTATION GROUP BIAS

Let Y_t be the true (unobserved) level of unemployment in month t. The "ratio estimator" y_t estimates Y_t by inflating the current month's sample observations up to population size. The rather complicated procedure for the ratio estimator is detailed in U.S. Bureau of the Census (1978). For present purposes, the important point is that the ratio estimator is the sum of the contributions from the eight rotation groups: $y_t = \sum_{i=1}^{8} y_{it}$.

The ratio estimator, however, is not used by itself to generate the published unemployment statistics. Instead, the Census Bureau uses a "composite estimator" defined by

$$y_t^c = \tfrac{1}{2}y_t + \tfrac{1}{2}(y_{t-1}^c + y_t^{t-1} - y_{t-1}^t),$$

where y_t^{t-1} denotes the ratio estimator for month t based on the six rotation groups that appeared also in month $t - 1$ and y_{t-1}^t denotes the ratio estimator for month $t - 1$ based on the rotation groups common to both months. The remainder of this section analyzes the properties of both the ratio and composite estimators under alternative models of rotation group bias.

2.1 Additive Model

Bailar (1975) assumed that the ith rotation group's contribution to the ratio estimator has expected value

$$E(y_{it}) = \tfrac{1}{8}Y_t + a_i, \tag{1}$$

where a_i is a constant additive bias. Then the expected value of the ratio estimator is

$$E(y_t) = Y_t + \sum_{i=1}^{8} a_i. \tag{2}$$

Except in the improbable case that $\sum_{i=1}^{8} a_i = 0$, the ratio estimator of the *level* of unemployment is biased. When estimating month-to-month *change*, though, the bias differences out so that

$$E(y_t - y_{t-1}) = Y_t - Y_{t-1}.$$

The analysis for the composite estimator is more complicated. As shown in Bailar's appendix, successive substitution leads to

$$y_t^c = \tfrac{1}{2}(y_t + y_t^{t-1} - y_{t-1}^t) + \tfrac{1}{4}(y_{t-1} + y_{t-1}^{t-2} - y_{t-2}^{t-1}) + \tfrac{1}{8}(y_{t-2} + y_{t-2}^{t-3} - y_{t-3}^{t-2}) + \cdots . \quad (3)$$

Since

$$E(y_t^{t-1}) = Y_t + \tfrac{4}{3}(a_2 + a_3 + a_4 + a_6 + a_7 + a_8)$$

and

$$E(y_{t-1}^t) = Y_{t-1} + \tfrac{4}{3}(a_1 + a_2 + a_3 + a_5 + a_6 + a_7),$$

then

$$E(y_t^{t-1} - y_{t-1}^t) = Y_t - Y_{t-1} + \tfrac{4}{3}(a_4 + a_8 - a_1 - a_5). \quad (4)$$

Taking expectations in (3) and using (4) produce

$$\begin{aligned} E(y_t^c) &= \tfrac{1}{2}\left[2Y_t - Y_{t-1} + \sum_{i=1}^{8} a_i + \tfrac{4}{3}(a_4 + a_8 - a_1 - a_5)\right] \\ &\quad + \tfrac{1}{4}\left[2Y_{t-1} - Y_{t-2} + \sum_{i=1}^{8} a_i + \tfrac{4}{3}(a_4 + a_8 - a_1 - a_5)\right] \\ &\quad + \tfrac{1}{8}\left[2Y_{t-2} - Y_{t-3} + \sum_{i=1}^{8} a_i + \tfrac{4}{3}(a_4 + a_8 - a_1 - a_5)\right] + \cdots \\ &= Y_t + \sum_{i=1}^{8} a_i + \tfrac{4}{3}(a_4 + a_8 - a_1 - a_5) \end{aligned}$$

and $E(y_t^c - y_{t-1}^c) = Y_t - Y_{t-1}$.

Again, the estimator of the level of unemployment is likely to be biased, but the bias differences out when estimating change. Thus the composite estimator, like the ratio estimator, yields unbiased estimation of change under the assumption of additive rotation group bias.

2.2 Multiplicative Model

The simplest alternative to the additive model is a multiplicative model in which

$$E(y_{it}) = (1 + b_i)(\tfrac{1}{8}Y_t). \quad (5)$$

Then the expected value of the ratio estimator is

$$E(y_t) = (1 + \bar{b})Y_t, \quad (6)$$

where $\bar{b} = \frac{1}{8}\sum_{i=1}^{8} b_i$. Unless $\bar{b} = 0$, this estimator is biased. Moreover, the bias does *not* difference out when estimating change. Instead,

$$E(y_t - y_{t-1}) = (1 + \bar{b})(Y_t - Y_{t-1}).$$

The ratio estimator tends to exaggerate or understate the magnitude of change according to whether $\bar{b}$ is positive or negative.

Again, the analysis for the composite estimator is more difficult. Let

$$\bar{b}_1 = \tfrac{1}{6}(b_1 + b_2 + b_3 + b_5 + b_6 + b_7)$$

and

$$\bar{b}_2 = \tfrac{1}{6}(b_2 + b_3 + b_4 + b_6 + b_7 + b_8).$$

Then

$$E(y_t^{t-1}) = (1 + \bar{b}_2)Y_t$$

and

$$E(y_{t-1}^t) = (1 + \bar{b}_1)Y_{t-1}.$$

Taking expectations in (3) and substituting in the expressions above for $E(y_t)$, $E(y_t^{t-1})$, and $E(y_{t-1}^t)$ yield

$$\begin{aligned} E(y_t^c) &= \tfrac{1}{2}[(1 + \bar{b})Y_t + (1 + \bar{b}_2)Y_t - (1 + \bar{b}_1)Y_{t-1}] \\ &\quad + \tfrac{1}{4}[(1 + \bar{b})Y_{t-1} + (1 + \bar{b}_2)Y_{t-1} - (1 + \bar{b}_1)Y_{t-2}] + \tfrac{1}{8}[(1 + \bar{b})Y_{t-2} \\ &\quad + (1 + \bar{b}_2)Y_{t-2} - (1 + \bar{b}_1)Y_{t-3}] + \cdots \\ &= [1 + \tfrac{1}{2}(\bar{b} + \bar{b}_2)]Y_t + \tfrac{1}{2}(\bar{b} + \bar{b}_2 - 2\bar{b}_1) \\ &\quad \times (\tfrac{1}{2}Y_{t-1} + \tfrac{1}{4}Y_{t-2} + \tfrac{1}{8}Y_{t-3} + \cdots). \end{aligned}$$

This in turn implies that

$$\begin{aligned} E(y_t^c - y_{t-1}^c) &= [1 + \tfrac{1}{2}(\bar{b} + \bar{b}_2)](Y_t - Y_{t-1}) + \tfrac{1}{4}(\bar{b} + \bar{b}_2 - 2\bar{b}_1) \\ &\quad \times [Y_{t-1} - (\tfrac{1}{2}Y_{t-2} + \tfrac{1}{4}Y_{t-3} + \tfrac{1}{8}Y_{t-4} + \cdots)]. \end{aligned}$$

Thus the composite estimator is biased for both level and change, and in both cases its expectation depends on the entire history of the true series Y. Therefore the bias might vary in complicated ways, possibly even changing sign, according to the current and recent values realized for Y.

2.3 Mixed Model

The additive and multiplicative models can be combined into a more general mixed model, where

$$E(y_{it}) = a_i + (1 + b_i)(\tfrac{1}{8}Y_t).$$

This mixed model specializes to the purely additive case if $b_i = 0$ and to the purely multiplicative case if $a_i = 0$. Straightforward extensions of the previous analyses show that in this model,

$$\begin{aligned} E(y_t) &= \Sigma a_i + (1 + \bar{b})Y_t, \quad (7) \\ E(y_t - y_{t-1}) &= (1 + \bar{b})(Y_t - Y_{t-1}), \\ E(y_t^c) &= \Sigma a_i + \tfrac{4}{3}(a_4 + a_8 - a_1 - a_5) \\ &\quad + [1 + \tfrac{1}{2}(\bar{b} + \bar{b}_2)]Y_t + \tfrac{1}{2}(\bar{b} + \bar{b}_2 - 2\bar{b}_1) \\ &\quad \times (\tfrac{1}{2}Y_{t-1} + \tfrac{1}{4}Y_{t-2} + \tfrac{1}{8}Y_{t-3} + \cdots), \end{aligned}$$

and

$$E(y_t^c - y_{t-1}^c)$$
$$= [1 + \tfrac{1}{2}(\bar{b} + \bar{b}_2)](Y_t - Y_{t-1}) + \tfrac{1}{4}(\bar{b} + \bar{b}_2 - 2\bar{b}_1)$$
$$\times [Y_{t-1} - (\tfrac{1}{2}Y_{t-2} + \tfrac{1}{4}Y_{t-3} + \tfrac{1}{8}Y_{t-4} + \cdots)].$$

In the mixed model, as in the multiplicative model, both the ratio and composite estimators give biased estimation of both level and change, and the biases for the composite estimator depend on past as well as current values of Y.

3. EMPIRICAL TESTS OF THE ADDITIVE MODEL

Previous discussions, such as the one on page 88 of U.S. Bureau of the Census (1978), have argued that rotation group bias may difference out in the estimation of month-to-month change. Bailar's theoretical analysis verified this claim under a purely additive model, but the previous section has shown that her results no longer apply if there is a multiplicative aspect to rotation group bias. Therefore it is important to determine whether the additive model is consistent with empirical evidence.

How can the additive model be formally tested? The approach presented here focuses on the difference between the unemployment estimates from the jth rotation group (inflated to population size) and the estimates from the full sample. Equations (1) and (2) for the additive model imply that this difference has expected value

$$E(8y_{jt} - y_t) = 8a_j - \Sigma a_i,$$

so the expected discrepancy is constant and unrelated to the level of unemployment. In contrast, Equations (5) and (6) for the multiplicative model imply

$$E(8y_{jt} - y_t) = (b_j - \bar{b})Y_t,$$

and the analogous equations for the mixed model imply

$$E(8y_{jt} - y_t) = 8a_j - \Sigma a_i + (b_j - \bar{b})Y_t. \quad (8)$$

In both of these models, the expected discrepancy varies with the level of unemployment (unless $b_j = \bar{b}$). Furthermore, the additive model is a special case of the mixed model with $b_i = 0$ for all i. Therefore, if the additive model is valid, the coefficient of Y_t in Equation (8) must be zero. Strong evidence that this coefficient is not zero should therefore lead to rejection of the additive model.

Before examining the evidence, it will be useful to consider some summary statistics on differences among rotation groups. These statistics are based on monthly CPS data from January 1974 through June 1983. Table 1 displays the average monthly ratio estimate of unemployment during this period and shows the average contribution from each rotation group. The second column gives the average value of $8y_{jt} - y_t$, the difference between the estimate from the jth rotation group and the estimate from the full sample.

As found in previous studies, the first rotation group tends to report more unemployment than does the rest of the sample. Unemployment as estimated from the first group averaged about 627,000 more than the full sample average of 7,593,000. The other conspicuous outliers among the rotation groups are the sixth and seventh, whose respective unemployment estimates averaged 290,000 and 392,000 less than the full sample average. Consequently, the first, sixth, and seventh rotation groups are likely to provide the most powerful tests of the additive model. As is clear from Equation (8), even if the multiplicative or mixed model is valid, $E(8y_{jt} - y_t)$ is related to Y_t only for rotation groups with $b_j \neq \bar{b}$. The first, sixth, and seventh groups seem the best candidates for meeting this criterion.

Table 1. Average Ratio Estimates of Unemployment (in thousands), Total and by Rotation Group, January 1974–June 1983

	Average Estimate	Average Rotation Group Discrepancy ($8y_{jt} - y_t$)
Ratio Estimate	7,593	
Contributions by Rotation Groups		
1	1,028	627
2	952	20
3	939	−85
4	964	120
5	958	67
6	913	−290
7	900	−392
8	941	−68

Source: Unpublished Census Bureau tabulations.

3.1 Some Informal Evidence

The additive model predicts a constant expected value for $8y_{jt} - y_t$, whereas the multiplicative and mixed models predict that $E(8y_{jt} - y_t)$ is linearly related to the true unemployment level Y_t. An informal way to check these predictions against the data is simply to ask whether the magnitude of the sample observations of $8y_{jt} - y_t$ tended to be greater in high-unemployment months than in low-unemployment months. Table 2 addresses this question by sorting the sample of 114 observations into three equally sized groups (the 38 observations with the lowest ratio estimates of unemployment, the 38 with the highest, and the 38 in between) and then presenting the low and high groups' sample means of $8y_{jt} - y_t$ for $j = 1, 6, 7$.

Table 2 reveals that in all three of these rotation groups, the averages of $8y_{jt} - y_t$ have greater magnitude in the high-unemployment months than in the low-unemployment months, a finding that is qualitatively consistent with the multiplicative model. But this evidence, though appealingly straightforward, is also technically flawed. Using the averages of

Table 2. Average Rotation Group Discrepancies (in thousands) by Level of Unemployment

	Average Values of $8y_{jt} - y_t$		
	$j = 1$	$j = 6$	$j = 7$
Low-Unemployment Months	527	−168	−335
High-Unemployment Months	670	−396	−440

$8y_{jt} - y_t$ in the high- and low-unemployment months to estimate the coefficient of Y_t in Equation (8) would amount to the well-known "method of group averages" for estimating a regression relationship when the explanatory variable is measured with error. This estimation technique, however, is consistent only when the grouping is independent of the measurement error (see Madansky 1959). Grouping the data according to y_t, the erroneous measure itself, clearly does not satisfy this condition. Therefore the remainder of this section develops a more rigorous test of the additive model.

3.2 A Formal Test

Equation (8) can be rewritten as

$$8y_{jt} - y_t = 8a_j - \Sigma a_i + (b_j - \bar{b})Y_t + \varepsilon_{jt}, \quad (9)$$

where the error term ε_{jt} has mean zero and reflects sampling error and other sources of random variation. Similarly, Equation (7) can be rewritten as

$$y_t = \Sigma a_i + (1 + \bar{b})Y_t + u_t, \quad (10)$$

with $E(u_t) = 0$. Rearranging (10) yields

$$Y_t = (y_t - \Sigma a_i - u_t)/(1 + \bar{b}).$$

Then substituting this expression for Y_t into Equation (9) produces

$$8y_{jt} - y_t = \left(8a_j - \frac{1 + b_j}{1 + \bar{b}} \Sigma a_i\right) + \left(\frac{b_j - \bar{b}}{1 + \bar{b}}\right) y_t + \left(\varepsilon_{jt} - \frac{b_j - \bar{b}}{1 + \bar{b}} u_t\right). \quad (11)$$

Equation (11) reexpresses $8y_{jt} - y_t$ as a regression function of the observed ratio estimate y_t, rather than of the unobserved true unemployment level Y_t. Two points about this equation deserve emphasis. First, if the additive model is valid, $b_i = 0$ for all i and then the coefficient of y_t, $(b_j - \bar{b})/(1 + \bar{b})$, also equals zero. Consequently, evidence that this coefficient is not zero should lead to rejection of the additive model. Second, the equation cannot be consistently estimated by ordinary least squares because of the correlation between the regressor y_t and elements of the error term.

Consistent estimation is possible, however, by the instrumental-variable method. In the present case, the level of insured unemployment, based on administrative data from the unemployment insurance program, is an ideal instrumental variable for y_t. Insured unemployment is highly correlated with y_t, but since it is generated from an entirely different source, it is not correlated with the measurement errors in y_t and y_{jt}. (The data on insured unemployment were obtained from various issues of the *Monthly Labor Review*.)

Table 3 displays the results of instrumental-variable estimation of Equation (11) for $j = 1$, 6, and 7. The findings are qualitatively similar to those in the informal analysis above. For all three rotation groups, the estimated coefficients suggest that the magnitude of $8y_{jt} - y_t$ tends to be larger when unemployment is high. Furthermore, two of the slope coefficient estimates are very large relative to their estimated standard errors. (See the Appendix for a discussion of serial correlation.) By two-tailed tests, the hypothesis that

Table 3. Estimated Coefficients and Standard Errors From Instrumental-Variable Estimation

	$j = 1$	$j = 6$	$j = 7$
Intercept			
$8a_j - [(1 + b_j)/(1 + \bar{b})] \Sigma a_i$	131.70	170.07	−160.18
	(175.67)	(150.64)	(154.68)
Slope Coefficient			
$(b_j - \bar{b})/(1 + \bar{b})$	.065	−.061	−.031
	(.023)	(.019)	(.020)

NOTE: Standard errors are in parentheses.

the coefficient of y_t equals zero is easily rejected at the .01 level for the first and sixth rotation groups and is nearly rejected at the .10 level for the seventh rotation group. In combination, these results constitute very strong evidence against a purely additive model of rotation group bias.

A referee has suggested, however, that the sizable estimated slope coefficients in Table 3 might result not from a multiplicative aspect to rotation group bias but rather from seasonal variation in additive biases. If the additive biases happen to be larger in months with higher seasonal unemployment, this could indeed produce results such as those above. This possibility can be explored by extending the additive bias model to include month-specific biases. The only change this induces in regression Equation (11) is to add a vector of month dummy variables as regressors. If this extended model is valid, however, the coefficient of y_t still should be zero. Nevertheless, when the extended model is estimated by instrumental variables, the results are very similar to those in Table 3. The coefficient estimates for y_t are .057 (with estimated standard error .020) for the first rotation group, −.060 (.019) for the sixth, and −.028 (.019) for the seventh. Thus, allowing for seasonal variation in the additive biases does not eliminate the relationship between $8y_{jt} - y_t$ and the level of unemployment. Instead, the evidence against a purely additive model remains quite strong.

Of course, it is still possible that some other factors (e.g., occasional changes in the sample design or survey instrument) have produced temporal variation in additive biases. Such possibilities could be explored in future research. It is worth noting, though, that to explain the above results such factors would have to magnify both the positive tendency of $8y_{1t} - y_t$ and the negative tendencies of $8y_{6t} - y_t$ and $8y_{7t} - y_t$ during periods of high unemployment. The multiplicative bias explanation stands out for predicting precisely this pattern.

4. SUMMARY

Previous analysis of rotation group bias in the CPS has concluded that if the biases are additive, the ratio and composite estimators of month-to-month change are unbiased. This article has shown that if the biases contain a multiplicative aspect, both estimators of change are biased, and the composite estimator's bias depends on the entire history of the true series. Empirical tests have produced convincing evidence that the size of the discrepancies between the unemployment reported by certain rotation groups and that reported by the full sample is related to the level of unemployment. This finding casts serious doubt on the additive model of rotation group bias.

Taken together, the theoretical and empirical results suggest that rotation group bias may lead to some bias in the estimation of change as well as level of unemployment. Moreover, the peculiar form this bias may take in the case of the composite estimator raises the question of whether that estimator is really preferable to the ratio estimator.

ACKNOWLEDGMENTS

I thank the staff of the Census Bureau for generous assistance with unpublished data. I also thank two anonymous referees and the participants in the Census Bureau/Bureau of Labor Statistics gross flows conference (especially Lawrence Summers) for their valuable comments.

APPENDIX

The standard error estimates are based on an assumption that the error term is serially uncorrelated, which cannot be strictly correct. For example, the individuals counted in y_{1t} in one month are included in y_{t+1} the next month. If these individuals' unemployment experience is serially correlated, then so is $8y_{1t} - y_t$. Nevertheless, the magnitude of this serial correlation is bound to be small, because most of the variance in $8y_{1t} - y_t$ arises from the $8y_{1t}$ component, which includes a completely different set of individuals every month.

Consider the following simple formalization. Suppose that $\text{var}(y_{it}) = \sigma^2$ for all i and t, $\text{cov}(y_{it}, y_{jt}) = 0$ for $i \neq j$, and $\text{cov}(y_{it}, y_{i+1,t+1}) = \rho\sigma^2$ for $i = 1, 2, 3, 5, 6, 7$ (i.e., for the rotation groups that reappear in the sample in $t + 1$). Then the variance of $8y_{1t} - y_t$ is

$$\begin{aligned} \text{var}(8y_{1t} - y_t) &= \text{var}(7y_{1t} - \sum_{i=2}^{8} y_{it}) \\ &= 49\sigma^2 + 7\sigma^2 \\ &= 56\sigma^2. \end{aligned}$$

The first-order autocovariance of $8y_{1t} - y_t$ is

$$\begin{aligned} &\text{cov}(8y_{1t} - y_t,\ 8y_{1,t+1} - y_{t+1}) \\ &\quad = \text{cov}\left(7y_{1t} - \sum_{i=2}^{8} y_{it},\ 7y_{1,t+1} - \sum_{i=2}^{8} y_{i,t+1}\right) \\ &\quad = \text{cov}(7y_{1t}, -y_{2,t+1}) + \text{cov}(-y_{2t}, -y_{3,t+1}) \\ &\qquad + \text{cov}(-y_{3t}, -y_{4,t+1}) + \text{cov}(-y_{5t}, -y_{6,t+1}) \\ &\qquad + \text{cov}(-y_{6t}, -y_{7,t+1}) + \text{cov}(-y_{7t}, -y_{8,t+1}) \\ &\quad = -7\rho\sigma^2 + 5\rho\sigma^2 \\ &\quad = -2\rho\sigma^2. \end{aligned}$$

Therefore the first-order serial correlation of $8y_{1t} - y_t$ equals $-2\rho\sigma^2/56\sigma^2 = -\rho/28$. Thus, even in the extreme case where $\rho = 1$, the magnitude of the first-order serial correlation is less than .04.

[Received August 1984. Revised July 1985.]

REFERENCES

Bailar, Barbara A. (1975), "The Effects of Rotation Group Bias on Estimates From Panel Surveys," *Journal of the American Statistical Association*, 70, 23–30.

Madansky, Albert (1959), "The Fitting of Straight Lines When Both Variables Are Subject to Error," *Journal of the American Statistical Association*, 54, 173–205.

McCarthy, Philip J. (1979), "Some Sources of Error in Labor Force Estimates From the Current Population Survey," in *Data Collection, Processing, and Presentation: National and Local*, Washington, DC: U.S. Government Printing Office (National Commission on Employment and Unemployment Statistics), pp. 1–39.

U.S. Bureau of the Census (1978), *The Current Population Survey: Design and Methodology*, Technical Paper 40, Washington, DC: U.S. Government Printing Office (Dept. of Commerce).