Evaluation of Error Components in a Simulation Based Evaluation of a Survey Procedure

MoonJung Cho
Cho.Moon@bls.gov *

John L. Eltinge†

Abstract
This paper considers sensitivity analyses intended to supplement standard simulation-based evaluations of survey procedures. The principal ideas are motivated by, and illustrated with, a study of bootstrap variance estimators for the U.S. International Price Program.

Key Words: Approximation error, bootstrap, independence assumption, sensitivity analysis, U.S. International Price Program, variance estimator

1. Introduction

For many large-scale surveys, a complete characterization of the survey process involves a complex assemblage of many factors, including the frame; the sample design; instrument design; field work; edit, imputation and weighting methods; and computational steps in calculation of point estimators, variance estimators, and inferential statistics like interval estimators. The statistical properties of the resulting survey procedure (e.g., bias, variance or confidence-interval coverage rates) depend on random processes that may affect each of the factors above.

Survey methodologists often use simulation studies to obtain some indications of the practical impact of some of these factors on specific survey procedures. However, due to the abovementioned complexities, the designs of these simulation studies generally use relatively simple approximations to the true survey design, and focus primary attention on sensitivity of results to changes in a small number of characteristics of the design or the underlying population structures. For example, in a study of nonresponse adjustments, one might use a relatively simple approximation to the underlying sample design and population units, and focus primary attention on differences in underlying nonresponse-probability models and prospective weighting-adjustment methods.

These simplifications can provide a very reasonable approach to study of survey procedures, but naturally do not provide direct indications of the extent to which reported results may be sensitive to differences between features of the true survey process and corresponding features of the simplified process represented in the simulation study. Due to the complexity of many true survey processes, it generally will not be realistic to provide a comprehensive sensitivity analysis, but it can be useful to develop systematic approaches to study these sensitivity issues for selected features. One possible approach is as follows.

A. Provide a mathematical characterization of the survey features for which one may wish to study deviations between the true survey process and the approximate process used in a simulation study.

B. For some of the factors identified in step (A), produce a numerical evaluation of the impact that specified deviations have on particular performance criteria, e.g., point estimator bias and variance; variance estimator bias and stability; and confidence interval coverage rates and mean widths.

C. Use results from steps (A) and (B) to identify specific types and magnitudes of deviations that warrant further empirical study. This follow-up empirical work may often involve a substantial investment of additional resources (e.g., through additional data collection or in-depth analysis of other data) and thus is likely to be feasible only for high-priority cases identified in step (B).

The remainder of this paper will focus attention on one possible approach to carrying out step (B). Specifically:

1. Suppose that standard simulation results are based on functions $\hat{\theta}(X)$ of a random vector X. For example, $\hat{\theta}(\cdot)$ may be a parameter point estimator or interval estimator, and X is a vector of the underlying sample observations, weights and other numerical values used in calculation of $\hat{\theta}(\cdot)$

*U.S. Bureau of Labor Statistics, Office of Survey Methods Research, 2 Massachusetts Avenue NE, Washington, DC 20212
†U.S. Bureau of Labor Statistics, Office of Survey Methods Research, 2 Massachusetts Avenue NE, Washington, DC 20212

2. Suppose further that the standard simulation condition under which we generate X can be embedded in a more general model $F_X(\gamma)$ where $\gamma = (\gamma_1, \gamma_2, \cdots, \gamma_K)'$ is a K-dimensional parameter vector that characterizes the multi-dimensional approximations used in the standard simulation study. In addition, let γ_0 be the value of the vector γ under standard simulation conditions, and suppose that we wish to study the impact of deviations from standard conditions in one particular dimension k. In other words, our standard simulation study presents the properties of $\hat{\theta}(X)$ under the distribution $F_X(\gamma_0)$ only. To study this, we define a neighborhood Γ_{0k} that contains the value γ_{0k}. For example in some applications, one would use $\Gamma_{0k} = [\gamma_{0k} - \varepsilon_k \;\; \gamma_{0k} + \varepsilon_k]$ for some specified $\varepsilon_k > 0$. We then consider the properties of $\hat{\theta}(X)$ for several members of the class of distributions $\{F_X(\gamma) : \gamma_i = \gamma_{0i}\, , \, i \neq k\, ; \, \gamma_k \in \Gamma_{0k}\}$. For example, if $K = 2$ and we wish to study the sensitivity of simulation results to deviations of γ_1 from its standard-case value of γ_{01}, while keeping γ_2 at its standard-case value of γ_{02}, we will be considering the class of distribution $\{F_X(\gamma) : \gamma_1 = \gamma_{01}\, , \, \gamma_2 = \gamma_{02}\}$ where $\Gamma_{01} = [\gamma_{01} - \varepsilon_1 \;\; \gamma_{01} + \varepsilon_1]$ for some specified ε_1. In the language of steps (A)-(C) above, the selected dimension k corresponds to a specific type of deviation, and the range of values in the neighborhood Γ_{0k} corresponds to the magnitude of deviation under consideration. In some cases, selection of the neighborhood Γ_{0k} may be based on prior empirical information. However, in other cases (including the current study), the selection of Γ_{0k} is simply an initial exploratory step and does not necessarily correspond to the values of γ_k that one may encounter in a specific application.

3. Expand the simulation study to consider properties of $\hat{\theta}(X)$ for several values of γ in a neighborhood of the standard-condition value γ_0 .

4. Compare results from part (3) across the values of γ in the specified neighborhood of γ_0 . This can provide some indication of the extent to which, respectively, the null-case simulation results (generated under conditions γ_0) do or do not provide a satisfactory approximation to the properties of $\hat{\theta}(X)$.

2. Application to Variance Estimation for the U.S. International Price Program (IPP)

2.1 U.S. International Price Program

This work originated with a study of the properties of bootstrap variance estimators that were considered for use in the IPP. See Chapter 15 of the Bureau of Labor Statistics Handbook of Methods, International Price Indexes, Bobbitt et al.(2007), Chen et al.(2007), and Cho et al.(2007) for general background on the IPP and on the prospective bootstrap variance estimators. For the current work four of features are of primary importance.

First, the IPP uses a heavily stratified multistage sample design; at the final stage of sampling, field economists collect monthly price quotes for specific items. For a given item i and month t, define a price quote p_{it}, and also define a "short term relative" quantity $r_{it} = p_{it}/p_{i,t-1}$, the ratio of the price quotes for item i in months t and $t-1$, respectively. Price-index estimators $\hat{\theta}_t$ are complex nonlinear functions of the STR terms from current and previous months, r_{it} as well as sampling weights, computed through several levels of aggregation.

Second, due to the complexity of the design and the estimation process, use of standard analytic approaches to develop linearization-based variance estimators and other analysis tools may be problematic. Instead, the IPP has studied the potential use of bootstrap and other resampling-based methods for variance estimation. The current paper will restrict attention to one bootstrap variance estimator $\hat{V}$, based on the general approach of Rao et al.(1992); Bobbitt et al.(2007) provides details of the implementation.

Third, simulation work for the current paper focused on variance estimation for price-index estimators $\hat{\theta}_t$ for relatively large subpopulations known as "Chapters," which are two-digit strata with principal emphasis on Chapter P07 ("edible vegetables, roots, and tubers") and P90 ("optical, photographic, measuring and medical instruments"). Chapter P07 had a single design stratum, while Chapter P90 had six design strata. In addition, the distributions of item-level STR values, $\hat{\theta}_{it}$, for item i and month t, were quite different in Chapter P07 and Chapter P90, respectively. For example, in Figure 2 of Cho et al.(2007), the $\hat{\theta}_{it}$ values display a relatively high degree of dispersion within P07, but they are largely concentrated around $\theta = 1$ within P90. In addition, define the proportion π_{gt} to equal the proportion of items i inside classification group g at month t for which $\hat{\theta}_{it} = 1$, i.e., for which the price quote was the same in months $t-1$ and t. Figures 1 and 2 of the current paper display box plots of π_{gt} for months 1 through 12 for classification groups g in, respectively, the Chapters P07 and P90 and for each month $t = 1, \cdots, 12$. Note especially that for P07, the π_{gt} values have relatively small means and medians (below 0.5 for each month) and display a relatively high level of dispersion. On the other hand, for P90, the π_{gt} values have much higher means and medians (above 0.75 for all months) and a somewhat lower degree of dispersion. In other words, within Chapter P90, prices are very often constant across months, and so the values θ_{it} often equal one. Within Chapter P07, prices are less often constant across months, and so the values $\hat{\theta}_{it}$ are less often equal to one.

Fourth, the simulation work in Chen et al.(2007) and Cho et al.(2007) was based on 1000 independently selected samples s of units, in which each sample s was selected from an import-panel frame used previously to draw samples for the IPP production surveys. For the simulation works of seven chapters (two-digit strata) in Cho et al.(2007) and the current paper, each sample s involved approximately 3,300 items.

2.2 Exploratory Analysis of Sensitivity of Variance Estimator Properties to Dependence of STR Values on Unit Size

To carry out a simulation-based evaluation of the bootstrap variance estimator $\hat{V}$, it was important to have the samples s described above, and to assign an STR value to each item contained in a given selected sample s. Specifically, in keeping with the approach in Chen et al.(2007) and Cho et al.(2007), let θ_{git} be a random variable selected from a distribution function $F_{gt}(\cdot)$.

In these previous papers, the STR values θ_{git} were selected independently across months t and across items i and were treated as independent of unit-level trading dollar value Z_{gi}, say. These independence assumptions can be reasonable under some conditions, but it is useful to study the extent to which the properties of $\hat{V}$ may be sensitive to moderate deviations from these conditions. To do this, let θ_{git} be generated independently as described above. Define the alternative STR values θ^*_{git} by

$$\theta^*_{git} = \beta_0 + \beta_1\theta_{git} + \beta_2 Z_{gi} \tag{1}$$

where we will consider several possible values of β_1 which correspond to different degrees of dependence between θ^*_{git} and θ_{git}. Choose β_0 and β_2 such that $E(\theta^*_{git}) = E(\theta_{git})$, $V(\theta^*_{git}) = V(\theta_{git})$, and $\beta_1 > 0$. Specifically, we need to find β_0 and β_2 to satisfy the following:

$$E(\theta_{git}) = E(\theta^*_{git}) = \beta_0 + \beta_1\, E(\theta_{git}) + \beta_2\, E(Z_{gi}) \tag{2}$$

$$V(\theta_{git}) = V(\theta^*_{git}) = {\beta_1}^2\, V(\theta_{git}) + {\beta_2}^2\, V(Z_{gi})\,. \tag{3}$$

Routine algebra then leads to the expressions

$$\beta_1 = \left[1 - \beta_2^2 \left\{\frac{V(Z_{gi})}{V(\theta_{git})}\right\}\right]^{\frac{1}{2}} \tag{4}$$

and

$$\beta_0 = (1-\beta_1)\, E(\theta_{git}) - \beta_2\; E(Z_{gi}) \tag{5}$$

where $(\beta_0, \beta_1, \beta_2)$ are fixed coefficients selected to ensure $E(\theta^*_{git}) = E(\theta_{git})$ and $V(\theta^*_{git}) = V(\theta_{git})$. We can vary β_1 to induce varying degrees of dependence between θ^*_{git} and θ_{git}. β_2 is a free parameter that we can vary to induce varying degrees of dependence between θ^*_{git} and Z_{gi}. Varying β_2 allows us to explore the impact of varying degrees of association between θ^*_{git} and size-based weight Z_{gi} .

Specifically, for the IPP application, we considered five possible values of β_1, equal to $1.0, 0.99, 0.90, 0.75$ and 0.50, respectively. Note that the condition $\beta_1 = 1$ corresponds to the assumption that $\hat{\theta}_{git}$ is independent of Z_{gi}, as in the above-cited previous simulation studies. At the other extreme, the conditions $\beta_1 = 0.75$ or $\beta_1 = 0.50$ would correspond to hypothetical cases in which the STR values $\hat{\theta}_{git}$ were strongly associated with the size measure Z_{gi}. Also, for the numerical work in the present paper, we replaced the theoretical values $E(Z_{gi})$ and $V(Z_{gi})$ with the sample mean and variance of the trading dollar values in each specified chapter (two-digit stratum). Similarly, we replaced $E(\theta_{git})$ and $V(\theta_{git})$ with the sample mean and variance for STR values in each specified combination of Chapter and month t. We then computed values of $\beta_2 > 0$ and β_0 from expressions (4) and (5), the selected values of β_1, and the four moments described above.

Thus, the coefficients β_0 and β_2 varied across different combination of Chapter and month, thereby ensuring that the moment equations (2) and (3) are satisfied at the Chapter $\times$ month level of aggregation. Finally, note that equation (1) introduces a specific type of deviation from the previous assumption of independence of STR from trading dollar value. Specifically, expression(1) induces an offset $\beta_2 Z_{gi}$ that is constant over the months t. One could explore alternative forms of deviation from the assumption of independence, but the details are beyond the scope of the present paper.

2.3 Comparison Criteria

Our primary interest is the properties of bootstrap variance estimator (Rao et al., 1992). Define $\hat{\theta}_{cs}$ to be an estimated STR for a Chapter c and a sample s where $s = 1, \cdots, 1000$. Define

$$\bar{\hat{\theta}}_{c.} = 1000^{-1} \sum_{s=1}^{1000} \hat{\theta}_{cs}$$

and $\tilde{V}_{c.} = (1000 - 1)^{-1} \sum_{s=1}^{1000} \left(\hat{\theta}_{cs} - \bar{\hat{\theta}}_{c.} \right)^2$.

Let $\hat{V}_{cs}$ be a bootstrap variance estimator of a two-digit stratum c and sample s. Define the estimator of relative bias

$$\tilde{V}_{c.}^{-1} \left(\bar{\hat{V}}_{c.} - \tilde{V}_{c.} \right) \tag{6}$$

where $\bar{\hat{V}}_{c.} = 1000^{-1} \sum_{s=1}^{1000} \hat{V}_{cs}$.

In addition, define the degrees of freedom term

$$\hat{d} = 2 \left(\bar{\hat{V}}_{c.} \right)^2 \left\{ \tilde{V} \left(\hat{V}_{cs} \right) \right\}^{-1} \tag{7}$$

where $\tilde{V} \left(\hat{V}_{cs} \right) = (1000 - 1)^{-1} \sum_{s=1}^{1000} \left(\hat{V}_{cs} - \bar{\hat{V}}_{c.} \right)^2$.

2.4 Numerical Results

Figure 3 and 4 display numerical values of the simulation-based relative bias (6) and degrees-of-freedom term (7), respectively, for Chapter P07. For both cases, the horizontal axis corresponds to 36 consecutive months. Note that the relative-bias results differ substantially across different choices β_1, and appear to move downward an asymptote close to -0.7 as β_1 declines from 1.0 to smaller positive values. In addition, the degrees-of-freedom term also declines as we consider smaller values of β_1. Figure 5 and 6 display the corresponding relative-bias and degrees-of-freedom results for Chapter 90. These results are qualitatively similar to those obtained for P07.

3. Discussion

For complex surveys, simulation-based evaluations require consideration of approximations X_A to the true population, design, estimation procedures, and adequacy of the resulting approximations to the true properties.

We considered a simulation study based on STR observations θ_{git}. Instead of an independent selection of STR, we simulated θ^*_{git} by inducing moderate dependence of STR on Z_{gi}, weight related to size. For each Chapter in our study, we compared the properties of θ^*_{git}, which were newly simulated, with the ones of θ_{git} across 1000 samples. We also compared relative bias, and degrees of freedom of the variance estimator across different values of β_1.

One could consider extending the current work in several dimensions, including the following. First, as noted in Section 2.2 expression(1) induces one specific level-shift type of dependence of STR values on trading dollar values. One could consider exploration of other possible deviations from the standard assumption of independence. For example, if in expression(1) one replaced $\beta_2 Z_{gi}$ with $\beta_2 Z_{gi} e_{git}$ for some smooth positive function $h(\cdot)$, where $\{e_{git}, t = 1, \cdots, 36\}$ are independent and identically distributed random variable, with mean zero and finite variance, then (conditional on appropriate definitions of β_0, β_1 and β_2) one would have induced a violation of independence through which trading dollar value conditionally affects the variance, but not the mean, of the STR values θ^*_{git}.

Second, as noted in Section 1, one could incorporate this current sensitivity analysis into a broader study that obtained empirical information on the magnitudes of potential deviations from assumed conditions; and that considered alternative estimators tuned to those the deviations of largest observed magnitudes. These extensions are of interest, but detailed development is beyond the scope of the current work.

4. Acknowledgment

The views expressed in this paper are those of the authors and do not necessarily reflect the policies of the U.S. Bureau of Labor Statistics. The authors thank Te-Ching Chen, Patrick Bobbitt, James Himelein, Steve Paben, and Lawrence Ernst for many helpful comments on the International Price Program; and thank Eric Sampson of the ASA for help with preparation of the graphs.

REFERENCES

Bobbitt, P.A., Cho, M.J., and Eddy, R.M. (2005), "Comparing Weighting Methods in the International Price Program," *2005 Proceedings of the American Statistical Association*, Government Statistics Section [CD-ROM], 1006-1014.

Bobbitt, P.A., Paben, S.P., Cho, M.J., Himelein, J.A., Chen, T-C., and Ernst, L.R. (2007), "Application of the Bootstrap Method in the International Price Program," *2007 Proceedings of the American Statistical Association*, Survey Research Methods [CD-ROM], 2910-2917.

Bureau of Labor Statistics. *BLS Handbook of Methods.* U.S. Department of Labor. Available at http://www.bls.gov/opub/hom/homtoc.htm

Chen, T-C., Bobbitt, P.A., Himelein, J.A., Paben, S.P., Cho, M.J., and Ernst, L.R.(2007), "Variance Estimation for International Price Program Indexes," *2007 Proceedings of the American Statistical Association*, Survey Research Methods Section [CD-ROM], 1427-1434.

Chen, T-C. (2007), "SAS Programming Steps," International Price Program, Technical Report, Statistical Methods Division, Office of Prices and Living Conditions, U.S. Bureau of Labor Statistics.

Cho, M.J., Chen, T-C, Bobbitt, P.A., Himelein, J.A., Paben, S.P., Ernst, L.R., and Eltinge, J. L.(2007), "Comparison of Simulation Methods Using Historical Data in the U.S. International Price Program," *2007 Proceedings of the American Statistical Association*, Third International Conference on Establishment Surveys [CD-ROM], 248-255.

Fisher, I.K. (1922), *The Making of Index Numbers*, Boston: Houghlin Mifflin and Co.

Paben, S.P. (2006), "Creating a Universe of Short-Term Price Relatives," International Price Program, Technical Report, Statistical Methods Division, Office of Prices and Living Conditions, U.S. Bureau of Labor Statistics.

Powers, R., Eltinge, J.L., and Cho, M.J. (2006), "Evaluation of the Detectability and Inferential Impact of Nonresponse Bias in Establishment Surveys," *2006 Proceedings of the American Statistical Association*, Survey Research Methods Section [CD-ROM], 3577-3583.

Rao, J.N.K., Wu, C.F.J., and Yue, K. (1992), "Some recent work on Resampling methods for complex surveys," *Survey Methodology*, Vol.18, No. 2, 209-217.

Wang, X. and Himelein J.A. (2005a), "Simulating response rates with logit regression model," International Price Program, Technical Report, Statistical Methods Division, Office of Prices and Living Conditions, U.S. Bureau of Labor Statistics.

Wang, X. and Himelein J.A. (2005b), "A Top-Down Approach of Modeling IPP Short Term Relatives," International Price Program, Technical Report, Statistical Methods Division, Office of Prices and Living Conditions, U.S. Bureau of Labor Statistics.

Wolter, K. (1985), *Introduction to Variance Estimation*, New York: Springer-Verlag.

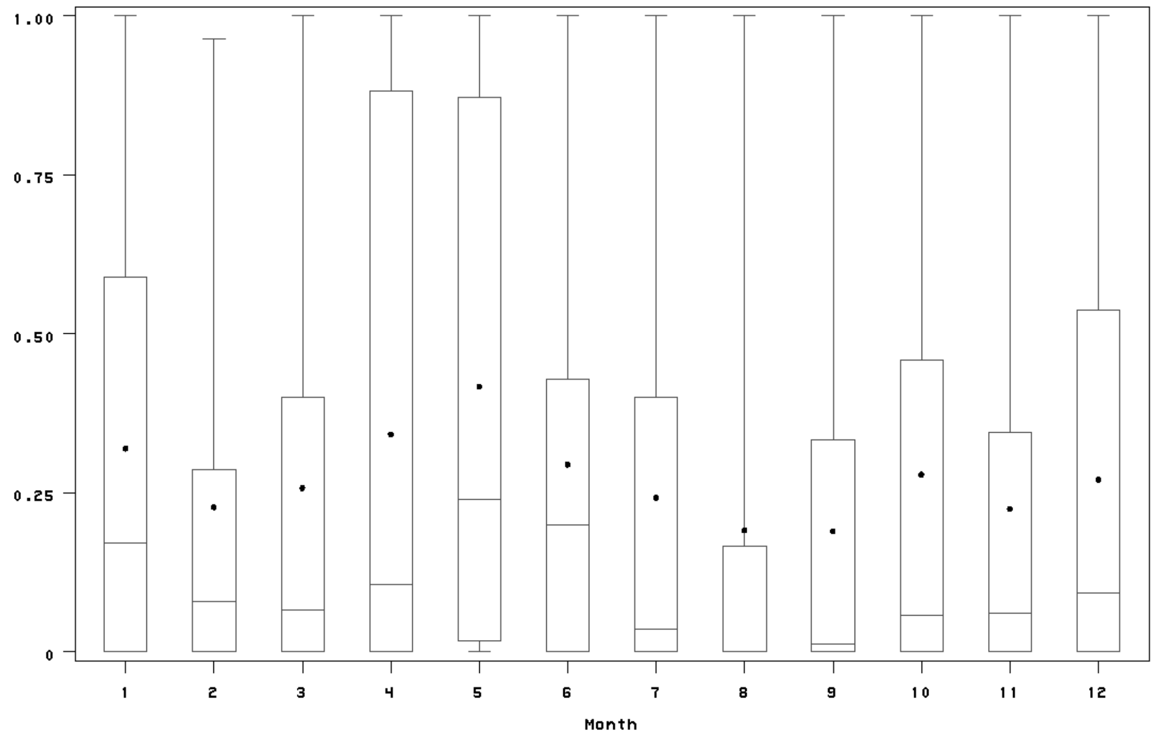

Figure 1: Boxplots of the Proportion of STR Values Equal to One for Classification Groups in Chapter P07: Historical Data from Months 1-12 over 13 years

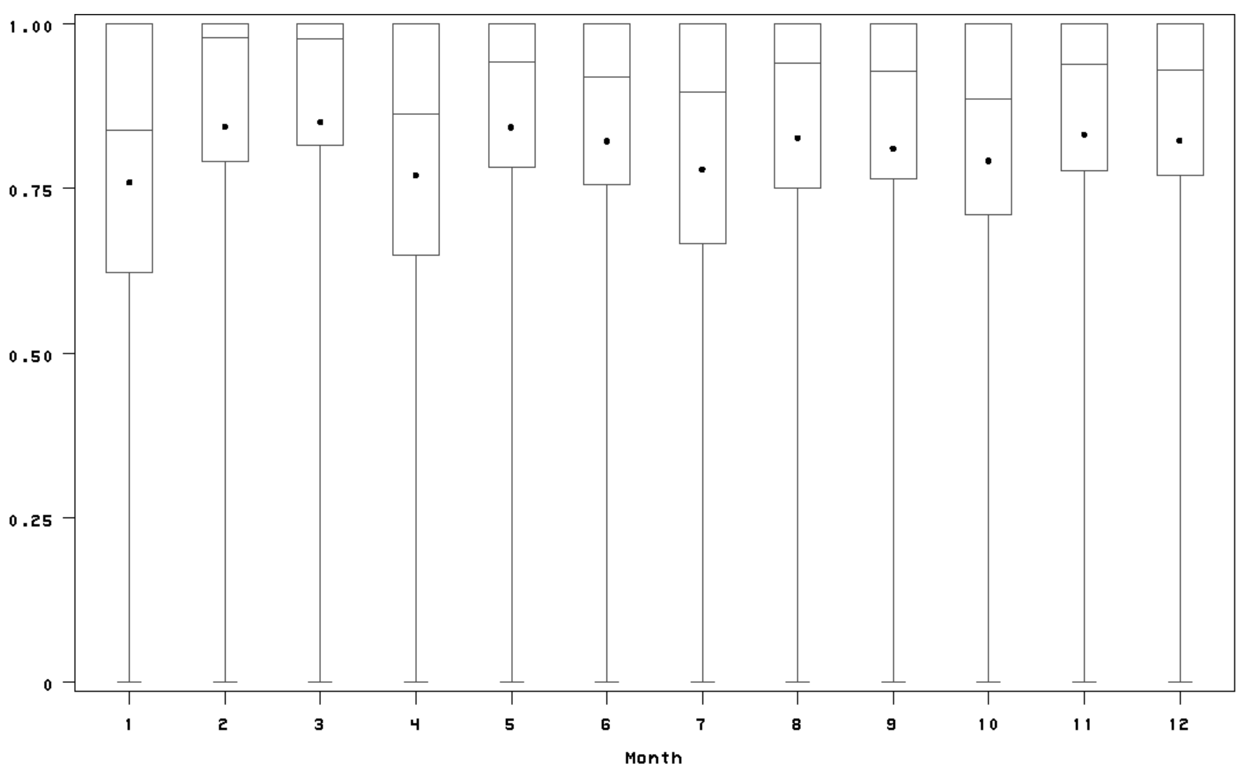

Figure 2: Boxplots of the Proportion of STR Values Equal to One for Classification Groups in Chapter P90: Historical Data from Months 1-12 over 13 years

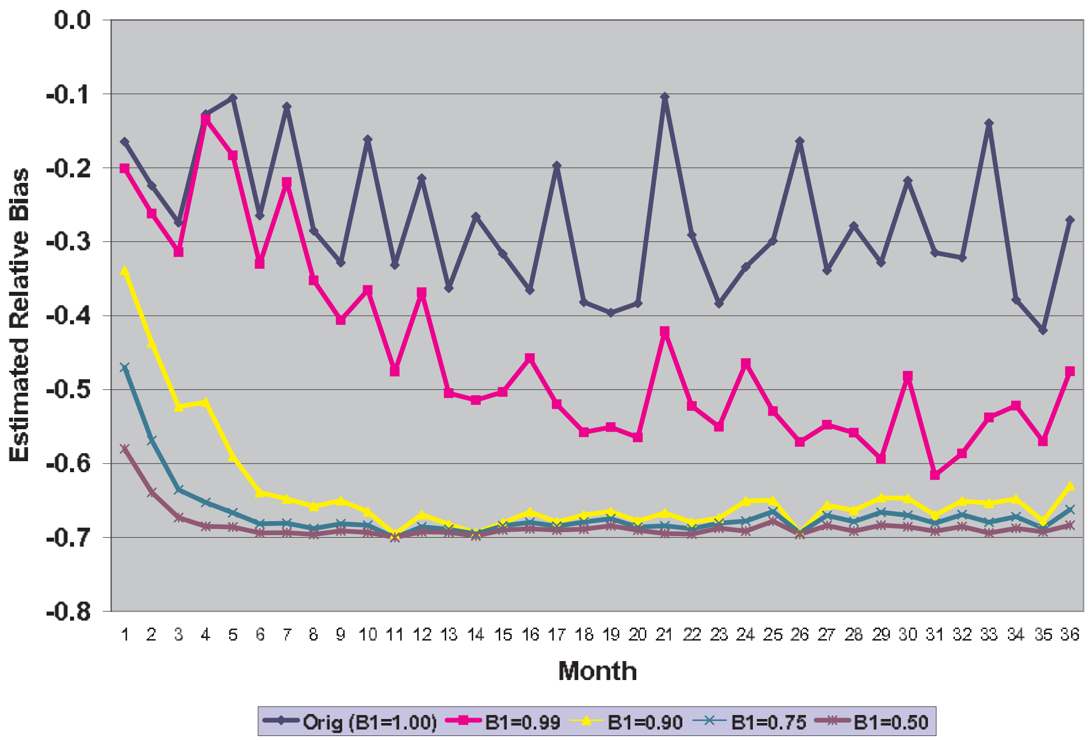

Figure 3: Comparison of Relative Bias for P07 across Five Different Values of β_1

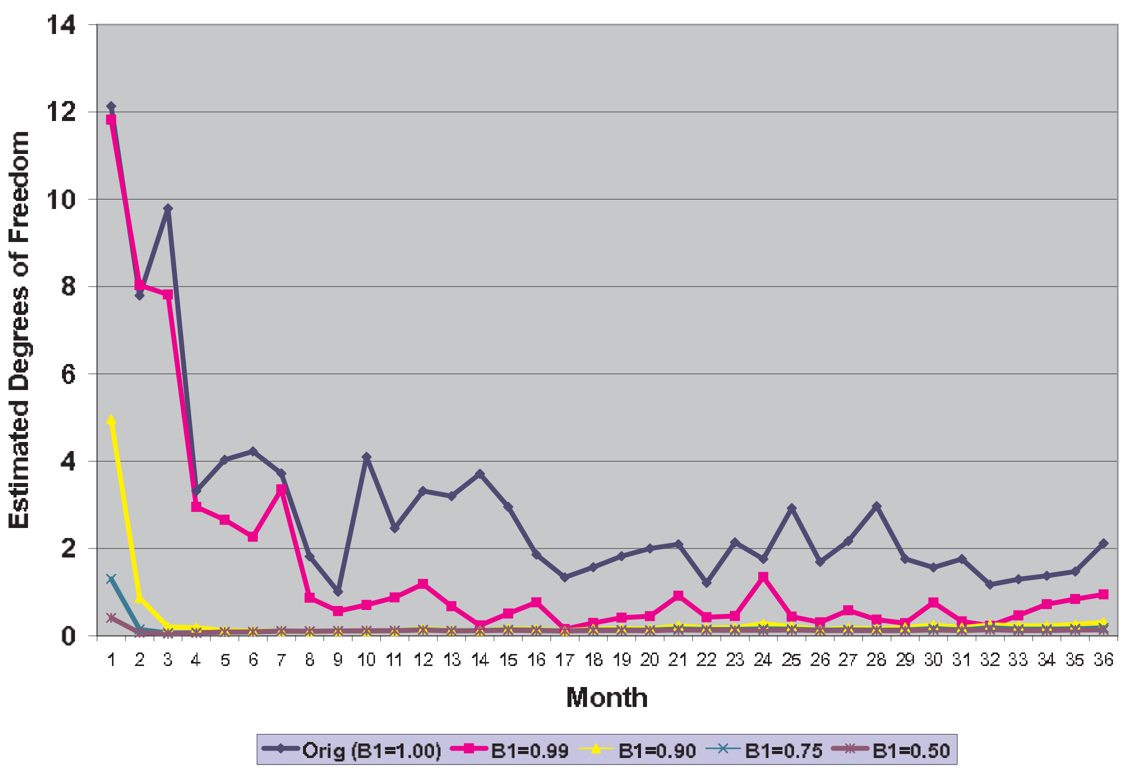

Figure 4: Comparison of $\hat{d}$ for P07 across Five Different Values of β_1

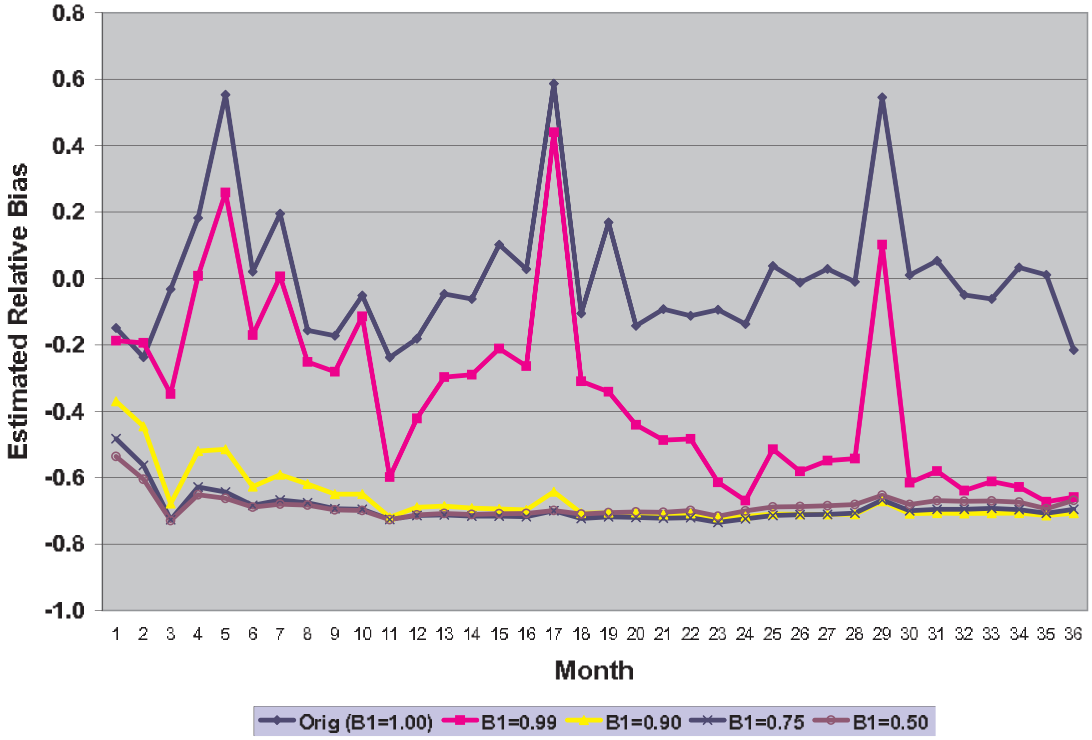

Figure 5: Comparison of Relative Bias for P90 across Five Different Values of β_1

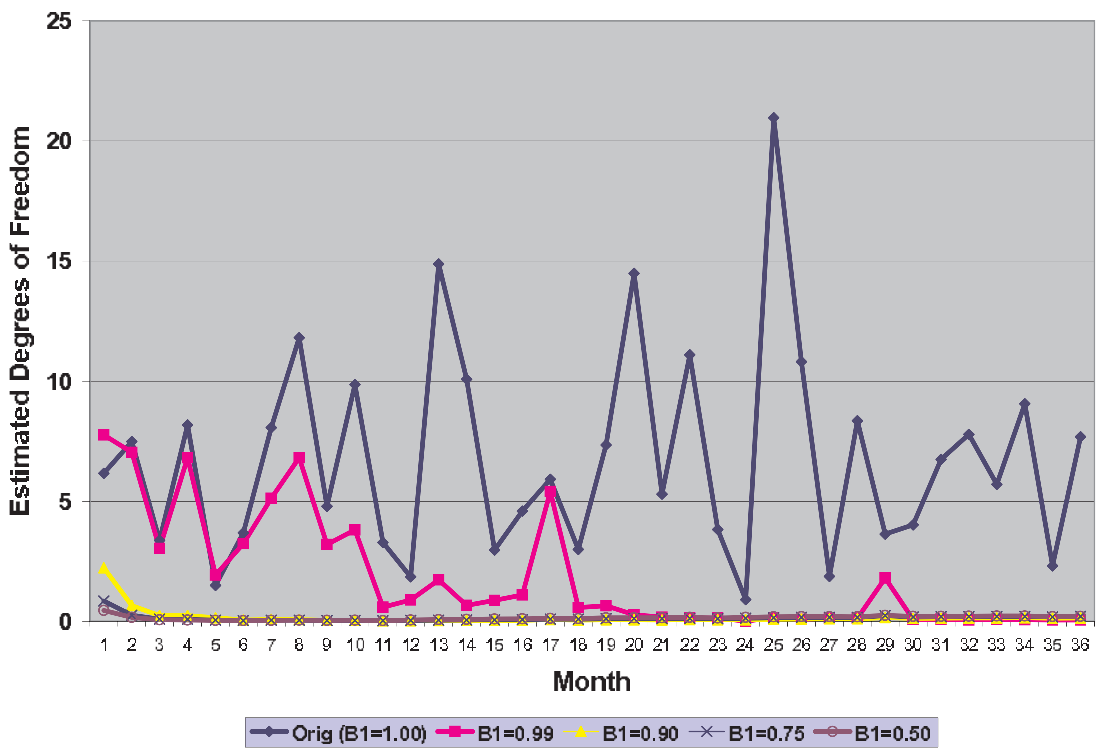

Figure 6: Comparison of $\hat{d}$ for P90 across Five Different Values of β_1