A GENERALIZED LEAST SQUARES WEIGHTING SYSTEM FOR THE CONSUMER EXPENDITURE SURVEY

Kimberly D. Zieschang, Bureau of Labor Statistics

1. Introduction

The Bureau of Labor Statistics, through a contract with the Bureau of the Census, collects information on consumer expenditures, income, and demographic characteristics from a two-component Consumer Expenditure (CE) Survey. The components are distinguished by survey instrument, the first being a three hour interview conducted with four rotating panels of consumer units (CU's) for five consecutive quarters, and the second being a two-week diary of expenditures to be completed in large part by the sampled consumer units. Each component in effect constitutes a separate survey, since there is no overlap between the respective samples, the only relationship between the two being that the samples for each are chosen from the same frame at the same time. There is considerable overlap in the kinds of information collected between the surveys, primarily though not exclusively demographic in nature. Some overlap also occurs in information collected on expenditures, though the surveys are oriented toward obtaining different kinds of expenditure information. The interview component is designed to provide information on major, infrequent purchases, "infrequent" meaning for practical purposes with a frequency of less than two weeks to a month. The diary component is oriented toward purchases with more than weekly or biweekly frequency, food purchases being one of the major categories of interest. However, the diary component is not explicitly limited to certain categories at present, in part to minimize confusion for the households completing the diary. This is the primary source of overlap in expenditure information collected.

In this paper the emphasis is on improving the accuracy, as measured by mean square error (MSE), of CE survey estimates of counts of consumer units partitioned into categories of demographic and economic interest, such as family composition and tenure status. The mode of accomplishing this increase in accuracy is develop a technique of weighting adjustment that uses ancillary information on the age, race, and sex composition of persons as part of a uniform, MSE minimizing estimation procedure for CU totals. The exposition first briefly reviews current practice for weighting adjustment and estimation in the CE surveys. We then introduce a method of weighting adjustment proposed by Don Luery (1980) and considered by Anthony Roman (1982) of the Bureau of the Census. This generalized least squares (GLS) method is extended to define a procedure that integrates overlapping demographic information from the surveys in a way that should reduce the MSE and improve the accuracy of key CU counts. The GLS procedure has been applied to thirteen quarters of data from the Consumer Expenditure survey covering 1980 I - 1983 IV and evaluated against the current Principal Person weighting procedure. The results of this study are described, followed by concluding remarks.

2. Current Procedures

The basic sampling weights for the CE surveys are determined according to the size measured in number of housing unit addresses of each of seven subframes from which sample addresses are chosen, relative to the sample size allocated to the subframe. These basic weights are adjusted by application of a "weighting control factor" to the basic weight, accounting among other things for unexpectedly clustered addresses, as for example, student housing in the "area segment" subframe. A further "noninterview factor" is applied to adjust for inability to obtain an interview for some sample units within cells defined by geographic area, tenure, household size, and race. All members of a respondent household at an address get this adjusted weight. Each member then receives a "second stage" adjustment to force the population of members classified into 48 age, race, and sex categories as estimated by the CE surveys in a given month to equal the U.S. population for that month in those age, race, and sex categories. These latter quantities are called "control totals" and we will refer to this as a weighting control procedure. The sources of control totals for the CE surveys are the U.S. Census and the Current Population Survey (CPS). A single "principal person" is chosen to represent the CU. The CU receives this person's weight. If the "principal person" is male, his weight and hence his CU's weight is multiplied by a "principal person factor" to adjust for a historical tendency for males to be underrepresented compared with females. The current stages of weighting beyond noninterview adjustment have two objectives: weighting control to add additional information to the estimation process, and the assignment of a weight to the CU, a sampling unit unique to the CE surveys. For additional information on the existing methodology see Alexander (1986).

The resulting "principal person" CU weights are then used for computing estimates of totals and averages for various quantities of interest from each of the surveys. No attempt is made at present to make an "integrated" selection between the two survey components of estimates of counts of persons or CU's within the demographic and economic categories for which average expenditures are currently computed. However, interest has arisen in making this "integration" because there is the need to know which survey, the CE diary or interview, is producing the best estimate for each of a number of counts. More generally, there is a need to combine information from both surveys in arriving at a single estimate more efficiently than merely "integrating" them. The alternative generalized least squares procedure described below is designed not only to perform the weighting "control" function of the current method, but this weighting "composition" function as well.

3. The Generalized Least Squares Procedure.

3.1 The Adjustment Equation and Computational Layout

The GLS procedure adjusts the sample weights after noninterview adjustment by minimizing theweighted squared adjustments subject to a set of linear "control" constraints the adjusted weights must satisfy. Antecedent studies on GLS in the weighting context are few. The least squares adjustment criterion was first proposed by Deming and Stephan (1940) though the raking algorithm they presented was based on an alternative "minimum discriminant information" criterion. GLS was proposed for use in CPS weighting research by Luery (1980). Based on Luery's work, the GLS methodology was subsequently proposed for a research project on survey weighting in the CE Survey by Roman (1982), and extended and applied in the present CE Survey context by Zieschang (1985a,b). GLS represents one of several possible ways to adjust survey weights to known control totals. The "least squared weight adjustment" criterion can be replaced by minimizing any other increasing convex function of weight adjustments to construct an adjustment algorithm. See Deming and Stephan (1940), Stephan (1942), Pugh, et al (1976), Alexander (1985), Alexander and Roebuck (1986), Fagan and Greenberg (1985), and Fienberg (1986) for some alternatives. The leading alternative methodology, Raking Ratio Estimation (RRE), is based on the minimum discriminant information adjustment criterion. RRE is currently used in CPS weighting, and has been extensively applied in weighting adjustment and related problems over the last twenty years. See Oh and Scheuren (1978), and the references therin. An important practical advantage of GLS over RRE in a survey production environment may be that it results in a finite rather than infinite algorithm for computation of adjustments to weighting cell entries. An exact solution is guaranteed for GLS under minimal consistency criteria for the constraints on the marginal totals of the adjusted weights.

The GLS problem can be expressed as

$$(1) \qquad \min_W \; (\Omega - W)'\Lambda^{-1}(\Omega - W) \quad \text{subject to} \quad MW = N$$

where Ω = the pre-adjustment sample weight vector of dimension n x 1, where n is the sample size;

W = the adjusted sample weight vector of dimension n;

Λ = the GLS weighting matrix of dimension nxn;

M = a kxn matrix whose columns provide the counts of k items for each of n CU's;

N = a kx1 vector of known "control" counts for each of the items corresponding to the rows of M.

In the usual single sample situation, the rows of M and N correspond to persons in each of a number of age, race, and sex categories, or any other item on which there is known control information that is also collected for each sample unit. The CE context is characterized by two samples. This implies dual control constraints, one set for each sample. In addition, it suggests the possibility of equating the survey counts of items on which no control information exists, but on which comparable information is collected in each sample. Both the "control" and "composition" aspects of weighting adjustment are straightforwardly accommodated in the constraint system MW = N by a suitable partitioning of matrices. Thus, let

$$\Omega = \begin{bmatrix} \Omega_1 \\ \Omega_2 \end{bmatrix}; \qquad \Lambda = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{12}' & \Lambda_{22} \end{bmatrix};$$

$$M = \begin{bmatrix} M_1^0 & 0 \\ 0 & M_2^0 \\ M_1^c & -M_2^c \end{bmatrix}; \qquad N = \begin{bmatrix} N^0 \\ N^0 \\ 0 \end{bmatrix};$$

where subscripts index the samples, superscript "0" refers to control items,and superscript "c" refers to composite items.

The solution of the minimization (1) is given by the following generic formula, with appropriate substitutions for the matrices Ω, W, Λ, M, and N as above for the two sample control/composition case:

$$(2) \qquad \hat{W} = \Omega + \Lambda M'(M\Lambda M')^{-1}(N - M\Omega).$$

It is shown in Zieschang (1985a) that when the elements of Λ_{ij}, i,j=1,2 are the known weight co-MSE's of units appearing in the samples, equation (2) is the outcome of the sample version of a heuristic procedure that that reduces the mean square error of sample estimates of all totals produced with the adjusted weights.

Let Λ be the weight MSE matrix, where $\Lambda_{ij} = E(\xi_i - \Omega_i)(\xi_j - \Omega_j)'$, i,j = 1,2, and ξ_i is an n_ix1 vector of ones. Each element of the weight MSE matrix Λ is given by

$$\begin{aligned} \Lambda_{ijk\ell} = 1 &- P(\text{unit } k \text{ appears in sample } i) \cdot \Omega_{ik} \\ &- P(\text{unit } \ell \text{ appears in sample } j) \cdot \Omega_{j\ell} \\ &+ P(\text{unit } k \text{ appears in sample } i, \text{ and unit } \ell \text{ appears in sample } j) \cdot \Omega_{ik}\,\Omega_{j\ell}. \end{aligned}$$

If Λ is not known beforehand, equation (2) underlies an operational procedure when Λ can be accurately estimated. To this end, consider the following assumptions

(A1) Ω_i, i=1,2 is unbiased; that is,

$$\Omega_{ik}^{-1} = P(\text{unit } k \text{ appears in sample } i);$$

(A2) Sampling is independent; that is,

$$P(\text{unit } k \text{ appears in sample } i \mid \text{unit } \ell \text{ appears in sample } j) =$$

$$P(\text{unit } k \text{ appears in sample } i),$$

so that

$$P(\text{unit } k \text{ appears in sample } i, \text{ and unit } \ell \text{ appears in sample } j)$$

$$= (\Omega_{ik}\Omega_{j\ell})^{-1}.$$

Under (A1) and (A2), the elements of Λ_{ii} are

$$\Lambda_{iik\ell} = 1 - \Omega_{ik}^{-1}\ \Omega_{ik} - \Omega_{i\ell}^{-1}\Omega_{i\ell}$$
$$+ (\delta_{k\ell}(\Omega_{ik}\Omega_{i\ell})^{-1} + (1-\delta_{k\ell})\Omega_{ik}^{-1})$$
$$\cdot\ \Omega_{ik}\Omega_{i\ell}$$
$$= 0 \quad \text{if} \quad k \neq \ell$$
$$= \Omega_{ik} - 1 \text{ if } k=\ell,$$

where $\delta_{k\ell} = 1$ if $k\neq\ell$ and $\delta_{k\ell} = 0$ otherwise. Hence Λ_{ii} = diag $(\Omega_i - \xi_i)$ where ξ_i is a vector of ones, and $\Lambda_{ij} = 0$ by independent sampling (A2), which generates a particularly simple and observable form for Λ in the weighting problem (2). This derivation is presented in greater detail in section B.1 of the available Appendix. It is noteworthy that when Ω is large, as in the CE context, this form is numerically very close to that of Luery and Roman, which is Λ_{ii} = diag Ω_i in the single sample context they consider.

While (A1) might to be granted as a reasonable prior about the sampling process, (A2) may generate some argument. In particular, the CE context is characterized by cluster sampling within primary statistical unit (PSU) or city/area. Hence within a survey, sample units within clusters will not be characterized well by (A2). However, in the CE few clusters have more than one CU, suggesting that this effect is likely to be quantitatively small. This research ignores these cluster effects. Also ignored are the effects of errors in adjustments for noninterview status, which may be correlated across CU's, and the correlations introduced by the controlled selection of PSU's and systematic random sampling of clusters within PSU.

Another way (A2) can be violated is if multiple observations are taken on a given sample unit within a weighting/estimation time interval. In the CE context, if weighting is done by quarter, as will be proposed, there will be two observations on the majority of consumer units appearing in the Diary survey, since diaries are collected weekly for two weeks per unit. Accounting for this multiple observations per unit effect is critical to the composition section of the constraints of problem (1) and is dealt with explicitly in this research, as detailed in the available Appendix section B.2. Section B.3 of the available Appendix contains a description of the computational methods by which the resulting non-diagonality of Λ is handled.

All computations were performed for both the full sample and the twenty balanced replicate samples used for variance calculations. For all but two of the thirteen test quarters considered in the empirical section below accuracy was sufficient to equate the Diary and Interview estimates of total consumer units to five decimal places on typical values of sixty five to seventy five million. For 1981 and 1982 quarter three, the total number of CU's was estimated with a fractional error between surveys of less than one millionth of the total. As a check for one of these cases, the ratio of minimum to maximum eighenvalues of $M\Lambda M'$ for 1982 quarter three were computed. These never fell below 10^{-5} for any replicate sample or the full file, indicating reasonably well-conditioned cross-products matrices.

3.2 Bounding Extreme Weight Adjustments

The empirical experience with weights generated by the GLS procedure reported below indicated little propensity to generate extremely large weights or proportional weight adjustments relative to those produced by the current principal person procedure. However, occasional extreme downward adjustments did occur and are conceivably indicative of high variance estimates of derivative statistics for some set of small subdomains of the CE universe. There are two methods of dealing with this problem; one is expedient and easy to implement and the other elegant and rather involved.

The expedient method simply recodes the adjusted weight when it falls outside a tolerance interval containing the unadjusted weight. In this study the tolerance interval was set, with lower bound only, at one fourth of the unadjusted weight. Upper bounds were not enforced because the 'pre stage 2' weight becomes progressively biased downward as population growth occurs between sample selections. Secularly rising proportional adjustments are therefore reasonable and upper limits on tolerance regions of weights potentially risky. Setting a lower bound only will bias up estimates of totals produced with the recoded weights; however, evidence from the empirical trial below clearly demonstrated that this bias is extremely small for the twenty five percent lower bound used here in the CE context. Another consideration involves the computation of variances and the appropriate tolerance interval lower bound for proportional adjustment in the replicates, given that this is set at one fourth of the unadjusted weight in the full file. The replicate proportional tolerance should be "looser" than that of the full file to produce accurate variance estimates, though how much "looser" is not straightforwardly determined. However, comparison of recoded CV's of estimates of CU subdomain size and mean family income before tax with CV's generated by the unrecoded GLS weights indicated very little effect here either. The expedient method therefore stands as an attractive and viable near term alternative to the current procedures.

The elegant method augments the control/composition constraints of problem (1) with inequality constraints enforcing tolerance regions around the unadjusted weights. This transforms solving problem (1) from a matter of simple matrix algebra into a task requiring quadratic programming methods.

Implementing the elegant method appears by no means impossible. However, the development of mathematical programming solutions to bounding the adjusted weights also increases the need to determine how these bounds are to be set to compute variances accurately. It is very likely that workable solutions to this problem can be derived from within the GLS framework, using the 'design matrix' $R'M'(M\Lambda M')^{-1}MR$, where $RR'=\Lambda$. The diagonals of this matrix are estimates of the variances of the proportional adjustments implied by the GLS adjusted weights, conditional on the sample of CU's, and might be used to set tolerance bounds. In the interim, notable improvements over the current principal person procedures are available using GLS weighting with recoding, as the evidence in the next section indicates.

3.3 Time Interval Selection and Control Subdomains

To evaluate of the performance of the GLS procedure in the context of the Consumer Expenditure Surveys, the time interval for weighting batches of consumer units was set at one quarter. Current principal person procedures are implemented on a monthly basis; however, this results in weighting batches of consumer units of between 300 and 400 for the Diary survey, or for rotation groups in the Interview. The number of CU's drops to 150-200 for the sample replicates that are weighted in parallel with the full sample. Batches of this size tend to have patchy coverage of CU-member age/race/sex characteristics that are to be controlled to known Census/CPS counts. This is not surprising, since there are 48 control categories within which these at most 300-400 CUs' weights are being adjusted. Current procedure deals with this by a large amount of collapsing of control cells into one another, sacrificing control detail. The primary purpose for monthly weighting adjustment is to insure that the Diary months and Interview rotation group/months are correctly scaled relative to one another in the aggregate. This purpose should be served well enough by controlling the CU member weights in each month or rotation/month so that they add to total population for that month.

To provide better coverage of the control cells quarterly weighting is preferable to monthly because the weighting batch is three times as large. Under current principal person procedures, going to a quarterly time interval would require sacrificing the aggregate population controls on Diary months and Interview rotation/months within quarter. On the other hand, a quarterly GLS procedure for CU member control cells can accomodate these aggregate monthly and rotation controls simply as additional linear constraints to be satisfied. It is therefore able to take advantage of the better control cell coverage of large quarterly batches of data while retaining essential controls at the month and rotation/month level.

Even for quarterly batches of data, these coverage considerations lead to an aggregated set of controls for the Diary Survey via a reduction in the number of member age groups from the available twelve to the six used in CE publications. Certain of the 48 detailed age/race/sex cells were empty for the Diary in either the full file or replicate samples in certain quarters. The 48 controls were applied to the aggregate of all rotations and months for each quarter in the Interview survey, resulting in weighting batches of between 4200 and 4600, a size sufficient to preclude empty control cells in the full or replicate samples over the thirteen quarter period of data used for this study. The quarterly age/race/sex cells used in this study for the Diary and Interview Surveys are detailed in Table 1. In addition, monthly total population controls were added for the Diary survey and rotation group/month total population controls were added for the Interview.

Table 1. Member Control Categories[a]

AGE	Black Male		Black Female		Non-black Male		Non-black Female	
	I	D	I	D	I	D	I	D
14-17	1		13		25		37	
18-21	2	1	14	7	26	13	38	19
22-24	3		15		27		39	
25-29	4	2	16	8	28	14	40	20
30-34	5		17		29		41	
35-39	6	3	18	9	30	15	42	21
40-44	7		19		31		43	
45-49	8	4	20	10	32	16	44	22
50-54	9		21		33		45	
55-59	10	5	22	11	34	17	46	23
60-64	11		23		35		47	
65+	12	6	24	12	36	18	48	24

[a] I = Interview
D = Diary

3.4 Composite Subdomains

Subdomains chosen for the rows of the M^c matrices of equation (2) include region of CU residence, sampling frame from which CU was drawn, tenure status of the CU, and four family types. These groupings are described in Table 2. Region, tenure, and family type were chosen as composite subdomains because of their use in CE publications. Additional family type detail plus a set of CU size categories would exhaust the set of demographic subdomains on which tables are published for the CE. Augmenting the composite classification chosen above to include these additional groups should be a useful extension of the current study. The frame classifications were included because of prior knowledge that the frames from which CU's are drawn are the same size (since they are identical) between surveys. Also, improved composite estimates of numbers of CU's in each subframe should aid in planning future CE samples, and in assessing ongoing sampling performance and identifying problems in the management of the sampling process.

Table 2. Consumer Unit Composite Subdomains

REGION		
NEAST	=	Northeast region
NCENTRAL	=	North-central region
SOUTH	=	Southern region
WEST	=	Western region
SAMPLING FRAME		
CEN70	=	1970 Census frame
SPECPLAC	=	Special places frame
ARSEG	=	Area segments frame
NEWCON	=	New construction frame
TENURE		
OWNER	=	Owner consumer units
RENTER & STUDENT	=	Renter consumer units including those in student housing
FAMILY TYPE		
ALL HW	=	All husband/wife consumer units
SPT1+<18	=	Single parent consumer units
SINGLE	=	Single person consumer units
OTHER	=	All other consumer units

4. Some Empirical Results

4.1 Data

The GLS weighting procedure described in the previous sections was applied to the CE Diary and Interview data for thirteen quarters covering 1980 IV through 1983 IV. For details not covered here and extensive tables of quantitative results, see Zieschang (1985b, 1986). The urban samples only were weighted, since rural CU's were eliminated from the universe from 1981 IV through 1983 IV. Urban controls included age/race/sex (A/R/S) population totals computed by the Bureau of the Census from a) updated 1970 census data and b) urban/rural population distributions obtained from the Current Population Survey (CPS). The controls for the 48 A/R/S cells were available monthly over the period covered. These were summed across months for each quarter and across A/R/S types for each month to obtain the quarterly A/R/S and monthly total population controls to be used by the quarterly GLS procedure. The summed monthly total population controls were divided by three for the Diary and by twelve for the Interview so that the controlled weights would sum to average total population for the quarter, in the first case across the three months within the quarter; and in the second across the three months and fourrotations. The quarterly sums of the monthly A/R/S controls were divided by three to obtain average quarterly population in each of the 48 cells.

Other data files originated from the BLS CE database. The 'pre-stage 2' weight Ω to be adjusted was computed as the product of the baseweight, weighting control number, and monthly noninterview factor. The M matrix of counts of persons or indicators of subdomain membership was generated in an obvious way from the AGE, RACE, SEX, and other variables on the database.

4.2 Control Errors

To provide a check on the success of the GLS and GLS-recoded procedures in meeting the control objective of weighting, the percentage deviation of the adjusted weights from the control counts was computed. For comparison purposes, these deviations were computed for the currently used monthly principal person weights in the database as well. The GLS weights hit the controls exactly, the errors of GLS recode were negligible, and the errors of the existing principal person weights are often substantial.

4.3 Composition Errors

The success with which the composition objective was met on the composite subdomains was measured by the 'arc-discrepancy' between weighted CU totals from the Diary and Interview surveys. This is computed as twice the difference in the two estimates divided by their sum. Arc-discrepancies for the composite subdomains were computed for the thirteen test quarters. As with meeting the control objective, GLS achieved the composition objective, equating the surveys exactly, GLS-recode generated negligible discrepancies between surveys.

4.4 Effects on Survey Discrepancies in Non-Composite Subdomains

Arc-discrepancies were also computed for subdomains related to BLS publication categories but not composited during weighting adjustment. These included six age of CU head categories (AGE<25, 25-34, 35-44, 45-54, 55-64, and AGE>=65), a tenure category for non-student housing renter status (RENTER), and five CU size categories (TWO PER, THREE PER, FOUR PER, FIVE PER, and SIX+). Also included are five husband & wife family/age of oldest child classes (HW ONLY, HW:OLD<6, HW:6-17, HWOLD>18, HW:OTHER), two single person CU/employment status classes (SING:OER and SING:1ER), three multi-person CU/employment status classes (CU>2:OER, CU>2:1ER, and CU>2:3+), and an incomplete income response class (INC NR). In all subdomains, GLS and GLS-recode usually were better than principal person on a quarter by quarter basis. Discrepancies in the incomplete income response category were large because income response is markedly different between the surveys, the most likely explanation involving the difference in survey instruments.

4.5 Effects on the Magnitude of Estimates of Subdomain Size and Mean Family Income before Tax

Because GLS is proposed as an alternative to the current principal person methods, it is of some interest to know if there is a systematic or pervasive difference in estimates produced with the adjusted weights. To evaluate this, two types of variable were considered. The first was subdomain size, the estimated total for assorted subgroups of consumer units. The second was mean family income before tax (FIBT) for the same assortment of CU subgroups. The behavior of mean FIBT should be indicative of the behavior of other important survey variables, notably detailed expenditures, because mean FIBT and mean expenditure by product classification are highly correlated for the most part.

The Diary and Interview ratios of estimates of CU population generated by GLS recode and principal person weights were computed for both the composite and a set of noncomposite subdomains that were introduced in section 4.5. GLS displayed a tendency to estimate lower CU counts than principal person across time and across subpopulations for either survey. For the Diary survey GLS produces noticeably larger estimates of CU's in husband and wife CU's not elsewhere classified (HW:OTHER), and CU's with Interview survey GLS estimates are noticeably larger for the special places sampling frame. While most of the other subpopulations are estimated lower by GLS in either survey, reductions from principal person estimates were slight for the Interview survey by comparison with the Diary.

Finally, ratios of mean FIBT produced by GLS and principal person weights were computed for complete reporters on the income question. In general, the relationships between these estimates derived from GLS and principal person weighting are opposite to those for population size. There is a very small but pervasive tendency for GLS estimates of mean FIBT to be higher than those for principal person, an effect slightly more pronounced for the Diary survey compared with the Interview.

4.6 Precision of Estimates of Subdomain Size and Mean Family Income before Tax

Ratios of GLS with principal person estimates of the coefficients of the variance reduction performance of GLS and Principle Person weighting. The standard deviations used in computing the CV's were computed using twenty independently (GLS or principal person) weighted replicate samples to generte twenty replicate estimates for each cell in the tables. For each quarter and subdomain the standard deviations were computed as the square root of the average squared difference between the replicate estimates and that of the full sample. The improvements in precision obtained with the GLS weights were rather striking, particularly for the composite subdomains in the Diary survey. Within the rather lengthy test period, GLS weighting improved the precision of estimates of CU population in every one of the subdomains considered.

The same statistics for mean FIBT were computed. These statistics reveal that over time GLS estimates were more precise than principal person for a majority of the composite subdomains composite subdomains in either survey. For the selection of noncomposite subdomains, Diary GLS mean FIBT estimates were about even with principal person in precision, while the Interview estimates were generally noticeably better using GLS. The measurement of the precision of GLS versus principal person estimates of totals and means is only one of two components of the accuracy of the estimates as measured by mean square error. The relative CV's just discussed provide information on the variance component of MSE, but not the squared bias component. GLS weighting as specified here is unbiased if the sample design and execution are unbiased. A great deal of effort is expended by BLS and the Census Bureau to collect unbiased samples, and to the extent this effort is successful, the GLS estimates are probably more accurate than the principal person, even in the infrequent event they are slightly less precise. Of course to the extent that these efforts to achieve unbiasedness fail, this assertion loses force. However, even in this case, GLS provides a well-ordered tableau within which issues of survey bias can be examined.

5. Concluding Remarks and Extensions

The results of the empirical evaluation of a quarterly GLS weighting procedure for the CE surveys have been for the most part highly favorable to GLS when compared with the existing monthly basis principal person procedures. Not only did the operational, recode variant of GLS display a high degree of numerical consistency with the control totals and in equating composite CU totals, it also demonstrably improved the precision with which the sizes of both the composite subdomains and a broad selection of other, noncomposite subdomains were estimated. For the analytical variable family income before tax, improvements in precision were minor at the subdomain level, but were of notable magnitude at the all CU level. The coefficients of variation of the Diary and Interview GLS estimates of mean FIBT averaged, respectively, eight and twelve percent below the same statistics computed for the corresponding principal person estimates. To the extent that these gains are inherited by estimates of mean expenditures, it is reasonable to expect that GLS weighting can improve not only the quality of the data published by the Division of Consumer Expenditure Studies, but also the Consumer Price Index by providing more precise expenditure weights for producing that series.

In assessing this study, another comparison for evaluating GLS could have been made, this being with the quarterly basis principal person weights also produced by the current processing system. Though undoubtedly providing additional information, this comparison was judged not worth the time and expense because (i) the quarterly principal person weights make no pretense of hitting any control totals at the monthly level, an important objective for weighting set by BLS statisticians, and (ii) at least partly for this reason, the quarterly weighting data is not used in any BLS publication.

There remains a great deal of work to be done in the area of survey weighting/estimation. The methods considered here condition on the current noninterview adjustment methodology. The evidence in the CE suggests that this methodology is not fully adequite for the task of capturing noninterview status, because noninterview adjusted person counts fall short of the controls in a pervasive if differential pattern across person types and consumer unit size classes. A plausible explanation for this that some addresses, at which one or more households may reside, are misclassified as vacant and excluded from the computed noninterview adjustment. Indeed, the use of population control information could be motivated by the need to address this "address undercoverage" issue. Along these lines, since the consumer unit is a subdivision of the Census Bureau's household unit, additional useful control information would include the total number of households by household size. Data from the decennial censuses and the Current Population Survey are now being used by the Census Bureau (U.S. Department of Commerce, 1985) to compute postcensal counts of households, though on an annual rather than quarterly basis as needed by the CE.

A few areas of further methodological research within the GLS framework should be highlighted. The first is to examine this approach from a model-based point of view as well as the sampling perspective adopted in this paper. For example, one suitable model would be

$$\Omega = \Lambda M'\beta + \varepsilon$$

$$M\Lambda M'\beta = N$$

where $W = \Lambda M'\beta$. Parallel with the GLS specification examined in the foregoing sections, it is assumed that $E(\varepsilon) = 0$ and $E(\varepsilon\varepsilon') = \Lambda$. Additional properties of the adjusted weights to those established here should flow from correct determination of the underlying model, bringing to bear the substantial statistical and econometric literature on constrained regression analysis. Some generalizations are suggested by the simple model above. It is reasonable to expect that Ω may be determined by regressors additional to those on which control or composition data are available. The longitudinal aspects of a survey can be handled through the covariance terms in the Λ matrix, as was done for the Diary survey in the empirical section of this study. A possibly less important technical matter is to determine efficient algorithms for solving the quadratic programming problem for bounding weight adjustments, and equally, the method of determining the bounds enforced.

Another area of useful inquiry is in the effects of measurement error in some or all rows of the M matrix. For example, it may be that for some households not all members are enumerated, causing errors in the rows of M corresponding to population control counts. On the other hand, by the luck of the draw, or a combination of this and "person undercoverage", the situation could occur when weighting cells are sufficiently narrowly defined that in a given sample there may be no persons at all corresponding to some control counts. Both of these problems suggest that "loose" rather than "exact" weighting control methods may be desirable in using the GLS principle in practice. "Loose control" versions of GLS can be defined, and are the subject of some current research within BLS. One advantage of such a "loose control" specification is that it effectively permits collapsing across control cells without the need for arbitrary "adjacency" and "sufficiency" criteria, as in currently used methods. Cells would be blended according to information on the covariance and relative bias (if any) contained in the specification of the weight co-MSE matrices in the GLS objective function.

Finally, this analysis can be extended to consider the composite estimation of population total expenditures as well as counts by simply incorporating these quantities into N_i^+ and M_i^+, $i = 1,2$, as additional rows. Additional issues arise in this context, such as the fineness of the aggregation of expenditures to be composited relative to the size of the computational problem to be handled. More fundamentally, reporting of these "analytic" quantities, as opposed to the more "enumerative" ones such as age, race, and sex of family members, is subject to non-negligible item nonresponse in addition to the unit nonresponse compensated for by the noninterview adjustment to the base weights. This introduces a bias not accounted for in the weight MSE matrix Λ of the quadratic weighting adjustment procedure considered here.

ACKNOWLEDGEMENTS

Significant contributions to this research were made by a number of people. Helpful and illuminating exchanges with Stephen Fienberg, Morris Hansen, Joseph Waksberg, Charles Alexander, Ben Tepping, James Chromy, and Fritz Scheuren have aided my own understanding of the statistical context within which weighting adjustment in general and the GLS procedure in particular can be placed. The advice, criticism, and guidance of Wesley Schaible, Richard Valliant, and Stewart Scott of the BLS Office of Research and Evaluation were invaluable. Curtis Jacobs and Paul Hsen of the BLS Division of Price Statistical Methods provided a wealth of information on Consumer Expenditure Survey procedures. Special thanks are due to Robert Gillingham and John Greenlees for their advice, support, and encouragement during the early stages of this project and for identifying the survey estimation problem GLS was designed to solve. Anna Sanders expertly processed the manuscript. I would like, finally, to thank Fritz Scheuren for providing the peer review, without ascribing to him or the others mentioned above any responsibility for positions taken or errors remaining herein.

REFERENCES

Alexander, C. (1985), "Alternatives to the Generalized Least Squares and Principal Person Methods for Consumer Expenditures Weighting," Statistical Methods Division, Bureau of the Census.

Alexander, C., and J. Roebuck (1986), "Comparison of Alternative Methods for Household Estimation," presented at the American Statistical Association Summer meetings, Chicago, Illinois.

Alexander, C. (1986), "The Present Consumer Expenditure Surveys Weighting Method" Population Controls in Weighting Sample Units, Bureau of Labor Statistics.

Deming, W.E., and F.F. Stephan (1940), "On a Least Squares Adjustment of a Sample Frequency when the Expected Marginal Totals are Known," The Annals of Mathematical Statistics 11, 427-444.

Fagan, J., and B. Greenberg (1985), "Algorithms for Making Tables Additive: Raking, Maximum Likelihood, and Minimum Chi-Square," Bureau of the Census, Statistical Research Division Report Series, Census-SRD-RR-85/15.

Fienberg, S.E. (1986), "Comments on Some Estimation Problems in the Consumer Expenditure Surveys," Population Controls in Weighting Sample Units, Bureau of Labor Statistics.

Luery, D. (1980), "An Alternative to Principal Person Weighting," Statistical Methods Division, Bureau of the Census.

Oh, H.L., and F. Scheuren (1978), "Multivariate Raking Ratio Estimation in the 1973 Exact Match Study," Proceedings of the Section on Survey Research Methods, American Statistical Association, 716-722.

Pugh, R.E., B. Tyler, and S. George (1976), "Computer-Based Procedure for the N-Dimensional Adjustment of Data--NJUST," Staff Paper No. 24, Office of Research and Statistics, Social Security Administration.

Roman, Anthony (1982), "Proposal for Weighting Research in the Consumer Expenditure Surveys," internal memorandum, Demographic Surveys Division, U.S. Bureau of the Census.

Stephan, F.F. (1942), "An Iterative Method of Adjusting Sample Frequency Tables when Expected Marginal Totals are Known," The Annals of Mathematical Statistics 13, 166-178.

U.S. Department of Commerce, Bureau of the Census (1985), "Estimates of Households for States: 1981 to 1984," Population Estimates and Projections, Current Population Reports, Series P. 25, No. 974, August.

Zieschang, K.D. (1985a) "Consumer Expenditure Survey GLS Weighting Study; I. Weighting and Estimation in the Consumer Expenditure Surveys," Division of Price and Index Number Research, Office of Prices and Living Conditions, Bureau of Labor Statistics.

Zieschang, K.D. (1985b), "Consumer Expenditure Survey GLS Weighting Study; II. Empirical Evaluation of GLS and Principal Person Procedures," Division of Price and Index Number Research, Bureau of Labor Statistics.

Zieschang, K.D. (1986), "Generalized Least Squares: An Alternative to Principal Person Weighting," Population Controls in Weighting Sample Units, Bureau of Labor Statistics.