WEIGHTING AND IMPUTATION METHODS FOR NONRESPONSE IN CPS GROSS FLOWS ESTIMATION

Arthur L. Hughes and Flora K. Peitzmeier
Bureau of Labor Statistics
Arthur L. Hughes, 441 G St., N.W., Washington, D.C. 20212

KEY WORDS: Nonignorable nonresponse, hot deck, multiple imputation, weighting cell adjustment, simulation

1. INTRODUCTION

1.1 Background

The Current Population Survey (CPS) is a monthly household survey conducted by the Bureau of the Census for the Bureau of Labor Statistics (BLS). Approximately 59,000 occupied housing units are in sample each month. Monthly estimates of employed, unemployed, and not-in-labor-force for the total U.S. civilian noninstitutional age 16+ population are the principle characteristics of interest that are produced from the CPS.

The CPS uses a rotating sample of 8 panels (or rotation groups), each containing about 1/8 of the total monthly sample (Hanson 1978). Each housing unit within a rotation group is interviewed for four consecutive months, not interviewed for eight months, then interviewed for four more consecutive months, before being dropped from the sample (called the 4-8-4 rotation design).

Monthly estimates of levels based on cross-sectional (stock) data can provide information on net changes in the labor force between months, but cannot provide information on the volatility of the change. One measure of volatility is called gross change (or gross flows). In the CPS, month-to-month gross flows is defined to be the aggregate change in labor force status (LFS) of the same individuals from one month to the next. Due to the 4-8-4 rotation group pattern, a maximum of 75 percent of the monthly sample will be common in any two consecutive months. Thus, individuals used to produce gross flows are contained in the 75 percent overlap sample.

Conceptually, gross flows can be expressed as a decomposition of the month-to-month net change. For example, suppose in the current month, the stock data exhibited a net increase of one-half million in unemployment from the previous month. The gross flows estimates could indicate that this increase was a result of one million people entering unemployment (inflows) and one-half million people leaving unemployment (outflows); or two-and-a-half million people entering unemployment and two million people leaving it. Even though the net change is the same, each scenario could, for example, result in different policy decisions being made on the funding of unemployment compensation, and employment and training programs in the U.S.

1.2 Factors Affecting the Quality of Gross Flows

Historically, estimates of month-to-month net change based on gross flows tend to indicate larger movements in the labor force than the same change based on stock data, causing inconsistency between the two sets of estimates. The major reasons for the differences are given below.

Reduced sample size - Currently about 60 percent of the monthly sample are available for use in estimating gross flows. Twenty-five (25) percent of the monthly sample cannot be used in estimating gross flows due to the 4-8-4 rotation pattern. This is due to the fact that one-fourth of the sample in any given month are either being interviewed for the first time or reentering the sample again after an eight month absence (being interviewed for the fifth time).

Of the 75% of the sample that could be used in estimating gross flows, around 15% do not match over two consecutive months due to the availability of data for only one month (Abowd and Zellner, 1984). The reasons include (Russell 1986, Hilaski 1968): movers (about 2%); coding, classification, and interviewer transcription errors (about 6%); refusals and not-at-homes (about 5%); and other reasons (includes within household coverage errors and sampling materials not arriving in time for interviewing, about 2%).

Rotation group bias - Bailar (1975) showed that the two rotation groups that are in sample for the first or fifth time tend to report more unemployment than the other six rotation groups. Even though it is not clear which panel produces more accurate labor force information, the presence of rotation group bias may lead to overestimation of gross flows.

Response Error - Problems such as the conditioning of responses over time, or the incorrect interpretation of the questions contributes to response error which affects gross flows. Response errors may not cause a serious problem in the estimates of month-to-month net change based on stock data since they often cancel out. However, these errors are cumulative in gross flows data, resulting in artificial movements in the labor force (Hogue and Flaim, 1986).

1.3 Purpose of Study

A special conference on gross flows was held by the Bureau of the Census and the Bureau of Labor Statistics in July 1984. The conference resulted in proposals for procedures that adjust for response or nonresponse error associated with gross flows at the macrodata level (Census and BLS, 1984).

The purpose of this paper is to examine microdata weighting and imputation procedures that adjust for individuals that are missing labor force information in only one of the two months (sometimes known as partial nonrespondents or partially cross-classified nonrespondents).

Section 2 describes the current CPS gross flows weighting procedure. Section 3 presents a description of the four alternative nonresponse adjustment procedures. Section 4 presents the application of these procedures on CPS data. Section 5 presents a description of the estimators used to evaluate the adjustment procedures. Section 6 presents the results, and Section 7 provides a summary, recommendations, and suggestions for future work.

2. THE CURRENT CPS GROSS FLOWS WEIGHTING PROCEDURE

The data file used to produce gross flows is created by matching the common rotation groups from two consecutive months. The common rotation groups are composed of housing units belonging to month-in-sample (MIS) 2, 3, 4, 6, 7, and 8 from the current month and MIS 1, 2, 3, 5, 6, and 7 from the previous month.

The current gross flows weighting procedure consists of computing factors that adjust the tabulated data to male and female population controls for the current month. Each factor (F_s) is computed as: $F_s=(8/6)(A_s/B_s)$, where A_s=sum of the current month's final weights of all records with sex=s in MIS 2, 3, 4, 6, 7, & 8 and B_s=sum of the current month's final weights of all **matched** records with sex=s in the same MIS (s=0 for females and 1 for males).

The final CPS weight (or second stage weight) is a cross-sectional weight that is equal to the product of the inverse of the probability of selection and several adjustment factors (Hanson 1978). These factors represent monthly adjustments for: subsampling of housing units, housing unit nonresponse, sampling of primary sampling units, and over or undercoverage in the population by age, sex, race, ethnicity, and state.

The factor F_s is comprised of two components. The first component (8/6) is required to account for the exclusion of the two rotation groups that are, by design, not common between consecutive months. The second component (A_s/B_s) is an adjustment for individuals belonging to the common rotation groups that responded in the current month but not the previous month. Individuals belonging to the common rotation groups that did not respond in the current month are accounted for in the nonresponse and post-stratification adjustments included in the final weight for the current month.

3. ALTERNATIVE WEIGHTING AND IMPUTATION PROCEDURES

3.1 Introduction

Four alternative procedures were evaluated based on a two month simulated file of respondents and partial nonrespondents. The imputation procedures considered were the carry-over, hot deck, and multiple imputation. The alternative weighting procedure considered was a weighting cell adjustment. All of the alternative procedures were performed within adjustment cells defined by the criteria presented in Section 3.2 below.

3.2 Formation of Adjustment Cells

Stasny (1987) indicated that the CPS nonrespondents exhibit a nonignorable missing data pattern (i.e., the labor force characteristics of the respondents are systematically different from the nonrespondents). In this study, adjustment cells were constructed such that ignorable nonresponse (or missing at random) could be assumed. Little and Rubin (1987) stated that, when using a weighting cell estimator, cells used to adjust for nonresponse should be formed such that 1) the probability of response is independent of the characteristic of interest within each adjustment cell (to reduce nonresponse bias) and 2) the mean square error of the weighting cell adjustment estimator is minimized (to control the within-cell variance).

3.3 Imputation Procedures

Carry-Over Imputation

Carry-over imputation only involves the use of data from the partial nonrespondents. For each partial nonrespondent, the LFS from the observed month is used to impute for the other month. No other adjustments are performed.

Hot Deck Imputation

Hot deck imputation uses information from respondents (those providing data for both months) to fill in the missing values for the partial nonrespondents. For each partial nonrespondent, the hot deck procedure was performed as follows:

o Select a respondent at random within each adjustment cell to serve as a donor for the partial nonrespondent's missing LFS.

o Impute the partial nonrespondent's missing LFS with the selected respondent's value.

Multiple Imputation

Rubin (1987) described an imputation procedure that, in practice, can be performed like the hot deck procedure, where the data produced can be used to estimate the uncertainty associated with imputing missing values. For each partial nonrespondent, the multiple imputation procedure was performed as follows:

o Select K(=3) respondents at random within each cell to serve as donors for the partial nonrespondent's missing LFS.

o Impute the partial nonrespondent's missing LFS with the K selected respondents' values, building K complete data sets.

The first data set consists of all of the respondents including partial nonrespondents with imputed LFS data received from donors selected in the first imputation process. The second and third data sets were formed in a similar fashion. The choice of K was determined somewhat arbitrarily, although it was decided to keep it small because of the large number of CPS records that are processed each month.

3.4 Weighting Cell Adjustment

In the weighting cell adjustment procedure, an estimate of the inverse of the probability of response is produced and used as the nonresponse adjustment factor (NRAF) for the respondents. The NRAF is then multiplied by the survey weight of each respondent to produce a new weight that accounts for both sampling and nonresponse. For the cth adjustment cell, the NRAF was calculated as

$$NRAF_c = \left[\sum_{i=1}^{n1c} W_{ic} + \sum_{j=1}^{n2c} W_{jc}\right] / \sum_{i=1}^{n1c} W_{ic}$$

where n1c=number of respondents, n2c=number of partial nonrespondents, and W_{ic} and W_{jc}=current month's final weight for the ith respondent and jth partial nonrespondent respectively, in adjustment cell c.

4. APPLICATION OF PROCEDURES IN THE CPS

4.1 Introduction and Description of Data Used

SAS programs were written to produce a simulated sample of respondents and partial nonrespondents and to apply the adjustment procedures presented in Sections 2 and 3 to this sample. Data used for this investigation came from a file containing November 1988 and December 1988 matched, matched interview and noninterview, and nonmatched records from the CPS. The match variables used were: household-ID number, line number (identifies a person within a household), age($\pm$ 1 year), sex, and race. Civilian persons belonging to both the November and December 1988 CPS sample, providing data for at least one month, and at least 16 years old in both months were considered as in scope for this study.

4.2 Formation of Adjustment Cells

As stated in Section 3.2, nonresponse in the CPS exhibits a nonignorable missing data pattern. To illustrate this, weighted estimates of the unemployment-to-population (UP) ratio for both respondents and partial nonrespondents are presented in Tables 4.1 and 4.2 below.

Table 4.1 Estimates of the UP Ratio for Respondents and Partial Nonrespondents (%)

	November		December	
	Resp	Dec-NR	Resp	Nov-NR
Total	3.3	4.6	3.1	4.6
Race				
Black	6.7	7.5	6.2	7.7
NonBlk	2.9	4.1	2.8	4.1
Sex				
Male	3.7	5.0	3.8	4.9
Female	2.9	4.2	2.5	4.4
Age				
16-19	7.3	8.5	6.6	9.4
20-24	4.9	5.7	4.4	6.5
25-34	2.8	4.4	3.0	4.0
35-54	2.8	4.7	2.8	3.2
55+	0.9	0.9	1.0	0.8
Education				
<HS	4.2	7.0	4.3	5.7
=HS	3.7	4.8	3.5	6.0
>HS	2.4	2.8	2.1	2.8
Martial status				
MSP	2.2	3.5	2.2	3.1
MSA/Sep	6.7	6.8	5.4	4.9
Wid/Div	2.6	4.4	2.5	3.6
Nev Mar	5.9	6.0	5.5	7.5

HS=high school, MSP=married-spouse-present, and MSA=married-spouse-absent.

Table 4.2 Estimates of the UP Ratio for November Respondents by Type of Response in December (%)

November	December	UP	n
Response	Response	3.1	73424
Response	Type-A	5.0	687
Response	Type-B,C	7.0	951
Response	D-nonmatch	3.7	2801
Response	R-nonmatch	4.5	2018

Type-A=eligible nonrespondents (refusals and not-at-homes), Type-B,C=vacant or demolished housing units, D-nonmatch= demographic nonmatch (age, sex, or race variables did not match), R-nonmatch= residual nonmatch (all other nonmatches), and n=sample size.

The data in Tables 4.1 and 4.2 show consistently higher UP ratios for the partial nonrespondents across almost all of the categories presented. In particular, note that the UP ratios for the Type-B,C partial nonrespondents in Table 4.2 are about twice that of those responding in both months. Hogue and Flaim (1986) showed monthly UP ratios from January 1978 to December 1980 for the total population that were also consistently higher for partial nonrespondents.

The major step in forming the adjustment cells involve satisfying the first criterion given in Section 3.2. Another way of stating this criterion is as follows: let R=1 if the individual is a respondent, and 0 otherwise; and let Y= LFS (employed, unemployed, or not in the labor force). Then we would like for R to be independent of Y; however, this cannot be tested directly because Y is not known for the partial nonrespondents. Instead, we look for variables (say $Z_1, Z_2, \ldots$) that are correlated with Y and whose values do not change (or change for only a few individuals) in two consecutive months. Once these variables are found, the hypothesis H_0: R is independent of Y given Z is tested. Table 4.3 show correlation estimates between November LFS and selected CPS variables for the respondents.

Table 4.3 Correlation Between November LFS and Selected Variables (n=73424)

Variable	ϕ
Dec LFS	.75
Race	.06
Sex	.18
Age	.36
Education	.21
Marital status	.13

The statistic ϕ, ($0 \leq \phi \leq 1$), is Cramer's measure of association (Marascuilo and McSweeney). The estimates given in Tables 4.3 and 4.4 (below) are based on unweighted responses in both months and were not adjusted by design effects.

As expected, LFS is highly correlated from one month to the next, and therefore should be a good candidate for defining the adjustment cells. To see what other variables could be considered in the adjustment cell definition, respondents were grouped into six cells according to their November or December LFS. Within each of these cells, correlations were computed between the same Z variables and LFS for the other month (Table 4.4).

Table 4.4 Correlation Between a Particular Month's LFS and Selected Variables Given the LFS for the Other Month

	November LFS			December LFS		
Variable	**E**	**U**	**N**	**E**	**U**	**N**
Race	.02	.09	.05	.02	.10	.06
Sex	.06	.12	.04	.05	.11	.03
Age	.10	.11	.15	.11	.10	.15
Ed	.06	.08	.04	.06	.06	.05
Marst	.05	.08	.11	.06	.07	.10

E=employed, U=unemployed, N=not in labor force, Ed=education, Marst=marital status.

For example, given November LFS=U, the correlation between December LFS and age is .11. While all of the correlations are relatively weak, the age variable exhibited the strongest correlation across the six LFS cells, especially for the two not-in-labor-force cells. Based on the results shown in Tables 4.3 and 4.4, it was decided to test the hypothesis H_0: response status and age are independent, within each of the six LFS cells. P-values (p) and ϕs based on the chi-square statistic (X^2) are shown below.

Table 4.5 Test of Independence Between Response Status and Age

	November			December		
LFS	p	ϕ	n	p	ϕ	n
E	.00	.05	49995	.00	.04	50025
U	.36	.04	2604	.02	.07	2515
N	.00	.06	27282	.00	.06	27783

The data in Table 4.5 are based on unweighted cross-tabulations of response status (response, partial nonresponse) by age (16-19, 20-24, 25-34, 35-54, 55+) within each of the six LFS cells shown in Table 4.4.

At the 5 percent level of significance, H_0 is not rejected for the November unemployed data (with 4 df, H_0 is rejected in any LFS cell if $X^2 > 9.49$). However, the large sample size in each of the six LFS groups may lead to misleading conclusions concerning any practical association between response status and age. If it is assumed that the data arise from a normal distribution, then zero correlation implies independence. Upon examination of the ϕ coefficients, it was concluded that, in all six cells, response status and age are independent; hence, nonresponse may be considered as ignorable.

Variables such as sex and race exhibited correlations in the U cells that are similar to those for age. However, the inclusion of these variables in the adjustment cell definition would significantly reduce the cell respondent sample size (<30 in some of the U cells) and the second criterion in Section 3.2 would not be satisfied. Therefore, thirty adjustment cells based on LFS and age were defined and are contained in Table A.

4.3 Selecting a Sample of Simulated Partial Nonrespondents

Simulated partial nonrespondents (total of 12,205) were selected from 73,424 respondents using systematic random sampling. The sampling fraction, f_c=n2c/(n1c+n2c), was defined to be the estimated partial nonresponse rate, where n1c=number of original respondents and n2c=number of original partial nonrespondents in adjustment cell c. The expected number of simulated partial nonrespondents in adjustment cell c was defined as $n2c^s$=(n1c)(f_c). For example, in adjustment cell 3 (Dec LFS=U and age=16-19), the values of n1c, n2c, f_c, and $n2c^s$ are 360, 54, 0.13, and 46.8 respectively. Forty-seven (47) out of 360 original respondents were selected and designated as partial nonrespondents, leaving 313 respondents in the simulated sample in adjustment cell 3.

Adjustment cells 1-15 contain all of the November nonrespondents while cells 16-30 contain all of the December nonrespondents. The partial nonrespondents appear in one and only one cell; however, there is a 100 percent overlap between respondents in cells 1-15 and respondents in cells 16-30 (Table A).

4.4 The Weighting Procedures

With the exclusion of the factor 8/6, the following values were generated using the current weighting procedure: F_0=1.0946, F_1=1.0925. Adjustment cells 1 through 15 were used for the weighting cell procedure described in Section 3.4. The NRAFs ranged from 1.0712 in cell 13, to 1.1907 in cell 6. The average NRAF=1.1117 (the factor 8/6 was not included in the NRAFs).

5. GROSS FLOWS ESTIMATION PROCEDURE

5.1 Estimates

Gross flows estimates were tabulated using observed and imputed labor force data from the simulated sample. Gross flows were also tabulated from the original set of respondents (n=73424) and served as a benchmark for comparison purposes. The December final weight was used for all respondents and partial nonrespondents to generate estimates based on the imputation procedures. These weights were adjusted by the appropriate factor and assigned to only the respondents when a weighting procedure was used. Of course, when imputation is performed in practice, these weights will not be available for the current month's partial nonrespondents. It was done here so that the results would not be confounded by the use of weights from different months or by some other procedure for determining the weights. For the current and weighting cell procedures, the gross flows estimator used was

$$F_{ij}^p = \sum_c \sum_r (W_{cr}\ NRAF_c^p) I_{ijrc}, \text{ where}$$

F_{ij}^p=gross flows estimate based on adjustment procedure p for the ith and jth LFS (i and j=employed, unemployed, or not-in-labor-force in November and December, respectively); I_{ijrc}=1 if the rth respondent in adjustment cell c belongs in LFS i and j, and 0 otherwise; W_{cr}=the December final weight for the rth respondent in adjustment cell c; and $NRAF_c^p$=the nonresponse adjustment factor in adjustment cell c for procedure p.

The above estimator was used for the carry-over and hot deck imputation procedures, where $NRAF_c^p$=1.0; I_{ijrc}=1 for both respondents and imputed partial nonrespondents with LFS i and j, and 0 otherwise; and W_{cr}=December final weight for both respondents and imputed partial nonrespondents.

For multiple imputation with K=3, the gross flows estimator used was

$$F_{ij}^p = \sum_{h=1}^{3} (1/3) F_{ijh}^p, \text{ where}$$

$$F_{ijh}^p = \sum_s (W_s) I_{ijhs};$$

I_{ijhs}=1 if the sth sample person (respondents and imputed partial nonrespondents) in imputation h belongs in LFS i and j, and 0 otherwise; and W_s=December final weight for sample person s. F_{ijh}^p=gross flows estimator based on the hth imputation.

For both the weighting and imputation procedures, the percent distribution of gross flows was tabulated using the estimator:

$$P_{ij}^p = [100 F_{ij}^p] / \sum_{ij}\sum F_{ij}^p, \text{ where}$$

$\sum_{ij}\sum F_{ij}^p$ = 126,668,865 for all procedures.

5.2 Estimates of Sampling Error and Bias

Measures of sampling variability were calculated using a random group estimator, with the six (6) sample rotation groups in common in both months as random groups. Using the rotation groups as random groups for estimating variances in the CPS has certain weaknesses, such as the presence of rotation group bias. However, the lack of design information on the public use data file prevents the development of more appropriate variance estimators by the BLS. Estimates of nonresponse bias were determined by assuming that the original matched sample has a 100 percent response rate.

5.2.1 Sampling Error Estimates

The form of the variance estimator is (Wolter 1985)

$$v(F_{ij}^p) = \sum_{\alpha=1}^{6} (1/30)(6F_{ij\alpha}^p - F_{ij}^p)^2,$$

$$\text{where} \quad F_{ij}^p = \sum_{\alpha=1}^{6} F_{ij\alpha}^p \text{ and}$$

$F_{ij\alpha}^p$=gross flows estimator based on adjustment procedure p for persons in LFS i and j belonging in the αth random group. Note: the term 1/30=1/[6(6-1)] for 6 random groups.

For the multiple imputation procedure, the gross flows variance estimator used was

$$\sum_{h=1}^{3}(1/3)\left[\sum_{\alpha=1}^{6}(1/30)(6F_{ijh\alpha}^{P}-F_{ijh}^{P})^{2}\right]$$

$$+\sum_{h=1}^{3}[1/(3-1)](F_{ijh}^{P}-F_{ij}^{P})^{2},$$

where $F_{ijh\alpha}^{P}$=gross flows estimate based on the hth imputation for persons in LFS i and j belonging in random group α, and F_{ijh}^{P}=gross flows estimator based on the hth imputation. The first term is the within-imputation variance; the second term is the between-imputation variance.

For all procedures, the percent relative standard error (Table C) was calculated using the estimator:

$$RSE(F_{ij}^{P}) = 100[v(F_{ij}^{P})]^{1/2}/F_{ij}^{P}.$$

5.2.2 Estimates of Bias

To measure the nonresponse bias, the percent relative bias (Table B) of the estimates was calculated using the estimator:

$RB(F_{ij}^{P}) = 100(F_{ij}^{P}-F_{ij})/F_{ij}$, where

F_{ij}^{P} is defined as before and F_{ij}=gross flows for persons in the original matched sample belonging to LFS i and j.

6. RESULTS

6.1 Estimates of Bias

As shown in Table B, the carry-over procedure clearly underestimates the inflows and outflows (by as much as 20.9%) and overstates the number of persons that do not change LFS (the UU flows are overstated by 18.0%). Estimates based on multiple imputation have generally the same or lower relative bias across all gross flows categories; compared to the current procedure, the average absolute relative bias is one-fourth as large. An explanation for this can be seen when the relative bias for each of the three imputations were examined. Here, the relative bias estimates for a given flow category are either all small or mixed with both large and small values; so when the average was taken, the bias nearly canceled out.

The weighting cell procedure also did quite well, reducing the average absolute relative bias by one-half over the current procedure. With the exception of the UE and NU flows, the hot deck also performed well compared to the current procedure.

6.2 Estimates of Sampling Error

Table C presents the relative standard errors (RSEs) of the gross flows estimates by adjustment procedure. No procedure appeared to do a good job in estimating the OR RSEs for the EU and UU flows. The multiple imputation RSEs are either about the same or somewhat higher than the RSEs from the other nonresponse procedures. A between imputation variance component was included in the multiple imputation variance estimator to represent variability due to nonresponse. This component (expressed as a percent of total variance) was about 12% for UE and NN, 9% for NU, 4% for EU, and <2% for the other flows.

7. SUMMARY, RECOMMENDATION, AND FUTURE WORK

7.1 Summary/Recommendation

An evaluation of five (5) microdata weighting and imputation procedures that adjusted for partial nonresponse in CPS gross flows estimation was conducted based on a simulation study. The results indicate that the multiple imputation procedure (with K=3) produced estimates with less bias. In addition, the RSEs were generally higher due to the inclusion of a between-imputation variance component that accounted for nonresponse variability.

Based on the data and the procedures used in this investigation, the authors recommend multiple imputation as the procedure to adjust for partial nonresponse in gross flows estimation. However, realizing the complexities involved in using several data sets to produce gross flows based on the multiple imputation procedure, the weighting cell adjustment is recommended as the interim replacement of the current procedure. Compared with the current procedure, the weighting cell procedure has an average absolute relative bias that is about one-half as large. In addition, the similarities between the current and the weighting cell procedures should make implementation an easier task.

Further work remains before a viable gross flows estimation procedure can be put in place. Some of the effort required to produce viable gross flows is discussed below.

7.2 Recommendations for Further Work

Along with the suggested recommendations for further work presented below, the method of maximum likelihood estimation in contingency tables with partial nonresponse as discussed by Little and Rubin (1987), and Stasny (1987) should be evaluated and compared to the multiple imputation and weighting cell procedures.

<u>Determine gross flows weights</u>

The current (previous) month's final weight for current (previous) month's nonrespondents are not available in practice because they are cross-sectional weights determined for only the respondents. A gross flows weighting procedure should be developed such that the weights adequately reflect the amount of nonresponse and population change that occurs in <u>both</u> the current and previous months.

<u>Develop a nonresponse adjustment procedure for eligible persons not responding in both months.</u>

By using the final weights, the procedures presented in this study indirectly accounted for individuals that did not respond in two consecutive months. However, this method is probably not optimal and alternatives should be investigated for these types of persons.

<u>Look at the consistency of bias and variability across time.</u>

In this investigation, only two consecutive months of data were used. Gross flows data covering an entire calendar year (or even an entire business cycle) should be examined in order to assess the impact of seasonal and cyclical behavior on the nonresponse adjustment procedures.

<u>Investigate the behavior of procedures by demographic characteristics</u>

In addition to the 3x3 tables showing aggregate labor force movements over two time periods, gross flows are produced by age, sex, race, industry, occupation, and other characteristics. The impact on the proposed partial nonresponse adjustment procedures on these tables should be examined.

Examine procedures that adjust for response error

Response error plays a major role in the quality of gross flows estimates. Response error adjustment methodologies suggested by participants of the Conference on Gross Flows in Labor Force Statistics (Census and BLS, 1984) and others should be investigated.

Examine procedures to reduce sampling error

Obtaining previous month's information directly from sample belonging to MIS 1 and 5 (retrospective interviewing) will significantly reduce the 25% sample loss, and would allow information on previous month nonrespondents and "new movers in" to be incorporated into the gross flows estimates. The drawback to this approach is the increased response burden and the potential for significant extended recall bias.

Look at defining adjustment cells given source of missing data

Unemployment-to-population ratios by record type differ markedly between the respondents and type B,C partial nonrespondents (Table 4.2). Work should be done to identify the characteristics of all of the type B,C partial nonrespondents in order to determine if modifications to adjustment cell definitions are necessary.

ACKNOWLEDGEMENT

The authors would like to thank Kennon R. Copeland, Sandra A. West, and Adnan M. Haider for their useful comments and suggestions on earlier versions of this paper.

REFERENCES

Abowd, John M. and Zellner, Arnold. (1984), "Application of Adjustment Techniques to U.S. Gross Flow Data." *Proceedings of the Conference on Gross Flows in Labor Force Statistics*, pp. 45-61.

Bailar, Barbara A. (1975), "The Effects of Rotation Group Bias on Estimates for Panel Surveys." *Journal of the American Statistical Association*, Vol. 70, No. 349, pp. 23-30.

Hanson, Robert H. (1978), "The Current Population Survey: Design and Methodology, Technical Paper 40." U.S. Bureau of the Census.

Hilaski, Harvey J. (1968), "The Status of Research on Gross Changes in the Labor Force." *Employment and Earnings*, 15, pp. 6-13.

Hogue, Carma R. and Flaim, Paul O. (1986), "Measuring Gross Flows in the Labor Force: An Overview of a Special Conference." *Journal of Business & Economic Statistics*, Vol. 4, No. 1, pp. 111-121.

Little, Roderick J.A. and Rubin, Donald B. (1987), *Statistical Analysis with Missing Data*. John Wiley and Sons.

Marascuilo, Leonard A. and McSweeney, Maryellen, *Nonparametric and Distribution-free Methods for the Social Sciences*. Kinko's Publishing Group.

Rubin, Donald B. (1987), *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons.

Russell, B. Gregory (March 4, 1986), "Special Study of Causes Contributing to CPS Month-to-Month Nonmatches." U.S. Bureau of the Census memorandum to Richard W. Swartz.

Stasny, Elizabeth A. (1987), "Some Markov-Chain Models for Nonresponse in Estimating Gross Labor Force Flows." *Journal of Official Statistics*, Vol. 3, No. 4, pp. 359-373.

U.S. Bureau of the Census and U.S. Bureau of Labor Statistics (1984), *Proceedings of the Conference on Gross Flows in Labor Force Statistics*.

Wolter, Kirk M. (1985), *Introduction to Variance Estimation*. Springer-Verlag.

Table A Dist of the Simulated Sample by Adjustment Cell

Cell	Dec	Age	Resp	NNR	Cell	Nov	Age	Resp	DNR
1	N	16-19	2317	262	**16**	N	16-19	2324	241
2	E	16-19	2279	248	**17**	E	16-19	2237	236
3	U	16-19	313	47	**18**	U	16-19	348	43
4	N	20-24	1404	168	**19**	N	20-24	1323	167
5	E	20-24	4268	535	**20**	E	20-24	4283	534
6	U	20-24	315	56	**21**	U	20-24	381	56
7	N	25-34	2380	252	**22**	N	25-34	2304	267
8	E	25-34	11716	1208	**23**	E	25-34	11736	1144
9	U	25-34	528	80	**24**	U	25-34	584	75
10	N	35-54	3590	344	**25**	N	35-54	3490	313
11	E	35-54	17745	1561	**26**	E	35-54	17861	1410
12	U	35-54	618	70	**27**	U	35-54	602	69
13	N	55+	13857	1000	**28**	N	55+	13725	956
14	E	55+	5635	444	**29**	E	55+	5773	393
15	U	55+	171	13	**30**	U	55+	165	13
Total			67136	6288				67136	5917

Dec=December LFS, Nov=November LFS, Resp=responded in both months, DNR=responded in Nov but did not respond in Dec, NNR=responded in Dec but did not respond in Nov, E=employed, U=unemployed, and N=not in labor force.

Table B Rel Bias of Gross Flows(%)

Flows		Adjustment Procedure				
Nov	Dec	CR	CO	HD	MI	WC
E	E	-0.3	0.6	-0.1	0.0	0.0
E	U	-4.1	-18.8	-1.0	0.4	-0.3
E	N	0.1	-15.3	2.1	-0.4	0.4
U	E	-0.9	-16.2	5.4	2.1	0.0
U	U	-4.8	18.0	-0.8	0.8	-1.4
U	N	-3.8	-18.6	0.5	0.8	-2.9
N	E	-2.7	-17.7	-1.1	-0.3	-2.2
N	U	-6.5	-20.9	5.5	-0.4	-2.9
N	N	1.2	1.1	0.0	0.0	0.4
Average Abs. Rel. Bias		2.7	14.1	1.8	0.6	1.2

Table C Rel Std Error of Gross Flows(%)

Flows		Adjustment Procedure					
Nov	Dec	OR	CR	CO	HD	MI	WC
E	E	0.9	0.9	0.9	0.9	0.9	0.9
E	U	5.3	7.1	7.1	7.6	6.6	7.1
E	N	8.5	8.6	8.6	8.6	8.8	8.6
U	E	5.7	5.9	5.9	5.9	5.9	5.9
U	U	4.3	3.4	3.5	3.5	3.6	3.5
U	N	8.0	8.5	8.5	8.9	8.5	8.5
N	E	3.7	3.4	3.4	3.3	4.3	3.3
N	U	5.7	5.3	5.3	5.1	6.5	5.3
N	N	1.2	1.2	1.1	1.1	1.1	1.2
Average		4.8	4.9	4.9	5.0	5.1	4.9

OR=Original matched sample of respondents, CR=Current method, CO=Carry over, HD=Hot deck, MI=Multiple imputation, and WC=Weighting cell. The percent distribution of gross flows based on OR is as follows: EE=60.3, EU=0.8, EN=1.7, UE=0.8, UU=1.8, UN=0.8, NE=1.3, NU=0.6, and NN=31.9.