ESTIMATES OF THE ERRORS IN CLASSIFICATION IN THE LABOR FORCE SURVEY AND THEIR EFFECT ON THE REPORTED UNEMPLOYMENT RATE

Joseph L. Gastwirth, George Washington University and Michael D. Sinclair, Response Analysis
Michael D. Sinclair, Response Analysis, 1060 State Road, Princeton, NJ 08542 [1]

Key Words: Missclassification Rates, Unemployment Rates, Diagnostic Tests

I. Introduction

In a recent article, Poterba and Summers (1995) used the data from the Census Bureau's reinterview program to estimate the misclassification rates of the Current Population Survey (CPS) and assessed their impact on estimates of labor market transition rates. The estimated misclassification rates were based on the assumption that a particular reinterview method, reconciliation, yields the "truth." Similar studies based on this assumption have been conducted by Poterba and Summers (1986) and Abowd and Zellner (1985). Biemer and Forsman (1992), Forsman and Schreiner (1991) and the U.S. Census Bureau (1963) have questioned this assumption. The purpose of this paper is to provide estimates of the misclassification rates from response errors in all interviews and reinterviews and to explore their impact on the reported unemployment and labor force rates. Our approach is based on extending the Hui and Walter (1980) paradigm for estimating error rates of medical diagnostic tests to trinomial classifications.

Under certain assumptions, Hui and Walter (H&W) developed a method for estimating the error rates associated with a new diagnostic screening test using a confirmatory test that also has unknown, but lower error rates. By treating the reinterview as the confirmatory test and the original survey as the screening test, this methodology can be used to estimate the error rates in the original survey and the reinterview and the prevalence rates of the trait screened for. The H&W method requires two subpopulations which have different prevalence rates for the characteristic. While the two tests may have different error rates, the error rates for each test are assumed equal in the two subpopulations. Furthermore, this procedure assumes that the errors from the two tests conditioned on the subject's true status are independent.

The H&W method was developed to evaluate dichotomous test outcomes and was used to examine data on labor force participation in Sinclair and Gastwirth (1996). Here, we have extended the approach to account for three classifications, unemployed, employed and not in the labor force (NLF). The basic model is presented in section II. The reinterview program data to which the model will be fitted are described in section III. The resulting error rates are given in section IV along with the "adjusted" unemployment rates which account for the estimated classification errors.

II. The Data and the Model

The CPS nonreconciled reinterview data consists of trinomial responses from both the original survey and the nonreconciled reinterview. We obtained yearly data for 1981 through 1986 and 1988 through 1990 which we divided into two subpopulations by sex. This data for a given subpopulation and year, is summarized in a 3 X 3 table. Let n_{ygij} denote the observed frequency counts of persons in the table, indexed as follows:

- y denotes the year, y=1 to 9 (1981-1986, 1988 - 1990) and
- g denotes subpopulation membership, g= 1 for females and 2 for males and
- i denotes the subject's classification by the original survey, i=1 for unemployed, i=2 for employed and i=3 for NLF, and
- j denotes the same subject's classification by the nonreconciled reinterview, j=1, 2 and 3.

For the true prevalence rates, we denote π_{ygi} as the prevalence rate among persons in subpopulation g and year y, having labor force status i, i=1,2,3. Note that the true labor force status rate of NLF, π_{yg3} is equal to $(1-\pi_{yg1} - \pi_{yg2})$, and that the true unemployment rate in year y for subpopulation g is equal to π_{yg1} divided by $(\pi_{yg1} + \pi_{yg2})$.

For the classification rates, β_{ygrij} are defined as the probability that the data collection process, r, r=1 for the original survey and r=2 for the nonreconciled reinterview will classify a person in year, y from subpopulation, g to be in category i, i=1,2 and 3 when the true status of the individual is category j. For example, β_{11131} denotes the probability that a 1981 (y=1) female person (g=1) is classified by the original survey (r=1) as NLF (i=3) when her true status is unemployed (j=1). The classification rates can be divided into two groups corresponding to those associated with a correct classification and those associated with an erroneous classification. Note that for each y, g and

[1] The authors wish to thank Irv Schreiner of the Census Bureau for the data used in this research and helpful discussion. We also wish to thank John Thompson, Henry Woltman and Jon Clark of the Census Bureau for their many helpful comments. This research was supported in part by a grant from the National Science Foundation.

r, the probability survey method r, classifies a truly unemployed person in year y from subpopulation g correctly as unemployed, is equal to β_{ygr11} = $(1-\beta_{ygr21} - \beta_{ygr31})$. The corresponding probabilities for employed and NLF are respectively, β_{ygr22} = $(1-\beta_{ygr12}-\beta_{ygr32})$, and $\beta_{ygr33} = (1-\beta_{ygr13} - \beta_{ygr23})$. Hence, the correct classification rates are simply determined by the error rates. The total sample size for year y and subpopulation g in the 3 X 3 table is denoted by $n_{yg..}$

The above data model has 14 parameters (six error rates for original survey, r=1, six error rates for the nonreconciled reinterview, r=2, and two unique prevalence rates) for each subpopulation and year. On the other hand, the 3 X 3 table for a given year and subpopulation has only 8 independent frequencies, or degrees of freedom. As a result, the model is overparameterized and the number of parameters must be reduced for estimation purposes. To accomplish this, we adopt the H&W model as discussed in section III.

III. Application of the Data Model and the CPS Reinterview Program

The Census Bureau's Current Population Survey Reinterview Program is conducted approximately two weeks after the initial survey to measure response errors and to evaluate interviewer performance. The sample design for the reinterview consists of the self-weighting random sample of households (Levy and Lemeshow, 1980) among the selected interviewer assignments. The sample size consists of about 1/18 of the monthly CPS sample of 50,000 to 60,000 household interviews. Two reinterview procedures are conducted. One-fifth to one-fourth of the sample households receive a response-variance type reinterview survey. The response-variance technique attempts to replicate the original CPS survey conditions so that the error rates for the reinterview are equal to those in the original survey. The remaining three-fourths to four-fifths of the sample cases participate in response-bias study. In the response-bias study an initial nonreconciled reinterview is conducted as in the response-variance study. In addition, the reinterviewer reconciles disagreements between the original and the initial reinterview responses by further discussion with the respondent. Hence, in the response-bias study up to two reinterview responses may be obtained from each subject including a nonreconciled reinterview response and a reconciled reinterview response.

In the response bias study, the reinterviewer is instructed not to look at the original survey responses until the initial reinterview is completed. Forsman and Schreiner (1991) indicate that the reinterviewers may change the initial reinterview responses to match the original response. To support this argument, several authors including Biemer and Forsman (1992), Poterba and Summers (1986), Schreiner (1980), Bailar (1968) and the U.S. Census Bureau (1963) observed that the rate of disagreement between the original responses and the initial reinterview responses were greater in the nonreconciled sample. Sinclair (1994) and Sinclair and Gastwirth (1996) showed that these differences were statistically significant. As a result, the reconciliation process creates a correlation between the original and nonreconciled reinterview responses in the reconciled sample. Hence, we decided to limit our analysis to the original and nonreconciled reinterview data from the response-variance study sample. We will assume for the purposes of this study that in the response-variance study sample that the errors from the original survey and the nonreconciled reinterview conditioned on the respondent's true status are independent.

To analyze the labor force data using the H&W model, we assume that each of the classification rates are equal in the two subpopulations, males and females, i.e. $\beta_{y1rij} = \beta_{y2rij}$. These groups are well known to have different labor force participation rates. At this stage, we assume that the classification rates for the original survey and the nonreconciled reinterview may be different and that the classification rates may differ by year. With this reduction, for the two subpopulations, in a given year, we now have a total of 12 error rate parameters and four prevalence rates yielding 16 parameters. Since two 3 X 3 tables contain a total of 16 degrees of freedom, estimation is possible.

The full CPS estimates of unemployment rate are published regularly by the Bureau of Labor Statistics (BLS). Since the reinterview is a sub-sample of the full CPS sample, the original survey estimates of the unemployment rate from the reinterview sample will differ from the BLS published results. Data processing procedures are also used on the full sample CPS that are not applied to the reinterview data. Therefore, we have used the CPS reinterview data primarily to estimate the error rates in the original survey. Using these error rate estimates, we estimate corrected Bureau of Labor Statistics (BLS) unemployment rates, where the term corrected means that the reported values have been adjusted to account for response errors in the survey. Note, the corrected prevalence rates {and hence, the unemployment rate, $\pi_{y1}/(\pi_{y1} + \pi_{y2})$ can be computed using the relationship below, where Obs π_{yi} are the BLS reported prevalence rates.

$$\begin{bmatrix} \hat{\pi}_{y1} \\ \hat{\pi}_{y2} \end{bmatrix} = \begin{bmatrix} 1-\hat{\beta}_{121}-\hat{\beta}_{131}-\hat{\beta}_{113} & \hat{\beta}_{112}-\hat{\beta}_{113} \\ \hat{\beta}_{121}-\hat{\beta}_{123} & 1-\hat{\beta}_{112}-\hat{\beta}_{132}-\hat{\beta}_{123} \end{bmatrix}^{-1} \begin{bmatrix} Obs\ \pi_{y1} - \hat{\beta}_{113} \\ Obs\ \pi_{y2} - \hat{\beta}_{123} \end{bmatrix}$$

IV. Data Analysis and Results

As a first step in preparing our final estimates, we developed the parameter estimates for each of nine yearly data tables using the SAS NLIN procedure with the Gauss-

Newton weighted least squares method. Given that the reinterview procedures remained constant during the period, we decided to test the hypothesis that each of the error rates remained equal across the years studied, i.e. $\beta_{ygrij} = \beta_{y'grij}$ for all years $y \neq y'$. In conjunction with the basic assumption, that the error rates for males and females are equal, i.e. $\beta_{y1rij} = \beta_{y2rij}$, this implies, $\beta_{ygrij} = \beta_{y'g'rij}$ for all $y \neq y'$ and $g \neq g'$.

Using the two sets of results, we conducted a likelihood ratio test to test the assumption that each of the error rates were the same for all years, given the base assumption that females and males exhibited the same error rates. The likelihood ratio statistic, -2 log λ with 96 degrees of freedom (144 parameters in the full model less 48 parameters in the reduced model) yielded a value of 84.06 with a p-value of 0.8027. Hence, the data is consistent with the reduced model, enabling us to use the reduced model estimates and to simplify the notation. We will now use β_{rij} to denote β_{ygrij} for all g and y.

The estimated error rates for the original survey and for the nonreconciled reinterview in response-variance study sample are presented in Tables 1 and 2, respectively, with their estimated standard errors. The estimated reinterview error rates in Table 2 are presented for comparative purposes and are similar to corresponding error rates estimates for the original survey. This similarity in the original and nonreconciled reinterview error rates is as it should be given the design of the response-variance procedures. Note that the proportion of truly unemployed persons who are classified as unemployed, $(1-\beta_{121}-\beta_{131})$, is 0.8397.

Table 1 also presents the estimates of the original survey error rates as obtained by Poterba and Summers (1986) using reinterview data (combined for both sexes) for the first half of 1981. For comparative purposes we conducted an analysis of the 75% sample reconciled reinterview data for same 1981-1990 period under the assumption that the reconciled responses were error free. These error rates are presented in Table 1 to illustrate how the estimated error rates from our method using only the nonreconciled data differ from those that would result for the same period assuming the reconciled reinterview was perfect.

Table 3 presents the reported BLS unemployment rates among those in the labor force for males and females combined in comparison to the estimated corrected unemployment rates based on: 1) our error rate estimates, 2) Poterba and Summers (1986) error rates and 3) error rates assuming the reconciled reinterview is perfect. Note that using our estimated error rates, the standard error of the unemployment rate estimates is about 0.27% (.00268).

If the results in the Table 3 are sorted by the value of the BLS reported unemployment rate, an apparent trend is observed in the bias in the original CPS estimates. Figure 1 clearly shows that the reported values tend to over-estimate the actual unemployment rate of persons in the labor force in low unemployment years (1989, 1988 and 1990) and to under-estimate the unemployment rate in high unemployment years (1982-1983). Furthermore, our method yields a noticeable difference in the magnitude and sign in the reported unemployment rate due to the misclassification then yielded by the misclassification rates obtained by Poterba and Summers (1986) or derived from assuming the reconciled reinterview is perfect (1981-1990 data).

In the screening test literature (Gastwirth, 1987) the fraction of positive classifications which are correct, called the predictive value of a positive test, is known to vary directly with the prevalence of the characteristic. This is why quite accurate diagnostic tests can have unacceptably high misclassification rates when populations with a low prevalence of a disease are screened with them. The analog of this measure in our context is the proportion of individuals classified as unemployed who are truly unemployed. This proportion is given in the third column of Table 3. Even though the range of reported unemployment rates is fairly narrow, the same relationship is observed.

V. Economic Implications Associated with The Corrected Estimates

The results in Figure 1 show that all three methods for adjusting the unemployment rate for misclassification error indicate that the degree of bias in the reported rate *varies* over the business cycle. In contrast with the previously used methods, which indicate that throughout the range of the data in Table 3 the reported rate is an overestimate, our approach suggests that the bias in the survey estimates is small in years when the unemployment rate is between 5.5% and 7.5%. Using our estimated misclassification rates the reported unemployment rate appears to be unbiased when the true unemployment rate is around 6.3% and yields an underestimate when the true rate is above this level and an overestimate when the true rate is below it.

The underestimation bias becomes quite noticeable when unemployment reaches 9%, while the overestimation bias is meaningful when unemployment is about 4%. Projecting slightly out of this range, shows that the bias would be about -0.60% when the unemployment rate is around 10% and would be +0.40% when the true rate is 4%.

VI. Discussion

In this paper we have presented an alternative method for estimating the error rates in the CPS survey. Our study differs from prior work in that we have followed the H&W approach to estimate the error rates by assuming that males and females will have the same error rates and that the errors in the original survey are independent of those in the nonreconciled reinterview. While the errors could be somewhat correlated, the assumption of independence is standard in data analysis of this type, see Bailar (1968). As for the equal error rate assumption, several of the authors cited in this paper have noted minor to moderate differences in the error rates between males and females under the assumption that the reconciled reinterview is perfect. In general, we feel that the equal error rate assumption is less restrictive than the assumption that the reconciled reinterview is perfect. Furthermore, the fact that the estimated error rates presented by Poterba and Summers (1986) are similar to ours is reassuring. Nevertheless, further research is needed to develop reinterview procedures and analytical techniques to relax the restrictive assumptions currently required in the analysis of this data.

Finally, it should be emphasized that all the estimates adjusting for misclassification are still in the research phase and that the error rates are not yet estimated with sufficient accuracy to adjust the regular survey data. While a new questionnaire and new interviewing procedures were introduced as of January 1994 (Bureau of Labor Statistics, 1993), which might have a slight impact on the misclassification rates, our finding that the impact of these errors varies with the business cycle will still hold.

Bibliography

Abowd John and Arnold. Zellner, "Estimated Gross Labor Force Flows," *Journal of Economic and Business Statistics* 3 (1985), 253-283

Bailar, B. A., "Recent Research in Reinterview Procedures," *Journal of the American Statistical Association* 63 (1968), 41-63

Biemer, Paul.P. and Gosta Forsman, "On the Quality of Reinterview Data with Application to the Current Population Survey," *Journal of the American Statistical Association,* 87 (1992) 915-923

Biemer, Paul. P., R. M. Groves,, L. E. Lyberg, N. A. Mathiowetz, and S. Sudman, *Measurement Errors in Surveys* (New York: John Wiley & Sons, 1991).Bureau of Labor Statistics, "Overhauling the Population Survey," *Monthly Labor Review* Vol. 116 No. 9 (1993).

Forsman, G. and Irv Schreiner, "The Design and Analysis of Reinterview: An overview," In Paul Biemer, et al., Eds., *Measurement Errors in Surveys* (New York: John Wiley and Sons, 1991).

Hui, Sui L. and Steven D. Walter, "Estimating the Error Rates of Diagnostic Tests," *Biometrics,* 36, (1980) 167-171.

Levy, Paul, and Stanley Lemeshow, *Sampling for Health Professionals*, (California: Lifetime Learning Publications, 1980)

Poterba, James M. and Lawrence H. Summers, "Reporting Errors and Labor Market Dynamics" *Econometrica,* 54(6), (1986) 1319-1338.

Poterba, James M. and Lawrence H. Summers, "Unemployment Benefits and Labor Market Transitions: A Multinominal Logit Model with Errors in Classification", *The Review of Economics and Statistics, 77(2)*, (1995) 207-216.

Schreiner, Irv, "Reinterview results from the CPS Independent Reconciliation Experiment (second quarter 1978 through third quarter 1979)," unpublished Bureau of the Census memorandum May 7, 1980.

Sinclair, Michael D. "Evaluating Reinterview Survey Methods for Measuring Response Errors," Phd Dissertation, George Washington University, September 1994

Sinclair Michael D. and Joseph L. Gastwirth "On Procedures for Evaluating the Effectiveness of Reinterview Survey Methods: Application to Labor Force Data," To appear in *the Journal of the American Statistical Associati*on September 1996.

U.S. Bureau of the Census, *The Current Population Survey Reinterview Program: Some Notes and Discussion*, Technical Paper No. 6 (Washington, D.C.: U.S. Government Printing Office 1963)

U.S. Bureau of the Census, *Evaluating Censuses of Population and Housing,* Statistical Training Document #ISP-TR-5 (Washington, D.C.: U.S. Government Printing Office, 1985).

Vacek, Pamela M., "The Effect of Conditional Dependence on the Evaluation of Diagnostic Tests," *Biometrics* 41, (1985) 959-968.

Table 1- Original Survey Error Rate Estimates

Original Survey Estimates Based on 1981-1986 and 1988-1992 CPS Nonreconciled Reinterview Data						
Error Rate Parameter	Description		Estimated Value			Estimated Standard Error
	Classified as	True Status	Our Method	P &S (1986)	Recon - Reint Perfect	Our Method
β_{121}	Employed	Unemployed	0.0407	0.0378	0.0188	0.01892
β_{131}	NLF	Unemployed	0.1196	0.1146	0.0838	0.01463
β_{112}	Unemployed	Employed	0.0049	0.0054	0.0017	0.00124
β_{132}	NLF	Employed	0.0100	0.0172	0.0098	0.00154
β_{113}	Unemployed	NLF	0.0110	0.0064	0.0034	0.00155
β_{123}	Employed	NLF	0.0205	0.0116	0.0053	0.00247

Table 2- Nonreconciled Reinterview Error Rate Estimates

Reinterview Survey Estimates Based on 1981-1986 and 1988-1992 CPS Nonreconciled Reinterview Response-Variance Study Sample Data				
Error Rate Parameter	Description		Estimated Value	Estimated Standard Error
	Classified as	True Status		
β_{221}	Employed	Unemployed	0.0333	0.01772
β_{231}	NLF	Unemployed	0.1128	0.01360
β_{212}	Unemployed	Employed	0.0057	0.00135
β_{232}	NLF	Employed	0.0145	0.00160
β_{213}	Unemployed	NLF	0.0157	0.00171
β_{223}	Employed	NLF	0.0248	0.00238

Table 3 - Implications of the Error Rate Estimates

Year y	BLS Reported Unemployment Rate	Prob Unemp. Given Classified Unemp.	Corrected Estimated of BLS Reported Unemployment Rate			Difference in Reported vs. Corrected
			Our Method	Poterba and Summers (1986)	Reconciled Data (1981-1990) Perfect	Our Method
1990	*5.44%*	*.8135*	5.27%	4.64%	4.53%	*0.17%*
1989	*5.20%*	*.8052*	4.99%	4.36%	4.26%	*0.21%*
1988	*5.43%*	*.8113*	5.25%	4.62%	4.51%	*0.18%*
1986	6.89%	.8503	6.97%	6.34%	6.16%	-0.08%
1985	7.09%	.8531	7.20%	6.57%	6.38%	-0.11%
1984	7.41%	.8581	7.56%	6.94%	6.73%	-0.15%
1983	9.47%	.8894	9.99%	9.37%	9.06%	-0.52%
1982	9.54%	.8902	10.08%	9.46%	9.15%	-0.54%
1981	7.50%	.8581	7.66%	7.03%	6.82%	-0.16%

Figure 1 - A Comparison of the Bias in the BLS Reported Unemployment Rates As Computed Using Three Methods