NONRESPONSE ADJUSTMENT IN CERTAINTY STRATA FOR AN ESTABLISHMENT SURVEY

Kirk Mueller, George Stamas, and Shail Butani
Kirk Mueller, U.S. Bureau of Labor Statistics, Suite 4985, 2 Massachusetts Avenue NE, Washington, D.C. 20212

KEY WORDS: Hot deck, imputation, link relative, separate ratio

1. INTRODUCTION

The Current Employment Statistics survey is conducted by the U.S. Bureau of Labor Statistics and State Employment Security Agencies to produce monthly estimates of employment, hours and earnings by industry for the U.S., States, and areas. This establishment survey is currently undergoing a redesign. Sample design research indicates that a simple but well executed probability design could considerably reduce mean squared error compared to samples selected using the current realized sampling rates. What this research has not considered is that the realized sampling rates are not deviations from a more optimum design as much as they are the result of low participation rates when units are first solicited for the survey, particularly among units in the largest employment size classes. This research compares bias of estimators resulting from nonresponse adjustment using information available on these nonparticipants from administrative records from the State Unemployment Insurance (UI) programs for earlier months, with more traditional survey methods. In other words, unlike most other studies, in this research we do not assume nonrespondents to be missing at random. For this study, we can evaluate the effectiveness of various imputation procedures since responses for every month for every unit are available from the administrative records consisting of UI accounts.

The data used in this study are discussed in Section 2. This will include a brief discussion of the CES survey and our test population. Section 3 presents the methods used in the various imputation routines. Section 4 describes the evaluation criteria used to analyze the results. Section 5 contains our results and comparisons of the imputation methods. Conclusions from this paper are contained in Section 6.

2. DATA

CES Estimation and Imputation

CES uses a simple sample design based on six employment size classes and detailed industry strata. Separate estimates are calculated for estimating cells, which are combinations of the sampling strata, using the link relative estimator, which is basically a ratio estimator. The link relative is expressed as

$$\hat{E}_c = \frac{\sum e_{i,c}}{\sum e_{i,p}} \times \hat{E}_p$$ where $\hat{E}$=estimated employment, e=sample unit employment, c is current month, p is previous month, and the sum is across a matched sample. In this study, all responding units have employment in each month, so all responses are by definition matched.

The CES sample is selected from a frame constructed from records of the State unemployment insurance programs. The employment series are benchmarked or reconciled once a year to employment counts from the same administrative records. CES employment estimates are benchmarked by summing employment from UI administrative records to a CES estimating cell level. The link relative estimator is then applied to the new level to produce estimates of subsequent months. This re-anchors the series of estimates to a more recent month.

The current CES estimator implicitly imputes monthly changes in the responding sample to nonrespondents within the same strata, as the nonrespondents are left out of the ratio of current sample employment to previous month employment. The implicit assumption is that nonrespondents behave as do the respondents in the same industry size strata. The goal here is to utilize information available from administrative records to impute for establishments that have not responded to the survey.

Test population

Our test population is administrative records from State Unemployment Insurance Programs for employers with only one worksite, and employees in April 92 and March 94. Data were collected for 6 states for the largest size class (employment of 250 or greater). The data included employment for 55, 2-digit Standard Industrial Classification (SIC) codes. Other 2-digit industries were not included in the analysis due to 100 percent response rates in the certainty strata or it's presence in only one of the 6 states. SIC 79 was dropped from all analysis due to an extremely large error for the hot deck random selection which dominated higher levels of aggregation. However, it's failure in this SIC should be noted.

We identified from this population the employers that participate in the CES. Employers that do not participate in CES remain non-respondents in the simulations. Month-to-month ratios of employment are calculated for each strata based on respondents.

Although the certainty cell was chosen due to the large amount of employment present in these cells, they are typically characterized by a small number of units in the population. Efforts to estimate employment for a large nonresponding unit by looking at trends of other large units might be inappropriate because these units are often in competition. Particularly in a flat or declining market, one employer might expand at the expense of other employers who lose business. Borrowing information from rivals could send imputed values in the wrong direction in these cases.

Benchmarking the estimates and the rules for borrowing from administrative data.

The files available for this analysis covered 24 consecutive months (April 1992 through March 1994). Because of the limited time frame and because one of imputation methods under consideration requires an over the year change, estimates have been benchmarked to June 1993. This creates 15 months of historical data and leaves 9 months to impute for the non-respondents. For the first six month of estimates, administrative information is available through the benchmark month. This was designed to mimic the CES benchmark process, with the difference being that we used June as the benchmark month rather than March. In preparing July through December 1993 estimates, the latest available administrative data would be June 1993. For the January through March 1994 estimates, the latest available administrative data would be September 1993.

3. METHODS

Imputation methods

The issue we are confronting is complete nonresponse or nonparticipation. In CES as in many other establishment surveys, our frame provides information on many items of interest for the unit--industry, location, previous employment, wages, etc. In a sense, even a complete nonresponse becomes an item nonresponse problem in CES. Therefore, we can compare item nonresponse procedures, such as mean imputation, hot deck - random selection, and hot deck - nearest neighbor, that have become standard with alternative methods we propose. For a more complete discussion of these imputation methods see Kalton (Mean imputation within classes, Random imputation within classes, Distance function matching). Specific details of these methods as programmed are provided in Robertson and Tou.

Imputation strata were defined by 2 digit industry and state. More detailed industry, Metropolitan Statistical Area (MSA), or county information could have been used to define more specific imputation strata. However, for the redesign, proposed sampling strata is defined by 2-digit industry and state.

While the end results of our simulations are employment levels, this study looks at the monthly change (expressed as a ratio) of employment as the variable to be imputed. We will compare four standard methods with two alternatives using administrative data. These methods have a presumed advantage of using more information about the nonrespondent and depending less on the absence of nonresponse bias (i.e. the sample being equally representative of respondents and nonrespondents). The models proposed are simple and require little data so that they do not challenge the resources available for conducting the monthly survey. Other models might provide better results but are not practical with time and resource constraints facing the survey.

Although we apply a chain of monthly links to arrive at our our estimates, since we have establishment response for the entire period, the chain of links simplifies to a link between the last administrative data and the current month. This simplified form is the what we will use when describing our estimators.

Establishment trend times the last observed value (UILT)-- The last administrative value available is multiplied by its over-the-month change from a year earlier to impute the current month.

$\hat{e}_{i,t} = \frac{e_{i,t-12}}{e_{i,l-12}} * e_{i,l}$, where $e_{i,t}$ is the establishment employment at time t, $e_{i,l}$ is the last available month of administrative data, and t > l.

Sample trend times the last observed value (UIST)-- In this situation, the last administrative value available is multiplied by the over-the-month change of the responding sample.

$\hat{e}_{i,t} = \frac{s_t}{s_l} * e_{i,l}$, where s_t and s_l are is the sum of employment across the participating sample at time t, and $e_{i,l}$ is the last available month of administrative data, and t > l.

The four standard methods are as follows:

Last observed value for the establishment (UILO)-- Also referred to as carry-over,

i.e. $\hat{e}_{i,t} = e_{i,l}$.

Mean imputation (MEAN)--Within the imputation strata, the ratio of the sum of current month employment to the sum of previous month employment across all respondents was multiplied by the previous month employment value of the nonrespondent. For the first month this previous month employment was reported, but for subsequent months it is imputed. Kalton refers to this as mean imputation within classes. Recall that as we are considering month-to-month ratios, this is essentially the current practice in CES.

Hot Deck-Random Selection (HDRS)--Randomly select a respondent from the imputation strata to represent the nonrespondent. The ratio of current month employment to previous month employment from this donor is multiplied by the previous month employment of the nonrespondent. This is referred to as random imputation within classes by Kalton.

Hot Deck-Nearest Neighbor (HDNN)--As with the hot deck random selection except rather than randomly select the donor, a respondent with the smallest difference in last reported employment is selected as donor. Referred to as distance function matching in Kalton.

All of these methods are applied within strata, and all impute a ratio of the current month employment to the previous month employment. In the case of donor records, the same donor was used for a nonrespondent over the entire estimation period.

Techniques like the hot deck were developed to reduce bias and to better estimate sampling error, but they only work well if the sample respondents are representative of the nonrespondents. Although certainty strata do not add to sampling error, nonrespondents of the units in these strata add to nonsampling errors. Application of a hot deck with random selection will add variability but it is from a different source, the random selection of donor records. The model behind the imputation process may contribute to bias and variability. The simulation here allows for comparison of bias estimated from a sample to the full response, or population, under the various imputation strategies described.

4. EVALUATION CRITERIA

The measures of error used here are the error (bias) and relative error (relative bias)

$$ERROR(BIAS) = \hat{E} - E$$

$$RELATIVE\ ERROR(RELATIVE\ BIAS) = \frac{\hat{E} - E}{E}$$

where $\hat{E}$ is the estimate of employment and E is the population employment.

5. Results

Because of space limitations results are presented for the last month of estimation, March 1994, rather than for all nine months. In this simulation, March is the estimate furthest away from the last available population value. Coincidentally, March is the official CES benchmark month. We present state total, as well as "national", which in this simulation is the sum of the 6 states, division and total estimates. We also present some summary results of the national 2-digit SICs, for each of the 55 industries used in the simulation.

The national total and division level results are contained in Table A. Total employment for the cell is given below the division heading in each cell. At the national total level, the carryover (UILO) method and the last UI value multiplied by sample trend (UIST) method outperform the others by a wide margin. UIST had a relative error of -0.14 percent while the relative error for UILO was -0.21 percent. The next smallest relative error was -0.77 percent for hot deck nearest neighbor (HDNN). The UILT and HDRS methods performed the worst with estimates near 2 percent. The one year of historical data used in the UILT does not provide for a useful model. At the division level UIST, in general, performs the best. It has the smallest errors in mining, manufacturing, wholesale trade, FIRE (Finance, Insurance and Real Estate), and services. In the construction and retail trade industries, it does not appear that any of the methods have substantially smaller errors than the UIST. In transportation and public utilities (TPU), however, UIST does not perform as well as most of the other methods. Error measures for state totals are given in Table B. The state total results are similar to the nation numbers. It appears that the UIST method is somewhat better than the other methods.

Establishment trend times the last observed does not perform well at this level (with the exception of Michigan where it does very well - - almost as well as UIST). UILT underestimates in every state except Michigan, and has the largest error in four states. This again demonstrates that the establishment of trend of one year ago is not a very good predictor of current establishment employment in a changing economy. New York yields the worst results across the six methods. UIST does produce a reasonable -0.67 percent error for New York, and is the only method with an error less than one percent.

Table A. Errors and relative errors for the six imputation methods, by Division (Employment level is given above the Error column for each division)

Total		
	2,905,627	
	Error	% Rel. Err.
HDNN	-22,457	-0.77
HDRS	-63,274	-2.18
MEAN	-32,261	-1.11
UILT	-57,130	-1.97
UIST	-4,040	-0.14
UILO	-6,177	-0.21

Mining		
	2,280	
	Error	% Rel. Err.
HDNN	18	0.79
HDRS	12	0.53
MEAN	22	0.96
UILT	21	0.92
UIST	8	0.35
UILO	-57	-2.50

Construction		
	25,622	
	Error	% Rel. Err
HDNN	-2,046	-7.99
HDRS	-2,643	-10.32
MEAN	-1,777	-6.94
UILT	-5,159	-20.14
UIST	-2,207	-8.61
UILO	-2,236	-8.73

Manufacturing		
	602,013	
	Error	% Rel. Err.
HDNN	-2,363	-0.39
HDRS	849	0.14
MEAN	-579	-0.10
UILT	-8,653	-1.44
UIST	325	0.05
UILO	-3,225	-0.54

TPU		
	81,472	
	Error	% Rel. Err.
HDNN	-663	-0.81
HDRS	-3,968	-4.87
MEAN	-1,612	-1.98
UILT	-1,236	-1.52
UIST	-1,750	-2.15
UILO	-25	-0.03

Wholesale		
	69,708	
	Error	% Rel. Err
HDNN	-1,716	-2.46
HDRS	-1,778	-2.55
MEAN	-1,543	-2.21
UILT	-3,092	-4.44
UIST	-1,321	-1.90
UILO	-1,840	-2.64

Retail		
	120,835	
	Error	% Rel. Err.
HDNN	-9,572	-7.92
HDRS	-38,144	-31.57
MEAN	-20,986	-17.37
UILT	233	0.19
UIST	645	0.53
UILO	4,131	3.42

FIRE		
	227,378	
	Error	% Rel. Err.
HDNN	-8,402	-3.70
HDRS	-5,506	-2.42
MEAN	-10,323	-4.54
UILT	-5,067	-2.23
UIST	149	0.07
UILO	309	0.14

Services		
	1,776,319	
	Error	% Rel. Err
HDNN	2,287	0.13
HDRS	-12,096	-0.68
MEAN	4,537	0.26
UILT	-34,177	-1.92
UIST	111	0.00
UILO	-3,234	-0.18

Ranking of the errors at both the state total and national division level, give strong evidence in support of the UIST method (Tables C and D). In an attempt to obtain an overview, the methods were ranked 1 to 6 (smallest to largest error), and the counts of those rankings for each method were placed in the tables. The weighted ranking column, shows the sum of ranks, i.e. 1 for smallest, 2 for next smallest, etc. The smaller this total, the better the method should be. Results are also generated for national 2 digit SICs. As mentioned previously, 55 two-digit SICs are considered in this analysis (Table E). At this level two UI based methods (carryover and UIST) outperform the others. All have a high occurrence of returning one of the three smallest errors in a 2-digit SIC. UIST had one of the smallest 3 errors in 71 percent of the industry groups. UILO had 73; the others were all 55 percent or less. In contrast, the three traditional imputation methods tend to have the largest errors.

Ranking the performance of the methods in state divisions (Table F) also lends support to the argument that the UIST method provides the best results. It has one of the three smallest errors in 58 percent of the state divisions. The UILO method also again does well with 56 percent of it's errors being among the 3 smallest. One change that should be noted is the improvement of the MEAN method in this case. In sixty-three percent of the state division level estimates, the MEAN imputation method yields the smallest three errors for the division. At the national 2-digit level, this was the case only 40 percent of the time.

Table B. Errors and relative errors for the six imputation methods, by State Total
(Employment level is given above the Error column for each State)

Florida		
	429,963	
	Error	% Rel. Err.
HDNN	2,303	0.54
HDRS	-228	-0.05
MEAN	2,374	0.55
UILT	-13,181	-3.07
UIST	-1,364	-0.32
UILO	2,229	0.52

Massachusetts		
	308,864	
	Error	% Rel. Err.
HDNN	895	0.29
HDRS	1,886	0.61
MEAN	1,658	0.54
UILT	-4,728	-1.53
UIST	1,820	0.59
UILO	3,537	1.15

Michigan		
	442,924	
	Error	% Rel. Err
HDNN	1,045	0.24
HDRS	-7,398	-1.67
MEAN	4,167	0.94
UILT	264	0.06
UIST	40	0.00
UILO	2,354	0.53

New Jersey		
	352,883	
	Error	% Rel. Err.
HDNN	2,061	0.58
HDRS	640	0.18
MEAN	805	0.23
UILT	-10,429	-2.96
UIST	1,257	0.36
UILO	-1,639	-0.46

New York		
	849,257	
	Error	% Rel. Err.
HDNN	-33,171	-3.91
HDRS	-55,975	-6.59
MEAN	-43,223	-5.09
UILT	-22,760	-2.68
UIST	-5,662	-0.67
UILO	-10,288	-1.21

Pennsylvania		
	521,736	
	Error	% Rel. Err
HDNN	4,410	0.85
HDRS	-2,199	-0.42
MEAN	1,958	0.38
UILT	-6,296	-1.21
UIST	-131	-0.03
UILO	-2,370	-0.45

Table C. Distribution of ranking of size of errors within "national" divisions

	Smallest Error 1	2	3	4	5	Largest Error 6	Weighted Ranking Sum	% of industries that method had one of 3 smallest errors
HDNN	0	3	2	2	1	0	25	63
HDRS	0	1	1	2	2	2	35	25
MEAN	1	2	0	2	2	1	29	38
UILT	1	0	2	1	0	4	35	38
UIST	5	1	1	0	1	0	15	88
UILO	1	1	2	1	2	1	23	50

Table D. Distribution of ranking of size of errors within state totals

	Smallest Error 1	2	3	4	5	Largest Error 6	Weighted Ranking Sum	% of industries that method had one of 3 smallest errors
HDNN	1	0	1	2	2	0	22	33
HDRS	2	0	1	1	0	2	21	50
MEAN	0	3	0	0	3	0	24	50
UILT	0	1	1	0	0	4	29	33
UIST	3	1	2	0	0	0	11	100
UILO	0	1	1	3	1	0	22	33

Table E. Distribution of ranking of size of errors within "national" 2-digit SICs

	Smallest Error 1	2	3	4	5	Largest Error 6	Weighted Ranking Sum	% of industries that method had one of 3 smallest errors
HDNN	1	9	11	16	11	8	215	38
HDRS	5	9	6	14	11	10	212	36
MEAN	4	6	12	17	10	6	206	40
UILT	16	9	5	4	6	15	185	55
UIST	13	9	17	2	8	6	166	71
UILO	16	13	11	5	4	6	151	73

Table F. Distribution of ranking of size of errors within state divisions

	Smallest Error 1	2	3	4	5	Largest Error 6	Weighted Ranking Sum	% of industries that method had one of 3 smallest errors
HDNN	10	6	4	8	7	8	149	47
HDRS	7	7	5	3	9	12	165	44
MEAN	8	7	12	8	6	2	132	63
UILT	9	4	5	2	8	15	170	42
UIST	9	5	11	9	7	2	135	58
UILO	4	14	6	9	7	3	139	56

6. Conclusions

In general, the UI based methods appear to provide slightly better results in this study, than do the traditional imputation methods. This is consistent with earlier research conducted by West, et. al. We would expect that using the most recent information available about a business would improve our imputation. How we use that information matters. Any establishment specific information that would improve our chances of estimating the employment should be used. Of the three UI based methods, the UIST method, using the most recent administrative information with sample trend to fill in the missing months yields the smallest relative error most of the time, while the carryover (UILO) method ranks second. But, as mentioned previously, this can vary depending upon the state, division or SIC.

It is possible that the best option for CES will be to apply different types of imputation methods depending on the estimating cell. Additional work needs to be done to develop criteria to decide which method to apply to a given cell. Regression models were fit to attempt to find differences in the 2-digit national estimates due to response rates. Very little difference between the methods was found in the models. It is also possible that simulations conducted in non-certainty strata on non-participants could provide additional information which would either help define criteria for use of the methods or find one method that stands out.

This analysis included employers in certainty strata for all industries, but only for employers with single worksites in one or more of six states. Before selecting an imputation method for this class of establishment the analysis should be expanded to include more states and multi-establishment employers over a longer time span.

References

Kalton, G. and Kasprzyk, D (1982). "Imputing for Missing Survey Responses" *American Statistical Association* 1982, *Proceedings of the Survey Research Methods Section.*

Tou, Albert and Robertson, Ken (1994). A Study of Imputation Effects on Estimation and Variance Estimation in the Occupational Employment Statistics, *American Statistical Association* 1994, *Proceedings of the Survey Research Methods Section.*

West, S., Butani, S., Witt, M., and Adkins, C. (1988). Alternative Imputation Methods for Employment Data, *American Statistical Association* 1988, *Proceedings of the Survey Research Methods Section.*

West, S., Kratzke, D., Robertson, K. (1993). Alternative Imputation Procedures for Item Non-response from New Establishments in the Universe, *American Statistical Association* 1993, *Proceedings of the Survey Research Methods Section.*