Non-response Adjustment in IRS' Taxpayer Compliance Studies

Karen Masken, Internal Revenue Service

The views expressed in this paper are my own personal views, not those of the IRS.

Introduction

The Internal Revenue Service (IRS) periodically conducts taxpayer compliance studies, the results of which are used in the strategic planning process and to develop workload selection formulas (i.e., which returns to audit). To conduct these studies, the standard IRS approach is to randomly select taxpayers and audit their returns to determine if there is any non-compliance. While the selected taxpayers probably do not view an audit of their tax return as the same as responding to a survey, the non-response, in the form of taxpayers not showing up for the audit, presents the same statistical issues as any other survey non-response. This paper takes a look at how the IRS currently deals with non-response and also at alternative methods that might be used to address this concern.

A recently completed compliance study examined taxpayers who claimed the Earned Income Tax Credit (EITC) for tax year 1999. The EITC is a refundable Federal income tax for the working poor. The maximum credit allowed for tax year 1999 was $3,816 and the credit was completely phased out for taxpayers with income greater than $30,580. In tax year 1999, 18.8 million taxpayers claimed $31.3 billion of EITC. The IRS estimates that between $9.7 billion and $11.1 billion of this was inappropriately claimed.

The sample for this study was a stratified random sample of 3,547 EITC claimants who were then subject to an audit. The sample was stratified on the presence of business income (none, positive, negative) and tax filing status (married filing jointly, head of household, other). In addition to these nine strata there was one for taxpayers whose credit amount was recalculated to be zero during processing.

Missing Data

There were 191 taxpayers in the EITC compliance study who failed to show up for the audit. This accounted for 5.5 percent of the sample, but represented 7.3 percent of the EITC population due to some having relatively large weights. For these "no-shows", the IRS has data on the amount the taxpayer claimed for the EITC from the sampling frame, but the amount of EITC allowed after the audit is missing. The following describes the three basic missing data mechanisms:

Missing Completely at Random (MCAR)
This assumes that the missing cases are just a random sample of the sample and are represented by the observed cases in general.

Missing at Random (MAR)
This assumes that the missing cases are not represented by the general sample of observed cases, but given certain characteristics, they are represented by observed cases with those same characteristics.

Not Missing at Random (NMAR)
This assumes that the missing cases are unique and are represented only by themselves.

The IRS has traditionally assumed that the mechanism for no-shows is NMAR (this corresponds to normal audit practice). The supposition is that people who do not show up for audits do so because they are non-compliant and therefore the entire EITC claim is disallowed. For the 1999 EITC study, there was some debate around whether this assumption was correct. The argument was made that there could be other reasons taxpayers did not show up for their audit. For example, the taxpayer might be intimidated, even if the claim is legitimate, or they may not have been able to get time off from work. In an effort to balance the opposing views, IRS used the MCAR point estimate as a lower bound and the NMAR point estimate as an upper bound.

While this approach does balance the opposing views, it introduced several problems. First, it seems peculiar to make such clearly opposite assumptions about the missing data mechanism. On one hand this approach assumed the no-shows look just like everyone else, on the other hand it assumed they look nothing like the observed taxpayers. If instead the assumption

was made that the mechanism is MAR, this could balance the opposing views, but not leave IRS in the difficult position of defending two opposing sets of assumptions. Another issue that arises from the assumptions of MCAR and NMAR is that one needs to use two different sets of weights. This introduces a variety of problems, not the least of which is that it allows analysts to arbitrarily choose which weights to use, leading to inconsistent results. Again, the assumption of MAR would mitigate this problem.

Analysis of Missing Cases

The first step was to look at the distribution of the missing cases. Table 1 shows the distribution, by strata, of both the missing cases and the sample for comparison. Note that the missing cases do not appear to be missing completely at random. In particular, taxpayers with no business income who file as head-of-household (strata 2b) account for a relatively high proportion of missing cases, while married taxpayers with negative business income (strata 4a) have a relatively low incidence of missing cases.

Table 1. Distribution by Strata of Missing Cases

Strata	Percent of Sample	Percent of Missing
1. Math Error EITC Filers	4	14
2. No Business Income		
a. Married Filing Jointly	8	3
b. Head-of-Household	27	39
c. Other (primarily Single)	8	13
3. Business Income > 0		
a. Married Filing Jointly	5	1
b. Head-of-Household	8	10
c. Other (primarily Single)	5	8
4. Business Income < 0		
a. Married Filing Jointly	21	2
b. Head-of-Household	10	6
c. Other (primarily Single)	4	3

When the assumption of MAR is made, characteristics that predict both non-compliance and missing cases should be used as the adjustment cells in order to reduce both bias and variance. Table 2 presents the compliance ratio (EITC allowed by auditor / EITC claimed by taxpayer) along with the percent of sample compared to percent of missing cases by filing status, gender and age. Note that the categories with lower than average compliance ratios have consistently higher proportions of missing cases compared to the sample proportion.

Table 2. Analysis of Missing Cases by Select Characteristics

	Compliance Ratio	Percent of Sample	Percent of Missing
Filing Status			
Married Filing Jointly	0.77	35	7
Head-of-Household	0.68	46	59
Male	*0.48*	*13*	*30*
Female	*0.77*	*33*	*29*
Other (primarily Single)	0.63	19	34
Male	*0.68*	*10*	*24*
Female	*0.58*	*9*	*10*
Age			
30 years or under	0.59	28	49
Over 30 years	0.75	72	51
Overall Compliance ratio	0.70		

Table 3 presents a comparison of the three mechanisms. The MCAR estimates are arrived at by excluding all missing cases and adjusting the strata weights. The MAR estimate is based on imputations using the hot deck method. Completed cases are randomly drawn from within the adjustment cells (strata, gender and age group), and their compliance ratio is applied to the EITC claimed on the missing case. As stated earlier, the NMAR assumes that the EITC allowed amount is zero for all missing cases (consistent with all missing cases being treated as non-compliant).

If the assumption of MCAR is incorrect, then the estimate will be biased and, in this context, is likely to overstate the amount of the credit allowed or understate the level of non-compliance (it is hard to imagine the no-shows are actually more compliant than the average taxpayer). Alternatively, if the assumption of NMAR is made and the model is misspecified, then the estimate will be biased and could result in understating the amount of EITC allowed or overstating the level of non-compliance. These points are illustrated in Table 3 (recall that the amount of EITC claimed for tax year 1999 was $31.3 billion). As expected, the MCAR estimate is higher than the MAR while the NMAR estimate is lower. It is also interesting to note that the NMAR point estimate does not fall within either of the other two 95% confidence intervals. This would suggest that the model is highly sensitive, which could be problematic if it is misspecified.

Table 3. Comparison of Estimated EITC Amount Allowed by Mechanism (in billions of dollars)

	Mechanism		
EITC Allowed	MCAR	MAR	NMAR
Lower 95% CL	21.4	21.2	20
Point Estimate	22.4	22.2	20.9
Upper 95% CL	23.3	23.1	21.8

Comparison of MAR Techniques

Once the assumption of MAR has been made, the issue becomes which technique to use to adjust for non-response. Of primary statistical concern is the fact that all adjustments add to the variability of the estimate, but this is not necessarily captured in normal variance calculations. Several techniques were used and the resulting point estimates, along with their respective coefficients of variation (CVs), are shown in Table 4 (as well as the MCAR and NMAR estimates for comparison). As expected, all of the MAR point estimates were similar to one another and fell between the MCAR and NMAR estimates. In addition, all the MAR estimates were closer to MCAR.

Post-stratification uses only complete cases and the weights are adjusted within the adjustment cells. While the post stratification technique has the largest CV of the MAR techniques, this variance is still underestimated. The weights are now random variables, but their variance is not captured in the CV. All of the remaining estimates are based on various imputation techniques that use all cases. Their variance is also underestimated since the imputation error is not captured in the CVs. The regression + random residual does capture more of this variance, as evidenced in the higher CV.

Table 4. Comparison of Estimated Amount of EITC Over-claimed Using Various Techniques (in billions of dollars)

	EITC Amount Overclaimed	CV
MCAR	8.9	2.20%
MAR		
Post-Stratification	9.1	2.20%
Hot Deck	9.1	2.14%
Last Obs. Carried Forward (LOCF)	9.2	2.15%
Regression	9.2	2.10%
Regression + Random Residual	9.0	2.14%
NMAR	10.4	2.27%

NOTE: These over-claim estimates are lower than those cited in the introduction because they also include cases with under-claims.

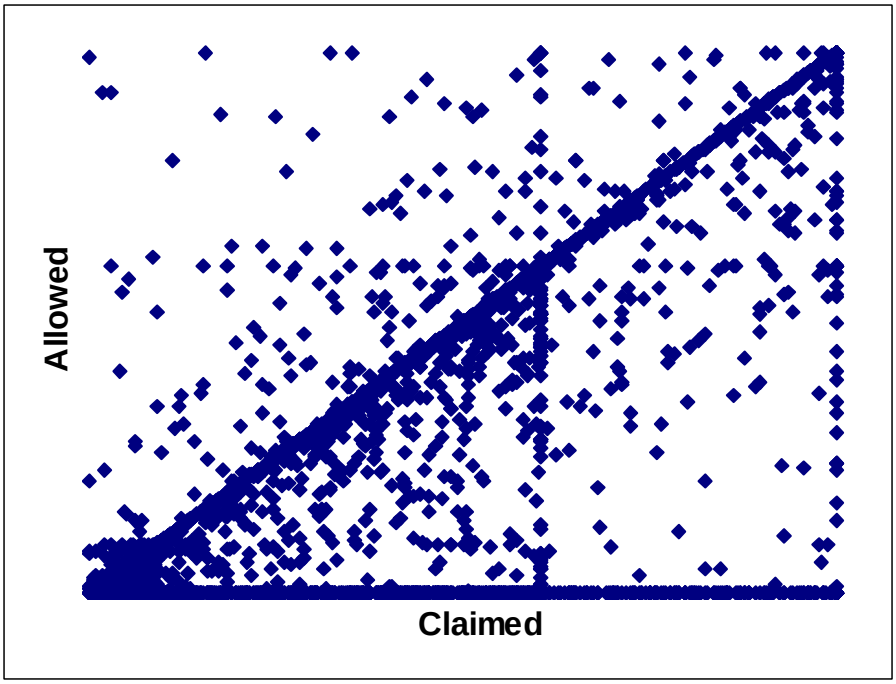
Issues in Modeling

One issue in developing imputation models is the lack of information the IRS has about specific individuals. A common misperception is that the IRS has a plethora of information about a taxpayer. The main source of information about taxpayers that the IRS utilizes is the information on the annual tax forms. While there is a lot of

financial information reported on a tax return, there is very little demographic information for individual taxpayers. In fact, filing status and number of dependents is just about all that is reported. The age and gender information used in this analysis for the adjustment cells is based on data from the Social Security Administration, but this data is also fairly limited.

Another common misperception is that the EITC amount claimed is a good predictor of the EITC amount allowed. The following scatter plot shows the unweighted plot from the EITC compliance study of the amount allowed versus the amount claimed (excluding the missing cases). This phenomena is typical of the IRS' compliance studies. The typical pattern is that a large proportion of taxpayers are allowed exactly what they claimed (appears as a solid line at a 45 degree angle), another large proportion of the claimed amounts are fully disallowed (solid line along the x-axis), a small proportion are allowed more than claimed by the taxpayer and the remainder are allowed a portion of what they claimed. The R-square for this study was 0.41.

Scatter plot of the EITC amount Allowed by the EITC Amount Claimed



Recommendations

While using the MCAR estimate as a lower bound and the NMAR estimate as an upper bound worked to illustrate the effects of opposing assumptions about the compliance characteristics of taxpayers who do not show up for an audit, this approach has some undesirable qualities. If the assumption of MCAR is incorrect then the estimate is biased and underestimates noncompliance. The estimated lower bound is too low. On the other hand, while the assumption of NMAR is plausible, it poses a quandary. If one assumes that the no-shows have completely different compliance characteristics than taxpayers who do show up, how does one know what these compliance characteristics are? The model, in the form of $Y = \beta X$ is : EITC Allowed = 0*EITC Claimed. This model is highly sensitive to the value of β, and setting that value to zero is arbitrary. This approach does provide a maximum upper bound, but it is likely a biased estimate, with the implication being that the upper bound is too high. The IRS approach of making contradictory assumptions about the no-shows to provide bounds for the estimate is rather difficult to defend.

One can argue that IRS should make one consistent assumption that the cases are missing at random and use a MAR technique. The resulting confidence limits around this single point estimate should then be used to determine the upper and lower bound of the overall compliance level. Using this approach, the estimate will fall between the MCAR and NMAR estimates, but it will tend to be much more precise and have tighter bounds.

The intent of this paper was to conceptually describe the assumptions the IRS currently makes in estimating compliance levels from its studies, along with their implications. There was no intent to revisit the estimates in the 1999 EITC compliance study, therefore, little emphasis was placed on researching adjustment cells. More research needs to be done to determine what the best predictors of both non-compliance and missing cases are. Part of this effort should include trying to gather and utilize more auxiliary data. For example, the IRS currently has access to the Federal Case Registry, which is a database that includes information on child custody orders. This data may prove useful as one dimension of the adjustment cells.

More imputation techniques should be explored as part of the MAR adjustment process. In particular, the use of multiple imputation, which would do a better job of capturing the variance introduced by imputation error, should be thoroughly explored. While several imputation techniques look promising for the EITC compliance studies, they may not work well for broader studies that cover a wider range of compliance issues. The IRS is nearing completion of data collection for a study of the reporting compliance behavior for the individual income tax (Form 1040). If an imputation method were used, it would require imputing every line item on the tax return, which would become unwieldy very quickly. Instead, IRS should explore using post-stratification for its broader studies and explore ways to capture the variance of the weights.

Finally, IRS should look into ways to cost effectively reduce non-response in its compliance studies.

References

Little, Roderick J.A. and Rubin, Donald B., *Statistical Analysis with Missing Data,* Second Edition, John Wiley & Sons, Inc. New York, NY, 2002.

Rubin, Donald B., *Multiple Imputation for Nonresponse in Surveys,* John Wiley & Sons, Inc. New York, NY, 1987.

Kish, Leslie, *Survey Sampling,* John Wiley & Sons, Inc. New York, NY, 1995.

Internal Revenue Service, *Compliance Estimates for Earned Income Tax Credit Claimed on 1999 Returns,* February 28, 2002.