OFFICE OF INSPECTOR GENERAL

Audit Report

Review of Statistical Methods Employed in the Financial Interchange Determination

This report was prepared by KPMG LLP, under contract to the U. S. Railroad Retirement Board, Office of Inspector General, and by acceptance, it becomes a report of the Office of Inspector General.

Original Signed by….

Martin J. Dickman
Inspector General
U.S. Railroad Retirement Board

Report No. 10-06
May 4, 2010

RAILROAD RETIREMENT BOARD

KPMG LLP
2001 M Street, NW
Washington, DC 20036

Independent Accountants' Report on Applying Agreed-Upon Procedures

Office of Inspector General
U.S. Railroad Retirement Board:

We have performed the procedures enumerated in the attached Exhibit, which were agreed to by the Office of Inspector General (OIG) of the U.S. Railroad Retirement Board (RRB), solely to assist you in evaluating the statistical methods employed in the Financial Interchange (FI) between RRB and the Social Security Administration and between RRB and the Centers for Medicare and Medicaid Services for the year ended September 30, 2008. RRB's management is responsible for the FI statistical methods and resulting calculations. This agreed-upon procedures engagement was conducted in accordance with attestation standards established by the American Institute of Certified Public Accountants and the standards applicable to attestation engagements contained in *Government Auditing Standards* issued by the Comptroller General of the United States. The sufficiency of these procedures is solely the responsibility of the RRB OIG. Consequently, we make no representation regarding the sufficiency of the procedures described in the attached Exhibit either for the purpose for which this report has been requested or for any other purpose. Our procedures and findings are summarized in the Exhibit.

We were not engaged to, and did not, conduct an examination, the objective of which would be the expression of an opinion on the statistical methods employed in the FI calculation as performed by RRB. Accordingly, we do not express such an opinion. Had we performed additional procedures, other matters might have come to our attention that would have been reported to you.

This report is intended solely for the information and use of the RRB OIG and RRB management, and is not intended to be and should not be used by anyone other than these specified parties.

KPMG LLP

April 23, 2010

Exhibit

Agreed-Upon Procedures and Findings for the Analysis of Statistical Methods Employed in the 57th Determination of the Financial Interchange (FI)

Procedure 1

Obtain the population of Railroad Retirement Act (RRA) beneficiaries and determine whether the RRB's Financial Interchange (FI) sample of beneficiaries is representative of the population.

a. Apply statistical analysis and modeling techniques to compare the characteristics of the FI systematic sample of annuitants with the overall population using all available information in the database provided by RRB, specifically focusing on data measures that are presumably correlated with the amount of benefits received by annuitants under the RRA.

b. Test the FI beneficiary sample (employee, spouse, and survivor beneficiaries), both pre- and post-sample stratification, to determine if the sample is representative of the population of annuitants receiving benefits under the FI.

Findings

We obtained the population of RRA beneficiaries and determined that the RRB's FI beneficiary sample is representative of the RRA beneficiary population.

a. We applied statistical analysis and modeling techniques to compare the characteristics of the RRB systematic sample of annuitants with the overall population of annuitants using all available information in the database, specifically focusing on data measures that are presumably correlated with the amount of benefits received by annuitants. We found that, though the RRB uses a systematic rather than random sampling methodology, the uniform distribution of social security numbers across beneficiaries yields an unbiased sampling result. (See further details below.)

b. We analyzed the RRB's FI beneficiary sample to determine if the sample is representative of the RRA beneficiary population and found that the sample is representative of the beneficiary universe. (See further detail below.)

Our findings[1] indicate that:

a. The RRB's methodology of using a 1 percent sample of beneficiaries—the sampled population is "cases" in which the retired employee's social security number (SSN) ends in "30"—constitutes a systematic sample, rather than a random sample. Based on generally accepted statistical sampling theory, a systematic sample with a random starting unit produces a random sample that is more evenly distributed over the population.[2] Thus, RRB could alternatively have used a random process to pick

[1] In examining the representativeness of the RRB annuitant sample to the "in force" universe, we assume that the majority of all relevant beneficiary information associated with benefit amounts is recorded in the database provided, and that the chance of hidden biases is small.

[2] We referred to Chapter 8 of Cochran, William B., *Sampling Techniques*, Wiley: New York, 1977 for a detailed discussion of systematic samples.

the last two digits of social security numbers, which would produce a stratified random sample with potentially more sampling precision. Given the practical constraints faced by the RRB, however, picking any last two SSN digits is an acceptable approach for the purpose of sample selection if there is no systematic correlation between the last two digits of social security numbers and retirement benefit amounts.

For all three RRA populations (non-disabled retired employees and their spouses (Retiree), Survivor, and Disability), there is evidence that the last two digits of the claim numbers are uniformly distributed. This implies that the systematic sample selected by RRB will be similar to a simple random sample selected based on a random number generator. When we test the RRB's FI sample against random samples (regardless of SSN), our findings indicate that for the Retiree population, RRB's systematic FI sample is relatively similar to a simple random sample, while those of the Survivor and Disability populations appear slightly skewed and could result in a potentially biased projection of the total benefit amount for those two populations. However, given that the last two digits of the claim numbers appear to be uniformly distributed for all three populations, and the correlations between these dollar amounts and the claim number (or the final benefit amount) appear to be relatively small, there is evidence that the potential bias in the benefit projection introduced by the RRB systematic FI sample would be relatively small.[3] Moreover, a simple random sampling approach would likely lead to a significantly different sample from year to year, and thus entail much more field work on the part of the Bureau of the Actuary, while the RRB systematic sampling approach would produce a relatively unchanged sample from year to year.

b. In analyzing each variable for systematic biases, we find that the most important variables tested are not statistically different from their population means, and for the few variables with means statistically different from the population means, the related dollar differences are so small relative to the mean benefit amounts as to be unlikely to bias the result.

We are limited in our ability to analyze the population data on a stratum-by-stratum basis. However, because our analysis indicates that the RRB's FI sample is representative of the overall population prior to the post-stratification, there is evidence that the various potential ways of stratifying and projecting the sample should lead to a very similar estimate of the total benefit amount. In other words, there is evidence that testing the overall randomness of the RRB's FI sample across all strata is sufficient in this case. This is supported by the theory that if the sample is similar enough to a simple random sample, the overall sample should be representative of the entire population. If this is the case, regardless of how the random sample and population is stratified, the subsample within each stratum should remain representative of the subpopulation within each stratum.

We find that the existing RRB documentation is fragmented and insufficient to enable a qualified external party to replicate the sampling process in the FI calculation. Moreover, the RRB Bureau of the Actuary does not retain clean versions of the November Master Benefit File sampling universe

[3] To quantify the magnitude of the bias, it would be necessary to compare the projection results calculated based on the RRB systematic sample with that computed based on a statistically random sample. Given that RRB does not have such a random sample with benefit amount calculated, it is not possible to quantify the exact magnitude of the potential bias.

each year for longer than the current one-year term. Therefore, it is not possible to examine more than the most recent year's FI sample in subsequent years, which is an important capability from an external analysis perspective.

Procedure 2

Obtain the universe of RRA annuitants and determine whether there is a statistically significant correlation between the last two digits of the social security numbers and the total employee benefit amount.

a. Test for any statistically significant correlations between the last two digits of social security numbers and various employee characteristics (such as demographic information) that are presumably correlated with the amount of benefits received. This analysis will further reinforce the testing on whether RRB's systematic FI sample selection will lead to an unbiased estimate of the total amount of benefits received.

Findings

We obtained the universe of RRA annuitants and determined that there is not a statistically significant correlation between the last two digits of the social security numbers and the total employee benefit amount.

a. We analyzed the universe of RRA annuitants for any statistically significant correlations between the last two digits of social security numbers and various employee characteristics that are presumably correlated with the amount of benefits received and found that the statistical correlations were insignificant. (See further detail below.)

We based our analysis of the universe of RRA annuitants on the following hypothesis: if we find in our analysis that the last two digits of the SSNs of RRB's FI beneficiary population are not correlated with any of the beneficiary demographic information that influence the amount of retirement benefit received, the test will provide statistical evidence that the RRB systematic FI sample selected based on the last two digits of employees' SSNs would not be skewed in any observable way, and thus would likely lead to an unbiased projection of the total amount of benefits received.

Our correlation analysis indicates that overall, there are no significant relationships between the SSNs/claim numbers and any of the relevant variables, as the few variables with statistically significant correlations were generally very close to zero. Moreover, the use of the age variable (one of the statistically significant correlated variables within the disability population) to stratify the sample further reduces the influence of age on the projection of the final benefit amount.

Agreed-Upon Procedures and Findings for the Analysis of Statistical Methods Employed in the 57th Determination of the Financial Interchange (FI)

Procedure 3

Determine whether the RRB benefits sample size overall and sample size within each post-stratum is sufficient to achieve the required sampling precision.

Findings

We determined that the RRB benefits sample size overall is sufficient to achieve the required sampling precision.[4] However, the sample size within each post-stratum would need to be increased to achieve a 5 percent precision limit. (See further detail below.)

In theory, if a simple random sample of sufficient size is selected using a probability selection method from a population, the distribution of any observable variable within the sample should be similar to that within the population.[5] The RRB sample size is over 4,000 employees, and thus is adequate to allow RRB to make precise projections of the total benefit amount over the entire population. However, if RRB needs to make equally precise projections of the benefit amount within each demographic stratum, then the RRB sample sizes for some strata would need to be larger to achieve a 95 percent confidence level with a 5 percent precision limit.

Procedure 4

Recalculate the projection formula used by RRB to project the FI sample to the universe and determine whether RRB's extrapolation implementation is consistent with the original sample design and sample stratification.

a. Compare formulas used in the RRB projections, including the point estimate and standard error calculations, to classical statistical formulas.

[4] The statistical sampling section in the AICPA audit guidance requires a targeted confidence level of 95 percent in financial statement audits when the risk of significant misstatement (RoSM) is high. In addition, the U.S. Government Accountability Office (GAO) Financial Audit Manual (FAM) also requires a targeted confidence level of 92 percent to 95 percent in financial statement audits of Federal Government Agencies when the risk of significant misstatement is high. To comply with this federal government audit standard, when the RoSM is high, we target a sampling precision of 5 percent at a confidence level of 95 percent.

[5] We referred to Chapter 1 of Cochran, William B., Sampling Techniques, Wiley: New York, 1977 for a detailed discussion of the advantages of a probability sampling.

Findings

We recalculated the projection formula used by RRB to project the FI sample to the universe and determined that RRB's extrapolation implementation is consistent with the original sample design and sample stratification.

a. We compared formulas used in the RRB projections and found consistency between those formulas and classical statistical formulas. (See further detail below.)

We analyzed the calculations performed by RRB and compared RRB's calculations with the appropriate formulae used in classical statistical sampling textbooks.[6] We found that the point estimate and variance formulas used by RRB are consistent with the formulas used for the stratified mean estimator; thus, we conclude that the projection produces a reliable total benefit amount as calculated by RRB.

We also used the mean and standard deviation computed from the sample data to calculate an independent projection. Both our point estimate and sampling error calculations based on the sample data match closely to that computed by RRB.

Procedure 5

Using the available data obtained from RRB:

a. Determine whether the FI beneficiary sample size and sample size within each post-stratum is sufficient to produce reliable estimates of population parameters. Compute an alternative projection of the total benefit amount using the same stratified mean estimator.

Findings

a. We analyzed the available data to determine whether the beneficiary sample size and sample size within each post-stratum is sufficient to produce reliable estimates of population parameters. In addition, we computed an alternative projection of the total benefit amount using the same stratified mean estimator and found that the RRB's projections of the point estimate and sampling error are consistent with the extrapolation results using the mean projection formula shown in statistical sampling textbooks. (See further detail below.)

We performed a sample size analysis and determined that the FI sample size used by RRB was adequate in order to produce a reliable overall projection of the total benefit amount over the entire RRA annuitant population; however, the size of some of the sample strata is lower than the required sample size at a 95 percent confidence level with a 5 percent precision limit.

[6] The stratified mean estimator and its projected variance are defined in Chapter 5 of Cochran, William B., *Sampling Techniques*, Wiley: New York, 1977.

To compare the RRB-calculated benefit amount to alternative projected results, we calculated the independent projections using the sample totals provided by RRB. We calculated a stratified mean estimate for the total benefit amount, which is approximately 0.07 percent[7] less than the projected benefit amount computed by RRB. This difference is due to a small number of missing records[8] and rounding errors. Thus RRB's projection produces a reliable total benefit amount.

We also used the statistical software STATA to project the standard error based on the sample benefit amount provided by RRB, computing the sample standard deviation of the benefit amount within each stratum using STATA, and then projecting the overall standard error by applying the sample ratio (or inflation factors). Based on our analysis, our calculated standard error is 0.7 percent less than RRB's calculated standard error. Thus, we find that RRB's projections of the standard error and sampling error are accurate.

Procedure 6

Compute the RRB's projections of the total amount of FI benefits using the following statistical extrapolation methods.

a. Compare alternative post-stratification and projection results (using methods specified below) with the RRB post-stratification and projection results to assess whether the RRB post-stratification minimizes the sampling error, and to determine if there is an alternate post-stratification of the RRB FI sample with a smaller sampling error. Compute the stratified difference, stratified separate ratio, and stratified combined regression estimators and use them to project the total benefit amount.

Findings

a. We compared alternative post-stratification and projection results with the RRB post-stratification and projection results and determined that the RRB post-stratification method minimizes the sampling error. Using the available data provided by RRB, and calculating the stratified difference, stratified separate ratio, and stratified combined regression estimators, we were not able to identify an alternate post-stratification of the RRB FI sample with a smaller sampling error. (See additional detail below.)

Based on the results of the correlation analysis of the sample, we performed a projection based on the stratified regression estimator.[9] Comparing the results to those calculated in procedures 4 and 5 above, we conclude that the mean estimator is a more precise sampling method than the regression estimator.

Unlike the regression and ratio estimators, the stratified mean estimator is theoretically unbiased and consistent. Given these desirable properties and based on our analysis of the stratified mean estimator, we

[7] 0.07 percent is equal to approximately $3.9 million, relative to the total benefit amount of about $5.9 billion.

[8] Specifically, the beneficiary universe data provided to us excluded records in OASI strata 23 and 24.

[9] The stratified regression estimator and its projected variance are defined in Chapter 7 of Cochran, William B., *Sampling Techniques*, Wiley: New York, 1977.

conclude that the stratified mean estimator used by RRB is a reliable approach to use for the projection of the total benefit amount.

Procedure 7

Analyze the data and sample extrapolation methodology used to calculate Hospital Insurance (HI) tax income (i.e. the amounts due the Centers for Medicare/Medicaid Services' trust fund) for active Railroad employees.

a. Verify the ratio extrapolation calculations based on the FI sample of RRB employees.

b. Determine whether the RRB's use of a ratio extrapolation method to calculate HI tax income calculations using a sample of covered employees produces an accurate estimate of the HI payroll amount with the required sampling precision.

Findings

We analyzed the data and sample extrapolation methodology used to calculate HI tax income for active Railroad employees.

a. We verified the ratio extrapolation calculations based on the sample of RRB employees by confirming that the gross and creditable sample sizes were identical and by replicating the HI calculation.

b. We determined that the RRB's use of a ratio extrapolation method to calculate the HI tax income using a sample of active Railroad employees produces an accurate estimate of the HI payroll amount with the required sampling precision. Given the statistically significant correlation among the gross payroll amount and creditable payroll amount, the large sample size, and the relatively small coefficients of variation, there is evidence that the ratio estimation method is appropriate, and the bias resulting from using the ratio estimation method for the extrapolation of HI tax amount is negligible. We are able to replicate the ratio estimate of the HI tax amount using the RRB FI sample data with a negligible difference due to rounding. Therefore, we find that the sample size and ratio estimation method used by RRB in the HI tax income calculations are appropriate.