Resolving Balance Complex Discrepancies in the Presence of Negative Data

Laura Bechtel[1], Nicole Czaplicki[1], Maria Garcia[1], Jeremy Knutson[1]
[1]U.S. Census Bureau, 4600 Silver Hill Rd, Washington, D.C. 20233

Economic data are often constrained by additivity conditions, where a set of item values (detail items) are required to sum to an associated total value. The set of detail items and their respective total is referred to as a balance complex. When these additivity constraints are not met, changes must be made to either the total or the set of details. Raking proportionally adjusts each detail item by the same amount. If each item's reporting error is random and has variance proportional to its value, then raking minimizes a chi-squared statistic. However, raking was developed for strictly positive data and can produce erroneous values when negative data are included. Modifications have been developed to address this situation, but implementation is not straightforward and does not always yield a feasible solution. In this paper, we develop separate linear and nonlinear programs that minimize loss functions under specified additivity constraints that work with negative data and include item costs. We apply the proposed methods to examples from the Quarterly Financial Report conducted by the U.S. Census Bureau, examining statistical properties of the resultant solutions.

Key Words: Consistency Edits, Raking, Imputation, Quadratic Programming, Linear Programming

1. Background

A balance complex is an additivity condition specifying the requirement that two or more item variables balance to a reported item total, i.e. $y = \sum_{i=1}^{D} x_i$ where D is the number of details, x_i's are known as the details and y is known as the total. If there are errors in the detail values then the balance complex fails, i.e. $y \neq \sum_{i=1}^{D} x_i$. The Economic Directorate at the Census Bureau ensures that data in records with failing balance complexes are adjusted before publishing relevant statistics. Sigman and Wagner (1987) and Luery and Sigman (2000) developed and implemented raking algorithms for resolving these discrepancies within the Plain Vanilla (PV) and the Standard Economic Processing System (StEPS) generalized editing and imputation systems respectively; the PV subsystem is used primarily to edit and impute Economic Census Data and StEPS is employed by many ongoing economic surveys.

[1] *Any views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.*

StEPS implements a raking algorithm for adjusting balance complexes in order to satisfy the requirement that the sum of items (details) in a balance complex balances to reported total. Raking proportionally adjusts detail values by calculating the ratio of the total to the reported details and multiplying each detail by this ratio so that the sum of these adjusted details balances to the corresponding total value. Under the assumption that the errors in the detail items are random with variance proportional to the value of the detail, raking minimizes a loss function in which the objective function is the chi-square "statistic" derived under these assumptions with the constraints as the additivity conditions (balance complex); see Deming (1943, Ch. 5).

The raking method was developed for positive data only. This method implicitly assumes that the reported *distribution* of details is accurate and should be preserved and that the ratio of the total to the summed details is "small." Consequently, raking can lead to erroneous values when data items are permitted to be negative or if there is subtraction in the balance complex. First, the assumption that the variance is proportional to the details may no longer hold. Second, the reported proportions will no longer be preserved, as raking changes all details in a balance complex in the same direction by the same amount. The Quarterly Financial Report (QFR), conducted in the Economic Directorate at the U.S. Census Bureau, is an example of a survey that consists of balance complexes that contain negative detail items and/or subtraction of details (see Section 2).

Luery and Sigman (2000) modified the raking algorithm described above to handle cases in which the detail items can have negative values. They proposed a more restrictive assumption by requiring that the variance be proportional to the **absolute value** of the reported detail (Luery and Sigman, 2000; Eltinge, 2003). Implementation of this modified raking method has not been straightforward, and we have found several situations where the solutions may be inconsistent and one situation where the solution is not correct because the adjusted details do not balance to the total.

The algorithm, as implemented in StEPS, imputes the detail items according to the following formula:

$$x_i' = \left(1 + sign(x_i)\frac{y - \sum x_i}{\sum abs(x_i)}\right) x_i \tag{1.1}$$

where y is the total, x_i is the original detail value for item i and x_i' is the imputed value for item i. When all items are real-valued (i.e., can be positive or negative), this formula works well. However, when x_i' <0 and item i cannot be negative, the balance complex must be resolved in another way. When this occurs, StEPS sets $x_i' = 0$ and adds the value it would have assigned to x_i' to the next item in the complex in addition to the value that item would have received by itself.

This algorithm is best illustrated through an example. In this example and throughout our research we do not assess the assumption that the ratio of the total to the summed details is "small." Consider the balance complex $y = x_1 + x_2 + x_3$ and $y = -200$, $x_1 = -12$, $x_2 = 59$, and $x_3 = -17$ where x_2 can only take non-negative values. Using (1.1), the algorithm initially imputes the values as $x_1' = -43$, $x_2' = -95$, and $x_3' = -62$. Since x_2 must be non-negative, it adds the -95 to x_3' and gives the final result of $x_1' = -43$, $x_2' = 0$, $x_3' = -157$. Although the imbalance has been solved, the solution no longer includes the positive contribution from the second item and therefore loses valuable information.

However, when the order of the items changes, a different result may be obtained. Suppose, we wrote the balance complex as $y = x_2 + x_1 + x_3$. The raking algorithm would now add the -95 it would have given to x_2, to x_1 giving a final result of $x_2' = 0$, $x_1' = -138$, $x_3' = -62$. This solution balances as well but distributes the total differently among the details just because the order of the details was changed.

Finally, suppose we wrote the balance complex as $y = x_1 + x_3 + x_2$. Since $x_2$is the last item, there is no place to put the -95 that would have been assigned to x_2. The "resolved" imbalance solution is $x_1' = -43$, $x_3' = -62$, $x_2' = 0$, which clearly does not add to the total $y = -200$.

The differences in the results from the StEPS raking algorithm given by different orderings of the items in the balance complex specification highlights an undesirable property of this method. Ideally, the solution should be invariant to the item order in the balance complex. However, this problem only arises under very specific circumstances. First, one or more detail items must be positive valued only. Secondly, the residual, defined as $y - \sum x_i$, must be negative. Finally, the absolute value of the residual must be greater than the sum of the absolute values of the details, $\sum abs(x_i) < abs(\, y - \sum x_i)$.

Clearly, when a survey has negative data, a simple extension of the raking algorithm is not straightforward and can yield an erroneous solution when the reported total is negative. Consequently, we needed to come up with an alternative to the raking method that is easy to implement, allows for real valued detail items, and provides valid solutions when the reported total can be negative. Currently, a new version of StEPS, aptly called StEPS II, is under development. One of the surveys scheduled to migrate to StEPS II is the QFR. We take this opportunity to explore alternative methods to raking when the total item and/or some of the detail items are allowed to be negative using data from the QFR. The QFR had additional requirements related to balance complexes for us to consider:

- Certain detail items should be held constant when resolving the balance complex
- The balance complex resolution method should be able to incorporate detail reliability information so that certain details are modified before others.

While conducting this research our goal was to find a method that was a satisfactory alternative to raking and flexible enough that the additional requirements could be addressed with our proposed method.

We present a brief description of the Quarterly Financial Report in Section 2. This is followed by Section 3 which describes the alternative methods that we considered for resolving failed balance complexes. Section 4 describes the implementation and evaluation of a case study using QFR data. We conclude in Section 5 with some recommendations for future research.

2. The Quarterly Financial Report

The Quarterly Financial Report (QFR) produces principal economic indicators that provide comprehensive and timely financial data, essential to calculation of key U.S. Government measures of national economic performance. Based upon a sample survey, the QFR presents estimated statements of income and retained earnings, balance sheets, and related financial and operating ratios for Manufacturing corporations with assets of $250,000 and over, and corporations in Mining, Wholesale Trade, Retail Trade, and Selected Service

Industries with assets of $50 million and over. For more information, see the "How the Data are Collected" section of the QFR website: https://www.census.gov/econ/qfr/collection.html.

There are three different forms used by QFR. Manufacturing companies with assets between $250 thousand and $50 million receive the short form. Companies with assets greater than $50 million receive one of two long forms, which collect more line items. One long form is for companies in manufacturing, mining, retail, and wholesale trade industries, the other long form is for companies in selected services industries. The items collected on both the long forms are the same, and thus for the remainder of this paper they will be collectively referred to singularly as "the long form." Figure 1 below presents an excerpt from the short form for *Total Liabilities And Stockholders' Equity*. In the form, the item number appears just before the response field. For example, item number 301 is presented just before the response field for question A1, *Short-term loans from banks*. These numbers are used in-house to describe the various items collected on the form. From this point forward we will use these numbers interchangeably with the descriptions on the form. Copies of the various QFR forms can be found at https://www.census.gov/econ/qfr/forms.html .

			AMOUNT (in thousands)
A	1. Short-term loans (original maturity of 1 year or less) from banks— Include overdrafts	301	$
	2. Other short-term loans (original maturity of 1 year or less)— Include commercial paper	304	$
B	3. Trade accounts and trade notes payable	306	$
C	4. Domestic income taxes accrued, prior and current years, net of payments— Include overpayments	309	$
D	Current portion of long-term debt— Classify noncurrent portion in line F.		
	1. Loans from banks	310	$
	2. Other long-term loans	313	$
E	All other current liabilities— Include accrued expenses and excise, sales, and payroll taxes[3].	315	$
F	Long-term debt due in more than one year— Classify current portion in line D.		
	1. Loans from banks	316	$
	2. Other long-term loans	319	$
G	All other noncurrent liabilities— Include deferred income taxes and minority stockholders' interest	320	$
H	1. Capital stock and other capital (less treasury stock)	326	$
	2. Retained earnings at end of quarter	322	$
	3. **Stockholders' equity— Sum of lines H-1 and H-2**	327	$
I	**TOTAL LIABILITIES AND STOCKHOLDERS' EQUITY— Sum of lines A-1 through G and H-3**	328	$

Figure 1: Liabilities and Stockholders' Equity Balance Sheet from Short Form

The QFR survey form contains two balance sheets: Schedule B1 (Assets) and Schedule B2 (Liabilities and Stockholders' Equity). On each of these balance sheets, it is important that the value entered on the (final) total line is the sum of the values entered for the line items above the total line. On the short form, there are eight items that must sum up to total assets and 12 items that must sum up to total liabilities and stockholders' equity. On the long form, there are 16 items that must sum up to total assets and 16 items that must sum up to total liabilities and stockholders' equity. A few of the detail items that go into the calculation of the totals are themselves the sum of other items on the balance sheet. For example, on both the long and the short forms, the line item "net property, plant, and equipment" (Item 219), which is a component of totals assets, is itself the sum of three other line items (216+217-218). Therefore, when the total assets balance complex is resolved either 219 needs to be held constant or 216, 217, and 218 need to be updated after the complex is brought into balance. We refer to this type of relationship as a nested balance complex, because the sum of the values from one complex is a detail in another complex.

Table 1: QFR Balance Complexes Subject to Raking After Imputation

Total		**Details**															
	$\boldsymbol{Y}$	$\boldsymbol{X_1}$	$\boldsymbol{X_2}$	$\boldsymbol{X_3}$	$\boldsymbol{X_4}$	$\boldsymbol{X_5}$	$\boldsymbol{X_6}$	$\boldsymbol{X_7}$	$\boldsymbol{X_8}$	$\boldsymbol{X_9}$	$\boldsymbol{X_{10}}$	$\boldsymbol{X_{11}}$	$\boldsymbol{X_{12}}$	$\boldsymbol{X_{13}}$	$\boldsymbol{X_{14}}$	$\boldsymbol{X_{15}}$	$\boldsymbol{X_{16}}$
B1 Long Form Items	Item 223	Item 201	Item 202	Item 203	Item 204	Item 205	Item 206	Item 207	Item 208	Item 209	Item 211	Item 212	Item 214	Item 215	Item 219	Item 220	Item 221
B1 Short Form Items	Item 223	Item 201	Item 202	Item 210	Item 213	Item 214	Item 215	Item 219	Item 222								
B2 Long Form Items	Item 328	Item 301	Item 302	Item 303	Item 305	Item 306	Item 307	Item 308	Item 310	Item 311	Item 312	Item 314	Item 316	Item 317	Item 318	Item 320	Item 327
B2 Short Form Items	Item 328	Item 301	Item 304	Item 306	Item 309	Item 310	Item 313	Item 315	Item 316	Item 319	Item 320	Item 322	Item 326				

Note: Items in green-shaded cells are kept constant for this research
Items in yellow-shaded cells can take negative values

Table 1 displays a representation of the QFR balance complexes subject to balancing after imputation. There is one total item represented by $\boldsymbol{Y}$, and 16 detail items are represented by $\boldsymbol{X_1}, \boldsymbol{X_2}, \ldots, \boldsymbol{X_{16}}$. The Schedule B1 Asset items from the long form and short form are displayed in rows 3 and 4, respectively. Item 223 denotes the reported total assets that must balance to the sum of the detail asset items. The Schedule B2 Liability and Stockholders' Equity items are displayed in rows 5 (long form) and 6 (short form), where Item 328 is the reported total and must balance to the sum of the reported details.

Although most of the items are positive valued only, some items, including item 309, "Domestic income taxes accrued, prior and current, net of payments – Include overpayments," can take either a positive or negative value.

The balance complexes in QFR have many detail items with very intricate relationships, so that there are often discrepancies between the sum of the details and their corresponding reported total. Furthermore, some companies may not have the requested detailed breakdowns readily available and may report an estimate of the value of the detailed item or leave it blank. This leaves the analysts, who are staff accountants and have specialized expertise, with the job of resolving these discrepancies by either calling the companies to ascertain a response or using their expertise to fill in the response. This can be quite cumbersome and time consuming. An alternative method for resolving out-of-balance complexes that relies on the same principles employed by the staff accountants would save a lot of time and resources and would eliminate a random error source

In addition to correcting reported out-of-balance complexes, QFR accounts for unit nonresponse via imputation. If previously reported data are available, this method uses statistical procedures utilizing previously reported data and data from current respondents of similar asset size and industry classification. Imputation is done item by item, and the imputed values of the detail items may not sum to the imputed value of the total item. In this case the imputed data need to be corrected so that the complex is in balance. Currently, the QFR uses raking, among other methods, to correct the balance complex discrepancies on imputed data. However, the current raking procedure is very vulnerable to errors, because it operates on *all* items in the balance complex and cannot hold any detail item constant. A delicate workaround has been put into place that only works because imputation is done for unit nonresponse. This workaround would not work if the QFR began imputation for partial response. Additionally, the workaround is not straightforward and adds several steps to raking the balance complex which increases the possibility of inadvertent errors during the imputation process. As we investigate alternatives to raking, we add the requirement of holding an item constant when balance complexes are nested. Our goal is to find a method for resolving out-of-balance complexes that is flexible enough to produce similar results to the currently-used raking method and at the same time can be modified to incorporate the preferences/rules used by the staff accountants in resolving the failed balance complexes. In the following section we describe the methods we considered.

3. Considered Methods

We investigated methods for finding optimal solutions to balance complex failures with the SAS® procedure PROC NLP, considering several different objective functions that can be used to meet the varying requirements a survey may have when resolving a balance complex.

For the methods outlined below, let c_i be the cost of changing detail i; x_i, x_i', and y retain the previous definitions. We assume that $y \neq 0 \; and \; \sum_{i=1}^{D} x_i \neq 0$.

Weighted Squared Difference (WSD): This approach is designed to minimize the weighted sum of the squared deviations of the perturbed values from the reported values. In this case the objective function is $f(\mathbf{x}') = \sum_{i=1}^{D} c_i (x_i - x_i')^2$. We include the reliability costs in the objective function as a means to control the frequency of which items are changed. This method is equivalent to raking as proposed by Luery and Sigman (2000) when $c_i = \frac{1}{|x_i|}$.

Weighted Absolute Difference (WAD): As an alternative to WSD, we looked at minimizing the weighted sum of absolute differences between the reported and perturbed detail item values. In this case, the objective function is $f(\mathbf{x}') = \sum_i c_i |x_i - x_i'|$.

Squared Difference of Ratios (SDR): This approach is designed to minimize the sum of the squared differences between the proportion of each detail to the summed total and the proportion of the perturbed detail to the reported total. Ultimately, the end goal is to preserve the reported distribution of the details to their summed total. In this case, the objective function is $f(\mathbf{x}') = \sum_{i=1}^{D} \left(\frac{x_i}{\sum_{i=1}^{D} x_i} - \frac{x_i'}{y} \right)^2$.

Ratio of Absolute Totals (RAT): The SDR method does not always preserve the distribution of the details to their summed total when negative data are present. This is because the magnitude of the differences will be large if there is a change of sign between the reported and perturbed values. The RAT approach compensates for distortion of the distribution that can occur when SDR is used. It also prevents any ratios in the objective function from being larger than 1. This approach minimizes the sum of the squared differences in the proportions of the absolute value of the reported details to the sum of absolute value of details and the proportions of the absolute perturbed values of each detail to the sum of the absolute value of the perturbed details. In this case, the objective function is $f(\mathbf{x}') = \sum_{i=1}^{D} (\frac{|x_i|}{\sum_{i=1}^{D} |x_i|} - \frac{|x_i'|}{\sum_{i=1}^{D} |x_i'|})^2$.

There are three core constraints:

- Most important, the details need to sum to the reported total y, i.e. $y = \sum_i x_i$
- Nonnegative items must have a solution that is greater than or equal to zero
- Input zero values should not be perturbed.

Additional constraints can be added depending on the requirements in the survey. In the following section, we describe an additional constraint that we implemented to meet the requirements of the QFR.

4. Case Study

For our case study we used data from the fourth quarter of 2012 through the third quarter of 2015 from the Quarterly Financial Report for a total of 12 statistical periods. We focused our research on reported data that fail at least one balance complex. This made it easy for us to compare our results to the analyst corrections, because we also had the analyst corrected versions of the data. We focused our research efforts on the B2 short form and long form balance complexes from Table 1, because these complexes included items that can contain negative values. The aggregate data are subject to sampling error. However, we are focusing on individual companies and sampling errors will not change our results or conclusions.

4.1. Implementation

When implementing the considered methods for the QFR, there are some survey-specific considerations that need to be addressed. First, an additional constraint needs to be added to the three core constraints outlined in Section 3 to ensure the additivity rules of the QFR survey are met. As mentioned earlier, some items need to be held constant. So, we add the constraint that an item's perturbed value must be equal to its reported value.

Our proposed method(s) will be used for two different scenarios that rely on two very different philosophies. In the first scenario, the analyst correction scenario, the optimized solution would be used in place of analyst corrections. In this scenario the philosophy is that some items' values are more reliable than others in general, and we would want to modify the values accordingly.

In the second scenario, the raking scenario, we are attempting to find an optimization method that can be used in place of raking. While the StEPS raking algorithm does not work perfectly with negative data, it does work well with positive data and is currently used in QFR. The philosophy in this case is that for the most part, each item is equally reliable and the distribution of the details should be maintained to the furthest extent possible.

For the analyst corrected scenario, we chose to employ costs according to how the staff accountants (analysts) choose to modify the balance complex items, independent of the order. Consequently, we consulted with them to understand how they decide which items to change when adjusting a failing balance complex. They provided us with four different levels, the verbal description of the costs, of how they make changes to the items as presented in columns 2-5 in Table 2. They explained how there are some items that are never changed (e.g. totals), while other items can be classified as rarely changed, sometimes changed, or most likely to be changed. The last column lists items that should be held constant by schedule and form type.

Table 2: Analyst Assigned Reliability Levels

	Analyst Assigned Reliability Categories				
	Most Likely to Change	Sometimes Change	Rarely Change	Never Change	Hold Constant
B1 Short Form Items	Item 215 Item 222	Item 201 Item 209	Item 202 Item 213	Item 214	Item 219
B1 Long Form Items	Item 215 Item 221	Item 201 Item 209	Item 202 Item 203 Item 204 Item 205 Item 206 Item 207 Item 208 Item 212	Item 211 Item 214 Item 220	Item 219
B2 Short Form Items	Item 315 Item 320	Item 301	Item 304 Item 306 Item 309 Item 310 Item 313 Item 316 Item 319	Item 322 Item 326	Item 319*
B2 Long Form Items	Item 314 Item 320	Item 301	Item 303 Item 306 Item 307 Item 308 Item 310 Item 312 Item 316 Item 318	Item 302 Item 305 Item 311 Item 317 Item 327	Item 317*

* Item 317 and Item 319 were held constant for research purposes only and are not held constant during production.

We use the analysts' insight to assign numeric costs to the QFR items, where the higher costs indicates higher reliability, meaning it is less likely to be changed. Table 3 presents the final numeric values used for the WSD and WAD optimization methods. We believed we could control the frequency at which a particular item would be changed by using costs, but it proved to be a bit more involved than that, because the costs really control the magnitude of the change and not necessarily the frequency. We began with a never-change cost of 200 and identified changes to "never change" items for over 90% of the resolved records. We kept increasing the cost for the "never change" items first to 1000, then 1,000,000 and so on until we reached 10 trillion when we felt that the frequency of changes to the "never change" items was at an acceptable level.

Table 3: Numeric Values of Costs

Categorization	**Cost**
Never Change	10,000,000,000,000
Rarely Change	50
Sometimes Change	25
Most Likely to Change	1

We only considered the WSD and WAD methods for the analyst corrected scenario, because applying costs when the goal is to maintain the distribution of details, as is the case in the raking scenario, is counterproductive.

For the raking scenario, we considered the methods that are designed to preserve the distribution of the details, specifically the RAT and SDR methods. Additionally we employed the WSD optimization method with a cost of $c_i = \frac{1}{|x_i|}$, because this should be nearly equivalent to the current raking method implemented in StEPS. Currently, StEPS does not have the capability to hold an item constant for the nested balance complexes. So, in our implementation for each method being compared to the StEPS raked value (RAK) we evaluated two sets of results, one where all items are allowed to change and the other with the specified item held constant.

Table 4 below summarizes the considered methods and under which scenarios they are applied.

Table 4: Methods Tested on Each Balance Complex by Objective

	Objective	
Method	**Analyst Corrected Scenario**	**Raking Scenario**
WSD	Y, with reliability costs	Y, with costs = $\frac{1}{|x_i|}$
WAD	Y, with reliability costs	N
SDR	N	Y
RAT	N	Y

4.2. Evaluation

We considered several summary statistics to evaluate how closely the perturbations match the results obtained by currently employed analyst corrections and by the raking procedures. We look at statistics that serve as a gauge for how close each method matches the desired results for each scenario. We use the analyst corrected values as our target in the evaluation, but note that this is only a guideline and not a gold standard. Analyst corrections can differ for the same case, and it would be impossible to match exactly how they correct the data.

First, we look at the change frequency percentages at which items are substantively changed for each method. We define a substantive change for observation j as a change were the rounded absolute difference of the perturbed value and the reported value is greater than zero, i.e. $round\left(\left|x_{i,j} - x'_{i,j}\right|\right) > 0$. We calculate the percentage of times each detail *i* changes in statistical period *t* using method $m\ as\ \ pct\Delta_{j,t,m} = \frac{\sum_{j=1}^{J} \Delta_{i,j,t,m}}{J}$ where J is the number of eligible cases with out-of-balance complexes and $\Delta_{i,j,t}$ is a zero-one indicator of a substantive change. If the optimization solution matched the analyst correction, these change frequency percentages would be the same.

A similar *rate* of change for each of the items being modified by the balance complex does not imply that the changes themselves are similar. We measure this feature using the relative perturbation measure, defined as

$$\mathrm{RelP}_{i,j,t} = 100 * \frac{Resolved\ item_{i,j,t} - Reported\ item_{i,j,t}}{Reported\ item_{i,j,t}},$$

and the average relative perturbation as

$$AvgRelP_{i,t} = \frac{1}{J}\sum_{j=1}^{J} \mathrm{RelP}_{i,j,t}.$$

For our recommended method, we are looking for the optimization method that results in average relative perturbations that are generally closest to the targeted scenarios average relative perturbations.

These statistics help us gauge how close each method's results are to the targeted scenario. However, it is also important that our recommended method would not make egregious changes to the estimates. After analyst review and correction is performed, the QFR data can still be machine-edited and changed for simple errors, so for this particular scenario it is very hard to gauge what the effect on the reported totals will be. However, raking occurs after imputation and just before estimation, so we are particularly interested in a method that will not have a severe effect on the totals estimates. In order to assess the effects on the estimates, we compare weighted total estimates of each detail item for each considered method along with StEPS raking for eligible observations that have out-of-balance complexes for each publication level industry, *pl*. These totals are defined as

$$\hat{X}_{i,t,m,pl} = \sum_{j=1}^{J} w_{i,t} x'_{i,j,t,m,pl}$$

where $w_{i,t}$ is the final weight used by the QFR. It is important to clarify the difference between a total estimate of a detail item and a total item. A total estimate is the aggregate of a **single** detail item across all eligible observations, J, and a total item should be the aggregate of **all** detail items for a single observation, j.

4.3. Analyst Corrected Scenario Results

Below we present box plots that summarize our findings for the change frequency percentages for each of the considered reliability levels presented in Table 2 comparing the results of the applicable considered methods to the analyst corrected values, labeled AC. The results for the long form and short form were very similar and for brevity's sake we will only present the short form results here. The long form results are available upon request. Figures 2 - 4 present box plots for the change frequency percentages for most likely change, sometimes, and rarely change details for the short form, respectively. These plots show that in general our considered methods change the details more frequently than the analysts do.

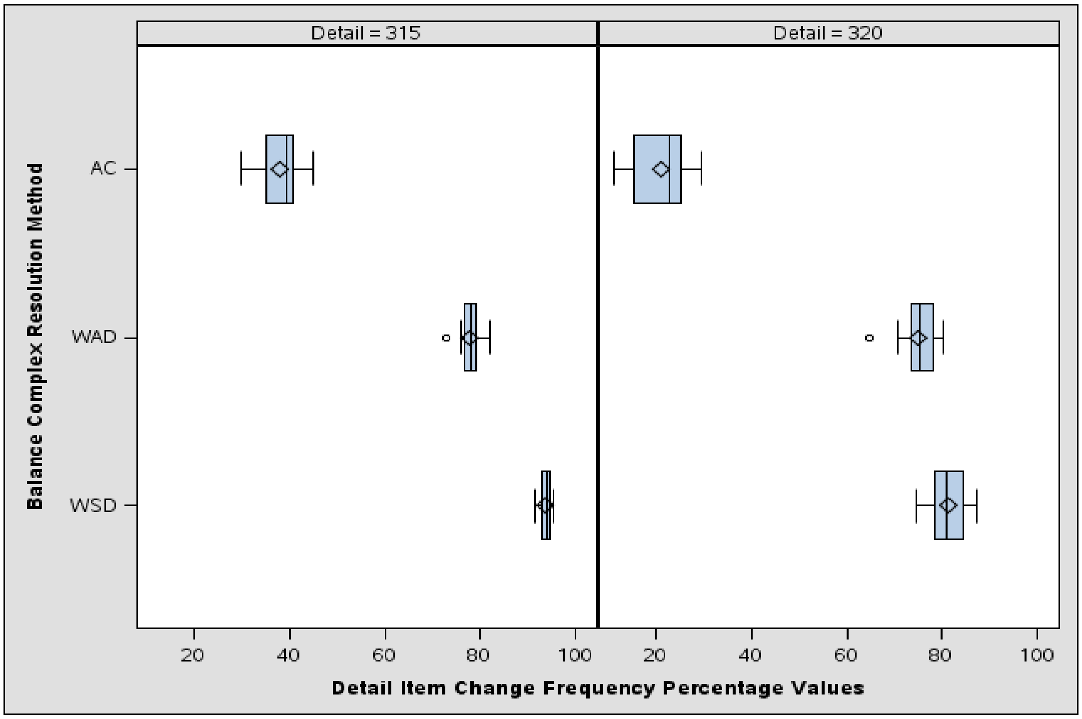

Figure 2: Analyst Corrected Scenario - Boxplots of the Change Frequency Percentages for Most Likely Change Items for the QFR Short Form 2012Q4-2015Q3[2]

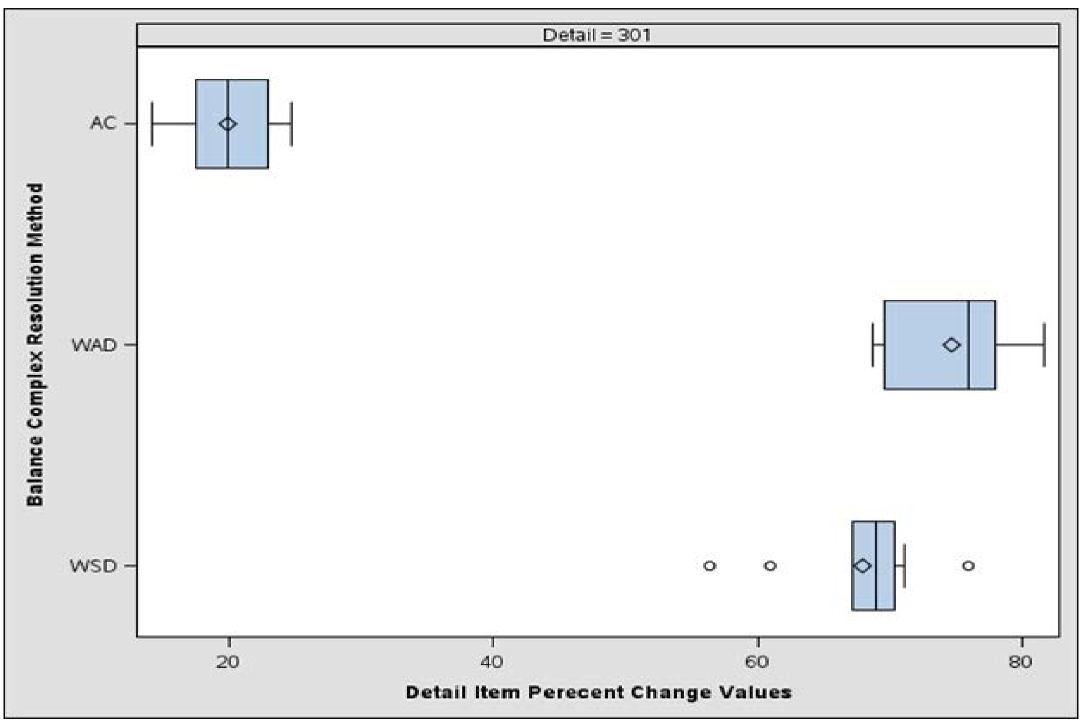

Figure 3: Analyst Corrected Scenario Boxplots of Change Frequency Percentages for QFR Short Form 2012Q4-2015Q3 Sometimes Change Detail

---

[2] While the data are subject to sampling error, the sampling errors do not affect the conclusions from these results.

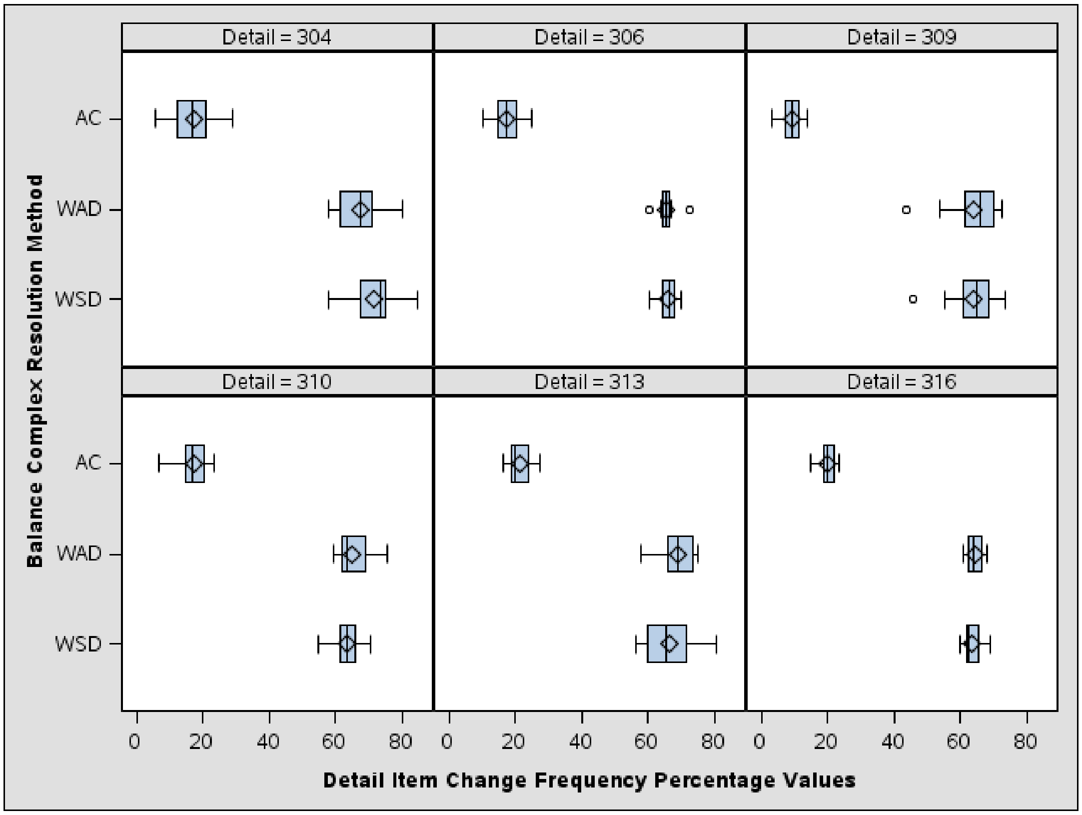

Figure 4: Analyst Corrected Scenario - Boxplots of Change Frequency Percentages for QFR Short Form 2012Q4-2015Q3 Rarely Change Details

Figure 5 presents box plots of the change frequency percentages for the two never change detail items (Items 322 and 326) for the short form. Here we see that our considered methods change the details less frequently than the analysts do. This is expected: we added an extremely high cost (reliability weight) to the never change items to ensure these items are not changed when solving the nonlinear optimization problem.

We confirmed that the item we held constant, Item 319, remained constant for all balance complexes and the change frequency percentages were all equal to zero. However, the analysts do not work under this constraint for this particular item. Consequently, we do not expect our considered methods change frequency percentages to match that of the analysts, and we do not present that comparison here.

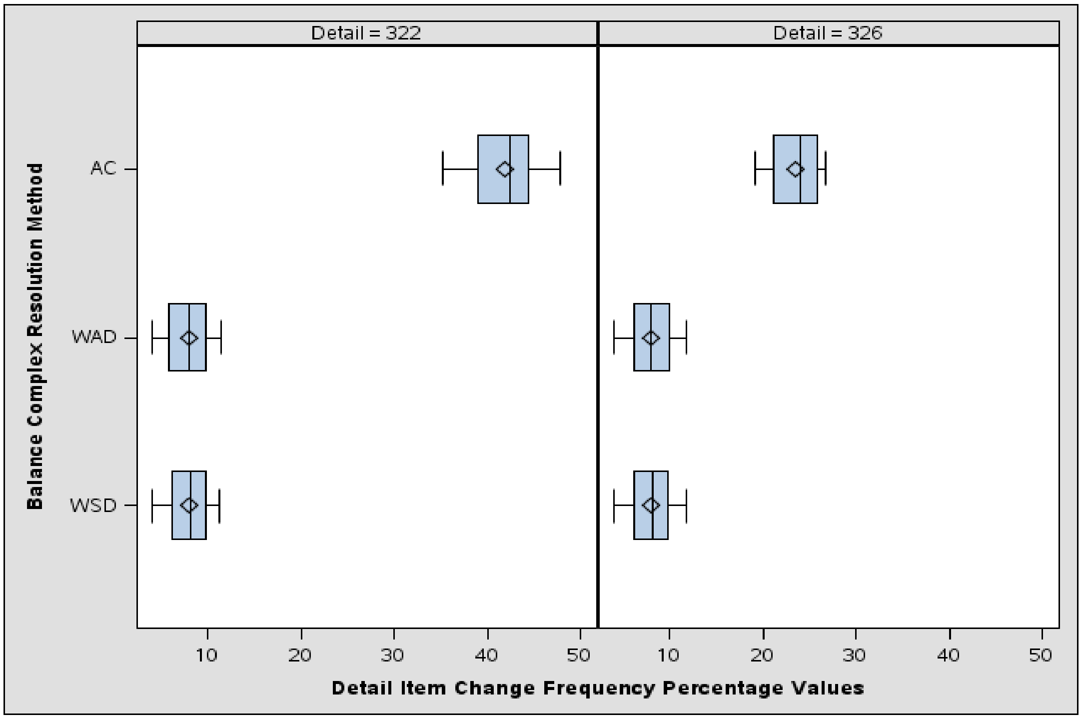

Figure 5: Analyst Corrected Scenario - Boxplots of Change Frequency Percentages for the QFR Short Form 2012Q4-2015Q3 Never Change Details

For this analysis we focused on getting the never change reliability weights to the appropriate level, and these results confirm that we selected an effective value. Looking at change frequency percentages for the rarely change reliability level items, we could use a similar process to assign the rarely change weight value. Ultimately, these change frequency percentages indicate that it is possible to meet the requirements outlined by the analysts.

In addition to the importance of obtaining similar change frequency percentages it is also important that the considered methods perturb the detail items on average in a way that reflects the reliability categories defined by the analysts. To assess this, we start by examining summary boxplots of the average relative perturbations for the most likely change and sometimes change reliability level detail items in Figures 6 and 7. While we are trying to match the requirements outlined by the analysts, we will look at the actual analyst corrections to see how their corrections are to the targeted requirements. We find that our considered methods, WSD and WAD, perturb the values by a higher magnitude than the analysts in these cases. These items have the lowest reliability costs and it is unsurprising that they are changed more frequently and the magnitude of the change is consistently larger.

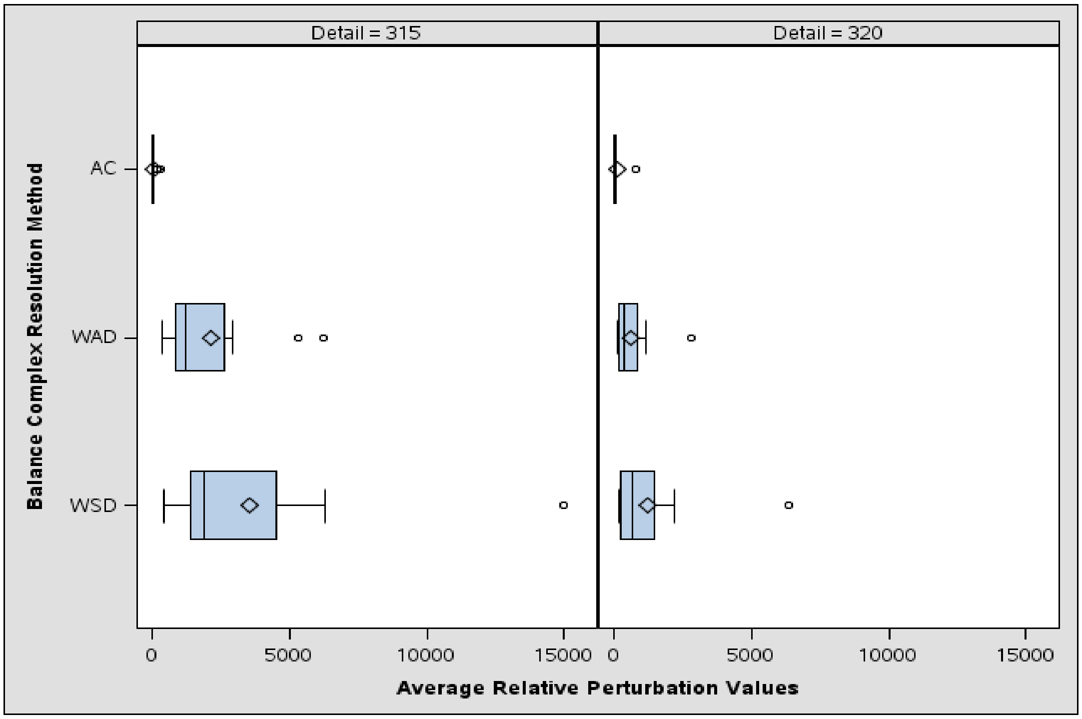

Figure 6: Analyst Corrected Scenario - Boxplots of the Average Relative Perturbation for the Most Likely Change QFR Short Form 2012Q4-2015Q3 Detail Items

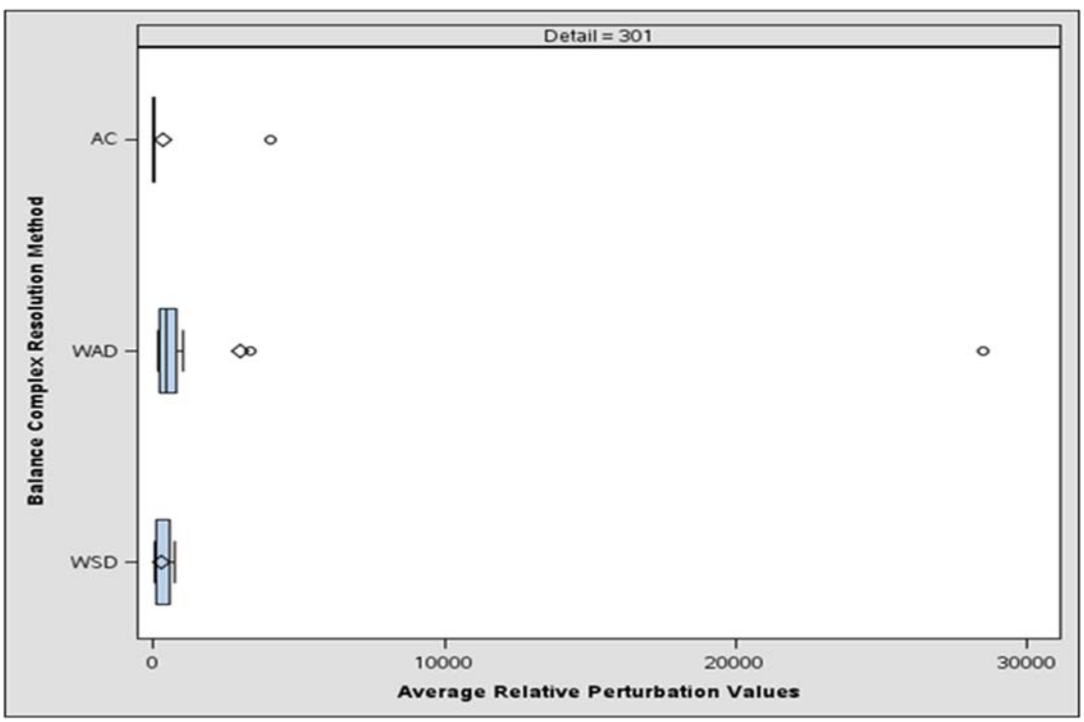

Figure 7: Analyst Corrected Scenario - Boxplots of the Average Relative Perturbation for the Sometimes Change QFR Short Form 2012Q4-2015Q3Detail Items

Next, we look at box plots of the average relative perturbations for the rarely change items. We find that the perturbations that result from the WSD method are similar in value to the analyst corrections for four of the six detail items. We present the box plots of items 304 and 306 in Figure 8 and 313 and 316 in Figure 9.

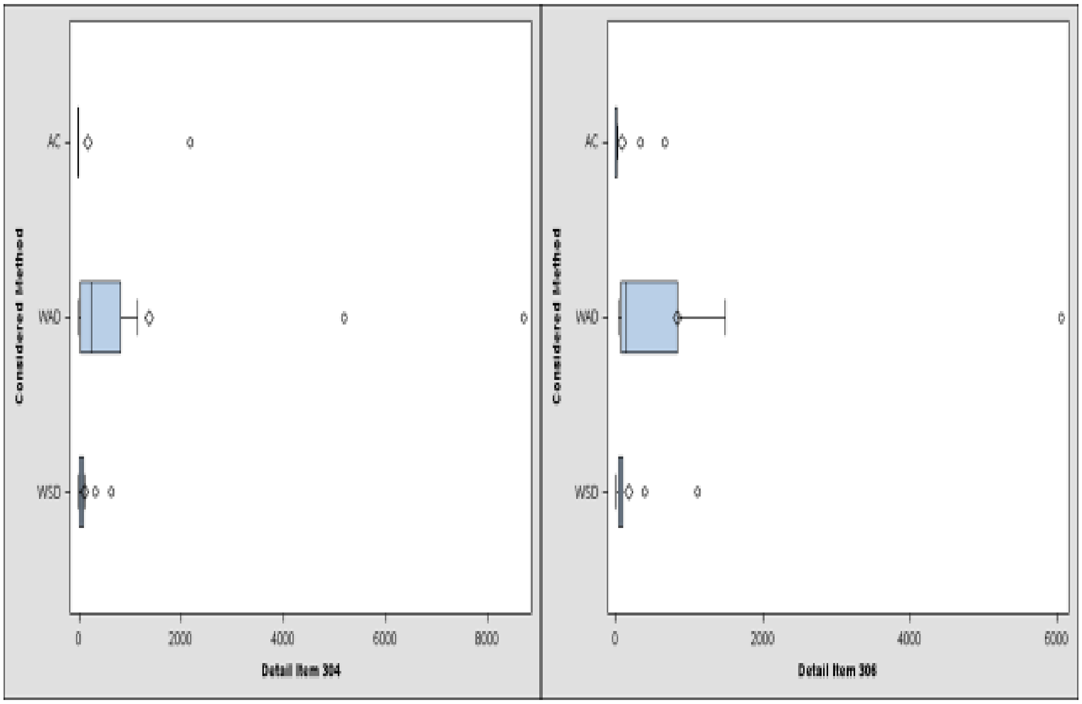

Figure 8: Analyst Corrected Scenario - Boxplots of the Average Relative Perturbation for Rarely Change QFR Short Form 2012Q4-2015Q3Detail Items 304 and 306

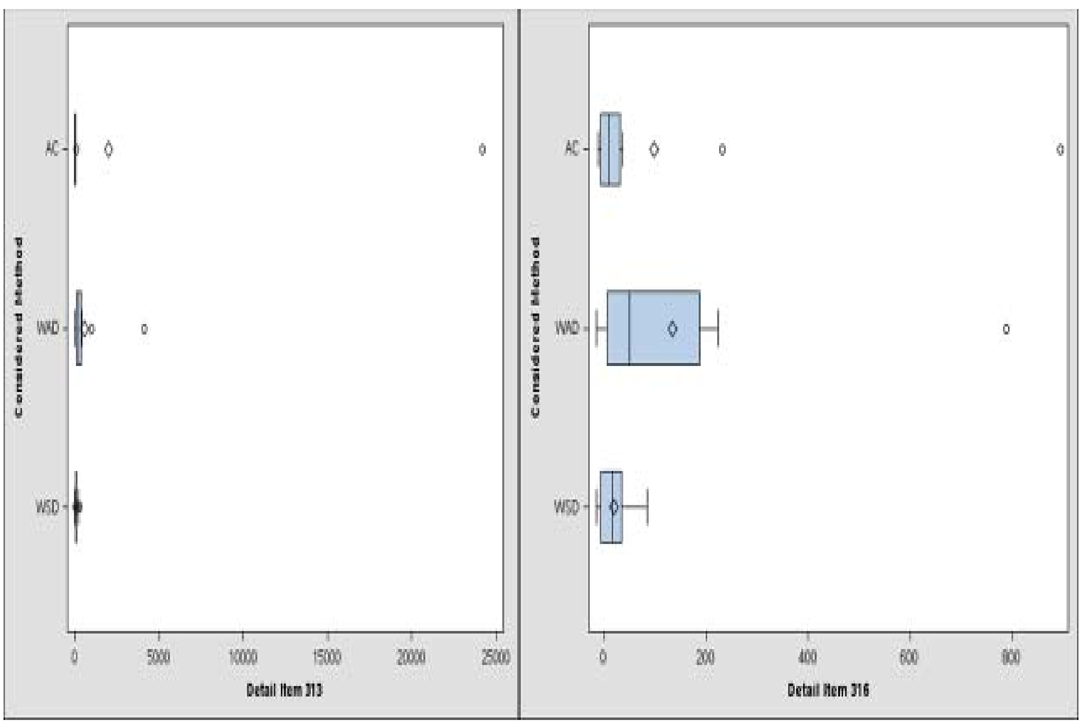

Figure 9: Analyst Corrected Scenario - Boxplots of the Average Relative Perturbation for Rarely Change QFR Short Form 2012Q4-2015Q3Detail Items 313 and 316

For two of the items, 309 and 310 present in Figure 10, the WSD and WAD methods result in larger perturbations than the analyst changes. Perhaps, modifying the assigned reliability cost values as discussed with the previous results may help to bring these average relative perturbations down to the analyst levels.

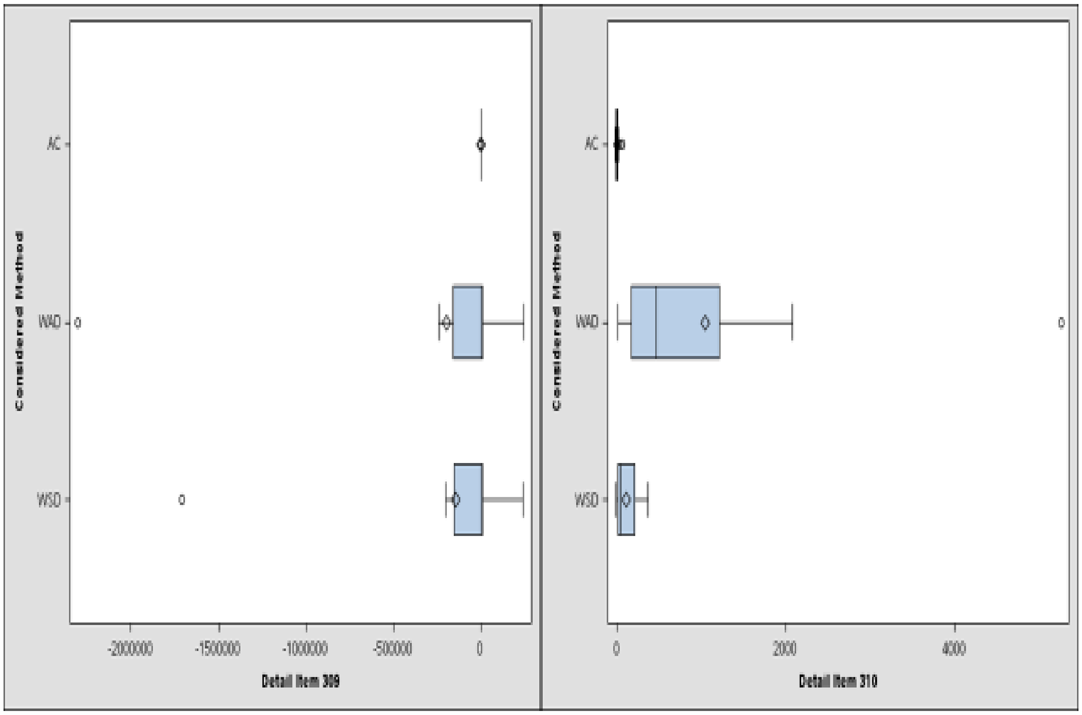

Figure 10: Analyst Corrected Scenario - Boxplots of the Average Relative Perturbation for Rarely Change QFR Short Form 2012Q4-2015Q3 Detail Items 309 and 310

We find results that are more consistent with the analysts' stated preferences for the average relative perturbations for the never change items, 322 and 326, presented in Figure 11. Here, we see that our considered method had comparable perturbations when compared to the analyst corrections for detail 326. However, we note that the WSD and WAD methods resulted in smaller perturbations than the analyst corrections. In this case the WSD and WAD methods are optimizing according the assigned reliability levels, but the analysts are not.

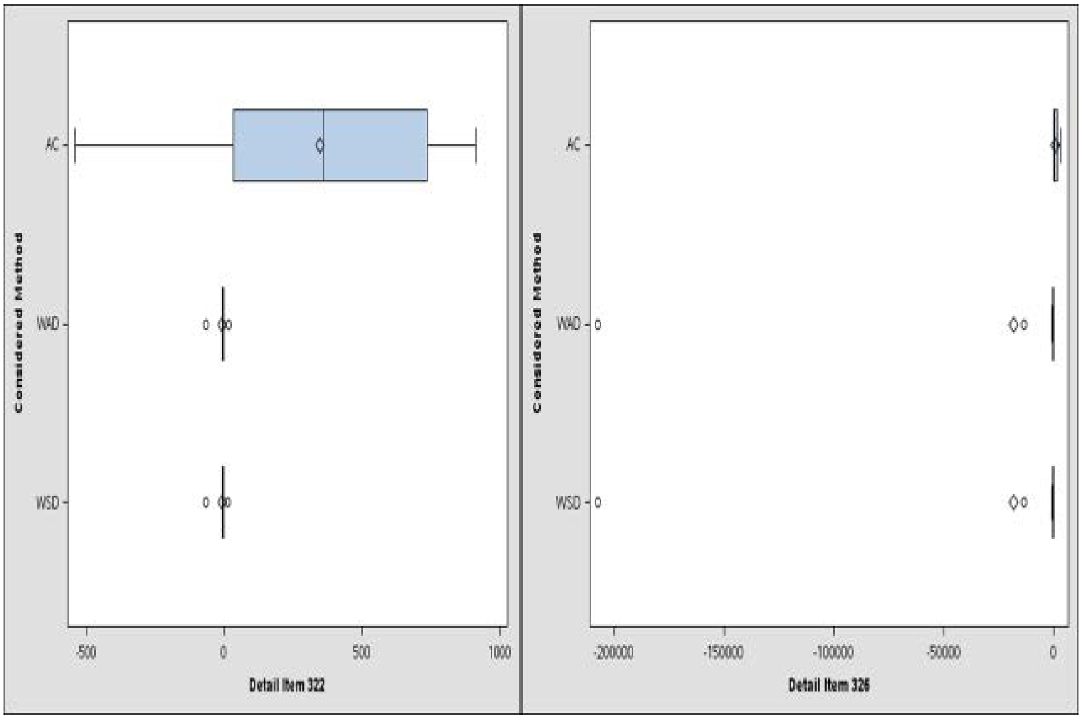

Figure 11: Analyst Corrected Scenario - Boxplots of the Average Relative Perturbation for Never Change QFR Short Form 2012Q4-2015Q3 Detail Items 322 and 326

After looking at the difference with the change frequency percentages and the average relative perturbations, we noted some discrepancies between the results from our considered methods and the analyst corrections that indicated perhaps we were meeting the requirements of the assigned reliability level, but the analysts did not seem to be correcting that data at that same level. We took these results to the analysts and they are revisiting the assigned reliability levels. In the meantime, we are reassessing the values of the costs assigned to the rarely change reliability level. However, we have shown that we can use these optimization methods to come up with solutions that can conform to the analysts' specifications.

4.4. Raking Scenario Results

For the methods that attempt to replicate the desirable properties of the raking algorithm, the results from the short form and the long form had similar patterns, and much like the previous section for brevity's sake we present the short form results here. Before we present the results for the raking scenario, recall that the StEPS raking algorithm does not allow an item to be held constant, but this is a requirement for the QFR balance complexes. The results we present in this section have item 319 held constant and the comparison to the raking results is not direct. However, we looked at these results without item 319 held constant and the differences were minimal.

Figures 12 and 13 present summary box plots for the change frequency percentages for all of the short form items (except 319) using the applicable considered methods compared to StEPS raking (RAK). Here most of the box plots show a similar pattern, where the RAK method and the WSD method are pretty well aligned and the RAT and SDR are distinctly different from the RAK and WSD methods. One exception is detail 316, where all of the methods have similar change frequency percentages. It does appear that raking has a slightly smaller change frequency percentage on average when compared to the WSD method. However, in general this difference appears to be less than five percent and would only be a big concern if the average relative perturbations also differed substantially.

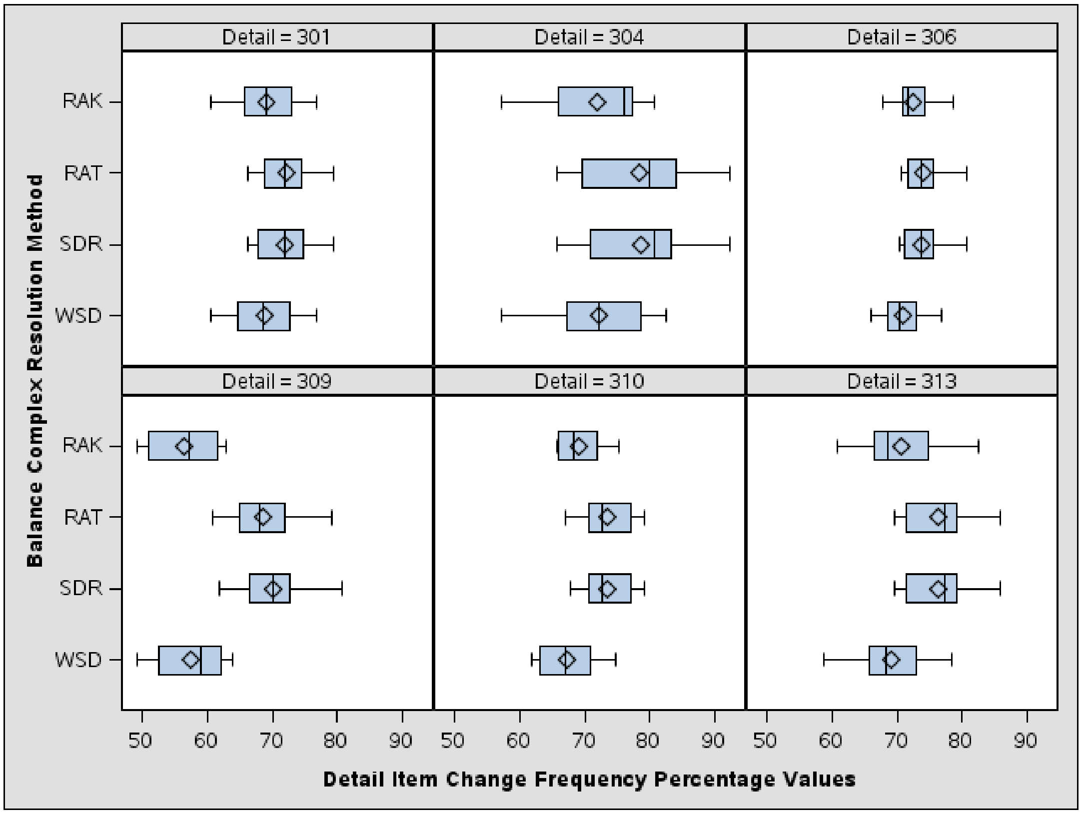

Figure 12: Raking Scenario - Boxplots of Change Frequency Percentages for the QFR Short Form 2012Q4-2015Q3 Details

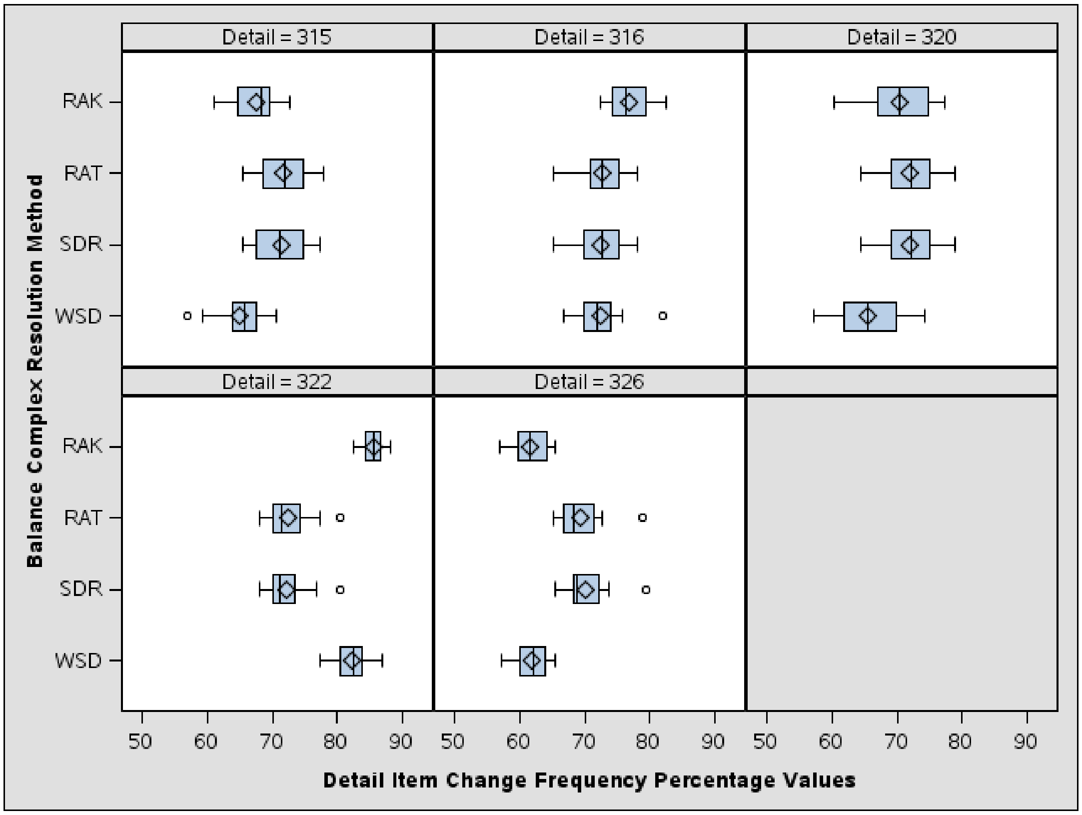

Figure 13: Raking Scenario - Boxplots of Change Frequency Percentages for the QFR Short Form 2012Q4-2015Q3 Details *Continued*

Looking at the average relative perturbations we found similar patterns for all of the detail items. The RAK and WSD methods had very similar average relative perturbations but the RAT and SDR methods were distinctly different and generally had larger magnitudes. Additionally, as shown in Table 5, which shows the average relative perturbations for detail item 310, the magnitude of the RAT and SDR perturbations is a lot larger than the magnitude of the RAK and WSD methods. The WSD average relative perturbations are slightly higher and more variable than the RAK average relative perturbations.

Table 5: Raking Scenario - Average Relative Perturbation Change QFR Short Form 2012Q4-2015Q3Detail Items 310

Statistical Period	SDR	RAT	WSD	RAK
2012Q4	29.50	42.82	23.83	12.59
2013Q1	107.15	109.58	16.49	21.15
2013Q2	390.96	-2.53	0.55	0.34
2013Q3	127.24	123.15	42.09	29.64
2013Q4	2681.31	2688.57	19.39	6.96
2014Q1	78.38	63.51	9.72	8.76
2014Q2	47.29	47.69	29.39	19.64
2014Q3	542.21	229.39	11.48	13.87

2014Q4	547.58	58.07	13.29	15.53
2015Q1	96.38	90.33	7.75	8.92
2015Q2	16.68	9.08	1.37	5.37
2015Q3	65.36	68.18	2.00	0.95

We created box plots summaries for detail item 310 of the RAK and WSD methods average relative perturbations, allowing us to visualize the difference between the two methods. Figure 14 presents these box plots and it can be seen that there is a good deal of overlap between the average relative perturbations.

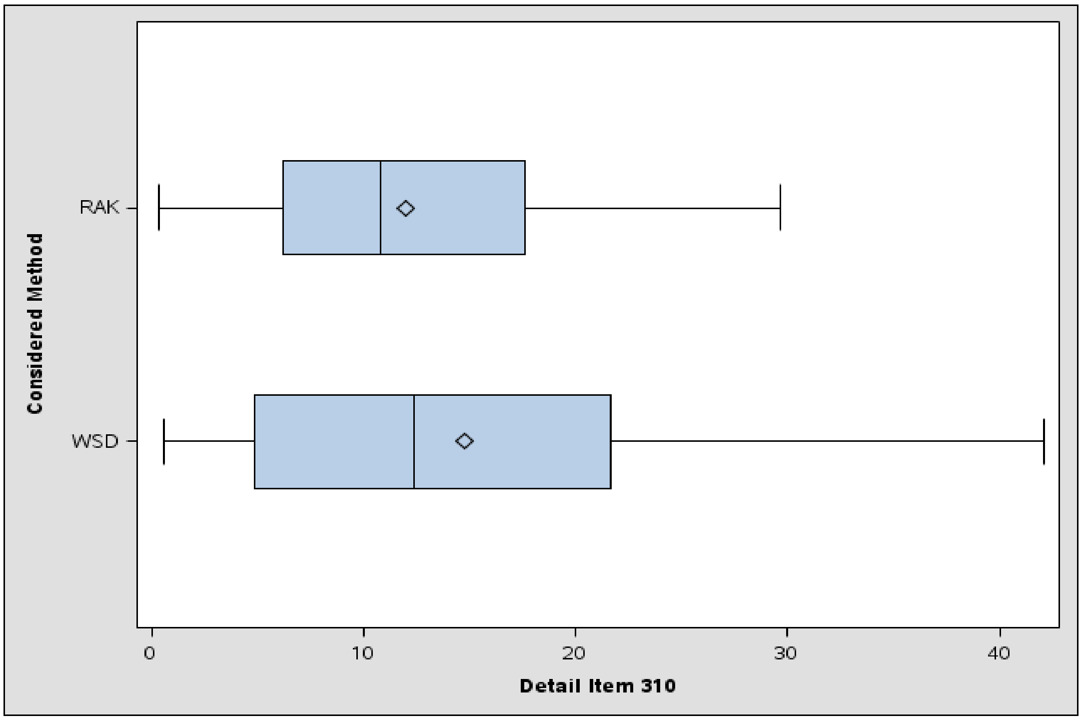

Figure 14:Raking Scenario - Boxplots of the Average Relative Perturbation Change Short QFR Short Form 2012Q4-2015Q3Items 310

Looking at similar box plots comparing only the RAK and WSD methods, we did find one detail item, item 326, where the average relative perturbations were drastically different. Figure 15 shows this difference. The WSD average relative perturbations are of a much larger magnitude than the RAK method. However, this is expected and not cause for concern because item 326 can take on negative values and the two methods differ in how they treat negative values.

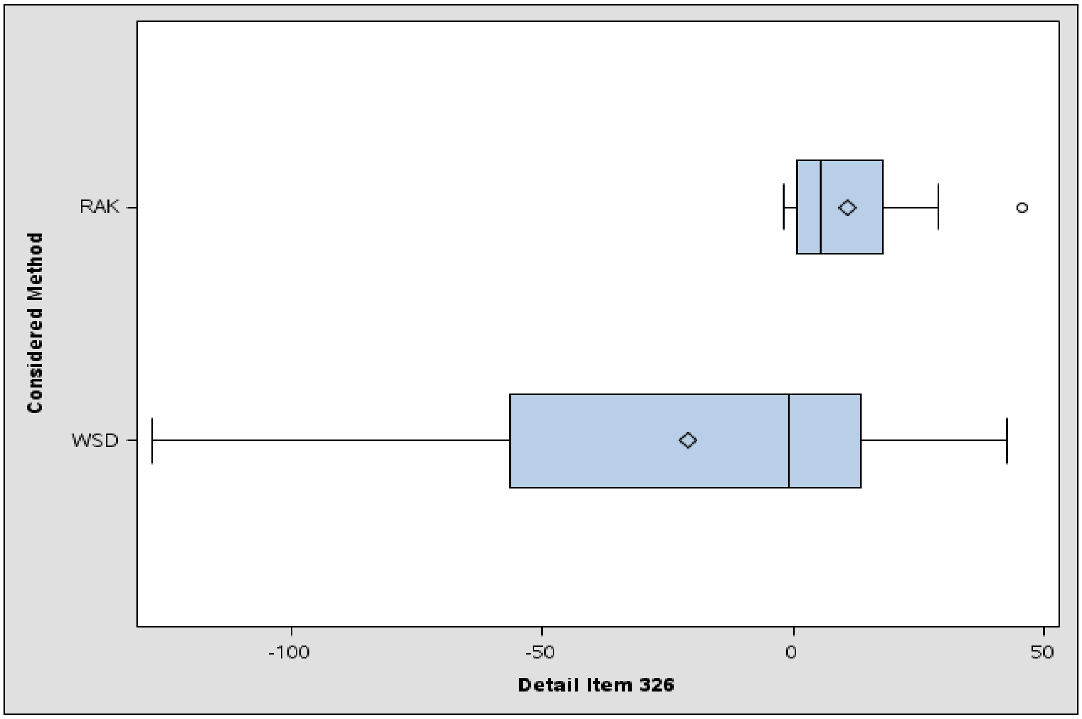

Figure 15 Raking Scenario - Boxplots of the Average Relative Perturbation Change QFR Short Form 2012Q4-2015Q3Detail Items 326:

Looking at the change frequency percentages and the average relative perturbations it does not appear that WSD method would be a severe departure from the currently implemented RAK method. It appears that the two methods are likely to provide comparable estimates of totals. To this, we looked at the weighted totals across observations that had out-of-balance complexes for all of the applicable methods to determine how well they tracked with the raking weighted total.

Again looking at detail item 310, we present the weighted totals of all observations with out-of balance complexes in Figure 16. Here it appears that the WSD method tracks nearly identical to the RAK method, suggesting that the change to the totals with the WSD method would be minimal.

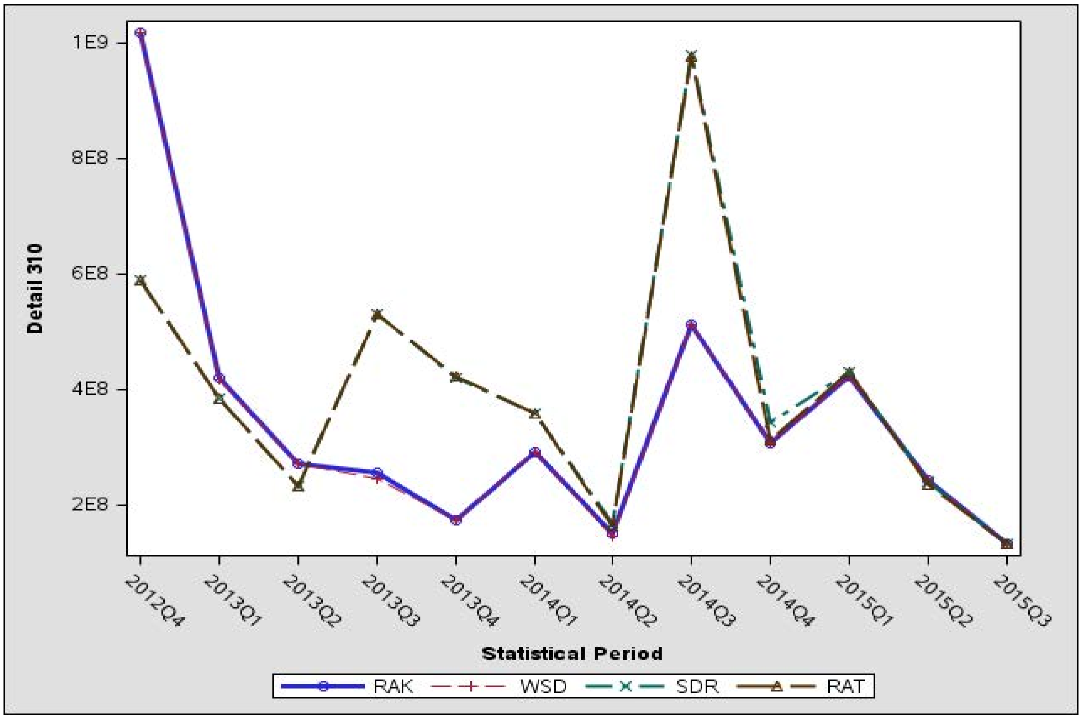

Figure 16:Raking Scenario – Plots of Weighted Totals for QFR Short Form 2012Q4-2015Q3Detail Items 310

There was one instance where the WSD and RAK methods totals did not appear to be identical and that was for item 322 shown in Figure 17. This is an item that can be negatively valued, thus we anticipate differences but the differences do not look severe.

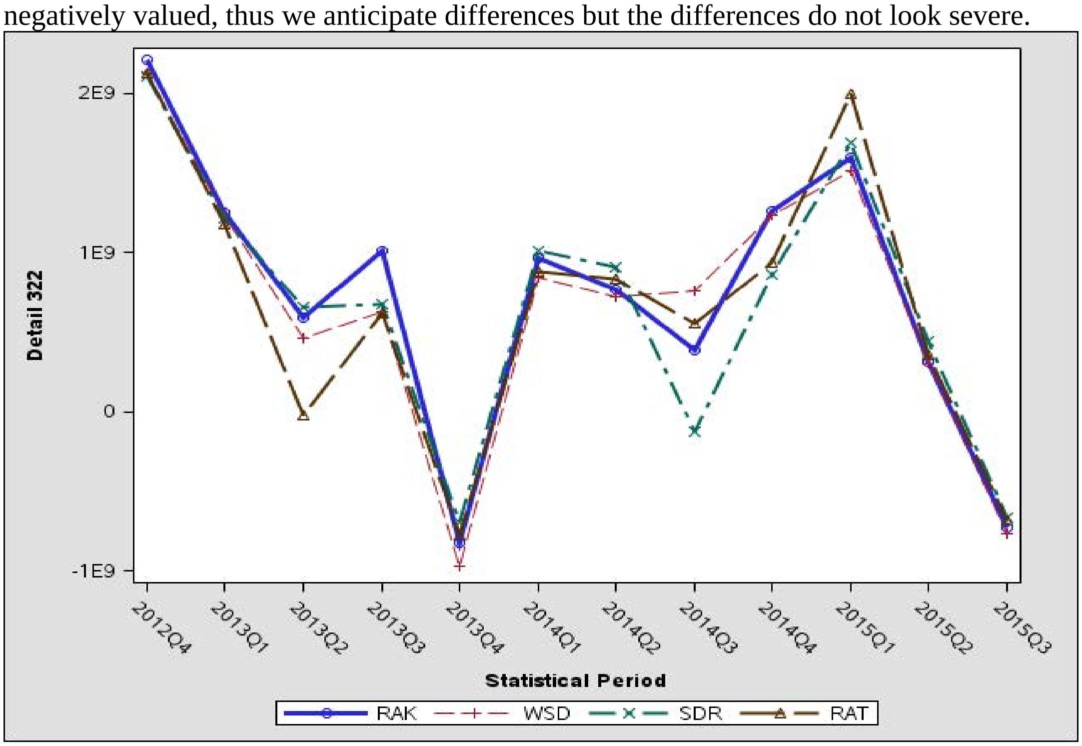

Figure 17:Raking Scenario – Plots of Weighted Totals for QFR Short Form 2012Q4-2015Q3Detail Items 322

5. Conclusion

At the beginning of this project we sought to find a method to resolve out-of-balance complexes that could contain real valued data items. However, as the project continued more requirements were provided to us and two different scenarios emerged. In the first scenario, we were attempting to match a procedure that has a human element to it and is therefore subject to a random analyst error. In the second scenario, we were trying to match a sometimes inconsistent or erroneous automated procedure. Of all of the methods that we considered, only the WSD method could be applied to both scenarios easily. This is particularly appealing, because we have found an automated procedure for adjusting failing balance complexes that is flexible enough that it can work with the constraints provided for both scenarios.

Narrowing down the method was just the beginning. There is still much work to be done regarding implementation and setting up the parameters so that transitioning to this new method is as seamless as possible. This includes adjusting the costs to more appropriately adjust the rarely change items. This also includes looking into why the WSD method can result in lower weighted totals than the RAK method does. Finally, we would like to continue this analysis on additional balance complexes like Assets (C223) from the QFR and balance complexes from other economic surveys.

Acknowledgements
The authors thank Katherine Jenny Thompson, James Hunt, Adam Fowler, and Brandi Maxson for their careful review and useful comments on earlier versions of this manuscript.

REFERENCES

Deming, W.E. 1943, Statistical Adjustment of Data. New York: Wiley.

Eltinge, J.,Kozlow, R. and Luery, D. 2003. "Imputation in Three Federal Statistical Agencies." Presentation at the 2003 Federal Economic Statistics Advisory Committee (FESAC).

Luery, D. and Sigman, R. 2000. "Raking When the Details are Positive and Negative." US Census Bureau unpublished manuscript, Washington DC.

Sigman, R. and Wagner, D. 1987. "Algorithms for Adjusting Survey Data that Fail Balance Edits." *Proceedings of the Section on Survey Methods*: American Statistical Association.