Revisions, Revisited: Data-Driven Approaches for Detection in Quarterly Financial Report Macro-Level Data

Laura Bechtel[1], Gregory Cepluch[1], Melissa McDaniel[1]
[1]U.S. Census Bureau, 4600 Silver Hill Rd, Washington, DC 20233

Abstract
In this paper, the Quarterly Financial Report (QFR) program investigates statistical methods for identifying substantial macro-level revisions of accounting-style data. Currently, macro-level relative revisions are identified as suspect if the absolute values of the differences are above a defined threshold, which is determined by subject matter expertise. Control charts are explored as a method to detect substantial revisions in the data. The inputs necessary for these control charts are not readily available for revision estimates. As a result, a focus is placed upon estimating the control chart parameters via the use of Generalized Variance Functions. Once these parameters are developed, various evaluation diagnostics are employed to assess their validity. Finally, the performances of the new and existing revision identification methodologies are compared.

Key Words: Revisions, Quarterly Financial Report, Process Control

1. Introduction

The U.S. Census Bureau[1] Quarterly Financial Report (QFR) is a Principal Federal Economic Indicator survey of U.S. corporations operating in the manufacturing, mining, wholesale trade, retail trade, and selected service industries. The QFR collects income statement (made up of line items such as sales, net income, etc.) and balance sheet (made up of line items such as cash, total liabilities, etc.) data from each surveyed company. These data are aggregated by industry and asset size, and published quarterly. Several key economic statistics (including percentage change in sales, total quarterly sales and total net income after taxes) are reported quarterly by QFR in two separate press releases: Retail Trade (RET) and Manufacturing, Mining, Wholesale Trade and Selected Service Industries (MMWS). The QFR data are initially released approximately 75 days after the end of each quarter. In addition, the QFR estimates of the prior four quarters of data are revised to incorporate changes due to late submissions and corrections.

In the event that a large disparity exists between the current and previous releases of an aggregated total, an investigation is initiated to ascertain its cause. Subject matter experts refer to a revision requiring such an investigation as a “substantial revision.” We adopt this terminology throughout the paper. For each particular industry/asset class where a substantial revision(s) was made, investigations involve delving into the micro-level data and identifying the particular companies whose survey submissions were revised. Due to the time-consuming nature of this process, it is desired that revisions flagged for investigation are in fact substantial. The current method to identify substantial revisions

[1] Any views expressed on statistical or methodological issues are those of the authors and not necessarily those of the U.S. Census Bureau.

flags a line item within an industry/asset class if the absolute value of the difference between the currently published aggregate value and the previously published aggregate value is above a threshold determined by subject matter experts. This threshold is constant across all items and industries, varying only for the retail trade sector.

To meet the needs of the QFR's data users, the survey is expanding to include more industries. The program currently used by analysts to identify substantial revisions needs to be updated to accommodate the addition of these new industries and any future modifications that may be necessary. The redesign of the program used to identify substantial revisions, along with the need to identify thresholds pertaining to the newly included industries, motivated the decision to investigate alternate statistically based approaches. While the existing methodology is adequate for the currently collected data, it is desirable to develop a statistical method for detecting these differences that does not rely as heavily on subject matter expertise.

2. The Quarterly Financial Report

The QFR is a voluminous publication. At present, there are approximately 140 industry-asset level combinations presented in the publication. For each of those individual combinations, there is both an income statement and a balance sheet data table. Some manufacturing industries' data (those with total assets of $25 million and over) are published in long format tables, consisting of 49 balance sheet items and 21 income statement items. All others are published in short format tables consisting of fewer line items. Within each industry/asset class data table, at the cell-level, the estimate pertains to the line item (these vary according to publication format as described above). Publication tables present cell-level estimates for the current quarter and the preceding revised four quarters (referred to as the five quarter spread). This leads to approximately 38,000 cells of estimates in each publication. Throughout the paper, we represent these cell-level estimates as $X_{c,q,r}$ where c is the industry/asset class by line item (we refer to this as the cell-level), q is the year/quarter statistical period, and r is the revision of the estimate. For the preliminary release of an estimate, r is zero, for the first revision of the estimate r is one, and for the fourth and final (published) revision of the estimate r is four.

Due to data constraints, the domain of our research is confined to the manufacturing, mining, and wholesale trade (MMW) industries of the QFR spanning the statistical periods from fourth quarter 2006 (2006Q4) through third quarter 2011 (2011Q3). Within our specified domain of analysis, the MMW sample is comprised of approximately 2,000 large corporations that are included in the survey with certainty and approximately 6,500 small and medium corporations, which are randomly selected for inclusion in the survey.

More details concerning the QFR survey design, methodology, and data limitations are available online in the source and accuracy statement of any publication table (http://www.census.gov/csd/qfr/).

3. Statistical Process Control Framework

The statistical process control (SPC) framework assumes that there exists a predefined process with some natural variation due to chance. In this context, a process is defined as a system with a set of inputs and an output (Montgomery, 2009, p. 13). The goal of SPC is to identify fluctuations in the process that are larger than expected given a stable random process or that are exhibiting a monotonic trend, thus indicating the process is out of control. In other words, the observed variation cannot be attributed to “chance” but instead is due to an “assignable cause.”

The control chart is one of the primary SPC tools used for monitoring whether a process is in control (Montgomery, 2009). The basics of the control chart can be seen below in Figure 1, which presents a fictitious in-control process. Here we have a time series plot, a centerline, and upper and lower control limits. The process fluctuates around the centerline in a random fashion within the control limits.

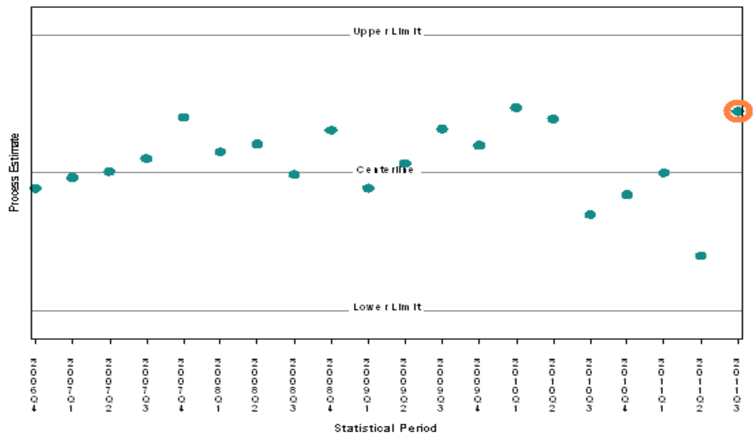

Figure 1: Fictitious Control Chart Example

In the manufacturing setting where SPC gained its popularity and is primarily used, the process is drawn from a single parametric distribution with known (or estimable) parameters. The process would be identified as out-of-control if one of the points lies outside of the control limits or if systematic behavior was identified within the control limits. An example of systematic behavior would be five or more observations consecutively appearing above or below the centerline. Once a process is identified as out-of-control, the goal is to fix the process or to find an assignable cause to return it to its in-control state. In our context, the focus is on the **final** point in the time series (circled in orange in Figure 1), which will be investigated by analysts if it falls outside of the control chart limits. For our purposes, trend detection is not considered in the review process.

We deviate from the traditional approach of estimating the control chart parameters by employing rolling averages instead of fixed limits based on historical data. Even though we are employing rolling averages, the goal is to attain something similar to a traditional control chart with stable control limits that are not sensitive to minor fluctuations in the data. Each statistical period, the control limits will change when using a rolling average even when the studied process is stable. Consequently, our rolling average uses a series of at least five adjacent quarters, allowing the limits to change only when there is a long-term shift. This approach contrasts with the rolling median used to develop control charts for monitoring response rates, where the goal is to have the control limits change as quickly as possible to reflect changes in the overall process (González and Oliver, 2012).

4. Methodology for Flagging Substantial Revisions

4.1 Current Methodology

By considering the revisions in an SPC framework, we are assuming that the difference between revisions is a process (the revision process) that has some expected random variation. Currently, a substantial revision is flagged if the absolute value of the difference between two adjacent revisions of a cell-level estimate is greater than a predetermined threshold k, i.e. where $|D_{c,q,r}| > k$ where $D_{c,q,r} = X_{c,q,r} - X_{c,q,r-1}$.

The QFR analysts investigate cell-level estimates that are flagged as substantial revisions. The primary purposes of an investigation are to establish that a given revision process was done correctly and to substantiate further questions regarding that revision. In the case of a major error – an error at the corporation level that would noticeably affect the publication totals – a correction will be made before the data are released. Generally, less than one percent of flagged revisions are actually major errors that result in immediate correction. Any minor corrections will be reflected in the revised estimate published the following quarter.

Before updating the methodology for identifying substantial revisions, there are several intricacies about the current method to consider. First, the threshold used for comparison in the current procedure is constant across all cells, line items, and revisions. Economic indicator programs such as the QFR are required to be able to explain the causes of large revisions. Using a single threshold allows the analysts to target the larger valued cell-level estimates that would be most likely to cause large revisions. Second, the subject matter experts only look at flagged substantial revisions for independent line items; that is, items that are not composed of sums and/or differences of other items. Last, flagging absolute differences using a single threshold implies that the subject matter experts believe the revisions are centered around zero. An observed centerline that is not close to zero might indicate that something systematic is going on either with late reporters or the imputation methodology.

4.2 Proposed Methodology

We propose applying control chart methodology to identify substantial revisions. Following the existing analyst review procedures, we will only attempt to identify an out-of-control process by looking at the final point in a presented time series to determine if it falls outside of the control limits. In the SPC framework, the identified flagged points would be investigated to determine the assignable cause of the out-of-control process, which in turn would be used to return the process back to an in-control state. However, our goal of flagging the substantial revisions is not necessarily to correct them. The control charts that will be produced will be analyzed so that if a particular process is considered out-of-control it will be investigated further; changing the offending figure only if it is found to be an error. The current procedure can be expressed as a control chart with a constant centerline of zero and constant (fixed) control limits. A goal of our research is to develop control charts with similar characteristics, but to use parameters that are specific to the estimates.

4.2.1 Developing Control Charts for the Revision Process

Generally, the control chart centerline and upper and lower bounds use the mean and standard deviation of the statistic of interest. In our case, the statistic of interest is the difference between adjacent revisions, $D_{c,q,r} = X_{c,q,r} - X_{c,q,r-1}$. The components of the control chart can be defined as:

- Lower Limit = $\hat{\mu}_{D_{c,q,r}} - L\hat{\sigma}_{D_{c,q,r}}$

- Centerline $=\hat{\mu}_{D_{c,q,r}}$
- Upper Limit $=\hat{\mu}_{D_{c,q,r}} + L\hat{\sigma}_{D_{c,q,r}}$

where $\hat{\mu}_{D_{c,q,r}}$ estimates $\mu_{D_{c,q,r}}$ (the process average), $\hat{\sigma}_{D_{c,q,r}}$ estimates $\sigma_{D_{c,q,r}}$ (the process standard deviation), and $D_{c,q,r} \sim N(\mu_{D_{c,q,r}}, \sigma^2{}_{D_{c,q,r}})$. L is commonly assigned a value of three ("3-sigma limits") so that approximately 99.7 percent of the process points would be expected to fall within the upper and lower limits (Montgomery, 2009, pp. 183-184). Our process points are likely not normally distributed and therefore are not expected to fall randomly about the (normal theory derived) centerline. In future applications, we may substitute an alternative value of L as a critical value, perhaps selecting bootstrapped percentiles in place of the parametric limits.

Initially, we considered three options for estimating the centerline:

$\overline{D}_{c,q,r}^{mean,t}$	the **mean** of the revision differences for item *x* over the *t* most recent statistical periods (t = 5, 8, 12, or 16)
$\overline{D}_{c,q,r}^{median,t}$	the **median** of the revision differences for item *x* over the *t* most recent statistical periods
0	the value used in the current procedure

Hereafter, we use an m superscript to indicate the averaging method used to obtain the rolling average (mean or median). In Section 5, we omit the results using $\overline{D}_{c,q,r}^{mean,t}$ or $\overline{D}_{c,q,r}^{median,t}$ centerline estimators because on average they were flagging many more substantial revisions when compared with the zero-valued centerline results.

Estimating the standard deviation of the revision differences is not quite so straightforward. We had three primary considerations:

1) Control chart theory assumes a constant process variance.
2) The QFR program does not compute or publish variance estimates for revisions, only select point estimates.
3) The QFR publishes a very large number estimates, each with varying reliability (variance of variance).

For reasons discussed in Section 4.4.2., we needed a smooth and stable variance estimator that could be used for *groups* of items. Our variance estimates needed to take into account the autocorrelation between revision estimates, which could vary by length of time between revisions. We expect that the variance for the *same* item could differ by revision (e.g., the final revised estimate *should* be less variable than the preliminary estimate), so it was necessary to develop variance estimates by item or item group and by revision.

We had delete-a-group jackknife estimates of variance for each of the point estimates, $X_{c,q,r}$: see Kott (2001) and Howe and Thompson (2005). For many of the line item totals, the replicate variance estimates are quite variable. A control chart whose limits vary greatly each time period is not useful for detection of out-of-control processes. To stabilize the variance estimates we developed Generalized Variance Functions (GVF) for sets of items, then used averaged point estimates as input to the variance models. We describe the model-development and validation procedure in Section 4.2.2 below.

After developing GVFs for groups of items, we estimate the variance of the revision difference as

$$\hat{\sigma}^{2\,m,t}_{\overline{D}_{c,q,r}} = f_{g,c,r}(\overline{X}^{m,t}_{c,q,r}) + f_{g,c,r-1}(\overline{X}^{m,t}_{c,q,r-1}) - 2\hat{\rho}\sqrt{f_{g,c,r}(\overline{X}^{m,t}_{c,q,r})(f_{g,c,r-1}(\overline{X}^{m,t}_{c,q,r-1})}$$

Where $f_{g,c,r}$ is the GVF associated with grouped items g for revision r, $\overline{X}^{m,t}_{c,q,r}$ is the rolling average estimate for item x obtained from t consecutive statistical periods (t = 5, 8, 12, or 16) and $\hat{\rho}$ is an assumed autocorrelation between revisions. We suspect that $\hat{\rho}$ is fairly high, but for a sensitivity analysis we looked at three levels: 0, 0.5, and 0.9.

4.2.2 Developing Estimates of Standard Deviation: Generalized Variance Functions

Generalized Variance Functions (GVFs) are simple mathematical expressions used to relate either the relative variance or variance to the expected value of an estimator. GVFs are used when it is not feasible to calculate direct variance estimates or the point estimates of variance are very unstable. Typically, they are used when a vast number of variance estimates are necessary.

Wolter (1985, pp. 205-206) outlines four main steps to the basic GVF procedure:

1. Group together all survey statistics that follow a common model.
2. Compute direct estimates of the variance (or relative variance) for several members of the group of statistics formed in Step 1.
3. Using the data from Step 2, compute estimates of the model parameters.
4. Obtain variance estimates for points whose direct variance estimates were not obtained in Step 2.

The first step is probably the most challenging. Examining the QFR data, we started by looking at necessary groupings and the available variances from which we could construct models. The first natural grouping for the QFR data was publication format: short vs. long. The second natural grouping was to classify items as either Balance Sheet or Income Statement items. Lastly, we characterize the items as real valued (e.g., income) or non-negative (e.g., sales). Table 1 presents the final GVF groups. Keeping in mind that we are conducting the GVF analysis on each revision of the estimate, we had 50 separate regression analyses to conduct.

Table 1: Final GVF Grouping used for each of the Five Revisions

Long Format / Short Format				
Income		Balance Sheet		
		Assets	Liabilities	
Real	Non-negative	Non-negative	Real	Non-negative

Step 2 was already performed for us. The QFR computes delete-a-group jackknife variance estimates.

After developing GVF groups, we needed to find a model to relate the relative variance or the variance and the expected value of the estimator. While constructing the data sets for modeling the GVFs, a couple of special circumstances came up. First, we removed the line items that the accountants do not look at when reviewing the substantial revisions. Typically, these line items are sums, differences, or ratios of other line items. Keeping them in our data sets would distort the relationship between the estimates and the variance estimates. Additionally, we needed to remove any cell-level estimates that had legitimate zero-valued variance estimates. This occurs in cells where most of the observations are selected with certainty and a small handful of cases are sampled, so that the rounded variance equals zero.

The first model we examined in each group g was $\sigma^2_{X_{c,q,r}} = \alpha_g + \beta_g / |X_{c,q,r}|$ for all $c \in g$, the "preferred" model in Wolter (1985, pp-202-204). However, this initial model was a very poor fit for the relationship between the cell-level estimates and their respective variance estimates. Valiant (1987) describes situations in which this model is appropriate, but these are not appropriate for the QFR. Therefore, we went back to the drawing board and conducted a simple linear regression using ordinary least squares (OLS) regression with $\sigma^2_{X_{c,q,r}} = \alpha_g + \beta_g * |X_{c,q,r}|$ for all $c \in g$. The R-squared value is very close to zero for a typical estimate group, and the root-MSE is very large, indicating that this model is a very poor fit for the data.

Scatter plots of the point estimates and their respective variance estimates indicated that the data needed to be transformed to reduce heteroscedasticity. Applying the natural log transformation seemed appropriate. Consequently, we applied the natural log transformation to the point estimate. We also transformed the variances by using the natural log transformation, so that the independent and dependent variables would be on the same scale. Next, we fit the following linear regression model (also recommended in Wolter, 1985, p. 203):

$$\ln(\sigma^2_{X_{c,q,r}}) = \alpha_g + \beta_g * \ln(|X_{c,q,r}|) \text{ for all } c \in g. \quad (1)$$

Applying the linear model (1) to the transformed data yielded vast improvements over the first model, although it is typical for there to be residual curvilinear pattern at the lower end of the distribution for a given group. Consequently, we evaluated the following transformations/ regression models:

$$\ln(\sigma^2_{X_{c,q,r}}) = \alpha_g + \beta_g * \ln^2(|X_{c,q,r}|) \text{ for all } c \in g \quad (2)$$

$$\ln(\sigma^2_{X_{c,q,r}}) = \alpha_g + \beta_g * \frac{1}{\ln(|X_{c,q,r}|)} \text{ for all } c \in g. \quad (3)$$

Models (2) and (3) did yield improved fits over model (1) within several of the groups. These slight improvements led us to consider adding additional terms to regression model (1) by evaluating $\ln\left(\sigma^2_{X_{c,q,r}}\right) = \alpha_g + \beta_{1g} * \ln\left(|X_{c,q,r}|\right) + \beta_{2g} * \ln^2\left(|X_{c,q,r}|\right)$ for all $c \in g$ as in Johnson and King (1987) and $\ln\left(\sigma^2_{X_{c,q,r}}\right) = \alpha_g + \beta_{1g} * \ln\left(|X_{c,q,r}|\right) + \beta_{2g} * \frac{1}{\ln\left(|X_{c,q,r}|\right)}$ for all $c \in g$. However, the additional terms really did not strengthen model (1). Note: we changed legitimate design-based variance estimates with the value of one to the value of 1.0001 so that the natural log transformation would not result in a value of zero.]

Although models (2) and (3) occasionally outperformed model (1), we liked the idea of using the same model for all groups for the sake of simplicity. A summary of the diagnostics for model (1) and the associated parameter estimates are shown in Table 2 below. The first column describes the grouping, the next two columns provide the parameter estimates that are closest to zero across all revisions, all of which were significant, and the last column provides the minimum R-squared value across all revisions. Root Mean Square Errors (MSE) are not presented here because they do not provide more information about the model than the R-squared values; they are available upon request. As you can see, we ended up with a wide range of R-squared values (falling as low as 0.49) that we could possibly improve upon using models (2) and (3) above, and/or by using weighted least squares (WLS) regression or iterative weighted least squares regression. These methods may be considered in future research.

Table 2: Summary of Final GVF Models

GVF Grouping	***$\hat{\alpha}_g$ Closest To Zero For All Revisions***	***$\hat{\beta}_g$ Closest To Zero For All Revisions***	***Smallest R-Squared For All Revisions***
Short Format Income Non-negative	1.77	1.44	0.72
Short Format Income Real-valued	5.38	1.14	0.58
Short Format Liabilities Real-valued	3.62	1.34	0.67
Short Format Liabilities Non-negative	8.88	0.94	0.49
Short Format Assets Non-negative	6.56	1.14	0.60
Long Format Income Non-negative	-4.84	1.72	0.73
Long Format Income Real-valued	0.63	1.32	0.51
Long Format Liabilities Real-valued	-4.17	1.84	0.78
Long Format Liabilities Non-negative	-3.71	1.82	0.75
Long Format Assets Non-negative	-4.18	1.86	0.88

Finally, we use the GVF parameters and average point estimates to obtain estimated variances for the revision differences. Using the parameter estimates presented in Table 2, we obtain cell-level variance estimates that correspond to the centerline estimate by applying the inverse transformation to the GVF estimates as follows: $\hat{\sigma}^2_{\bar{X}^{m,t}_{c,q,r}} = e^{(\hat{\alpha}_g + \hat{\beta}_g * \ln(|\bar{X}^{m,t}_{c,q,r}|)}$ for all $c \in g$.

We use the GVF variance estimates to derive the variance of the averaged differences as $\hat{\sigma}^2_{\bar{D}^{m,t}_{c,q,r}} = \hat{\sigma}^2_{\bar{X}^{m,t}_{c,q,r}} + \hat{\sigma}^2_{\bar{X}^{m,t}_{c,q,r-1}} - 2 * \hat{\rho} * \sqrt{\hat{\sigma}^2_{\bar{X}^{m,t}_{c,q,r}} * \hat{\sigma}^2_{\bar{X}^{m,t}_{c,q,r-1}}}$. Note that these variance estimates are **cell-level** variance estimates that use the same GVF model within in a GVF group. It might be possible to develop GVF expressions for the revision differences themselves. However, these expressions would assume that the estimate levels, the estimate coefficients of variation, and the autocorrelations remain stable over time. There is little evidence supporting these assumptions. First, revised estimates do change, and the direction can and does differ, and the last assumption is difficult to validate. The approach described here maximizes flexibility in the event that the GVFs do not appear to be correctly capturing variation for a given series.

4.2.3 Lower Limit, Centerline, and Upper Limit Estimates

Initially, we developed control charts for each cell-level revision difference. Cell-level control charts may flag too many cases, especially lower valued cell-level estimates, because the control limits are developed based on those estimates. In the case where cell-level control charts flag too many substantial revisions we considered grouping the control chart limits. Figure 2 shows how many control limits are generated for each of the different methods described in this paper. The top table in the figure shows the current method, which has one Upper Limit (UL), Centerline (CL), and Lower Limit (LL) for all cell-level estimates. Then we discuss cell-level limits in the second table, and it can be seen that there are *n* different sets of limits where *n* is the number of cell-level estimates (there are thousands). Finally, there are ten GVF group limits (for each revision difference), one set of limits for each of the groups we defined for the GVFs in Table 1.

Current Method			
Cell 1	Cell 2	...	Cell n
UL = k CL = 0 LL = -k	UL = k CL = 0 LL = -k	... ...	UL = k CL = 0 LL = -k

Cell-Level Control Charts			
Cell 1	Cell 2	...	Cell n
UL_1 CL_1 LL_1	UL_2 CL_2 LL_2	... ...	UL_n CL_n LL_n

Group-Level Control Charts								
Group 1				...	Group 10			
Cell 1	Cell 2	...	Cell n_1	...	Cell 1	Cell 2	...	Cell n_{10}
UL_1 CL_1 LL_1	UL_1 CL_1 LL_1	UL_1 CL_1 LL_1	UL_1 CL_1 LL_1	... ...	UL_{10} CL_{10} LL_{10}	UL_{10} CL_{10} LL_{10}	UL_{10} CL_{10} LL_{10}	UL_{10} CL_{10} LL_{10}

Figure 2: Representation of Upper and Lower Limits for Alternative Sets of Control Charts

To create these GVF group-level control limits, we borrowed an idea from a selective editing paper by Lawrence and McDavitt (1994). For selective editing, each individual item has a local score that measures its deviation from its expected value. Each unit will end up having multiple local scores – one for each item being edited. The local scores for each item are combined to create a single global score for the entire unit, which is compared to a critical value to determine if the unit contributes significantly to the total. Lawrence and McDavitt proposed taking the maximum of the local scores or averaging the local scores as a global score. To obtain our GVF group-level control limits, we considered four different functions to combine cell-level variance estimates into a single GVF grouped variance estimator:

- $Min(\hat{\sigma}^2_{\bar{D}^{m,t}_{c,q,r}})$ - minimum of all cell-level revision difference variances in GVF group *g*
- $Mean(\hat{\sigma}^2_{\bar{D}^{m,t}_{c,q,r}})$- simple mean of all cell-level revision difference variances in GVF group *g*
- $Median(\hat{\sigma}^2_{\bar{D}^{m,t}_{c,q,r}})$ - median of all cell-level revision difference variances in GVF group *g*
- $Max(\hat{\sigma}^2_{\bar{D}^{m,t}_{c,q,r}})$ - maximum of all cell-level revision difference variances in GVF group *g*.

Using the minimum variance estimate of the GVF group should yield the least conservative variance estimate (and the narrowest control limits), whereas the maximum yields the most conservative variance estimate (and the widest control limits). The mean variance estimate would account for the range of the variance estimate levels within a particular GVF group, but would be sensitive to large or small values. The median would compensate for this tail-estimate sensitivity.

5. Evaluation of the Control Charts Versus the Current Method

Our objective is to provide the analysts with a statistically based tool that will flag substantial revisions. To determine if the considered control chart methods could be used in place of the current method, we compare the results of the various methods by considering the substantial revisions flagged by the current method[2] as the "gold standard." Any cell flagged by the current method is a ***true substantial revision*** and any cell not flagged by the current method is a ***true non-substantial revision.***

For each grouping established for the GVFs, we prepared a table like Table 3 below to compare the performance of the control charts to the current method. Each row represents a control chart applied to the cell-level revision difference estimates, for the given values of $\hat{\rho}$ and *t*. A two-way classification table is presented in the fourth through sixth columns. The rows provide counts for the considered control chart method and the columns represent counts for the current method. We looked at the following measures, which have been modified from the definitions presented in Thompson and Sigman (1999):

- ***Type I Error Rate***: The proportion of cells that are true non-substantial revisions that are flagged as substantial by the control chart method.
- ***Type II Error Rate***: The proportion of cells that are true substantial revisions that are not flagged as substantial by the control chart method.
- ***Hit Rate***: The proportion of flagged revisions that are true substantial revisions.

Using our criteria, we are looking for a control chart method where the Type I and Type II error rates are close to zero and where the hit rates are close to one. Our analysis differs slightly from Thompson and Sigman (1999) in the sense that the same cell cannot be in more than one test. Therefore, it is possible to achieve Type I and Type II error rates of zero simultaneously.

Table 3 presents results aggregated across all cell-level control charts for one GVF group: Long Format Income Non-negative, flagging substantial revisions in the 2011Q3 statistical period. These results did vary slightly by GVF grouping, but the final takeaway was the same for all GVF groups. For this particular GVF group, Type I error rates are not close enough to zero: the control chart method is flagging at least double the number of cases currently being flagged. We omit results where averages are computed with $t = 5$ or 8, because the driving force behind our varying results was $\hat{\rho}$ and smaller values of *t* produced slightly less favorable results. The Type II error rates are far too high for all values of $\hat{\rho}$ and *t*. Moreover, the hit rates are too low to make this control chart method a viable option. In short, the cell-level control charts flag too many substantial revisions.

Note that the results in Table 3 are for cell-level control charts using a zero-valued centerline. To illustrate the discrepancies between the control charts and the current method seen in Table 3, we look at three different cell-level control charts from the Long Format Income Non-Negative group in statistical period 2011Q3.

[2] The publication data available to us had already been corrected for major errors. When looking at revisions flagged by the current method in this paper, we may be missing cases where these corrections have been made.

Table 3: Comparisons for the Long Format Income Non-negative Group with Zero-Valued Centerline

$\hat{\rho}$	*t*	***Control Chart: Substantial?***	***True Substantial Revision***	***True Non-substantial Revision***	***Type I Error***	***Type II Error***	***Hit Rate***
0	16	Yes	2	16	0.057	0.750	0.111
		No	6	264			
0.5	16	Yes	2	24	0.086	0.750	0.077
		No	6	256			
0.9	16	Yes	4	68	0.243	0.500	0.056
		No	4	212			
0	12	Yes	2	17	0.061	0.750	0.105
		No	6	263			
0.5	12	Yes	2	25	0.089	0.750	0.074
		No	6	255			
0.9	12	Yes	4	67	0.239	0.500	0.056
		No	4	213			

Figure 3 provides an example of when the control chart and the current method agree that the 2011Q3 (circled in orange) first revision was substantial, because it is outside both sets of bounds. Note that our control chart limits are narrower because they are derived from the cell-level data presented in the time series plot.

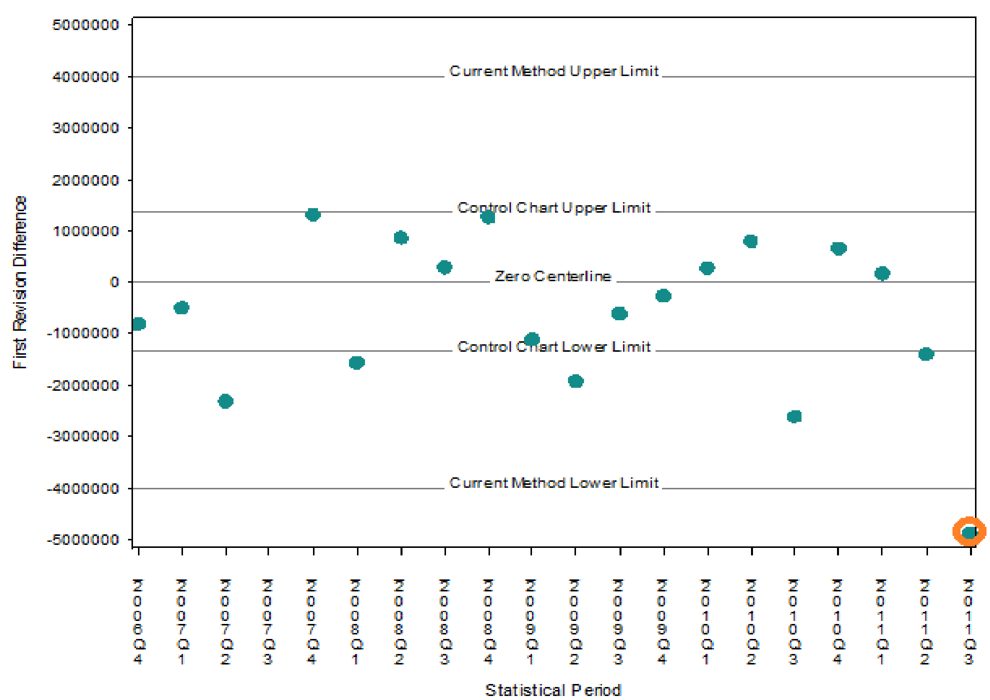

Figure 3: Current Method Limits and Control Chart Limits Where the Control Chart and the Current Method Flag a Substantial Revision

Figure 4 provides an example of when the control chart does not flag the 2011Q3 first revision as substantial and the current method does. Here it becomes more obvious that the cell-level control limits are data-driven. The cell-level control limits account for the large variability observed in this time series, while most of the data points are outside of current method control limits. This suggests that our GVFs are accurately representing the variability in the estimates.

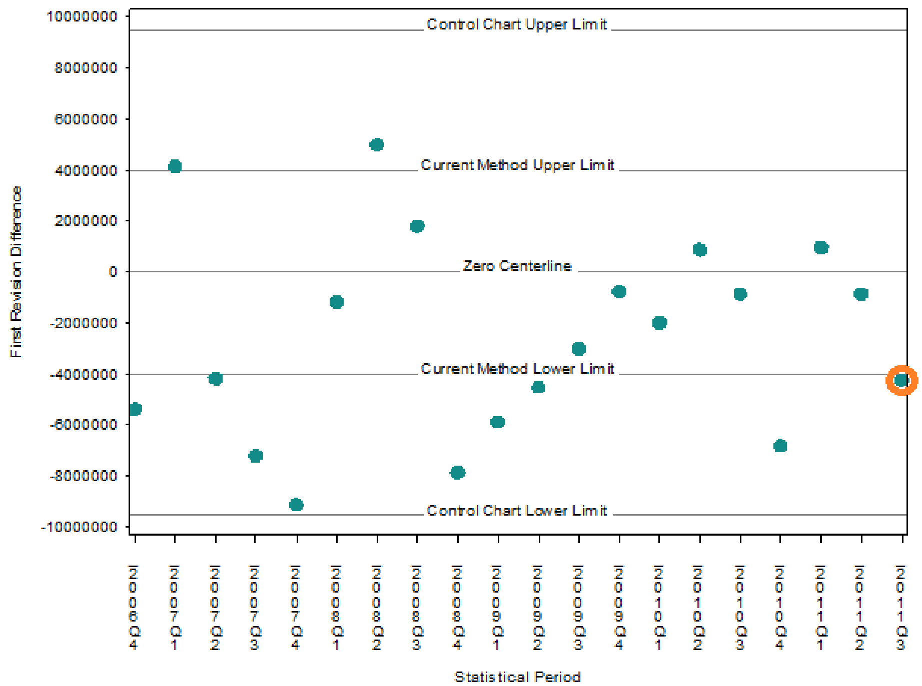

Figure 4: Control Chart Limits Overlaid with the Current Method Limits Where the Control Chart is Not Flagging a Currently Flagged Substantial Revision

Our final cell-level control chart in Figure 5 below displays the case where the control chart flags a substantial revision that is not currently being flagged. Looking at this control chart it can be easily seen this is one of the smaller cell-level estimates in the Long Format Income Non-negative group that we are less concerned with flagging. This indicated to us that our cell-level control charts may be too dynamic (i.e., too sensitive to the input data). When we looked deeper into our data, we noted that a lot of our over-flagging occurs with the smaller cell-level estimates, which tend to be more variable than their larger cell-level counterparts. To address the sensitivity to small estimates demonstrated by the cell-level control charts, we develop GVF group-level control charts.

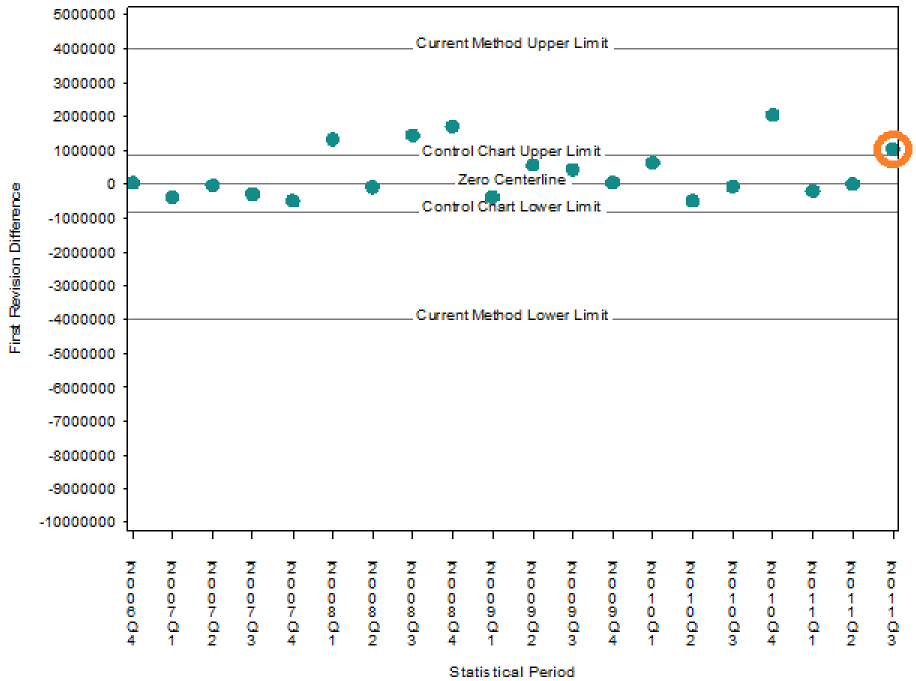

Figure 5: Control Chart Limits Overlaid with the Current Method Limits Where the Control Chart is Flagging a Substantial Revision that is not Currently Flagged

Table 4 below presents results that are analogous to the results we presented in Table 3except that the control limits used were obtained using the *Mean* GVF group-level variance estimates. Now, we are getting close to our ideal results: Row #4 displays Type I and Type II Error Rates of zero and a Hit Rate of one, indicating the results are exactly the same as the current method.

Table 4: Comparisons for the Long Format Income Non-negative Group with Zero-Valued Centerline and *Mean* GVF Group-Level Variance

$\hat{\rho}$	***T***	***Control Chart: Substantial?***	***True Substantial Revision***	***True Non-substantial Revision***	***Type I Error***	***Type II Error***	***Hit Rate***
0	16	Yes	5	0	0.000	0.375	1.000
		No	3	280			
0.5	16	Yes	8	4	0.014	0.000	0.667
		No	0	276			
0.9	16	Yes	8	6	0.021	0.000	0.571
		No	0	274			
0	12	Yes	8	0	0.000	0.000	1.000
		No	0	280			
0.5	12	Yes	8	4	0.014	0.000	0.667
		No	0	276			
0.9	12	Yes	8	6	0.021	0.000	0.571
		No	0	274			

Table 5: Functions That Match the GVF Group-Level Control Chart Method with the Current Method

GVF Grouping	***Best Function for GVF Grouping Variances***
Short Format Income Non-negative	Mean
Short Format Income Real-valued	Max
Short Format Liabilities Real-valued	Mean
Short Format Liabilities Non-negative	Max
Short Format Assets Non-negative	Mean
Long Format Income Non-negative	Mean
Long Format Income Real-valued	Max
Long Format Liabilities Real-valued	Between Median and Mean
Long Format Liabilities Non-negative	Between Median and Mean
Long Format Assets Non-negative	Between Median and Mean

We obtained similar results for a majority of the GVF groups (7 of the 10 GVF groups) using either the maximum or mean functions to combine the variance estimates. These results are summarized in Table 5. For the three GVF groups where the best function is "Between Median and Mean," the median function flagged many more substantial revisions than the current method, while the mean function did not flag as many. We believe that there may be a compromise between these two functions that will work well. Another possibility is that we are encountering retransformation bias caused by the log transformation of the point estimates, which could be remedied by following the steps outlined in Miller (1984); this will be left to future research.

6. Conclusion

While we still have a lot of work to do before we can hand the analysts a program to flag substantial revisions, we most certainly have found a method that can work. The GVF group-level control charts appear to be the best of both worlds. They are statistically based and data-driven, but they are stable enough that they do not flag too many substantial revisions. Pleasantly, using the GVF group-level control charts, we obtained the same or very similar results to the current method; confirming the intuition of the accountants who developed the current method.

Control charts provide important information about the process being considered. For example, while developing the appropriate control chart method, we quickly steered away from using the averaged (mean or median) revision differences to estimate the centerline, to using a zero-valued centerline. We changed our approach because we found the revision differences are *not* centered around zero. However, we expect the revision differences to be centered around zero if they are entirely attributable to random reporting changes. As mentioned earlier, this may indicate that something systematic is occurring with the late reporters, the imputation methodology, or a combination of both. We would like to investigate this phenomenon in the future.

In order to develop a production ready program, we need to conduct some further research. First, we would like to incorporate a fix to account for retransformation bias as seen in Miller (1984) which may give us better fitting models for our variance grouping functions. Additionally, it would be preferable to conduct our analysis over multiple statistical periods to ensure that our GVF group-level variance functions are consistent across time. Once we address these issues, it will be necessary to repeat this evaluation for the Retail portion of the QFR.

Finally, when we have completed all of this planned research we are confident that we will have a new method for identifying substantial revisions. The biggest advantage of the new method will be that it is data-driven and will objectively reflect the ever-changing economy.

Acknowledgements

The authors thank Demetria Hanna, James Hunt, Xijian Liu, and Katherine Jenny Thompson for their valuable contributions to this research project.

References

González, Y. and Oliver, B. (2012). Producing Control Charts to Monitor Response Rates for Business Surveys in the Economic Directorate of the U.S. Census Bureau. *Proceedings of the Federal Committee on Statistical Methodology Research Conference*.

Howe, A. and Thompson, K. J. (2005) Investigation of Alternative Variance Estimators for the Quarterly Financial Report. *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Johnson, E. G. and King, B. F. (1987), "Generalized Variance Functions for a Complex Sample Survey," *Journal of Official Statistics*, 3, 235-250.

Lawrence, D. and McDavitt, C. (1994). Significance Editing in the Australian Survey of Average Weekly Earnings. *Journal of Official Statistics*, **10**, pp. 437-447.

Kott, P. (2001). The Delete-A-Group Jackknife. *Journal of Official Statistics,* **17**, pp. 521-526.

Miller, D. (1984). Reducing transformation bias in cure fitting, The American Statistician, 38, pp. 124-126.

Montgomery, Douglas C. (2009) *Introduction to Statistical Quality Control*. John Wiley.

Thompson, K.J. and Sigman, R.S. (1999). Statistical Methods for Developing Ratio Edit Tolerances for Economic Data.
Journal of Official Statistics, **15**, pp. 517 – 535.

Valliant, R. (1987). Generalized variance functions in stratified two-stage sampling, JASA, 82. pp. 499-508.

Wolter, K. (1985). *Introduction to Variance Estimation*. New York: Springer-Verlag.