

Ratio Edits Based on Statistical Tolerance Intervals

Derek S. Young¹ and Thomas Mathew²

The role of statistical tolerance intervals for developing ratio edit tolerances in a parametric setup is investigated. The performance of the methodology is assessed for the normal and Weibull distributions. The numerical results show that in terms of Type I and Type II errors, statistical tolerance intervals exhibit better performance compared to other ratio edit procedures available in the literature. The methodology is illustrated using 2010 and 2011 data from the Annual Survey of Manufacturers.

Key words: Outliers; resistant; robust; tolerance limits; trimming; Winsorization.

1. Introduction

Ratio edit tolerances are bounds used for identifying errors in the data obtained by Economic Census Programs so that they can be flagged for further review. The tolerances represent upper and lower bounds on the ratio of two highly correlated items and are used for outlier detection; that is, to identify units that are inconsistent with the rest of the data. Some texts dedicated to the general topic of outlier detection include [Barnett and Lewis \(1994\)](#), [Rousseeuw and Leroy \(2003\)](#), and [Aggarwal \(2013\)](#). A number of outlier detection methods are also available in the literature and can be used for developing ratio edit tolerances; we refer to [Thompson and Sigman \(1999\)](#) and [Rais \(2008\)](#) for a review and comparison of these methods as they apply to the ratio edit problem. [Thompson and Sigman \(1999\)](#) compared different methods for generating ratio edit tolerances, which focused on “Type I” and “Type II” errors. A Type I error flags a ratio value as inconsistent or wrong when it is not so. A Type II error flags an inconsistent ratio as consistent or correct. [Thompson and Sigman \(1999\)](#) recommended a stepwise approach for developing ratio edit tolerances, while [Thompson and Adeshiyan \(2003\)](#) discussed the effects of ratio edit and imputation procedures on data quality for the 1997 Economic Census. Both articles also emphasized the importance of incorporating subject-matter expertise when developing the ratio edits.

¹ Department of Statistics, University of Kentucky, 725 Rose Street, Lexington, KY 40536, USA. Email: derek.young@uky.edu

² Department of Mathematics and Statistics, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, U.S.A. Email: thomas.mathew@census.gov

Acknowledgments: This research was conducted while the first author was a Research Mathematical Statistician in the Center for Statistical Research and Methodology, U.S. Census Bureau. The authors wish to thank Jenny Thompson and Eric Slud of the U.S. Census Bureau, three anonymous referees, and the Associate Editor for their numerous helpful comments on this paper. This report is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

The issue of outliers or large data values in surveys has been addressed in the literature. [Tambay \(1988\)](#) presents an empirical study comparing various methodologies for identifying level outliers and/or trend outliers in subannual economic surveys. [Latouche and Berthelot \(1992\)](#) focus on respondent follow-ups to units that may have an important effect on statistical estimates. The authors present and compare three score functions as a way to identify suspicious units according to their potential effect on the estimates. [Kokic and Bell \(1994\)](#) discuss the setting where a number of unusually large observations fall in the survey sample, which may grossly overestimate population totals. They proceed to specify a cutoff criterion so that an optimal level can be found for Winsorizing the data. As discussed in [Rivest and Hidiroglou \(2004\)](#), Winsorization is widely used to curb the effect of outliers when computing survey estimates. Winsorized estimates have a downward bias and smaller variance relative to their non-Winsorized analogues. When aggregating survey estimates, these effects result in larger biases and less precision than standard aggregated estimates. Hence, [Rivest and Hidiroglou \(2004\)](#) propose using a “corrected” Winsorized estimate.

While not investigated here, we note a few other novel outlier detection methods that could be investigated for performing ratio edits. [Hido et al. \(2011\)](#) present an approach to identify outliers in a test dataset based on a training dataset comprised solely of inliers, which is accomplished by using the ratio of the two dataset densities as an outlier score. [Yuen and Mu \(2012\)](#) use a Bayesian linear regression setup to compute probabilities that an observation is an outlier. Finally, [Chawla and Gionis \(2013\)](#) present a generalization to the k -means algorithm as a way to simultaneously cluster and discover outliers in a dataset.

The purpose of our investigation is to examine the role of statistical tolerance intervals in the process of developing ratio edit tolerances. A statistical tolerance interval provides bounds that will capture a specified proportion or more of a sampled population with a given confidence level; we refer to the book by [Krishnamoorthy and Mathew \(2009\)](#) for a detailed discussion of the topic. Since ratio edit tolerances provide a range for the acceptable ratios, a statistical tolerance interval can do the same provided that such an interval is constructed using the good ratios; that is, using the data after deleting the ratios that are inconsistent or problematic. An advantage of using a statistical tolerance interval is that such an interval, by construction, controls the Type I error at a specified level, similar to what is done in hypothesis testing. The Type II error performance can then be studied and compared with other ratio edit tolerance intervals available in the literature, as described in [Thompson and Sigman \(1999\)](#).

Our approach consists of computing statistical tolerance intervals based on the “good” part of the data; that is, after trimming the data so that potentially bad ratios are excluded from the tolerance interval computation. We have no clear guidance on the percentage of trimming to be done, which should perhaps be done using the input of a subject-matter expert. In the case of a nearly symmetric distribution, we recommend trimming both tails of the distribution, unless there is reason to believe that the contamination is only in one tail. We report numerical results for a two-sided tolerance interval for the case of a normal distribution, computed after trimming both tails. Type I and Type II error probabilities are reported and compared with the ratio edit tolerances available in the literature. We also report results for a one-sided upper statistical tolerance limit for the case of a Weibull distribution, computed after trimming is done only in the right tail. The overall conclusion

is that the statistical tolerance interval approach has a considerable edge over the available ratio edit tolerances in terms of controlling Type I and Type II error probabilities. Furthermore, for several standard parametric distributions (including the normal and Weibull distributions considered in our work), analytic expressions or accurate approximations are available for the limits that define a statistical tolerance interval. In other words, they are easy to compute and we refer to [Krishnamoorthy and Mathew \(2009\)](#) for further details.

Before describing the methodology for computing a statistical tolerance interval, we want to make a brief comment on the terminology used in this article and in the literature. As already noted, ratio edit tolerances are thresholds used for identifying ratio edit failures and are determined through a wide range of possible outlier detection methods; however, they are not defined or determined using the same criteria that define a statistical tolerance interval. On the other hand, statistical tolerance limits are bounds that capture at least a specified proportion of the sampled population with a given confidence level. Since both notions are traditionally referred to as “tolerance limits,” we will make it clear through the context which type of “tolerance” is being discussed.

We begin our discussion with a review of outlier detection methods that are used for ratio edits and then investigate the role of statistical tolerance intervals for the same.

2. Outlier Detection Methods for Ratio Edits

There are numerous procedures for outlier detection in the literature; for example, see the texts by [Iglewicz and Hoaglin \(1993\)](#), [Barnett and Lewis \(1994\)](#), and [Rousseeuw and Leroy \(2003\)](#). The focus of this study is not to provide an exhaustive comparison of those procedures, but rather to compare our approach with the standard methods used in setting ratio edit tolerances. In this section, we discuss three common approaches that have been employed by the U.S. Census Bureau.

2.1. Robust Control Limits

[Shewhart \(1939\)](#) provided the first thorough treatment of control charts as a way to monitor a quality characteristic of a process over time. Control charts (also called Shewhart charts) are a simple, yet powerful way to visualize variability in a process. They can be used to identify shifts in a process or when a process goes out of control, where this latter setting is essentially an outlier detection problem. The outliers are identified by placing control limits on the data. Let μ_T and σ_T denote the mean and standard deviation, respectively, of a statistic of interest $T \equiv T(X)$ for the process being monitored. Then lower and upper control limits are given by $\mu_X - L\sigma_X$ and $\mu_X + L\sigma_X$, respectively. Here, L controls how far one will allow the process to vary from the mean before determining that it has gone “out of control.” Typically, we set $L = 3$, which is the 3σ -limit rule of thumb often used for outlier detection. A more contemporary treatment of control chart methodology can be found in [Montgomery \(2013\)](#).

While ratio data is usually not time ordered (even though the ratios themselves may be constructed using the same variable measured at two different time points), we can still apply a similar type of control limit methodology. As discussed in [Thompson and Sigman \(1999\)](#), we can use robust estimates of the population mean and standard deviation to

construct control limits, which in turn will be the ratio edit tolerances. The robust estimates are based on trimming and Winsorizing, which we now describe in more detail for any general univariate setting.

Suppose we have observed data x_1, \dots, x_n and let $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ denote the ordered data. The (symmetric) α -trimmed mean for the data is given by

$$\bar{x}_\alpha = \frac{1}{n - 2 \lceil \alpha n \rceil} \sum_{i=\lceil \alpha n \rceil+1}^{n-\lceil \alpha n \rceil} x_{(i)}, \quad (1)$$

where $\lceil \cdot \rceil$ is the ceiling function and $0 < \alpha < 1$. As noted in [Tukey and McLaughlin \(1963\)](#), the Winsorized variance is a consistent estimator of the variance of (1). The Winsorized variance is given by

$$s_{W_\alpha}^2 = \frac{1}{n - 2 \lceil \alpha n \rceil} \sum_{i=\lceil \alpha n \rceil+1}^{n-\lceil \alpha n \rceil} (x_{(i)} - \bar{x}_{W_\alpha})^2, \quad (2)$$

where

$$\bar{x}_{W_\alpha} = \frac{1}{n} \left(\sum_{i=\lceil \alpha n \rceil+1}^{n-\lceil \alpha n \rceil} x_{(i)} + \lceil \alpha n \rceil (x_{(\lceil \alpha n \rceil+1)} + x_{(n-\lceil \alpha n \rceil)}) \right) \quad (3)$$

is the Winsorized mean. It is easy to modify the above formulas to accommodate asymmetric trimming and Winsorizing, which includes one-sided trimming and Winsorizing as special cases. Finally, the interval based on robust control limits is given by

$$(\bar{x}_\alpha - Ls_{W_\alpha}, \bar{x}_\alpha + Ls_{W_\alpha}). \quad (4)$$

For ratio data, [Thompson and Sigman \(1999\)](#) use $L = 2$ to set a more liberal rule and $L = 3$ to set a more conservative rule regarding the number of cases flagged for review.

Many robust measures of location and scale could be investigated to construct analogues to the robust control limits in Equation (4). For example, one might simply consider the median or an M -estimator for a robust estimate of location, while the median absolute deviation or Gini's mean difference could be used for a robust estimate of scale. These may result in more informative limits for a particular application. However, our focus is on comparing some of the more common methods used in setting ratio edit tolerances (e.g., Equation (4)) with the tolerance interval approach that we discuss in Section 3.

2.2 Fence-Based Methods

In exploratory data analysis, the interquartile range (IQR) can be used to identify potential outliers in a univariate dataset. The IQR is a resistant measure of dispersion defined as $Q_3 - Q_1$, where Q_1 and Q_3 are the first and third quartiles, respectively. As discussed in [Hoaglin et al. \(1986\)](#), the resistant rule flags values as outliers if they fall outside the interval

$$(Q_1 - k\text{IQR}, Q_3 + k\text{IQR}), \quad (5)$$

for some non-negative constant k . Thompson and Sigman (1999) studied the use of Equation (5) as a way to set ratio edit tolerances and referred to the above rule as *resistant fences*. They referred to the specific rules of setting the values of k equal to 1.5, 2.0, and 3.0 as inner, middle, and outer fences, respectively. Note that the inner-fences rule is almost always employed when identifying univariate outliers on a boxplot.

Thompson (1999) explored a variation of resistant fences for asymmetric distributions. *Asymmetric fences* are elongated in the direction of the skewness of the distribution. Denoting the median by \tilde{x} , the asymmetric-fences method replaces the IQR in Equation (5) with distances from \tilde{x} . Specifically, the asymmetric-fences rule flags values as outliers if they fall outside the interval

$$(Q_1 - k^*(\tilde{x} - Q_1), Q_3 + k^*(Q_3 - \tilde{x})). \quad (6)$$

For asymmetric fences, Thompson (1999) refers to values of k^* equal to 3.0, 4.0, and 6.0 as inner, middle, and outer fences. Note that these rules are just twice the value of k used for the resistant-fences rule.

2.3 Hidiroglou-Berthelot Method

The methodology introduced by Hidiroglou and Berthelot (1986) is a ratio edit procedure that uses a *centering transformation* of the ratios followed by a *magnitude transformation*. Here is a brief description of the procedure.

Let $(x_1, y_1), \dots, (x_n, y_n)$ be observations of the variables of interest and $r_i = x_i/y_i$, $i = 1, \dots, n$ denote the n ratios to be analyzed. Moreover, let \tilde{r} denote the median of the ratios. Define

$$s_i = \begin{cases} (r_i/\tilde{r}) - 1, & \text{if } r_i \geq \tilde{r} \\ 1 - (\tilde{r}/r_i), & \text{if } r_i < \tilde{r} \end{cases} \quad (7)$$

and

$$e_i = s_i \times (\max\{x_i, y_i\})^U, \quad (8)$$

where $0 \leq U \leq 1$. As noted by Hidiroglou and Berthelot (1986), the quantity U “provides control on the importance associated with the magnitude of the data”; see also Thompson (2007). The values $U = 0.30$ and $U = 0.50$ are recommended in Belcher (2003), Sigman (2002), and Thompson (2007).

Next, let e_{Q_1} , \tilde{e} and e_{Q_3} denote, respectively, the first quartile, the median, and the third quartile of the e_i 's. Now define $d_{Q_1} = \max\{\tilde{e} - e_{Q_1}, |A\tilde{e}|\}$ and $d_{Q_3} = \max\{e_{Q_3} - \tilde{e}, |A\tilde{e}|\}$, which involve a constant A . The value $A = 0.05$ is recommended in Hidiroglou and Berthelot (1986). Ratios outside the interval

$$(\tilde{e} - Cd_{Q_1}, \tilde{e} + Cd_{Q_3}) \quad (9)$$

are flagged as outliers, where C will determine the width of the interval. Various values of C have been assessed in the literature; see Sigman (2002) and Thompson (2007). For our study, we use $C \in \{4, 10, 15\}$ since these provide a good representation of values found in

the literature. We note that an appropriate choice of U , A , and C is necessary before the procedure can be implemented.

The Hidirolou-Berthelot method and the fence-based methods were both applied to microlevel ratio editing for the Annual Survey of Government Finances in Cornett et al. (2006). The authors found that the middle-fences rule and the Hidirolou-Berthelot method provided better results for their application, which they explain is partly influenced by how the edit cells were formed. These methods (including some multivariate methods) were also investigated for macroediting using survey estimates from the U.S. Census Bureau's Annual Capital Expenditures Survey in Thompson (2007). That paper found that the Hidirolou-Berthelot method performed the best, since it is designed to develop flexible limits when the ratios are "highly volatile." Thompson (2007) also underscores how it is difficult to develop a "one method fits all" approach to ratio editing, especially at the macrolevel. Thus it is important to emphasize that these methods, including the approach we present, are all possible tools for setting ratio edit tolerances and final determination should be done in coordination with a content-matter expert.

3. Statistical Tolerance Limits

By definition, a P/γ tolerance interval captures a specified proportion P (called the *content* of the tolerance interval) or more of a population with a given confidence level γ . A tolerance interval is computed using a random sample and the confidence level γ reflects the sampling variability. More formally, suppose a tolerance interval is to be computed for the distribution of a random variable X and let $\mathbf{X} = (X_1, X_2, \dots, X_n)$ denote a random sample of size n . A P/γ two-sided tolerance interval, say $(L(\mathbf{X}), U(\mathbf{X}))$, computed using the random sample \mathbf{X} , satisfies

$$P_X(P_X[L(\mathbf{X}) \leq X \leq U(\mathbf{X})|\mathbf{X}] \geq P) = \gamma. \quad (10)$$

The above condition states that with confidence level γ , the interval $(L(\mathbf{X}), U(\mathbf{X}))$ contains a proportion P or more of the distribution of X . As already noted, the confidence level γ reflects the sampling variability in the random sample \mathbf{X} . The quantities $L(\mathbf{X})$ and $U(\mathbf{X})$ are referred to as the tolerance limits. A one-sided tolerance interval, having only an upper or lower limit, can be similarly defined.

In this article, we use a two-sided tolerance interval for a normal distribution and a one-sided upper tolerance limit for a Weibull distribution. We shall now give expressions for the corresponding approximate tolerance limits. For a univariate normal distribution with unknown mean and unknown variance, let \bar{X} and S^2 denote the sample mean and sample variance based on a sample of size n . Then a two-sided tolerance interval for the normal distribution is given by $\bar{X} \pm kS$, where the quantity k , referred to as a tolerance factor, has the approximate expression (see chap. 2 in Krishnamoorthy and Mathew 2009)

$$k = \left(\frac{(n-1)\chi_{1;p}^2(1/n)}{\chi_{n-1;1-\gamma}^2} \right)^{1/2}. \quad (11)$$

Here, $\chi_{1;p}^2(1/n)$ denotes the 100 P th percentile of a noncentral chi-square distribution with 1 degree of freedom (df) and noncentrality parameter $1/n$, while $\chi_{n-1;1-\gamma}^2$ denotes the 100 $(1-\gamma)$ th percentile of a central chi-square distribution with $(n-1)$ df.

Now consider a random variable X following a Weibull distribution with scale parameter θ and shape parameter β , whose density is given by

$$f_X(x) = \frac{\beta}{\theta^\beta} x^{\beta-1} \exp \left[- \left(\frac{x}{\theta} \right)^\beta \right]. \quad (12)$$

Let $\hat{\theta}$ and $\hat{\beta}$ denote the maximum likelihood estimates of θ and β , respectively, based on a random sample of size n . An approximate P/γ upper tolerance limit for the Weibull distribution is given by

$$\exp \left\{ \ln(\hat{\theta}) - \frac{t_{n-1;1-\gamma}(-\sqrt{n} \ln \{ -\ln(1-P) \})}{\hat{\beta}\sqrt{n-1}} \right\}, \quad (13)$$

where $t_{n-1;1-\gamma}(-\sqrt{n} \ln \{ -\ln(1-P) \})$ is the $100(1-\gamma)$ th percentile of a non-central t distribution with $(n-1)$ df and non-centrality parameter $-\sqrt{n} \ln \{ -\ln(1-P) \}$. The above approximation is due to [Bain and Engelhardt \(1981\)](#).

3.1 Statistical Tolerance Limits for Ratio Edits

If the data are roughly symmetric, an upper and lower tolerance bound may be needed to identify extremes in both tails of the data. However, ratio data are often right skewed. Thus, [Thompson and Sigman \(1999\)](#) suggest first omitting extreme observations of the untransformed data followed by a modified power transformation of the remaining data to obtain approximate symmetry.

There is some additional flexibility and insight gained by using statistical tolerance limits as an alternative to traditional ratio edit tolerance procedures. For example, we typically do not need to be concerned about transforming the data to near symmetry since approximate tolerance intervals have been developed for a wide range of distributions; see, for example, [Krishnamoorthy and Mathew \(2009\)](#). Also, the content and confidence levels of a tolerance interval allow us to reflect the uncertainty of what we are trying to capture with these intervals. Such uncertainty is not directly quantified by the traditional ratio edit tolerance procedures.

For the tolerance-limit approach, we first temporarily trim the data based on a user-specified trimming level. The assumption is that the remaining data behave similarly to the “true” uncontaminated distribution. The trimmed dataset is then used to calculate statistical tolerance limits, which can extend beyond the extremes of the trimmed data. Thus, some of the initially trimmed data may be retained as “good” data if they fall within the statistical tolerance limits, or the statistical tolerance limits may indicate that further data should be classified as ratio edit failures.

Another benefit to using statistical tolerance intervals is that the limits can never be negative for distributions with nonnegative support, regardless of the confidence and content levels specified. However, robust control limits and fence-based limits can yield negative lower bounds. While one can simply truncate the lower limits from these methods at zero, we do not have to specify this additional assumption when using statistical tolerance intervals.

4. Numerical Study

We now compare the performance of statistical tolerance limits with the traditional outlier procedures for determining ratio edits. All simulations in this section and calculations

for the example in the next section are performed using the R programming language (R Development Core Team 2013). Moreover, statistical tolerance limits are calculated using the R add-on package *tolerance* (Young 2010).

We compare the performance of the statistical tolerance limits with the ratio edit tolerances in two ways. First, we compute the average width of each procedure to comment on the relative conservatism of each procedure. Next, we compute the proportion of misclassified ratios with respect to each procedure's limits. We are interested in the proportion of false hits and misses, which are basically Type I and Type II error rates, respectively. Specifically, let X be a ratio. Then

$$\text{Type I Error Rate} = \Pr\{X \text{ flagged as "bad"} | X \text{ is "good"}\} \quad (14)$$

$$\text{Type II Error Rate} = \Pr\{X \text{ flagged as "good"} | X \text{ is "bad"}\} \quad (15)$$

Note that in the literature on outliers, the Type I and Type II errors defined above are rates of swamping and masking, respectively; we refer to Barnett and Lewis (1994) for further discussion on swamping and masking effects. We also note that some researchers may prefer to switch the definitions of Type I and Type II errors given above, unlike in a hypothesis-testing situation where Type I and Type II errors have universally accepted definitions. We chose the definitions given in (14) and (15) since they have already been used in the literature; cf. sec. 4.1 of Thompson and Sigman (1999).

In the case of a heavily skewed distribution, the region of outliers will typically be in the direction of the skewness. Therefore, instead of exploring simulated data where transformations could get the data close to symmetry, we will explore using one-sided trimming on the raw data in the direction of the skewness followed by a robust one-sided limit.

Our simulations assess the efficacy of one-sided tolerance limits and two-sided tolerance intervals for determining ratio edits. For the one-sided setting, we use a two-component mixture of Weibull distributions to simulate contamination in the upper tail of the data. For the two-sided setting, we use a three-component mixture of normals to simulate contamination in both tails of the data. It should be noted that mixture distributions (e.g., the contaminated normal model) have been used in the literature to assess the performance of editing procedures for survey data; see Ghosh-Dastidar and Schafer (2006). For each set of simulations, three scenarios were considered: well-separated components (i.e., a big gap between the "good" ratios and the "bad" ratios), moderate overlapping, and heavy overlapping.

Let $\text{Wei}(\theta, \beta)$ be the Weibull distribution with scale parameter θ and shape parameter β . Let $N(\mu, \sigma^2)$ be the normal distribution with mean μ and variance σ^2 . The distributions we use for the one-sided contaminated simulations are:

- (Well Separated): $0.95 * \text{Wei}(1, 15) + 0.05 * \text{Wei}(50, 100)$
- (Moderate Overlapping): $0.95 * \text{Wei}(1, 15) + 0.05 * \text{Wei}(20, 60)$
- (Heavy Overlapping): $0.95 * \text{Wei}(1, 15) + 0.05 * \text{Wei}(5, 40)$

The distributions we use for the two-sided contaminated simulations are:

- (Well Separated): $0.90 * N(1000, \sqrt{50}) + 0.05 * N(500, \sqrt{50}) + 0.05 * N(1500, \sqrt{50})$

- (Moderate Overlapping): $0.90*N(1000, \sqrt{50}) + 0.05*N(750, \sqrt{50}) + 0.05*N(1250, \sqrt{50})$
- (Heavy Overlapping): $0.90*N(1000, \sqrt{50}) + 0.05*N(900, \sqrt{50}) + 0.05*N(1100, \sqrt{50})$

The following outlines the general simulation performed for our study:

1. Simulate n ratios, X_1, \dots, X_n , from one of the contaminated models discussed above. Denote this sequence of ratios by \mathbf{X} .
2. Apply the traditional methods (i.e., the methods in Section 2) to \mathbf{X} and calculate the ratio edit tolerances based on these approaches.
3. Use trimming at the $\alpha \in \{0.01, 0.02, \dots, 0.10, 0.15\}$ levels on \mathbf{X} . Call these trimmed datasets \mathbf{X}^α .
4. Using \mathbf{X}^α , compute a normal statistical tolerance interval if contamination is assumed in both tails, or a one-sided upper Weibull statistical tolerance limit if contamination is assumed only in the right tail.
5. For each method and with respect to \mathbf{X} , calculate the proportion of good ratios falling outside of the tolerance limits (Type I error), and the proportion of bad ratios falling within the tolerance limits (Type II error).
6. Calculate the width of the statistical tolerance interval and the intervals determined by the traditional methods. For the one-sided setting, the one-sided upper tolerance limit will be taken as the width since an absolute lower limit of 0 is assumed for the data.
7. Repeat the above B times. For each method, average the Type I error rates, Type II error rates, and interval widths to get Monte Carlo estimates of each quantity.

For our simulations, we generate $n \in \{300, 1000\}$ ratios $B = 10,000$ times and compute P/γ tolerance intervals at the 90/90 and 95/95 levels. Recall from Section 3 that P is the content of the tolerance interval and γ is its confidence level. For the methods discussed in Section 2, we specify values for the constants (which we refer to as “Factors” in the summary tables) based on the references cited within.

Tables 1–3 give the simulation results for the three contamination structures considered for $n = 1,000$. The general results are similar for $n = 300$, which are reported in Tables 6–8 in the Appendix. We only report the results for a subset of the trimming levels used, but the trend in the average widths and errors as α changes is apparent. When the contamination structure is well separated or moderately overlaps with respect to the “good” data and a trimming level is selected close to the amount of contamination (5% for our simulations), then the statistical tolerance interval approach performs the best, namely meaning that the Type I error comes close to the nominal $(1 - \gamma)$ level. Note the results in bold in the tables, which pertain to the temporary trimming done at the true percentage of contamination. Regardless of the contamination structure, this approach does a good job of controlling the Type I errors as long as the level of trimming does not heavily exceed the content level P of the tolerance interval.

For the robust control limits, larger values of L yield smaller Type I errors, but larger Type II errors. Using $L \in \{2.0, 2.5, 3.0, 3.5\}$, we see there is generally a wide spread in the Type I and Type II errors. Again, we note that Thompson and Sigman (1999) use $L = 2$ for a more liberal rule and $L = 3$ for a more conservative rule regarding the number of cases

Table 1. Simulation results for the well-separated contamination structure ($n = 1,000$)

Factors	One-sided			Two-sided		
	Average width	Type I error	Type II error	Average width	Type I error	Type II error
L						
2.0	29.6914	0.1383	0.0000	Robust control limits	0.2019	0.0000
2.5	34.3543	0.1011	0.0000		0.1109	0.0000
3.0	39.0171	0.0741	0.0000		0.0559	0.0000
3.5	43.6799	0.0543	0.0000		0.0259	0.0000
k						
1.5	51.5592	0.0323	0.0000	Resistant fences	0.0024	0.0000
2.0	60.9528	0.0174	0.0000		0.0002	0.0000
3.0	79.7400	0.0050	0.0002		0.0000	0.0000
k^*						
3.0	59.8644	0.0188	0.0000	Asymmetric fences	0.0025	0.0000
4.0	72.0264	0.0085	0.0000		0.0002	0.0000
6.0	96.3504	0.0017	0.3340		0.0000	0.0000
(U, C)						
(0.3, 4)	50.4094	0.0361	0.0000	Hidiroglou-Berthelot	0.0042	0.0000
(0.3, 10)	94.1750	0.0023	0.1147		0.0000	0.0015
(0.3, 15)	155.5960	0.0003	0.9794		0.0000	0.3112
(0.5, 4)	46.2717	0.0470	0.0000		0.0041	0.0000
(0.5, 10)	80.9931	0.0057	0.0001		0.0000	0.0006
(0.5, 15)	109.1108	0.0014	0.5276		0.0000	0.2400
α						
0.01	46.4975	0.0453	0.0000	90/90 tolerance limits	0.0000	0.0000
0.05	36.3107	0.0895	0.0000		0.0592	0.0000
0.10	28.9472	0.1454	0.0000		0.1873	0.0000
0.15	24.7898	0.1917	0.0000		0.2763	0.0000
α						
0.01	63.8249	0.0143	0.0000	95/95 tolerance limits	0.0000	0.0001
0.05	48.2107	0.0407	0.0000		0.0252	0.0000
0.10	37.4035	0.0827	0.0000		0.1126	0.0000
0.15	31.6277	0.1215	0.0000		0.1901	0.0000

Table 2. Simulation results for the moderately overlapping contamination structure ($n = 1,000$)

Factors	One-sided		Two-sided	
	Average width	Type I error	Type II error	Type I error
L				
2.0	29.6914	0.1383	0.0000	0.2019
2.5	34.3542	0.1011	0.0000	0.1109
3.0	39.0170	0.0741	0.0002	0.0559
3.5	43.6799	0.0543	0.0024	0.0259
			Robust control limits	
			127.5003	0.0001
			159.3754	0.0003
			191.2504	0.0010
			223.1255	0.0029
k				
1.5	51.5592	0.0323	0.0679	0.0024
2.0	60.9528	0.0174	0.6918	0.0002
3.0	79.7400	0.0050	0.9797	0.0000
			Resistant fences	
			305.5721	0.0271
			381.9652	0.1220
			534.7512	0.6328
k^*				
3.0	59.8644	0.0188	0.5919	0.0025
4.0	72.0264	0.0085	0.9773	0.0002
6.0	96.3504	0.0017	0.9797	0.0000
			Asymmetric fences	
			305.5721	0.0280
			381.9652	0.1250
			534.7512	0.6314
(U, C)				
(0.3, 4)	50.2838	0.0361	0.0408	0.0042
(0.3, 10)	102.4551	0.0023	0.9797	0.0000
(0.3, 15)	125.2692	0.0003	0.9866	0.0000
(0.5, 4)	46.2483	0.0470	0.0078	0.0041
(0.5, 10)	81.8213	0.0057	0.9797	0.0000
(0.5, 15)	112.6471	0.0014	0.9797	0.0000
			Hidiroglou-Berthelot	
			303.1756	0.0191
			872.6594	0.9098
			1322.8198	0.9798
			303.7215	0.0190
			843.6258	0.9296
			1324.5677	0.9798
α				
0.01	40.2427	0.0685	0.0004	0.0054
0.05	34.8773	0.0980	0.0000	0.0870
0.10	28.9446	0.1454	0.0000	0.1874
0.15	24.7898	0.1917	0.0000	0.2763
			90/90 tolerance limits	
			280.4024	0.0146
			171.6823	0.0005
			131.7994	0.0001
			108.7582	0.0001
α				
0.01	53.6683	0.0280	0.1215	0.0009
0.05	45.9078	0.0469	0.0064	0.0398
0.10	37.3994	0.0827	0.0001	0.1126
0.15	31.6277	0.1215	0.0000	0.1901
			95/95 tolerance limits	
			336.9388	0.0535
			206.3752	0.0017
			158.5197	0.0003
			130.8940	0.0001

Table 3. Simulation results for the heavily overlapping contamination structure ($n = 1,000$)

Factors	One-sided			Two-sided		
	Average width	Type I error	Type II error	Average width	Type I error	Type II error
L						
2.0	28.9717	0.1450	0.1818	123.5955	0.2160	0.2226
2.5	33.4823	0.1071	0.3390	154.4944	0.1222	0.3247
3.0	37.9929	0.0793	0.5393	185.3932	0.0638	0.4421
3.5	42.5035	0.0588	0.7394	216.2921	0.0307	0.5644
k						
1.5	50.9635	0.0336	0.9533	297.9491	0.0030	0.8351
2.0	60.2380	0.0182	0.9795	372.4364	0.0002	0.9535
3.0	78.7868	0.0053	0.9797	521.4109	0.0000	0.9798
k^*						
3.0	58.9201	0.0199	0.9791	297.9491	0.0032	0.8341
4.0	70.8468	0.0091	0.9797	372.4364	0.0003	0.9521
6.0	94.7000	0.0019	0.9797	521.4109	0.0000	0.9798
(U, C)						
(0.3, 4)	49.5845	0.0377	0.9287	288.0104	0.0051	0.7931
(0.3, 10)	98.5630	0.0025	0.9797	1024.5677	0.0000	0.9798
(0.3, 15)	118.6151	0.0003	0.9962	1024.5677	0.0000	0.9836
(0.5, 4)	45.5290	0.0489	0.8369	288.2080	0.0050	0.7942
(0.5, 10)	80.1847	0.0062	0.9797	1024.5677	0.0000	0.9798
(0.5, 15)	106.9169	0.0015	0.9797	1024.5677	0.0000	0.9807
α						
0.01	36.6272	0.0871	0.4750	184.5406	0.0649	0.4381
0.05	31.9656	0.1188	0.2784	152.0293	0.1283	0.3155
0.10	27.8326	0.1564	0.1511	125.9865	0.2075	0.2295
0.15	24.3998	0.1966	0.0817	105.4296	0.2912	0.1718
α						
0.01	48.0975	0.0405	0.9146	221.7488	0.0266	0.5859
0.05	41.4319	0.0631	0.6961	182.7508	0.0674	0.4318
0.10	35.7166	0.0924	0.4345	151.5282	0.1293	0.3142
0.15	31.0461	0.1261	0.2475	126.8878	0.2040	0.2326

flagged for review. While the overall simulation results for $n = 300$ and $n = 1,000$ were similar, we note that the sample size does affect the errors for the robust control limits; that is, for larger n , the Type I error rates increase, while the Type II error rates decrease.

The fence-based methods are typically more conservative with respect to the statistical tolerance interval approach. As the contamination structure mixes more with the good data, we note that the Type II errors for the fence-based methods increase significantly with respect to the Type II errors for the tolerance intervals. We also note that the summaries are very similar for the two fence-based methods under the two-sided setting. This is expected given the symmetry of the generated data.

For most of the common values of the Hidioglou-Berthelot method, we see that their performance is comparable to the statistical tolerance interval approach (at the 90/90 and 95/95 levels) under the well-separated case. The exceptions are when $(U, C) = (0.3, 10)$ and $(U, C) = (0.3, 15)$. Again, as the contamination structure mixes more with the good data, we note that the Type II errors increase significantly with respect to the Type II errors for the tolerance intervals.

Overall, the simulation results show that as more of the contaminated data mixes with the good data, masking becomes more prevalent. This results in intervals that do not (or cannot) exclude the contaminated data, which in turn increases the Type II errors for all procedures. When assessing the methods of Section 2, we simply used common levels found in the literature. Different results would obviously be obtained by adjusting the user-specified constants. But for a given set of data, the intuition may not always be apparent as to the trade-off in terms of the types of errors. However, the intuition with the values specified in the tolerance interval approach (i.e., α , P , and γ) are all clear. Informative choices of these levels will help control both types of errors, thus suggesting the utility of statistical tolerance intervals as a way to set ratio edit tolerances.

5. Annual Survey of Manufacturers

The Annual Survey of Manufacturers (ASM) collects data for the years between the Economic Census, which is conducted in the years ending in 2 and 7. The annual survey data are estimates derived from a statistically selected sample from all manufacturing establishments with one or more paid employees. The collection mode for this survey is through paper and internet reporting. Examples of statistics that the ASM reports for different manufacturing sectors include employment, payroll, operating expenses, value of shipments, and inventories.

In order to make the results of this example accessible and reproducible for the reader, our analysis uses the *Statistics for Industry Group and Industries* file for the years 2010 and 2011. The data can be accessed from the U.S. Census Bureau's website for the ASM found at <http://www.census.gov/manufacturing/asm/index.html>. The statistics are reported at various North American Industry Classification System (NAICS) levels. We use the lowest level reported, which is the six-digit NAICS industry grouping. We note that since this is officially published data, it has already gone through the U.S. Census Bureau's editing process. Our intent is to highlight the implementation of the statistical tolerance interval approach on this edited macrodata, which would typically be followed by a subject-matter expert's analysis of the flagged values.

We study six ratios for this example. Many of the variables comprising the ratios are reported in U.S. dollars, such as payroll, materials, and inventories. For all such quantities, the values are reported in \$1,000 on the summary file. The ratios we study, as well as the abbreviations we use, are:

- **PR/NE**: annual payroll/number of employees;
- **MU/TS**: materials used/total value of shipments;
- **ME/MB**: materials and supplies at end of the year/materials and supplies at beginning of the year;
- **WH/WA**: all production worker's hours (in 1,000 hours)/production worker's average per year (i.e., the number of employees on payroll on certain days of the month specified by the ASM);
- **IE/IB**: total inventories at the end of the year/total inventories at the beginning of the year; and
- **WE/WB**: work-in-process inventories at the end of the year/work-in-process inventories at the beginning of the year.

The total number of industries for each dataset is 321. However, some ratios are not calculated since one or both of the values for an industry are withheld due to estimates not meeting publication or disclosure standards set by the U.S. Census Bureau.

We first determine whether a normal or Weibull distribution is most appropriate for each 5% trimmed ratio dataset. While we only explore these two distributions, there are no restrictions on which parametric distributions to investigate – especially if knowledge is available from a subject-matter expert. Regardless, we first use the Kolmogorov-Smirnov test to assess whether data from corresponding years follow the same distribution. Four of the ratios (PR/NE, MU/TS, ME/MB, and IE/IB) yield p -values well over 0.15, while the other two ratios (WH/WA and WE/WB) have p -values below 0.05.

For the four ratios that have statistically similar distributions between the two years, we temporarily pool each pair of ratio datasets. We use the Shapiro-Wilk test for normality and the chi-square goodness-of-fit for testing the Weibull assumption. We then select the distribution of which corresponding test had the higher p -value. While these are two different tests, this is merely a simple approach to decide upon a distribution.

For the two ratios that are significantly different, we proceed similarly with testing the normality or the Weibull assumption. However, we keep each year's data separate and run the tests on these datasets. We then choose the distribution of which the test yielded the higher p -value between the two datasets for a given ratio.

After determining to proceed with the Weibull or normal assumption, we then compute one-sided tolerance limits or two-sided tolerance intervals, respectively. We consider the 90/90 and 95/95 levels with an initial trimming of 5%. We also perform a relative comparison between the 2010 and 2011 ratios. Specifically, we compare the proportions of how an industry is classified (i.e., as being “good” or an “outlier”) from 2010 and 2011. These quantities give us an indication of how stable the classifications are from 2010 to 2011 with respect to the calculated limits.

For the PR/NE ratios, we also calculate the other limits discussed in this article. We found that the Weibull distribution is appropriate for both the 2010 and 2011 data. Thus, we calculate 90/90 and 95/95 one-sided upper Weibull tolerance limits. The results are

Table 4. Comparison of the one-sided upper limits for the PR/NE data

Method	Upper limit (2010)	Upper limit (2011)	Outlier to good	Good to outlier	Good to good
Robust control limits					
L					
2.0	68.9846	71.0740	0.0031	0.0094	0.8840
2.5	73.7211	76.0442	0.0125	0.0031	0.9154
3.0	78.4577	81.0143	0.0031	0.0031	0.9498
3.5	83.1942	85.9844	0.0063	0.0031	0.9687
Resistant fences					
k					
1.5	84.5603	86.1832	0.0000	0.0063	0.9749
2.0	92.9164	94.7511	0.0000	0.0000	0.9875
3.0	109.6287	111.8871	0.0000	0.0000	1.0000
Asymmetric fences					
k^*					
3.0	88.3639	88.8199	0.0031	0.0031	0.9812
4.0	97.9879	98.2668	0.0000	0.0000	0.9875
6.0	117.2360	117.1606	0.0000	0.0000	1.0000
Hidiroglou-Berthelot					
(U, C)					
(0.3, 4)	87.9928	87.2365	0.0000	0.0063	0.9781
(0.3, 10)	105.6065	107.3582	0.0000	0.0000	1.0000
(0.3, 15)	105.6065	107.3582	0.0000	0.0000	1.0000
(0.5, 4)	83.0819	84.1461	0.0063	0.0157	0.9561
(0.5, 10)	105.6065	107.3582	0.0000	0.0000	1.0000
(0.5, 15)	105.6065	107.3582	0.0000	0.0000	1.0000
Tolerance limits					
P/γ					
90/90	65.8670	67.7087	0.0157	0.0094	0.8464
95/95	70.1582	72.1881	0.0063	0.0094	0.8934

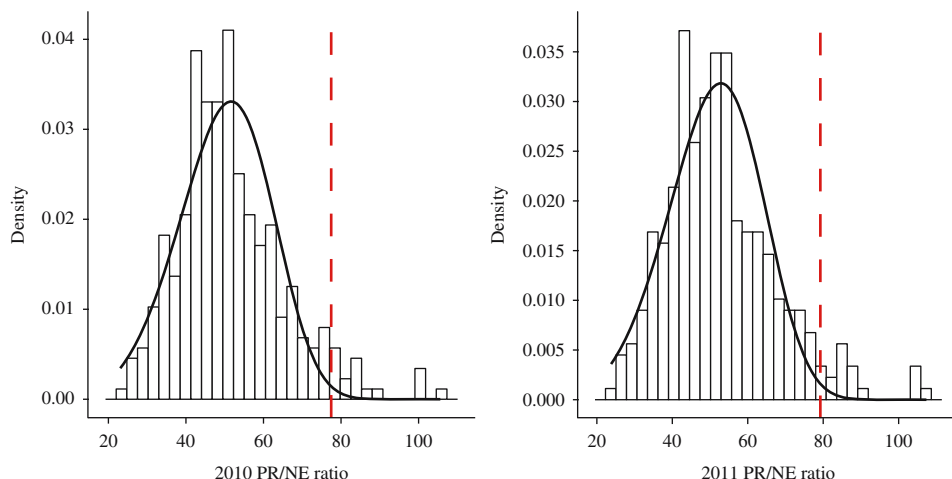


Fig. 1. Histograms of the PR/NE ratios for (a) 2010 and (b) 2011. The dashed line represents the 5% trimming threshold and the solid line is the Weibull density curve fit to the trimmed data

reported in Table 4. We see that the resistant fences provide fairly conservative limits. As such, the proportion of points classified as “good” to “good” is close to or at 1 and this conservatism is likely not desirable. As the histograms in Figure 1 show, there are clearly a few ratios above the value of 90 that may be candidates for editing. The robust control limits and the tolerance interval procedures would flag these values for possible editing, whereas the other approaches produce fairly conservative limits. Given the ability to better control Type I and Type II errors with the statistical tolerance intervals, their use here gives this approach a significant edge over the other procedures.

Scatterplots of the payroll versus the number of employees for each year are given in Figure 2. As can be seen, each year shows a strong correlation (which is approximately $+0.93$ for each year). Values flagged using the 90/90 and 95/95 tolerance limits are color coded accordingly. One thing to note is that as the correlation strengthens, the resulting tolerance limits will be “tighter” around the data.

Results for the other five ratios are similar to those reported for the PR/NE ratios. Hence, we only focus on the tolerance interval results. Table 5 gives the one-sided tolerance limit or two-sided tolerance interval results depending on the distributional assumption made. For the 90/90 limits, approximately 70% to 85% of the data stay within the limits across years, while for the 95/95 limits, these same percentages range from approximately 80% to 90%. These percentages give an indication of those industries that have essentially remained stable between 2010 and 2011. If one wants to develop certain summary statistics between the two years, then those industries that fell outside of the limits in one or both years could be candidates for editing. Moreover, they could be indicative of changes that occurred within that particular industry.

6. Discussion

The criterion used in developing a statistical tolerance interval indicates that it is a natural choice for computing bounds that can be used to perform ratio edits; that is, in order to flag ratios that are inconsistent or problematic. In our work, we have demonstrated this in the

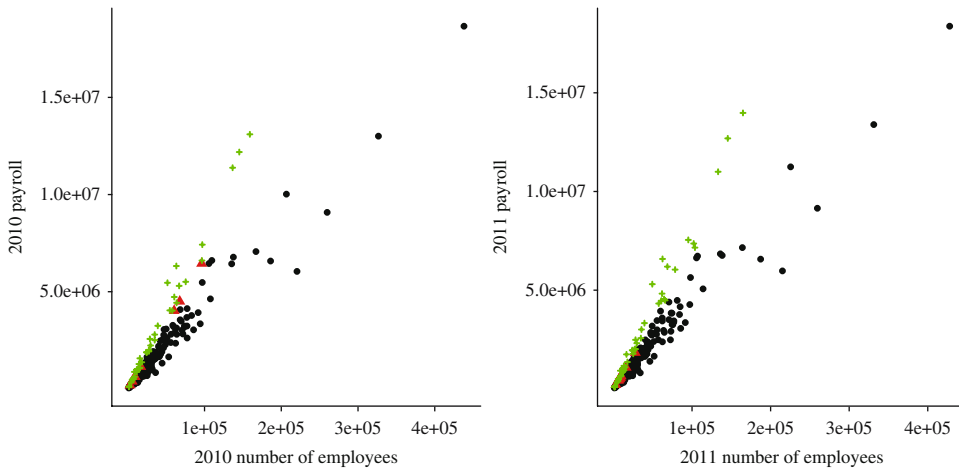


Fig. 2. Scatterplots of the payroll (in \$1,000) versus the number of employees for (a) 2010 and (b) 2011. The triangles are values greater than the 95/95 upper tolerance limit, while the plusses and triangles are values greater than the 90/90 upper tolerance limit

case of the normal distribution (where the problematic ratios can appear in either tail of the distribution) and in the case of the Weibull distribution (where the problematic ratios appear only in the right tail). A comparison with other ratio edit procedures shows that the statistical tolerance-interval approach has a significant edge over the existing procedures in terms of controlling Type I and Type II errors. The approach also depends on an initial level of trimming. As noted in Section 1, there is no clear guidance on choosing a percentage of trimming to perform, so one should seek input from a subject-matter expert. Our approach can certainly be adopted for other distributions; see [Krishnamoorthy and Mathew \(2009\)](#) for details on the development of tolerance intervals for a variety of distributions.

We also acknowledge that the ratio editing process is often complex and includes numerous rules that are typically dependent on the type of survey. Moreover, ratio editing at the microlevel and macrolevel often use different approaches, with the latter setting not as well studied in the literature. We illustrated the statistical tolerance interval approach on ASM data at the macrolevel, but the approach is applicable to the microlevel setting. We are not suggesting a panacea for setting ratio edit tolerances in all survey settings; however, we are suggesting that statistical tolerance intervals can be useful in informing ratio editing processes.

We note that both of the variables used in the computation of a ratio can have values that are outliers, and yet the ratio will not be flagged as an outlier. This can obviously happen when values of both variables are too small or too large, so that the outlyingness gets cancelled when we take the ratio. A simple example is if a small business reports 400 trillion dollars in payroll for ten million employees, then the PR/NE ratio would be consistent with those displayed in [Figure 1](#). In view of this, it is essential to have outlier detection methods that are applicable to bivariate data, or to multivariate data when data are available on several variables. A Mahalanobis distance based outlier detection method (cf. [Franklin et al. \(2000\)](#) and [Thompson \(2007\)](#)) may not adequately flag the outliers, since the outlyingness of a single variable (or a few variables) may be cancelled out by the magnitudes of the other variables. We believe a rectangular tolerance region that provides simultaneous tolerance intervals on each variable is required. Such a tolerance region is currently under investigation.

Table 5. Comparison of the statistical tolerance intervals for the other five ratio datasets. The second column gives the distributional assumption, which determined whether we calculated a one-sided tolerance limit or two-sided tolerance interval

Ratio	Dist.	P/γ	2010 limits	2011 limits	Outlier to good	Good to outlier	Good to good
MU/TS	Weibull	90/90	0.5914	0.6053	0.0221	0.0063	0.8360
		95/95	0.6385	0.6546	0.0126	0.0189	0.8991
ME/MB	Normal	90/90	(0.9339, 1.2204)	(0.9555, 1.2504)	0.1169	0.1039	0.7143
		95/95	(0.9037, 1.2507)	(0.9243, 1.2815)	0.0714	0.0649	0.8149
WH/WA	Weibull	90/90	2.0850	2.0952	0.0313	0.0251	0.8464
		95/95	2.1065	2.1182	0.0157	0.0251	0.8903
IE/IB	Normal	90/90	(0.9461, 1.2143)	(0.9712, 1.2186)	0.1266	0.1266	0.6899
		95/95	(0.9178, 1.2425)	(0.9452, 1.2447)	0.0791	0.1076	0.7911
WE/WB	Normal	90/90	(0.8513, 1.3252)	(0.9068, 1.2868)	0.1303	0.1303	0.6775
		95/95	(0.8013, 1.3752)	(0.8667, 1.3269)	0.0912	0.0879	0.7818

Appendix: Additional Simulation Results

Table 6. Simulation results for the well-separated contamination structure ($n = 300$)

Factors	One-sided		Two-sided	
	Average width	Type I error	Average width	Type II error
L				
2.0	35.1594	0.0950	Robust control limits 127.8109	0.1996
2.5	41.1869	0.0638		0.1095
3.0	47.2145	0.0429		0.0554
3.5	53.2420	0.0289		0.0258
k				
1.5	51.3763	0.0330	Resistant fences 304.5684	0.0028
2.0	60.7281	0.0179		0.0002
3.0	79.4316	0.0053		0.0000
k^*				
3.0	59.5974	0.0197	Asymmetric fences 304.5684	0.0033
4.0	71.6895	0.0091		0.0003
6.0	95.8738	0.0020		0.0000
(U, C)				
(0.3, 4)	51.4053	0.0370	Hidiroglou-Berthelot 595.6948	0.0051
(0.3, 10)	101.8578	0.0026		0.0039
(0.3, 15)	156.3003	0.0004		0.3038
(0.5, 4)	46.9528	0.0480	1000.5404	0.0000
(0.5, 10)	85.1081	0.0062	596.8402	0.0050
(0.5, 15)	118.1210	0.0016	831.7058	0.0000
			954.3806	0.2579
α				
0.01	48.5763	0.0399	90/90 tolerance limits 507.6096	0.0000
0.05	38.0012	0.0808		0.0000
0.10	30.1621	0.1342	223.7015	0.0000
0.15	25.7931	0.1790	136.2889	0.1721
			112.6394	0.2582

Table 7. Simulation results for the moderately overlapping contamination structure ($n = 300$)

Factors	One-sided			Two-sided		
	Average width	Type I error	Type II error	Average width	Type I error	Type II error
L						
2.0	35.1591	0.0950	0.0001	127.8100	0.1996	0.0001
2.5	41.1866	0.0638	0.0021	159.7626	0.1095	0.0003
3.0	47.2141	0.0429	0.0284	191.7151	0.0554	0.0011
3.5	53.2415	0.0289	0.1795	223.6676	0.0258	0.0032
k						
1.5	51.3763	0.0330	0.1136	304.5681	0.0028	0.0285
2.0	60.7281	0.0179	0.6064	380.7101	0.0002	0.1264
3.0	79.4316	0.0053	0.9270	532.9941	0.0000	0.6217
k^*						
3.0	59.5974	0.0197	0.5271	304.5681	0.0033	0.0313
4.0	71.6895	0.0091	0.8880	380.7101	0.0003	0.1349
6.0	95.8738	0.0020	0.9286	532.9941	0.0000	0.6171
(U, C)						
(0.3, 4)	50.7645	0.0370	0.0822	329.2378	0.0051	0.0211
(0.3, 10)	111.4427	0.0026	0.9280	1020.4076	0.0000	0.8680
(0.3, 15)	123.4679	0.0004	0.9561	1270.2145	0.0000	0.9313
(0.5, 4)	46.7444	0.0480	0.0197	329.7336	0.0050	0.0210
(0.5, 10)	92.8964	0.0062	0.9232	1056.5267	0.0000	0.8836
(0.5, 15)	117.3996	0.0016	0.9287	1271.0021	0.0000	0.9311
α						
0.01	42.0401	0.0607	0.0015	287.8207	0.0050	0.0190
0.05	36.2829	0.0891	0.0001	180.5547	0.0740	0.0007
0.10	30.1462	0.1343	0.0000	136.2762	0.1721	0.0001
0.15	25.7931	0.1790	0.0000	112.6387	0.2582	0.0001
α						
0.01	57.1274	0.0221	0.3816	348.3759	0.0008	0.0717
0.05	48.6076	0.0393	0.0386	218.6988	0.0310	0.0030
0.10	39.6391	0.0711	0.0010	165.2392	0.0975	0.0004
0.15	33.4736	0.1069	0.0000	136.7524	0.1700	0.0002

Table 8. Simulation results for the heavily overlapping contamination structure ($n = 300$)

Factors	One-sided		Two-sided	
	Average width	Type I error	Type II error	Type I error
L				
2.0	33.7435	0.1043	0.3568	0.2140
2.5	39.4462	0.0714	0.6078	0.1210
3.0	45.1488	0.0490	0.8155	0.0634
3.5	50.8515	0.0336	0.9080	0.0308
			Robust control limits	
			123.7923	0.2261
			154.7404	0.3294
			185.6885	0.4452
			216.6366	0.5674
k				
1.5	50.7531	0.0343	0.8999	0.0035
2.0	59.9802	0.0187	0.9272	0.0003
3.0	78.4344	0.0056	0.9280	0.0000
			Resistant fences	
			296.9006	0.8242
			371.1257	0.9164
			519.5760	0.9311
k^*				
3.0	58.6094	0.0208	0.9251	0.0040
4.0	70.4553	0.0098	0.9279	0.0004
6.0	94.1470	0.0022	0.9323	0.0000
			Asymmetric fences	
			296.9006	0.8205
			371.1257	0.9129
			519.5760	0.9310
(U, C)				
(0.3, 4)	50.3828	0.0387	0.8765	0.0061
(0.3, 10)	103.0961	0.0028	0.9291	0.0000
(0.3, 15)	112.2477	0.0004	0.9825	0.0000
(0.5, 4)	45.9667	0.0500	0.8010	0.0060
(0.5, 10)	88.3975	0.0066	0.9280	0.0000
(0.5, 15)	107.4762	0.0017	0.9344	0.0000
			Hidiroglou-Berthelot	
			378.4962	0.7847
			971.1992	0.9311
			971.2298	0.9518
			377.1936	0.7854
			971.2298	0.9311
			971.2298	0.9457
α				
0.01	38.3546	0.0773	0.5560	0.0575
0.05	33.2305	0.1088	0.3305	0.1165
0.10	28.9141	0.1452	0.1825	0.1921
0.15	25.3551	0.1842	0.0987	0.2730
			90/90 tolerance limits	
			189.7963	0.4602
			156.6118	0.3346
			130.0538	0.2448
			109.1294	0.1831
α				
0.01	51.3207	0.0324	0.9143	0.0213
0.05	43.8078	0.0535	0.7789	0.0571
0.10	37.7291	0.0801	0.5315	0.1135
0.15	32.8047	0.1116	0.3196	0.1838
			95/95 tolerance limits	
			229.7280	0.6194
			189.6976	0.4623
			157.6943	0.3400
			132.4919	0.2532

7. References

- Aggarwal, C.C. 2013. *Outlier Analysis*. New York: Springer.
- Bain, L.J. and M. Engelhardt. 1981. "Simple Approximate Distributional Results for Confidence and Tolerance Limits for the Weibull Distribution Based on Maximum Likelihood Estimators." *Technometrics* 23: 15–20. DOI: <http://dx.doi.org/10.1080/00401706.1981.10486231>
- Barnett, V. and T. Lewis. 1994. *Outliers in Statistical Data*, 3rd ed. Wiley Series in Probability and Mathematical Statistics. Chichester: John Wiley & Sons.
- Belcher, R. 2003. "Application of the Hidioglou-Berthelot Method of Outlier Detection for Periodic Business Surveys." In Proceedings of the Survey Methods Section: Statistical Society of Canada Annual Meeting, June, 2003. 25–30 Halifax, Nova Scotia, Canada. Available at: http://www.ssc.ca/survey/documents/SSC2003_R_Belcher.pdf
- Chawla, S. and A. Gionis. 2013. "*k*-means-: A Unified Approach to Clustering and Outlier Detection." In Proceedings of the 2013 SIAM International Conference on Data Mining: Society for Industrial and Applied Mathematics, May, 2013. 187–197 Austin, Texas, USA. Available at <http://epubs.siam.org/doi/pdf/10.1137/1.9781611972832.21>
- Cornett, E., J.F. McLaughlin, and C.R. Hogue. 2006. "A Comparison of Two Ratio Edit Methods for the Annual Survey of Government Finances." In Proceedings of the Section on Survey Research Methods: American Statistical Association, August, 2006. 2878–2883 Seattle, WA, USA. Available at: <http://www.amstat.org/sections/SRMS/Proceedings/y2006/Files/JSM2006-000199.pdf>
- Franklin, S., S. Thomas, and M. Brodeur. 2000. "Robust Multivariate Outlier Detection Using Mahalanobis' Distance and Modified Stahel-Donoho Estimators." In Proceedings of the Second International Conference on Establishment Surveys (ICES-II), Survey Methods for Businesses, Farms, and Institutions, June, 2000. 697–706 Buffalo, NY, USA. Available at: <http://www.amstat.org/meetings/ices/2000/proceedings/S33.pdf>
- Ghosh-Dastidar, B. and J.L. Schafer. 2006. "Outlier Detection and Editing Procedures for Continuous Multivariate Data." *Journal of Official Statistics* 22: 487–506.
- Hidioglou, M.A. and J.-M. Berthelot. 1986. "Statistical Editing and Imputation for Periodic Business Surveys." *Survey Methodology* 12: 73–83.
- Hido, S., Y. Tsuboi, H. Kashima, M. Sugiyama, and T. Kanamori. 2011. "Statistical Outlier Detection Using Direct Density Ratio Estimation." *Knowledge and Information Systems* 26: 309–336. DOI: <http://dx.doi.org/10.1007/s10115-010-0283-2>
- Hoaglin, D.C., B. Iglewicz, and J.W. Tukey. 1986. "Performance of Some Resistant Rules for Outlier Labeling." *Journal of the American Statistical Association* 81: 991–999. DOI: <http://dx.doi.org/10.1080/01621459.1986.10478363>
- Iglewicz, B. and D.C. Hoaglin. 1993. *How to Detect and Handle Outliers*, vol. 16. Milwaukee, WI: American Society for Quality Control.
- Kokic, P.N. and P.A. Bell. 1994. "Optimal Winsorizing Cutoffs for a Stratified Finite Population Estimator." *Journal of Official Statistics* 10: 419–435.
- Krishnamoorthy, K. and T. Mathew. 2009. *Statistical Tolerance Regions: Theory, Applications, and Computation*. Hoboken, NJ: Wiley.

- Latouche, M. and J.-M. Berthelot. 1992. "Use of a Score Function to Prioritize and Limit Recontacts in Editing Business Surveys." *Journal of Official Statistics* 8: 389–400.
- Montgomery, D.C. 2013. *Introduction to Statistical Quality Control*, 7th ed. Hoboken, NJ: Wiley.
- R Development Core Team. 2014. *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. Available at: <http://www.R-project.org/> ISBN 3-900051-07-0 (accessed February 13, 2015)
- Rais, S. 2008. "Outlier Detection for the Consumer Price Index." In Proceedings of the Survey Methods Section: Statistical Society of Canada Annual Meeting, May, 2008. 1–10 Ottawa, Ontario, Canada. Available at: http://www.ssc.ca/survey/documents/SSC2068_5_Rais.pdf.
- Rivest, L.-P. and M. Hidirolou. 2004. "Outlier Treatment for Disaggregated Estimates." In Proceedings of the Section on Survey Research Methods: American Statistical Association, August, 2004. 4248–4256 Toronto, Ontario, Canada. Available at: <http://www.amstat.stat.org/sections/SRMS/Proceedings/y2004/files/Jsm2004-000149.pdf>
- Rousseeuw, P. and A.M. Leroy. 2003. *Robust Regression and Outlier Detection*. Hoboken, NJ: Wiley Series in Probability and Mathematical Statistics.
- Shewhart, W.A. 1939. *Statistical Method from the Viewpoint of Quality Control*. Washington, DC: Dover.
- Sigman, R.S. 2002. "Statistical Methods Used to Detect Cell-Level and Respondent-Level Outliers in the 2002 Economic Census of the Services Sector." In Proceedings of the Section on Survey Research Methods: American Statistical Association, August 2002. 3566–3573 Minneapolis, MN, USA. Available at: <https://www.amstat.org/sections/SRMS/Proceedings/y2005/Files/JSM2005-000465.pdf>
- Tambay, J.-L. 1988. "An Integrated Approach for the Treatment of Outliers in Sub-Annual Economic Surveys." In Proceedings of the Section on Survey Research Methods: American Statistical Association, 229–234. Available at: <http://www.amstat.org/sections/SRMS/Proceedings/papers/1988-040.pdf>
- Thompson, K.J. 1999. "Ratio Edit Tolerance Development Using Variations of Exploratory Data Analysis (EDA) Resistant Fences Methods." In Proceedings of the 1999 Federal Committee on Statistical Methodology Research Conference, November 1999. 1–10. Arlington, VA, USA. Available at: https://fcsml.sites.usa.gov/files/2014/05/VII-B_Thompson_FCSM1999.pdf.
- Thompson, K.J. 2007. "Investigation of Macro Editing Techniques for Outlier Detection in Survey Data." In Proceedings of the Third International Conference on Establishment Surveys (ICES-III), Survey Methods for Businesses, Farms, and Institutions, June 2007. 1186–1193. Montreal, Quebec, Canada. Available at <http://www.amstat.org/meetings/ices/2007/proceedings/ICES2007-000071.pdf>.
- Thompson, K.J. and S.A. Adeshiyan. 2003. "Data Quality Effects of Alternative Edit Parameters." *Journal of Data Science* 1: 1–25.
- Thompson, K.J. and R.S. Sigman. 1999. "Statistical Methods for Developing Ratio Edit Tolerances for Economic Data." *Journal of Official Statistics* 15: 517–535.

- Tukey, J.W. and D.H. McLaughlin. 1963. "Less Vulnerable Confidence and Significance Procedures for Location Based on a Single Sample: Trimming/Winsorization 1." *Sankhyā: The Indian Journal of Statistics (Series A)* 25: 331–352.
- Young, D.S. 2010. "Tolerance: An R package for Estimating Tolerance Intervals." *Journal of Statistical Software* 36: 1–39. Available at: <http://www.jstatsoft.org/v36/i05/> (accessed February 13, 2015)
- Yuen, K.-V. and H.-Q. Mu. 2012. "A Novel Probabilistic Method for Robust Parametric Identification and Outlier Detection." *Probabilistic Engineering Mechanics* 30: 48–59. DOI: <http://dx.doi.org/10.1016/j.probengmech.2012.06.002>

Received August 2013

Revised April 2014

Accepted June 2014