

Detecting outliers in weighted univariate survey data

Anna Pauliina Sandqvist*

October 27, 2015
Preliminary Version

Abstract

Outliers and influential observations are a frequent concern in all kind of statistics, data analysis and survey data. Especially, if the data is asymmetrically distributed or heavy-tailed, outlier detection turns out to be difficult as most of the already introduced methods are not optimal in this case. In this paper we examine various non-parametric outlier detection approaches for (size-)weighted growth rates from quantitative surveys and propose new respectively modified methods which can account better for skewed and long-tailed data. We apply empirical influence functions to compare these methods under different data specifications.

JEL Classification: C14

Keywords: Outlier detection, influential observation, size-weight, periodic surveys

1 Introduction

Outliers are usually considered to be extreme values which are far away from the other data points (see, e.g., Barnett and Lewis (1994)). Chambers (1986) was first to differentiate between *representative* and *non-representative* outliers. The former are observations with correct values and are not considered to be unique, whereas non-representative outliers are elements with incorrect values or are for some other reasons considered to be unique. Most of the outlier analysis focuses on the representative outliers as non-representatives values should be taken care of already in (survey) data editing.

The main reason to be concerned about the possible outliers is, that whether or not they are included into the sample, the estimates might differ considerably. The data points with substantial influence on the estimates are called *influential observations* and they should be

*Authors address: KOF Swiss Economic Institute, ETH Zurich, Leonhardstrasse 21 LEE, CH-8092 Zurich, Switzerland. Authors e-mail: sandqvist@kof.ethz.ch

identified and treated in order to ensure unbiased results. Quantitative data, especially from non-probability sample surveys, is often being weighted by size-weights. In this manner, unweighted observations may not itself be outliers (in the sense that they are not extreme values), but they might become influential through their weighting.

If the data generating process (DGP) is unknown, as it is usually the case with survey data, parametric outlier detection methods are not applicable. In this paper we investigate non-parametric methods of influential observation detection in univariate data from quantitative surveys, especially in size-weighted growth rates respectively ratios. This kind of data sources could be, for example, retail trade and investment surveys. As survey data is often skewed, we examine outlier detection in case of asymmetrically distributed data and data from heavy-tailed distributions. Further speciality of survey data is, that there are usually much more smaller units than bigger ones, and those are also more volatile. Yet, as the outlier detection (and treatment) involves in general lot of subjectivity, we pursue an approach in which no (or hardly any) parameters need to be defined but which would still be able to deal with data from different non-symmetric and fat-tailed distributions. We present some new respectively modified outlier detection methods to address these topics. We compare the presented outlier detection approaches using the empirical influence functions and analyse how they perform compared to the approach of Hidioglou and Berthelot (1986). Furthermore, we focus on analysing influential observations in (sub)samples with only small number of observations. Moreover, as many surveys are periodic, a dynamic and effective procedure is pursued.

This paper is organized as follows. In the next section, we describe the most relevant outlier detection approaches and discuss the location and scale estimates. In the third section we present the outlier detection methods and in the fifth section we analyse them with the empirical influence functions. Finally, the sixth section concludes.

2 Overview of outlier detection approaches

Various non-parametric as well as parametric techniques to detect outliers have already been introduced. A further distinction between the procedures is if it is applied only once (block procedure) or consecutively/sequentially. Sequential methods are sometimes preferred if multiple outliers are expected as in some block approaches the *masking effect* (explained later) could hinder to recognize multiple outliers. However, sequential approaches are usually computationally intensive and therefore, more time-consuming.

Commonly applied non-parametric approaches include rule-based methods like those based on the *relative distance*. The relative distance of the observation is defined in general as the absolute value of the distance to a location estimate l divided by a scale estimate s :

$$rd_i = \frac{|y_i - l|}{s}$$

If this test statistic exceeds some value k , which is defined by the researcher/data analyst,

then the observation is considered to be an outlier. Also a *tolerance interval* of $l - k_l s, l + k_h s$ can be applied where k_l and k_h can also differ in order to be able to account for the skewness of the data.

A further outlier detection method has been developed by Hidiroglou and Berthelot (1986) (the HB-method). Their method is applied to data transformed as a ratio from the previous period to the current one, which is then transformed to access the possible skewness of the data. According to this method outliers will be those observations whose period-to-period growth rates diverge significantly from the general trend of other observations of the same subset. This method is based on the idea having an acceptance boundary that differs according to the size of the observation as the change ratios tend to be more volatile for smaller observations and therefore, the smaller units are more probable to be identified as an outlier. This is the so called size masking effect. However, the selection of the values for the parameters can be difficult.

The parametric approaches underlie distributional assumption and are therefore rather seldom applied to survey data, as an assumption of specific DGP is generally unrealistic. However, if the data can be transformed to be approximately normally distributed, some of these methods might be then useful. Grubbs (1950) introduce an outlier test which assumes that the data is normally distributed. The test statistic is defined as follows:

$$G = \frac{\max(|y_i| - \bar{y})}{s}$$

where \bar{y} is the sample mean and s is the standard deviation of the sample. However, for the Grubbs test the number of outliers must be defined beforehand and it works only in cases of one outlier, two outliers on one side or one outlier on each side. The Grubbs test is also not suitable for small samples as the test statistic can per definition not be larger than $(n - 1)/\sqrt{n}$ where n is the sample size. A generalized version of the Grubbs test was introduced by Tietjen and Moore (1972) which can be used to test for multiple outliers. However, also here the exact number of outliers need to be chosen beforehand.

As the efficiency of standard deviation as an estimator of scale might be poor for samples containing only few observations, alternative measures have been examined. Dean and Dixon (1951) show that the range is a highly efficient measure of the dispersion for ten or fewer observations. They introduce a simple Q-test for outliers in samples with 3-30 observations, where the test statistics Q can be written as:

$$Q = (y_n - y_{n-1}) / (y_n - y_1)$$

If Q exceeds the critical value, the observation is considered to be an outlier. However, the range is sensitive to outliers and therefore, in case of multiple outliers this test is not optimal. Also the distance to nearest neighbour is not a suitable measure in that case.

2.1 Estimators for location and scale

Mean and standard deviation are the mostly used estimators for location and scale. Yet, if multiple outliers exists, one should be careful using mean and standard deviation as estimates of sample location and scale in the outlier detection techniques as they are very sensitive to extreme data points and could thus hinder the identification of outliers. This is called the *masking effect*. Both mean and standard deviation have a breakdown point of zero, what means that they "break" down already if only one outlying observation is in the sample.

An alternative is to apply more robust measures of location and scale, which are less sensitive to outliers. For example, median has a breakdown point of 50 percent and is therefore a good measure of location even if the data is skewed. A further robust estimators for location include the Hodges-Lehmann estimator (HL) (Hodges and Lehmann (1963)), trimmed mean, Winzorized mean as well as Huber M-estimator of a location parameter.

An alternative scale measure is *Median absolute deviation* (MAD) which is defined as

$$MAD = \text{median}_i |y_i - \text{median}_j(y_j)|$$

and it has the highest possible breakdown point of 50 percent. However, MAD is based on the idea of symmetrically distributed data as it corresponds the symmetric interval around the median containing 50 percent of the observations. A further alternative for a scale estimator is the *interquartile range* (IQR), i.e. the difference between upper and lower quantiles. The interquartile range is not based on a symmetric assumption, however it has a lower breakdown point of 25 percent. In addition, Rousseeuw and Croux (1993) introduce two alternatives for MAD, which are similar to it but do not assume symmetrically distributed data. They propose a scale estimator based on different order statistics. Their estimator is defined as

$$Q_n = d\{|x_i - x_j|; i < j\}_{(k)} \quad (1)$$

where d is a constant and $k = \binom{h}{2} \approx \binom{n}{2}/4$ whereas $h = [n/2] + 1$ corresponds about the half of the number of observations. This estimator is also suitable for asymmetrically distributed data and has a breakdown point of 50 percent.

3 Methods

Let y_i be the (year-on-year) growth rate of a company i and w_i be its size weight. Then the simple point estimator for the weighted mean of the sample with n observations is as follows

$$\hat{\mu}_n = \frac{\sum w_i y_i}{\sum w_i} \quad (2)$$

Our goal is the detect outliers, which have considerable influence on this estimator. As in this context size weights are being used, we need to analyse both unweighted and weighted

observations. Furthermore, we take into account that the observations might be asymmetrically distributed as it is often the case with survey data and/or the distribution of the data exhibits heavy tails. However, no specific distribution assumption can and will be made.

In general we have different possibilities to deal with skewness of the data in outlier detection: transform observations to be approximately symmetrically distributed, to use different scale measure each side of the median (based on skewness-adjusted acceptance interval) or apply an estimator of scale which is robust also in the case of asymmetric data.

3.1 Baseline approach

We use the method of Hidioglou and Berthelot (1986) as our baseline approach. They use in their method the following transformation technique

$$s_i = \begin{cases} 1 - q_{0.5}/r_i & \text{if } 0 < r_i < q_{0.5} \\ r_i/q_{0.5} - 1 & \text{if } r_i \geq q_{0.5} \end{cases}$$

where $r_i = y_{i,t+1}/y_{i,t}$ and $q_{0.5}$ is the median of r values. Thus, one half of the transformed values are less than zero and the other half greater than zero. However, this transformation does not ensure symmetric distribution of the observations. In the second step the size of the observations is taken into account:

$$E_i = s_i \{ \max([y_i(t), y_i(t+1)]) \}^U$$

where $0 \leq U \leq 1$. The parameter U controls the importance of the magnitude of the data. The acceptance interval of Hidioglou and Berthelot (1986) is as follows

$$\{E_{med} - cD_{Q1}, E_{med} + cD_{Q3}\}$$

where $D_{Q1} = \max\{E_{med} - E_{Q1}, |aE_{med}|\}$ and $D_{Q3} = \max\{E_{Q3} - E_{median}, |aE_{med}|\}$ and E_{med} is the median of the HB-statistics E , E_{Q1} and E_{Q3} are the first and third quartile and a is a parameter, which takes usually the value of 0.05. The idea of the term $|aE_{med}|$ is to hinder that the interval becomes very small in the case that the observations are clustered around some value. The constant c defines the width of the acceptance interval. However, for this approach the parameters U , a and especially c need to be more or less arbitrarily chosen or estimated. We use this method as baseline approach with following parameter values: $a = 0.05$, $c = 7$ and $U = 1$. The parameter c could also take different values for each side of the interval and thus, account better for asymmetric data.

3.2 Transformation approach

One way to deal with asymmetric data is to use transformation which produces then (approximately) symmetrically distributed observations. Then, in general, methods assuming normally

distributed data could be applied. In this approach we transform the data to be approximately symmetrically distributed. For this, we apply a transformation based on Yeo and Johnson (2000) work:

$$t_i = \begin{cases} ((y_i + 1)^\lambda - 1)/\lambda & \text{for } y_i \geq 0, \lambda \neq 0 \\ -((-y_i + 1)^{2-\lambda} - 1)/(2 - \lambda) & \text{for } y_i < 0, \lambda \neq 2 \end{cases}$$

where y_i are growth rates in our case. However, we apply the same λ parameter for positive as well as negative values. This ensures that positive and negative observations are treated same way. We choose following parameter value: $\lambda = 0.8$. The transformed values are then multiplied by their size weight w_i :

$$t_i^w = t_i w_i^V$$

The parameter $0 \leq V \leq 1$ controls the importance of the weights like in the approach of Hidiroglou and Berthelot (1986). A tolerance interval based on the interquartile range is then applied as these transformed values t_i^w should be approximately symmetrically distributed: $q_{0.25} - 3 IQR, q_{0.75} + 3 IQR$. We refer to this approach as *tr-approach*.

3.3 Right-left scale approach

In this approach we apply different scale measures for each side of the median to account for the skewness of the data. Hubert and Van Der Veecken (2008) propose *adjusted outlyingness* (AO) measure based on different scale measures on both sides of the median:

$$AO_i = \begin{cases} \frac{y_i - q_{0.5}}{w_u - q_{0.5}} & \text{if } y_i > q_{0.5} \\ \frac{q_{0.5} - y_i}{q_{0.5} - w_l} & \text{if } y_i < q_{0.5} \end{cases}$$

where w_l and w_u are the lower and upper whisker of the adjusted boxplot as in Hubert and Vandervieren (2008):

$$\begin{aligned} w_l &= q_{0.25} - 1.5^{-4MC} IQR; w_u = q_{0.75} + 1.5^{3MC} IQR & \text{for } MC \geq 0 \\ w_l &= q_{0.25} - 1.5^{-3MC} IQR; w_u = q_{0.75} + 1.5^{4MC} IQR & \text{for } MC < 0 \end{aligned}$$

where MC equals to the *medcouple*, a measure of skewness, introduced by Brys et al. (2004):

$$MC(Y_n) = \frac{\text{med}_{y_i < \text{med}_n < y_j} h(y_i, y_j)}{\text{med}_n}$$

where med_n is the sample median and

$$h(y_i, x_j) = \frac{(x_j - \text{med}_n) - (\text{med}_n - x_i)}{x_j - x_i}$$

However, the denominator in AO can become quite small if the values are clustered near to each other (on one side), and then, more observations could be recognized as outliers as actually the case is. On the other hand, the denominator can also become very large, if, for example, one (or more) clear outlier is in the data and thus, other observations may not be declared as outliers but are kind of masked by this effect. Therefore, we adjust the denominator to account for these cases where it could become relatively low or high as follows:

$$rl_i = \begin{cases} \frac{y_i - q_{0.5}}{\min(\max(w_u - q_{0.5}, \text{range}(y)/n_y^{0.6}), \text{range}_y/n_y^{0.6})} & \text{if } y_i \geq q_{0.5} \\ \frac{y_i - q_{0.5}}{\min(\max(q_{0.5} - w_l, \text{range}(y)/n_y^{0.6}), \text{range}_y/n_y^{0.6})} & \text{if } y_i < q_{0.5} \end{cases}$$

In the second step, this measure is also multiplied by the size weight:

$$rl_i^w = rl_i * w_i^V$$

where the parameter $0 \leq V \leq 1$ again controls the importance of the weights. The interquartile range interval $q_{0.25} - 3 IQR, q_{0.75} + 3 IQR$ is applied to identify the outliers. We refer to this approach as *rl-approach*.

3.4 Q_n -approach

This approach is based on the measure of relative distance. We use the Q_n estimator of Rousseeuw and Croux (1993) as scale estimate, which is more robust scale estimator than MAD for asymmetrically distributed data. In the first step the distance to median divided by Q_n is calculated for unweighted values:

$$Q_i = \frac{y_i - \text{median}(y)}{Q_n(y)}$$

This value is then multiplied by the size weight of the observation and the parameter $0 \leq V \leq 1$ is introduced to control the influence of the weights:

$$Q_i^w = Q_i * w_i^V$$

We apply the following interquartile interval for Q_i^w : $q_{0.25} - 3 IQR, q_{0.75} + 3 IQR$. We refer to this approach as *Qn-approach*.

4 Empirical influence function analysis

With the *empirical influence function* (EIF) the influence of a single observation on the estimator can be studied. To examine how an additional observation respectively multiple additional observations affect the different estimators, we make use of the empirical influence functions. In

this way, we can compare how the considered outlier detection approaches behave in different data sets. We define the EIF as follows:

$$EIF_{\hat{\mu}}(w_0 y_0) = \hat{\mu}(w_1 y_1, \dots, w_{n-1} y_{n-1}, w_0 y_0)$$

where y_0 is the additional observation.

For this analysis we use randomly generated data to examine how these methods work for data with different levels of skewness/asymmetry and heavy-tailness in various sample sizes. First, the weights of the observations w_i are drawn from exponential distribution with $\lambda = 1/4$ truncated at 20 and round up to make sure no zero values are possible. Therefore, we have more smaller companies and 1 to 20 is the range of the size weight classes we take into account. Thereafter, the actual observations y_1, \dots, y_{n-1} , which are in this case growth rates respectively ratios, are randomly generated from normal distribution with mean = 0 and standard deviation = 3. In the following figures we present the EIFs of the estimators $\hat{\mu}_{HB}$, $\hat{\mu}_{tr}$, $\hat{\mu}_{rl}$ and $\hat{\mu}_{Qn}$ for various sample sizes. The $\hat{\mu}_{n-1}$ denotes the weighted mean of the original sample without the additional observation. In the following figures the x-axis denotes the unweighted values (y_0).

We let the additional observation y_0 to take values between -50 and 50. To analyse the effect of the additional observation first in the simplest possible case, we first examine the effect among unweighted observations i.e. we adjust all size weights w_i to be one, also for the added observation. As long as the EIF is horizontal, it means that there is no change in the estimator for those y_0 values, i.e. the additional observation is considered to be an outlier (and deleted) and therefore, there is no effect on the estimator. The results are presented in Figure 1. The methods seem to behave quite similarly expect the case of sample size equal to 20. In this case the HB-method seems to detect also an other or different outlier as the estimator $\hat{\mu}_{HB}$ do not correspond the original estimate $\hat{\mu}_{n-1}$ in the horizontal areas.

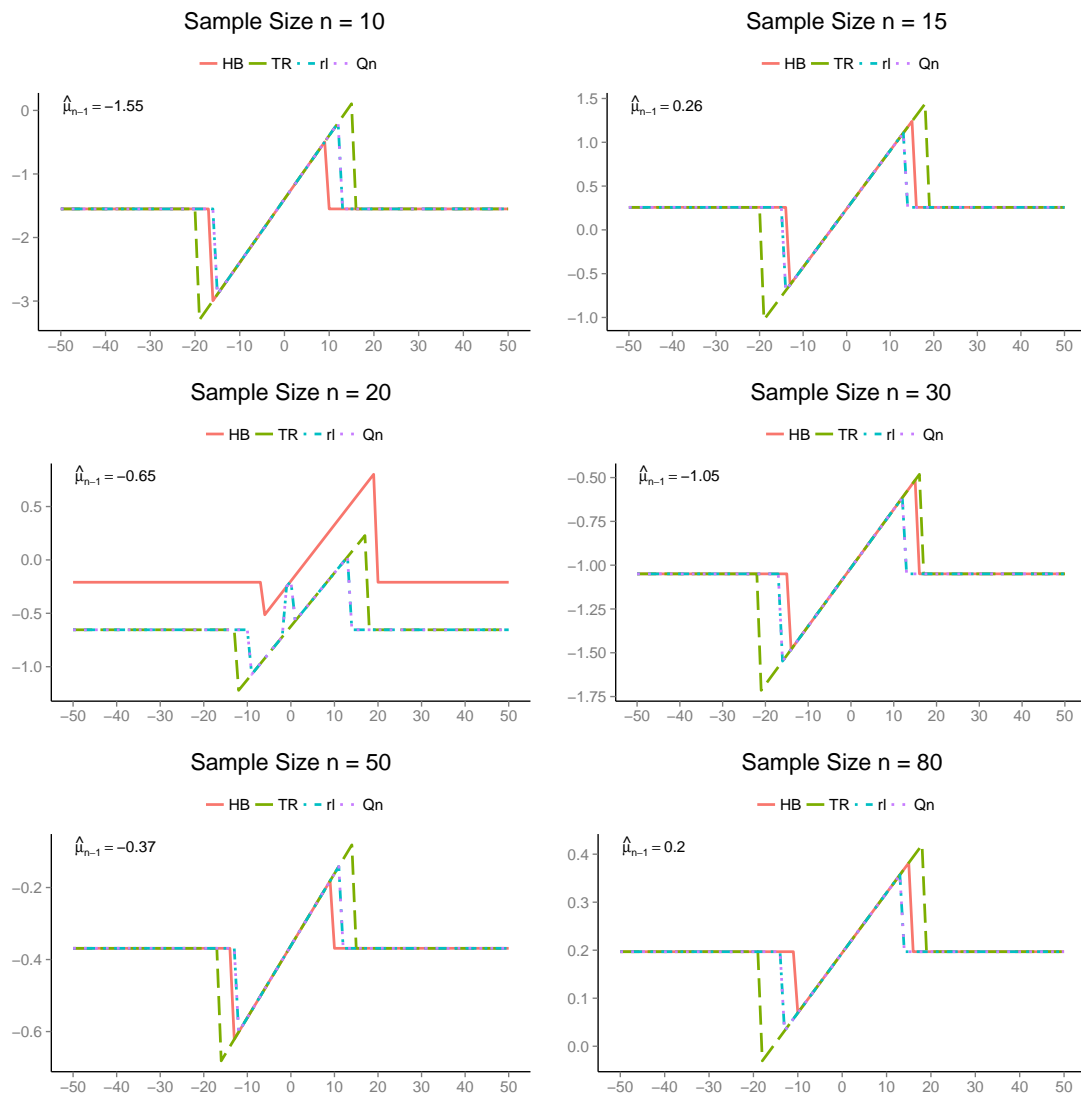


Figure 1: Influence of one additional observation on the estimators in case of unweighted observations

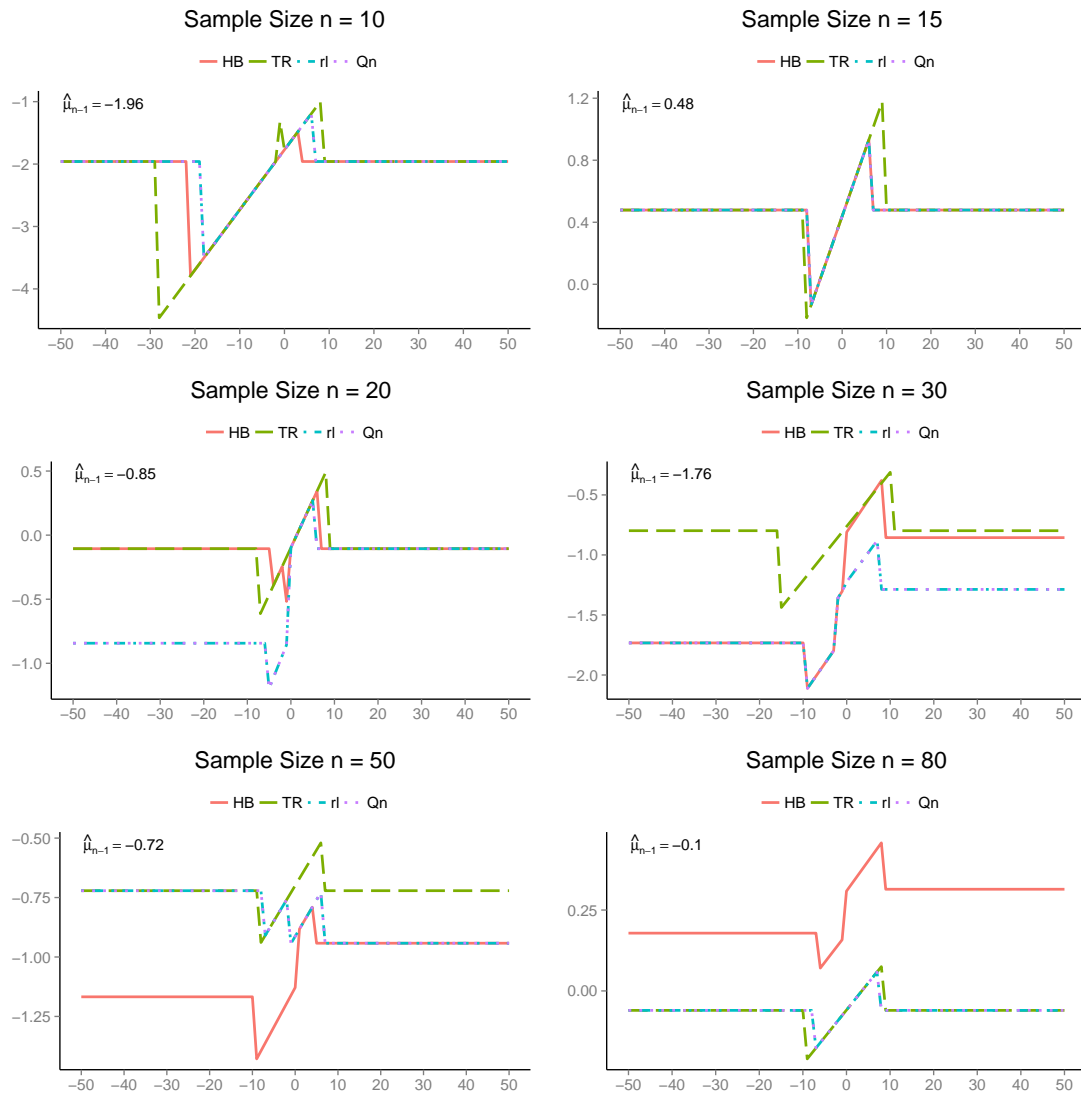


Figure 2: Influence of one additional observation on the estimators

In Figure 2 the same plots are presented for the standard case with the size weights. Here, we assume $w_0 = 5$ so that the weighted values lie between -250 and 250. We can observe that the estimators behave quite similarly as long as the sample size is small but are affected by the one additional observation $w_0 y_0$ quite differently as the sample size gets bigger. For all, some methods tend to identify also other outliers besides the additional one, as they diverge from the $\hat{\mu}_{n-1}$ in the horizontal areas, like the TR-approach for sample sizes $n = 20$ and $n = 30$ and the HB-method for sample sizes $n = 50$ and $n = 80$.

Furthermore, also the effect of multiple increasing replacements can be studied with EIF. In

this case, the EIF can be written as follows:

$$EIF_{\hat{\mu}} = \hat{\mu}(w_0 y_0, \dots, w_0 (y_0 * j + 1), w_{j+1} y_{j+1}, \dots, w_{n-1} y_{n-1}), j = 1, \dots, (n/2)$$

where j denotes the number of replacements. Here, we assume that $y_0 = 5$ and $w_0 = 5$. Otherwise, the same data is used as in the previous case of single additional observation. This exercise demonstrates the case of right skewed data. In Figure 3, the number of replacements j is on the x-axis. The vertical ticks mark the weighted mean $\hat{\mu}$ of each sample. If the curves are beneath those ticks, this indicates that some of the positive observations are being considered as outliers and therefore, the estimator is lower than the original one. On the other hand, if the curve is above the ticks, then this shows that this method is labelling some negative or low observations as outliers. This seems to be the case for the HB-method for samples sizes $n = 20$ or greater, what implies that the HB-method tends to be in some cases too sensitive if the data is (very) right skewed.

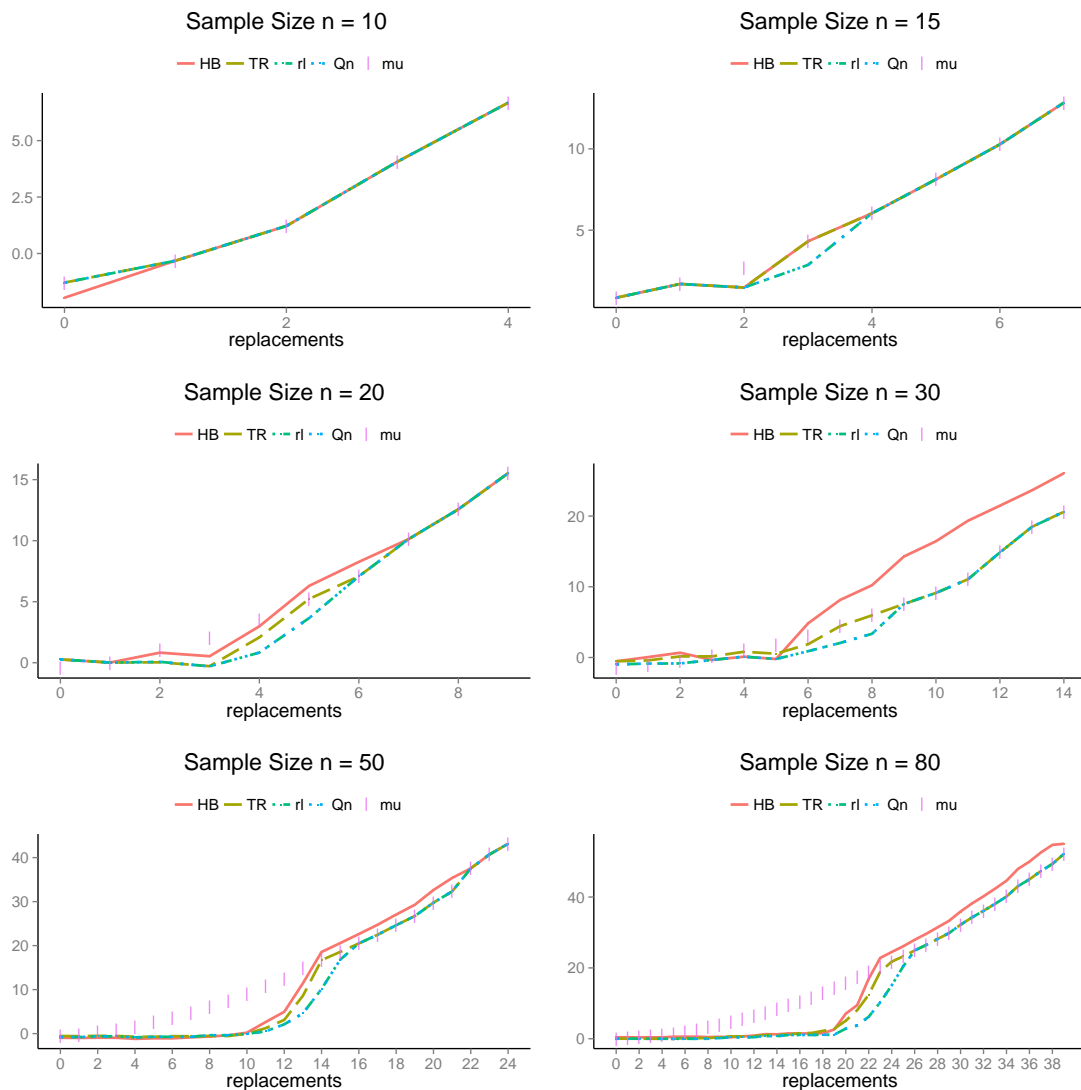


Figure 3: Effect of multiple (positive) replacements on the estimators

To analyse the various outlier detection approaches in case of long-tailed data, we repeat the empirical influence function analysis for randomly generated data from truncated non-standard t -distribution (mean = 0, standard deviation = 3, lower bound = -100, upper bound = 1000, degrees of freedom = 3). In Figure 4 the EIFs for the considered estimators are presented. In this case, the EIFs of the estimators tend to divers more from each other than in the case of normally distributed data. There is no clear picture, how the estimators behave for this data sets, but the HB-method seems to work well for big samples size ($n = 80$), whereas one of the other estimators (depending on the sample size which one) seems to work more appropriate for smaller sample sizes. In Figure 5 the plots for the case of replacements by increasing outliers are

presented. Also here, the the HB-method appears to be sensitive with the right skewed data for some sample sizes, however it is less pronounced than in the case of normally distributed data.

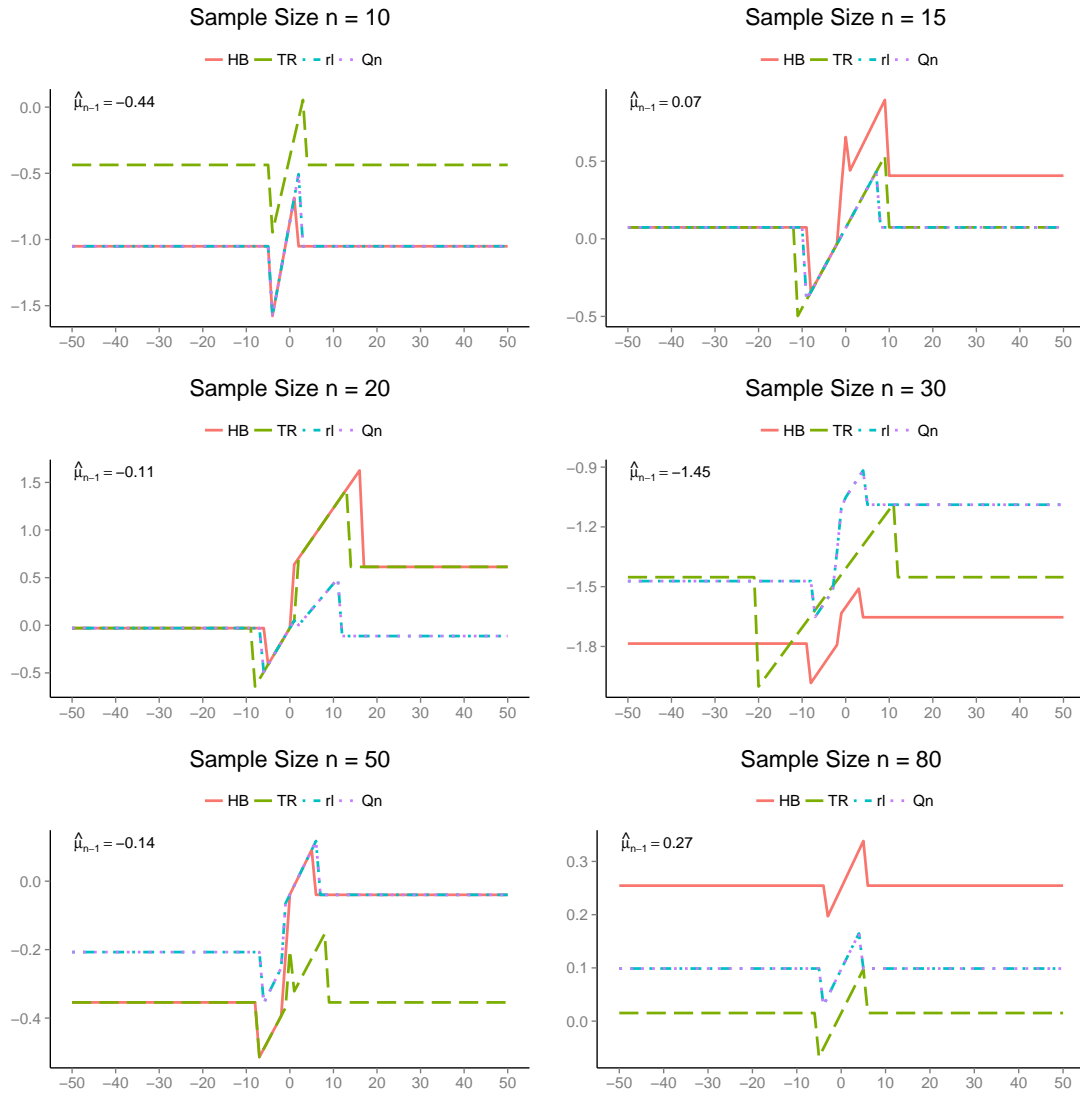


Figure 4: Influence of one additional observation on the estimators in case of fat-tailed data

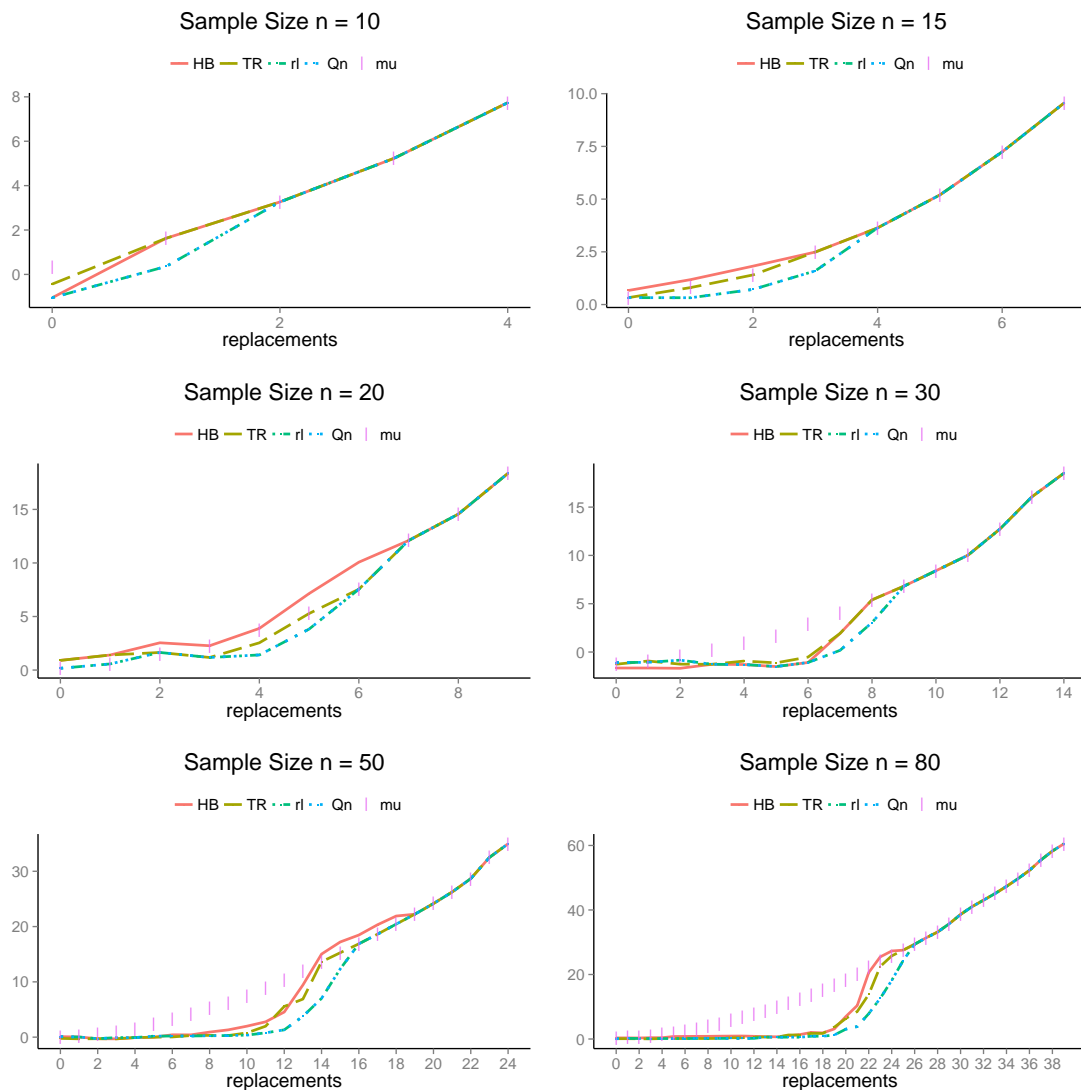


Figure 5: Effect of multiple (positive) replacements on the estimators in case of fat-tailed data

5 Simulation analysis

In this section we conduct a simulation study. To examine how these methods work for data with different levels of skewness/asymmetry and heavy-tailness, we analyse the methods for data sets generated from two different distributions with varying parameters. Within each simulation, we generate 1000 samples of sizes 15, 30, 50 and 100 observations. Again, the weights of the observations w_i are drawn from exponential distribution with $\lambda = 1/6$ truncated at 20 and round up to make sure no zero values are possible. The weight categories are usually

derived from turnover or employees of a company. Thereafter, to generate data with fat tails, the actual observations y_i are generated from truncated non-standard t-distribution distribution with varying values for degrees of freedom (lower bound -, 100, upper bound 1000, mean equal to 5, standard deviation equal to 20). For observation with smaller weight (1 to 9) higher standard deviation (+10) is applied in order to be able to account for the fact that the smaller observations tend to have more volatile growth rates.

To get a better view how these methods work depending on the tail heaviness of the data distribution, we compare the percentage of observations declared as outliers by each method over the different levels of heavy-tails, i.e number of degrees of freedom. In Figure 6 we can observe that the HB-method identifies the most outliers across all sample sizes as well as numbers of degrees of freedom whereas the tr-approach finds always the fewest outliers. Furthermore, all the methods tend to identify more outliers when the data has heavier tails across all the sample sizes, otherwise the average percentages seem to be quite constant for all methods across all sample sizes.

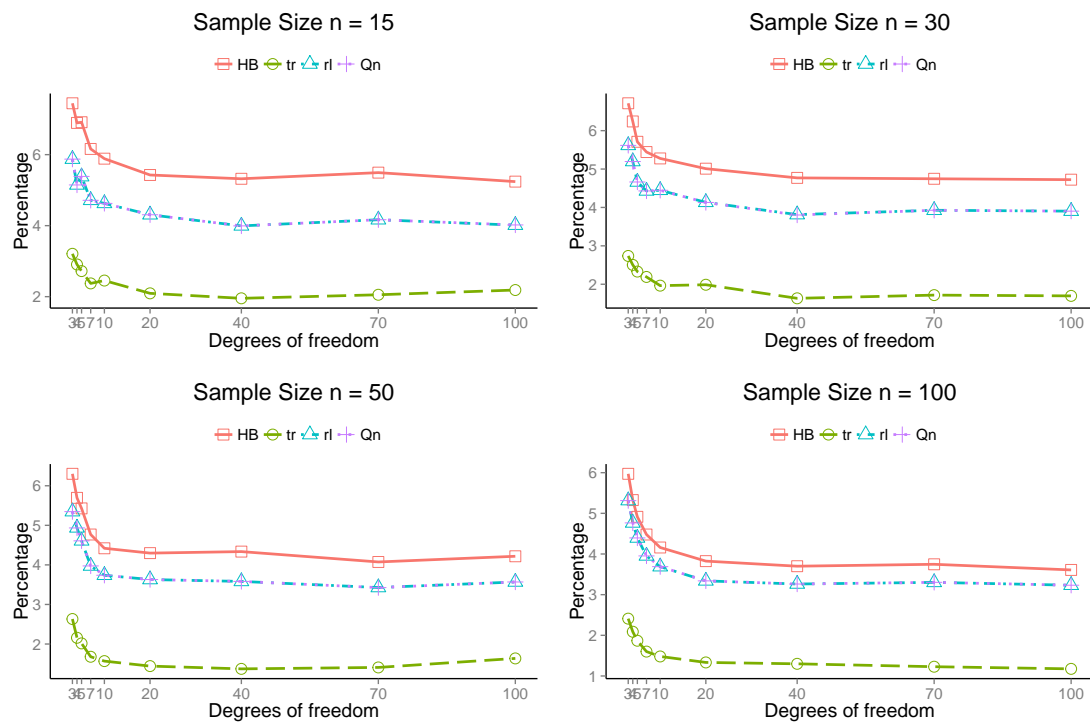


Figure 6: Average percentage of observation declared as outliers depending on the heavy-tailness of the data

To produce asymmetric data sets (of ratios), we generated data as a sum of two random variables drawn from at zero truncated normal (standard deviation equal to 0.01 for big companies, and 0.02 for smaller companies) and exponential distribution where as the mean of the normal

distribution was set to between 0.6 and 1.0 and the mean of the exponential distribution was adjusted so that the total mean stays constant at 1.05 or respectively 5% . Therefore, the data is the more skewed the lower the mean of the normal distribution. In the following plots, the x-axis denotes the mean of the normal distribution and therefore, at very left is the data point with highest level of skewness.

Again, the average percentage of outliers labelled as outliers by the four methods are compared over the level of skewness of the data. In Figure 7 we can see that the tr-approach detects also here the fewest outliers over all the sample sizes. For the smallest sample size, the HB-method finds the most outliers, whereas for the biggest sample size the rl- and Qn-approaches detect the most outlying observations.

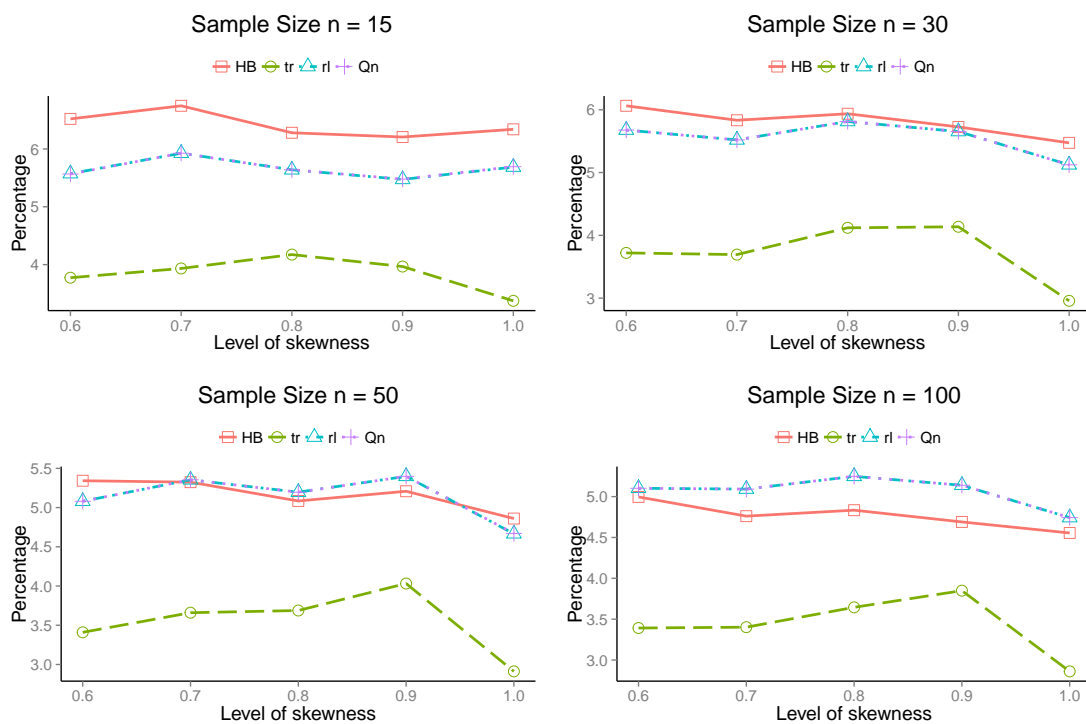


Figure 7: Average percentage of observation declared as outliers depending on the skewness of the data

All together, it seems that in case of asymmetric data, the HB-approach do not necessarily detect more outliers than rl- or Qn-approach but rather labels different observations as outliers, as these methods lead partly to different estimates as shown in Figure 3 and 5.

6 Conclusions

In this paper we analyse outlier detection in size-weighted quantitative univariate (survey) data. We pursue to find a non-parametric method with predefined parameter values which can account for asymmetrically distributed data as well as data from heavy-tailed distributions. New respectively modified outlier detection approaches are being introduced for this purpose. We analyse these methods and compare them to each other and to the method of Hidioglou and Berthelot (1986) with help empirical influence functions. In general, all the studied methods seem to work well both in case of asymmetric as well long-tailed data. We find some evidence that the HB-method (with fixed parameter c value) can be too sensitive in case of asymmetric respectively right-skewed data and detecting too many outliers in the left tail of the data. The other methods seem to be able deal better with right-skewed data.

References

- Barnett, V. and Lewis, T. (1994). *Outliers in statistical data*. Wiley, New York, 3. edition.
- Brys, G., Hubert, M., and Struyf, A. (2004). A robust measure of skewness. *Journal of Computational and Graphical Statistics*, 13(4):996–1017.
- Chambers, R. L. (1986). Outlier robust finite population estimation. *Journal of the American Statistical Association*, 81(396):1063–1069.
- Dean, R. B. and Dixon, W. J. (1951). Simplified Statistics for Small Numbers of Observations. *Analytical Chemistry*, 23(4):636–638.
- Grubbs, F. E. (1950). Sample Criteria for Testing Outlying Observations. *The Annals of Mathematical Statistics*, 21(1):27–58.
- Hidioglou, M. and Berthelot, J.-M. (1986). Statistical Editing and Imputation for Periodic Business Surveys. *Survey Methodology*, 12:73–83.
- Hodges, J. L. and Lehmann, E. L. (1963). Estimates of Location Based on Rank Tests. *The Annals of Mathematical Statistics*, 34(2):598–611.
- Hubert, M. and Van Der Veeken, S. (2008). Outlier detection for skewed data. *Journal of Chemometrics*, 22(3-4):235–246.
- Hubert, M. and Vandervieren, E. (2008). An adjusted boxplot for skewed distributions. *Computational Statistics and Data Analysis*, 52(12):5186–5201.
- Rousseeuw, P. J. and Croux, C. (1993). Alternatives to the Median Absolute Deviation. *Journal of the American Statistical Association*, 88(424):1273–1283.
- Tietjen, G. L. and Moore, R. H. (1972). Some Grubbs-type statistics for the detection of several outliers. *Technometrics*, 14(3):583–597.
- Yeo, I.-K. and Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, 87(4):954–959.

A Appendix: Robustness