

Journal of Official Statistics, Vol. 37, No. 1, 2021, pp. 97-119, http://dx.doi.org/10.2478/JOS-2021-0005

Identifying Outliers in Response Quality Assessment by Using Multivariate Control Charts Based on Kernel Density Estimation

Jiayun Jin¹ and Geert Loosveldt¹

When monitoring industrial processes, a Statistical Process Control tool, such as a multivariate Hotelling T^2 chart is frequently used to evaluate multiple quality characteristics. However, research into the use of T^2 charts for survey fieldwork-essentially a production process in which data sets collected by means of interviews are produced-has been scant to date. In this study, using data from the eighth round of the European Social Survey in Belgium, we present a procedure for simultaneously monitoring six response quality indicators and identifying outliers: interviews with anomalous results. The procedure integrates Kernel Density Estimation (KDE) with a T^2 chart, so that historical "in-control" data or reference to the assumption of a parametric distribution of the indicators is not required. In total, 75 outliers (4.25%) are iteratively removed, resulting in an in-control data set containing 1,691 interviews. The outliers are mainly characterized by having longer sequences of identical answers, a greater number of extreme answers, and against expectation, a lower item nonresponse rate. The procedure is validated by means of ten-fold crossvalidation and comparison with the minimum covariance determinant algorithm as the criterion. By providing a method of obtaining in-control data, the present findings go some way toward a way to monitor response quality, identify problems, and provide rapid feedbacks during survey fieldwork.

Key words: Kernel density estimation; Hotelling T^2 chart; multivariate control charts; response quality; ten-fold cross-validation.

1. Introduction

In a process such as bottling beer, data including the weight of bottles is usually collected and monitored, in order to check whether the process is operating as expected, and to detect underfilled or overfilled bottles. A proven tool for pursuing this goal is known as Shewhart control chart (Shewhart 1931), which presents the fluctuating patterns of the data collected from a process over time. It allows control limits to be calculated using statistical equations, and a process is defined as "in control" if fluctuations fall within the control limits, otherwise the process is defined as "out of control". In domains such as manufacturing, public health, and financial services, control charts have been extensively used to assess process performance and to identify and examine defects (MacCarthy and Wasusri 2002).

¹Catholic University of Leuven, Centre for Sociological Research, Parkstraat 45, bus 3601, 3000 Leuven, Belgium. Emails: jiayun.jin@kuleuven.be, and geert.loosveldt@soc.kuleuven.be

Survey fieldwork–although it by nature varies according to the mode of data collection–is essentially a production process, in which data sets collected by means of interviews are produced. However, the use of control charts for identifying "out-of-control" signals–such as the interviews or interviewers with abnormal results–has attracted little attention. The limited work available on control charts in a survey research context has only focused on univariate charts, which only monitor one variable at a time. For example, researchers have used control charts to monitor interview duration (Jin et al. 2019; Jans et al. 2013; Sirkis et al. 2011; Peng and Feld 2011), item nonresponse rate (Sirkis et al. 2010). However, there are various measurements that are informative about survey data quality. In this situation, using univariate control charts to monitor each variable individually has been found to be time consuming, misleading (MacGregor and Kourti 1995), and sometimes unfeasible in practice (Ferrer 2007).

In our previous work, we introduced a projection pursuit method known as Principal Component Analysis Mix (Chavent et al. 2014) to analyze a combination of numerical and categorical response quality indicators (Jin and Loosveldt 2020). We showed that the obtained principal components are capable of identifying and characterizing interviews that are not in control (hereafter termed as outliers). In the present study, we focus on the use and performance of the Hotelling T^2 control chart (T^2 chart) in identifying outliers based on multiple numerical response quality indicators. A T^2 chart monitors the T^2 statistic, which measures the squared Mahalanobis distance from an observation to the target values of the monitored variables, or the mean estimated from the data if the target values are unknown.

Conventionally, the control limit of a T^2 chart is calculated based on the percentile of an F distribution, assuming that the monitored variables follow a multivariate normal distribution (Mason and Young 2002). However, if this assumption does not hold, the control limit may generate many false alarms (Chou et al. 2001; Phaladiganon et al. 2011). In order to relax the multivariate normality assumption, in line with Chou et al. (2001), we use the kernel density estimation (KDE) technique to fit the distribution of the T^2 statistic.

Assuming that a set of historical in-control data is available, the control limit of the T^2 chart can immediately be established using the percentile of the derived kernel density curve of the T^2 statistic. Information about the in-control data and the obtained control limit can then be carried forward to actively monitor new observations from the process. The assumption of in-control data being available is reasonable and thus remarkably common in industrial applications such as chemical batch processes, because the operating conditions of the processes can be ensured as "normal" (see e.g., Ahsan et al. 2018; Kini and Madakyaru 2016; Costa et al. 2015; Phaladiganon et al. 2011). The vast majority of industrial applications for control charts are accordingly characterized as "Phase II" monitoring. This mainly revolves around prospective monitoring of new observations from a process, treating the in-control process parameters as known.

The production of interviews, however, is more complex due to the variations in interviewers, respondents, interview environments, and the involvement of human elements (such as attitude and motivation). Consequently, historical in-control data is usually not directly available, but needs to be first determined by conducting "Phase I" monitoring retrospectively, so that possible outliers can be detected and removed. In this

case, the control limit is unknown and needs to be estimated. Due to the impact that extreme outliers may have on the T^2 statistic and the estimated control limit, outliers are usually removed in an iterative fashion in Phase I (Montgomery 2009; Chakraborti et al. 2008; Ferrer 2007; Mason and Young 2002).

2. Situating the Research

The present study aims to address the research question of how to assess response quality and identify outlying interviews based on a set of quality indicators, without making assumptions about the underlying distributions of the indicators and without the availability of historical in-control data. We define outlying interviews as the minority of those in a data set that are characterized by response behaviour different from the majority. Accordingly, they are potentially problematic interviews resulting from undesirable response behavior.

By skewing the distribution parameters, outliers are known to have a masking effect (one outlier masks a second outlier: a false negative) and a swamping effect (one outlier swamps a second outlier: a false positive) on the whole data set. Accordingly, when identifying outliers, methods that are resistant to the distorting effect of outliers–generally known as robust methods–are needed. In the field of statistics, most of the robust methods for identifying outliers in multivariate data fall into two categories: first, methods based on Mahalanobis distance (in which our approach is situated) and second, methods based on projection pursuit, of which principal component analysis is a well-known special case.

Methods based on Mahalanobis distance have as the main requirements, obtaining robust estimates of the mean and the covariance matrix to compute a robust Mahalanobis distance, and determining a threshold to separate outliers from non-outliers (Rocke and Woodruff 1996). The most widely-used method in this category seems to be the minimum covariance determinant (MCD) (Rousseeuw 1984, 1985), usually implemented by the fast-MCD algorithm (Rousseeuw and Driessen 1999). MCD is a highly robust estimator of multivariate location and scatter. Its objective is to find *h* out of *n* observations (h > n/2) where the covariance matrix has the smallest determinant. The MCD estimates of location and scatter are then respectively the mean vector and the covariance matrix of these *h* observations (this method is not discussed in depth in this article, but for a detailed introduction please refer to Rousseeuw 1984, 1985). However, MCD and fast-MCD suffer from low computation efficiency, which limits their applicability in high-dimensional or large-scale problems.

In the current article, we present an iterative procedure that integrates the use of a T^2 chart with the non-parametric KDE technique. This procedure is designed in a different way than other Mahalanobis-based methods with regard to fulfilling the above presented requirements for obtaining robust estimators (of multivariate mean and covariance) and determining the threshold. We exclude outliers one at a time, and re-estimate the mean, covariance matrix, and the threshold with each successive outlier excluded. This is repeated until no outliers can be found. The iterative nature of the procedure ensures that the influence of outliers on the estimates of the mean and the covariance matrix is reduced, and a robust Mahalanobis distance (T^2 statistic in our context) is ultimately obtained. The

threshold (control limit in our context) is determined based on the percentile values of the fitted kernel distribution in each iteration.

To return to the survey context, this procedure will ultimately separate a set of incontrol data from the outliers. The obtained in-control data can (1) provide an understanding of the multiple quality indicators and their correlation structure for an incontrol survey process, and (2) serve as the reference data to be used for real-time monitoring from the beginning of new and comparable survey fieldwork in Phase II. The identified outliers, on the other hand, will pinpoint which response quality indicators are more important for detecting potentially problematic interviews. To evaluate the reliability and validity of the procedure and the results concerning the identified outliers and in-control data, a k-fold cross-validation is conducted. We also compare the results with those obtained using the MCD estimator, in order to validate our procedure from another perspective.

To illustrate the methods and address the research question, we use six numerical indicators of response quality that were constructed in our previous work (Jin and Loosveldt 2020). The indicators were developed based on the respondent satisficing theory (Krosnick 1991; Simon 1956). This theory proposes that respondents sometimes provide a "satisfactory" instead of an "optimal" answer, in order to minimize the cognitive effort needed. As indicators of reduced response quality, we accordingly select response styles that are deemed to be the consequences or expressions of satisficing behavior. We first use four of the five response styles detailed in the work of Yan et al. (2004), which comprise acquiescence, nondifferentiated answers, the selection of middle responses, and the selection of extreme responses. Moreover, we also consider item nonresponse, which measures respondents' repeated use of "don't know" answers, refusals, and no answers (Loosveldt and Beullens 2017).

The response styles and constructed indicators are listed in Table 1. The data used to calculate the indicators are from the eighth round of the European Social Survey (ESS8), collected in Belgium during 2016 and 2017. A total of 1,766 respondents participated in the ESS8, with a response rate of 56.8% (AAPOR Response Rate 1) via Computer Assisted Personal Interviews (CAPI).

Acquiescence refers to the tendency of a respondent to agree with all the items regardless of content. The indicator of acquiescence is constructed based on three pairs of items from the ESS8 questionnaire. These item pairs are worded in opposite directions toward the same construct. Detailed information about these items are displayed in Table 8, (Section 6, Appendix). In line with previous research (Rammstedt et al. 2017; Billiet and McClendon 2000), for each respondent an average score is computed by averaging all six items. Nondifferentiation is the tendency of a respondent to give identical answers to a block of items. In line with Loosveldt and Beullens (2017), to measure nondifferentiation we calculate the standard deviation of all responses of a respondent to one block of items and the maximum sequence of identical responses. In the ESS8 questionnaire, two blocks of items are suitable to calculate these two indicators. They are listed in Table 9 (Section 6, Appendix). For each of the two blocks of items, the standard deviation (*sd*) and the maximum identical sequence (*max*) of the responses are calculated separately for each respondent. The average values for the two blocks of items are used as the measurements of these two indicators. In the same way, the selection of extreme answers and middle

Response styles	Indicator	Meaning	Range of the values
Acquiescence	ave	The average score of oppositely-worded items that measure one construct (smaller values indicate more acquies- cence)	1-5
Non- differentiation	sd	The standard deviation of all responses to one block of Items (smaller values indicate more non-differentiation)	0-4.41
	max	The maximum sequence of identical responses to one block of items (larger values indicate more non-differen- tiation)	1-7
Extreme answers	perc.ex	The proportion of extreme answers (larger values indicate more extreme answers)	0-0.86
Middle answers	perc.mi	The proportion of middle answers (larger values indicate more middle answers)	0-0.93
Item nonresponse rate	perc. unanswered	The amount of item nonresponse divided by the number of all applicable items (larger values indicate a higher item nonresponse rate)	0-0.11

Table 1. Indicators of response quality (modified from Jin and Loosveldt 2020).

answers, respectively measured by the proportion of extreme answers and middle answers, are also calculated based on these two blocks of items. The last indicator is item nonresponse rate, calculated as the ratio of item nonresponse ("don't know", refusal, and no answer) to the total number of the applicable items. For a more detailed introduction to the calculation of the indicators, we refer to our previous work (Jin and Loosveldt 2020).

The selected indicators are used only as examples to approach our research question. It is notable that although not the focus of the present study, justifying the choice of the indicators plays an important role in the subsequent analysis using multivariate control charts.

The article is structured as follows. The next section details the methods involved in addressing the research question. We then present the results of applying the procedure (Subsection 4.1), insights into the in-control data and the outliers (Subsection 4.2), and the validation results (Subsection 4.3). Lastly, the findings and limitations are discussed and suggestions for future work are offered.

3. Methodology

3.1. Hotelling T^2 Chart Based on Kernel Density Estimation

The most widely used tool for monitoring multivariate processes is a multivariate control chart based on the Hotelling T^2 statistic (Hotelling 1947). Suppose that a data set x is collected from an in-control process over a period of time, and contains n observations

with each observation being characterized by p quality indicators. The Hotelling T^2 statistic for each observation x_i is calculated with the formula:

$$T_i^2 = (x_i - \bar{x})S^{-1}(x_i - \bar{x})^T$$
(1)

where \bar{x} is the sample $(p \times 1)$ mean vector and S is the sample $(p \times p)$ covariance matrix of the in-control data set x. In this case, the Hotelling T^2 statistic measures the distance between the values of the variables for an observation and the values of the mean variables of the data set, which come from the process itself as "the voice of the process" (Jans et al. 2013; Scherkenbach 1986). An alternative is to use the target values of the mean vectors and the covariance matrix, referred to as "the voice of the customer", in order to check whether the process is operating within the target values specified by a "customer". In this article, we focus on the estimates based on our sample, assuming that no expert or customer opinions are available with regard to the "gold" values for the response quality indicators and their correlation structure.

The next step is to assess the control limit of the T^2 chart, for which the $100(1 - \alpha)$ percentile of the distribution of T^2 statistic is used. Here, α is the Type I error rate (or false alarm rate), indicating the proportion of in-control observations that are falsely identified as outliers. Since the distribution of the T^2 statistic is unknown, we use the KDE approach to estimate it and determine the control limit as an upper percentile of the derived kernel distribution function.

Given the obtained $T_1^2, T_2^2, \ldots, T_n^2$ for the *n* observations x_1, x_2, \ldots, x_n , the kernel density estimator of the distribution of T^2 statistics is specified as:

$$\widehat{f}_{h}(t) = \frac{1}{n} \sum_{i=1}^{n} K[\frac{(t-T_{i}^{2})}{\widehat{h}}]$$
(2)

where *K* is the kernel function and \hat{h} is the smoothing parameter (termed bandwidth). For *K*, we consider the most commonly used function: the standard normal density function. In addition, in line with Chou et al. (2001), we use Polansky and Baker's (2000) two-stage estimation procedure to compute \hat{h} . The kernel distribution function of T^2 is constructed as

$$\widehat{F_h}(t) = \int_0^t \widehat{f_h}(T^2) dT^2$$
(3)

The control limit of T^2 based on the KDE (*CL_k*), determined as the 100(1 - α) percentile of the distribution function $\widehat{F_h}(t)$, can be calculated as

$$CL_k = \widehat{F_h}(t)^{-1}(1-\alpha) \tag{4}$$

The choice of the bandwidth parameter is crucial, as it controls the degree of smoothing applied to the data (T^2) , and accordingly may influence the variability of the estimated control limit of the KDE-based T^2 chart. The verification of our selection of Polansky and Baker's procedure is discussed in the appendix, Section 6.

3.2. A Dynamic Procedure for Removing Outliers

It should be remembered that the control limit discussed above is estimated based on a set of historical in-control data. As already noted, the historical in-control data is usually not directly available in surveys, and the data therefore needs to be determined by detecting and removing possible outliers. The extreme outliers, however, may easily affect the T^2 statistic and the estimated control limit. To address this, an iterative procedure is developed in order to remove one outlier at a time, and then to re-estimate the T^2 statistic. and the control limit based on KDE without this outlier. This operation is continued until a data set is obtained that can be considered as in-control.

One problem associated with using a non-parametric technique such as KDE in the Phase I analysis is that each time an outlier is removed, there will always be around $n\alpha$ values of T^2 above the newly calculated control limit (Mason and Young 2002). The reason for this is that the KDE technique is solely based on the data, and always finds the $100(1 - \alpha)$ percentile of the fitted distribution of the T^2 statistic. Observations with a T^2 above this percentile will hence always (inevitably) be identified as outliers. Consequently, a "stop rule" is needed in order to terminate the re-estimation of the control limit after a certain number of iterations, when a suitable control limit can be determined. Mason and Young (2002) proposed considering the difference between two consecutive estimated control limits. If this difference is large, it implies that the fitted kernel density curves change significantly with the outliers dropped. A very small difference, on the other hand, implies that the desired control limit has already been obtained, which should be carried forward rather than being re-estimated.

In the present study, we consider the deterministic part of the variation in the estimates of CL_k , rather than the difference between (only) two consecutive estimates. To do so, we perform locally weighted regression, or Loess (Cleveland 1979; Cleveland and Devlin 1988), to smooth the estimated CL_k at different iterations. Loess has the advantage of being a nonparametric visual technique, which provides a graphical assessments of the relationship between variables without any prior specification of the form of the relationship. When the Loess curve has nearly horizontal tangents, it implies that the estimates of CL_k are stable in terms of an approximately zero rate of change, and the desired control limit is obtained.

Specifically, at each iteration, Loess builds up a polynomial regression to describe the variation in CL_k on the nearest neighbors using weighted least squares. The degree of polynomial regression (almost always first or second: that is, respectively locally linear or locally quadratic), the fraction of data points that are considered as the neighborhood (span), and the weighting function, are thus the choices that should be made to perform Loess. In this study, we use the default settings in the R software *loess* function, with a span of 0.75, quadratic local regression, and tri-cubic weighting. For a detailed explanation of the Loess algorithm, please refer to Cleveland and Devlin (1988).

Assuming the smoothed value for CL_k at iteration *i* is $\hat{g}(i)$, the complete procedure for identifying and removing outliers is then summarized as follows (with the relevant equation numbers shown):

- a. Construct the T^2 statistic based on the data set x using Equation (1).
- b. Fit the kernel density estimation using Equation (2).
- c. Calculate the control limit $(CL_k^{(1)})$ of T^2 using Equation (3) and (4). d. Remove the observation with the largest T^2 value.

- e. Reconstruct the T^2 statistic based on the remaining data and calculate a new control limit $(CL_k^{(2)})$.
- f. Repeat steps d and e until the Loess curve of CL_k exhibits a zero slope, approximately expressed as $|\hat{g}(i) \hat{g}(i-1)| < c$, where c is set as 0.001.
- g. $CL_k^{(i-1)}$ is used as the suitable control limit CL_k^* . The observations removed before this control limit is obtained are considered as outliers, and the remaining ones form the in-control data set.

The Loess curve is only fitted after $CL_k^{(4)}$ is obtained, because a quadratic local regression requires at least three points. Moreover, the use of *c* is only for the algorithm to start and converge. For the determination of the desired estimate of CL_k , the choice of *c* should be of no real consequence. The reason is that once the iterations start, the visual inspection of the fitted Loess curve will help a customer or user to determine whether and where (in terms of iterations) the curve has become close to being horizontal and stable.

We emphasize that at the end of the Phase I analysis, approximately $n \times \alpha$ observations are larger than the obtained control limit CL_k^* , but are tolerated. In Phase II, however, any new observation with a corresponding T^2 statistic larger than CL_k^* , will be considered as out of control. The probability of this new observation actually being in-control-the false alarm rate-is α .

4. Results

4.1. Identifying and Removing Outliers

Our data set includes 1,766 interviews described by six response quality indicators. The procedure developed in Subsection 3.2 is applied to the data set to extract the in-control data by identifying and removing outliers. Figure 1 presents the utilization of the T^2 chart for our data set. The data points (1,766) on the x-axis are all the interviews (respondents) ordered by the respondent ID number, and the y-axis shows the T^2 statistic. The control limit obtained from the KDE with a false alarm rate (α) of 0.01 is 29.9967, represented by the red line in Figure 1.

Some 17–as expected based on the calculation of $1,766 \times 0.01$ –interviews fall outside the control limit. As described in Subsection 3.2, the outliers are removed one at a time in order to diminish the impact the extreme outliers may have on the T^2 statistic and the estimated control limit, until the stop rule is satisfied. With regard to the stop rule, in the present study we specify *c* as 0.001 to start the iterations.

The estimated control limit based on all interviews is 29.9967, as shown in Figure 2, and is decreased as the most extreme outliers are removed, one at each iteration. The curve in the figure is the Loess curve fit to the data (obtained using the R function *loess*, with a span of 0.75, quadratic regression, and tri-cubic weighting function). The Loess curve flattens out after around 70 iterations, indicating a nearly stable estimate of the control limit. The desired estimate of the control limit ($\alpha = 0.01$) is considered to have been achieved at the 76th iteration ($Cl_k^{(76)}$ 23.1993), because the difference between the fitted smooth value of $\hat{g}(76)$ (23.1931) and $\hat{g}(77)$ (23.1927) is only -0.0004, which satisfies our stop rule.

At the 76th iteration (with 75 outliers having been removed), the T^2 chart for the remaining 1,691 interviews with $CL_k^{(76)}$ is presented in Figure 3. As expected, for



Fig. 1. T2 chart for all interviews (n = 1,766) with a KDE-based control limit ($\alpha = 0.01$).

n = 1,691 and $\alpha = 0.01$, some 17 observations are still outside the control limit. As explained in Subsection 3.2, these 17 outliers, however, are acceptable given $\alpha = 0.01$, and should be kept in the in-control data set. In total, around 4.25% (75/1,766) of the total interviews are considered as outliers in our data set. The 1,691 interviews form the in-control data set.



Fig. 2. The estimated KDE-based control limit at different iterations ($\alpha = 0.01$).



Fig. 3. T2 chart for in-control interviews (n = 1,691) with a KDE-based control limit ($\alpha = 0.01$).

4.2. The Obtained In-Control Data and the Identified Outliers

Based on the obtained in-control data, the means of the quality Indicators \bar{x} and the correlation matrix (and covariance matrix *S*) can be computed in order to understand the in-control process. As shown in Table 2, for a block of seven items (measured on an 11-point scale), in-control interviews on average have 2.246 identical answers in sequence, 0.315 extreme answers (0.045×7), and 1.253 (0.179×7) middle answers, with the standard deviation of the responses being 1.619. The average acquiescence score is 2.606 and the average item nonresponse rate is 5%. With regard to the correlation structure of the response quality indicators, Table 2 also shows the Pearson's correlation coefficients between them. It is clear that each indicator is significantly correlated with at least one of the other indicators at a 0.05 significance level.

More importantly, once the means and the covariance matrix of the response quality indicators are available, new interviews from similar processes can be monitored in real time. For instance, assume the quality indicators for a new interview is x_{new} , then the corresponding T^2 statistic will be $(x_{new} - \bar{x})S^{-1}(x_{new} - \bar{x})^T$. If this statistic is beyond 23.1993, the new interview is considered as out of control with a 0.01 false alarm rate. New interviews with relevant information regarding the quality indicators can accordingly be checked immediately, rather than when the survey fieldwork has been completed.

Moving from the profile of the in-control group to that of the outlier group, we then compare the two groups on each of the six quality indicators using box plots (Figure 4) and Welch's t-tests (Table 3). The goal is to obtain information about which indicators are more important with regard to characterizing the outliers.

Remarkable differences between the two groups can be seen in the variables *perc.ex* and *max*. Interviews that are identified as outliers, compared with the in-control interviews, tend to have a greater number of extreme answers (48.1% versus 4.5%) and longer

	Means		Correlation matrix				
		max	sd	ave	perc. ex	perc. mi	perc. unanswered
max	2.246	1.000					
sd	1.619	-0.326*	1.000				
ave	2.606	-0.016	-0.048	1.000			
perc.ex	0.045	0.023	0.588*	-0.041	1.000		
perc.mi	0.179	0.142*	-0.116*	-0.024	-0.100*	1.000	
perc. unanswered	0.050	-0.028	-0.327*	0.107*	-0.338*	-0.311*	1.000

Table 2. The means and Pearson's correlation coefficients of quality indicators for the in-control data (n = 1,691).

Note: * p < 0.05

sequences of identical answers (3.973 versus 2.246). This is in line with the expectation that satisficing is responsible for respondent behavior that results in inferior data quality.

It is, however, unexpected but interesting to note that the outlier group on average has a lower rate of item nonresponse compared with the in-control group (respectively 2.6% versus 5%), and differentiates more in the answers to the measured questions in terms of a larger standard deviation of the responses (respectively 1.913 versus 1.619). This reveals that the interviews are identified based on a complete picture of the data quality to be assessed, represented by all the indicators as a whole. Outliers are not always linked to poor performance on every individual indicator, each of which describes a specific aspect of the quality.



Fig. 4. Box plots of quality indicators for the outlier group and the in-control group.

	Mean		Welch's t-test statistic	<i>p</i> -value	
	In-control data (1,691)	Outliers (75)			
max	2.246	3.973	13.557	0.000	
sd	1.619	1.913	2.249	0.027	
ave	2.606	2.561	-0.547	0.586	
perc.ex	0.045	0.481	15.201	0.000	
perc.mi	0.179	0.176	-0.091	0.927	
perc.unanswered	0.050	0.026	- 12.210	0.000	

Table 3. Comparing the outliers with the in-control data.

In addition, the notable difference in the item nonresponse rate between the two groups is probably a consequence of using the mean of the data set to construct the Hotelling T^2 statistic. As described in Subsection 3.1, we use the center of the data set as the ideal, and measure whether an interview is too far away from it. The center of the data set, which is the "voice of the process", however, is not necessarily the same as one might expect. For example, a 5% item nonresponse rate may be higher than the goal a customer would set for the indicator. If a lower target value for this indicator was used, interviews with a higher item nonresponse rate would be "punished" by having a larger Hotelling T^2 statistic value. As an attempt to construct the statistic in a case in which the least amount of information is available, we relax the need for "gold" values for the mean vector and covariance matrix concerning the response quality indicators, by using "the voice of the process" rather than a pre-specified "voice of the customer". Moreover, from a psychological perspective, the lower item nonresponse rate in the outlier group can also be the consequence of satisficing behavior. The reason is that some respondents may use a mental shortcut to simply respond with answers that are not well considered to every question, rather than making sufficient cognitive effort to determine the answers.

4.3. Validation Tests

The above sections detail how Phase I monitoring was carried out, by applying the procedure introduced in Subsection 3.2 to the data set. As a result, we identified and labeled 75 out of a total of 1,766 interviews as outliers, and the others (1,691) as in control. We also illustrate how the obtained in-control data would enable future Phase II monitoring by providing the in-control parameters: the control limit, the mean vector, and the covariance matrix. The current section details how, in order to test the procedure and verify the obtained results, we conduct a ten-fold cross-validation procedure to examine whether the results are stable when different subsets are used for training. To validate our procedure from another perspective, we also applied the MCD algorithm to our data set and compared the results with those previously obtained using the KDE-based T^2 procedure.

4.3.1. The Ten-Fold Cross-Validation Test

The ten-fold cross-validation test is performed in four steps as below:

a. Shuffle the entire data set randomly and split it into ten equal sized (or nearly equal sized) folds.

Fold	Size	Control limit		Mean vector					Accuracy
			max	sd	ave	perc. ex	perc. mi	perc. unanswerd	
1	177	23.501	2.247	1.634	2.603	0.050	0.180	0.050	1.000
2	177	28.151	2.305	1.627	2.599	0.060	0.176	0.049	0.983
3	176	26.421	2.272	1.643	2.598	0.057	0.177	0.049	0.977
4	177	26.839	2.296	1.631	2.606	0.058	0.178	0.049	0.977
5	176	23.827	2.254	1.618	2.611	0.044	0.179	0.050	0.977
6	177	23.302	2.244	1.613	2.603	0.043	0.179	0.050	0.977
7	176	23.751	2.243	1.629	2.615	0.049	0.179	0.050	1.000
8	177	23.302	2.243	1.623	2.605	0.046	0.178	0.050	1.000
9	176	24.394	2.269	1.629	2.601	0.050	0.176	0.050	1.000
10	177	23.772	2.268	1.629	2.603	0.053	0.180	0.050	0.994
In	nitial est	imate	2.246	1.618	2.606	0.045	0.179	0.050	

Table 4. Summary of ten-fold cross-validation results.

- b. Repeat the procedure in Subsection 3.2 on nine folds to train the KDE-based control limit, the mean vector, and the covariance matrix of the obtained in-control data.
- c. Compute the Hotelling T^2 statistic values for the remaining tenth fold based on the parameter estimates, and predict as outliers the interviews with the corresponding Hotelling T^2 values larger than the control limit.
- d. Repeat step b and c until each fold has been taken as the test fold.

In the end, each interview is used as training data nine times and is tested once to predict the "label" (outlier or in-control).

The data set (1,766) is split into ten folds with each fold containing 176 or 177 interviews for test and 1,590 (1,766 - 176) or 1,589 (1,766 - 177) for training. For each test fold, Table 4 details the size of the fold, the estimated control limit, and the mean vector based on the remaining training folds. Also presented in Table 4 is the accuracy rate for each test fold, assuming that the initial labels obtained based on the whole data set are correct. The results show that although the estimated control limit tends to fluctuate (from 23.302 to 28.151, with 23.8 being the median), the estimated mean vector of the six indicators is fairly stable and close to the initial estimate based on the whole data set. The accuracy obtained for each test fold is high, ranging from 0.997 to 1.

Pulling the test folds together, Table 5 shows a confusion matrix, which cross-tabulates the predicted labels with the initial correct labels for the whole data set. For most of the interviews, the predicted labels are the same as the initial labels. The procedure, for example, correctly predicts 1,683 of the 1,691 interviews that are in control (99.53%), as well as 63 of the 75

-			
		Correct	labels
		in-control	outlier
Predicted labels	in-control outlier	1,683 8	12 63

Table 5. Confusion matrix for the ten-fold cross-validation test.

Note: Correct labels are the initial labels obtained based on the whole dat set.

interviews that are outliers (84%). The overall prediction accuracy is about 98.87% ((1,683 + 63)/1,766). To sum up, the ten-fold cross-validation results imply that our procedure is stable and validated, as the results concerning the identification of in-control interviews and outliers are almost the same as when the whole data set is used as training data.

4.3.2. The Minimum Covariance Determinant (MCD) Algorithm

We now move on to the minimum covariance determinant (MCD), which is widely recognized as a highly robust estimator of the multivariate mean and covariance matrix (Hubert et al. 2018). We make a direct comparison between the results obtained from the MCD and the KDE-based T^2 procedure: a good agreement between the results would be interpreted as validating the KDE-based T^2 method.

Table 6 presents the robust means obtained by using the MCD algorithm. Comparing these with what was obtained previously using the KDE-based T^2 procedure, no marked difference can be found, except for a slight difference in *perc.ex*. Based on the MCD estimates of the mean and the covariance matrix, we then compute the MCD-based T^2 statistic. Figure 5 shows the MCD-based T^2 value versus the corresponding KDE-based T^2 value for each of the 1,766 interviews. Interviews that were previously labeled as outliers using the KDE-based T^2 procedure are marked in red. Most of the interviews either cluster in the lower left-hand corner or scatter in the upper right-hand part of the figure, sugesting that the two methods will give the same assessment of most of the interviews with regard to whether they are in-control or outliers. This is further supported by a significant and (extremely) high positive correlation (r = 0.932) between the T^2 values from the two methods.

We should recall that the outliers identified by the KDE-based T^2 procedure account for 4.25% of the total interviews. If we, for the time being, flag the top 4.25 percentile interviews with the largest MCD-based T^2 value as outliers (separated by the horizontal line in Figure 5), we can gain more insight into the results of the two methods by comparing the "outliers" and "in-control" interviews from MCD with those from the KDE-based T^2 procedure.

The results from the two methods are more similar to each other in the identification of in-control interviews than in the identification of outliers. Around 77.33% (55/75) of the MCD outliers are also identified by the KDE-based T^2 procedure, whereas around 98.81% (1,671/1,691) of the MCD in-control interviews are also identified as in control by the KDE-based T^2 procedure. Overall, around 97.73% (1,726/1,766) of the interviews are assigned the same label using the two methods. Given the good agreement between the two methods, we conclude that the results of the KDE-based T^2 procedure are validated by the MCD algorithm.

	Indicator mean						
Procedure	max	sd	ave	prec.exe	perc.mi	perc.unanswered	
${\text{MCD}}$ KDE-based T^2	2.205 2.246	1.497 1.618	2.616 2.606	0.009 0.045	0.179 0.179	0.052 0.050	

Table 6. Robust means obtained using the MCD algorithm.



Fig. 5. T2 values computed based on MCD estimates of the mean and the covariance matrix.

Table 7. Confusion matrix for comparing the identification of outliers using MCD and KDEbased T2 procedure.

		KDE-based T^2		
		in-control	outlier	
MCD	in-control outlier	1,671 20	20 55	

Note: MCD outliers are the top 4.25 percentile interviews with the largest MCD-based T^2 value.

5. Conclusion and Discussion

This study started from the observation that although the multivariate Hotelling T^2 control chart (T^2 chart) has been widely used for monitoring industrial production processes, its application in survey fieldwork – essentially a production process in which data based on interviews is produced – has been scant to date. In this article, we have presented a procedure based on Kernel Density Estimation (KDE) and the Hotelling T^2 chart. The procedure can identify interviews that are potentially problematic interviews by simultaneously monitoring a set of response quality indicators, without relying on historical in-control data or referring to the assumption of a parametric distribution of the indicators. Although illustrated here by using the data collected from the eighth round of the European Social Survey (ESS) in Belgium using computer-assisted personal interviews, the procedure is not limited to any particular survey mode. It could also be applied to assess the response quality of interviews collected using other survey modes.

First, we constructed the Hotelling T^2 statistic based on six numerical indicators of response quality. These six indicators, representing various response styles (acquiescence, non-differentiation, the selection of middle answers, the selection of extreme answers, and item nonresponse), were used as examples of response quality indicators. Second, we used the KDE to fit the distribution of the T^2 statistic in order to relax the conventional, yet strict, normality assumption of the indicators. The control limit of the T^2 chart was subsequently determined based on the percentile of the derived kernel curve of the T^2 statistic. Lastly, we presented a procedure for extracting the in-control data by iteratively removing the outliers until the desired KDE-based control limit was determined as being achieved. To guide the decision, we produced a locally weighted regression (Loess) curve to describe the variations in the estimates of the control limit, and searched for the iteration around which the curve had a (nearly) horizontal tangent. Our analysis was treated as a Phase I analysis, so as to avoid the need for prior knowledge of the chart parameters. This was also the rationale behind the presented procedure.

By applying the procedure to our data set, around 4.25% of the total interviews are considered as outliers, resulting in an in-control data set comprising the remaining interviews. The outlier group pinpoints a few indicators that are seen to be more important for fieldwork monitoring. The outliers are characterized by having longer sequences of identical answers and a greater number of extreme answers. This finding is in line with the basic idea that satisficing behavior would lead to compromised data quality (Krosnick 1991). In addition, the outliers also have on average a lower item nonresponse rate and differentiate more in the answers to the measured questions. While contrary to expectation at first sight, the lower item nonresponse rate in the outlier group is well interpretable as some respondents may incline to respond to every question without making sufficient cognitive effort to carefully consider the answers. The greater number of extreme answers in the outlier group points to the advantages of a multivariate framework over a univariate one (that examines one indicator at a time). The outliers are identified based on the entirety of the indicators, and do not necessarily exhibit poor performance on every individual indicator. As documented by MacGregor and Kourti (1995), the quality of a product can only be represented by the simultaneous consideration of all the measured indicators. The individual indicator, however, does not adequately define the product quality by itself. In the meantime, it cannot be ruled out that the choice of the target values for the indicators when constructing the Hotelling T^2 statistic may have influenced the results regarding the identified outliers. In cases of no prior knowledge or expert opinions regarding the target values for the indicators, the mean of the data set is used as the reference to measure whether an interview is too far away from the ideal. The mean of the data set, taken from the process, could differ from the target value that would be set as the goal. An alternative would be to use "the voice of the customer" (Jans et al. 2013; Scherkenbach 1986), concerned about the response quality indicators, if available.

The obtained in-control data set, on the other hand, enables a basic understanding of the in-control process by providing the in-control parameters: the mean vector and covariance matrix for the response quality indicators, and the estimated control limit. For example, for a block of seven items, the in-control interviews on average are found to have 2.246 identical answers in sequence. More importantly, as noted in Montgomery (2009), the obtained in-control parameters lead to the implementation of control charts in Phase II: the

subsequent on-line, real-time monitoring phase based on new incoming data from the survey. Although often relying on the timely availability of the monitored data (Kreuter et al. 2010), the capability to monitor new interviews during rather than after fieldwork is crucial for survey practitioners to identify problems and give rapid feedback. The idea of using control charts to monitor key variables and guide interventions with a view to adapting fieldwork to the problems identified is essentially the same as that of responsive designs (Groves and Heeringa 2006) and adaptive designs (Wagner 2008; Schouten et al. 2017).

The minimum covariance determinant (MCD) is another Mahalanobisbased method that is widely recognized as a highly robust estimator of the multivariate mean and the covariance matrix. Applying MCD to our data set and comparing the results with those obtained using the proposed KDE-based T^2 procedure, no marked difference is found in either the estimated mean or the T^2 statistic (computed based on the mean and covariance matrix). Although the robust MCD estimator is very helpful in detecting multivariate outliers, the commonly-used cutoff value – based on the asymptotic chi-square distribution of the robust distances – has been found to tend to flag far too many outliers (Hubert and Debruyne 2010). The KDE-based T^2 procedure, by contrast, does not rely on any underlying distribution, but instead searches for the cutoff (control limit) through the iterative process of removing outliers one at a time. However, it is worthwhile to note that KDE involves determination of the bandwidth parameter regulating the scale of smoothing in advance, and the estimated control limits can be sensitive to that.

To sum up, this article underlines the suitability of using a T^2 chart based on nonparametric techniques such as KDE to assess interview response quality and identify outliers. First, survey researchers and practitioners could simultaneously monitor all quality indicators. This is preferable to monitoring one indicator at a time, which could lead to unsatisfactory results (Bersimis et al. 2005). Also, thanks to the KDE technique integrated in our procedure, the indicators do not need to follow multivariate normal distribution. Second, the in-control interviews extracted from the data set can be used to estimate unknown in-control parameters (i.e., the mean vector and covariance matrix of the in-control process) and obtain the desired control limit that can be used in Phase II monitoring.

Our work clearly has some limitations. First, the present study only uses data for Belgium from the European Social Survey (ESS), and this may have influenced the results regarding the estimated KDE-based control limit and the important response quality indicators. Although these results might not be generalizable to other data sets, the methods and procedure we present—the integration of a T^2 chart with KDE—is not limited to the Belgian data, and could be applied to other data sets. Second, we have only considered a limited number of indicators of response quality. Future studies would benefit from starting with a broader range of indicators of response quality. Third, the indicators used in the present study are all numerical indicators. Investigations into the use of control charts for a mix of numerical and categorical indicators (e.g., whether or not a respondent exhibits acquiescence bias) can be considered as the next step in the exploration of the use of control charts to assess survey data quality. Moreover, in addition to the ten-fold cross-validation test and the comparison with the results from the MCD, using simulations

(where the data distribution is contaminated with another distribution) is an alternative way to evaluate the performance of the KDE-based T^2 procedure.

Despite the limitations of the exclusive selection of a data set from one country and the use of a limited number of numerical indicators, we believe that the results presented in this article could be a starting point for using statistical process control tools to simultaneously monitor multiple response quality indicators and identify interviews with outlying performance on the indicators.

6. Appendix

6.1. Sensitivity of the KDE-Based Control Limits to the Choice of Smoothing Parameter

Figure 6 illustrates the control limits of KDE-based T^2 chart using different values of bandwidths. We consider a simplified scenario where simulated data are described by six variables following a skewed normal distribution (skewness parameter is set to five). We clearly observe on the simulation results that in the range from 0.01 to around 0.04, the bandwidth parameter introduces important variations on the estimated control limits. The two-stage bandwidth computed using Polansky and Baker's procedure, represented by the red circle in Figure 6, is in a range where the corresponding control limits are already stable.



Fig. 6. Control limits of KDE-based T2 chart with different bandwidths.

6.2. Items Used in the ESS8 Questionnaire for the Calculations of the Response Quality Indicators

Construct	Pair	Items	Positively or negatively keyed	Label
Beliefs about inequality	1	E1	+	Large differences in people's incomes are acceptable to properly reward differences in talent and effort.
		E2	_	For a society to be fair, differences in people's standard of living should be small.
	2	E1	+	Large differences in people's incomes are acceptable to properly reward differences in talent and effort.
		B33	_	The government should take measures to reduce differences in income levels.
Attitudes towards target	3	E17	+	Many people with very low income receive lower benefits than they are legally entitled to.
groups and recipients of benefits and services		E18	_	Many people manage to obtain benefits and services to which they are not entitled.

Table 8. Items used to measure acquiescence.

Table 9. Items used to measure nondifferentiation, selection of extreme answers and selection of middle answers.

Block	Variable	Item label	11 point scale $(0-10)$
	B6	How much you personally trust the [country] parlia- ment?	No trust at all – Complete trust
	B7	How much you personally trust the legal system?	No trust at all – Complete trust
	B8	How much you personally trust the police?	No trust at all – Complete trust
1	B9	How much you personally trust politicians?	No trust at all – Complete trust
	B10	How much you personally trust political parties?	No trust at all – Complete trust
	B11	How much you personally trust the European Parlia- ment?	No trust at all – Complete trust

Block	Variable	Item label	11 point scale (0-10)
	B12	How much you personally trust the United Nations?	No trust at all – Complete trust
	B26	In politics people sometimes talk of "left" and "right". Where would you place yourself?.	Left – Right
2	B27	All things considered, how satisfied are you with your life as a whole nowadays?	Extremely dissatisfied – Extremely satisfied
	B28	On the whole how satisfied are you with the present state of the economy in (country)?	Extremely dissatisfied – Extremely satisfied
	B29	About the (country) govern- ment, how satisfied are you with the way it is doing its job?	Extremely dissatisfied – Extremely satisfied
	B30	On the whole, how satisfied are you with the way decocracy works in (country)?	Extremely dissatisfied – Extremely satisfied
	B31	What do you think overall about the state of edu- cation in (country) nowa- days?	Extremely bad – Extremely good
_	B32	What do you think overall about the state of health services in (country) nowadays?	Extremely bad – Extremely good

Table 9. Continued

7. References

- Ahsan, M., M. Mashuri, H. Kuswanto, D.D. Prastyo, and H. Khusna. 2018. "Multivariate Control Chart Based on PCA Mix for Variable and Attribute Quality Characteristics." *Production & Manufacturing Research* 6(1): 364–384. DOI: https://doi.org/10.1080/ 21693277.2018.1517055.
- Bersimis, S., J. Panaretos, and S. Psarakis. 2005. "Multivariate Statistical Process Control Charts and the Problem of Interpretation: A Short Overview and Some Applications in Industry." In *Proceedings of the 7th Hellenic European Conference on Computer Mathematics and Its Applications*. Athens, Greece. September 22–24, 2005. Available at: https://arxiv.org/ftp/arxiv/papers/0901/0901.2880.pdf (accessed September 2020).
- Billiet, J.B., and M.J. McClendon. 2000. "Modeling Acquiescence in Measurement Models for Two Balanced Sets of Items." *Structural Equation Modeling* 7(4): 608–628. DOI: https://doi.org/10.1207/S15328007SEM0704_5.

- Chakraborti, S., S. Human, and M. Graham. 2008. "Phase I Statistical Process Control Charts: An Overview and Some Results." *Quality Engineering* 21(1): 52–62. DOI: https://doi.org/10.1080/08982110802445561.
- Chavent, M., V. Kuentz-Simonet, A. Labenne, and J. Saracco. 2014. "Multivariate Analysis of Mixed Data: The R Package PCAmixdata." Available at: https://arxiv.org/abs/1411.4911 (accessed September 2020).
- Chou, Y.-M., R.L. Mason, and J.C. Young. 2001. "The Control Chart for Individual Observations from a Multivariate Non-Normal Distribution." *Communications in Statistics–Theory and Methods* 30(8-9): 1937–1949. DOI: https://doi.org/10.1081/-STA-100105706.
- Cleveland, W.S. 1979. "Robust Locally Weighted Regression and Smoothing Scatterplots." *Journal of the American Statistical Association* 74(368): 829–836. DOI: http://doi.org/10.1080/01621459.1979.10481038.
- Cleveland, W.S., and S.J. Devlin. 1988. "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting." *Journal of the American Statistical Association* 83(403): 596–610. DOI: http://doi.org/10.1080/01621459.1988.10478639.
- Corsetti, G., M. Giammatteo, and A. Martini. 2010. "Monitoring Process and Nonsampling Errors Control in PLUS Sample Survey." In *Proceedings of the European Conference on Quality in Official Statistics (Q2010)*. Helsinki, Finland. May 4–6, 2010. Available at: https://q2010.stat.fi/media/presentations/session24/corsetti_giammatteo_martini_q2010_v2_paper.pd (accessed September 2020).
- Costa, F.S., R.H. Pedroza, D.L. Porto, M.V. Amorim, and K.M. Lima. 2015. "Multivariate Control Charts for Simultaneous Quality Monitoring of Isoniazid and Rifampicin in a Pharmaceutical Formulation Using a Portable Near Infrared Spectrometer." *Journal of the Brazilian Chemical Society* 26(1): 64–73. DOI: http://dx.doi.org/10.5935/0103-5053.20140214.
- Ferrer, A. 2007. "Multivariate Statistical Process Control Based on Principal Component Analysis (MSPC-PCA): Some Reflections and a Case Study in An Autobody Assembly Process." *Quality Engineering* 19(4): 311–325. DOI: https://doi.org/10.1080/ 08982110701621304.
- Groves, R.M., and S.G. Heeringa. 2006. "Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 169(3): 439–457. DOI: https://doi.org/10.1111/j.1467-985X.2006.00423.x.
- Hotelling, H. 1947. "Multivariate Quality Control–illustrated by the Air Testing of Sample Bombsights." In *Techniques of Statistical Analysis*, edited by C. Eisenhart, M. Hastay, and W. Wallis, 111–184. New York: McGraw-Hill.
- Hubert, M., and M. Debruyne. 2010. "Minimum Covariance Determinant." Wiley Interdisciplinary Reviews: Computational Statistics 2(1): 36–43. DOI: https://doi.org/ 10.1002/wics.61.
- Hubert, M., M. Debruyne, and P.J. Rousseeuw. 2018. "Minimum Covariance Determinant and Extensions." Wiley Interdisciplinary Reviews: Computational Statistics 10(3): E1421. DOI: https://doi.org/10.1002/wics.1421.
- Jans, M., R. Sirkis, and D. Morgan. 2013. "Managing Data Quality Indicators With Paradata Based Statistical Quality Control Tools: The Keys to Survey Performance." In

Improving Surveys With Paradata. Analytic Uses of Process Information, Edited by F. Kreuter, 191–229. John Wiley & Sons, Inc.

- Jin, J., and G. Loosveldt. 2020. "Assessing Response Quality by Using Multivariate Control Charts for Numerical and Categorical Response Quality Indicators." *Journal of Survey Statistics and Methodology*. ISSN: 2325-0984. DOI: https://doi.org/10.1093/ jssam/smaa012.
- Jin, J., C. Vandenplas, and G. Loosveldt. 2019. "The Evaluation of Statistical Process Control Methods to Monitor Interview Duration During Survey Data Collection." Sage Open 9(2). DOI: https://doi.org/10.1177/2158244019854652.
- Kini, K.R., and M. Madakyaru. 2016. "Multivariate Statistical Based Process Monitoring Using Principal Component Analysis: An Application to Chemical Reactor." *International Journal of Control Theory and Applications* 9(39): 303–311.
- Kreuter, F., M. Couper, and L. Lyberg. 2010. "The Use of Paradata to Monitor and Manage Survey Data Collection." In *Proceedings of the Joint Statistical Meetings*, 282–296. Vanouver, Canada. August 2–4, 2010. Available at: http://sampieuchair.ec. unipi.it/wp-content/uploads/2018/10/Couper-etal.pdf (accessed September 2020).
- Krosnick, J.A. 1991. "Response Strategies for Coping With the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5(3): 213–236. DOI: https://doi.org/10.1002/acp. 2350050305.
- Loosveldt, G., and K. Beullens. 2017. "Interviewer Effects on Non-differentiation and Straightlining in the European Social Survey." *Journal of Official Statistics* 33(2): 409–426. DOI: https://doi.org/10.1515/jos-2017-0020.
- MacCarthy, B., and T. Wasusri. 2002. "A Review of Non-standard Applications of Statistical Process Control (SPC) Charts." *International Journal of Quality & Reliability Management* 19(3): 295–320. DOI: http://doi.org/10.1108/ 02656710210415695.
- MacGregor, J.F., and T. Kourti. 1995. "Statistical Process Control of Multivariate Processes." *Control Engineering Practice* 3(3): 403–414. DOI: https://doi.org/10.1016/ 0967-0661(95)00014-L.
- Mason, R., and J. Young. 2002. Multivariate Statistical Process Control With Industrial Applications. ASA-SIAM Series on Statistics and Applied Probability. Society for Industrial / Applied Mathematics. DOI: https://doi.org/10.1137/1.9780898718461.
- Montgomery, D.C. 2009. *Introduction to Statistical Quality Control*. New York: John Wiley & Sons.
- Peng, D., and K. Feld. 2011. "Quality Control in Telephone Survey Interviewer Monitoring." Survey Practice 4(2). DOI: https://doi.org/10.29115/SP-2011-0011.
- Phaladiganon, P., S.B. Kim, V.C. Chen, J.-G. Baek, and S.-K. Park. 2011. "Bootstrapbased T2 Multivariate Control Charts." *Communications in Statistics–Simulation and Computation* 40(5): 645–662. DOI: https://doi.org/10.1080/03610918.2010.549989.
- Polansky, A.M., and E.R. Baker. 2000. "Multistage Plug-in Bandwidth Selection for Kernel Distribution Function Estimates." *Journal of Statistical Computation and Simulation* 65(1-4): 63–80. DOI: https://doi.org/10.1080/00949650008811990.
- Rammstedt, B., D. Danner, and M. Bosnjak. 2017. "Acquiescence Response Styles: a Multilevel Model Explaining Individual-level and Country-level differences."

Personality and Individual Differences 107:190–194. DOI: https://doi.org/10.1016/j.-paid.2016.11.038.

- Rocke, D.M., and D.L. Woodruff. 1996. "Identification of Outliers in Multivariate Data." *Journal of the American Statistical Association* 91(435): 1047–1061. DOI: https://doi.org/10.2307/2291724.
- Rousseeuw, P.J. 1984. "Least Median of Squares Regression." *Journal of the American Statistical Association* 79(388): 871–880. DOI: https://doi.org/10.1080/01621459. 1984. 10477105.
- Rousseeuw, P.J. 1985. "Multivariate Estimation With High Breakdown Point." In Mathematical Statistics and Applications: Proceedings of the Fourth Pannonian Symposium on Mathematical Statistics and Probability, 283–297. Bad Tatzmannsdorf, Austria. September 4–10, 1985. Available at: https://wis.kuleuven.be/stat/robust/papers/publications-1985/rousseeuw-multiv Ariateestimationhighbreakdown-1985.pdf (accessed September 2020).
- Rousseeuw, P.J., and K.V. Driessen. 1999. "A Fast Algorithm for the Minimum Covariance Determinant Estimator." *Technometrics* 41(3): 212–223. DOI: https://doi.org/10.1080/00401706.1999.10485670.
- Scherkenbach, W.W. 1986. *The Deming Route to Quality and Productivity: Road Maps and Roadblocks*. Washington, DC: CEE Press.
- Schouten, B., A. Peytchev, and J. Wagner. 2017. *Adaptive Survey Design*. Boca Raton, FL: Chapman & Hall/CRC.
- Shewhart, W.A. 1931. *Economic Control of Quality of Manufactured Product*. London: Macmillan.
- Simon, H.A. 1956. "Rational Choice and the Structure of the Environment." *Psychological Review* 63(2): 129–138. DOI: https://doi.org/10.1037/h0042769.
- Sirkis, R., M. Jans, J. Dahlhamer, R. Gindi, and B. Duffey. 2011. "Using Statistical Process Control to Understand Variation in Computer-assisted Personal Interviewing Data." In *Proceedings of the Survey Research Methods Section*, 477–489. Miami Beach, FL. July 30–August 4, 2011. Available at: http://www.asasrms.org/ Proceedings/y2011/Files/300484_65051.pdf (accessed September 2020).
- Wagner, J.R. 2008. "Adaptive Survey Design to Reduce Nonresponse Bias." PhD Thesis, University of Michigan. Available at: https://deepblue.lib.umich.edu/bitstream/handle/2027.42/60831/jameswag_1.pdf?sequence=1&isAllowed=y (accessed January 2021).
- Yan, T., R. Tourangeau, and Z. Arens. 2004. "When Less is More: Are Reluctant Respondents Poor Reporters?" In *Proceedings of the Survey Research Methods Section*. Toronto, Canada. August 8-12, 2004. Available at: http://www.asasrms.org/ Proceedings/y2004/files/Jsm2004-000169.pdf. (accessed September 2020).

Received June 2019 Revised August 2020 Accepted September 2020