JISE

# A robust wavelet based profile monitoring and change point detection using S-estimator and clustering

## Orod Ahmadi[1*], Hamid Shahriari[1] and Yaser Samimi[1]

[1]*Industrial Engineering Department, K.N. Toosi University of Technology, Tehran, Iran*

*oroarm@yahoo.com, hshahriari@kntu.ac.ir, y_samimi@kntu.ac.ir*

## Abstract

Some quality characteristics are well defined when treated as response variables and are related to some independent variables. This relationship is called a profile. Parametric models, such as linear models, may be used to model profiles. However, in practical applications due to the complexity of many processes it is not usually possible to model a process using parametric models. In these cases, non-parametric methods are used to model the processes. One important and applicable non- parametric method used to model complicated profiles is wavelet transformation. Use of wavelet transformation requires estimation of the in control profile in phase I. Classical estimators are usually used in phase I to estimate the in control profile, using wavelet transformation. However, the presence of outliers in data in phase I may affects classical estimators. In this research a robust estimator of the in control profile based on clustering is proposed which is insensitive to the presence of outliers. As well as estimating the in control profile in phase I it is of interest to determine the change point of the process in phase II. In this work a procedure for estimating the change point of complicated profiles in phase II is also introduced. This suggested method does not require normality assumption of the error terms. Aggregation of the proposed robust estimator with the change point detection method results a procedure for detecting the change point. Simulation studies show that the proposed method is robust in presence of outliers compare to the classical methods of profile monitoring and change point detection.

**Keywords:** Change point detection, clustering, profile monitoring, robust estimation, Wavelet transformation.

## 1-Introduction

In some statistical process monitoring applications, it is assumed that the quality characteristic of interest is well modelled using univariate or multivariate distribution. According to Stover and Brill (1998), Kang and Albin (2000) and Mestek et al. (1994) in some cases it is better to relate the quality characteristic of interest, say $Y$, to an independent variable $x$. This relationship may be represented as:

$$Y = f(x) + \varepsilon \tag{1}$$

Where $f(x)$ is the mean response function and $\varepsilon$ is the random error term with mean zero and variance $\sigma^2$. This is also called a profile. Monitoring a profile is usually performed in two phases. In phase I, $m$ random samples of size $n$ are taken from the process. Each sample includes paired observations on response and independent variable. Let $(x_1, Y_{1j}), (x_2, Y_{2j}), ..., (x_n, Y_{nj})$ be the $j^{th}$ random sample $j = 1, 2, ..., m$ in which

$$Y_{ij} = f(x_i) + \varepsilon_{ij}; \; i = 1, 2, ..., n; \; j = 1, 2, ..., m \tag{2}$$

where $\varepsilon_{ij}$s are random error terms.

Usually in practice the outlying samples are eliminated and the profile is estimated in phase I. This estimated profile is called here a reference profile. In phase II samples of size $n$ are taken sequentially from the process and a profile is estimated. This estimated profile is then compared with the reference profile. When these two estimated profiles are significantly different the process is considered to be out of control. Noorossana et al. (2016) studied the effect of estimation error in phase I, on the monitoring of simple linear profiles in phase II. Woodall and Montgomery (2014) provided some recent research of profile monitoring. Qi et al. (2016) provided a unified framework to monitor generalized linear profiles in phase II using weighted likelihood ration test Grasso et al. (2016) presented a profile monitoring approach based on curve registration information. Khedmati and Akhavan Niaki (2016) proposed a control chart for monitoring simple linear profiles in multistage processes. Woodall (2017) summarized some works on profile monitoring and the effect of estimation error on phase II monitoring.

In profile monitoring three problems including the functional form of $f(x)$, the distribution of the error terms and the method used to monitor the process in phase II must be investigated. The functional form of the profile is usually known to be linear with normal error terms and with unknown parameters. Control charts to identify the reference profile and to monitor the linear profile in phase II are introduced by Kim et al. (2003). A change point model to monitor the linear profile in phase II is proposed by Mahmoud et al. (2007). A control chart based on the sequential change point model is proposed to monitor the linear profile by Zhou et al. (2006). In monitoring a linear profile, it is important to test the stability of the intercept and the slope of the model. Regarding the second problem, the distribution of the error terms is usually assumed to be normal with mean zero and variance $\sigma^2$. Based on normality assumption of the error terms, different methods of analysing and monitoring the profiles are proposed. Different methods of monitoring are usually used in phase II. Profile monitoring methods in phase II are usually some generalizations of the Shewhart type control charts. In these methods, a test statistic is defined and its value for each sample taken in phase II is computed. This test statistic reflects the deviation of each profile observed in phase II from the reference profile defined in phase I. According to Montgomery (2012), Shewhart type control charts use only the current information to determine whether the process is in control or not. As a result, these control charts are insensitive to detect small to moderate changes in model parameters. Other control charts such as CUSUM and EWMA, which use the information from the past samples, detect the small changes more rapidly. In process monitoring it is more important to determine the time when a process becomes out of control. Thus, instead of using the traditional Shewhart type control charts, some authors suggested other methods for determining when the process becomes out of control in phase II. These are generally considered as the change point methods. A change point model for monitoring the linear profiles is proposed by Zhou et al. (2006) and Mahmoud et al. (2003). The change point model for monitoring the nonlinear profiles is used by Chicken et al. (2009). In the context of profile monitoring, the change point model is stated as the fallowing:

$$Y_{ij} = \begin{cases} f_0(x_i) + \varepsilon_{ij} & j \le \tau \; i = 1, 2, ..., n \\ f_1(x_i) + \varepsilon_{ij} & j > \tau \; i = 1, 2, ..., n \end{cases} \tag{3}$$

where $Y_{ij}$ is the response variable for the $i^{th}$ observation from the $j^{th}$ sample, $\tau$ is the unknown

change point and $\varepsilon_{ij}$ s are usually assumed to be i.i.d. normal random variables with mean zero and variance $\sigma^2$. In equation 3, the $f_0$ (.) is the in control profile which is usually being estimated in phase I, while the $f_1$ (.) is the out of control profile. Paynabar et al. (2016) introduced a change point approach to monitor multivariate profiles in phase I. They proposed a new modelling, monitoring and diagnosis framework for phase-I analysis of multichannel profiles. Shadman et al. (2015) introduced a change point model to monitor Generalized Linear Model (GLM) profiles in phase I. Shadman et al. (2017) proposed a change point approach to monitor GLM profiles in phase II based on Rao score test.

In some cases, such as signal processing, the nature of the process is so complicated and fitting a parametric model to the data is inappropriate. The Nortel's antenna manufacturing system is studied by Jeong and Lu (2006) as a process with complicated profile. In cases of complicated profiles, it is better to use some nonparametric methods to estimate and to analyse the profiles. Two frequently used nonparametric methods for analysing and estimating complicated profiles are smoothing spline and wavelet transformation. Chen et al. (2015) introduced a nonparametric method of analysing profiles using clustering. In this method p-spline regression is used to model nonparametric profiles. At the end of the clustering step, profiles are classified as in control profiles and out of control profiles. For more details of using spline method for monitoring profiles, one may refer to Gardner et al. (1997). It is noted by Chicken et al. (2009) that using spline method for monitoring an unsmooth profile is not appropriate. The spline method cannot reveal the non-smoothing feature of a profile. So, when profiles are complicated and unsmooth it is better to use other nonparametric methods. As stated by Ogden (1997) wavelet transformation is a nonparametric method which is used for analysing and monitoring complicated and unsmooth profiles. In this study wavelet transformation is used to decompose complicated and unsmooth functions into simpler functions which are called wavelets. So, the coefficients of these simpler functions are monitored instead of the parameters of the complicated function.

Regarding the assumptions of the distribution of the error terms, in some cases the distribution of the error terms may not be normal. In addition to the non -normality there may exist some outliers which are far from the main bulk of the observations. Existence of these atypical error terms leads to some unexpected values for the response variable. Non-normality of the error terms as well as the presence of outliers result the inefficiency of normal based methods. So, it is very important to use some other methods of estimation, insensitive to deviation from normality assumption and the presence of outliers. These estimation methods are called robust estimation in literature. Some authors such as Jeong and Lu (2006) assumed known in control profile and normality of the error terms and only monitored the profile in phase II. While others such as Chicken (2009) applied the classical methods of estimation in phase I. Kamranrad and Amiri (2016) proposed a combined control chart for monitoring simple linear profiles in the presence of outliers and autocorrelation within observation. They used robust Holt-Winter model to design this control charts. Hakimi et al. (2017) proposed robust approaches for estimating the parameters of logistic regression profiles. The main purpose of their study is to develop approaches for monitoring logistic regression profiles which is insensitive to the presence of outliers. They used weighted maximum likelihood estimation, redescending M-estimator and a combined approach to reduce the effect of outliers.

In this research a robust method for estimating the in control profile in phase I is proposed based on the wavelet transformation. This estimation method is based on the S-estimator as a robust estimator of the mean vector of a multivariate distribution. The S-estimator is presented briefly in section 3. The proposed method is used when the data are non-normal, include outliers and the profile itself is complicated.

In summary, in this research a robust method of estimating complicated profiles in phase I based on wavelet transformation and a non-parametric procedure for change point detection are proposed. These methods are used when the error terms are not necessarily normal. They are robust to the presence of outliers. This paper is organized as follow: In section 2 some preliminaries about wavelet transformation are provided. The proposed method of estimating the profile in phase I based on the S-

estimator is developed in section 3. A non-parametric procedure for finding the change is provided in section 4. Performance evaluation of the proposed method for monitoring a complicated profile is given in section 4. In section 5, the results are discussed and the conclusions are made.

## 2-Wavelet transformation

Wavelets are simple functions with compact support which are used to generate orthonormal basis. This orthonormal basis may be used to express a function in $L_R^2$, the space of square integrable functions. A function $f(x)$ is said to be square integrable over interval $[a,b]$, if

$$\|f\|^2 = \int_a^b f^2(x)dx < \infty \tag{4}$$

The wavelets and their applications in statistics are introduced in Ogden (1997). Different families of wavelets may be chosen to express a function. One of the frequently used is the Haar family which is also used in this research. Let $\phi(x)$ and $\psi(x)$ be the father and the mother wavelets, then $\phi_{jk}(x) = 2^{j/2}\phi(2^j x - k)$ and $\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)$ are the translation and dilation versions of the $\phi(x)$ and $\psi(x)$, respectively for integers $j$ and $k$. Then for any $j_0 \geq 0$ the set of functions $\left\{ \phi_{j_0,k}(x), \; \psi_{j_0,k}(x); j \geq j_0, k \in Z \right\}$ form an orthonormal basis to express any function in $L_R^2$. In this work, the smallest value of $j_0$ is chosen for forming an orthonormal basis. The function $f$ is usually unknown and must be estimated using sample data. Suppose $n = 2^J$ observations $(x_i, Y_i)$; $i = 1, 2, ..., n$ are given for some positive integer $J$. In this study it is assumed that the $x_i$ values are equally spaced over the interval $[0,1]$. If $f \in L_R^2$ then its finite approximation using wavelets is as:

$$f(x) \approx \sum_{k=0}^{2^{j_0}-1} c_{j_0,k}\phi_{j_0,k}(x) + \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} d_{j_0,k}\psi_{j_0,k}(x) \tag{5}$$

In which $c_{j,k} = \left\langle f(x), \phi_{j,k}(x) \right\rangle$ and $d_{j,k} = \left\langle f(x), \psi_{j,k}(x) \right\rangle$ are the inner product of the function $f(x)$ with $\phi_{j,k}(x)$ and $\psi_{j,k}(x)$, respectively. The coefficients $c_{j,k}$ and $d_{j,k}$ are called the wavelet coefficients. The function $f(x)$ could be represented in terms of wavelet coefficients. Let $\mathbf{f} = (f(x_1), f(x_2), ..., f(x_n))'$ be an $n \times 1$ vector. By using the pyramid algorithm in Ogden (1997) and equation 5, the wavelet coefficients are then defined as follows:

$$\boldsymbol{\theta} = \mathbf{Wf} \tag{6}$$

where, $\mathbf{W}$ is an $n \times n$ orthogonal matrix determined based on the father and the mother wavelets and $\boldsymbol{\theta} = (c_{0,0}, d_{0,0}, ..., d_{J-1,2^{J-1}-1})'$ is an $n \times 1$ vector of the wavelet coefficients. equation 6 is called Discrete Wavelet Transformation, DWT.

According to Chicken et al. (2009) and for the sake of simplicity, the components of vector $\boldsymbol{\theta}$ renamed as:

$$\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_n)' \tag{7}$$

In practice the components of vector $\mathbf{f}$ are unknown and only the response values $Y_i$ s are observed. Let $\mathbf{y} = (Y_1, Y_2, ..., Y_n)'$ be the vector of response values and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)'$ be the vector of error terms. The matrix form of equation 1 may be then represented as:

$$\mathbf{y} = \mathbf{f} + \boldsymbol{\varepsilon} \tag{8}$$

Pre-multiplying the two sides of equation 8 by the previously defined matrix $\mathbf{W}$ results the

following:
$$\mathbf{Wy} = \mathbf{Wf} + \mathbf{W\varepsilon} \tag{9}$$

Let $\hat{\boldsymbol{\theta}} = \mathbf{Wy}$ and $\boldsymbol{\upsilon} = \mathbf{W\varepsilon}$, then:

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\theta} + \boldsymbol{\upsilon} \tag{10}$$

where $\hat{\boldsymbol{\theta}}$ is the vector of estimated wavelet coefficients and $\boldsymbol{\upsilon}$ is the vector of the transformed error terms. So, by using wavelet transformation one may transform the data to obtain the estimated wavelet coefficients according to equation 10. In this approach, instead of estimating the parameters of the mean response function, the vector of wavelet coefficients is estimated. Let $f_0(x)$ be the in control profile and $\boldsymbol{\theta}_0$ be its wavelet coefficients vector. $\boldsymbol{\theta}_0$ is estimated based on $m$ random samples of size $n$ in phase I. Let also $\hat{\boldsymbol{\theta}}_j$; $j = 1, 2, ..., m$ denotes the vector of estimated wavelet coefficients for the $j^{th}$ sample taken in phase I. The classical estimator of $\boldsymbol{\theta}_0$ , widely used in phase I, is defined as:

$$\bar{\hat{\boldsymbol{\theta}}} = \frac{\sum_{j=1}^{m} \hat{\boldsymbol{\theta}}_j}{m} \tag{11}$$

When there are no outliers in the $m$ random samples, the estimator in equation 11 is an unbiased and consistent estimator for $\boldsymbol{\theta}_0$. However, in practical situations some outliers may exist in the dataset. The presence of outliers in phase I affects the performance of the classical estimators such as sample mean vector. It is stated by Rousseeuw and Leroy (1987) that outliers have much more influence in multivariate studies than in univariate. The estimation in phase I using wavelet coefficients, is of multivariate nature. So, it is better to use robust multivariate estimators in phase I. In section 3 a robust multivariate estimator for estimating $\boldsymbol{\theta}_0$ is proposed.

## 3-Proposed robust estimator of $\boldsymbol{\theta}_0$ in phase I

Classical methods of parameter estimation are usually based on the assumption that the data do not include outliers. However, when the data are taken from a process, the existence of outliers is inevitable. These outliers influence the classical estimators. In situations where the data are suspected of being contaminated it is better to use robust estimators. For each random sample the vector of estimated wavelet coefficients is computed and shown by $\hat{\boldsymbol{\theta}}_j$; $j = 1, 2, ..., m$. After computing these vectors, some measure of central tendency of them must be computed. So, it is important to choose some estimators in this phase as a measure of central tendency of $\hat{\boldsymbol{\theta}}_j$; $j = 1, 2, ..., m$. As stated before, sample mean vector given in Equation 11 is usually used in this step to estimate $\boldsymbol{\theta}_0$. Considering the dynamic nature of the process and the possibility of generating outliers, the sample mean vector does not estimate $\boldsymbol{\theta}_0$ precisely. As an example consider the following simulation study which shows the effects of outliers on the estimator.

**Example 1:**

In this example $m = 2000$ random samples of size $n = 2$ are generated from the general model introduced in equation 2, in which $f(x) = \sqrt{x(1-x)} \sin\left(\frac{2\pi \times 1.05}{x + 0.05}\right)$; $0 \leq x \leq 1$. This $f(x)$ is with varying frequency called the Doppler function, shown in figure 1. So, the model may be represented as:

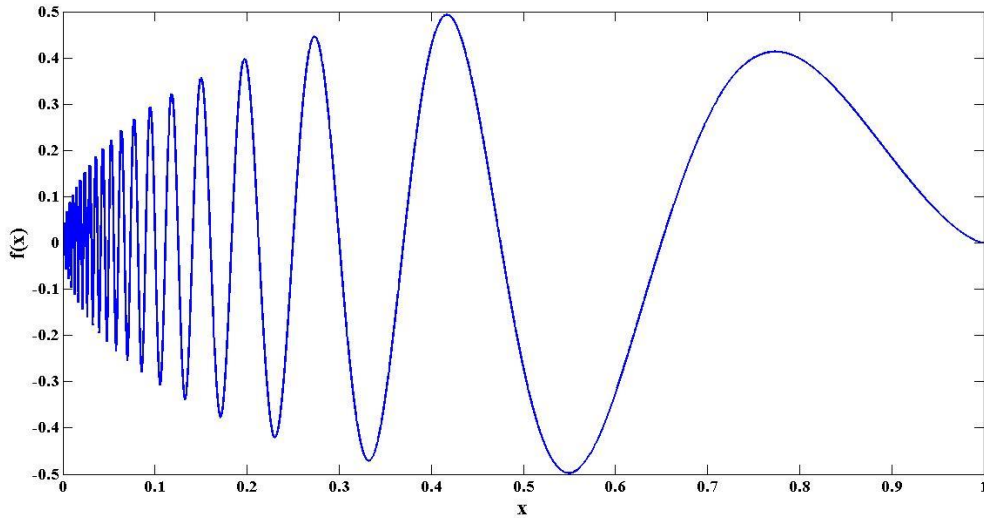$$Y_{ij} = Doppler(x) + \varepsilon_{ij}; \ i = 1, 2; \ j = 1, 2, \ldots, 2000 \tag{12}$$



**Fig 1.** Doppler function

In the literature of profile monitoring it is usually assumed that the error terms are normally distributed. However, possibility of presence of outliers requires that the error terms distribution be considered mixed normal. More formally $\varepsilon_{ij}$ s are distributed according to:

$$F_\varepsilon(x) = 0.9\Phi(x) + 0.1\Phi(x - 10) \tag{13}$$

where $\Phi(x)$ is the standard normal CDF. Equation 13 may be summarized as:

$$D(\varepsilon_{ij}) = 0.9N(0,1) + 0.1N(10,1); \ i = 1, 2; \ j = 1, 2, \ldots, 2000 \tag{14}$$

Where $D(\varepsilon)$ is distribution of $\varepsilon$. Considering Equation 13, $10\%$ of the error terms are outliers. The scatter plot of the error terms for the $2000$ generated samples of size 2 are displayed in figure 2.
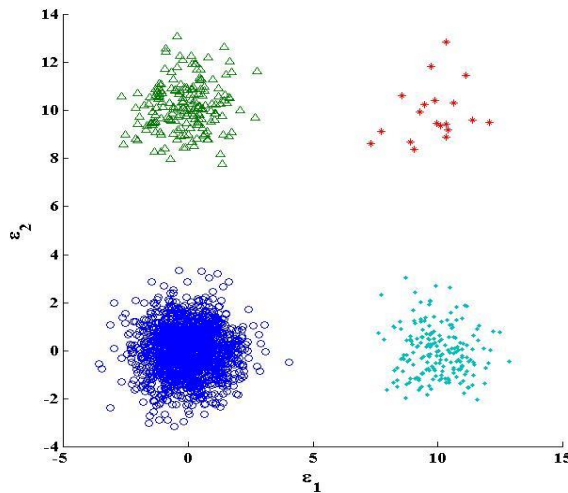


**Fig 2.** Scatter plot of the error terms in the presence of outliers

The presence of outliers results four clusters of error terms. One of these clusters corresponds to the

172

cases where none of the error terms are outliers. This cluster is formed around the origin and is shown by "O" marker. Other three clusters correspond to the cases where one or both of the error terms are outliers. According to equation 9, the responses are transformed to obtain the estimated wavelet coefficients. The scatter plot of the estimated wavelet coefficients for $2000$ samples is shown in figure 3.
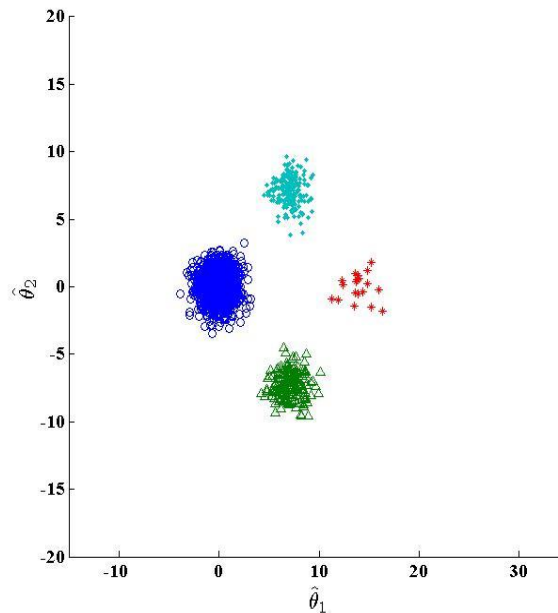


**Fig 3.** Scatter plot of estimated wavelet coefficients in the presence of outliers

Figure 3 reveals that the outliers constitute four clusters for the estimated wavelet coefficients, $\hat{\theta}_j; j = 1, 2, ..., 2000$. One of these clusters corresponds to the samples with no outliers. While, the others correspond to the contaminated samples. The clustering of estimated wavelet coefficients is the consequence of the clustering of the error terms due to the presence of outliers. So, the sample mean vector of the estimated wavelet coefficients, $\overline{\hat{\theta}}$ is inappropriate to be used for estimating the in control wavelet coefficients, $\theta_0$. ∎

The vectors $\hat{\theta}_j; j = 1, 2, ..., m$ are n-variate random vectors with mean $\theta_0$ when the data is not contaminated. When the data includes outliers, $\overline{\hat{\theta}}$ may no longer be used to estimate $\theta_0$. So, robust estimator for $\theta_0$ must be defined. Different multivariate robust estimators are introduced in the literature. The robust estimators of the mean vector of the multivariate distribution are classified into three categories by Rock and Woodruff (1993). The first class of estimators is based on the projection pursuit concept stated by Donoho (1982), Huber (1985) and Stahel (1981). The second class is the combinatorial estimators such as Minimum Volume Ellipsoid (MVE) proposed by Rousseeuw (1985). The third class is the iterative estimators such as M-estimators and S-estimators introduced in Maronna et al. (2006). All of these estimators require the use of some robust starting point. The combinatorial estimators such as MVE are usually used as a robust starting point. These iterative estimators are called compound estimators which are considered for further investigation in this work. Some criteria are usually used to measure the robustness of the estimators. One of these criteria,

173

widely used, is the breakdown point. Roughly speaking the breakdown point is the maximum proportion of contamination an estimator tolerates. It is obvious that the classical estimators such as sample mean vector do not tolerate any proportion of outliers. So, the breakdown point of estimator given in Equation 11 is zero and is not suitable to be used when the data is contaminated. According to Maronna et al. (2006), the breakdown point of the M-estimators of mean vector is equal to $1/(p+1)$ where $p$ is the number of variates. So, if the M-estimators are used in phase I to estimate $\theta_0$, the dimension of $\theta_0$ would be $n$ and the resulting breakdown point is equal to $1/(n+1)$. In the context of profile monitoring $n$ is usually large and the breakdown point of the M-estimators of $\theta_0$ tends to zero, $1/(n+1) \rightarrow 0$. So, the M-estimators are not suitable for estimating $\theta_0$ in phase I. According to Rousseeuw and Yohai (1984) the S-estimators have desirable properties including the breakdown point of $0.5$, when tuned properly. So, it seems suitable to use S-estimator in phase I to estimate $\theta_0$. However, computing the S-estimators require a robust initial estimate for $\theta_0$. As mentioned before the MVE estimator is usually used as a starting point for computing the S-estimators. Computing the MVE estimator requires taking of numerous sub samples of size $n+1$ from the $\hat{\theta}_j; j = 1, 2, ..., m$. It is stated by Rousseeuw and Leroy (1987) that the required number of sub samples for computing MVE estimator increases exponentially as a function of $n$. The high dimensional nature of $\hat{\theta}_j; j = 1, 2, ..., m$ requires taking of many sub samples from $\hat{\theta}_j$s and computation of their corresponding ellipsoid containing half of the $\hat{\theta}_j$s. So, computation of this estimator is highly time consuming. In addition to the computation time, the MVE estimator may not introduce robust starting point for high dimensional data. In this research the clustering property of the estimated wavelet coefficients is used to obtain the initial estimate of $\theta_0$. By applying this initial estimate, an iterative algorithm is used to compute the S-estimate of $\theta_0$. As presented in Example 1, the presence of outliers creates clusters of estimated wavelet coefficients. Thus, it is rational to determine these clusters in advance. Then, the cluster corresponds to the typical samples must be identified. In $n$-dimensional space one can determine the ellipsoid containing half of $\hat{\theta}_j; j = 1, 2, ..., m$. According to Maronna et al. (2006) the volume of this ellipsoid is proportional to the determinant of sample covariance matrix of $\hat{\theta}_j$s for each cluster. The ellipsoids correspond to the clusters of atypical samples are far from the ellipsoid representing the bulk of the $\hat{\theta}_j$s. The cluster related to ellipsoid with the smallest volume is considered as the cluster of the typical samples. The sample mean vector and the sample covariance matrix of $\hat{\theta}_j$s of this cluster are considered as the initial robust estimators of the mean vector and the covariance matrix of $\hat{\theta}_j$s, respectively. These initial estimates are used to compute S-estimate of $\hat{\theta}_j$. Representations of these clusters are provided in example 2.

**Example 2:**

It is seen in example 1 that the outliers create four clusters of estimated wavelet coefficients shown in figure 3. For each cluster the ellipsoid containing half of the points is introduced. These ellipsoids as well as the four clusters are shown in figure 4.
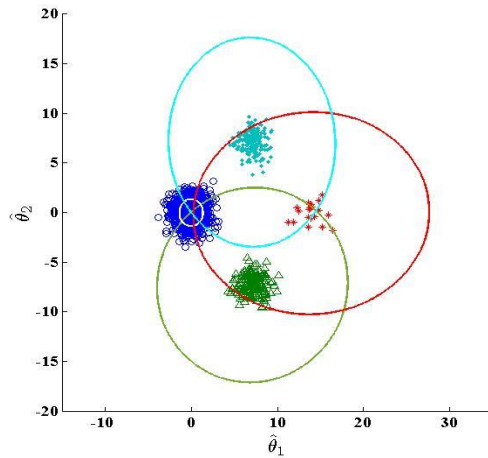
**Fig 4.** Scatter plot of estimated wavelet coefficients in the presence of outliers and the corresponding ellipsoids

The smallest ellipsoid in figure 4 shown by white color corresponds to the cluster of the typical samples. The mean vector of the data points in this cluster is considered as the initial robust estimate of $\theta_0$. By using this initial estimate and performing the algorithm of S-estimator, the estimate of $\theta_0$ will be obtained. ∎

Aptness of the proposed method requires the identification of the number of clusters. For high dimensional data sets one must use an appropriate method to find the number of clusters. Different approaches for finding the number of clusters are reviewed and proposed by Sugar and James (2003). Usually $\hat{\theta}_j$s come from a multimodal distribution. So, for determining the number of clusters one must test whether $\hat{\theta}_j$s come from a unimodal distribution or not. If this test is rejected then $\hat{\theta}_j$s must be clustered into two new groups. A procedure for identifying the number of clusters called G-means is suggested by Hamerly and Elkan (2003). In G-means method the k-means procedure is used which does not cluster the data properly when the data is contaminated. A clustering method called alternative c-means is proposed in Wu and Yang (2002). The alternative c-means method uses a distance norm which is robust to the presence of outliers. This method is applied in this research for identifying the appropriate number of clusters. Computation steps of the proposed method are provided in the followings.

## 3-1-Computation steps for the proposed method

Suppose $m$ random samples of size $n$ observations are given where the $j^{th}$ sample contains $n$ measurements on the response and the independent variables. Let $\{(x_1, Y_{1j}), (x_2, Y_{2j}), ..., (x_n, Y_{nj})\}$ for $j = 1, 2, ..., m$, denote the $j^{th}$ random sample.

a) Choose a wavelet family for transforming response variable, and a significance level $\alpha$ for hypothesis testing.
b) For the $j^{th}$ random sample, estimate the wavelet coefficient vector using equation 9. Denote the resulting wavelet coefficient vector by $\hat{\theta}_j$.

c)

175

c1) Compute the median for $\hat{\boldsymbol{\theta}}_j$s. Let $\mathbf{c}$ be this median. Compute two centres as $\mathbf{c} \pm \mathbf{e}\sqrt{\dfrac{2\gamma}{\pi}}$ where $\gamma$ is the largest eigenvalue of the sample covariance matrix of $\hat{\boldsymbol{\theta}}_j$s and $\mathbf{e}$ is the corresponding eigenvector. Denote these two centres by $\mathbf{z}_1$ and $\mathbf{z}_2$.

c2) Perform clustering about these centers. The method used here is called alternative c-means. This method of clustering is provided in *Appendix A*. By completing the clustering algorithm, two centre denoted as $\mathbf{c}_1 = \mathbf{z}_1^l$ and $\mathbf{c}_2 = \mathbf{z}_2^l$ and two corresponding clusters of $\hat{\boldsymbol{\theta}}_j$s will be obtained.

c3) Let $\boldsymbol{\delta} = \mathbf{c}_1 - \mathbf{c}_2$ be an $n$ dimensional vector which connects the two centres $\mathbf{c}_1$ and $\mathbf{c}_2$. Project the $\hat{\boldsymbol{\theta}}_j$s that clustered in Step c2 onto $\boldsymbol{\delta}$ such as $U_j = \dfrac{\langle \hat{\boldsymbol{\theta}}_j, \boldsymbol{\delta} \rangle}{\langle \boldsymbol{\delta}, \boldsymbol{\delta} \rangle}$; $j = 1, 2, ..., m$ where $\langle ., . \rangle$ denotes the inner product of two vectors. Thus, $U_j$s constitute a one dimensional random sample.

c4) Perform a normality test for $U_j$s at significance level $\alpha$. In this study the Anderson-Darling normality goodness-of-fit test is used. If the normality assumption is rejected keep the original centre, $\mathbf{c}$, and discard $\mathbf{c}_1$ and $\mathbf{c}_2$. Otherwise, keep the centres $\mathbf{c}_1$ and $\mathbf{c}_2$ as the centers of the two clusters $I_1$ and $I_2$, respectively. Then go to step c1, to partition each cluster into two clusters. When no newer clusters are found, step c is fathomed and the corresponding clusters are found. Let $k_c$ be the number of clusters found in this step.

d) Let $\hat{\boldsymbol{\theta}}_{i1}, \hat{\boldsymbol{\theta}}_{i2}, ..., \hat{\boldsymbol{\theta}}_{im_i}$ denote the elements of the $i^{th}$ cluster obtained in step c4 for $i = 1, 2, ..., k_c$ where $m_i$ is the number of elements of the $i^{th}$ cluster. For each of the $k_c$ clusters, define the ellipsoid containing half of all $\hat{\boldsymbol{\theta}}_j$ vectors. Compute $MED_i^2$ as:

$$MED_i^2 = \underset{j=1,2,...,m}{med} \left\{ d^2 \left( \hat{\boldsymbol{\theta}}_j; \underset{h=1,2,...,m_i}{med} (\hat{\boldsymbol{\theta}}_{ih}), \underset{h=1,2,...,m_i}{cov} (\hat{\boldsymbol{\theta}}_{ih}) \right) \right\}; \quad i = 1, 2, ..., k_c \tag{15}$$

where $med(.)$ and $cov(.)$ denote the sample median vector and the sample covariance matrix and $d^2(\mathbf{x}; \mathbf{b}, \mathbf{C}) = (\mathbf{x} - \mathbf{b})'\mathbf{C}^{-1}(\mathbf{x} - \mathbf{b})$. The volume of the $i^{th}$ ellipsoid is proportional to

$$V_i = \sqrt{\left| MED_i^2 \times \underset{h=1,2,...,m_i}{cov} \left( \hat{\boldsymbol{\theta}}_{ih} \right) \right|}; \quad i = 1, 2, ..., k_c \tag{16}$$

e) Compute $V_i$ for cluster $i$ using equation 16 and find the cluster with minimum value for $V_i$. Use the elements of the selected cluster and apply the algorithm provided in *Appendix B* to compute $\hat{\boldsymbol{\theta}}_{00}$ and $\hat{\boldsymbol{\Sigma}}_0$.

f) Compute the S-estimate of $\boldsymbol{\theta}_0$ using $\hat{\boldsymbol{\theta}}_{00}$ and $\hat{\boldsymbol{\Sigma}}_0$ as the starting points, by applying the algorithm provided in Rock and Woodruff (1993).

The S-estimation algorithm provides estimators for the covariance matrix of $\hat{\boldsymbol{\theta}}_j$s as well as $\boldsymbol{\theta}_0$. Let $\hat{\boldsymbol{\theta}}_0$ and $\hat{\boldsymbol{\Sigma}}$ show the S-estimates of $\boldsymbol{\theta}_0$ and the covariance matrix of $\hat{\boldsymbol{\theta}}_j$, respectively. It is stated in

Wasserman (2006) that when the error terms in equation 2 are uncorrelated then, $\hat{\boldsymbol{\theta}}_j; \, j = 1, 2, ..., m$ are also uncorrelated. The mean of the diagonal elements of $\hat{\boldsymbol{\Sigma}}$ may be considered as a robust estimator of the error terms variance. So, the proposed method estimates $\boldsymbol{\theta}_0$ and the error terms variance with no sensitivity to the presence of outliers. The application of this proposed method is shown in example 3.

**Example 3:**
In this example $m = 500$ random samples of size $n$ are generated from the following model.
$$Y_{ij} = Doppler(x_i) + \varepsilon_{ij}; \; i = 1, 2, ..., 128; \; j = 1, 2, ..., 50 \tag{17}$$
Measurements are made on $n = 128$ points on each profile. In model represented in Equation 17 error terms are distributed according to a mixed normal distribution as the following:
$$D(\varepsilon_{ij}) = 0.9972 N(0, 0.05^2) + 0.0028 N(10, 0.05^2) \tag{18}$$
In equation 18, only $0.28\%$ of the error terms are outliers. Each sample contains 128 error terms. So, the probability of each sample being contaminated is equal to $1 - (1 - 0.0028)^{128} = 0.3$. Therefore, the contamination ratio in this example is about 30%. After generating 50 samples, the vector of estimated wavelet coefficients for each sample is computed using Equation 9. Let $\hat{\boldsymbol{\theta}}_j; \, j = 1, 2, ..., 50$ denote these vectors. By applying the proposed method, the vector of wavelet coefficients, $\boldsymbol{\theta}_0$, is then estimated. The sample mean vector of $\hat{\boldsymbol{\theta}}_j; \, j = 1, 2, ..., 50$ is also computed to estimate $\boldsymbol{\theta}_0$. After estimating $\boldsymbol{\theta}_0$, the in control profile is estimated by using the inverse discrete wavelet transformation. The two estimates of the in control profile are shown in figure 5.
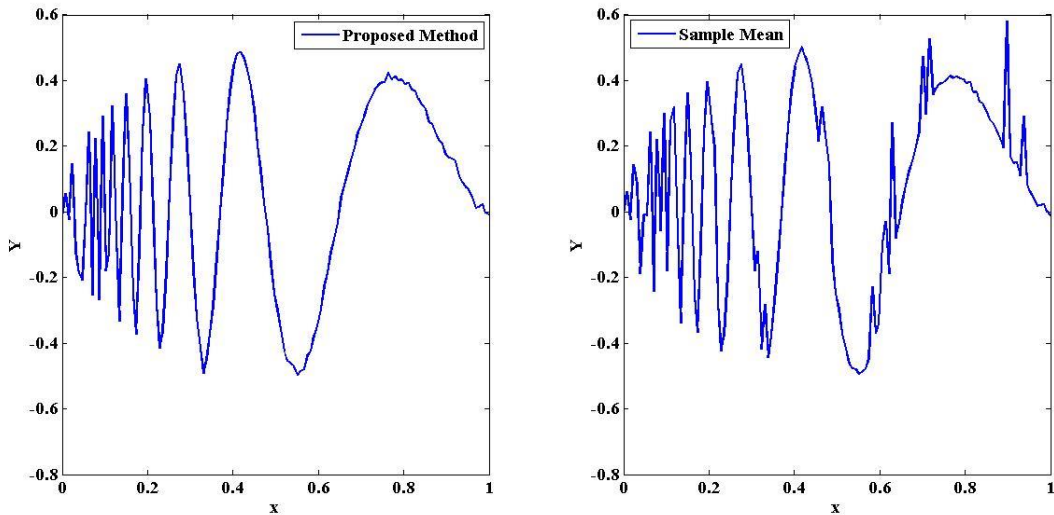


**Fig 5.** Estimates of in control profile in the presence of outliers using proposed method and sample mean

Figure 5 shows that the proposed method estimates the in control profile more precisely than the sample mean in presence of outliers. This method is used in section 4 to monitor the process in phase II.

# 4-Proposed sequential monitoring of profiles in phase II

As mentioned in section 1, the main purpose of phase II of profile monitoring is to detect the out of control situations. More formally, consider the model given in equation 3. In this model it is assumed that the process is in control in phase II and the functional form of the profile is $f_0(x)$ by time $\tau$. At $\tau$, the process changes and the functional form of the profile becomes $f_1(x)$. As stated before many change point models have been introduced in the area of profile monitoring. One of the most suitable methods of monitoring profiles is introduced in Chicken et al. (2009). This method is based on the notion of sequential probability ratio test. This method is not only based on the normality of the error terms but also requires taking $m$ random samples from the in control process in phase I when estimating $\theta_0$. However, in some practical situations the error terms may not follow a normal distribution. So, it is desired to use some other methods insensitive to the normality assumption. A method is now being proposed to monitor the profiles sequentially in phase II. More formally in phase II we assume that the samples are taken from the process represented according to the following model:

$$Y_{ij} = \begin{cases} f_0(x_i) + \varepsilon_{ij} & m \le j \le \tau \\ f_1(x_i) + \varepsilon_{ij} & j > \tau \end{cases} \quad i = 1, 2, ..., n \tag{19}$$

where $Y_{ij}$ is the $i^{th}$ observation from the $j^{th}$ sample, $f_0(x)$ is the in control profile and $f_1(x)$ is the out of control profile and $\tau$ is the change point. Measurements are taken on $n$ points on each sample. Error terms in equation 19 are assumed to be i.i.d. random variables with mean zero and unknown variance $\sigma^2$ and not necessarily normal. For the $t^{th}$ sample taken sequentially in phase II, the following statistic is computed:

$$w_t = \frac{\left\| \hat{\theta}_t - \theta_0 \right\|^2}{\sigma^2}; \quad t = m+1, m+2, ... \tag{20}$$

where $\hat{\theta}_t$ is the estimated wavelet coefficients vector of the $t^{th}$ sample. In many practical applications one must estimate $\theta_0$ and $\sigma^2$ based on the $m$ samples taken in phase I. The possibility of the data being contaminated requires use of the estimators proposed in section 3. So, for the $t^{th}$ sample the following statistic must be computed:

$$w_t = \frac{\left\| \hat{\theta}_t - \hat{\theta}_0 \right\|^2}{\hat{\sigma}^2} \tag{21}$$

$w_t$ is simply the squared Mahalanobis distance of $\hat{\theta}_t$ from $\hat{\theta}_0$ with respect to the covariance matrix $\hat{\sigma}^2 \mathbf{I}_{n \times n}$. For an in control process in phase II, $\hat{\theta}_t$ s are IID with mean vector $\theta_0$. So, $w_t$ s are also IID. When the process goes out of control at time $\tau$, then $\hat{\theta}_t$; $t = \tau+1, \tau+2, ...$ will be distributed differently. So, before the change point $w_t$; $t = m+1, m+2, ..., \tau$ are distributed according to $F_0$, while after the change point $w_t$; $t = \tau+1, \tau+2, ...$ come from a different distribution say $F_1$, ($F_0 \ne F_1$). Therefore, finding the change point in the process is equivalent to testing the following hypothesis in phase II. Let $F^t$ be the distribution function of $w_t$; $t = m+1, m+2, ...$ . Then, the null hypothesis for no change in model distribution is

$$H_0 : F^t = F_0 ; \quad t = m+1, m+2, ... \tag{22}$$

which is tested against the alternative hypothesis.

$$H_1 : \begin{cases} F^t = F_0 & t = m+1, m+2, ..., \tau \\ F^t = F_1 & t = \tau, \tau+1, ... \end{cases} \tag{23}$$

The problem of finding the change point, $\tau$, in the distribution of the sequence of independent variables is considered in Huskova and Chochola (2010). Their method is used here to monitor the change in the distribution of $w_t$ s in phase II for finding the change point for the profiles. This method is characterized by the following rule:

$$t_{Stop} = m + \inf\left\{ k \geq 1 : |Q(m,k)| \geq 3.0722 q\left(\frac{k}{m}\right) \right\} \tag{24}$$

where $t_{Stop}$ is the time at which the change in process is detected and $Q(m,k)$ is the detector function. This detector function is defined as

$$Q(m,k) = \frac{\sqrt{12}}{\sqrt{m}} \sum_{i=1}^{k} \left( \hat{F}_m\left(w_{m+i}\right) - \frac{1}{2} \right); \ k = 1, 2, ... \tag{25}$$

where $m$ is the number of samples used in phase I to estimate the in control profile, $\hat{F}_m(x)$ is the estimated distribution function obtained in phase I. In Equation 24, $q(x)$ is a boundary function defined as

$$q(x) = (1+x)\left(\frac{x}{x+1}\right)^{0.49} ; \ x > 0 \tag{26}$$

For monitoring a process in phases I and II the following steps must be taken.

1. In phase I, take $m$ random samples of size $n$ observations from the process. Estimate the vector of wavelet coefficients for each sample using equation 9.
2. Estimate $\theta_0$ and $\sigma^2$ using the proposed method discussed in section 3.
3. For each sample taken in phase I, compute the statistic given in equation 21. Let $w_1, w_2, ..., w_m$ be these computed statistics. Based on $w_1, w_2, ..., w_m$ compute $\hat{F}_m(x)$.
4. In phase II, take random samples of size $n$ sequentially.
5. For each sample taken in phase II, compute the statistic given in equation 21. Let $w_t$ ; $t = m+1, m+2, ...$ be these statistics.
6. Compute the detector function given in equation 25 and the boundary function given in equation 26 using $w_t$ s from step 5.
7. If equation 24 is satisfied for the sample taken at time $t$, signal the change and stop the algorithm, otherwise go to step 5.

In the proposed method one needs to compute $\hat{F}_m(x)$, using $w_1, w_2, ..., w_m$. The presence of outliers in the data taken in phase I, contaminates the computed values of $w_1, w_2, ..., w_m$. This is identified by the values for $w_1, w_2, ..., w_m$ exceeding the limit $med\{w_t ; t = 1, 2, ..., m\} + 3MAD\{w_t ; t = 1, 2, ..., m\}$ where $MAD$ is the median absolute deviation from the median. Now, the distribution function $\hat{F}_m(x)$ could be computed using the rest of $w_t$ s. The performance of the proposed method is investigated by means of simulation in section 5.

## 5-Performance evaluation of the proposed change point detection method

In this section the performance of the proposed method of detecting the change point in phase II is evaluated by means of simulation. All simulations are performed using MATLAB ® software. The

proposed method is studied under two different scenarios. In first scenario, the error terms are assumed to be distributed according to a mixed t-student distribution. In second scenario a mixed normal error term is assumed. In each case the two different approaches of the proposed method are compared. The difference between these two approaches is the procedure used in phase I to estimate $\theta_0$. The first approach, which is introduced in section 4 is shown here by $A1$. In this approach the robust estimators introduced in section 3 are used in phase I to estimate $\theta_0$. While, in second approach, shown here by $A2$, the sample mean vector of $\hat{\theta}_j$; $j = 1, 2, ..., m$ is used to estimate $\theta_0$. Two types of changes in functional form of the profile including structural break and structural change which are further explained in the next subsection are considered. These types of changes are also called constant and linear shifts by some authors.

## 5-1-Simulation study for non-normal error terms

In this section the performance of the proposed method is evaluated for the two types of changes, structural break and structural change.

### 5-1-1-Structural break

In each simulation run, $m = 20$ random samples of size $n = 16$ are generated from the following contaminated model in phase I

$$Y_{ij} = Doppler(x_i) + \varepsilon_{ij}; \ i = 1, 2, ..., 16; \ j = 1, 2, ..., 20 \tag{27}$$

where $\varepsilon_{ij}$ s are distributed according to a mixed t-student distribution as

$$D(\varepsilon_{ij}) = (1 - \lambda)t_{(25)} + \lambda t_{(3)} \tag{28}$$

where $\lambda$ is the proportion of contamination of the error terms, $t_{(\upsilon)}$ is the t distribution with $\upsilon$ degrees of freedom. After generating these 20 samples, $\theta_0$ is estimated using the approaches $A1$ and $A2$. In phase II random samples of size $n = 16$ are sequentially generated from the following model:

$$Y_{ij} = \begin{cases} Doppler(x_i) + \varepsilon_{ij} & 21 \leq j \leq 70 \\ Doppler(x_i) + 0.05 + \varepsilon_{ij} & j \geq 71 \end{cases} \ i = 1, 2, ..., 16 \tag{29}$$

where $\varepsilon_{ij}$ s are distributed according to $t_{(25)}$ and $\tau = 70$ is the change point. After generating each sample in phase II, the condition given in equation 24 is evaluated to test whether any change has occurred or not. Let $T_i$ denotes the time where a change is detected in the $i^{th}$ simulation run. For a given value of $\lambda$ this procedure is repeated 5000 times and the mean squared error is computed for the two approaches $A1$ and $A2$:

$$MSE = \frac{\sum_{i=1}^{5000} (T_i - 70)^2}{5000} \tag{30}$$

The proportion of contaminated samples is used to determine the value of $\lambda$. Each sample taken in phase I contains $n$ error terms. Each error term is contaminated with probability of $\lambda$. So, the proportion of contaminated samples, $\lambda_s$, is equal to:

$$\lambda_s = 1 - (1 - \lambda)^n \tag{31}$$

In simulation studies, $\lambda_s = 0(0.01)0.3$. Results are shown in figure 6.
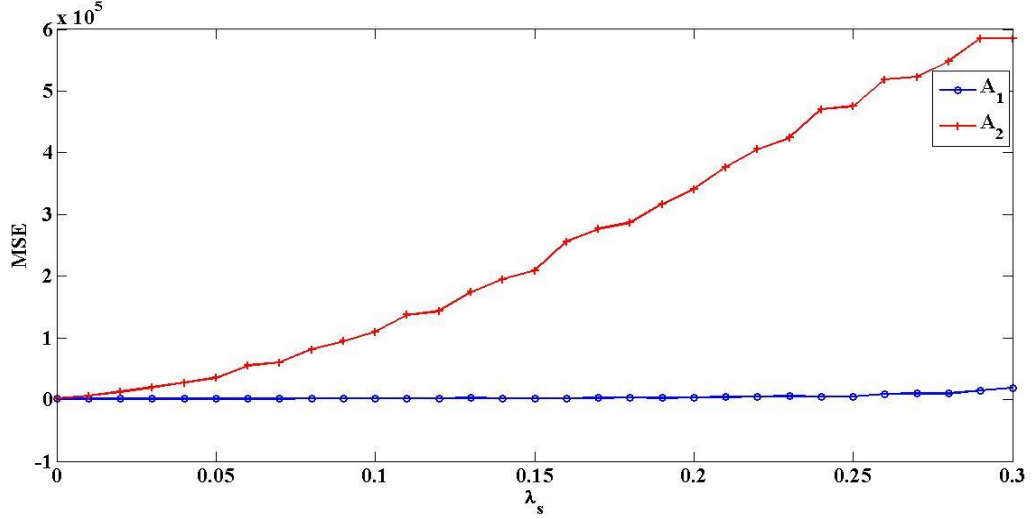
**Fig 6.** Mean squared error for detecting structural break with mixed t distribution

Investigation of figure 6 indicates that the $A1$ approach detects the shift in the process sooner than the $A2$ approach. As well as this property, the $A1$ approach is insensitive to the changes in the proportion of contamination, $\lambda_s$ . This is due to using a robust estimator in phase I for $A1$ approach. The $A2$ approach is highly sensitive to the value of $\lambda_s$ . Only when error terms in phase I are not contaminated, $\lambda_s = 0$ , the $A2$ approach performs similar to $A1$ approach.

### 5-1-2-Structural change

In each simulation run, $m = 20$ random samples of size $n = 16$ are generated from the following model in phase I:

$$Y_{ij} = Doppler(x_i) + \varepsilon_{ij}; \; i = 1,2,...,16; \; j = 1,2,...,20 \tag{32}$$

where $\varepsilon_{ij}$ s are the mixed t distributed random variables used to model the contamination scenario and is defined as before:

$$D(\varepsilon_{ij}) \square (1-\lambda)t_{(25)} + \lambda t_{(3)} \tag{33}$$

$\lambda_s = 0(0.01)0.3$ is also applied here. After generating $m = 20$ samples, the vector of wavelet coefficients $\boldsymbol{\theta}_0$ is estimated using the proposed robust estimator for the $A1$ approach, and the sample mean vector for the $A2$ approach. In phase II, random samples of size $n = 16$ are sequentially generated according to the following model:

$$Y_{ij} = \begin{cases} Doppler(x_i) + \varepsilon_{ij} & 21 \leq j \leq 70 \\ Doppler(x_i) + 0.1x_i + \varepsilon_{ij} & j \geq 71 \end{cases} \; i = 1,2,...,16 \tag{34}$$

where $\varepsilon_{ij}$ s are $t_{(25)}$ random variables and $\tau = 70$ is the change point. After generating each sample in phase II, the change point detection test is performed to determine whether any change has

181

occurred or not. This procedure is repeated 5000 times and the *MSE* is computed for each approaches, $A1$ and $A2$,. Results are shown in figure 7.
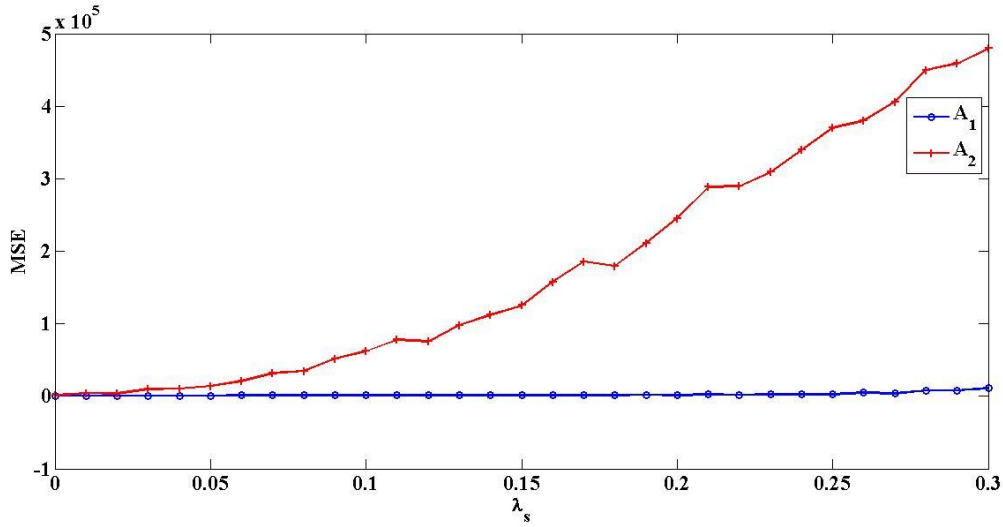


**Fig 7.** Mean squared error for detecting structural change with mixed t distribution

Figure 7 indicates that the $A1$ approach detects the structural change more precisely than the $A2$ approach. When the value of the contamination proportion, $\lambda_s$, increases the *MSE* of the $A1$ approach remains almost constant, while the *MSE* of the $A2$ approach increases with increase in $\lambda_s$.

## 5-2-Simulation study for normal error terms

In this section the performance of the proposed method is evaluated when error terms are normally distributed.

### 5-2-1-Structural break

In each simulation run $m = 20$ random samples of size $n = 16$ are generated from the following model:

$$Y_{ij} = Doppler(x_i) + \varepsilon_{ij}; \ i = 1, 2, ..., 16; \ j = 1, 2, ..., 20 \tag{35}$$

where $\varepsilon_{ij}$ s are distributed according to a mixed normal distribution as:

$$D(\varepsilon_{ij}) = (1 - \lambda)N(0, 0.05^2) + \lambda N(\mu, 0.05^2) \tag{36}$$

According to equation 31, the values for $\lambda$ are selected as $\lambda = 0.0066, 0.0138$ to assure the values of $0.1$ and $0.2$ for $\lambda_s$, respectively. The magnitude of the contamination of the error terms is $\mu = -1(0.1)1$. After generating 20 samples, $\theta_0$ is estimated using the $A1$ and $A2$ approaches. In phase II random samples of size $n = 16$ are sequentially generated from the following model:

$$Y_{ij} = \begin{cases} Doppler(x_i) + \varepsilon_{ij} & 21 \le j \le 70 \\ Doppler(x_i) + 0.05 + \varepsilon_{ij} & j \ge 71 \end{cases} \quad i = 1, 2, ..., 16 \tag{37}$$

where $\varepsilon_{ij}$ s are distributed according to $N(0, 0.05^2)$ distribution and $\tau = 70$ is the change point. For each combination of $\lambda_s$ and $\mu$ this procedure is repeated 5000 times and the *MSE* is computed for the approaches $A1$ and $A2$. Results for $\lambda_s = 0.1$ are illustrated in figure 8.
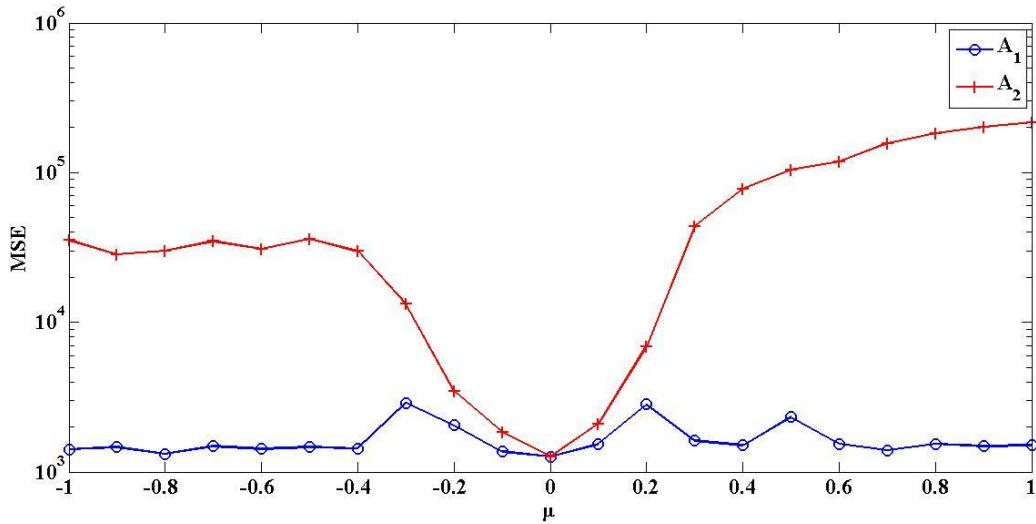
**Fig 8.** Mean squared error for detecting structural break with mixed normal distribution for $\lambda_s = 0.1$

For the sake of more clarity the vertical axis is shown in logarithmic scale. Results for $\lambda_s = 0.2$ are illustrated in figure 9.
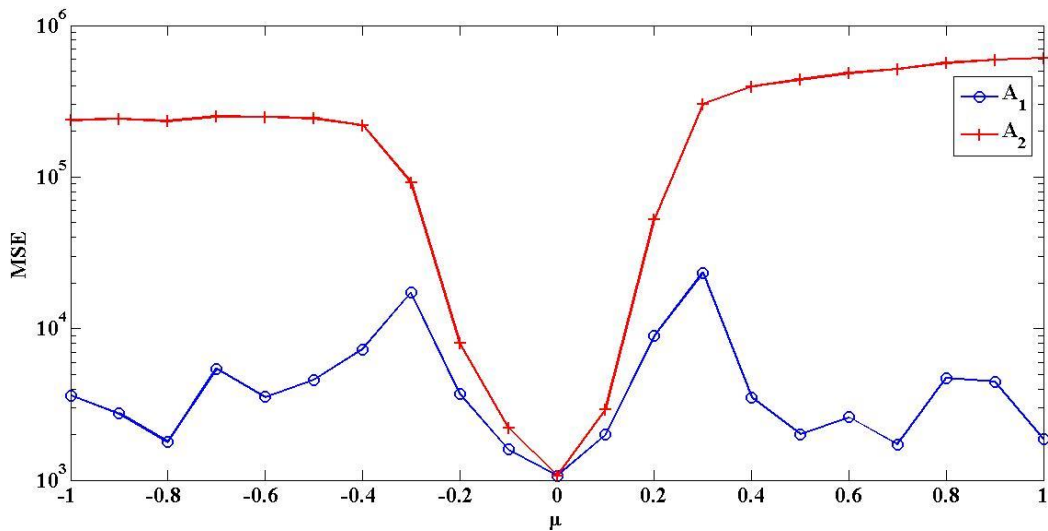


**Fig 9.** Mean squared error for detecting structural break with mixed normal distribution for $\lambda_s = 0.2$.

Investigation of figures 8 and 9 indicates that the $A1$ approach detects the structural break in process sooner than the $A2$ approach.

### 5-2-2-Structural change

Simulation settings for this scenario are the same as the ones in subsection 5.1.2. except that the error terms are normally distributed according to $N(0, 0.05^2)$. Results for $\lambda_s = 0.1$ and $\lambda_s = 0.2$ are illustrated in Figure 10 and 11, respectively.
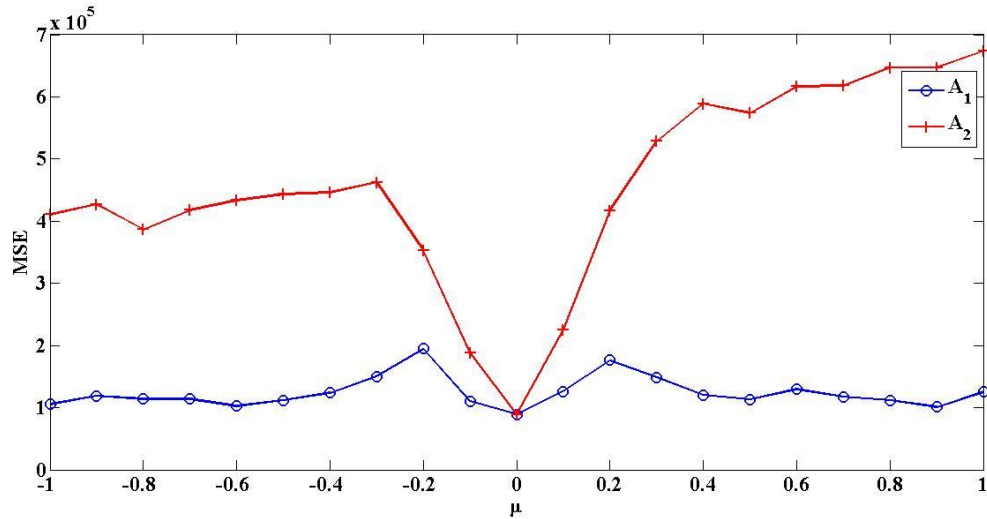
**Fig 10.** Mean squared error for detecting structural change with mixed normal distribution for $\lambda_s = 0.1$
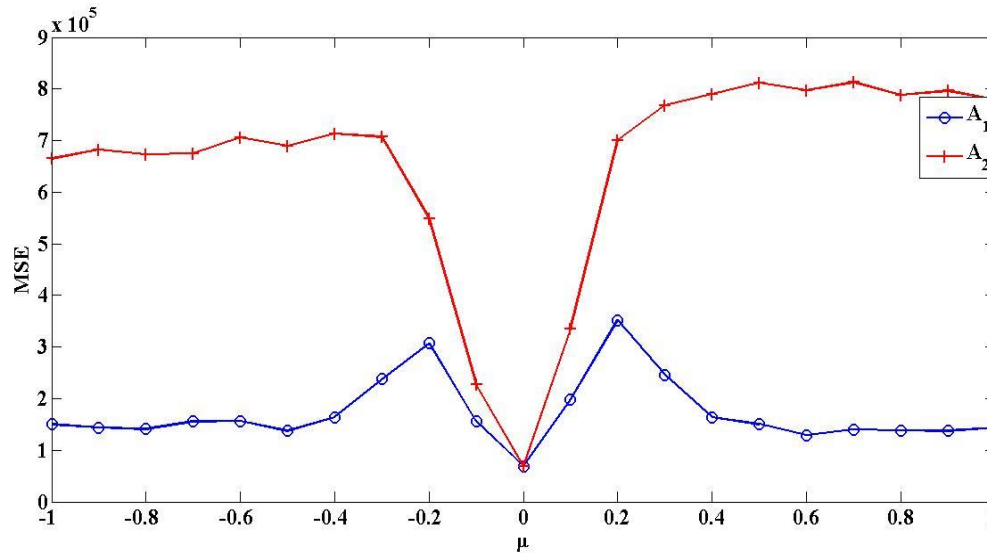


**Fig 11.** Mean squared error for detecting structural change with mixed normal distribution for $\lambda_s = 0.2$

## 6-Conclusions

In this work a robust method for monitoring complicated and unsmooth profiles is proposed. This method is based on wavelet transformation. Using wavelet transformation, the data are transformed to obtain the estimated wavelet coefficients. These estimated wavelet coefficients are used to estimate the in control wavelet coefficients vector, $\theta_0$. Under the normality assumption of the error terms, the classical estimators such as sample mean vector of the estimated wavelet coefficients are usually used to estimate $\theta_0$. However, the presence of outliers in the data in phase I affects the accuracy of the classical estimator of $\theta_0$. It is shown here that the presence of outliers in phase I results different clusters of estimated wavelet coefficients. This property of the estimated wavelet coefficients helps to

develop a robust estimator for $\theta_0$ in phase I. This estimator is an integration of the robust clustering and the S-estimator. This estimator is shown to be less sensitive to the presence of outliers compare to the classical estimator of $\theta_0$. As well as estimating the in control profile in phase I, a procedure for finding the change point in phase II is also proposed. Change point detection methods are usually based on normality assumption of the error terms. However, in some practical situations the error terms may not follow a normal distribution. The proposed method of detecting the change point does not require the assumption of normality for the error terms. In the suggested method, for each sample a test statistic based on the proposed robust estimator of $\theta_0$ is computed. The empirical distribution of this statistic is used to find the change point. Simulation study shows that the proposed method detects the change point more accurately than the classical methods in presence of outliers. In summary the proposed methods for estimating the in control profile in phase I and monitoring the process in phase II, are used when the profile is complicated and the data is contaminated. The proposed method for finding the change point in phase II does not require the normality of the error terms.

## Acknowledgements

## References

Chen, Y., Birch, J. B., & Woodall, W. H. (2015). A phase I cluster-based method for analyzing nonparametric profiles. *Quality and Reliability Engineering International.* 31(8), 1675-1689.

Chicken, E., Pignatiello, J., & Simpson, J. (2009). Statistical process monitoring of non-linear profiles using wavelets. *Journal of Quality Technology*, 41-198-212.

Donoho, D. (1982). *Breakdown properties of multivariate location estimators*. Ph. D. diss, Harvard University.

Gardner, M., Lu, J., Gyurcsik, R., Davis, J., & Mozumder, P. (1997). Equipment fault detection using spatial signatures. *IEEE Transaction on Components, Packing and Manufacturing Technology.* 20, 295-303.

Grasso, M., Menafoglio, A., Colosimo, B. M., & Secchi, P. (2016). Using curve-registration information for profile monitoring. *Journal of Quality Technology.* 48(2), 99-127.

Hakimi, A., Amiri, A., & Kamranrad, R. (2017). Robust approaches for monitoring logistic regression profiles under outliers. *International Journal of Quality & Reliability Management.* 34(4), 494-507.

Hamerly, G., & Elkan, C. (2003). Learning the k in k-means. *In proceeding of seventeenth Annual Conference of Neural Information Processing Systems (NIPS),* 281-288.

Huber, P. (1985). Projection pursuit. *The Annals of Statistics*, 13, 435-475.

Huskova, M., & Chochola, O. (2010). Simple sequential procedures for change in distribution. *Nonparametric and Robustness in Modern Statistical Inference and Time Series Analysis,* 7, 95-104.

Jeong, M., & Lu, J. (2006). Wavelets based SPC procedure for complicated functional data.

*International Journal of Production Research.* 44, 729-744.

Kamranrad, R., & Amiri, A. (2016). Robust Holt-Winter based control chart for monitoring autocorrelated simple linear profiles with contaminated data. *Scientia Iranica.* 23(3), 1345-1354.

Kang, L., & Albin, S. (2000). On-line monitoring when the process yields a linear profile. *Journal of Quality Technology.* 32, 418-426.

Khedmati, M., & Akhavan Niali, S. T. (2016). Monitoring simple linear profiles in multistage processes by a MaxEWMA control chart. *Computers & Industrial Engineering.* 98, 125-143.

Kim, K., Mahmoud, M., & Woodall, W. (2003). On the monitoring of linear profiles. *Journal of Quality Technology.* 35,317-328.

Mahmoud, M., Parker, P., Woodall, W., & Hawkins, D. (2007). A change point method for linear profile data. *Quality and Reliability Engineering International.* 23, 247-268.

Maronna, R., Martin, R., & Yohai, V. (2006). *Robust Statistics: Theory and Methods.* New York: Wiley.

Mestek, O., Pavlik, J., & Suchanek, M. (1994). Multivariate control charts: control charts for calibration curves. *Fresenius' Journal of Analytical Chemistry.* 350, 344-351.

Montgomery, D. C. (2012). *Introduction to Statistical Quality Control,* 7th ed., New York: Wiley.

Noorossana, R., Aminmadani, M., & Saghaei, A. (2016). Effect of phase I estimation error on the monitoring of simple linear profiles in phase II. *The International Journal of Advanced Manufacturing Technology.* 84 (5-8), 873-884.

Ogden, R. (1997). *Essential Wavelets for Statistical Applications and Data Analysis.* Birkhauser, Boston.

Paynabar, K., Qiu, P., & Zou, C. (2016). A change point approach for phase-I analysis in multivariate profile monitoring and diagnosis. *Technometrics.* 58(2), 191-204.

Qi, D., Wang, Z., Zi, X., & Li, Z. (2016). Phase II monitoring of generalized linear profiles using weighted likelihood ratio charts. *Computers & Industrial Engineering.* 94, 178-187.

Rock, D., & Woodruff. D. (1993). Computation of robust estimators of multivariate location and shape. *Statistica Neerlandica.* 47, 27-42.

Rousseeuw, P. (1985). Multivariate estimation with high breakdown point. *Mathematical Statistics and its Application,* B, 283-297.

Rousseeuw, P., & Leroy, A. (1987). *Robust Regression and Outlier Detection.* New York: Wiley.

Rousseeuw, P., & Yohai, V. (1984). *Robust regression by means of S-estimators,* in: Franke, J., Hardle, W., & Martin, D. (Eds.), Robust and Nonlinear Time Series Analysis, Springer, pp. 256-272.

Shadman, A., Mahlooji, H., Yeh, A., & Zou, C. (2015). A change point method for monitoring generalized linear profiles in phase I. *Quality and Reliability Engineering International.* 31(8), 1367-1381.

Shadman, A., Zou, C., Mahlooji, H., & Yeh, A.B. (2017). A change point method for phase II monitoring of generalized linear profiles. *Communications in Statistics-Simulation and Computation.* 46 (1), 559-578.

Stahel, W. (1981). *Robust schatzungen: infinitestimale optimalitatund und schatzungen von kovariazmatrizen.* Ph. D. diss, ETH Zurich.

Stover, F., & Brill, R. (1998). Statistical quality control applied to ion chromatography calibrations. *Journal of Chromatography*. 804,37-43.

Sugar, C. & James, G. (2003). Finding the numbers of clusters in a dataset. *Journal of the American Statistical Association*, 98(463), 750-763.

Wasserman, L. (2006). *All of Non Parametric Statistics.* Springer.

Woodall, W. H. (2017). Bridging the gap between theory and practice in basic statistical process monitoring. *Quality Engineering.* 29(1), 2-15.

Woodall, W. H., & Montgomery, D. C. (2014). Some current directions in the theory and application of statistical process monitoring. *Journal of Quality Technology.* 46(1), 78-94.

Wu, K., & Yang, M. (2002). Alternative c-means clustering algorithm. *Pattern Recognition,* 35, 2267-2278.

Zhou, C., Zhang, Y., & Wang, Z. (2006). A control chart based on a change point model for monitoring linear profiles. *IIE Transactions,* 38, 1093-1103.

## Appendix A: Alternative c-means Algorithm

According to Wu and Yang (2002) for clustering $m$ vectors of estimated wavelet coefficients the following steps must be accomplished:

Set $l = 0$.

**Step 1:** First of all vectors of estimated wavelet coefficients must be divided into two groups, $I_1$ and $I_2$. For this purpose consider a sample of size $m$, $\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_m$ to be clustered into two clusters. Let

$$\beta = \left( \frac{\sum_{j=1}^{m} \left\| \hat{\theta}_i - \bar{\hat{\theta}} \right\|^2}{m} \right)^{-1} \tag{A.1}$$

where $\|.\|$ is the Euclidian norm of a vector and $\bar{\hat{\theta}}$ is the sample mean vector of estimated wavelet coefficients. Define $I_1$ and $I_2$ to be the two different index set where index $j \in I_i$; $i = 1, 2$ if

$$j \in \begin{cases} I_1 & if \quad \exp\left(-\beta \left\| \hat{\theta}_j - \mathbf{z}_1 \right\|^2\right) > \exp\left(-\beta \left\| \hat{\theta}_j - \mathbf{z}_2 \right\|^2\right) \\ I_2 & if \quad \exp\left(-\beta \left\| \hat{\theta}_j - \mathbf{z}_1 \right\|^2\right) < \exp\left(-\beta \left\| \hat{\theta}_j - \mathbf{z}_2 \right\|^2\right) \end{cases} \quad j = 1, 2, ..., m \tag{A.2}$$

Using equation A.2 each vector $\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_m$ is assigned to one of the index sets, $I_1$ and $I_2$.

**Step 2:** In this step the centroid of each cluster is computed using robust distance norm. Find $\mathbf{z}_i^{l+1}$ using

$$\mathbf{z}_i^{l+1} = \frac{\sum_{j \in I_1} \left( \exp\left(-\beta \left\| \hat{\theta}_j - \mathbf{z}_i^l \right\|^2\right) \right) \hat{\theta}_j}{\sum_{j \in I_1} \left( \exp\left(-\beta \left\| \hat{\theta}_j - \mathbf{z}_i^l \right\|^2\right) \right)}; \; i = 1, 2; \; \mathbf{z}_i^0 = \mathbf{z}_i \tag{A.3}$$

Stop: If $\max_{i=1,2} \left\| \mathbf{z}_i^{l+1} - \mathbf{z}_i^l \right\| < \zeta$ then stop where $\zeta$ is the pre specified stopping tolerance. Otherwise $l \leftarrow l + 1$ and go to Step 1 and update the sets $I_1$ and $I_2$.

## Appendix B

Let $\mathbf{b}_j$; $j = 1, 2, ..., m_*$ be the elements of the selected cluster to compute the starting point for the S-estimator where $m_*$ represents the number of elements in selected cluster. Estimate the mean vector and the covariance matrix using:

$$\hat{\theta}_{*0} = mad(\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_{m_*})$$

$$\hat{\Sigma}_{*0} = \frac{MED_*^2}{\chi^2(0.5; n)} \mathbf{C}_* \tag{B.1}$$

$$\mathbf{C}_* = cov\left(\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_{m_*}\right)$$

where $\chi^2(1-\alpha; n)$ is the lower quantile of the chi squared distribution with $n$ degrees of freedom and

$$MED_*^2 = \underset{j=1,2,\ldots,m}{med} \left\{ d^2 \left( \hat{\boldsymbol{\theta}}_j ; \hat{\boldsymbol{\theta}}_{*0}, \mathbf{C}_* \right) \right\} \tag{B.2}$$

Put weights $\upsilon_j; j = 1, 2, \ldots, m$ on $\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2, \ldots, \hat{\boldsymbol{\theta}}_m$ to get a more accurate estimator. Based on these weights the final estimator will be:

$$\hat{\boldsymbol{\theta}}_{00} = \frac{\displaystyle\sum_{j=1}^{m} \upsilon_j \hat{\boldsymbol{\theta}}_j}{\displaystyle\sum_{j=1}^{m} \upsilon_j}$$

$$\hat{\boldsymbol{\Sigma}}_0 = \frac{\displaystyle\sum_{j=1}^{m} \upsilon_j (\hat{\boldsymbol{\theta}}_j - \hat{\boldsymbol{\theta}}_{00})(\hat{\boldsymbol{\theta}}_j - \hat{\boldsymbol{\theta}}_{00})'}{\displaystyle\sum_{j=1}^{m} \upsilon_j - 1} \tag{B.3}$$

where

$$\upsilon_j = \begin{cases} 1 & d_2(\hat{\boldsymbol{\theta}}_j; \hat{\boldsymbol{\theta}}_{*0}, \hat{\boldsymbol{\Sigma}}_{*0}) \leq \chi^2(0.975; n) \\ 0 & o.w. \end{cases} \tag{B.4}$$