

# An Alternative Hotelling $T^2$ Control Chart Based on Minimum Vector Variance (MVV)

Sharipah S.S.Yahaya

College of Art and Sciences, Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia

Tel: 60-4-928-6932 E-mail: sharipah@uum.edu.my

Hazlina Ali

College of Art and Sciences, Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia

Tel: 60-4-928-6919 E-mail: hazlina@uum.edu.my

Zurni Omar

College of Art and Sciences, Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia

Tel: 60-4-928-6919 E-mail: zurni@uum.edu.my

Received: May 8, 2011

Accepted: July 7, 2011

doi:10.5539/mas.v5n4p132

*The research is financed by FRGS. No. 11764(Sponsoring information)*

## Abstract

The performance of traditional Hotelling  $T^2$  control chart using classical estimators in Phase I suffers from masking and swamping effects. To alleviate the problem, robust location and scale estimators are recommended. This paper proposed a robust Hotelling  $T^2$  control chart for individual observations based on minimum vector variance (MVV) estimators as an alternative to the traditional multivariate  $T^2$  control chart for Phase II data. MVV is a new robust estimator which possesses the good properties as in minimum covariance determinant (MCD) with better computational efficiency. Through simulation study, we evaluate the performance of the proposed chart in terms of probability of detection and false alarm rates, then compared with the performance of the traditional charts and the chart issued from MCD estimators. The results showed that MVV control chart has competitive performance relative to MCD and traditional control charts even under certain location parameter shifts in Phase I data.

**Keywords:** Hotelling's  $T^2$  chart, Minimum covariance determinant (MCD) estimator, Minimum vector variance (MVV) estimator, Reweighted MCD

## 1. Introduction

In many industrial settings it is frequently required to monitor more than one interrelated variables. The Hotelling's  $T^2$  control chart is one of the multivariate statistical tools which is widely used to detect the presence of special-causes of variation by monitoring a nominal mean vector  $\mu$ . This chart is popular as it possesses almost all the desirable characteristics for a multivariate control chart such as ease of application, flexibility, sensitivity to small process changes, and the availability of software for application (Mason & Young, 2002). Like any other control charts for monitoring the variability in a process, its construction consists of Phase I and Phase II (Alt, 1985) which are also referred to as retrospective and prospective analysis respectively (Woodall & Montgomery, 1999). Phase I focuses on analyzing historical data to determine whether the process is in control by estimating the in-control parameters of the process and the control limits. While in Phase II, the centre of attention is on monitoring on-line data to quickly detect shifts in the process from the in-control parameter values estimated in Phase I. Unusual observations in Phase I can lead to the inflation of control limits and reduction of power to detect process changes in Phase II. Therefore a successful Phase II analysis depends on a successful Phase I analysis in estimating in-control mean, variance, and covariance parameters.

The preliminary data set collected in retrospective analysis involves either initial subgroups or individual observations. In many situations, data are collected according to the rational subgroups concept. Nevertheless, sometimes data come in the form of individual observations especially when the production rate is too slow to conveniently collect subgroup size greater than one. For individual multivariate observations, the parameter estimates for the mean vector and covariance matrix in Phase I are based on pooling all the observations (Jackson, 1985; Tracy et al., 1992; Wierda, 1994; Lowry & Montgomery, 1995). However, this approach will cause the variance estimates to inflate if the special-cause of variation present in Phase I, as demonstrated by Sullivan and Woodall (1996;1998), Vargas (2003) and Williams (2004). To overcome the problem, alternative estimation methods have been proposed in the literature. One of the approaches is to calculate the  $T^2$  statistic based on successive differences variance-covariance matrix estimator (Holmes & Mergen, 1993; Sullivan & Woodall, 1996;1998; and Williams et al., 2006). Though this approach is effective in detecting shifts in the mean

vector, it fails to detect outliers as shown in Vargas (2003). Another approach is to use robust estimators of the process parameters.

A notable use of robust estimator is in identifying the deviation of data, or outliers. Compared to the classical statistics, the use of robust statistics will give a clearer variability description between an outlier and 'good data', whereas the classical statistics will vaguely show the difference (Hampel et al., 1986). There are two approaches to deal with outliers. The first approach is to identify and remove outliers before using the remaining good data points to calculate the classical estimators. The other approach is to use the robust estimators in place of classical estimators (Beckman & Cook, 1983). A wide range of robust estimators of multivariate location and scatter is available, see Maronna and Zamar (2002) and Maronna et al. (2006) for a review. Nonetheless, the minimum volume ellipsoid (MVE) and minimum covariance determinant (MCD) estimators introduced by Rousseeuw (1985) have received a considerable attention by scientific community and widely used in practice. The advantage of using MVE estimators is that, they have high breakdown point of approximately 50% and also affine equivariant (Lopuhaa & Rousseeuw, 1991, pg 236). However the computation of MVE estimator is very expensive and it may not even be computationally feasible (Hadi, 1992, pg 762). In addition, there is no fast algorithm known to compute the estimator. This is due to the fact that MVE has poor rate of convergence (Lopuhaa & Rousseeuw, 1991, pg 237) and fails to cope with large sample of more than 30 (Rousseeuw & van Driessen, 1999, pg 213). To alleviate the complexity of MVE, Rousseeuw (1985) also introduced the minimum covariance determinant (MCD) method. MVE and MCD estimators have the same characteristics with respect to affine equivariance, high breakdown value and bounded influence function properties (Rousseeuw & Leroy, 1987). The only difference is in the criteria used where MVE minimizes the volume of the ellipsoid on  $h = (n + p - 1) / 2$  data, while MCD minimizes the determinant of the covariance matrix based on the  $h$  data.

The MCD estimator is more attractive than MVE because it has a better convergence rate of  $n^{-1/2}$  compared to  $n^{-1/3}$  of MVE (Butler et al., 1993; Croux & Haesbroeck, 1999) and MCD gives the exact solution (Hadi, 1992; Hubert et al., 2005).

Lopuhaa and Rousseeuw (1991) realized that the efficiency of high breakdown methods can be quite low, and proposed reweighted MCD (RMCD) estimator to alleviate the problem. Croux and Haesbroeck (1999) employed the reweighted version and noticed that this approach maintains the breakdown point of the initial MCD estimators, while attaining a better efficiency. To compute the initial MCD estimator and its reweighted, various algorithms have been suggested. Most of the algorithms attempt to increase the computational efficiency because to obtain approximate values of these estimators is not only expensive, but could be impossible for large sample sizes with large number of quality characteristics (dimensions). Nevertheless, the main contribution in this domain is the Fast MCD algorithm proposed by Rousseeuw and van Driessen (1999) and improved by Hubert et al. (2005) which is available in many computer packages such as Matlab, R, SAS, and S-Plus. However, Fast MCD is not without limitation. For example, the use of minimum covariance determinant as the objective function in data concentration process can be computationally laborious especially when the data set is of high dimension. On the other hand, as Angiulli and Pizzuti (2005) have pointed out, the computational efficiency is as important as effectiveness. Furthermore, as noted by Fauconnier and Haesbroeck (2009), Fast MCD algorithm may return different results when used repeatedly in the same or in different statistical packages and could be more critical when  $n/p$  are small. To overcome the weaknesses of Fast MCD algorithm, Herwindiati (2006)

proposed minimum vector variance (MVV) as an alternative measure of multivariate data concentration. Herwindiati et al. (2007) revealed that MVV was successfully used as an objective function in Fast MCD algorithm to substitute the MCD criterion. The findings showed that MVV is computationally more efficient than Fast MCD and as effective as Fast MCD in labelling outliers. Therefore MVV is one of the methods where the algorithm itself is referred to as the estimator as discussed in Werner (2003). A detail explanation about this method can be referred in Section 2.

The study on the significant role of MVE, MCD and RMCD estimators in scientific application can be easily found in the literature specifically in the construction of robust Hotelling  $T^2$  chart. Vargas (2003) and Jensen et al. (2007) introduced robust control charts based on MVE and MCD estimators for multivariate individual observations. They identified and removed the outliers in Phase I analysis and then calculate the classical estimators using the remaining good data points for Phase II analysis. Through this approach, the computability and breakdown point of the estimators become more important, but statistical efficiency is not as crucial because the highly robust estimators will eventually be replaced by classical estimators in Phase II analysis (Jensen et al., 2007). They noticed some drawbacks when MVE and MCD are used in Phase I. The  $T^2$  issued from MVE did not perform under large sample size. Conversely,  $T^2$  issued from MCD needs a larger sample size if large number of outliers is suspected to ensure that MCD estimator does not breakdown and lose its ability especially when monitoring with more variables ( $p$ ).

To abate the problems, Chenouri et al. (2009) proposed robust Hotelling  $T^2$  chart based on RMCD estimator. Besides possessing the nice properties of MCD estimator, this estimator is not unduly influenced by outliers and

is more efficient than MCD. However, their approach is different from Vargas (2003) and Jensen et al. (2007) whereby they used RMCD estimators in place of classical estimators in constructing Hotelling  $T^2$  chart for Phase II data. Using the same approach as Chenouri et al. (2009), Alfaro & Ortega (2009) make comparison study for the performance of Hotelling  $T^2$  control chart based on robust estimators of MCD, MVE, RMCD, and trimmed estimator. They concluded their work by recommending the use of  $T^2$  based on trimmed estimator and RMCD when there are few outliers in the production process because of their ability in controlling false alarm rates. However, in the manufacturing of products which emphasizes more on outliers detection than the false alarms generated, the  $T^2$  based on MCD can be considered as better alternatives. This is due to the fact that the Hotelling  $T^2$  control charts based on MCD performed well in terms of probability of outliers detection. Theoretically, if the percentage of outliers' detection increases, the chart should also able to control the overall false alarm rate,  $\alpha$  (Jensen et al., 2007). However the finding in Alfaro and Ortega (2009) shows a conflict between the percentage of outliers detection and the ability of robust control chart in controlling the overall false alarm rate when using robust charts under certain conditions.

In this study, we proposed a robust Hotelling  $T^2$  control chart using a new robust estimator known as minimum vector variance (MVV). The ability of MVV estimator when used in Hotelling  $T^2$  control chart for Phase I has never been tested before. Djauhari et al. (2008) had suggested to use MVV estimator in Hotelling's statistic as a new criterion because MVV estimators have the same breakdown point as MCD estimators and also possess affine equivariant properties. Therefore we are inspired to greater efforts by their suggestion. This study integrates the MVV estimators in the Hotelling  $T^2$  control chart for Phase II data using the same approach as Chenouri et al. (2009) and Alfaro and Ortega (2009) for monitoring the multivariate observations. Even though RMCD was observed to be better than MCD in controlling the false alarm rate, in this study, comparisons are made based on the initial MCD since the algorithm for the proposed method follows the algorithm of the initial MCD. We want to compare the algorithm in its original state and diagnosing problems that might arise using the proposed method (MVV) in constructing Hotelling  $T^2$  control chart. If there is a need to improve the method, this will be continued in the next study. The performance of our proposed control chart was evaluated in the case where there are no changes in the covariance structure. Performance evaluation measured the effectiveness in terms of the probability of outlier(s) detection and false alarm rate (type I error) on Phase II data. It is worthwhile to investigate on the performance using both of them because these measures are closely related (Ramaker et al., 2004). When the data comes from an in-control process the false alarm rate should be close to a nominal value,  $\alpha$ . In this study,  $\alpha$  is set to be equals to 0.05. When data comes from an out-of-control process then the probability of detection should be large to ensure that the chart is able to monitor on-line data and quickly detect shifts in the process of Phase II.

Organization of the remaining part of the paper is as follows. Section 2 discusses about formal definition and the properties of the MVV estimator. In Section 3, we formally introduce a robust control chart based on the MVV estimators. We also demonstrate the Monte Carlo method to estimate the distribution of Hotelling  $T^2$  statistic needed for the computation of control limit and discuss how the performance evaluation is done. The results of the analysis are presented in Section 4. Finally, discussion and conclusion are given in the last section.

## 2. Minimum Vector Variance (MVV) Estimator

MVV and MCD estimators have the same characteristics with respect to breakdown point, affine equivariant properties, and that their algorithms also display the same structures. For details see, Herwindiati (2006), Herwindiati et al. (2007) and Djauhari et al. (2008). The only difference between the two is the way in which the concentration step is generated; MCD uses covariance determinant (CD), while MVV uses vector variance (VV).

By definition, vector variance (VV) is the sum of square of all elements on the diagonal of covariance matrix. If  $X$  is a random vector of  $p$  dimension with  $\Sigma$  as its covariance matrix, then VV of  $X$  is  $T_r(\Sigma^2)$ . Its value indicates the degree of how multivariate distribution is scattered. The larger the value of VV the more scattered the distribution around its mean vector in a subspace of dimension  $q \leq p$ . It is equal to zero if and only if the distribution degenerates at the mean vector. The use of VV instead of CD as multivariate data concentration measures have several advantages (Djauhari, 2007). First, its computation is very efficient even for covariance matrix of large size because VV is a quadratic form while CD is a multilinear form. Thus, the number of operations of VV is smaller than CD since VV is of order  $O(p^2)$  and CD is of order  $O(p^3)$ . Second, vector variance does not depend on non-singularity of the covariance matrix like covariance determinant. The singularity problem usually arises when the number of variable  $p$  is larger than number of sample size  $n$ . Another advantage of VV was illustrated by Djauhari (2007) via comparing the power of vector variance-based test with covariance determinant-based test. In general, both tests have similar performance when  $p$  is small such as  $p = 2$ . However, the power of VV is greater than CD to larger shift of covariance structure when  $p$  and  $n$  are large. The main statistical method used in the estimation of MVV is Mahalanobis squared distances (MSD) which is defined as

$$d_i^2 = (x_i - \mu)^t, \quad i = 1, 2, \dots, n.$$

Consider a data set  $X = \{\bar{X}_1, \bar{X}_2, \dots, \bar{X}_n\}$  of  $p$ -variate observations and let  $H \subseteq X$ . The optimal number of the data involved in the computation of MVV estimator (i.e.  $M_{MUV}$  and  $S_{MUV}$  is  $h = (n + p - 1) / 2$  data points which generate a covariance matrix,  $S_{MUV}$  having minimum  $Tr(S_{MUV}^2)$  among all possible sets of  $h$  data. The location and scatter estimators are given by Equation (1) and (2) respectively as follows,

$$M_{MUV} = \frac{1}{h} \sum_{i \in H} \bar{X}_i \quad (1)$$

$$S_{MUV} = \frac{1}{h-1} \sum_{i \in H} (\bar{X}_i - M_{MUV})(\bar{X}_i - M_{MUV})' \quad (2)$$

To compute the MVV estimators, we propose the new MVV algorithm by combining the C-step from Herwindiati et al. (2007). The C-step is similar to the one in the Fast MCD algorithm, except that the computation of covariance determinant is replaced by the vector variance. The complete algorithm is described as below:

### Stage 1: Creating Initial Subsets.

This step must be repeated 500 times

1. Draw a random subset  $(H_0)$  with number of observations,  $h = p + 1$ . Compute the mean vector  $\bar{x}_{H_0}$  and covariance matrix  $S_{H_0}$ .

$$\bar{x}_{H_0} = \text{average}(H_0) \text{ and } S_{H_0} = \text{cov}(H_0)$$

2. Compute the MSDs  $d_0^2(i) = (x_i - \bar{x}_{H_0})^t S_{H_0}^{-1} (x_i - \bar{x}_{H_0})$  for  $i = 1, \dots, n$ .

3. Sort these MSDs in ascending order,  $d_0^2(\pi(1)) \leq d_0^2(\pi(2)) \leq \dots \leq d_0^2(\pi(n))$ . This ordering defines a permutation  $\pi$  on the index set.

4. Take a new subset  $H_1 = \{\pi(1), \dots, \pi(h)\}$  where  $h = \left\lfloor \frac{n+p-1}{2} \right\rfloor$ , then calculate  $\bar{x}_{H_1}$ ,  $S_{H_1}$ ,  $Tr(S_{H_1}^2)$  and compute MSD, where

$$d_1^2(i) = (x_i - \bar{x}_{H_1})^t S_{H_1}^{-1} (x_i - \bar{x}_{H_1}) \text{ for } i = 1, \dots, n.$$

5. Repeat step 3 and 4 for  $H_2$
6. Sort the 500 values of  $Tr(S_{H_2}^2)$  in ascending order, then select 10 subsets of  $H_2$  which have the lowest  $Tr(S_{H_2}^2)$ . These subsets are treated as the initial subsets and their mean vectors,  $\bar{x}_{H_2}$  and covariance matrices,  $S_{H_2}$  will be used in Stage 2.

### Stage 2: Concentration Steps (C-step)

This process will be repeated until the convergence for each of the 10 subsets is achieved. The convergence is obtained when  $Tr(S_{k-1}^2) = Tr(S_k^2)$ , where  $k$  is the number of iterations.

1. Compute the MSDs by using  $\bar{x}_{H_2}$  and  $S_{H_2}$ , where
- $$d_2^2(i) = (x_i - \bar{x}_{H_2})^t S_{H_2}^{-1} (x_i - \bar{x}_{H_2}) \text{ for } i = 1, \dots, n.$$
2. Repeat step 3 and 4 in Stage 1 until  $Tr(S_{k-1}^2) = Tr(S_k^2)$ . If  $Tr(S_{k-1}^2) > Tr(S_k^2)$  the process is continued. This process will be repeated until convergence is achieved.
  3. When convergence is achieved for all the 10 subsets, choose the subset  $(H^*)$  that generates the lowest  $Tr(S_{H_k}^2)$ . From  $H^*$ , calculate  $\bar{x}_{H^*} = M_{MVV}$  and  $S_{H^*} = S_{MVV}$  as the location and scatter estimators for MVV respectively.

### 3. Proposed Hotelling $T^2$ Control Chart

Suppose that  $x_i = \{x_{i1}, x_{i2}, \dots, x_{in}\}$  is the  $p$ -variate random sample of  $n$  observations of preliminary data set in Phase I. Assume that  $x_i$  are independent and follow a multivariate normal distribution with mean vector  $\mu$  and covariance matrix  $\Sigma$ . If  $\mu$  and  $\Sigma$  are unknown then we need to estimate them using an in-control data set. The process of identifying the in-control data set from  $x_i$  is referred to as Phase I operation.

From the preliminary data set, compute  $\bar{x}$  and  $S$ . Using this estimates, compute  $T^2(i)$  for  $i = 1, \dots, n$ , such that

$$T^2(i) = (x_i - \bar{x})S^{-1}(x_i - \bar{x})^T \quad (3)$$

To get in-control data set, define outliers by using UCL based on Beta distribution as follows,

$$UCL_1 \sim \left[ \frac{(n-1)^2}{n} \right] B_{\left( \frac{p}{2}, \frac{n-p-1}{2} \right)} \quad (4)$$

Remove the observations which are considered as outliers when  $T^2(i) > UCL_1$ . The sample with the outliers removed ( $n_c$ ) is then used to compute the new estimates,  $\bar{x}_N$  and  $S_N$ . Using these estimates, compute  $T^2(g)$  statistic for Phase II observation, where  $x_g \notin x_i$ , such that

$$T^2(g) = (x_g - \bar{x}_N)S_N^{-1}(x_g - \bar{x}_N)^T \quad (5)$$

By using the desired  $\alpha, p$  and  $n_c$  values, compute the upper control limit using  $F$  distribution such that

$$UCL \sim \left[ \frac{p(n_c+1)(n_c-1)}{n_c(n_c-p)} \right] F_{(p, n_c-p)} \quad (6)$$

However, this standard approach is only effective in eliminating very extreme outliers and detecting large shift in the mean vector in small sample sizes, but it fails to detect more moderate outliers especially when number of variables increased (Vargas, 2003; Jensen et al., 2007; Chenouri et al., 2009). To alleviate the problem of this procedure, we proposed using MVV estimator in Phase I data,  $x_i$ . Since the estimator is known to be free from outliers due to its estimation process, they could be readily used as in-control estimators in Phase II. Let  $x_g = \{x_{n+1}, x_{n+2}, \dots\}$  where  $x_g \notin x_i$  and  $M_{MVV}$  and  $S_{MVV}$  represent the MVV mean vector and covariance matrix estimators, respectively. We define a robust Hotelling's  $T^2$  for Phase II data,  $x_g$  based on these MVV estimates by

$$T_{MVV}^2(g) = (x_g - M_{MVV})S_{MVV}^{-1}(x_g - M_{MVV})^T \quad (7)$$

### 3.1 Estimation of control limits for the proposed Hotelling $T^2$ control chart

The application of robust estimators in place of the mean and covariance structure in  $T^2$  chart (5) will cause the distributional properties of the traditional  $T^2(g)$  to change (Williams et al., 2006). To demonstrate the performance of  $T_{MVV}^2(g)$  in equation (7), we need a better understanding about its distribution in order to obtain appropriate control limits i.e. UCL. Since the distribution of  $T_{MVV}^2$  is unknown, we apply Monte Carlo simulations to estimate the quantiles of the  $T_{MVV}^2(g)$ , for several combinations of sample sizes and dimensions (refer to Table 1). Even, the finite sample distribution of the MCD estimators is still questionable for applications, thus the distribution of  $T_{MCD}^2(g)$  is also unknown (Vargas, 2003; Jensen et al., 2007; Chenouri et al., 2009; Alfaro & Ortega, 2009). For that reason, we also use quantile in estimating the distribution of  $T_{MCD}^2(g)$  obtained via Monte Carlo method.

This study focuses on multiple dimensions with reasonable values of sample size  $n$ . The range of values are listed in Table 1. To ensure a smooth computation for MCD, the setting of the starting value for  $n$  follows the rule of thumb, where  $\frac{n}{p} > 5$  (Rousseeuw & van Zomeren, 1990).

In order to estimate the 95% quantile of  $T_{MVV}^2(g)$  for a given Phase I sample size  $n$  and dimension  $p$ , we generate  $K = 5000$  samples of size  $n$  from a standard multivariate normal distribution  $MVN_p(0, I_p)$ . For each data set of size  $n$ , we compute the MVV mean vector and covariance matrix estimates,  $M_{MVV}(k)$  and  $S_{MVV}(k)$ ,  $k = 1, \dots, K$ . In addition, for each data set, we randomly generate a new observation  $x_{g,k}$  treated as a Phase II observation from  $MVN_p(0, I_p)$  and calculate the corresponding  $T_{MVV}^2(g, k)$  values as given by equation (7). The empirical distribution function of  $T_{MVV}^2(g)$  is based on the simulated values

$$T_{MVV}^2(g, 1), T_{MVV}^2(g, 2), \dots, T_{MVV}^2(g, K) \quad (8)$$

We sort  $T_{MVV}^2(g, k)$  values in ascending order, and the UCL is the 95% quantile of the 5000 statistics.

### 3.2 Performance evaluation.

In order to assess the performance of  $T_{MUV(g)}^2$  control charts, various conditions were created by manipulating number of observations ( $n$ ), number of dimensions ( $p$ ), and level of contamination by using different proportion of outliers ( $\varepsilon$ ) and several mean shifts values ( $\mu_1$ ). The performance of the robust Hotelling  $T^2$  charts is judged based on the false alarm rate and probability of detection in the process behavior from Phase II data. Thus, for Phase II observations, we simulate 1000 new datasets of different sample sizes ( $n$ ) and dimensions ( $p$ ) in Table 1. To determine the false alarm rate and probability of detection, we randomly generate a Phase II observation with in-control and out of control parameters respectively from Phase I, and calculate the proposed robust Hotelling  $T^2$  statistics. The false alarm rate or probability of detection is estimated as the proportion of statistic values that are above the control limits of 1000 replications.

For Phase I, we simulate data of various conditions created by manipulating the number of observations, dimensions and levels of contamination. To examine the effect of contamination on the charts' performance, we have considered a contaminated model by using a mixture of normal suggested by Alfaro & Ortega (2009)

$$(1-\varepsilon)N_p(\mu_0, \sum_0) + \varepsilon N_p(\mu_1, \sum_1) \quad (9)$$

where  $\varepsilon$  is the proportion of outliers,  $\mu_0$  and  $\sum_0$  are the in-control parameters while  $\mu_1$  and  $\sum_1$  are the out-of-control parameters. In this study we assume contamination with shift in the mean but no changes in covariance structure, therefore, the covariance matrix  $\sum_0$  and  $\sum_1$  in Equation (9) represent the identity matrix of  $p$  dimensions ( $I_p$ ). To check on these conditions, we consider  $\varepsilon$  to be 0, 0.1 or 0.2. While for the probability of detecting a change which depends on the shift in the mean vector, we set  $\mu_1$  to be a vector of size  $p$  with value of 0 (when there is no change), 3 or 5 (which is a good leverage point). Manipulation on the mean shifts and percentage of outliers generate 5 different types of contaminated distributions as listed below which were categorized as ideal, mildly contaminated, moderately contaminated and extremely contaminated as follows,

- 1)  $N_p(0, I_p)$  -Ideal (no contamination)
- 2)  $(0.9)N_p(0, I_p) + (0.1)N_p(3, I_p)$  -mild contamination
- 3)  $(0.8)N_p(0, I_p) + (0.2)N_p(3, I_p)$  -Moderate contamination
- 4)  $(0.9)N_p(0, I_p) + (0.1)N_p(5, I_p)$  -Moderate contamination
- 5)  $(0.8)N_p(0, I_p) + (0.2)N_p(5, I_p)$  -Extreme contamination

Each of these model was paired with different combinations of sample sizes,  $n$ , and number of dimensions,  $p$  (refer to Table 1) to create various conditions which are capable of highlighting the strengths and weaknesses of the charts. Next, in Phase II, we simulate data from multivariate normal distribution  $MCN_p(\mu_1, I_p)$ , where  $\mu_1$  is the shift in the mean vector with values similarly assigned to Phase I (i.e. 0, 3, and 5). The performance of the proposed chart is then compared with robust Hotelling  $T^2$  chart using  $MCD$  ( $T_{MCD}^2$ ) and the traditional Hotelling  $T^2$  control charts. For the traditional chart we employed two approaches; first approach denoted as  $T_0^2$  is without cleaning the outliers as being adopted by Alfaro and Ortega (2009) and the second approach, which is known as the standard approach, cleans the outliers once ( $T_s^2$ ). Each of these charts was tested on 5 types of contaminations on 23 combinations of  $n$  and  $p$  which totalled up to 115 conditions. For each condition, the false alarm rates and probability of detection were determined. The programs and simulations were run using MATLAB 7.8.0 (R2009a). The algorithm of MVV was executed using the MATLAB 7.8.0 (R2009a), while Fast MCD algorithm using *mc cov.m* in the LIBRA package under MATLAB 7.8.0 (R2009a).

### 4. Simulation Results

In this section, we compared the performance of  $T_{MVV}^2$ ,  $T_{MCD}^2$ ,  $T_0^2$  and  $T_s^2$  control chart in terms of probability of detection and false alarm rate. The results of the investigation are presented in figures and tables for the probability of detection and false alarm rates respectively.

#### 4.1 Probability of detection of outliers

The graphs illustrating the performance of the four charts in terms of probability of detection are exhibited in Figure 1-5. Each figure represents different dimension ( $p$ ). For each condition, the performance of the control chart is regarded as better in detecting changes when the value of the probability is closer to 1. Under bivariate case ( $p = 2$ ) as presented in Figure 1, initially  $T_s^2$  showed better detection than other charts at mild and moderate contamination. However, the good performance of  $T_s^2$  only sustain at  $n = 10, 25$ . Once the value of  $n$  and  $p$  increased, which can be clearly observed in Figure 2-5, the line representing  $T_{MVV}^2$  consistently at the



highest location in the graphs with the probability value of approximately 1, and overlapping with  $T_{MCD}^2$  line under most of the conditions. There are instances when the  $T_{MCD}^2$  line started with lower values creating gaps between the two lines but merged later on when the  $n$  values increased. This situation occurs when the sample size is small with 20% outliers and mean shift 3.

Overall, the  $T_{MVV}^2$  and  $T_{MCD}^2$  control charts consistently achieved high probability in detecting outliers. One can observe that, the lines representing  $T_0^2$  and  $T_S^2$  charts are always at the lowest and second lowest respectively, creating a very wide gap between the other two lines ( $T_{MVV}^2$  and  $T_{MCD}^2$ ). This pattern repeats even within the same dimension for  $p > 5$ . Result on the  $T_S^2$  chart reveals that the chart performs so well when the number of outliers is small (small  $p$  and low percentage of contamination), but underperforms when the number of outliers gets larger (large  $p$  and high percentage of contamination). This weakness can be mitigated by the use of robust Hotelling  $T^2$  chart.

#### 4.2 False alarm rates

The performance of a chart is not purely judged by its ability in detecting outliers, but also in controlling the false alarm rate. False alarm rate is the probability of out-of-control signal when a process is in control. The value becomes large if the process is unstable due to increase in variability. Inflated false alarm rate can lead to unnecessary process adjustments and loss of confidence in the control chart as a monitoring tool (Chang & Bai, 2004). Hence, a method which can control the false alarm rate to the desired level is necessary.

Table 2-6 which recorded the false alarm rates for each condition are arranged based on the ascending number of dimensions (variables) namely  $p = 2, 5, 10, 15$  and 20 with  $\alpha = 0.05$ . The first column in each table displays the number of sample sizes, followed by the percentage of outliers and non centrality values respectively in the second and third column. The last four columns record the false alarm rates of the control charts investigated in this study; namely  $T_0^2$ ,  $T_S^2$ ,  $T_{MCD}^2$  and  $T_{MVV}^2$ . The control chart is considered to be in control of its false alarm if the empirical value is close to the nominal value,  $\alpha$ .

For the bivariate ( $p = 2$ ) case presented in Table 2, the overall results on false alarm rates show that  $T_{MVV}^2$  outperforms the other control charts in especially when the sample size is very small ( $n = 10$ ). Even though the results for and  $T_S^2$  under most conditions are well controlled, however under ideal condition (no contamination) the chart failed to control the false alarm, causing the rate to inflate to 0.1000. The  $T_{MCD}^2$  and  $T_0^2$  control charts are badly affected when the sample size is very small, which are verified by the rates of false alarm which are far below the nominal value except for ideal condition. When the percentage of outliers increased to 20% we observed that the rates for  $T_{MVV}^2$ ,  $T_{MCD}^2$  and  $T_0^2$  charts dwindle as the sample size increased, but the  $T_S^2$  chart is still in control of its false alarm. The performance of the robust  $T_{MVV}^2$  chart is much better than the  $T_{MCD}^2$ . The  $T_{MVV}^2$  chart performs so well in controlling false alarm rates except when the percentage of outliers is large. When the dimension increased to  $p = 5$ , the rates of false alarm for both traditional charts ( $T_0^2$  and  $T_S^2$ ) improved. Refer to Table 3. Nevertheless, the rates for  $T_S^2$  chart under ideal condition are still high (very far above the nominal value). We also notice improvements in the robust  $T_{MVV}^2$  charts especially when the percentage of outliers is large. In contrast, the false alarm rates for  $T_{MCD}^2$  chart worsen with values as small as 0.0020.

Table 4 displays the false alarm rates for the case of  $p = 10$ . While there are noticeable improvements in most of the conditions for  $T_0^2$ ,  $T_{MVV}^2$  and  $T_{MCD}^2$  charts, most of the rates for  $T_S^2$  seem to be deviating away from the nominal value. Even under the influence of extreme contamination, the rates of  $T_{MVV}^2$  chart are no less than 0.022 value. However, the rates for  $T_{MCD}^2$  chart are still far below the nominal level despite the improvement. Under the case of  $p = 15$ , as can be clearly observed in Table 5, all the charts show better results than the previous case. Great improvement could be detected in  $T_S^2$  chart under ideal condition and  $T_{MVV}^2$  under extreme contamination, but  $T_{MCD}^2$  chart is still unable to control its false alarm rates under the latter condition. As we scrutinized the false alarm rates for  $p = 20$  in Table 6, we discover sporadic improvements under different conditions. There is no obvious improvement pattern could be observed. However, we can clearly observe that  $T_{MCD}^2$  chart performs badly in controlling false alarm rate in all conditions.

#### 4.3 Real Data Analysis

To illustrate the Hotelling  $T^2$ , a company which involved in the production of advanced for the aircraft industry has provided us with data on spoilers. Spoilers are vital devices in an airplane. Their function is to increase lifts when the airplane is flying. The products are used in civilian, defence, and space applications, which cannot

compromise any mistakes, albeit a minor one. Thus, careful monitoring is required to ensure that no variation occur in the process. Any slight mistake could jeopardize a human life.

For the purpose of this study, a sample of 47 products ( $n = 47$ ) which consists of several features such as trim edge ( $X_1$ ), trim edge spar ( $X_2$ ), and drill hole ( $X_3$ ) was furnished to us by the company. Out of the total, 21 products were collected from 2009, while the rest were from 2010. Hence, we decided to use the 2009 products as Phase I historical data, and considered the products from 2010 as future data in this study. The details of the historical and future data are displayed in Table 7 and 9 respectively. The products consist of 3 quality variables (dimensions) namely trim edge, trim edge spar, and drill hole. Estimates for the location vector ( $\bar{X}$ ) and scatter matrix ( $S$ ) are presented in Table 8. The calculation of the upper control limits (UCLs) based on the estimates are

presented in the last column of the table. The values of the  $T^2$  statistics based on the above estimators appear in the last four columns of Table 9. The graphical presentation of the corresponding control charts are put on view in Figure 6.

When comparing the values of the  $T^2$  statistics in Table 9 with the corresponding control limits in Table 8, we observe that the three statistics  $T_{MVV}^2$ ,  $T_{MCD}^2$  and  $T_s^2$  signal observations 20, 22 and 25 as out-of-control but  $T_0^2$  only signals 20 and 25 as out-of-control observations and fails to signal observation 22. The result for  $T_0^2$  is as expected since the analysis on the probability of detection using simulated data showed that  $T_0^2$  was not as effective as the other charts in detecting outliers. Chart (a), (b), (c) and (d) in Figure 6 represent the control chart for  $T_0^2$ ,  $T_s^2$ ,  $T_{MCD}^2$  and  $T_{MVV}^2$  respectively. Even though the performance of  $T_s^2$  chart in this example is on par with the proposed  $T_{MVV}^2$  chart and also  $T_{MCD}^2$  chart, but the outcome could be due to the small number of quality characteristics (dimension) of the product. As revealed in the simulation study,  $T_s^2$  performed well in detecting outliers under low dimension (not more than 5) only, but underperformed when the dimension increased to above 5.

## 5. Conclusion

Hotelling  $T^2$  chart is well accepted as a reliable method to monitor production; however, under conditions of non normality, this chart is known to be underperformed. Alternative on the Hotelling  $T^2$  statistic particularly on the location and scatter measures are recommended in order to produce a reliable chart regardless of the conditions. This study proposed an alternative to the the Hotelling  $T^2$  chart by using a robust estimator known as minimum variance vector (MVV) for its location and scatter measures. MVV not only has all the properties of the well known minimum covariance determinant (MCD) such as high breakdown point and affine equivariant, but also has better computational efficiency. The performance of our proposed robust Hotelling  $T^2$  chart using MVV in terms of false alarm rate and probability of detection were compared with the robust Hotelling  $T^2$  chart using MCD ( $T_{MCD}^2$ ) and the traditional Hotelling  $T^2$  charts ( $T_0^2$  and  $T_s^2$ ).

Investigation on the  $T_0^2$  and  $T_{MCD}^2$  by Alfaro and Ortega (2009) showed a conflicting result between the percentage of outliers detection and the overall false alarm rate such that when the probability of detection increased, the false alarm rates inflate away from the nominal value. However, our proposed chart,  $T_{MVV}^2$  perform so well in terms of detecting outliers and also in controlling false alarm rates. Even though the traditional Hotelling  $T_s^2$  chart performs so well in terms of controlling false alarm rates, but this chart fails to achieve good probability of detection especially when the number of quality characteristics is large. On contrary, the Hotelling  $T_{MCD}^2$  chart performs wonderfully in detecting outliers, however the chart fails terribly in controlling false alarm rates. With its good performance in terms of detecting outliers and controlling false alarm rates, plus the good properties of its statistics, Hotelling  $T_{MVV}^2$  chart is indeed a good alternative to the multivariate control chart.

## References

- Angiulli, F. & Pizzuti, C. (2005). Outlier mining and large high-dimensional data sets. *IEEE Transaction Knowledge Data Engineering*, 17(2), 203–215.
- Alfaro, J. L. & Ortega, J. Fco. (2009). A comparison of robust alternatives to Hotelling's  $T^2$  control chart. *Journal of Applied Statistics*, 36(12), 1385–1396, doi:10.1080/02664760902810813, <http://dx.doi.org/10.1080/02664760902810813>
- Alt F.B. (1985). Multivariate quality control. In S.Kotz and N.L. Johnson NL (Eds.). *Encyclopedia of Statistical Sciences*, (6), New York:Wiley.
- Beckman, R., & Cook, R. (1983). Outlier.....s. *Technometrics*, 25, 119–149.



- Butler, R.W., Davies, P.L., & Jhun, M. (1993). Asymptotics for the minimum covariance determinant estimator, *The Annals of Statistics*, 21, 1385-1400.
- Chang, S.C. & Bai, D.S. (2004). A multivariate  $T^2$  control chart for skewed populations using weighted standard deviations. *Qual. Reliab. Engng. Int.*, 20, 31-46.
- Chenouri, S., Steiner, S.H. & Mulayath, A. (2009). A multivariate robust control chart for individual observations, *Journal of Quality Technology*, 41(3), 259-271.
- Croux, C. & Haesbroeck, G. (1999). Influence function and efficiency of the minimum covariance determinant scatter matrix estimator, *Journal of Multivariate Analysis* 71, 161-190.
- Djauhari, M.A., Mashuri, M. & Herwindiati, D. E. (2008). Multivariate process variability monitoring, *Communications in Statistics - Theory and Methods*, 37(11), 1742 — 1754, doi: 10.1080/03610920701826286, <http://dx.doi.org/10.1080/03610920701826286>
- Djauhari, M.A. (2007). A measure of multivariate data concentration. *Journal of Applied Probability and Statistics*, 2, 139-155.
- Fauconnier, C. & Haesbroeck, G. (2009). Outliers detection with the minimum covariance determinant estimator in practice. *Statistical Methodology*, 6(4), 363-379, doi:10.1016/j.stamet.2008.12.005, <http://dx.doi.org/10.1016/j.stamet.2008.12.005>
- Hadi, A.S. (1992). Identifying multivariate outlier in multivariate data. *Journal of Royal Statistical Society B*, 53, 761-771.
- Hampel F.R., Ronchetti E.M., Rousseeuw P.J. and Stahel W.A. (1986). *Robust Statistics. The Approach Based on Influence Functions*. New York: John Wiley & Sons.
- Herwindiati, D.E. (2006). A new criterion in robust estimation for location and covariance matrix, and its application for outlier labeling. Unpublished PhD's thesis, Institut Teknologi Bandung, Bandung.
- Herwindiati, D.E., Djauhari, M.A., & Mashuri, M. (2007). Robust Multivariate Outlier Labeling. *Communication in Statistics-Computation and Simulation*, 36: 1287-1294, doi: 10.1080/03610910701569044, <http://dx.doi.org/10.1080/03610910701569044>
- Holmes, D. S., & Mergen, A. E. (1993). Improving the Performance of the  $T^2$  Control Chart. *Quality Engineering*, 5, 619-625.
- Hubert, M., Rousseeuw, P.J., & van Aelst, S. (2005). Multivariate outlier detection and robustness. *Handbook of Statistics*, 24, Elsevier B.V., 263-302.
- Jackson, J.E. (1985). Multivariate quality control. *Communications in statistics: Theory and Methods*, 11, 2657-2688
- Jensen, W.A., Birch, J.B. & Woodall, W.H. (2007). High breakdown estimation methods for Phase I multivariate control charts, *Qual. Reliab. Eng. Int.* 23, pp. 615-629. doi: 10.1002/qre.837, <http://dx.doi.org/10.1002/qre.837>
- Lopuhaa, H.P., & Rousseeuw, P.J. (1991). Breakdown points of affine equivariance estimators of multivariate location and covariance matrices. *Annal of Statistics*, 19, 229-248.
- Lowry, C.A. and Montgomery, D.C. (1995). A review of multivariate control charts. *IIE Transactions*, 27, 800-810.
- Maronna, R.A., Martin, R.D. & Yohai, V.J.(2006). *Robust Statistics: Theory and Methods*. New York: John Wiley & Sons.
- Maronna, R.A. & Zamar, R. (2002). Robust estimation of location and dispersion for high-dimensional datasets. *Technometrics*, 44, 307-317.
- Mason, R. L., & Young, J. C. (2002). *Multivariate Statistical Process Control with Industrial Applications*. Philadelphia: ASA-SIAM.
- Ramaker, Henk-Jan., van Sprang, Eric N.M., Westerhuis, J. A. & Smilde, A. K. (2004). The effect of the size of the training set and number of principal components on the false alarm rate in statistical process monitoring. *Chemometrics and Intelligent Laboratory Systems*, 73, 181- 187. doi:10.1016/j.chemolab.2003.12.015, <http://dx.doi.org/10.1016/j.chemolab.2003.12.015>
- Rousseeuw, P.J. (1985). Multivariate estimation with high breakdown point. In W.Grossman, G. Pflug, I. Vincze & W.Wertz (Eds), *Mathematical Statistics and Applications*, B, (p. 283-297). D. Reidel Publishing Company.
- Rousseeuw, P. J. & Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. New York: John Wiley & Sons.
- Rousseeuw, P.J. & van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41, 212-223.
- Rousseeuw, P.J. & van Zomeren, B.C. (1990). Unmasking multivariate outliers and leverage points. *Journal of American Statistical Association*, 85(411), 633-639.
- Sullivan, J.H. & Woodall, W.H. (1996). A comparison of multivariate control charts for individual measurements. *Journal of Quality Technology*, 28 (4), 398-408.
- Sullivan, J.H. & Woodall, W.H. (1998). Adapting control charts for the preliminary analysis of multivariate observations, *Commun. Stat. Simulation Comput.*, 27, 953-979.

- Tracy, N. D., Young, J. C. & Mason, R. L. (1992). Multivariate control charts for individual observations. *Journal of Quality Technology*, 24, 88-95.
- Vargas, J.A. (2003). Robust estimation in multivariate control charts for individual observation, *Journal of Quality Technology*, (35), 367-376.
- Werner, M. (2003). Identification of Multivariate Outliers in Large Data Sets. Unpublished PhD's thesis, University of Colorado, Denver.
- Wierda, S.J. (1994). Multivariate statistical process control-Recent results and directions for future research, *Statistica Neerlandica*, 48, 147-168.
- Williams, J.D. (2004). Contributions to Profile Monitoring and Multivariate Statistical Process Control, Unpublished PhD's dissertation, Virginia Polytechnic Institute & State University.
- Williams, J. D., Woodall, W. H., Birch, J. B. & Sullivan, J. H. (2006). Distribution of Hotelling's  $T^2$  statistic based on the successive differences estimator, *Journal of Quality Technology*, 38(3), 217-229.
- Woodall, W. H. and Montgomery, D. C. (1999). Research issues and ideas in statistical process control. *Journal of Quality Technology*, 31(4), 376-386.

Table 1. The values of  $n$  and  $p$ 

$p$	$n$
2	10, 25, 50, 100, 200, 500
5	30, 50, 100, 200, 500
10	50, 100, 200, 500
15	80, 100, 200, 500
20	100, 200, 300, 500

Table 2. False alarm rates of the corresponding control charts with dimension,  $p = 2$ 

Sample Size (n)	% outliers ( $\epsilon$ )	Mean shift ( $\mu_1$ )	Control Charts			
			$T^2_0$	$T^2_S$	$T^2_{HCD}$	$T^2_{MVR}$
10	0	0	0.0530	0.1000	0.0520	0.0520
	10%	3	0.0170	0.0650	0.0290	0.0450
		5	0.0160	0.0630	0.0250	0.0450
		3	0.0180	0.0540	0.0210	0.0330
	20%	5	0.0180	0.0480	0.0110	0.0330
25	0	0	0.0590	0.0980	0.0480	0.0530
	10%	3	0.0290	0.0600	0.0280	0.0390
		5	0.0230	0.0670	0.0290	0.0390
		3	0.0280	0.0470	0.0090	0.0190
	20%	5	0.0240	0.0390	0.0050	0.0190
50	0	0	0.0560	0.0920	0.0580	0.0540
	10%	3	0.0200	0.0480	0.0230	0.0350
		5	0.0160	0.0490	0.0230	0.0340
		3	0.0210	0.0370	0.0080	0.0180
	20%	5	0.0160	0.0340	0.0060	0.0170
100	0	0	0.0550	0.0930	0.0460	0.0490
	10%	3	0.0210	0.0470	0.0200	0.0300
		5	0.0160	0.0490	0.0200	0.0290
		3	0.0210	0.0350	0.0050	0.0150
	20%	5	0.0160	0.0350	0.0040	0.0150
200	0	0	0.0580	0.0950	0.0600	0.0690
	10%	3	0.0210	0.0510	0.0310	0.0490
		5	0.0180	0.0470	0.0310	0.0500
		3	0.0200	0.0410	0.0050	0.0280
	20%	5	0.0180	0.0360	0.0020	0.0280
500	0	0	0.0500	0.0880	0.0520	0.0630
	10%	3	0.0190	0.0480	0.0270	0.0490
		5	0.0160	0.0390	0.0260	0.0480
		3	0.0170	0.0370	0.0040	0.0230
	20%	5	0.0160	0.0350	0.0040	0.0230

Table 3. False alarm rates for independent case with dimension,  $p = 5$ 

Sample Size (n)	% outliers ( $\epsilon$ )	Mean shift ( $\mu_1$ )	Control Charts			
			$T^2_0$	$T^2_5$	$T^2_{WCD}$	$T^2_{MVF}$
30	0	0	0.0460	0.0790	0.0430	0.0500
	10%	3	0.0280	0.0510	0.0100	0.0300
		5	0.0260	0.0530	0.0100	0.0330
	20%	3	0.0300	0.0620	0.0050	0.0210
		5	0.0320	0.0620	0.0000	0.0200
50	0	0	0.0530	0.0790	0.0650	0.0490
	10%	3	0.0270	0.0480	0.0130	0.0350
		5	0.0260	0.0460	0.0130	0.0370
	20%	3	0.0260	0.0560	0.0040	0.0220
		5	0.0250	0.0520	0.0020	0.0230
100	0	0	0.0540	0.0740	0.0320	0.0380
	10%	3	0.0290	0.0510	0.0140	0.0300
		5	0.0280	0.0420	0.0140	0.0320
	20%	3	0.0300	0.0520	0.0020	0.0170
		5	0.0290	0.0490	0.0020	0.0190
200	0	0	0.0430	0.0740	0.0410	0.0390
	10%	3	0.0250	0.0450	0.0200	0.0350
		5	0.0240	0.0420	0.0200	0.0350
	20%	3	0.0270	0.0460	0.0010	0.0220
		5	0.0270	0.0440	0.0010	0.0220
500	0	0	0.0390	0.0620	0.0420	0.0430
	10%	3	0.0200	0.0410	0.0160	0.0360
		5	0.0190	0.0350	0.0170	0.0370
	20%	3	0.0210	0.0430	0.0030	0.0190
		5	0.0200	0.0420	0.0030	0.0190

Table 4. False alarm rates for dimension,  $p = 10$ 

Sample Size (n)	% outliers ( $\epsilon$ )	Mean shift ( $\mu_1$ )	Control Charts			
			$T^2_0$	$T^2_5$	$T^2_{WCD}$	$T^2_{MVF}$
50	0	0	0.0570	0.0920	0.0530	0.0520
	10%	3	0.0410	0.0690	0.0210	0.0370
		5	0.0380	0.0670	0.0210	0.0380
	20%	3	0.0420	0.0720	0.0080	0.0250
		5	0.0410	0.0720	0.0020	0.0220
100	0	0	0.0550	0.0780	0.0420	0.0450
	10%	3	0.0330	0.0570	0.0190	0.0390
		5	0.0340	0.0550	0.0200	0.0350
	20%	3	0.0350	0.0560	0.0030	0.0240
		5	0.0340	0.0520	0.0030	0.0230
200	0	0	0.0430	0.0730	0.0540	0.0520
	10%	3	0.0330	0.0530	0.0200	0.0390
		5	0.0320	0.0520	0.0200	0.0420
	20%	3	0.0340	0.0500	0.0020	0.0250
		5	0.0340	0.0490	0.0020	0.0240
500	0	0	0.0510	0.0750	0.0490	0.0540
	10%	3	0.0330	0.0540	0.0220	0.0390
		5	0.0330	0.0520	0.0230	0.0390
	20%	3	0.0340	0.0580	0.0040	0.0260
		5	0.0340	0.0550	0.0040	0.0230

Table 5. False Alarm rate for dimension,  $p = 15$ 

Sample Size (n)	% outliers ( $\epsilon$ )	Mean shift ( $\mu_1$ )	Control Charts			
			$T^2_{\bar{X}}$	$T^2_S$	$T^2_{MCD}$	$T^2_{MVF}$
80	0	0	0.0450	0.0650	0.0580	0.0560
	10%	3	0.0330	0.0550	0.0260	0.0470
		5	0.0330	0.0560	0.0230	0.0430
	20%	3	0.0350	0.0560	0.0060	0.0270
		5	0.0360	0.0540	0.0030	0.0320
100	0	0	0.0430	0.0680	0.0490	0.0520
	10%	3	0.0330	0.0590	0.0240	0.0450
		5	0.0330	0.0610	0.0240	0.0430
	20%	3	0.0330	0.0540	0.0030	0.0250
		5	0.0330	0.0560	0.0020	0.0220
200	0	0	0.0440	0.0620	0.0540	0.0470
	10%	3	0.0290	0.0520	0.0330	0.0420
		5	0.0280	0.0520	0.0310	0.0410
	20%	3	0.0310	0.0560	0.0040	0.0200
		5	0.0300	0.0550	0.0040	0.0240
500	0	0	0.0530	0.0690	0.0460	0.0470
	10%	3	0.0370	0.0540	0.0270	0.0390
		5	0.0370	0.0530	0.0260	0.0390
	20%	3	0.0390	0.0530	0.0060	0.0260
		5	0.0380	0.0520	0.0060	0.0290

Table 6. False Alarm rate for dimension,  $p = 20$ 

Sample Size (n)	% outliers ( $\epsilon$ )	Mean shift ( $\mu_1$ )	Control Charts			
			$T^2_{\bar{X}}$	$T^2_S$	$T^2_{MCD}$	$T^2_{MVF}$
100	0	0	0.0410	0.0720	0.0500	0.0530
	10%	3	0.0340	0.0550	0.0210	0.0400
		5	0.0330	0.0560	0.0240	0.0420
	20%	3	0.0310	0.0570	0.0060	0.0300
		5	0.0300	0.0570	0.0000	0.0240
200	0	0	0.0450	0.0660	0.0490	0.0510
	10%	3	0.0320	0.0510	0.0220	0.0370
		5	0.0340	0.0520	0.0220	0.0380
	20%	3	0.0360	0.0550	0.0050	0.0310
		5	0.0380	0.0550	0.0020	0.0250
300	0	0	0.0430	0.0680	0.0390	0.0440
	10%	3	0.0400	0.0480	0.0210	0.0350
		5	0.0390	0.0490	0.0210	0.0340
	20%	3	0.0410	0.0470	0.0060	0.0240
		5	0.0420	0.0470	0.0050	0.0220
500	0	0	0.0520	0.0690	0.0560	0.0530
	10%	3	0.0390	0.0560	0.0280	0.0400
		5	0.0380	0.0570	0.0280	0.0400
	20%	3	0.0390	0.0550	0.0030	0.0350
		5	0.0390	0.0570	0.0040	0.0320

Table 7. Historical data set (Phase I data)

Product No.	Trim edge (x1)	Trim edge spar (x2)	Drill hole (x3)
1	-0.0011	0.0003	0.0128
2	0.0011	0.0021	0.0246
3	0.0252	0.0308	0.0378
4	-0.0017	0.0109	0.0177
5	-0.0005	-0.0010	0.0106
6	0.0016	-0.0059	0.0128
7	0.0004	0.0001	0.0062
8	0.0078	0.0003	0.0159
9	0.0076	0.0089	0.0097
10	0.0020	0.0005	0.0071
11	0.0108	0.0011	0.0092
12	0.0039	0.0034	0.0425
13	0.0060	-0.0033	0.0160
14	0.0066	0.0100	0.0056
15	0.0045	-0.0067	0.0147
16	0.0110	-0.0207	0.0337
17	0.0047	0.0059	0.0065
18	0.0077	0.0003	0.0191
19	0.0015	0.0123	0.0124
20	0.0011	0.0038	0.0104
21	0.0056	0.0065	0.0063

Table 8. Estimates of location vector, covariance matrix, and UCL.

Types of Control Chart	Location Vector ( $\bar{\bar{X}}$ )	Scatter Matrix (S)	Upper Control Limit (UCL)
$T^2_0$	[0.00504 0.00284 0.01579]	$\begin{bmatrix} 0.000004 & 0.000002 & 0.000003 \\ 0.000002 & 0.000009 & 0.000001 \\ 0.000003 & 0.000001 & 0.000011 \end{bmatrix}$	11.035
$T^2_3$	[0.00365 0.00256 0.01209]	$\begin{bmatrix} 0.000001 & 0.000000 & 0.000000 \\ 0.000000 & 0.000003 & -0.000001 \\ 0.000000 & -0.000001 & 0.000003 \end{bmatrix}$	11.798
$T^2_{NCD}$	[0.00414 0.00207 0.01096]	$\begin{bmatrix} 0.000002 & 0.000000 & 0.000000 \\ 0.000002 & 0.000009 & -0.000002 \\ 0.000000 & -0.000002 & 0.000003 \end{bmatrix}$	21.946
$T^2_{MVF}$	0.00336 0.00354 0.00913	$\begin{bmatrix} 0.000001 & 0.000001 & 0.000000 \\ 0.000001 & 0.000003 & 0.000000 \\ 0.000000 & 0.000000 & 0.000001 \end{bmatrix}$	41.298



Table 9. The Hotelling  $T^2$  values for the future (Phase II ) data

Product No.	x1	x2	x3	$T^2$	$T^2$	$T^2_{\text{theo}}$	$T^2_{\text{theo}}$
1	0.0041	0.0087	0.0129	0.55822	1.42418	1.76591	4.39082
2	0.0047	0.0109	0.0124	0.90026	2.54922	2.46944	5.16947
3	0.0031	0.0057	0.0096	0.49916	0.49356	0.34367	0.29917
4	0.0035	-0.0020	0.0101	0.54633	1.01572	0.54563	1.50640
5	0.0040	-0.0028	0.0125	0.45922	0.95881	0.45797	3.78687
6	0.0031	0.0008	0.0061	0.90130	1.74802	1.25274	2.24213
7	-0.0019	0.0101	0.0112	3.09329	4.13719	4.44043	6.53612
8	0.0009	0.0039	0.0082	0.80608	1.28839	0.68370	1.05554
9	-0.0052	0.0090	0.0203	7.36021	9.68427	14.97663	26.04990
10	-0.0008	0.0110	0.0184	3.61976	5.80349	9.74168	19.17603
11	-0.0021	0.0139	0.0170	5.38392	8.08967	11.87166	19.63128
12	-0.0017	0.0092	0.0061	2.73870	4.79492	2.97882	8.13879
13	-0.0010	0.0133	0.0138	3.80577	5.68902	7.40398	11.38954
14	-0.0030	0.0002	0.0053	2.05480	6.34679	3.30863	9.14983
15	0.0016	0.0134	0.0151	2.50731	5.02274	6.80538	12.38812
16	0.0027	0.0086	0.0070	1.19755	1.89797	1.06789	2.05633
17	0.0004	0.0086	0.0087	1.57979	2.26296	1.75966	2.87650
18	-0.0036	0.0136	0.0129	5.79103	7.96571	9.28168	13.92929
19	-0.0028	0.0003	0.0078	1.83044	4.70032	2.41775	4.87909
20	0.0120	0.0123	0.0768	<b>38.13972</b>	<b>190.29688</b>	<b>214.92329</b>	<b>894.51844</b>
21	-0.0015	0.0004	0.0115	1.26507	2.33013	1.54862	2.06405
22	0.0009	0.0232	0.0202	8.41812	<b>19.77199</b>	<b>24.65515</b>	<b>45.24620</b>
23	-0.0035	0.0088	0.0107	3.75884	5.16445	4.87934	7.53275
24	0.0016	0.0061	0.0066	1.06020	1.75635	0.93200	2.23575
25	-0.0228	-0.0466	0.0231	<b>42.84468</b>	<b>134.62223</b>	<b>68.63065</b>	<b>116.02933</b>
26	0.0037	-0.0038	0.0147	0.48316	1.39455	0.77959	7.32655

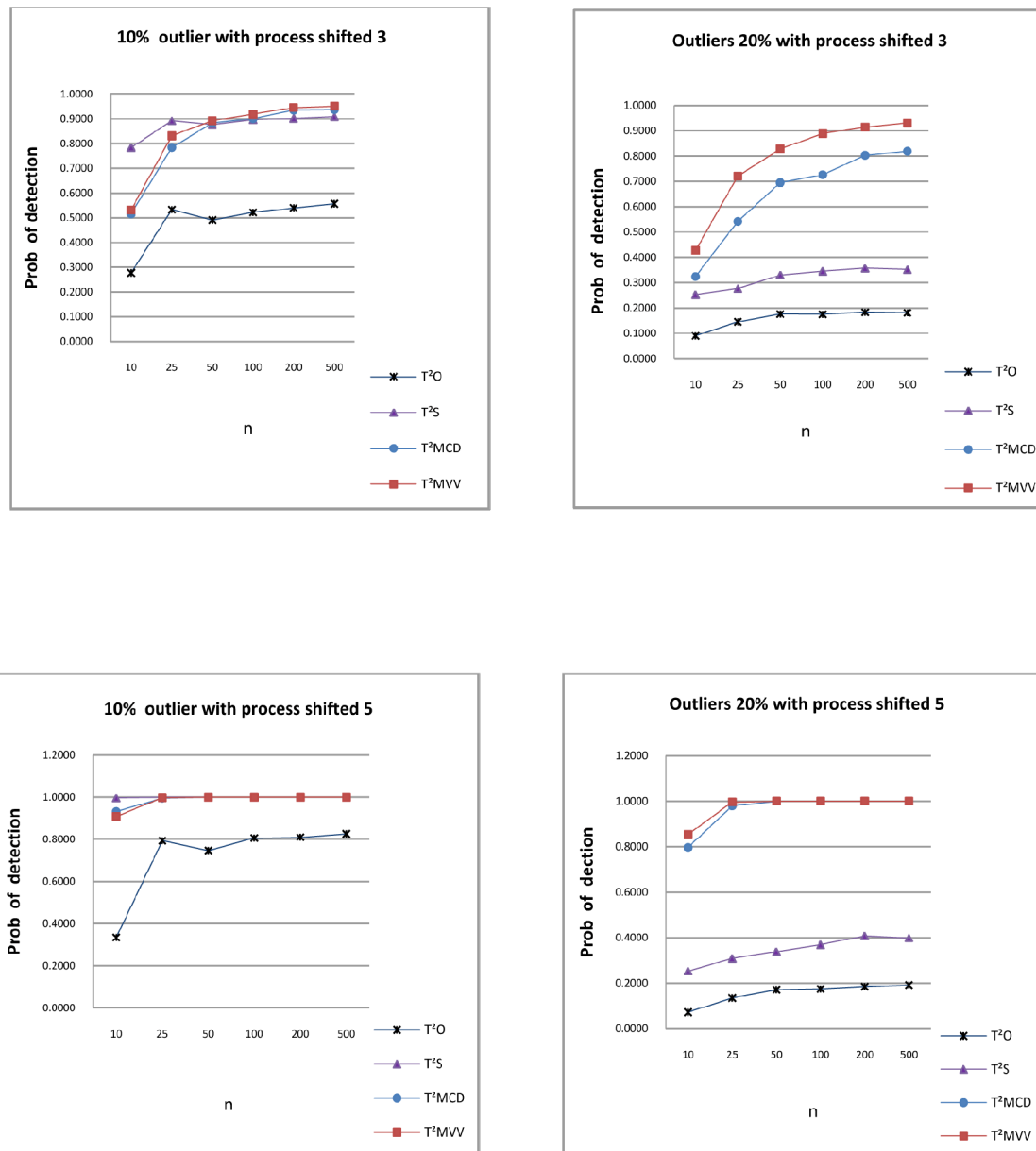
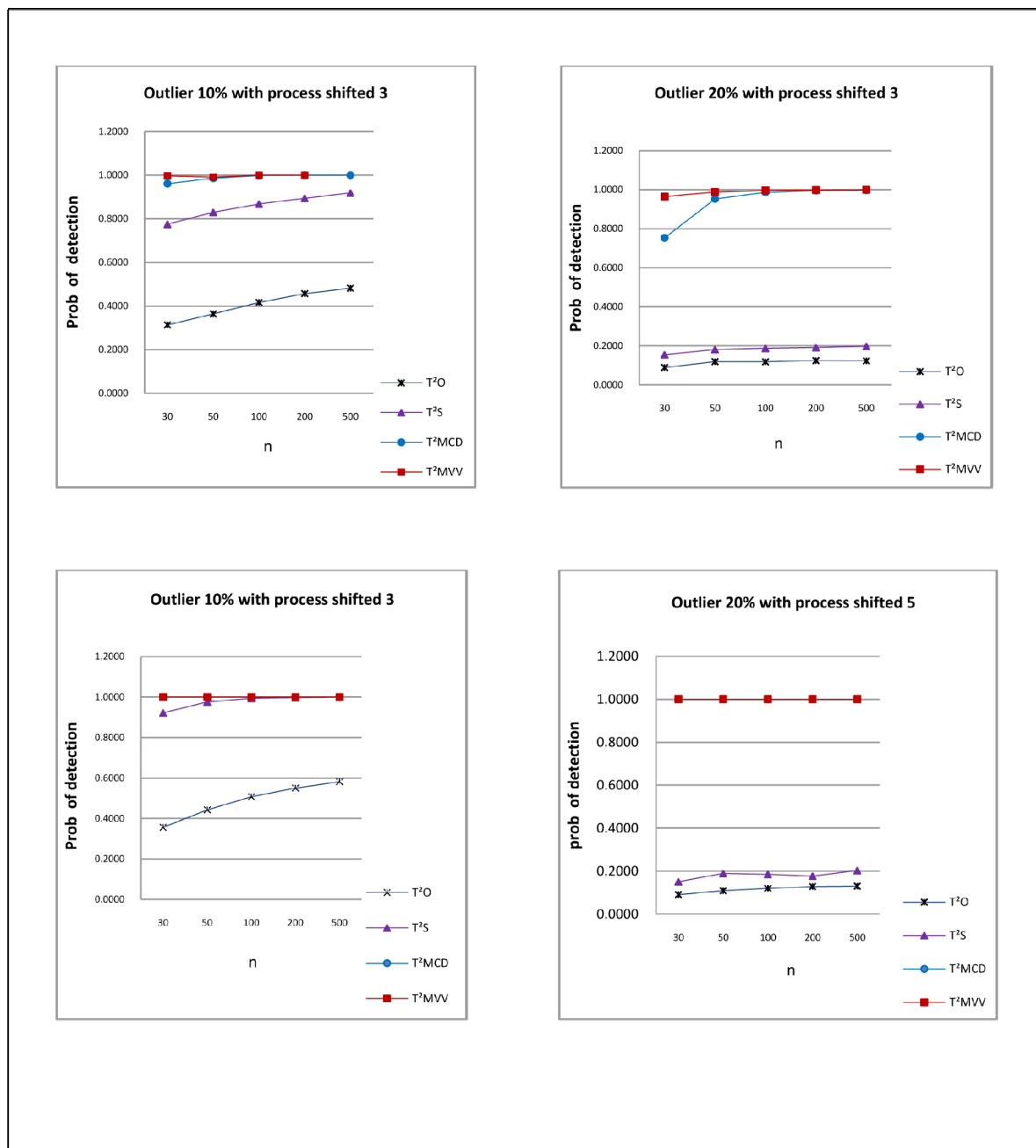


Figure 1: Probability of signal when  $p=2$ .

Figure 2: Probability of signal when  $p=5$ .

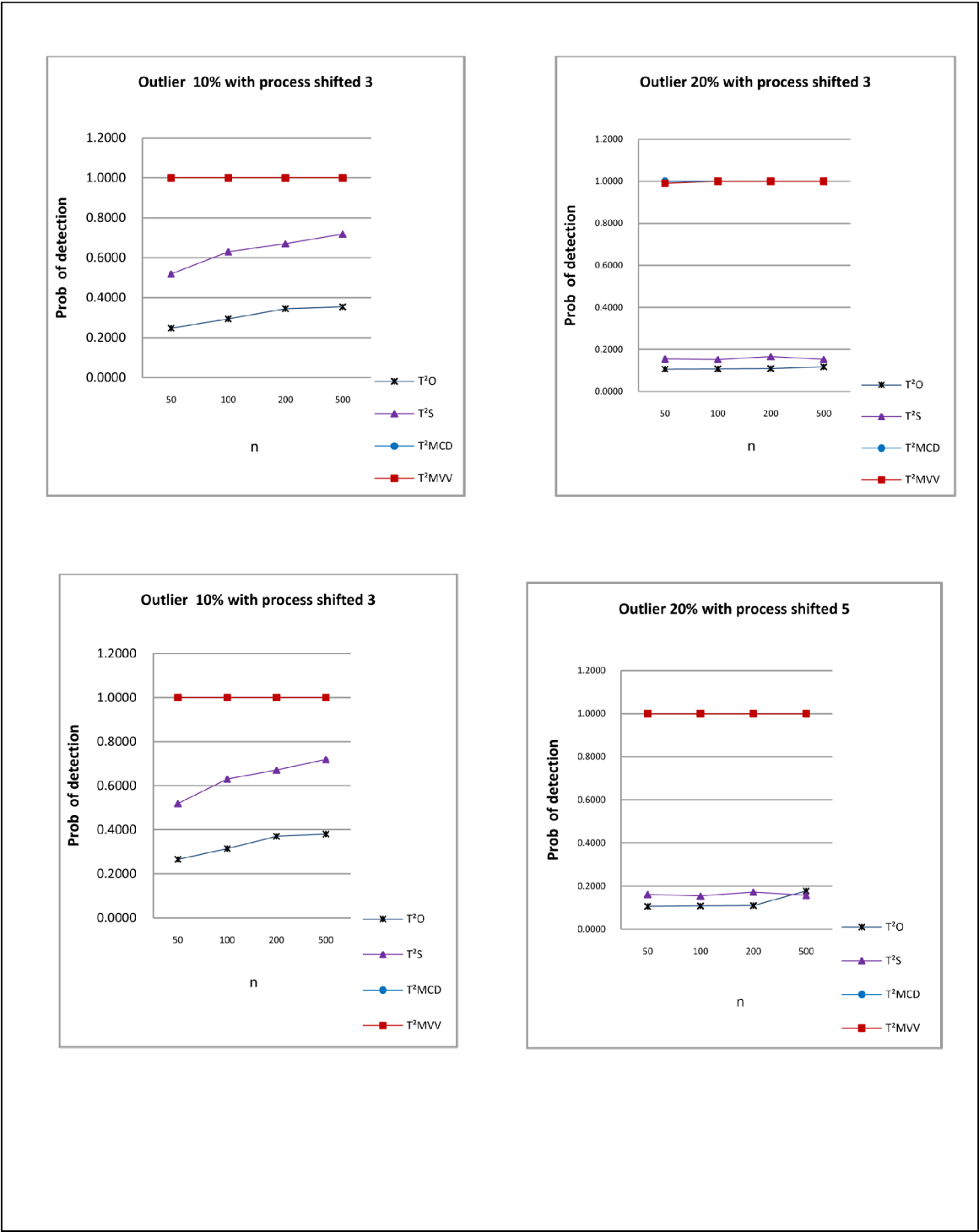
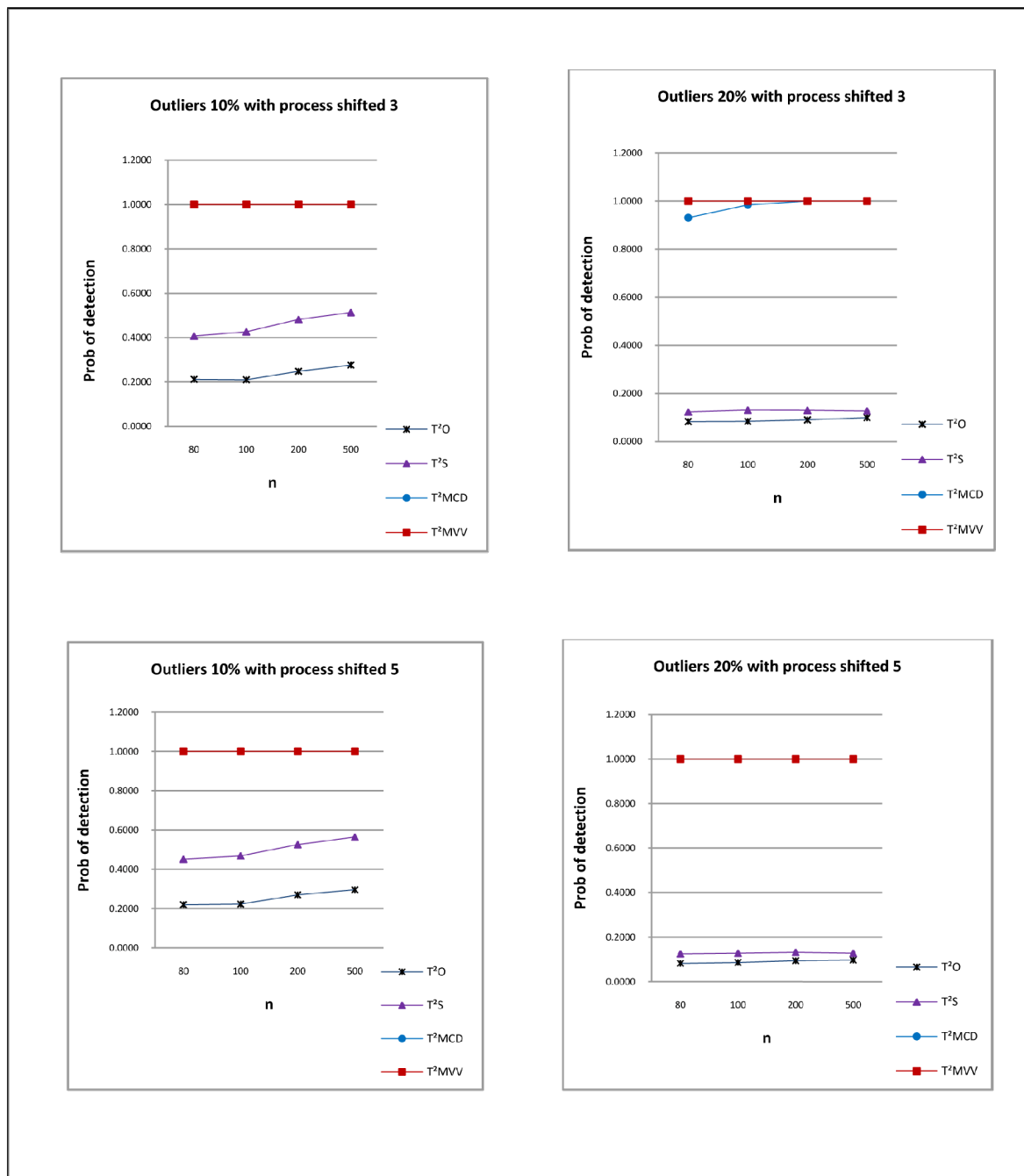


Figure 3: Probability of signal when  $p=10$ .

Figure 4: Probability of signal when  $p=15$ .



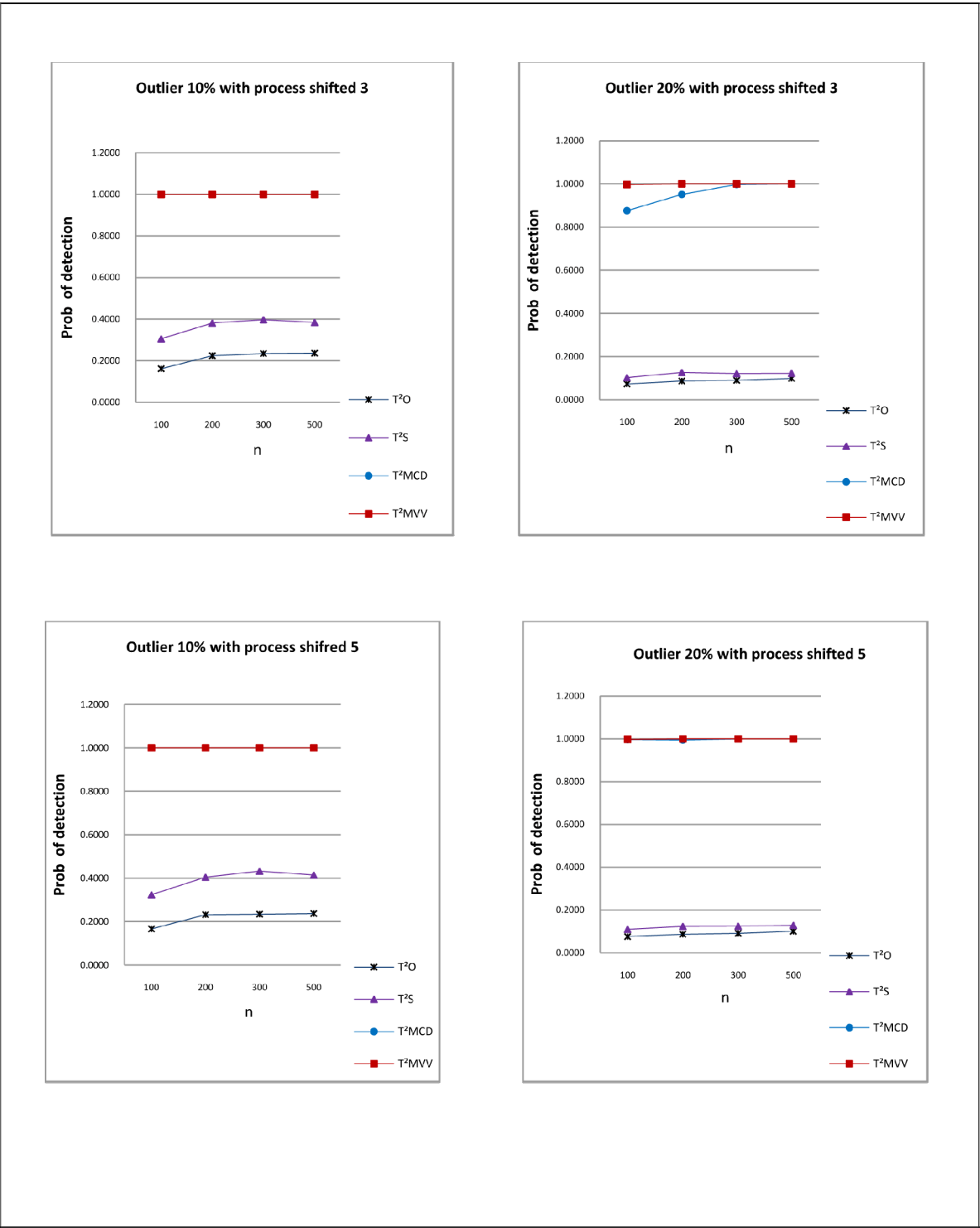


Figure 5: Probability of signal when  $p=20$ .

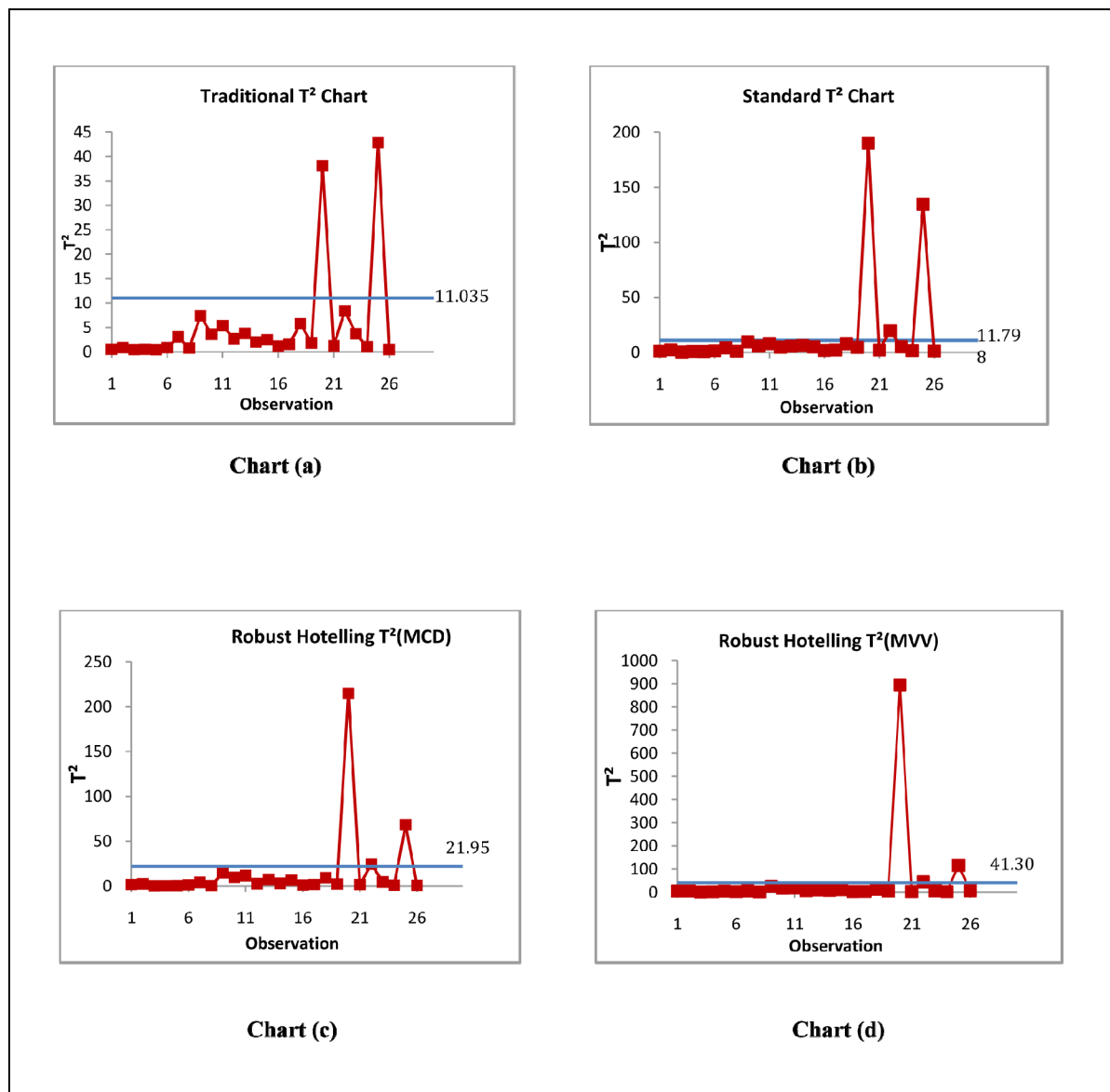


Figure 6. Hotelling  $T^2$  Control Charts