A Nonparametric Multivariate CUSUM Procedure For Detecting Shifts In All Directions

Peihua Qiu and Douglas Hawkins
School of Statistics
University of Minnesota
313 Ford Hall
224 Church St. SE
Minneapolis, MN 55455

Abstract

The fairly limited range of tools for multivariate statistical process control generally rest on the assumption that the data vectors follow a multivariate normal distribution – an assumption that is rarely satisfied. In this paper, we discuss detecting possible shifts in the mean vector of a multivariate measurement of a statistical process when the multivariate distribution of the measurement is non-Gaussian. A nonparametric cumulative sum (CUSUM) procedure is suggested, which is based both on the order information among the measurement components and on the order information between the measurement components and their in-control means. It is shown that this procedure is effective in detecting a wide range of possible shifts. Several numerical examples are presented to evaluate its performance. This procedure is also applied to a dataset from an aluminum smelter.

Key Words: Antiranks; CUSUM; Distribution-Free; Order Statistics; Robustness; Statistical Process Control.

1 Introduction

Statistical process control (SPC) when measurements are multivariate is an important problem with broad applications in industry (see e.g., Mason *et al.* 1997). Multivariate CUSUM procedures provide a tool for detecting possible changes in the distribution of the multivariate measurement of a statistical process. In the literature, a number of multivariate CUSUM procedures have been proposed. Most of them are based on the assumption that the multivariate measurement has a joint normal distribution, which is true (or approximately true) in some cases but may not be true in some other cases (e.g., when the measurements are counts). Therefore it is important to suggest some multivariate CUSUM procedures which do not depend on the normal distribution assumption. Stoumbos *et al.* (2000) concurred, pointing out that "nonparametric procedures for multivariate problems is an open field with great potential." The main objective of this paper is to propose a multivariate CUSUM procedure which could detect all possible shifts in the mean vector of a multivariate process without the normal distribution assumption.

According to Woodall (2000), process control has two phases. In Phase I, a set of process data is gathered and analyzed. Any 'patterns' in this data set indicate a lack of statistical control and lead to adjustments and fine tuning. Usually the early readings are more variable than the later as a result of adjustments and fine-tuning. There may be outliers, indicative of isolated special causes; these too are diagnosed and steps are taken to prevent their recurrence. Once all such bugs are fixed, we are left with a clean set of data, gathered under stable operating conditions and illustrative of the actual process performance. This set is then used to estimate the in-control distribution of the measurement. Like most SPC procedures in the literature, the current procedure focuses on Phase II problems in which the in-control distribution of the measurement is assumed to be known (it is actually estimated from Phase I experiments) and our major purpose is to detect any changes in the measurement distribution of a statistical process.

Let $\mathbf{X}(i) = (X_1(i), X_2(i), \cdots, X_p(i))'$ denote p measurements of a process at the i-th time point. When the process is in-control, $\mathbf{X}(i)$ is assumed to be independent and identically distributed across different time points with joint distribution function $F(\mathbf{x})$. The process becomes out-of-control if the measurement distribution changes to another distribution after an unknown time point. In this paper we focus specifically on possible shifts in the mean vector of the measurement distribution. We will suppose that a preliminary transformation has been applied to the data, so that the variables all have a mean of zero. As a practical matter, it is a good idea to make the components as exchangeable as possible, for example by scaling them to have a standard deviation of 1. As a consequence of the pre-centering, the in-control mean vector $\mu = \mathbf{0} = (0, 0, \cdots, 0)'$. Subsequently, the process may go out of control through its mean vector μ shifting at some unknown instant from $\mathbf{0}$, and other properties of the process are assumed to be unchanged after the shift.

There are many existing procedures for detecting shifts in the mean vector of $\mathbf{X}(i)$ if the distribution of $\mathbf{X}(i)$ is multivariate normal. Hotelling's (1947) control chart signals a shift in the mean vector when $T^2 = \mathbf{X}(i)'\Sigma_{\mathbf{X}}^{-1}\mathbf{X}(i) > h$ where h is a control limit and $\Sigma_{\mathbf{X}}$ denotes the covariance matrix of $\mathbf{X}(i)$. Woodall and Ncube (1985) suggested detecting the shift by monitoring CUSUMs of individual measurements simultaneously. Healy (1987) suggested a CUSUM procedure based on a linear combination of individual measurements. Hawkins (1991) extended Healy's method by proposing a CUSUM procedure based on regression adjustment among p individual measurements to detect a shift in several known directions. Pignatiello and Runger (1990) proposed a similar procedure to detect a shift by using several aimed CUSUMs. Crosier (1988) suggested two multivariate CUSUM procedures: one is a CUSUM of T and the other signals a shift when $S_i'\Sigma_{\mathbf{X}}^{-1}S_i > h$, where $S_i = 0$ if $C_i \leq k$ and $S_i = (S_{i-1} + \mathbf{X}(i))(1 - k/C_i)$ otherwise, $C_i = \{(S_{i-1} + \mathbf{X}(i))'\Sigma_{\mathbf{X}}^{-1}(S_{i-1} + \mathbf{X}(i))\}^{1/2}$, $S_0 = 0$, and k is a constant. (The control limit h appearing at several different places may represent different constants. We use the same notation for simplicity.) From his simulation studies, Crosier (1988) concluded that his second procedure generally performed better than the first. Lowry *et al.* (1992) extended the univariate exponentially weighted moving average control procedure (e.g., Lucas and Saccucci 1990) to the multivariate case. In the literature, the performance of a CUSUM procedure is often measured by the average run length (ARL), which is the average number of samples needed for the procedure to signal the shift.

In many (perhaps most) applications, the distribution of $\mathbf{X}(i)$ may not be normal. When this is the case, the actual in-control ARL value is generally not equal to the specified in-control ARL value if a conventional multivariate CUSUM procedure is used. Consequently the chance that the conventional multivariate CUSUM procedure signals because of randomness (namely, the false-alarm rate) would be different from what we would expect. Such difference could be substantial in many situations.

A natural idea to overcome this difficulty is to find a transformation in Phase I of the process control such that the transformed data follow a normal distribution when the process is in-control. However, it is notoriously difficult to find such a transformation in the multivariate case. Even worse, if any of the measurements are discrete, it is impossible to transform them to normal no matter how many data we have.

Qiu and Hawkins (2001) suggested an alternative idea based on the cross-sectional antirank vector: the vector of the indices of the order statistics of the individual measurements at each time point. For example, if a data vector is (-1, 5, 0, 3, 1, -2), then the antirank vector is (6, 1, 3, 5, 4, 2). Procedures based on antiranks are distribution-free in the sense that all their properties depend on the distribution of the antirank vector only. The Qiu-Hawkins procedure, however, can not detect shifts in or close to one specific direction in which the components of the shift are all the same but not zero. We call this direction the equal-component direction hereafter.

In applications, the direction of a possible shift is often unknown. By using the above antirank-based procedure, it is possible to miss a shift simply because the shift is in or close to the equal-component direction. In this paper we suggest a new procedure which can detect shifts in all possible directions without making the normal distribution assumption. The new procedure is based on both the order information of the measurement components and the order information between the measurement components and their in-control means. It is distribution-free in the sense that all its properties depend only on the ordering of the measurement components and on the ordering between the measurement components and their in-control means. It is therefore appropriate to use in the case when the potential shifts in the mean vector of the process can occur in all possible directions and the distribution of the multivariate measurement might be non-Gaussian.

The remaining part of the article is organized as follows. In next section, the new procedure is introduced in detail along with some technical remarks. A simulation study is presented in Section 3 regarding the numerical performance of the procedure. The new procedure is then applied to a dataset of process-control readings from an aluminum smelter in Section 4. Some remarks conclude the article in Section 5.

2 The Proposed Multivariate CUSUM Procedure

As discussed in Section 1, we focus on detecting shifts in the mean vector of the multivariate measurements of a statistical process. The process is out-of-control if the mean vector $\mu = (\mu_1, \mu_2, \cdots, \mu_p)'$ of its multivariate measurement violates the hypothesis $H_0 : \mu_1 = \mu_2 = \cdots = \mu_p = 0$ after a specific time point. Clearly, one can split this hypothesis into two components $H_0^{(1)} : \mu_1 = \mu_2 = \cdots = \mu_p$; and $H_0^{(2)} : \sum_{j=1}^{p} \mu_j = 0$. The null hypothesis $H_0^{(1)}$ is related to the ordering of $\{\mu_j, j = 1, 2, \cdots, p\}$. When $H_0^{(1)}$ is true, $H_0^{(2)}$ is related to the magnitudes of $\{\mu_j, j = 1, 2, \cdots, p\}$.

A formal definition of the antirank vector is that the vector $\mathbf{A}(i) = (A_1(i), A_2(i), \cdots, A_p(i))'$ is called an antirank vector of $\mathbf{X}(i)$ if $\mathbf{A}(i)$ is a permutation of $(1, 2, \cdots, p)'$ such that $X_{A_1(i)}(i) \leq X_{A_2(i)}(i) \leq \cdots \leq X_{A_p(i)}(i)$ are the order statistics of $\{X_j(i), j = 1, 2, \cdots, p\}$. So the first component of the antirank vector identifies which element of $\mathbf{X}$ is the smallest, and the last component identifies which is the largest. Each component of the antirank vector follows a multinomial distribution whose possible values are $\{1, 2, \ldots, p\}$. Clearly, the distribution of $\mathbf{A}(i)$ would be changed by any shift violating $H_0^{(1)}$. For example, if the mean of, say, X_3 decreases, then the first component of the antirank vector should show an increased frequency of the value 3. The distribution of the antiranks, however, would not be changed if all the measurement components increase or decrease by the same amount, since this would not change the relative magnitudes of the components.

The Qiu-Hawkins method is therefore unable to detect this 'equal shift' situation. In this paper we suggest a remedy. We notice that in order to detect all possible shifts by a procedure based on the antirank vector alone, the definition of the antirank vector needs to be modified. The modified antirank vector should be sensitive to shifts that violate either $H_0^{(1)}$ or $H_0^{(2)}$ or both. Based on this observation, we define $\mathbf{B}(i) = (B_1(i), B_2(i), \cdots, B_p(i), B_{p+1}(i))'$ as the antirank vector of $\mathbf{Y}(i) := (X_1(i), X_2(i), \cdots, X_p(i), 0)'$, a combination of the measurement vector $\mathbf{X}(i)$ and the common in-control mean of 0. Therefore $\mathbf{B}(i)$ is a permutation of $(1, 2, \cdots, p, p+1)'$ such that $Y_{B_1(i)}(i) \leq Y_{B_2(i)}(i) \leq \cdots \leq Y_{B_{p+1}(i)}(i)$ are the order statistics of the components of $\mathbf{Y}(i)$. Clearly the antirank vector $\mathbf{B}(i)$ contains the order information of $\{\mu_1, \mu_2, \ldots, \mu_p, 0\}$. It is sensitive to the ordering of $\{\mu_j, j = 1, 2, \cdots, p\}$ and to the ordering between $\{\mu_j, j = 1, 2, \cdots, p\}$ and 0 as well. Therefore a CUSUM procedure based on this antirank vector has the potential to detect all possible shifts.

There are many ways to use the antirank vector, so we will start with a scheme based on its first component. For $1 \leq j \leq p+1$, define

$$\eta_{1,j}(i) = I(B_1(i) = j) \tag{2.1}$$

and $\eta_1(i) := (\eta_{1,1}(i), \eta_{1,2}(i), \ldots, \eta_{1,p+1}(i))'$, where $I(\cdot)$ is an indicator function which equals 1 if its argument is "true" and 0 otherwise. That is, $\eta_{1,j}(i)$ is defined as an indicator of the event that the j-th component of $\mathbf{Y}(i)$ takes the smallest value among all $p+1$ components at time point i. Under H_0, suppose that $E(\eta_{1,j}(i)) = d_j$, for $j = 1, 2, \cdots, p+1$. Then the probability distribution of $B_1(i)$ is $\{d_j, j = 1, 2, \cdots, p+1\}$. It can be checked that this distribution will change to some other distribution after any shift in the mean vector of the process as long as the null distribution $F(\mathbf{x})$ of the multivariate measurement has a positive support around its mean vector. Therefore a CUSUM based on $B_1(i)$ can be used to detect shifts that violate H_0. Based on Crosier's (1988) results, we suggest the following procedure for detecting shifts. Let

$$\begin{cases} \mathbf{S}_n^{(1)} &= \mathbf{0} \\ \mathbf{S}_n^{(2)} &= \mathbf{0}, \qquad\qquad\qquad\qquad \text{if } C_n \leq k \\ \mathbf{S}_n^{(1)} &= (\mathbf{S}_{n-1}^{(1)} + \eta_{\mathbf{1}}(n))(C_n - k)/C_n \\ \mathbf{S}_n^{(2)} &= (\mathbf{S}_{n-1}^{(2)} + \mathbf{d})(C_n - k)/C_n, \qquad \text{if } C_n > k, \end{cases}$$

$$\begin{aligned} C_n \quad = \quad & [(\mathbf{S}_{n-1}^{(1)} - \mathbf{S}_{n-1}^{(2)}) + (\eta_{\mathbf{1}}(n) - \mathbf{d})]' diag(\frac{1}{S_{n-1,1}^{(2)} + d_1}, \cdots, \frac{1}{S_{n-1,p+1}^{(2)} + d_{p+1}}) \\ & [(\mathbf{S}_{n-1}^{(1)} - \mathbf{S}_{n-1}^{(2)}) + (\eta_{\mathbf{1}}(n) - \mathbf{d})], \end{aligned}$$

where $\mathbf{d} = (d_1, d_2, \cdots, d_{p+1})'$, $\mathbf{S}_0^{(1)} = \mathbf{S}_0^{(2)} = \mathbf{0}$ and $k \geq 0$ is a constant. Define

$$y_n = (\mathbf{S}_n^{(1)} - \mathbf{S}_n^{(2)})' diag(1/S_{n,1}^{(2)}, \cdots, 1/S_{n,p+1}^{(2)})(\mathbf{S}_n^{(1)} - \mathbf{S}_n^{(2)}). \tag{2.2}$$

Then

$$y_n > h \tag{2.3}$$

signals a shift where $h > 0$ is a control limit. The constant k is sometimes called 'allowance' constant. In the special case $k = 0$, $\mathbf{S}_{n,j}^{(1)}$ is the observed count of the event $(B_1(i) = j)$ as of time point n, and $\mathbf{S}_{n,j}^{(2)} = nd_j$ is the corresponding expected count as of time point n, for $j = 1, 2, \ldots, p$, and therefore $y_n = \sum_{j=1}^{p+1}(S_{n,j}^{(1)} - S_{n,j}^{(2)})^2/S_{n,j}^{(2)}$ is the conventional Pearson's χ^2 statistic which measures the difference between the observed counts of the events $(B_1(i) = j)$ and the expected counts of the same events under H_0. While this exact connection no longer holds if $k > 0$, this special case perhaps motivates the chart statistic y_n.

This CUSUM is clearly distribution-free in that it depends only on the distribution of the first antirank $B_1(i)$, and it is also simple to use. There is, of course, some price to pay for the additional capability to detect equal-component out-of-controls; this price is a slightly reduced ability to detect a shift in a single component compared to the procedures based on $\mathbf{A}(i)$. In Section 3 we will show that there is some performance loss, but that the loss is small.

The CUSUM procedure (2.2)-(2.3) is based on the first antirank $B_1(i)$. Similar procedures can be constructed from any single component or any combination of several components of $\mathbf{B}(i)$.

Let $(B_{j_1}(i), B_{j_2}(i), \ldots, B_{j_q}(i))'$ be any q components of $\mathbf{B}(i)$ with $j_1 < j_2 < \cdots < j_q$. Suppose that the in-control distribution of $(B_{j_1}(i), B_{j_2}(i), \ldots, B_{j_q}(i))'$ has been determined in sample space $S(j_1, j_2, \ldots, j_q) = \{(i_1, i_2, \ldots, i_q) : i_1, i_2, \ldots, i_q$ are q different integers in $(1, 2, \ldots, p, p+1)\}$. This sample space has $P_{q,p} = (p+1)p \ldots (p-q+2)$ elements. Similar to (2.1), we can define a $P_{q,p}$-dimensional random vector $\eta_{j_1,j_2,\ldots,j_q}(i)$ with its j-th component equal to 1 when $(B_{j_1}(i), B_{j_2}(i), \ldots, B_{j_q}(i))'$ takes the value of the j-th element in $S(j_1, j_2, \ldots, j_q)$ and 0 otherwise. Then a multivariate CUSUM can be constructed similar to (2.2)-(2.3) with $\eta_1(i)$ replaced by $\eta_{j_1,j_2,\ldots,j_q}(i)$ and p by $P_{q,p}$. In Section 3, we will demonstrate with numerical examples that the first antirank is particularly effective in detecting a downward shift in a single arbitrary and unknown component and the last antirank is attractive for detecting an upward shift in one of the components. In all cases, the CUSUM procedure based on the first and last antiranks is effective.

When all or some of the measurements are discrete, ties may exist in the components of $\mathbf{Y}(i)$. In such case, the antirank vector $\mathbf{B}(i)$ and consequently $\eta_{\mathbf{1}}(i)$ defined in (2.1) may not be well defined. We suggest using the following strategy to overcome this difficulty. Suppose that $Y_{j_1}(i), Y_{j_2}(i), \ldots, Y_{j_r}(i)$ form a tie and their values reach the minimum among all $p+1$ components at time point i. Then instead of defining $\eta_{\mathbf{1}}(i)$ by (2.1) we define $\eta_{1,j}(i) = \frac{1}{r} I(j \in \{j_1, j_2, \cdots, j_r\})$. It can be seen that no information about the ordering of the components of $\mathbf{Y}(i)$ is lost by using this definition and the results are also reproducible.

The 'allowance' constant k used in (2.2)-(2.3) should be chosen from the interval $[0, \max_{\ell=1}^{p+1} \frac{\sum_{j \neq \ell} d_j}{d_\ell})$. If k lies above this interval, then the CUSUM procedure will restart at each time point and consequently the specified in-control ARL property can not be achieved.

3 A Simulation Study

In this section, we present some simulation results regarding the performance of the CUSUM procedure (2.2)-(2.3). We assume that $p = 4$ and the in-control joint distribution of $\mathbf{X}(i)$ is $N(\mathbf{0}, I_4)$. In such a case, the in-control distribution of $B_1(i)$ is $multinomial(0.2344, 0.2344, 0.2344, 0.2344, 0.0624)$. We first compare the procedure (2.2)-(2.3) with the procedure of Qiu and Hawkins (2001). The former is called the current procedure and the latter is called the QH01 procedure in this section. The in-control ARL values of both procedures are fixed at 200. The constant k in both procedures is chosen to be 0.5. The control limits are 12.488 and 8.029 for the current and QH01 procedures, respectively. Remember that the current procedure depends only on the distribution of the first antirank $B_1(i)$. So its control limit value can be searched by a simulation in which the CUSUM is constructed from a series of i.i.d. random vectors with multinomial distribution (0.2344,0.2344,0.2344,0.2344,0.0624). These random vectors can be generated by a multinomial random number generator. From our experience, it is much faster to search the control limit value in this way than to search its value based on a CUSUM constructed from the original measurements. When the process undergoes a shift in mean, the distribution of each antirank changes to some other multinomial distribution whose probabilities are given by some straightforward multivariate normal probability calculations.

We first study the equal-component scenario, assuming that the mean vector has become $(a, a, a, a)'$, with a taking the values 0, 0.2, 0.4, 0.6, 0.8 and 1. Figure 3.1(a) shows the ARL's given by 10,000 simulations. The QH01 procedure, as expected, has no ability whatever to detect this shift. The current procedure, however, is quite effective for a values above 0.4. This is because the mean shift substantially increases the probability that the four components of the data will all be positive.

Next consider an intermediate case - where the mean vector shifts to $(a, 1, 1, 1)'$. The case $a = 1$ reduces to the previous one. The results of simulating 10,000 runs are shown in Figure 3.1(b). The QH01 procedure is ineffective when a is close to 1 (as expected) but performs quite well when a is near zero. This is because the situation $a = 0$ is (to the QH01 procedure) essentially the same as a downward shift in μ_1, all other means remaining the same. The current procedure however has a fairly consistent and much better performance for all a values.

Figure 3.1(c) shows a situation in which the first component of the mean vector shifts, but in an upward direction, while Figure 3.1(d) shows the situation for which the QH01 procedure was designed - a downward shift in just the first component of the mean vector. In the first case, the current procedure retains a small superiority. In the second case, it is inferior to the QH01 method, though the difference is substantial only if the shift is relatively small.

As a summary, we can have the following conclusions from the above example. (1) The current procedure can detect shifts in the equal-component direction. (2) When the shifts are off but close to that direction, the current procedure often outperforms the QH01 procedure, especially when the shifts are far away from the origin. (3) When only a small number of measurement components have shifts in their means, the benefit to use the current procedure is small. Only in special circumstances does the current procedure give substantially worse performance than the QH01 procedure.

As mentioned in Section 2, the CUSUM procedure (2.2)-(2.3) can be constructed using any antirank, or any combination of antiranks. Intuitively, if we are to use a single antirank, it should be either the first (to check for a downward shift in a single component) or the last (to check for an upward shift in a single component). If we are willing to envisage using more than one antirank, then the logical choice would seem to be the first and the last.

Table 3.1 supports this reasoning. It shows the out-of-control ARL's for all possible single-antirank and two-antirank CUSUMs in our four-component setting using a variety of shifts. In the table, P1 denotes the CUSUM procedure based on the first antirank $B_1(i)$ only, P12 denotes the CUSUM procedure based on the first and second antiranks, and the other notations starting with "P" are similarly defined. In particular, P15 refers to a scheme using the first and last antiranks. In all CUSUMs, $k = 0.5$ and the in-control ARL is 200. Getting the ARL's involves calculating the multinomial probabilities for each out-of-control setting and then simulating these multinomials to get the antiranks. For example the shift in mean to $(-2, 0, 0, 0)$ changes the multinomial distribution of $B_1(i)$ from (0.2344, 0.2344, 0.2344, 0.2344, 0.0624) to (0.8217, 0.0585, 0.0585, 0.0585, 0.0028), showing that the first antirank will usually be 1 and practically never be 5.

In all settings, the best single-antirank scheme was either P1 or P5. The two-antirank scheme P15 was not always the best of the two-antirank schemes, but it was never the worst either and its ARL was uniformly small.

The other two-antirank schemes do well for the specific situation for which they make sense, but not so well elsewhere. So P12 is very good for the shifts (-2,-2,0,0) and (2,2,2,0) since in the first case the first two antiranks will most commonly be 1 and 2, and in the second 4 and 5, but is not very effective at a shift of (-2, -2, -2, 0)

Table 3.1 also shows situations in which the out-of-control ARL is larger than the in-control ARL of 200. This situation arises with single-antirank CUSUMs that are 'looking in the wrong direction'. So the P1 scheme is worse than no scheme at all if the shift is predominantly downward, as is P5 if the shift is predominantly upward. This situation is conceptually like that of a "biased" statistical test.

The next example explores this phenomenon. In Figure 3.2(a), the shift in the mean vector of the multivariate measurement has the form $(\mu_1, \mu_2, \mu_3, \mu_4) = (a, 0, 0, 0)$ where a varies among -1, -0.8, -0.6, -0.4, -0.2 and 0. Three procedures P1, P5 and P15 are considered. In these procedures, $k = 0.5$ and the in-control ARL is fixed at 200 as before. From the construction of P1 and P5, it is clear that the performance of P5 when $(\mu_1, \mu_2, \mu_3, \mu_4) = (a, 0, 0, 0)$ is exactly the same as the performance of P1 when $(\mu_1, \mu_2, \mu_3, \mu_4) = (-a, 0, 0, 0)$. The out-of-control ARL values of the three procedures are presented in the plot. We can see that when the shifts are small (namely, when a = -0.2 or -0.4 in the plot), the out-of-control ARL values of P1 are larger than 200. But the out-of-control ARL values of P1 when a = 0.2 and 0.4 (which are the out-of-control ARL values of P5 when a = -0.2 and -0.4, respectively, presented in the plot) are smaller than 200. In other words, the out-of-control ARL of P1 is not symmetric about the upward and downward shifts. This can be explained by the fact that the upward and downward shifts with a same magnitude change the null distribution of $B_1(i)$ differently, which is obvious from the definition of $B_1(i)$. When the shift is large ($|a|$ is large in the plot), both out-of-control ARL values of P1 and P5 are smaller than 200. So this bias phenomenon may not get our attention in such cases although P1 is still biased as demonstrated by the plot. When the magnitude of the shift is small, this phenomenon becomes obvious because the out-of-control ARL of P1 can be larger than the specified in-control ARL value. From the plot, we can see that our recommended procedure P15 performs consistently well in all cases.

The bias phenomenon is not new. Hawkins (1987) showed that the conventional univariate CUSUM procedure was biased with respect to shifts in the scale parameter of the measurement distribution (see also Hawkins and Olwell 1998, Chapter 7). It was shown that the out-of-control ARL of that procedure was much larger than the specified in-control ARL value when the variance of the measurement distribution increased a small amount.

Figure 3.2(c) presents the out-of-control ARL values of P1, P5 and P15 when $(\mu_1, \mu_2, \mu_3, \mu_4) = (a, a, a, 0)$ and a varies among -3, -2.5, -2, -1.5, -1, -0.5 and 0. A similar phenomenon to that demonstrated by Figure 3.2(a) can be seen from this plot except that the bias of P1 seems more serious in these cases because the shift $(a, a, a, 0)$ when a is negative does not change the distribution

Table 3.1: The out-of-control ARL values and their standard errors (numbers in parentheses) based on 10,000 replications for several shifts and several CUSUM procedures based on one or two different antiranks. The in-control ARL is fixed at 200 and $k = 0.5$ for each procedure. P1 denotes the CUSUM procedure based on the first antirank $B_1(i)$, P12 denotes the CUSUM procedure based on the first and second antiranks $B_1(i)$ and $B_2(i)$, and the other notations started with "P" are similarly defined.

	$(\mu_1, \mu_2, \mu_3, \mu_4)$				
	(-2,0,0,0)	(-2,-1,0,0)	(-2,-2,0,0)	(-2,-2,-1,0)	(-2,-2,-2,0)
P1	8.31 (.04)	12.60 (.07)	18.05 (.07)	31.49 (.19)	238.13 (2.30)
P2	76.47 (.78)	22.90 (.21)	11.90 (.08)	17.42 (.13)	15.76 (.10)
P3	113.79 (1.18)	64.43 (.65)	32.48 (.31)	17.69 (.16)	15.55 (.12)
P4	116.83 (1.21)	46.34 (.43)	26.58 (.20)	26.32 (.21)	17.77 (.13)
P5	80.36 (.94)	15.22 (.14)	10.16 (.08)	4.29 (.04)	3.21 (.02)
P12	5.47 (.04)	4.88 (.03)	3.70 (.02)	5.07 (.03)	6.26 (.04)
P13	6.31 (.04)	7.70 (.06)	7.98 (.06)	6.05 (.04)	5.34 (.03)
P14	6.01 (.04)	7.29 (.05)	7.22 (.05)	9.17 (.07)	10.30 (.08)
P15	5.84 (.04)	4.72 (.03)	4.23 (.03)	2.59 (.02)	2.18 (.02)
P23	12.10 (.22)	3.69 (.05)	2.55 (.03)	1.49 (.02)	1.15 (.01)
P24	16.73 (.21)	9.12 (.08)	6.27 (.04)	7.05 (.05)	6.28 (.04)
P25	7.82 (.10)	3.84 (.04)	2.99 (.02)	2.11 (.01)	1.83 (.01)
P34	20.29 (.26)	13.77 (.17)	8.37 (.08)	4.90 (.05)	3.73 (.04)
P35	8.75 (.10)	4.39 (.04)	3.23 (.03)	2.10 (.01)	1.82 (.01)
P45	6.95 (.07)	3.66 (.03)	2.88 (.02)	1.96 (.01)	1.64 (.07)
	$(\mu_1, \mu_2, \mu_3, \mu_4)$				
	(2,0,0,0)	(2,1,0,0)	(2,2,0,0)	(2,2,1,0)	(2,2,2,0)
P1	79.94 (.93)	14.99 (.14)	9.96 (.08)	4.24 (.04)	3.20 (.03)
P2	114.03 (1.20)	45.69 (.42)	26.58 (.20)	26.49 (.21)	18.01 (.13)
P3	112.55 (1.17)	64.69 (.66)	32.40 (.31)	17.86 (.15)	15.47 (.12)
P4	76.00 (.77)	22.79 (.20)	12.08 (.08)	17.57 (.13)	15.93 (.10)
P5	8.42 (.04)	12.63 (.07)	18.05 (.07)	31.54 (.19)	235.15 (2.24)
P12	6.87 (.07)	3.65 (.03)	2.86 (.02)	1.95 (.01)	1.64 (.01)
P13	8.75 (.10)	4.39 (.04)	3.23 (.03)	2.10 (.01)	1.82 (.01)
P14	7.68 (.10)	3.86 (.04)	2.98 (.02)	2.11 (.01)	1.82 (.01)
P15	5.93 (.04)	4.72 (.04)	4.14 (.03)	2.59 (.02)	2.18 (.02)
P23	19.92 (.26)	13.78 (.16)	8.43 (.08)	4.88 (.05)	3.70 (.04)
P24	16.52 (.21)	9.07 (.08)	6.32 (.04)	7.06 (.05)	6.29 (.04)
P25	6.09 (.04)	7.20 (.05)	7.24 (.05)	9.31 (.07)	10.45 (.08)
P34	11.36 (.21)	3.67 (.05)	2.56 (.03)	1.52 (.02)	1.16 (.01)
P35	6.31 (.04)	7.70 (.06)	7.98 (.06)	6.05 (.04)	5.34 (.03)
P45	5.50 (.04)	4.91 (.03)	3.79 (.02)	5.21 (.03)	6.30 (.04)

of $B_1(i)$ so much as the shift $(a, 0, 0, 0)$ does.

The choice of the constant k also bears on this bias. Letting $k = 0.25$, the corresponding results of Figure 3.2(a) are presented in Figure 3.2(b) and the corresponding results of Figure 3.2(c) are presented in Figure 3.2(d). It can be seen that the out-of-control ARL of P1 when the absolute value of a is small is decreased and consequently the bias of P1 is also lessened. Therefore using a small k ameliorates the bias.

The 'allowance' constant k in CUSUM design depends on the target shift. Generally speaking, large k is appropriate for detecting large shifts and small k for detecting small shifts (cf. Hawkins and Olwell 1998, Chapter 6). This is also true for the procedure (2.2)-(2.3) although purely theoretical derivation of the optimal k is not clear. In the next example, we focus on the procedure P1 of which the in-control ARL is fixed at 200 and the constant k is one of 0.1, 0.3 and 0.5. The dimension p could be either 2 or 3 or 4. The in-control distributions of $B_1(i)$ when p= 2, 3 and 4 are (.375,.375,.25), (.2917,.2917,.2917,.1249) and (.2344,.2344,.2344,.2344,.0624), respectively. They are the distributions of $B_1(i)$ when the distribution of the measurement vector $\mathbf{X}(i)$ is $N(\mathbf{0}, I_p)$ for p= 2, 3 and 4. The distribution of $B_1(i)$ after a shift is denoted by $\mathbf{d}^* = (d_1^*, d_2^*, \cdots, d_p^*)$. For each p, four shifts are considered. The first three components of $\mathbf{d}^*$ are assumed to be (1.0,0.0,0.0), (0.9,0.1,0.0), (0.7,0.2,0.1) and (0.5,0.3,0.2), respectively, for these four shifts. The remaining components of $\mathbf{d}^*$ are assumed to be 0. The four shifts are ordered from the largest to the mildest in magnitude. The out-of-control ARL values of the three procedures are presented in Table 3.2.

Let us first check the performance of the procedure when p=3. In three different cases with $k = 0.1$, 0.3 and 0.5, it can be seen that the procedure performs the best when $k = 0.5$ for the first shift. For the second and third shifts, P1 performs the best when $k = 0.3$. For the last shift, P1 performs the best when $k = 0.1$. For $p = 2$ and 4, similar results can be found. Therefore it is indeed true that small k is good for detecting relatively small shifts and large k is good for detecting relatively large shifts. If we have a target shift in mind, then the optimal value of k can be searched by the programs provided in the paper such that the out-of-control ARL value reaches the minimum among all combinations of h and k resulting in a given in-control ARL value.

Table 3.2: The out-of-control ARL values and their standard errors (numbers in parentheses) of the procedure P1 based on 10,000 replications. The in-control ARL is fixed at 200. The constant k varies among 0.1, 0.3 and 0.5. The dimension p is either 2 or 3 or 4.

$p=2$			
(d_1^*, d_2^*, d_3^*)	$k=0.1$	$k=0.3$	$k=0.5$
(1,0,0)	7.60 (.09)	6.88 (.06)	6.96 (.05)
(.9,.1,0)	9.77 (.14)	9.25 (.13)	9.67 (.15)
(.7,.2,.1)	19.18 (.42)	23.41 (.56)	28.80 (.84)
(.5,.3,.2)	92.15 (3.09)	126.13 (4.24)	159.55 (5.85)
$p=3$			
$(d_1^*, d_2^*, d_3^*, d_4^*)$	$k=0.1$	$k=0.3$	$k=0.5$
(1,0,0,0)	7.46 (.09)	6.35 (.06)	6.34 (.05)
(.9,.1,0,0)	8.88 (.12)	7.89 (.11)	8.04 (.11)
(.7,.2,.1,0)	14.58 (.27)	14.16 (.30)	17.28 (.43)
(.5,.3,.2,0)	31.43 (.68)	56.05 (1.86)	86.47 (3.24)
$p=4$			
$(d_1^*, d_2^*, d_3^*, d_4^*, d_5^*)$	$k=0.1$	$k=0.3$	$k=0.5$
(1,0,0,0,0)	7.28 (.09)	6.36 (.05)	6.60 (.04)
(.9,.1,0,0,0)	8.72 (.12)	7.46 (.09)	7.96 (.08)
(.7,.2,.1,0,0)	11.14 (.17)	11.92 (.18)	14.32 (.24)
(.5,.3,.2,0,0)	18.72 (.30)	23.06 (.45)	41.69 (.93)

4 An Application

In this section, we apply the CUSUM procedure (2.2)-(2.3) to a data set (kindly supplied to us by Len Homer) from an aluminum smelter. The data set contains 5 variables – the content of SiO_2, Fe_2O_3, MgO, CaO, and Al_2O_3 (labelled as $x1, x2, x3, x4$ and $x5$ below) in the charge. All these measures are relevant to the operation of the smelter. Stability of the alumina level and calcium oxide level is desirable. The silica, ferric oxide and magnesium oxide levels are affected by the raw materials and are potential covariates to be taken into account in a fully-fledged multivariate scheme. The data set comprised 135 vectors. We used the first 95 vectors to calibrate the models and the remaining 40 vectors to test.

The standardized data (after each measurement component is subtracted by the sample mean and divided by the sample standard deviation both computed from the first 95 vectors) and their density plots are shown in Figure 4.1. As the figure makes abundantly clear, the marginals are not normal (e.g., the density of x_1 has a long right tail), and so the vector cannot be multivariate normal. It is therefore inadvisable to apply normal-based methods such as Crosier's CUSUM to these data.

Visually, Figure 4.1 does not suggest step changes in the measures over the course of the data set, so it is interesting to see what multivariate methods can indicate. Figure 4.2(a) shows the CUSUM of the first antirank along with its decision interval. Figure 4.2(b) shows the corresponding results of the CUSUM based on the first and last antiranks. In both procedures, k is fixed at 1 and h is computed to be 9.25 and 28.63, respectively, such that the in-control ARL is 200. The first antirank CUSUM (AR1) broke out of its decision interval briefly at three different places, which did not support a step change leading to a change in the first antirank. The first-and-last antirank CUSUM (AR15), however, broke through its decision interval convincingly, immediately into the test set, and remained well above for the remainder of the history. This result therefore shows a marked change, which seems to have occurred at the very beginning of the test data. To verify this change, we calculated the antirank distributions (estimated by the relative frequencies) for both parts of the data (i.e., the first 95 and the remaining 40 vectors). The two distributions are quite different. For example, $P(B_1(i) = 4, B_6(i) = 5)$ equals 0.042 for the first 95 vectors and it is 0.325 for the remaining 40 vectors.

5 Concluding Remarks

We have presented a nonparametric CUSUM procedure for detecting shifts in the location vector of the multivariate measurement of a statistical process. This procedure is effective in detecting a wide range of possible shifts. It is based on the order information between the measurement components and their in-control means as well as the order information among the measurement components. We have shown with numerical examples that it outperforms the procedure based on $\mathbf{A}(i)$ when the shifts are in or close to the equal-component direction, especially when such shifts are far away from the origin. When only a small number of measurement components have shifts in their means, the current procedure does not perform much better than the procedure based on $\mathbf{A}(i)$. Sometimes it performs even a little bit worse than the latter procedure although their difference is generally small. We applied the current method to an aluminum smelter dataset in which the normal distribution assumption is obviously inappropriate. Our method works well in this case. At this moment formulas for in-control and out-of-control run length distributions of the current procedure are still unavailable yet, which needs much further research. But computer algorithms for computing the in-control and out-of-control ARL values are provided in the paper and the corresponding softwares are available from the authors upon request.

Acknowledgments We thank the editor and a referee for many helpful comments and suggestions. This research was partially supported by the National Science Foundation under grant 9803622 and by a Grant-in-Aid of Research, Artistry and Scholarship of University of Minnesota.

References

Crosier, R.B. (1988), "Multivariate generalizations of cumulative sum quality-control schemes," *Technometrics*, **30**, 291-303.

Hawkins, D.M. (1987), "Self-starting cusums for location and scale," *The Statistician*, **36**, 299-315.

Hawkins, D.M. (1991), "Multivariate quality control based on regression-adjusted variables," *Technometrics*, **33**, 61-75.

Hawkins, D.M., and Olwell, D.H. (1998), *Cumulative Sum Charts and Charting for Quality Improvement*, Springer, New York.

Healy, J.D. (1987), "A note on multivariate CUSUM procedures," *Technometrics,* **29**, 409-412.

Hotelling, H. (1947), "Multivariate quality control - illustrated by the air testing of sample bomb sights," in *Techniques of Statistical Analysis* (C. Eisenhart, M.W. Hastay, and W.A. Wallis eds.), McGraw Hill, New York.

Lowry, C.A., Woodall, W.H., Champ, C.W., and Rigdon, S.E. (1992), "A multivariate exponentially weighted moving average control chart," *Technometrics*, **34**, 46-53.

Lucas, J.M., and Saccucci, M.S. (1990), "Exponentially weighted moving average control schemes: properties and enhancements," *Technometrics*, **32**, 1-12.

Mason, R.L., Champ, C.W., Tracy, N.D., Wierda, S.J., and Young, J.C. (1997), "Assessment of multivariate process control techniques," *Journal of Quality Technology*, **29**, 140-143.

Pignatiello, J.J., Jr., and Runger, G.C. (1990), "Comparisons of multivariate CUSUM charts," *Journal of Quality Technology*, **22**, 173-186.

Qiu, P., and Hawkins, D. (2001), "A rank based multivariate CUSUM procedure," *Technometrics*, **43**, 120-132.

Stoumbos, Z.G., Reynolds, M.R., Ryan, T.P., and Woodall, W.H. (2000), "The state of statistical process control as we proceed into the 21st century," *Journal of the American Statistical Association*, **95**, 992-998.

Woodall, W.H. (2000), "Controversies and contradictions in statistical process control," *Journal of Quality Technology*, **32**, 341-350.

Woodall, W.H., and Ncube, M.M. (1985), "Multivariate CUSUM quality-control procedures," *Technometrics*, **27**, 285-292.

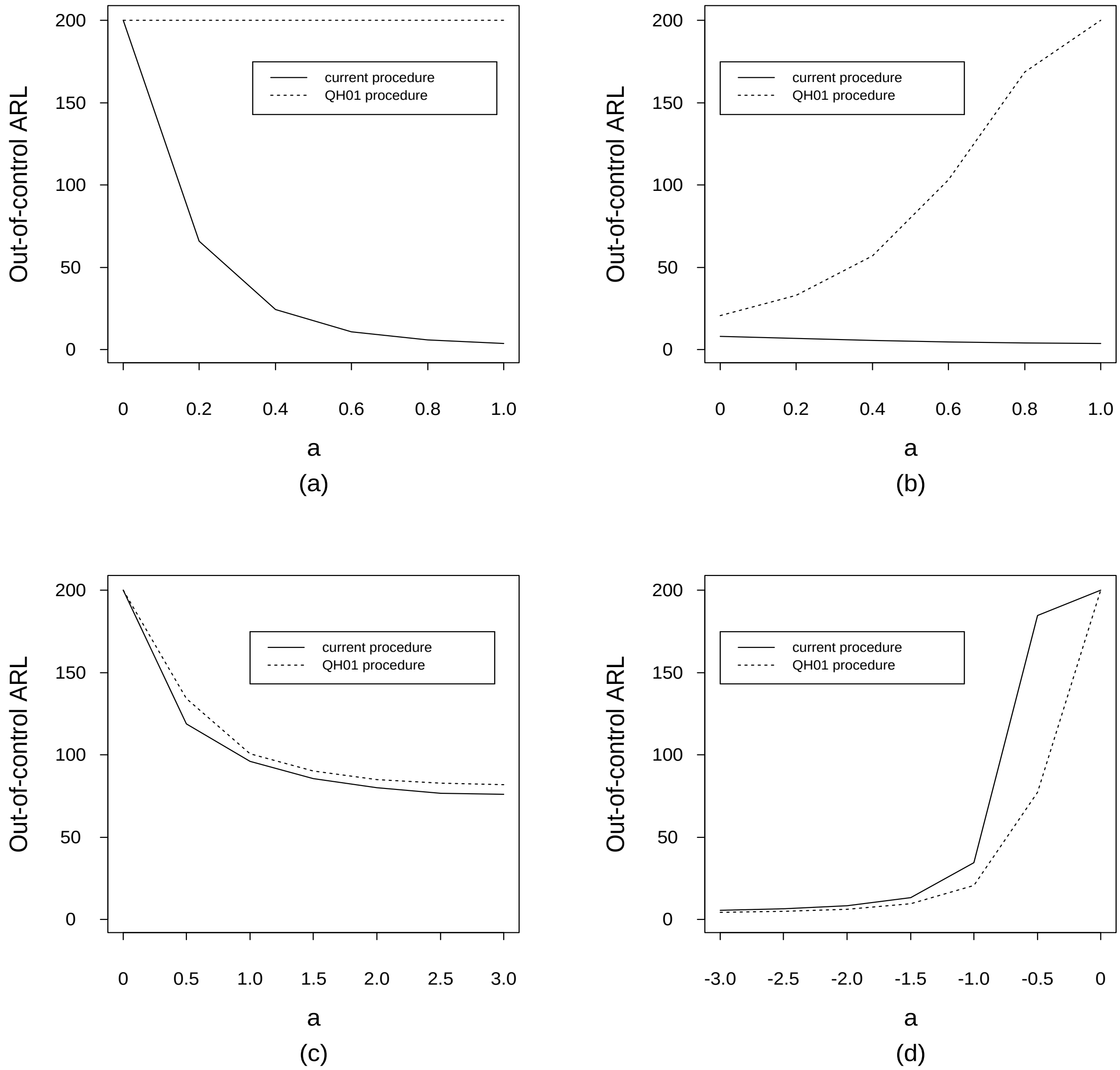

Figure 3.1: The out-of-control ARL values of the current and QH01 procedures. The in-control ARL values of the two procedures are fixed at 200. The constant k and the control limit h are 0.5 and 12.488 for the current procedure, and 0.5 and 8.029 for the QH01 procedure. (a) $(\mu_1, \mu_2, \mu_3, \mu_4)' = (a, a, a, a)'$ and a varies among 0, 0.2, 0.4, 0.6, 0.8 and 1; (b) $(\mu_1, \mu_2, \mu_3, \mu_4)' = (a, 1, 1, 1)'$ and a varies among 0, 0.2, 0.4, 0.6, 0.8 and 1; (c) $(\mu_1, \mu_2, \mu_3, \mu_4)' = (a, 0, 0, 0)'$ and a varies among 0, 0.5, 1.0, 1.5, 2.0, 2.5 and 3; (d) $(\mu_1, \mu_2, \mu_3, \mu_4)' = (a, 0, 0, 0)'$ and a varies among -3.0, -2.5, -2.0, -1.5, -1.0, -0.5 and 0.

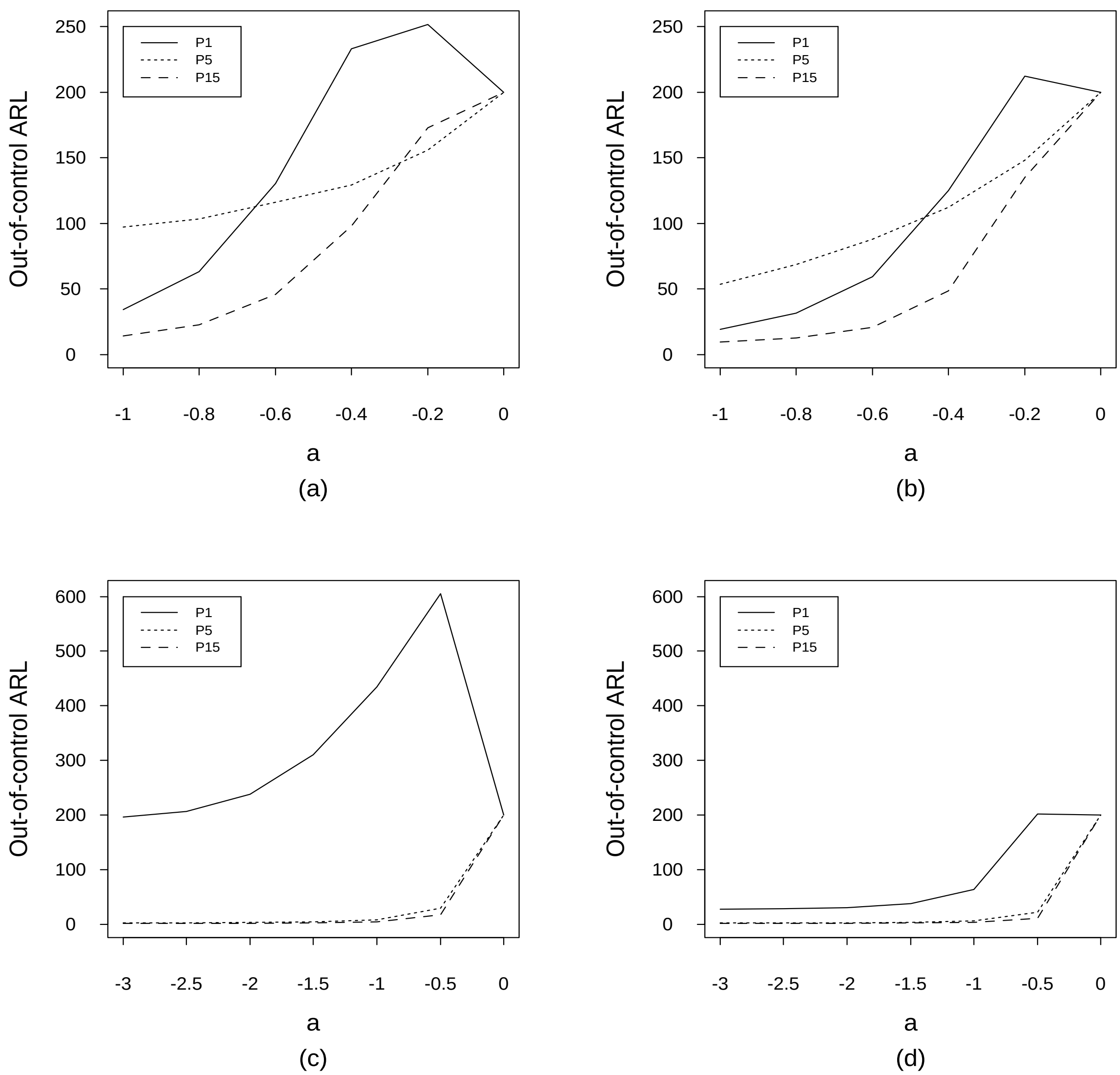

Figure 3.2: The out-of-control ARL values of P1, P5 and P15. The in-control ARL values of the three procedures are all 200. (a)-(b) $(\mu_1, \mu_2, \mu_3, \mu_4)' = (a, 0, 0, 0)'$ and a varies among -1,-0.8, -0.6,-0.4,-0.2 and 0; (c)-(d) $(\mu_1, \mu_2, \mu_3, \mu_4)' = (a, a, a, 0)'$ and a varies among -3,-2.5,-2,-1.5,-1,-0.5 and 0. The value of k is 0.5 in plots (a) and (c) and is 0.25 in plots (b) and (d).

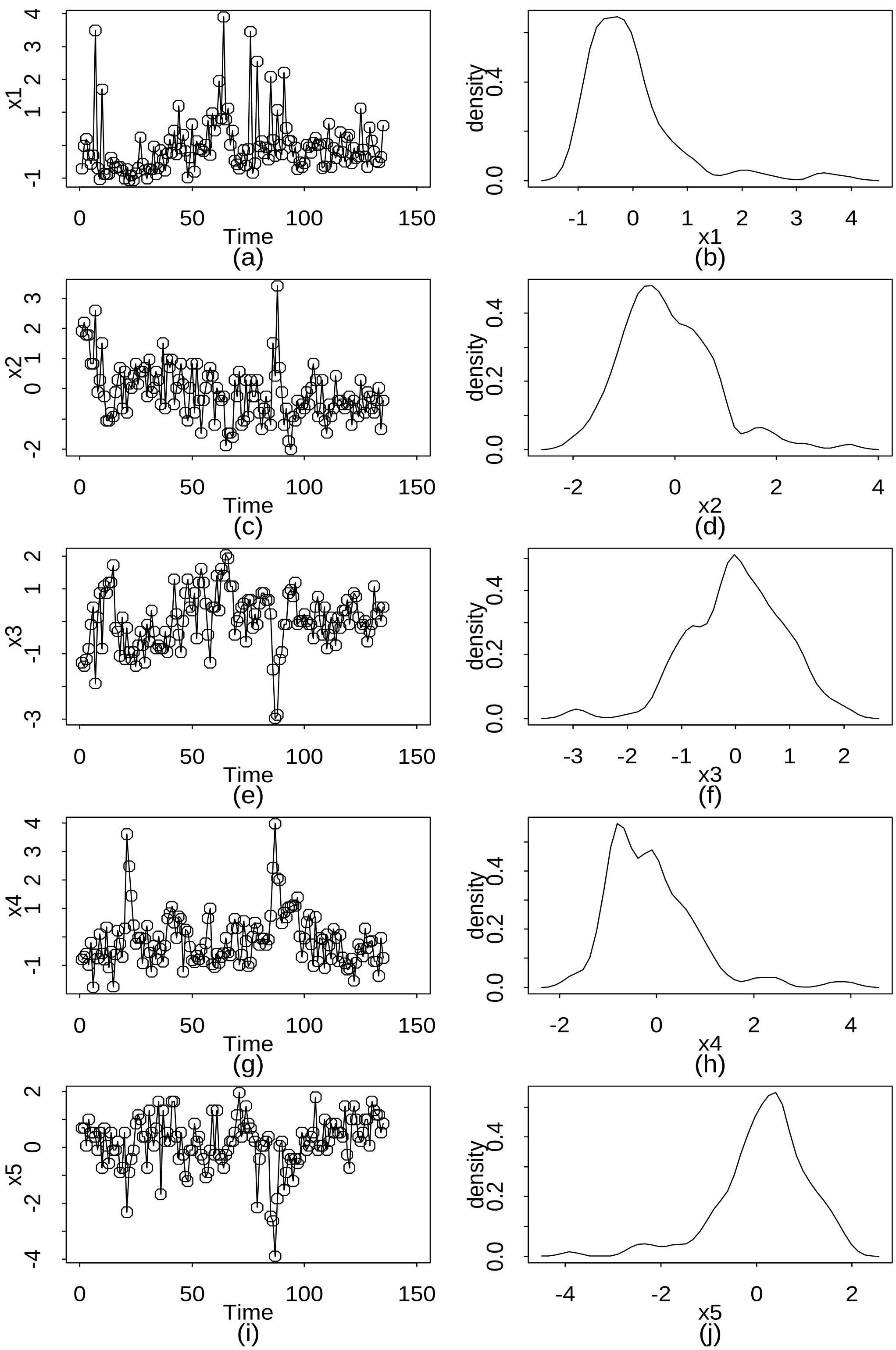

Figure 4.1: The aluminum smelter data. Plots (a), (c), (e), (g) and (i) present the standardized data. Plots (b), (d), (f), (h) and (j) show the corresponding densities.

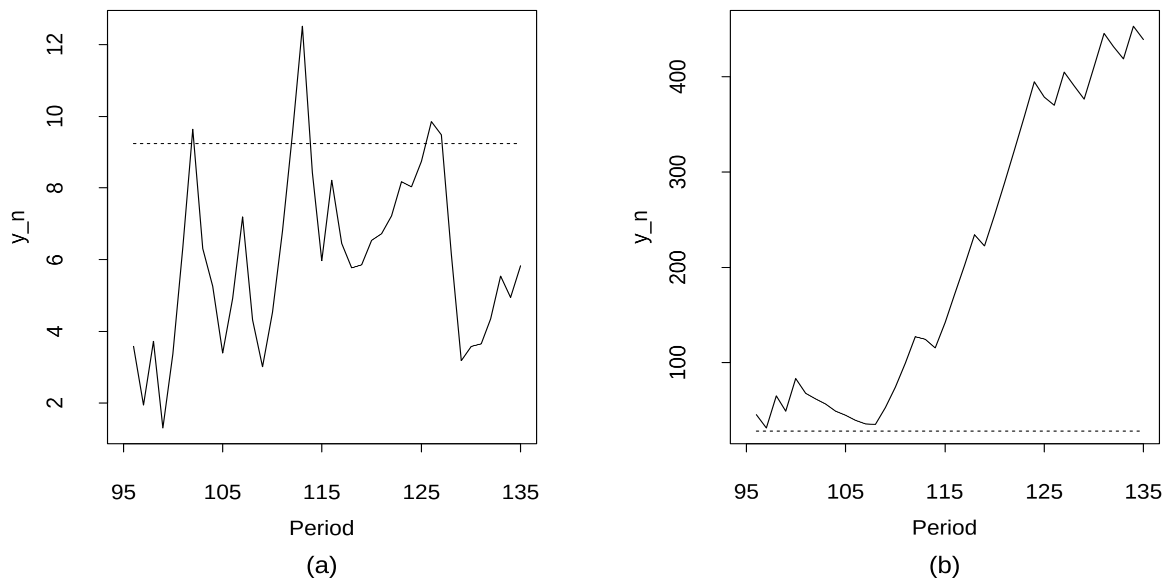

Figure 4.2: (a) The CUSUM criterion y_n of the procedure (2.2)-(2.3) based on the first antirank. (b) The corresponding results of the CUSUM procedure based on the first and last antiranks. The dotted line in each plot indicates the control limit value.