

---

Faculty of Mathematical Sciences



University of Twente  
The Netherlands

---

---

P.O. Box 217  
7500 AE Enschede  
The Netherlands

Phone: +31-53-4893400

Fax: +31-53-4893114

Email: [memo@math.utwente.nl](mailto:memo@math.utwente.nl)

[www.math.utwente.nl/publications](http://www.math.utwente.nl/publications)

---

MEMORANDUM No. 1623

Parametric Control Charts

W. ALBERS, W.C.M. KALLENBERG AND S. NURDIATI

APRIL 2002

ISSN 0169-2690

# Parametric Control Charts

Willem Albers, Wilbert C.M. Kallenberg and Sri Nurdiati  
Faculty of Mathematical Sciences  
University of Twente  
P.O.Box 217, 7500 AE Enschede  
The Netherlands

## Abstract

Standard control charts are based on the assumption that the observations are normally distributed. In practice, normality often fails and consequently the false alarm rate is seriously in error. Application of a nonparametric approach is only possible with many Phase I observations. Since nowadays such very large sample sizes are usually not available, there is need for an intermediate approach by considering a larger parametric model containing the normal family as a submodel. In this paper control limits are presented in such larger parametric models, with emphasis on the so called normal power family. Correction terms are derived, taking into account that the parameters are estimated. Simulation results show that the control limits are accurate, not only in the considered parametric family, but also for common distributions outside the parametric family, thus covering a broad class of distributions.

*Keyword and phrases:* statistical process control, Phase II control limits, second order unbiasedness, normal power family, model error, stochastic error.

*2000 Mathematics Subject Classification:* 62 F 12, 62 P 30, 65 C 05.

---

<sup>0</sup>This research was supported by the Technology Foundation STW, applied science division of NWO and the technology programme of the Ministry of Economic Affairs.

# 1 Introduction

Let  $X$  be a monitoring random variable in a production process which provides an out-of-control signal when  $X$  is larger than a certain control limit. Let  $p$  be the false alarm rate, that is the probability of concluding that the process is out-of-control when the process is in control, and let, for example, the in-control situation be modeled by assuming that  $X$  has a standard normal distribution. Then the out-of-control signal is given if  $X > u_p$ , with  $\Phi(u_p) = 1 - p$ , where  $\Phi$  is the standard normal distribution function.

However, if in fact  $X$  follows a normal distribution with expectation  $\mu$  and variance  $\sigma^2$ , possibly different from 0 and 1,  $P(X > u_p) = \bar{\Phi}(\frac{u_p - \mu}{\sigma})$ , where  $\bar{\Phi} = 1 - \Phi$ . For instance, when  $\mu = 0.5$  and  $\sigma = 1.5$ , we get for  $p = 0.001$  that  $P(X > u_p) = 0.042$  and hence the false alarm rate is 42 times as large as it should be. We conclude that misspecification of the model can lead to enormous errors.

To avoid such errors we take a larger model. Indeed, it is hard to believe that we know  $\mu$  and  $\sigma$ . Therefore, very often it is assumed that  $X$  belongs to the class of normal distributions with unknown parameters  $\mu$  and  $\sigma$ . Based on observations from the past, so called Phase I observations, the parameters  $\mu$  and  $\sigma$  have to be estimated. In Albers and Kallenberg (2000, 2001) it is shown that simply plugging in the estimators of  $\mu$  and  $\sigma$  leads to inaccurate results, unless very large sample sizes are used. Since nowadays short production runs are more and more in demand, such large sample sizes are usually not available. Fortunately, simple but efficient correction terms can be derived, leading to control charts performing according to the required criterion; see Albers and Kallenberg (2000, 2001).

Therefore, as long as normality describes the behavior of  $X$  rather well, the corrected control charts can be applied in practice. But the same problem of misspecification may arise here. It is well known that the probability of incorrectly producing an out-of-control signal may be seriously in error when the distributional form of the observations differs from normality, see e.g. Chan et al. (1988), Pappanastos and Adams (1996), table 7 on page 222.

Basically, the problem is that the normal approximation may be fair for the central part of the distribution, but produces large relative errors in the tails. And, in view of the small values of  $p$  typically used, the tails are what we are dealing with. Again a larger model is needed, providing more flexibility.

In order to cover a broad class of distributions we consider a large parametric family, starting for instance with normality and adding one or more parameters. Of course, one may object that the larger parametric family again does not cover all distributions and that a nonparametric approach should be taken. As typically the 0.999-quantile should be estimated, it is clear that with, say, 100 observations in Phase I such a quantile cannot be estimated nonparametrically. Therefore, we look for classes of distributions which are on the one hand sufficiently rich, but on the other hand not that large that estimation is impossible or too inaccurate.

As a consequence, assumptions should be made about the relation between the behavior of  $X$  in the far tail (where we should estimate an appropriate quantile) and the behavior of  $X$  "somewhat more to the middle", where we have observations and can do the estimation. The link between these parts of the distribution is given by the parametric model. If only such "minimal" relations are given, it is natural to restrict the estimators of the additional parameters (other than those for location and scale) to the fraction of the order statistics corresponding to the part of the distribution which is modeled. If the whole distribution is modeled, we will use all observations to determine the estimators.

In general terms such an intermediate approach between the classical assumption of normality and the nonparametric estimation of upper quantiles very far in the tail of the distribution seems to give opportunity to a lot of parametric models. Indeed, there are many possibilities,

like random or deterministic mixtures, particular extensions of the normal family with one or more parameters such as the normal power family, orthonormal families etc.

However, it turns out that several standard extensions of the family of normal distributions lead to very substantially more complicated control limits. The reason for this phenomenon is the large number of side conditions involved. The problem starts with the quantiles in the parametric family, prominently emerging in the control limits. They should be not too complicated functions of the parameters, because the next step is plugging in estimators of the parameters into the control limit and hence also into these quantiles. If this were the end of the process, complicated forms were not that problematic, because numerical evaluation would suffice.

Unfortunately, due to the estimation serious stochastic error comes in. In order to correct for it a purely numerical solution is impossible and therefore a second order asymptotic approximation is applied. Such a strategy has been successfully used in many statistical problems, where, as in the present situation, numerical work (alone) cannot give sufficient insight. To derive appropriate correction terms both the quantiles and the estimators of the parameters should be analytically manageable. Moreover, a rather delicate analysis is needed, because first order asymptotics are not sufficiently accurate. (Note that the false alarm rate is about 0.001 and hence very small quantities are involved.)

It turns out that the normal power family provides a tractable model, although even in that case the derivation of the correction terms is no sinecure and more importantly, the final result is still complicated. For practical implementation of the method, a final numerical approximation of the corrected control limit is presented, which behaves rather well and can be used straightforwardly.

Using the normal power family gives more flexibility and hence an improvement over simply using normality and ignoring the well-known facts that in practice normality often fails and that this causes big errors in the false alarm rate. On the other hand, it is not our claim that the normal power family always is the "right" model. But firstly, if control limits based on normality are applied, this should implicitly mean that the distribution is approximately normal, and in that case the normal power family is certainly appropriate, since normality is a submodel of the normal power family. Secondly, by the extension to the normal power family many other commonly used distributions show up, which are covered sufficiently well by members of the normal power family. Thirdly, the normal power family is not so large that accurate estimation is only possible with huge sample sizes as in the nonparametric approach.

In Section 2 a general exposition of the problem is given with a discussion on model error due to misspecification and stochastic error implied by estimation. Section 3 presents special classes of models, thus covering a lot of common distributions, and the corresponding model error when the observations are in such a class, while taking normality as the supposed model. It can be concluded that this model error can be very large, thus showing the need for a larger parametric model.

Having reduced the model error by taking a larger parametric model, the next step is to avoid a large stochastic error due to estimation of the parameters. In Section 4 estimators are presented and criteria are given for comparison of the behavior of the false alarm rate, which due to the estimation is no longer a number, but a random variable. Starting from Section 4 emphasis is on the normal power family. The required correction terms are derived in Section 5, thus reducing the stochastic error.

A simulation study is performed to see how well the asymptotic results come true for finite sample sizes. It turns out that the corrected control limits work very well, not only when sampled from distributions belonging to the normal power family, but also for other distributions like random and deterministic mixtures. This provides a great improvement of the false alarm rate

based on the normality assumption, which may fail seriously. The loss when really having normal observations is small. The simulation results are presented in Section 6.

In view of our theoretical and numerical results we recommend to apply the corrected control limit based on the normal power family, using as estimator of the extra shape parameter the estimator linking the very far tail to the behavior somewhat more to the middle, invoking the 0.95 and 0.75 sample quantiles. This recommendation is presented in detail in Section 7, where also a discussion is given of the results in view of the main questions stated in Section 2.

## 2 The main questions

Let  $X_1, \dots, X_n, X_{n+1}$  be i.i.d. random variables (r.v.'s) with distribution function  $F$ , that is, we consider the in-control situation. The r.v.'s  $X_1, \dots, X_n$  are the data from Phase I on which the estimators of the unknown parameters are based and  $X_{n+1}$  is the monitoring characteristic. The monitoring r.v. may be based on  $m$  observations, but here we consider the situation  $m = 1$  of individual measurements to avoid additional complications, thus facilitating the explanation of the basic arguments. The case  $m > 1$  will be treated in a next paper.

The true, but unknown distribution function  $F$  is modeled by a parametric family  $\{G_\theta : \theta \in \Theta\}$ . Apart from this general parametric model, we consider also the *restricted model*, where the parameter space is restricted to a subset  $\Theta_0$ . Hopefully  $F$  equals  $G_\theta$  for some  $\theta \in \Theta$ , or even for some  $\theta \in \Theta_0$ , but this is not necessarily true. For any distribution function  $H$  we write  $\bar{H} = 1 - H$  and  $H^{-1}$  and  $\bar{H}^{-1}$  for the respective inverse functions.

If  $F$  equals  $G_\theta$  and  $\theta$  is known, the control limit equals  $\bar{G}_\theta^{-1}(p)$ . Often  $F$  is unknown and two problems arise: (i)  $F$  may be (slightly) outside the parametric family  $\{G_\theta : \theta \in \Theta\}$  (or  $\{G_\theta : \theta \in \Theta_0\}$  if the supposed model is the restricted model) and (ii)  $\theta$  is unknown. This leads to two kinds of errors, the *model error* and the *stochastic error*. The latter is due to estimation of  $\theta$  in the control limit by its estimator  $\hat{\theta}$ , as is clearly seen by splitting up the total error:

$$P = P\left(X_{n+1} > \bar{G}_{\hat{\theta}}^{-1}(p)\right) = \bar{F}\left(\bar{G}_{\hat{\theta}}^{-1}(p)\right) = p + ME + SE \quad (2.1)$$

with the model error  $ME$  given by

$$ME = \bar{F}\left(\bar{G}_{\hat{\theta}}^{-1}(p)\right) - p = \bar{F}\left(\bar{G}_\theta^{-1}(p)\right) - \bar{F}\left(\bar{F}^{-1}(p)\right) \quad (2.2)$$

for some suitably chosen point  $\theta \in \Theta$  (or  $\theta \in \Theta_0$  if the supposed model is the restricted model), and the stochastic error  $SE$  defined by

$$SE = \bar{F}\left(\bar{G}_{\hat{\theta}}^{-1}(p)\right) - \bar{F}\left(\bar{G}_\theta^{-1}(p)\right). \quad (2.3)$$

(On how  $\theta$  is "suitably chosen" we come back below.)

The larger the parametric family, the smaller in general the  $ME$ , but the more difficult the estimation problem (more parameters involved) and hence the larger  $SE$ .

We want to control both  $ME$  and  $SE$ . To establish this, we consider three cases:

1. We are in the *restricted model*. A typical example is that our observations are normally distributed.
2. We are in the *general parametric model*. This model contains the restricted model. A typical example is an extension of the normal family with a few extra parameters. The general parametric model is denoted by  $\{G_\theta : \theta \in \Theta\}$ . Here  $\theta = (\theta_1, \dots, \theta_k)$ , say, while in the restricted model only  $\theta_1, \dots, \theta_l$  vary, for some  $0 \leq l < k$ , and  $\theta_{l+1}, \dots, \theta_k$  are fixed.

Without loss of generality, we let these fixed values be 0. Hence the restricted model is presented by  $\{G_\theta : \theta = (\theta_1, \dots, \theta_l, 0, \dots, 0) \in \Theta\}$ . In other words each  $\theta \in \Theta_0$  is of the form  $\theta = (\theta_1, \dots, \theta_l, 0, \dots, 0)$ .

3. We are outside the general parametric model.

Of course, the model error  $ME$  depends on the true model and the supposed model.

**Example 2.1** Consider the very simple situation of the standard normal distribution as the restricted model. Then  $ME = \bar{F}(u_p) - p$  in the restricted model. Let in the general parametric model the quantile function be given by the mixture  $(1 - \theta)\Phi^{-1} + \theta K^{-1}$  for some distribution function  $K$ . When we estimate the 0.75-quantile, the suitably chosen point  $\theta \in \Theta$  is determined by  $(1 - \theta)\Phi^{-1}(0.75) + \theta K^{-1}(0.75) = F^{-1}(0.75)$  and  $ME = \bar{F}\left(\bar{G}_\theta^{-1}(p)\right) - p$  in the general model. If the restricted model is the true model ( $F = \Phi$ ), then both  $ME$ 's are 0 (note that the suitably chosen  $\theta$  equals 0, implying  $G_\theta = \Phi$ ). When the general model is the true model ( $F = G_\theta$  for some  $\theta$ ) the  $ME$  in the general model equals 0, but the  $ME$  in the restricted model in general does not equal 0.

Given a true model, we speak of the *restricted*  $ME$  and  $SE$  if the supposed model is the restricted model and we speak of the *general*  $ME$  and  $SE$  if the supposed model is the general model. The "suitably chosen point  $\theta \in \Theta$ " can be characterized as follows. The parameter  $\theta_i, i = 1, \dots, k$ , is assumed to be a functional of  $F$ , given by  $\theta_i = T_i(F)$ . Tacitly it is assumed that the parameterization is such that  $\theta_i = T_i(G_\theta)$ . In this way the parameter has a meaning also outside the restricted or general model and this meaning coincides with the concept of the parameter within these models. For instance, the functional of  $F$  may correspond to the 0.75-quantile of  $F$ . In the restricted and general model the 0.75-quantile then leads to the suitable parameter values (see also Example 2.1).

In view of the notation of the restricted model it is implied that  $T_i(F) = 0$  for  $i = l+1, \dots, k$ , when  $F$  belongs to the restricted model.

Several functionals may be associated with the parameter  $\theta_i, i = 1, \dots, k$ . The choice is determined by the estimation of the parameter as is exemplified in the following example.

**Example 2.2** Consider the random mixture model  $X = \mu + \sigma\{(1 - W)Z_0 + WZ_1\}$ , where  $W$  is independent of  $Z_0$  and  $Z_1$ ,  $P(W = 1) = 1 - P(W = 0) = \gamma$  and where  $Z_0$  and  $Z_1$  have mean zero, variance one and distribution functions  $K_0$  and  $K_1$ , respectively. The parameter  $\theta$  is denoted by  $(\mu, \sigma, \gamma)$ . Let  $K_0 = \Phi$  and let  $K_1$  be the distribution function of the standardized Student-distribution with 6 degrees of freedom. The parameters  $\mu$  and  $\sigma$  are estimated by the usual location and scale estimators  $\bar{X} = n^{-1} \sum X_i$  and  $S = \sqrt{S^2}$  with  $S^2 = (n - 1)^{-1} \sum (X_i - \bar{X})^2$ . The related functionals are  $\mu = \int x dF(x)$  and  $\sigma = \sqrt{\int (x - \mu)^2 dF(x)}$ .

To estimate  $\gamma$  we may use the moment estimator based on the fourth moment. As the fourth moment of  $Z_1$  equals 6, this leads to the estimator  $\hat{\gamma}_1$  corresponding to the functional  $\gamma_1$ , where

$$\hat{\gamma}_1 = \frac{n^{-1} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{S}\right)^4 - 3}{3}, \gamma_1 = \frac{\int \left(\frac{x - \mu}{\sigma}\right)^4 dF(x) - 3}{3}.$$

Writing  $X_{m:n}$  for the  $m^{th}$  order statistic of  $X_1, \dots, X_n$ , another estimator  $\hat{\gamma}_2$ , with corresponding functional  $\gamma_2$ , is obtained by equating a sample 90%-quantile to the 90%-quantile of

the distribution, giving, with  $m=[0.9n]$ , where  $[x]$  denotes the entier of  $x$ ,

$$\hat{\gamma}_2 = \frac{\frac{m}{n} - \Phi\left(\frac{X_{m:n} - \bar{X}}{S}\right)}{K_1\left(\frac{X_{m:n} - \bar{X}}{S}\right) - \Phi\left(\frac{X_{m:n} - \bar{X}}{S}\right)}, \gamma_2 = \frac{0.9 - \Phi\left(\frac{F^{-1}(0.9) - \mu}{\sigma}\right)}{K_1\left(\frac{F^{-1}(0.9) - \mu}{\sigma}\right) - \Phi\left(\frac{F^{-1}(0.9) - \mu}{\sigma}\right)}.$$

The idea is to relate the (unknown) true distribution  $F$  to a member of the general parametric model by fitting several characteristics of  $F$  to the nominated member of the family. Hopefully, this results in similar (extreme) quantiles of  $F$  and the chosen member of the family as well. The more characteristics are fitted, the better presumably the approximation in the far tail and hence the smaller the  $ME$ , but, on the other hand, the more difficult the estimation of those parameters, that is the larger the  $SE$ .

Moreover, not only the number of parameters is important, but also *which* characteristics are used. Since the estimator determines the suitably chosen point in the parameter space, the model error for distributions outside the supposed model depends on the estimator. In other words, one estimator will give a better fit w.r.t the required extreme  $p$ -quantile in the parametric model than another. On the other hand, a lower  $ME$  may involve a larger  $SE$ . For instance, the perfect fit in terms of  $ME$  is obtained by taking  $\theta = \theta(F)$ , such that  $\bar{G}_\theta^{-1}(p) = \bar{F}^{-1}(p)$ . However, such  $\theta(F)$  cannot be estimated with, say, 100 observations. (In fact, with this estimator we would not use the parametric family at all: it boils down to the nonparametric approach.) A similar reasoning holds when using less extreme quantiles. For instance, using the preceding quantile-approach to estimate  $\gamma$ , there is with  $n = 100$  observations no difference between the estimators using the 0.99-quantile or the 0.999-quantile. This shows that for some characteristics the  $SE$  will be very large.

Furthermore, additional functionals should be such that really different characteristics are involved.

The main questions treated in the paper are the following.

1. When the *restricted* model is true, that is  $F = G_\theta$  with  $\theta = (\theta_1, \dots, \theta_l, 0, \dots, 0)$ , both the restricted and the general  $ME$  are equal to 0. Hence, we only have to control  $SE$ . As a rule, the general  $SE$  will be larger than the restricted  $SE$ . How large is this difference? Can the restricted and general  $SE$  be reduced by (simple) corrections? How large are the corrected general  $SE$  and the corrected restricted  $SE$ , when the restricted model is true?
2. When the general model is true, the general  $ME$  equals 0, but as a rule the restricted  $ME$  does not. How bad can the restricted  $ME$  be? How does this balance with the larger general  $SE$ ? Can the restricted and general  $SE$  be reduced by (simple) corrections? How large are the corrected restricted  $SE$  and the corrected general  $SE$ , when the general model is true?
3. When we are outside the general model, the  $ME$ 's can in principle be very large if we are far away from the general model. In the present paper we concentrate on the parametric approach and hence we consider only situations outside the general model where the general  $ME$  is not too big. In a forthcoming paper the nonparametric approach will be considered and then also departures farther away from the general model are considered. But, as already observed in the introduction, the nonparametric approach only makes sense if some flexibility is allowed w.r.t.  $n$  and/or  $p$ : for e.g.  $n = 100$  and  $p = 0.001$  as given quantities, nothing can be done.

How large are the uncorrected and corrected general  $SE$  and the uncorrected and corrected



restricted  $SE$ , when we are outside the general model and the general  $ME$  is not too big? How is the total error when applying the corrected control limit with as supposed model the general one and how does this compare with the total error when applying the corrected control limit with as supposed model the restricted one?

### 3 Suitable models

The following classes of models are candidates for the (restricted and) general parametric model. Often, the restricted model will be the normal family, but adaptation to another choice, as for instance the family of exponential distributions, can be easily made.

We will take location and scale parameters  $\mu$  and  $\sigma$ . In the restricted model these are the only parameters. In terms of  $\theta = (\theta_1, \dots, \theta_l, \theta_{l+1}, \dots, \theta_k)$  this means that  $l = 2, \theta_1 = \mu$  and  $\theta_2 = \sigma$ . The additional parameters  $\theta_3, \dots, \theta_k$  are denoted by  $\gamma$ . In this paper we will restrict mainly to  $k = 3$ , that is,  $\gamma$  is a one-dimensional parameter, but see also Remark 3.1. The distribution function  $G_\theta$  is written as  $G_\theta(x) = K_\gamma\left(\frac{x-\mu}{\sigma}\right)$ . Hence, the (uncorrected) control limit has the following form

$$\hat{\mu} + \hat{\sigma} \overline{K_{\hat{\gamma}}}^{-1}(p).$$

The models are defined in such a way that varying tail behavior can be described. Especially, heavier tails than those of the normal distribution are of interest. In terms of high upper quantiles this means larger values than the normal upper quantiles.

The conditions for an appropriate general model are rather comprehensive. Therefore, several classical ways of extending the normal model turn out to cause (technical) difficulties.

In order to make the necessary (bias) corrections we need to evaluate (first and second) moments of  $\hat{\mu} + \hat{\sigma} \overline{K_{\hat{\gamma}}}^{-1}(p) - \left(\mu + \sigma \overline{K_\gamma}^{-1}(p)\right)$  up to high precision. This implies that either  $\overline{K_\gamma}^{-1}(p)$  should be analytically tractable as function of  $\gamma$ , or we should have a very precise approximation of  $\overline{K_\gamma}^{-1}(p)$  by a simple function of  $\gamma$ . A particular problem arises in the well known random and deterministic mixture models, because negative values of  $\gamma$  are meaningless there. Hence  $\hat{\gamma}$  is restricted to nonnegative values, which can often only be achieved by adding a suitable indicator function to the definition of the estimator. Due to the required precision this causes great (technical) problems, aggravated by the fact that  $\hat{\gamma}$  (and also the indicator function) is tied up with  $\hat{\mu}$  and  $\hat{\sigma}$ . Apart from that, the truncation of negative values also introduces an artificial bias near  $\gamma = 0$ , which is also rather unattractive.

#### (i) *Random Mixture*

In the family of normal distributions the distribution function is given by  $K\left(\frac{x-\mu}{\sigma}\right)$  with  $K$  fixed and given by  $K = \Phi$ . In the random mixture model we make two extensions:  $K$  may be replaced by another (fixed) distribution and, more importantly, an additional parameter is added, replacing the fixed  $K$  by  $K_\gamma = (1 - \gamma)K_0 + \gamma K_1$  with  $K_0$  and  $K_1$  (fixed) distribution functions with corresponding expectation 0 and variance 1. Under the restricted model, we have  $\gamma = 0$  and hence the distribution  $K_0$ . The r.v.  $X$  can be written as

$$X = \mu + \sigma \{(1 - W)Z_0 + WZ_1\},$$

where  $W$  is independent of  $Z_0$  and  $Z_1$ ,  $P(W = 1) = 1 - P(W = 0) = \gamma$  and where  $Z_0$  and  $Z_1$  have distribution functions  $K_0$  and  $K_1$ , respectively. Often we will consider  $K_0 = \Phi$ , also see Example 2.2.

(ii) *Deterministic Mixture*

Since we are focussed on quantiles here, it seems appropriate to consider mixtures of quantiles. In the family of normal distributions the quantile is given by  $\mu + \sigma K^{-1}(t)$  with  $K = \Phi$ . We replace the fixed  $K^{-1}$  by  $K_\gamma^{-1} = c(\gamma) \{(1 - \gamma) K_0^{-1} + \gamma K_1^{-1}\}$  with  $K_0$  and  $K_1$  distribution functions with corresponding expectation 0 and variance 1 and where  $c(\gamma)$  is a normalizing factor such that the variance corresponding to  $K_\gamma$  equals 1. The r.v.  $X$  can be written as

$$X = \mu + \sigma c(\gamma) \{(1 - \gamma) Z_0 + \gamma Z_1\}$$

with  $Z_0 = K_0^{-1}(U)$ ,  $Z_1 = K_1^{-1}(U)$  and  $U$  a r.v. with a uniform distribution on  $(0, 1)$ . Note that  $Z_0$  and  $Z_1$  have distribution functions  $K_0$  and  $K_1$ , respectively, but that they are anything but independent; in fact these r.v.'s are comonotone. It is easily seen that the normalizing constant  $c(\gamma)$  is given by

$$c(\gamma) = \left\{ (1 - \gamma)^2 + \gamma^2 + 2\gamma(1 - \gamma)\rho \right\}^{-1/2} \quad \text{with } \rho = \int_0^1 K_0^{-1}(t)K_1^{-1}(t)dt. \quad (3.1)$$

Because of the technical difficulties with the random and deterministic mixture models, we deal in this paper mainly with the following model.

(iii) *Normal Power family*

In the family of normal distributions the quantile is given by  $\mu + \sigma K^{-1}(t)$  with  $K = \Phi$ . We replace  $K^{-1}$  by  $K_\gamma^{-1}(t) = c(\gamma) |\Phi^{-1}(t)|^{1+\gamma} \text{sign}(\Phi^{-1}(t))$ , where  $\gamma > -1$  and where  $c(\gamma)$  is a normalizing constant given by  $c(\gamma) = \left\{ E|Z|^{2(1+\gamma)} \right\}^{-1/2} = \pi^{1/4} 2^{-(1+\gamma)/2} \Gamma(\gamma + \frac{3}{2})^{-1/2}$  with  $Z$  a r.v. with a standard normal distribution. For  $\gamma > -1$  let  $Z_\gamma = c(\gamma) |Z|^{1+\gamma} \text{sign}(Z)$ , then

$$X = \mu + \sigma Z_\gamma.$$

We mention two further well known models. Also in these cases analytic evaluation of the estimators of the additional parameter(s) up to the required precision is difficult.

(iv) *Tukey's  $\lambda$ -family*

The r.v  $X$  is given by

$$X = \mu + \sigma c(\lambda) \left\{ U^\lambda - (1 - U)^\lambda \right\},$$

where  $U$  has a uniform distribution on  $(0, 1)$  and  $c(\lambda)$  is a normalizing constant such that  $\frac{X-\mu}{\sigma}$  has variance 1, that is

$$c(\lambda) = \left[ \frac{2}{2\lambda + 1} \left\{ 1 - \frac{\lambda}{2} B(\lambda, \lambda) \right\} \right]^{-1/2},$$

in which  $B$  is the beta-function. For  $\lambda = 0$ , we define  $X$  in a continuous way, leading to the logistic distribution for  $\frac{X-\mu}{\sigma}$ . The choice  $\lambda = 0.14$  gives a distribution close to the standard normal distribution, especially for upper  $t$ -quantiles with  $t$  from 0.2 to 0.005, cf. also Chan et al. (1988) page 118. Therefore, we take this distribution as the restricted model. Introducing the parameter  $\gamma$ , given by

$$\gamma = 0.14 - \lambda,$$

we refer to  $K_\gamma$  as the distribution function corresponding to the r.v.

$$c(0.14 - \gamma) \left\{ U^{0.14 - \gamma} - (1 - U)^{0.14 - \gamma} \right\}.$$

(v) *Orthonormal family*

In the normal family  $\Phi\left(\frac{X - \mu}{\sigma}\right)$  has a uniform distribution. Starting from a uniform distribution an orthonormal family of densities w.r.t Lebesgue measure on  $(0, 1)$  is defined by

$$f(y, \gamma) = c^*(\gamma) \exp \left\{ \sum_{j=1}^k \gamma_j \pi_j(y) \right\},$$

where  $c^*(\gamma)$  is a normalizing constant such that the integral of  $f$  equals 1, and where  $\pi_j$  is the  $j^{\text{th}}$  Legendre polynomial on  $(0, 1)$ . Let  $Y$  be a r.v. having density  $f(y, \gamma)$  and let  $E(\gamma)$  and  $c(\gamma)^{-1}$  be the expectation and standard deviation of  $\Phi^{-1}(Y)$ . The r.v.  $X$  is given by

$$X = \mu + \sigma c(\gamma) \left\{ \Phi^{-1}(Y) - E(\gamma) \right\}.$$

**Remark 3.1** Model (v) is the only one which explicitly offers the possibility for more than one additional parameter beyond  $\mu$  and  $\sigma$ . However, if desired, the mixtures in (i) and (ii) obviously can be taken for more than just two  $K_i$ 's, while e.g. in (iv) a generalized version of Tukey's  $\lambda$ -family can be used, see, for instance, Hušková (1988).

### The Model Error for the suitable models

The restricted model is of the form

$$X = \mu + \sigma Z$$

with  $Z$  having a fixed distribution. Often the r.v.  $Z$  has a standard normal distribution, leading to the family of normal distributions as the restricted model.

In the general parametric model we have

$$X = \mu + \sigma Z_\gamma$$

with  $Z_\gamma$  having distribution function  $K_\gamma$ , say.

In view of (2.2), the model error in the situation that the supposed model is the restricted one and the true distribution is from the general model is therefore given by

$$ME = \bar{K}_\gamma \left( \bar{K}_0^{-1}(p) \right) - p. \quad (3.2)$$

We elaborate  $ME$  for the various models. Some numerical evaluations of the model error, also in the situation where the supposed model is the general parametric model and the true distribution is outside this model, are presented in Section 6.

(i) *Random Mixture*

It is easily seen that (3.2) equals

$$ME = \gamma \left\{ P \left( Z_1 > \bar{K}_0^{-1}(p) \right) - p \right\}.$$

For instance, if the restricted model is the family of normal distributions (that is  $Z_0$  has a standard normal distribution),  $p = 0.001$  and  $Z_1$  has a standardized Student distribution with 6 degrees of freedom, then  $ME = 0.00356\gamma$ .

(ii) *Deterministic Mixture*

Here (3.2) can be rewritten as

$$ME = t - p \quad \text{with } t \text{ given by } c(\gamma) \left\{ (1 - \gamma) \bar{K}_0^{-1}(t) + \gamma \bar{K}_1^{-1}(t) \right\} = \bar{K}_0^{-1}(p).$$

For instance, let the restricted model be the family of normal distributions and let  $K_1$  correspond to the standardized Student distribution with 6 degrees of freedom. Taking  $p = 0.001$ , we get  $\rho = 0.99$  in (3.1) and hence for  $\gamma = \frac{1}{2}$  we have to solve

$$\bar{K}_0^{-1}(t) + \bar{K}_1^{-1}(t) = 6.165$$

implying  $t = 0.00292$  and hence  $ME = 0.00192$ . For  $\gamma = 1$  we obtain  $ME = 0.00356$ .

(iii) *Normal Power family*

In this model (3.2) reduces to

$$ME = \bar{\Phi} \left( \left( \frac{u_p}{c(\gamma)} \right)^{\frac{1}{1+\gamma}} \right) - p.$$

For  $p = 0.001$  and  $\gamma = -\frac{1}{4}$  we get  $c(-\frac{1}{4}) = 1.0783$  and hence  $ME = -0.00098$ . For  $p = 0.001$  and  $\gamma = \frac{1}{2}$  we get  $c(\frac{1}{2}) = \frac{1}{2} \sqrt[4]{2\pi}$  and hence  $ME = 0.00558$ , while for  $p = 0.001$  and  $\gamma = 1$  we have  $c(1) = 3^{-1/2}$  and thus  $ME = 0.00935$ .

(iv) *Tukey's  $\lambda$ -family*

According to the definition of the parameter  $\gamma$  (3.2) gives

$$ME = P \left( U^{0.14-\gamma} - (1-U)^{0.14-\gamma} > \frac{\nu_p}{c(0.14-\gamma)} \right) - p,$$

where  $\nu_p$  is defined by

$$P \left( U^{0.14} - (1-U)^{0.14} > \frac{\nu_p}{c(0.14)} \right) = p,$$

or, more explicitly,  $\nu_p = c(0.14) \times \left\{ (1-p)^{0.14} - p^{0.14} \right\}$ . For  $p = 0.001$  we get  $\nu_{0.001} = 3.0469$ .

For  $\lambda = 1$  (and hence  $\gamma = -0.86$ ) we get the uniform distribution on  $(-1, 1)$  leading to

$$ME = P \left( 2U - 1 > \frac{\nu_p}{\sqrt{3}} \right) - p = \max \left\{ \frac{1}{2} - \frac{\nu_p}{2\sqrt{3}}, 0 \right\} - p.$$

Note that  $ME = -p$  if  $\nu_p > \sqrt{3}$ . Hence, for  $p = 0.001$  we have  $ME = -0.001$ . The distribution corresponding to  $\lambda = 0$  is the (standardized) logistic distribution. In that case we obtain

$$ME = P \left( \frac{\sqrt{3}}{\pi} \log \frac{U}{1-U} > \nu_p \right) - p = \left( 1 + e^{\nu_p \pi / \sqrt{3}} \right)^{-1} - p.$$

For  $p = 0.001$  this results in  $ME = 0.00296$ .

(v) *Orthonormal family*

The model error can be written as

$$ME = P \left( Y > \Phi \left( u_p c(\gamma)^{-1} + E(\gamma) \right) \right) - p$$

with  $Y$  having density  $f(y, \gamma)$ . Let  $p = 0.001$ . Taking  $k = 1$  and  $\gamma = 0.3$  we obtain  $E(\gamma) = 0.2886$ ,  $c(\gamma) = 1.0179$  and  $ME = -0.00029$ . For  $k = 2$  and  $\gamma = (-0.1, -0.4)$  we get  $E(\gamma) = -0.0634$ ,  $c(\gamma) = 1.2872$  and  $ME = 0.00218$ . Choosing  $k = 3$  and  $\gamma = (-0.1, -0.1, 0.1)$  we have  $E(\gamma) = -0.0787$ ,  $c(\gamma) = 1.0696$  and  $ME = 0.00113$ .

## Conclusion

From the considered models it is seen that the model error, assuming as restricted model the family of normal distributions, can be quite large. Often the model error is several factors larger than the prescribed  $p$ . Therefore, if normality fails, there is a need for a larger model, thus reducing the model error.

**Remark 3.2** From the calculations of the model error it is seen that the model error can be negative, implying a lower false alarm rate. From the point of view of the false alarm rate this looks nice. However, this will certainly have harmful consequences for the out-of-control behavior. For instance, when  $\gamma = -\frac{1}{4}$  in the normal power family,  $ME = -0.00098$ . To illustrate what this might imply for the out-of-control behavior, we consider the more simple situation of a control chart for a standard normal distribution with  $p = 0.001$  and  $p = 0.001 - 0.00098 = 0.00002$ . Then the expected run length to detect a shift 2 equals 7.3 for  $p = 0.001$  and no fewer than 57.0 for  $p = 0.00002$ . Hence, both positive and negative model errors should be reduced.

## 4 Estimators and criteria for comparison

The parameters  $\mu$  and  $\sigma$  will be estimated here by the usual location and scale estimators  $\bar{X} = n^{-1} \sum X_i$  and  $S = \sqrt{S^2}$  with  $S^2 = (n-1)^{-1} \sum (X_i - \bar{X})^2$ . We also sometimes write  $\hat{\mu}$  and  $\hat{\sigma}$  for  $\bar{X}$  and  $S$ , respectively. As the actual distribution function  $F$  drifts away from  $\Phi$ , it becomes feasible that more robust estimators are used. However, we do not go too deeply into these kinds of problems, since they are known to exist already for a long time and have no specific relation to the present setup.

For the additional parameter(s) we may use (standard) estimators based on all observations. However, we only believe in our model mildly. Especially, we hope that the behavior in the ordinary tail has some relevance for the information in the far tail, where we need to go. Hence, we want to rely on the largest  $k$  order statistics only. The same type of reasoning is applied in extreme-value theory, see e.g. Dekkers and de Haan (1989) and Hall and Weissman (1997).

Note that a restriction as symmetry, for instance occurring in the normal power family, can be taken for granted, since we are only interested in fitting the tail of the distribution. Especially when dealing with the largest  $k$  order statistics only, symmetry plays no role and hence is no restriction at all.

Using the supposed relation of the behavior in the far tail to the behavior in the ordinary tail, we may estimate the additional parameter(s) by equating one or more sample quantiles in the ordinary tail to the corresponding quantiles of the distribution. Quantiles in the far tail are then estimated by plugging in the estimates of the parameter(s) in the parametric form of these quantiles in the far tail, thus using the relation between the far tail and the ordinary tail.

For the normal power family, model (iii), we calculate an estimator based on all observations and an estimator based on quantiles. In principle one can do the same job for the other models as well. However, for the models (iv) and (v) estimators are mostly implicit and not that easy to deal with, while for the mixture models indicator functions appear and hence technical problems arise when calculating correction terms. Therefore, the other considered models are mainly used to provide examples outside the normal power family.

Within the normal power family, the family of normal distributions corresponds to  $\gamma = 0$ , which is an inner point of the parameter space, in contrast to the random and deterministic mixture models. Hence, we do not suffer from the problems which complicate the mixture models.

For the estimator based on all observations we take a moment estimator. To determine moment estimators it should be remarked that  $EZ_\gamma = 0$ ,  $EZ_\gamma^2 = 1$ ,  $EZ_\gamma^3 = 0$  and hence the fourth

moment is applied for getting the moment estimator. Moreover, to facilitate the estimation we use a different parameterization of the normal power family in such a way that the additional parameter is the functional corresponding to the fourth moment (see also the discussion in Section 2 on the "suitably chosen point  $\theta \in \Theta$ "). Therefore, we replace  $\gamma$  by

$$\gamma_1^* = EZ_\gamma^4 = \frac{\sqrt{\pi}\Gamma(2\gamma + 5/2)}{\Gamma(\gamma + 3/2)^2}. \quad (4.1)$$

As a consequence, our first estimator is given by

$$\hat{\gamma}_1^* = n^{-1} \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{S} \right)^4. \quad (4.2)$$

The second estimator is based on quantiles in the ordinary tail and defined by

$$\hat{\gamma}_2^* = \frac{X_{[n+1-qn]:n} - \bar{X}}{X_{[n+1-rn]:n} - \bar{X}}, \quad (4.3)$$

where  $[x]$  denotes the entier of  $x$  and  $0 < q, r < \frac{1}{2}$ . The corresponding functional leads to the reparametrization (with  $u_t = \bar{\Phi}^{-1}(t)$ ,  $0 < t < \frac{1}{2}$ )

$$\gamma_2^* = \frac{\bar{K}_\gamma^{-1}(q)}{\bar{K}_\gamma^{-1}(r)} = \left( \frac{u_q}{u_r} \right)^{1+\gamma}. \quad (4.4)$$

**Remark 4.2** The estimator based on only one quantile may lead to problems, since the corresponding functional  $K_\gamma^{-1}(1-q) = c(\gamma) (\Phi^{-1}(1-q))^{1+\gamma}$  is not monotone in  $\gamma$  for e.g.  $q = 0.1$ .

**Remark 4.3** Implementation of the estimators in the control limit requires rewriting of  $K_\gamma^{-1}(t)$  in terms of the new parameterizations. Writing  $\gamma_i^* = h_i(\gamma)$ ,  $i = 1, 2$  the (uncorrected) control limit becomes

$$\hat{\mu} + \hat{\sigma} \bar{K}_{h_i^{-1}(\hat{\gamma}_i^*)}^{-1}(p). \quad (4.5)$$

The function  $h_2^{-1}$  is simply given by

$$h_2^{-1}(x) = \frac{\log(x)}{\log\left(\frac{u_q}{u_r}\right)} - 1. \quad (4.6)$$

For commonly used values of  $q$  and  $r$ , the estimator  $\gamma_2^*$  will be positive. If desired, the extension of the domain of  $h_2^{-1}$  to the whole real line is obtained by replacing  $\log(x)$  by  $\log(|x|)$  in (4.6).

Next observe that due to estimation the (observed) false alarm rate is a random variable  $P$ , given by (2.1). We want  $P$  to be close to  $p$ . To compare the stochastic  $P$  and the deterministic  $p$ , several criteria can be used. The most obvious one is to consider  $EP$  and to compare it with  $p$ . Another possibility is to investigate the average run length, leading to a comparison of  $E\left(\frac{1}{P}\right)$  and  $\frac{1}{p}$ . Consideration of the probability that the run length is at most some specified value  $k$ , leads to comparison of  $E\left\{1 - (1 - P)^k\right\}$  with  $1 - (1 - p)^k$ .

More generally, we consider a function  $g(p)$ , estimate it by  $g(P)$  and compare  $Eg(P)$  with  $g(p)$ . In particular, we focus on the previous functions

$$g(p) = p, g(p) = \frac{1}{p}, g(p) = 1 - (1 - p)^k. \quad (4.7)$$

Typical values of interest for  $k$  are small fractions of the average run length, that is  $k = [\delta/p]$  with small  $\delta$ .

Other criteria, like the exceedance probability can be considered as well. However, in this paper we restrict attention to the bias as criterion.

## 5 Correction terms

It is shown in Albers and Kallenberg (2000, 2001) that, even when normality holds, without correction  $Eg(P)$  may differ substantially from  $g(p)$ , unless very large sample sizes are used. On the other hand, when normality holds, simple corrections can be developed, leading to control limits that meet for common sample sizes the required conditions posed by the criteria involved. Here we derive similar corrections in the present set-up in order to get  $Eg(P)$  close to  $g(p)$ .

We insert a correction term  $c_u = c_u(\hat{\theta})$ , to be determined later on, leading to the control limit  $\hat{\mu} + \hat{\sigma} \left\{ \overline{K}_{\hat{\gamma}}^{-1}(p) + c_u(\hat{\theta}) \right\}$ , which implies that the false alarm rate is now given by

$$\begin{aligned} P \left( X_{n+1} > \hat{\mu} + \hat{\sigma} \left\{ \overline{K}_{\hat{\gamma}}^{-1}(p) + c_u \right\} \right) &= \overline{F} \left( \hat{\mu} + \hat{\sigma} \left\{ \overline{K}_{\hat{\gamma}}^{-1}(p) + c_u \right\} \right) \\ &= \overline{F}_0 \left( \overline{K}_{\hat{\gamma}}^{-1}(p) + V + \frac{\hat{\sigma}}{\sigma} c_u \right), \end{aligned}$$

where  $F_0$  is the distribution function of  $\sigma^{-1}(X_{n+1} - \mu)$  and where

$$V = \frac{\hat{\mu} + \hat{\sigma} \overline{K}_{\hat{\gamma}}^{-1}(p) - \left\{ \mu + \sigma \overline{K}_{\gamma}^{-1}(p) \right\}}{\sigma} = \frac{\hat{\mu} - \mu}{\sigma} + \frac{\hat{\sigma}}{\sigma} \overline{K}_{\hat{\gamma}}^{-1}(p) - \overline{K}_{\gamma}^{-1}(p).$$

It is seen from the definitions of our estimators  $\hat{\mu}$ ,  $\hat{\sigma}$  and  $\hat{\gamma}$  that without loss of generality we may assume that  $\mu = 0$  and  $\sigma = 1$  and that  $X_1, \dots, X_n, X_{n+1}$  have distribution function  $F_0$ . Therefore, the false alarm rate reads as  $\overline{F}_0 \left( \overline{K}_{\hat{\gamma}}^{-1}(p) + V + \hat{\sigma} c_u \right)$  with

$$V = \hat{\mu} + \hat{\sigma} \overline{K}_{\hat{\gamma}}^{-1}(p) - \overline{K}_{\gamma}^{-1}(p). \quad (5.1)$$

Similar to (2.1) we have, for some suitably chosen point  $\gamma$ ,

$$\begin{aligned} g \left( \overline{F}_0 \left( \overline{K}_{\hat{\gamma}}^{-1}(p) + V + \hat{\sigma} c_u \right) \right) & \quad (5.2) \\ &= g(p) + \left[ g \left( \overline{F}_0 \left( \overline{K}_{\hat{\gamma}}^{-1}(p) \right) \right) - g(p) \right] \\ &+ \left[ g \left( \overline{F}_0 \left( \overline{K}_{\hat{\gamma}}^{-1}(p) + V + \hat{\sigma} c_u \right) \right) - g \left( \overline{F}_0 \left( \overline{K}_{\hat{\gamma}}^{-1}(p) \right) \right) \right]. \end{aligned}$$

The aim of the correction term is to reduce the bias in the stochastic error

$$g \left( \overline{F}_0 \left( \overline{K}_{\hat{\gamma}}^{-1}(p) + V + \hat{\sigma} c_u \right) \right) - g \left( \overline{F}_0 \left( \overline{K}_{\hat{\gamma}}^{-1}(p) \right) \right).$$

Typically we have that  $EV$  and  $EV^2$  are of order  $O(n^{-1})$ , and  $E|V|^k = o(n^{-1})$  for  $k \geq 2$ . The correction term  $c_u$  for correcting the bias typically is of order  $O(n^{-1})$ . Let  $F_0$  have density

$f_0$ . Taylor expansion of the stochastic error gives, writing  $\doteq$  for "approximately equal to" and ignoring terms of order  $V^k$  for  $k \geq 3$  and terms of order  $c_u(\hat{\sigma} - 1)$ ,  $c_u V$  and  $(c_u)^k$  for  $k \geq 2$ ,

$$\begin{aligned} & g\left(\bar{F}_0\left(\bar{K}_\gamma^{-1}(p) + V + \hat{\sigma}c_u\right)\right) - g\left(\bar{F}_0\left(\bar{K}_\gamma^{-1}(p)\right)\right) \\ & \doteq -g'\left(\bar{F}_0\left(\bar{K}_\gamma^{-1}(p)\right)\right) f_0\left(\bar{K}_\gamma^{-1}(p)\right) (V + c_u) \\ & + \frac{1}{2}V^2 \left[ g''\left(\bar{F}_0\left(\bar{K}_\gamma^{-1}(p)\right)\right) f_0^2\left(\bar{K}_\gamma^{-1}(p)\right) - g'\left(\bar{F}_0\left(\bar{K}_\gamma^{-1}(p)\right)\right) f_0'\left(\bar{K}_\gamma^{-1}(p)\right) \right]. \end{aligned} \quad (5.3)$$

In order to get the bias close to 0, we equate the right-hand side of (5.3) to 0 and hence a starting point for the correction term is  $c_{u1}$ , given by

$$c_{u1} = -EV + \frac{1}{2}EV^2 \left[ \frac{g''}{g'}\left(\bar{F}_0\left(\bar{K}_\gamma^{-1}(p)\right)\right) f_0\left(\bar{K}_\gamma^{-1}(p)\right) - \frac{f_0'}{f_0}\left(\bar{K}_\gamma^{-1}(p)\right) \right]. \quad (5.4)$$

Next we shall simplify the expression from (5.4). First, as concerns  $g$ , consider the functions given in (4.7) with in the latter case  $\delta (\doteq kp)$  small. Then  $\frac{g''}{g'}(p)$  equals 0,  $-\frac{2}{p}$  and  $-(k-1)/(1-p) \doteq -k \doteq -\frac{\delta}{p}$ , respectively. Assuming that  $F_0$  is close to  $K_\gamma$ , we get

$$\frac{g''}{g'}\left(\bar{F}_0\left(\bar{K}_\gamma^{-1}(p)\right)\right) f_0\left(\bar{K}_\gamma^{-1}(p)\right) \doteq \tilde{\lambda} \frac{f_0}{F_0}\left(\bar{K}_\gamma^{-1}(p)\right) \quad (5.5)$$

with  $\tilde{\lambda} = 0, -2$  and  $-\delta$ , respectively. Moreover, for a wide class of  $F$ 's we have

$$\frac{f_0}{F_0} \doteq -\frac{f_0'}{f_0}, \quad (5.6)$$

see Andrews (1973) for more details and examples.

Together (5.5) and (5.6) enable us to reduce (5.4) to the simpler correction term  $c_{u2}$ , given by

$$c_{u2} = -EV - \frac{1}{2}\lambda EV^2 \frac{f_0'}{f_0}\left(\bar{K}_\gamma^{-1}(p)\right)$$

with  $\lambda = 1, -1$  and  $1 - \delta$ , respectively.

Note that  $c_{u2}$  involves unknown parts, which in turn need to be estimated. Write  $k_\gamma$  for the density of  $Z_\gamma$ . The density  $f_0$  is supposed to be close to  $k_\gamma$ . The parameter  $\gamma$  in this  $k_\gamma$  will be estimated. Furthermore,  $EV$  and  $EV^2$  depend on the unknown  $F_0$ . Here again  $F_0$  is replaced by  $K_\gamma$  and it suffices to estimate  $\gamma$  in here as well. Finally, we get as our correction term

$$c_u = c_u(\hat{\gamma}) = -\widehat{EV} - \frac{1}{2}\lambda \widehat{EV^2} \frac{k_\gamma'}{k_\gamma}\left(\bar{K}_{\hat{\gamma}}^{-1}(p)\right).$$

The rule for providing a signal that the process may be out-of-control reads as

$$X > \hat{\mu} + \hat{\sigma} \left\{ \bar{K}_{\hat{\gamma}}^{-1}(p) + c_u(\hat{\gamma}) \right\}.$$

Note that the correction term depends on the supposed model. If the supposed model is the restricted model, then we can take  $\gamma = 0$  and no estimation is needed in the correction term. (The correction term itself is still needed!) In the general parametric model we have to estimate  $\gamma$ .



We present the correction terms for the normal power family, model (iii). Here  $V$  is of the form

$$\hat{\mu} + \hat{\sigma} \overline{K}_{h_i^{-1}(\hat{\gamma}_i^*)}^{-1}(p) - \overline{K}_{h_i^{-1}(\gamma_i^*)}^{-1}(p).$$

The main part of the calculations to find the correction terms concerns  $\widehat{EV}$  and  $\widehat{EV}^2$ . Because very complicated expressions are involved, we replace  $\overline{K}_{\gamma}^{-1}(p)$ ,  $V$  and  $V^2$  by suitable quadratic approximations. The derivation and resulting formulas are given in the Appendix. The control limits using the estimators  $\hat{\gamma}_1^*$  and  $\hat{\gamma}_2^*$  including the correction terms are given by

$$\hat{\mu} + \hat{\sigma} \left\{ \overline{K}_{h_i^{-1}(\hat{\gamma}_i^*)}^{-1}(p) - \widehat{EV} + \frac{\lambda \widehat{EV}^2}{2} \frac{u_p^2 + h_i^{-1}(\hat{\gamma}_i^*)}{\{1 + h_i^{-1}(\hat{\gamma}_i^*)\} c(h_i^{-1}(\hat{\gamma}_i^*)) u_p^{1+h_i^{-1}(\hat{\gamma}_i^*)}} \right\}$$

for  $i = 1, 2$  and with  $\widehat{EV}$  and  $\widehat{EV}^2$  replaced by, respectively, the right-hand sides of (A.8) and (A.9) for  $\hat{\gamma}_1^*$  and (A.14) and (A.15) for  $\hat{\gamma}_2^*$ .

## 6 Simulation and other numerical results

A simulation study is performed to see to what kind of improvement the several steps lead: firstly, extending the restricted model to a larger model and secondly, the application of the correction terms in the larger model. For comparison we also consider the restricted model with correction terms. In the simulation study we want to cover the restricted model, the general parametric model and distributions outside the general parametric model (but not too far away from it).

As criterion we take  $EP$  and compare this to  $p$ . In terms of Section 4 this means that we take  $g(p) = p$  in the simulation study. Similar results as presented here hold for the other functions of Section 4, cf. (4.7). In the simulations we always choose  $p = 0.001$  and for  $n$  we take 100, 250 and 500. Our main attention is on  $n = 100$ , since nowadays large samples are usually not available. The columns with  $n = 250, 500$  give an impression of the rate of convergence. The number of repetitions in the simulation study equals 100,000.

For the restricted model we use the family of normal distributions, while the general parametric model in the simulation study is the normal power family. For the distributions outside the general parametric model we take distributions from the models described in Section 3.

More specifically, in the simulation study we use the following distributions (with  $\mu = 0$  and  $\sigma = 1$ ):

$\Phi$ : standard normal distribution function;

$K_{\gamma}$ : normal power distribution function with  $\gamma = -0.5, -0.25, 0(= \Phi), 0.25, 0.5, 0.75, 1$ , see Section 3 (iii);

$T$ : standardized Student distribution function with 6 degrees of freedom;

$RM$ : random mixture:  $\frac{1}{2}\Phi + \frac{1}{2}T$ , see Section 3 (i);

$DM$ : deterministic mixture, given by:  $c(\frac{1}{2}) \{ \frac{1}{2}\Phi^{-1}(U) + \frac{1}{2}T^{-1}(U) \}$ , see Section 3 (ii);

$TU$ : Tukey's  $\lambda$  with  $\lambda = -0.1, \lambda = 0$  (logistic distribution),  $\lambda = 0.14$ , see Section 3 (iv);

$O$ : Orthonormal family with  $k = 3$  and  $(\gamma_1, \gamma_2, \gamma_3) = (-0.1, -0.1, 0.1)$ , see Section 3 (v).

We also calculate numerically the first and second order approximations and the model errors. For convenience the unit in the tables equals 0.001, thus, for instance, 1.25 means 0.00125.

The control limit depends on the supposed model, the estimators of the parameters and whether we make a correction or not. We consider the following cases.

(i) *Supposed model: normality, no correction.*

Assuming that we are in the restricted model of normality and applying no correction for using estimators, the control limit simply equals

$$\bar{X} + Su_p.$$

The first order approximation of the expected (observed) false alarm rate is given by  $\bar{F}_0(u_p)$  with  $F_0$  the distribution function of  $\sigma^{-1}(X_{n+1} - \mu)$  and hence the first order approximation is related to the model error by  $\bar{F}_0(u_p) = p + ME$ , cf. (2.2). It easily follows from (5.3) with  $g(p) = p$ ,  $\bar{K}_\gamma^{-1}(p) = u_p$  and  $c_u = 0$ , that the second order approximation of  $EP$  equals

$$\bar{F}_0(u_p) - f_0(u_p)EV - \frac{1}{2}EV^2 f'_0(u_p)$$

with, cf. (5.1),

$$V = \bar{X} + Su_p - u_p.$$

Hence the second order approximation reduces to, cf. also (A.5),

$$\bar{F}_0(u_p) + \frac{(\mu_4 - 1) f_0(u_p) u_p}{8n} - \frac{f'_0(u_p) \{1 + \mu_3 u_p + \frac{1}{4}(\mu_4 - 1) u_p^2\}}{2n},$$

where  $\mu_k$  is the  $k^{th}$  moment under  $F_0$ .

The simulation results, the model error and the first and second order approximations are presented in Table 1.

*Table 1 Simulated expected (observed) false alarm rate without correction for the estimation of the mean and variance and corresponding first and second order approximation assuming normality as model, with in the column ME the model error. The unit in the table is 0.001.*

$F_0$	simulation			1 <sup>st</sup> order appr.	ME	2 <sup>st</sup> order appr.		
	$n = 100$	$n = 250$	$n = 500$			$n = 100$	$n = 250$	$n = 500$
$\Phi$	1.36	1.13	1.07	1.00	0.00	1.33	1.13	1.07
$K_{-0.5}$	0.00	0.00	0.00	0.00	-1.00	0.00	0.00	0.00
$K_{-0.25}$	0.07	0.03	0.03	0.02	-0.98	0.05	0.03	0.03
$K_{0.25}$	4.43	3.96	3.81	3.66	2.66	4.39	3.95	3.81
$K_{0.5}$	7.70	7.01	6.80	6.58	5.58	7.67	7.02	6.80
$K_{0.75}$	10.31	9.45	9.15	8.86	7.86	10.30	9.44	9.15
$K_1$	12.12	11.06	10.71	10.35	9.35	12.17	11.08	10.71
$T$	5.31	4.88	4.73	4.56	3.56	5.37	4.89	4.73
$RM$	3.33	3.00	2.89	2.78	1.78	3.34	3.00	2.89
$DM$	3.45	3.15	3.03	2.92	1.92	3.43	3.13	3.02
$TU(-0.1)$	6.61	6.08	5.89	5.71	4.71	6.66	6.09	5.90
$TU(0)$	4.28	3.91	3.79	3.67	2.67	4.26	3.90	3.79
$TU(0.14)$	1.22	0.98	0.91	0.85	-0.15	1.19	0.98	0.91
$O$	2.80	2.40	2.25	2.13	1.13	2.76	2.38	2.26

It is seen from Table 1 that the false alarm rate may be completely wrong if we are not in the restricted model, that is if normality does not hold. This confirms previous results of e.g. Chan et al. (1988) and Pappanastos and Adams (1996). They only give the total error. By

splitting up the total error in the model error and the stochastic error more insight is obtained about the roles of misspecification on the one hand and estimation of the location and scale parameter on the other hand.

The model error  $ME$  is the difference between the first order approximation and  $0.001 (= p)$ . Note that the  $ME$ 's of the Tukey distributions differ from those presented in Section 3, due to the fact that for the Tukey distributions in Section 3  $TU(0.14)$  is used as supposed model instead of normality. It is illuminating that, while in the middle of the distribution there is hardly any difference between  $TU(0.14)$  and the standard normal distribution, the far tails are different. The effect of this difference is clearly seen by comparing  $TU(0.14)$  and  $\Phi$  in Table 1. Often it is also stated that in the middle there is not much difference between normal and logistic distributions. Comparison of  $TU(0)$  and  $\Phi$  in Table 1 shows that for the problem at hand sloppy inspection of the data in the middle, leading to the conclusion that "normality is not so bad" may produce serious errors.

Apart from the misspecification the effect of the estimation is also not negligible. Due to the estimation the total error becomes substantially larger in case of positive  $ME$ 's, which may occur in practice more often than negative  $ME$ 's. In the latter situation the total error is compensated by the (positive) stochastic error. This, for instance, occurs for  $TU(0.14)$ . However, such compensation is in an uncontrolled way and may be far too small (see  $K_{-0.25}$ ) or may lead to overcompensation ( $TU(0.14)$ ,  $n = 100$ ).

Clearly, the first order approximation gives an impression of  $EP$ , the expected (observed) false alarm rate, but is not very precise. The second order approximation gives a very good prediction of the simulation.

(ii) *Supposed model: normality, with correction.*

Assuming that we are in the restricted model of normality and applying the suitable correction term for using estimators, the control limit equals

$$\bar{X} + Su_p + \frac{Su_p}{4n} (u_p^2 + 3).$$

The correction term is a second order term and hence the first order approximation is the same as in Table 1. The second order approximation equals

$$\bar{F}_0(u_p) - f_0(u_p) \frac{u_p}{4n} \left\{ u_p^2 + 3 - \frac{1}{2}(\mu_4 - 1) \right\} - \frac{1}{2} f_0'(u_p) \frac{1 + \mu_3 u_p + \frac{1}{4}(\mu_4 - 1) u_p^2}{n}.$$

It is seen that indeed for  $F_0 = \Phi$  the second order approximation gives exactly  $p$ .

The simulation results, the model error and the first and second order approximations are presented in Table 2. Note that the model error and the first order approximation are the same as in Table 1.

Table 2 Simulated expected (observed) false alarm rate with correction for the estimation of the mean and variance and corresponding first and second order approximation assuming normality as model, with in the column ME the model error. The unit in the table is 0.001.

$F_0$	simulation			1 <sup>st</sup> order appr.	ME	2 <sup>st</sup> order appr.		
	$n = 100$	$n = 250$	$n = 500$			$n = 100$	$n = 250$	$n = 500$
$\Phi$	1.01	1.00	1.00	1.00	0.00	1.00	1.00	1.00
$K_{-0.5}$	0.00	0.00	0.00	0.00	-1.00	0.00	0.00	0.00
$K_{-0.25}$	0.10	0.03	0.02	0.02	-0.98	0.03	0.03	0.02
$K_{0.25}$	3.63	3.66	3.66	3.66	2.66	3.65	3.66	3.66
$K_{0.5}$	6.72	6.65	6.61	6.58	5.58	6.71	6.63	6.61
$K_{0.75}$	9.25	9.03	8.94	8.86	7.86	9.28	9.03	8.95
$K_1$	11.06	10.64	10.51	10.35	9.35	11.17	10.68	10.51
$T$	4.64	4.62	4.60	4.56	3.56	4.73	4.63	4.60
$RM$	2.81	2.81	2.79	2.78	1.78	2.86	2.81	2.80
$DM$	2.90	2.91	2.92	2.92	1.92	2.90	2.91	2.92
$TU(-0.1)$	5.85	5.78	5.75	5.71	4.71	5.94	5.80	5.76
$TU(0)$	3.62	3.65	3.66	3.67	2.67	3.62	3.65	3.66
$TU(0.14)$	0.87	0.85	0.84	0.85	-0.15	0.87	0.85	0.85
$O$	2.18	2.15	2.14	2.13	1.13	2.16	2.14	2.14

It is seen from the row denoted by  $\Phi$  in Table 2 that the correction for estimation of the parameters works very well, cf. also Albers and Kallenberg (2000) for more details. Comparison of Table 1 and Table 2 shows that the correction reduces the stochastic error not only for the normal distribution, but as well for the other distributions, thus bringing the total error closer to the model error. Nevertheless, if normality fails still the false alarm rate may be completely wrong due to the model error. As a consequence of the bias correction, the first and second order approximation are much closer to each other. Again, the second order approximation gives a very good prediction of the simulation.

(iii) *Supposed model: normal power family, no correction, estimator  $\hat{\gamma}_1^*$ .*

Assume that our observations are from the normal power family and that we use the estimator based on all observations  $\hat{\gamma}_1^*$ . Then the control limit without correcting for the estimation equals

$$\bar{X} + S\bar{K}_{h_1^{-1}(\hat{\gamma}_1^*)}^{-1}(p).$$

The first order approximation of the false alarm rate is given by  $\bar{F}_0\left(\bar{K}_\gamma^{-1}(p)\right)$  with  $\gamma$  the suitably chosen point given by  $\gamma = h_1^{-1}(\mu_4)$ . It is related to the model error by  $\bar{F}_0\left(\bar{K}_\gamma^{-1}(p)\right) = p + ME$ , cf. (2.2). It easily follows from (5.3) with  $g(p) = p$  and  $c_u = 0$ , that the second order approximation of  $EP$  equals

$$\bar{F}_0\left(\bar{K}_\gamma^{-1}(p)\right) - f_0\left(\bar{K}_\gamma^{-1}(p)\right)EV - \frac{1}{2}EV^2f_0'\left(\bar{K}_\gamma^{-1}(p)\right)$$

with  $EV$  and  $EV^2$  given by (A.6) and (A.7), respectively. The involved moments  $\mu_i$  should be taken under  $F_0$ .

The simulation results, the model error and the first and second order approximations are presented in Table 3.

Table 3 Simulated expected (observed) false alarm rate using  $\hat{\gamma}_1^*$  as estimator without correction for the estimation of the mean, the variance and the parameter  $\gamma$  and corresponding first and second order approximation assuming the normal power family as model, with in the column ME the model error. The unit in the table is 0.001.

$F_0$	simulation			1 <sup>st</sup> order appr.	ME	2 <sup>st</sup> order appr.		
	$n = 100$	$n = 250$	$n = 500$			$n = 100$	$n = 250$	$n = 500$
$\Phi$	2.26	1.46	1.22	1.000	0.000	2.02	1.41	1.20
$K_{-0.5}$	2.30	1.46	1.22	1.000	0.000	2.02	1.41	1.20
$K_{-0.25}$	2.14	1.41	1.20	1.000	0.000	1.89	1.36	1.18
$K_{0.25}$	2.42	1.54	1.26	1.000	0.000	2.31	1.52	1.26
$K_{0.5}$	2.64	1.65	1.32	1.000	0.000	2.78	1.71	1.36
$K_{0.75}$	2.87	1.76	1.40	1.000	0.000	3.55	2.02	1.51
$K_1$	3.09	1.88	1.48	1.000	0.000	4.81	2.52	1.76
$T$	3.42	2.18	1.73	0.91	-0.09	—	—	—
$RM$	3.01	1.98	1.61	0.89	-0.01	—	—	—
$DM$	3.06	2.02	1.67	1.16	0.16	—	—	—
$TU(-0.1)$	3.45	2.17	1.70	1.00	0.00	28.04	11.81	6.41
$TU(0)$	3.04	1.98	1.63	1.24	0.24	3.46	2.12	1.68
$TU(0.14)$	2.11	1.33	1.10	0.90	-0.10	1.85	1.28	1.09
$O$	3.05	1.94	1.63	1.33	0.33	2.73	1.89	1.61

Compared to Tables 1 and 2 it is seen from Table 3 that the model error is tremendously reduced by considering the normal power family as parametric model. This is obvious for the distributions belonging to the normal power family, but also holds for the distributions outside the normal power family. In this sense the estimator based on the fourth moment does a very good job by selecting an appropriate parameter value  $\gamma$  in the normal power family and thus reducing the model error from 3.6 to  $-0.1$  ( $T$ ), 1.8 to  $-0.1$  ( $RM$ ) etc.

On the other hand, the stochastic error may be (very) large. This is well understood from the second order approximation, where the eighth moment appears, see (A.6) and (A.7). Hence, it may be expected that for distributions with heavy tails the nonrobustness of the estimator based on the fourth moment causes a large stochastic error. Indeed, this is seen in Table 3 by looking at for instance  $K_{0.75}$ ,  $K_1$ ,  $T$ ,  $RM$ ,  $DM$ ,  $TU(-0.1)$  and  $TU(0)$ . (Note that for  $T$ ,  $RM$  and  $DM$  the sixth and eighth moment are infinite and that for  $TU(-0.1)$  the eighth moment equals 42546, implying a nonexistent or an inaccurate second order approximation in those cases.)

Nevertheless, when comparing the uncorrected control limits (Table 1 and 3) a great improvement in the total error is achieved in most cases, while the loss is not large when the restricted model of normality holds. The second order approximation gives a pretty good prediction of the simulation, unless heavy tails occur with large sixth and eighth moments.

(iv) *Supposed model: normal power family, with correction, estimator  $\hat{\gamma}_1^*$ .*

The control limit in this case equals

$$\bar{X} + S \left\{ \bar{K}_{h_1^{-1}(\hat{\gamma}_1^*)}^{-1}(p) - \widehat{EV} + \frac{\lambda \widehat{EV}^2}{2} \frac{u_p^2 + h_1^{-1}(\hat{\gamma}_1^*)}{\{1 + h_1^{-1}(\hat{\gamma}_1^*)\} c(h_1^{-1}(\hat{\gamma}_1^*)) u_p^{1+h_1^{-1}(\hat{\gamma}_1^*)}} \right\}$$

with  $\widehat{EV}$  and  $\widehat{EV}^2$  replaced by the right-hand sides of (A.8) and (A.9), respectively.

The correction term is a second order term and hence the first order approximation is the same as in Table 3. The second order approximation equals, with  $\gamma$  the suitably chosen point given by  $\gamma = h_1^{-1}(\mu_4)$ ,

$$\begin{aligned} & \bar{F}_0 \left( \bar{K}_\gamma^{-1}(p) \right) - f_0 \left( \bar{K}_\gamma^{-1}(p) \right) \{ E_{F_0} V - E_\gamma V \} \\ & - \frac{1}{2} \left\{ f_0' \left( \bar{K}_\gamma^{-1}(p) \right) E_{F_0} V^2 - f_0 \left( \bar{K}_\gamma^{-1}(p) \right) \frac{k_\gamma'}{k_\gamma} \left( \bar{K}_\gamma^{-1}(p) \right) E_\gamma V^2 \right\}, \end{aligned}$$

where  $E_{F_0} V$  and  $E_\gamma V$  refer to the right-hand side of (A.6) inserting  $\mu_i = EX_1^i$  and  $\mu_i = EZ_\gamma^i$ , respectively, and  $E_{F_0} V^2$  and  $E_\gamma V^2$  refer to the right-hand side of (A.7) inserting  $\mu_i = EX_1^i$  and  $\mu_i = EZ_\gamma^i$ , respectively. It immediately follows that for  $F_0 = K_\gamma$  the second order approximation gives  $p$ .

The simulation results, the model error and the first and second order approximations are presented in Table 4.

*Table 4 Simulated expected (observed) false alarm rate using  $\hat{\gamma}_1^*$  as estimator with correction for the estimation of the mean, the variance and the parameter  $\gamma$  and corresponding first and second order approximation assuming the normal power family as model, with in the column ME the model error. The unit in the table is 0.001.*

$F_0$	simulation			1 <sup>st</sup> order appr.	ME	2 <sup>st</sup> order appr.		
	$n = 100$	$n = 250$	$n = 500$			$n = 100$	$n = 250$	$n = 500$
$\Phi$	1.12	1.04	1.01	1.00	0.00	1.00	1.00	1.00
$K_{-0.5}$	1.03	1.01	1.01	1.00	0.00	1.00	1.00	1.00
$K_{-0.25}$	1.09	1.02	1.01	1.00	0.00	1.00	1.00	1.00
$K_{0.25}$	1.15	1.02	1.01	1.00	0.00	1.00	1.00	1.00
$K_{0.5}$	1.19	1.03	1.00	1.00	0.00	1.00	1.00	1.00
$K_{0.75}$	1.21	0.98	0.98	1.00	0.00	1.00	1.00	1.00
$K_1$	1.16	0.91	0.92	1.00	0.00	1.00	1.00	1.00
$T$	2.06	1.66	1.46	0.91	-0.09	—	—	—
$RM$	1.82	1.55	1.40	0.89	-0.11	—	—	—
$DM$	1.99	1.56	1.41	1.16	0.16	—	—	—
$TU(-0.1)$	2.02	1.59	1.40	1.00	0.00	26.47	11.18	6.09
$TU(0)$	1.70	1.45	1.37	1.24	0.24	2.16	1.60	1.42
$TU(0.14)$	1.05	0.91	0.91	0.90	-0.10	0.81	0.86	0.88
$O$	1.49	1.38	1.36	1.33	0.33	1.28	1.31	1.32

It is seen from Table 4 that the correction for estimation of the parameters works very well in all cases. The great reduction of the model error (see Table 3) is followed by a substantial reduction of the stochastic error. The appropriate parameter value is picked up by the estimator (both within and outside the parametric family) and the correction term, based on the behavior in the parametric family, is very useful not only inside the parametric family (reducing the total error e.g. from 200% to 20% for  $n = 100$ ) but also outside it. For instance, the total error goes from 242% to 106% ( $T, n = 100$ ) or from 112% to 5% ( $TU(0.14), n = 100$ ).

When the higher moments of the distribution are not very large, the second order approximation gives a pretty good prediction of the simulation. However, for heavy tailed distributions the second order approximation does not exist or gives a bad approximation. This is due to the application of the moment estimator, based on the fourth moment, being quite nonrobust. As is seen in (A.5) the second moment of  $\hat{\gamma}_1^*$  contains already high moments as  $\mu_8$  and  $\mu_6$ . It has to

be expected that lower order terms, involving for instance the third and fourth moment of  $\hat{\gamma}_1^*$ , contain even higher moments, like  $\mu_{12}$  etc. Although in the expansion these terms have  $n^2$  in the denominator they may have great impact for such large values of  $\gamma$  as  $\gamma = 1$ , since these very high moments increase very fast. For instance, for  $\gamma = 0, \frac{1}{2}, 1$  we have  $\mu_8 = 105, 1603, 25025$ , respectively.

The corrected control limit based on the estimator  $\hat{\gamma}_1^*$  works very well and much better than the one based on the normality assumption. Nevertheless, we do consider also the estimator  $\hat{\gamma}_2^*$ . The first reason is that the control limit based on  $\hat{\gamma}_1^*$  is computationally more complicated due to the fact that in contrast to  $h_2^{-1}(\hat{\gamma}_2^*)$  the function  $h_1^{-1}(\hat{\gamma}_1^*)$  is not explicitly given. As these functions take a prominent place in the control limit already in the first order term, their values should be calculated rather precisely, which is easy for  $h_2^{-1}(\hat{\gamma}_2^*)$  and difficult for  $h_1^{-1}(\hat{\gamma}_1^*)$ .

The second reason is that  $\hat{\gamma}_2^*$  is based on order statistics and comes conceptually closer to the nonparametric approach, thus being an interesting intermediate step between assuming much knowledge about the distribution (control chart based on normality) and no assumptions at all (nonparametric control chart using very extreme order statistics). For the latter we need a huge amount of data, while the estimator  $\hat{\gamma}_2^*$  uses more common order statistics and can be applied for the sample sizes that we meet in practice.

(v) *Supposed model: normal power family, no correction, estimator  $\hat{\gamma}_2^*$ .*

Assume that our observations are from the normal power family and that we use the estimator  $\hat{\gamma}_2^*$  based on quantiles in the ordinary tail. Then the control limit without correcting for the estimation equals

$$\bar{X} + SK_{h_2^{-1}(\hat{\gamma}_2^*)}^{-1}(p).$$

The first order approximation of the false alarm rate is given by  $\bar{F}_0\left(\bar{K}_\gamma^{-1}(p)\right)$  with  $\gamma$  the suitably chosen point given by

$$\gamma = \frac{\log\left(\frac{\bar{F}_0^{-1}(q)}{\bar{F}_0^{-1}(r)}\right)}{\log\left(\frac{u_q}{u_r}\right)} - 1. \quad (6.1)$$

It is related to the model error by  $\bar{F}_0\left(\bar{K}_\gamma^{-1}(p)\right) = p + ME$ , cf. (2.2). It easily follows from (5.3) with  $g(p) = p$  and  $c_u = 0$ , that the second order approximation of  $EP$  equals

$$\bar{F}_0\left(\bar{K}_\gamma^{-1}(p)\right) - f_0\left(\bar{K}_\gamma^{-1}(p)\right)EV - \frac{1}{2}EV^2f_0'\left(\bar{K}_\gamma^{-1}(p)\right)$$

with  $EV$  and  $EV^2$  given by (A.12) and (A.13), respectively. The involved moments  $\mu_i$  should be taken under  $F_0$ .

In the simulation we take  $q = 0.05$  and  $r = 0.25$ . The simulation results, the model error and the first and second order approximations are presented in Table 5.

Table 5 Simulated expected (observed) false alarm rate using  $\hat{\gamma}_2^*$  as estimator without correction for the estimation of the mean, the variance and the parameter  $\gamma$  and corresponding first and second order approximation assuming the normal power family as model, with in the column  $ME$  the model error. The unit in the table is 0.001.

$F_0$	simulation			1 <sup>st</sup> order appr.	$ME$	2 <sup>st</sup> order appr.		
	$n = 100$	$n = 250$	$n = 500$			$n = 100$	$n = 250$	$n = 500$
$\Phi$	2.69	1.62	1.28	1.00	0.00	2.26	1.52	1.25
$K_{-0.5}$	3.45	1.92	1.42	1.00	0.00	3.15	1.89	1.43
$K_{-0.25}$	2.82	1.69	1.31	1.00	0.00	2.50	1.63	1.30
$K_{0.25}$	2.39	1.53	1.24	1.00	0.00	2.15	1.48	1.23
$K_{0.5}$	2.38	1.51	1.24	1.00	0.00	2.11	1.46	1.22
$K_{0.75}$	2.37	1.50	1.23	1.00	0.00	2.12	1.46	1.22
$K_1$	2.36	1.51	1.24	1.00	0.00	2.18	1.48	1.24
$T$	4.67	3.77	3.39	3.08	2.08	4.56	3.73	3.38
$RM$	3.66	2.77	2.43	2.16	1.16	3.44	2.71	2.42
$DM$	3.85	2.90	2.60	2.28	1.28	3.61	2.85	2.55
$TU(-0.1)$	4.92	3.98	3.57	3.25	2.25	4.79	3.93	3.56
$TU(0)$	3.93	3.00	2.63	2.33	1.33	3.75	2.94	2.61
$TU(0.14)$	2.48	1.41	1.09	0.81	-0.19	2.11	1.35	1.07
$O$	2.28	1.24	0.91	0.69	-0.31	1.82	1.15	0.91

Compared to Tables 1 and 2 it is seen from Table 5 that the difference between the expected (observed) false alarm rate and the required value 0.001 is seriously reduced by considering the normal power family as parametric model. Moreover, when the restricted model of normality holds, the loss is not large. Comparison with Table 3 shows that in the normal power family  $\hat{\gamma}_2^*$  is (slightly) better than  $\hat{\gamma}_1^*$  for positive  $\gamma$ 's and (slightly) worse for negative  $\gamma$ 's. For the distributions outside the normal power family the model error  $ME$  is substantially reduced compared to the model error obtained when assuming normality, but not as good as in case of  $\hat{\gamma}_1^*$ . Consequently, the performance under these distributions is worse than in Table 3, although the stochastic error is smaller than when using  $\hat{\gamma}_1^*$ . The estimator  $\hat{\gamma}_2^*$  is more robust than  $\hat{\gamma}_1^*$ . As a consequence, in contrast to Table 3, here the second order approximation gives a very good prediction of the simulation throughout the table. This makes it much easier to analyze and predict the behavior of the control chart under all kinds of distributions.

(vi) *Supposed model: normal power family, with correction, estimator  $\hat{\gamma}_2^*$ .*

The control limit in this case equals

$$\bar{X} + S \left\{ \bar{K}_{h_2^{-1}(\hat{\gamma}_2^*)}^{-1}(p) - \widehat{EV} + \frac{\lambda \widehat{EV}^2}{2} \frac{u_p^2 + h_2^{-1}(\hat{\gamma}_2^*)}{\{1 + h_2^{-1}(\hat{\gamma}_2^*)\} c(h_2^{-1}(\hat{\gamma}_2^*)) u_p^{1+h_2^{-1}(\hat{\gamma}_2^*)}} \right\}$$

with  $\widehat{EV}$  and  $\widehat{EV}^2$  replaced by the right-hand sides of (A.14) and (A.15), respectively. The correction term is a second order term and hence the first order approximation is the same as in Table 5. The second order approximation equals, with  $\gamma$  the suitably chosen point given by (6.1),

$$\begin{aligned} & \bar{F}_0 \left( \bar{K}_\gamma^{-1}(p) \right) - f_0 \left( \bar{K}_\gamma^{-1}(p) \right) \{E_{F_0} V - E_\gamma V\} \\ & - \frac{1}{2} \left\{ f_0' \left( \bar{K}_\gamma^{-1}(p) \right) E_{F_0} V^2 - f_0 \left( \bar{K}_\gamma^{-1}(p) \right) \frac{k_\gamma'}{k_\gamma} \left( \bar{K}_\gamma^{-1}(p) \right) E_\gamma V^2 \right\}, \end{aligned}$$



where  $E_{F_0}V$  and  $E_\gamma V$  refer to the right-hand side of (A.12) inserting  $\mu_i = EX_1^i$ ,  $F_0$  etc and  $\mu_i = EZ_\gamma^i$ ,  $K_\gamma$  etc., respectively, and  $E_{F_0}V^2$  and  $E_\gamma V^2$  refer to the right-hand side of (A.13) inserting  $\mu_i = EX_1^i$ ,  $F_0$  etc. and  $\mu_i = EZ_\gamma^i$ ,  $K_\gamma$  etc., respectively. It immediately follows that for  $F_0 = K_\gamma$  the second order approximation gives  $p$ .

The simulation results, the model error and the first and second order approximations are presented in Table 6.

*Table 6 Simulated expected (observed) false alarm rate using  $\hat{\gamma}_2^*$  as estimator with correction for the estimation of the mean, the variance and the parameter  $\gamma$  and corresponding first and second order approximation assuming the normal power family as model, with in the column ME the model error. The unit in the table is 0.001.*

$F_0$	simulation			1 <sup>st</sup> order appr.	ME	2 <sup>st</sup> order appr.		
	$n = 100$	$n = 250$	$n = 500$			$n = 100$	$n = 250$	$n = 500$
$\Phi$	1.23	1.05	1.02	1.00	0.00	1.00	1.00	1.00
$K_{-0.5}$	0.73	0.93	0.97	1.00	0.00	1.00	1.00	1.00
$K_{-0.25}$	0.98	1.01	1.00	1.00	0.00	1.00	1.00	1.00
$K_{0.25}$	1.18	1.05	1.01	1.00	0.00	1.00	1.00	1.00
$K_{0.5}$	1.41	1.06	1.01	1.00	0.00	1.00	1.00	1.00
$K_{0.75}$	1.45	1.05	1.01	1.00	0.00	1.00	1.00	1.00
$K_1$	1.53	1.07	1.01	1.00	0.00	1.00	1.00	1.00
$T$	2.85	3.02	3.04	3.08	2.08	2.74	2.97	3.01
$RM$	2.06	2.14	2.15	2.16	1.16	1.96	2.09	2.12
$DM$	2.10	2.19	2.23	2.28	1.28	1.92	2.15	2.21
$TU(-0.1)$	3.07	3.21	3.23	3.25	2.25	2.91	3.15	3.18
$TU(0)$	2.16	2.20	2.27	2.33	1.33	1.96	2.20	2.25
$TU(0.14)$	1.67	0.92	0.84	0.81	-0.19	0.92	0.85	0.83
$O$	0.90	0.78	0.73	0.69	-0.31	0.83	0.73	0.71

Table 6 shows that the correction for estimation of the parameters works very well, leading to  $EP$  (very) close to  $p$  for distributions from the normal power family. Comparison with Table 2 shows that the corrected control limit based on  $\hat{\gamma}_2^*$  performs much better than the corrected standard control limit based on normality. The stochastic error for  $\hat{\gamma}_2^*$  is (for  $n = 100$ , which is of main interest) lower than that of  $\hat{\gamma}_1^*$ . However, the model error is (sometimes much) larger and hence for larger  $n$  the control limit based on  $\hat{\gamma}_1^*$  may perform better outside the normal power family, see Tables 4 and 6. Nevertheless, both the model error and the stochastic error are reduced to an acceptable level by this more simple control limit. The second order approximation gives a very good prediction of the simulation, thus making it easy to analyze and predict the behavior of this control chart for all kind of distributions.

## 7 Recommendation and discussion

We start the discussion with answering the questions posed in Section 2. As restricted model we take normality, while for the general parametric model we consider the normal power family.

- (i) *Restricted model true*

**How large is the difference between the general  $SE$  and the restricted  $SE$ ? Can the restricted and general  $SE$  be reduced by (simple) corrections? How large are the corrected general  $SE$  and the corrected restricted  $SE$ ?**

The simulated expected restricted  $SE$  when applying the uncorrected control limit and its first and second order approximations are found in Table 1 on the row containing  $\Phi$  by subtracting  $p$  from the corresponding entry. Similarly, again applying the uncorrected control limit the simulated expected general  $SE$  and its approximations are found in Tables 3 and 5 on the rows containing  $\Phi$ . It is seen that without correction indeed the general  $SE$  is larger than the restricted  $SE$ .

However, by making the appropriate corrections the differences disappear: all the corrected versions are close to 0. Unfortunately, the corrections in the normal power family do not have such a simple form as under normality. At the end of this section a more simple correction term in case of the recommended estimator  $\hat{\gamma}_2^*$  will be discussed.

Apart from considering the bias one may also look at the variance. This will be treated in another paper.

We may conclude that the larger general  $SE$  in the uncorrected control limit can be accurately repaired by an appropriate correction term.

(ii) *General parametric model true*

**How bad can the restricted  $ME$  be? How does this balance with the larger general  $SE$ ? Can the restricted and general  $SE$  be reduced by (simple) corrections? How large are the corrected restricted  $SE$  and the corrected general  $SE$ , when the general model is true?**

The restricted  $ME$  is extensively discussed in Section 3 and furthermore numerically calculated in Tables 1 and 2. The conclusion is clear: the restricted  $ME$  can be quite large.

In the situation where no correction takes place we consider the balance between the larger general  $SE$  and the possibly large restricted  $ME$ . Therefore, we compare the results of Tables 3 and 5 with those of Table 1 for the distributions  $K_{-0.5}, K_{-0.25}, K_{0.25}, K_{0.5}, K_{0.75}$  and  $K_1$ . It is seen that the  $ME$  has a larger effect than the  $SE$ . Moreover, the  $SE$  can be corrected, while the  $ME$  remains.

The restricted  $SE$  refers to the situation where the supposed model is normality. Therefore, the corresponding correction terms are based on this assumption. The correction is not tailored to the normal power family. How well the correction still helps in these kinds of situations is seen by comparing the lines containing  $K_{-0.5}, K_{-0.25}, K_{0.25}, K_{0.5}, K_{0.75}$  and  $K_1$  in Table 2 and Table 1. The  $SE$  reduces also for these distributions, but due to the  $ME$  the total error is still (very) large.

The corrected general  $SE$  is quite small for the second estimator, based on the ordinary tail, see the lines containing  $K_{-0.5}, K_{-0.25}, K_{0.25}, K_{0.5}, K_{0.75}$  and  $K_1$  in Table 6. Also the correction for the first estimator, based on the fourth moment, works very well, see the lines containing  $K_{-0.5}, K_{-0.25}, K_{0.25}, K_{0.5}, K_{0.75}$  and  $K_1$  in Table 4.

(iii) *Outside the general parametric model*

**How large are the uncorrected and corrected general  $SE$  and the uncorrected and corrected restricted  $SE$ , when we are outside the general model and the general  $ME$  is not too big? How is the total error when applying the corrected control limit with as supposed model the general parametric one and how does this compare with the total error when applying the corrected control limit with as supposed model the restricted one.**

From the lines containing  $T, RM, DM, TU(-0.1), TU(0), TU(0.14), O$  in Tables 3, 4, 5 and 6 it is seen that the uncorrected general  $SE$  is substantially improved by application of the correction terms (even although they are derived for the parametric model). Also in the restricted case the correction is very helpful, see the lines containing  $T, RM, DM, TU(-0.1), TU(0), O$  in Tables 1 and 2.

With respect to the total error the lines containing  $T, RM, DM, TU(-0.1), TU(0), TU(0.14), O$  in Tables 1-6 show the following. The total error is often very large when using the classical normal control limit, due to a high model error (and a stochastic error, when no correction is applied). As a first step the total error is considerably reduced by application of the normal power family with, secondly, a further great improvement by taking the correction terms into account.

We conclude that the classical normal control limit is unreliable with often a very large total error, mainly due to the large model error. Although the correction for estimating the mean and the variance is very useful in reducing the stochastic error (even when normality does not hold), the model error often dominates and causes a large total error.

The (corrected) control limits based on the normal power family work very well, also outside the normal power family. The one based on  $\hat{\gamma}_2^*$  has the advantage of being far more simple, although the correction term is still rather complicated, and has a more reliable second order approximation, which make it possible to predict in an accurate way the behavior of this control chart under all kind of distributions.

### A more simple approximation

For practical implementation the correction term in the control limit given by (A.16) is rather complicated. Therefore, we present a more simple approximation, which can be applied straightforwardly, because the control limit is completely explicit. Basically, this control limit is derived from (A.16) by fitting polynomials in  $\gamma$  and  $u_p$  to

$$A_2, n \left[ EV - A_2 \left\{ \frac{F_0^{-1}(1 - q_n)}{F_0^{-1}(1 - r_n)} - \frac{F_0^{-1}(1 - q)}{F_0^{-1}(1 - r)} \right\} \right] \text{ and } n \times \text{coefficient of } \lambda.$$

This leads to the following control limit

$$\hat{\mu} + \hat{\sigma} \left\{ \bar{K}_{\hat{\gamma}}^{-1}(p) - C1(\hat{\gamma}) C2(\hat{\gamma}) - \frac{C3(\hat{\gamma})}{n} + \lambda \frac{C4(\hat{\gamma})}{n} \right\}, \quad (7.1)$$

where  $\hat{\mu} = n^{-1} \sum X_i$  and  $\hat{\sigma} = S = \sqrt{S^2}$  with  $S^2 = (n-1)^{-1} \sum (X_i - \bar{X})^2$ ,  $\bar{K}_{\gamma}^{-1}(p) = \pi^{1/4} 2^{-(1+\gamma)/2} \Gamma(\gamma + \frac{3}{2})^{-1/2} u_p^{1+\gamma}$  refers to the normal power family (see Section 3 (iii)),  $\lambda = 1, -1, 1 - \delta$  according to  $g(p) = p, g(p) = \frac{1}{p}, g(p) = 1 - (1-p)^k$ , respectively (see Section 5) and where

$$\hat{\gamma} = 1.1218 \log \left( \frac{X_{[0.95n+1]:n} - \bar{X}}{X_{[0.75n+1]:n} - \bar{X}} \right) - 1, \quad (7.2)$$

$$C1(\gamma) = -1.23 - 0.63\gamma + 0.73\gamma^2 + 0.74u_p - 0.08\gamma u_p - 0.14\gamma^2 u_p,$$

$$C2(\gamma) = \left( \frac{\Phi^{-1} \left( \frac{[0.95n+1]}{n+1} \right)}{\Phi^{-1} \left( \frac{[0.75n+1]}{n+1} \right)} \right)^{1+\gamma} - 2.4387^{1+\gamma},$$

$$C3(\gamma) = -10.86 - 27.77\gamma - 22.36\gamma^2 + 4.72u_p + 9.98\gamma u_p + 7.29\gamma^2 u_p,$$

$$C4(\gamma) = -87.23 - 147.89\gamma - 104.29\gamma^2 + 40.25u_p + 63.69\gamma u_p + 44.47\gamma^2 u_p.$$

The following table presents the simulation results, the model error and the first and second order approximations for the control limit given by (7.1) and (7.2).

Table 7 Simulated expected (observed) false alarm rate using  $\hat{\gamma}_2^*$  as estimator and the simplified version of the control limit with correction for the estimation of the mean, the variance and the parameter  $\gamma$  and corresponding first and second order approximation assuming the normal power family as model, with in the column *ME* the model error. The unit in the table is 0.001.

$F_0$	simulation			1 <sup>st</sup> order appr.	<i>ME</i>	2 <sup>st</sup> order appr.		
	$n = 100$	$n = 250$	$n = 500$			$n = 100$	$n = 250$	$n = 500$
$\Phi$	1.19	1.08	1.02	1.00	0.00	1.06	1.02	1.01
$K_{-0.5}$	0.76	0.94	0.97	1.00	0.00	1.06	1.02	1.01
$K_{-0.25}$	1.23	1.04	1.01	1.00	0.00	1.08	1.03	1.02
$K_{0.25}$	1.21	1.06	1.02	1.00	0.00	1.02	1.01	1.00
$K_{0.5}$	1.39	1.06	1.01	1.00	0.00	1.01	1.00	1.00
$K_{0.75}$	1.43	1.06	1.02	1.00	0.00	1.03	1.01	1.00
$K_1$	1.47	1.07	1.02	1.00	0.00	1.08	1.03	1.02
$T$	2.91	3.10	3.06	3.08	2.08	2.80	2.99	3.03
$RM$	2.09	2.18	2.15	2.16	1.16	2.02	2.11	2.13
$DM$	2.20	2.22	2.25	2.28	1.28	1.99	2.18	2.22
$TU(-0.1)$	3.11	3.29	3.21	3.25	2.25	2.96	3.17	3.19
$TU(0)$	2.22	2.31	2.31	2.33	1.33	2.02	2.22	2.26
$TU(0.14)$	1.70	0.93	0.85	0.81	-0.19	0.98	0.88	0.84
$O$	0.91	0.78	0.75	0.69	-0.31	0.86	0.75	0.72

The first order approximation and the model error are the same as in Tables 5 and 6. It is seen from Tables 6 and 7 that the simulation results and the second order approximation are quite similar. In this way the last step is made ending with a simple explicit control limit with small total error in the normal power family and reasonable total error for distributions outside this family.

**In view of the results of this paper we recommend the control limit based on the normal power family using as estimator  $\hat{\gamma}_2^*$  with the corresponding correction terms as presented in (7.1) and (7.2).**

## 8 Appendix

In this appendix we present the derivation and formulas of the control limits with correction terms based on the normal power family using as estimators  $\hat{\gamma}_1^*$  and  $\hat{\gamma}_2^*$ . Note that we take  $\mu = 0$  and  $\sigma = 1$  and do the calculations under  $F_0$ .

We start with  $\hat{\gamma}_1^*$ . Since  $\gamma_1^*$  is the suitably chosen point in the parameter space, we get  $\gamma_1^* = EX_1^4$ . In order to do the calculations for the correction terms we make the following quadratic approximation in  $\gamma$  (with exact fit at the points  $\gamma = -0.5, 0, 1$ )

$$\begin{aligned} \bar{K}_\gamma^{-1}(p) \doteq & (0.3849u_p^2 + 1.4927\sqrt{u_p} - 2u_p)\gamma^2 \\ & + (0.1925u_p^2 - 1.4927\sqrt{u_p} + u_p)\gamma + u_p. \end{aligned} \quad (\text{A.1})$$

The function  $h_1$  behaves as an exponential function and hence  $h_1^{-1}$  can be well approximated

by

$$h_1^{-1}(\gamma_1^*) \doteq 0.7363 \log\left(\frac{\gamma_1^*}{3}\right). \quad (\text{A.2})$$

Inserting (A.2) in (A.1) and applying a second order Taylor expansion gives the approximation

$$\overline{K}_{h_1^{-1}(\hat{\gamma}_1^*)}^{-1}(p) - \overline{K}_{h_1^{-1}(\gamma_1^*)}^{-1}(p) \doteq A_1(\hat{\gamma}_1^* - \gamma_1^*) + B_1(\hat{\gamma}_1^* - \gamma_1^*)^2$$

with

$$\begin{aligned} A_1 &= \frac{\log \gamma_1^*}{\gamma_1^*} (0.4173u_p^2 + 1.6185\sqrt{u_p} - 2.1686u_p) \\ &\quad + \frac{1}{\gamma_1^*} (-0.3168u_p^2 - 2.8772\sqrt{u_p} + 3.1188u_p), \\ B_1 &= \frac{\log \gamma_1^*}{(\gamma_1^*)^2} (-0.2087u_p^2 - 0.8093\sqrt{u_p} + 1.0843u_p) \\ &\quad + \frac{1}{(\gamma_1^*)^2} (0.3671u_p^2 + 2.2479\sqrt{u_p} - 2.6437u_p). \end{aligned}$$

For the calculation of  $EV$  we approximate  $V$  by

$$\hat{\mu} + (\hat{\sigma} - 1) \overline{K}_{h_1^{-1}(\hat{\gamma}_1^*)}^{-1}(p) + A_1(\hat{\gamma}_1^* - \gamma_1^*) + B_1(\hat{\gamma}_1^* - \gamma_1^*)^2 + (\hat{\sigma} - 1) A_1(\hat{\gamma}_1^* - \gamma_1^*), \quad (\text{A.3})$$

while in  $EV^2$  we approximate  $V^2$  by

$$\begin{aligned} \hat{\mu}^2 + (\hat{\sigma} - 1)^2 \left( \overline{K}_{h_1^{-1}(\hat{\gamma}_1^*)}^{-1}(p) \right)^2 + A_1^2(\hat{\gamma}_1^* - \gamma_1^*)^2 + 2\hat{\mu}(\hat{\sigma} - 1) \overline{K}_{h_1^{-1}(\hat{\gamma}_1^*)}^{-1}(p) \\ + 2\hat{\mu}A_1(\hat{\gamma}_1^* - \gamma_1^*) + 2(\hat{\sigma} - 1) \overline{K}_{h_1^{-1}(\hat{\gamma}_1^*)}^{-1}(p) A_1(\hat{\gamma}_1^* - \gamma_1^*). \end{aligned} \quad (\text{A.4})$$

The following formulas give approximations for the moments of the estimators  $\hat{\mu}, \hat{\sigma}$  and  $\hat{\gamma}_1^*$  assuming that  $X_1, \dots, X_n$  are i.i.d. r.v.'s with distribution function  $F_0$ , thus having expectation 0 and variance 1. For the derivation of the correction terms in case of the normal power family, we may restrict to  $F_0 = K_\gamma$  for some  $\gamma$ . However, we need more general formulas in Section 6, and therefore we only assume expectation 0 and variance 1 (and existence of further appropriate moments). Writing  $\mu_i = EX_1^i$ , we have

$$\begin{aligned} E\hat{\mu} &= 0, \quad E(\hat{\sigma} - 1) \doteq -\frac{\mu_4 - 1}{8n}, \\ E(\hat{\gamma}_1^* - \mu_4) &\doteq \frac{-2\mu_6 + 3\mu_4^2 - 5\mu_4 + 8\mu_3^2 + 6}{n}, \\ E\hat{\mu}^2 &= \frac{1}{n}, \quad E(\hat{\sigma} - 1)^2 \doteq \frac{\mu_4 - 1}{4n}, \\ E(\hat{\gamma}_1^* - \mu_4)^2 &\doteq \frac{\mu_8 - 4\mu_4\mu_6 + 4\mu_4^3 - \mu_4^2 + 16\mu_3^2 - 8\mu_3\mu_5 + 16\mu_3^2\mu_4}{n}, \\ E\hat{\mu}(\hat{\sigma} - 1) &\doteq \frac{\mu_3}{2n}, \quad E\hat{\mu}(\hat{\gamma}_1^* - \mu_4) \doteq \frac{\mu_5 - 4\mu_3 - 2\mu_4\mu_3}{n}, \\ E(\hat{\sigma} - 1)(\hat{\gamma}_1^* - \mu_4) &\doteq \frac{\mu_6 - \mu_4}{2n} - \frac{\mu_4(\mu_4 - 1)}{n} - \frac{2\mu_3^2}{n}. \end{aligned} \quad (\text{A.5})$$

Hence, we get (note that by definition of the "suitably chosen parameter"  $\gamma_1^* = \mu_4$ )

$$\begin{aligned}
EV &\doteq -\frac{\mu_4 - 1}{8n} \overline{K}_{h_1^{-1}(\gamma_1^*)}^{-1}(p) \\
&+ A_1 \frac{-2\mu_6 + 3\mu_4^2 - 5\mu_4 + 8\mu_3^2 + 6}{n} \\
&+ B_1 \frac{\mu_8 - 4\mu_4\mu_6 + 4\mu_4^3 - \mu_4^2 + 16\mu_3^2 - 8\mu_3\mu_5 + 16\mu_3^2\mu_4}{n} \\
&+ A_1 \left\{ \frac{\mu_6 - \mu_4}{2n} - \frac{\mu_4(\mu_4 - 1)}{n} - \frac{2\mu_3^2}{n} \right\}
\end{aligned} \tag{A.6}$$

and

$$\begin{aligned}
EV^2 &\doteq \frac{1}{n} + \frac{\mu_4 - 1}{4n} \left( \overline{K}_{h_1^{-1}(\gamma_1^*)}^{-1}(p) \right)^2 \\
&+ A_1^2 \frac{\mu_8 - 4\mu_4\mu_6 + 4\mu_4^3 - \mu_4^2 + 16\mu_3^2 - 8\mu_3\mu_5 + 16\mu_3^2\mu_4}{n} \\
&+ \frac{\mu_3}{n} \overline{K}_{h_1^{-1}(\hat{\gamma}_1^*)}^{-1}(p) + 2A_1 \frac{\mu_5 - 4\mu_3 - 2\mu_4\mu_3}{n} \\
&+ 2\overline{K}_{h_1^{-1}(\gamma_1^*)}^{-1}(p) A_1 \left\{ \frac{\mu_6 - \mu_4}{2n} - \frac{\mu_4(\mu_4 - 1)}{n} - \frac{2\mu_3^2}{n} \right\}.
\end{aligned} \tag{A.7}$$

Returning to the normal power family, we have  $\mu_i = 0$  for odd  $i$  and

$$\begin{aligned}
\mu_4 = \gamma_1^* &= EZ_\gamma^4 = \frac{\sqrt{\pi}\Gamma(2\gamma + \frac{5}{2})}{\Gamma(\gamma + \frac{3}{2})^2}, \\
\mu_6 &= EZ_\gamma^6 = \frac{\pi\Gamma(3\gamma + \frac{7}{2})}{\Gamma(\gamma + \frac{3}{2})^3}, \\
\mu_8 &= EZ_\gamma^8 = \frac{\pi^{3/2}\Gamma(4\gamma + \frac{9}{2})}{\Gamma(\gamma + \frac{3}{2})^4}.
\end{aligned}$$

The estimated versions of  $EV$  and  $EV^2$  are

$$\begin{aligned}
\widehat{EV} &\doteq -\frac{\hat{\gamma}_1^* - 1}{8n} \overline{K}_{h_1^{-1}(\hat{\gamma}_1^*)}^{-1}(p) + \hat{A}_1 \frac{-2\hat{\mu}_6 + 3(\hat{\gamma}_1^*)^2 - 5\hat{\gamma}_1^* + 6}{n} \\
&+ \hat{B}_1 \frac{\hat{\mu}_8 - 4\hat{\gamma}_1^*\hat{\mu}_6 + 4(\hat{\gamma}_1^*)^3 - (\hat{\gamma}_1^*)^2}{n} + \hat{A}_1 \left\{ \frac{\hat{\mu}_6 - \hat{\gamma}_1^*}{2n} - \frac{\hat{\gamma}_1^*(\hat{\gamma}_1^* - 1)}{n} \right\}
\end{aligned} \tag{A.8}$$

and

$$\begin{aligned}
\widehat{EV}^2 &\doteq \frac{1}{n} + \frac{\hat{\gamma}_1^* - 1}{4n} \left( \overline{K}_{h_1^{-1}(\hat{\gamma}_1^*)}^{-1}(p) \right)^2 + \hat{A}_1^2 \frac{\hat{\mu}_8 - 4\hat{\gamma}_1^*\hat{\mu}_6 + 4(\hat{\gamma}_1^*)^3 - (\hat{\gamma}_1^*)^2}{n} \\
&+ 2\overline{K}_{h_1^{-1}(\hat{\gamma}_1^*)}^{-1}(p) \hat{A}_1 \left\{ \frac{\hat{\mu}_6 - \hat{\gamma}_1^*}{2n} - \frac{\hat{\gamma}_1^*(\hat{\gamma}_1^* - 1)}{n} \right\},
\end{aligned} \tag{A.9}$$

where  $\hat{A}_1$  and  $\hat{B}_1$  are obtained from  $A_1$  and  $B_1$ , respectively, by inserting  $\hat{\gamma}_1^*$  for  $\gamma_1^*$ , and where  $\hat{\mu}_6$  and  $\hat{\mu}_8$  are obtained from  $\mu_6$  and  $\mu_8$ , respectively, by inserting  $h_1^{-1}(\hat{\gamma}_1^*)$  for  $\gamma$ . The control limit including the correction term is given by

$$\hat{\mu} + \hat{\sigma} \left\{ \overline{K}_{h_1^{-1}(\hat{\gamma}_1^*)}^{-1}(p) - \widehat{EV} + \frac{\lambda \widehat{EV}^2}{2} \frac{u_p^2 + h_1^{-1}(\hat{\gamma}_1^*)}{\{1 + h_1^{-1}(\hat{\gamma}_1^*)\} c(h_1^{-1}(\hat{\gamma}_1^*)) u_p^{1+h_1^{-1}(\hat{\gamma}_1^*)}} \right\} \tag{A.10}$$

with  $\widehat{EV}$  and  $\widehat{EV}^2$  replaced by the right-hand sides of (A.8) and (A.9), respectively.

**Remark A.1** To represent the restricted model of normality take in (A.10)  $\hat{\gamma}_1^*$  identically equal to  $\gamma_1^* = 3$  and hence  $h_1^{-1}(\hat{\gamma}_1^*)$  identically equal to 0, according to the fact that we do not need to estimate  $\gamma$ . Moreover, by the same reason disregard the contribution of  $\overline{K}_{h_1^{-1}(\hat{\gamma}_1^*)}^{-1}(p) - \overline{K}_{h_1^{-1}(\gamma_1^*)}^{-1}(p) \doteq A_1(\hat{\gamma}_1^* - \gamma_1^*) + B_1(\hat{\gamma}_1^* - \gamma_1^*)^2$  in  $\widehat{EV}$  and  $\widehat{EV}^2$ . Then (A.10) reduces to  $\hat{\mu} + \hat{\sigma}u_p + (4n)^{-1}\hat{\sigma}u_p\{1 + \lambda(u_p^2 + 2)\}$ , which, (as it should be!) is the correction term for the control limit based on normality, cf. Albers and Kallenberg (2000). (Note that here we use  $\hat{\sigma} = S$  and not its bias-corrected version, explaining here the “extra” term  $1/(4n)$  compared to (3.6) in Albers and Kallenberg (2000).).

For the second estimator  $\hat{\gamma}_2^*$ , given in (4.3), insertion of (4.6) in (A.1) and application of a second order Taylor expansion gives the approximation

$$\overline{K}_{h_2^{-1}(\hat{\gamma}_2^*)}^{-1}(p) - \overline{K}_{h_2^{-1}(\gamma_2^*)}^{-1}(p) \doteq A_2(\hat{\gamma}_2^* - \gamma_2^*) + B_2(\hat{\gamma}_2^* - \gamma_2^*)^2$$

with

$$A_2 = \frac{\log \gamma_2^*}{\gamma_2^*} \left\{ \frac{0.7698u_p^2 + 2.9854\sqrt{u_p} - 4u_p}{\left(\log \frac{u_q}{u_r}\right)^2} \right\} + \frac{1}{\gamma_2^*} \left\{ \frac{-0.5773u_p^2 - 4.4781\sqrt{u_p} + 5u_p}{\log \frac{u_q}{u_r}} \right\}$$

and

$$B_2 = \frac{\log \gamma_2^*}{(\gamma_2^*)^2} \left\{ \frac{-0.3849u_p^2 - 1.4927\sqrt{u_p} + 2u_p}{\left(\log \frac{u_q}{u_r}\right)^2} \right\} + \frac{1}{(\gamma_2^*)^2} \left\{ \frac{0.3849u_p^2 + 1.4927\sqrt{u_p} - 2u_p}{\left(\log \frac{u_q}{u_r}\right)^2} + \frac{0.2887u_p^2 + 2.2391\sqrt{u_p} - 2.5u_p}{\log \frac{u_q}{u_r}} \right\}.$$

The following formulas give approximations for the moments of the estimators  $\hat{\mu}, \hat{\sigma}$  and  $\hat{\gamma}_2^*$  assuming that  $X_1, \dots, X_n$  are i.i.d. r.v.'s with distribution function  $F_0$ , thus having expectation 0 and variance 1. We need these more general formulas in Section 6. Writing  $x_q = F_0^{-1}(1 - q)$ ,  $q_n = 1 - \frac{[n+1-qn]}{n+1}$ , we have for  $q \leq r$  (if  $q > r$  replace  $q(1 - r)$  by  $r(1 - q)$ )

$$E(\hat{\gamma}_2^* - \gamma_2^*) \doteq \frac{F_0^{-1}(1 - q_n)}{F_0^{-1}(1 - r_n)} - \frac{F_0^{-1}(1 - q)}{F_0^{-1}(1 - r)} - \frac{q(1 - q)}{2nx_r} \frac{f_0'(x_q)}{f_0^3(x_q)} + \frac{x_q r(1 - r)}{2nx_r^2} \frac{f_0'(x_r)}{f_0^3(x_r)} - \frac{1}{nx_r^2} \left\{ \frac{q(1 - r)}{f_0(x_q)f_0(x_r)} - \frac{\int_{x_q}^{\infty} yf_0(y)dy}{f_0(x_q)} - \frac{\int_{x_r}^{\infty} yf_0(y)dy}{f_0(x_r)} + 1 \right\} + \frac{x_q}{nx_r^3} \left\{ \frac{r(1 - r)}{[f_0(x_r)]^2} - 2 \frac{\int_{x_r}^{\infty} yf_0(y)dy}{f_0(x_r)} + 1 \right\},$$

$$\begin{aligned}
E(\hat{\gamma}_2^* - \gamma_2^*)^2 &\doteq \frac{1}{nx_r^2} \left\{ \frac{q(1-q)}{[f_0(x_q)]^2} - 2 \frac{\int_{x_q}^{\infty} y f_0(y) dy}{f_0(x_q)} + 1 \right\} \\
&+ \frac{x_q^2}{nx_r^4} \left\{ \frac{r(1-r)}{[f_0(x_r)]^2} - 2 \frac{\int_{x_r}^{\infty} y f_0(y) dy}{f_0(x_r)} + 1 \right\} \\
&- \frac{2x_q}{nx_r^3} \left\{ \frac{q(1-r)}{f_0(x_q) f_0(x_r)} - \frac{\int_{x_q}^{\infty} y f_0(y) dy}{f_0(x_q)} - \frac{\int_{x_r}^{\infty} y f_0(y) dy}{f_0(x_r)} + 1 \right\}, \\
E\hat{\mu}(\hat{\gamma}_2^* - \gamma_2^*) &\doteq \frac{1}{nx_r} \left\{ \frac{\int_{x_q}^{\infty} y f_0(y) dy}{f_0(x_q)} - 1 - \frac{x_q}{x_r} \left[ \frac{\int_{x_r}^{\infty} y f_0(y) dy}{f_0(x_r)} - 1 \right] \right\}, \\
E(\hat{\sigma} - 1)(\hat{\gamma}_2^* - \gamma_2^*) &\doteq \frac{1}{2nx_r} \left\{ \frac{\int_{x_q}^{\infty} (y^2 - 1) f_0(y) dy}{f_0(x_q)} - \mu_3 \right\} \\
&- \frac{x_q}{2nx_r^3} \left\{ \frac{\int_{x_r}^{\infty} (y^2 - 1) f_0(y) dy}{f_0(x_r)} - \mu_3 \right\}.
\end{aligned} \tag{A.11}$$

Approximating  $V$  in a similar way as in (A.3) we get

$$EV \doteq -\frac{\mu_4 - 1}{8n} \bar{K}_{h_2^{-1}(\gamma_2^*)}^{-1}(p) + A_2(C_{21} + C_{22}) + B_2 C_{23} \tag{A.12}$$

with

$$\begin{aligned}
C_{21} &= \frac{F_0^{-1}(1-q_n)}{F_0^{-1}(1-r_n)} - \frac{F_0^{-1}(1-q)}{F_0^{-1}(1-r)} - \frac{q(1-q)}{2nx_r} \frac{f_0'(x_q)}{f_0^3(x_q)} + \frac{x_q r(1-r)}{2nx_r^2} \frac{f_0'(x_r)}{f_0^3(x_r)} \\
&- \frac{1}{nx_r^2} \left\{ \frac{q(1-r)}{f_0(x_q) f_0(x_r)} - \frac{\int_{x_q}^{\infty} y f_0(y) dy}{f_0(x_q)} - \frac{\int_{x_r}^{\infty} y f_0(y) dy}{f_0(x_r)} + 1 \right\} \\
&+ \frac{x_q}{nx_r^3} \left\{ \frac{r(1-r)}{[f_0(x_r)]^2} - 2 \frac{\int_{x_r}^{\infty} y f_0(y) dy}{f_0(x_r)} + 1 \right\}, \\
C_{22} &= \frac{1}{2nx_r} \left\{ \frac{\int_{x_q}^{\infty} (y^2 - 1) f_0(y) dy}{f_0(x_q)} - \mu_3 \right\} - \frac{x_q}{2nx_r^2} \left\{ \frac{\int_{x_r}^{\infty} (y^2 - 1) f_0(y) dy}{f_0(x_r)} - \mu_3 \right\}
\end{aligned}$$



and

$$C_{23} = \frac{1}{nx_r^2} \left\{ \frac{q(1-q)}{[f_0(x_q)]^2} - 2 \frac{\int_{x_q}^{\infty} y f_0(y) dy}{f_0(x_q)} + 1 \right\} \\ + \frac{x_q^2}{nx_r^4} \left\{ \frac{r(1-r)}{[f_0(x_r)]^2} - 2 \frac{\int_{x_r}^{\infty} y f_0(y) dy}{f_0(x_r)} + 1 \right\} \\ - \frac{2x_q}{nx_r^3} \left\{ \frac{q(1-r)}{f_0(x_q) f_0(x_r)} - \frac{\int_{x_q}^{\infty} y f_0(y) dy}{f_0(x_q)} - \frac{\int_{x_r}^{\infty} y f_0(y) dy}{f_0(x_r)} + 1 \right\}.$$

By a similar approximation as in (A.4) for  $V^2$  we have

$$EV^2 \doteq \frac{1}{n} + \frac{\mu_4 - 1}{4n} \left( \overline{K}_{h_2^{-1}(\hat{\gamma}_2^*)}^{-1}(p) \right)^2 + \frac{\mu_3}{n} \overline{K}_{h_2^{-1}(\hat{\gamma}_2^*)}^{-1}(p) + A_2^2 C_{23} \\ + 2A_2 \left( C_{24} + \overline{K}_{h_2^{-1}(\hat{\gamma}_2^*)}^{-1}(p) C_{22} \right) \quad (\text{A.13})$$

with

$$C_{24} = \frac{1}{nx_r} \left\{ \frac{\int_{x_q}^{\infty} y f_0(y) dy}{f_0(x_q)} - 1 - \frac{x_q}{x_r} \left[ \frac{\int_{x_r}^{\infty} y f_0(y) dy}{f_0(x_r)} - 1 \right] \right\}.$$

Returning to the normal power family, we have  $\mu_3 = 0$  and  $\mu_4 = \gamma_1^*$  given by (4.1). Moreover,

$$x_q = c(\gamma) u_q^{1+\gamma}, \quad f_0(x_q) = \frac{\varphi(u_q)}{c(\gamma)(1+\gamma)u_q^\gamma}, \quad f_0'(x_q) = - \frac{\varphi(u_q) \{u_q^{1-2\gamma} + \gamma u_q^{-1-2\gamma}\}}{\{c(\gamma)(1+\gamma)\}^2}, \\ \int_{x_q}^{\infty} y f_0(y) dy = EZ_\gamma 1_{Z_\gamma > x_q} = c(\gamma) EZ^{1+\gamma} 1_{Z > u_q} = c(\gamma) \int_{u_q}^{\infty} x^{1+\gamma} \varphi(x) dx \\ \int_{x_q}^{\infty} (y^2 - 1) f_0(y) dy = E(Z_\gamma^2 - 1) 1_{Z_\gamma > x_q} = E \left\{ [c(\gamma) Z^{1+\gamma}]^2 - 1 \right\} 1_{Z > u_q} \\ = \int_{u_q}^{\infty} \left\{ [c(\gamma) x^{1+\gamma}]^2 - 1 \right\} \varphi(x) dx$$

The estimated versions of  $EV$  and  $EV^2$  are

$$\widehat{EV} \doteq - \frac{\hat{\mu}_4 - 1}{8n} \overline{K}_{h_2^{-1}(\hat{\gamma}_2^*)}^{-1}(p) + \hat{A}_2 \left( \hat{C}_{21} + \hat{C}_{22} \right) + \hat{B}_2 \hat{C}_{23} \quad (\text{A.14})$$

and

$$\widehat{EV}^2 \doteq \frac{1}{n} + \frac{\hat{\mu}_4 - 1}{4n} \left( \overline{K}_{h_2^{-1}(\hat{\gamma}_2^*)}^{-1}(p) \right)^2 + \hat{A}_2^2 \hat{C}_{23} + 2\hat{A}_2 \left( \hat{C}_{24} + \overline{K}_{h_2^{-1}(\hat{\gamma}_2^*)}^{-1}(p) \hat{C}_{22} \right), \quad (\text{A.15})$$

where  $\hat{A}_2, \hat{B}_2$ , etc. are obtained from  $A_2, B_2$ , etc. by inserting  $\hat{\gamma}_2^*$  for  $\gamma_2^*$ , and where  $\hat{\mu}_4$  is obtained from  $\mu_4$  by inserting  $h_2^{-1}(\hat{\gamma}_2^*)$  for  $\gamma$ . The control limit including the correction term is given by

$$\hat{\mu} + \hat{\sigma} \left\{ \overline{K}_{h_2^{-1}(\hat{\gamma}_2^*)}^{-1}(p) - \widehat{EV} + \frac{\lambda \widehat{EV}^2}{2} \frac{u_p^2 + h_2^{-1}(\hat{\gamma}_2^*)}{\{1 + h_2^{-1}(\hat{\gamma}_2^*)\} c(h_2^{-1}(\hat{\gamma}_2^*)) u_p^{1+h_2^{-1}(\hat{\gamma}_2^*)}} \right\} \quad (\text{A.16})$$

with  $\widehat{EV}$  and  $\widehat{EV}^2$  replaced by the right-hand sides of (A.14) and (A.15), respectively.

**Remark A.2** The same considerations as in Remark A.1 lead also in (A.16) to the correction term in the restricted model of normality.

## References

- Albers, W. and Kallenberg, W.C.M. (2000). Estimation in Shewhart control charts. Technical Report 1559, University of Twente.
- Albers, W. and Kallenberg, W.C.M. (2001). Are estimated control charts in control? Technical Report 1569, University of Twente.
- Andrews, D.F. (1973). A general method for the approximation of tail areas. *Ann. Statist.* **1** 367-372.
- Chan, L.K., Hapuarachchi, K.P. and Macpherson, B.D. (1988). Robustness of  $\overline{X}$  and  $R$  charts. *IEEE Trans. Reliability* **37** 117-123.
- Dekkers, A.L.M. and de Haan L. (1989). On the estimation of the extreme-value index and large quantile estimation. *Ann. Statist.* **17** 1795-1832.
- Hall, P. and Weissman, I. (1997). On the estimation of extreme tail probabilities. *Ann. Statist.* **25** 1311-1326.
- Hušková, M. (1988). Adaptive parameter estimation for generalized Tukey's  $\lambda$ -family. *Statistics* **19** 15-26.
- Pappanastos, E.A. and Adams, B.M. (1996). Alternative designs of the Hodges-Lehmann control chart. *J. Qual. Technol.* **28** 213-223.