Accurate test limits using the bootstrap

W. Albers, W.C.M. Kallenberg, R. Schmölzer
Department of Applied Mathematics
University of Twente
P.O. Box 217, 7500 AE Enschede
The Netherlands

Abstract During inspection, mass-produced items are compared to given specification limits. The presence of measurement errors forces to set test limits which are slightly more strict. It is important to make this choice accurately: test limits which are too strict cause unnecessary loss of yield, whereas choices which are too liberal lead to consumer losses exceeding prescribed bounds. From earlier work, such accurate test limits are available. Using rather elaborate corrections, these limits have also been adapted to the often more realistic case where (some of) the underlying parameters need to be estimated. In this paper it is investigated how well such corrections can be achieved using the bootstrap method instead.

Keywords and phrases: specification limits, consumer loss, inspection.

1991 Mathematics Subject Classification: 62 E 20, 62 E 25, 62 N 10

1 Introduction

In many large scale production processes, like those occurring in semiconductor industry, relations between process characteristics on one hand and product characteristics on the other, are not sufficiently well understood to successfully apply methods which control and reduce variability. Instead one has to rely on inspection: each item produced is measured and compared to a given specification limit. Typically, the measurement process is not infallible: we measure

$$\tilde{X} = X + U, \tag{1.1}$$

where X is the true value of the characteristic measured and U is the measurement error. According to standard assumptions, X and U are independent and normally distributed: $X \sim N(\mu_X, \sigma_X^2)$ and $U \sim N(0, \sigma_U^2)$, with typically $\sigma = \sigma_U/\sigma_X \ll 1$.

Without loss of generality we assume that a product is nonconforming if X exceeds a given specification limit s. Common values for $\pi = P(X > s)$ are in the range from 0.1% to 15%; we shall suppose at least that $s > \mu_X$. In view of the presence of U, products are only allowed to pass inspection if $\tilde{X} < t$, where the test limit t is slightly more strict than s. The selection of t is a delicate matter. Usually an upper bound γ, ranging from e.g. 1 to 100 ppm(= parts per million), is prescribed for the consumer loss

$$CL(t) = P(X > s, \tilde{X} < t), \tag{1.2}$$

which is the probability that a product is both nonconforming and accepted. (In certain applications $P(\tilde{X} < t)$ can become substantially lower than 1. Then one may prefer to use $CR = P(X > s|\tilde{X} < t)$ instead of CL, see Albers, Arts and Kallenberg (1995) for results in this case.) If t is too close to s, the requirement $CL(t) \leq \gamma$ will be violated; however, if t is too conservative, the yield $YD = P(\tilde{X} < t)$ will be lowered unnecessarily, which can be quite costly. Informal approaches used in practice, like setting $t = 0.9s$ or $t = s - 3\sigma_U$, clearly cannot cope with this problem.

What we are actually looking for is of course t such that $CL(t) = \gamma$. This exact solution can be obtained numerically, using that X and $\tilde{X}$ have a bivariate normal distribution: $(X, \tilde{X}) \sim N(\mu_X, \mu_X, \sigma_X^2, \sigma_X^2 + \sigma_U^2, \rho)$, with $\rho = (1 + \sigma_U^2/\sigma_X^2)^{-\frac{1}{2}}$. However, an easier approach is possible, as will become transparent shortly. Instead we shall base ourselves on the accurate approximations derived by Albers, Kallenberg and Otten (1994a). Let

$$\sigma = \frac{\sigma_U}{\sigma_X},\ Y = -\frac{U}{\sigma_U},\ \bar{X} = \frac{X - \mu_X}{\sigma_X},\ \bar{s} = \frac{s - \mu_X}{\sigma_X}, \tag{1.3}$$

then $\bar{X}$ and Y are independent and standard normal, while $\bar{s} > 0$. Together with (1.1) and (1.2) it follows that, for $a = (s - t)/\sigma_U$,

$$\begin{aligned} CL(t) &= P(Y > a, \bar{s} < \bar{X} < \bar{s} + \sigma(Y - a)) \\ &= \int_a^\infty \{\Phi(\bar{s} + \sigma(y - a)) - \Phi(\bar{s})\}\phi(y)dy, \end{aligned} \tag{1.4}$$

where Φ and ϕ stand for the standard normal distribution function (df) and its density, respectively.

As σ is small, the integral in (1.4) can be expanded, leading to

$$CL(t) = \sigma\phi(\bar{s})g_1(a) + \tfrac{1}{2}\sigma^2\phi'(\bar{s})g_2(a) + \ldots, \tag{1.5}$$

where $g_k(a) = \int_a^\infty (y-a)^k\phi(y)dy,\ k = 0, 1, 2, \ldots$. In this way, Albers, Kallenberg and Otten (1994a) derive the following j-th order ($j = 1, 2$) approximations $t_j = s - a_j\sigma_U$ to the exact t (satisfying $CL(t) = \gamma$)

$$a_1 = g_1^{-1}\left(\frac{\gamma}{\sigma\phi(\bar{s})}\right)\ ,\ a_2 = a_1 - \frac{\sigma\bar{s}}{2}(a_1^2 + 1 - a_1k(a_1)), \tag{1.6}$$

in which $k = \phi/(1 - \Phi)$. It turns out that especially a_2 is extremely accurate: $(CL(t_2) - \gamma)/\gamma$ is completely negligible for practical purposes.

However, all of the above relies on the often unrealistic assumption that μ_X, σ_X and σ_U are known. In practice we need estimators $\hat{\mu}_X, \hat{\sigma}_X$ and $\hat{\sigma}_U$ for these parameters, leading to stochastic test limits $\hat{t}_j = s - \hat{a}_j\hat{\sigma}_U$, where $\hat{a}_j = a_j(\hat{\mu}_X, \hat{\sigma}_X, \hat{\sigma}_U)$. The effect of plugging in estimators instead of true values, is considerable, unless huge sample sizes are involved. Unlike with numerically obtained results, the explicit structure of the approximations given by (1.6) allows to correct for this effect. In fact, let $\hat{c} = c(\hat{\mu}_X, \hat{\sigma}_X, \hat{\sigma}_U)$ be a - typically small - correction to $\hat{a}_j$, resulting in a test limit $\hat{t} = s - \hat{\sigma}_U(\hat{a}_j + \hat{c})$, then

$$CL(\hat{t}) = CL(\hat{t}_j) + \sigma\phi(\bar{s})\hat{c}g_1'(\hat{a}_1) + \ldots. \tag{1.7}$$

Now suppose for example that we want to apply a bias correction. In principle, we could apply it to $\hat{t}_2$. However, this seems to be a bit half-hearted, as our criterion stipulates that $\widehat{CL} = CL(\hat{t})$ should be as close as possible to γ. Therefore, we correct for the bias $ECL(\hat{t}_2) - \gamma$ caused by applying the plug-in method to t_2, and not for the bias of the test limit itself. Nevertheless, as a correction for the bias $E\hat{t}_2 - t_2$ may be easier to obtain than one for $ECL(\hat{t}_2) - \gamma$, one might want to use the first as a simple step towards the second. Note, however, that $ECL(\hat{t}_2) = \gamma + (E\hat{t}_2 - t_2)CL'(t_2) + \frac{1}{2}E(\hat{t}_2 - t_2)^2CL''(t_2) + \ldots$, which reveals that correction for $ECL(\hat{t}_2) - \gamma$ not only requires the bias of $\hat{t}$, but also its MSE, as these two quantities are typically of the same order. Hence, we shall focus here

on direct correction for the bias $ECL(\hat{t}_2) - \gamma$. From (1.7) it is immediate that for this purpose we should select (using that $g_1'(x) = -(1 - \Phi(x))$)

$$c = c(\mu_X, \sigma_X, \sigma_U) = \frac{ECL(\hat{t}_2) - \gamma}{\sigma\phi(\bar{s})(1 - \Phi(a_1))}\,. \tag{1.8}$$

At this point it is important to notice the different nature of numerator and denominator in (1.8). The denominator is given explicitly and, using (1.3) and (1.6), the step from c to $\hat{c}$ is easily made by going to $(\hat{\sigma}_U/\hat{\sigma}_X)\phi((s - \hat{\mu}_X)/\hat{\sigma}_X) \times (1 - \Phi(\hat{a}_1))$. The numerator, however, is still unknown. In Albers, Kallenberg and Otten (1994a), considerable effort is needed to obtain an explicit expansion which approximates this bias to sufficient precision. The bias is subsequently estimated by replacing μ_X, σ_X and σ_U in this expansion by the corresponding estimators.

This phenomenon is not special for the example above, but occurs for each type of correction applied through (1.7). Hence it seems worthwhile to obtain an easier alternative for the difficult part in the procedure, which is the estimation of the numerator of (1.8). In this paper we shall investigate the possibilities of the bootstrap method for this purpose. That this method can indeed offer an easier way, is evident from the previous example: obtaining a bootstrap estimator for the bias of $CL(\hat{t}_2)$ is rather straightforward. Besides the simplicity, an additional advantage of bootstrapping is its generality: if new situations arise or new types of corrections are called for, a recipe like (1.7) remains applicable, whereas an explicit evaluation of approximate correction terms has to be performed anew for each additional application. To balance matters, however, we should remark that each application of the bootstrap requires the determination of an appropriate resampling scheme, as well as the verification, empirically and/or theoretically, that the method indeed works.

The paper is organized as follows. Sections 2 and 3 are devoted to bias correction. In section 2 the special case is considered where only σ_U is unknown, while section 3 deals with the general case. Section 4 addresses the situation where the correction is meant to ensure that $P(CL(\hat{t}) > \gamma) \leq \alpha$ for some small, prescribed α. Some material on the theoretical background has been gathered in an Appendix.

2 Bias correction under unknown error variance

The study of the case where μ_X, σ_X^2 and σ_U^2 are all unknown, will be postponed till section 3. Here we shall consider the case where only the measurement error variance σ_U^2 is unknown. There are two reasons for doing this, the first of which is quite obvious: the exposition in the one-parameter case is more simple and

hence more transparent. But another aspect is that this case is of interest in its own right. The estimation of σ_U^2 requires a series of duplicated measurements

$$X_{ij} = X_i + U_{ij},\ i = 1, \ldots, n,\ j = 1, 2. \tag{2.1}$$

These test data typically form part of a set of production data $\tilde{X}_i = X_i + U_i,\ i = 1, \ldots, m$ (cf.(1.1)). Quite often, m is so much larger than n that the sampling error in $\mu_{\tilde{X}}$ and $\sigma_{\tilde{X}}^2$ is negligible. As $\mu_X = \mu_{\tilde{X}}$ and $\sigma_X^2 = \sigma_{\tilde{X}}^2 - \sigma_U^2$, while $\sigma_U \ll \sigma_X$, it follows that μ_X and σ_X^2 can be assumed to be known in this situation. Then it indeed remains to estimate σ_U^2, which is done by

$$\hat{\sigma}_U^2 = \frac{1}{2n} \sum_{i=1}^{n} (X_{i2} - X_{i1})^2, \tag{2.2}$$

where $n\hat{\sigma}_U^2/\sigma_U^2$ is $\chi^2(n)$-distributed.

For known μ_X and σ_X, we can without loss of generality assume that $\mu_X = 0$ and $\sigma_X = 1$. It then follows from (1.3) that $\sigma = \sigma_U, \bar{X} = X$ and $\bar{s} = s$. Moreover, for $a_j (j = 1, 2)$ from (1.6) we now simply have $a_j = a_j(\sigma), \hat{a}_j = a_j(\hat{\sigma})$. To keep track of the various standard deviations that play a role, it will be convenient to introduce the following notation here:

$$CL(\sigma_1, \sigma_2) = P(X > s, X - \sigma_1 Y < s - \sigma_2 a_2(\sigma_2)), \tag{2.3}$$

where X and $Y = -U/\sigma_1$ are independent and standard normal (cf.(1.3)). Hence, in the notation of section 1, we have $CL(t_2) = CL(\sigma, \sigma)$ and $CL(\hat{t}_2) = CL(\sigma, \hat{\sigma})$. Moreover, we know from that section that $CL(\sigma_1, \sigma_2)$ can be evaluated through (1.6) and the $N(0, 0, 1, 1, (1 + \sigma_1^2)^{-\frac{1}{2}})$-distribution, while $CL(\sigma_j, \sigma_j) = \gamma$, $j = 1, 2$, to high precision. Finally, do observe that by definition $CL(\sigma, \hat{\sigma})$ is the conditional probability $P(X > s, X - \sigma Y < s - \hat{\sigma} a_2(\hat{\sigma}) | \hat{\sigma})$, as should be the case, since we are interested in the consumer loss resulting from judging new items, based on the obtained estimated test limit $\hat{t}_2 = s - \hat{\sigma} a_2(\hat{\sigma})$.

After these preliminaries, we now are in position to start investigating the use of the bootstrap in removing the bias

$$b = ECL(\sigma, \hat{\sigma}) - \gamma \tag{2.4}$$

from $CL(\hat{t}_2) = CL(\sigma, \hat{\sigma})$. We shall proceed in several steps, each time presenting some simulation results to demonstrate that things work out well. For theoretical background and technical details which make clear that and why the bootstrap bias of $CL(\hat{t})$ is indeed of the right order, we refer to the Appendix.

The estimator $\hat{\sigma}^2$ from (2.2) can be written as $\hat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} W_i^2$, where $W_i = (X_{i2} - X_{i1})/\sqrt{2} = (U_{i2} - U_{i1})/\sqrt{2}$ has df $G(x) = \Phi(x/\sigma)$. Let G_n be the

empirical df corresponding to the sample $W_1, \ldots, W_n$. Then we have the basic bootstrap recipe (cf e.g. Efron and Tibshirani (1993)):

$$
\begin{array}{l}
* \text{ a sample } W_1, \ldots, W_n \text{ from } G \text{ is given} \\
* \text{ evaluate } \hat{\sigma} = (n^{-1} \sum_{i=1}^{n} W_i^2)^{\frac{1}{2}} \text{ and } CL(\sigma, \hat{\sigma}) \\
* \text{ repeat for } j = 1, \ldots, B \text{ the following steps:} \\
\quad * \text{ take a sample } W_{1j}, \ldots, W_{nj} \text{ from } G_n \\
\quad * \text{ evaluate } \sigma_j^* = (n^{-1} \sum_{i=1}^{n} W_{ij}^2)^{\frac{1}{2}} \text{ and } CL(\sigma, \sigma_j^*) \\
* \text{ evaluate } \hat{b}_1 = B^{-1} \sum_{j=1}^{B} CL(\sigma, \sigma_j^*) - CL(\sigma, \hat{\sigma})
\end{array}
\tag{2.5}
$$

Note that we could also have applied the parametric bootstrap, by using $\Phi(x/\hat{\sigma})$ rather than G_n for resampling. However, we shall restrict attention to the basic recipe here and leave such other possibilities unexplored.

The idea is that $\hat{b}_1$ estimates $b_1 = E\{CL(\sigma, \sigma_1^*)|G_n\} - CL(\sigma, \hat{\sigma})$, which should be close to $b = E\{CL(\sigma, \hat{\sigma}) - CL(\sigma, \sigma)\}$ from (2.4). (Here and in the sequel we neglect the difference between $CL(\sigma, \sigma)$ and γ). Hence

$$
CL_B(\hat{t}_2) = CL(\sigma, \hat{\sigma}) - \hat{b}_1 = 2CL(\sigma, \hat{\sigma}) - \frac{1}{B} \sum_{j=1}^{B} CL(\sigma, \sigma_j^*) \tag{2.6}
$$

should be approximately unbiased.

To investigate this empirically for given $\sigma,\ \pi = P(X > s), \gamma$ and n we apply ns simulations. In each simulation step (2.5) is performed for a given B, leading to a realization $CL_B(\hat{t}_2)$ in (2.6). The corresponding averages and standard deviations are collected in Table 2.1 for a variety of configurations (σ, π, γ), using $n = 80$, $ns = 1000$ and $B = 100$.

Table 2.1 Averages and standard deviations (between brackets) of $CL_B(\hat{t}_2)$ in ppm for various choices of $\sigma,\ \pi$ and γ, based each time on $n = 80$ observations, $ns = 1000$ simulations and $B = 100$ bootstrap replications per simulation step.

σ	$(\pi, \gamma) = (0.15,\ 20\text{ppm})$	$(\pi, \gamma) = (0.10,\ 40\text{ ppm})$	$(\pi, \gamma) = (0.01,\ 100\text{ ppm})$
0.01	19.6 (10.9)	39.8 (28.4)	100.2 (8.2)
0.10	19.2 (18.1)	39.1 (29.6)	99.7 (30.5)
0.20	17.9 (20.5)	38.8 (34.6)	103.2 (39.2)

From Table 2.1 it follows that $CL_B(\hat{t}_2)$ is indeed reasonably unbiased. Hence it makes sense to proceed with the next step: it follows from (1.8) and (2.4) that a_2 should be corrected here by $c = b_1/\{\sigma\phi(s)(1 - \Phi(a_1))\}$. Using the bootstrap from (2.5), b_1 can be estimated by $\hat{b}_1$, leading to

$$
\hat{c}^{(1)} = \hat{b}_1/\{\sigma\phi(s)(1 - \Phi(a_1))\}. \tag{2.7}
$$

Let $\hat{t}_B^{(1)} = \hat{t}_2 - \hat{\sigma}\hat{c}^{(1)} = s - \hat{\sigma}(\hat{a}_2 + \hat{c}^{(1)})$, then $CL(\hat{t}_B^{(1)})$ should be approximately unbiased. In Table 2.2 we collect for about the same (σ, π, γ) and (n, ns, B) as in Table 2.1 (only $B = 100$ becomes $B = 250$) averages and standard deviations of $CL(\hat{t}_B^{(1)})$, $\hat{b}_1$ and $CL(\hat{t}_2)$. It is clear that $CL(\hat{t}_B^{(1)})$ performs much better than the uncorrected $CL(\hat{t}_2)$ with respect to unbiasedness.

Table 2.2 Averages and standard deviations (between brackets) of $CL(\hat{t}_B^{(1)})$, $\hat{b}_1$ and $CL(\hat{t}_2)$ for various (σ, π, γ) and $(n, ns, B) = (80, 1000, 250)$.

σ	$CL(\hat{t}_B^{(1)})$		$\hat{b}_1$		$CL(\hat{t}_2)$	
			$(\pi, \gamma) = (0.15, 20\text{ppm})$			
0.01	19.9	(9.9)	2.5	(1.1)	22.5	(10.9)
0.10	18.1	(14.3)	7.6	(14.4)	27.0	(22.7)
0.20	17.6	(15.5)	9.5	(6.5)	29.6	(30.5)
			$(\pi, \gamma) = (0.10, 40\text{ppm})$			
0.01	40.0	(14.9)	2.8	(1.3)	42.8	(15.5)
0.10	39.5	(25.0)	9.8	(4.8)	50.3	(32.3)
0.20	37.1	(28.1)	12.5	(6.9)	51.1	(41.4)
			$(\pi, \gamma) = (0.01, 100\text{ppm})$			
0.01	99.9	(8.3)	0.3	(0.5)	100.2	(8.2)
0.10	100.4	(31.0)	4.4	(2.4)	104.6	(31.2)
0.20	100.6	(36.1)	6.6	(3.1)	106.7	(37.0)

Although the results up to this point are encouraging, there remains at least one additional step to be taken before the method can be applied in practice. For, in (2.5) and (2.7), and hence in both $CL_B(\hat{t}_2)$ and $CL(\hat{t}_B^{(1)})$, the true σ has been used. That is fine while performing simulations, but not for actual application, as σ is considered unknown. To remedy this, the obvious modification consists of evaluating $CL(\hat{\sigma}, \sigma_j^*)$ instead of $CL(\sigma, \sigma_j^*)$ in (2.5), leading to

$$\hat{b}_2 = B^{-1} \sum_{j=1}^{B} CL(\hat{\sigma}, \sigma_j^*) - \gamma \tag{2.8}$$

as a replacement for $\hat{b}_1$ from (2.5). Again the idea is that $\hat{b}_2$ from (2.8) estimates $b_2 = E(CL(\hat{\sigma}, \sigma_1^*)|G_n) - \gamma$, which should be close to b from (2.4). In analogy to (2.7), we use

$$\hat{c}^{(2)} = \hat{b}_2 / \{\hat{\sigma}\phi(s)(1 - \Phi(\hat{a}_1))\} \tag{2.9}$$

to arrive at $\hat{t}_B^{(2)} = \hat{t}_2 - \hat{\sigma}\hat{c}^{(2)} = s - \hat{\sigma}(\hat{a}_2 + \hat{c}^{(2)})$. In Table 2.3 we present some values of $CL(\hat{t}_B^{(2)})$ and $CL(\hat{t}_2)$, which show that the picture for $CL(\hat{t}_B^{(2)})$ is exactly the same as for $CL(\hat{t}_B^{(1)})$.

Table 2.3 Averages and standard deviations (between brackets) of $CL(\hat{t}_B^{(2)})$ and $CL(\hat{t}_2)$ for various (σ, π, γ) and $(n, ns, B) = (80, 1000, 100)$.

σ	$CL(\hat{t}_B^{(2)})$		$CL(\hat{t}_2)$	
	$(\pi, \gamma) = (0.15, 20\text{ppm})$			
0.01	20.3	(11.0)	22.8	(11.8)
0.10	19.8	(17.4)	26.6	(21.6)
0.20	20.2	(19.9)	29.2	(26.1)
	$(\pi, \gamma) = (0.10, 40\text{ppm})$			
0.01	40.7	(16.0)	43.5	(16.5)
0.10	41.5	(30.4)	50.6	(34.6)
0.20	38.3	(31.7)	49.6	(37.2)
	$(\pi, \gamma) = (0.01, 100\text{ppm})$			
0.01	100.1	(8.4)	100.4	(8.3)
0.10	99.3	(29.3)	103.8	(29.8)
0.20	101.3	(37.2)	107.6	(38.4)

Hence in the present problem the bootstrap is indeed suitable for practical use: the average relative errors in $CL(\hat{t}_2)$ from Table 2.3 are typically at least five times as large as those in $CL(\hat{t}_B^{(2)})$. Incidentally, in Table 4 from Albers, Kallenberg and Otten (1994b), a similar comparison is made between $CL(\hat{t}_2)$ and the CL which is corrected using the explicitly evaluated expansion of the bias. Not surprisingly, this latter version is even better than $CL(\hat{t}_B^{(2)})$. Nevertheless, the improvement using the rather unsophisticated approach based on the bootstrap, is still quite satisfactory.

At the end of this section, we draw attention to the following simplification. For both corrected versions $\hat{t}_B^{(j)}$, $j = 1, 2$, we take $\hat{t}_2$ as our starting point, as it is more precise than the first order approximation $\hat{t}_1 = s - \hat{a}_1\hat{\sigma}$. However, in determining the correction itself, which is of a lower order, such precision is not needed. Hence for this part, we can replace $CL(\sigma_1, \sigma_2)$ from (2.3) by $\widetilde{CL}(\sigma_1, \sigma_2) = P(X > s, X - \sigma_1 Y < s - \sigma_2 a_1(\sigma_2))$, which is slightly easier to evaluate. Then $\hat{b}_2$ from (2.8) in its turn is replaced by $\hat{b}_3 = B^{-1}\sum_{j=1}^{B}(\widetilde{CL}(\hat{\sigma}, \sigma_j^*)) - \widetilde{CL}(\hat{\sigma}, \hat{\sigma})$. Simulations confirm that the resulting $CL(\hat{t}_B^{(3)})$ works just as well as $CL(\hat{t}_B^{(2)})$.

As announced earlier, some theoretical background is offered in the Appendix. There it is demonstrated why $E\{CL(\sigma, \sigma_1^*)|G_n\} - CL(\sigma, \hat{\sigma})$ and $E(CL(\hat{\sigma}, \sigma_1^*)|G_n) - \gamma$ are indeed both close to $ECL(\sigma, \hat{\sigma}) - \gamma$.

3 Bias correction when all parameters are unknown

Again consider production data $\tilde{X}_i = X_i + U_i,\ i = 1, \ldots, m$. A subsample of size n from these data has been measured twice, leading to $X_{ij} = X_i + U_{ij},\ i = 1, \ldots, n,\ j = 1, 2$ (cf. (2.1)), with $X_{i1} = \tilde{X}_i$ (and hence $U_{i1} = U_i$) for $i = 1, \ldots, n$. Unlike the situation in the previous section, we now do not assume m to be so much larger than n that estimation effects in $\hat{\mu}_X$ and $\hat{\sigma}_X^2$ can be ignored. In addition to $\hat{\sigma}_U^2$ from (2.2) we use, denoting averaging over an index by a "$\bullet$",

$$\hat{\mu}_X = \tilde{X}_\bullet,\ \hat{\sigma}_X^2 = \frac{1}{(m-1)} \sum_{i=1}^{m} (\tilde{X}_i - \tilde{X}_\bullet)^2 - \hat{\sigma}_U^2. \tag{3.1}$$

From (1.2) it follows that for $\hat{t}_2 = s - \hat{a}_2 \hat{\sigma}_U$,

$$CL(\hat{t}_2) = P(\bar{X} > \bar{s},\ \bar{X} - \sigma Y < \bar{s} - \hat{\sigma}_U \hat{a}_2 / \sigma_X), \tag{3.2}$$

with $\sigma,\ Y,\ \bar{X}$ and $\bar{s}$ as in (1.3). In analogy to (2.3), we shall denote the expression from (3.2) by $CL(\mu_X, \sigma_X, \sigma_U, \hat{\mu}_X, \hat{\sigma}_X, \hat{\sigma}_U)$. Likewise, $CL(t_2)$ will be $CL(\mu_X, \sigma_X, \sigma_U, \mu_X, \sigma_X, \sigma_U)$, which again is indistinguishable from γ. To correct the bias

$$b = ECL(\mu_X, \sigma_X, \sigma_U, \hat{\mu}_X, \hat{\sigma}_X, \hat{\sigma}_U) - \gamma \tag{3.3}$$

(cf. (2.4)), a similar approach as in the previous section leads to

$$\hat{b}_2 = B^{-1} \sum_{j=1}^{B} CL(\hat{\mu}_X, \hat{\sigma}_X, \hat{\sigma}_U, (\mu_X^*)_j,\ (\sigma_X^*)_j,\ (\sigma_U^*)_j) - \gamma \tag{3.4}$$

(cf. (2.8)), in which the parameters have been replaced by the estimators, and the estimators in their turn by bootstrap estimators. Thus we obtain the test limit $\hat{t}_B^{(2)} = \hat{t}_2 - \hat{\sigma}_U \hat{c}^{(2)} = s - \hat{\sigma}_U (\hat{a}_2 + \hat{c}^{(2)})$, where

$$\hat{c}^{(2)} = \hat{b}_2 \Big/ \left\{ \frac{\hat{\sigma}_U}{\hat{\sigma}_X} \phi \left(\frac{s - \hat{\mu}_X}{\hat{\sigma}_X} \right) (1 - \Phi(\hat{a}_1)) \right\}. \tag{3.5}$$

To obtain the bootstrap estimators required, the recipe determined by (2.5) and (2.8) is extended as follows. In addition to the empirical df G_n of $G(x) =$

$\Phi(x/\sigma_U)$, let H_m be the empirical df based on $\tilde{X}_1, \ldots, \tilde{X}_m$, corresponding to $H(x) = \Phi((x-\mu_X)/\{\sigma_X^2 + \sigma_U^2\}^{\frac{1}{2}})$. Then:

$$
\begin{array}{l}
* \text{ samples } W_1, \ldots, W_n \text{ from } G \text{ and } \tilde{X}_1, \ldots, \tilde{X}_m \text{ from } H \text{ are given} \\
* \text{ evaluate } \hat{\sigma}_U = (n^{-1}\sum_{i=1}^n W_i^2)^{\frac{1}{2}}, \hat{\mu}_X = \tilde{X}_\bullet \text{ and} \\
\quad \hat{\sigma}_X = \{(m-1)^{-1}\sum_{i=1}^m (\tilde{X}_i - \tilde{X}_\bullet)^2 - \hat{\sigma}_U^2\}^{\frac{1}{2}} \\
* \text{ repeat for } j = 1, \ldots, B \text{ the following steps:} \\
\quad * \text{ take samples } W_{1j}, \ldots, W_{nj} \text{ from } G_n \text{ and } \tilde{X}_{1j}, \ldots, \tilde{X}_{mj} \\
\quad\quad \text{from } H_m \\
\quad * \text{ evaluate } (\sigma_U^*)_j = (n^{-1}\sum W_{ij}^2)^{\frac{1}{2}}, (\mu_X^*)_j = \tilde{X}_{\bullet j}, \\
\quad\quad (\sigma_X^*)_j = \{(m-2)^{-1}\sum_{i=1}^m (\tilde{X}_{ij} - \tilde{X}_{\bullet j})^2 - (\sigma_U^*)_j^2\}^{\frac{1}{2}} \text{ and} \\
\quad\quad CL(\hat{\mu}_X, \hat{\sigma}_X, \hat{\sigma}_U, (\mu_X^*)_j, (\sigma_X^*)_j, (\sigma_U^*)_j) \\
* \text{ evaluate } \hat{b}_2 = B^{-1}\sum_{j=1}^B CL(\hat{\mu}_X, \hat{\sigma}_X, \hat{\sigma}_U, (\hat{\mu}_X^*)_j, (\sigma_X^*)_j, (\sigma_U^*)_j) - \gamma
\end{array}
\tag{3.6}
$$

For some remarks on the theoretical background, we once more refer to the Appendix. Here we just highlight the use of a factor $(m-2)^{-1}$, rather than $(m-1)^{-1}$, in $(\sigma_X^*)_j$. Note that the conditional expectation of the sample variance $(m-1)^{-1}\sum_{i=1}^m (\tilde{X}_{ij} - \tilde{X}_{\bullet j})^2$ equals the variance of the underlying df, which here boils down to $\int x^2 dH_m - (\int x dH_m)^2 = m^{-1}\sum_{i=1}^m \tilde{X}_i^2 - \tilde{X}_\bullet^2 = ((m-1)/m)\hat{\sigma}_{\tilde{X}}^2$. Hence $E((m-2)^{-1}\sum_{i=1}^m (\tilde{X}_{ij} - \tilde{X}_{\bullet j})^2 | H_m) = \{(m-1)^2/(m(m-2))\}\hat{\sigma}_{\tilde{X}}^2 = \{1 + \mathcal{O}(m^{-2})\}\hat{\sigma}_{\tilde{X}}^2$. Thus the use of $(m-2)^{-1}$ in $(\sigma_X^*)_j$ ensures unbiasedness to order m^{-2} instead of m^{-1}. The latter order does not suffice, as the bias correction we are looking for, precisely is meant to compensate for terms of this order.

To close this section, we present Table 3.1, in which $CL(\hat{t}_B^{(2)})$ is compared to $CL(\hat{t}_2)$. The conclusions are completely similar to those from Table 2.3: $CL(\hat{t}_B^{(2)})$ performs quite well, which is not the case for $CL(\hat{t}_2)$.

Table 3.1 Averages and standard deviations (between brackets) of $CL(\hat{t}_B^{(2)})$ and $CL(\hat{t}_2)$ for various (σ, π, γ) and $(n, m, ns, B) = (80, 240, 1000, 100)$.

σ	$CL(\hat{t}_B^{(2)})$		$CL(\hat{t}_2)$	
	$(\pi, \gamma) = (0.15, 20\text{ppm})$			
0.01	20.3	(10.8)	22.9	(11.5)
0.10	20.6	(19.1)	27.6	(23.6)
0.20	20.0	(20.2)	29.0	(26.9)
	$(\pi, \gamma) = (0.10, 40\text{ppm})$			
0.01	40.3	(15.8)	43.5	(16.5)
0.10	41.7	(30.0)	51.5	(34.5)
0.20	40.8	(33.4)	53.2	(40.4)
	$(\pi, \gamma) = (0.01, 100\text{ppm})$			
0.01	100.2	(28.6)	106.9	(31.5)
0.10	100.0	(41.7)	110.7	(44.8)
0.20	102.3	(48.7)	115.6	(53.3)

Again, we can also look at the situation where $CL(\hat{t}_2)$ is not compared to $CL(\hat{t}_B^{(2)})$, but to the CL which relies on the explicitly evaluated expansion of the bias for the three-parameter case. For this purpose, consult Table 3.4.1 from Otten (1995) and the remarks at the end of section 3 in Albers, Kallenberg and Otten (1994a). Once more the picture is similar to that from the previous section: although the bootstrap performs quite well, it is still outperformed by the version based on explicit evaluation of the correction terms, as expected.

4 Controlling the exceedance probability

In the previous two sections we have corrected $\hat{t}_2$ for the bias b by introducing $\hat{t}_B^{(2)}$. In view of the symmetry of the asymptotic normal distribution, this typically means that $P(CL(\hat{t}_B^{(2)}) > \gamma)$ will be close to $\frac{1}{2}$. Hence, γ will be exceeded quite often. Moreover, as the tables from sections 2 and 3 reveal, the standard deviations involved are rather large, and thus the exceedances will be large as well. To remedy this, we can impose the more stringent condition

$$P(CL(\hat{t}) > \gamma) \leq \alpha, \tag{4.1}$$

for some small prescribed α, e.g. 5 or 10%.

In addition to the bias b, let τ^2 be the variance of $CL(\hat{t}_2)$. Then for $\hat{t} = s - \hat{\sigma}_U(\hat{a}_2 + \hat{c})$ we have that $CL(\hat{t})$ is, in view of (1.7),

$$AN(\gamma + b + c\sigma\phi(\bar{s})g_1'(a_1), \tau^2), \tag{4.2}$$

where "AN" stands for asymptotically normal. It follows that equality is achieved to first order in (4.1) if $1 - \Phi(\{-b - c\sigma\phi(\bar{s})g_1'(a_1)\}/\tau) = \alpha$, i.e.

$$c = (b + \tau u_\alpha)/\{\sigma\phi(\bar{s})(1 - \Phi(a_1))\}, \tag{4.3}$$

where $u_\alpha = \Phi^{-1}(1-\alpha)$ is the upper α-quantile of the standard normal df. Note that (4.3) indeed reduces to (1.8) for $\alpha = \frac{1}{2}$.

From here, it is rather evident how to proceed: use $CL_j^* = CL(\hat{\mu}_X, \hat{\sigma}_X, \hat{\sigma}_U, (\mu_X^*)_j, (\sigma_X^*)_j, (\sigma_U^*)_j)$ in (3.6) not merely to obtain the bias estimator $\hat{b}_2$, but also for the sample variance $\hat{\tau}^2 = (B-1)^{-1}\sum_{j=1}^B (CL_j^* - CL_\bullet^*)^2$. Then, in analogy to (3.5), apply $\hat{t}_E = \hat{t}_2 - \hat{\sigma}_U\hat{c} = s - \hat{\sigma}_U(\hat{a}_2 + \hat{c})$, where

$$\hat{c} = (\hat{b}_2 + \hat{\tau}u_\alpha)/\left\{\frac{\hat{\sigma}_U}{\hat{\sigma}_X}\,\phi\left(\frac{s-\hat{\mu}_X}{\hat{\sigma}_X}\right)(1 - \Phi(\hat{a}_1))\right\}. \tag{4.4}$$

Typically, $\hat{\tau}$ is considerably larger than $\hat{b}_2$, which can be explained asymptotically by noting that both $\hat{\tau}^2$ and $\hat{b}_2$ consist of terms which are of order either m^{-1} or n^{-1}. Especially for larger m and n, the contribution of $\hat{b}_2$ to the correction becomes negligible and it can be omitted from (4.4).

Since we are correcting here for terms of order $n^{-\frac{1}{2}}$ (and $m^{-\frac{1}{2}}$) rather than of order n^{-1} (and m^{-1}), larger sample sizes n (and m) are required in this situation. E.g. in Albers, Kallenberg and Otten (1994b), n is increased from 40 or 80 in the bias case to 500 or 2000 in the present case. In view of the fact that now each simulation step entails a number of bootstrap replications, we shall be a bit more moderate and increase n from 80 in sections 2 and 3 to $n = 250$ here. Likewise, m will be increased from 240 to 500.

In Table 4.1 we compare the realized exceedance probability $\hat{\alpha}$ to a prescribed value α of 10% (cf. (4.1)). We cover both the case where only σ_U^2 is considered unknown (cf. section 2) and the one where all parameters are unknown (cf. section 3). In the latter case, m enters the picture. Moreover, we also consider both the situation with and without bias correction $\hat{b}_2$.

Table 4.1 Realized exceedance probabilities (in percentages) $\hat{\alpha}$ for a prescribed value $\alpha = 10\%$, for various (σ, π, γ) and $(n, ns, B) = (250, 1000, 100)$. Results are given with and without bias correction $\hat{b}_2$ and with and without (μ_X, σ_X) known. For the case of unknown (μ_X, σ_X), the choice $m = 500$ is used.

	(μ_X, σ_X) known		(μ_X, σ_X) unknown	
σ	without $\hat{b}_2$	with $\hat{b}_2$	without $\hat{b}_2$	with $\hat{b}_2$
		$(\pi, \gamma) = (0.15, 20\text{ppm})$		
0.01	11.7	9.3	10.4	7.4
0.10	12.4	9.1	10.7	8.3
0.20	9.3	6.4	9.2	7.0
		$(\pi, \gamma) = (0.10, 40\text{ppm})$		
0.01	9.7	8.0	10.8	8.9
0.10	11.5	8.9	9.7	6.8
0.20	10.0	8.1	8.4	5.9
		$(\pi, \gamma) = (0.01, 100\text{ppm})$		
0.01	10.9	10.1	6.6	4.5
0.10	13.3	12.1	11.9	8.6
0.20	13.2	11.4	14.8	10.8

From Table 4.1 we conclude that the agreement between $\hat{\alpha}$ and α is quite reasonable, especially when we take into account that the value of $n^{\frac{1}{2}}$ considered, is still rather small. It is also clear that the bias correction $\hat{b}_2$ indeed loses its importance: the values of $\hat{\alpha}$ obtained without it, are not really inferior to those obtained when including it.

Appendix

$*$ Comparison of b from (2.4), $\hat{b}_1$ from (2.5) and $\hat{b}_2$ from (2.8).

In b, $E(\hat{b}_1|G_n)$ and $E(\hat{b}_2|G_n)$ we encounter $ECL(\sigma, \hat{\sigma})$, $E(CL(\sigma, \sigma_1^*)|G_n)$ and $E(CL(\hat{\sigma}, \sigma_1^*)|G_n)$. In order to compare these expectations, we begin by noting, in analogy to (1.4), that $CL(\sigma_1, \sigma_2)$ from (2.3) equals, for $\tilde{a} = \sigma_2 a_2(\sigma_2)/\sigma_1$,

$$\int_{\tilde{a}}^{\infty} \{\Phi(s + \sigma_1(y - \tilde{a})) - \Phi(s)\}\phi(y)dy. \tag{A.1}$$

Together with (1.5) and (1.6), it follows from (A.1) that

$$CL(\sigma_1, \sigma_2) = \sigma_1 \phi(s) g_1 \left(\frac{\sigma_2}{\sigma_1} g_1^{-1} \left(\left(\frac{\gamma}{\sigma_1 \phi(s)} \right) \Big/ \left(\frac{\sigma_2}{\sigma_1} \right) \right) \right) + \dots , \tag{A.2}$$

which in turn implies that $CL(\sigma_1, \sigma_2) - \gamma = \sum_{j=1}^{2} A_j (\frac{\sigma_2}{\sigma_1} - 1)^j + \ldots$, for certain coefficients A_j. Let

$$S = S(\sigma_1, \sigma_2) = \sigma_2^2/\sigma_1^2 - 1, \tag{A.3}$$

then $\sigma_2/\sigma_1 - 1 = (1+S)^{\frac{1}{2}} - 1 = \frac{1}{2}S - \frac{1}{8}S^2 + \ldots$, and thus $(\sigma_2/\sigma_1 - 1)^2 = \frac{1}{4}S^2 + \ldots$. Consequently,

$$CL(\sigma_1, \sigma_2) - \gamma = \tfrac{1}{2}A_1 S(\sigma_1, \sigma_2) + (\tfrac{1}{4}A_2 - \tfrac{1}{8}A_1)S^2(\sigma_1, \sigma_2) + \ldots . \tag{A.4}$$

Hence the required comparison of the various expected consumer losses can be carried out by obtaining the first and second moments of the corresponding S from (A.3). For $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^{n} W_i^2$ from (2.2), we know that $n\hat{\sigma}^2/\sigma^2$ is $\chi^2(n)$-distributed, and hence $ES(\sigma, \hat{\sigma}) = 0$, $ES^2(\sigma, \hat{\sigma}) = 2n^{-1}$. For $(\sigma_1^*)^2 = n^{-1}\sum_{i=1}^{n} W_{i1}^2$ from (2.5) it follows that

$$E((\sigma_1^*)^2 | G_n) = \int x^2 dG_n = n^{-1}\sum_{i=1}^{n} W_i^2 = \hat{\sigma}^2. \tag{A.5}$$

Likewise, $\mathrm{var}((\sigma_1^*)^2 | G_n) = n^{-1}E((W_{11}^2 - \hat{\sigma}^2)^2 | G_n) = n^{-1}(n^{-1}\sum_{i=1}^{n} W_i^4 - \hat{\sigma}^4) = 2n^{-1}\hat{\sigma}^4(1 + o_p(1))$. Consequently, $E(S(\hat{\sigma}, \sigma_1^*) | G_n) = 0$, $E(S^2(\hat{\sigma}, \sigma_1^*) | G_n) = 2n^{-1} + \ldots$. Finally, (A.5) implies that $E(S(\sigma, \sigma_1^*) | G_n) - S(\sigma, \hat{\sigma}) = 0$, whereas $E(S^2(\sigma, \sigma_1^*) | G_n) - S^2(\sigma, \hat{\sigma}) = \sigma^{-4}\mathrm{var}((\sigma_1^*)^2 | G_n) = 2n^{-1} + \ldots$. Together with (A.4), these moment results imply that, to sufficient precision

$$E(CL(\sigma, \sigma_1^*) | G_n) - CL(\sigma, \hat{\sigma}) = E(CL(\hat{\sigma}, \sigma_1^*) | G_n) - \gamma = ECL(\sigma, \hat{\sigma}) - \gamma. \tag{A.6}$$

Then (2.5) and (2.8) imply that $b_1 = E(\hat{b}_1 | G_n)$ and $b_2 = E(\hat{b}_2 | G_n)$ to first order equal $b = ECL(\sigma, \hat{\sigma}) - \gamma$ from (2.4). □

∗ Comparison of b from (3.3) and $\hat{b}_2$ from (3.4).

Basically, the argument here is just an extension of the one given above for the one-parameter case. In analogy to (A.2), we can derive

$$CL(\mu_X, \sigma_X, \sigma_U, \hat{\mu}_X, \hat{\sigma}_X, \hat{\sigma}_U) = \sigma\phi(\bar{s}) g_1 \left(X g_1^{-1} \left(\frac{\gamma}{\sigma\phi(\bar{s})} \cdot \frac{Y}{X} \right) \right), \tag{A.7}$$

where $X = \hat{\sigma}_U/\sigma_U$ and $Y = \hat{\sigma}_X\phi(\bar{s})/\{\sigma_X\phi((s - \hat{\mu}_X)/\hat{\sigma}_X)\}$. After straightforward but tedious computations, (A.7) in analogy to (A.4) leads to an expansion involving not only S from (A.3) in the form $S(\sigma_U, \hat{\sigma}_U)$, but also $S(\sigma_X, \hat{\sigma}_X)$ and $V(\mu_X, \hat{\mu}_X, \sigma_X) = (\mu_X - \hat{\mu}_X)/\sigma_X$. Evaluation of the bias b from (3.3) now entails evaluation of the first and second (mixed) moments of $S(\sigma_U, \hat{\sigma}_U)$, $S(\sigma_X, \hat{\sigma}_X)$ and $V(\mu_X, \hat{\mu}_X, \sigma_X)$.

Likewise, to obtain $b_2 = E(\hat{b}_2|G_n, H_m)$ for b_2 from (3.4), we need the corresponding moments of $S(\hat{\sigma}_U, (\sigma^*_U)_1)$, $S(\hat{\sigma}_X, (\sigma^*_X)_1)$ and $V(\hat{\mu}_X, (\mu^*_X)_1, \hat{\sigma}_X)$, conditional on G_n and H_m. Again, it is straightforward but laborious to demonstrate that for each pair of moments we have agreement to order n^{-1} and m^{-1}. The only noteworthy point is that the regular definition $(\sigma^*_X)^2_j = (m-1)^{-1}\sum_{i=1}^m (\tilde{X}_{ij} - \tilde{X}_{\bullet j})^2 - (\hat{\sigma}^*_U)^2_j$ would produce $E(S(\hat{\sigma}_X, (\sigma^*_X)_1)|G_n, H_m) = -m^{-1}$, which does not agree with $E(S(\sigma_X, \hat{\sigma}_X)) = 0$ to order m^{-1}. Replacement of $(m-1)^{-1}$ by $m/(m-1)^2$ in this definition ensures that $E(S(\hat{\sigma}_X, (\sigma^*_X)_1)|G_n, H_m) = 0$ as well, while the use of $(m-2)^{-1}$ leads to $E(S(\hat{\sigma}_X, (\sigma^*_X)_1)|G_n, H_m) = \{m(m-2)\}^{-1}$, which is sufficiently small. In this way, $b_2 = E(\hat{b}_2|G_n, H_m)$ and b agree to sufficient precision to justify the use of $\hat{b}_2$.

The last step above incidentally may leave a somewhat awkward feeling: should not the bootstrap work "automatically", without requiring adaptation of estimators, however minor? The answer is that this is indeed the case and that we have brought the need for adaptation upon ourselves by using the sample variance $\hat{\sigma}^2_{\tilde{X}}$ rather than $\tilde{\sigma}^2_{\tilde{X}} = m^{-1}\sum_{i=1}^m (\tilde{X}_i - \tilde{X}_{\bullet})^2 = (m-1)\hat{\sigma}^2_{\tilde{X}}/m$. The latter choice is a substitution estimator, meaning that if $\sigma^2_{\tilde{X}} = \theta(H) = \int x^2 dH - (\int x dH)^2$, then $\tilde{\sigma}^2_{\tilde{X}} = \theta(H_m)$. Using in the bootstrap $\theta(H^*_{mj}) = (\tilde{\sigma}^*_{\tilde{X}})^2_j$ leads to pairs $(\sigma^2_{\tilde{X}}, \tilde{\sigma}^2_{\tilde{X}})$ and $(\tilde{\sigma}^2_{\tilde{X}}, (\tilde{\sigma}^*_{\tilde{X}})^2_1)$ which again show the desired agreement in their behavior to the order required. Nevertheless we prefer to work with the more common unbiased choice $\hat{\sigma}^2_{\tilde{X}}$, accepting the minor accompanying adaptation of the procedure.

References

Albers, W., Arts, G.R.J. and Kallenberg, W.C.M. (1995). "Test limits for correlated measurements under prescribed consumer risk" Technical Report 1282, University Twente.

Albers, W., Kallenberg, W.C.M. and Otten, G.D. (1994a). "Setting test limits under prescribed consumer loss". *Metrika* **41**, 163-181.

Albers, W., Kallenberg, W.C.M. and Otten, G.D. (1994b). "Accurate test limits with estimated parameters". *Technometrics* **36**, 92-101.

Efron, B. and Tibshirani, R.J. (1993). *"An introduction to the bootstrap".* Chapman and Hall, New York.

Otten, G.D. (1995). *"Statistical test limits in quality control".* Ph.D.Thesis, University Twente.