Box-Cox transformations versus data modelling — robust charts in Statistical Process Control*

Fernanda Figueiredo
CEAUL and Faculdade de Economia da Universidade do Porto

and

M. Ivette Gomes
CEAUL and DEIO (FCUL), Universidade de Lisboa

Abstract. In Statistical Process Control we often assume that the observations come from a normal population. Then, to monitor the process, we implement the traditional control charts for normal data. Despite the advantages of the use of the normal distribution, the performance of these charts may be adversely affected if we have disturbances in the data, such as contamination, outliers, or even small deviations from the normal. It is thus important to advance with "robust control charts" to monitor the process parameters, giving the expected false alarm rate whenever these parameters are close to the target values, although with a data distribution no longer normal. In this paper we propose either an *adequate modelling* of our data or alternatively the use of a *Box-Cox transformation* to approximately normalize the collected data. Any of these approaches should then be followed by the use of robust statistics, the *total median* and the *total range* to control the mean and the deviation, respectively, of the process under analysis.

AMS 2000 subject classification. 62P30, 93C83, 93B17, 62G35, 62F40.
Keywords and phrases. *Quality control charts; "Robust" control charts; Data modelling; Box-Cox transformations; Statistical Process Control.*

1 Introduction and control statistics

Control charts are one of the basic tools in Statistical Process Control (SPC), widely used to monitor industrial processes on line production. Most of those control charts assume an underlying normal process, and such an hypothesis is often inappropriate in practice, due to contamination, outliers, or even small deviations from the normal. Those charts exhibit then a rate of false alarms much higher or lower than expected. It is thus important to advance with

*Research partially supported by FCT / POCTI / FEDER.

"robust control charts" to monitor the process parameters, so that we do not have either a very high or a very low false alarm rate whenever the parameters to be controlled are close to the target values, although the data is no longer normal. Indeed, before implementing the traditional control charts for normal data we must obviously confirm such an assumption. If the data leads us to reject the normality, the traditional charts are inappropriate and the process monitoring must be carried out using other kind of control charts or a different approach to the process control. Some authors have already addressed the problem of monitoring non-normal data. Among them we mention Spedding and Rawlings (1994), Wu (1996), Borror et al. (1999), Stoumbos and Reynolds (2000), Ryan and Faddy (2001) and Shore (2001). Here, we suggest either a preliminar careful modelling of the process through the available data and the construction of control charts for the specific model or the use of an adequate Box-Cox transformation of the process data (Box and Cox, 1964). Any of these approaches should then be followed by the use of a robust statistic to monitor the mean value of the quality characteristic under study. The standard deviation should also be separately and adequately controlled through the use of a "robust control chart".

Given an observed sample of size n, $(x_1, x_2, \cdots, x_n)$ from a model F, let us denote $(x_{i:n},\ 1 \le i \le n)$ the observed sample of associated ascending order statistics (o.s.), and $(x_i^*,\ 1 \le i \le n)$ the associated bootstrap sample. As usual, the notation $X_{i:n}$ and X_i^* is used for the corresponding random variables (r.v.'s).

Given the original observations from the process to be controlled, which we assume to be positive, as usually happens with a great diversity of industrial processes' measures, in the areas of engineering, insurance and reliability, for instance (see Hawkins and Olwell, 1998), the Box-Cox transformed sample is given by

$$z_i = \begin{cases} (x_i^\lambda - 1)/\lambda & \text{if} \quad \lambda \neq 0 \\ \ln x_i & \text{if} \quad \lambda = 0 \end{cases}, \quad 1 \le i \le n. \tag{1.1}$$

To monitor the mean value of the process we propose here, as an alternative to the classical $\overline{X}$-chart (or Mean-chart), the use of the total median (Figueiredo and Gomes, 2001; 2003).

$$TMd := \sum_{i=1}^{n} \sum_{j=i}^{n} \alpha_{ij} \ \frac{X_{i:n} + X_{j:n}}{2} =: \sum_{i=1}^{n} a_i X_{i:n},$$
$$\alpha_{ij} = P\left(BMd = \frac{x_{i:n} + x_{j:n}}{2}\right), \tag{1.2}$$

where BMd denotes the bootstrap median. The total median is thus a linear combination of o.s., in which the weights decrease as we approach the tails of the underlying model, enabling thus an interesting resistance to changes in the underlying model, particularly when we compare this statistic with the classical one, the sample mean $\overline{X}$.

To monitor the standard deviation, and as an alternative to the classical R and S-charts, we propose the use of the total range, TR, given by

$$TR := \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \beta_{ij} \left(X_{j:n} - X_{i:n}\right) =: \sum_{i=1}^{n} b_i X_{i:n},$$
$$\beta_{ij} = P\left(BR = x_{j:n} - x_{i:n} | BR \neq 0\right), \; 1 \leq i < j \leq n, \quad (1.3)$$

where BR denotes the bootstrap range. Again, the total range may be written as a linear combination of the sample o.s., i.e., it is a systematic statistic (Figueiredo, 2002, 2003). For the standard deviation control, we have also used, apart from the total range TR, the Median of the Absolute Deviations (MAD) to the median, i.e.,

$$MAD := \underset{1 \leq i \leq n}{Md} \left|X_i - Md\right|, \quad (1.4)$$

where Md denotes the sample median, together with modified versions of the sample standard deviation S, given by

$$S^* := \sqrt{\sum_{i=1}^{n} a_i \left(X_{i:n} - \overline{X}\right)^2} \quad (1.5)$$

and

$$S^{**} := \sqrt{\sum_{i=1}^{n} a_i \left(X_{i:n} - TMd\right)^2}, \quad (1.6)$$

where a_i are the weights in (1.2).

In section 2 of this paper we shall review the importance of the use of these statistics in SPC, to monitor either the process mean or the standard deviation. These statistics are of easy manipulation and computation, in opposition to the usual M and R robust estimators (Lax, 1985; Tatum, 1997), and such a fact leads us to advance with the use, in practice, of control charts associated to them. In section 3, some of these control charts are applied both to original data from known parents and to the same data, transformed through a convenient normalized Box-Cox transformation. We have herewith considered a set of models covering a large region of the (β, τ)-plane. The skewness coefficient β is given by

$$\beta := \frac{\mu_3}{\mu_2^{3/2}}, \quad (1.7)$$

with μ_r denoting the r-th central moment of F. The tail-weight coefficient τ is given by

$$\tau := \frac{1}{2} \frac{\frac{F^{\leftarrow}(0.99)-F^{\leftarrow}(0.5)}{F^{\leftarrow}(0.75)-F^{\leftarrow}(0.5)} + \frac{F^{\leftarrow}(0.5)-F^{\leftarrow}(0.01)}{F^{\leftarrow}(0.5)-F^{\leftarrow}(0.25)}}{\frac{\Phi^{\leftarrow}(0.99)-\Phi^{\leftarrow}(0.5)}{\Phi^{\leftarrow}(0.75)-\Phi^{\leftarrow}(0.5)}}, \tag{1.8}$$

where $F^{\leftarrow}$ and $\Phi^{\leftarrow}$ denote the inverse functions of F and of the standard normal distribution function Φ, respectively.

Note that we consider that, whenever we are controlling the process at the target mean value μ_0 (or standard deviation σ_0), a control chart based on a statistic W is said "robust" if the alarm rate is as close as possible to the pre-assigned α-risk, whenever the model changes but the process parameter is kept at the respective target. The results of the study undertaken in section 3 lead us to advocate the use of Box-Cox transformations, prior to the implementation of the TMd-chart. Indeed, while the TMd-chart is "quasi-robust" when applied to specific parents, which are obviously unknown and need to be adequately identified in practice, it is "totally robust" when we first apply an *a priori* Box-Cox transformation to the observations, in order to change them into normal ones. Relatively to the standard deviation charts, the application of an *a priori* Box-Cox transformation to the data does not work as well as before and does not lead us to results as good as the ones we have obtained with the charts specifically devised for the original parent. Only the MAD chart may be considered "robust", but it exhibits a low efficiency in the detection of an out-of-control situation; the TR chart is non "robust" either when we use an *a priori* Box-Cox transformation of the data or the original data; the S^* and S^{**} charts are less sensitive to departures from the normality assumption than the TR chart, but they cannot also at all be considered "robust". Anyway, they are more resistant to changes in the underlying model than the classical S and R charts. In section 4 we present a design for the implementation of this new approach to the monitoring of the mean value of non-normal processes on the basis of Box-Cox transformations, in order to motivate its future use. We also draw some overall conclusions.

2 The use of robust statistics in SPC

In Table 1, and for rational subgroups of size $n = 5, 10, 15$ and 20, we present the values of the weights a_i and b_i needed for the computation of the observed value of the total median, of the total range, of S^* and of S^{**} in (1.2), (1.3), (1.5) and (1.6), respectively. The blank values are either zero or given by the relations

$$\sum_{i=1}^{n} a_i = 1, \quad \sum_{i=1}^{n} b_i = 0, \quad a_i = a_{n-i+1} \text{ and } b_i = -b_{n-i+1}, \quad 1 \le i \le n.$$

The coefficients a_i and b_i are *distribution-free*, i.e., they are independent of the underlying model F and depend only on the sample size n.

	$n=5$		$n=10$		$n=15$		$n=20$	
i	a_i	b_i	a_i	b_i	a_i	b_i	a_i	b_i
1	0.058	-0.737	0.001	-0.652	0.000	-0.645	0.000	-0.643
2	0.259	-0.263	0.019	-0.241	0.000	-0.239	0.000	-0.237
3	0.366		0.078	-0.079	0.004	-0.082	0.000	-0.083
4			0.168	-0.022	0.021	-0.026	0.001	-0.027
5			0.234	-0.004	0.063	-0.007	0.007	-0.008
6					0.125	-0.001	0.023	-0.002
7					0.183		0.055	
8					0.208		0.099	
9							0.143	
10							0.172	

Table 1: Values of the coefficients a_i and b_i for sample sizes n=5, 10, 15, 20.

To control the standard deviation, the above mentioned statistics were standardized, through the division by a scale c_n, so that the control statistics have a unit mean value whenever the underlying model is normal. The scales c_n are provided in Table 2, for the same rational subgroups $n = 5, 10, 15$ and 20.

n	S	R	TR	S^*	S^{**}	MAD
5	0.940	2.326	1.975	0.830	0.585	0.555
10	0.973	3.078	2.611	0.907	0.434	0.616
15	0.982	3.472	3.036	0.931	0.336	0.639
20	0.987	3.734	3.320	0.945	0.295	0.647

Table 2: Scale constants, c_n, for sample sizes $n = 5, 10, 15, 20$.

2.1 Control of a mean

The use of the total median in SPC was discussed extensively in Figueiredo and Gomes (2001, 2003). In order to get information on the resistance of the $\overline{X}$ and the TMd control statistics to departures from the normal model, we have computed, either analitically or through simulation techniques, their alarm rates associated to targets $\mu_0 = 0$ and $\sigma_0 = 1$, without loss of generality, and 3-sigma charts. We have considered both symmetric (standard normal, Student-t and Logistic) and asymmetric (standard Weibull(θ), gamma(θ), inverse gaussian and lognormal) models. For details on these models, quite common in practice, see Hawkins and Olwell (1998) and Johnson et al.

(1994, 1995). The inverse Gaussian model, $GI(\mu, \zeta)$, has a probability density function (p.d.f.) $f(x; \mu, \zeta) = \sqrt{\zeta/(2\pi x^3)} \exp\left(-\zeta(x-\mu)^2/(2x\mu^2)\right)$, $x \geq 0$. The lognormal model herewith considered, denoted $LN^*(\delta)$, has null mean value, unit variance, and δ is the standard deviation of the generating normal process.

The alarm rates are presented in Table 3, for both charts, and for rational subgroups of size $n = 5, 10, 15$ and 20. It is clear from Table 3, partially pictured in Figure 1, that there is a reasonably high variability of alarm rates for both charts when the model is no longer normal (much more evident for the $\overline{X}$-chart). Neverthless, for the usual rational subgroup size, $n = 5$, the differences to the normal-case are much smaller whenever we consider the TMd-chart, even for asymmetric models with a high tail-weight, like the χ^2_1. We have separated symmetric and asymmetric distributions. We have ordered the symmetric distributions by the tail-weight coefficient τ in (1.8). The asymmetric distributions have been ordered by the skewness coefficient β in (1.7). The set of models considered enables us to get values of β between 0 and 33.5 and values of τ between 1 and 3.5.

	$\overline{X}$-chart				TMd-chart			
Distribution	$n=5$	$n=10$	$n=15$	$n=20$	$n=5$	$n=10$	$n=15$	$n=20$
$D0: Normal$	.00280	.00277	.00266	.00275	.00276	.00283	.00271	.00277
$D1: t_{20}$	.00315	.00299	.00297	.00280	.00268	.00236	.00206	.00206
$D2: t_{10}$	.00387	.00335	.00320	.00309	.00264	.00181	.00156	.00141
$D3: Logistic$	.00413	.00356	.00326	.00310	.00273	.00161	.00129	.00108
$D4: t_3$	.01070	.00910	.00836	.00791	.00197	.00020	.00005	.00003
$D5: W(2)$	.00316	.00303	.00292	.00283	.00301	.00281	.00330	.00410
$D6: \chi^2_{20}$	.00365	.00318	.00299	.00327	.00295	.00227	.00234	.00237
$D7: \chi^2_{10}$	.00442	.00356	.00329	.00330	.00316	.00193	.00205	.00268
$D8: Ga(2)$	.00663	.00490	.00426	.00381	.00389	.00134	.00113	.00210
$D9: LN^*(0.5)$	.00844	.00622	.00520	.00464	.00360	.00096	.00057	.00096
$D10: Exponential$	.00931	.00665	.00571	.00489	.00433	.00101	.00039	.00064
$D11: \chi^2_1$	.01280	.00934	.00752	.00669	.00431	.00057	.00013	.00003
$D12: Ga(0.5)$	.01285	.00922	.00761	.00674	.00420	.00060	.00016	.00005
$D13: GI(1,1)$	.01354	.01010	.00815	.00717	.00390	.00050	.00010	.00003
$D14: LN^*(1)$	.01611	.01416	.01272	.01183	.00302	.00018	.00002	.00000
$D15: W(0.5)$	.01916	.01693	.01497	.01356	.00283	.00010	.00001	.00000
Control limits	±1.342	$\pm$ 0.949	±0.775	±0.671	±1.390	±1.021	±0.857	±0.751

Table 3: Alarm rates of the 3-sigma limits control charts.

The way the TMd statistic has been devised led us to think first that the TMd-chart would have, in an out-of-control situation, a worse performance than the $\overline{X}$-chart. Astonishingly, the TMd-chart has been even able to overpass the $\overline{X}$-chart for a great diversity of models (Figueiredo and Gomes, 2003), providing a slightly faster alarm signal whenever the process is out-of-control. Such a situation occurs for instance when we are controlling the mean of a logistic or a lognormal process, and use control limits of the statistics placed at the respective simulated $\chi_{0.001}$ and $\chi_{0.999}$ quantiles. These quantiles are given in Table 4. For more detailed tables see Figueiredo (2002).

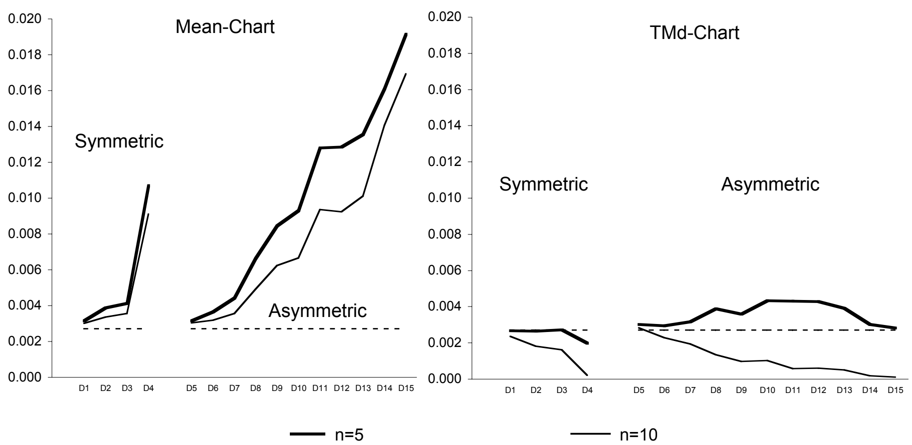

Figure 1: Alarm rates of the 3-sigma $\overline{X}$ and TMd control charts for rational subgroups of size $n = 5$ and $n = 10$.

	Distribution							
	Normal		*Logistic*		$LN^*(0.5)$		$LN^*(1)$	
	$\overline{X}$	TMd	$\overline{X}$	TMd	$\overline{X}$	TMd	$\overline{X}$	TMd
LCL	−1.382	−1.442	−1.475	−1.452	−0.985	−1.028	−0.616	−0.760
UCL	+1.382	+1.425	+1.454	+1.434	+1.946	+1.709	+3.140	+1.811

Table 4: Lower and upper control limits of the $\overline{X}$ and TMd charts, placed at the respective $\chi_{0.001}$ and $\chi_{0.999}$ quantiles.

We here summarize some of the general conclusions:

1. The total median TMd is highly efficient and "robust" for small-to-moderate rational subgroup sizes, the most usual ones in SPC, and should be used in practice, as an alternative to the $\overline{X}$-chart. Notice however that, for very small sample sizes, the sample median exhibits the best performance, and for large sample sizes and not too heavy tail models the sample mean may exhibit the best performance.

2. Together with the consideration of the TMd control statistic, the use of a rational subgroup size $n = 5$ is highly advisable in practice. This choice is mainly due to the fact that in non-normal situations, if n is much larger than 5, the alarm rate of such a control chart can be much smaller than expected.

2.2 Control of a standard deviation

In Figure 2 we present the alarm rates of the classical S and R-charts as well as of the S^*, S^{**}, MAD and TR charts. It is again clear from Figure 2 that there is a reasonably high variability of alarm rates for all the charts when the model is no longer normal.

For the detection of an increase in the standard deviation of the process we may say that the MAD is quite "robust", in the sense of resistant to changes

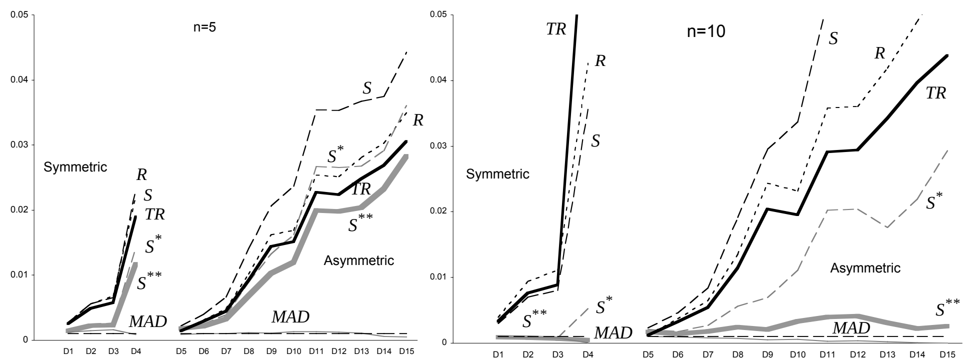

Figure 2: Alarm rates of the upper-control charts for the standard deviation ($n = 5, 10$), with a control limit placed at the .999 quantile.

in the model, not a long way from the S^{**}, whenever the rational subgroup size increases. Contrarily to what happens to the total median, the total range does not exhibit a "robust" behaviour to departures from the normal model.

The large scale simulation study undertaken in Figueiredo (2003), to compare the efficiency and robustness of the above mentioned estimators of the population standard deviation, enables us to draw the following conclusions:

1. The median absolute deviation from the median is, for large sample sizes, the most efficient and robust estimator among the ones considered. Anyway, it is the one with the worst overall performance in out-of-control set-ups.

2. The robustness of all the new proposed standard deviation estimators, although not significant, must be paid for by a reduction in their efficiency. However, even for the normal model, they turn out to be reasonably efficient.

3. For skewed distributions, the total range TR is an efficient estimator of the standard deviation for sample sizes smaller than 15.

4. For symmetric contaminated distributions, S^* and S^{**} are the most efficient and robust estimators for small-to-moderate sample sizes (up to 17), being the TR estimator preferable only for very small sample sizes ($n \leq 5$).

5. The sample standard deviation is more efficient than the others only for symmetric distributions very close to the normal, such as the logistic and the Student-t distribution with a high number of degrees of freedom, although it is not at all robust.

As an overall conclusion of this section we may say that the TMd (TR)-chart is in general the one which provides the highest power, i.e., the smallest Average

Time to Signal (ATS) for most of the models and for the control of a target $\mu_0 = 0$ (σ_0=1). For this reason, and in a situation of compromise between efficiency and robustness, we strongly advise the use of the total median and the total range for controlling the mean and the standard deviation of an industrial process, respectively.

3 The use of Box-Cox transformations, prior to on-line control

Most of the data analysis' statistical procedures, including analysis of variance, regression analysis and control charting, assume the hypothesis of independent and normally distributed observations. And these assumptions rarely hold true in practice. In the literature, many data-transformations have been suggested in order to reach normality, including the Box-Cox transformations in (1.1). For $\lambda = 0$ we obtain the commonly used log-transformation. For most of the usual models we can obtain explicit expressions for the parameters $\mu_\lambda = \left(E(X^\lambda) - 1\right)/\lambda$ and $\sigma_\lambda^2 = Var(X^\lambda)/\lambda^2$, as functions of λ. We may thus construct uniformly most powerful unbiased ($UMPU$) tests and more accurate confidence limits.

3.1 Why non-normal processes?

Despite the usefullness of the normal model in SPC, most of the sets of continuous data related to the most diversified industrial processes exhibit often asymmetry and tails heavier than the normal tail. The normal tail is of an exponential type, i.e. $1 - \Phi(x) \sim \alpha \exp(-x^2)$, with an exponent equal to 2, greater than 1, i.e., the normal tail has a penultimate light tail behaviour in a context of Extreme Value Theory (Beirlant et al., 1995; Gomes and de Haan, 1999). Data from areas as telecommunication traffic, insurance, finance and reliability usually exhibit both skew p.d.f.'s and heavy tails (see, for instance Hawkins and Olwell, 1998). Many failure times are well modeled by the exponential distribution; however, in some practical situations, the assumption of a constant failure rate is very strong, and the use of a gamma distribution or a Weibull distribution, which have monotone failure rates, may be an appropriate alternative. Sometimes it is also convenient to have an initially decreasing and then increasing failure rate, and the lognormal and the inverse gaussian distributions may be then more appropriate. It should also be pointed out that the estimation in the inverse gaussian model is very simple, and this distribution has a non-zero asymptotic failure rate; this has provided a strong argument for its use rather than the lognormal.

3.2 Estimation of λ and efficiency of a Box-Cox transformation

In this paper we compare different methods of estimation of the parameter λ in (1.1). The first two methods are the ones suggested in Box and Cox (1964): the maximum likelihood (ML) estimate and an adequate explicit approximation for the ML-estimate. Alternatively to these two methods of estimation, we shall consider here an additional method, denoted R^2-method, and advocate its use in practice. This was due to some peculiarities of the explicit "ML"-estimates in 3.2.2 and to convergence problems of the numerical procedures used to solve the ML-equation providing the ML-estimate in 3.2.1.

Let us in the sequel assume that we have access to a sample $(x_1, x_2, \cdots, x_n)$, and denote $(z_1, z_2, \cdots, z_n)$ the Box-Cox transformed data. Denoting f the p.d.f. of the original data x, defined in the domain D_f, we have for a fixed value of λ in (1.1), the p.d.f.

$$g(z;\lambda) = \begin{cases} (1+\lambda z)^{1/\lambda - 1} f\left((1+\lambda z)^{1/\lambda}\right) & \text{if } \lambda \neq 0 \text{ and } (1+\lambda z)^{1/\lambda} \in D_f \\ e^z f(e^z) & \text{if } \lambda = 0 \text{ and } e^z \in D_f \end{cases} .$$

3.2.1 The ML-estimate

Assuming $\lambda \neq 0$, and denoting,

$$\widehat{\mu}_\lambda := \frac{1}{n}\sum_{i=1}^{n} z_i =: \overline{z}, \quad \widehat{\sigma}^2_\lambda := \frac{1}{n}\sum_{i=1}^{n} (z_i - \overline{z})^2, \quad z_i = \frac{x_i^\lambda - 1}{\lambda}, 1 \leq i \leq n,$$

in order to obtain $\widehat{\lambda}$, we need to maximize

$$\ln L(\lambda, \widehat{\mu}_\lambda, \widehat{\sigma}^2_\lambda; x_1, \cdots, x_n) = -n \ln \widehat{\sigma}_\lambda + (\lambda - 1)\sum_{i=1}^{n} \ln x_i - n \ln \sqrt{2\pi} - n/2.$$

The ML-equation is then

$$\frac{1}{\widehat{\sigma}^2_\lambda}\sum_{i=1}^{n} (z_i - \overline{z})^2 \ \frac{dz_i}{d\lambda} - \sum_{i=1}^{n} \ln x_i = 0, \tag{3.1}$$

which needs to be solved numerically. Equivalently, we may solve the equation

$$n\sum_{i=1}^{n} (z_i - \overline{z})\left((1+\lambda z_i)\ln(1+\lambda z_i) - \lambda z_i\right) - \lambda \sum_{i=1}^{n} (z_i - \overline{z})^2 \sum_{i=1}^{n} \ln(1+\lambda z_i) = 0. \tag{3.2}$$

3.2.2 An explicit "ML"-estimate

The alternative estimate of λ, also proposed by Box and Cox (1964), is based on a previous transformation of our original data,

$$w_i = \frac{x_i}{gm(x)} - 1, \text{ with } gm(x) := \sqrt[n]{\prod_{i=1}^{n} x_i}, \tag{3.3}$$

the geometric mean of our sample $x_i,\ 1 \le i \le n$. They then estimate λ, through the maximum likelihood technique, so that

$$z_i = \frac{(w_i+1)^\lambda - 1}{\lambda}, \quad 1 \le i \le n,$$

are Normal$(\mu_\lambda, \sigma_\lambda)$, but they consider the approximation

$$\frac{(w_i+1)^\lambda - 1}{\lambda} = w_i + \sum_{j=2}^{\infty} \frac{\prod_{l=1}^{j-1}(\lambda - l)}{j!} w_i^j \approx w_i - \alpha w_i^2 + \tfrac{2}{3}\alpha\left(\alpha + \tfrac{1}{2}\right) w_i^3,\ \alpha = \frac{1-\lambda}{2}.$$

They finally get

$$\widehat{\lambda} = 1 - 2\ \frac{3gm(x)(m_3 + 2m_2 d) - m_4 - 3m_3 d - 3m_2 d^2}{7m_4 + 24m_3 d + 24m_2 d^2 - 3m_2^2}, \tag{3.4}$$

where

$$d := \overline{x} - gm(x), \quad m_1 = \overline{x}, \quad m_p = \tfrac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^p, \quad 2 \le p \le 4.$$

3.2.3 The R^2-estimate

This method is essentially based on the probability paper technique. Consider the ordered sample $(x_{1:n} \le x_{2:n} \le \cdots \le x_{n:n})$ and the transformed values $z_{i:n} = \left(x_{i:n}^\lambda - 1\right)/\lambda,\ 1 \le i \le n$. Let us then assume that z_i are approximately normal with mean value μ_λ and standard deviation σ_λ. Then, with $p_i = i/(n+1)$, the points $(z_{i:n}, \Phi^{\leftarrow}(p_i),\ 1 \le i \le n)$, should approximately be over a straight line. Such a consideration leads us to suggest as a possible estimate of λ the value of λ which maximizes R^2, with R given by

$$R = \frac{\sum_{i=1}^{n}(z_{i:n} - \overline{z})\left(\Phi^{\leftarrow}(p_i) - \overline{\Phi^{\leftarrow}(p_i)}\right)}{\sqrt{\sum_{i=1}^{n}(z_{i:n} - \overline{z})^2}\ \sqrt{\sum_{i=1}^{n}\left(\Phi^{\leftarrow}(p_i) - \overline{\Phi^{\leftarrow}(p_i)}\right)^2}}.$$

The R^2-estimate of λ is then solution of the equation

$$\sum_{i=1}^n (z_{i:n}-\overline{z})^2 \;\sum_{i=1}^n \Phi^{\leftarrow}(p_i)\,\frac{d}{d\lambda}\,(z_{i:n}-\overline{z}) \\ -\sum_{i=1}^n (z_{i:n}-\overline{z})\,\Phi^{\leftarrow}(p_i)\;\sum_{i=1}^n (z_{i:n}-\overline{z})\,\frac{d}{d\lambda}\,(z_{i:n}-\overline{z}) = 0, \quad (3.5)$$

which also needs to be solved numerically.

3.2.4 An explicit R^2-estimate

Under this context we may similarly apply to our data the transformation defined in (3.3), and consider then the approximation

$$z_{i:n} \approx w_{i:n} + \frac{\lambda-1}{2}\; w_{i:n}^2,\;\; 1 \le i \le n.$$

Then, with the notation,

$$A=\sum_{i=1}^n w_{i:n},\quad B=\sum_{i=1}^n w_{i:n}^2,\quad C=\sum_{i=1}^n w_{i:n}^3,\quad D=\sum_{i=1}^n w_{i:n}^4,$$

$$E=\sum_{i=1}^n w_{i:n}\Phi^{\leftarrow}(p_i),\quad F=\sum_{i=1}^n w_{i:n}^2\Phi^{\leftarrow}(p_i),\quad G=\sum_{i=1}^n \Phi^{\leftarrow}(p_i),\quad H=\sum_{i=1}^n (\Phi^{\leftarrow}(p_i))^2$$

we get the following explicit estimate for λ:

$$\widehat{\lambda} = 1 + 2\;\frac{n\,(B\times F - C\times E) + A\,(E\times G - A\times F) + B\,(A\times E - B\times G)}{n\,(E\times D - C\times F) + B\,(A\times F - E\times B) + G\,(B\times C - A\times D)}. \quad (3.6)$$

3.2.5 Comparison of the methods

To evaluate the efficiency of the estimators of λ herewith proposed, and to analyze the performance of the Box-Cox transformations, we have again considered a reasonably large set of models with different degrees of skewness and tail-weight. Indeed, in this comparison we have taken the following models, with support $[0,\infty)$: the standard exponential, χ^2 distributions with 1, 3, 5, 10 and 20 degrees of freedom, lognormal distributions, $LN(\mu,\sigma)$, with mean value $\mu = 0$ and standard deviation $\sigma = 0.5$ and 1, Weibull(δ,θ) models with scale $\delta = 1$ and shape $\theta = 0.5$ and 2 (denoted $W(\theta) \equiv W(1,\theta)$), Gamma$(\delta,\theta)$ models with scale $\delta = 1$ and shape $\theta = 0.5$ and 2 (denoted $Ga(\theta) \equiv Ga(1,\theta)$) and inverse Gaussian models, $GI(\mu,\zeta)$. We have also considered underlying normal models, with a mean value equal to 5, and unit variance; this was done essentially to evaluate the estimates in terms of deviation from the known target $\lambda = 1$. A similar evaluation may be undertaken through the estimates obtained for lognormal parents, because then we know that λ should be equal to 0. For each model we have generated 100 samples of size 100 and we have estimated λ through any of the proposed methods. We have also got the p-value associated to the Shapiro-Wilk normality test statistic, associated to the transformed sample. Both for the 100 λ-estimates and p-values have we computed their average and standard deviation.

For the ML and the R^2 estimation, and for all the models, we have obtained large p-values (between 0.257 and 0.655). Such a fact assesses the good performance of the Box-Cox transformation. For each distribution, the normality assumption of the transformed data has been rejected for not more than 5% of the samples, except in the case of very skewed distributions such as the chi-square with one degree of freedom, χ_1^2, and for some inverse gaussian distributions. For the lognormal distributions we have obtained values $\widehat{\lambda}$ very close to zero, as expected. For the normal distribution we have obtained λ-estimates very close to one, also as expected. When we consider either the ML or the R^2 method, we get very satisfactory results, as may be seen in Table 5, where we present the ML and R^2-estimates of λ and of the *p-value* for different models, after discarding the values of the samples for which we have rejected the normality of the transformed data. The differences between these two methods do not appear to be significant, and the R^2 estimates are easier to obtain.

	ML-estimates			R^2 -estimates		
Distribution	$\widehat{\lambda}$	$IC_{95\%}\ \lambda$	$\widehat{\text{p-value}}$	$\widehat{\lambda}$	$IC_{95\%}\ \lambda$	$\widehat{\text{p-value}}$
χ_{20}^2	0.346	(0.291,0.401)	0.611	0.369	(0.311,0.427)	0.616
χ_{10}^2	0.340	(0.307,0.374)	0.607	0.341	(0.307,0.375)	0.602
χ_5^2	0.300	(0.272,0.323)	0.626	0.313	(0.288,0.338)	0.634
χ_3^2	0.290	(0.270,0.311)	0.560	0.310	(0.288,0.331)	0.545
χ_1^2	0.211	(0.204,0.218)	0.315	0.231	(0.223,0.238)	0.317
$LN(1,0.5)$	0.016	(-0.021,0.052)	0.638	0.013	(-0.027,0.052)	0.614
$LN(1,1)$	-0.004	(-0.022,0.015)	0.674	-0.001	(-0.020,0.018)	0.669
$W(2)$	0.518	(0.494,0.542)	0.485	0.551	(0.525,0.578)	0.480
$Exp(1)$	0.271	(0.259,0.282)	0.473	0.284	(0.272,0.296)	0.472
$W(0.5)$	0.130	(0.125,0.136)	0.533	0.138	(0.132,0.144)	0.531
$GI(5,25)$	-0.028	(-0.067,0.011)	0.560	-0.028	(-0.069,0.012)	0.544
$Ga(0.5)$	0.204	(0.197,0.211)	0.337	0.226	(0.218,0.234)	0.330
$Ga(2)$	0.312	(0.292,0.331)	0.531	0.311	(0.290,0.332)	0.512
$GI(5,10)$	-0.061	(-0.084,-0.038)	0.530	-0.060	(-0.084,-0.036)	0.516
$GI(1,1)$	-0.079	(-0.100,-0.058)	0.450	-0.084	(-0.107,-0.062)	0.450
$GI(2,1)$	-0.107	(-0.121,-0.092)	0.370	-0.115	(-0.131,-0.099)	0.370
$GI(5,1)$	-0.136	(-0.149,-0.124)	0.318	-0.146	(-0.159,-0.132)	0.315

Table 5: ML and R^2 estimates of λ, and *p-value* of Shapiro-Wilk test of normality.

When we consider any of the explicit methods, in (3.4) and in (3.6), there is a high number of average p-values smaller then 0.01. Consequently, we do not advise the consideration of such explicit methodologies.

Remark 3.1. *If we consider as a measure of performance of the Box-Cox normalization method, the minimum p-value obtained among the large class of models herewith considered, we may advance that if we think sensible that such a value should be larger than 0.25, it is necessary to base our estimation on samples of size equal to at least 100.*

3.3 Power of control charts after a Box-Cox data transformation

In this study, to evaluate the binomium robustness/efficiency of the Box-Cox transformation, we have considered an *a priori* application of a Box-Cox transformation in order to obtain normal data, before monitoring the process data. After having assessed in section 3.2, by simulation, the performance of this kind of transformation to normalize data from some symmetric as well as asymmetric distributions, we think sensible to use such a transformation in practice, and to proceed with the monitoring of the transformed data, as approximately normal. However, the bad results obtained whenever controlling the standard deviation of the transformed data, lead us to consider here only the monitoring of the mean value, both trough the $\overline{X}$ and the TMd charts.

We are first going to consider Weibull and Gamma processes, $W(\delta,\theta)$ and $Ga(\delta,\theta)$, respectively, both generalizing the exponential model ($\theta = 1$), i.e., we shall assume that our characteristic quality X has a Weibull distribution function (d.f.)

$$F_W(x;\delta,\theta) = 1 - e^{-(x/\delta)^\theta},\ x \geq 0, \quad \delta > 0, \quad \theta = 0.5, 1.0 \text{ and } 2,$$

or a gamma p.d.f.

$$f_G(x;\delta,\theta) = \frac{1}{\delta^\theta\ \Gamma(\theta)}\ x^{\theta-1}\ e^{-x/\delta},\ x \geq 0, \quad \delta > 0, \quad \theta = 0.5, 1 \text{ and } 2.$$

We next present the values of μ_λ and σ^2_λ, both for Weibull and gamma models.

Model	μ_λ	σ^2_λ
$W(\delta,\theta)$	$\frac{1}{\lambda}\left(\delta^\lambda\Gamma\left(\frac{\lambda}{\theta}+1\right)-1\right)$	$\frac{\delta^{2\lambda}}{\lambda^2}\left(\Gamma\left(\frac{2\lambda}{\theta}+1\right)-\Gamma^2\left(\frac{\lambda}{\theta}+1\right)\right)$
$Ga(\delta,\theta)$	$\frac{1}{\lambda}\left(\frac{\delta^\lambda\Gamma(\lambda+\theta)}{\Gamma(\theta)}-1\right)$	$\frac{\delta^{2\lambda}}{\lambda^2}\left(\frac{\Gamma(2\lambda+\theta)}{\Gamma(\theta)}-\frac{\Gamma^2(\lambda+\theta)}{\Gamma^2(\theta)}\right)$

The parameters λ, μ_λ and σ_λ were adequately estimated (see section 4) and transformed standardized data were controlled through 3-sigma $\overline{X}$ and TMd control charts. We here compare the values of the power function of the charts $\overline{X}$ and TMd, applied to the transformed "normal" data, with the power function of the $UMPU$ chart based on the maximum likelihood estimator of δ. We have chosen an α-risk equal to 0.002. The charts were implemented for rational subgroups of size $n = 5$. Without loss of generality, we have assumed $\delta = 1$ when the process is in-control. All the remaining parameters were considered known and fixed (location placed at 0 and shape parameter $\theta = \theta_0 = 0.5,\ 1$ (Exponential), and 2). For the original data we have implemented $UMPU$ charts based on the exact distribution of the statistics referred in Table 6; given an α-risk, the upper (UCL) and lower (LCL) control limits of these charts verify the conditions

$$\begin{cases} G_{\text{in-control}}(UCL) - G_{\text{in-control}}(LCL) = 1 - \alpha \\ \\ \eta_{\text{out-of-control}} \geq \eta_{\text{in-control}} = \alpha, \end{cases}$$

where G, $\eta = P(\text{chart signal})$ and α denote the d.f., the power function and the alarm rate, respectively, of the control statistic. The charts were implemented for rational subgroups of size $n = 5$, and the control limits were computed so that we have an alarm rate of 0.0020. Apart from the Weibull and gamma distributions, we have also considered lognormal and inverse gaussian distributions, and we have implemented control charts to monitor the process mean. Without loss of generality, we have assumed $\mu = 1$ and $\sigma = \sigma_0 = 0.5$ or 1, when the process is in-control.

Distribution	$Exp(\delta)$	$W(\delta,\theta_0)$	$Ga(\delta,\theta_0)$	$GI(\mu,1)$	$LN(\mu,\sigma_0)$
Control statistic	$\overline{X}$	$\left(\frac{1}{n}\sum_{i=1}^{n} X_i^{\theta_0}\right)^{1/\theta_0}$	$\overline{X}$	$\overline{X}$	$\overline{\ln X}$

Table 6: $UMPU$ control statistics for the original data.

To evaluate the efficiency of these control charts, $\overline{X}$, TMd and $UMPU$, we have computed the power function for some magnitudes of shift in the parameter δ or μ. For the $UMPU$ charts we have used the sampling distribution of the control statistic to compute the power function; for the charts implemented for the transformed data we have evaluated it by simulation, through a sample of 1,000,000 values of the control statistic, which allow us to obtain results with four-digits accuracy.

If we apply an *a priori* Box-Cox transformation to the data, the TMd chart is "totally robust" to the normality assumption, while the $\overline{X}$ chart is only "quasi-robust", as we can observe through the values presented in Tables 7 and 9, when we look at the entries for $\delta = 1$ and $\mu = 1$, respectively. We can also observe that the power function values of the charts implemented for the transformed data are similar, and we are led to conclude that these charts are reasonably efficient comparatively to the efficiency of the charts implemented for the original data.

	$W(\delta, 0.5)$		$Exp(\delta)$		$W(\delta, 2)$		$Ga(\delta, 0.5)$		$Ga(\delta, 2)$	
δ	$\overline{X}$	TMd	$\overline{X}$	TMd	$\overline{X}$	TMd	$\overline{X}$	TMd	$\overline{X}$	TMd
0.1	.0696	.0577	.7034	.6027	1.0000	1.0000	.1196	.0957	.9999	.9996
0.3	.0105	.0098	.0642	.0540	.7883	.6938	.0143	.0142	.4590	.3883
0.5	.0040	.0041	.0105	.0101	.1170	.0956	.0048	.0054	.0664	.0556
0.7	.0021	.0023	.0028	.0030	.0125	.0117	.0022	.0026	.0101	.0092
0.9	.0017	.0020	.0014	.0016	.0021	.0021	.0014	.0019	.0022	.0022
1.0	.0017	.0020	.0017	.0021	.0016	.0019	.0016	.0020	.0019	.0020
1.1	.0020	.0024	.0031	.0036	.0056	.0063	.0018	.0024	.0036	.0037
1.3	.0033	.0037	.0104	.0111	.0494	.0494	.0037	.0047	.0180	.0183
1.5	.0054	.0060	.0266	.0269	.1661	.1605	.0078	.0093	.0563	.0548
1.7	.0085	.0093	.0531	.0526	.3393	.3223	.0144	.0167	.1230	.1183
1.9	.0125	.0134	.0894	.0873	.5178	.4909	.0235	.0266	.2144	.2035
2.0	.0149	.0161	.1121	.1088	.5975	.5675	.0288	.0325	.2652	.2512
3.0	.0488	.0500	.3781	.3574	.9520	.9349	.1093	.1159	.7209	.6907

Table 7: Efficiency of the $Mean$ and TMd charts for the transformed data ($n = 5$).

δ	$W(\delta, 0.5)$	$Exp(\delta)$	$W(\delta, 2)$	$Ga(\delta, 0.5)$	$Ga(\delta, 2)$
0.1	.1160	.9041	1.0000	.2321	1.0000
0.3	.0172	.1360	.9439	.0268	.5768
0.5	.0063	.0244	.2247	.0081	.0958
0.7	.0033	.0068	.0262	.0038	.0150
0.9	.0022	.0025	.0037	.0022	.0030
1.0	.0019	.0021	.0019	.0020	.0020
1.1	.0021	.0024	.0046	.0022	.0032
1.3	.0031	.0080	.0478	.0046	.0160
1.5	.0049	.0225	.1789	.0104	.0538
1.7	.0081	.0491	.3718	.0206	.1233
1.9	.0124	.0874	.5651	.0361	.2197
2.0	.0149	.1112	.6473	.0455	.2736
3.0	.0541	.4046	.9679	.1825	.7442

Table 8: Efficiency of the $UMPU$ charts implemented for the original data ($n = 5$).

The alarm rates of the $\overline{X}$ and the TMd statistics when we change the model are pictured in Figure 3, as a function of the shape parameter θ in a Weibull model. We may then easily notice the resistance of the TMd statistic to changes in the model, comparatively to the classical $\overline{X}$ statistic.

Regarding the robustness property we have at the moment no doubt that a statistic like the total median, TMd, is quite resistant to changes in the underlying model. We should then expect a high decrease in the power function of a

	$LN(\mu,0.5)$			$LN(\mu,1)$			$GI(\mu,1)$		
μ	$\overline{X}$	TMd	$UMPU$	$\overline{X}$	TMd	$UMPU$	$\overline{X}$	TMd	$UMPU$
0.1	1.0000	.9999	.9999	.9994	.9990	.9989	1.0000	1.0000	1.0000
0.3	.9824	.9765	.9789	.9133	.8930	.8625	.2506	.2010	.4019
0.5	.7613	.7317	.7211	.5044	.4702	.3741	.0182	.0157	.0383
0.7	.2328	.2138	.1838	.1100	.0994	.0521	.0035	.0034	.0074
0.9	.0111	.0105	.0067	.0076	.0070	.0027	.0014	.0015	.0026
1.0	.0020	.0020	.0020	.0021	.0021	.0020	.0015	.0019	.0020
1.1	.0029	.0030	.0043	.0018	.0019	.0034	.0026	.0032	.0024
1.3	.0196	.0176	.0280	.0052	.0050	.0103	.0078	.0092	.0081
1.5	.1032	.0880	.1383	.0152	.0133	.0290	.0179	.0205	.0226
1.7	.3569	.3062	.4335	.0401	.0336	.0735	.0317	.0349	.0462
1.9	.7325	.6662	.8006	.0953	.0783	.1644	.0484	.0517	.0765
2.0	.8739	.8232	.9164	.1403	.1145	.2331	.0580	.0609	.0933
3.0	1.0000	1.0000	1.0000	.9489	.9095	.9838	.1548	.1510	.2611

Table 9: Efficiency of the $Mean$ and TMd charts for the transformed data, and efficiency of the $UMPU$ charts implemented for the original data ($n = 5$).

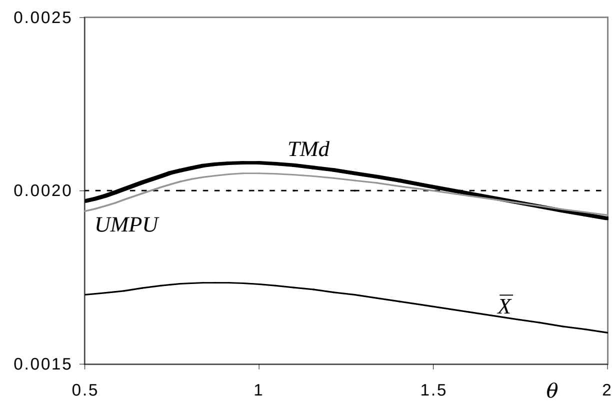

Figure 3: Alarm rates of the 3-sigma $\overline{X}$ and TMd control charts and the $UMPU$ based on a ML estimate of δ in Weibull(δ, θ) models.

control chart based on such a statistic, but is not really what has happened for the families of simulated models, as may be partially seen in the tables presented.

Relatively to the charts for the standard deviation, and as mentioned before, a Box-Cox transformation applied to the data before monitoring the process do not lead us to obtain robust control charts, being however the S^*, the S^{**} and the MAD chart the ones with higher robustness, when implemented for the transformed data.

Remark 3.2. *Assuming we have an a priori knowledge of the underlying model, the $UMPU$-chart, applied to the original data, performs usually better than any statistic applied to the transformed data, but not a long way from it. Note that the knowledge of the underlying model is not obviously the usual situation in practice, where a preliminary data analysis should be done together with the testing of the postulated model.*

4 Some overall conclusions

Since modelling our data adequately is obviously a delicate problem in Statistics, and the ATS of the charts based on the Box-Cox transformed data and based on the original data (together with the knowledge of the specific underlying model, assumed to be known), do not differ significantly, we suggest, as a possible alternative to data modelling, the use of the Box-Cox transformation, whenever we have in mind monotoring the mean value of an industrial process.

4.1 Design of the experiment

In a practical situation in which we have available an *a priori* sample of size k observations, say $k = 500$, we advise the use of the bootstrap methodology based on the resampling of our sample in sub-samples of size $m = 100$, in order to get more precise estimates of λ, μ_λ and σ_λ. The results so far describeded in this paper, lead us to propose the following way of monitoring our data:

1. Collect a reasonable large data set associated to any relevant quality characteristic X, say $(X_1, X_2, \cdots, X_k)$, $k = 500$, a sample we consider to be representative of such a quality characteristic whenever the industrial process is *in control*.

2. On the basis of such a sample and using a *bootstrap* procedure, generate a sufficiently large number B, say $B = 5000$, of subsamples of size $m = 100$. Using either the maximum likelihood method in (3.1) or the R^2 method in (3.5), compute B partial estimates, $\widehat{\lambda}_i$, $\widehat{\mu}_{\lambda i}$ and $\widehat{\sigma}_{\lambda i}$, $1 \le i \le B$, of the unknown parameters λ, μ_λ and σ_λ, respectively and consider their average, $\widehat{\widehat{\lambda}} \equiv \widehat{\widehat{\lambda}}_B = \sum_{i=1}^{B} \widehat{\lambda}_i/B$, $\widehat{\widehat{\mu}}_\lambda$ and $\widehat{\widehat{\sigma}}_\lambda$, as the overall estimates to be used.

3. To control the on-line production, collect rational subgroups of size $n = 5$ along time, apply first the Box-Cox transformation in (1.1), with λ replaced by $\widehat{\widehat{\lambda}}$, and consider the standardized data $\left(z_i - \widehat{\widehat{\mu}}_\lambda\right)/\widehat{\widehat{\sigma}}_\lambda$. The data obtained may then be regarded as approximately standard normal, being then safe to use control limits adequate for such a model.

4. To control the mean value of the industrial process, use the total median TMd-chart instead of the usual $\overline{X}$-chart.

5. To control the standard deviation, the TR-statistic should be used, but it is more efficient, if possible, to use the original data, and a procedure similar to the one developed in 2., but for the estimation of the unknown model parameters.

Remark 4.1. *As time goes on, it is sensible to use the larger quantity of data available in-control, and re-evaluate the estimates needed.*

4.2 Some overall comments

- Normality tests should always be applied to our data, in order to detect possible departures from the common normality assumption. Indeed control charts based on a normal model exhibit a very poor performance when the underlying parents are either asymmetric or symmetric with heavy tails.

- It is in general possible to find an adequate Box-Cox transformation, which enables us to deal with the transformed data as approximately normal, applying then procedures based on such a common model. Such procedures have revealed to be highly efficient when we are monitoring the mean value of the process. The estimation of the parameter λ in the Box-Cox transformation in (1.1) needs to be done with care, and is one of the main points of the proposed methodology. In practice we advocate the use of the bootstrap methodology, together with either the ML or the R^2 estimates of λ (see the algorithm above). Taylor series expansions enable us to get explicit estimates of λ, but they are poor and should not be used in practice.

- After the proper identification of the transformation, we should monitor the mean value and standard deviation of the transformed data, through control charts for normal parents, based on "robust" statistics, like the ones described in this paper.

- The power of the proposed charts, applied to the transformed data, is reasonably high, sometimes higher than the one achieved when we identify the model underlying our data, and use control charts specifically devised for such a type of model.

4.3 Data modelling versus Box-Cox transformations

- Data modelling may be preferable to the use of Box-Cox transformations provided that it is clear how to specify properly the underlying model, that the estimation of the unknown model parameters is not problematic and that it is easy to find a $UMPU$ test for the unknown parameter of interest. Such an approach has revealed to be particularly useful when we need monitoring the process standard deviation.

- However, there is always the possibility of mispecification of the model. And even if we assume that a proper specification of the model has been achieved, the estimation of the unknown parameters may be problematic, at least as problematic as the estimation of the unknown λ parameter in the Box-Cox transformation (1.1).

- The gain in terms of power, when we use $UMPU$-charts for a specific parameter of a known parent, whenever such a $UMPU$ statistic is easily available, is not usually significant, comparatively to the monitoring of the mean value of the transformed data through a "robust" statistic like the total median, TMd, together with the associated normal control levels.

- For the control of a mean value:
  - whenever we decide modelling our data, we suggest the TMd-chart for rational subgroups of size $n = 5$, as an efficient and "quasi-robust" control chart for the mean value. However, in situations of non-normality the alarm rate of the TMd-chart can be much smaller than expected, particularly as n increases, and must thus be used with care;
  - if we previously apply a normalizing Box-Cox transformation to the data, the TMd chart is "totally robust" to the normality assumption, provided we consider again rational subgroups of size $n = 5$. Indeed, the transformed data are only approximately normal, the $\overline{X}$-chart to control the mean is no longer $UMPU$, and the efficiency of the TMd chart for transformed data is reasonably high — in some cases similar to the one achieved by the charts implemented for the original data.
- For the control of a standard deviation:
  - we suggest the TR-chart, implemented for the original data. This chart is reasonably efficient and although cannot be considered "robust", it presents higher robustness than the usual S and R charts. It is thus important to model the data adequately and to obtain accurate estimates of the unknown parameters, whenever the process may be considered under control;
  - even doing an apriori application of a Box-Cox transformation to the original data in order to obtain normal values, the charts herewith proposed exhibit quite variable alarm rates, and cannot thus be considered "robust". Anyway, the S^*, the S^{**} and the MAD charts are the ones which exhibit a higher degree of "robustness". Indeed, our study suggests, in terms of efficiency/robustness, the use of the proposed charts for the standard deviation (TR, S^* and S^{**}), instead of the usual ones (S and R), both to monitor the original or the transformed data.

References

[1] Beirlant, J., Broniatowski, M., Teugels, J. and P. Vynckier (1995). The mean residual life function at great age: applications to tail estimation. *J. Statist. Planning and Inference* **45**, 21-48.

[2] Borror, C. N., Montgomery, D. C. and G. C. Runger (1999). Robustness of the EWMA control chart to non-normality. *J. Quality Technology* **31**, 309-316.

[3] Box, G. E. P. and D. R. Cox (1964). An analysis of transformations. *J. Royal Statist. Soc.* , 211-252.

[4] Figueiredo, F. (2002). *Controlo Estatístico da Qualidade e Métodos Robustos*. Ph. D. Thesis. Faculty of Science, Univ. Lisbon.

[5] Figueiredo, F. (2003). Robust estimators for the standard deviation based on a bootstrap sample. *Proc. Young Statisticians European Meeting*.

[6] Figueiredo, F. and M. I. Gomes (2001). The total median in Statistical Quality Control. *Revista de Estatística* 2001-2, 143-144.

[7] Figueiredo, F. and M. I. Gomes (2003). *The total median is Statistical Quality Control*. Notas e Comunicações CEAUL 6/2003. Submitted.

[8] Gomes, M. I. and L. de Haan (1999). Approximation by penultimate extreme value distributions. *Extremes* **2**:1, 71-85.

[9] Hampel, F. R., Ronchetti, E. M., Rousseew, P. J. and W. Stahel (1986). *Robust Statistics: the Approach Based on Influence Functions*. Wiley, New York.

[10] Hawkins, M. and D. H. Olwell (1998). *Cumulative Sum Charts and Charting for Quality Improvement*. Springer-Verlag.

[11] Hoaglin, D. C., Mosteller, F. and L. G. Tukey (1983). *Understanding Robust and Exploratory Data Analysis*. Wiley, New York.

[12] Johnson, N. L., Kotz, S. and N. Balakrishnan (1994). *Continuous Univariate Distributions*, Vol. 1, 2nd edition, Wiley, New York.

[13] Johnson, N. L., Kotz, S. and N. Balakrishnan (1995). *Continuous Univariate Distributions*, Vol. 2, 2nd edition, Wiley, New York.

[14] Lax, D. A. (1985). Robust estimators of scale: finite sample performance in long-tailed symmetric distributions. *J. Amer. Statist. Assoc.* **80**, 736-741.

[15] Ryan, T. P. and B. J. Faddy (2001). The effect of non-normality on the performance of CUSUM procedures. In Lenz, H. L. and P.-Th. Wilrich (eds.), *Frontiers in Statistical Quality Control* 6, Physica-Verlag, 176-193.

[16] Shore, H. (2001). Process control for non-normal populations based on an inverse normalizing transformation. In Lenz, H. L. and P.-Th. Wilrich (eds.), *Frontiers in Statistical Quality Control* **6**, Physica-Verlag, 194-206.

[17] Spedding, T. A. and P. L. Rawlings (1994). Non-normality in statistical process control measurements. *International J. Quality and Reliability Management* **11**:6, 27-37.

[18] Stoumbos, Z. G. and M. R. Jr. Reynolds (2000). Robustness to non-normality and autocorrelation of individuals control charts. *J. Statist. Comp. and Simulation* **66**, 145-167.

[19] Tatum, L. G. (1997). Robust estimation of the process standard deviation for control charts. *Technometrics* **39**, 127-141.

[20] Wu, Z. (1996). Asymmetric control limits of the X-bar chart for skewed process distributions. *International J. Quality* **13**, 9, 49-60.