5-1-2011

A Simulation Study of the Relative Efficiency of the Minimized Integrated Square Error Estimator (L2E) For Phase I Control Charting

John N. Dyer

Georgia Southern University, jdyer@georgiasouthern.edu

A Simulation Study of the Relative Efficiency of the Minimized Integrated Square Error Estimator (L2E) For Phase I Control Charting

John N. Dyer
Georgia Southern University,
Statesboro, GA, USA

Parameter estimates used in control charting, the sample mean and variance, are based on maximum likelihood estimation (MLE). Unfortunately, MLEs are not robust to contaminated data and can lead to improper conclusions regarding parameter values. This article proposes a more robust estimation technique; the minimized integrated square error estimator (L2E).

Key words: Phase I control charting, SPC, L2E, MLE, parameter estimation.

Introduction

Process monitoring using control charts is the most common method used in statistical process control (SPC). In the literature two phases of control charting are distinguished: Phase I and Phase II control charting. Phase I control charting consists of two stages: Stage 1, the retrospective stage, and Stage 2, the prospective stage (Koning & Does, 2000). During Phase I, the appropriate control charting methods must be determined, and the appropriate process parameters estimated (Jones, 2002).

The techniques associated with Phase I include analyzing sample data using gauge repeatability and reliability (GR&R) studies to investigate measuring system accuracy and variability, using capability indices to determine if a process is capable of producing within specification, using histograms and probability plots to verify distributional assumptions, using outlier detection tools (Ramsey & Ramsey, 2007) to detect and remedy special causes of variation in the process, and obtaining reliable estimates of the process parameters (Montgomery, 1997). Thus, part of Phase I can

be considered a data editing process wherein outlying or contaminated data are removed from the sample to enable estimation of the appropriate process parameters.

Phase II control charting is the actual use of the desired control chart to monitor and control a process in regards to changes in the process parameters (Woodall, 2000), distributional changes, and the randomness of the process. The construction of a Phase II control chart is based on the parameter estimates obtained in Phase I. Common Phase II control charts include the following (applied to either individual process observations or subgroups): the Shewhart-type, the exponential weighted moving average (EWMA), and the cumulative sum (CUSUM), among others (Dyer, Adams & Conerly, 2003).

It is crucial that the data collected in Phase I are good data, meaning, free from outliers (contaminated data) and representative of typical process data with no special causes of variability. Contaminated data can lead to unreliable parameter estimates which, in turn, lead to improper conclusions regarding distribution assumptions, process capability and control chart design. The use of most control charts requires the estimation of the mean, μ, and standard deviation, σ (or a function thereof), of the in-control (IC) process. A process is said to be IC when only common cause variation is present, otherwise it is considered out-of-control (OC).

John N. Dyer is an Associate Professor of Information Systems in the College of Business Administration, with research streams in statistics, information systems, and finance. Email him at: jdyer@georgiasouthern.edu.

The estimates used for the true process mean, μ, and standard deviation, σ, are typically sample statistics, specifically, the sample mean, $\bar{x}$, and the sample standard deviation, s, obtained from the good data. The sample statistics used in Phase I control charting are based on the principle of maximum likelihood estimation, that is, the sample mean and sample variance are maximum likelihood estimates (MLEs) of μ and σ^2, respectively.

Some of the practical deficiencies of MLEs are their lack of resistance to outliers and their general non-robustness with respect to model misspecification (Rudemo, 1982). For example, consider the following 5 data values: 4, 5, 6, 7 and 100, and estimates based on MLEs (the sample mean and standard deviation). The sample mean and variance of all five data values are 24.4 and 1,781, respectively. If the data value of 100 is identified as an outlier and removed, then the new MLEs for the mean and variance are 5.5 and 1.69, respectively. Although the magnitude of the outlier is absurdly large, it is obvious that the MLEs cannot resist the influence of the large value. The values of the new MLEs are dramatically different, but they are more representative of the true nature of the data values. Recall, one emphasis of Phase I control charting is to identity and remove outliers, hence providing reliable estimates of the true process parameters. It should also be noted that, although MLEs are nonresistant to outliers, they are typically preferred because of their constructive nature as well as their asymptotic optimality properties.

To overcome the deficiencies of MLEs and better enable the practitioner to obtain reliable parameter estimates, this article proposes the use of a specific nonparametric density estimation technique using a form of the integrated square error (ISE) estimator, also called L2E. Scott (2001) provides the theoretical construct of the L2E and the interested reader is encouraged to review the article.

In this study, the L2E technique is shown to provide parameter estimates that are robust to contaminated data and to be constructive in nature. For example, considering the full data set previously discussed, the L2E estimates of the mean and variance (obtained through a simply executed Excel spreadsheet algorithm) are 5.5 and 2.25, respectively. Notice how the L2E estimates are robust to the inclusion of the outlier.

Although Scott (2001) introduced the L2E as an estimator of process parameters, evidences the estimator's robustness to outliers in large data sets, and shows its constructive nature, this research explores the properties of the L2E as an alternative estimator to MLE across a broad range of sample sizes and a broad range of data contamination affecting the mean alone, the variance alone, and the mean and variance together. This study also compares the absolute difference between MLE and L2E over the range of sample sizes and contaminations (mean, variance, and mean-variance), and shows that the L2E estimates are as good as MLE estimates in almost all cases. Additionally, the relative efficiency of MLE versus L2E estimates is compared across all cases and it is shown that the L2E estimates are more robust in most cases than MLE estimates.

The literature related to Phase I control charting for univariate processes is limited. Readers are referred to Chou & Champ (1995), Koning & Does (2000), Newton & Champ (1997), Sullivan & Woodall (1996), and Woodall (2000). Surprisingly, the focus of the majority of the literature is devoted to methods for multivariate Phase I SPC (Alt & Smith, 1988; Sullivan & Woodall, 1994; Sullivan, Barrett & Woodall, 1995; Woodall, 2000).

Overview of the Phase I Environment

During Phase I, process data are collected and analyzed to enable Phase II control charting. After the data are collected, the SPC method can be considered as the combination of Phase I and Phase II applications. The general SPC method can be thought of in terms of four design steps. The first three steps occur in the Phase I environment and step 4 occurs in the Phase II environment.

Step 1:

Identify the desired control chart (for monitoring individual observations or subgroup data), the required parameters, and the desired IC average run length (ARL). The IC ARL is the average number of samples taken until an IC

process indicates a statistic outside of the control limits.

Step 2:

Determine the subgroup size, n, and the number of subgroups, m, which will be used to estimate the parameters of the IC process. Obtain a reference sample of m subgroups of size n ≥ 1 observations.

Step 3:

Ensure that the reference sample is representative of the IC process, simultaneously estimating the required control chart parameters using a robust technique, such as, L2E (recommended herein) or an iteratively robust technique like MLE.

Step 4:

Apply the desired control chart to an ongoing process, monitoring, controlling and adjusting the process as it evolves.

In Step 1, the typical choice of control chart is related to the desire for quick detection of extreme changes in process parameters versus eventual detection of minor changes in process parameters (Dyer, Adams & Conerly, 2003; Lin & Adams, 1996). The Shewhart-type control charts are commonly used for the former, and the EWMA and CUSUM control charts are used for the latter. The choice of the IC ARL in Step 1 involves practical and economic considerations, depending largely on the costs associated with false alarms versus concealment of true process changes (Dyer, Adams & Conerly, 2003).

In Step 2, the subgroup size (n) is a function of the sampling frequency, the process output rate, and practical considerations and limitations regarding time and costs. Marsaglie, Maclaren & Bray (1964) provide a discussion of the selection of an appropriate subgroup size (n) and sampling frequency to design control charts. The choice of the number of subgroups (m) is most likely an economic consideration (Jones, 2002). If contaminated data exist in the reference sample, the parameter estimates obtained can be adversely affected if MLEs are used to obtain parameter estimates (L2E to a lesser degree). Small reference samples tend to

magnify the adverse effects of estimation. A widely accepted heuristic is that $m = 30$ subgroups from a process will provide reasonable estimates (Jones, 2002); Quesenberry (1993) suggests at least $m = 100$ subgroups of size $n = 5$ to estimate the parameters for the Shewhart-type control chart. Jones, Champ & Rigdon (2001) showed that an m much greater than 100, up to $m = 400$, is often required when designing an EWMA control chart.

In Step 3, the reference sample obtained in Step 2 is analyzed in order to estimate the unknown parameters and to determine the state of the process (IC versus OC). This is also the stage when distributional and randomness assumptions are verified, as well as when GR&R and capability studies are conducted. Concerning parameter estimation, if MLEs are used, the resulting values are the estimates used to construct an initial control chart with limits set according to the desired IC ARL in Step 1. In Stage 2, the control charts are used for prospective monitoring of the reference sample to determine departures from the estimated parameters. The control charts are primarily used to detect contaminated data or nonrandom process output, that is, data resulting from special cause variation.

Step 3 is often an iterative process, wherein contaminated data are identified (to the degree possible) and removed using a control chart based on the initial parameter estimates (MLEs). Any contaminated data identified are investigated and removed, new MLEs are obtained, a new control chart is constructed using the MLE values and more contaminated data are removed.

The process of parameter estimation and control chart removal of contaminated data continues until sufficient experience has been accumulated so that the IC parameters are effectively considered to be known through estimation. It should also be noted that if a large degree of contaminated data exist in the reference sample (as a percent of the sample size), or the magnitude of contaminated data is large (measured in terms of shifts in the process mean or variability), then the initial control limits may be inflated to a point where the contaminated data are hidden and unidentifiable. If this is the case, the Phase II parameter

estimates will be unreliable. If L2E estimates are used instead, it will be shown that the iterative process in Step 3 might be minimized by providing a more robust set of parameter estimates in the first iteration, which will lead to a more robust set of control limits, thus enabling more efficient detection and removal of contaminated data.

Methodology
The L2E Estimation Technique

The L2E estimation criterion for the two-parameter normal density technique requires the minimization of the L2E function with respect to the parameters μ and σ. (See Scott (2001) for the derivation of the general L2E criterion and specification of the two-parameter normal density.) Suppose a sample of size $n \geq 1$ is drawn from a normal distribution with mean, μ, and standard deviation, σ. Let the n sample data be represented by $x_1, x_2, \ldots, x_n$, and let the univariate normal density be denoted by $\phi(x|\mu, \sigma)$. The minimization of the normal L2E function (equation 1) with respect to μ and σ produces the L2E estimates, $\hat{\mu}, \hat{\sigma}$, that is, the estimation criterion is shown as:

$$L2E\left(\hat{\mu}, \hat{\sigma}\right) = \arg \min_{\mu, \sigma} \left[\frac{1}{2\sigma\sqrt{\pi}} - \frac{2}{n}\sum_{i=1}^{n} \phi\left(x_i|\mu, \sigma\right) \right].$$
(1)

Observe that the L2E minimizes a function of the sum of the densities; however, the MLE can be shown to maximize a function of the product of the densities. For values of x extremely distant from μ, the density value approaches zero. As a result, the L2E utilizes only the largest portion of the data that matches the model (good data), that is, x values located within a reasonable distance of $\mu \pm 3\sigma$. In effect, the L2E criterion ignores contaminated data, hence generally providing more robust parameter estimates. Because MLE must account for all the data, the fits often blur the distinction between good data and contaminated data (Scott, 2001). In cases wherein there are no contaminated data, the L2E and MLE estimates are nearly equal. It can be shown through consistency theory that, for a large sample of uncontaminated data, MLE is a very good estimator (Mood, Graybill & Boes, 1970); other estimators, such as the L2E may be just as good, but not better. In this study the L2E is shown to be just as good when the reference sample is uncontaminated and better in almost all simulated cases when contamination exists.

Results
Comparison with MLEs

Unfortunately there are few example data sets that cover the range of samples sizes and contamination types and levels described herein. Montgomery (1997) provides some of the most referenced data sets in SPC research, but unfortunately none of these have sufficient examples required to cover the 96 cases of sample sizes and contamination types and levels described in this article. Simulation results are therefore used to investigate the behavior of the L2E estimates across a broad range of sample sizes as well as types and levels of data contamination. In lieu of borrowing an example data set, the simulation results are used to reveal the behavior of the L2E estimates over a broad range of cases and an example application is provided to assist the user in applying the L2E technique.

Regarding the simulation results, Tables 1a and 1b reveal average L2E and MLE estimates for μ and σ^2 (σ^2 reported as σ) based on averaging 10,000 simulations of $n = 100$ normal pseudo-random variables representing differing levels and degrees of good versus contaminated data. (A complete description of the simulation design is provided in the Appendix.) The good data (IC process) are random variables representing a normal ($\mu = 0$, $\sigma = 1$) process, N(0, 1). The contaminated data are drawn from a normal process with parameters that vary from the IC process. Levels of contamination refer to the number of contaminated data values (cn) in a sample of size $n = 100$ and degrees of contamination refer to whether the contaminated data has experienced a mean shift alone, a shift in the standard deviation alone, or a shift in both the mean and standard deviation. Contamination levels in Tables 1a and 1b correspond to $n = 5$, 15, 25 and 45. Degrees of contamination

correspond to the following shifts (for cn = 5, 15, 25 and 45):

- Mean shifts (alone) of μ = 0.5, 1.0, 2.0, and 3.0 (16 cases)
- Standard deviation shifts (alone) of σ = 1.5, 2.0, 2.5, 3.0 (16 cases)
- Simultaneous mean and standard deviation shifts representing combinations of all mean and standard deviation shifts alone (64 cases).

Tables 1a and 1b display simulation results providing 96 comparisons for average L2E versus MLE estimates of μ and σ. For the IC process data (n = 100 random variables generated from a N(0, 1) process, the resulting simulation based estimates are $\hat{\mu}$(L2E) = 0.0006, $\hat{\mu}$(MLE) = 0.0007, $\hat{\sigma}$(L2E) = 0.9988, and $\hat{\sigma}$(MLE) = 1.0015. In Tables 1a and 1b the estimates of μ and μ are shown as $\hat{\mu}$(L2E), $\hat{\mu}$(MLE), $\hat{\sigma}$(L2E) and $\hat{\sigma}$(MLE). For all mean shifts and standard deviation shifts alone, the mathematical expectation and standard deviation (based on the levels and degrees of contamination) match the simulated MLE results.

In deriving the expected value, let X_n be the mixture of two normally distributed samples of size n, where X_{n-cn} is the uncontaminated distribution with $E(X_{n-cn}) = \mu_{n-cn}$, and X_{cn} is the contaminated distribution with $E(X_{cn}) = \mu_{cn}$ (recall, cn is the number of contaminated data values in the combined sample of size n). In this case, the $E(X_n)$ is the weighted average expectation of each distribution of data, where the weights are the sample sizes from each distribution relative to the total sample size. Thus,

$$E(X_n) = \frac{n-cn}{n} E(X_{n-cn}) + \frac{cn}{n} E(X_{cn}).$$

In the case where the uncontaminated data distribution has $E(X_{n-cn})$ = 0,

the $E(X_n) = \frac{cn}{n} E(X_{cn})$. For example, for X_{cn}~N(3, 1) where n = 100 and cn = 45, the $E(X_n) = \frac{45}{100}(3) = 1.35$. This value matches the simulated value given by $\hat{\mu}$ (MLE) in Table 1b.

All simulated values for $\hat{\mu}$(MLE) (for both mean and standard deviation shifts alone) match the mathematical expectations. This is expected given that $\hat{\mu}$(MLE) is location invariant to distributional changes due to shifts in either the mean or standard deviation. The same can be observed for the standard deviation estimates, $\hat{\sigma}$ (MLE), where

$$\sigma(X_n) = \sqrt{\frac{n-cn}{n} Var(X_{n-nc}) + \frac{cn}{n} Var(X_{nc})}$$

when $E(X_n) = 0$ and $Var(X_{n-nc}) = 1$. All simulated values for $\hat{\sigma}$ (MLE) (for standard deviation shifts alone) match $\sigma(X_n)$. This is expected because $\hat{\sigma}$ (MLE) is scale invariant to distributional changes due to shifts in the mean alone or the standard deviation alone. For cases where the mixed distribution has experienced both a mean shift and a standard deviation shift, $\hat{\sigma}$ (MLE) is not scale invariant; hence, the variance is not the weighted average of mixed variance components.

Simulation Result Comparison with MLEs

The simulation results reveal that in all cases

$$\{abs(\hat{\mu}(L2E) - \mu) \leq abs(\hat{\mu}(MLE) - \mu)\},$$

and in 95% of cases

$$\{abs(\hat{\sigma}(L2E) - \sigma) \leq abs(\hat{\sigma}(MLE) - \sigma)\}.$$

That is, the L2E estimates in almost all cases are as good (and often much better) as the MLE

estimates. This attests to the contention that the L2E estimators are as robust, or more robust, than MLE estimators.

Observe in Tables 1a and 1b that $\overset{\wedge}{\mu}$(L2E) is robust for most shifts in μ_{cn}, for all $\frac{cn}{n} \leq .45$, and more robust than $\overset{\wedge}{\mu}$(MLE) in all cases. The relative efficiency measures in Tables 2a and 2b indicate that the worst cases are those with large $\frac{cn}{n}$, for $\mu_{cn} \geq 2$. When $\mu_{n-cn} = 0$, the relative efficiency for either mean estimator is defined as RE_{μ} = 1- abs($\overset{\wedge}{\mu}$) where $\overset{\wedge}{\mu}$ is the estimate of μ_{n-cn} and is the mean of the IC process. Table 3 displays the percent frequency distribution of relative efficiency measures for all cases simulated. Notice that, for mean shifts alone, 57% of $\overset{\wedge}{\mu}$(L2E) have RE_{μ} > 0.80 versus 44% of $\overset{\wedge}{\mu}$(MLE).

For shifts in the mean and standard deviation (simultaneously), the frequency of RE_{μ} > 0.80 is 81% for $\overset{\wedge}{\mu}$(L2E) and only 43% for $\overset{\wedge}{\mu}$(MLE). It appears that $\overset{\wedge}{\mu}$(L2E) is most robust when both a mean and standard deviation shift has occurred.

The relative efficiency for a standard deviation estimate is defined as

$$RE_{\sigma} = 1 - \left[\frac{abs\left(\widehat{\sigma} - \sigma_{n-cn}\right)}{\sigma_{n-cn}} \right],$$

where $\overset{\wedge}{\sigma}$ is the estimate of σ_{n-cn}, the standard deviation of the IC process. Because σ_{n-cn}= 1 in all simulation cases, RE_{σ} =1-abs($\overset{\wedge}{\sigma}$-1). Again, observe in Tables 1a and 1b that $\overset{\wedge}{\sigma}$(L2E) is robust for most shifts in σ_{cn}, for all $\frac{cn}{n} \leq .45$, and particularly when μ_{n-cn} < 1. Notice also that $\overset{\wedge}{\sigma}$(L2E) is more robust than $\overset{\wedge}{\sigma}$(MLE) in 95%

of all cases. It appears that $\overset{\wedge}{\sigma}$(MLE) is less robust when all of μ_{n-cn}, σ_{n-cn}, and $\frac{cn}{n}$ are large. The relative efficiency measures in Tables 2a and 2b also indicates that these are the worst cases for $\overset{\wedge}{\sigma}$(L2E). Note in Table 3 that, for standard deviation shifts alone, 87% of $\overset{\wedge}{\sigma}$(L2E) have RE_{σ} > 0.80 versus 50% of $\overset{\wedge}{\sigma}$(MLE). For shifts in both the mean and standard deviation (simultaneously), the frequency of RE_{σ} > 0.80 is 69% for $\overset{\wedge}{\sigma}$(L2E) and only 31% for $\overset{\wedge}{\sigma}$(MLE). It appears that $\overset{\wedge}{\sigma}$(L2E) is more robust when only a shift in the standard deviation has occurred.

L2E Application Example

As noted, one advantage of using MLE is its constructive nature. In other words, it is simple to average a collection of data values or calculate the standard deviation. The L2E estimates are also constructive in nature, but require optimization techniques. Specifically, the L2E function given by equation 1 must be formulated and minimized subject to constraints. This can be readily accomplished in a spreadsheet environment with little or no knowledge of programming or minimization techniques. The authors suggest using Microsoft Excel and the spreadsheet add-in Solver. The data can be displayed in the spreadsheet, the L2E function can be formulated using the data and functions of the data as input, and the Solver function can be invoked to provide the L2E estimates via Solver's built-in optimization algorithm.

The data can represent individual observations or subgroup averages. If individual observations are used, then the resulting L2E estimates are those for process μ and σ. If subgroup averages are used, the resulting L2E estimates are those for μ and $\sigma/\sqrt{n}$ (standard error of the mean, SE). In the latter case, multiplying the estimate of SE by $\sqrt{n}$ yields the estimate for σ. For practitioners familiar with optimization, the L2E estimation problem can be

Table 1a: L2E and MLE Estimates of μ and σ

CSS	μ	0.0	0.0	0.0	0.0	0.5	0.5	0.5	0.5	0.5	1.0	1.0	1.0	1.0	1.0
	σ	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
5	$\hat{\mu}$ (L2E)	0.00	0.00	0.00	0.00	0.02	0.01	0.01	0.00	0.00	0.04	0.02	0.01	0.01	0.00
	$\hat{\mu}$ (MLE)	0.00	0.00	0.00	0.00	0.02	0.02	0.02	0.02	0.03	0.05	0.05	0.05	0.05	0.05
	$\hat{\sigma}$ (L2E)	1.00	1.01	1.02	1.02	0.99	1.01	1.01	1.02	1.02	1.01	1.01	1.02	1.02	1.02
	$\hat{\sigma}$ (MLE)	1.03	1.07	1.12	1.17	1.00	1.03	1.07	1.12	1.18	1.02	1.05	1.09	1.14	1.19
15	$\hat{\mu}$ (L2E)	0.00	0.00	0.00	0.00	0.07	0.04	0.02	0.01	0.01	0.13	0.07	0.04	0.03	0.02
	$\hat{\mu}$ (MLE)	0.00	0.00	0.00	0.00	0.08	0.08	0.08	0.08	0.08	0.15	0.15	0.15	0.15	0.15
	$\hat{\sigma}$ (L2E)	1.04	1.07	1.08	1.09	1.00	1.05	1.07	1.09	1.09	1.05	1.07	1.08	1.09	1.10
	$\hat{\sigma}$ (MLE)	1.08	1.20	1.33	1.47	1.01	1.10	1.21	1.34	1.48	1.06	1.14	1.25	1.37	1.51
25	$\hat{\mu}$ (L2E)	0.00	0.00	0.00	0.00	0.12	0.07	0.04	0.03	0.02	0.23	0.14	0.08	0.05	0.03
	$\hat{\mu}$ (MLE)	0.00	0.00	0.00	0.00	0.12	0.12	0.12	0.12	0.12	0.25	0.25	0.25	0.25	0.25
	$\hat{\sigma}$ (L2E)	1.08	1.13	1.17	1.19	1.01	1.09	1.14	1.17	1.19	1.08	1.13	1.16	1.18	1.19
	$\hat{\sigma}$ (MLE)	1.14	1.31	1.51	1.72	1.02	1.16	1.33	1.52	1.73	1.09	1.22	1.38	1.57	1.77
45	$\hat{\mu}$ (L2E)	0.00	0.00	0.00	0.00	0.23	0.15	0.10	0.07	0.05	0.44	0.30	0.21	0.15	0.11
	$\hat{\mu}$ (MLE)	0.00	0.00	0.00	0.00	0.22	0.22	0.22	0.22	0.22	0.45	0.45	0.45	0.45	0.45
	$\hat{\sigma}$ (L2E)	1.16	1.16	1.39	1.47	1.01	1.19	1.31	1.40	1.48	1.12	1.26	1.36	1.43	1.49
	$\hat{\sigma}$ (MLE)	1.24	1.53	1.82	2.13	1.02	1.26	1.54	1.84	2.14	1.11	1.34	1.60	1.89	2.19

viewed in the instructional form given by objective: minimize

$$L2E = \frac{1}{2\hat{\sigma}\sqrt{\pi}} - \frac{2}{n}\sum_{i=1}^{n}\phi\left(x_i|\hat{\mu},\hat{\sigma}\right)$$

by changing the values $\hat{\mu}$, $\hat{\sigma}$ subject to constraints: $\hat{\sigma} > 0$.

Figure 1 displays the author's spreadsheet in functional form, before using

Solver to minimize the L2E function. The data values 4, 5, 6, 7, 100 are input into column B, cells B11 to B15. The MLE sample mean and standard deviation, from the MLE variance, (24.4, 42.7) are calculated and displayed in column A, cells A5 and A6, respectively, using the built-in Excel function formulas shown in Figure 2. Figure 2 displays the same spreadsheet in formula/function view, allowing replication of cell formulas by the practitioner. Figure 1, column A, cells A11 to A15, contain the calculated normal probability density function

(Npdf) values resulting from the built-in Excel function shown in Figure 2. Because the Npdf function requires input values for the mean and standard deviation, the MLE estimates are initially used, and these values are temporarily input into the L2E estimate cells, column B, cells B5 and B6. Cells B5 and B6 will eventually be overwritten and contain the L2E estimates, as provided by Solver. Figure 1, cell A2, displays the L2E function value that is to be minimized, and Figure 2 displays the formula given by equation 1 as a function of both the

Table 1b: L2E and MLE Estimates of μ and σ

CSS	μ	2.0	2.0	2.0	2.0	2.0	3.0	3.0	3.0	3.0	3.0
	σ	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
5	$\hat{\mu}$ (L2E)	0.04	0.03	0.02	0.01	0.01	0.02	0.02	0.02	0.01	0.01
	$\hat{\mu}$ (MLE)	0.10	0.10	0.10	0.10	0.10	0.15	0.15	0.15	0.15	0.15
	$\hat{\sigma}$ (L2E)	1.04	1.02	1.02	1.02	1.02	1.04	1.03	1.03	1.02	1.02
	$\hat{\sigma}$ (MLE)	1.09	1.12	1.15	1.20	1.25	1.19	1.22	1.25	1.29	1.34
15	$\hat{\mu}$ (L2E)	0.16	0.10	0.07	0.04	0.03	0.09	0.08	0.07	0.05	0.03
	$\hat{\mu}$ (MLE)	0.30	0.30	0.30	0.30	0.30	0.45	0.45	0.45	0.45	0.45
	$\hat{\sigma}$ (L2E)	1.16	1.12	1.10	1.10	1.10	1.18	1.15	1.12	1.11	1.11
	$\hat{\sigma}$ (MLE)	1.23	1.30	1.39	1.51	1.63	1.47	1.53	1.61	1.71	1.82
25	$\hat{\mu}$ (L2E)	0.34	0.22	0.14	0.09	0.06	0.28	0.20	0.15	0.11	0.07
	$\hat{\mu}$ (MLE)	0.50	0.50	0.50	0.50	0.50	0.75	0.75	0.75	0.75	0.75
	$\hat{\sigma}$ (L2E)	1.30	1.24	1.21	1.21	1.21	1.46	1.33	1.26	1.24	1.23
	$\hat{\sigma}$ (MLE)	1.32	1.43	1.57	1.74	1.92	1.64	1.73	1.85	1.99	2.16
45	$\hat{\mu}$ (L2E)	0.86	0.61	0.42	0.30	0.21	1.24	0.95	0.66	0.44	0.31
	$\hat{\mu}$ (MLE)	0.90	0.90	0.90	0.90	0.90	1.35	1.35	1.35	1.35	1.35
	$\hat{\sigma}$ (L2E)	1.52	1.54	1.54	1.55	1.57	2.09	2.01	1.87	1.75	1.69
	$\hat{\sigma}$ (MLE)	1.41	1.59	1.82	2.08	2.35	1.80	1.95	2.14	2.36	2.61

Table 2a: Relative Efficiency of L2E and MLE Estimates of μ and σ

CSS	μ	0.0	0.0	0.0	0.0	0.5	0.5	0.5	0.5	0.5	1.0	1.0	1.0	1.0	1.0
	σ	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
5	$\hat{\mu}$ (L2E)	1.00	1.00	1.00	1.00	0.98	0.99	0.99	1.00	1.00	0.96	0.98	0.99	0.99	1.00
	$\hat{\mu}$ (MLE)	1.00	1.00	1.00	1.00	0.98	0.98	0.98	0.98	0.97	0.95	0.95	0.95	0.95	0.95
	$\hat{\sigma}$ (L2E)	1.00	0.99	0.98	0.98	0.99	0.99	0.99	0.98	0.98	0.99	0.99	0.98	0.98	0.98
	$\hat{\sigma}$ (MLE)	0.97	0.93	0.88	0.83	1.00	0.97	0.93	0.88	0.82	0.98	0.95	0.91	0.86	0.81
15	$\hat{\mu}$ (L2E)	1.00	1.00	1.00	1.00	0.93	0.96	0.98	0.99	0.99	0.87	0.93	0.96	0.97	0.98
	$\hat{\mu}$ (MLE)	1.00	1.00	1.00	1.00	0.92	0.92	0.92	0.92	0.92	0.85	0.85	0.85	0.85	0.85
	$\hat{\sigma}$ (L2E)	0.96	0.93	0.92	0.91	1.00	0.95	0.93	0.91	0.91	0.95	0.93	0.92	0.91	0.90
	$\hat{\sigma}$ (MLE)	0.92	0.80	0.67	0.53	0.99	0.90	0.79	0.66	0.52	0.94	0.86	0.75	0.63	0.49
25	$\hat{\mu}$ (L2E)	1.00	1.00	1.00	1.00	0.88	0.93	0.96	0.97	0.98	0.77	0.86	0.92	0.95	0.97
	$\hat{\mu}$ (MLE)	1.00	1.00	1.00	1.00	0.88	0.88	0.88	0.88	0.88	0.75	0.75	0.75	0.75	0.75
	$\hat{\sigma}$ (L2E)	0.92	0.87	0.83	0.81	0.99	0.91	0.86	0.83	0.81	0.92	0.87	0.84	0.82	0.81
	$\hat{\sigma}$ (MLE)	0.86	0.69	0.49	0.28	0.98	0.84	0.67	0.48	0.27	0.91	0.78	0.62	0.43	0.23
45	$\hat{\mu}$ (L2E)	1.00	1.00	1.00	1.00	0.77	0.85	0.90	0.93	0.95	0.56	0.70	0.79	0.85	0.89
	$\hat{\mu}$ (MLE)	1.00	1.00	1.00	1.00	0.78	0.78	0.78	0.78	0.78	0.55	0.55	0.55	0.55	0.55
	$\hat{\sigma}$ (L2E)	0.84	0.84	0.61	0.53	0.99	0.81	0.69	0.60	0.52	0.88	0.74	0.64	0.57	0.51
	$\hat{\sigma}$ (MLE)	0.76	0.76	0.18	-0.13	0.98	0.74	0.46	0.16	-0.14	0.89	0.66	0.40	0.11	-0.19

sample size (*n*) in cell B8 and the summed Npdf values. Prior to invoking the Solver function, the L2E function value (shown in Figure 1) is calculated using the MLE mean and standard deviation, but referencing the cells for the L2E mean and standard deviation. Figure 3 displays the Solver dialogue box referencing (1) the minimized L2E value cell (A2) as the target cell to minimize, (2) the cells to be changed to produce the minimum L2E value (B5 and B6), and (3) the constraint requiring the standard deviation to be non-negative. Selecting the Solve button invokes Solver to produce the L2E estimates of μ and σ whose values will overwrite the MLE values temporarily stored in cells B5 and B6. After solving for the L2E estimates, the actual value of the minimized L2E function is of no practical use and can be discarded. The L2E estimates of the mean and standard deviation (based on this example) are 5.5 and 1.5, respectively.

Table 2b: Relative Efficiency of L2E and MLE Estimates of μ and σ

CSS		μ	2.0	2.0	2.0	2.0	2.0	3.0	3.0	3.0	3.0	3.0
		σ	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
5		$\hat{\mu}$ (L2E)	0.96	0.97	0.98	0.99	0.99	0.98	0.98	0.98	0.99	0.99
		$\hat{\mu}$ (MLE)	0.90	0.90	0.90	0.90	0.90	0.85	0.85	0.85	0.85	0.85
		$\hat{\sigma}$ (L2E)	0.96	0.98	0.98	0.98	0.98	0.96	0.97	0.97	0.98	0.98
		$\hat{\sigma}$ (MLE)	0.91	0.88	0.85	0.80	0.75	0.81	0.78	0.75	0.71	0.66
15		$\hat{\mu}$ (L2E)	0.84	0.90	0.93	0.96	0.97	0.91	0.92	0.93	0.95	0.97
		$\hat{\mu}$ (MLE)	0.70	0.70	0.70	0.70	0.70	0.55	0.55	0.55	0.55	0.55
		$\hat{\sigma}$ (L2E)	0.84	0.88	0.90	0.90	0.90	0.82	0.85	0.88	0.89	0.89
		$\hat{\sigma}$ (MLE)	0.77	0.70	0.61	0.49	0.37	0.53	0.47	0.39	0.29	0.18
25		$\hat{\mu}$ (L2E)	0.66	0.78	0.86	0.91	0.94	0.72	0.80	0.85	0.89	0.93
		$\hat{\mu}$ (MLE)	0.50	0.50	0.50	0.50	0.50	0.25	0.25	0.25	0.25	0.25
		$\hat{\sigma}$ (L2E)	0.70	0.76	0.79	0.79	0.79	0.54	0.67	0.74	0.76	0.77
		$\hat{\sigma}$ (MLE)	0.68	0.57	0.43	0.26	0.08	0.36	0.27	0.15	0.01	-0.16
45		$\hat{\mu}$ (L2E)	0.14	0.39	0.58	0.70	0.79	-0.24	0.05	0.34	0.56	0.69
		$\hat{\mu}$ (MLE)	0.10	0.10	0.10	0.10	0.10	-0.35	-0.35	-0.35	-0.35	-0.35
		$\hat{\sigma}$ (L2E)	0.48	0.46	0.46	0.45	0.43	-0.09	-0.01	0.13	0.25	0.31
		$\hat{\sigma}$ (MLE)	0.59	0.41	0.18	-0.08	-0.35	0.20	0.05	-0.14	-0.36	-0.61

Table 3: Percent Frequency of L2E and MLE Estimates of μ and σ within a Range of Relative Efficiency

Range of Relative Efficiency		μ Shifts Alone		σ Shifts Alone		μ and σ Shifts			
		$\hat{\mu}$(L2E)	$\hat{\mu}$(MLE)	$\hat{\sigma}$(L2E)	$\hat{\sigma}$(MLE)	$\hat{\mu}$(L2E)	$\hat{\mu}$(MLE)	$\hat{\sigma}$(L2E)	$\hat{\sigma}$(MLE)
0.90	1.00	38%	19%	56%	19%	67%	19%	47%	17%
0.80	0.90	19%	25%	31%	31%	14%	25%	23%	14%
0.70	0.80	19%	13%	6%	13%	6%	13%	9%	20%
0.60	0.70	6%	6%	6%	19%	5%	6%	13%	19%
0.50	0.60	6%	13%	0%	13%	3%	13%	5%	19%
<0.50		12%	18%	0%	6%	5%	24%	3%	11%

Figure 1: Functional Form Excel Spreadsheet
(Prior to using Solver function)

	A	B	C
1	Min L2E Value		
2	-0.00772		
3	Estimator		
4	MLE	L2E	
5	24.4	24.4	Mean
6	42.3	42.7	St Dev
7			
8	Sample Size: n =	5	
9			
10	Npdf	Data	
11	0.00834	4	
12	0.00843	5	
13	0.00851	6	
14	0.00860	7	
15	0.00195	100	

Figure 2: Formula/Function View Excel Spreadsheet

	A	B	C
1	Min L2E Value		
2	=(((2*B6*((PI())^0.5))^-1))-2*SUM(A11:A15)/B8		
3	Estimator		
4	MLE	L2E	
5	=AVERAGE(B11:B15)	24.4	Mean
6	=STDEV(B11:B15)	42.7	St Dev
7			
8	Sample Size: n =	=COUNT(B11:B15)	
9			
10	Npdf	Data	
11	=NORMDIST(B11,B5,B6,FALSE)	4	
12	=NORMDIST(B12,B5,B6,FALSE)	5	
13	=NORMDIST(B13,B5,B6,FALSE)	6	
14	=NORMDIST(B14,B5,B6,FALSE)	7	
15	=NORMDIST(B15,B5,B6,FALSE)	100	

Figure 3: Solver Dialogue Box Referencing the
Minimized L2E Value Cell
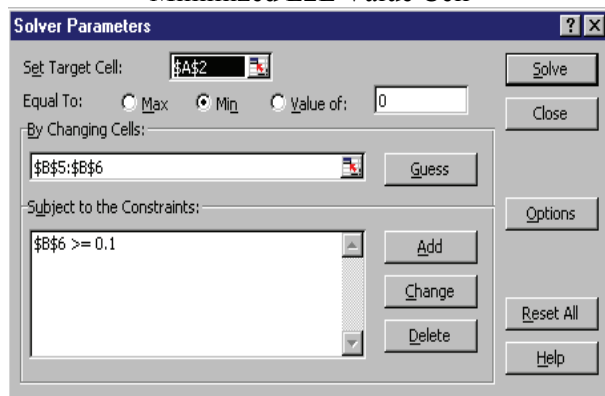
Conclusion

The importance of Phase I control charting was discussed, particularly the estimation of appropriate parameters to enable Phase II control charting. The general SPC method was shown to be a collection of steps that include both Phase I and Phase II control charting. For the Phase I environment, the minimized integrated square error estimator, L2E, was introduced as a robust parameter estimation technique and suggested as an alternative to MLEs.

Regarding managerial implications, the L2E estimation technique was described and shown to be easily constructed and applied in a spreadsheet environment. It was also shown to be a robust alternative to MLE estimation and just as simple to apply. The study also provided insights to the importance of clean data when constructing control charts based off of the Phase I processes and how the L2E estimator can facilitate robust parameter estimation required in SPC applications.

A simulation study revealed that the L2E estimates of μ and σ for a normal distribution are as good, and in most cases better, than MLE estimates when the reference sample is contaminated by shifts in the mean, the variance, or both the mean and variance. Tables based on the simulation results compare the absolute and relative performance of both the L2E and MLE estimators. Finally, an example was provided to enable an SPC practitioner, with little or no knowledge of programming or optimization, to readily apply the L2E technique.

Although this article discussed the application of L2E estimators in the SPC environment (assuming a univariate normal distribution), the technique can also be adapted to enable robust parameter estimation when discrete (Poisson) or multivariate processes are to be monitored and controlled. Additionally, the L2E is only one of several nonparametric density estimators that can be considered in the Phase I environment. Other estimators that might be of research interest include M-Estimators and estimators based on Hellinger's distance criterion.

References

Alt, F. B., & Smith, N. D. (1988). Multivariate process control. In *Handbook of Statistics 7*, P. R. Krishnaiah & C. R. Rao (Eds.), 333-351. New York: Elsevier.

Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, *71*, 353-360.

Chou, S., & Champ, C. W. (1995). A comparison of two phase I control charts. *Proceeding of the Quality and Productivity Section of the American Statistical Association*, 31-35.

Dyer, J. N., Adams, B. M., & Conerly, M. D. (2003). The reverse moving average control chart and comparisons of forecast based monitoring schemes. *Journal of Quality Technology*, 35(*2*), 135-155.

Jones, L. A. (2002). The statistical design of EWMA control charts with estimated parameters. *Journal of Quality Technology*, *34*, 277-288.

Jones, L. A., Champ, C. W., & Rigdon, S. E. (2001). The performance of exponentially weighted moving average control charts with estimated parameters. *Technometrics*, *43*, 156-167.

Kinderman, A. J., & Ramage, J. G. (1976). Computer generation of normal random variables. *Journal of the American Statistical Association*, *71*, 893-896.

Koning, A. J., & Does, R. J. M. M. (2000). CUSUM charts for preliminary analysis of individual observations. *Journal of Quality Technology*, *32*, 122-132.

Lin, W. S. W., & Adams B. M. (1996). Combined control charts for forecast-based monitoring schemes. *Journal of Quality Technology*, *28*, 289-301.

Marsaglia, G. (1964). Generating a variable from the tail of a normal distribution. *Technometrics*, *6*, 101-102.

Marsaglia, G., & T. A. Bray (1964). A convenient method for generating normal variables. *SIAM Review*, *6*, 260-264.

Marsaglia, G., MacLaren, M. D., & Bray, T. A. (1964). A fast procedure for generating normal random variables. *Communications of the ACM*, *7*, 4-10.

Montgomery, D. C. (1997). *Introduction to statistical quality control*, *3rd Ed*. New York, NY: John Wiley.

Mood, M. M., Graybill, A. G., & Boes, D. C. (1970). *Introduction to the theory of statistics*, *3rd Ed*. New York, NY: McGraw-Hill.

Newton, P. B., & Champ, C. W. (1997). Probability limits for Shewhart Phase I xbar-charts. *Proceeding of the Southeast Decision Sciences Institute Annual Conference*, Atlanta, Georgia, 234-236.

Quesenberry, C. P. (1993). The effect of sample size on estimated limits for $\overline{X}$ and X control charts. *Journal of Quality Technology*, *25*, 237-247.

Ramsey, P. H., & Ramsey, P. P. (2007). Optimal trimming and outlier elimination. *Journal of Modern Applied Statistical Methods*, *6*(*2*), 355-360.

Rudemo, M. (1982). Empirical choice of histogram and kernel density estimators. *Scandinavian Journal of Statistics*, *9*, 65-78.

Scott, W. S. (2001). Parametric statistical modeling by minimum integrated square error. *Technometrics*, *43*, 274-285.

Sullivan, J. H., & Woodall, W. H. (1996). A control chart for preliminary analysis of individual observations. *Journal of Quality Technology*, *28*, 265-278.

Sullivan, J. H., & Woodall, W. H. (1994). A comparison of multivariate control charts for individual observations. *Technical Report*, *Applied Statistics Program*, University of Alabama, 265-278.

Sullivan, J. H., Barrett, J. D., & Woodall, W. H. (1995). Interpreting the retrospective T2 control chart based on individual observations. *Indian Association for Productivity, Quality, and Reliability Transactions*, *20*, 183-195.

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

Appendix: Simulation Description

The simulation program was designed and compiled using Microsoft Visual Basic 6.0, executed in Microsoft Excel 2000 using normal random variates generated and imported from

Microsoft FORTRAN PowerStation for Windows, Version 4.0, FORTRAN 90. Each simulation was conducted according to steps provided below. A series of 100 N(0, 1) random variates was generated by FORTRAN MSIMSL subroutine RNNOA.

Routine RNNOA generates pseudorandom numbers from a standard normal (Gaussian) distribution using an acceptance/rejection technique due to (Kinderman & Ramage, 1976). In this method, the normal density is represented as a mixture of densities over which a variety of acceptance/rejection methods due to (Marsaglia, 1964), (Marsaglia & Bray, 1964), and (Marsaglia, Maclaren & Bray, 1964) are applied. The final parameter estimates for each of the 96 cases were based on 10,000 simulations, which provided a maximum margin of error of 0.02 in estimation of the MLE means, with 95% confidence. These variates were the simulated observations, X_i's, for each of the cases investigated.

Step 1:

a. For estimation of the mean (the 16 cases of a mean shift only), a shift in the mean was induced in the simulated observations affecting cn of the $n = 100$ variates. The values of $cn = 5, 15, 25$ and 45 (levels of contamination), and the magnitudes of shifts were $\mu_{cn} = 1.50$, 2.00, 2.50 and 3.00 (degrees of contamination). Every combination of cn and σ_{cn} produced the 16 cases.

b. For estimation of the standard deviation (the 16 cases of a standard deviation shift only), a shift in the standard deviation was induced in the simulated observations affecting cn of the $n = 100$ variates. Again, the values of $cn = 5, 15, 25$ and 45, and the magnitudes of shifts were $\sigma_{cn} = 1.50, 1.00, 2.00$ and 3.00. Every combination of cn and μ_{cn} produced the 16 cases.

c. For estimation of the mean and standard deviation (the 64 cases of both a mean and standard deviation shift), a shift in each parameter was induced in the simulated observations affecting cn of the $n = 100$ variates. Again, the values of $cn = 5, 15, 25$ and 45, and the magnitudes of shifts were $\mu_{cn} = 1.50$, 2.00, 2.50, 3.00 and $\sigma_{cn} = 1.50, 1.00$, 2.00 and 3.00. Every combination of cn, μ_{cn}, and σ_{cn} produced the 64 cases.

Step 2:

The individual L2E and MLE estimates of μ and σ (10,000 for each estimate, per case) were calculated using the procedures described in the article.

Step 3:

The average L2E and MLE estimates of μ and σ for each case was obtained by averaging over the 10,000 individual estimates for each estimator.